Université de Montréal

Optimal Portfolio Selection with Transaction Costs

Par
N'Golo Koné

Département de sciences économiques
Faculté des arts et sciences

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de Philosophiæ Doctor (Ph.D.)
en sciences économiques

Mai, 2020

Université de Montréal
Faculté des Études Supérieures et Postdoctorales

Cette thèse intitulée :

Optimal Portfolio Selection with Transaction Costs

présentée par :
N'Golo Koné

a été évaluée par un jury composé des personnes suivantes :

Bénoît Perron,	president du jury
Marine Carrasco,	directrice de recherche
René Garcia,	membre du jury
Tony Wirjanto,	examinateur externe

Thèse acceptée le 29 Juillet 2020

à mon Épouse Assiata, ma fille Katchiénin, mon feu père Zonzérigué et ma mère Yogo

Remerciements

J'ai une profonde gratitude envers ma directrice de recherche Marine Carrasco, pour sa présence continue, sa patience, ses multiples conseils en tout genre, son implication et sa contribution, pour m'avoir encouragé et d'avoir cru en moi tout au long de cette thèse. Merci infiniment pour tout.

Je remercie infiniment Georges Dionne, pour son soutien, son aide et pour sa disponibilité surtout durant les deux dernières années de ma thèse. Je remercie également Benoit Perron et René Garcia pour leur disponibilité, leur aide et leur conseil tout au long de cette thèse sans oublier l'ensemble des professeurs, chercheurs ainsi que le personnel administratif du département de sciences économiques de l'Université de Montréal et du CIREQ. J'aimerais aussi exprimer toute ma gratitude à Tony Wirjanto pour sa disponibilité comme examinateur externe.

J'exprime également ma reconnaissance à mes collègues de promotion qui ont été d'une aide précieuse tout au long de cette thèse: Marlène Koffi, Fatim Diabagaté, Magnim Farouh, Guy Leonel Siwe, Lucienne Talba, Idriss Tsafack et Samuel Gingras. Je remercie particulièrement mes amis Louphou Coulibaly et Souleymane Zerbo pour leur soutien et leur disponibilité continue. Je ne saurai oublier tous les étudiants de doctorat du département de sciences économiques de l'Université de Montréal dont les discussions enrichissantes m'ont permis d'approfondir mes connaissances et d'améliorer le contenu de cette thèse.

Mes pensées et ma reconnaissance vont à ma mère Yogo, mes frères et soeurs pour le soutien moral. À mon épouse Wovana Assiata, je témoigne ma profonde gratitude pour son soutien continu surtout dans les moments difficiles, pour sa grande compréhension, pour sa patience et d'avoir toujours été pour moi une source d'inspiration tout au long de ce processus. Sans oublier ma merveilleuse fille Katchiénin Grâce dont la joie de vivre constitue pour moi une source d'inspiration continue.

Résumé

Le choix de portefeuille optimal d'actifs a été depuis longtemps et continue d'être un sujet d'intérêt majeur dans le domaine de la finance. L'objectif principal étant de trouver la meilleure façon d'allouer les ressources financières dans un ensemble d'actifs disponibles sur le marché financier afin de réduire les risques de fluctuation du portefeuille et d'atteindre des rendements élevés. Néanmoins, la littérature de choix de portefeuille a connu une avancée considérable à partir du 20ieme siècle avec l'apparition de nombreuses stratégies motivées essentiellement par le travail pionnier de Markowitz (1952) qui offre une base solide à l'analyse de portefeuille sur le marché financier. Cette thèse, divisée en trois chapitres, contribue à cette vaste littérature en proposant divers outils économétriques pour améliorer le processus de sélection de portefeuilles sur le marché financier afin d'aider les intervenants de ce marché.

Le premier chapitre, qui est un papier joint avec Marine Carrasco, aborde un problème de sélection de portefeuille avec coûts de transaction sur le marché financier. Plus précisément, nous développons une procédure de test simple basée sur une estimation de type GMM pour évaluer l'effet des coûts de transaction dans l'économie, quelle que soit la forme présumée des coûts de transaction dans le modèle. En fait, la plupart des études dans la littérature sur l'effet des coûts de transaction dépendent largement de la forme supposée pour ces frictions dans le modèle comme cela a été montré à travers de nombreuses études (Dumas and Luciano (1991), Lynch and Balduzzi (1999), Lynch and Balduzzi (2000), Liu and Loewenstein (2002), Liu (2004), Lesmond et al. (2004), Buss et al. (2011), Gârleanu and Pedersen (2013), Heaton and Lucas (1996)). Ainsi, pour résoudre ce problème, nous développons une procédure statistique, dont le résultat est indépendant de la forme des coûts de transaction, pour tester la significativité de ces coûts dans le processus d'investissement sur le marché financier. Cette procédure de test repose sur l'hypothèse que le modèle estimé par la méthode des moments généraliés (GMM) est correctement spécifié. Un test commun utilisé pour évaluer cette hypothèse est le J-test proposé par Hansen (1982). Cependant, lorsque le paramètre d'intérêt se trouve au bord de l'espace paramétrique, le J-test standard souffre d'un rejet excessif. De ce fait, nous proposons une procédure en deux étapes pour tester la sur-identification lorsque le paramètre d'intérêt est au bord de l'espace paramètrique. Empiriquement, nous appliquons nos procédures de test à la classe des anomalies utilisées par Novy-Marx and Velikov (2016). Nous montrons que les coûts de transaction ont un effet significatif

sur le comportement des investisseurs pour la plupart de ces anomalies. Par conséquent, les investisseurs améliorent considérablement les performances hors échantillon en tenant compte des coûts de transaction dans le processus d'investissement.

Le deuxième chapitre aborde un problème dynamique de sélection de portefeuille de grande taille. Avec une fonction d'utilité exponentielle, la solution optimale se révèle être une fonction de l'inverse de la matrice de covariance des rendements des actifs. Cependant, lorsque le nombre d'actifs augmente, cet inverse devient peu fiable, générant ainsi une solution qui s'éloigne du portefeuille optimal avec de mauvaises performances. Nous proposons deux solutions à ce problème. Premièrement, nous pénalisons la norme des poids du portefeuille optimal dans le problème dynamique et montrons que la stratégie sélectionnée est asymptotiquement efficace. Cependant, cette méthode contrôle seulement en partie l'erreur d'estimation dans la solution optimale car elle ignore l'erreur d'estimation du rendement moyen des actifs, qui peut également être importante lorsque le nombre d'actifs sur le marché financier augmente considérablement. Nous proposons une méthode alternative qui consiste à pénaliser la norme de la différence de pondérations successives du portefeuille dans le problème dynamique pour garantir que la composition optimale du portefeuille ne fluctue pas énormément entre les périodes. Nous montrons que, sous des conditions de régularité appropriées, nous maîtrisons mieux l'erreur d'estimation dans le portefeuille optimal avec cette nouvelle procédure. Cette deuxième méthode aide les investisseurs à éviter des coûts de transaction élevés sur le marché financier en sélectionnant des stratégies stables dans le temps. Des simulations ainsi qu'une analyse empirique confirment que nos procédures améliorent considérablement la performance du portefeuille dynamique.

Dans le troisième chapitre, nous utilisons différentes techniques de régularisation (ou stabilisation) empruntées à la littérature sur les problèmes inverses pour estimer le portefeuille diversifié tel que définie par Choueifaty (2011). En effet, le portefeuille diversifié dépend du vecteur de volatilité des actifs et de l'inverse de la matrice de covariance du rendement des actifs. En pratique, ces deux quantités doivent être remplacées par leurs contrepartie empirique. Cela génère une erreur d'estimation amplifiée par le fait que la matrice de covariance empirique est proche d'une matrice singulière pour un portefeuille de grande taille, dégradant ainsi les performances du portefeuille sélectionné. Pour résoudre ce problème, nous étudions trois techniques de régularisation, qui sont les plus utilisées : le rigde qui consiste à ajouter une matrice diagonale à la matrice de covariance, la coupure spectrale qui consiste à exclure les vecteurs propres associés aux plus petites valeurs propres, et Landweber Fridman qui est une méthode itérative, pour stabiliser l'inverse de matrice de covariance dans le processus d'estimation du portefeuille diversifié. Ces méthodes de régularisation impliquent un paramètre de régularisation qui doit être choisi. Nous proposons donc une méthode basée sur les données pour sélectionner le paramètre de stabilisation de manière optimale. Les solutions obtenues sont comparées à plusieurs stratégies telles que le portefeuille le plus diversifié, le portefeuille cible, le portefeuille de variance minimale et la stratégie naïve 1 / N à l'aide du ratio de Sharpe

dans l'échantillon et hors échantillon.

Mots-clés: Sélection de portefeuille, test d'évaluation de l'effet des coûts de transaction, test de sur-identification, utilité récursive, choix de portefeuille dynamique, marché de grand taille, efficacité asymptotique, régularisation, diversification maximale.

Abstract

The optimal portfolio selection problem has been and continues to be a subject of interest in finance. The main objective is to find the best way to allocate the financial resources in a set of assets available on the financial market in order to reduce the portfolio fluctuation risks and achieve high returns. Nonetheless, there has been a strong advance in the literature of the optimal allocation of financial resources since the 20th century with the proposal of several strategies for portfolio selection essentially motivated by the pioneering work of Markowitz (1952) which provides a solid basis for portfolio analysis on the financial market. This thesis, divided into three chapters, contributes to this vast literature by proposing various economic tools to improve the process of selecting portfolios on the financial market in order to help stakeholders in this market.

The first chapter, a joint paper with Marine Carrasco, addresses a portfolio selection problem with trading costs on stock market. More precisely, we develop a simple GMM-based test procedure to test the significance of trading costs effect in the economy regardless of the form of the transaction cost. In fact, most of the studies in the literature about trading costs effect depend largely on the form of the frictions assumed in the model (Dumas and Luciano (1991), Lynch and Balduzzi (1999), Lynch and Balduzzi (2000), Liu and Loewenstein (2002), Liu (2004), Lesmond et al. (2004), Buss et al. (2011), Gârleanu and Pedersen (2013), Heaton and Lucas (1996)). To overcome this problem, we develop a simple test procedure which allows us to test the significance of trading costs effect on a given asset in the economy without any assumption about the form of these frictions. Our test procedure relies on the assumption that the model estimated by GMM is correctly specified. A common test used to evaluate this assumption is the standard J-test proposed by Hansen (1982). However, when the true parameter is close to the boundary of the parameter space, the standard J-test based on the χ^2 critical value suffers from overrejection. To overcome this problem, we propose a two-step procedure to test overidentifying restrictions when the the parameter of interest approaches the boundary of the parameter space. In an empirical analysis, we apply our test procedures to the class of anomalies used in Novy-Marx and Velikov (2016). We show that transaction costs have a significant effect on investors' behavior for most anomalies. In that case, investors significantly improve out-of-sample performance by accounting for trading costs.

The second chapter 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. 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. 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. 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.

In the third chapter, we use various regularization (or stabilization) techniques borrowed from the literature on inverse problems to estimate the maximum diversification as defined by Choueifaty (2011). In fact, the maximum diversification portfolio depends on the vector of asset volatilities and the inverse of the covariance matrix of assets distribution. In practice, these two quantities need to be replaced by their sample counterparts. This results in estimation error which is amplified by the fact that the sample covariance matrix may be close to a singular matrix in a large financial market, yielding a selected portfolio far from the optimal one with very poor performance. To address this problem, we investigate three regularization techniques, such as the ridge, the spectral cut-off, and the Landweber-Fridman, to stabilize the inverse of the covariance matrix in the investment process. These regularization schemes involve a tuning parameter that needs to be chosen. So, we propose a data-driven method for selecting the tuning parameter in an optimal way. The resulting regularized rules are compared to several strategies such as the most diversified portfolio, the target portfolio, the global minimum variance portfolio, and the naive 1/N strategy in terms of in-sample and out-of-sample Sharpe ratio.

Keywords: Portfolio selection, test for trading costs effect, testing overidentifying restrictions, recursive utility, Dynamic portfolio selection, Large Market, Asymptotic efficiency, Regularization, Maximum diversification.

Contents

List of Tables

List of Figures

Chapter 1

Test for Trading Costs Effect in a Portfolio Selection Problem with Recursive Utility*

1.1 Introduction

The problem of optimal allocation of economic resources is far from being a recent issue. It is a problem which already existed in the world before the first century[1]. Nonetheless, there has been a strong advances in the literature of the optimal allocation of financial resources since the 20th century with the proposal of several strategies for portfolio selection, especially with the seminal work of Markowitz (1952) which offers an essential basis to portfolio selection in a single period. However, his quadratic form utility function hypothesis has been strongly criticized and many alternative utility functions such as power utility and exponential utility have emerged in the literature of portfolio optimization. Moreover, Epstein and Zin (1989, 1991) develop a more flexible version of the basic power utility model. This new version of utility retains the desirable scale-independence of the power utility[2] but breaks the link between the elasticity of intertemporal substitution and the coefficient of relative risk aversion. Campani et al. (2015) use a closed-form approximation solution to a portfolio selection problem to show the importance of disentangling the intertemporal substitution from the risk aversion. Regarding the large advantages of this class of preferences and their ability to explain financial variables, we use recursive utility to characterize investors' preferences in our economy. Hence, our work is related to

*This chapter is co-authored with Marine Carrasco. We thank René Garcia and Benoit Perron for their helpful comments. Carrasco thanks SSHRC for partial financial support.

[1]For instance, in circa 400 A.D. Rabbi Issac Bar Aha recommended that one should always divide his wealth equally into three parts: land, merchandise and cash at hand.

[2]such as the relative risk aversion coefficient and the elasticity of intertemporal substitution are constant

the previous literature of portfolio optimization with recursive preferences (see Campbell and Viceira (2002), Campbell et al. (2004),Campani et al. (2015)). More importantly, all those studies are carried out in a frictionless framework. Nonetheless, financial frictions in the form of liquidity costs, taxes, and transaction costs may affect investors' behavior on the financial market. For instance, an investor will have an incentive to invest in a more liquid asset compared to a less liquid asset. Indeed, according to Acharya and Pedersen (2005) the wealth problem which arises in the financial market due to the low market return at a given time can be amplified if selling investors hold illiquidity assets at this time. In fact, the asset illiquidity[3] could be seen as the potential loss because one cannot sell it at the price previously thought at a short notice. Moreover, investors will tend to have high preference for assets which require less costs to be invested in. Therefore, one needs to examine the rule played by those frictions in a portfolio selection problem with recursive preferences. We address this issue in this paper treating trading costs as the only friction in the financial market since assets illiquidity costs could also be seen as a certain transaction cost (Acharya and Pedersen (2005)). Our paper is then related to the vast literature about transaction costs and portfolio selection problems (see Dumas and Luciano (1991), Lynch and Balduzzi (1999), Lynch and Balduzzi (2000), Liu and Loewenstein (2002), Liu (2004), Lesmond et al. (2004), Buss et al. (2011), Gârleanu and Pedersen (2013), Novy-Marx and Velikov (2016) among others). However, most of the studies in the literature about trading costs effect depend largely on the form of the frictions assumed in the model. Indeed, with proportional or fixed costs, the optimal investment policy is shown to be in the form of a no-trade region so that trade occurs only when the proportion of wealth invested in the risky asset is outside this region (Dumas and Luciano (1991), Lynch and Balduzzi (1999), Lynch and Balduzzi (2000), Liu and Loewenstein (2002), Liu (2004), Buss et al. (2011)). Nevertheless, the optimal investment policy is no longer in the form of a no-trade region with quadratic trading costs since the investor trades at each period in small quantities (Heaton and Lucas (1996), Gârleanu and Pedersen (2013)). Moreover, Lynch and Balduzzi (1999) compute the utility cost due to the presence of these frictions and obtain an utility cost close to 4% with proportional costs and about 15% when added fixed costs to the proportional one.

In this paper, to overcome this problem, we develop a simple test procedure which allows us to test the significance of trading costs effect on a given asset in the economy without any assumption about the form of these frictions. The most interesting property of this test procedure is that our results do not depend on the form of the trading costs in our model. To our knowledge, this paper seems to be the first one to propose a statistical test for trading costs effect in the context of portfolio selection. Our test boils down to test the nullity of a parameter which is at the boundary of the parameter space under the null. Its asymptotic distribution is non standard and is derived using results by Andrews (1999). In the empirical application, we apply our test procedure to the class of

[3]See Amihud (2002) for a more general definition of the asset illiquidity.

anomalies used in Novy-Marx and Velikov (2016). We obtain that transaction costs have significant effect for most of anomalies considered in particular those whose trading costs exceed 1% of the gross return. Not surprisingly, trading costs do not have a significant effect when the risky asset is assumed to be the market portfolio.

Our test procedure relies on the assumption that the model is correctly specified. We wish to test this assumption using Hansen's J-test for overidentifying restrictions. However, when the true parameter is close to the boundary of the parameter space, the standard J-test based on the χ^2 critical value suffers from overrejection. To overcome this problem, we propose a two-step procedure to test overidentifying restrictions when the the parameter of interest approaches the boundary of the parameter space. This paper is related to the work of Ketz (2017) who proposes a J-test based on adjusted critical values and a modified J-test. We find by simulations that our two-step procedure has good small sample properties.

We measure the economic gain using a proportional trading costs in our model by comparing the out-of-sample performance to the model which ignores trading costs in the portfolio selection process. For this purpose we use several statistics such as the certainty equivalent (CE), the Sharpe ratio (SR) and the portfolio mean. We find that our model significantly outperforms the null model (in terms of the CE, the SR and the portfolio mean) for strategies whose trading costs have been shown to have significant effect according to our test procedure.

The rest of the paper is organized as follows. The model economy and the first order conditions from optimization problem are presented in Section 1.2. In Section 1.3, we develop a GMM-based test procedure to test whether trading costs have a significant effect. A two-step procedure for testing overidentifying restrictions in the GMM estimation is proposed in Section 1.4. Section 1.5 presents the empirical analysis where the test developed in Section 1.3 is applied to the twenty-three anomalies used in Novy-Marx and Velikov (2016). In Section 1.6, we evaluate the out-of-sample performance of our model based on several statistics such as the CE, the SR and the portfolio mean. Our conclusion and remarks are presented in Section 1.7.

1.2 The model and the first order conditions for the optimization problem

In this section we will start by the model economy before talking about the optimization problem.

1.2.1 The model

We consider a simple economy with two assets in which an investor can trade:

1. One risk-free asset (a bond) with a constant rate R^f. In general R^f will be calibrated to be the mean of the one-month Treasury-Bill rate observed in a monthly data.

2. One risky asset with a gross return R_{t+1} assumed to be predictable using a vector of instrumental variables D_t available at time t.

Novy-Marx and Velikov (2016) argue that the cost of trading in the low-turnover strategies such as the value-weighted strategies is quite low, and generally less than 10 basis points (bp) per month so that the utility cost associated with ignoring transaction costs in this case is sometimes negligible (see Lynch and Balduzzi (1999)). However, trading costs can be significantly higher for strategies that trade disproportionately in high transaction costs stocks such as anomalies based on idiosyncratic volatility or distress. In fact, trading costs on those strategies are more than 20 bp in average and exceed sometimes 1% per month (see Novy-Marx and Velikov (2016)). Since the incentive to find strategies based on anomalies are high, the risky asset considered here will be a portfolio based on one of the large array of the well-known anomalies used in Novy-Marx and Velikov (2016) instead of using only the market portfolio in which trading costs are very small.

An anomaly is defined as a strategy that generates significant positive alpha relative to a given asset pricing model. Note that the alpha is a measure of the active return on an investment, the performance of that investment compared to a suitable market index. It can be shown that in an efficient market, the expected value of the alpha coefficient is zero in the capital asset pricing model (CAPM) and a positive value of this parameter implies that the investment has a return in excess of the reward for the assumed risk.

Anomalies considered here are the twenty-three strategies used in Novy-Marx and Velikov (2016). Table 1.11 in Appendix B provides the list of anomalies considered and the average monthly trading costs on those strategies as presented by Novy-Marx and Velikov (2016).

We consider a finite-life horizon investor with recursive preferences as introduced in Epstein and Zin (1989, 1991).

The investor's utility function is defined recursively by the following equation:

$$U_t = \left[(1-\beta) C_t^{\rho} + \beta \left(E_t U_{t+1}^{1-\gamma} \right)^{\frac{\rho}{1-\gamma}} \right]^{\frac{1}{\rho}} \tag{1.1}$$

where $\beta \in (0, 1)$ is the rate of time preferences, γ is the coefficient of relative risk aversion which controls for investor's attitude over the states of the economy. $\Psi = \frac{1}{1-\rho}$ controls for intertemporal consumption allocation and will be considered as a measure of the elasticity of intertemporal substitution (EIS). U_t is the utility level at time t which is a function of the current consumption C_t and the future expected utility given time t information.

Recursive utilities help us to distinguish the relative risk aversion from the elasticity of intertemporal substitution. This property of separability of these two parameters is very useful when one is interested in a portfolio selection problem (see Campani et al. (2015)).

Moreover, in a simple numerical analysis of a 10 years horizon investor, we obtain as in Campani et al. (2015) that:

- Investors tend to take more risk for greater values of the EIS.
- The optimal investment decision is more affected by the EIS than the relative risk aversion.

These numerical results point out the importance of the property of separability of the relative risk aversion from the EIS in a portfolio optimization problem and justify the use of recursive utilities in this framework.

Because investors in general face some frictions such as liquidity costs, taxes, transaction costs, which can affect their behavior on financial market, it is important to incorporate these frictions when one is interested in a portfolio selection problem. For instance, Dumas and Luciano (1991), Lynch and Balduzzi (1999), Lynch and Balduzzi (2000), Liu and Loewenstein (2002), Liu (2004) show that realistic proportional or fixed costs cause optimal portfolio rebalancing frequency to decline considerably. Lesmond et al. (2004) also argue that the large gross spreads observed on momentum trades creates an "illusion of profit opportunity when in fact, none exists" because of the presence of trading costs. The same argument has been pointed out by Novy-Marx and Velikov (2016) who show that with trading costs in financial market, a strategy can have a significant positive alpha relative to the explanatory assets without significantly improving the investment opportunity set. Therefore, it is important not to ignore trading costs when one is particularly interested on investors behavior on financial markets. Hence, we assume that investors face transaction costs when trading on the risky asset and the transaction costs are assumed to be the only source of frictions in the financial market.Trading costs could be seen as all costs incurred by investors in the process of buying or selling an asset on the stock market. Hence, trading costs include brokerage fees, cost of analysis, information cost and any expenses incurred in the process of deciding upon and placing an order. Delay in execution which cause prices at which one trades to be different from those at which one planned to trade maybe included as well.

Let denote by y_t the proportion of the risky-asset that the investor holds at time t in the share of portfolio value. y_t is constrained to be between 0 and 1 to avoid short position in the financial market. The short position is a directional trading or investment strategy where investors sell shares of borrowed stocks in the open market. This is a realistic assumption since individual investors typically face high costs in taking short position and institutional investors are often precluded by their clients from taking short positions (Lynch and Balduzzi, 1999).

A portfolio will be defined as a list of weights y_t, $1 - y_t$ that represents the amount of capital to be invested in the risky asset and the bond respectively.

If we denote by $\bar{R}_{t+1}$ the return net of transaction costs on the risky asset in the optimal portfolio, then the total return on the optimal portfolio is given by:

$$R_{p,t+1} = y_t \left(\bar{R}_{t+1} - R^f \right) + R^f \tag{1.2}$$

From this equation it follows that in the frictionless economy, the gross return on the portfolio becomes $R_{p,t+1} = y_t \left(R_{t+1} - R^f \right) + R^f$.

We also assume that at each period of time the investor consumes a fraction of his current income. Thus, if A_t is his income at time t and C_t the consumption level then we define $k_t = \frac{C_t}{A_t}$ so that k_t varies (is random) as in Lynch and Balduzzi (2000). This assumption is more realistic than the one in Campbell and Viceira (2002) who assume a constant consumption-wealth ratio over time $\frac{C_t}{A_t} = b$.

We also assume in our model that the investor does not receive labor income, so he finances consumption entirely from financial wealth. Indeed, an external source of income to the financial market could affect investors' behavior toward risk and biased transaction costs effect on a portfolio selection problem as well as the result of our test procedure. Hence, assuming only the financial income in the model is a convenient assumption when one is interested in trading costs effect.

Therefore, the law of motion of his total wealth is:

$$A_{t+1} = (A_t - C_t) R_{p,t+1} = A_t (1 - k_t) R_{p,t+1} \tag{1.3}$$

where $R_{p,t+1}$ is the gross return on the optimal portfolio defined by equation (1.2).

1.2.2 First-order conditions for consumption-investment optimization problem

The agent maximizes his utility defined in (1.1) subject to the constraint (1.3). The Bellman equation associated with this optimization problem is given as follows:

$$J_t(A_t, I_t) = \max_{C_t \geq 0, y_t \in [0,1]} \left\{ (1-\beta) C_t^{\rho} + \beta \left[E_t J_{t+1}(A_{t+1}, I_{t+1})^{\lambda} \right]^{\frac{\rho}{\lambda}} \right\}^{\frac{1}{\rho}} \tag{1.4}$$

(see Epstein and Zin (1989, 1991)) where $I_t = (R_t, D_t)$ is a vector of state variables, D_t is the dividend yield, $J_t(A_t, I_t)$ is the value function of the optimization problem, and $\lambda = 1 - \gamma$. Because we do not model directly the transaction cost in the economy, investors do not have prior information about those frictions when selecting the optimal strategy.

Garlappi and Skoulakis (2010) show that, under the homothetic recursive preferences, the value function that solves (1.4) is given by:

$$J_t(A_t, I_t) = (1-\beta)^{\frac{1}{\rho}} V_t(I_t) A_t \tag{1.5}$$

where

$$V_t(I_t) = \left\{ 1 + \left[\beta \left(\min_{y_t \in [0,1]} E_t \left[R_{p,t+1}(y_t)^{\lambda} V_t(I_{t+1})^{\lambda} \right] \right)^{\frac{\rho}{\lambda}} \right]^{\frac{1}{1-\rho}} \right\}^{\frac{1-\rho}{\rho}} \tag{1.6}$$

with

$$V_T(I_T) = 1 \tag{1.7}$$

and the optimal consumption-to-wealth ratio is given by $k_t = V_t(I_t)^{\frac{-\rho}{\lambda}}$ with $\lambda = 1-\gamma$. Such a decomposition of the value function of the consumption-investment problem proves that the optimal portfolio and the optimal consumption problems can be solved separately and makes the numerical resolution easier to implement. The optimal consumption problem can then be analyzed independently from the portfolio optimization problem.

The first-order condition for the optimal consumption optimization problem as analyzed by Epstein and Zin (1989, 1991) is given by:

$$E_t \left[\beta^{\frac{\lambda}{\rho}} \left(\frac{C_{t+1}}{C_t} \right)^{\frac{\lambda}{\rho}(\rho-1)} R_{p,t+1}^{\frac{\lambda}{\rho}} \right] = 1 \tag{1.8}$$

for $t = 1, ..., T-1$ where $R_{p,t+1}$ is the gross return on the selected portfolio and given by equation (1.2). Equation (1.8) gives us a description of the investor's behavior in terms of intertemporal consumption allocation as a function of his preference parameters such as the relative risk aversion coefficient $1-\lambda$, the elasticity of intertemporal substitution $\frac{1}{1-\rho}$ and the discount factor β. It gives us information about how investors are going to smooth consumption over their life-cycle. According to this relation, the cost of reducing consumption today corresponds to the benefice of investing this amount in the available set of assets in the financial market in order to consume more tomorrow or to smooth consumption.

For the portfolio optimization, trading costs on the risky asset constrain investors in their investment decisions since they are obliged to make a trade-off between future gain and the cost due to the presence of those frictions in the economy. In fact, the expected gain for reallocating the portfolio may be smaller than the marginal cost due to the presence of trading costs so that conventional asset pricing relationships are transformed into inequality conditions (see Luttmer (1999), Brunnermeier et al. (2012)). Hence, the first-order condition when solving the optimal portfolio selection problem is given by the following relation:

$$E_t\left[\frac{\lambda}{\rho}\left(\frac{C_{t+1}}{C_t}\right)^{\frac{\lambda}{\rho}(\rho-1)} R_{p,t+1}^{\frac{\lambda}{\rho}-1} R_{t+1}^{p}\right] \leq 0 \tag{1.9}$$

for $t = 1, ..., T-1$ where $R_{t+1}^{p} = \bar{R}_{t+1} - R^{f}$ is the return in excess of the risk-free rate that the investor expected to obtain when he took some positions in the stock market and faced transaction costs in this market. It is a form of compensation for investors who tolerate the extra risk, compared to that of a risk-free asset or a benchmark in a given investment (it is the risk premium). This quantity becomes $\left(R_{t+1} - R^{f}\right)$ in the frictionless economy with R_{t+1} the gross return observed on the risky asset. The relation in (1.9) describes investors' optimal decisions in the financial market as a function of the model parameters. Because we assume only two assets available in the model, the first order condition in (1.9) is a non arbitrage condition between the risky asset and the risk-free asset in the financial market. Without trading costs in the economy, (1.9) becomes the standard Euler Equation from the portfolio optimization problem as presented in Epstein and Zin (1991) (see the relation (1.15) for instance). Hence, the expected benefit from investing only in the risky asset should correspond to the expected gain when one uses only the risk-free asset in the financial market. Nonetheless, when we account for trading costs, the non-arbitrage condition given by (1.9) implies that the expected gain from investing only in the risky asset is less than the expected benefit obtained by investing only in the risk-free asset because of the cost faced by investors when trading in the risky asset.

The consumption-investment optimization problem also implies the following terminal condition:

$$J_T(A_T, I_T) = (1-\beta)^{\frac{1}{\rho}} A_T \tag{1.10}$$

This terminal condition is very useful when one is interested in a numerical solution to the portfolio selection. It helps us to solve the problem by backward induction.

The system of Euler Equation (1.8) and Inequality (1.9) are going to be used as a set of moment conditions to estimate the parameters of the model and to construct tests.

1.3 Testing trading cost effect using GMM estimation

Our goal in this section is to develop a GMM-based test procedure which allows us to test the significance of the transaction costs effect in the economy.

1.3.1 The GMM procedure to estimate the parameter of interest

Assume that we have the following moment conditions

$$E\left[g\left(Z_t,\theta\right)\right]=0 \tag{1.11}$$

where θ is a $L\times 1$ vector of parameters, K =dim(g), and $Z_t=(w_t,x_t)$ where w_t contains model variables and x_t is the vector of instruments.

To test if trading costs have a significant effect on a given asset, we first transform the first order conditions obtained in (1.8) and (1.9) as follows:

$$E_t\left\{\left[\beta\left(\frac{C_{t+1}}{C_t}\right)^{\rho-1}R_{p,t+1}\right]^{\frac{\lambda}{\rho}}-1\right\}=0 \tag{1.12}$$

$$E_t\left\{\frac{\lambda}{\rho}\left(\frac{C_{t+1}}{C_t}\right)^{\frac{\lambda}{\rho}(\rho-1)}R_{p,t+1}^{\frac{\lambda}{\rho}-1}R_{t+1}^{p}+\delta\right\}=0 \tag{1.13}$$

where $\delta\in\mathbb{R}^+$ can be seen as a parameter which captures the transaction costs effect in the economy. This makes sense because in the frictionless case, (1.9) is satisfied with equality so that $\delta=0$. Hence, the test procedure we are going to propose in the next subsection will be about the significance of the parameter δ. Thus, for a given risky asset in the economy, a significant parameter δ means that investors have to account for trading costs in this asset when they have to include it in their optimal portfolio. However, when δ is not statistically significant then trading costs could be ignored in the portfolio selection process without significant consequences in terms of utility cost.

Let

$$g(Z_t,\theta)=\begin{pmatrix}\left[\beta\left(\frac{C_{t+1}}{C_t}\right)^{\rho-1}R_{p,t+1}\right]^{\frac{\lambda}{\rho}}-1\\ \frac{\lambda}{\rho}\left(\frac{C_{t+1}}{C_t}\right)^{\frac{\lambda}{\rho}(\rho-1)}R_{p,t+1}^{\frac{\lambda}{\rho}-1}R_{t+1}^{p}+\delta\end{pmatrix}\otimes\begin{pmatrix}1\\ x_t\end{pmatrix} \tag{1.14}$$

where $\theta=(\delta,\beta,\lambda/\rho,\rho)'$, $Z_t=\left\{\frac{C_{t+1}}{C_t},R_{p,t+1},\bar{R}_{t+1},x_t'\right\}$ where x_t is a vector of instruments (elements of the information set available at time t), C_t the level of consumption at time t, $R_{p,t+1}$ the gross return on the optimal portfolio, and $\bar{R}_t$ the gross return net of trading costs on the risky asset. Then, we use (1.11) with g defined by (1.14) as the set of moment conditions to estimate $\theta=(\delta,\psi')'$ by a two-step GMM procedure where $\psi=(\beta,\lambda/\rho,\rho)'$ with β the discount factor, $\gamma=1-\lambda$ the relative risk aversion coefficient, and $EIS=\frac{1}{1-\rho}$. In our test procedure ψ will be treated as an identified vector of nuisance parameters. Since we want to estimate θ by a two-step GMM procedure based on (1.11), let $G_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} g(Z_t, \theta)$ denote the empirical counterpart of the moment conditions defined in (1.11) where T is the sample size. Let us also denote by $l_T(\theta, \hat{W}) = -\frac{T}{2} G_T(\theta)' \hat{W} G_T(\theta)$ the GMM objective function where $\hat{W}$ is a random symmetric positive definite matrix such that $\hat{W} \xrightarrow{P} W$ with W a non-random symmetric positive definite matrix.

Let $\hat{\theta}$ denote the two-step GMM estimator of θ using (1.14) as the set of moment conditions. We obtain this estimator using the following procedure.

In the first step, we estimate θ by GMM using $W = I$ so that we obtain the first step estimator, $\hat{\theta}(I) = \text{argmax}_\theta \, l_T(\theta, I)$.

We then estimate $S = E(g(Z_t, \theta_0)g(Z_t, \theta_0)')$ by $\hat{S} = \frac{1}{T}\sum_{t=1}^{T} g_t(\hat{\theta}(I))g_t(\hat{\theta}(I))'$ where $g_t(\theta) = g(Z_t, \theta)$ so that the second step GMM estimator is given by $\hat{\theta} = \text{argmax}_\theta \, l_T(\theta, \hat{S}^{-1})$.

1.3.2 Testing the significance of the transaction cost effect

Our objective in this part is to propose a procedure to test whether the transaction costs have a significant effect on investor's welfare (in terms of utility cost) based on the two-step GMM estimation presented above.

An interesting property of this test procedure is that our results do not depend on any form given to trading costs in the model as it has been done in the literature. Indeed, the conclusions of most of the studies in the literature about trading costs effect depend largely on the form of frictions assumed in the model. Here, unlike in the previous literature, only the trading costs computation method could affect our results instead of the form assumed to the trading costs. Since, we would like to propose a procedure which allows us to test whether trading costs have a significant effect, we are only interested on the significance of δ treating ψ as an identified vector of nuisance parameters.

For this purpose we formulate the following hypothesis:

$H_0 : \delta = 0$ vs $H_1 : \delta > 0$

where $\delta \in \mathbb{R}^+$ is the parameter which informs us about the transaction cost effect in our economy. Using a compact form, the test hypothesis becomes:

$H_0 : H\theta = 0$ vs $H_1 : H\theta > 0$

where $H = (1, 0, 0, 0)$ and $\theta = (\delta, \beta, \lambda/\rho, \rho)'$ the vector of parameters to be estimated by GMM.

To implement this test, one needs to derive the asymptotic distribution for $\hat{\delta}$ under the null hypothesis.

Let us first introduce some useful notations.

Notations

Let K be the number of moment conditions, $G(\theta) = E(g(Z_t, \theta))$ and $\Gamma = \frac{\partial G(\theta_0)}{\partial \theta'}$ be the $K \times 4$ matrix of right partial derivatives of $G(\theta)$ at θ_0.

Let $l_T(\theta) = -T G_T(\theta)' \widehat{S}^{-1} G_T(\theta)/2$ and $\hat{\theta} = \arg\max_{\theta \in \Theta} l_T(\theta)$.

To derive asymptotic distributions under the null hypothesis, we also need a set of assumptions.

Assumption A.

1. $Z_t = \left\{ \frac{C_{t+1}}{C_t}, R_{p,t+1}, \bar{R}_{t+1}, x_t^{'} \right\}$ is a stationary and ergodic process.

2. $\theta_0 \in \Theta = \left\{ \theta \in \mathbb{R}^4 : \theta = (\delta, \beta, \lambda/\rho, \rho)^{'}, \delta \geq 0, 0 \leq \beta \leq 1, \|\theta_j\| \leq M_j,\ j \leq 4 \right\}.$

3. Identification: $G(\theta) = 0$ if and only if $\theta = \theta_0$.

4. Dominance: (i) $E(\sup_{\Theta} \|g(Z_t, \theta)\|) < \infty$
   (ii) $E(\sup_{\mathcal{N}} \|\frac{\partial g(Z_t,\theta)}{\partial \theta'}\|) < \infty$ where $\mathcal{N}$ is a neighborhood of θ_0 and $\frac{\partial g(Z_t,\theta)}{\partial \theta'}$ denotes the $K \times 4$ matrix of right partial derivatives of $g(Z_t, \theta)$.

5. $\widehat{S} \xrightarrow{P} S$ where $S = E\left(g(Z_t, \theta_0) g(Z_t, \theta_0)'\right)$ is a finite positive definite matrix.

6. Γ is full column rank.

Assumption A1 is a standard assumption in macroeconometrics. Assumption A2 proposes a reparametrization of the model so that the resulting moment condition $g(Z_t, \theta)$ is continuous in θ, moreover Θ is assumed to be compact which guarantees the consistency of the GMM estimator. The other assumptions are standard and can be found in textbooks (see for instance Hayashi (2000)) except that g is not assumed to be differentiable for all $\theta \in \Theta$ but only right differentiable.

A standard and convenient assumption in literature is that the true parameter θ_0 is an interior of the parameter space. Indeed, it allows the use of the mean value theorem useful to establish the asymptotic normality of $\hat{\theta}$. When the true parameter θ_0 is an interior point of Θ and Assumption A is satisfied, the following results hold (see Hayashi (2000)):

- $\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, (\Gamma' S^{-1} \Gamma)^{-1})$.

- $Wald_{stat} = \frac{T(\hat{\delta} - \delta_0)^2}{\widehat{\sigma}_\delta^2} \xrightarrow{\mathcal{L}} \chi^2(1)$, where $\widehat{\sigma}_\delta^2$ is a consistent estimator of $\sigma_\delta^2 = H\left(\Gamma' S^{-1} \Gamma\right)^{-1} H'$.

- $J = T G_T(\hat{\theta})' \widehat{S}^{-1} G_T(\hat{\theta}) \xrightarrow{\mathcal{L}} \chi^2(K-4)$ where K is the number of moment conditions and 4 the number of estimated parameters.

However, in our economic application, the true parameter θ_0 is not an interior point of Θ under the null hypothesis $H_0 : \delta = 0$. When the true parameter is on the boundary, the asymptotic distribution of $\hat{\theta}$ is no longer a standard distribution (see Andrews (1999)).

The following proposition establishes the asymptotic distribution of the Wald test statistic under the null hypothesis.

Proposition 1 *Let* $\widehat{\sigma}_{\delta}^{2}$ *denote a consistent estimator of the asymptotic variance of* $\hat{\delta}$. *Assume that Assumption A holds and that* θ_0 *is such that* $\delta = 0$ and $(\theta_2, \theta_3, \theta_4)$ are interior points of the parameter space. *Then,*

$$W = \frac{T\hat{\delta}^2}{\widehat{\sigma}_{\delta}^{2}} \xrightarrow{\mathcal{L}} \frac{1}{2}\chi^2(0) + \frac{1}{2}\chi^2(1)$$

where $\chi^2(0)$ *is the Dirac distribution at the origin and* $\chi^2(1)$ *is a chi-square distribution with one degree of freedom.*

Remark: A consistent estimator for the asymptotic variance of $\hat{\delta}$ will be obtained based on a bootstrap method. As noted by Andrews (1999), the standard bootstrap does not generate consistent estimators of the asymptotic standard errors of extremum estimator when the true parameter is on the boundary. Hence, we use a version of the bootstrap procedure in which bootstrap samples of size T_1 $(< T)$ rather than T, are employed (for more details about this procedure, see Andrews (1999, p.1371)).

The asymptotic distribution of the Wald test under H_0 is a mixture of a chi-square with one degree of freedom and a mass-point at zero. Its critical values are 1.642, 2.706, and 5.412 for significance level 10%, 5% and 1% respectively, see Carrasco and Gregoir (2002). When the true parameter is an interior point of the parameter space, the asymptotic distribution of the Wald test becomes a chi-square distribution with one degree of freedom $\chi^2(1)$ instead of a mixed distribution so that its critical values are given by 2.71, 3.84, and 6.63 for significance levels 10%, 5%, and 1%. We see that the correct critical values are smaller that those given by the $\chi^2(1)$, hence using mistakenly the $\chi^2(1)$ critical value would yield a test that lacks of power. To prove Proposition 1, we use results from Lemma 1 in Appendix A1. Its proof given in Appendix A1 draws from results by Andrews (1999).

This test procedure is based on the GMM estimation of the parameters assuming the model is correctly specified. To test the validity of the moment conditions, it is customary to test overidentifying restrictions.

1.4 Testing overidentifying restrictions

In this section we are going to propose a two-step procedure which helps us to test overidentifying restrictions when one component of the parameter of interest may be at the boundary of its parameter space. Nonetheless, we start by a simulation exercise which shows that the standard J-test performs poorly when a component of the parameter of interest is near or at the boundary of its parameter space.

1.4.1 J-test when the true parameter is near or at the boundary of the parameter space

When the number of moment conditions exceeds the number of unknown parameters to be estimated by GMM, one can test the model validity by testing overidentifying restrictions before any inference in the resulting estimation. A common test used for this purpose is the J-test proposed by Hansen (1982) and one of the assumptions underlying this test is that the true parameter is an interior point of the parameter space. In this situation, Hansen's J-statistic satisfies $J = TG_T(\hat{\theta})'\hat{S}^{-1}G_T(\hat{\theta}) \xrightarrow{\mathcal{L}} \chi^2(K-L)$ where K is the number of moment conditions and L the number of estimated parameters. However, when the true parameter is on the boundary of the parameter space, Ketz (2017) shows that the standard J-test suffers from overrejection since $J \xrightarrow{\mathcal{L}} \chi^2(K-L) - \hat{\lambda}'\Upsilon\hat{\lambda}$ (see Ketz (2017) for details). Moreover, this test statistic suffers from the same problem near the boundary of the parameter space. A simple way to control the nominal size of the J-test in such a situation is to use an adjusted critical value. Ketz (2017) also proposes a modified J-statistic which has the same asymptotic distribution as the standard J-test under the null hypothesis. However, the J-test implemented with the adjusted critical value tends to outperform the modified J-test when we are too close to the boundary of the parameter space. Therefore, we propose a simple two step procedure to test overidentifying restrictions.

Let's start by a simple simulation exercise to understand how the J-test behaves nearly or at the boundary of the parameter space.

The simulation model is specified as follows:

$$y_i = \psi + \delta x_i + u_i$$

with $E(u_i) = 0$. x_i is assumed to be the only endogenous regressor and is specified by the following equation:

$$x_i = \pi_0 + \pi_1 z_{1i} + \pi_2 z_{2i} + v_i$$

where $E(v_i z_i) = 0$, $z_i = (1, z_{1i}, z_{2i})'$ is the vector of instruments used to estimate $\theta = (\delta, \psi)'$ by the standard GMM estimation method where $\psi \in \mathbb{R}$. We assume that $\delta \in \mathbb{R}^+$, this implies that when $\delta = 0$, we are at the boundary of the parameter space so that the standard asymptotic theory fails (see Andrews (1997)).

The data are generated using the following assumption

$$\begin{pmatrix} u_i \\ v_i \\ z_{1i} \\ z_{2i} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1 & 0.5 & \sigma_1 & \sigma_2 \\ 0.5 & 1 & 0 & 0 \\ \sigma_1 & 0 & 1 & 0.2 \\ \sigma_2 & 0 & 0.2 & 1 \end{bmatrix}\right)$$

with $\psi = \pi_0 = \pi_1 = \pi_2 = 1$ for several values of δ so that depending on the chosen value of δ, we could be close to the boundary of the parameter space or not. These different values for δ help us to make a power analysis of the test procedure when δ approaches the boundary of the parameter space. σ_1 and σ_2 are the correlations between u and z_1 and z_2 respectively. We set $\sigma_1 = 0$ and generate data for different values of σ_2 so that we obtain several alternative hypotheses for overidentifying restrictions test. The sample size is chosen to be 250 as in Ketz (2017) and we replicate the procedure 100,000 times.

Table 1.1 gives us the rejection frequency of the null hypothesis for testing the overidentifying restrictions in the standard GMM estimation (under the null hypothesis of J-test, here that means $\sigma_2 = 0$), using the J-statistic for different critical values at the significant level 5% when $\delta = 0$. The column 2 gives the rejection frequency under the null hypothesis when we use the standard critical value (from $\chi^2(1)$) and the column 3 gives the same quantity using the adjusted critical value (from $0.5\chi^2(1) + 0.5\chi^2(2)$ obtained by Ketz (2017)).This result shows that at the boundary of the parameter space, the J-statistic using the $\chi^2(K - L)$ distribution as the asymptotic distribution over-rejects the null hypothesis of the J-test. Hence, one needs to adjust the critical value of the standard J-statistic in this context in order to control the size. Nonetheless, what happen with the J-test when the true parameter is not at the boundary but near to the boundary?

Table 1.1: Simulation results for two critical values at the significant level 5%

	J-stat	J-stat1
Critical values (5%)	$\chi^2(1)$	$0.5\chi^2(1) + 0.5\chi^2(2)$
Empirical size	0.1015	0.0547

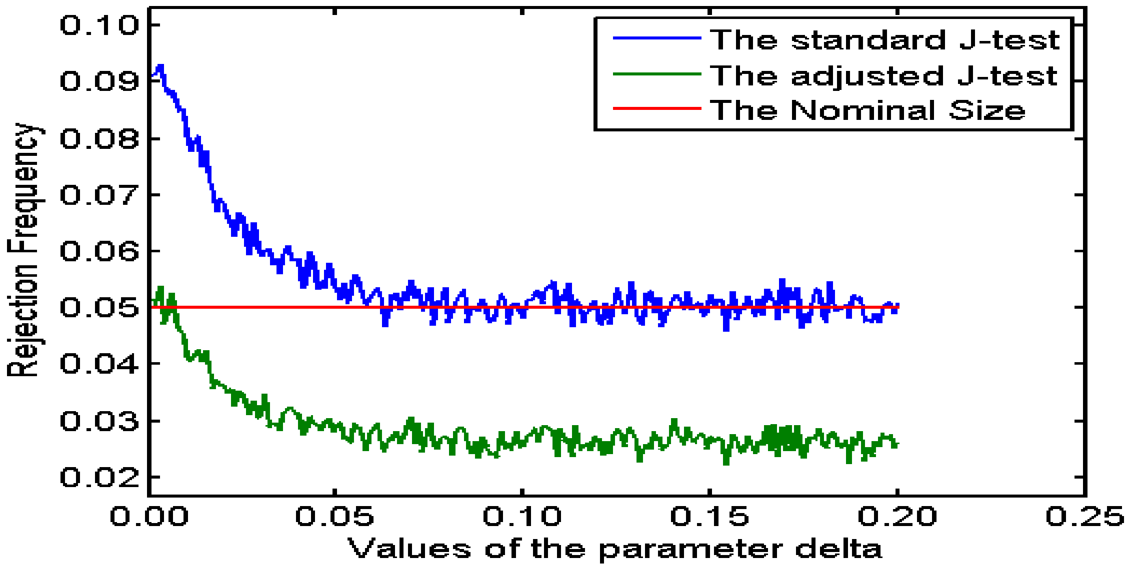
Table 1.14 gives us the same result as in Table 1.1 for several values of the parameter δ. Those results inform us about the ability of the J-test to control the size when δ approaches the boundary of the parameter space[4]. We can notice through our simulation results in Table 1.14 that the J-test overrejects the null hypothesis for values of δ close to the boundary of the parameter space when using the standard critical value. In fact, Andrews (1997) argues that asymptotic theories about the estimated parameter $\hat{\theta}$ fail when we are close to the boundary of the parameter space so that asymptotic distributions of several test statistics including the J-statistic become non standard. On the other hand, the J-statistic based on the adjusted critical value has better empirical size than the standard J-test as observed in Tables 1.14. Nonetheless, when the parameter δ becomes large enough, the standard J-test starts to control the empirical size which is not the case of the J-test based on the adjusted critical value (see Table 1.14). The same results can be observed in Figure 1.1 which gives the rejection frequency of the J-test based on the adjusted critical value and the standard one. The adjusted J-test controls the nominal size of the J-test only when the true value of the parameter δ is very close to the boundary of the parameter space and the standard J-test overrejects the null hypothesis in that case.

[4]based on both the standard critical value and the adjusted critical value

Hence, there seems to exist a certain value of the parameter δ starting from which ($\delta = 0.04$ in this simulation exercise) the standard J-test controls the empirical size when testing overidentifying restrictions. Therefore, the standard J-test will be used for values of δ greater than this threshold and we will use the adjusted critical value for the J-test otherwise. However, because this threshold is unknown, we don't know when to use the standard critical value and when to use the adjusted critical value in practice.

In the next subsection, we propose a simple procedure to test overidentifying restrictions.

Figure 1.1: The null rejection probability of the J-test for two critical values at the significance level 5%



1.4.2 A two-step procedure to test overidentifying restrictions

In this part of our analysis, we are going to propose a two-step method for testing overidentifying restrictions in our GMM estimation procedure. In the first step we will test the significance of δ based on a first step estimation. In the second step, we will use this information to decide whether to use the standard critical value or the adjusted critical value of Ketz (2017) to implement the J-test.

Nonetheless, to implement the test about the nuisance parameter in the first step, we need to have a first step consistent estimator of the parameter δ. So, we need some assumptions about the set of moment conditions used in our GMM estimation procedure.

Let $g = \begin{pmatrix} g_1 \\ g_2 \end{pmatrix}$ where $E\left[g(Z_t, \theta)\right] = 0$ is the set of moment conditions to be used in our estimation process, g_1 is a $k_1 \times 1$ vector and g_2 a $k_2 \times 1$ vector. Hence, we obtain that

$$E\left[g(Z_t,\theta)\right]=E\begin{pmatrix} g_1(Z_t,\theta) \\ g_2(Z_t,\theta) \end{pmatrix}=\begin{pmatrix} E\left[g_1\left(Z_t,\theta\right)\right] \\ E\left[g_2(Z_t,\theta)\right] \end{pmatrix}$$

To implement correctly our procedure, let us start by the following assumption.

Assumption B. $E\left[g_1(Z_t,\theta)\right]=0$ if $\theta=\theta_0$ with $k_1\geq L$ where L is the number of parameters to be estimated by GMM.

Assumption B implies that there is a subset of moment conditions which are correctly specified in order to identify the parameter θ so that we can obtain a consistent first step estimator of θ denoted by $\tilde{\theta}$ based only on $E\left[g_1(Z_t,\theta)\right]$. Moreover, the criticism raised by Guggenberger and Kumar (2012) in the standard GMM setting does not apply to our two step procedure because we do not use the J-test as pretest. Using these two assumptions, we describe our procedure as follows:

Step 1: Test the following hypothesis about the unknown nuisance parameter δ:
H_0: $\delta=0$ vs H_1: $\delta>0$ at the significance level $\alpha_1\in(0,1)$. The test of this step is implemented based on the assumption B so that we can obtain a consistent estimator of θ using only $E\left[g_1(Z_t,\theta)\right]=0$ as the set of moment conditions in the GMM process. The test statistic used to test the null hypothesis in this situation is the Wald test statistic given by:

$$W=\frac{T\tilde{\delta}^2}{\tilde{\sigma}_{\delta}^2}$$

where $\tilde{\sigma}_{\delta}^2$ is a consistent estimator of the asymptotic variance of $\tilde{\delta}$ and T is the number of observations used in the estimation process. Using the result of Proposition 1 under assumption A, we obtain that under the null hypothesis

$$W\xrightarrow{\mathcal{L}}\frac{1}{2}\chi^2(0)+\frac{1}{2}\chi^2(1)$$

where critical values have been given in Carrasco and Gregoir (2002) by 1.642, 2.706, and 5.412 for significant level 10%, 5%, and 1% respectively. We also simulate critical values for several significant level and the results of this simulation are in Table 1.13 in Appendix B.

Step 2: In this step we use the J-test to test overidentifying restrictions in our GMM estimation based on the entire available set of moment conditions $g(Z_t,\theta)$ at the significance level $\alpha_2\in(0,1)$. In fact, as mentioned before when one element of the vector of parameters to be estimated by GMM is close to the boundary of its parameter space, the asymptotic distribution of the J-statistic could be different from the standard one depending on the value of this unknown nuisance parameter. Therefore, information obtained at the first step about the unknown nuisance parameter will be used to decide if we have to use the standard critical value or the adjusted critical value. Hence, if we denote by c_{α_1} the critical value of the test implemented in the first step then in the second step the J-test is implemented as follows:

- *If* $W > c_{\alpha_1}$ *then the critical value of the J-test is the standard one* $\chi^2_{\alpha_2}(K-L)$

- *If* $W \leq c_{\alpha_1}$ *then the critical value of the J-test is* $(0.5\chi^2(K-L)+0.5\chi^2(K-L+1))_{\alpha_2}$

However, to well implement the J-test based on a two-step procedure, we need first to answer two important questions in order to have good results from our method.

Firstly, how to choose the nominal size α_2 given α_1 in order to control the global size α of the J-test implemented using a two-step procedure? To implement the J-test in the second step of our procedure, we use a Bonferroni-type correction as in Romano et al. (2014) to account for the fact that with some probability $\alpha_1 \in [0,1)$ the unknown nuisance parameter may not lie in the critical region, where α_1 is the nominal size of the test at the first step. In other words, we need to adjust the critical value of the second step by using a nominal size different from the usual nominal size of the J-test (the global size of the J-test α) see also Dufour and Kiviet (1996) for the same adjustment in a two-step test procedure. More precisely, if we denote by α the global size of the J-test, then the correct size of the test in the second step in order to control the global size of the J-test implemented using our two step procedure is given by $\alpha_2 = \alpha - \alpha_1$ where $\alpha_1 \in [0, \alpha)$.

Secondly, how to choose also the nominal size of the first step α_1 in order to have good results in terms of power analysis of the two-step J-test? Romano et al. (2014) argue that large values of the nominal size in the first step leads to somewhat reduced average power but lower values of α_1 do not make a noticeable differences in terms of average power of the two-step procedure. Moreover, in their simulation exercise, the nominal size in the first step is chosen to be 0.5%. In our simulation exercise, several values of α_1 will be used including 0.5%. Nonetheless, since we do not propose a specific way for selecting the nominal size of the first step, it may be important in future research to find a consistent way to choose this nominal size to improve the power of the two-step J-test.

Now we are going to use the same theoretical model as in Ketz (2017) to make a simulation exercise for our two step procedure of the J-test. We assume that $\sigma_1 = 0$ so that the data is generated only for different value of σ_2 which is, as mentioned before, the correlation coefficient between u and z_1. $\sigma_1 = 0$ implies that $E(uz_1) = 0$ so that we have a subset of moment conditions ($E(uz_1) = E(u) = 0$) correctly specified which could help us to consistently estimate $\theta = (\delta, \psi)'$ by GMM without using the third moment condition ($E(uz_2) = 0$).

Let $\tilde{\theta}$ be the estimator of θ obtained using only $E(uz_1) = 0$ and $E(u) = 0$. This estimation using a subset of moment conditions will be used to test the nullity of the parameter δ in the first step of our procedure at the significance level α_1 with $\alpha_1 \in [0, \alpha)$ where α is the global size of the J-test.

The first simulation exercise we did is a power analysis of the J-test implemented in one step procedure based on two different critical values. We did this analysis for a global significance level of $\alpha = 5\%$ across several values of σ_2.

The second simulation exercise is also about a power analysis of the J-test implemented in a two step procedure for the same significant levels across several values of σ_2. We use several values of the nominal size $\alpha_1 \in [0, \alpha)$ of the first step of our procedure to see how it could affect the power of the J-test implemented in a two-step procedure and its ability to control the empirical size.

We did our simulations for two different values of the parameter δ, $\delta = 0$ and $\delta = 0.04$. The results of these simulations are summarized in Tables 1.2 and 1.3 where columns 2 and 3 give us the results of the J-test implemented using the standard critical value and the adjusted critical value respectively. Column 4 contains the result about the modified J-test proposed by Ketz (2017) and columns 5, 6, 7, and 8 give the simulation results of the J-test implemented using a two-step procedure for several values of α_1. This simulations are done with 100,000 replications.

Table 1.2: Empirical rejection rate for $\delta = 0$ and $\alpha = 5\%$

σ_2	J-stat	J-stat1	Modified J-test	J-stat2 $\alpha_1 = 0.5\%$	J-stat2 $\alpha_1 = 1\%$	J-stat2 $\alpha_1 = 1.5\%$	J-stat2 $\alpha_1 = 2\%$
0.0	0.1024	0.0543	0.0504	0.0503	0.0448	0.0420	0.0346
-0.3	0.9997	0.9980	0.9489	0.9981	0.9990	0.9980	0.9970
-0.2	0.9324	0.8860	0.7021	0.9218	0.8623	0.8512	0.8304
-0.1	0.4790	0.3663	0.3204	0.4500	0.33210	0.3050	0.3206
0.1	0.2623	0.1932	0.2615	0.2620	0.1704	0.1528	0.1808
0.2	0.7519	0.6602	0.7362	0.7379	0.6282	0.6209	0.6120
0.3	0.9842	0.9686	0.9780	0.9871	0.9764	0.9567	0.9531

Table 1.3: Empirical rejection rate for $\delta = 0.04$ and $\alpha = 5\%$

σ_2	J-stat	J-stat1	Modified J-test	J-stat2 $\alpha_1 = 0.5\%$	J-stat2 $\alpha_1 = 1\%$	J-stat2 $\alpha_1 = 1.5\%$	J-stat2 $\alpha_1 = 2\%$
0.0	0.0560	0.0305	0.0504	0.0509	0.0457	0.0400	0.0165
-0.3	0.9932	0.9870	0.9489	0.9925	0.9897	0.9902	0.9870
-0.2	0.8259	0.7395	0.7021	0.8105	0.7967	0.7835	0.6589
-0.1	0.2945	0.2077	0.3204	0.3189	0.2560	0.2435	0.2269
0.1	0.2525	0.1770	0.2615	0.2597	0.1947	0.1919	0.1880
0.2	0.7470	0.6548	0.7362	0.7279	0.6789	0.6787	0.6707
0.3	0.9812	0.9650	0.9780	0.9805	0.9785	0.9760	0.9447

The first important thing we can notice from Tables 1.2 and 1.3 is that our two-step procedure out-performs the J-test based on the adjusted critical value for a wide choice of the nominal size of the first step of the two-step procedure. For instance with $\alpha_1 = 0.5\%$ as in Romano et al. (2014), the two-step test outperforms the adjusted J-test at the boundary or near to the boundary of parameter space. This result holds for the set of alternatives considered in the simulation exercise. However, the choice of the nominal size of the first step procedure is an important element to have good properties for the two-step procedure. In particular we can notice in general that the two-step procedure tends to have poor results in terms of power analysis for greater values of α_1. This result has been pointed out by Romano et al. (2014).

The second thing we can notice from our simulation results is that the power of our two-step J-test declines with the value of σ_2 in absolute value. We can also notice through this simulation result that our two-step procedure gives similar results as the modified J-test in terms of power analysis near the boundary of the parameter space. Nonetheless, our procedure outperforms the modified J-test at the boundary of the parameter space.

In the next section we are going to apply our test procedures to a problem of portfolio selection with trading costs.

1.5 Empirical Analysis

In this section we are going to apply empirically the test procedures described in Sections 1.3 and 1.4 in a context of portfolio selection with trading costs.

1.5.1 Data and data sources

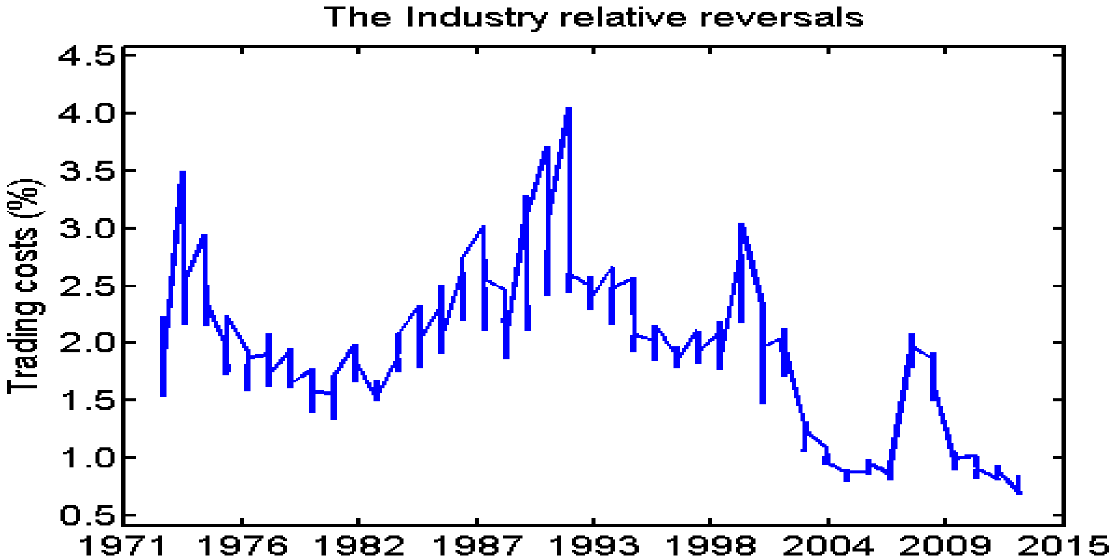
In our empirical analysis, we use monthly data from July 1973 to December 2013. The monthly rate of the return on the value-weighted NYSE index is used as a proxy for the return on the market portfolio. The one-month Treasury-Bill (T-Bill) rate is used as a proxy for the risk-free rate and R^f is calibrated to be the mean of the one-month Treasury-Bill rate observed in the data. The consumption C_t is taken to be the U.S. real per capita consumption of nondurable goods and services, and is constructed using data from the Federal Reserve Bank of St Louis database. The monthly CPI inflation corresponding to the definition of the consumption adopted is also used to deflate the stock return and the risk-free rate. The return on the market portfolio and the interest rate are from the Fama-French database and the CPI are from the Federal Reserve Bank of St. Louis database. The returns on the risky assets (here anomalies) are from Robert Novy-Marx Data Library.

1.5.2 Descriptive statistics about assets returns

In this analysis we use two measures of per capita consumption. The first one is the expenditure on nondurable goods and the second one is the expenditure on nondurable goods and services. All the nominal variables such as nominal asset returns and nominal consumption are converted into real variables using the CPI inflation index corresponding to the definition of consumption adopted. The real consumptions are put into per capita terms using total civilian population from the Federal Reserve Bank of St. Louis database.

Figure 1.2 gives us the estimated trading costs on a specific anomaly namely the industry relative reversals from July 1973 to December 2013 based on data from Novy-Marx. Those trading costs are expressed as a percentage of the gross return. The average trading costs in this strategy is about 1.86% of the gross return with strong fluctuations between 1973 and 2013. As we can see it from the graph, these frictions represent an important part of the gross return. So, one should not ignore trading costs when we are interested in a portfolio selection problem including this asset.

Figure 1.2: Trading costs in the portfolio based on the Industry relative reversals



We compute some statistics such as the empirical mean and standard deviations on some variables of interest used in the estimation process. Table 1.4 summarizes those statistics. Columns 2 and 3 of this table contain the empirical mean of each variable in column 1. Quantities in brackets are empirical standard deviation. The difference between columns 2 and 3 comes from the measure of CPI index (which changes with the measure of per capita consumption) used to transform nominal variables into real variables. Note that M_t is the real return on the market portfolio and c_{t+1}/c_t is the real consumption growth. For anomalies, we use the real returns net of transaction costs for assets whose trading costs exceed 0.50% of the gross return (see Table 1.11 in Appendix B).

The first thing we can notice from those descriptive statistics is that returns on stock market are substantially more volatile than the consumption growth and the bond market is not very volatile. The second interesting thing about the Table 1.4 is that real returns

Table 1.4: Descriptive statistics

Variables	**Mean** **Nondurable and services**	**Mean** **Nondurable**
c_{t+1}/c_t	1.0009 (0.0039)	1.0006 (0.0072)
M_t	1.0082 (0.0488)	1.0149 (0.0811)
Bonds	1.0043 (0.0039)	1.0074 (0.0060)
Failure probability	0.9997 (0.0767)	1.0007 (0.1274)
Idiosyncratic volatility	0.9986 (0.0663)	0.9985 (0.1153)
Momentum	1.0046 (0.0643)	1.0080 (0.111)
PEAD (CAR3)	1.0035 (0.0283)	1.0059 (0.0486)
Industry momentum	0.9961 (0.0534)	0.9923 (0.0938)
Industry relative reversals	0.9948 (0.0424)	0.9892 (0.0718)
High frequency combo	1.0028 (0.033)	1.0031 (0.0578)
Short run reversals	0.9905 (0.0505)	0.9820 (0.0870)
Seasonality	0.9938 (0.0403)	0.9890 (0.0670)
Industry Relative Reversals (Low volatility)	1.0029 (0.0359)	1.0037 (0.0591)

appear to be relatively stable when services are added to consumption measure indicating that the inflation index does not differ much across these measures of consumption.

1.5.3 Estimation results and testing

Our goal in this subsection is to estimate $\theta = (\delta, \psi')'$ by the two-step GMM procedure developed in Section 1.3 in order to test the significance of the transaction costs effect in the economy. To test whether trading costs in a given strategy have a significant effect, we use the result of Proposition 1 to test whether the parameter δ is significant or not. We estimate the parameter θ using the set of moment conditions obtained through the first order conditions (see (1.12)-(1.13)) of the optimal consumption-investment selection problem.

The gross return on the optimal portfolio $R_{p,t+1}$ which appears in the set of moment conditions is approximated by the gross return on the market portfolio M_{t+1} as it has been done by Epstein and Zin (1991).

The risk premium R^p_{t+1} in the relation (1.9) has been approximated by $R_{t+1} - R_f$ in Epstein and Zin (1991) where R_{t+1} is the gross return in the risky asset without accounting for trading costs. In this situation, the relation in (1.9) becomes as follows

$$E_t\left[\left(\frac{C_{t+1}}{C_t}\right)^{\frac{\lambda}{\rho}(\rho-1)} M_{t+1}^{\frac{\lambda}{\rho}-1}\left(R_{t+1}-R_f\right)\right]=0 \tag{1.15}$$

It implies that the parameter δ in the relation (1.13) which informs us about the effect of the trading costs in the economy is equal to zero. However, because we account for trading costs in the economy, instead of using $R_{t+1}-R_f$ to approximate the risk premium R^p_{t+1}, we use $\bar{R}_{t+1} - R_f$ as a proxy of this variable where $\bar{R}_{t+1}$ is the gross return net of trading costs in the risky asset. For comparison purposes, we also estimate our model using $R_{t+1} - R_f$ as a proxy of the risk premium (Epstein and Zin, 1991). We expect from this estimation that the parameter δ to be non-significant in this case.

The set of instruments used in our estimation procedure is given by: $x_{tl} = (1, \frac{c_t}{c_{t-1}}, ..., \frac{c_{t-l}}{c_{t-l-1}}, M_t, ..., M_{t-l})$ with $l \in \mathbb{N}^*$ and $x_{tl} \subset \mathcal{F}_t$ where $\mathcal{F}_t$ is investors' information set. We estimate models for several sets of instruments that means for several values of $l \in \mathbb{N}^*$ across the consumption and the risk premium measures. Several models have been estimated depending on the strategy used for the risky asset. Because the main objective through this paper is to test whether trading costs have a significant effect, we only report the estimation results about the trading cost parameter δ. The first panel in Table 1.5 contains our estimation results when the consumption is measured by the nondurable goods. The second panel in Table 1.6 provides results when the consumption is measured by the nondurable goods and services. The results of these two panels are obtained using x_{t2} ($l = 2$) as the set of instruments in our estimation process. We also estimate models with a second set of instruments x_{t3} ($l = 3$) and the results of this estimation are given in Tables 1.15 and 1.16 in Appendix B.

To test overidentifying restrictions in our estimation procedure, we use the two-step procedure proposed in Section 1.4 because there is an unknown nuisance parameter (the trading costs parameter) which could be at the boundary of its parameter space. In fact, as we saw in Section 1.4 (by simulations), when the true parameter is close to the boundary of the parameter space, the standard J-test proposed by Hansen (1982) over-rejects. An adjusted critical value has been used to overcome this problem. However, this procedure out-performs only when the nuisance parameter (which is unknown) is close to its parameter space. So, because the nuisance parameter is unknown, we use a two step procedure to implement the J-test. To implement the J-test using our two-step procedure we need to have a first step consistent estimator based on a subset of the moment conditions as pointed out in Section 1.4. In the first step, we estimate θ by GMM using the following moment conditions:

$$E\left[g_1(Z_t,\theta)\right]=0$$

with

$$g_1(Z_t,\theta)=\begin{pmatrix}\left[\beta\left(\frac{C_{t+1}}{C_t}\right)^{\rho-1}R_{p,t+1}\right]^{\frac{\lambda}{\rho}}-1\\ \frac{\lambda}{\rho}\left(\frac{C_{t+1}}{C_t}\right)^{\frac{\lambda}{\rho}(\rho-1)}R_{p,t+1}^{\frac{\lambda}{\rho}-1}R_{t+1}^{p}+\delta\end{pmatrix}\otimes w_{tp}$$

where w_{tp} is a set of instruments given by $w_{tp}=(1,\frac{c_t}{c_{t-1}},...,\frac{c_{t-p}}{c_{t-p-1}},M_t,...,M_{t-p})'$ with $p=1$. The results of this first estimation are given in Tables 1.17 and 1.18 in Appendix B. This first estimation will be used to test $H_0:\delta=0$ vs $H_1:\delta>0$ at the significance level $\alpha_1\in[0,\alpha)$ based on Proposition 1 given in Section 1.3 where α is the global size of the J-test. The critical values used to implement the first step test procedure are given in Table 1.13 depending on the nominal size of this test α_1. Information about the trading costs parameter from the first step estimation will be used in the second step to implement the J-test based on all moment conditions available. An important thing to notice is that the size of the J-test implemented in the second step is $\alpha_2=\alpha-\alpha_1$ to control the global size of the two-step J-test.

Columns 2 and 4 in Tables 1.5 and 1.6 provide the estimate results of the parameter δ across consumption measures and the risk premium. Quantities in brackets are statistics used to test whether δ is significant or not. Those quantities are computed using the following formula $\frac{T\hat{\delta}^2}{\hat{\sigma^2}_\delta}$ where $\hat{\delta}$ is from the two-step GMM estimation and $\hat{\sigma^2}_\delta$ a consistent estimator of the asymptotic variance of $\hat{\delta}$. These statistics are compared for significance levels 10%, 5%, and 1% respectively to 1.642, 2.706, and 5.412 (see Carrasco and Gregoir (2002)) based on the result of Proposition 1. Finally, columns 3 and 5 in Tables 1.5 and 1.6 summarize results on two-step J-statistics used to test overidentifying restrictions in our GMM procedure. Here, values in brackets are p-values associated with statistics. If the first step test gives us a significant value of the trading costs parameter, we test overidentifying restrictions by comparing the p-value to the nominal size α_2 instead of α. Here the nominal size of the first-step α_1 is chosen to be 0.5% with the global size of the J-test implemented using a two step procedure given by $\alpha=5\%$. The choice of the nominal size of the first step of 0.5% was motivated by the simulation of Section 1.4.

The first thing we can notice is that the trading costs parameter δ is not significant for the strategy based on the market portfolio. This result has been obtained using several set of instruments and for different measures of the consumption and the risk premium (see Tables 1.5-1.6 and 1.15-1.16). Hence, trading costs have no effect when the risky asset in the economy is assumed to be the market portfolio. We could explain this result essentially by the fact that trading costs on the market portfolio are quite low so that the utility cost becomes negligible as well as the effect in the optimal portfolio. Therefore, if

Table 1.5: GMM estimation result for testing trading costs effect

Nondurable goods ($l = 2$)				
Strategy	***Net return***		***Gross return***	
	$\hat{\delta}$	J test	$\hat{\delta}$	J test
Market Portfolio	1.3664e-05 (9.7245e-06)	9.028 (0.172)	1.0428e-05 (5.6759e-06)	9.057 (0.1704)
Size	0.0064013* (1.6657)	25.28+ (0.0003033)	0.0054747 (1.2442)	25.28+ (0.0003035)
Gross Profitability	0.0051534* (1.769)	4.775 (0.573)	0.0045914 (1.6171)	5.346 (0.5003)
Asset growth	3.5763e-13 (9.5228e-21)	7.674 (0.263)	5.1455e-13 (1.75e-20)	8.531 (0.2017)
Piotroski's F-score	0.0064663* (2.1903)	17.21+ (0.008531)	0.0042506 (0.97976)	17.41+ (0.007879)
PEAD (SUE)	0.0041693* (1.6842)	7.182 (0.3043)	6.5098e-13 (3.3749e-20)	9.515 (0.1466)
Industry Momentum	0.016371** (4.3862)	6.811 (0.3387)	1.3407e-14 (4.4593e-24)	10.35 (0.1106)
Industry Relative Reversals	0.029408** (5.312)	5.274 (0.5092)	3.2885e-13 (1.1809e-21)	1.827 (0.9349)
High Frequency Combo	0.0089251** (3.0787)	5.34 (0.501)	4.4611e-12 (3.4795e-19)	42.58 (1.413e-07)
Short-run reversals	0.038244*** (5.8189)	5.787 (0.4475)	0.0041816 (0.55006)	6.299 (0.3906)
Seasonality	0.024196*** (10.0722)	12.43 (0.05295)	5.6079e-12 (2.0994e-18)	13.23+ (0.03953)
Industry Relative Reversals (Low volatility)	0.0090724** (3.7962)	3.284 (0.1936)	8.5873e-13 (3.272e-20)	23.09+ (0.0007676)

$*$ 10%, $**$ 5%, $***$ 1%, + rejected at 4.5%

the market portfolio is used as the risky asset in our economy, the relation defined in (1.9) is satisfied with equality as in a frictionless setting even if we account for trading costs. In fact, the optimal investment policy with the market portfolio becomes very close to that we obtain in the frictionless setting in such a way that the utility cost due to the presence of trading costs becomes negligible. For instance, in a simple numerical analysis on a 10-years horizon investor who faces quadratic transaction costs, we find the following result when the market portfolio is assumed to be the risky asset in the economy. The average optimal investment policy in the risky asset is about 39.86% which is very close to the same quantity obtained in the frictionless setting (about 40.40%). These quantities have been obtained using the numerical procedure developed in Appendix A3. Moreover,

Table 1.6: GMM estimation result for testing trading costs effect (continued)

Nondurable goods and services ($l = 2$)				
Strategy	***Net return***		***Gross return***	
	$\hat{\delta}$	J test	$\hat{\delta}$	J test
Market Portfolio	1.0897e-12 (1.8736e-19)	10.16 (0.1181)	1.8e-12 (5.5e-19)	10.57 (0.1026)
Size	0.0028094 (1.1089)	18.46^{+} (0.00519)	0.0023344 (0.78437)	18.46^{+} (0.005173)
Gross Profitability	0.0028013 (1.3623)	5.574 (0.4726)	0.002528 (0.86494)	3.914 (0.6883)
Asset growth	0.0002716 (0.019201)	9.161 (0.1647)	6.9036e-13 (1.0931e-19)	9.387 (0.1529)
Piotroski's F-score	0.0036007^{*} (2.1995)	16.83^{+} (0.009932)	0.0023875 (1.0298)	16.91^{+} (0.009616)
PEAD (SUE)	0.0025953^{*} (1.9174)	6.897 (0.3305)	2.0229e-14 (9.8233e-23)	8.932 (0.1774)
Industry Momentum	0.0090454^{***} (5.6283)	10.48 (0.1057)	4.6873e-12 (1.7489e-18)	11.11 (0.08514)
Industry Relative Reversals	0.014861^{**} (4.3446)	5.8 (0.446)	5.0304e-12 (3.7817e-18)	14.74^{+} (0.02239)
High Frequency Combo	0.0042342^{**} (3.5827)	9.251 (0.1599)	7.6272e-12 (2.3702e-18)	48.91^{+} (7.757e-09)
Short-run reversals	0.020137^{*} (1.8545)	2.385 (0.8811)	0.0018176 (0.41066)	8.809 (0.1846)
Seasonality	0.0124^{***} (8.0757)	11.9 (0.06428)	5.4295e-12 (5.9229e-18)	16.62^{+} (0.0108)
Industry Relative Reversals (Low volatility)	0.0039186^{*} (1.8283)	5.975 (0.426)	2.0517e-13 (4.5296e-21)	26.92^{+} (0.0001496)

$*$ 10%, $**$ 5%, $***$ 1%, $+$ rejected at 4.5%

when trading costs have no effect for a given strategy, the second Euler equation in (1.9) could be rewritten as in (1.15) so that one can combine (1.8) and (1.15) to obtain the following set of equations:

$$E_t \left\{ \beta^{\frac{\lambda}{\rho}} \left(\frac{C_{t+1}}{C_t} \right)^{\frac{\lambda}{\rho}(\rho-1)} R_{p,t+1}^{\frac{\lambda}{\rho}-1} R_{jt+1} \right\} = 1 \tag{1.16}$$

for $j = 1, 2$. R_{jt+1} is the gross return of the asset j where $j = 1$ is the risk-free asset with $R_{1t+1} = R_f$ and $j = 2$ corresponds to the risky asset in the economy with $R_{2t+1} = R_{t+1}$ (see Appendix A2 for more details about (1.16)).

The second thing we can notice from Tables 1.5 and 1.6 (column "gross return") is that the parameter δ is not significant for any model when the risk premium R^p_{t+1} is measured by $R_{t+1} - R_f$ where R_{t+1} the gross return on the risky asset. In fact, when $R_{t+1} - R_f$ is used to measure the risk premium, this implies that we do not account for trading costs in our economy model. Hence, the non-arbitrage condition in the financial market given by the relation (1.9) is satisfied with equality so that the parameter δ which appears in the second moment condition defined in (1.13) is zero.

Nonetheless, using $\bar{R}_{t+1} - R_f$ to measure the risk premium in the estimation process, we account for trading costs in the economy model so that the parameter δ could be different from zero in the relation (1.13). We can notice for example that trading costs have a significant effect for most of strategies whose trading costs exceed 30 bp (0.03% of the gross return) in particular all anomalies with trading costs more than 1% of their gross return have a significant trading costs effect (see Tables 1.5-1.6 and 1.15-1.16). Hence, for each of those strategies, the relation defined in (1.9) is satisfied with a strict inequality. In this situation the expected benefit by investing in the risky asset only is less than the benefit in the portfolio based on the risk-free asset because of the costs incurred by investors when taking positions on the risky asset. In fact, trading costs considerably reduce the selected portfolio profitability due to their negative effect on investors risk premium. Therefore, we record an important loss in consumers' income so that investors are obliged to reduce their consumption, this in turn creates a utility loss. Moreover, we also find through a simple numerical analysis that trading costs substantially reduce the optimal holdings in the risky asset. With trading costs in the economy, the relation in (1.16) does not hold for all assets and we obtain the following results about the first order conditions (see Appendix A2):

$$E_t \left\{ \beta^{\frac{\lambda}{\rho}} \frac{\lambda}{\rho} \left(\frac{C_{t+1}}{C_t} \right)^{\frac{\lambda}{\rho}(\rho-1)} R_{p,t+1}^{\frac{\lambda}{\rho}-1} R_{jt+1} \right\} = \frac{\lambda}{\rho} + \delta \omega_{it} \beta^{\frac{\lambda}{\rho}} \tag{1.17}$$

for $i, j = 1, 2$ with $i \neq j$ where δ is a positive parameter which shows us the effect of trading costs in the economy, $\omega_{1t} = y_t$, $\omega_{2t} = 1 - y_t$, $R_{1t+1} = R_f$, $R_{2t+1} = R_{t+1}$. If δ is not significant, Equation (1.17) becomes (1.16) and the analytical solution to the optimal portfolio in this case is close to the one we have in the frictionless setting. However, when δ is significant, the analytical solution obtained using (1.16) is not the optimal one in the presence of trading costs. So, one needs to care about such a friction in the portfolio selection problem.

Most models which exhibit a significant trading cost effect when the consumption is measured by the nondurable good have also significant trading costs effect with non-durable goods and services. Nonetheless, the intensity of the effect differs across these two measures of the consumption. Moreover, the number of models with significant trading cost effect tends to increase with the number of instruments. In fact, when the number of instruments increases, estimation variance becomes smaller in such a way that the results of the tests become more accurate even if estimation bias could increase. But we can notice that estimation results are very similar across the set of instruments used in the estimation process. Hence, our results seem robust to the number of instruments used in the estimation process.

According to the results in Tables 1.5-1.6 about the two-step J-test, we reject only for two models when the transaction costs are included in the model. However, when the transaction costs are ignored in the model, the two-step J-test rejects our estimation for several models in particular when the services are adding to nondurable goods.

1.5.4 Comparison with the literature

The parameter δ in the relation (1.13) help us to test for a given asset if trading costs have a significant effect on investors' behavior. In fact, a significant δ helps us through the relation (1.17) to see how inefficient will be the analytical solution of the portfolio selection problem obtained based on the relation (1.16). In such a situation ignoring trading costs could have disaster consequences.

The parameter δ could be seen as the average adjustment to bring to the net return of trading costs so that the relation (1.9) will be satisfied with equality. In this context, this parameter could be considered as a proxy for the trading costs faced by investors in the financial market. Using δ as a proxy of trading costs simplified inferences about trading costs effect in terms of utility costs based on Proposition 1.

Table 1.19 contains in column 3 the estimates of the trading costs parameter obtained by our GMM estimation. We also report in the column 2 the average trading costs on these strategies provided by Novy-Marx and Velikov (2016). We notice that our estimation for these strategies are quite close to the average trading costs obtained by Novy-Marx and Velikov (2016) using different estimation methods.

Indeed, Novy-Marx and Velikov (2016) evaluate trading costs on anomalies using a Bayesian Gibbs Sampler on a generalized Roll (1984) model of stocks price dynamics. While we did our estimation based on a standard GMM procedure using the return net of trading costs obtained by Novy-Marx and Velikov (2016). In addition to the computation of trading costs, our estimation procedure allows us to test whether such costs have a significant effect on investors' actions in the financial market.

1.6 Economic benefits from accounting for trading costs

In this section we are going to measure the economic gain an investor can obtain when he accounts for trading costs in the portfolio selection process. This analysis will be done by comparing the out-of-sample performance of portfolio from our model (with trading costs) to the null model which ignores trading costs in the investment process.

So, assumed that we have monthly data-set of size T_1. We also consider a finite life horizon (T_2 months with $T_2 < T_1$) investor who reallocates his portfolio at the end of each month of his life cycle. Then we use the first $T_1 - T_2$ information on the data-set to estimate unknown parameters about the vector of state variables in the optimization problem. Those estimations will help us to be able to implement the numerical procedure developed in Appendix A3 in order to obtain the portfolio rule at each period of time. Hence, at each time period of his life cycle ($t = T_1 - T_2 + 1, ..., T_1$), our investor finds portfolio weights to maximize the expected utility. The investor then holds those assets for a given period (a month), realizes gains and losses and recomputes optimal portfolio weights for the next period. This procedure is repeated for each time period through the investor's life cycle generating a time series of out-of-sample portfolio returns to evaluate the performance of the models. We compute optimal portfolios for two different models one with trading costs and the other one which ignores the trading costs. For this purpose, we need to assume a given form to the transaction cost in our model. A standard way to parametrize those frictions is to model them as proportional to the amount of rebalancing.

Let f_t denote the transaction cost per dollar of portfolio value. Then, we model f_t as follows:

$$f_t = \phi_p \mid y_t - \hat{y}_t \mid$$

where $\hat{y}_t$ is the proportion of the risky asset inherited from the previous period and given by:

$$\hat{y}_t = \frac{y_{t-1}(1 - k_{t-1})A_{t-1}(1 - f_{t-1})R_t}{A_t} = \frac{y_{t-1}R_t}{y_{t-1}(R_t - R^f) + R^f}$$

k_t is the fraction of the current income allocated to the consumption at time t and ϕ_p is the proportional cost parameter associated with the risky asset (see Lynch and Balduzzi (2000) for more details). A_t is the investor's income at time t defined according to the law of motion in Equation (1.3) with $R_{p,t+1}$ such that $R_{p,t+1} = (1 - f_t)\left[y_t(R_{t+1} - R^f) + R^f\right]$ instead of (1.2) to account for the transaction cost. We still assume that the risky asset is one of the anomalies used in Novy-Marx and Velikov (2016) so that the parameter ϕ_p is given in Table 1.11 for each strategy. For example, when the risky asset in the economy is taken to be the industry-relative reversals (IRR) (one of the anomalies used in Novy-Marx and Velikov (2016)), the proportional cost parameter ϕ_p is 1.78% with a significant trading cost effect according to our empirical results obtained in Section 1.5.

Several statistics such as the mean of the portfolio return (Mean) and its standard deviation (SD), the Sharpe Ratio (SR) will be used to evaluate the out-of-sample performance of our portfolio selection process. The SR is obtained using the following relation:

$$SR = \frac{E(Portfolio) - R_f}{\sigma_{Portfolio}}$$

Because $E(Portfolio)$ and $\sigma_{Portfolio}$ are unknown, we estimate those quantities by their empirical counterpart from the sample of the optimal portfolio returns.

We report these statistics in Tables 1.7 and 1.8 (for a 10 years horizon investor with T=120) for two different anomalies. More importantly, we obtain Table 1.7 by using the parameter calibrated on the industry-relative reversals as the risky asset (ϕ_p=1.78%) and Table 1.8 with the parameter calibrated on the asset growth (ϕ_p=0.11%). The Panel A of those two tables gives statistics when we account for trading costs in the portfolio selection problem and the panel B contains the same statistics for the null model. Moreover, we compute those statistics for two different values of the EIS (see column 1 of each table) when the relative risk aversion is set to $\gamma = 6$. We report the out-of-sample mean of the optimal portfolio in column 2, the out-of-sample volatility given by the standard deviation in column 3 and the out-of-sample excess return per unit of deviation in column 4.

Table 1.7: Out-of-sample performance analysis for the Industry-relative reversals with $\gamma = 6$

EIS	Mean	SD	SR
Panel A: With trading costs in the model			
0.8	0.16	0.0105	0.1349
2	0.09	0.0107	0.0715
Panel B: Ignoring trading costs in the model			
0.8	0.08	0.010	0.0487
2	0.03	0.0127	0.0231

The first thing we can notice from Table 1.7 is that our model outperforms the null model in terms of the portfolio mean and the Sharpe ratio. For instance, we can see that the Sharpe ratio obtained in our model when the $EIS = 0.8$ is 0.1349, about 2.76 times the Sharpe ratio of the null model. A similar result is obtained with $EIS = 2$. According to the SR there is a large economic gain from accounting for trading costs in the investment process when the risky asset is the IRR. In other words accounting for the trading costs in the investment process helps investors to increase their portfolio performance in terms of the SR. This finding is consistent with the result of the empirical analysis about the effect of trading costs for this strategy. Indeed, we found empirically that the trading costs have a significant effect when the risky asset is assumed to the IRR. Thus, investors have to care about trading costs in such a situation in order to optimally behave in the financial market.

The second finding is that the effect of ignoring trading costs on the portfolio performance is more important for $EIS = 2$ than for $EIS = 0.8$. In fact, the effect of ignoring transaction costs is amplified by the fact that investors with $EIS > 1$ tend to be more aggressive on the financial market. More precisely, when $EIS < 1$, consumers' income effect is larger than their substitution effect so that investors prefer to consume more today and participate less to the financial market. In this situation, the effect of the transaction costs on the portfolio performance is attenuated by the fact that investors do not want to take risks in the financial market. However, $EIS > 1$ implies that the substitution effect is stronger than the income effect and investors prefer participation to the financial market in order to smooth the consumption in the future. This will amplify the effect of the trading costs on the the optimal portfolio performance.

Table 1.8: Out-of-sample performance analysis for the Asset growth with $\gamma = 6$

EIS	Mean	SD	SR
Panel A: With trading costs in the model			
0.8	0.33	0.0110	0.3029
2	0.39	0.0114	0.3400
Panel B: Ignoring trading costs in the model			
0.8	0.31	0.0101	0.2974
2	0.38	0.0116	0.3254

When the risky asset in the economy is assumed to be the asset growth (Ag) (as in Table 1.8) the proportional cost parameter ϕ_p is 0.11%. We found through the empirical analysis that trading costs on this strategy do not have a significant effect on the investment decision according to the test procedure developed in Section 1.3.

We can notice from Table 1.8 that no significant difference in terms of out-of-sample performance exists between our model and the null model. In fact, as we saw it in Section 1.5, trading costs have no effect on the investment decision for this strategy. Thus, the optimal investment policies from our model become very close to those of the null model in such a way that no significant difference exists between these two models.

The results of Tables 1.7 and 1.8 imply that if trading costs have no effect on investment decision according to our test procedure of Section 1.3, using trading costs in the portfolio selection process does not significantly improve the out-of-sample performance (see Table 1.8). In this context investors could ignore those frictions in their investment process to simplify their optimization problem. However, when a significant trading cost effect is obtained through the test procedure of Section 1.3, investors need to account for trading costs in the portfolio selection process in order to improve the out-of-sample performance of the optimal portfolio (see Table 1.7).

We also use an utility based statistic which is the certainty equivalent (CE) return. This is the most relevant metric to set up the out-of-sample performance since it quantifies benefits based on investors' preferences. Here the CE represents the annualized risk-free return that gives the investor the same utility as the portfolio obtained without trading costs in the model. It is a form of compensation which makes the investor indifferent between the portfolios from our model and those of the null model. When the CE > 0 investors ask a certain compensation to be added to the null model in order to obtain the same utility as in the model with trading costs. This implies that there is a gain from accounting for trading costs in the investment process. However, when the CE $\cong 0$, we conclude that there is no significant economic gain from accounting for trading costs in the portfolio selection process.

Table 1.9: The Certainty Equivalent for two models with $\gamma = 6$

EIS	The industry relative reversals	The Asset Growth
0.8	0.0300	0.00367
2	0.0700	0.00524

Table 1.9 reports the CE as defined above for two models across two different values of the EIS. We can notice through this table that the CE is very close to zero when the risky asset in the economy is assumed to be the Ag. Hence, the transaction cost does not improve significantly the investor's utility compared to his utility provided in a frictionless setting. This result is due to the fact that trading costs have no effect on investment decision for this strategy as we saw it from our empirical results given in Tables 1.5 and 1.6. However, we obtain an important CE for the model with the IRR. According to this statistic, investors have to take into account trading costs in their investment process in order to improve the out-of-sample performance of the optimal portfolio in terms of the CE.

We also observe that the CE is larger for $EIS = 2$ compared to what we obtain for $EIS = 0.8$. This result means that investors with large EIS (for instance $EIS > 1$) ask more compensation in order to be indifferent between our model and the null one. Thus, as observed for the SR, the trading costs effect on the portfolio performance seems to be important for greater values of the EIS. This analysis about the trading cost effect on the portfolio performance also justifies the importance of distinguishing the relative risk aversion from the EIS.

1.7 Conclusion

In this paper we analyze a portfolio optimization problem of a recursive preference investor who faces trading costs on stock market. In this context, we consider a simple economy with two assets including a risky asset and a risk-free asset.

We develop a simple test procedure based on a two-step GMM estimation which allows us to test whether trading costs have a significant effect on investors welfare in the economy. An interesting property of this test procedure is that the results do not depend on the form of the trading costs assumed in the model. We also propose a two-step procedure to test overidentifying restrictions when one component of the parameter of interest could be at the boundary of its parameter space. We find through a simulation exercise that our two-step procedure has good properties for a wide choice of the nominal size of the first step of the procedure. Our procedure outperforms the J-test based on the adjusted critical value and the modified J-test proposed by Ketz (2017) when the nominal size of the first step is taken to be $\alpha_1 = 0.5\%$.

In an empirical analysis we apply our test procedures to the class of anomalies used in Novy-Marx and Velikov (2016). Not surprisingly, we find that trading costs have no effect when the risky asset is assumed to be the market portfolio. Nonetheless, trading costs have a significant effect in terms of utility costs for most of anomalies from Novy-Marx and Velikov (2016) in particular those whose trading costs exceed 1% of the gross return. Thus, it is important not to ignore such a friction when making investment decisions.

We measure the economic gain using a proportional trading costs in our model by comparing the out-of-sample performance to the model which ignores trading costs in the portfolio selection process. For this purpose we use several statistics such as mean, SD, SR, and the CE. We obtain through this analysis that the investor significantly improves the out-of-sample performance only when a significant trading costs effect is obtained according to our test procedure of Section 1.3.

1.8 Appendix A

1.8.1 Appendix A1: Proof of Proposition 1

Proposition 1 is a corollary from Lemma 1 below.

Lemma 1. Assume that Assumption A holds and that θ_0 is such that $\delta = 0$ and $(\theta_2, \theta_3, \theta_4)$ are interior points of the parameter space. Then, the following results hold:

1. $\hat{\theta}$ is a consistent estimator of θ_0, i.e. $\hat{\theta} = \theta_0 + o_p(1)$.

2. $l_T(\theta)$ admits a quadratic expansion in θ given by:

$$l_T(\theta) = l_T(\theta_0) + \frac{1}{2} X_T' \mathcal{I} X_T - \frac{1}{2} q_T\left(\sqrt{T}(\theta - \theta_0)\right) + R_T(\theta)$$

   where $\mathcal{I} = \Gamma' S^{-1} \Gamma$, $X_T = \mathcal{I}^{-1}\Gamma' S^{-1}\sqrt{T} G_T(\theta_0)$, $q_T(\lambda) = (\lambda - X_T)' \mathcal{I}(\lambda - X_T)$, $\lambda \in \mathbb{R}^4$ and for all $\gamma_T \to 0$, $\sup_{\theta\in\Theta, \|\theta-\theta_0\|\leq\gamma_T}\left[\frac{|R_T(\theta)|}{(1+\left\|\sqrt{T}(\theta-\theta_0)\right\|)^2}\right] = o_p(1)$.

3. Let $\Lambda = \mathbb{R}^+ \times \mathbb{R}^3$. Let $\hat{\lambda}_T = \inf_{\lambda\in\Lambda} q_T(\lambda)$. Then, $\sqrt{T}(\hat{\theta} - \theta_0) = \hat{\lambda}_T + o_p(1)$.

4. Let $q_\delta(\lambda_\delta) = (\lambda_\delta - Z_\delta)^2 / (H\mathcal{I}^{-1}H')$ where $Z_\delta \sim N(0, H\mathcal{I}^{-1}H')$ and $q_\delta\left(\hat{\lambda}_\delta\right) = \inf_{\lambda_\delta\geq 0} q_\delta(\lambda_\delta)$. Then, $\sqrt{T}\hat{\delta} \xrightarrow{d} \hat{\lambda}_\delta$.

5. $\sqrt{T}\hat{\delta} \xrightarrow{d} \hat{\lambda}_\delta = Z_\delta I(Z_\delta \geq 0)$ so that $\hat{\delta}$ has a half-normal asymptotic distribution.

Proof of Lemma 1

1. Proof of consistency: As $g(Z_t, \theta)$ is continuous in θ and Θ is compact, the minimum $\hat{\theta}$ exists. Moreover, $\{g(Z_t, \theta_0)\}$ is continuous in Z_t and hence is stationary ergodic by Assumption A1. $G_T(\theta)$ satisfies a uniform law of large numbers by Assumption A4(i) (see for instance Hayashi (2000)). Therefore, $\hat{\theta}$ is a consistent estimator of θ_0.

2. We need to check the conditions GMM1*, GMM2, and GMM3 of Andrews (1997).

GMM1*:

GMM1*(a) requires that GMM1(a), GMM1(C), and GMM1(e) hold. GMM1(a) holds because $G_T(\theta)$ satisfies the law of large numbers, hence $G_T(\theta) \xrightarrow{P} G(\theta)$.

GMM1(c), namely $G(\theta_0) = 0$, is satisfied by Assumption A3.

GMM1(e) follows from the fact that $\widehat{S}$ does not depend on θ and $\widehat{S}$ is a consistent estimator of S by Assumption A5.

GMM1*(b) and (c) hold because the domain of $G(\theta)$ includes a set Θ^+ that satisfies conditions (i) and (ii) of Assumption 1*(a). Moreover, each element of the K vector valued function $G_T(\theta)$ has continuous right derivatives of order one on Θ^+ with probability 1.

GMM1*(d) holds because $\partial G_T(\theta)/\partial\theta$ converges in probability to $\partial G(\theta)/\partial\theta$ uniformly in θ on $\mathcal{N}$ by Assumption A4(ii).

GMM1*(e) holds because under Assumption A, $\partial G_T(\theta_0)/\partial\theta' \xrightarrow{P} \partial G(\theta_0)/\partial\theta' = \Gamma$.

GMM2:

Because $\{g(Z_t, \theta_0)\}$ is a martingale difference sequence (see Equations (12) and (13)) and the existence of S, we have a central limit theorem:

$$\sqrt{T} G_T(\theta_0) \xrightarrow{d} \mathcal{N}(0, S),$$

hence $\sqrt{T} G_T(\theta_0) = O_p(1)$ and GMM2 holds.

GMM3 is the same as our assumption A6.

By Theorem 7 of Andrews (1997), the expansion of $l_T(\theta)$ given in point 2 holds.

3. Point 3 follows from Theorem 3(a) of Andrews (1999). We need to check Assumptions 2 to 6 of Andrews (1999). By Theorem 7 of Andrews (1997), Assumptions GMM1, GMM2, and GMM3 imply Assumptions 1-3 of Andrews (1999). Assumption 4 (consistency) of Andrews (1999) follows from the point 1. Assumption 5 of Andrews (1999) holds with $B_T = b_T = \sqrt{T}$ and $\Lambda = \mathbb{R}^+ \times \mathbb{R}^3$. Assumption 6 of Andrews (1999) holds because the cone Λ is convex.

4. Point 4 follows from Theorem 4 and Corollary 1 of Andrews (1999).

5. Point 5 follows from the minimization of $q_\delta(\lambda)$.

Appendix A2: Justification of Equations (1.16) and (1.17).

In a frictionless economy (1.9) becomes

$$E_t\left[\left(\frac{C_{t+1}}{C_t}\right)^{\frac{\lambda}{\rho}(\rho-1)} R_{p,t+1}^{\frac{\lambda}{\rho}-1}\left(R_{t+1}-R_f\right)\right]=0. \tag{1.18}$$

So by multiplying (1.18) by y_t (the weight of the risky asset in the optimal portfolio), we have that

$$E_t\left[\left(\frac{C_{t+1}}{C_t}\right)^{\frac{\lambda}{\rho}(\rho-1)} R_{p,t+1}^{\frac{\lambda}{\rho}-1} y_t\left(R_{t+1}-R_f\right)\right]=0 \tag{1.19}$$

and using the fact that $y_t\left(R_{t+1}-R_f\right)=R_{p,t+1}-R_f$, (1.19) becomes

$$E_t\left[\left(\frac{C_{t+1}}{C_t}\right)^{\frac{\lambda}{\rho}(\rho-1)} R_{p,t+1}^{\frac{\lambda}{\rho}-1}\left(R_{p,t+1}-R_f\right)\right]=0 \tag{1.20}$$

which gives that

$$E_t\left[\left(\frac{C_{t+1}}{C_t}\right)^{\frac{\lambda}{\rho}(\rho-1)} R_{p,t+1}^{\frac{\lambda}{\rho}}\right]=E_t\left[\left(\frac{C_{t+1}}{C_t}\right)^{\frac{\lambda}{\rho}(\rho-1)} R_{p,t+1}^{\frac{\lambda}{\rho}-1} R_f\right]. \tag{1.21}$$

After substituting (1.21) in (1.8) we obtain the following result

$$E_t\left[\beta^{\frac{\lambda}{\rho}}\left(\frac{C_{t+1}}{C_t}\right)^{\frac{\lambda}{\rho}(\rho-1)} R_{p,t+1}^{\frac{\lambda}{\rho}-1} R_f\right]=1. \tag{1.22}$$

Moreover, by multiplying (1.18) by $1-y_t$ (the weight of the risk-free asset in the optimal portfolio) and using the same technique as before, we obtain that

$$E_t\left[\beta^{\frac{\lambda}{\rho}}\left(\frac{C_{t+1}}{C_t}\right)^{\frac{\lambda}{\rho}(\rho-1)} R_{p,t+1}^{\frac{\lambda}{\rho}-1} R_{t+1}\right]=1. \tag{1.23}$$

Therefore, (1.22) and (1.23) imply (1.16).

When we include transaction costs in the model, we have that

$$E_t\left[\frac{\lambda}{\rho}\left(\frac{C_{t+1}}{C_t}\right)^{\frac{\lambda}{\rho}(\rho-1)} R_{p,t+1}^{\frac{\lambda}{\rho}-1}\left(R_{t+1}-R_f\right)\right]=-\delta. \tag{1.24}$$

So (1.17) is obtained by replacing (1.18) by (1.24) and by using the same procedure as before.

1.8.2 Appendix A3: The numerical procedure.

The results of Section 1.6 are obtained using the same numerical procedure as in Lynch and Balduzzi (1999) and Lynch and Balduzzi (2000). Here, we explain how the optimization problem defined in (1.6) and (1.7) can be numerically solved. First, we need to discretize all the state variables in this optimization problem.

Discrete approximation of the set of state variables

Because the proportion of the portfolio in the risky asset y_t is assumed to be in $[0, 1]$ for $t = 1, ..., T$, we need to discretize this set into a grid of points. Thus, as in Lynch and Balduzzi (1999) and Lynch and Balduzzi (2000), the following grid of points on the interval [0,1] will be used to discetize y_t for all $t = 1, ..., T$: $y = \{0.00, 0.02, 0.04, ..., 0.96, 0.98, 1.00\}$ so that we obtain 50 discrete points for this variable.

Let $d_t = \log(1 + D_t)$, $r_t = \log(1 + R_t)$ with D_t the dividend yield and R_t the risky asset return. We assume that the vector of state variables $Q_t = (r_t, d_t)'$ follows a VAR model:

$$Q_{t+1} = b + AQ_t + \epsilon_{t+1}$$

where $b = (b_1, b_2)'$, $\epsilon_t = (e_{1t}, e_{2t})' \sim$ iid $\mathcal{N}(0, \Sigma)$, and $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$. We also assume that d_t is the only state variable in the VAR model that means : $a_{j1} = 0, j = 1, 2$. This last assumption about the investment opportunity set implies that the dividend yield is sufficient to well predict the risky-asset return (Fama and French (1988), Lynch and Balduzzi (1999) and Lynch and Balduzzi (2000)). Hence, the VAR model becomes:

$$\begin{cases} r_{t+1} = b_1 + a_{12}d_t + e_{1,t+1} \\ d_{t+1} = b_2 + a_{22}d_t + e_{2,t+1} \end{cases}$$

and this model will be estimated by OLS using data from U.S financial market.

But, since d_t depends on asset prices at the end of period t, the value of that regressor at the end of period $t + 1$ reflects changes in asset prices during $t + 1$ as does r_{t+1} so $E\left(e_{1,t+1} | d_{t+1}, d_t\right) \neq 0$ (see Stambaugh (1999)). Consequently, OLS estimators of coefficients of the first equation in the VAR model although consistent are biased and have sampling distributions that differ from those in the standard setting. Stambaugh (1999) shows that this bias is given by:

$$E\left(\hat{a}_{12} - a_{12}\right) = \frac{\sigma_{e_1 e_2}}{\sigma^2_{e_2}} E\left(\hat{a}_{22} - a_{22}\right)$$

It is a positive bias since the bias in $\hat{a}_{22}$ is negative and that the unexpected return $e_{1,t+1}$

is negatively correlated with the innovation in the dividend yield $e_{2,t+1}$. Empirically the value of $\frac{\sigma_{e_1e_2}}{\sigma^2_{e_2}}$ is in the order of 10 to 20 so that the magnitude of the positive bias in $\hat{a}_{12}$ is many times the negative bias in $\hat{a}_{22}$. A bias-corrected OLS estimator has been proposed in the literature in particular by Stambaugh (1999) using a Bayesian approach. However, Lewellen (2004) shows that this correction can substantially understate, in some circumstances, dividend yield's predictive ability since this approach implicitly discards any information we have about $\hat{a}_{22} - a_{22}$. Hence, using the fact that the slope in a predictive regression is strongly correlated with the dividend yield's auto-correlation, Lewellen (2004) proposes the following bias-adjusted estimator:

$$\hat{a}_{12adj} = \hat{a}_{12} - \frac{\sigma_{e_1e_2}}{\sigma^2_{e_2}} \left(\hat{a}_{22} - a_{22}\right)$$

Because the dividend yield is a persistent variable, even if we do not know $\hat{a}_{22} - a_{22}$, a lower bound can be put on it using $a_{22} \approx 1$ which gives us an upper bound on the bias in $\hat{a}_{12}$.

Based on those estimations, the following procedure is used to have a discrete approximation for Q_t. First, the dividend yield is discretized as a first order autoregressive process (Tauchen and Hussey, 1991) to obtain a discrete process of nineteen points. For the return on the risky asset, we use the fact that the VAR model implies the following expression for the stock returns:

$$r_{t+1} = b_1 + a_{12}d_t + \nu e_{2,t+1} + u_{t+1}$$

where ν is the regression coefficient from regressing e_1 on e_2 and u is an i.i.d. normally distributed random variable with 0 mean and unknown variance σ^2_u, and assumed to be uncorrelated with e_2. The quadrature method is used to have a discrete distribution for u (with three points) calibrating σ^2_u by an estimator which is given by:

$$\hat{\sigma^2_u} = \frac{1}{T-1}\sum_{t=1}^{T} \hat{u_t}^2 = \frac{1}{T-1}\sum_{t=1}^{T} \left(e_{1t} - \hat{e_{1t}}\right)^2$$

Then, we can have a discrete distribution for r_{t+1} for each $\{d_t, d_{t+1}\}$ since $e_{2,t+1} = d_{t+1} - b_2 - a_{22}d_t$, so $r_{t+1} = b_1 + a_{12}d_t + \nu\left(d_{t+1} - b_2 - a_{22}d_t\right) + u_{t+1}$. Hence, we obtain a discrete process for the asset return distribution with $19 \times 19 \times 3 = 1083$ which will be used to implement our numerical procedure. More details about this numerical method can be found in Lynch and Balduzzi (1999) and Lynch and Balduzzi (2000).

The estimation is done using data from the Federal Reserve Bank of St. Louis database. The VAR model gives us the following results in Table 1.10.

The result from regressing e_1 on e_2 provides $\hat{\nu} = -11.9831$ with a standard error of 0.9298 and the unknown variance σ^2_u is calibrated by $\hat{\sigma^2_u}$ so that we obtain 0.0064.

$\frac{\sigma_{e_1e_2}}{\sigma^2_{e_2}}$ is calibrated to be -11.98 using data so that $\hat{a}_{12adj}$ is given by 0.2607.

Table 1.10: The VAR model estimation results

Estimation results	$a_{\cdot}$	$b_{\cdot}$	Adjusted R^2
r_{t+1}	0.3191 (0.0719)	-0.0240 (0.0082)	0.025
d_{t+1}	0.9951 (0.0025)	3.88e-4 (2.87e-4)	0.9925

We can now compute the investor's optimal investment strategy by solving the optimization problem for some values of the preferences parameters and transaction costs parameter.

1.9 Appendix B

Table 1.11: The list of anomalies and average transaction costs on each anomaly

Anomalies	Average trading costs (%)	Signal
Size	0.04	Market equity
Gross profitability	0.03	Gross profitability
Value	0.05	Book-to-market equity
ValProf combo	0.06	Sum of firms' ranks in univariate sorts on book-to-market and gross profitability
Accrusals	0.09	Accruals
Asset growth	0.11	Asset growth
Investment	0.10	Investment
Piotroski's F-score	0.11	Piotroski's F-score
Net issuance	0.20	Net stock issuance
Return-on-book equity	0.38	Return-on-book equity
Failure probability	0.61	Failure probability
ValMomProf combo	0.43	Sum of firms' ranks in univariate sorts on book-to-market,gross profitability, and momentum
ValMom combo	0.41	Sum of firms' ranks in univariate sorts on book-to-market and momentum
Idiosyncratic volatility	0.52	Idiosyncratic volatility, measured as the residuals of regressions of their past three months' daily returns on the daily returns of the Fama-French three factors
Momentum	0.65	Prior year's stock performance excluding the most recent month
PEAD (SUE)	0.46	Standardized unexpected earnings (SUE)
PEAD (CAR3)	0.57	Cumulative three-day abnormal return around announcement (days minus one to one)
Industry momentum	1.22	Industry past month's return

Source: from Novy-Marx and Velikov (2016)

Table 1.12: List of anomalies and average transaction costs on each anomaly (Continued)

Anomalies	Average trading costs (%)	Signal
Industry-relative reversals	1.78	Difference between a firm's prior month's return and the prior month's return of their industry
High-frequency combo	1.45	Sum of firms' ranks in the univariate sorts on industry relative reversals and industry momentum
Short-run reversals	1.65	Prior month's returns
Seasonality	1.46	Average return in the calendar month over the preceding five years
Industry-relative-reversals (Low volatility)	1.06	Industry relative reversals, restricted to stocks with idiosyncratic volatility lower than the NYSE median for the month

Source: from Novy-Marx and Velikov (2016)

Table 1.13: Critical values for Proposition 1 with several significant levels

Significant level (%)	0.5	1.5	2	2.5	3	3.5	4	4.5	6	6.5
Critical value	6.6262	4.709	4.223	3.844	3.545	3.292	3.068	2.878	2.416	2.293

Table 1.14: The empirical size for different value of δ and $\alpha = 5\%$

δ	J-stat	J-stat1
0.001	0.0988	0.0539
0.002	0.0963	0.0535
0.003	0.0934	0.0521
0.004	0.0916	0.0522
0.005	0.0875	0.0467
0.006	0.0859	0.0474
0.007	0.0886	0.0483
0.008	0.0847	0.0473
0.009	0.0783	0.0428
0.010	0.0853	0.0451
0.020	0.0670	0.0347
0.030	0.0592	0.0316
0.040	0.0562	0.0296
0.050	0.0557	0.0281
0.060	0.0493	0.0272
0.070	0.0488	0.0265
0.080	0.0538	0.0270
0.090	0.0495	0.0259
0.100	0.0470	0.0236
0.200	0.0502	0.0255
0.300	0.0529	0.0279
0.400	0.0491	0.0265
0.500	0.0475	0.0249
0.600	0.0493	0.0260

Table 1.15: GMM estimation result for testing trading costs effect

Strategy	Net return		Gross return	
Nondurable goods ($l = 3$)	$\hat{\hat{\delta}}$	J test	$\hat{\hat{\delta}}$	J test
Market Portfolio	1.1202e-5 (6.9561e-6)	11.31 (0.3339)	7.4085e-6 (3.0402e-6)	11.31 (0.3339)
Size	0.006213* (1.9759)	28.19+ (0.001682)	0.0053058 (1.4745)	28.2+ (0.001678)
Gross Profitability	0.0060357** (2.8466)	8.569 (0.5734)	0.0054511 (2.4264)	8.492 (0.5809)
Asset growth	0.00075158 (0.054149)	9.813 (0.4571)	2.7102e-11 (6.6775e-17)	10.19 (0.4243)
Piotroski's F-score	0.006984* (2.6078)	19.43+ (0.03516)	0.0047469 (1.2549)	19.63+ (0.03292)
PEAD (SUE)	0.0050633* (2.6548)	8.428 (0.5871)	4.8668e-14 (2.6289e-22)	10.26 (0.4181)
Industry Momentum	0.015866*** (6.1015)	12.3 (0.2653)	8.9095e-12 (2.3689e-18)	13.13 (0.2166)
Industry Relative Reversals	0.029084*** (6.0485)	8.48 (0.5821)	3.3939e-12 (8.1042e-19)	15.87 (0.1033)
High Frequency Combo	0.0090193** (4.6537)	10.57 (0.392)	2.9413e-14 (1.7572e-23)	44.6+ (2.565e-16)
Short-run reversals	0.037915*** (7.1833)	9.867 (0.4523)	0.0042125 (0.6927)	11.57 (0.3147)
Seasonality	0.023264*** (10.7918)	13.58 (0.193)	4.4464e-13 (1.3648e-20)	15.84 (0.1043)
Industry Relative Reversals (Low volatility)	0.0084407** (4.5955)	10.9 (0.3651)	6.7388e-12 (2.2958e-19)	28.12+ (0.001724)

$*$ 10%, $**$ 5%, $***$ 1%, + rejected at 4.5%

Table 1.16: GMM estimation result for testing trading costs effect

Nondurable goods and services ($l = 3$)				
Strategy	***Net return***		***Gross return***	
	$\hat{\delta}$	J test	$\hat{\delta}$	J test
Market Portfolio	1.5086e-11 (4.1184e-17)	13.03 (0.2219)	1.761e-12 (5.6067e-19)	13.03 (0.2218)
Size	0.003591* (2.1682)	20.89+ (0.02184)	0.0031195 (1.6788)	20.94+ (0.02153)
Gross Profitability	0.0035227* (2.2638)	7.147 (0.7115)	0.0032024 (2.2041)	7.96 (0.6327)
Asset growth	0.000987 (0.31822)	11.93 (0.29)	1.9722e-12 (1.2802e-18)	12.21 (0.2712)
Piotroski's F-score	0.0038143* (2.5751)	19.48+ (0.0346)	0.0025587 (1.2394)	19.6+ (0.03325)
PEAD (SUE)	0.0025297* (2.0475)	9.248 (0.5088)	1.5506e-12 (7.7305e-19)	11.45 (0.3238)
Industry Momentum	0.009534*** (5.6769)	10.04 (0.4369)	2.9285e-12 (8.9548e-19)	12.22 (0.2755)
Industry Relative Reversals	0.01484*** (5.7915)	7.521 (0.6756)	2.1991e-14 (8.4784e-23)	17.59 (0.0623)
High Frequency Combo	0.0043807** (3.9421)	9.754 (0.4623)	2.7008e-12 (4.1934e-19)	49.04 (4.003e-7)
Short-run reversals	0.019843*** (6.0628)	7.331 (0.6939)	0.0012486 (0.17565)	8.172 (0.06121)
Seasonality	0.01262*** (9.2385)	13.53 (0.1954)	9.8411e-13 (2.2143e-19)	19.34+ (0.03619)
Industry Relative Reversals (Low volatility)	0.0041517** (3.6257)	8.837 (0.5477)	1.9125e-12 (4.7762e-19)	28.15+ (0.001709)

$*$ 10%, $**$ 5%, $***$ 1%, + rejected at 4.5%

Table 1.17: GMM estimation result for testing trading costs effect

Nondurable goods ($l = 1$)				
Strategy	***Net return***		***Gross return***	
	$\hat{\hat{\delta}}$	J test	$\hat{\hat{\delta}}$	J test
Market Portfolio	0.002353 (0.17381)	6.29^{+} (0.04306)	0.0023659 (0.16608)	5.85 (0.05366)
Size	0.0073 (1.543)	22.75^{+} (0.0002)	0.0063348 (1.1955)	22.74^{+} (1.153e-05)
Gross Profitability	0.0056 (1.280)	2.128 (0.345)	0.0050697 (1.1253)	2.155 (0.3405)
Asset growth	3.7017e-12 (6.712e-19)	3.805 (0.1492)	1.64e-13 (9.8206e-22)	5.283 (0.07125)
Piotroski's F-score	0.0047834 (1.1356)	13.14^{+} (0.001402)	0.0025528^{+} (0.32826)	13.34 (0.00127)
PEAD (SUE)	0.0033 (0.9505)	3.765 (0.1522)	2.2118e-14 (2.4805e-23)	6.916^{+} (0.03149)
Industry Momentum	0.0134^{**} (3.224)	4.563 (0.1021)	6.3617e-14 (5.2885e-23)	7.294^{+} (0.02607)
Industry Relative Reversals	0.0303^{*} (2.4073)	1.97 (0.3735)	3.0753e-13 (4.9125e-21)	6.493^{+} (0.03891)
High Frequency Combo	0.008^{**} (2.858)	4.342 (0.114)	1.0676e-12 (1.1622e-20)	41.16^{+} (1.154e-09)
Short-run reversals	0.041^{*} (2.299)	1.6 (0.4494)	0.0061406 (0.7371)	2.767 (0.2507)
Seasonality	0.025^{***} (8.525)	10.48^{+} (0.0053)	2.7857e-12 (4.8852e-19)	10.83^{+} (0.004442)
Industry Relative Reversals (Low volatility)	0.01^{*} (2.043)	3.284 (0.1936)	1.5166e-12 (6.6958e-20)	17.99^{+} (0.0001243)

$*$ 10%, $**$ 5%, $***$ 1%, + rejected at 5%

Table 1.18: GMM estimation result for testing trading costs effect

Nondurable goods and services ($l = 1$)				
Strategy	***Net return***		***Gross return***	
	$\hat{\hat{\delta}}$	J test	$\hat{\hat{\delta}}$	J test
Market Portfolio	3.2804e-12 (1.6485e-18)	6.77^{+} (0.03387)	3.23e-12 (1.7249e-18)	7.657^{+} (0.02174)
Size	0.00273 (0.64431)	10.7^{+} (0.00474)	0.0022781 (0.63711)	14.59^{+} (0.0006783)
Gross Profitability	0.00344 (0.79065)	1.4 (0.4966)	0.0031298 (0.71032)	1.405 (0.4954)
Asset growth	0.0012 (0.3975)	5.379 (0.06791)	1.4792e-13 (3.4053e-21)	4.112 (0.128)
Piotroski's F-score	0.0035 (1.41)	7.395^{+} (0.02479)	0.0021802 (0.81194)	10.29^{+} (0.00583)
PEAD (SUE)	0.00256 (1.3346)	2.693 (0.2602)	9.2752e-13 (1.445e-19)	4.594 (0.1006)
Industry Momentum	0.00795^{**} (3.064)	4.42 (0.1097)	1.0147e-12 (4.3657e-20)	6.958^{+} (0.03084)
Industry Relative Reversals	0.0149^{*} (1.96)	2.367 (0.3061)	4.2774e-13 (1.7953e-20)	11.02^{+} (0.004038)
High Frequency Combo	0.00357^{*} (1.672)	3.406 (0.1822)	4.7894e-13 (6.0685e-21)	48.09^{+} (3.615e-11)
Short-run reversals	0.021^{*} (1.8231)	1.986 (0.3705)	0.0019814 (0.27817)	2.665 (0.2638)
Seasonality	0.0131^{***} (7.0885)	9.598^{+} (0.00824)	9.7158e-13 (1.7835e-19)	13.23^{+} (0.001341)
Industry Relative Reversals (Low volatility)	0.0041 (1.609)	3.572 (0.1676)	2.3293e-12 (4.2284e-19)	22.66^{+} (1.198e-05)

$*$ 10%, $**$ 5%, $***$ 1%, + rejected at 5%

Table 1.19: Comparison of trading costs

Anomalies	Average trading costs	Our estimate trading costs
Size	0.04	0.27
Gross profitability	0.03	0.34
Asset growth	0.11	0.12
Piotroski's F-score	0.11	0.35
PEAD (SUE)	0.46	0.33
Industry momentum	1.12	1.34
Industry relative reversals	1.78	1.49
High-frequency combo	1.45	1.0
Short-run reversals	1.65	2.1
Seasonality	1.46	1.31
Industry Relative Reversals (Low volatility)	1.06	1.0

Chapter 2

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

2.1 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

*I am greatly indebted to Marine Carrasco for her invaluable guidance. I am grateful to Benoit Perron, Georges Dionne, 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] 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.

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 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. 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. DeMiguel et al. (2009) propose a general framework in terms of a norm-constrained minimum-variance portfolio. Brandt et

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

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. 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 2.2 presents the economy and shows that the dynamic portfolio selection problem can be seen as a linear inverse problem. Section 2.3 imposes a constraint on the portfolio weights in the dynamic problem and derives a closed-form solution to this new problem. In Section 2.4, we impose a temporal stability constraint on the dynamic portfolio optimization. Section 2.5 gives some asymptotic properties of the selected strategy and proposes data-driven methods to select the optimal tuning parameter. Section 2.6 presents some simulation results and an empirical study. Section 2.7 concludes the paper.

2.2 The model and an empirical fact

2.2.1 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

$$R_{p,t+1} = \omega_t' \bar{R}_{t+1} + R_f \left(1 - \omega_t' 1_N\right) = R_f + \omega_t' r_{t+1}. \tag{2.1}$$

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

$$A_{t+1} = A_t R_{p,t+1} = A_t (R_f + \omega_t' r_{t+1}) \tag{2.2}$$

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

$$V(0, A_0) = \max_{\{\omega_s\}_0^{T-1}} E_0(U(A_T)), \tag{2.3}$$

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:

$$V(t, A_t) = \max_{\omega_t} E_t \{V(t, A_{t+1})\} = \max_{\omega_t} E_t \left\{V(t+1, A_t(R_t^f + r_{t+1}\omega_t^{'}))\right\}, \tag{2.4}$$

$t = 0, ..., T-1$ 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 (2.4) is given by

$$\omega_t = \gamma_t \Sigma_t^{-1} \mu_t, \tag{2.5}$$

with $\gamma_t = \left(\gamma A_t R_f^{T-t-1}\right)^{-1}$ for $t = 0, ..., T-1$, which can be seen as 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 (2.5) 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

$$\widehat{\omega}_t = F\left(\gamma, R_f, \{\widehat{\mu}_s\}_{s=0}^{t}, \left\{\widehat{\Sigma}_s\right\}_{s=0}^{t}\right). \tag{2.6}$$

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.

2.2.2 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 (2.5) or equivalently by the following equation

$$\Sigma_t\omega_t = \gamma_t\mu_t. \tag{2.7}$$

Equation (2.7) can be seen as an inverse problem because it can be written as follows

$$\Sigma_t\omega_t = \eta_t, \tag{2.8}$$

where $\eta_t = \gamma_t\mu_t$. Equation (2.8) 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 (2.7) does not admit a unique solution. To ensure the uniqueness of the solution of Equation (2.7) 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

[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.

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 (2.8) by dampening the explosive effect of the inversion of the singular values of $\hat{\Sigma}_t$. It consists in replacing the sequence $\{1/\lambda_j\}$ of explosive inverse singular values by a sequence $\{q(\alpha, \lambda_j)/\lambda_j\}$ 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.

2.2.3 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 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 2.2 for more details) which may be considered five risky assets and the Kenneth French data library provides information about the returns of those assets.

We estimate Equation (2.5) 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 2.3 and 2.4 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.

2.3 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

$$V(0, A_0) = \max_{\left\{\omega_s : \|\omega_s\|_2^2 \leq d_s\right\}_0^{T-1}} E_0(U(A_T)) \tag{2.9}$$

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

2.3.1 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 (2.9) is given by

$$V(t, A_t) = \max_{\left\{\omega_t : \|\omega_t\|_2^2 \leq d_t\right\}} E_t \left\{V(t+1, A_{t+1})\right\} = \max_{\left\{\omega_t : \|\omega_t\|_2^2 \leq d_t\right\}} E_t \left\{V(t+1, A_t(R_f + \omega_t' r_{t+1}))\right\} \tag{2.10}$$

We then obtain the following result for this optimization problem.

Proposition 1. *Under the assumptions about the economy stated in Section 2.2.1,* *the solution of (2.10) can be obtained by solving the following unconstrained problem*

$$\max_{\{\omega_t\}} \left\{ \exp\left(\lambda_t \|\omega_t\|^2\right) E_t \left[V\left(t+1, A_t\left(R_f + \omega_t' r_{t+1}\right)\right)\right]\right\} \tag{2.11}$$

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 (2.11) can be written as follows

$$\omega_t^* = \left(\gamma A_t R_f^{T-t-1}\right)^{-1} \left(\Sigma_t + \alpha_t I_N\right)^{-1} \mu_t \tag{2.12}$$

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 (2.11) 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 (2.4) 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 (2.5). $\frac{\alpha_t}{2}\left(\gamma R_f^{T-t-1}\right)^{T-t} A_t \|\omega_t\|^2$ 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 (2.12) is that with an appropriate constraint on the portfolio weights, we solve the problem of ill-posedness that arises when trying to estimate (2.5). Imposing such a constraint may thus improve the performance of the estimated portfolio. The solution of this corollary is in fact a particular regularized version (the Ridge regularization) of the optimal solution obtained in (2.5). 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 (2.12) 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

$$\hat{\omega}_{\alpha_t}^{RdgP} = \hat{\gamma}_t \left(\hat{\Sigma}_t + \alpha_t I_N\right)^{-1} \hat{\mu}_t \tag{2.13}$$

2.3.2 Comments on the result of the first procedure

The first thing to note about this method is that the selected strategy in (2.12) 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.

2.4 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.

2.4.1 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

$$V(0, A_0) = \max_{\left\{\omega_s : \|\omega_s - \omega_{s-1}\|_2^2 \leq d_s\right\}_0^{T-1}} E_0(U(A_T)). \tag{2.14}$$

In the recursive form we have that

$$V(t, A_t) = \max_{\left\{\omega_t : \|\omega_t - \omega_{t-1}\|_2^2 \leq d_t\right\}} E_t \left\{ V(t+1, A_t(R_f + \omega_t^{'} r_{t+1})) \right\} \tag{2.15}$$

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

$$\gamma_t^{-1} \left(\Sigma_t + \alpha_t I_N\right) \omega_t = \mu_t + \alpha_t \omega_{t-1}. \tag{2.16}$$

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 (2.16).

Proposition 2. *The optimal solution of the optimization problem in (2.15) 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*

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

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 2.5.5.

2.4.2 Imposing an 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 asset 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

$$V(t, A_t) = \max_{\left\{\omega_t : \|\omega_t - \omega_{t-1}\|_1 \leq d_t\right\}} E_t \left\{ V(t+1, A_t(R_f + \omega_t' r_{t+1})) \right\}. \tag{2.17}$$

Note that, unlike what we obtained in Subsection 2.4.1, 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 (2.5 as follows)

$$\omega_t = \gamma_t \frac{\theta_t}{1-\mu_t'\theta_t} \tag{2.18}$$

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

$$\hat{\theta}_{\alpha_t} = arg \min_{\theta_t} \frac{1}{n} \left\|1_n - R_t\theta_t\right\|_2^2 + \alpha_t \left\|\theta_t - \theta_{t-1}\right\|_1 \tag{2.19}$$

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. In fact, 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, in fact, 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

$$V(t, A_t) = \max_{\left\{\omega_t : \left\|\omega_t\right\|_1 \leq d_{1t}, \left\|\omega_t - \omega_{t-1}\right\|_1 \leq d_{2t}\right\}} E_t\left\{V(t+1, A_t(R_f + \omega_t' r_{t+1}))\right\}. \tag{2.20}$$

2.5 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 2.4.2 to easily derive our asymptotic properties.

2.5.1 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

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

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 conditions in assumption A can be found in Carrasco et al. (2007) and Carrasco (2012). Moreover, 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*

$$\left\|\hat{\omega}_{\alpha_t}^{RdgP} - \omega_t\right\| = o_p(1) \tag{2.21}$$

if $\max_{0\leq j\leq t-1}\left\{\frac{N^{3/2}}{\alpha_j\sqrt{n}} + \sqrt{N}\alpha_j^{min(\frac{\tau_j}{2},1)}\right\} \to 0$, $\sqrt{N}\alpha_t^{\min(\frac{\tau_t}{2},1)} \to 0$ *and* $\frac{N}{\alpha_t\sqrt{n}} \to 0$ *as* $n \to \infty$ *where* $\hat{\omega}_{\alpha_t}^{RdgP}$ *is the estimated version of the selected portfolio obtained by imposing the L2 norm on the portfolio weights.*

Proof. In the appendix.

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

2.5.2 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:

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

The Sharpe ratio of the optimal portfolio at period t as defined in Equation (2.5) is thus given by

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

However, as mentioned in Section 2.2, 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*

$$s_t\left(\hat{\omega}_{\alpha_t}\right)^2 = s_t\left(\omega_t\right)^2 + O_p\left[\left(\frac{N}{\alpha_t\sqrt{n}} + \|\theta_t - \theta_{\alpha_t}\|\right)\right], \tag{2.22}$$

for the RdgP and the L2TSP if $\frac{N}{\alpha_t\sqrt{n}} \to 0$ as n goes to infinity where ω_t is the optimal portfolio at the time point t given by the equation (2.5) 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 several things. First, it 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. Second, $\alpha_t\sqrt{n}$ may go to infinity faster than the number of assets in the financial market. Therefore, the number of assets may be limited asymptotically compared with the estimation window. More importantly, under the regularity condition $\frac{N}{\alpha_t\sqrt{n}} \to 0$ the result of Proposition 4 can be rewritten as follows

$$s_t\left(\hat{\omega}_{\alpha_t}\right)^2 = s_t\left(\omega_t\right)^2 + o_p\left(1\right) \tag{2.23}$$

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 the next subsection).

2.5.3 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

$$MSE\left(\hat{\omega}_{\alpha_t}\right) = \frac{1}{Nn} E\left[\left\|\hat{\Sigma}_t\left(\hat{\omega}_{\alpha_t} - \omega_t\right)\right\|_2^2\right] \tag{2.24}$$

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*

$$MSE\left(\hat{\omega}_{\alpha_t}^{L2TSP}\right) \sim \frac{N^2}{n^2\alpha_t^2} + \frac{N}{n}\alpha_t^{\min(\tau_t,2)} \tag{2.25}$$

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

$$MSE\left(\hat{\omega}_{\alpha_t}^{L2TSP}\right) \leq MSE\left(\hat{\omega}_{\alpha_t}^{RdgP}\right) \tag{2.26}$$

if $\max_{0\leq j\leq t-1}\left\{\frac{N^{3/2}}{\alpha_j\sqrt{n}} + \sqrt{N}\alpha_j^{min(\frac{\tau_j}{2},1)}\right\} \to 0$, $\sqrt{N}\alpha_t^{\min(\frac{\tau_t}{2},1)} \to 0$ *as* $n \to \infty$, *where we denote by* $\hat{\omega}_{\alpha_t}^{RdgP}$ *the optimal strategy obtained by Ridge regularization and* $\hat{\omega}_{\alpha_t}^{L2TSP}$ *the solution obtained with the L2 temporal stability constraint.*

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. 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 (2.19).

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 $\xi_{\Omega_t}^2$ 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 close 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).

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*

$$MSE\left(\hat{\omega}_{\alpha_t}^{L1TSP}\right) \sim N\alpha_t^2 \left(s_t/\xi_{\Omega_t}^2\right) + N\alpha_t \left\|\theta_t - \theta_{t-1}\right\| \quad (2.27)$$

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.

2.5.4 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*

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

if $\max_{0 \le j \le t-1}\left\{\frac{N^{3/2}}{\alpha_j \sqrt{n}} + \sqrt{N}\alpha_j^{min\left(\frac{\tau_j}{2},1\right)}\right\} \to 0$, $\max\left(\sqrt{N}, \alpha_t \frac{\sqrt{n}}{\sqrt{N}}\right)\alpha_t^{\min\left(\frac{\tau_t}{2},1\right)} \to 0$, *and* $\frac{N^{5/2}}{\alpha_t n} + \frac{N^{3/2}}{\sqrt{n}} + N^{3/2}\alpha_t^{\min\left(\frac{\tau_t}{2},1\right)} \to 0$ *as* n *goes to infinity.*

Proof. In Appendix.

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 $\varphi \in (0, 1)$.

2.5.5 Data-driven Method for Selecting the Tuning Parameter

Sections 2.3 and 2.4 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.

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*

$$\left(1-\mu^{'}\beta_t\right)^4\gamma_t^{-2}E\left[\left(\mu^{'}\left(\hat{\omega}_t-\omega_t\right)\right)^{'}\left(\mu^{'}\left(\hat{\omega}_t-\omega_t\right)\right)\right]\sim\frac{1}{n}E\left[\left\|1_n^{'}R_t\left(\hat{\theta}_t-\theta_t\right)\right\|^2\right] \quad (2.28)$$

if $\max_{0\leq j\leq t-1}\left\{\frac{N^{3/2}}{\alpha_j\sqrt{n}}+\sqrt{N}\alpha_j^{min(\frac{\tau_j}{2},1)}\right\}\to 0$ *as* $n\to\infty$

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, 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:

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

(ii) Mallow's C_L where:

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

with,

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

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

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

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

$$\hat{\theta}_{\alpha_t} = arg \min_{\theta_t} \frac{1}{n} \left\| 1_n - R_t \theta_t \right\|_2^2 + \alpha_t C\left(\theta_t, \theta_{t-1}\right) \tag{2.29}$$

where

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

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)

$$\hat{\tilde{\theta}}_{\alpha_t} = arg \min_{\tilde{\theta}_t} \frac{1}{n} \left\| y_t - R_t \tilde{\theta}_t \right\|_2^2 + \alpha_t C\left(\tilde{\theta}_t\right) \tag{2.30}$$

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 (2.30) 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 (2.30).

2.6 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.

2.6.1 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 (2.5). 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 $t = 1, ..., n + T$

$$r_{it} = b_{i1}f_{1t} + b_{i2}f_{2t} + b_{i3}f_{3t} + \epsilon_{it} \tag{2.31}$$

$f_t = (f_{1t}, f_{2t}, f_{3t})'$ 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 Σ_ϵ=diag$(\sigma_1^2, ..., \sigma_N^2)$ with the diagonal elements drawn from a uniform distribution between 0.10 and 0.30 to yield an average cross-sectional volatility of 20%.

In the compact form (2.31) can be written as follows:

$$R = BF + \epsilon \tag{2.32}$$

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 2.1. 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 2.1: 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 $\mathrm{SR}(\omega_t)$ be the Sharpe ratio associated with the optimal portfolio ω_t, then $\mathrm{SR}(\omega_t)$ is given as follows

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

To set up 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 2.2, 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 2.4 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 2.5 and 2.6 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 2.5 and 2.6, 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 2.7. 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 2.7 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 2.8, 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 2.1. 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 2.2. 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 2.1). This feature of the RdgP portfolio observed in Figure 2.2 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 2.9. 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 2.10) 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 2.3 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 2.4, 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 2.11, 2.12 and 2.13. 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 2.14 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 2.5.

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 2.15 and Figure 2.5 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 (2.5), 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 2.6. 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.

2.6.2 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. 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 2.16 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 2.17 and 2.18, 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 2.16 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 2.19). 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.

In Figure 2.7, 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 2.8.

2.7 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.

2.8 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 (2.29) is equivalent to the optimization problem in (2.30) for the L2 norm.

Proof of Lemma 4 The first order condition of (2.29) is given as follows

$$\begin{aligned}
FOC2.29 &= \frac{2}{n} R_t'\left(1_n - R_t\theta_t\right) + 2\alpha_t\left(\theta_t - \theta_{t-1}\right) = 0 \\
&= \frac{1}{n} R_t'\left(1_n - R_t\theta_t\right) + \alpha_t\left(\theta_t - \theta_{t-1}\right) = 0.
\end{aligned}$$

The first order condition of (2.30) is

$$\begin{aligned}
FOC2.30 &= \frac{2}{n} R_t'\left(y_t - R_t\tilde{\theta}_t\right) + 2\alpha_t\tilde{\theta}_t = 0 \\
&= \frac{1}{n} R_t'\left\{1_n - R_t\theta_t - R_t\left(\theta_t - \theta_{t-1}\right)\right\} + \alpha_t\left(\theta_t - \theta_{t-1}\right) = 0 \\
&= \frac{1}{n} R_t'\left(1_n - R_t\theta_t\right) + \alpha_t\left(\theta_t - \theta_{t-1}\right) = FOC2.29.
\end{aligned}$$

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 $P\left[\frac{X_n}{Y_n} \leq B_\epsilon\right] \geq 1 - \epsilon \ \forall n \geq N_\epsilon$.*

2.8.1 Proof of Proposition 1

Let's first look at a one period problem. Using the same assumptions as in Section 2.2, the optimal selection problem will be given as follows

$$\max_{\{\omega:\|\omega\|^2\le d\}} E\left(-\exp(-\gamma A_1)\right) = \max_{\{\omega:\|\omega\|^2\le d\}} E\left(V_1\right). \tag{2.33}$$

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, (2.33) becomes as follows

$$\max_{\{\omega:\|\omega\|^2\le d\}} \left\{-\exp\left[-\gamma A_0\left(R_f + \omega'\mu - \frac{A_0\gamma}{2}\omega'\Sigma\omega\right)\right]\right\} \tag{2.34}$$

which is equivalent of solving the following problem

$$\max_{\{\omega:\|\omega\|^2\le d\}} \left\{\gamma A_0\left(R_f + \omega'\mu - \frac{A_0\gamma}{2}\omega'\Sigma\omega\right)\right\} \tag{2.35}$$

or equivalently,

$$\max_{\{\omega\}} \left\{\gamma A_0\left(R_f + \omega'\mu - \frac{A_0\gamma}{2}\omega'\Sigma\omega\right) - \lambda\|\omega\|^2\right\} \tag{2.36}$$

because $\gamma A_0 \ge 0$ by assumption, with $\lambda > 0$ the Lagrange multiplier associated with $\|\omega\|^2 \le d$. Let α be the positive constant solution of $\lambda = \frac{\alpha}{2}\gamma^2 A_0^2$, then (2.36) becomes as follows

$$\max_{\{\omega\}} \left\{\gamma A_0\left(R_f + \omega'\mu - \frac{\alpha}{2}\gamma A_0\|\omega\|^2\right) - \frac{(A_0\gamma)^2}{2}\omega'\Sigma\omega\right\}. \tag{2.37}$$

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

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

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 2.2

$$V(t, A_t) = \max_{\left\{\omega_t : \|\omega_t\|^2 \leq d_t\right\}} E_t \left\{V(t+1, A_{t+1})\right\} = \max_{\left\{\omega_t : \|\omega_t\|^2 \leq d_t\right\}} E_t \left\{V(t+1, A_t(R_f + \omega_t' r_{t+1}))\right\} \quad (2.38)$$

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

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

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 $t = 0$

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

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

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

Therefore,

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

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 $t=1$

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

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

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

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

Solving this optimization problem is equivalent to solving

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

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 $\lambda_0=\frac{\left(\gamma R_f\right)^2 A_0^2\alpha_0}{2}$, we obtain that

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

(2.10) 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,

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

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,

$$
V(T-1, A_{T-1}) = \max_{\left\{\omega_{T-1}:\|\omega_{T-1}\|^2 \leq \delta_{T-1}\right\}} \left\{-\exp\left\{-\gamma\left(A_{T-1}R_f + A_{T-1}\omega_{T-1}^{'}\mu - \frac{\gamma A_{T-1}^2}{2}\omega_{T-1}^{'}\Sigma\omega_{T-1}\right)\right\}\right\} \quad (2.3
$$

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

$$
\max_{\left\{\omega_{T-1}:\|\omega_{T-1}\|^2 \leq \delta_{T-1}\right\}} \left\{A_{T-1}R_f + A_{T-1}\omega_{T-1}^{'}\mu - \frac{\gamma A_{T-1}^2}{2}\omega_{T-1}^{'}\Sigma\omega_{T-1}\right\}. \quad (2.40)
$$

If we denote by λ_{T-1} the Lagrange multiplier associated with $\|\omega_{T-1}\|^2 \leq \delta_{T-1}$, we have that, solving (2.40) with respect to ω_{T-1} is equivalent of solving the following unconstrained problem by assuming that λ_{T-1} is given

$$
\max_{\omega_{T-1}} \left\{A_{T-1}R_f + A_{T-1}\omega_{T-1}^{'}\mu - \frac{\gamma A_{T-1}^2}{2}\omega_{T-1}^{'}\Sigma\omega_{T-1} - \lambda_{T-1}\|\omega_{T-1}\|^2\right\}. \quad (2.41)
$$

Moreover, solving (2.41) with respect to ω_{T-1} is also equivalent of solving the following unconstrained problem with respect to ω_{T-1}

$$
\begin{aligned}
&\max_{\omega_{T-1}} \left\{-\exp\left(-\gamma\left(A_{T-1}R_f + A_{T-1}\omega_{T-1}^{'}\mu - \frac{\gamma A_{T-1}^2}{2}\omega_{T-1}^{'}\Sigma\omega_{T-1} - \frac{\lambda_{T-1}}{\gamma}\|\omega_{T-1}\|^2\right)\right)\right\} \\
&= \max_{\omega_{T-1}} \left\{\exp\left(\lambda_{T-1}\|\omega_{T-1}\|^2\right)\left\{-\exp\left(-\gamma\left(A_{T-1}R_f + A_{T-1}\omega_{T-1}^{'}\mu - \frac{\gamma A_{T-1}^2}{2}\omega_{T-1}^{'}\Sigma\omega_{T-1}\right)\right)\right\}\right\} \\
&= \max_{\omega_{T-1}} \left\{\left\{\exp\left(\lambda_{T-1}\|\omega_{T-1}\|^2\right) E_{T-1}\left\{V(T, A_T)\right\}\right\}\right\}
\end{aligned}
$$

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

Therefore, the solution of

$$V\left(T-1, A_{T-1}\right) = \max_{\left\{\omega_{T-1}:\|\omega_{T-1}\|^2 \leq \delta_{T-1}\right\}} E_{T-1}\left\{V(T, A_T)\right\} \quad (2.42)$$

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 (2.41) 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

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

with

$$\begin{aligned}
V^*\left(T-1, A_{T-1}\right) &= \underbrace{E_{T-1}\left\{-\exp\left(-\gamma A_T\right)\right\}}_{\omega_{T-1}=\omega_{T-1}^*} \\
&= -\exp\left\{-\gamma A_{T-1}R_f - \gamma\left(A_{T-1}\left(\omega_{T-1}^*\right)'\mu - \frac{\gamma A_{T-1}^2}{2}\left(\omega_{T-1}^*\right)'\Sigma\omega_{T-1}^*\right)\right\} \\
&= -\exp\left\{-\gamma A_{T-1}R_f - \gamma\left(\frac{1}{\gamma}\mu'\left(\Sigma + \alpha_{T-1}I_N\right)^{-1}\mu - \frac{1}{\gamma}\mu'\left(\Sigma + \alpha_{T-1}I_N\right)^{-1}\Sigma\left(\Sigma + \alpha_{T-1}I_N\right)^{-1}\mu\right)\right\} \\
&= -\exp\left\{-\gamma A_{T-1}R_f + f_{T-1}\left(\mu, \Sigma, \gamma, R_f, \alpha_{T-1}\right)\right\}
\end{aligned}$$

with

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

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

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.

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

Hence, since

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

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

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

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

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

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

2.8.2 Proof of Proposition 2

Using the same procedure as in the proof of Proposition 1, one can easily show that solving (2.15) is equivalent to solving the following non-constrained problem

$$\max_{\{\omega_t\}} \left\{ \underbrace{\exp\left(\lambda_t \|\omega_t - \omega_{t-1}\|^2\right)}_{B} E_t \left[V\left(t+1, A_t\left(R_f + \omega_t' r_{t+1}\right)\right)\right] \right\} \tag{2.43}$$

for $t = 0, ..., T-1$ with the following terminal condition $V(T, A_T) = -\exp(-\gamma A_T)$. And solving (2.43) at each period from $T-1$ one can easily obtain the following first order condition

$$A_t \gamma \left(\Sigma + \alpha_t I_N\right) \omega_t = \mu + \alpha_t \omega_{t-1} \tag{2.44}$$

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

$$A_0 \gamma \left(\Sigma + \alpha_0 I_N\right) \omega_0 = \mu. \tag{2.45}$$

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

$$\hat{\omega}_0 = \hat{\gamma}_0 \hat{\Sigma}_{\alpha_0}^{-1} \hat{\mu}_0. \tag{2.46}$$

At the $t = 1$ by combining (2.44) and (2.46) 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

$$\hat{\omega}_1 = \hat{\gamma}_1 \hat{\Sigma}_{\alpha_1}^{-1} \left[\hat{\mu}_1 + \alpha_1 \hat{\gamma}_0 \hat{\Sigma}_{\alpha_0}^{-1} \hat{\mu}_0\right]. \tag{2.47}$$

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

$$\hat{\omega}_2 = \hat{\gamma}_2 \hat{\Sigma}_{\alpha_2}^{-1} \left[\hat{\mu}_2 + \alpha_2 \hat{\gamma}_1 \hat{\Sigma}_{\alpha_1}^{-1} \hat{\mu}_1 + \alpha_1 \alpha_2 \hat{\gamma}_0 \hat{\gamma}_1 \hat{\Sigma}_{\alpha_0}^{-1} \hat{\Sigma}_{\alpha_1}^{-1} \hat{\mu}_0\right]. \tag{2.48}$$

Therefore, we have that

$$\hat{\omega}_t = \hat{\gamma}_t \hat{\Sigma}_{\alpha_t}^{-1} \tilde{\mu}_t \tag{2.49}$$

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

$$\hat{\Sigma}_{\alpha_t} = \hat{\Sigma}_t + \alpha_t I_N \tag{2.50}$$

and

$$\tilde{\mu}_t = \hat{\mu}_t + \sum_{j=0}^{t-1}\left(\prod_{i=j}^{t-1}\hat{\gamma}_i\alpha_{i+1}\hat{\Sigma}_{\alpha_i}^{-1}\right)\hat{\mu}_j \tag{2.51}$$

2.8.3 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*

$$\left\|\hat{\theta}_{\alpha_t} - \theta_t\right\| = o_p(1) \tag{2.52}$$

$$\left\|\mu'\left(\hat{\theta}_{\alpha_t} - \theta_t\right)\right\| = o_p(1) \tag{2.53}$$

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

Proof of Lemma 1

$$\left\|\hat{\theta}_{\alpha_t} - \theta_t\right\| = \left\|\hat{\theta}_{\alpha_t} - \theta_{\alpha_t} + \theta_{\alpha_t} - \theta_t\right\| \tag{2.54}$$

where $\theta_{\alpha_t} = \Omega_{\alpha_t}^{-1}\mu_t$. By (2.54), we have that

$$\left\|\hat{\theta}_{\alpha_t} - \theta_t\right\| \leq \underbrace{\left\|\hat{\theta}_{\alpha_t} - \theta_{\alpha_t}\right\|}_{(A)} + \underbrace{\left\|\theta_{\alpha_t} - \theta_t\right\|}_{(B)}. \tag{2.55}$$

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 $\mu_t = \Omega_t\theta_t$

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

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

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

Then,

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

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

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 $\left\|(\Omega_{\alpha_t})^{-1}\Omega_t\theta_t\right\| \leq \|\theta_t\| = O(1)$. Hence, we obtain the following relation

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

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

Hence, we have that

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

where

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

Therefore if $\frac{N}{\alpha_t\sqrt{n}} \to 0$ 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*

$$\left\|\hat{\gamma}_t^{-1} - \gamma_t^{-1}\right\| = o_p(1) \tag{2.56}$$

if $\max_{0\leq j\leq t-1}\left\{\frac{N^{3/2}}{\alpha_j\sqrt{n}} + \sqrt{N}\alpha_j^{min(\frac{\tau_j}{2},1)}\right\} \to 0$ *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

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

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

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

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

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

Hence,

$$\left\|\hat{\gamma}_1^{-1}-\gamma_1^{-1}\right\|=O_p\left(\frac{N^{3/2}}{\alpha_0\sqrt{n}}+\sqrt{N}\alpha_0^{min(\frac{\tau_0}{2},1)}+\frac{N}{\sqrt{n}}\right)$$

Therefore, if $\frac{N^{3/2}}{\alpha_0\sqrt{n}}+\sqrt{N}\alpha_0^{min(\frac{\tau_0}{2},1)}\to 0$, $P(1)$ 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

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

Hence,

$$\left\|\hat{\gamma}_{t+1}^{-1}-\gamma_{t+1}^{-1}\right\|=O_p\left(\frac{N^{3/2}}{\alpha_t\sqrt{n}}+\sqrt{N}\alpha_t^{min(\frac{\tau_t}{2},1)}+\frac{N}{\sqrt{n}}\right)$$

Therefore, if $\frac{N^{3/2}}{\alpha_t\sqrt{n}}+\sqrt{N}\alpha_t^{min(\frac{\tau_t}{2},1)}\to 0$, $P(t+1)$ is true.

The rest of the proof of Proposition 3

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

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

$$A=\underbrace{\frac{\hat{\theta}_t-\theta_t}{\left(1-\hat{\mu}_t'\hat{\theta}_t\right)\left(1-\mu'\theta_t\right)}}_{a}-\underbrace{\frac{\left[\hat{\theta}_t\left(\mu'\theta_t\right)-\theta_t\left(\hat{\mu}_t'\hat{\theta}_t\right)\right]}{\left(1-\hat{\mu}_t'\hat{\theta}_t\right)\left(1-\mu'\theta_t\right)}}_{b}. \tag{2.57}$$

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

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

$$\frac{\hat{\theta}_t - \theta_t}{\left(1-\hat{\mu}_t'\hat{\theta}_t\right)\left(1-\mu'\theta_t\right)} = \frac{\hat{\theta}_t - \theta_t}{\left(1-\mu'\theta_t\right)^2} + O_p\left(\left(\hat{\theta}_t - \theta_t\right)\mu'\left(\hat{\theta}_t - \theta_t\right)\right). \tag{2.58}$$

The second terms in (2.57) can be developed according to Carrasco et al. (2019) as follows

$$\frac{\hat{\theta}_t\left(\mu'\theta_t\right) - \theta_t\left(\hat{\mu}_t'\hat{\theta}_t\right)}{\left(1-\hat{\mu}_t'\hat{\theta}_t\right)\left(1-\mu'\theta_t\right)} = \frac{\left(\hat{\theta}_t - \theta_t\right)\mu'\theta_t - \theta_t\left(\hat{\mu}_t - \mu\right)'\left(\hat{\theta}_t - \theta_t\right) - \theta_t\left(\hat{\mu}_t - \mu\right)'\theta_t - \theta_t\mu'\left(\hat{\theta}_t - \theta_t\right)}{\left(1-\hat{\mu}_t'\hat{\theta}_t\right)\left(1-\mu'\theta_t\right)} \tag{2.59}$$

By (2.58) and because $\left|\mu'\theta_t\right| < 1$, we have that

$$\frac{\left(\hat{\theta}_t - \theta_t\right)\mu'\theta_t}{\left(1-\hat{\mu}_t'\hat{\theta}_t\right)\left(1-\mu'\theta_t\right)} = \frac{\left(\hat{\theta}_t - \theta_t\right)\mu'\theta_t}{\left(1-\mu'\theta_t\right)^2} + O_p\left(\left(\hat{\theta}_t - \theta_t\right)\mu'\left(\hat{\theta}_t - \theta_t\right)\right) \tag{2.60}$$

$$\frac{\theta_t\mu'\left(\hat{\theta}_t - \theta_t\right)}{\left(1-\hat{\mu}_t'\hat{\theta}_t\right)\left(1-\mu'\theta_t\right)} = \frac{\theta_t\mu'\left(\hat{\theta}_t - \theta_t\right)}{\left(1-\mu'\theta_t\right)^2} + O_p\left(\left(\hat{\theta}_t - \theta_t\right)\mu'\left(\hat{\theta}_t - \theta_t\right)\right) \tag{2.61}$$

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

$$\frac{\hat{\theta}_t\left(\mu'\theta_t\right) - \theta_t\left(\hat{\mu}_t'\hat{\theta}_t\right)}{\left(1-\hat{\mu}_t'\hat{\theta}_t\right)\left(1-\mu'\theta_t\right)} = \frac{\left(\hat{\theta}_t - \theta_t\right)\mu'\theta_t}{\left(1-\mu'\theta_t\right)^2} - \frac{\theta_t\mu'\left(\hat{\theta}_t - \theta_t\right)}{\left(1-\mu'\theta_t\right)^2} + O_p\left(\left(\hat{\theta}_t - \theta_t\right)\mu'\left(\hat{\theta}_t - \theta_t\right)\right) \tag{2.62}$$

$$+O_p\left[\left(1 + \left\|\hat{\theta}_t - \theta_t\right\|\right)\sqrt{\frac{N}{n}}\right] \tag{2.63}$$

Hence, by assumption A, we obtain that

$$A = \frac{\hat{\theta}_t - \theta_t}{(1-\mu'\theta_t)} + \frac{\theta_t \mu' \left(\hat{\theta}_t - \theta_t\right)}{(1-\mu'\theta_t)^2} + O_p\left(\left(\hat{\theta}_t - \theta_t\right)\mu'\left(\hat{\theta}_t - \theta_t\right)\right) + o_p(1) \tag{2.64}$$

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

$$\gamma_t A \approx (\hat{\omega}_t - \omega_t) = \frac{\gamma_t\left(\hat{\theta}_t - \theta_t\right)}{(1-\mu'\theta_t)} + \frac{\gamma_t\theta_t \mu' \left(\hat{\theta}_t - \theta_t\right)}{(1-\mu'\theta_t)^2} + o_p(1) \tag{2.65}$$

$$\|\hat{\omega}_t - \omega_t\| \leq \left\|\frac{\gamma_t\left(\hat{\theta}_t - \theta_t\right)}{(1-\mu'\theta_t)}\right\| + \left\|\frac{\gamma_t\theta_t \mu' \left(\hat{\theta}_t - \theta_t\right)}{(1-\mu'\theta_t)^2}\right\| + o_p(1) \tag{2.66}$$

$$\begin{aligned} \left\|\frac{\gamma_t\left(\hat{\theta}_t - \theta_t\right)}{(1-\mu'\theta_t)}\right\| + \left\|\frac{\gamma_t\theta_t \mu' \left(\hat{\theta}_t - \theta_t\right)}{(1-\mu'\theta_t)^2}\right\| &= \frac{\gamma_t}{(1-\mu'\theta_t)}\left\|\hat{\theta}_t - \theta_t\right\| + \frac{\gamma_t}{(1-\mu'\theta_t)^2}\left\|\theta_t\mu'\left(\hat{\theta}_t - \theta_t\right)\right\| & (2.67) \\ &\leq \frac{\gamma_t}{(1-\mu'\theta_t)}\left\|\hat{\theta}_t - \theta_t\right\| + \frac{\gamma_t}{(1-\mu'\theta_t)^2}\|\theta_t\|\left\|\mu'\left(\hat{\theta}_t - \theta_t\right)\right\| & (2.68) \end{aligned}$$

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

$$\|\hat{\omega}_t - \omega_t\| = o_p(1). \tag{2.69}$$

2.8.4 Proof of Proposition 4

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

$$s(\hat{\omega}_{\alpha_t}) = \frac{\mu'\hat{\theta}_t}{\left(\hat{\theta}_t'\Sigma\hat{\theta}_t\right)^{1/2}}. \tag{2.70}$$

(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,

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

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

This implies that

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

Therefore,

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

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

$$\left\| \mu' \left(\hat{\theta}_{\alpha_t} - \theta_t \right) \right\| = O_p \left(\frac{N^{\frac{3}{2}}}{\alpha_t \sqrt{n}} + \sqrt{N} \|\theta_{\alpha_t} - \theta_t\| \right)$$

which implies that

$$\mu' \hat{\theta}_{\alpha_t} = \mu' \theta_t + O_p \left[\sqrt{N} \left(\frac{N}{\alpha_t \sqrt{n}} + \|\theta_{\alpha_t} - \theta_t\| \right) \right]. \tag{2.71}$$

(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,

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

Hence,

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

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

Hence, by assumption A we obtain that

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

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

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

Therefore,

$$\hat{\theta}_t'\Sigma\hat{\theta}_t = \theta_t'\Sigma\theta_t + O_p\left[N\left(\frac{N}{\alpha_t\sqrt{n}} + \|\theta_{\alpha_t} - \theta_t\|\right)\right]. \tag{2.72}$$

Combining (2.71) and (2.72) we obtain that,

$$s\left(\hat{\omega}_{\alpha_t}\right)^2 = \frac{\left(\mu'\hat{\theta}_t\right)^2}{\hat{\theta}_t'\Sigma\hat{\theta}_t} = \frac{\left(\mu'\theta_t\right)^2}{\theta_t'\Sigma\theta_t} + O_p\left[\left(\frac{N}{\alpha_t\sqrt{n}} + \|\theta_{\alpha_t} - \theta_t\|\right)\right] \tag{2.73}$$

$$= s\left(\omega_t\right)^2 + O_p\left[\left(\frac{N}{\alpha_t\sqrt{n}} + \|\theta_{\alpha_t} - \theta_t\|\right)\right] \tag{2.74}$$

2.8.5 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

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

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

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

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

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

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

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

$$\begin{aligned}\left\|\hat{\Sigma}_t\left(\hat{\omega}_{\alpha_t}-\omega_{\alpha_t}\right)\right\|_2 &= O_p\left(\gamma_t\frac{N}{\sqrt{n}}.\sqrt{\frac{N}{n}}.\frac{1}{\alpha_t}+\gamma_t.\frac{N}{\sqrt{n}}.\frac{1}{\alpha_t}.\sqrt{N}+\gamma_t.\sqrt{\frac{N}{n}}\right)\\ &= O_p\left(\frac{N^{3/2}}{\alpha_t\sqrt{n}}\right).\end{aligned}$$

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

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

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

2.8.6 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 $\kappa_t > 1$ $\left\|\theta_{S_t^c}\right\|_1 \leq \kappa_t \|\theta_{S_t} - \theta_{t-1}\|_1$.

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

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

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,

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

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

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

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

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

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

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,

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

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

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.

$$\frac{\left\| R_t\left(\hat{\theta}_{\alpha_t} - \theta_t\right)\right\|^2}{n} + \alpha_t \left\|\hat{\theta}_{\alpha_t} - \theta_{t-1}\right\|_1 \leq \frac{2u_t' R_t\left(\hat{\theta}_{\alpha_t} - \theta_t\right)}{n} + \alpha_t \|\theta_t - \theta_{t-1}\|_1 \qquad (2.75)$$

We have that

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

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

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

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

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

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,

$$\frac{\left\| 1_n - R_t\hat{\theta}_{\alpha_t} \right\|^2}{n} + \alpha_t \left\| \hat{\theta}_{\alpha_t} - \theta_{t-1} \right\|_1 \leq \frac{\left\| 1_n - R_t\theta_t \right\|^2}{n} + \alpha_t \left\| \theta_t - \theta_{t-1} \right\|_1 \tag{2.76}$$

which is always true because we have that

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

Therefore, (2.75) and (2.76) are equivalent. Then, using (2.75), we have that

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

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

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

And using B(i), we obtain that

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

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

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

by triangular inequality. Hence,

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

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,

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

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

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

Therefore,

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

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

2.8.7 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*

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

if $\max_{0\leq j\leq t-1}\left\{\frac{N^{3/2}}{\alpha_j\sqrt{n}}+\sqrt{N}\alpha_j^{min(\frac{\tau_j}{2},1)}\right\}\rightarrow 0$ *as* $n\rightarrow\infty$

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

Proof of proposition 6.1

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

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

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 $\|\delta\|=O\left(1\right)$

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

Because,

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

we have that,

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

Using Proof of Lemma 1, we have the following result

$$\begin{aligned}\left\|\mu'\left(\hat{\theta}_{\alpha_t} - \theta_{\alpha_t}\right)\right\| &= O_p\left(\frac{N^{3/2}}{\alpha_t n} + \sqrt{\frac{N}{n}} + \frac{N^2}{\alpha_t n} + \frac{N}{\sqrt{n}}\right)\\ &= O_p\left(\frac{N^2}{\alpha_t n} + \frac{N}{\sqrt{n}}\right).\end{aligned}$$

Hence, using those two relations we obtain that

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

Using Lemma 3 we obtain that

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

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*

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

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,

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

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*

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

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

Proof of proposition 6.2

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

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

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

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

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

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

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

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

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

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

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 $\delta^{'}\hat{\omega}_{\alpha_t}$.

2.8.8 Proof of Proposition 7

We start by the fact that

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

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

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

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

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

where

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

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

with

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

Hence,

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

Therefore,

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

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

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

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

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

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

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

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

Hence,

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

Under the assumption that $\frac{N}{\alpha_t\sqrt{n}} \to 0$, (k1) becomes as follows

$$(k1) = O_p\left[\frac{N^{5/2}}{\alpha_t n^{3/2}}+\frac{N^{3/2}}{n}+\frac{N}{\sqrt{n}}\alpha_t^{min(\frac{\tau_t}{2},1)}\right].$$

Therefore,

$$\begin{aligned}rest\left(\alpha_t\right) &= O_p\left[\frac{N^{5/2}}{\alpha_t n^{3/2}}+\frac{N^{3/2}}{n}+\frac{N}{\sqrt{n}}\alpha_t^{min(\frac{\tau_t}{2},1)}+\frac{N^2}{\alpha_t n}+\frac{N}{\sqrt{n}}+\sqrt{N}\alpha_t^{min(\frac{\tau_t}{2},1)}\right]\\ &= O_p\left[\frac{N^{5/2}}{\alpha_t n^{3/2}}+\frac{N^2}{\alpha_t n}+\frac{N}{\sqrt{n}}+\frac{N^{3/2}}{n}+\left(\sqrt{N}+\frac{N}{\sqrt{n}}\right)\alpha_t^{min(\frac{\tau_t}{2},1)}\right]\end{aligned}$$

2.9 Tables and Figures

Table 2.2: 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 2.3: 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 2.4: 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 2.5: 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 (29.1304)	622.9837 (76.1754)	1629.7 (202.0158)	3346.2 (461.6792)	15842 (3019.2)	27065 (6250.2)	72148 (21181)
2	220.8378 (29.3274)	616.1889 (73.7110)	1629.6 (199.8609)	3376.7 (459.1104)	15545 (2896.0)	25885 (5797.1)	72336 (21344)
5	221.7853 (29.3883)	627.2623 (75.0828)	1621.9 (200.7195)	3382.3 (466.2505)	15233 (2844.9)	26193 (5791.3)	69306 (19456)
7	222.2615 (29.3190)	641.9389 (75.9786)	1621.8 (202.5438)	3260.8 (457.5931)	15043 (2836.5)	26714 (5720.8)	68237 (19483)
9	220.6481 (29.0634)	632.5309 (74.0857)	1628.0 (200.7087)	3271.2 (461.0739)	14895 (2773.8)	26274 (5662.3)	66849 (19114)
11	218.7989 (29.1129)	626.0016 (73.3732)	1602.1 (193.5361)	3266.8 (461.2739)	15451 (2840.9)	26382 (5790.8)	66520 (18701)

Table 2.6: 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	$\left(\hat{\lambda}_{\max}/\hat{\lambda}_{\min}\right) / \left(\lambda_{\max}/\lambda_{\min}\right)$						
	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 2.7: 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 2.8: 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 2.9: 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 2.10: 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	0.3532	0.6689	0.7919	0.1067	0.9140	0.1317	0.1249
L2TSP	0.8539	0.556	0.5350	0.1192	0.940	0.1845	0.1352
L1TSP	0.9648	0.5237	0.604	0.0774	0.0943	0.0668	0.0874
BP	0.4161	3.8001	6.1319	6.4807	7.0534	9.0807	9.3808

Table 2.11: 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 2.12: 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 2.13: 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 2.14: 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 2.15: 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 2.16: 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 2.17: 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 2.18: 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 2.19: 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

Figure 2.1: 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 $n = 120$.

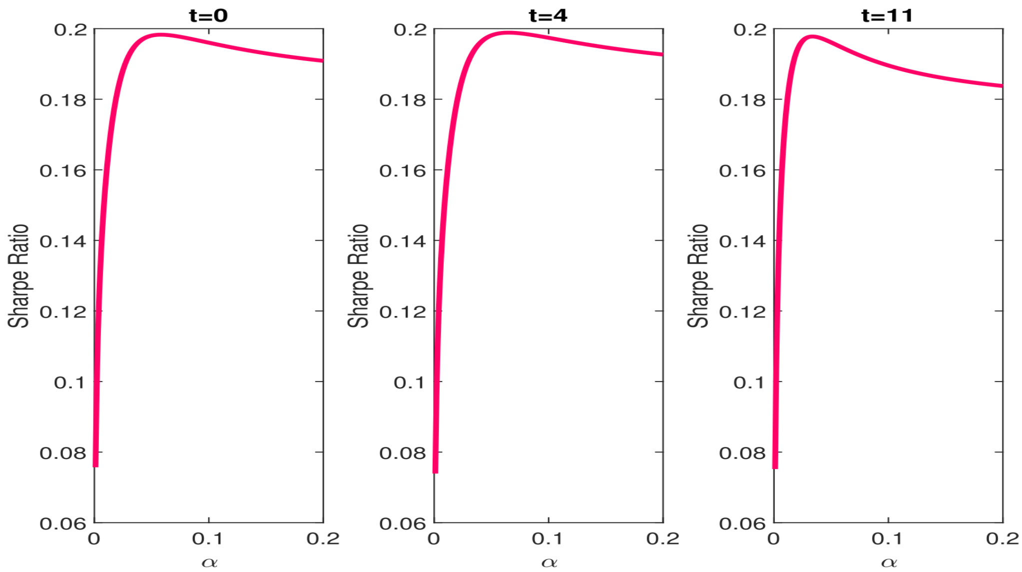


Figure 2.2: The GCV criterion as a function of the tuning parameter for the Ridge regularization using a single sample when $N = 60$ and $n = 120$.

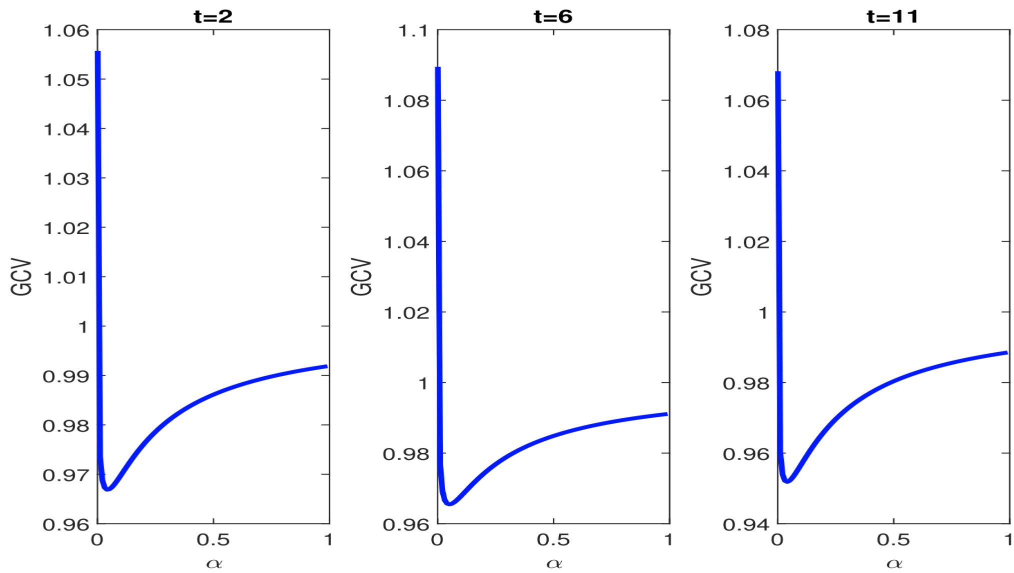


Figure 2.3: The transaction cost as a function of the tuning parameter for the L1TSP and L2TSP. We obtain this figure for $N = 20$, $T = 12$ with an estimation window of $n = 120$.

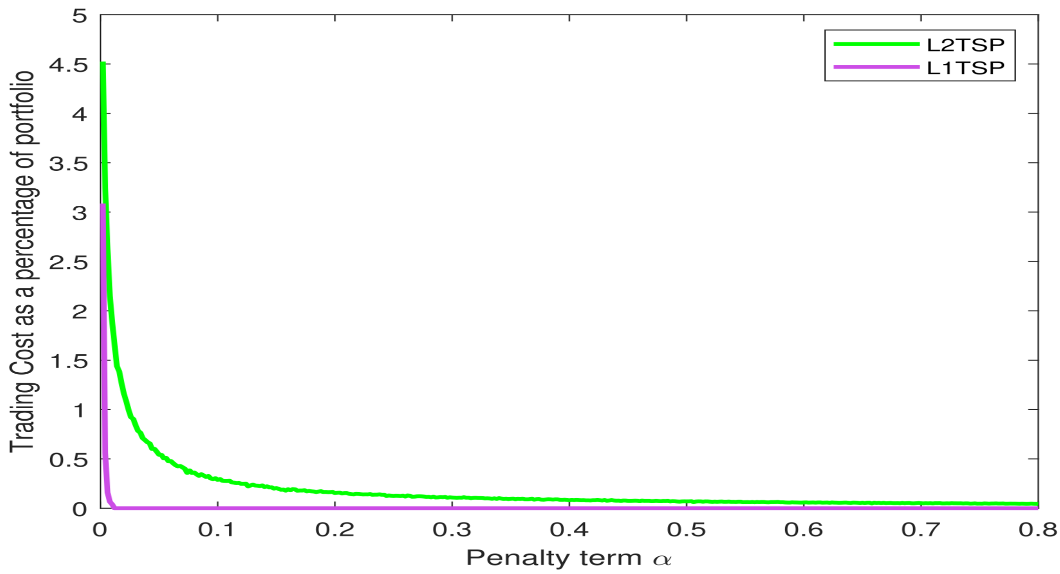


Figure 2.4: The average stability rate as a function of the tuning parameter for the L1TSP and L2TSP. We obtain this figure for $N = 20$, $T = 12$ with an estimation window of $n = 120$.

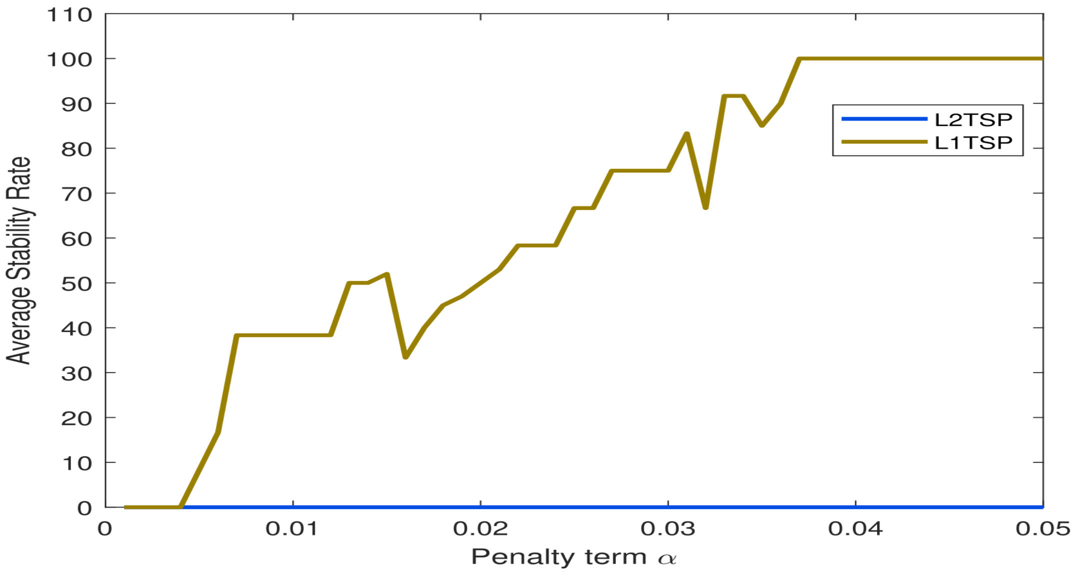


Figure 2.5: The Average Optimal selected tuning parameter for the RdgP, the L2TSP and the L1TSP over the life cycle when $N = 60$ and $n = 120$.

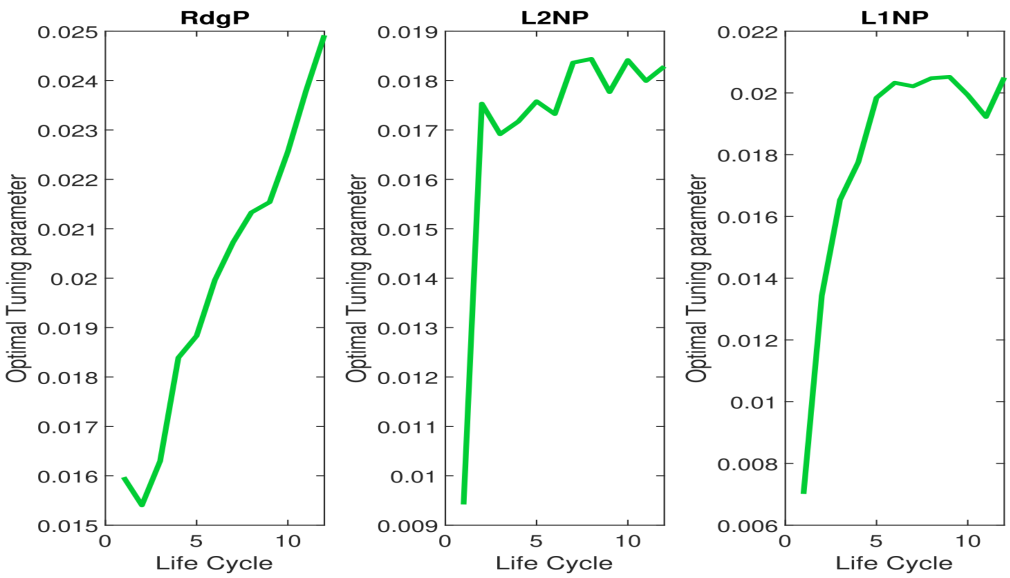


Figure 2.6: The Mean Squared Error of the selected strategy over the life cycle for the RdgP and the L2TSP with $N = 20$, $T = 24$ month and an estimation window of $n = 120$.

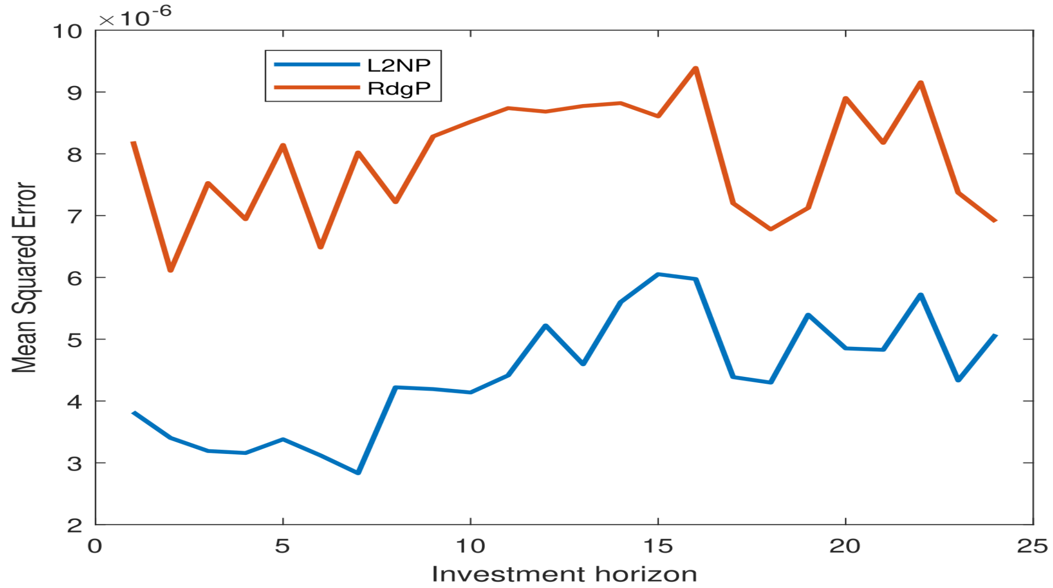


Figure 2.7: The optimal wealth over the life cycle for our procedures. We obtain this figure using the 30 industry portfolios with an estimation window of $n = 120$.

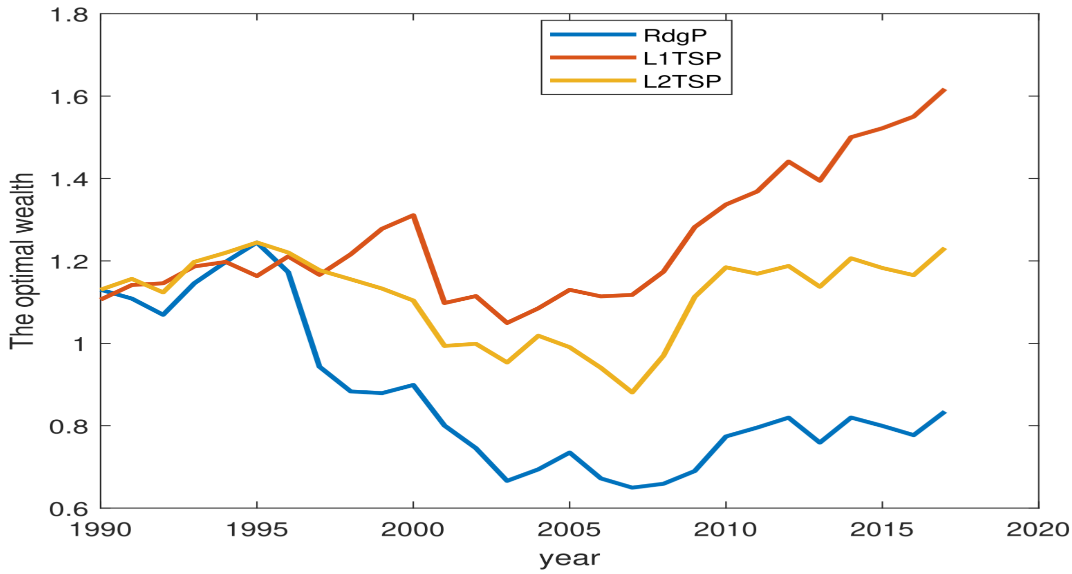
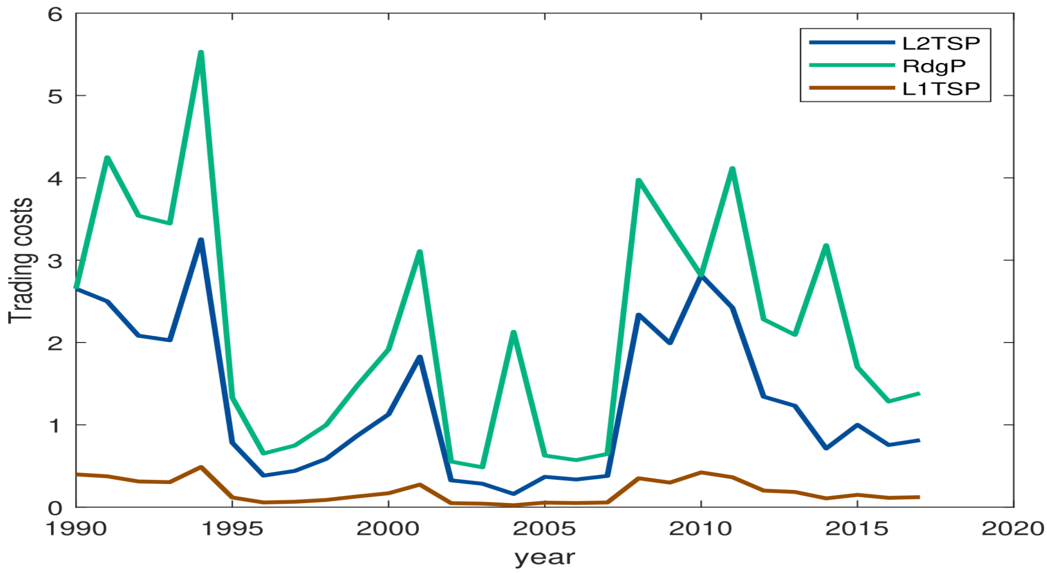


Figure 2.8: The re-balancing cost over the life cycle. We obtain this figure using the 30 industry portfolios with an estimation window of $n = 120$.



Chapter 3

Regularized Maximum Diversification Investment Strategy*

3.1 Introduction

Since the seminal work of Markowitz (1952) which offers an essential basis to portfolio selection, diversification issues have been in the center of many problems in the financial market. According to Markowitz's portfolio theory, a portfolio is diversified if its variance could not be reduced any further at the same level of the expected return.The fundamental objective of this diversification is to construct a portfolio with various assets that earns the highest return for the least volatility which may be a good alternative to the market cap-weighted portfolios. In fact, there is evidence that market portfolios are not as efficient as assumed by Sharpe (1964) in the Capital Asset Price Model (CAPM). The CAPM model as introduced by Sharpe (1964) implies that the tangency portfolio is the only efficient one and should produce the greatest returns relative to risk. Nonetheless, several empirical studies have shown that investing in the minimum variance portfolio yields better out-of-sample results than does an investment in the tangency portfolio (for instance see Haugen and Baker (1991), Choueifaty et al. (2013), Lohre et al. (2014)).

Even if these surprising results seem to be due to the high estimation risk associated with the expected returns (according to Kempf and Memmel (2006)), the efficiency of the market capitalization weighted index has been questioned motivating numerous investment alternatives (see Arnott et al. (2005)), Clarke et al. (2006), Maillard et al. (2010)). Subsequently, Choueifaty (2011) introduced the concept of maximum diversification, via a formal definition of portfolio diversification: the diversification ratio (DR) and claimed that portfolios with maximal DRs were maximally diversified and provided an efficient alternative to market cap-weighted portfolios.

This optimal maximum diversification portfolio is shown to be a function of the in-

*I am greatly indebted to Marine Carrasco for her invaluable guidance. I am grateful to Georges Dionne for his helpful comments.

verse of the covariance matrix of asset returns (see Theron and Van Vuuren (2018)), which is unknown and need to be estimated. So, solving for the maximum diversification portfolio leads to estimate the covariance matrix of returns and take its inverse. This results in estimation error, amplified by the fact that the number of securities is typically very high in the selected diversified portfolio, and these security returns are highly correlated in general. The resulting estimation errors may affect negatively the performance of the maximum diversification selected portfolio. Therefore, Choueifaty et al. (2013) propose the most diversified portfolio (MDP) by imposing a non-negative constraint on the maximum diversification problem[1]. However, this ad hoc constraint suggests that the MDP is unlikely to represent the final word of diversification. Without the ability to short securities it may be impossible to unlock the full range of uncorrelated risk sources present in the market (see Maguire et al. (2014)). In this paper we propose a more general method to control for estimation error in the covariance matrix of asset returns without restricting the ability to short sell in the financial market. This method is fundamentally based on different ways to stabilize the inverse of the covariance matrix in the selected portfolio and is particularly useful when the number of assets in the financial market increases considerably compared with the estimation window. More precisely, as in Carrasco (2012) and Carrasco and Tchuente (2015) we investigate three regularization techniques such as the spectral cut-off, the Tikhonov and the Landweber Fridman to stabilize the inverse of the covariance matrix. This procedure has been used by Carrasco et al. (2019) to stabilize the inverse of the covariance matrix in the mean-variance portfolio.

These regularization schemes involve a tuning parameter which needs to be chosen efficiently. So, we propose a data-driven method for selecting the tuning parameter in an optimal way i.e. in order to minimize the distance between the inverse of the estimated covariance matrix and the inverse of the true covariance matrix.

We show, under appropriate regularity conditions, that the selected strategy by regularization is asymptotically efficient with respect to the diversification ratio for a wide choice of the tuning parameter. Meaning that, even if the optimal diversified portfolio is unknown, there exists a feasible portfolio obtained by regularization capable of reaching similar level of performance in terms of the diversification ratio.

To evaluate the performance of our procedures we implement a simulation exercise based on a three-factor model calibrated on real data from the US financial market. We obtain by simulation that our procedure significantly improve the performance of the proposed strategy with respect to the Sharpe ratio. Moreover, the regularized rules are compared to several strategies such as the most diversified portfolio, the target portfolio, the global minimum variance portfolio, and the naive 1/N strategy in terms of in-sample and out-of-sample Sharpe ratio. To confirm our simulations, we do an empirical analysis using Kenneth R. French's 30-industry portfolios and 100 portfolios formed on size and

[1]The objective is to reduce the effect of estimation error on the performance of selected maximum diversification portfolio

book-to-market.

The rest of the paper is organized as follows. Section 3.2 presents the economy. The regularized portfolio is presented in section 3.3. Section 3.4 gives some asymptotic properties of the selected strategy and proposes data-driven methods to select the optimal tuning parameter. Section 3.5 presents some simulation results and an empirical study. Section 3.6 concludes the paper.

3.2 The model

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

We assume that the excess returns $r_{t+1} = R_{t+1} - R_f 1_N$ are independent and identically distributed with the mean and the covariance matrix given by μ and $\Sigma = \{\sigma_{i,j}\}_{i,j \in N}$ respectively. Let $\omega = (\omega_1, ..., \omega_N)^{'}$ be the vector of portfolio weights that represents the amount of the capital to be invested in the risky assets and the remain $1 - \omega^{'} 1_N$ is allocated to the risk-free asset. Short-selling is allowed in the financial market, i.e. some of the weights ω_i could be negative. Let $\sigma = (\sigma_{1,1}, ..., \sigma_{N,N})^{'}$ be the vector of asset volatilities.

According to Choueifaty (2011), the diversification ratio (DR) of any portfolio ω is given by

$$DR(\omega) = \frac{\omega^{'}\sigma}{\sqrt{\omega^{'}\Sigma\omega}} \tag{3.1}$$

which is the ratio of weighted average of volatilities divided by the portfolio volatility.

Using the relation in Equation (3.1), the maximum diversification portfolio is obtained by solving the following optimization problem

$$\max_{\omega} DR(\omega). \tag{3.2}$$

Since the DR is invariant by scalar multiplication (for instance see Choueifaty et al. (2013)), solving the problem in Equation (3.2) is equivalent of solving this new problem according to Theron and Van Vuuren (2018)

$$\min_{\omega^{'}\sigma=1} \frac{1}{2}\omega^{'}\Sigma\omega. \tag{3.3}$$

This new optimization problem is very close to the global minimum variance portfolio. The only difference is that the constraint $\omega' 1 = 1$ in the global minimum variance problem is replaced by $\omega' \sigma = 1$. The optimal solution of this new optimization problem is given by

$$\omega = \frac{\Sigma^{-1}\sigma}{\sigma'\Sigma^{-1}\sigma}. \tag{3.4}$$

The optimal solution in (3.4) is unknown because it depends on the covariance matrix of asset returns and the vector of volatilities which are unknown and need to be estimated from available data set. We need in particular to estimate the covariance of matrix and take its inverse. The sample covariance may not be appropriate because it may be nearly singular, and sometimes not even invertible. The issue of ill-conditioned covariance matrix must be addressed because inverting such matrix increases dramatically the estimation error and then makes the maximum diversification portfolio unreliable. Many techniques have been proposed in the literature to stabilize the inverse of the covariance matrix in the optimal solution in (3.4). According to Carrasco et al. (2007) an interesting way to stabilize the inverse of the covariance matrix consists of dampening the explosive effect of the inversion of the singular values of $\hat{\Sigma}$. It consists in replacing the sequence $\{1/\lambda_j\}$ of explosive inverse singular values by a sequence $\{q(\alpha, \lambda_j)/\lambda_j\}$ 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.

Here, we implement a regularization approach to estimate the optimal solution in (3.4) using three regularization schemes based on three different ways of inverting the covariance matrix of asset returns. These regularization techniques are the spectral cut-off, the Tikhonov and the Landweber Fridman. The spectral cut-off regularization scheme is based on principal components whereas the Tikhonov's one is based on Ridge regression (also called Bayesian shrinkage) and the last one is an iterative method.

3.3 The regularized portfolio

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

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

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

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

3.3.1 Tikhonov regularization (TH)

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

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

3.3.2 The spectral cut-off (SC)

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

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

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

3.3.3 Landweber Fridman regularization (LF)

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

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

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

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

The basic idea behind this procedure is similar to spectral cut-off but with a smooth bias function.

See Carrasco et al. (2007) for more details on these regularization techniques. The regularized diversified portfolio for a given regularization scheme is

$$\hat{\omega}_\alpha = \frac{\hat{\Sigma}^\alpha \hat{\sigma}}{\hat{\sigma}'\hat{\Sigma}^\alpha \hat{\sigma}} = \left(\hat{\sigma}'\hat{\Sigma}^\alpha \hat{\sigma}\right)^{-1} \hat{\Sigma}^\alpha \hat{\sigma}. \tag{3.5}$$

This regularized portfolio depends on an unknown tuning parameter which needs to be selected in an optimal way.

3.4 Asymptotic properties of the selected portfolio

In this section we will look at the efficiency of the regularized portfolio with respect to the diversification ratio. We will also propose a data driven method to select the tuning parameter.

3.4.1 Efficiency of the regularized diversified portfolio

To obtain the efficiency of the selected portfolio, we need to impose some regularity conditions, in particular we will need the following assumption to show the efficiency.

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

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

Under assumption A, the following proposition presents information about the asymptotic property of the diversification ratio associated with the selected portfolio.

Proposition 1. *Under assumption A we have that*

$$DR(\hat{\omega}_\alpha) \rightarrow_p DR(\omega_t), \tag{3.6}$$

if $\frac{N}{\alpha\sqrt{T}} \rightarrow 0$ *as* T *goes to infinity.*

Proof. In the appendix.

Comment on proposition 1. The regularity condition behind Proposition 1 implies several things. First, $\alpha\sqrt{T} \rightarrow +\infty$ implies that the estimation window should go to infinity faster than the optimal tuning parameter goes to zero. Second, $\frac{N}{\alpha\sqrt{T}} \rightarrow 0$ implies that $\alpha\sqrt{T}$ should go to infinity faster than the number of assets in the financial market. Hence, the number of assets should be limited asymptotically compared with the estimation window. The regularity condition $\frac{N}{\sqrt{T}} \rightarrow 0$ seems to be more restrictive than assuming that $\frac{N}{T} \rightarrow Constant$. One way to avoid this regularity condition will be to assume that the covariance matrix of assets distribution is a trace class operator. Proposition 1 shows that the regularized diversified portfolio is asymptotically efficient in terms of the diversification ratio for a wide choice of the tuning parameter. Meaning that, even if the optimal diversified portfolio is unknown, there exists a feasible portfolio obtained by regularization capable of reaching similar level of performance in terms of the diversification ratio.

3.4.2 Data-driven Method for Selecting the Tuning Parameter

We show in the previous sections that the selected portfolio depends on a certain smoothing parameter $\alpha \in (0, 1)$. We have derived the efficiency 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, we propose a data-driven selection procedure to obtain an approximation of this parameter.

Our objective here is to select the tuning parameter which minimizes the distance between the inverse of the estimated covariance matrix and the inverse of the true covariance matrix. According to Ledoit and Wolf (2003), most of the existing shrinkage estimators from finite-sample statistical decision theory as well as in Frost and Savarino (1986) break down when $N \geq T$ because their loss functions involve the inverse of the sample covariance matrix which is a singular matrix in this situation. Therefore, to avoid this problem, they propose a loss function that does not depend on this inverse.This loss function is a quadratic measure of distance between the true and the estimated covariance matrices based on the Frobenius norm. Unlike in Ledoit and Wolf (2003), we will use a loss function which depends on the inverse of the covariance matrix under the assumption that the true covariance matrix is invertible. One important thing to notice here is that the regularized covariance matrix is always invertible even if $N \geq T$ meaning that our loss function exists even for $N \geq T$. In fact, we know that the optimal diversified portfolio as given by Equation (3.4) depends on the inverse of the covariance matrix of assets distribution. And because our objective is to stabilize the inverse of this covariance matrix in the estimated portfolio by regularization, we propose here to use a loss function that minimizes a quadratic distance between the regularized inverse and the theoretical covariance matrix.

The loss function we consider here is given by

$$\mu' \left[\left(\hat{\Sigma}^{\alpha} - \Sigma^{-1} \right)' \Sigma \left(\hat{\Sigma}^{\alpha} - \Sigma^{-1} \right) \right] \mu \tag{3.7}$$

where μ is the expected excess return. The choice of this specific quadratic distance is useful to obtain a criterion that can easily be approximated by generalized cross validation approach.

Hence, the objective is to select the tuning parameter which minimizes

$$E \left\{ \mu' \left[\left(\hat{\Sigma}^{\alpha} - \Sigma^{-1} \right)' \Sigma \left(\hat{\Sigma}^{\alpha} - \Sigma^{-1} \right) \right] \mu \right\} \tag{3.8}$$

which implies that

$$\hat{\alpha} = arg \min_{\alpha \in H_T} E \left\{ \mu' \left[\left(\hat{\Sigma}^{\alpha} - \Sigma^{-1} \right)' \Sigma \left(\hat{\Sigma}^{\alpha} - \Sigma^{-1} \right) \right] \mu \right\} \tag{3.9}$$

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

We denote by $\Omega = E \left(r_t r_t' \right) = E \left(X' X \right) / T$ and $\beta = \Omega^{-1} \mu = E \left(X' X \right)^{-1} E \left(X' 1_T \right)$ where r_t, $t = 1, \cdots, T$ are the observations of the excess returns and X the $T \times N$ matrix with tth row given by r_t'.

Assumption B

For some $\nu > 0$, we have that

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

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

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

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

β_α is the regularized version of β.

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 2. *Under assumptions A and B we have that*

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

if $\frac{1}{\alpha^2 T} \to 0$ *and* $\sqrt{N}\alpha^{\min(\frac{\nu}{2},1)} \to 0$ *as* T *goes to infinity.*

I will only show the first part of this proposition. The second part comes from proposition 1 in Carrasco et al. (2019).

Proof. In the appendix.

From proposition 2, it follows that minimizing $E\left\{\mu' \left[\left(\hat{\Sigma}^\alpha - \Sigma^{-1}\right)' \Sigma \left(\hat{\Sigma}^\alpha - \Sigma^{-1}\right)\right]\mu\right\}$ is equivalent to minimizing

$$\frac{1}{T} E\left\|X\left(\hat{\beta}_\alpha - \beta\right)\right\|^2 \tag{3.10}$$

$$+\frac{\left(\mu'\left(\beta_\alpha - \beta\right)\right)^2}{\left(1 - \mu'\beta\right)}. \tag{3.11}$$

Terms (3.10) and (3.11) depend on the unknown β and hence need to be approximated. To approximate (3.10), we use results on cross-validation from Craven and Wahba (1978), Li (1986, 1987), and Andrews (1991) among others.

The rescaled MSE

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

can be approximated by generalized cross validation criterion:

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

Using the fact that

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

(3.11) can be estimated by plug-in:

$$\frac{\left(1'_T (M_T(\alpha) - I_T) X\hat{\beta}_{\tilde{\alpha}}\right)^2}{T^2\left(1 - \hat{\mu}'\hat{\beta}_{\tilde{\alpha}}\right)} \tag{3.12}$$

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

The optimal value of τ is defined as

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

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

3.5 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 most diversified portfolio proposed by Choueifaty and Coignard (2008). More precisely, in this section, we focus our attention on how our procedure performs in terms of the Sharpe ratio and the diversification ratio. To end this section, we analyze the out-of-sample performance of the selected portfolio from each procedure we have proposed.

3.5.1 Data

In our simulations and empirical analysis, various forms of monthly data will be used from July 1980 to June 2016. The one-month Treasury-Bill (T-Bill) rate is used as a proxy for the risk-free rate and R_f is calibrated to be the mean of the one-month Treasury-Bill rate observed in the data. We use monthly returns of Fama-French three factors and of 30 industry portfolios from the Kenneth R. French data library in order to calibrate unknown parameters of the simulation model. In the empirical study, we also use monthly data for the 100 portfolios formed on size and book-to-market from the Kenneth R. French data Library.

3.5.2 Simulation

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

$$r_{it} = b_{i1}f_{1t} + b_{i2}f_{2t} + b_{i3}f_{3t} + \epsilon_{it} \tag{3.13}$$

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

In the compact form (3.13) can be written as follows:

$$R = BF + \epsilon \tag{3.14}$$

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

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

The mean of the excess return is given by $\mu = B\mu_f$. The return on the risk-free asset R_f is calibrated to be the mean of the one-month T-B observed in the data from July 1980 to June 2016. The calibrated parameters used in our simulation process are given in Table 3.1. The gross return on the risk-free asset calibrated on the data is given by $R_f = 1.0036$. Once generated, the factor loadings are kept fixed over replications, while the factors differ from simulations and are drawn from a trivariate normal distribution.

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

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

To set up 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

Table 3.1: 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

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 consider the following portfolio selection procedures: the sample-based diversified portfolio (SbDP), the most diversified portfolio (MDP) proposed by Choueifaty et al. (2013), the global minimum variance portfolio (GMVP), the ridge regularized diversified portfolio (RdgDP), the spectral cut regularized diversified portfolio (SCDP), the Landweber-Fridman regularized diversified portfolio (LFDP), the equal-weighted portfolio which is also called the naive portfolio (XoNP) which allocates a constant amount 1/N in each asset, and the target (or the maximum Sharpe ratio) portfolio (TgP). We perform 1000 simulations and estimate our statistics over replications. We obtain the following result about the actual Sharpe ratio.

Table 3.2: 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 $n = 120$, over 1000 replications. True SR is the true actual Sharpe ratio.

Strategies	Number of risky assets						
	10	20	40	60	80	90	100
SbDP	0.1549	0.0906	0.0889	0.0779	0.0652	0.0719	0.0704
XoNP	0.2604	0.2604	0.2415	0.2525	0.2406	0.2461	0.2467
GMVP	0.2227	0.2338	0.2098	0.2298	0.1710	0.1640	0.1449
MDP	0.2514	0.2545	0.2410	0.2544	0.1778	0.1821	0.1935
TgP	0.2608	0.2818	0.2662	0.2687	0.2026	0.1925	0.1699
RdgDP	0.2587	0.2785	0.2817	0.2907	0.2947	0.2830	0.2991
SCDP	0.2592	0.2872	0.2993	0.2898	0.2746	0.2887	0.2853
LFDP	0.2605	0.2765	0.2840	0.2870	0.2850	0.2912	0.2980
True SR	0.2626	0.2922	0.3393	0.3379	0.3592	0.3477	0.3657

Table 3.2 contains the results about the average monthly Sharpe ratio obtained in the

simulation process. We can notice that the sample based diversified portfolio performs very poorly in terms of maximizing the Sharpe ratio in the investment process for large number of assets in the financial market. This result is essentially due to the fact that the estimation error from estimating the vector of assets volatilities is amplified by using the sample covariance matrix of assets distribution which is close to a singular matrix when N becomes too large compared with the sample size. Hence, even if this strategy is supposed to be the maximum diversification's one with the highest Sharpe ratio, we can notice that the SbDP is dominated by several other strategies such as the GMVP, the XoNP, and the TgP. Therefore, this strategy cannot be consider as the maximum diversification strategy in practice. To solve this problem, Choueifaty et al. (2013) proposes the most diversified portfolio (MDP) which is obtained by maximum the diversification ratio under a non-negative constraint on the portfolio weights. This additional constraint in the investment process may help to reduce the effect of estimation error on the performance of the selected portfolio. The result of the most diversified portfolio can be found in Table 3.2. By imposing the non-negative constraint, investors considerably improve the performance of the selected portfolio in terms of the Sharpe ratio. This new strategy even out-performs the global minimum variance portfolio. However, this procedure is still dominated by the target portfolio and the equal weighted portfolio meaning that much remains to be done about finding the maximum diversification strategy in practice. One explanation to this result is that imposing the non-negative constraint on the portfolio weight may limit the ability of the selected portfolio to be fully diversified. Hence, one needs to find a more general estimation procedure for the maximum diversified portfolio that allows for short selling.

For this purpose, I propose a new way to estimate the optimal diversified portfolio by stabilizing the inverse of the sample covariance matrix without imposing a non-negative constraint on the portfolio weights in the investment process. Three different regularization methods are considered in this paper based on three different ways to compute the inverse of the covariance matrix that appears in the optimal selected portfolio. The results of these methods can also be found in Table 3.2. The first thing to point out about these results is that the regularized diversified portfolio out-performs the most diversified portfolio in terms of maximizing the Sharpe ratio. For instance, we obtain an average Sharpe ratio of 0.2514, 0.2587, 0.2592, and 0.2605 for the MDP, the RdgDP, the SCDP, and the LFDP respectively when only 10 assets are considered in the economy. The difference in terms of the actual Sharpe ratio performance between our procedure and the most diversified portfolio significantly increases with the number of assets in the financial market. For example, for 100 assets, the average Sharpe ratio is about 0.1935, 0.2991, 0.2853, and 0.2980 for the MDP, the RdgDP, the SCDP, and the LFDP respectively. These results may be due to the fact that when the number of assets in the economy increases, the degree of diversification of the selected strategy may deteriorate with non-negative constraints on the investment process which may reduce the ability to find a strategy that performs the Sharpe ratio. Moreover, the regularized diversified portfolio out-performs the target strategy and the equal-weighted portfolio when the number of assets in the financial market exceeds 40. Nonetheless, for 10 assets in the economy, the target portfolio outperforms the RdgDP and the SCDP but is dominated by the LFDP. With 20 assets the target portfolio dominates the RdgDP and the LFDP and is dominated by the SCDP. The equal-weighted portfolio out-performs some regularized strategies such as the RdgDP and the SCDP only for 10 assets in the financial market. The fact that the regularized strategies give very interesting results in terms of maximizing the Sharpe ratio (compared with the existing strategies) for large N is because these methods are essentially used to address estimation issues in large dimensional problems.

3.5.3 Empirical study

In this subsection, we investigate the performance of our procedures empirically. We apply our method to several sets of portfolios from Kenneth R. French's website. In particular, we apply our procedure to the following portfolios: the 30-industry portfolios and the 100 portfolios formed on size and book-to-market. We allow investors to re-balance their portfolios every month. This implies that the optimal portfolio is constructed at the end of each month for a given estimation window M by maximizing the diversification ratio. The investor holds this optimal portfolio for one month, realizes gains and losses, updates information, and then recomputes optimal portfolio weights for the next period using the same estimation window. This procedure is repeated each month, generating a time series of out-of-sample returns. This time series can then be used to analyze the out-of-sample performance of each strategy based on several statistics such as the out-of-sample Sharpe ratio. For this purpose, we use data from July 1980 to June 2018.

Table 3.3 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. The empirical results in this table confirm what we have obtained in the simulation part. According to this result, by stabilizing the inverse of the covariance matrix in the maximum diversification portfolio, we considerably improve the performance of the selected strategy in terms of maximizing the Sharpe ratio. Therefore, our regularized strategies outperform the most diversified strategy, the target portfolio, and the global minimum variance portfolio for each data set. The most-diversified strategy outperforms the global minimum variance portfolio but is dominated by the Equal-Weight portfolio for each data set. These results of the most-diversified portfolio can essentially be explained by the fact that by imposing a non-negative constraint in the investment process, one cannot fully diversify the optimal portfolio.

Tables 3.4 and 3.5 contain the Fama-French monthly regression coefficients for the 100 portfolios formed on size and book-to-market and the 30-industry portfolios respectively. Monthly data are used from July 1990 to June 2018. According to Table 3.4 only the return on the Equal-Weight portfolio can be explained by the Fama-French three-factor model for the 100 portfolios formed on size and book-to-market. The return obtained with the other strategies such as the regularized portfolios and the most diversified portfolio can be explained only with the return on the market portfolio (a one-factor model) through a positive relation. However, the return of the most diversified portfolio and the global minimum variance portfolio can be explained with a two factors model when the optimal strategy is obtained using the 30-industry portfolios. The return of the other strategies such as the regularized portfolios, the Equal-Weight portfolio, and the target portfolio can be explained by the Fama-French three-factor model.

3.6 Conclusion

This paper addresses the estimation issue that exists in the maximum diversification portfolio framework in the large financial market. We propose to stabilize the inverse of the covariance matrix in the optimal diversified portfolio using regularization techniques from inverse problem literature. These regularization techniques namely the ridge, the spectral cut-off, and Landweber-Fridman involve a regularization parameter or penalty term whose optimal value is selected to minimize the expected distance between the inverse of the estimated covariance matrix and the inverse of the true covariance matrix. We show, under appropriate regularity conditions, that the selected strategy by regularization is asymptotically efficient with respect to the diversification ratio for a wide choice of the tuning parameter. Meaning that, even if the optimal diversified portfolio is unknown, there exists a feasible portfolio obtained by regularization capable of reaching a similar level of performance in terms of the diversification ratio.

To evaluate the performance of our procedures we implement a simulation exercise based on a three-factor model calibrated on real data from the US financial market. We obtain by simulation that our procedure significantly improves the performance of the selected strategy with respect to the Sharpe ratio. Moreover, the regularized rules are compared to several strategies such as the most diversified portfolio, the target portfolio, the global minimum variance portfolio, and the naive 1/N strategy in terms of in-sample and out-of-sample Sharpe ratio. To confirm our simulations, we do an empirical analysis using Kenneth R. French's 30-industry portfolios and 100 portfolios formed on size and book-to-market. According to this empirical result, by stabilizing the inverse of the covariance matrix in the maximum diversification portfolio, we considerably improve the performance of the selected strategy in terms of maximizing the Sharpe ratio.

3.7 Proofs

3.7.1 Proof of Proposition 1

By definition we have that

$$DR(\hat{\omega}_\alpha) = \frac{\hat{\omega}_\alpha^{'}\sigma}{\sqrt{\hat{\omega}_\alpha^{'}\Sigma\hat{\omega}_\alpha}}.$$

Let us first look at $\hat{\omega}_\alpha^{'}\Sigma\hat{\omega}_\alpha$

$$\begin{aligned}
\hat{\omega}_\alpha^{'}\Sigma\hat{\omega}_\alpha &= \left[(\hat{\omega}_\alpha - \omega) + \omega\right]^{'} \Sigma \left[(\hat{\omega}_\alpha - \omega) + \omega\right] \\
&= \omega^{'}\Sigma\omega + \underbrace{(\hat{\omega}_\alpha - \omega)^{'} \Sigma (\hat{\omega}_\alpha - \omega)}_{(a)} + 2\underbrace{(\hat{\omega}_\alpha - \omega)^{'} \Sigma\omega}_{(b)}.
\end{aligned}$$

Now we are going to look at the properties of (a) and (b). We know that

$$\hat{\omega}_\alpha = \left(\underbrace{\hat{\sigma}^{'}\hat{\Sigma}^\alpha\hat{\sigma}}_{(c)}\right)^{-1} \underbrace{\hat{\Sigma}^\alpha\hat{\sigma}}_{(d)}.$$

$$\begin{aligned}
(c) &= \sigma^{'}\hat{\Sigma}^\alpha\sigma + (\hat{\sigma} - \sigma)^{'} \hat{\Sigma}^\alpha (\hat{\sigma} - \sigma) + 2(\hat{\sigma} - \sigma)^{'} \hat{\Sigma}^\alpha\sigma \\
\hat{\Sigma}^\alpha &= \left(\hat{\Sigma}^\alpha - \Sigma^\alpha + \Sigma^\alpha\right).
\end{aligned}$$

$$\begin{aligned}
\left\|(\hat{\sigma} - \sigma)^{'} \hat{\Sigma}^\alpha (\hat{\sigma} - \sigma)\right\| &= \left\|\frac{(\hat{\sigma} - \sigma)^{'}}{\sqrt{N}} \left(\frac{\hat{\Sigma}}{N}\right)^\alpha \frac{(\hat{\sigma} - \sigma)}{\sqrt{N}}\right\| \\
&= O_p\left(\frac{\|\hat{\sigma} - \sigma\|^2}{N\alpha}\right) \\
&= O_p\left(\frac{\left\|\frac{\hat{\sigma}-\sigma}{\sqrt{N}}\right\|^2}{\alpha}\right).
\end{aligned}$$

By assumption A $\left\|\frac{\sigma}{\sqrt{N}}\right\| = O(1)$. Hence, we obtain that

$$\begin{aligned}
\left\|(\hat{\sigma} - \sigma)^{'} \hat{\Sigma}^\alpha\sigma\right\| &= \left\|\frac{(\hat{\sigma} - \sigma)^{'}}{\sqrt{N}} \left(\frac{\hat{\Sigma}}{N}\right)^\alpha \frac{\sigma}{\sqrt{N}}\right\| \\
&= O_p\left(\frac{\|\hat{\sigma} - \sigma\|}{\sqrt{N}\alpha}\right) \\
&= O_p\left(\frac{\left\|\frac{\hat{\sigma}-\sigma}{\sqrt{N}}\right\|}{\alpha}\right).
\end{aligned}$$

Using those information combine with the fact that $\hat{\Sigma}^{\alpha} = \hat{\Sigma}^{\alpha} - \Sigma^{\alpha} + \Sigma^{\alpha}$, we have that

$$(c) = \sigma^{'}\Sigma^{\alpha}\sigma + \sigma^{'}\left(\hat{\Sigma}^{\alpha} - \Sigma^{\alpha}\right)\sigma + O_p\left(\frac{\left\|\frac{\hat{\sigma}-\sigma}{\sqrt{N}}\right\| + \left\|\frac{\hat{\sigma}-\sigma}{\sqrt{N}}\right\|^2}{\alpha}\right).$$

$$\begin{aligned}
\left\|\sigma^{'}\left(\hat{\Sigma}^{\alpha} - \Sigma^{\alpha}\right)\sigma\right\| &= \left\|\frac{\sigma^{'}}{\sqrt{N}}\left[\left(\frac{\hat{\Sigma}}{N}\right)^{\alpha} - \left(\frac{\Sigma}{N}\right)^{\alpha}\right]\frac{\sigma}{\sqrt{N}}\right\| \\
&\leq \left\|\frac{\sigma}{\sqrt{N}}\right\|^2 \left\|\left(\frac{\hat{\Sigma}}{N}\right)^{\alpha} - \left(\frac{\Sigma}{N}\right)^{\alpha}\right\| \\
&= O_p\left(\left\|\left(\frac{\hat{\Sigma}}{N}\right)^{\alpha} - \left(\frac{\Sigma}{N}\right)^{\alpha}\right\|\right).
\end{aligned}$$

$$\left\|\left(\frac{\hat{\Sigma}}{N}\right)^{\alpha} - \left(\frac{\Sigma}{N}\right)^{\alpha}\right\| \leq \left\|\left(\frac{\Sigma}{N}\right)^{\alpha}\right\| \left\|\left(\frac{\hat{\Sigma}}{N}\right)^{\alpha}\right\| \left\|\frac{\hat{\Sigma}}{N} - \frac{\Sigma}{N}\right\|.$$

Hence,

$$\left\|\left(\frac{\hat{\Sigma}}{N}\right)^{\alpha} - \left(\frac{\Sigma}{N}\right)^{\alpha}\right\| = O_p\left(\frac{\left\|\frac{\hat{\Sigma}}{N} - \frac{\Sigma}{N}\right\|}{\alpha}\right)$$

which implies that

$$(c) = \sigma^{'}\Sigma^{\alpha}\sigma + O_p\left(\frac{\left\|\frac{\hat{\Sigma}}{N} - \frac{\Sigma}{N}\right\| + \left\|\frac{\hat{\sigma}-\sigma}{\sqrt{N}}\right\| + \left\|\frac{\hat{\sigma}-\sigma}{\sqrt{N}}\right\|^2}{\alpha}\right).$$

As $T \to \infty$ we have that $\alpha \to 0 \Rightarrow$

$$(c) = \sigma^{'}\Sigma^{-1}\sigma + O_p\left(\frac{\left\|\frac{\hat{\Sigma}}{N} - \frac{\Sigma}{N}\right\| + \left\|\frac{\hat{\sigma}-\sigma}{\sqrt{N}}\right\| + \left\|\frac{\hat{\sigma}-\sigma}{\sqrt{N}}\right\|^2}{\alpha}\right).$$

Using the Assumption A combined with Theorem 4 of Carrasco and Florens (2000), we have that

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

Moreover, since $\left\|\frac{\hat{\sigma}-\sigma}{\sqrt{N}}\right\|^2 = O_p\left(\frac{1}{T}\right)$ by assumption A, we have that

$$(c) = \sigma^{'}\Sigma^{-1}\sigma + O_p\left(\frac{1}{\alpha\sqrt{T}}\right).$$

$$\begin{aligned}
(d) &= \hat{\Sigma}^{\alpha}\hat{\sigma} \\
&= \hat{\Sigma}^{\alpha}\sigma + \hat{\Sigma}^{\alpha}\left(\hat{\sigma}-\sigma\right) \\
&= \Sigma^{\alpha}\sigma + \left(\hat{\Sigma}^{\alpha} - \Sigma^{\alpha}\right)\sigma + \hat{\Sigma}^{\alpha}\left(\hat{\sigma}-\sigma\right).
\end{aligned}$$

Since $\alpha \to 0$ as $T \to \infty$, we have that

$$(d) = \Sigma^{-1}\sigma + \left(\hat{\Sigma}^{\alpha} - \Sigma\right)\sigma + \hat{\Sigma}^{\alpha}\left(\hat{\sigma}-\sigma\right).$$

We know that

$$\begin{aligned}
\left\|\hat{\Sigma}^{\alpha}\left(\hat{\sigma}-\sigma\right)\right\| &= \left\|\left(\frac{\hat{\Sigma}}{N}\right)^{\alpha}\frac{\left(\hat{\sigma}-\sigma\right)}{N}\right\| \\
&\leq \left\|\left(\frac{\hat{\Sigma}}{N}\right)^{\alpha}\right\|\left\|\frac{\left(\hat{\sigma}-\sigma\right)}{N}\right\| \\
&= O_p\left(\frac{1}{\alpha\sqrt{TN}}\right).
\end{aligned}$$

Using the fact that,

$$\begin{aligned}
\left\|\left(\hat{\Sigma}^{\alpha} - \Sigma\right)\sigma\right\| &= \left\|\left\{\left(\frac{\hat{\Sigma}}{N}\right)^{\alpha} - \left(\frac{\Sigma}{N}\right)^{\alpha}\right\}\frac{\sigma}{N}\right\| \\
&\leq \left\|\left(\frac{\hat{\Sigma}}{N}\right)^{\alpha} - \left(\frac{\Sigma}{N}\right)^{\alpha}\right\|\left\|\frac{\sigma}{N}\right\| \\
&= O_p\left(\frac{\left\|\frac{\hat{\Sigma}}{N} - \frac{\Sigma}{N}\right\|}{\alpha\sqrt{N}}\right) \\
&= O_p\left(\frac{1}{\alpha\sqrt{TN}}\right)
\end{aligned}$$

we obtain that

$$(d) = \Sigma^{-1}\sigma + O_p\left(\frac{1}{\alpha\sqrt{TN}}\right).$$

Under the assumption that $\frac{1}{\alpha\sqrt{T}} \to 0$, we have that

$$\hat{\omega}_\alpha = \omega + o_p(1). \tag{3.15}$$

By assumption A we have that $\|\Sigma\| = O(N)$. Therefore, using (3.15), we obtain that

$$\hat{\omega}_\alpha^{'} \Sigma \hat{\omega}_\alpha = \omega^{'} \Sigma \omega + o_p(1) \tag{3.16}$$

if $\frac{N}{\alpha\sqrt{T}} \to 0$. Therefore,

$$DR\left(\hat{\omega}_\alpha\right) \to_p DR\left(\omega_t\right).$$

3.7.2 Proof of Proposition 2

$$(A) = \mu^{'} \left[\left(\hat{\Sigma}^\alpha - \Sigma^{-1} \right)^{'} \Sigma \left(\hat{\Sigma}^\alpha - \Sigma^{-1} \right) \right] \mu$$

We also know that $\mu = \hat{\mu} + (\mu - \hat{\mu})$, so

$$\begin{aligned}
(A) &= \mu^{'} \left[\left(\hat{\Sigma}^\alpha - \Sigma^{-1} \right)^{'} \Sigma \left(\hat{\Sigma}^\alpha - \Sigma^{-1} \right) \right] \mu \\
&= \left[\hat{\Sigma}^\alpha (\mu - \hat{\mu}) + \left(\hat{\Sigma}^\alpha \hat{\mu} - \Sigma^{-1} \mu \right) \right]^{'} \Sigma \left[\hat{\Sigma}^\alpha (\mu - \hat{\mu}) + \left(\hat{\Sigma}^\alpha \hat{\mu} - \Sigma^{-1} \mu \right) \right] \\
&= \left(\hat{\Sigma}^\alpha \hat{\mu} - \Sigma^{-1} \mu \right)^{'} \Sigma \left(\hat{\Sigma}^\alpha \hat{\mu} - \Sigma^{-1} \mu \right) + \left[\hat{\Sigma}^\alpha (\mu - \hat{\mu}) \right]^{'} \Sigma \left[\hat{\Sigma}^\alpha (\mu - \hat{\mu}) \right] \\
&+ 2 \left[\hat{\Sigma}^\alpha (\mu - \hat{\mu}) \right]^{'} \Sigma \left(\hat{\Sigma}^\alpha \hat{\mu} - \Sigma^{-1} \mu \right)
\end{aligned}$$

Let denote by $x = \Sigma^{-1}\mu$ and $\hat{x} = \hat{\Sigma}^\alpha \hat{\mu}$, hence,

$$(A) = (\hat{x} - x)^{'} \Sigma (\hat{x} - x) + \left[\hat{\Sigma}^\alpha (\mu - \hat{\mu}) \right]^{'} \Sigma \left[\hat{\Sigma}^\alpha (\mu - \hat{\mu}) \right] + 2 \left[\hat{\Sigma}^\alpha (\mu - \hat{\mu}) \right]^{'} \Sigma (\hat{x} - x)$$

Since, $\|\mu - \hat{\mu}\|^2 = O_p\left(\frac{N}{T}\right)$, $\left\| \left(\frac{\hat{\Sigma}}{N} \right)^\alpha \right\|^2 = O_p\left(\frac{1}{\alpha^2}\right)$, we have that

$$\begin{aligned}
\left\| \left[\hat{\Sigma}^\alpha (\mu - \hat{\mu}) \right]^{'} \right\| &= \left\| \left[\left(\frac{\hat{\Sigma}}{N} \right)^\alpha \frac{(\mu - \hat{\mu})}{N} \right]^{'} \right\| \\
&\leq \left\| \left(\frac{\hat{\Sigma}}{N} \right)^\alpha \right\| \left\| \frac{(\mu - \hat{\mu})}{N} \right\| \\
&= O_p\left(\frac{1}{\alpha\sqrt{TN}} \right)
\end{aligned}$$

$$\begin{aligned}\left[\hat{\Sigma}^{\alpha}\left(\mu-\hat{\mu}\right)\right]^{'}\Sigma\left[\hat{\Sigma}^{\alpha}\left(\mu-\hat{\mu}\right)\right] &= O_p\left(\left\|\left[\hat{\Sigma}^{\alpha}\left(\mu-\hat{\mu}\right)\right]^{'}\right\|^2\|\Sigma\|\right)\\ &= O_p\left(\frac{\|\Sigma\|}{\alpha^2 TN}\right)\end{aligned}$$

Using the fact that $\|\Sigma\| = O(N)$ by assumption A, we obtain that

$$\left[\hat{\Sigma}^{\alpha}\left(\mu-\hat{\mu}\right)\right]^{'}\Sigma\left[\hat{\Sigma}^{\alpha}\left(\mu-\hat{\mu}\right)\right] = O_p\left(\frac{N}{\alpha^2 TN}\right) = O_p\left(\frac{1}{\alpha^2 T}\right)$$

$$\begin{aligned}\hat{x}-x &= \hat{\Sigma}^{\alpha}\hat{\mu}-\Sigma^{-1}\mu\\ \hat{\mu} &= (\hat{\mu}-\mu)+\mu \Rightarrow\\ \hat{x}-x &= \hat{\Sigma}^{\alpha}\left(\hat{\mu}-\mu\right)+\left(\hat{\Sigma}^{\alpha}-\Sigma^{-1}\right)\mu \Rightarrow\end{aligned}$$

$$\left[\hat{\Sigma}^{\alpha}\left(\mu-\hat{\mu}\right)\right]^{'}\Sigma\left(\hat{x}-x\right) = \left[\hat{\Sigma}^{\alpha}\left(\mu-\hat{\mu}\right)\right]^{'}\Sigma\left[\hat{\Sigma}^{\alpha}\left(\mu-\hat{\mu}\right)\right] + \left[\hat{\Sigma}^{\alpha}\left(\mu-\hat{\mu}\right)\right]^{'}\Sigma\left(\hat{\Sigma}^{\alpha}-\Sigma^{-1}\right)\mu$$

$$\begin{aligned}\left(\hat{\Sigma}^{\alpha}-\Sigma^{-1}\right)\mu &= \left(\frac{\hat{\Sigma}}{N}\right)^{\alpha}\left[\frac{\Sigma}{N}-\frac{\hat{\Sigma}}{N}\right]\left(\frac{\Sigma}{N}\right)^{-1}\frac{\mu}{N}\\ &= O_p\left(\frac{1}{\alpha\sqrt{TN}}\right)\end{aligned}$$

which implies that

$$\begin{aligned}(A) &= \left(\hat{x}-x\right)^{'}\Sigma\left(\hat{x}-x\right)+O_p\left(\frac{2}{\alpha^2 T}\right)\\ &= \left(\hat{x}-x\right)^{'}\Sigma\left(\hat{x}-x\right)+O_p\left(\frac{1}{\alpha^2 T}\right)\end{aligned}$$

Therefore, we obtain that

$$\begin{aligned}&E\left\{\mu^{'}\left[\left(\hat{\Sigma}^{\alpha}-\Sigma^{-1}\right)^{'}\Sigma\left(\hat{\Sigma}^{\alpha}-\Sigma^{-1}\right)\right]\mu\right\}\\ \sim\ &E\left\{\left(\hat{x}-x\right)^{'}\Sigma\left(\hat{x}-x\right)\right\}\end{aligned}$$

if $\frac{1}{\alpha^2 T} \to 0$.

The proof of the second part of proposition 2 comes from proposition 1 in Carrasco et al. (2019).

3.8 Tables

Table 3.3: Out-of-sample performance in terms of the Sharpe ratio applied on the 30 industry portfolios (FF30) and the 100 portfolios formed on size and book-to-market (FF100) with a rolling window of 120.

Strategies		XoNP	GMVP	MDP	TGP	RdgP	LFP	SCP
FF30	Excess return	0.0110	0.01134	0.0121	0.017	0.0149	0.014	0.014
	Volatility	0.0540	0.0630	0.058	0.076	0.063	0.057	0.061
	Sharpe ratio	0.204	0.180	0.209	0.224	0.237	0.246	0.2295
FF100	Excess return	0.0103	0.0127	0.015	0.0173	0.0200	0.0201	0.0203
	Volatility	0.0485	0.075	0.088	0.091	0.0772	0.0770	0.078
	Sharpe ratio	0.212	0.1693	0.1705	0.1901	0.2590	0.2610	0.2602

Table 3.4: Fama-French Monthly Regression Coefficients for the 100 portfolios formed on size and book-to-market from July 1990 to June 2018.

Strategies	Market	HML	SMB	Intercept
Rdg-regularized Portfolio	0.9168	0.079	-0.139	0.0075
	(0.000)	(0.531)	(0.302)	(0.057)
LF- regularized Portfolio	0.823	0.174	-0.1651	0.0125
	(0.000)	(0.153)	(0.204)	(0.001)
SC-regularized Portfolio	1.02	-0.127	-0.133	0.0077
	(0.000)	(0.177)	(0.189)	(0.010)
Most-Diversified Portfolio	0.72	0.13	0.098	0.007
	(0.000)	(0.344)	(0.506)	(0.002)
Equal-Weight-Portfolio	1.002	0.5104	0.33	0.0001
	(0.000)	(0.000)	(0.000)	(0.815)
Global-Minimum-Variance Portfolio	0.416	-0.125	0.155	0.0094
	(0.000)	(0.319)	(0.247)	(0.000)
Target-Portfolio	0.43	0.144	0,207	0.010
	(0.000)	(0.367)	(0.226)	(0.000)

Table 3.5: Fama-French Monthly Regression Coefficients for the 30-industry portfolios from July 1990 to June 2018.

Strategies	Market	HML	SMB	Intercept
Rdg-regularized Portfolio	1.03	0.24	0.36	0.0007
	(0.000)	(0.003)	(0.000)	(0.767)
LF- regularized Portfolio	0.93	0.22	0.25	0.0046
	(0.000)	(0.003)	(0.001)	(0.042)
SC-regularized Portfolio	0.86	0.27	0.21	0.0054
	(0.000)	(0.000)	(0.031)	(0.053)
Most-Diversified Portfolio	0.46	-0.285	0.070	0.002
	(0.000)	(0.000)	(0.391)	(0.001)
Equal-Weight-Portfolio	0.983	0.061	0.265	0.0013
	(0.000)	(0.006)	(0.000)	(0.050)
Global-Minimum-Variance Portfolio	0.46	-0.146	0.077	0.0021
	(0.000)	(0.008)	(0.188)	(0.017)
Target-Portfolio	0.54	-0.44	-0.21	0.013
	(0.000)	(0.000)	(0.019)	(0.000)

Conclusion Générale

Dans cette thèse divisée en trois chapitres nous proposons divers outils économétriques pour améliorer le processus de sélection de portefeuilles sur le marché financier afin d'aider les intervenants de ce marché.

Dans le premier chapitre nous analysons un problème d'optimisation de portefeuille dynamique d'un investisseur à préférences récursives faisant face à des coûts de transactions sur le marché boursier. Plus précisément, nous développons dans ce chapitre une procédure de test simple basée sur une estimation de type GMM pour évaluer l'effet des coûts de transaction dans le processus d'investissement sans une forme particulière présumée pour ces frictions dans l'économie. Nous montrons que la distribution asymptotique de la statistique de test ne dépend pas d'une forme particulière des coûts de transactions dans le modèle de choix de portefeuille. Une procédure de test implémentée en deux étapes a été proposée pour évaluer la sur-identification lorsque le paramètre d'intérêt est au bord de l'espace des paramètres. Empiriquement, nous appliquons nos procédures de test à la classe d'anomalies considérées par Novy-Marx and Velikov (2016). On obtient que les coûts de transaction affectent significativement le comportement d'investissement pour la plupart des anomalies. Par conséquent, les investisseurs améliorent considérablement les performances hors échantillon en tenant compte de ces coûts de transaction dans la prise de décision sur le marché financier.

Dans le deuxième chapitre, nous analysons un problème dynamique de choix de portefeuille de grande taille en développant une nouvelle méthode économétrique pour estimer la solution optimale. Premièrement, nous pénalisons la norme des poids attribués aux actifs dans le portefeuille optimal et nous obtenons une forme analytique qui pourrait être obtenue par une régularisation de type ridge, qui consiste à ajouter une matrice diagonale à la matrice de covariance. Cependant, cette méthode contrôle partiellement l'erreur d'estimation dans la solution optimale car elle ignore l'erreur d'estimation du rendement moyen des actifs, qui peut également être importante lorsque le nombre d'actifs sur le marché financier augmente considérablement. Nous proposons une méthode alternative qui consiste à pénaliser la norme de la différence de pondérations successives du portefeuille dans le problème dynamique pour garantir que la composition optimale du portefeuille ne fluctue pas énormément entre les périodes. Nous montrons, sous des conditions de régularité appropriées, que nous maîtrisons mieux l'erreur d'estimation dans le portefeuille optimal avec cette nouvelle procédure. Pour évaluer la performance de nos procédures, nous faisons une simulation à l'aide d'un modèle à trois facteurs calibrés sur les données réelles du marché financier américain. Les simulations sont confirmées par d'un cas empirique utilisant 30 et 100 portefeuilles d'industries américaine.

Dans le troisième chapitre, nous utilisons diverses techniques de régularisation (ou stabilisation) empruntées à la littérature sur les problèmes inverses pour estimer le portefeuille diversifié tel que défini par Choueifaty (2011). Ici, nous appliquons les trois techniques de régularisation, qui sont les plus utilisées : le ridge qui consiste à ajouter une matrice diagonale à la matrice de covariance, la coupure spectrale qui consiste à exclure les vecteurs propres associés aux plus petites valeurs propres, et Landweber Fridman qui est une méthode itérative, pour stabiliser l'inverse de matrice de covariance dans le processus d'estimation du portefeuille diversifié. Les solutions obtenues sont comparées à plusieurs stratégies telles que le portefeuille le plus diversifié, le portefeuille cible, le portefeuille de variance minimale et la stratégie naïve 1 / N à l'aide du ratio de Sharpe dans l'échantillon et hors échantillon.

Bibliography

Acharya, Viral V and Lasse Heje Pedersen, "Asset pricing with liquidity risk," *Journal of Financial Economics*, 2005, *77* (2), 375–410.

Amihud, Yakov, "Illiquidity and stock returns: cross-section and time-series effects," *Journal of Financial Markets*, 2002, *5* (1), 31–56.

Andrews, Donald WK, "cross-validation in regression with heteroskedastic errors," *Journal of Econometrics*, 1991, *47*, 359–377.

———, "Estimation When a Parameter Is on a Boundary: Theory and Application," Technical Report, Yale University, Cowles Foundation Discussion paper No.1153 1997.

———, "Estimation When a Parameter Is on a Boundary," *Econometrica*, 1999, *67* (6), 1341–1383.

Arnott, Robert D, Jason Hsu, and Philip Moore, "Fundamental indexation," *Financial Analysts Journal*, 2005, *61* (2), 83–99.

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.

Brunnermeier, Markus K, Thomas M Eisenbach, and Yuliy Sannikov, "Macroeconomics with financial frictions: A survey," Technical Report, National Bureau of Economic Research 2012.

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

Buss, Adrian, Raman Uppal, and Grigory Vilkov, "Asset prices in general equilibrium with transactions costs and recursive utility," Technical Report, Working paper, Goethe University Frankfurt 2011.

Campani, Carlos Heitor, René Garcia, and Abraham Lioui, "Approximate Analytical Solutions for Consumption/Investment Problems under Recursive Utility and Finite Horizon." PhD dissertation, EDHEC Business School 2015.

Campbell, John Y and Luis M Viceira, *Strategic asset allocation: portfolio choice for long-term investors*, Oxford University Press, USA, 2002.

———, **George Chacko, Jorge Rodriguez, and Luis M Viceira**, "Strategic asset allocation in a continuous-time VAR model," *Journal of Economic Dynamics and Control*, 2004, *28* (11), 2195–2214.

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 Jean-Pierre Florens**, "Generalization of GMM to a continuum of moment conditions," *Econometric Theory*, 2000, *16* (6), 797–834.

——— **and Stéphane Gregoir**, "Policy evaluation in macroeconometric doubly stochastic models," *Annales d'Économie et de Statistique*, 2002, pp. 73–109.

———**, 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.

Choueifaty, Yves, "Methods and systems for providing an anti-benchmark portfolio," June 7 2011. US Patent 7,958,038.

——— **and Yves Coignard**, "Toward maximum diversification," *The Journal of Portfolio Management*, 2008, *35* (1), 40–51.

———**, Tristan Froidure, and Julien Reynier**, "Properties of the most diversified portfolio," *Journal of investment strategies*, 2013, *2* (2), 49–70.

Clarke, Roger G, Harindra De Silva, and Steven Thorley, "Minimum-variance portfolios in the US equity market," *The Journal of Portfolio Management*, 2006, *33* (1), 10–24.

Craven, Peter and Grace Wahba, "Smoothing noisy data with spline functions," *Numerische mathematik*, 1978, *31* (4), 377–403.

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.

———, ———, **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.

Dufour, Jean-Marie and Jan F Kiviet, "Exact tests for structural change in first-order dynamic models," *Journal of Econometrics*, 1996, *70* (1), 39–68.

Dumas, Bernard and Elisa Luciano, "An exact solution to a dynamic portfolio choice problem under transactions costs," *The Journal of Finance*, 1991, *46* (2), 577–595.

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.

Epstein, Larry G and Stanley E Zin, "Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework," *Econometrica*, 1989, pp. 937–969.

——— **and** ———, "Substitution, risk aversion, and the temporal behavior of consumption and asset returns: An empirical analysis," *Journal of Political Economy*, 1991, pp. 263–286.

Fama, Eugene F and Kenneth R French, "Dividend yields and expected stock returns," *Journal of Financial Economics*, 1988, *22* (1), 3–25.

Frost, Peter A and James E Savarino, "An empirical Bayes approach to efficient portfolio selection," *Journal of Financial and Quantitative Analysis*, 1986, *21* (3), 293–305.

Garlappi, Lorenzo and Georgios Skoulakis, "Solving consumption and portfolio choice problems: The state variable decomposition method," *Review of Financial Studies*, 2010, *23* (9), 3346–3400.

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

Guggenberger, Patrik and Gitanjali Kumar, "On the size distortion of tests after an overidentifying restrictions pretest," *Journal of Applied Econometrics*, 2012, *27* (7), 1138–1160.

Hansen, Lars Peter, "Large Sample Properties of Generalised Method of Moments Estimators," *Econometrica*, 1982, *50*, 1029–1054.

Haugen, Robert A and Nardin L Baker, "The efficient market inefficiency of capitalization–weighted stock portfolios," *The Journal of Portfolio Management*, 1991, *17* (3), 35–40.

Hayashi, Fumio, *Econometrics*, NJ: Princeton University Press, 2000.

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.

Kempf, Alexander and Christoph Memmel, "Estimating the global minimum variance portfolio," *Schmalenbach Business Review*, 2006, *58* (4), 332–348.

Ketz, Philipp, "Testing Overidentifying Restrictions When the True Parameter Vector Is near or at the Boundary of the Parameter Space," Technical Report, Paris School of Economics 2017.

Kress, Rainer, "Singular integral equations," in "Linear Integral Equations," Springer, 1999, pp. 94–124.

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.

Lesmond, David A, Michael J Schill, and Chunsheng Zhou, "The illusory nature of momentum profits," *Journal of Financial Economics*, 2004, *71* (2), 349–380.

Lewellen, Jonathan, "Predicting returns with financial ratios," *Journal of Financial Economics*, 2004, *74* (2), 209–235.

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.

Liu, Hong, "Optimal consumption and investment with transaction costs and multiple risky assets," *The Journal of Finance*, 2004, *59* (1), 289–338.

——— **and Mark Loewenstein**, "Optimal portfolio selection with transaction costs and finite horizons," *Review of Financial Studies*, 2002, *15* (3), 805–835.

Lohre, Harald, Heiko Opfer, and Gabor Orszag, "Diversifying risk parity," *Journal of Risk*, 2014, *16* (5), 53–79.

Luttmer, Erzo GJ, "What level of fixed costs can reconcile consumption and stock returns?," *Journal of Political Economy*, 1999, *107* (5), 969–997.

Lynch, Anthony W and Pierluigi Balduzzi, "Transaction costs and predictability: Some utility cost calculations," *Journal of Financial Economics*, 1999, *52* (1), 47–78.

——— **and** ———, "Predictability and transaction costs: The impact on rebalancing rules and behavior," *The Journal of Finance*, 2000, *55* (5), 2285–2309.

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.

Maguire, Phil, Philippe Moser, Kieran O'Reilly, Conor McMenamin, Robert Kelly, and Rebecca Maguire, "Maximizing positive porfolio diversification," in "2014 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr)" IEEE 2014, pp. 174–181.

Maillard, Sébastien, Thierry Roncalli, and Jérôme Teïletche, "The properties of equally weighted risk contribution portfolios," *The Journal of Portfolio Management*, 2010, *36* (4), 60–70.

Markowitz, Harry, "Portfolio selection," *The Journal of Finance*, 1952, *7* (1), 77–91.

———, "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.

Novy-Marx, Robert and Mihail Velikov, "A taxonomy of anomalies and their trading costs," *Review of Financial Studies*, 2016, *29* (1), 104–147.

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.

Roll, Richard, "A simple implicit measure of the effective bid-ask spread in an efficient market," *The Journal of Finance*, 1984, *39* (4), 1127–1139.

Romano, Joseph P, Azeem M Shaikh, and Michael Wolf, "A Practical Two-Step Method for Testing Moment Inequalities," *Econometrica*, 2014, *82* (5), 1979–2002.

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

Sharpe, William F, "Capital asset prices: A theory of market equilibrium under conditions of risk," *The journal of finance*, 1964, *19* (3), 425–442.

Stambaugh, Robert F, "Predictive regressions," *Journal of Financial Economics*, 1999, *54* (3), 375–421.

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

Tauchen, George and Robert Hussey, "Quadrature-based methods for obtaining approximate solutions to nonlinear asset pricing models," *Econometrica*, 1991, pp. 371–396.

Theron, Ludan and Gary Van Vuuren, "The maximum diversification investment strategy: A portfolio performance comparison," *Cogent Economics & Finance*, 2018, *6* (1), 1427533.

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.