Université de Montréal

Choix de portefeuille de grande taille et mesures de risque pour preneurs de décision pessimistes

par
Codjo Nérée Gildas Maxime NOUMON

Département de sciences économiques
Faculté des arts et des sciences

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

Août, 2013

Université de Montréal
Faculté des arts et des sciences

Cette thèse intitulée :

Choix de portefeuille de grande taille et mesures de risque pour preneurs de décision pessimistes

présentée par :

Codjo Nérée Gildas Maxime NOUMON

a été évaluée par un jury composé des personnes suivantes :

... président-rapporteur
... directrice de recherche
... directeur de recherche
... membre du jury
... examinateur externe
... vice-doyen, représentant du doyen de la FAS

Thèse acceptée le ...

A mes parents, et à Dieu !

Remerciements

Ma profonde gratitude et vibrant hommage vont tout d'abord à mes directeurs Marine Carrasco et Marc Henry. Pour leur disponibilité, leur implication et leur contributions, tout au long de ce long cheminement qu'est une thèse.

Je suis infiniment reconnaissant envers tout le corps professoral du Département de sciences économiques de l'Université de Montréal pour les connaissances transmises. En particulier Benoît Perron, Lyndon Moore, Ilze Kalnina et Rui Castro pour leurs conseils et encouragements. J'aimerais aussi exprimer toute ma gratitude à Dalibor Stevanovic pour sa disponibilité comme examinateur externe. Je remercie Basma Majerbi pour ces commentaires sur le second chapitre, et pour tout ce que j'ai appris au cours de ma visite à la Peter B. Gustavson School of Business de l'Université de Victoria. Mes Remerciements vont également à Iftekhar Hasan pour les fructueuses discussions concernant mon deuxième chapitre.

J'aimerais remercier le Département de sciences économiques de l'Université de Montréal, les laboratoires de recherches et les institutions qui m'ont soutenu aussi bien financièrement et qu'en nature : le Centre Interuniversitaire de Recherche en Analyse des Organisations (CIRANO), le Centre Interuniversitaire de Recherche en économie Quantitative (CIREQ) et la Banque Laurentienne.

Chaleureux remerciements à ma mère Angèle, mon père Benoît, mes frères et soeurs, Doris, Roméo, Odia et Bérénice pour leur amour, leur patience, et soutient dans les moments difficiles. Un grand merci à Ulrich Hounyo et Michel Kpade pour leur sages conseils dans les moments d'incertitudes. Merci à mes nombreux collègues et amis de l'université de Montreal que je ne pourrai pas tous citer ici, pour les bons moments passés ensembles. Pour finir, je remercie toutes les personnes qui ont contribué d'une manière ou d'une autre à cette thèse.

Résumé

Cette thèse de doctorat consiste en trois chapitres qui traitent des sujets de choix de portefeuilles de grande taille, et de mesure de risque. Le premier chapitre traite du problème d'erreur d'estimation dans les portefeuilles de grande taille, et utilise le cadre d'analyse moyenne-variance. Le second chapitre explore l'importance du risque de devise pour les portefeuilles d'actifs domestiques, et étudie les liens entre la stabilité des poids de portefeuille de grande taille et le risque de devise. Pour finir, sous l'hypothèse que le preneur de décision est pessimiste, le troisième chapitre dérive la prime de risque, une mesure du pessimisme, et propose une méthodologie pour estimer les mesures dérivées.

Le premier chapitre améliore le choix optimal de portefeuille dans le cadre du principe moyenne-variance de Markowitz (1952). Ceci est motivé par les résultats très décevants obtenus, lorsque la moyenne et la variance sont remplacées par leurs estimations empiriques. Ce problème est amplifié lorsque le nombre d'actifs est grand et que la matrice de covariance empirique est singulière ou presque singulière. Dans ce chapitre, nous examinons quatre techniques de régularisation pour stabiliser l'inverse de la matrice de covariance : le ridge, spectral cut-off, Landweber-Fridman et LARS Lasso. Ces méthodes font chacune intervenir un paramètre d'ajustement, qui doit être sélectionné. La contribution principale de cette partie, est de dériver une méthode basée uniquement sur les données pour sélectionner le paramètre de régularisation de manière optimale, i.e. pour minimiser la perte espérée d'utilité. Précisément, un critère de validation croisée qui prend une même forme pour les quatre méthodes de régularisation est dérivé. Les règles régularisées obtenues sont alors comparées à la règle utilisant directement les données et à la stratégie naïve 1/N, selon leur perte espérée d'utilité et leur ratio de Sharpe. Ces performances sont mesurée dans l'échantillon (in-sample) et hors-échantillon (out-of-sample) en considérant différentes tailles

d'échantillon et nombre d'actifs. Des simulations et de l'illustration empirique menées, il ressort principalement que la régularisation de la matrice de covariance améliore de manière significative la règle de Markowitz basée sur les données, et donne de meilleurs résultats que le portefeuille naïf, surtout dans les cas le problème d'erreur d'estimation est très sévère.

Dans le second chapitre, nous investiguons dans quelle mesure, les portefeuilles optimaux et stables d'actifs domestiques, peuvent réduire ou éliminer le risque de devise. Pour cela nous utilisons des rendements mensuelles de 48 industries américaines, au cours de la période 1976-2008. Pour résoudre les problèmes d'instabilité inhérents aux portefeuilles de grandes tailles, nous adoptons la méthode de régularisation spectral cut-off. Ceci aboutit à une famille de portefeuilles optimaux et stables, en permettant aux investisseurs de choisir différents pourcentages des composantes principales (ou dégrées de stabilité). Nos tests empiriques sont basés sur un modèle International d'évaluation d'actifs financiers (IAPM). Dans ce modèle, le risque de devise est décomposé en deux facteurs représentant les devises des pays industrialisés d'une part, et celles des pays émergents d'autres part. Nos résultats indiquent que le risque de devise est primé et varie à travers le temps pour les portefeuilles stables de risque minimum. De plus ces stratégies conduisent à une réduction significative de l'exposition au risque de change, tandis que la contribution de la prime risque de change reste en moyenne inchangée. Les poids de portefeuille optimaux sont une alternative aux poids de capitalisation boursière. Par conséquent ce chapitre complète la littérature selon laquelle la prime de risque est importante au niveau de l'industrie et au niveau national dans la plupart des pays.

Dans le dernier chapitre, nous dérivons une mesure de la prime de risque pour des préférences dépendent du rang et proposons une mesure du degré de pessimisme, étant donné une fonction de distorsion. Les mesures introduites généralisent la mesure de prime de risque dérivée dans le cadre de la théorie de l'utilité espérée, qui est fréquemment violée aussi bien dans des situations expérimentales que dans des situations réelles. Dans la grande famille des préférences considérées, une attention particulière est accordée à la CVaR (valeur à risque conditionnelle). Cette dernière mesure de risque est de plus en plus utilisée pour la construction de portefeuilles et est préconisée pour compléter la VaR (valeur à risque) utilisée depuis 1996 par le comité de Bâle. De plus, nous fournissons le cadre statistique nécessaire pour faire de l'inférence sur les mesures proposées. Pour finir, les propriétés des estimateurs proposés sont évaluées à travers une étude Monte-Carlo, et une illus-

tration empirique en utilisant les rendements journaliers du marché boursier américain sur de la période 2000-2011.

Mots-clés : Choix de portefeuille, analyse moyenne-variance, erreur d'estimation, régularisation, modèle de sélection, portefeuille d'actifs domestiques, risque de devise, modèle d'évaluation d'actifs financiers, prime de risque, préférences dépendant du rang, pessimisme, valeur à risque conditionnelle, processus empirique.

Abstract

This thesis consists of three chapters on the topics of portfolio choice in a high-dimensional context, and risk measurement. The first chapter addresses the estimation error issue that arises when constructing large portfolios in the mean-variance framework. The second chapter investigates the relevance of currency risk for optimal domestic portfolios, evaluates their ability of to diversify away currency risk, and study the links between portfolio weights stability and currency risk. Finally, under the assumption that decision makers are pessimistic, the third chapter derives the risk premium, propose a measure of the degree of pessimism, and provide a statistical framework for their estimation.

The first chapter improves the performance of the optimal portfolio weights obtained under the mean-variance framework of Markowitz (1952). Indeed, these weights give unsatisfactory results, when the mean and variance are replaced by their sample counterparts (plug-in rules). This problem is amplified when the number of assets is large and the sample covariance is singular or nearly singular. The chapter investigates four regularization techniques to stabilizing the inverse of the covariance matrix : the ridge, spectral cut-off, Landweber-Fridman, and LARS Lasso. These four methods involve a tuning parameter that needs to be selected. The main contribution is to derive a data-based method for selecting the tuning parameter in an optimal way, i.e. in order to minimize the expected loss in utility of a mean-variance investor. The cross-validation type criterion derived is found to take a similar form for the four regularization methods. The resulting regularized rules are compared to the sample-based mean-variance portfolio and the naive 1/N strategy in terms of in-sample and out-of-sample Sharpe ratio and expected loss in utility. The main finding is that regularization to covariance matrix significantly improves the performance of the mean-variance problem and outperforms the naive portfolio, especially in ill-posed cases, as suggested by

our simulations and empirical studies.

In the second chapter, we investigate the extent to which optimal and stable portfolios of domestic assets can reduce or eliminate currency risk. This is done using monthly returns on 48 U.S. industries, from 1976 to 2008. To tackle the instabilities inherent to large portfolios, we use the spectral cut-off regularization described in Chapter 1. This gives rise to a family of stable global minimum portfolios that allows investors to select different percentages of principal components for portfolio construction. Our empirical tests are based on a conditional International Asset Pricing Model (IAPM), augmented with the size and book-to-market factors of Fama and French (1993). Using two trade-weighted currency indices of industrialized countries currencies and emerging markets currencies, we find that currency risk is priced and time-varying for global minimum portfolios. These strategies also lead to a significant reduction in the exposure to currency risk, while keeping the average premium contribution to total premium approximately the same. The global minimum weights considered are an alternative to market capitalization weights used in the U.S. market index. Therefore, our findings complement the well established results that currency risk is significantly priced and economically meaningful at the industry and country level in most countries.

Finally, the third chapter derives a measure of the risk premium for rank-dependent preferences and proposes a measure of the degree of pessimism, given a distortion function. The introduced measures generalize the common risk measures derived in the expected utility theory framework, which is frequently violated in both experimental and real-life situations. These measures are derived in the neighborhood of a given random loss variable, using the notion of local utility function. A particular interest is devoted to the CVaR, which is now widely used for asset allocation and has been advocated to complement the Value-at-risk (VaR) proposed since 1996 by the Basel Committee on Banking Supervision. We provide the statistical framework needed to conduct inference on the derived measures. Finally, the proposed estimators are assessed through Monte Carlo study and illustrated using U.S. stock market data.

Keywords : Portfolio selection, mean-variance analysis, estimation error, regularization, model selection, optimal domestic portfolio, currency risk, International asset pricing,risk premium, rank-dependent preferences, pessimism, conditional value-at-risk, empirical process.

Table des matières

Table des figures

Liste des tableaux

Introduction

La présente thèse se situe dans le cadre générale des problèmes de choix de portefeuille et de mesures de risque. Elle propose des outils statistiques pour améliorer l'implémentation des règles d'investissement de portefeuille, et pour mesurer le risque dans le cas particulier où le preneur de décision est pessimiste. Les deux premiers chapitres s'intéressent aux problèmes de choix de portefeuilles, tandis que le troisième chapitre se propose de mesurer la prime de risque, sous l'hypothèse que preneur de décision est pessimiste.

Le premier chapitre est basé sur article écrit conjointement avec Marine Carrasco, dans lequel nous considérons le cadre de l'analyse moyenne-variance de Markowitz (1952). Dans son cadre théorique, Markowitz fait l'hypothèse que les rendements moyens et la matrice de covariance, qui interviennent dans la règle optimale de l'investisseur, sont connus. Cependant, dans la pratique, ces moments sont inconnus et doivent être estimés. Les problèmes importants causés par l'incertitude dans ces paramètres ont été soulignés par de nombreux auteurs, voir par exemple Kan et Zhou (2007). Résoudre le problème de moyenne-variance nécessite l'estimation de la matrice de covariance des rendements et de prendre son inverse. Il en résulte un problème inverse mal posé et des erreurs d'estimation, amplifiées par deux faits. Tout d'abord, le nombre de titres est généralement très élevé et, deuxièmement, les rendements des actifs peuvent être fortement corrélés. Dans l'échantillon, ces instabilités sont exprimées par le fait que la frontière minimum variance obtenu de l'échantillon est un estimateur fortement biaisée de la frontière de la population (Kan et Smith (2008)) ; hors - échantillon, les règles d'investissement qui en résultent sont caractérisées par des performances très médiocres. Pour résoudre ces problèmes, diverses solutions ont été proposées. Certains auteurs comme Frost et Savarino (1986), ont adopté un approche bayésienne. D'autre auteurs, comme Ledoit et Wolf

(2003, 2004a, b), ont utilisé le shrinkage, qui consiste à remplacer la matrice de covariance par une moyenne pondérée de la covariance de l'échantillon et une matrice issue d'un modèle avec peu d'erreur d'estimation. Tu et Zhou (2009) prennent une combinaison de la règle naïve 1 / N portefeuille avec des portefeuilles de Markowitz. Alternativement, Brodie, Daubechies, De Mol, Giannone et Loris (2008) et Fan, Zhang et Yu (2012) utilisent une méthode appelée Lasso qui consiste à imposer une contrainte sur la somme des valeurs absolues (l_1-norme) des pondérations du portefeuille. La contrainte l_1 généralise la contrainte de vente à découverte de Jagannathan et Ma (2003), et génère des portefeuilles parcimonieux dont le degré de parcimonie dépend d'un paramètre de réglage. Récemment, un cadre général et unifié a été proposé par DeMiguel, Garlappi, Nogales et Uppal (2009) en termes de portefeuille minimum variance, avec des contraintes utilisant différentes familles de normes, qui incluent toutes les règles citées ci-dessus. Par ailleurs, une nouvelle approche prometteuse introduite par Brandt, Santa-Clara et Valkanov (2009) éviter les difficultés dans l'estimation des rendements des actifs moments en modélisant directement la pondération du portefeuille dans chaque actif en fonction des caractéristiques de l'actif.

Dans ce premier chapitre, nous examinons diverses techniques de régularisation (ou stabilisation) empruntées à la littérature sur les problèmes inverses. En effet, l'inversion d'une matrice de covariance peut être considérée comme la résolution d'un problème inverse. Les problèmes inverses sont rencontrés dans de nombreux domaines et ont été largement étudiés, voir Carrasco, Florens, et Renault (2007) pour une revue. Ici, nous appliquons les 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 être complet, nous considérons également une forme de Lasso où on pénalise la norme l_1 des pondérations optimales de portefeuille. Ces différentes techniques de régularisation ont été utilisées et comparées dans le cadre de prévision des séries temporelles macroéconomiques utilisant un grand nombre de prédicteurs (Bai et Ng, 2008 ; De Mol, Giannone et Reichlin, 2008). Les quatre méthodes envisagées impliquent qu'un paramètre de régularisation qui doit être sélectionné. Jusqu'à présent très peu a été dit sur la façon de choisir le paramètre de réglage pour aboutir à une sélection optimale du portefeuille. Par exemple en utilisant le Lasso, Brodie et al. (2008), Fan et al. (2012) montrent qu'en faisant varier le paramètre de régularisation, on pourrait construire un portefeuille avec une parcimonie souhaitable, mais ne donnent pas une règle systématique sur la façon de le sélectionner dans la pratique. Ledoit et Wolf (2004) choisissent le paramètre de réglage afin de minimiser l'erreur quadratique moyenne de la matrice de covariance, mais cette approche n'est pas optimale pour la sélection du portefeuille. DeMiguel et al. (2009) calibrent la limite supérieure de la norme sur les poids du portefeuille minimum-variance, en minimisant la variance, ou en maximisant le rendement hors-échantillon du portefeuille. Leurs calibrages peuvent être améliorées en considérant un compromis optimal entre le risque du portefeuille et le rendement.

L'objectif du premier chapitre est de dériver une méthode axée sur les données pour la sélection du paramètre de régularisation de manière optimale. Nous adoptons le cadre de Kan et Zhou (2008), et supposons que l'investisseur est caractérisé par une fonction d'utilité moyenne-variance et voudrait minimiser la perte espérée due à l'utilisation d'une stratégie de portefeuille donnée. Notre approche englobe différents problèmes tels que le portefeuille de variance minimum considéré dans DeMiguel et al (2009) et Fan et al. (2012), le portefeuille moyenne-variance pris en compte dans Brodie et al (2009) et le portefeuille tangent. La perte espérée ne peut pas être dérivée analytiquement. Notre contribution est de fournir une estimation de la perte d'utilité anticipée qui utilise seulement les observations. Cette estimation est une version corrigée du biais, du critère de validation croisée généralisée. L'avantage de notre critère est qu'il s'applique à toutes les méthodes mentionnées ci-dessus et donne une base pour comparer les différentes méthodes.

Le second chapitre se place également dans le cadre des problèmes de choix de portefeuille. Il explore dans quelle mesure les portefeuilles optimaux d'actifs domestiques réduisent l'exposition au risque de change, en utilisant un modèle d'évaluation d'actifs financiers. Dans leur version internationale ces modèles contiennent des facteurs représentant le risque de devises, en plus du facteur de risque de marché du modèle standard (Sharpe, 1964 ; Lintner, 1965). Ce résultat est théoriquement prouvé sous l'hypothèse de la défaillance bien connue de la parité du pouvoir d'achat (Solnik, 1974 ; Adler et Dumas, 1983). Sur le marché américain des actions, ce résultat théorique est soutenu par un grand nombre d'évidences empiriques, tant au niveau des pays et qu'au niveau de l'industrie (De Santis et Gérard, 1998 ; Carrieri, Errunza et Majerbi, 2006a ; Francis Hasan et Hunter, 2008). Bien que les preuves du fait que le risque de change est un facteur primé sont accablants, très peu est dit sur les conséquences d'un tel résultat pour les investisseurs nationaux, en particulier, si le risque de change pourrait être réduit ou éliminé par une stratégie de portefeuille donnée, sur des actifs nationaux. En effet, la tarification du risque de change, implique que le risque de change est systématique et que les investisseurs exigent d'être récompensés par une prime pour y être exposés. Différentes stratégies de portefeuille sont susceptibles d'avoir différentes expositions au risque de change. Ceci vient du fait que ces stratégies utilisent des pondérations différentes sur les actifs dont l'exposition varie selon les industries et dans le temps (Francis, Hasan et Hunter, 2008). Par conséquent, face à des prix du risque communs, différentes stratégies sont également susceptibles de conduire à différentes primes de risque de change.

Choisir des portefeuilles composés exclusivement d'actifs domestiques, dans un contexte international est défendable pour au moins trois raisons. Tout d'abord, les investisseurs ont typiquement un avantage dans la négociation des actions dans leur pays, pour de nombreuses raisons telles que l'informations de qualité supérieure, le biais de la réglementation domestique à l'encontre des investisseurs étrangers (Choe, Kho, et Stulz, 2005). Deuxièmement, investir à l'étranger pourrait s'avérer pas nécessaire, car, dans certains cas, il est possible d'obtenir les gains de la diversification internationale à travers la diversification domestique (Errunza, Hogan et Hung, 1999). Enfin, bien que les avantages de la diversification mondiale ont été largement documentés, et sont toujours d'actualité (par exemple, Solnik, 1974 ; Christoffersen et al, 2012 ;. Driessen et Laeven, 2007), les dernières décennies ont connu une augmentation du niveau d'intégration internationale des marchés. Entre autres choses, ces tendances ont conduit à une augmentation des risques de marchés intérieurs des fluctuations des taux de change, et à la réduction du potentiel de diversification à l'international (voir par exemple Li, Sarkar et Wang (2003) [1]). Cette réduction des avantages de la diversification internationale, implique que l'investissement national est de plus en plus pertinent à considérer et à étudier.

Dans ce chapitre nous considérons des portefeuilles de variance minimale (DeMiguel, Garlappi, Nogales et Uppal, 2009 ; Jagannathan et Ma, 2003). Cette stratégie correspond à l'hypothèse que la principale préoccupation de l'investisseur est de réduire le risque global du portefeuille. Les portefeuilles

1. Bien que les marchés internationaux sont de plus en plus intégré cette n'élimine pas les avantages de la diversification de nouveaux investissements sur le marché.

sont construits aux États-Unis en utilisant un ensemble d'investissement composé de 48 portefeuilles d'industrie, et les pondérations optimales sont obtenues de manière récursive basée sur M mois précédents afin d'imiter le comportement dynamique d'un investisseur. Par ailleurs, comme mentionné dans le chapitre 1, lors de la construction des portefeuilles moyenne-variance, l'inverse de la matrice de covariance entre les rendements des actifs doit être estimée et inversée. Ceci peut conduire à des pondérations optimales instables, et peut compromettre les gains de diversification (DeMiguel Garlappi et Uppal, 2009 ; Carrasco et Noumon, 2013). Pour cette raison, nous considérons également dans le Chapitre 2, une famille de portefeuilles globaux minimaux qui sont régularisés pour faire face aux instabilités qui se posent dans les portefeuilles de grande taille. Parmi les techniques de régularisation disponibles, et cité plus haut, nous choisissons la coupure spectrale qui est étroitement liée à l'analyse des composantes principales, et constitue une manière de contrôler la stabilité des poids de portefeuilles.

L'objectif du deuxième chapitre est donc de déterminer si les portefeuilles optimaux d'actifs nationaux sont moins exposés au risque de change, et si ces portefeuilles conduisent à des primes qui non-négligeable . De plus, puisqu'une stabilisation est appliquée aux pondérations optimales considérées, notre méthodologie nous permet également d'étudier comment les stratégies stables d'investissement sont affectées par le risque de change. Les études les plus proches de la nôtre sont De Santis, Gérard et Hillion (1999) et Francis, Hasan et Hunter (2008). Nous utilisons un modèle similaire à celui utilisé dans ce dernier papier : la version conditionnelle du modèle de Fama et French (1993), augmenté du rendement sur deux indices de taux de change, représentant les pays industrialisés et les marchés émergents. Nos résultats sont en accord avec cette dernière étude. Cependant notre étude se distingue, par le fait qu'il étend le cadre de Francis Hasan et Hunter à celui des portefeuilles optimaux d'industries américaines. De plus, notre étude ne se limite pas aux 36 industries qui sont les plus susceptibles d'être exposés au risque de change, mais considère les 48 portefeuilles d'industrie qui fournissent une couverture exhaustive du marché boursier américain. En ce qui concerne la première étude, De Santis, Gérard et Hillion (1999) utilisent des stratégies d'allocation d'actifs dynamiques, pour des investisseurs universels.[2] Contrairement à ces auteurs, notre étude se concentre entièrement

2. Pour ces investisseurs, l'univers d'investissement contient des actions indices boursiers, et les dépôts euro-devises à court terme.

sur les portefeuilles domestiques. De plus ces auteurs décomposent le risque de change en ses composantes européenne et non européenne, alors que nous désagrégeons le risque de change en ses composantes correspondant aux les marchés industrialisés et émergents.

Les résultats du deuxième chapitre, peuvent être résumés en trois points principaux. Tout d'abord, nous obtenons que le risque de change est aussi un facteur primé et varie dans le temps, pour les portefeuilles domestiques diversifiés de manière optimale, aux États-Unis. Deuxièmement, nous constatons que les portefeuilles optimisés sur les avoirs intérieurs réduisent considérablement l'exposition moyenne des industries américaines au risque de change, tant par la taille que par la volatilité. Troisièmement, pour tous les portefeuilles stables de variance minimales considérés, la prime de devise est non-négligeable, mais reste inchangée en termes de leur contribution à la prime totale. Ces résultats sont pertinents à bien des égards. Premièrement, comme les règles d'investissement construits, représentent une alternative aux poids de la capitalisation boursière standards utilisés pour tester la tarification du risque de change, nos conclusions fournissent des preuves supplémentaires de la tarification du risque de change aux États-Unis. Par ailleurs, la réduction des expositions obtenus peuvent être considérées comme des mesures additionnelles de l'avantage de la diversification domestique. Deuxièmement, nos résultats ont des implications en terme de stratégies de couverture. Plus précisément, les portefeuilles optimaux domestiques ont le potentiel de réduire les coûts de couverture du risque de change, puisque ces règles conduisent à réduire la quantité et la fréquence des opérations de couverture et de rééquilibrage.

Contrairement au deuxième chapitre qui mesure la prime de risque dans le cadre d'un modèle d'évaluation d'actifs financiers, le troisième chapitre mesure la prime de risque en faisant une hypothèse sur les préférences du preneur de décision, et considère que le risque est représenté par une variable aléatoire X. Ce chapitre est motivé par le fait que dans beaucoup de situations réelles et expérimentales, les preneurs de décision tendent à adopter un comportement pessimiste, qui consiste à amplifier la probabilité d'un événement négatif. Ces faits sont supportés aussi bien par les violations des axiomes de la théorie de l'utilité espérée (le paradoxe d'Allais, et de saint Petersburg), et par des faits expérimentaux. Ainsi, nous faisons l'hypothèse que le preneur de décision est caractérisé par des préférences du type rang-dépendant (RDU) introduit par Quiggin (1982), Schmeidler (1989) et Yaari

(1987). Cette hypothèse particulièrement opportune au vu de leur utilisation croissante en finance et en assurance.

Sous l'hypothèse des préférences RDU, le preneur de décision évalue les alternatives risquées avec une fonction d'utilité sur les réalisations et une fonction de distorsion qui transforme la distribution de probabilité de ces réalisations. Différentes hypothèses sur la fonction de distorsion conduisent à différents comportements. Lorsque cette fonction est concave, cela correspond à la définition du pessimisme de Basset et al (2004). En effet, dans ce cadre les probabilités des réalisations les moins favorables sont accentuées.

Cette littérature est étroitement liée à celle des mesures de risque. Plutôt qu'une définition issue de la théorie économique du choix dans l'incertitude, cette dernière introduit des mesures de risque qui satisfont des propriétés désirables. L'un des travaux les plus importants dans cette lignée est l'article de Artzner et al (1999) qui introduit la classe des mesures de risque cohérentes satisfaisant 4 propriétés désirables : la monotonicité, l'invariance par translation, l'homogénéité et la sous-additivité.

Plusieurs estimateurs ont été proposés pour l'estimation des mesures du type RDU. Par exemple Gourieroux et Liu (2006a) estiment l'allocation efficace lorsque les mesures de distorsion définissent la fonction objective et les contraintes ; Gourieroux et Liu (2006b) proposent un cadre unifié pour l'analyse des mesures de distorsion et de leur sensibilité ; Scaillet (2004) considère l'approche non paramétrique pour estimer la CVaR et ses sensibilités ; Bassett et al. (2004) utilisent une fonction d'utilité de Choquet pour l'optimisation des portefeuilles pessimistes, en résolvant un problème de régression quantile.

Ce dernier chapitre apporte deux contributions au cadre décrit dans les paragraphes précédents. La première contribution est la dérivation de la prime de risque associée aux décideurs pessimistes. Ceci est effectué en surmontant deux difficultés. Premièrement, la séparation de la notion de pessimisme de celle de l'aversion pour le risque, est réputée être difficile à accomplir empiriquement (Quiggin (1982)). La deuxième difficulté vient de la non linéarité de la fonction de distorsion par rapport aux probabilités. En effet, ceci empêche l'utilisation immédiate de la définition de la prime de risque selon Pratt (1964), qui fait l'hypothèse que les décideurs maximisent l'utilité espérée. Par conséquent dériver la prime de risque nécessite de trouver la fonction d'utilité qu'utilise le preneur de décision, et qui correspond à un critère de maximisation d'utilité espérée. La première difficulté est surmontée en supposant que le décideur est neutre au risque et utilise une fonction d'utilité linéaire. Le deuxième question peut être résolue en supposant que la fonction des préférences est dérivable par rapport aux probabilités. Sous cette hypothèse la fonction d'utilité requise peut être dérivée et correspond à la fonction d'utilité locale proposée par Machina (1982). La résolution de ces questions conduit une mesure de la prime de risque et du coefficient d'aversion pour le risque, qui dépendent uniquement du pessimisme du décideur. En particulier, le coefficient d'aversion pour le risque peut être utilisé pour comparer et résumer le degré de pessimisme d'un décideur. La seconde contribution de ce chapitre est de proposer des outils statistiques pour estimer les mesures de risque dérivées sous l'hypothèse des préférences du type RDU, principalement sous leur forme de DRM. Précisément nous considérons le cas général d'une utilité RDU et le cas plus particulier de la CVaR. Nous adoptons une approche non paramétrique et établissons des résultats de convergence et de distributions asymptotiques.

Chapitre 1

Optimal portfolio selection using regularization

1.1 Introduction

In his seminal paper of 1952, Markowitz stated that the optimal portfolio selection strategy should be an optimal trade-off between return and risk instead of an expected return maximization only. In his theoretical framework, Markowitz made the important assumption that the beliefs about the future performance of asset returns are known. However in practice these beliefs have to be estimated. The damage caused by the so-called parameter uncertainty has been pointed out by many authors, see for instance Kan and Zhou (2007). Solving the mean-variance problem leads to estimate the covariance matrix of returns and take its inverse. This results in a ill-posed problem and in estimation error, amplified by two facts. First, the number of securities is typically very high and second, these security returns may be highly correlated. In-sample, these problems are reflected by the fact that the sample minimum-variance frontier is a highly biased estimator of the population frontier as shown by Kan and Smith (2008) ; out-of-sample, the resulting rules are characterized by very poor performances as extensively documented by DeMiguel, Garlappi, and Uppal (2007). To tackle these issues, various solutions have been proposed. Some authors have taken a Bayesian approach, see Frost and Savarino (1986). Some have used shrinkage, more precisely Ledoit and Wolf (2003, 2004a,b) propose to replace the covariance matrix by a weighted average of the sample covariance and some structured matrix.

Tu and Zhou (2009) take a combination of the naive 1/N portfolio with the Markowitz portfolio. Alternatively, Brodie, Daubechies, De Mol, Giannone, and Loris (2008) and Fan, Zhang, and Yu (2009) use a method called Lasso which consists in imposing a constraint on the sum of the absolute values (l_1-norm) of the portfolio weights. The l_1- constraint generalizes the short sale constraint of Jagannathan and Ma (2003) and generates sparse portfolios which degree of sparsity depends on a tuning parameter. Recently, a general and unified framework has been proposed by DeMiguel, Garlappi, Nogales, and Uppal (2009) in terms of norm-constrained minimum-variance portfolio that nests all the rules cited above. A new promising approach introduced by Brandt, Santa-Clara, and Valkanov (2009) avoid the difficulties in the estimation of asset returns moments by modelling directly the portfolio weight in each asset as a function of the asset's characteristics.

In this chapter, we investigate various regularization (or stabilization) techniques borrowed from the literature on inverse problems. Indeed, inverting a covariance matrix can be regarded as solving an inverse problem. Inverse problems are encountered in many fields and have been extensively studied, see Carrasco, Florens, and Renault (2007) for a review. Here, we will apply the three regularization techniques that are the most used : the ridge which consists in adding a diagonal matrix to the covariance matrix, the spectral cut-off which consists in discarding the eigenvectors associated with the smallest eigenvalues, and Landweber Fridman iterative method. For completeness, we also consider a form of Lasso where we penalize the l_1 norm of the optimal portfolio weights. These various regularization techniques have been used and compared in the context of forecasting macroeconomic time series using a large number of predictors by among others Bai and Ng (2008), and De Mol, Giannone, and Reichlin (2008). The four methods under consideration involve a regularization (or tuning) parameter which needs to be selected. Little has been said so far on how to choose the tuning parameter to perform optimal portfolio selection. For example using the Lasso, Brodie et al. (2008), Fan et al. (2009) show that by tuning the penalty term one could construct portfolio with desirable sparsity but do not give a systematic rule on how to select it in practice. Ledoit and Wolf (2004) choose the tuning parameter in order to minimize the mean-square error of the shrinkage covariance matrix, however this approach may not be optimal for portfolio selection. DeMiguel et al. (2009) calibrate the upper bound on the norm of the minimum variance portfolio weights, by minimizing portfolio variance or by maximizing the last period out-of-sample portfolio return. Their calibrations may be improved by considering an optimal trade-off between portfolio risk and return.

The main objective of this chapter is to derive a data-driven method for selecting the regularization parameter in an optimal way. Following the framework of Kan and Zhou (2008), we suppose that the investor is characterized by a mean-variance utility function and would like to minimize the expected loss incurred in using a particular portfolio strategy. The mean-variance investor approach nests different problems such as the minimum variance portfolio considered in DeMiguel et al. (2009) and Fan et al. (2009), the mean-variance portfolio considered in Brodie et al. (2009), and the tangency portfolio. The expected loss can not be derived analytically. Our contribution is to provide an estimate of the expected loss in utility that uses only the observations. This estimate is a bias-corrected version of the generalized cross-validation criterion. The advantage of our criterion is that it applies to all the methods mentioned above and gives a basis to compare the different methods.

The rest of the chapter is organized as follows. Section 2 reviews the mean-variance principle. Section 3 describes three regularization techniques of the inverse of the covariance matrix. Section 4 discusses stabilization techniques that take the form of penalized least-squares. Section 5 derives the optimal selection of the tuning parameter. Section 6 presents simulations results and Section 7 empirical results. Section 8 concludes.

1.2 Markowitz paradigm

Markowitz (1952) proposes the mean-variance rule, which can be viewed as a trade-off between expected return and the variance of the returns. For a survey, see Brandt (2010). Consider N risky assets with random return vector R_{t+1} and a risk-free asset with known return R_t^f. Define the excess returns $r_{t+1} = R_{t+1} - R_t^f$. We assume that the excess returns are independent identically distributed with mean and covariance matrix denoted by μ and Σ, respectively. The investor allocates a fraction x of wealth to risky assets and the remainder $(1 - 1_N' x)$ to the risk-free asset, where 1_N denotes a $N-$vector of ones. The portfolio excess return is therefore $x' r_{t+1}$. The investor is assumed to choose the vector x to maximize the mean-variance expected utility function

$$U(x) = x'\mu - \frac{\gamma}{2} x' \Sigma x \tag{1.1}$$

where γ is the relative risk aversion. The optimal portfolio is given by

$$x^* = \frac{1}{\gamma}\Sigma^{-1}\mu. \tag{1.2}$$

In practice the optimal portfolio that maximizes (1.1) is obtained by first estimating the expected return μ and the covariance matrix Σ of return, and then plug them into (1.2) to obtained the so called plug-in rule. For example, an investor can base his strategy on the T previous observed returns data $\Phi_T = \{r_1, r_2, \cdots, r_T\}$ to form a portfolio for period $T+1$. In particular, an inverse of the covariance matrix is needed. The choice of the sample covariance to form the plug-in rule may not be appropriate because it may be nearly singular and sometimes not 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 mean variance solution unreliable. Many regularization techniques can stabilize the inverse. They can be divided into two classes : regularization directly applied to the covariance matrix and regularization expressed as a penalized least-squares.

1.3 Regularization as approximation to an inverse problem

1.3.1 Inverse problem

Let r_t, $t = 1, \cdots, T$ be the observations of asset returns and R be the $T \times N$ matrix with tth row given by r_t'. Let $\Omega = E\left(r_t r_t'\right) = E\left(R'R\right)/T$.

$$\begin{aligned}
\Sigma^{-1}\mu &= \left(\Omega - \mu\mu'\right)^{-1}\mu \\
&= \left(\Omega^{-1} + \frac{\Omega^{-1}\mu\mu'\Omega^{-1}}{1 - \mu'\Omega^{-1}\mu}\right)\mu \\
&= \frac{\Omega^{-1}\mu}{1 - \mu'\Omega^{-1}\mu}
\end{aligned}$$

where the second equality follows from the updating formula for an inverse matrix (see Greene, 1993, p.25). Hence

$$x^* = \frac{\Omega^{-1}\mu}{\gamma\left(1 - \mu'\Omega^{-1}\mu\right)} = \frac{\beta}{\gamma\left(1 - \mu'\beta\right)} \tag{1.3}$$

where

$$\beta = \Omega^{-1}\mu = E\left(R'R\right)^{-1} E\left(R'1_T\right). \tag{1.4}$$

It is customary to replace the unknown expectation μ by the sample average $\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} r_t$ and the covariance Σ by the sample covariance $\hat{\Sigma} = \left(R - 1_T\hat{\mu}'\right)'\left(R - 1_T\hat{\mu}'\right)/T \equiv \tilde{R}'\tilde{R}$. Replacing μ and Σ by their sample counterparts, one obtains the sample based optimal allocation $\widehat{x} = \hat{\Sigma}^{-1}\hat{\mu}/\gamma$. Jobson and Korkie (1983, Equation (15)) and later Britten-Jones (1999) showed that $\widehat{x}$ can be rewritten as

$$\widehat{x} = \hat{\beta}/\left(\gamma\left(1 - \hat{\mu}'\hat{\beta}\right)\right)$$

where $\hat{\beta}$ is the OLS estimate of β in the regression

$$1 = \beta' r_{t+1} + u_{t+1}$$

or equivalently

$$1_T = R\beta + u \tag{1.5}$$

where R is the $T \times N$ matrix with rows composed of r_t'. In other words, one should not center r_t in the calculation of x^*. Finding β can be thought of as finding the minimum least-squares solution to the equation :

$$R\beta = 1_T. \tag{1.6}$$

It is a typical inverse problem.

The ill-posedness of the previous problem depends on the characteristics of the matrix $\hat{\Omega} = R'R/T$. Two difficulties may occur : the assets could be highly correlated (i.e. the population covariance matrix Σ is nearly singular) or the number of assets could be too large relative to the sample size (i.e. the sample covariance is (nearly) singular even though the population covariance is not). In such cases, $\hat{\Omega}$ typically has some singular values close to zero resulting in an ill posed problem, such that the optimization of the portfolio becomes a challenge. These difficulties are summarized by the condition number which is the ratio of the maximal and minimal eigenvalue of $\hat{\Omega}$. A large condition number leads to unreliable estimate of the vector of portfolio weights x.

The inverse problem literature, that usually deals with infinite dimensional problems, has proposed various regularization techniques to stabilize the solution to (1.6). For an overview on inverse problems, we refer the readers to Kress (1999) and Carrasco, Florens, and Renault (2007). We will consider here the three most popular regularization techniques : ridge, spectral cut-off, and Landweber Fridman. Each method will give a different estimate of β, denoted $\hat{\beta}_\tau$ and estimate of x^*, denoted $\widehat{x}_\tau = \hat{\beta}_\tau / \left(\gamma \left(1 - \hat{\mu}' \hat{\beta}_\tau\right)\right)$.

The $T \times N$ matrix R can be regarded as an operator from $\mathbb{R}^N$ (endowed with the inner product $\langle v, w \rangle = v'w$) into $\mathbb{R}^T$ (endowed with the inner product $\langle \phi, \varphi \rangle = \phi' \varphi / T$). The adjoint of R is R'/T. Let $\left(\hat{\lambda}_j, \hat{\phi}_j, \hat{v}_j\right)$, $j = 1, 2, ..., N$ be the singular system of R, i.e. $R\hat{\phi}_j = \hat{\lambda}_j \hat{v}_j$, $R'\hat{v}_j / T = \hat{\lambda}_j \hat{\phi}_j$, moreover $\left(\hat{\lambda}_j^2, \hat{\phi}_j\right)$ are the eigenvalues and orthonormal eigenvectors of $R'R/T$ and $\left(\hat{\lambda}_j^2, \hat{v}_j\right)$ are the nonzero eigenvalues and orthonormal eigenvectors of RR'/T. If $N < T$, it is easier to compute $\hat{\phi}_j$ and $\hat{\lambda}_j^2$, $j = 1, ..., N$ the orthonormal eigenvectors and eigenvalues of the matrix $R'R/T$ and deduce the spectrum of RR'/T. Indeed, the eigenvectors of RR' are $\hat{v}_j = R\widehat{\phi}_j / \hat{\lambda}_j$ associated with the same nonzero eigenvalues $\hat{\lambda}_j^2$. Let $\tau > 0$ be a regularization parameter.

1.3.2 Ridge regularization

The Ridge regression has been introduced by Hoerl and Kennard (1970) as a more stable alternative to the standard least-squares estimator with potential lower risk. It consists in adding a diagonal matrix to $R'R/T$.

$$\hat{\beta}_\tau = \left(\frac{R'R}{T} + \tau I\right)^{-1} \frac{R' 1_T}{T}, \tag{1.7}$$

$$\hat{\beta}_\tau = \sum_{j=1}^{N} \frac{\hat{\lambda}_j}{\hat{\lambda}_j^2 + \tau} \left(1_T' \widehat{v}_j\right) \hat{\phi}_j.$$

This regularization has a Bayesian interpretation, see i.e. De Mol et al. (2008).

1.3.3 Spectral cut-off regularization

This method discards the eigenvectors associated with the smallest eigenvalues.

$$\hat{\beta}_\tau = \sum_{\widehat{\lambda}_j^2 > \tau} \frac{1}{\widehat{\lambda}_j} \left(1_T' \widehat{v}_j\right) \widehat{\phi}_j.$$

Interestingly, $\widehat{v}_j$ are the principal components of $\hat{\Omega}$, so that if r_t follows a factor model, $\widehat{v}_1$, $\widehat{v}_2$,... estimate the factors.

1.3.4 Landweber-Fridman regularization

Let c be a constant such that $0 < c < 1/\|R\|^2$ where $\|R\|$ is the largest eigenvalue of R. The solution to (1.6) can be computed iteratively as

$$\psi_k = \left(I - c\frac{R'R}{T}\right)\psi_{k-1} + c\frac{R'1_T}{T},\ k = 1, 2, ..., 1/\tau - 1$$

with $\psi_0 = cR'1_T/T$. Alternatively, we can write

$$\hat{\beta}_\tau = \sum \frac{1}{\widehat{\lambda}_j}\left\{1 - \left(1 - c\widehat{\lambda}_j^2\right)^{1/\tau}\right\}\left(1_T'\widehat{v}_j\right)\widehat{\phi}_j.$$

Here, the regularization parameter τ is such that $1/\tau - 1$ represents the number of iterations. The three methods involve a regularization parameter τ which needs to converge to zero with T at a certain rate for the solution to converge.

1.3.5 Explicit expression of estimators

For the three regularizations considered above, we have

$$R\widehat{\beta}_\tau = M_T(\tau)\,1_T$$

with

$$M_T(\tau)\,w = \sum_{j=1}^{T} q\left(\tau, \hat{\lambda}_j^2\right)\left(w'\widehat{v}_j\right)\widehat{v}_j$$

for any $T-$ vectors w. Moreover, $trM_T(\tau) = \sum_{j=1}^{T} q\left(\tau, \hat{\lambda}_j^2\right)$. The function q takes a different form depending on the type of regularization. For Ridge, $q\left(\tau, \hat{\lambda}_j^2\right) = \hat{\lambda}_j^2/\left(\hat{\lambda}_j^2 + \tau\right)$. For Spectral cut-off, $q\left(\tau, \hat{\lambda}_j^2\right) = I\left(\hat{\lambda}_j^2 \geq \tau\right)$. For Landweber Fridman, $q\left(\tau, \hat{\lambda}_j^2\right) = 1 - \left(1 - c\hat{\lambda}_j^2\right)^{1/\tau}$.

1.3.6 Related estimator : Shrinkage

In this subsection, we compare our methods with a popular alternative called shrinkage. Shrinkage can also be regarded as a form of regularization. Ledoit and Wolf (2003) propose to estimate the returns covariance matrix by a weighted average of the sample covariance matrix $\hat{\Sigma}$ and an estimator with a lot of structure F, based on a model. The first one is easy to compute and has the advantage to be unbiased. The second one contains relatively little estimation error but tends to be misspecified and can be severely biased. The shrinkage estimator takes the form of a convex linear combination : $\delta F + (1-\delta)\hat{\Sigma}$, where δ is a number between 0 and 1. This method is called shrinkage since the sample covariance matrix is shrunk toward the structured estimator. δ is referred to as the shrinkage constant. With the appropriate shrinkage constant, we can obtain an estimator that performs better than either extreme (invertible and well-conditioned).

Many covariance matrices F could be used. Ledoit and Wolf (2003) suggested the single factor model of Sharpe (1963) which is based on the assumption that stock returns follow the model (Market model) :

$$r_{it} = \alpha_i + \beta_i r_{0t} + \varepsilon_{it}$$

where residuals ϵ_{it} are uncorrelated to market returns r_{0t} and to one another, with a constant variance $Var(\epsilon_{it}) = \delta_{ii}$. The resulting covariance matrix is

$$\Phi = \sigma_0^2 \beta\beta' + \Delta$$

Where σ_0^2 is the variance of market returns and $\Delta = diag(\delta_{ii})$. σ_0^2 is consistently estimated by the sample variance of market returns, β by OLS, and δ_{ii} by the residual variance estimate. A consistent estimate of Φ is then

$$F = s_0^2 bb' + D.$$

Instead of using the F derived from a factor model, one can use the constant correlation model[1] (Ledoit and Wolf (2004a)) or the identity matrix $F = I$ (Ledoit and Wolf (2004b)). They give comparable results but are easier to compute.

In the particular case where the shrinkage target is the identity matrix, the

1. All the pairwise covariances are identical.

shrinkage method is equivalent to Ridge regularization since the convex linear combination $\delta I + (1-\delta)\hat{\Sigma}$ can be rewritten :

$$\Sigma_{Shrink} = c\left(\hat{\Sigma}+\alpha I\right),$$

and

$$\Sigma_{Shrink}^{-1} = c^{-1}\left(\hat{\Sigma}+\alpha I\right)^{-1},$$

where c is a constant. Once the shrinkage target is determined one has to choose the optimal shrinkage intensity δ^*. Ledoit and Wolf (2004b) propose to select δ^* so that it minimizes the expected L^2 distance between the resulting shrinkage estimator $\Sigma_{Shrink} = \hat{\delta}^* F + (1-\hat{\delta}^*)\hat{\Sigma}$ and the true covariance matrix Σ. The limitation of this criterion is that it only focuses on the statistical properties of Σ, and in general could fail to be optimal for the portfolio selection.

1.4 Regularization scheme as penalized least-square

The traditional optimal Markowitz portfolio x^* is obtained from (1.3) and

$$\beta = \arg\min_{\beta} E\left[\left|1-\beta' r_t\right|^2\right]$$

If one replaces the expectation by the sample average, the problem becomes :

$$\hat{\beta} = \arg\min_{\beta} \|1_T - R\beta\|_2^2 \tag{1.8}$$

As mentioned before, the solution of this problem may be very unreliable if $R'R$ is nearly singular. To avoid having explosive solutions, we can penalize the large values by introducing a penalty term applied to a norm of β. Depending on the norm we choose, we end up with different regularization techniques.

1.4.1 Bridge method

For $\varsigma > 0$ the Bridge estimate is given by

$$\widehat{\beta}_\tau = \arg\min_{\beta} \|1_T - R\beta\|_2^2 + \tau\sum_{i=1}^{N}|\beta_i|^{\varsigma}$$

where τ is the penalty term.

The Bridge method includes two special cases. For $\varsigma = 1$ we have the Lasso regularization, while $\varsigma = 2$ leads to the Ridge method. The term $\sum_{i=1}^{N} |x_i|^{\varsigma}$ can be interpreted as a transaction cost. It is linear for Lasso, but quadratic for the ridge. The portfolio will be sparse as soon as $\varsigma \leq 1$. The objective function is strictly convex when $\varsigma > 1$, convex for $\varsigma = 1$ and no longer convex for $\varsigma < 1$. The case with $\varsigma < 1$ is considered in Huang, Horowitz, and Ma (2008), but will not be examined any further here.

1.4.2 Least Absolute Shrinkage and Selection Operator (LASSO)

The Lasso regularization technique introduced by Tibshirani (1996) is the l_1-penalized version of the problem (1.8). The Lasso regularized solution is obtained by solving :

$$\widehat{\beta}_{\tau} = \arg\min_{\beta} \|1_T - R\beta\|_2^2 + \tau \|\beta\|_1 .$$

The main feature of this regularization scheme is that it induces sparsity. It has been studied by Brodie, Daubechies, De Mol, Giannone, and Loris (2008) to compute portfolio involving only a small number of securities. For two different penalty constants τ_1 and τ_2 the optimal regularized portfolio satisfies : $(\tau_1 - \tau_2)\left(\left\|\beta^{[\tau_2]}\right\|_1 - \left\|\beta^{[\tau_1]}\right\|_1\right) \geq 0$ then the higher the l_1-penalty constant (τ), the sparser the optimal weights. So that a portfolio with non negative entries corresponds to the largest values of τ and thus to the sparsest solution. In particular the same solution can be obtained for all τ greater than some value τ_0.

Brodie et al. consider models without a riskfree asset. Using the fact that all the wealth is invested ($x'1_N = 1$), they use the equivalent formulation for the objective function as :

$$\|1_T - Rx\|_2^2 + 2\tau \sum_{i \text{ with } x_i < 0} |x_i| + \tau$$

which is equivalent to a penalty on the short positions. The Lasso regression then regulates the amount of shorting in the portfolio designed by the optimization process, so that the problem stabilizes. For a value of τ sufficiently large, all the components of x will be nonnegative, thus excluding short-selling. This gives a rationale for the finding of Jagannathan and Ma (2003). Jagannathan and Ma found that imposing the no short-selling constraint improves the performance of portfolio selection. This constraint acts as a regularization on the portfolio weights. The general form of the l_1-penalized regression with linear constraints is :

$$\widehat{\beta}_\tau = \arg\min_{\beta \in H} \|b - A\beta\|_2^2 + \tau \|\beta\|_1$$

H is an affine subspace defined by linear constraints. The regularized optimal portfolio can be found using an adaptation of the homotopy / LARS algorithm as described in Brodie et al. (2008). In appendix A, we provide a detailed description of this algorithm.

1.4.3 Ridge method

Interestingly, the ridge estimator described in (1.7) can be written alternatively as a penalized least-squares with l_2 norm. The Ridge regression is then given by

$$\widehat{\beta}_\tau = \arg\min_{\beta} \|1_T - R\beta\|_2^2 + \tau \|\beta\|_2^2 \tag{1.9}$$

Contrary to the Lasso regularization, the Ridge does not deliver a sparse portfolio, but selects all the securities with possibly short-selling.

1.5 Optimal selection of the regularization parameter

In order to compare different portfolio rules, a natural objective function can be established as the average out-of-sample performance. As proposed by Khan and Zhou (2007), the out-of-sample performance for a given rule $\hat{x}_\tau$ can be measured as

$$U(\hat{x}_\tau) = \hat{x}_\tau' \mu - \frac{\gamma}{2} \hat{x}_\tau' \Sigma \hat{x}_\tau,$$

which represents the expected utility conditional on the weights being chosen as $\hat{x}_\tau$. Since $\hat{x}_\tau$ is a function of the observed historical returns $\Phi_T = \{r_1, r_2, \cdots, r_T\}$, it is a random variable. As a result, $U(\hat{x}_\tau)$ it is also random variable and it is natural to evaluate a portfolio rule based on its expected out-of-sample performance $E[U(\hat{x}_\tau)]$, where the expectation is taken with respect to the true distribution of Φ_T.

1.5.1 Loss function of estimated allocation

Given the previous argument, the standard statistical decision theory can be used, and we can define the loss function of using rule $\hat{x}_\tau$ as :

$$\begin{aligned} L_T(\tau) &= U(x^*) - U(\hat{x}_\tau) \\ &= (x^* - \hat{x}_\tau)' \mu + \frac{\gamma}{2} \left(\hat{x}_\tau' \Sigma \hat{x}_\tau - x^{*\prime} \Sigma x^* \right) \\ &= (x^* - \hat{x}_\tau)' (\mu - \gamma \Sigma x^*) + \frac{\gamma}{2} (\hat{x}_\tau - x^*)' \Sigma (\hat{x}_\tau - x^*) \\ &= \frac{\gamma}{2} (\hat{x}_\tau - x^*)' \Sigma (\hat{x}_\tau - x^*) . \end{aligned} \tag{1.10}$$

The loss $L_T(\tau)$ depends on the realizations of the historical returns data Φ_T. It is natural to think that the investor would like to select the parameter τ to minimize the expected loss function $EL_T(\tau) = \frac{\gamma}{2} E(\hat{x}_\tau - x^*)' \Sigma (\hat{x}_\tau - x^*)$ in order to account for the average losses involving actions taken under various outcomes of Φ_T. Note that the optimal τ will also maximize the expected out-of-sample performance $E(U(\hat{x}_\tau))$.

The risk function $EL_T(\tau)$ has been used by many authors to rank various portfolios rules, where the portfolio with the lowest risk is the preferred [e.g. Jorion (1986), Frost and Savarino (1986), Khan and Zhou (2007)]. A distinct feature of our study, is that the risk function also depends on the regularization parameter that we seek to select optimally.

In the remainder of this session, our goal is to give a convenient expression for the criterion $E(L_T(\tau))$. Consider $\widehat{x}_\tau = \hat{\beta}_\tau / \left(\gamma \left(1 - \hat{\mu}' \hat{\beta}_\tau \right) \right)$ where $\hat{\beta}_\tau$ is given by

$$\widehat{\beta}_\tau = \widehat{\Omega}_\tau^{-1} R' 1_T / T \tag{1.11}$$

where $\widehat{\Omega}_\tau^{-1}$ is a regularized inverse of $\hat{\Omega} = R'R/T$. Using the notation (1.3), the optimal allocation x^* can be written as $\beta / (\gamma (1 - \mu' \beta))$.

The criterion (1.10) involves

$$\begin{aligned} \gamma (\hat{x}_\tau - x^*) &= \frac{\hat{\beta}_\tau}{1 - \hat{\mu}' \hat{\beta}_\tau} - \frac{\beta}{1 - \mu' \beta} \\ &= \frac{\hat{\beta}_\tau - \beta}{\left(1 - \hat{\mu}' \hat{\beta}_\tau\right)(1 - \mu' \beta)} - \frac{\hat{\beta}_\tau (\mu' \beta) - \beta \left(\hat{\mu}' \hat{\beta}_\tau \right)}{\left(1 - \hat{\mu}' \hat{\beta}_\tau\right)(1 - \mu' \beta)} . \end{aligned} \tag{1.12}$$

Note that $|\mu' \beta| < 1$ by construction. To evaluate (1.10), we need to evaluate the rate of convergence of the different terms in its expansion. To do so, we assume that β (or equivalently μ) satisfies some regularity condition. This condition given in Assumption A is similar in spirit to the smoothness condition of a function in nonparametric regression estimation for instance. However contrary to a smoothness condition that would concern only β, this condition relates the properties of β to those of Ω. It implies that β belongs to the range of $\Omega^{\nu/2}$. This type of conditions can be found in Carrasco, Florens, and Renault (2007) and Blundell, Chen, and Kristensen (2007) among others.

Assumption A.
(i) For some $\nu > 0$, we have

$$\sum_{j=1}^{N} \frac{\langle \mu, \phi_j \rangle^2}{\lambda_j^{2\nu+4}} < \infty$$

where ϕ_j and λ_j^2 denote the eigenvectors and eigenvalues of Ω.
(ii) Σ is Hilbert Schmidt (its eigenvalues are square summable).

Assumption A(i) is equivalent to $\sum_{j=1}^{N} \frac{\langle \beta, \phi_j \rangle^2}{\lambda_j^{2\nu}} < \infty$ because $\beta = \Omega^{-1}\mu$.

Assumption A implies in particular that $\|\beta\|^2 < \infty$. Let β_τ be defined as $\hat{E}\left(\hat{\beta}_\tau | R\right)$ where $\hat{E}\,(.|R)$ is the orthogonal projection on R.

Proposition 1. *Under Assumption A and assuming N and T go to infinity, we have*

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

The proof of Proposition 1 is given in Appendix A.1. The rest $rest(\tau)$ of the approximation in Proposition 1 is evaluated by simulation, and is generally negligible compared to the left-hand side (less than 1%). Given that the rest does not have a closed-form analytical expression, we ignore this term in subsequent optimality results derivations.

1.5.2 Cross-validation

From Proposition 1, it follows that minimizing $E\left(L_T\left(\tau\right)\right)$ is equivalent to minimizing

$$\frac{1}{T}E\left\|R\left(\hat{\beta}_\tau-\beta\right)\right\|^2 \tag{1.13}$$

$$+\frac{\left(\mu'\left(\beta_\tau-\beta\right)\right)^2}{\left(1-\mu'\beta\right)}. \tag{1.14}$$

Terms (1.13) and (1.14) depend on the unknown β and hence need to be approximated. Interestingly, (1.13) is equal to the prediction error of model (2.18) plus a constant and has been extensively studied. To approximate (1.13), we use results on cross-validation from Craven and Wahba (1979), Li (1986, 1987), and Andrews (1991) among others.

The rescaled MSE

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

can be approximated by generalized cross validation criterion :

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

Using the fact that

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

(1.14) can be estimated by plug-in :

$$\frac{\left(1_T'\left(M_T\left(\tau\right)-I_T\right)R\hat{\beta}_{\tilde{\tau}}\right)^2}{T^2\left(1-\hat{\mu}'\hat{\beta}_{\tilde{\tau}}\right)} \tag{1.15}$$

where $\hat{\beta}_{\tilde{\tau}}$ is an estimator of β obtained for some consistent $\tilde{\tau}$ ($\tilde{\tau}$ can be obtained by minimizing $GCV\left(\tau\right)$). Note that the expression of (1.15) does not presume anything on the regularity of β (value of ν).

The optimal value of τ is defined as

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

where $H_T = \{1, 2, ..., T\}$ for spectral cut-off and Landweber Fridman and $H_T = (0, 1)$ for Ridge. In our simulations the bias term contribute on average to less than 4% to the value of our criterion. At the optimal τ, this contribution falls down to 1% and the optimal τ obtained are approximately the same with and without the bias term.

The Lasso estimator does not take the simple form (1.11). However, Tibshirani (1996) shows that it can be approximated by a ridge type estimator and suggests using this approximation for cross-validation. Let $\tilde{\beta}(\tau)$ be the Lasso estimator for a value τ. By writing the term $\sum |\beta_j|$ as $\sum \beta_j^2 / |\beta_j|$, we see that $\tilde{\beta}(\tau)$ can be approximated by

$$\beta^* = \left(R'R + \tau(c) W^-(\tau)\right)^{-1} R' 1_T$$

where c is the upper bound $\sum |\beta_j|$ in the constrained problem equivalent to the penalized Lasso and $W(\tau)$ is the diagonal matrix with diagonal elements $\left|\tilde{\beta}_j(\tau)\right|$, W^- is the generalized inverse of W and $\tau(c)$ is chosen so that $\sum_j \left|\beta_j^*\right| = c$. Since $\tau(c)$ represents the Lagrangian multiplier on the constraint $\sum_j \left|\beta_j^*\right| \leq c$, we always have this constraint binding when $\tau(c) \neq 0$ (ill-posed cases). Let

$$p(\tau) = tr\left\{R\left(R'R + \tau(c) W^-(\tau)\right)^{-1} R'\right\}.$$

The generalized cross-validation criterion for Lasso is

$$GCV(\tau) = \frac{1}{T} \frac{\left\|1_T - R\widetilde{\beta}(\tau)\right\|^2}{\left(1 - p(\tau)/T\right)^2}.$$

Tibshirani (1996) shows in simulations that the above formula gives good results.

1.5.3 Optimality

Let $L_T^*(\tau) = \frac{1}{T}\left\|R\left(\hat{\beta}_\tau - \beta\right)\right\|^2 + \frac{(\mu'(\beta_\tau - \beta))^2}{1 - \mu'\beta}$, hence $L_T(\tau) = L_T^*(\tau) + rest(\tau)$ (where $rest(\tau)$ is defined as $L_T(\tau) - L_T^*(\tau)$). Let $R_T^*(\tau) = EL_T^*(\tau)$.

Assumption B. (i) u_t in the regression $1 = \beta' r_t + u_t$ is independent, identically distributed with mean $(1 - \beta'\mu)$ and $E(u_t^2) = \omega^2$. Moreover,

$E(u_t r_t) = 0$.
(ii) $Eu_t^{4m} < \infty$,
(iii) $\sum_{\tau \in H_T} (TR_T^*(\tau))^{-m} \to 0$ for some natural number m.
(iv) N diverges to infinity as T goes to infinity and $NT^{-1/(\nu+2)}$ goes to zero.

Note that the model in Assumption B, (i) should not be regarded as an economic model. It is an artificial regression for which the assumptions on u_t are consistent with our assumptions on r_t and the fact that $\beta = E(R'R)^{-1} E(R'1)$.

Using the same argument as in Li (1987, (2.5)), it can be shown that a sufficient condition for B(iii) for spectral cut-off and $m = 2$ is

$$\inf_{\tau \in H_T} TR_T^*(\tau) \to \infty.$$

This condition is satisfied under Assumption A (see Lemma 2 in Appendix).

Proposition 2. *Under Assumptions A and B, our selection procedure for τ in the case of spectral cut-off is asymptotically optimal in the sense that*

$$\frac{L_T(\hat{\tau})}{\inf_{\tau \in H_T} L_T(\tau)} \to 1 \text{ in probability.}$$

The proof of Proposition 2 draws from that Li (1987) for discrete sets H_T. However, it requires some adjustments for two reasons. First, our residual u_t does not have mean zero. On the other hand, $E(u_t r_t) = 0$ and we exploit this equality in our proof. Second, it is usual to prove optimality by assuming that the regressors are deterministic or alternatively by conditioning on the regressors. Here, we can not condition on the regressors because, given r_t, u_t is not random. Hence, we have to proceed unconditionally. The proof for the asymptotic optimality of SC is presented in Appendix A.1. The proof for LF uses a similar approach. The asymptotic optimality for the ridge using Li (1986) is left for future research.

1.6 Simulations

In this section, we use simulated data to assess the performance of the proposed investment strategies. The naive 1/N portfolio is taken as a benchmark to which we compare the regularized rules. The comparisons are made in terms of in-sample performance as in Fan and Yu (2009). For a wide range of number of observations and level of aversion to risk, we examine the in-sample expected loss in utility and the sharpe ratio. The expected loss in utility is also referred to as the actual loss since it is computed using the true covariance matrix Σ. The out-of-sample performances in terms of Sharpe ratio are instead provided in the empirical study.

1.6.1 A three-factor model

We use a three-factor model to assess the in-sample performance of our strategies through a Monte Carlo study. Precisely, we suppose that the N excess returns of assets are generated by the model :

$$r_{it} = b_{i1}f_{1t} + b_{i2}f_{2t} + b_{i3}f_{3t} + \varepsilon_{it} \quad for \quad i = 1, \cdots, N \tag{1.16}$$

or in a contracted form :

$$R = BF + \varepsilon$$

where b_{ij} are the factors loading of the i^{th} asset on the factor f_j, ε_i is the idiosyncratic noise independent of the three factors and independent of each other.

We assume further a trivariate normal distribution for the factor loading coefficients and for the factors : $b_i \sim N(\mu_b, \Sigma_b)$ and$f_t \sim N(\mu_f, \Sigma_f)$. The ε_i are supposed to be normally distributed with level σ_i drawn from a uniform distribution, so their covariance matrix is $\Sigma_\varepsilon = diag(\sigma_1^2, \cdots, \sigma_N^2)$. As a consequence the covariance matrix of returns is given by :

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

The parameters μ_f, Σ_f, μ_b and Σ_b used in the model (1.16) are calibrated to market data from July 1980 to June 2008. The data sets used consist of 20 years monthly returns of Fama-French three factors and of 30 industry portfolio from French data library. As pointed out in Fan et al. (2008) a natural idea for estimating Σ is to use the least-squares estimators of B, Σ_f and Σ_ϵ and obtain a substitution estimator :

$$\hat{\Sigma} = \hat{B}\hat{\Sigma}_f\hat{B}' + \hat{\Sigma}_\varepsilon$$

where $\hat{B} = RF'(FF')^{-1}$ is the matrix of estimated regression coefficients, Σ_f is the covariance matrix of the three Fama-French factors. These three factors are the excess return of the proxy of the market portfolio over the one-month treasury bill, the difference of return between large and small capitalization, that capture the size effect, and the difference of returns between high and low book-to-market ratios, that capture the valuation effect. We choose idiosyncratic noise to be normally distributed with standard deviation σ_i uniformly distributed between 0.01 and 0.03. The calibrated values are such that the generated asset returns exhibit three principal components. This means in practice that the covariance matrix of the generated returns have three dominant eigenvalues. Once generated, the factor loadings are kept fixed throughout replications, while the factors differ from simulations to simulations and are drawn from the trivariate distribution. Table 1.1 summarizes the calibrated mean and covariance matrix for the factors and the factors loadings.

Parameters for factor loadings				Parameters for factor returns			
μ_b		Σ_b		μ_f		Σ_f	
0.9919	0.0344	0.0309	0.0005	0.0060	0.0019	0.0003	-0.0005
0.0965	0.0309	0.0769	0.0042	0.0014	0.0003	0.0009	-0.0003
0.1749	0.0005	0.0042	0.0516	0.0021	-0.0005	-0.0003	0.0012

TABLE 1.1 – Calibrated parameters used in simulations

1.6.2 Estimation methods and tuning parameters

We start by a series of simulations to assess the performance of the different strategies proposed. This is done relative to the benchmark naive 1 over N strategy and the sample based Markowitz portfolio that is well known to perform poorly. The portfolios considered are the naive equally weighted portfolio (1oN), the sample-based mean variance portfolio (M), the Lasso portfolio (L), the ridge-regularized portfolio (Rdg), the spectral cut-off regularized portfolio (SC) and the Landweber-Fridman portfolio (LF) as summarized in Table 1.2.

The three regularization techniques introduced to improve the optimality of the sample-based Markowitz portfolio involve a regularization parameter τ and they correspond to the sample-based Markowitz portfolio for $\tau = 0$. So our approach can be considered as a generalization that aims to stabilize while improving the performance of the sample-based mean-variance

#	Model	Abbreviations
1	Naive evenly weighted portfolio	1oN
2	Sample-based mean variance portfolio	M
3	Lasso Portfolio	L
4	Optimal Ridge portfolio	Rdg
5	Optimal Spectral cut-off Portfolio	SC
6	Optimal Landweber-Fridman Portfolio	LF

Table 1.2 – List of investment rules

portfolio. Here we give some insights about the effect from tuning different regularization parameters.

The ridge, the spectral cut-off and the Landweber-Fridman schemes have a common feature that they transform the eigenvalues of the returns covariance matrix so that the resulting estimate has a more stable inverse. This transformation is done with a damping function $q(\tau, \lambda)$ specific to each approach as introduced previously.

The Ridge is the easiest regularization to implement and recovers the sample-based mean-variance minimizer for $\tau = 0$.

For SC, minimizing GCV with respect to τ is equivalent to minimizing with respect to p, the number of eigenvalues ranked in decreasing order. The higher the number of eigenvectors kept, the closer we are to the sample based Markowitz portfolio. For values of τ lower than the smallest eigenvalue, the SC portfolio is identical to the classical sample-based portfolio.

The Landweber-Fridman regularization technique can be implemented in two equivalent ways. Either we perform a certain number l of iterations or we transform the eigenvalues using the function $q(\frac{1}{l}, \lambda)$. Consequently, a larger number of iterations corresponds to smaller value the penalty term τ that belongs to the interval $]0, 1[$. Besides, for a large number of iterations ($\tau \approx 0$) the regularized portfolio $\hat{x}_\tau$ obtained becomes very close to the sample-based optimal portfolio $\hat{x}$. In the Landweber-Fridman case we seek the optimal number of iterations so that $\hat{x}_\tau$ is the closest to the theoretically optimal rule x^*. In the ill-posed case we typically have a very few number of iterations which corresponds to a value of τ close to one. That is, $\hat{x}_\tau$ is far from the Markowitz allocation $\hat{x}$ known to perform very poorly.

In the context of Lasso regularization, the effect of tuning the penalty τ in the l_1-penalized regression has been extensively studied by Brodie et al.

(2009). Our approach is different in the fact that we are interested in the rule that maximizes the expected out-of-sample utility of a mean-variance investor. An additional distinction is that our rules are function of the parameter $\hat{\beta}_\tau$ derived using the unconstrained version of the Homotopy/Lars Algorithm (see Appendix A.2 for a detailed description). For a given value of the penalty term, the algorithm determine the number of assets (from 1 to N) to be included in the portfolio as well as the weights associated up to a normalization.

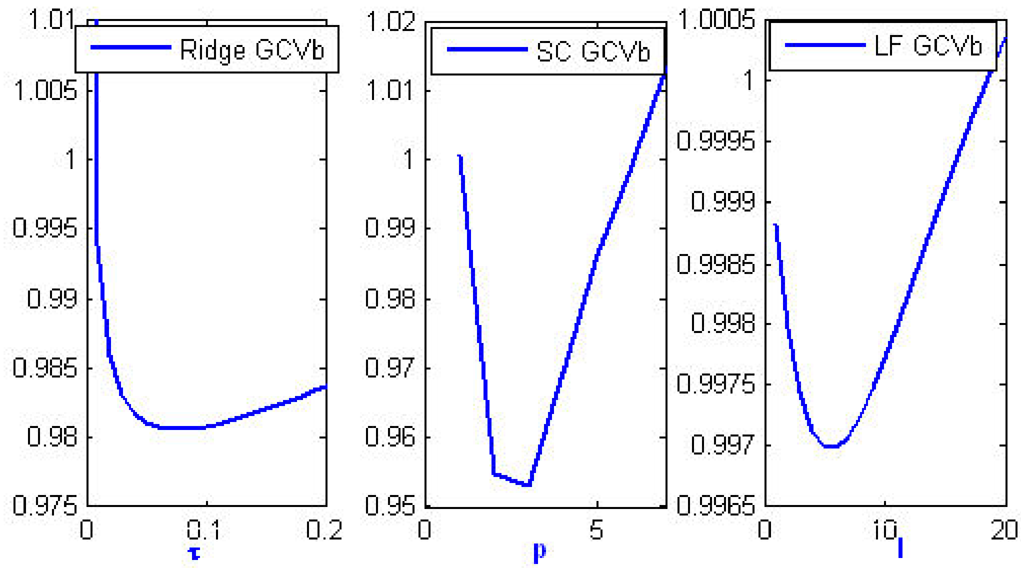
The effect of tuning τ can also be captured by the shape of the bias corrected version of the GCV criterion, GCV_b, plotted in Figure 1.1 for a single sample with $N = 100$ and $T = 120$. In our computations, the GCV_b for the Rdg the SC, and the LF portfolios are minimized with respect to τ, the number p of eigenvectors kept in the spectral decomposition of returns covariance matrix, and the number of iterations l, respectively. For all the regularization schemes, the function GCV_b have a convex shape which is a particularly interesting feature since it guarantees the unicity of the optimal parameter τ. Another interesting pattern of the GCV_b is that its curve gets steeper and gives higher values for the parameters corresponding to the sample-based investment rule : ridge penalty term close to 0, large number of eigenvectors kept for SC or large number of iterations for LF. This suggests that the performance of the regularized rules are always improved relative to the sample-based rule.

1.6.3 In-sample performance

We perform 1000 replications. In each of the replications, model (1.16) is used along with the parameters in Table 1.1 to generate $T = 120$ monthly observations of asset excess returns. We consider four different values for the number of assets traded, namely $N \in \{20, 40, 60, 100\}$. These values correspond to a ill-posed case with a large number of assets and a number of observations relatively small. The case $N = 100$ is the worse, while the other less ill-posed cases give us insights about how our method perform in general. Indeed, Table 1.9 displays some characteristics of the minimum and the maximum eigenvalues of the sample covariance matrix over replications. The smallest eigenvalue $\lambda_{\min}$ is typically very small (of the order 10^{-5}) relative to the largest eigenvalue $\lambda_{\max}$, due to the factor structure of the generating model. The ill-posedness is better measured by the condition number, $\lambda_{\max}/\lambda_{\min}$, and the relative condition number defined as the ratio of the empirical

FIGURE 1.1 – Shape of GCVs for Rdg, SC, and LF

The Figure displays GCV_b as function of regularization parameters for the Ridge, the SC, and the LF schemes. We consider a single sample with N=100 assets and T=120 observations of asset returns generated by the three-factor model described in Section 1.6.1. For this particular sample, minima are obtained for $\tau = 0.08, p = 3, l = 6$



condition number to the theoretical condition number. The bigger the condition number, the more ill-posed the case. Table 1.9 shows the evolution of the ill-posedness as the number N of assets increases. In effect, the empirical condition number goes from 0.89 to 12.3 times the value of the theoretical condition number.

We compare actual loss in utility for different levels of risk aversion and across the different rules listed in Table 1.2. The different degrees of risk aversion are reflected by the parameter of risk aversion γ chosen in $\{1, 3, 5\}$. In addition, we also consider the actual Sharpe ratio $SR(\hat{x}_\tau) = \frac{\hat{x}_\tau' \mu}{\sqrt{\hat{x}_\tau \Sigma \hat{x}_\tau}}$ as a performance criterion to compare strategies. This is particularly relevant since most investors are interested by the reward to the risk they take by investing in risky assets. The results on empirical Sharpe ratio $SR_T(\hat{x}_\tau) = \frac{\hat{x}_\tau' \hat{\mu}}{\sqrt{\hat{x}_\tau' \hat{\Sigma} \hat{x}_\tau}}$ are not reported here because they are not reliable. Essentially, they correspond to values which are overly optimistic and then are subject to the criticisms made by Khan and Smith (2008) and Fan et al. (2009) concerning empirical risk : the empirical risk is under evaluated when no constraints are imposed on the portfolio.

We report descriptive statistics on actual loss in utility and on actual Sharpe ratio across replications for $N = 20, 40, 60$, and 100 respectively in Tables 1.3, 1.4, 1.5, and 1.6. Smaller loss in utility and larger Sharpe ratio correspond to higher portfolio performance. The bias corresponding to the actual SR uses the theoretical Sharpe ratio given by $SR\left(x^{*}\right) = \sqrt{\mu'\Sigma^{-1}\mu} \equiv SR^{*}$. It appears that the Markowitz sample-based strategy leads to substantial loss in utility and, as stressed in the literature, does not provide the best actual Sharpe ratio. However, using the Rdg, SC or LF optimal parameter lead to significant improvement in terms of actual loss in utility and actual Sharpe ratio. In almost all the cases the regularized rules outperform the 1/N with respect to these criteria except the SC for $N = 40$, where the expected loss is 0.0188 for 1oN and $0.0176, 0.019$ and 0.0165 for Rdg, SC, and LF, respectively. An interesting point is that the good behavior of the regularized rules seems not to depend on the number of assets considered. Meaning that they can be expected to work well in-sample, irrespective of the degree of ill-posedness of the underlying inverse problem. Large numbers of assets give better results, which correspond to the case where a treatment is necessary. Indeed, on the one hand for $N = 20$ the higher average loss in utility for regularized rules is 0.0123 and is 0.0178 for 1oN ; for $N = 40$ the higher loss in utility is 0.019 compared to 0.0188 for 1oN. On the other hand, for $N = 60$ the worst performance is 0.0077 compared to 0.0194 for 1oN ; for $N = 100$ the higher loss is obtained is 0.007 compared to 0.0202 for the naive rule. The simulations then reveal that the regularized rules proposed outperform the 1oN by a larger margin for larger number of assets. The performance of the rules Rdg, SC, and LF are confirmed by the actual Sharpe ratio for which we obtain similar results.

Concerning the Lasso, for all the value for N and γ, the regularized portfolios obtained performs better than the sample-based Markowitz portfolio but is still far from what is theoretically optimal. The adaptation of GCV criterion proposed by Tibshirani (1996) does not provide a good approximation to the Lasso penalty term that minimizes the expected loss in utility, in presence of a large number of assets relative to the sample size. These results can be explained by the fact that our procedure usually selects a large portion of the available assets, as it appears in Table 1.7, so that the instability of the inverse problem remains unsolved and the performance of the resulting sparse portfolio deteriorates.

1.6.4 Monte Carlo assessment of GCV_b

A question that we seek to answer through simulations is whether the corrected version of generalized cross-validation criterion (GCV_b) provides a good approximation to the theoretically optimal τ that minimizes the expected loss in utility. To address this issue, we use the 1000 samples generated in the previous section ($T = 120$ and $N \in \{20, 40, 60, 100\}$). For each of the samples, we compute the GCV_b as a function of τ and determine its minimizer $\hat{\tau}$. We provide some statistics for $\hat{\tau}$ in Table 1.8.

To compute the MSE of $\hat{\tau}$, we need to derive the true optimal regularization parameter τ_0. To do so, we use our 1000 samples to approximate τ_0 as the minimizer $\hat{\tau}_0$ of the sample counterpart of the expected loss in utility corresponding to the use of the regularized rule $\widehat{x}(\tau) = \hat{\Sigma}_\tau^{-1}\hat{\mu}/\gamma$:

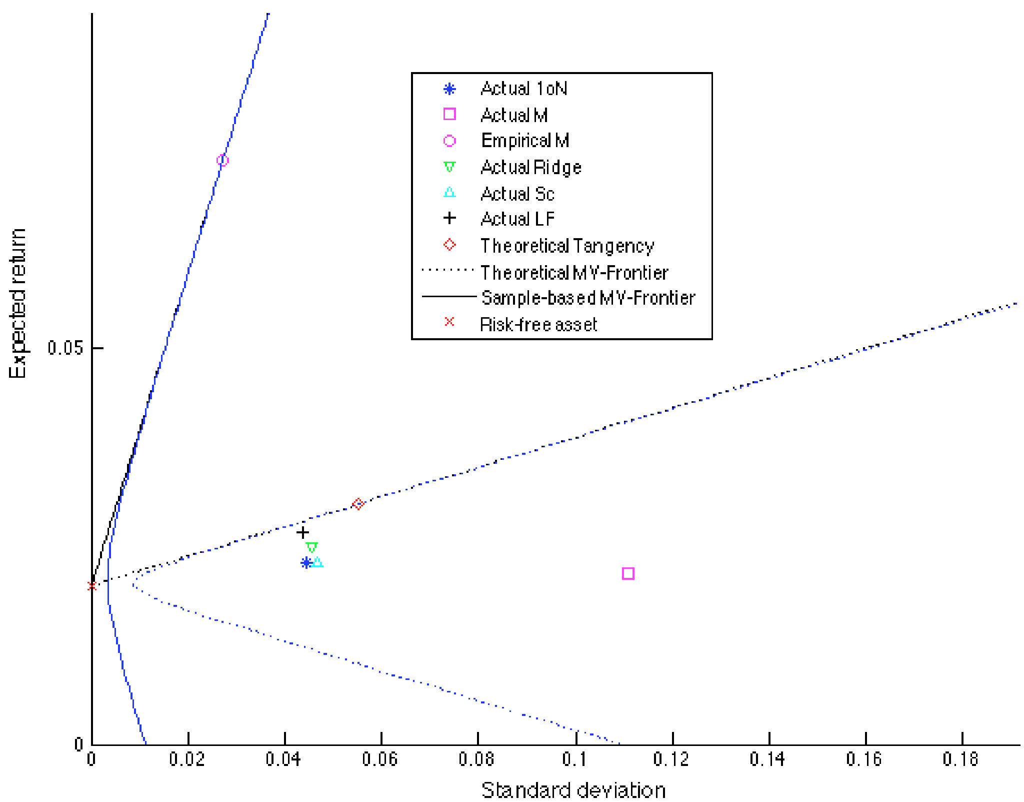
$$\hat{E}\left[(\widehat{x}(\tau) - x^*)' \, \Sigma \, (\widehat{x}(\tau) - x^*)\right]$$

where $\hat{E}$ is an average over the 1000 replications, Σ the theoretical covariance matrix, $\hat{\Sigma}_\tau^{-1}$ the regularized inverse to the sample covariance matrix and $x^* = \Sigma^{-1}\mu/\gamma$ the theoretical optimal allocation. This first step provides us with an estimation of the true parameter which is a function of the number of assets N and the sample size T under consideration and does not depend on γ. Simulations reported in Table 1.8 show that the minimizers of GCV_b are relatively good approximations to the parameters that minimize the expected loss in utility of the mean-variance strategy. For each regularization scheme the true optimal parameter is approximated by the value that minimizes the sample counterpart of the expected loss in utility $\hat{E}$.

In general, regularization parameters have a relatively high volatility across replications especially for the LF. As it appears in the Table 1.8, this is mainly due to the presence of outliers in the tails. In contrast, the value provided as a minimizer of the GCV_b are relatively accurate for the Rdg and SC in the sense that they are relatively close to the estimations of their theoretical value. A very intuitive fact concerning the ridge penalty term is that it increases with the number of assets in the portfolio, reflecting that the penalty intensity increase with the degree of ill-posedness. In the SC case, the GCV_b criterion selects on average value of p close to 3, the number of factors used, which is also the minimizer of the expected loss in utility in most cases.

To visualize the effectiveness of the rule obtained from the regularization parameter that minimizes GCV_b, we plot the corresponding Rdg, SC, and LF strategies in the mean-variance plane. All the rules are computed on a single sample of 100 asset returns and $T = 120$ monthly returns. A comparison is made with three benchmarks : the M, 1oN, and the theoretical tangency portfolio. Figure 1.2 shows how estimation errors can affect the optimality of the mean-variance optimizer. We can see a huge discrepancy between the theoretically optimal mean-variance and the sample-based optimal Markowitz portfolio. However, each of our regularized strategies get closer to the theoretically optimal rule. The main message from Figure 1.2 is that regularization reduces the distance between the sample-based tangency portfolio and the theoretical tangency portfolio.

FIGURE 1.2 – Effect of regularization on the mean-variance frontier
The figure displays regularized portfolios in Table 1. We consider a sample of 120 observations and 100 asset returns generated using the three-factor model in Equation (1.16). Optimal parameters are obtained by minimizing the GCV_b. The sample-based MV portfolio is overly optimistic, while its actual performance is the worse.



1.7 Empirical Application

In this section, we adopt the rolling sample approach used in MacKinlay and Pastor (2000) and in Kan and Smith (2008). Given a dataset of size T and a window size M, we obtain a set of $T - M$ out-of-sample returns, each generated recursively using the M previous returns. For each rolling window, the portfolio constructed are held for one year (Brodie et al. (2009)) which in our computations leads essentially to the same results as when holding optimal rules for one month (DeMiguel et al. (2007, 2009)). The time series of out-of-sample returns obtained can then be used to compute out-of-sample performance for each strategy. As pointed out by Brodie et al. (2009), this approach can be seen as an investment exercise to evaluate the effectiveness of an investor who bases his strategy on the M last periods returns. For each estimation window, we minimize the GCV_b criterion to determine the optimal tuning parameter for the ridge, the spectral cut-off and the Landweber-Fridman. The investment rules listed in Table 1.2 are then compared with respect to their out-of-sample Sharpe ratios for sub-periods extending over M years. The obtained values reflect the performance in terms of the reward to risk an investor would have if he were trading over the considered period.

We apply our methodology to two sets of portfolios from French web site : the 48 industry portfolios (FF48) and 100 portfolios formed on size and book-to-market (FF100), ranging respectively from July 1969 to June 2009 and from July 1963 to June 2009. Following our methodology for FF48, the optimal portfolios listed in Table 1.2 are constructed at the end of June every year from 1974 to 2009 for a rolling window of size $M = 60$ months and from 1979 to 2009 for $M = 120$. The risk-free rate is taken to be the one-month T-bill rate. Given the estimation windows considered, the portfolio construction problem can be considered as ill-posed for the two datasets as reflected by the condition numbers in Table 1.10. The rolling sample covariance matrices, tend to have very small eigenvalues and large condition numbers, and the situation is worse for the cases where the magnitude of N is of a comparable order of magnitude as the estimation window M.

We use an appropriate range for τ to carry out our optimizations depending on the regularization scheme. For the Ridge we use a grid on $[0, 1]$ with a precision of 10^{-2}, for the SC the parameter p ranges from 1 to $p_{max} = N-1$ while for LF we use a maximal number of iterations equal to $l_{\max} = 300$. For FF48 and $M = 60$ our first portfolios are constructed in June 1974. From the $T \times N$ matrix of excess returns R, empirical mean $\hat{\mu}$, regularized inverse $\widehat{\Omega}_{\tau}^{-1}$ and $\widehat{\beta}_{\tau}$ are computed using historical returns from July 1969 to June 1974. We then deduce the optimal portfolio corresponding to each type of regularization as a function of the minimizer of the corrected version of the *GCV* criterion. The portfolio obtained is kept from July 1974 to June 1975 and its returns recorded. We repeat the same process using data from July 1970 to June 1975 to predict portfolio return from July 1975 to June 1976. The simulated investment exercise is done recursively until the last set of portfolio constructed at the end of June 2009. The different steps are essentially the same for the FF100 except that we only consider a rolling window of $M = 120$, so that the first portfolio is constructed in June 1973.

Panel A, B, and C of Table 1.11 indicate that regularized portfolios are a more stable alternative to the Markowitz sample-based portfolio. Compared to the naive strategy, the obtained results are relatively good. The performance of the proposed rules are better in ill-posed cases, ($N = 48$, $M = 60$ and $N = 100$, $M = 120$) where the best performance of our regularized rule is higher than the performance the out-of-sample Sharpe ratio provided by the 1/N rule in all the break-out periods and in the whole period of study.

The Lasso strategy using the approximated GCV_b does not provide considerable improvement upon the Markowitz portfolio as noticed in the in-sample simulation exercise. Efforts remain to be done to address the unsatisfactory performances of the GCV-based Lasso portfolio. Among possible alternatives, a promising approach is a regularized version of the post-Lasso introduced by Belloni and Chernozhukov (2013). This approach is a two-stage procedure that consists in applying the Rdg, the SC or the LF scheme, with optimal regularization parameters, [2] to subsets of assets selected by the Lasso. The open question is how to select, ex-ante, the optimal subset size that leads to the maximum out-of-sample performance. Table 1.12 shows the maximum level of out-of-sample attainable by tuning the number of assets to be kept, after the optimal regularized rules have been used. It supports the fact that the two-stage procedure offers additional room to improvement. We plan to investigate this approach in future work.

2. Optimal parameters are obtained by minimizing the GCV_b for each optimal subset containing a given number of assets.

1.8 Conclusion

In this chapter, we address the issue of estimation error in the framework of the mean-variance analysis. We propose to regularize the portfolio choice problem 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 implied expected loss in utility of a mean-variance investor. We show that this is equivalent to select the penalty term as the minimizer of a bias-corrected version of the generalized cross validation criterion.

To evaluate the effectiveness of our regularized rules, we ran some simulations using a three-factor model calibrated to real data and an empirical study using French's 48 industry portfolios and 100 portfolios formed on size and book-to-market. The rules are essentially compared with respect to their expected loss in utility and Sharpe ratios. The main finding is that in ill-posed cases a regularization to covariance matrix drastically improves the performance of mean-variance problem, very often provides better results than the existing asset allocation strategies and outperforms the naive portfolio especially in ill-posed cases.

Our methodology can be used as well for any investment rule that requires an estimate of the covariance matrix and given a performance criterion. The appeal of the investment rules we propose is that they are easy to implement and constitute a valid alternative to the existing rules in ill-posed cases, as demonstrated by our simulations.

Table 1.3 – Statistics on actual loss in utility and actual Sharpe ratio from optimal strategies for $N = 20$ and $T = 120$

The table displays distribution characteristics of actual loss in utility and actual Sharpe ratio from optimal strategies using a three-factor model for $N = 20$ and $T = 120$ over 1000 replications. The risk aversion parameter γ is in $\{1, 3, 5\}$

Actual Loss in utility for Optimal strategies							
γ	Statistics	1oN	M	L	Rdg	SC	LF
	Mean	0.0178	0.1516	0.0256	0.0072	0.0123	0.008
	Std	0	0.0706	0.0705	0.0262	0.0143	0.009
$\gamma = 1$	q1	0.0178	0.1036	0.0056	0.0029	0.0058	0.004
	median	0.0178	0.1372	0.009	0.0037	0.0069	0.004
	q3	0.0178	0.1835	0.0155	0.0058	0.0121	0.007
	Mean	0.0054	0.0505	0.0085	0.0024	0.0041	0.003
	Std	0	0.0235	0.0235	0.0087	0.0048	0.003
$\gamma = 3$	q1	0.0054	0.0345	0.0019	0.001	0.0019	0.001
	median	0.0054	0.0457	0.003	0.0012	0.0023	0.001
	q3	0.0054	0.0612	0.0052	0.0019	0.004	0.002
	Mean	0.0032	0.0303	0.0051	0.0014	0.0025	0.002
	Std	0	0.0141	0.0141	0.0052	0.0029	0.002
$\gamma = 5$	q1	0.0032	0.0207	0.0011	0.0006	0.0012	7E-04
	median	0.0032	0.0274	0.0018	0.0007	0.0014	8E-04
	q3	0.0032	0.0367	0.0031	0.0012	0.0024	0.001
Actual Sharpe ratio for Optimal strategies							
	Bias	-0.0268	-0.1	-0.065	-0.0193	-0.0264	-0.0207
	Std	0	0.0537	0.0422	0.0093	0.0156	0.0095
	rmse	0.0268	0.1136	0.0776	0.0214	0.0307	0.0227
	q1	0.1533	0.0796	0.1037	0.1587	0.153	0.1579
	median	0.1533	0.0942	0.12	0.1625	0.1535	0.1601
	q3	0.1533	0.107	0.137	0.166	0.1645	0.1634

Table 1.4 – Statistics on actual loss in utility and actual Sharpe ratio from optimal strategies for $N = 40$ and $T = 120$

Distribution characteristics of actual loss in utility and actual Sharpe ratio from optimal strategies using a three-factor model for $N = 40$ and $T = 120$ over 1000 replications. The risk aversion parameter γ is in $\{1, 3, 5\}$

Actual Loss in utility for Optimal strategies							
γ	Statistics	1oN	M	L	Rdg	SC	LF
	Mean	0.0188	0.6857	0.1342	0.0176	0.019	0.0165
	Std	0	0.2806	0.9555	0.0154	0.007	0.0078
$\gamma = 1$	q1	0.0188	0.4844	0.0137	0.0075	0.0147	0.0105
	median	0.0188	0.6295	0.0247	0.0107	0.0167	0.0135
	q3	0.0188	0.8348	0.0544	0.0211	0.0209	0.0208
	Mean	0.0057	0.2286	0.0447	0.0059	0.0063	0.0055
	Std	0	0.0935	0.3185	0.0051	0.0023	0.0026
$\gamma = 3$	q1	0.0057	0.1615	0.0046	0.0025	0.0049	0.0035
	median	0.0057	0.2098	0.0082	0.0036	0.0056	0.0045
	q3	0.0057	0.2783	0.0181	0.007	0.007	0.0069
	Mean	0.0034	0.1371	0.0268	0.0035	0.0038	0.0033
	Std	0	0.0561	0.1911	0.0031	0.0014	0.0016
$\gamma = 5$	q1	0.0034	0.0969	0.0027	0.0015	0.0029	0.0021
	median	0.0034	0.1259	0.0049	0.0021	0.0033	0.0027
	q3	0.0034	0.167	0.0109	0.0042	0.0042	0.0042
Actual Sharpe ratio for Optimal strategies							
	Bias	-0.0286	-0.1165	-0.0645	-0.0216	-0.0212	-0.0205
	Std	0.0000	0.0486	0.0312	0.0067	0.0122	0.0053
	rmse	0.0286	0.1262	0.0717	0.0226	0.0245	0.0211
	q1	0.1556	0.0693	0.1095	0.1603	0.1538	0.1584
	median	0.1556	0.0823	0.1209	0.1637	0.1641	0.1642
	q3	0.1556	0.0918	0.1339	0.1668	0.1752	0.1682

TABLE 1.5 – Statistics on actual loss in utility and actual Sharpe ratio from optimal strategies for $N = 60$ and $T = 120$

Distribution characteristics of actual loss in utility and actual Sharpe ratio from optimal strategies using a three-factor model for $N = 60$ and $T = 120$ over 1000 replications. The risk aversion parameter γ is in $\{1, 3, 5\}$

Actual Loss in utility for Optimal strategies							
γ	Statistics	1oN	M	L	Rdg	SC	LF
	Mean	0.0194	2.3523	0.3576	0.0045	0.0077	0.0057
	Std	0	1.0135	1.1909	0.0007	0.0029	0.001
$\gamma = 1$	q1	0.0194	1.6382	0.0068	0.0039	0.0067	0.0053
	median	0.0194	2.1435	0.0166	0.0045	0.007	0.0056
	q3	0.0194	2.8687	0.0729	0.005	0.0075	0.0058
	Mean	0.0058	0.7841	0.1192	0.0015	0.0026	0.0019
	Std	0	0.3378	0.397	0.0002	0.001	0.0003
$\gamma = 3$	q1	0.0058	0.5461	0.0023	0.0013	0.0022	0.0018
	median	0.0058	0.7145	0.0055	0.0015	0.0023	0.0019
	q3	0.0058	0.9562	0.0243	0.0017	0.0025	0.0019
	Mean	0.0035	0.4705	0.0715	0.0009	0.0015	0.0011
	Std	0	0.2027	0.2382	0.0001	0.0006	0.0002
$\gamma = 5$	q1	0.0035	0.3276	0.0014	0.0008	0.0013	0.0011
	median	0.0035	0.4287	0.0033	0.0009	0.0014	0.0011
	q3	0.0035	0.5737	0.0146	0.001	0.0015	0.0012
Actual Sharpe ratio for Optimal strategies							
	Bias	-0.0344	-0.1764	-0.1274	-0.0240	-0.0294	-0.0259
	Std	0.0000	0.0471	0.0784	0.0048	0.0124	0.0071
	rmse	0.0344	0.1825	0.1496	0.0245	0.0319	0.0269
	q1	0.1527	-0.0384	0.0573	0.1589	0.1508	0.1549
	median	0.1527	0.0308	0.0864	0.1630	0.1510	0.1565
	q3	0.1527	0.0503	0.1089	0.1669	0.1694	0.1680

Table 1.6 – Statistics on actual loss in utility and actual Sharpe ratio from optimal strategies for $N = 100$ and $T = 120$

Distribution characteristics of actual loss in utility and actual Sharpe ratio from optimal strategies using a three-factor model for $N = 100$ and $T = 120$ over 1000 replications. The risk aversion parameter γ is in $\{1, 3, 5\}$.

Actual Loss in utility for Optimal strategies							
γ	Statistics	1oN	M	L	Rdg	SC	LF
$\gamma = 1$	Mean	0.0202	138.74	290.92	0.005	0.0038	0.0069
	Std	0	126.51	7799.7	0.0029	0.002	0.0026
	q1	0.0202	60.586	0.0065	0.0024	0.0024	0.0038
	median	0.0202	98.993	0.0246	0.0032	0.003	0.0083
	q3	0.0202	171.42	1.6148	0.0082	0.0048	0.009
$\gamma = 3$	Mean	0.0059	46.247	96.972	0.0017	0.0013	0.0023
	Std	0	42.17	2599.9	0.001	0.0007	0.0009
	q1	0.0059	20.195	0.0022	0.0008	0.0008	0.0013
	median	0.0059	32.998	0.0082	0.0011	0.001	0.0028
	q3	0.0059	57.14	0.5383	0.0027	0.0016	0.003
$\gamma = 5$	Mean	0.0036	27.748	58.183	0.001	0.0008	0.0014
	Std	0	25.302	1559.9	0.0006	0.0004	0.0005
	q1	0.0036	12.117	0.0013	0.0005	0.0005	0.0008
	median	0.0036	19.799	0.0049	0.0006	0.0006	0.0017
	q3	0.0036	34.284	0.323	0.0016	0.001	0.0018
Actual Sharpe ratio for Optimal strategies							
	Bias	-0.0355	-0.1838	-0.1358	-0.0181	-0.021	-0.0251
	Std	0	0.0241	0.0584	0.0072	0.0159	0.0064
	rmse	0.0355	0.1853	0.1478	0.0195	0.0263	0.026
	q1	0.1532	-0.0127	0.0386	0.1626	0.1525	0.1584
	median	0.1532	0.0057	0.0618	0.1737	0.1744	0.1593
	q3	0.1532	0.0228	0.0833	0.177	0.1763	0.1709

Table 1.7 – Statistics on the number of assets kept by the Lasso

The figure shows the distribution characteristics of the optimal number of assets selected by the Lasso procedure in the in-sample study

	Number of assets			
Statistics	20	40	60	100
Mean	18.20	36.94	55.26	88.90
Std	1.61	1.60	2.36	3.89
q1	18	36	54	86
median	18	37	55	89
q3	19	38	57	91

Table 1.8 – Optimal regularization parameters for SC, LF, and Rdg

Distribution characteristics of the optimal regularization parameters. The number of iterations performed is 1000 for $T = 120$ monthly observations. We consider N assets, $N \in \{20, 40, 60, 100\}$.

Rule	Rdg (τ)				SC(p)				LF(l)			
N	20	40	60	100	20	40	60	100	20	40	60	100
$\hat{\tau}_0$	0.006	0.018	0.01	0.05	2	3	3	3	14	8	149	32
Mean	0.008	0.007	0.023	0.072	2.42	2.93	2.3	2.23	68.48	129.6	113	24.18
Std	0.007	0.006	0.016	0.073	1.94	1.95	1.51	1.17	82.35	105.3	107	20.46
rmse	0.007	0.012	0.021	0.076	1.98	1.95	1.66	1.4	98.7	160.8	113	32.94
q1	0.002	0.002	0.01	0.01	1	1	1	2	12	22	15	8
median	0.006	0.005	0.01	0.02	2	3	1	2	16	110	17	9
q3	0.013	0.012	0.04	0.14	3	3	3	2	103	220	251	50

TABLE 1.9 – Statistical properties of the sample covariance matrices eigenvalues

Panel A displays statistical properties of the maximum eigenvalue and the minimum eigenvalue derived from the sample covariance matrices in the in-sample study. Panel B shows the distribution characteristics of the condition number of the sample covariance matrices obtained over replications in the in-sample study.

Panel A : Eigenvalues								
	$\hat{\lambda}_{\min}$				$\hat{\lambda}_{\max}$			
N	20	40	60	100	20	40	60	100
λ	1.52E-05	2.79E-05	7.83E-06	1.53E-05	0.0346	0.0708	0.1039	0.19
Mean	1.26E-05	1.77E-05	3.80E-06	1.50E-06	0.025	0.061	0.1166	0.211
Std	1.78E-06	2.75E-06	6.95E-07	3.97E-07	0.0006	0.001	0.0013	0.002
q1	1.13E-05	1.58E-05	3.30E-06	1.23E-06	0.0245	0.0604	0.1157	0.209
median	1.24E-05	1.77E-05	3.76E-06	1.47E-06	0.025	0.0611	0.1166	0.211
q3	1.37E-05	1.95E-05	4.26E-06	1.75E-06	0.0254	0.0617	0.1175	0.212
Panel B : Condition numbers								
	$\hat{\lambda}_{\max}/\hat{\lambda}_{\min}$				$(\hat{\lambda}_{\max}/\hat{\lambda}_{\min})/(\lambda_{\max}/\lambda_{\min})$			
N	20	40	60	100	20	40	60	100
Mean	2027.9497	3527.8	31767.5	152232	0.8904	1.3879	2.392	12.3
Std	297.58027	573.51	6058.67	47231.5	0.1307	0.2256	0.4562	3.816
q1	1820.7269	3121.4	27225.8	120726	0.7994	1.228	2.05	9.753
median	2008.3814	3454.6	31054.8	143773	0.8818	1.3591	2.3384	11.61
q3	2198.8976	3875.3	35266.7	171208	0.9654	1.5246	2.6555	13.83

TABLE 1.10 – Statistics on the eigenvalues and condition numbers of the rolling window sample covariance matrices.

Statistics on the eigenvalues and condition numbers of the rolling window sample covariance matrices. The estimation window considered are M= 60, 120, and the datasets F48 and FF100.

Data set	M	Statistics	λ_{min}	λ_{max}	$\lambda_{max}/\lambda_{min}$
		Mean	8.93E-06	0.1197	1.77E+04
		Std	4.08E-06	0.0443	1.56E+04
FF48	60	q1	6.36E-06	0.0870	8.82E+03
		median	8.05E-06	0.1157	1.37E+04
		q3	1.14E-05	0.1557	2.14E+04
		Mean	8.58E-05	0.1160	1.68E+03
		Std	3.66E-05	0.0293	8.79E+02
FF48	120	q1	5.76E-05	0.0940	7.38E+02
		median	7.35E-05	0.1136	1.61E+03
		q3	0.0001237	0.1399	2.49E+03
		Mean	5.23E-06	0.2697	5.74E+04
		Std	1.52E-06	0.0645	2.78E+04
FF100	120	q1	4.14E-06	0.2265	3.79E+04
		median	5.07E-06	0.2527	5.14E+04
		q3	6.17E-06	0.3301	6.73E+04

Table 1.11 – Out-of-sample Sharpe ratio for regularized rules applied to industry portfolios

Out-of-sample performance in terms of Sharpe ratio of the optimal rules applied to FF48 and FF100 using a rolling window of length $M = \{60, 120\}$. The optimal regularized rules are obtained using the regularization parameters that minimizes the GCV.

Panel A : Out-of-sample SR for FF48, and M=60months						
Period	1oN	M	L	Rdg	Sc	LF
07/74 - 06/79	0.1334	0.0797	0.1252	0.1243	0.3006	0.2620
07/79 - 06/84	0.1070	-0.2879	0.0004	0.0874	-0.0186	0.1227
07/84 - 06/89	0.1823	0.1302	0.1895	0.2974	0.3131	0.2158
07/89 - 06/94	0.1077	0.0332	-0.1140	0.0490	0.2856	0.1526
07/94 - 06/99	0.2910	0.0433	-0.2273	0.3370	0.2954	0.3379
07/99 - 06/04	0.0767	0.0375	-0.0202	0.0602	0.0673	0.0699
07/04 - 06/09	-0.0119	-0.2494	-0.0294	-0.0212	-0.0987	-0.0111
07/74-06/09	0.1266	-0.0305	-0.0108	0.1334	0.1635	0.1642

Panel B : Out-of-sample SR for FF48, and M=120months						
Period	1oN	M	L	Rdg	Sc	LF
07/79 - 06/89	0.1468	-0.0183	0.0465	0.1311	0.1282	0.1247
07/89 - 06/99	0.2002	-0.0568	-0.0656	0.2427	0.1664	0.2373
07/99 - 06/09	0.0269	-0.0485	-0.0462	-0.0555	-0.0992	-0.0465
07/79 - 06/09	0.1246	-0.0412	-0.0218	0.1061	0.0651	0.1052

Panel C : Out-of-sample SR for FF100, and M=120months						
Period	1oN	M	L	Rdg	Sc	LF
07/79 - 06/89	0.1468	-0.0183	0.0465	0.1311	0.1282	0.1247
07/89 - 06/99	0.2002	-0.0568	-0.0656	0.2427	0.1664	0.2373
07/99 - 06/09	0.0269	-0.0485	-0.0462	-0.0555	-0.0992	-0.0465
07/79 - 06/09	0.1246	-0.0412	-0.0218	0.1061	0.0651	0.1052

Table 1.12 – Out-of-sample performance in terms of Sharpe ratio for optimal rules combined with the Lasso and applied to FF48

Out-of-sample performance in terms of Sharpe ratio for optimal rules applied to FF48 using a rolling window length of 60 months. The optimal number of assets over different subsets of assets obtained using the Lasso as a first step. The combined rules use the Lasso a first step procedure and are obtained by minimizing GCV over different sets of assets.

Period	**1oN**	**M**	**L**	**Rdg**	**Sc**	**LF**	**L-Rdg**	**L-Sc**	**L-LF**
07/74 - 06/79	0.1334	0.0797	0.1252	0.1287	0.3006	0.2620	0.3576	0.3506	0.3624
07/79 - 06/84	0.1070	-0.2879	0.0004	0.0879	-0.0186	0.1227	0.4541	0.3826	0.1653
07/84 - 06/89	0.1823	0.1302	0.1895	0.2979	0.3131	0.2158	0.4590	0.4937	0.3763
07/89 - 06/94	0.1077	0.0332	-0.1140	0.0490	0.2856	0.1526	0.2961	0.3551	0.3556
07/94 - 06/99	0.2910	0.0433	-0.2273	0.3372	0.2954	0.3379	0.4847	0.5193	0.4862
07/99 - 06/04	0.0767	0.0375	-0.0202	0.0610	0.0673	0.0699	0.2258	0.2723	0.2542
07/04 - 06/09	-0.0119	-0.2494	-0.0294	-0.0206	-0.0987	-0.0111	0.2369	0.3200	0.3568
07/69 - 06/03	0.1266	-0.0305	-0.0108	0.1344	0.1635	0.1642	0.3592	0.3848	0.3367

Bibliographie

[1] Andrews, D., 1991. "Asymptotic optimality of generalized C_L, cross-validation, and generalized cross-validation in regression with heteroskedastic errors," *Journal of Econometrics*, 47, 359-377.

[2] Bai, J. and S. Ng, 2008. "Forecasting economic time series using targeted predictors," *Journal of Econometrics*, 146, 304-317.

[3] Belloni, A., Chernozhukov, V., 2013. "Least squares after model selection in high-dimensional sparse models," *Bernoulli*, 19, 521-547.

[4] Blundell, R., X. Chen, and D. Kristensen, 2007. "Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves," *Econometrica*, 75, 1613-1669.

[5] Brandt, Michael W., 2010. "Portfolio choice problems," *Handbook of Financial Econometrics*, 1, 269-336.

[6] Brandt, M. W., P. Santa-Clara, and R. Valkanov, 2009. "Parametric Portfolio Policies : Exploiting Characteristics in the Cross-Section of Equity Returns," *The review of Financial Studies*, 22, 3411-3447.

[7] Britten-Jones, M., 1999. "The sampling error in estimates of mean-variance efficient portfolio weights," *Journal of Finance,* 54, 655-671.

[8] Brodie, J., I. Daubechies, C. De Mol, D. Giannone, I. Loris, 2009. "Sparse and stable Markowitz portfolios,"Proceedings of the National Academy of Sciences of the USA, 106, 12267-12272.

[9] Carrasco , M., 2012. "A regularization approach to the many instrument problem," *Journal of Econometrics*, 170, 383-398.

[10] Carrasco, M., J-P. Florens, 2000. "Generalization of GMM to a Continuum of Moment Conditions," *Econometric Theory*, 16, 797-834.

[11] Carrasco, M., J-P. Florens, E. Renault, 2007. "Linear Inverse Problems and Structural Econometrics : Estimation Based on Spectral Decomposition and Regularization," *Handbook of Econometrics*, 6, 5633-5751.

[12] Craven, P. and G. Wahba, 1979. "Smoothing noisy data with spline functions : Estimating the correct degree of smoothing by the method of the generalized cross-validation," *Numer. Math.*, 31, 377-403.

[13] DeMiguel, V., L. Garlappi, F. J. Nogales, R. Uppal, 2009. "A Generalized Approach to Portfolio Optimization : Improving Performance by Constraining Portfolio Norms," *Management Science*, 55, 798-812.

[14] DeMiguel, V., L. Garlappi, R. Uppal, 2007. "Optimal Versus Naive Diversification : How Inefficient is the 1/N Portfolio Strategy ?," *The review of Financial studies*, 22, 1915-1953.

[15] De Mol, C., D. Giannone, and L. Reichlin, 2008. "Forecasting using a large number of predictors : Is Bayesian shrinkage a valid alternative to principal components ?," *Journal of Econometrics*, 146, 318-328.

[16] Fan, J., Y. Fan and J. Lv, 2008. "High dimensional covariance matrix estimation using a factor model," *Journal of Econometrics*, 147, 186-197.

[17] Fan, J., Zhang, J., Yu, K., 2012. "Vast portfolio selection with gross-exposure constraints," *Journal of the American Statistical Association*, 107, 592-606.

[18] Frost, A. and E. Savarino, 1986. "An Empirical Bayes Approach to Efficient Portfolio Selection," *Journal of Financial and Quantitative Analysis*, 21, 293-305.

[19] Hoerl, A.E. and R.W. Kennard, 1970. "Ridge regression : biased estimation for nonorthogonal problems," *Technometrics*, 12, 55-67.

[20] Huang J., J. Horowitz, and S. Ma, 2008. "Asymptotic properties of bridge estimators in sparse high-dimensional regression models," *Annals of Statistics*, 36, 587-613.

[21] Jagannathan, R. and T. Ma, 2003. "Risk Reduction in Large Portfolios : Why Imposing the Wrong Constraints Helps," *Journal of Finance*, 58, 1651-1683.

[22] Jobson, J.D. and B. Korkie, 1983. "Statistical Inference in Two-Parameter Portfolio Theory with Multiple Regression Software," *Journal of Financial and Quantitative Analysis*, 18, 189-197.

[23] Kan, R. and D. R. Smith, 2008."The distribution of the Sample Minimum-Variance Frontier," *Management Science*, 54, 1364-1380.

[24] Kan, R. and G. Zhou, 2007. "Optimal portfolio choice with parameter uncertainty," *Journal of Financial and Quantitative Analysis*, 42, 621-656.

[25] Ledoit, O. and M. Wolf, 2003. "Improved estimation of the covariance matrix of stock returns with an application to portfolio selection," *The Journal of Empirical Finance*, 10, 603-621.

[26] Ledoit, O. and M. Wolf, 2004a. "Honey, I shrunk the sample covariance matrix," *The Journal of Portfolio Management*, 30, 110-119.

[27] Ledoit, O. and M. Wolf, 2004b. "A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices," *Journal of Multivariate Analysis*, 88, 365-411.

[28] Li, K.C., 1986. "Asymptotic optimality of C_L and generalized cross-validation in ridge regression with application to spline smoothing," *The Annals of Statistics*, 14, 1101-1112.

[29] Li, K-C., 1987. "Asymptotic optimality for C_p, C_L, cross-validation and generalized cross-validation : Discrete Index Set," *The Annals of Statistics*, 15, 958-975.

[30] MacKinlay, A.C. and L. Pastor, 2000. "Asset Pricing Models : Implications for Expected Returns and Portfolio Selection," *The Review of financial studies*, 13, 883-916.

[31] Markowitz, H. M., 1952. "Portfolio selection," *The Journal of Finance*, 7, 77-91.

[32] Michaud, R.O., 1989. "The Markowitz Optimization Enigma : Is Optimal Optimized," *Financial Analyst Journal*, 45, 31-42.

[33] Stock, J. and M. Watson, 2002. "Forecasting Using Principal Components from a Large Number of Predictors," *Journal of the American Statistical Association*, 97, 1167-1179.

[34] Tibshirani, R., 1996. "Regression Shrinkage and Selection via the Lasso," *J. Roy. Statist. Soc.*, Series B, 58, 267-288.

[35] Tu, J. and G. Zhou, 2010. "Markowitz Meets Talmud : A combination of Sophisticated and Naive Diversification Strategies," *Journal of Financial Economics*, 99, 204-215.

[36] Varian, H., 1996. "A Portfolio of Nobel Laureates : Markovitz, Miller and Sharpe," *The Journal of economic Perspectives*, 7, 159-169.

[37] Whittle, P., 1960. "Bounds for the moments of linear and quadratic forms in independent variables," *Theory Probab. Appli.*, 5, 302-305.

Chapitre 2

Optimal domestic portfolios and the exchange rate risk

2.1 Introduction

Under the well recognized failure of purchasing power parity (PPP), currency risk must be a priced factor (Solnik, 1974 ; Adler and Dumas, 1983). In the U.S. equity market, this theoretical result is supported by a large body of empirical evidence, both at the country level and at the industry level (De Santis and Gerard, 1998 ; Carrieri, Errunza, and Majerbi, 2006a ; Francis, Hasan, and Hunter, 2008). While evidence on the pricing of the exchange rate risk is overwhelming, very little is said about the implications of such result for domestic investors, in particular, if currency risk could be reduced or eliminated through a given portfolio strategy over domestic assets. Indeed, the pricing of currency risk, implies that currency risk is systematic and that investors require to be rewarded by a premium for bearing it. Different portfolio strategies are likely to have different exposures to currency risk. This is because they use different weights on assets whose exposures vary across industries and over time (Francis, Hasan, and Hunter, 2008). Therefore, when exposed to common prices of risk, different strategies are also likely to lead to different premiums for currency risk.

The case for portfolios consisting exclusively of domestic assets, can be made in at least three points. First, there is a consensus among financial economists and practitioners that a country's domestic investors have an advantage in trading stocks in their country over foreign investors, for many reasons such as superior information, the bias of domestic regulator against foreign investors (Choe, Kho, and Stulz, 2005). For instance, these reasons have been used to explain the home bias puzzle, of the predominant share of domestic investments in investors portfolios, (see French and Poterba, 1991 ; Cooper and Kaplanis,1994 ; and Kang and Stulz, 1997). Second, investing abroad might be unnecessary, since in some cases, it is possible to exhaust the gains from international diversification through home-made diversification (Errunza, Hogan and Hung, 1999). Finally, although the benefits of global diversification have been extensively documented, and are still relevant, (e.g. Solnik, 1974 ; Christoffersen et al., 2012 ; Driessen and Laeven, 2007), the last decades have experienced an increased level of integration in international markets. Among other things, these trends have led to an increase in the exposures of domestic markets to foreign exchange rate movements, and to the reduction of the potential for international diversification (See e.g. Li, Sarkar, and Wang (2003) [1]). This shrinking benefit from international diversification, implies that investing domestically is becoming increasingly relevant to consider and study. Besides, the assessment of the benefits of domestic diversification pointed out in previous studies can be made more precise, by examining how optimal combinations of domestic assets are exposed to international sources of risk, in particular to the exchange rate risk.

In testing how domestic investors are impacted by currency risk, we need to specify the strategies implemented, and how optimal weights are determined. For example, when studying the impact of the European Monetary Union (EMU) and non-EMU currency risks on international portfolio choices, De Santis, Gérard, and Hillion (1999) considered dynamic asset allocation strategies, for universal investors [2]. In contrast to these authors, this chapter focuses entirely on domestic portfolios. We consider global minimum-variance portfolios, that have been extensively documented to perform better than mean-variance portfolios (DeMiguel, Garlappi, Nogales, and Uppal, 2009 ; Jagannathan and Ma, 2003). This strategy corresponds to the assumption that the investor's primary concern is to minimize the overall portfolio risk. The portfolios are constructed in the U.S. using an investment set that consists of 48 industry portfolios, and the optimal weights are obtained recursively based on M previous months to mimic the dynamic behavior of a mean-variance

1. Though international markets are becoming more integrated this does not eliminate the diversification benefits of emerging market investment.

2. For such investors, the investment universe contains stock market indices, short-term, and euro-currency deposits.

investor.

When constructing mean-variance portfolios, the inverse of the covariance matrix between all asset returns, is needed. However, the size of the investment set and estimation errors can lead to instabilities in the inverse, the optimal weights, and offset the gain from diversification as investigated in chapter 1 (e.g DeMiguel, Garlappi, and Uppal, 2009). For this reason, we also consider a family of global minimum portfolios that are regularized to deal with the instabilities arising in large portfolios. Among the available regularization techniques (see Kress, 1999 ; Carrasco, Florens, and Renault, 2007), we choose the spectral cut-off which is closely related to principal component and factor models. This technique stabilizes optimal portfolios weights by considering reduced number of the covariance matrix's principal components.

The objective of this chapter is to investigate whether optimal portfolios of domestic assets are associated with lower exchange risk exposures, and if such portfolios lead to premiums that are economically important. Since a stabilization is applied to the optimal weights considered, our methodology also allows us to investigate how stable investment strategies are impacted by currency risk. Even though our entire focus is on the global minimum portfolio strategy and its stable extensions, it is worth mentioning, that this chapter, does not promote a particular rule, or that domestic investment outperforms international investment.

Currency risk for cross-sections of asset returns, is better assessed in the context of international asset pricing models (IAPM). Thus to achieve our objective, we use the optimal portfolios constructed to estimate and test a conditional IAPM under the assumption the U.S. industries are integrated with the world stock market. The model we use is similar to the five-factor conditional IAPM of Francis, Hasan, and Hunter (2008), which is the conditional version of the three Fama and French (1993) model augmented with the changes in the trade-weighted exchange rates indices, representing industrialized economies and emerging markets currency risk. In the model we adopt, we substitute the US market index with the world equity market index, to be in accordance with our assumption that U.S. industries are integrated with the world market. The model is suited to our objective, since it has demonstrated its methodological superiority by establishing [3] that cur-

3. Thus resolving the puzzle posed by previous studies that currency risk is priced at aggregate equity market but not at the industry level.

rency risk is priced and time-varying at the industry level and is associated with an economically important premium.

Our findings can be summarized by three main points. First, we find that currency is also priced and time-varying for cross-sections of optimally diversified domestic portfolios in the US. Second, we find that portfolios optimized over domestic assets substantially reduce the average industry exposure to currency risk across U.S. industries, both in size and volatility. Third, for all the global minimum portfolios considered, the currency premium contribution to the total premium is economically meaningful, but approximately remains the same in terms of their contributions to total premium.

These results are relevant for many reasons. First, the constructed investment rules represent an alternative to the standard market capitalization weights used to test the pricing of currency risk. Therefore, our finding provide additional evidence for the pricing of currency risk in the U.S.. Besides, the reduction in exposures obtained can be seen as additional measures of the benefit of domestic diversification. Second, our results provides insights for hedging strategies. Indeed, hedging strategy requires the choice of the amount and type of the hedging instrument, which in turn depends only on the size and volatility of the exposures. Thus the fact that exposures are substantially reduced and less volatile, while the contribution to premium is roughly the same, implies investors can greatly benefit from optimal regularized portfolios, that allows them to reduce their exposures to currency risk while keeping the same level of reward. In other words, optimal domestic portfolios have the potential to reduce currency risk hedging costs, as investor reduce the amount and frequency of hedging rebalancing.

The studies closest to ours are De Santis, Gérard, and Hillion (1999) and Francis, Hasan, and Hunter (2008). Our results are consistent with the latter study, and extend the industry portfolios level framework to optimal portfolio of U.S. industries. In addition, in this chapter, we did not restrict ourselves to the 36 industries, but considers the 48 industries portfolios that provide an exhaustive coverage of the U.S. stock market. The former paper considers universal investors and decomposes currency risk into its EMU and non-EMU components, while we disaggregate currency risk in its industrialized and emerging markets components.

The rest of the chapter is organized as follows. Section 2 describes the model and methodology used to construct optimal minimum portfolios, and to measure currency exposures and premiums. Section 3 presents the data used as inputs in our estimations, and provides some statistics. Empirical evidence on the pricing of currency risk for industry portfolios and global minimum portfolios are discussed in Section 4. Section 5 performs some diagnostics and robustness tests, and finally Section 6 concludes.

2.2 Model and Methodology

2.2.1 Investment strategies

The investment universe considered consists of N industry portfolios in the U.S.. We use $r_{i,t}$ and $r_{p,t}$ to denote the excess returns (over the risk-free asset) on the i^{th} industry and the returns on a given optimal portfolios p, respectively. The N-vector r_t is the excess returns on the N industries. All returns are between time $t-1$ and time t.

The returns on optimal portfolios $r_{p,t}$ at time t, are obtained from optimal portfolio weights, w_{t-1}, constructed using observed industry returns up to period $t-1$, and the realized return at time t. These time-varying weights are obtained recursively using a rolling window of $M = 100$ monthly returns. In other words, investors first use the last M observations on industry returns, to construct the weights to be used the next period. Then investment window is rolled over one month forward for the next investment decision. In the period t, the realized return on portfolio p, using the weights w_{t-1} is :

$$r_{p,t} = w'_{t-1} r_t = \sum_{i=1}^{N} w_{i,t-1} r_{i,t} \tag{2.1}$$

The investment strategies considered are the equally weighted portfolio and the Markowitz's global minimum variance portfolio. These portfolios represent an alternative to the market portfolio that stems the equilibrium relationship underlying CAPM models. Precisely, instead of market capitalization weights, the weights that we use stems form optimization procedures. A particular case of portfolio, are single industry portfolios, considered by Francis, Hasan, and Hunter (2008) as cross section in the estimation of an IAPM. For example in the case of industry i portfolio, the weight w_{t-1} has 1 for its i^{th} component and zero everywhere else.

The 1 over N equally diversified

As described in Chapter 1, the 1/N rule consists in investing equally in each asset of given investment universe, that is $w_{i,t-1} = \frac{1}{N}$, for all asset i. This rule is not derived from a model, and then has no estimation errors. As a result, the equally weighted rule tends to perform better than most existing rules for large number of assets, and has recently been established as a benchmark investment rule by DeMiguel, Garlappi, and Uppal (2007).

The global minimum portfolio

In this chapter we focus on the minimum-variance portfolio strategy, as extensive empirical evidence shows that this rule usually performs better out-of-sample than any other mean-variance portfolio (DeMiguel, Garlappi, Nogales, and Uppal, 2009 ; Jagannathan and Ma, 2003). This strategy has the attractive feature that it does not require an estimate of the mean of asset returns, which has been documented to be difficult to estimate (See Merton, 1980). It corresponds to a risk-minimizing equity investor, whose primary concern is to minimize the overall risk of his portfolio. Let Σ_{t-1} be the sample covariance matrix based on industry returns up to $t-1$. The global minimum [4] variance portfolio weights are function of the inverse Σ_{t-1}^{-1} of Σ_{t-1} and are given by [5] :

$$\omega_{t-1} = \frac{\Sigma_{t-1}^{-1} 1_N}{1_N' \Sigma_{t-1}^{-1} 1_N} \tag{2.2}$$

where 1_N is the N-vector of ones.

The inverse of the sample covariance can exhibit explosive effects, when the investment universe is large, and lead to low levels of portfolio performance. Many solutions have been proposed to tackle this issue (See Chapter 1 for a review). Here, we also stabilize the inverse of the covariance matrix, through a statistical technique called regularization, but we only consider the spectral cut-off. Other regularizations such as the ridge and Landweber-Fridman could have been considered. However, these regularizations often lead to similar performance, when the optimal regularization parameter is

4. The strategies is unrestricted, have no position limits, nor transaction costs.
5. This analytical expression is straightforward and is obtained by solving for ω, in the optimization problem $Min\,\omega' H_{t-1}^{-1}\omega,\ st : \omega' 1_N = 1$.

selected. The spectral cut-off is closely related to principal component analysis and factor models, and has appealing statistical and economic interpretations. In the next paragraph, we recall how the spectral cut-off improves the poor performance of portfolio weights involving an estimate of the covariance matrix.

The global minimum portfolios and spectral cut-off

As a positive definite matrix, the covariance matrix Σ can be written as $P\Lambda P'$. Where Λ is the diagonal matrix with the eigenvalues $\{\lambda_j\}_{j=1}^N$ of Σ on its diagonal, arranged in decreasing order ; P is the matrix such that the j^{th} column, P_j, is the eigenvector associated with λ_j. This means that $\Sigma P_j = \lambda_j P_j$. The inverse of Σ that appears in (2.2) is given by :

$$\Sigma_{t-1}^{-1} = P \begin{pmatrix} \frac{1}{\lambda_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\lambda_2} & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \frac{1}{\lambda_N} \end{pmatrix} P' \tag{2.3}$$

Equation (2.3) shows the issue that may arise when inverting a large covariance matrix of assets returns. Indeed, in such a case the smallest eigenvalue, λ_N, tends to converge toward zero (e.g. Florens, Carrasco, and Renault, 2007), leading to unstable portfolio weights.

In general, the regularized inverse of the sample covariance matrix is obtained using function $q(\tilde{\tau}, \cdot)$ that dampens the possible explosive effects stemming from inverting eigenvalues. The dampening function $q(\tilde{\tau}, \cdot)$ is parametrized by $\tilde{\tau}$, which represents the amount of regularization introduced [6]. Precisely, the $\frac{1}{\lambda_j}$ are replaced by $\frac{q(\tilde{\tau},\lambda_j)}{\lambda_j}$ in Equation (2.3), which leads to :

$$\Sigma_{\tilde{\tau},t-1}^{-1} = P \begin{pmatrix} \frac{q(\tilde{\tau},\lambda_1)}{\lambda_1} & 0 & \cdots & 0 \\ 0 & \frac{q(\tilde{\tau},\lambda_2)}{\lambda_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \frac{q(\tilde{\tau},\lambda_N)}{\lambda_N} \end{pmatrix} P' \tag{2.4}$$

6. The dampening functions for the ridge regularization and the Landweber-Fridman regularization are $q(\tau, \lambda_j) = \frac{\lambda_j}{\lambda_j+\tau}$ and $q(\tau, \lambda_j) = 1 - (1 - \lambda_j)^{\frac{1}{\tau}}$, respectively, and are not considered in this chapter.

In the case of the spectral cut-off, stability is guaranteed by ruling out the eigenvectors associated with eigenvalues smaller than a threshold given by $\tilde{\tau}$. We have :

$$q(\tilde{\tau}, \lambda_j) = 1_{\lambda_j > \tilde{\tau}} \tag{2.5}$$

Denote by $l^{\tilde{\tau}}$ the number of eigenvalues satisfying $\lambda_j > \tilde{\tau}$. Using (2.4) and (2.5), we have :

$$\Sigma^{-1}_{\tilde{\tau},t-1} = P \begin{pmatrix} \frac{1}{\lambda_1} & & & 0 \\ & \ddots & & \\ & & \frac{1}{\lambda_l} & \\ 0 & & & 0 \end{pmatrix} P' = \sum_{j=1}^{l^{\tilde{\tau}}} \frac{1}{\lambda_j} P_j P'_j. \tag{2.6}$$

The parameter $l^{\tilde{\tau}}$ is the number of eigenvector used. Similarly, given a certain percentage τ of eigenvectors to be kept, $l^{\tilde{\tau}}$ can be replaced by l_τ, the integer part of $\tau * N$.

The corresponding regularized covariance matrix is :

$$\Sigma^{-1}_{\tau,t-1} = \sum_{j=1}^{l_\tau} \frac{1}{\lambda_j} P_j P'_j. \tag{2.7}$$

The strategy obtained by using the spectral cut-off, with l_τ principal components is :

$$\omega_{t-1}(\tau) = \frac{\Sigma^{-1}_{\tau,t-1} 1_N}{1'_N \Sigma^{-1}_{\tau,t-1} 1_N} \tag{2.8}$$

In this chapter, we interpret the weights $\omega_{t-1}(\tau)$, as being the strategy adopted by the investors who uses a certain percentage of the PCs to make a decision. This percentage also represents the percentage of total variation in the investment set, and is related to the stability of portfolio weights as measured by the turnover [7] :

7. The turnover is a measure of stability in portfolio weights, the lower the turnover the stabler the portfolio. It is equal to the sum of the absolute value of the rebalancing trades across the N available assets and over the $T - M - 1$ trading dates, normalized by the total number of trading dates.

$$TO = \frac{1}{T-M-1} \sum_{t=M}^{T-1} \sum_{j=1}^{N} (|w_{j,t+1} - w_{j,t+}|) \tag{2.9}$$

where $w_{j,t+}$ is the portfolio weight before rebalancing at $t+1$, and $w_{j,t+1}$ is the portfolio weight at time $t+1$ after rebalancing.

The lower the percentage τ, the more stable the rule, and so τ can also be interpreted as the degree of stability required by investors. Indeed, as displayed in Table 2.2, the turnover which measures the stability of portfolio weights is a decreasing function of the percentage of PCs used.

In Chapter 1, the parameter τ is selected to achieve specific performance criterion. The present chapter does not treat the optimal selection of τ, as its objective is to investigate the ability of optimal and stable portfolios to reduce the exposures to currency risk, and how the adoption of such strategies affects currency premiums.

2.2.2 Conditional International Asset Pricing Models

Our empirical model falls is an international asset pricing model. This section presents the theoretical foundations of such models as well as some transformations proposed in the literature to make estimations feasible.

In the sequel we adopt the following notations. Ω_{t-1} denotes the set of information available to the investor at the end of time $t-1$; $E_{t-1}(\cdot)$ and $\text{cov}_{t-1}(\cdot)$ are the expectation and covariance operator conditioned on Ω_{t-1}, respectively.

The standard Conditional International Asset Pricing Models

The model we use in our investigations is based on the seminal model proposed by Adler and Dumas (1983). The failure of PPP assumed in their model, implies that investors in different countries use different price indices to evaluate their investments. As a result, optimal portfolios are also different across countries, and asset pricing models contain a priced currency risk factor.

In a world with $L+1$ countries, and a set of N securities, the conditional version of the international asset pricing model expresses the expected excess returns on each asset i, as a function of a global market risk premium and inflation risk premium :

$$E_{t-1}(r_{i,t}) = \delta_{\omega,t-1}\text{cov}_{t-1}(r_{i,t}, r_{\omega,t}) + \sum_{j=1}^{L} \delta_{j,t-1}\text{cov}_{t-1}(r_{i,t}, \pi_{j,t}) \qquad (2.10)$$

where $r_{i,t}$ and $r_{\omega,t}$ are the excess returns on asset i and world market portfolio, in period t, respectively ; $\pi_{j,t}$ is the inflation rate of country j expressed in U.S. dollar, the reference currency ; $\delta_{\omega,t}$ and $\delta_{j,t}$ are the prices of the world market risk and inflation risk, respectively ; $\text{cov}_{t-1}(r_{i,t}, r_{\omega,t})$ measures the exposure of asset i to world market risk and represents market risk component of the standard CAPM (Sharpe, 1964 ; Lintner, 1965) ; while $\text{cov}_{t-1}(r_{i,t}, \pi_{j,t})$ measures the exposure of asset i to both inflation risk and exchange rate risk associated with country j. This last measure of exposure also represents the additional source of risk induced by PPP deviation.

Two general transformations of the model in Equation (2.10) were adopted in the literature to make estimations feasible. The first transformation consists in replacing $\pi_{j,t}$ by the real exchange rate [8], which is a more empirically tractable variable, and ensures that the adjustment for inflation is also accounted for, (see Carrieri et al. 2006a). This means that the real exchange rate is more robust to relatively high levels of inflation. Therefore in the estimations of Section 2.4, we adopt the changes in real exchange rate indices as currency risk factors. In the sequel, the real exchange rate of country j, with respect to U.S dollar is denoted by $e_{j,t}$. For $i = 1, \cdots, N$, replacing the inflation rate of country j by the real exchange rate between the reference currency and the currency of country j, in Equation (2.10) leads to :

$$E_{t-1}(r_{i,t}) = \delta_{w,t-1}\text{cov}_{t-1}(r_{i,t}, r_{w,t}) + \sum_{j=1}^{L} \delta_{j,t-1}\text{cov}_{t-1}(r_{i,t}, e_{j,t}), \qquad (2.11)$$

The time-varying exposure of asset i to factor k is measured by the time-varying beta of portfolio i relative to factor k, $\beta_{i,kt-1}$, defined as the condi-

8. An alternative to real exchange rate is the nominal exchange rate, under the assumption that U.S. inflation rate is non-stochastic so that the only random component of π_j comes from the relative changes in the real exchange rate between the reference currency and the currency of country j. This assumption is violated when changes in domestic inflation are is not negligible relative to exchange rate fluctuations (See Dumas and Solnik, 1995 ; De Santis and Gérard, 1998).

tional covariance risk normalized by the conditional variance of the returns on factor k. For the currency source of risk j,

$$\beta_{i,j,t-1} = \frac{\text{cov}_{t-1}(r_{i,t}, e_{j,t})}{\text{var}_{t-1}(e_{j,t})}. \tag{2.12}$$

If the risk j is priced for the asset i, then investors holding that asset require a premium for bearing the systematic source of risk j. The premium associated with the source of risk j is the product between the conditional price of risk of factor j, $\delta_{j,t-1}$, and the conditional covariance $\text{cov}_{t-1}(r_{i,t}, e_{j,t})$. In the model (2.11), the market risk premium (MRP), and the currency risk premium (CRP) are respectively :

$$MRP_{t-1,i} = \delta_{\omega,t-1}\text{cov}_{t-1}(r_{i,t}, r_{\omega,t}),\ and\ CRP_{t-1,i} = \sum_{j=1}^{L} \delta_{j,t-1}\text{cov}_{t-1}(r_{i,t}, e_{j,t}).$$

The second transformation of (2.10) is obtained by replacing the exchange rates of the L currencies, $\{e_i\}_{i=1}^{L}$, with a reduced number [9], C, of currency exchange rates, $\{e_c\}_{c=1}^{C}$, or by composite exchange rate measures [10]. Following Carrieri, Errunza, and Majerbi (2006a) and Francis, Hasan, and Hunter (2008), we use two exchange rate indices representing U.S. major partners currency risk (MJ), and emerging market (EM) currency risk. Therefore, the total currency premium, that consists of the emerging market currency risk and the major currency risk is expressed as :

$$CRP_{t-1,i} = \sum_{j=mj,\ em} \delta_{j,t-1}\text{cov}_{t-1}(r_{i,t}, e_{j,t}) \tag{2.13}$$

where $e_{mj,t}$ and $e_{em,t}$ are the changes in the two real exchange rate indices of the major currencies and the EM currencies, respectively, vis-à-vis the U.S. dollar, the reference currency.

The different transformations of model (2.10) mentioned above lead to model (2.14) to which refer as the standard model :

$$E_{t-1}(r_{i,t}) = \delta_{\omega,t-1}\text{cov}_{t-1}(r_{i,t}, r_{\omega,t}) + \sum_{k=mj,\ em} \delta_{k,t-1}\text{cov}_{t-1}(r_{i,t}, e_{k,t}). \tag{2.14}$$

9. Among others, see De Santis and Gérard, 1998 ; De Santis, Gérard, and Hillion, 2003.
10. See among others, Jorion, 1991 ; Ferson and Harvey, 1993, 1994 ; Choi, Hirachi, and Takezawa, 1998 ; Carrieri, Errunza, and Majerbi, 2006a.

The models described above have mainly been estimated to investigate the pricing of currency risk across countries. While empirical evidence on the pricing of currency risk at the industry level have long been inconclusive, Francis et al. (2008) demonstrate that this was due to a methodological weakness, rather than effective use of hedging at the industry level. We next review their model which is closely linked to ours.

The conditional Fama-French three-factor model augmented with currency factors

Francis, Hasan, and Hunter (2008) use a five-factor conditional asset pricing model applied to 36 U.S. industry portfolios. Their model is the conditional version of the Fama and French (1993) model, augmented with two currency indices. The five factors used are the value-weighted U.S. market portfolio in excess of the risk-free rate, $r_{m,t}$; the returns on the size factor (SMB), $r_{smb,t}$; the returns on the book-to-market factor (HML), $r_{hml,t}$; and the MJ and EM currency factors. The model specifies the expected excess returns on each industry as the sum of the products of the time-varying betas with the expected excess returns on each of the risk factors considered :

$$E_{t-1}(r_{i,t}) = \sum_{k=m,smb,hml} \beta_{i,k,t-1} E_{t-1}(r_{k,t}) + \sum_{k=mj,\ em} \beta_{i,k,t-1} E_{t-1}(e_{k,t}) \quad (2.15)$$

where $r_{i,t}$ is the return at time t on the i^{th} industry in excess of the return on the risk free asset ; $E_{t-1}(r_{k,t})$, for $k = m, smb, hml$, is the conditionally expected excess return on the k^{th} equity factor ; $\beta_{i,k,t-1}$ is the time-varying exposure of portfolio i relative to factor k defined in (2.12) ; $E_{t-1}(e_{mj,t})$ and $E_{t-1}(e_{em,t})$ are the conditionally expected change in the the MJ and EM currency factors, respectively.

Note that model (2.15) which is equivalent to (2.16), uses the same currency indices as model (2.13), while the worldwide market index is replaced by the Fama and French factors.

$$E_{t-1}(r_{i,t}) = \sum_{k=m,smb,hml} \delta_{k,t-1} \mathrm{cov}_{t-1}(r_{i,t}, r_{j,t}) + \sum_{k=mj,\ em} \delta_{k,t-1} \mathrm{cov}_{t-1}(r_{i,t}, e_{j,t}) \quad (2.16)$$

Model used to evaluate currency risk exposures and currency premiums

In order to evaluate the currency exposures of the portfolio strategies considered, we use the standard model in Equation (2.14) as a starting point. The implicit assumption that we make, is that U.S. industries are integrated with the rest of that world, and are globally priced. This assumption also has a practical aspect, in the sense that the optimal portfolios we construct from U.S. industry portfolios might be highly correlated with the U.S. market portfolio, which is replaced by the world market portfolio. Besides, in order to capture the cross-section variation in U.S. industries, we augment model (2.14) with two of the three Fama-French factors : the returns on the size factor and the returns on the book-to-market factor. The introduction of FF factors will ensure that it is unlikely that currency risk is priced because it is a proxy of an omitted factor. This also prevents us from introducing a missing variable bias in our estimations, as these factors have been shown to be priced factors for cross sections of U.S. industry portfolios, in previous studies. As a result, for each industry i :

$$E_{t-1}(r_{i,t}) = \sum_{k=w,smb,hml} \delta_{k,t-1}\mathrm{cov}_{t-1}(r_{i,t}, r_{k,t}) + \sum_{k=mj,\ em} \delta_{k,t-1}\mathrm{cov}_{t-1}(r_{i,t}, e_{k,t}) \tag{2.17}$$

By multiplying both side of the previous equation by weights constructed using information up to $t-1$, and summing over all N industries we have, using (2.1)

$$E_{t-1}(r_{p,t}) = \sum_{k=w,smb,hml} \delta_{k,t-1}\mathrm{cov}_{t-1}(r_{p,t}, r_{k,t}) + \sum_{k=mj,\ em} \delta_{k,t-1}\mathrm{cov}_{t-1}(r_{p,t}, e_{k,t}) \tag{2.18}$$

Our model differs from the model in Francis et al. (2008) by the fact that we considered the world market factor, instead of the U.S. market factor. This is justified by our assumption that U.S. industries are priced globally. As shown in Equation (2.18), the price of the sources of risk $\delta_{k,t-1}$ remains unchanged. However, the exposures to risk factors, in particular to currency risk exposures, are modified. The objective of this chapter is to investigate the extent to which the adoption of global minimum portfolios of U.S. industries affects the exposures to and premiums for currency risk.

2.2.3 Empirical methodology

The model in Equation (2.13) has to hold for every assets. Therefore, the following system of asset pricing restrictions has to be satisfied for each portfolio p, at each point in time :

$$\begin{aligned}
E_{t-1}(r_{p,t}) &= \sum_{k=w,smb,hml} \delta_{k,t-1}\mathrm{cov}_{t-1}(r_{p,t}, r_{k,t}) + \sum_{k=mj,\ em} \delta_{k,t-1}\mathrm{cov}_{t-1}(r_{p,t}, e_{k,t}) \\
E_{t-1}(r_{w,t}) &= \sum_{k=w,smb,hml} \delta_{k,t-1}\mathrm{cov}_{t-1}(r_{w,t}, r_{k,t}) + \sum_{k=mj,\ em} \delta_{k,t-1}\mathrm{cov}_{t-1}(r_{w,t}, e_{k,t}) \\
E_{t-1}(r_{smb,t}) &= \sum_{k=w,smb,hml} \delta_{k,t-1}\mathrm{cov}_{t-1}(r_{smb,t}, r_{k,t}) + \sum_{k=mj,\ em} \delta_{k,t-1}\mathrm{cov}_{t-1}(r_{smb,t}, e_{k,t}) \\
E_{t-1}(r_{hml,t}) &= \sum_{k=w,smb,hml} \delta_{k,t-1}\mathrm{cov}_{t-1}(r_{hml,t}, r_{k,t}) + \sum_{k=mj,\ em} \delta_{k,t-1}\mathrm{cov}_{t-1}(r_{hml,t}, e_{k,t}) \\
E_{t-1}(e_{mj,t}) &= \sum_{k=w,smb,hml} \delta_{k,t-1}\mathrm{cov}_{t-1}(e_{mj,t}, r_{k,t}) + \sum_{k=mj,\ em} \delta_{k,t-1}\mathrm{cov}_{t-1}(e_{mj,t}, e_{k,t}) \\
E_{t-1}(e_{em,t}) &= \sum_{k=w,smb,hml} \delta_{k,t-1}\mathrm{cov}_{t-1}(e_{em,t}, r_{k,t}) + \sum_{k=mj,\ em} \delta_{k,t-1}\mathrm{cov}_{t-1}(e_{em,t}, e_{k,t})
\end{aligned}$$

Denote by r_t^p, the 6×1 vector of excess returns $r_t^p = (r_{p,t}, r_{w,t}, r_{smb,t}, r_{hml,t}, e_{mj,t}, e_{em,t})'$. The following system of equations can be used to estimate the conditional version of the ICAPM :

$$r_t^p = \sum_{k=w,smb,hml} \delta_{k,t-1} h_{k,t} + \sum_{k=mj,\ em} \delta_{k,t-1} h_{k,t} + \varepsilon_t, \quad \varepsilon_t |\, \Omega_{t-1} \sim N(0, H_t) \tag{2.19}$$

where $\varepsilon_t = (\varepsilon_{p,t}, \varepsilon_{w,t}, \varepsilon_{smb,t}, \varepsilon_{hml,t}, \varepsilon_{mj,t}, \varepsilon_{em,t})'$ is the vector of residuals ; H_t is the 6×6 conditional covariance matrix of asset returns. Denote by n the order of the subscript k in the list $\{w, smb, hml, mj, em\}$, then the conditional covariance matrix between each asset and the return on the k^{th} factor, $h_{k,t}$, corresponds the $(n+1)^{th}$ column of H_t ; for example $h_{em,t}$ is the last column of H_t, represents the conditional covariance of asset returns with emerging market currency index, and measures the exposure to emerging market currency risk.

To allow for time variation in the prices of risk, we use a set of instruments or information variables, well known in the literature of asset pricing to have a predictive power for equity returns and for the change in the currency indices. These instruments include a constant and are observed at the end of time $t-1$. The prices of risk are modeled by a linear relationship for all risk factors except for the market price of risk, which is constrained to be positive [11] in Adler and Dumas (1980) model. For this reason, we specify the world market price as the exponential of a linear relationship (See Bekaert and Harvey, 1995 ; De Santis and Gérard, 1997, 1998).

We denote by $Z_{s,t-1}$ the set of information variables related to equities, and by $Z_{f,t-1}$ the instruments used to predict changes in the currency factors. The instruments and their choice is discussed in detail in Section 2.3. We have the following equations :

$$\begin{aligned} &\delta_{w,t-1} = \exp(\kappa'_w Z_{s,t-1}) \\ &\delta_{j,t-1} = \kappa'_j Z_{s,t-1}, \ j = smb, hml \\ &\delta_{j,t-1} = \kappa'_j Z_{f,t-1}, \ j = mj, em \end{aligned} \tag{2.20}$$

An additional advantage of specifying the dynamics of the prices of risk as a function of instruments, is that the hypothesis zero and constant price of risk are easily testable. For instance, testing the significance of the price of factor j, amounts to testing whether all the element of κ_j are jointly zero ; while testing for time-variation in the prices, is equivalent to testing the null hypothesis that the coefficients of non constant instruments are jointly equal to zero. In Section 2.4, we test these hypothesis using robust Wald statistics.

Finally, we adopt a parsimonious GARCH process for the residuals to accommodate the GARCH-in-mean feature found in most tests of asset pricing models (Ding and Engle ,1994 ; De Santis and Gérard, 1997,1998). We assume that the conditional second moments follow a diagonal GARCH process [12](Bollerslev et al., 1988) and that the conditional covariance matrix H_t is covariance stationary, with H_0 its unconditional mean. Thus H_t can be written as :

$$H_t = (\iota\iota' - aa' - bb') * H_0 + aa' * \varepsilon_{t-1}\varepsilon'_{t-1} + bb' * H_{t-1} \tag{2.21}$$

where ι is the 6×1 vectors of ones, a and b are 6×1 vectors of parameters, and $*$ is the Hadamard (element-by-element) matrix product.

11. The price of risk is the weighted average of the coefficients of risk aversion of all national investors.

12. The variances in H_t depend only on past squared residuals and an autoregressive component, while the covariances depend upon past cross product of residuals and an autoregressive component.

The model we estimate consists of Equations (2.19), (2.20) and (2.21). The set of unknown parameters Ψ [13], are estimated, under conditional normality assumption, the log-likelihood function is :

$$\ln L(\Psi) = -\frac{Ts}{2}\ln 2\pi - \frac{1}{2}\sum_{t=1}^{T}\ln|H_t(\Psi)| - \frac{1}{2}\sum_{t=1}^{T}\varepsilon_t(\Psi)' H_t(\Psi)^{-1}\varepsilon_t(\Psi). \quad (2.22)$$

To make our estimations robust to the frequent violation of the conditional normality assumption in financial time series, we use the quasi-maximum likelihood (QML) approach proposed (Bollerslev and Wooldridge, 1992) to compute all our tests. The system is estimated using the Broyden, Fletcher, Goldfarb, and Shanno (1985) and the Berndt, Hall, Hall, and Hausman (1974) algorithms.

2.3 Data and summary statistics

In this section, we describes the data on the industry portfolios, optimal portfolios, risk factors and instruments used in the empirical analysis. This section also investigate the performance of the optimal rules studied in terms of their Sharpe ratio and turnover.

2.3.1 Description of industry portfolios and GMPs

The global minimum portfolios (GMPs) constructed are based on all 48 U.S. industry portfolios representing an exhaustive coverage [14] of the U.S. stock market, during the period January 1976-September 2008 (393 observations). The monthly returns on industry portfolios are obtained from Ken French's website [15]. The one-month eurodollar deposit quoted in London at the last day of the month, is extracted from Datastream and is used as the conditionally risk-free rate (e.g. Carrieri, Errunza, and Majerbi, 2006a ; De

13. The parameter to estimate are the coefficient for the risk prices, κ_j, for $j \in \{w, smb, hml, mj, em\}$, in the equation for the mean ; the parameter a and b in the equations for the volatility ; the unconditional covariance matrix H_0.

14. Francis, Hasan, and Hunter (2008) only considered the 36 industries that are the most likely to affected by currency risk via foreign outputs or exports of domestic outputs, foreign inputs, foreign competition, foreign clientele, or relation to other industries that are affected by exchange rate movements.

15. http : //mba.tuck.dartmouth.edu/pages/faculty/ken.french/data$_l$ibrary.html.

Santis and Gérard, 1998). We compute the monthly excess returns by subtracting the eurodollar rate from the monthly returns on each portfolios.

Table 2.1 and Panel A in Table 2.2 display summary statistics, for industry portfolio returns, and GMP returns, respectively. The returns on the GMPs are computed recursively, using a rolling windows of $M = 100$, so we lose 100 observations. All the statistics are then computed after May 1984 to match time frame of the GMPs constructed. Industry portfolios are extensively described in Francis, Hasan, and Hunter (2008), so we focus more on the GMPs introduced in this chapter. From Table 2.2, when all the PCs are used ($\tau = 100\%$), the GMP yields on average returns lower than the risk-free rate, and corresponds to the worse performance, in terms of Sharpe ratio, among the portfolios considered. The $GMP(100\%)$ portfolio also leads to a very high turnover, which is symptomatic of a high level of instability in the portfolio weights. As discussed in Chapter 1, these poor performance are a reflection of the negative impact that estimation errors and dimensionality have on Markowitz portfolios' performance. By varying the percentage τ of PCs used, from 90% to 10%, the stability of the weights increases, and the performance of the GMP is improved both in terms of average excess returns, volatility and turnover. For instance, all portfolios outperform the GMP(100%), while the portfolios $GMP(\tau)$, with $\tau \in \{60\%, 50\%, 40\%, 30\%, 20\%, 10\%\}$, are associated with a lower volatility and higher returns compared to the benchmark 1oN.

In all cases, the Bera-Jarque test statistic strongly rejects the null hypothesis of normally distributed returns, and almost all GMPs exhibit a lack of autocorrelations, using the Ljung-Box portmanteau statistic. Evidence of autocorrelation are found in two GMPs at the 5%, and in four GMPs at the 10% significance level. To complete the statistical analysis of the GMPs, Table 2.2 also shows the presence of autocorrelation in the squared returns of the GMPs, at the 1% level for 3 portfolios and at the 5% level for 4 portfolios. This suggest that the GARCH specification that we adopt is appropriate, especially for the series displaying such evidence.

In terms of domestic performance, this preliminary analysis reveals that regularization leads to substantial improvements. The open question is whether these improvements also translates to the exposures to and premiums for currency risk. This question is relevant since the exchange rate risk has been documented to be statistically and economically significant at the country and industry level in the U.S.. We explore this question in Section 4, where the performance of GMPs, in terms of currency risk exposures and premiums, are compared to the average performance across industries.

2.3.2 Description of the risk factors

The world market returns are computed from MSCI total returns indices, while the size and the book-to-market factors are extracted from French's website. The currency factors are represented by the percentage changes in two trade-weighted [16] indices of real bilateral exchange rates. These indices are extracted from the Federal Reserve Board website [17], in U.S. dollar against the currencies of two groups of U.S. trading partners. The first group of countries, the major trading partners, consists of 16 industrialized countries. The second group of countries, consists of 19 other important trading partners (OITP), mostly emerging market economies. In the sequel, the former and latter groups will be referred to as EM and MJ, respectively. Since all returns are expressed in U.S. dollar, we convert [18] the extracted indices in foreign currencies per U.S. dollar, so that a positive changes in their values reflects an appreciation of the dollar.

Panel B in Table 2.2 contains statistics for the risk factors described above. Changes in the EM and MJ currency indices exhibit different patterns. While their ranges are similar and approximately centered around 5% per month, the MJ risk is more volatile than the EM currency risk. Besides, we also found evidence of normality in the changes in the MJ currency index, whereas the normality assumption is strongly rejected for EM currency risk. These differences in statistical properties suggest that the exposures and premiums to the corresponding currency risks are likely to differs. Finally, Panel C of Table 2.2 shows that the pairwise correlations are low among the five risk factors considered, with a maximum of 0.336 for the correlation between the returns on the world market and the returns on the book-to-market factor (SMB). This means that the risk factors included in our IAPM model are not proxies for each other, and then non redundant.

16. The index weights are time-varying, and derived from U.S. export shares and from U.S. and foreign import shares.

17. http : //www.federalreserve.gov/releases/h10/summary/.

18. If I_t is a currency index expressed U.S. dollar against Foreign currencies, the return on the U.S dollar, between $t-1$ and t is $\frac{I_{t-1}-I_t}{I_t}$.

2.3.3 Description of the information instruments

In order to preserve the comparability of our study, with that of Francis, Hasan, and Hunter (2008), we use the same instruments as theirs to model the dynamics of the prices of risk. All instruments are lagged one period relative to factor returns, and have been shown by Francis, Hasan, and Hunter to have significant explanatory power for the returns on the risk factors considered. More specifically, for equity related factors, the instruments include a constant, the change in the U.S. term premium (DUSTP) measured by the difference in yields of the treasury constant-maturity 10-year and three month bills, the U.S. default premium (USDP) measured by the spread between the yields on Moody's Baa-rated and Aaa-rated corporate bonds, and the Fed Funds rate (FED), which is indicative of monetary policy. The changes in the exchange rate factors, are instead predicted by the ratio of imports to GDP (MtoGDP), the ratio of exports to GDP (XtoGDP), and the Fed funds rate. The variables MtoGDP and XtoGDP are in percent of GDP, and are computed using data from the International Financial Statistics (IFS) database of the International Monetary Fund. Table 2.3 contains statistics for both equity and exchange rates related instruments.

Table 2.1 – Summary statistics of returns on industries and risk factors

The table reports summary statistics for the 48 value-weighted industry portfolios from Fama and French (1997), computed over the period May 1984-September 2008 (293 obs.). Returns are expressed in U.S. dollar, in percent per month, and are in excess of the risk-free rate. The risk free rate is taken to be the one-month eurodollar deposit rate. BJ is the Bera-Jarque statistic to test whether the returns are normally distributed. $Q(12)$ and $Q(12)^2$ are the Ljung-Box tests for autocorrelation of order 12 for the excess returns and the excess returns squared, respectively. * and ** denote statistical significance at the 5% and 1% levels, respectively.

Panel A : Summary statistics of portfolio excess returns

Industry		Mean	Std	Min	Max	BJ	$Q(12)$	$Q(12)^2$
Agric	Agriculture	1.298	5.925	-28.81	28.56	159.232**	11.577	11.275
Food	Food Products	1.247	4.5	-17.78	19.34	63.164**	18.376	42.452**
Soda	Candy and Soda	1.117	6.582	-26.03	38.9	218.64**	19.461	33.028**
Beer	Beer and Liquor	1.362	5.404	-19.74	22.1	44.142**	15.681	26.482**
Smoke	Tobacco Products	1.548	6.623	-24.96	32.46	123.512**	7.573	56.553**
Toys	Recreation	0.94	6.78	-34.51	23.1	56.958**	11.705	16.036
Books	Printing and Publishing	1.302	6.843	-31.89	21.08	75.905**	18.662	9.647
Hshld	Consumer Goods	1.075	5.237	-22.59	16.83	15.951**	17.536	22.111*
Clths	Apparel	1.015	4.611	-21.67	18.71	76.337**	7.993	14.745
Fun	Entertainment	1.188	6.181	-30.86	24.62	126.695**	24.389*	14.484
MedEq	Medical Equipment	1.123	5.112	-20.56	16.32	24.567**	9.724	15.316
Drugs	Pharmaceutical Products	1.169	4.971	-19.1	16.37	8.01**	15.739	17.013
Chems	Chemicals	1.004	5.335	-27.96	22.01	131.56**	8.333	15.276
Rubbr	Rubber and Plastic Products	1.129	5.584	-30.49	19.52	172.226**	14.504	11.737
Txtls	Textiles	1.026	6.114	-32.63	22.86	181.92**	24.468*	21.571*
BldMt	Construction Materials	1.123	5.599	-27.8	19.27	109.332**	10.773	8.768
Cnstr	Construction	1.186	7.073	-31.21	24.28	35.297**	20.995*	8.606
Steel	Steel Works Etc	0.89	7.463	-30.86	30.67	98.572**	5.959	31.705**
FabPr	Fabricated Products	0.755	6.912	-28.91	25.58	108.733**	15.154	12.468
Mach	Machinery	1.01	6.103	-31.37	18.68	112.923**	11.14	12.981
ElcEq	Electrical Equipment	1.369	6.095	-32.09	18.29	85.161**	8.957	12.771
Autos	Automobiles and Trucks	0.855	6.435	-28.11	19.17	43.418**	11.052	21.262*
Aero	Aircraft	1.368	6.548	-30.28	25.33	142.048**	14.991	21.038*
Ships	Shipbuilding. Railroad Equipment	0.988	6.916	-32.22	22.01	54.201**	5.794	8.639
Guns	Defense	1.439	6.485	-30.08	32.87	160.014**	11.729	13.788
Gold	Precious Metals	1.038	10.94	-31.09	78.02	659.384**	11.688	5.099
Mines	Non-Metallic and Industrial Metal Mining	1.024	7.036	-33.65	20.53	66.818**	3.803	20.898
Coal	Coal	1.392	10.008	-38.04	44.04	66.135**	18.168	132.048**
Comps	Computers	1.005	7.618	-32.78	24.46	54.202**	11.27	281.828**
Chips	Electronic Equipment	1.178	7.889	-31.77	26.8	67.502**	7.118	175.385**
LabEq	Measuring and Control Equipment	1.103	7.312	-30.13	22.48	22.064**	11.202	74.169**
Paper	Business Supplies	1.021	5.409	-26.16	24.11	152.845**	11.139	28.718**
Boxes	Shipping Containers	1.117	5.697	-28.32	20.39	137.759**	5.305	27.228**
Rtail	Retail	1.16	5.504	-29.29	16.86	71.472**	17.502	15.382
Meals	Restaraunts, Hotels, Motels	1.111	5.549	-23.97	16.06	62.888**	19.391	6.29
Banks	Banking	1.184	5.665	-24.05	20.47	67.184**	11.327	12.441
Hlth	Healthcare	1.417	7.22	-31.5	21.66	35.797**	12.642	23.384*
Oil	Petroleum and Natural Gas	1.29	5.506	-18.27	24.4	32.500**	10.572	48.298**
Util	Utilities	1.029	3.921	-12.42	12.02	10.913**	21.193*	29.425**
Telcm	Communication	0.97	4.864	-15.45	22.12	41.801**	23.802*	87.696**
PerSv	Personal Services	1.08	6.029	-28.23	24.68	60.317**	26.623**	32.361**
BusSv	Business Services	1.35	6.768	-27.56	24.12	27.015**	11.954	89.567**
Trans	Transportation	1.111	5.549	-23.97	16.06	62.888**	19.391	6.29
Whlsl	Wholesale	1.184	5.665	-24.05	20.47	67.184**	11.327	12.441
Insur	Insurance	1.196	5.065	-16.84	23.35	35.433**	17.372	23.202*
RlEst	Real Estate	0.76	6.041	-25.98	20.76	93.977**	36.026**	25.026*
Fin	Trading	1.367	6.121	-26.02	18.54	61.211**	22.455*	126.075**
Other	Almost Nothing	0.89	6.285	-26.31	20.08	56.982**	4.691	40.105**

Table 2.2 – Summary statistics of returns on investment rule and risk factors

Panel A reports summary statistics for GMPs, where $GMP(\tau)$ represents global minimum portfolios constructed using τ percent of PCs. Panel B contains the risk factors. Equity related factors consists of the returns on the world market in excess of the one-month eurodollar rate (WM) computed using MSCI total returns indices from Datastream, and two Fama and French (1993) factors : the returns on small minus big firms (SMB), and the returns on high minus low book-to-market value firms (HML). The currency factors are represented by percentage changes in the real US Treasury trade-weighted exchange rate index made up of the currencies of 16 developed countries that are the main trading partners of the US (MJ) and the real index of the currencies of the other important trading partners (OITP) from the emerging economies (EM). Data on both indices are in foreign currency per US dollar, and are obtained from the Federal Reserve Board. * and ** denote statistical significance at the 5% and 1% levels, respectively.

Panel A : Summary statistics of portfolio excess returns in the whole sets of assets

Investment rule		Mean	Std	Sharpe ratio	Turnover	Min	Max	BJ	$Q(12)$	$Q(12)^2$
1oN	Equally weighted portfolio	0.489	4.461	0.110	0.000	-30.499	12.141	951.599**	16.198	4.224
GMP	Global Minimum portfolio	-0.053	4.312	-0.012	2.36e14	-19.694	11.645	41.517**	16.039	5.889
GMP(90%)	GMP with 90% of the PCs	0.090	4.200	0.021	1.075	-20.081	11.321	72.121**	17.122	5.203
GMP(80%)	GMP with 80% of the PCs	0.227	3.846	0.059	0.935	-17.900	9.936	86.938**	21.088	7.629
GMP(70%)	GMP with 70% of the PCs	0.351	3.601	0.097	0.748	-19.274	8.926	145.900**	16.314	2.953
GMP(60%)	GMP with 60% of the PCs	0.472	3.639	0.130	0.724	-16.938	14.897	78.711**	15.343	31.601**
GMP(50%)	GMP with 50% of the PCs	0.594	3.588	0.166	0.648	-15.670	16.191	63.793**	17.834	51.804**
GMP(40%)	GMP with 40% of the PCs	0.607	3.800	0.160	0.496	-19.316	15.780	165.291**	14.492	27.382**
GMP(30%)	GMP with 30% of the PCs	0.722	3.730	0.194	0.446	-19.461	14.487	176.209**	20.486	21.672*
GMP(20%)	GMP with 20% of the PCs	0.602	3.867	0.156	0.221	-19.651	11.402	231.209**	22.693*	11.314
GMP(10%)	GMP with 10% of the PCs	0.428	3.998	0.107	0.205	-22.529	9.993	334.556**	20.559	6.219

Panel B : Summary statistics of returns on the risk factors

Risk Factors		Mean	Std	Min	Max	BJ	$Q(12)$	$Q(12)^2$
WR	World market equity index	0.371	4.228	-19.227	10.583	76.161**	10.648	7.829
SMB	Size factor in the U.S.	-0.010	3.255	-17.901	19.885	702.677**	11.485	131.828**
HML	Book-to-market factor in the U.S.	0.296	3.054	-13.467	12.962	115.567**	11.264	365.710**
MJ	Major currency index	0.096	1.737	-4.909	5.376	2.679	47.744**	11.219
EM	Emerging market currency index	0.040	1.046	-5.222	4.109	570.885**	44.700**	84.279**

Panel C : Unconditional correlations

Risk Factors	WM	SMB	HML	MJ	EM
WM	1.000				
SMB	0.084	1.000			
HML	-0.360	-0.411	1.000		
MJ	0.256	-0.048	-0.088	1.000	
EM	0.145	0.133	-0.081	0.182	1.000

Table 2.3 – Summary statistics for instruments

The set of equity related instruments includes a constant, the change in the U.S. term premium (DUSTP), the difference in the yields of the Treasury constant-maturity 10-year and the three-month bills ; the US term premium, the spread between the yields on Moodys Baa- and Aaa-rated corporate bonds ; the Fed funds rate (FED). The instrumental variables related to currency factors are, MtoGDP, the percentage of US imports to gross domestic product (GDP), XtoGDP, the percentage of US exports to GDP, and the fed fund rate. All variables are in percent per month, lagged one month relative to factor returns. Data are monthly from May 1984 to September 2008 (293 observations).

Panel A : Summary statistics					
Instrument	Mean	Std	Min	Max	ρ_1
DUSTP	-0.003	0.255	-0.990	0.800	0.343
USDP	0.946	0.273	0.550	1.760	0.947
FFR	0.431	0.195	0.082	0.965	0.980
MtoGDP	0.893	0.175	0.622	1.443	0.929
XtoGDP	0.583	0.090	0.364	0.823	0.872

Panel C : Unconditional correlations					
INSTR	DUSTP	USDP	FFR	MtoGDP	XtoGDP
DUSTP	1.000				
USDP	0.042	1.000			
FFR	-0.073	0.232	1.000		
MtoGDP	-0.031	-0.164	-0.477	1.000	
XtoGDP	0.056	-0.564	-0.317	0.627	1.000

2.4 Empirical evidence

This section presents the results of our empirical tests. Subsection 2.4.1 performs a number of specification tests on the conditional IAPM, and assesses the significance and time-variation in the prices of risk. In Subsection 2.4.2, we measure the exposures and premiums associated with currency risks. Finally, we discuss in Subsection 2.4.3 the performances of global minimum portfolios in terms of currency exposures and premiums. Overall, the results presented for industry portfolios are consistent with Francis, Hasan, and Hunter (2008).

2.4.1 Specification tests

We test the statistical significance of and time-variation in the market prices of risk and currency prices of risk. We report in Table 2.4 and Table 2.6 the p-values for the robust Wald test statistics, for industry portfolios and global minimum portfolios, respectively.

As already mentioned, we disaggregate currency risk in its industrialized countries (MJ) and emerging market (EM) components, and the results presented concern the significance of currency risk for each components and jointly.

For all industry portfolios, as in Francis et al. (2008), the null hypothesis of both zero and constant price of risk, for the MJ currency risk, is rejected at the 1% significance level. Meaning that the major currency is significantly priced and time-varying at the 1% significance level. Concerning the EM currency risk, the price of risk is significant at the 10% significance level for 27 out of 36 industries in the U.S., which is consistent with the 32 out of 36 in Francis et al. (2008). While the currency prices of risk are jointly time-varying, the null hypothesis of constant EM price of risk cannot be rejected at any significance level.

The results obtained for industry portfolios also translates to global minimum portfolios, in the case of the MJ risk. Indeed, the MJ price of currency risk remains significant and time-varying at the 1% significance level, for the 1oN and for all GMPs. However, for the EM risk, the results across global minimum portfolios vary. EM remains significantly priced when 100% of the PCs is used. However, for $\tau = 80\%, 40\%, 20\%$, and 10% the EM currency risk is no longer significant at the 5% significance level. This suggests that emerging currency risk can be diversified away, when investors use less than 100% of the PCs. As for industry portfolios, we find that EM and MJ currency risks are jointly time-varying, for GMPs and for the 1oN portfolio.

2.4.2 Economic relevance of currency risk for GMPs

To analyze how GMPs are affected by currency risk, the statistical significance of the prices of risk must be complemented by an assessment of the economic significance of the currency risk premiums to the portfolio expected returns. This is because even if priced, currency risk can still play a negligible economic role, since it competes with equity related risks. The economic relevance is typically assessed, by considering the exposures to currency risk, and the contribution of currency premium to total premium, in absolute terms (see among other, De Santis and Gérard, 1998 ; Carrieri, Errunza, and Majerbi, 2006a). We compare our results to the average across all 48 industries, as our objective is to investigate how optimal domestic portfolios are affected by currency risk.

Table 2.7 and Panel A in Table 2.9 display statistics on the exposures to various sources of risks, for industry portfolios and optimal portfolios, respectively. Exposures are measured by the respective betas, as defined in Equation (2.12). We obtain lower mean absolute exposures to the MJ currency risk for 9 GMPs out of 11 portfolios, compared to the average exposures across the U.S. industries. For the EM risk, absolute exposures are lower for all GMPs. These results suggest that stable global minimum portfolios, can substantially reduce the exposure to EM and MJ currency risks. With the average MJ exposure reduced by up to 50% (for $\tau = 60\%$), and the average EM exposure reduced by up to 70% (for $\tau = 70\%$).

The estimated premiums for equity and currency related factors are displayed in Table 2.8, and Panel B of Table 2.9. We find that, at the U.S. industry level, the currency premium accounts for approximately 13.05% of the total premium across all 48 industries, and 13.48% across the set S_1 of the 36 industries that are the most exposed to currency risk. This is comparable to the 11.7% [19] average contribution across the set S_1, provided by Francis, Hasan, and Hunter (2008), and to the 20% figure provided by Carrieri, Errunza, and Majerbi (2006) for the aggregate U.S stock market.

The results in Panel B of Table 2.9, provide strong evidence that currency risk is also economically relevant for optimal domestic portfolios. First, these tables reveal that GMPs are associated with reduced risk premia, for equity sources of risk, as well as currency sources of risk. Second, the average contribution to total premium ranges from 12.94% to 21.94%, with the minimum achieved for the global minimum portfolios with $\tau = 60\%$. This minimum is approximately equal to the average percentage over all industries. Therefore this result suggests that the premiums associated with each source of risk (equity and currency) are reduced by about the same proportion, so that in relative terms, the contribution of currency premiums is unchanged.

19. The slight difference could be attributed to the period spanned by our sample and the difference in our specifications.

TABLE 2.4 – Hypothesis tests of the pricing of currency and other risks

The table reports p-values of tests of the null hypothesis that the conditional prices of the risk factors, defined in Table 2.2 are zero and for the hypothesis of constant price (not time-varying). This amounts to test that coefficients in Equation (2.20) are zero, or that the coefficient associated with non-constant instruments are zero. All hypotheses are based on Wald tests made robust to non-normality of the residuals. The instruments for the equity factors are the one-period lagged USDP, USTP, and FED variables. For the currency risks, the instruments are MtoGDP, XtoGDP, and FED. Instruments are defined in Table 2.3.

Panel A : Hypothesis tests for the pricing of equity risk equity risk					
H_0 :	World market risk not time-varying $\kappa_{w,j}=0$, for $j>1$	SMB risk is not significant $\kappa_{smb,j}=0$, for $j>0$	SMB risk is not time-varying $\kappa_{smb,j}=0$, for $j>1$	HML risk is not significant $\kappa_{hml,j}=0$, for $j>0$	HML risk is not time-varying $\kappa_{hml,j}=0$, for $j>1$
df	3	4	3	4	3
Agric	0.022	0.001	0.000	0.011	0.242
Food	0.043	0.001	0.000	0.003	0.367
Soda	0.021	0.001	0.000	0.003	0.157
Beer	0.058	0.001	0.001	0.006	0.428
Smoke	0.022	0.000	0.000	0.002	0.308
Toys	0.040	0.000	0.000	0.001	0.200
Fun	0.039	0.001	0.000	0.002	0.110
Books	0.051	0.113	0.095	0.080	0.331
Hshld	0.088	0.012	0.005	0.051	0.289
Clths	0.053	0.003	0.002	0.002	0.228
MedEq	0.019	0.001	0.000	0.003	0.167
Drugs	0.009	0.002	0.001	0.001	0.177
Chems	0.035	0.001	0.001	0.003	0.247
Rubbr	0.052	0.002	0.001	0.025	0.486
Txtls	0.022	0.001	0.000	0.004	0.202
BldMt	0.066	0.001	0.001	0.004	0.260
Cnstr	0.037	0.001	0.000	0.005	0.205
Steel	0.096	0.000	0.000	0.002	0.211
FabPr	0.032	0.000	0.000	0.000	0.111
Mach	0.050	0.001	0.000	0.002	0.154
ElcEq	0.051	0.001	0.000	0.005	0.258
Autos	0.034	0.000	0.000	0.001	0.108
Aero	0.054	0.001	0.001	0.018	0.322
Ships	0.025	0.004	0.002	0.019	0.226
Guns	0.060	0.001	0.001	0.013	0.385
Gold	0.029	0.001	0.000	0.003	0.175
Mines	0.026	0.002	0.001	0.002	0.123
Coal	0.009	0.000	0.000	0.003	0.075
Comps	0.051	0.182	0.206	0.068	0.503
Chips	0.070	0.000	0.000	0.013	0.186
LabEq	0.057	0.000	0.000	0.003	0.201
Paper	0.050	0.000	0.000	0.004	0.207
Boxes	0.037	0.001	0.000	0.004	0.209
Rtail	0.023	0.001	0.000	0.003	0.283
Meals	0.036	0.001	0.001	0.003	0.294
Banks	0.049	0.001	0.000	0.009	0.274

(Continued on next page)

Table 2.4 (Continued) Hypothesis tests of the pricing of the currency risks

Panel B : Hypothesis tests for the pricing of currency risks					
H_0 :	MJ currency risk not significant $\kappa_{mj,j}=0$, for $j>0$	MJ currency risk not time-varying $\kappa_{mj,j}=0$, for $j>1$	EM currency risk real currency risk $\kappa_{em,j}=0$, for $j>0$	EM currency risk not time-varying $\kappa_{em,j}=0$, for $j>1$	Joint currency risk not time-varying $\kappa_{mj,j}=0$ and $\kappa_{em,j}=0$ for $j>0$
df	3	4	3	4	3
Agric	0.002	0.001	0.141	0.490	0.004
Food	0.002	0.001	0.238	0.383	0.004
Soda	0.002	0.001	0.014	0.342	0.003
Beer	0.002	0.001	0.019	0.278	0.002
Smoke	0.003	0.001	0.001	0.504	0.007
Toys	0.002	0.001	0.426	0.548	0.002
Fun	0.002	0.001	0.043	0.504	0.005
Books	0.005	0.002	0.085	0.512	0.010
Hshld	0.010	0.004	0.201	0.599	0.018
Clths	0.002	0.001	0.032	0.225	0.002
MedEq	0.003	0.001	0.025	0.427	0.005
Drugs	0.002	0.001	0.091	0.654	0.005
Chems	0.002	0.001	0.045	0.327	0.003
Rubbr	0.002	0.001	0.112	0.470	0.005
Txtls	0.002	0.001	0.089	0.285	0.003
BldMt	0.002	0.001	0.008	0.304	0.003
Cnstr	0.003	0.001	0.001	0.303	0.004
Steel	0.002	0.001	0.025	0.585	0.007
FabPr	0.002	0.001	0.063	0.406	0.003
Mach	0.002	0.001	0.190	0.468	0.004
ElcEq	0.003	0.001	0.033	0.358	0.004
Autos	0.003	0.001	0.093	0.392	0.004
Aero	0.006	0.003	0.000	0.556	0.018
Ships	0.002	0.001	0.250	0.527	0.004
Guns	0.002	0.001	0.000	0.422	0.004
Gold	0.002	0.001	0.072	0.287	0.002
Mines	0.002	0.001	0.077	0.266	0.002
Coal	0.001	0.000	0.060	0.208	0.001
Comps	0.003	0.001	0.000	0.533	0.008
Chips	0.002	0.001	0.000	0.426	0.003
LabEq	0.003	0.001	0.061	0.518	0.004
Paper	0.003	0.002	0.000	0.414	0.004
Boxes	0.002	0.001	0.074	0.383	0.003
Rtail	0.003	0.001	0.279	0.386	0.004
Meals	0.003	0.001	0.018	0.304	0.004
Banks	0.002	0.001	0.104	0.292	0.003

Table 2.5 – Quasi-maximum likelihood estimates of the conditional International CAPM with time-varying prices of risk

Estimates are based on monthly continuously compounded returns from Mai 1984 through September 2008. Panel A and B present the results for a single estimation system ($\tau = 60\%$), from the 59 estimations performed. Each mean equation relates the asset excess return $r_{i,t}$ to its world covariance risks $\mathrm{cov_{t-1}(r_{i,t}, r_{k,t})}$ for $k \in \{w, smb, hml, mj, em\}$. The prices of risk are functions of a number of equity and currency risk related instruments, denoted $Z_{s,t-1}$ and $Z_{f,t-1}$, respectively. These instruments are described in Table 2.3.

$$r_t^p = \sum_{k=w,smb,hml} \delta_{k,t-1} h_{k,t} + \sum_{k=mj,\ em} \delta_{k,t-1} h_{k,t} + \varepsilon_t, \quad \varepsilon_t |\, \Omega_{t-1} \sim N(0, H_t)$$

where r_t^p, the 6×1 vector of excess returns $r_t^p = (r_{p,t}, r_{w,t}, r_{smb,t}, r_{hml,t}, e_{mj,t}, e_{em,t})'$ and

$$\delta_{w,t-1} = \exp(\kappa_w' Z_{s,t-1}); \quad \delta_{j,t-1} = \kappa_j' Z_{s,t-1},\ j = smb, hml; \quad \delta_{j,t-1} = \kappa_j' Z_{f,t-1},\ j = mj, em.$$

The conditional covariance matrix is parametrized as $H_t = (\iota\iota' - aa' - bb') * H_0 + aa' * \varepsilon_{t-1}\varepsilon_{t-1}' + bb' * H_{t-1}$, where * denotes the Hadamard matrix product, a and b are 6×1 vectors of constants, and ι is an 6×1 unit vector. SE represent the QML standard errors. ** and *** denote statistical significance at the 5% and 1% levels, respectively

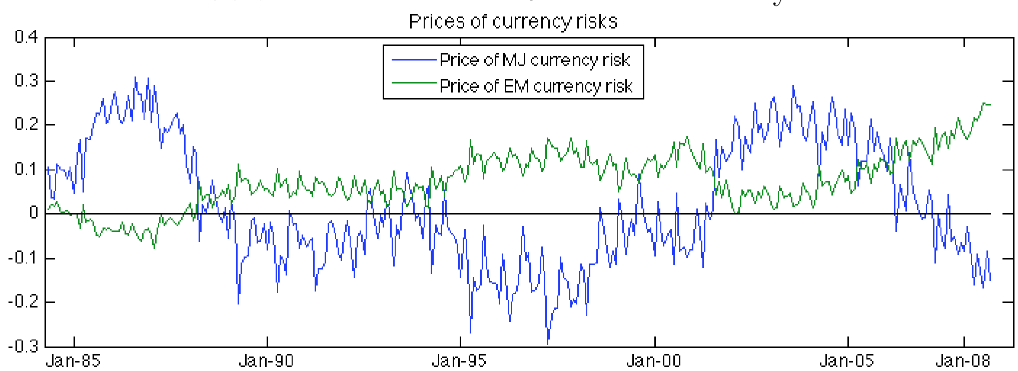
Panel A : Parameter estimates - mean equations								
	(a) Price of Equity related risks							
	Constant	SE	DUSTP	SE	USDP	SE	FFR	SE
κ_w	-3.179***	0.757	-1.746**	0.728	-0.601	1.039	0.924	1.144
κ_{smb}	0.026	0.053	0.148*	0.078	0.121**	0.052	-0.307***	0.076
κ_{hml}	0.139**	0.059	0.013	0.073	0.004	0.053	-0.207**	0.082
	(b) Price of currency related risks							
	Constant	SE	MtoGDP	SE	XtoGDP	SE	FFR	SE
κ_{mj}	0.794***	0.219	0.465**	0.197	-1.881***	0.446	-0.218	0.193
κ_{em}	0.024	0.324	0.079	0.278	-0.018	0.484	0.024	0.276
Likelihood function	-3.848e+03							

Panel B : Parameter estimates - Covariance process						
	GMP(60%)	WM	SMB	HML	EM	MJ
a_i	0.114	0.187***	0.291***	0.275***	0.025	0.714***
SE	0.117	0.042	0.057	0.020	0.053	0.035
b_i	-0.395	0.943***	0.919***	0.937***	0.992***	0.284**
SE	0.475	0.019	0.034	0.007	0.007	0.120

2.4.3 How are GMPs affected by currency risk

As shown in Equations (2.17) and (2.18), adopting a given portfolio strategy p, primarily affects the covariance risk, $\text{cov}_{t-1}(r_{p,t}, e_{j,t})$ (or equivalently the exposure $\beta_{p,j,t-1}$) and leaves the prices of risk $\delta_{j,t-1}$, to which all assets are exposed, unchanged. The resulting premiums are determined by multiplying the covariance risk of portfolio p, and the prices of the risk considered.

FIGURE 2.1 – Prices of MJ and EM currency risks



The reduction in the currency risk exposures, discussed above are supported by Figure 2.3 that gives a visual summary of the reduction from industry portfolios, to GMPs. The exposures of GMP(70%) and GMP(60%) are not only substantially reduced, but the resulting exposures have a very low volatility, and fluctuate around their means within a narrow band. EM and MJ currency risks exhibit different patterns. The volatility reduction in the exposure to MJ risks appears to be higher than the reduction in the EM exposure volatility. As a consequence, especially for EM currency risk, GMPs exposures to currency risk, often display constant signs.

Contrary to the market price of risk, currency prices of risk are not restricted to be positive. Figure 2.1 reveals that the EM and MJ prices of currency risk take alternatively positive and negative values, over time. As a result, even though the exposures to currency risk have a constant sign, the associated premiums fluctuate around zero, as shown in Figure 2.4 for industry portfolios, and in Figure 2.5 for optimal portfolios. Besides, Figure 2.5 indicates that, optimal portfolios associated to stable weights tend to lead to less volatile currency premiums.

To summarize our findings, global minimum variance portfolios can substantially reduce exposures to currency risks. However, while the reduction in exposures is very important (more than half), the corresponding reduction in the contribution to total premium (in absolute terms), is approximately unchanged. Finally, better performances are obtained through the use of regularization techniques that stabilizes portfolios weights. This means that, in order to fully benefit from domestic diversification, in terms of currency risk exposures, the choice of the investment rules is important, but stable weights are more likely to provide better results. These characteristics of GMP currency exposures, are especially attractive in terms of hedging strategies. Indeed, optimal stable GMPs have the attractive feature that they can reduce the size and the volatility of the exposures to the exchange rate risk. This implies that risk-minimizing investors that use stable rules, will incur lower hedging cost, because these rule require smaller amount of hedging with less frequent rebalancing.

2.5 Diagnostics and robustness tests

The results presented in this chapter are based on a rolling window of size $M = 100$. We also considered different values for M that we did not present here. Larger values for M correspond to better portfolio performance, as they have less estimation errors. However, large M also correspond to low amount of data left for our estimations ($T - M$). Therefore, there is a trade-off involved in the choice of M. The fact that we considered all the possible values for the regularization parameter τ in our estimations, make it irrelevant to consider different ranges for M.

Table 2.5 and Table 2.10 show that the conditional covariance matrix is appropriately described by the multivariate GARCH process and that our model is well specified. In Table 2.5, almost all parameters in the vectors a and b are statistically significant, and the point estimates indicate that all the variances and covariance processes in H_t are stationary. [20] Table 2.10 contains diagnostic statistics on the standardized residuals ($\varepsilon_t h_t^{-1/2}$) and on the standardized residuals squared ($\varepsilon_t^2 h_t^{-1}$), for industry portfolios and GMPs. The normality hypothesis is strongly rejected by the Bera-Jarque test in almost all portfolios. This typical rejection of the normality assumption do not affect the validity of our procedures since we adopted the QML approach.

20. Each process in H is covariance stationary if $a_i a_j + b_i b_j < 1 \; \forall i, j$ (See Theorem 1 in Bollerslev (1986)).

Table 2.6 – Hypothesis tests for the pricing of risk factors in GMPs

The table reports the p-values of tests of the null hypothesis that the conditional prices of the risk factors, defined in Table 2.2 are zero and for the hypothesis of constant price (not time-varying). For each sources of risk, this amounts to test that coefficients in Equation (2.20) are jointly zero, or that the coefficient associated with non-constant instruments are zero. All hypotheses are based on Wald tests made robust to non-normality of the residuals. The instruments for the equity factors are the one-period lagged DUSTP, USDP, and FED. For the currencies, the instruments are FED, MtoGDP, and XtoGDP.

Panel A : Hypothesis tests for the pricing of equity risk equity risk					
H_0 :	World market risk not time-varying $\kappa_{w,j} = 0$, for $j > 1$	SMB risk is not significant $\kappa_{smb,j} = 0$, for $j > 0$	SMB risk is not time-varying $\kappa_{smb,j} = 0$, for $j > 1$	HML risk is not significant $\kappa_{hml,j} = 0$, for $j > 0$	HML risk is not time-varying $\kappa_{hml,j} = 0$, for $j > 1$
df	3	4	3	4	3
1oN	0.401	0.000	0.000	0.005	0.155
Gmin(100%)	0.026	0.000	0.000	0.003	0.109
Gmin(90%)	0.024	0.000	0.000	0.002	0.096
Gmin(80%)	0.023	0.001	0.000	0.002	0.109
Gmin(70%)	0.033	0.001	0.000	0.004	0.116
Gmin(60%)	0.027	0.001	0.000	0.006	0.155
Gmin(50%)	0.036	0.003	0.002	0.005	0.242
Gmin(40%)	0.040	0.001	0.001	0.005	0.304
Gmin(30%)	0.034	0.000	0.000	0.003	0.324
Gmin(20%)	0.029	0.000	0.000	0.001	0.182
Gmin(10%)	0.042	0.000	0.000	0.003	0.308

Panel B : Hypothesis tests of the pricing of currency risks					
H_0 :	MJ currency risk not significant $\kappa_{mj,j} = 0$, for $j > 0$	MJ currency risk not time-varying $\kappa_{mj,j} = 0$, for $j > 1$	EM currency risk real currency risk $\kappa_{em,j} = 0$, for $j > 0$	EM currency risk not time-varying $\kappa_{em,j} = 0$, for $j > 1$	Joint currency risk not time-varying $\kappa_{mj,j} = 0$ and $\kappa_{em,j} = 0$ for $j > 0$
df	3	4	3	4	3
1oN	0.014	0.006	0.030	0.631	0.028
Gmin(100%)	0.002	0.001	0.000	0.584	0.006
Gmin(90%)	0.002	0.001	0.039	0.503	0.005
Gmin(80%)	0.002	0.001	0.085	0.488	0.004
Gmin(70%)	0.002	0.001	0.029	0.479	0.004
Gmin(60%)	0.002	0.001	0.013	0.547	0.008
Gmin(50%)	0.003	0.001	0.000	0.462	0.008
Gmin(40%)	0.003	0.001	0.078	0.529	0.007
Gmin(30%)	0.005	0.002	0.000	0.322	0.010
Gmin(20%)	0.003	0.001	0.118	0.218	0.002
Gmin(10%)	0.005	0.002	0.245	0.275	0.005

TABLE 2.7 – Time-varying currency betas for industry portfolios

The table reports the sample mean of the time-varying currency factor betas estimated from the system of Equations (2.19)-(2.21). Mean abs is the mean of the absolute betas. All 48 industries are considered. Because of the limited space only betas for the set S_1 of the industries are the most likely exposed to currency risk, is displayed (Francis et al. (2008)). The currency factors are represented by percentage changes in the real US Treasury trade-weighted exchange rate index made up of the currencies of 16 developed countries that are the main trading partners of the US (MJ) and the real index of the currencies of the other important trading partners from the emerging economies (EM). The indices are in foreign currency per US dollar.

Time varying currency betas										
	MJ					EM				
Industry	Mean Abs	Mean	Std	Min	Max	Mean Abs	Mean	Std	Min	Max
Agric	0.247	0.247	0.015	0.195	0.297	0.554	0.535	0.312	-0.518	1.99
Food	0.178	0.178	0.02	0.125	0.239	0.346	-0.04	0.483	-1.858	1.547
Soda	0.127	0.11	0.166	-0.115	0.673	0.481	0.076	0.688	-2.581	3.346
Beer	0.395	0.395	0.026	0.315	0.467	0.796	0.65	0.721	-1.793	3.501
Smoke	0.135	0.114	0.141	-0.151	0.521	0.692	0.195	0.974	-3.032	4.062
Toys	0.11	0.11	0.030	0.000	0.218	0.866	0.789	0.59	-2.276	2.328
Fun	0.019	0.008	0.023	-0.069	0.071	0.701	0.653	0.466	-1.725	2.494
Books	0.116	-0.088	0.107	-0.342	0.17	0.533	0.308	0.608	-2.835	2.077
Hshld	0.116	0.115	0.039	-0.1	0.262	0.575	0.417	0.579	-2.443	3.445
Clths	0.135	-0.072	0.145	-0.585	0.324	1.129	0.919	1.032	-5.189	5.162
MedEq	0.094	0.094	0.017	0.059	0.158	0.457	0.428	0.289	-0.79	1.303
Drugs	0.311	0.311	0.02	0.239	0.397	0.285	-0.063	0.395	-1.651	1.182
Chems	0.243	0.243	0.054	0.119	0.397	1.163	1.152	0.451	-0.838	2.583
Rubbr	0.174	0.174	0.064	0.017	0.321	0.894	0.84	0.586	-2.589	2.88
Txtls	0.224	-0.224	0.044	-0.43	-0.098	1.373	1.28	0.886	-3.985	4.049
BldMt	0.212	0.212	0.075	0.027	0.393	0.946	0.89	0.597	-2.415	3.235
Cnstr	0.17	0.17	0.01	0.143	0.202	1.034	1.034	0.368	0.011	1.496
Steel	0.375	0.375	0.045	0.284	0.477	1.829	1.815	0.789	-1.033	4.187
FabPr	0.232	0.231	0.089	-0.169	0.513	2.036	1.967	1.062	-2.364	5.386
Mach	0.334	0.334	0.028	0.239	0.438	1.695	1.694	0.6	-0.216	2.699
ElcEq	0.336	0.336	0.038	0.245	0.464	0.815	0.794	0.418	-1.215	1.939
Autos	0.119	0.119	0.042	-0.019	0.271	1.213	1.081	0.92	-2.859	4.347
Aero	0.091	0.068	0.092	-0.165	0.453	1.149	1.057	0.752	-2.581	5.128
Ships	0.347	-0.347	0.1	-0.674	-0.051	0.888	0.77	0.684	-2.876	3.426
Guns	0.101	-0.101	0.026	-0.174	0.01	0.515	0.195	0.687	-3.196	3.228
Gold	0.782	0.782	0.044	0.409	0.894	1.297	1.179	0.898	-2.171	4.66
Mines	0.353	0.353	0.033	0.29	0.438	2.038	2.032	0.723	-0.721	3.538
Coal	0.478	0.478	0.069	0.263	0.676	2.319	2.224	1.353	-2.844	6.215
Comps	0.514	0.514	0.052	0.389	0.69	1.717	1.55	1.162	-2.697	5.333
Chips	0.311	0.311	0.059	0.074	0.555	1.918	1.827	1.086	-1.664	5.653
LabEq	0.467	0.467	0.059	0.248	0.636	1.692	1.66	0.83	-1.653	3.792
Paper	0.261	0.261	0.017	0.171	0.308	0.735	0.67	0.534	-2.074	3.057
Boxes	0.371	0.371	0.03	0.267	0.521	1.174	1.129	0.682	-2.147	4.111
Rtail	0.21	-0.21	0.029	-0.273	-0.127	0.395	0.318	0.339	-1.519	1.287
Meals	0.084	0.021	0.102	-0.239	0.274	0.994	0.964	0.524	-1.557	2.255
Banks	0.037	0.003	0.049	-0.168	0.155	0.639	0.336	0.787	-2.936	5.132
Panel B : Averages across industries										
Average in S_1	0.245	0.180				1.052	0.926			
Average	0.224	0.154				1.001	0.862			
Standard deviation	0.158	0.229				0.498	0.571			
t-stat $H_0 : avg = 0$	8.367	3.990				11.902	8.923			

Table 2.8 – Time-varying currency premiums for industry portfolios

Panel A reports the sample mean risk premiums associated with each risk factor, the sum of the individual currency premiums, the total risk premium, and the percentage of the total risk premium attributed to the currency risk premiums, respectively. The risk premium is the product of the respective price of risk and corresponding conditional covariance risk for each period. The total risk premium is the sum of the three equity and two currency risk premiums, in each period. For comparison across premiums, all premiums are reported as the mean of the absolute period-by-period premium. All risk factors are described in Table 2.2. Panel B contains summary statistics across all 48 industries. Because of the limited space only premiums for the set S_1 of the industries are the most likely exposed to currency risk, is displayed.

Industry	Market	SMB	HML	MJ	EM	Total Currency	Total premium	Currency as a % of total
Panel A : Summary of absolute time-varying risk premiums								
Agric	0.467	0.257	0.209	0.081	0.038	0.119	1.052	11.965
Food	0.362	0.124	0.161	0.059	0.029	0.088	0.735	13.546
Soda	0.482	0.097	0.145	0.038	0.048	0.086	0.81	10.018
Beer	0.4	0.126	0.196	0.134	0.066	0.2	0.921	23.611
Smoke	0.344	0.147	0.198	0.049	0.07	0.12	0.809	15.105
Toys	0.631	0.333	0.278	0.036	0.058	0.094	1.335	7.855
Fun	0.7	0.309	0.402	0.007	0.048	0.055	1.466	3.97
Books	0.545	0.117	0.196	0.033	0.038	0.071	0.928	8.168
Hshld	0.506	0.044	0.154	0.034	0.048	0.082	0.786	10.68
Clths	0.729	0.338	0.299	0.04	0.093	0.132	1.498	9.591
MedEq	0.442	0.188	0.376	0.031	0.029	0.06	1.066	5.991
Drugs	0.431	0.073	0.341	0.103	0.023	0.126	0.972	12.826
Chems	0.589	0.092	0.147	0.084	0.08	0.164	0.992	17.538
Rubbr	0.56	0.339	0.228	0.059	0.064	0.124	1.251	11.022
Txtls	0.579	0.349	0.152	0.075	0.092	0.167	1.247	14.175
BldMt	0.626	0.202	0.192	0.074	0.067	0.141	1.161	13.152
Cnstr	0.677	0.351	0.242	0.056	0.068	0.124	1.393	9.666
Steel	0.776	0.523	0.543	0.124	0.122	0.245	2.088	13.348
FabPr	0.644	0.503	0.303	0.077	0.137	0.214	1.663	14.523
Mach	0.662	0.425	0.431	0.111	0.119	0.23	1.748	14.225
ElcEq	0.746	0.115	0.445	0.111	0.054	0.165	1.472	11.475
Autos	0.725	0.218	0.138	0.04	0.092	0.132	1.212	11.573
Aero	0.674	0.159	0.142	0.031	0.096	0.127	1.103	12.213
Ships	0.517	0.25	0.16	0.116	0.061	0.177	1.103	17.545
Guns	0.342	0.111	0.208	0.035	0.059	0.094	0.755	13.744
Gold	0.439	0.336	0.107	0.26	0.09	0.35	1.231	28.309
Mines	0.649	0.29	0.125	0.119	0.165	0.284	1.348	22.161
Coal	0.611	0.419	0.284	0.17	0.164	0.334	1.649	22.766
Comps	0.856	0.514	1.055	0.169	0.17	0.338	2.763	14.283
Chips	0.748	0.628	0.944	0.106	0.18	0.286	2.605	13.471
LabEq	0.702	0.637	0.762	0.156	0.123	0.278	2.38	13.906
Paper	0.564	0.126	0.146	0.089	0.084	0.173	1.01	18.651
Boxes	0.607	0.114	0.215	0.124	0.097	0.221	1.156	19.331
Rtail	0.563	0.165	0.348	0.068	0.027	0.095	1.17	8.366
Meals	0.55	0.169	0.173	0.025	0.065	0.09	0.983	9.969
Banks	0.64	0.17	0.154	0.012	0.052	0.064	1.028	6.527
Panel B : Averages across industries								
Average in S_1	0.586	0.260	0.294	0.082	0.081	0.163	1.302	13.480
Average	0.573	0.249	0.291	0.074	0.077	0.151	1.265	13.050
Standard deviation	0.125	0.150	0.219	0.051	0.038	0.078	0.468	5.135
Minimum	0.249	0.044	0.070	0.007	0.023	0.055	0.623	3.970
Maximum	0.856	0.637	1.055	0.260	0.180	0.350	2.763	28.309

Table 2.9 – Time-varying currency betas for the 1oN and global minimum portfolios

Panel A reports the mean of the time-varying currency factor betas estimated from the system of Equations (2.19)-(2.21). The results displayed are for the benchmark 1oN, global minimum portfolios $GMP(\tau)$, where τ represents the percent of PCs used. Mean abs is the mean of the absolute betas. The currency factors are described in Table 2.2, and are in foreign currency per US dollar. Panel B reports the sample mean risk premium associated with each risk factor, the sum of the individual currency premiums, the total risk premium, and the percentage of the total risk premium attributed to the currency risk premiums. For comparison across premiums, all premiums are reported as the mean of the absolute period-by-period premium. Numbers in bold face are lower that their corresponding average across all 48 industries portfolios, and the lowest value is indicated by *.

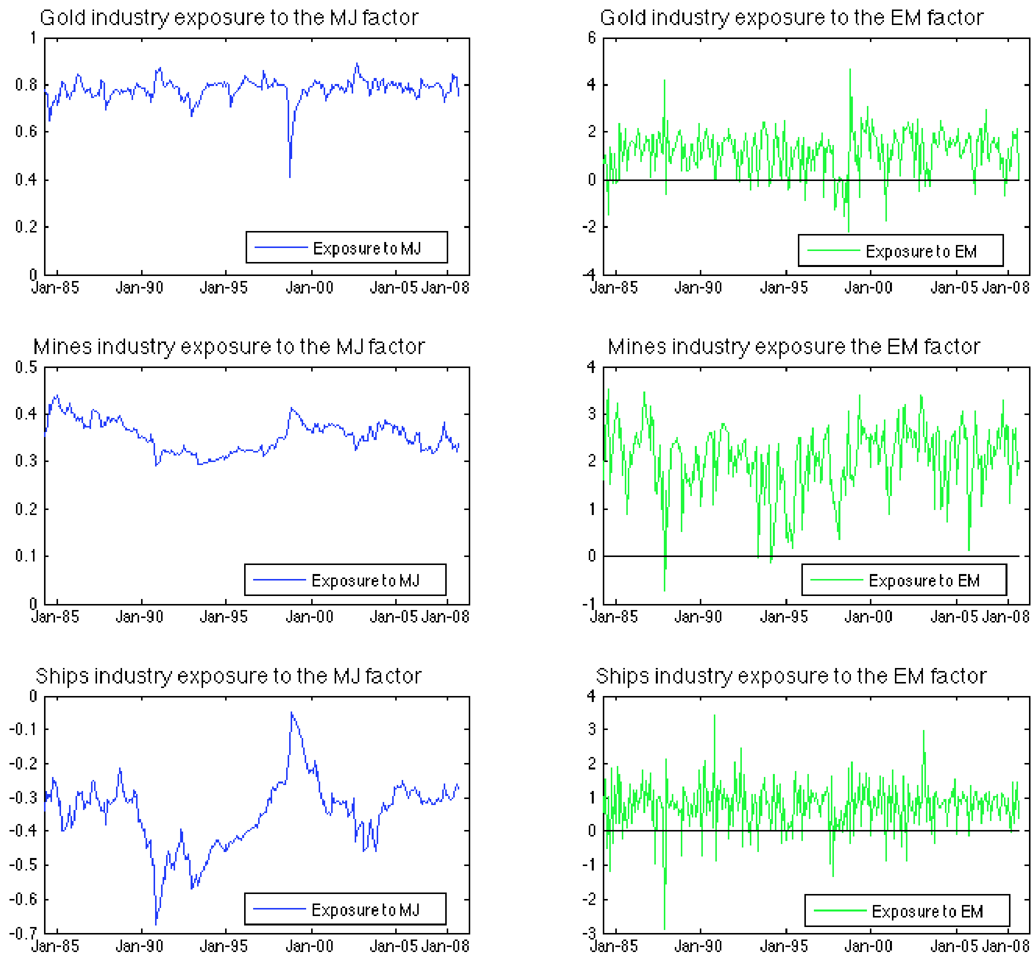
Panel A : Estimates of time-varying currency betas for optimal portfolios

	MJ					EM				
Portfolio	Mean Abs	Mean	Std	Min	Max	Mean Abs	Mean	Std	Min	Max
Cross Ind. Avg.	0.224	0.154				1.001	0.862			
1oN	**0.151**	**0.143**	0.121	-0.155	0.558	**0.927**	**0.860**	0.599	-2.602	3.398
Gmin(100%)	**0.165**	**0.165**	0.019	0.126	0.201	**0.669**	**0.665**	0.321	-0.163	1.44
Gmin(90%)	**0.205**	**0.205**	0.009	0.149	0.249	**0.516**	**0.500**	0.275	-0.576	1.227
Gmin(80%)	**0.141**	**0.141**	0.009	0.100	0.173	**0.473**	**0.451**	0.281	-0.686	1.204
Gmin(70%)	**0.122**	**0.122**	0.003	0.102	0.135	**0.293***	**0.279***	0.166	-0.328	0.708
Gmin(60%)	**0.113***	**0.113***	0.007	0.083	0.154	**0.408**	**0.390**	0.233	-0.516	1.402
Gmin(50%)	**0.163**	0.163	0.016	0.108	0.227	**0.445**	**0.397**	0.319	-0.82	2.028
Gmin(40%)	**0.225**	0.225	0.016	0.179	0.285	**0.627**	**0.602**	0.368	-0.926	2.225
Gmin(30%)	0.274	0.274	0.034	0.181	0.426	**0.560**	**0.491**	0.429	-1.459	2.315
Gmin(20%)	0.259	0.259	0.059	0.063	0.596	**0.693**	**0.578**	0.613	-2.348	2.902
Gmin(10%)	**0.202**	0.201	0.097	-0.035	0.699	**0.698**	**0.580**	0.612	-2.174	2.628

Panel B : Estimates of time-varying risk premiums for optimal portfolios

Portfolio	WM	SMB	HML	MJ	EM	Total Currency	Total premium	Currency as a % of total
Cross Ind. Avg.	0.573	0.249	0.291	0.074	0.077	0.151	1.265	13.050
1oN	**0.336**	**0.211**	**0.219**	**0.046**	**0.065**	**0.111**	**0.877**	14.999
Gmin(100%)	**0.340**	**0.047**	**0.036**	**0.055**	**0.050**	**0.104**	**0.527**	21.055
Gmin(90%)	**0.342**	**0.032**	**0.019**	**0.068**	**0.034**	**0.102**	**0.495**	21.602
Gmin(80%)	**0.378**	**0.022**	**0.043**	**0.047**	**0.032**	**0.080**	**0.523**	15.960
Gmin(70%)	**0.352**	**0.029**	**0.038**	**0.040**	**0.020**	**0.060**	**0.479**	13.251
Gmin(60%)	**0.375**	**0.026**	**0.072**	**0.037**	**0.030**	**0.067**	**0.540**	**12.942***
Gmin(50%)	**0.356**	**0.070**	**0.063**	**0.053**	**0.043**	**0.096**	**0.585**	17.753
Gmin(40%)	**0.376**	**0.061**	**0.073**	0.074	**0.041**	**0.115**	**0.625**	20.145
Gmin(30%)	**0.418**	**0.089**	**0.098**	0.088	**0.058**	**0.146**	**0.752**	21.936
Gmin(20%)	**0.424**	**0.109**	**0.117**	0.085	**0.055**	0.140	**0.789**	19.372
Gmin(10%)	**0.496**	**0.109**	**0.158**	**0.065**	**0.054**	**0.119**	**0.881**	14.652

FIGURE 2.2 – Industry portfolio exposures to MJ and EM currency risks
The Figure displays graphs of time-varying MJ and EM currency betas for a subsample of industries. The MJ currency risk is represented by percentage changes in the real US Treasury trade-weighted exchange rate index of the currencies of the main trading partners of the United States. The EM component of currency risk, is a similar risk factor made up of the currencies of other important trading partners from the emerging economies. Both indices are expressed as foreign currency per US dollar.



(Continued on next page)

Figure 2.2 (Continued)

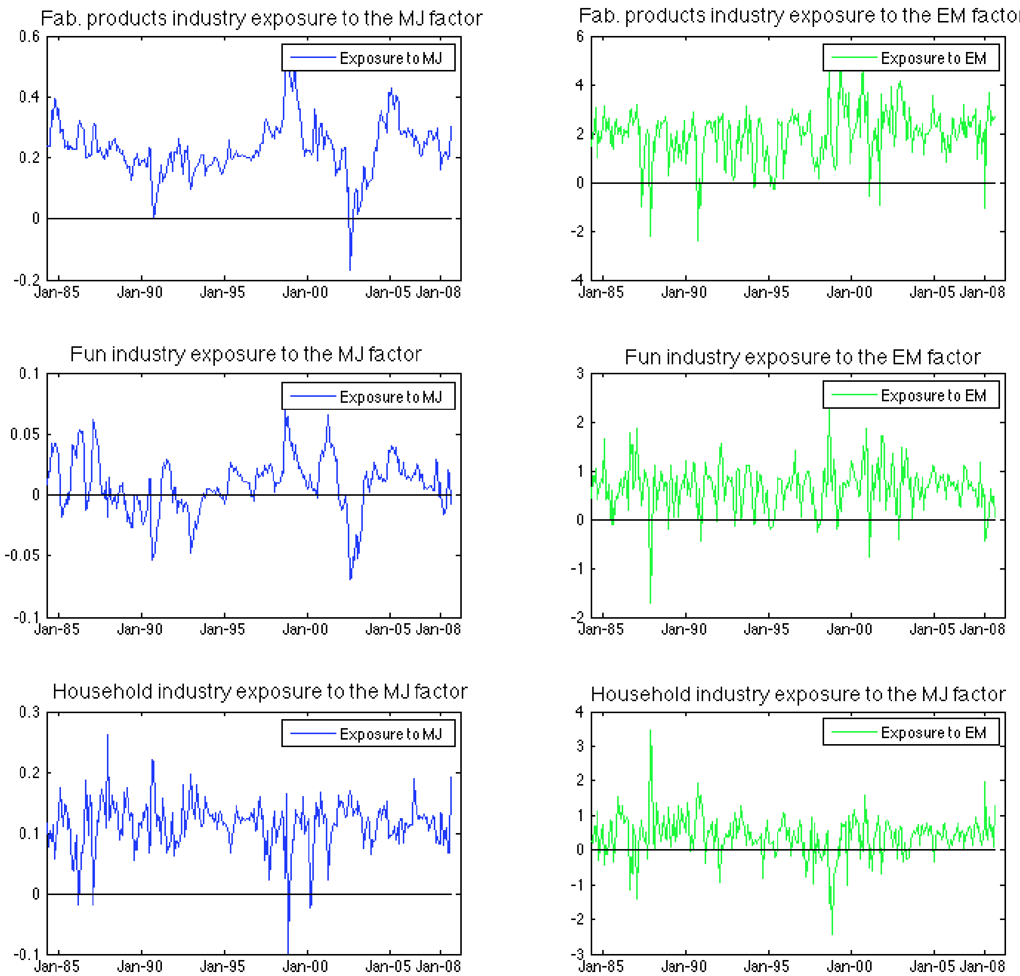
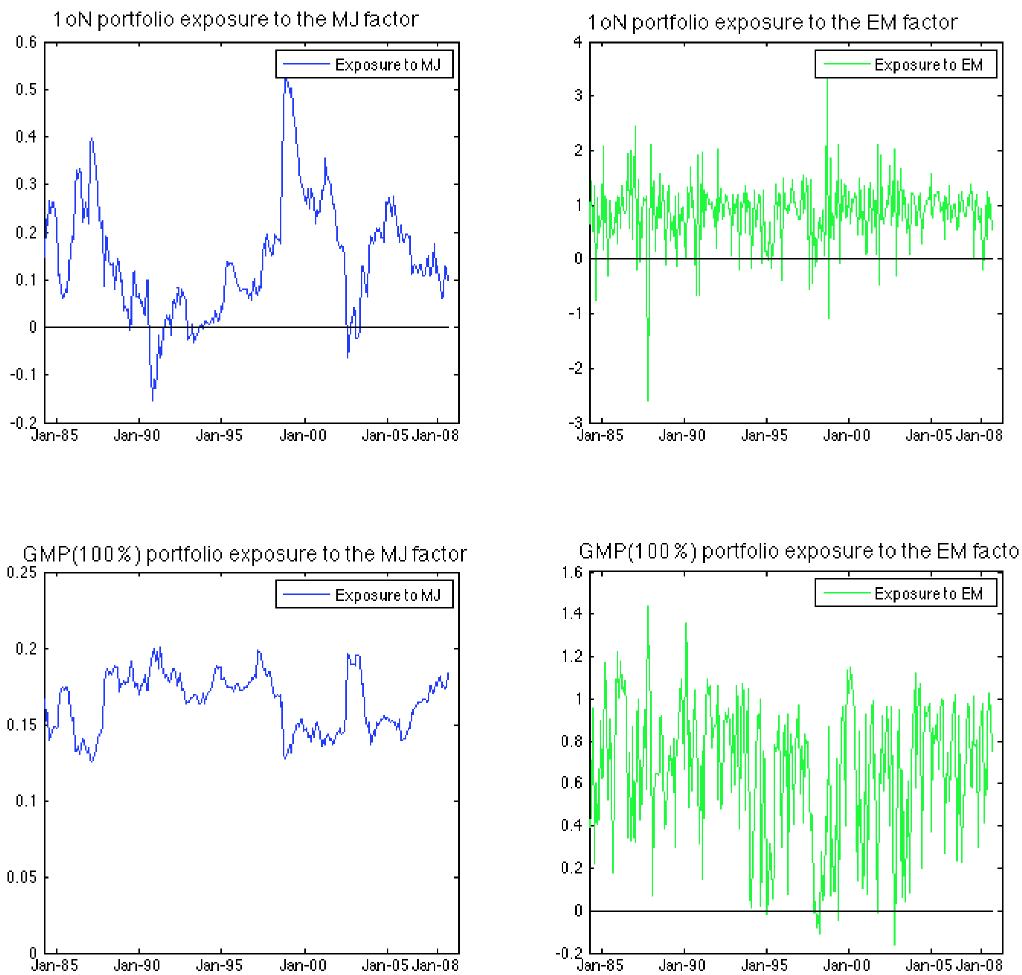


FIGURE 2.3 – Time-varying currency betas for optimal portfolios

The Figure plots the exposures of 4 optimal portfolios to MJ and EM currency risks : the equally weighted portfolio (1oN) and global minimum portfolios $GMP(\tau)$ for $\tau = 100\%, 70\%$, and 60%. The MJ currency risk is represented by percentage changes in the real US Treasury trade-weighted exchange rate index of the currencies of the main trading partners of the United States. The EM component of currency risk, is a similar risk factor made up of the currencies of other important trading partners from the emerging economies. Both indices are expressed as foreign currency per US dollar.



(Continued on next page)

Figure 2.3 (Continued)

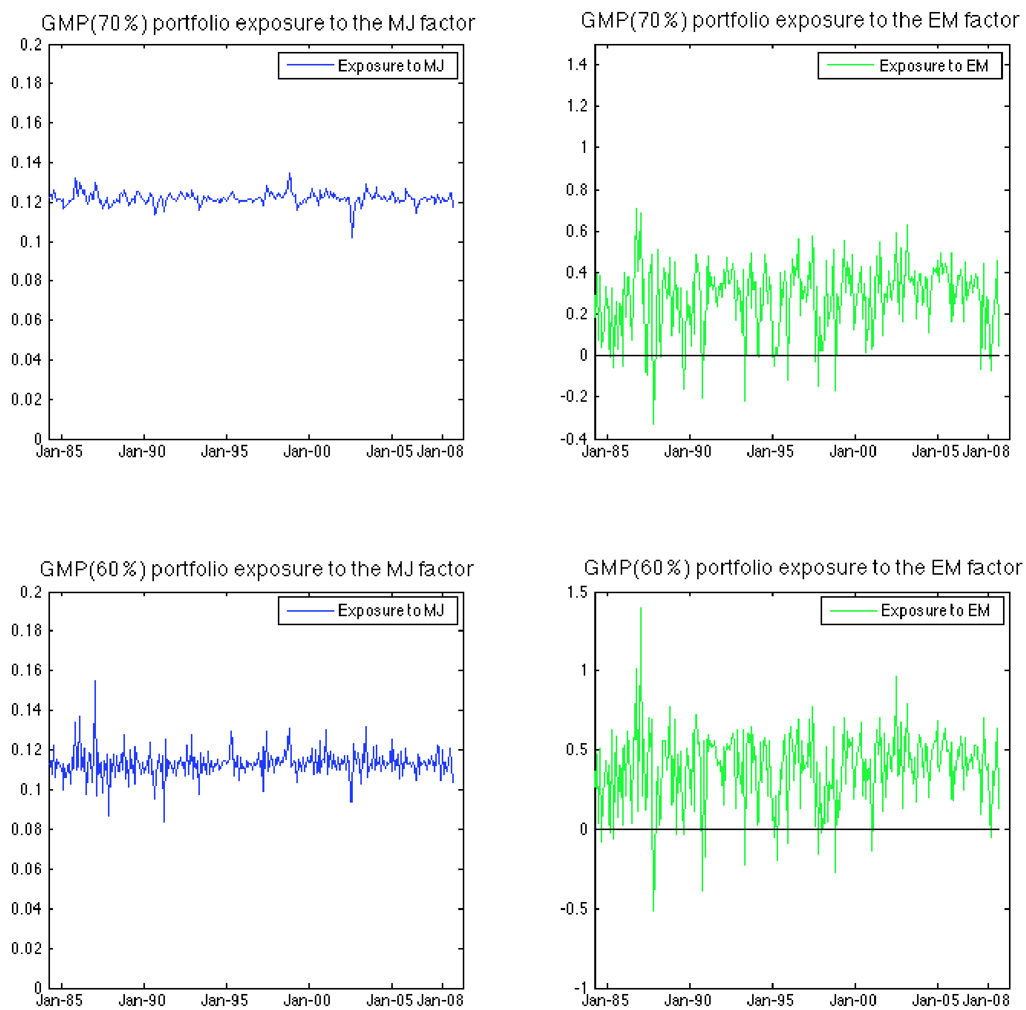


Figure 2.4 – Currency premium for selected industries

The Figure plots the currency risk premiums for a selected number of industry portfolios. The MJ currency risk is represented by percentage changes in the real US Treasury trade-weighted exchange rate index of the currencies of the main trading partners of the United States. The EM component of currency risk, is a similar risk factor made up of the currencies of other important trading partners from the emerging economies. Both indices are expressed as foreign currency per US dollar. The total currency premium is the sum of the premiums for currency risk and equity related risks

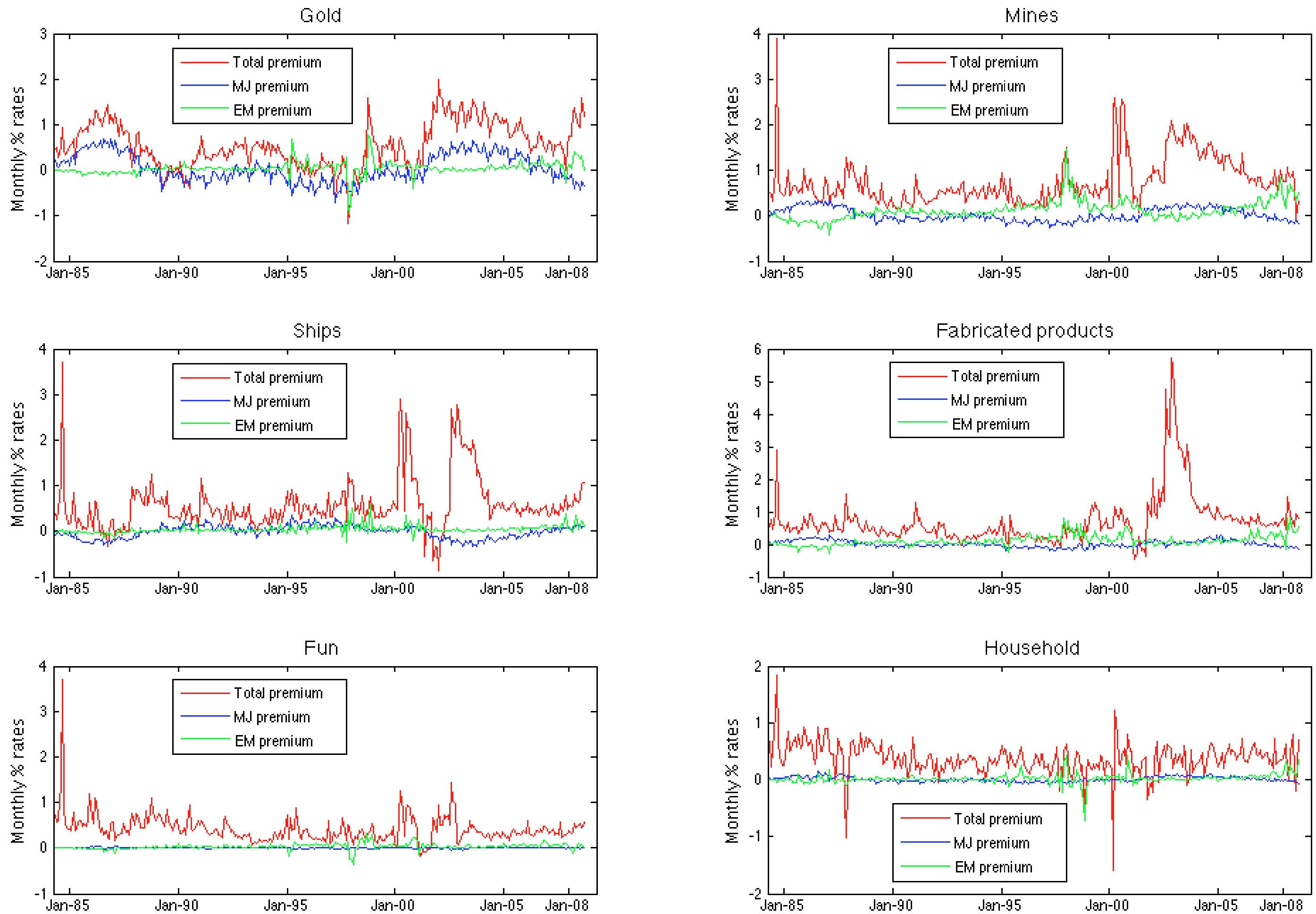
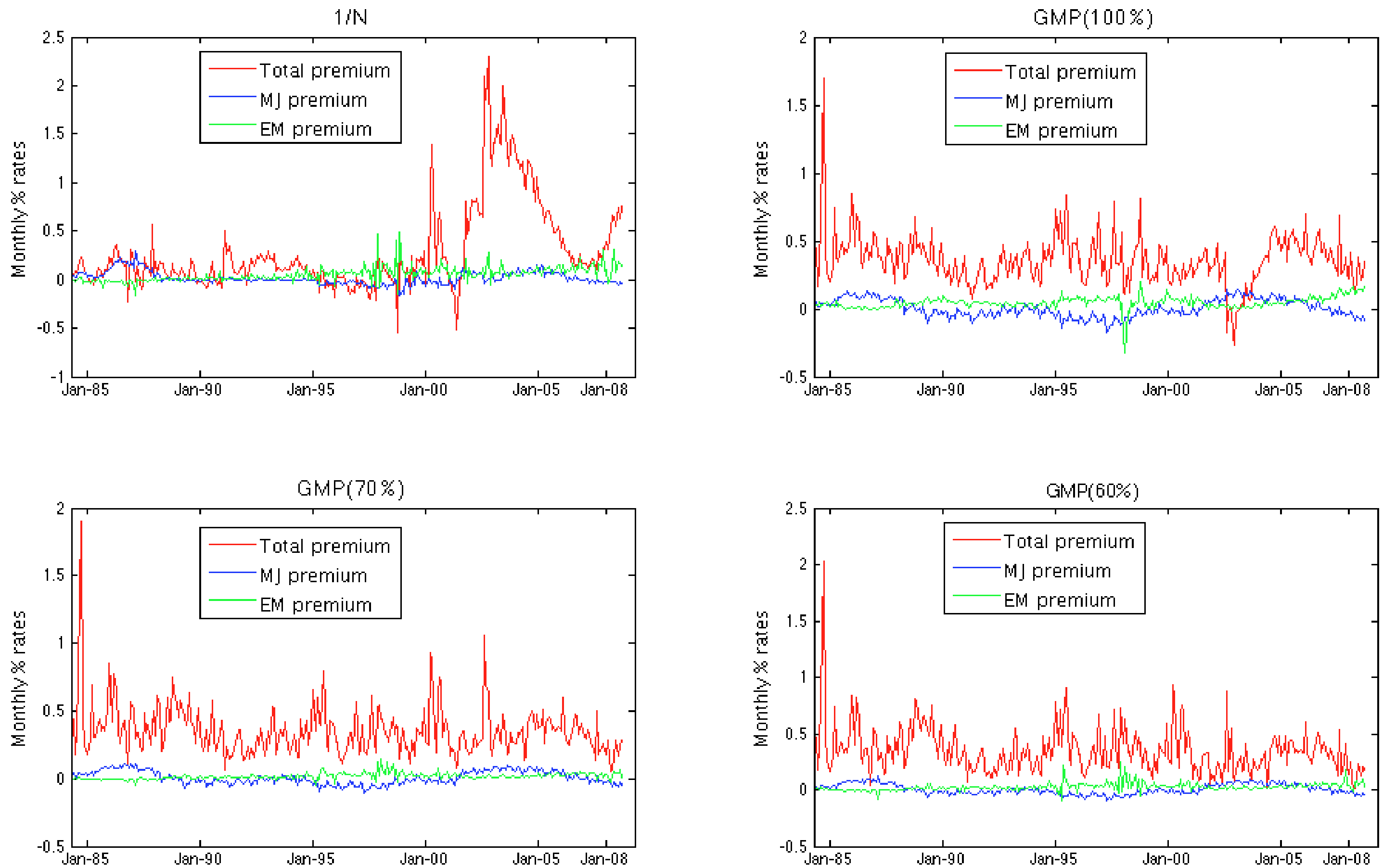


Figure 2.5 – Currency premium for optimal portfolios

The Figure plots the currency risk premiums of four optimal portfolios to MJ and EM currency risk. The equally weighted portfolio (1oN) and global minimum portfolios $GMP(\tau)$ for $\tau = 100\%, 70\%$, and 60%. The MJ currency risk is represented by percentage changes in the real US Treasury trade-weighted exchange rate index of the currencies of the main trading partners of the United States. The EM component of currency risk, is a similar risk factor made up of the currencies of other important trading partners from the emerging economies. Both indices are expressed as foreign currency per US dollar. The total currency premium is the sum of the premiums for currency risk and equity related risks



The Ljung-Box portmanteau statistic is used to test the null hypothesis of zero autocorrelation, for a maximum of 12 lags, in both the standardized residuals and the standardized residuals squared. For both industry portfolios and global minimum portfolios, we cannot reject the null hypothesis of zero mean and zero autocorrelation, for the standardized residuals. Besides, no dynamic is left in the squared residuals. These results show that the GARCH(1,1) specification captures the dynamics of the conditional second moments, for industry portfolios and global minimum portfolios.

Finally, we perform the robustness test also considered in De Santis and Gérard (1998) and Francis, Hasan, and Hunter (2008). The test consists in including the instruments described in Table 2.10, as exogenous variables in the conditional mean equation (2.17). Indeed, if the model is well specified, then the risk factors used to estimate the expected returns should remove all time variation in the portfolio returns, so that the instruments have no explanatory power in the presence of these factors. This test can alternatively be performed by regressing the unstandardized residuals from Equation (2.17) on a constant and the instruments. The p-value for the Wald tests, in the fifth column of the Table 2.10. indicates that the hypothesis of non predictability of unstandardized residual cannot be rejected in any portfolios considered.

2.6 Conclusion

The chapter investigates whether domestic optimal portfolios have a positive externality in terms of exchange rate risk exposure reduction, and how the currency premium is impacted. We restrict ourselves to the global minimum portfolio (GMP) strategy as this rule corresponds to the lowest risk among mean-variance portfolios, and to better out-of-sample performance. In addition, we considered a family of GMPs that extends the GMP, by considering different percentage of the principal components (PCs) of the covariance matrix of asset returns. The lower the percentage of PCs kept, the stabler the portfolio, as measured by the turnover. Thus this chapter also allowed us to investigate the links between portfolio stability and currency risk.

Our finding can be summarized in three main points. First, we find that currency risk remains consistently priced, time-varying for cross-sections of optimal domestic portfolios. However, under some level of stability, it is possible to completely eliminate the emerging markets component of currency risk, so that emerging markets currency risk is not significant. Second, although most industry portfolios are exposed to currency risk, either directly or indirectly, our results also suggest that global minimum portfolios have the potential to significantly reduce the exposure to currency risk by more than half its size. Third, for GMPs, currency risk is economically relevant and the currency premium contribution to total premium, remains approximately equal to the average across industries.

The assessment of the relevance of currency risk for GMPs is done in the framework of a conditional IAPM. Thus, the second chapter provides additional evidence on the pricing of currency risk, in the US. This work also has important implications in terms of hedging strategies for domestic investors. Precisely, adopting stable weights is likely to produce low and stable exposures to currency risks, then lead lower hedging cost.

Finally, we leave two important questions for future work. The first one, is related to the selection of the number of PCs (or level of stability) that lead to lower currency exposures. Ideas such as the one proposed in Chapter 1 could be used here. Second, the focus of the chapter was entirely on domestic portfolios, leaving opened the question of how optimal domestic portfolios are related to optimal portfolios diversified internationally, in terms of currency premiums and exposures. The answer to this last question, will contribute to the current debate of achieving the gains of international diversification without investing abroad.

Table 2.10 – Industry residual diagnostics

The table reports diagnostic tests on the residuals, ε_{it}, of the conditional mean model in Equation (2.19). The portfolios considered are 48 industry portfolios (In Panel B), and global minimum portfolios (In Panel A). The tests are conducted on the residuals standardized by the conditional standard deviation of the industry returns obtained from Equation (2.21). BJ is the Bera-Jarque test of the null hypothesis that the standardized residuals are normally distributed. $Q(12)$ is the Q-statistic of the test of the null hypothesis that the standardized residuals are not autocorrelated up to the 12th lag, and $Q^2(12)$ tests for autocorrelation in the squared residuals (i.e., a test for remaining heteroskedasticity in the residuals). The test in the fifth column tests the null hypothesis that the raw residuals from the conditional mean equation of the industry returns are not predictable, using the lagged instruments DUSTP, MtoGDP, USDP, XtoGDP, and FED (See Table 2.5). **And * represent significance at the 1% and 5% levels, respectively.

Panel A : Industry portfolio residuals					
Industry portfolio residuals standardized by conditional standard deviation					Unstandardized
	H_0 : Mean=0	H_0 : Not Normal (JB=0)	H_0 : Not autcorrelated $Q(12)$	H_0 : squared residuals not autocorrelated $Q^2(12)$	residuals. H_0 : Residuals are not Predictable
Agric	0.150	0.001	0.698	0.258	0.943
Food	0.034	0.001	0.515	0.239	0.498
Soda	0.443	0.001	0.349	0.740	0.990
Beer	0.029	0.002	0.442	0.937	0.091
Smoke	0.140	0.001	0.972	0.396	0.284
Toys	0.887	0.001	0.576	0.940	0.382
Fun	0.341	0.001	0.290	0.924	0.623
Books	0.694	0.043	0.023	0.315	0.123
Hshld	0.114	0.001	0.611	0.788	0.550
Clths	0.760	0.001	0.253	0.995	0.528
MedEq	0.015	0.001	0.475	0.812	0.928
Drugs	0.008	0.006	0.104	0.534	0.501
Chems	0.504	0.001	0.543	0.766	0.831
Rubbr	0.569	0.001	0.336	0.972	0.630
Txtls	0.911	0.001	0.100	0.998	0.422
BldMt	0.566	0.001	0.667	0.993	0.429
Cnstr	0.661	0.001	0.006	0.460	0.925
Steel	0.777	0.001	0.874	0.832	0.798
FabPr	0.981	0.001	0.546	0.648	0.459
Mach	0.606	0.001	0.075	0.987	0.940
ElcEq	0.138	0.001	0.409	0.996	0.557
Autos	0.973	0.001	0.445	0.861	0.181
Aero	0.687	0.001	0.590	0.544	0.882
Ships	0.879	0.001	0.950	0.966	0.839
Guns	0.348	0.001	0.667	0.089	0.939
Gold	0.926	0.001	0.194	0.433	0.311
Mines	0.841	0.001	0.534	0.821	0.837
Coal	0.660	0.001	0.685	0.944	0.701
Comps	0.343	0.041	0.547	0.022	0.689
Chips	0.366	0.001	0.990	0.100	0.301
LabEq	0.450	0.001	0.608	0.060	0.826
Paper	0.701	0.001	0.100	0.002	0.736
Boxes	0.464	0.001	0.989	0.935	0.369
Rtail	0.130	0.001	0.223	0.943	0.628
Meals	0.576	0.001	0.398	0.979	0.801
Banks	0.663	0.001	0.855	1.000	0.310

(Continued on next page)

Table 2.10 (Continued)

Panel B : Optimal portfolio residuals					
Optimal portfolio residuals standardized by conditional standard deviation					Unstandardized
	H_0 : Mean=0	H_0 : Not Normal (JB=0)	H_0 : Not autocorrelated $Q(12)$	H_0 : squared residuals not autocorrelated $Q^2(12)$	residuals. H_0 : residuals are not Predictable
1oN	0.187	0.001	0.574	0.999	0.644
GMP(100%)	0.945	0.001	0.164	0.834	0.690
GMP(90%)	0.885	0.001	0.266	0.963	0.894
GMP(80%)	0.679	0.001	0.134	0.977	0.862
GMP(70%)	0.355	0.001	0.322	0.999	0.877
GMP(60%)	0.139	0.001	0.437	0.013	0.788
GMP(50%)	0.059	0.054	0.100	0.000	0.503
GMP(40%)	0.072	0.001	0.310	0.031	0.572
GMP(30%)	0.016	0.004	0.259	0.197	0.557
GMP(20%)	0.083	0.001	0.079	0.935	0.473
GMP(10%)	0.361	0.001	0.369	0.998	0.759

Bibliographie

[1] Adler, M. and Dumas, B., 1983. "International Portfolio Choice and Corporation Finance : A Synthesis," *The Journal of Finance*, 38, 925-985.

[2] Bailey, W. and Y.P. Chung, 1995. "Exchange rate Fluctuations, Political Risk and Stock Returns : Some Evidence from an Emerging Market," *Journal of Financial and Quantitative Analysis*, 30, 541-560.

[3] G. and C.R. Harvey, 1995. "Time Varying World Market Integration," *Journal of Finance*, 50, 403-444.

[4] Berndt, E. K., B.H. Hall, Robert Hall and Jerry Hausman, 1974. "Estimation and Inference in Nonlinear Structural Models," *Annals of Economics and Social Measurement*, 3, 653-665.

[5] Campbell, J. Y., K. Serfaty-De Medeiros, and Luis M. Viceira, 2010. "Global Currency Hedging," *Journal of Finance*, 65, 87-121.

[6] Brodie, J., I. Daubechies, C. De Mol, D. Giannone, I. Loris, 2009. "Sparse and stable Markowitz portfolios,"Proceedings of the National Academy of Sciences of the USA, 106, 12267-12272.

[7] Carrieri, F., 2001. "The Effects of Liberalization on Market and Currency Risk in the EU," *European Financial Management*, 7, 259-290.

[8] Cooper, I., Kaplanis, E., 1994. "Home bias in equity portfolios inflation hedging, and international capital market equilibrium," *Review of Financial Studies*, 7, 45-60.

[9] Chaieb, I., and Errunza, V., 2007. "International asset pricing under segmentation and PPP deviations," *Journal of Financial Economics*, 86, 543-578.

[10] Carrieri, F., Errunza, V. and Majerbi, B., 2006a. “Does Emerging Market Exchange Risk Affect Global Equity price,” *Journal of Financial and Quantitative Analysis*, 41, 511-540.

[11] Carrieri, F., Errunza, V. and Majerbi, B., 2006b. “Local Risk Factors in Emerging Markets : Are there separately priced ?,” *Journal of Empirical Finance*, 13, 444-461.

[12] Choi, J.J., T. Hiraki, and N. Takezawa, 1998. “Is Foreign Exchange Risk Priced in the Japanese Stock Market ?,” *Journal of Financial and Quantitative Analysis*, 33, 361-382.

[13] Choe, H., Kho, B. C., and Stulz, R. M., 2005. “Do domestic investors have an edge ? The trading experience of foreign investors in Korea,” *Review of Financial Studies*, 18, 795-829.

[14] Christoffersen, P., Errunza, V., Jacobs, K., Jin X., 2012. “Is the Potential for International Diversification Disappearing ?,” *Review of Financial Studies*, 25, 3711-3751.

[15] De Santis, G., Gérard, B., Hillion, P., 2003. “The Relevance of Currency Risk in the EMU,” *Journal of Economics and Business*, 55, 427-462.

[16] De Santis, G., Gérard, B., 1998. “How Big is the Premium for Currency Risk,” *Journal of Financial Economics*, 49, 375-412.

[17] DeMiguel, V., L. Garlappi, F. J. Nogales, R. Uppal, 2009. “A Generalized Approach to Portfolio Optimization : Improving Performance by Constraining Portfolio Norms,” *Management Science*, 55, 798-812.

[18] DeMiguel, V., L. Garlappi, R. Uppal, 2007. “Optimal Versus Naive Diversification : How Inefficient is the 1/N Portfolio Strategy ?,” *The review of Financial studies*, 22, 1915-1953.

[19] Doukas, J., P. Hall and L. Lang, 1999. “The Pricing of Currency Risk in Japan,” *Journal of Banking and Finance*, 23, 1-20.

[20] Dumas, B., Solnik, B., 1995. “The World Price of Foreign Exchange Risk,” *The Journal of Finance*, 50, 445-479.

[21] Driessen, J., Laeven, L., 2007. "International portfolio diversification benefits : Cross-country evidence from a local perspective," *Journal of Banking & Finance*, 31, 1693-1712.

[22] Ehling, P., S.B. Ramos, 2006. "Geographic versus industry diversification : Constraints matter," *Journal of Empirical Finance*, 13, 396-416.

[23] Errunza, V., Hogan, K. and Hung, M.-W., 1999. "Can the Gains from International Diversification Be Achieved without Trading Abroad ?," *The Journal of Finance*, 54, 2075-2107.

[24] Ferson, W. and C.R. Harvey, 1993. "The Risk and Predictability of International Equity Returns," *Review of Financial Studies*, 6, 527-566.

[25] Ferson, W.E. and C.R. Harvey, 1994. "Sources of Risk and Expected Returns in Global Equity Markets," *Journal of Banking and Finance*, 18, 775-803.

[26] Francis, B., Hasan, I., and Hunter, D. M., 2008. "Can hedging tell the full story ? Reconciling differences in United States aggregate- and industry-level exchange rate risk premium," *Journal of Financial Economics*, 90, 169-196.

[27] Goetzmann, W., Lingfeng Li, and K. Geert Rouwenhorst, 2005. "Long Term Global Market Correlations," *The Journal of Business*, 78, 1-38.

[28] Harvey, C.R., 1995. "Global Risk exposure to a Trade-Weighted Foreign Currency Index," *NBER Working paper*, Duke University.

[29] Jorion, P., 1991. "The Pricing of Exchange Rate Risk in the Stock Market," *Journal of Financial and Quantitative Analysis*, 26, 361-376.

[30] Kang, J. K., Stulz, R., 1997. "Why is there a home bias ? An analysis of foreign portfolio equity ownership in Japan," *Journal of Financial Economics*, 46, 3-28.

[31] Karolyi A. and R.M. Stulz, 2002. "Are Financial Assets Priced Locally or Globally ?," Working Paper, Ohio State University.

[32] French, K. R., Poterba, J. M., 199. "Investor Diversification and International Equity Markets," *The American Economic Review* , 81, 222-226.

[33] Levy, H., Sarnat, M., 1970. "International diversification in investment portfolios," *American Economic Review*, 60, 668- 675.

[34] Li, K., Sarkar, A., Wang, Z., 2003. "Diversification benefits of emerging markets subject to portfolio constraints," *Journal of Empirical Finance*, 10, 57-80.

[35] Merton, R. C. 1980. "On Estimating the Expected Return on the Market : An Exploratory Investigation," *Journal of Financial Economics*, 8, 323-61.

[36] Soenen L. A., 1985. "The optimal currency cocktail - a tool for strategic foreign exchange risk management," *Management international Review*, 25, 12-22.

[37] Solnik, B., 1974. "An Equilibrium Model of the International Capital Market," *Journal of Economic Theory*, 8, 500-524.

[38] Solnik, B., 1995. "Why Not Diversify Internationally Rather Than Domestically ?" *Financial Analysts Journal*, 51, 89-94.

[39] Stulz, R.M., 1981. "A Model of International Asset Pricing," *Journal of Financial Economics*, 9, 383-406.

[40] White, H., 1982. "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1-25.

[41] Phylaktis, K., Ravazzolo, F., 2004. "Currency Risk in Emerging Equity Markets," *Emerging Market Review*, 5, 317-339.

Chapitre 3

Estimating the local risk premium for pessimistic decision makers

3.1 Introduction

The notion of risk premium is one of the most widely used concept of the modern theory of decision making under uncertainty and risk. When deriving the risk premium under a given preferences assumption, the expected utility (EU) framework has been mostly used. However, frequent violations of EU axioms in both experimental and real-life situations (e.g Allais paradox) have been reported. Some of these limitations of the EU theory can be tackled, by considering the rank-dependent expected utility (RDEU or simply RDU) framework, introduced by Quiggin (1982), Schmeidler (1989), and Yaari (1987).

Under the RDU assumption, the decision maker evaluates risky alternatives with a utility function over outcomes and a distortion function that transforms the probability distribution of outcomes. Different assumptions on the distortion function leads to different behaviors of the decision maker. In this chapter, we are interested in the behavior of increasing pessimism of Quiggin (1982), and we call it pessimism as Bassett et al. (2004). This notion of pessimism is induced by a concave distortion function that accentuates the probability weights associated with least favorable outcomes.

Defining a notion of risk premium for RDU involves two difficulties. First, a well known characteristic of rank-dependent preferences stressed in the literature is that pessimism is rather difficult to distinguish empirically from risk aversion (e.g. Quiggin (1982)). To isolate the risk premium only due to the pessimism of the decision maker, we consider risk neutral decision makers with linear utility function over outcomes. The second issue comes from the nonlinearity in probability introduced by the distortion function. Indeed, the original definition of Pratt (1964) assumes that decision makers are expected utility maximizers, and use a specific utility function over wealth. Therefore, deriving the risk premium requires finding the utility function used by decision makers to evaluate outcomes, and which corresponds to an expected utility maximization criterion. The second issue can be solved by assuming that the preference functional over alternative probability distributions is smooth, so that the required utility function can be derived, as the local utility function introduced by Machina (1982). Indeed, Machina established that any smooth preference functional can have, around a given risk, an expected utility representation and thereby, inherit the basic concepts, tools and results of expected utility analysis. By solving the mentioned difficulties, the risk premium and coefficient of risk aversion that we obtain are only due to the pessimistic behavior of the decision maker. In particular, the coefficient of risk aversion can be used to compare and summarize the degree of pessimism.

The literature on RDU is closely related to that of risk measurement. Instead of a definition from the economic theory of choice under uncertainty, the latter literature introduces risk measures satisfying desirable properties. One of the most prominent endeavor in such a direction is the seminal chapter of Artzner et al. (1999), who introduced the class of coherent risk measures satisfying four desirable properties : monotonicity, invariance with respect to drift, homogeneity and subadditivity. In turn, these properties lead to appealing representations as weighted sum of quantiles (Kusuoka (2001)), when comonotonicity and law invariance are further assumed. This last representation describes the wider family of distortion risk measures (DRM), that distort the probability measure while using a linear utility. Therefore, DRMs are directly linked to rank-dependent utility functions, as already pointed out by many authors (e.g. Wang (1996, 2000), Gourieroux and Liu (2006)).

Considering rank-dependent utilities is particularly relevant in view of their appealing behavioral interpretation, and their increasing use in Finance and Insurance. For example, they are used as objective function in portfolio optimization to generalize the mean-variance analysis, [e.g. Rockafeller and

Uryasev (2000), Krokhmal et al. (2002), Adam et al. (2008)], and for fixing the reserves needed to balance or hedge a risky investment. The most famous member of DRMs is certainly the Conditional Value-at-Risk (CVaR), advocated by Artzner et al. (1999) to replace the Value-at-risk (VaR) proposed since 1996 by the Basel Committee on Banking Supervision. In effect, contrarily to the VaR, the CVaR also satisfies the subadditivity property and takes into account the magnitude of the loss when it occurs.

Many statistical tools have been developed to analyze rank-dependent preferences, mainly as distortion risk measures. Gourieroux and Liu (2006a) estimated the efficient portfolio allocation when distortion risk measures define the objectives and the constraints ; Gourieroux and Liu (2006b) provided a unified statistical framework for the analysis of distortions risk measures and of their sensitivity ; Scaillet (2004) considered a nonparametric method to estimate the CVaR and its sensitivities using kernel estimators ; Bassett et al. (2004) used Choquet utility for pessimistic portfolio optimization by solving a linear quantile regression problem. Although considerable effort has been devoted to provide tools for the statistical study of rank-dependent preferences, to the best of our knowledge, our study is the first attempt to provide such tools for the analysis the risk premium associated with rank-dependent utilities.

The contribution of the chapter is twofold. First, we derive a notion of risk premium that we call local risk premium and propose a measure of the degree of pessimism for decision makers with rank-dependent utility. By doing so, we generalize the commonly used notion of risk premium derived under the expected utility framework. To achieve this result, we derived the local utility function using Machina's approach and obtained the associated risk premium from the original definition of Pratt (1964). This leads to a measure with the appealing property of depending only on the distribution of the risk considered and on a distortion function. Our second contribution is to provide statistical procedures for the nonparametric analysis of the local utility functional associated with rank-dependent preferences and the corresponding risk premium. More specifically, estimators are proposed, and the consistency and the asymptotic properties of these estimators are established. This is done for a general rank-dependent utility with given distortion function, and for CVaR preferences.

The rest of the chapter is organized as follows. Section 2 sets some notations and defines the notions, the concepts and tools used throughout the chapter. Section 3 derives the local utility of rank-dependent preferences around a given distribution, and the corresponding risk premium and measure of pessimism. In Section 4, estimation procedures are proposed for the measures derived in Section 3, along with consistency and asymptotic normality results. Section 5 assesses the proposed estimators through Monte Carlo simulations, and presents an empirical illustration. Finally, Section 6 presents future works and concludes.

3.2 Notations and basic framework

3.2.1 Notations

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and Y denote the value of a given random prospect (profit and loss variable) defined on $(\Omega, \mathcal{F}, \mathbb{P})$. There is a profit if Y is positive and a loss otherwise. The loss and profit variable X associated with Y is $X = -Y$. We denote by F_X the cumulative distribution function (c.d.f) of X and by Q_X its quantile function, respectively defined by $F_X(x) = P(X \leq x)$ and $Q_X(t) = \inf\{x : F_X(x) > t\}$. We call L^2 the set of real-valued random variables with finite second moment. All preference functionals will be considered on the subset $\mathcal{D}$ of L^2, containing loss variables with a density relative to the Lebesgue measure.

Let X and Z be two elements of $\mathcal{D}$. Denote $X \succeq_1 Z$ (resp $X \succeq_2 Z$), X dominates Z according to first-order (resp. second-order) stochastic dominance. Then $X \succeq_1 Z$ (resp. $X \succeq_2 Z$) if and only if $F_X(x) \leq F_Z(x)$ (resp. $\int_{-\infty}^{x} F_X(u)du \leq \int_{-\infty}^{x} F_Z(u)du$), for all x. X and Z are comonotonic if there exist a random variable U and two increasing functions ϕ and ψ such that $X = \phi(U)$ and $Z = \psi(U)$ almost surely. A functional Φ defined on $\mathcal{D}$ is called comonotonic additive (CA) if $\Phi(X + Z) = \Phi(X) + \Phi(Z)$ when X and Z are comonotonic. Φ is said to be law invariant if $\Phi(X) = \Phi(Z)$ when X and Z have the same distribution.

Let $\{x_t\}_{t=1}^{T}$ be a given i.i.d sample of size T from the random variable X, where x_t denotes the observation of x at time t; $x_{(k)}$ denotes the order statistic of order k, that is the k^{th} element when the x_t are arranged in increasing order. The empirical c.d.f and the empirical quantile function of X, based on the given sample, are respectively $\hat{F}_T(x) = \frac{1}{T}\sum_{t=1}^{T} 1_{x_t \leq x}$ for all $x \in \mathbb{R}$ and $\hat{Q}_T(u) = \inf\left\{x : \hat{F}_T(x) \geq u\right\}$ for all $u \in [0, 1]$. For a fixed real number x, we denote by $t_l(x)$ the number $T\hat{F}_T(x)$ of observations before x, which can also be written as $t_l(x) = \sup\left\{t \,:\, x_{(t)} \leq x\right\}$.

3.2.2 Basic framework

The basic framework of the chapter involves the notion of rank-dependent preferences and the notion of local utility required to derive the local risk premium. This Section defines these notions.

Rank-dependent utility function

The rank-dependent expected utility (RDU) or anticipated utility theory is an extension of the classical expected utility theory, were the distribution of the random variable is replaced by a distorted version. If the decision maker evaluates outcomes with a utility function u, and transforms probability assessments with a distortion function h, his RDU is obtained by averaging the utility-equivalent of monetary outcomes, $u(-Q_X(1-t))$ [1], using a weighting function $dh(t)$.

The function h is used to attach probability to outcomes depending on their utility ranking and reflects the decision maker's attitude toward the risk distribution. A concave distortion function h corresponds to decreasing weights associated to outcomes ranked in increasing order of utility. Therefore, with such distortion, the implicit likelihood of least-favorable outcomes is accentuated, while the likelihood of the most-favorable outcomes is depressed. RDU preferences with concave distortion then corresponds to a "pessimistic" behavior, as argued by Bassett et al. (2004). Following this argument, we define pessimism in the following way :

Definition 1. *Pessimism*
A decision maker with rank-dependent preferences is said to be pessimistic if the distortion function h he uses to assess probabilities is concave.

The notion of pessimism in Definition 1 also corresponds to Quiggin's (1982) notion of increasing pessimism. A more general definition of pessimism introduced by this author, but that we do not consider in this chapter, imposes on the distortion h, the condition $h(\alpha) \leq \alpha$, for every probability level α.

1. Since X is the loss variable, the $-Q_X(1-t)$, with $t \in [0,1]$, represent the quantiles of the profit in monetary terms.

The functional form of the RDU theory was introduced by Quiggin (1979) as a generalization to the expected utility theory, following the violation of the dominance axiom reported in many studies (e.g. Kahneman and Tvesky (1979)). Early axiomatization of the RDU theory, include Quiggin (1982) and Yaari (1987). In the sequel, we adopt Definition 2 as the definition of rank-dependent preferences. For expository purposes, we consider the particular case of linear utility function ($u(x) = x$), which corresponds to a risk-neutral decision maker. In addition the choice of a linear function will be key in defining a measure of pessimism. As we will see subsequently, a risk neutral decision maker still pays a premium if he is pessimistic, that is, if his distortion function is concave and nonlinear.

Definition 2. *Rank-dependent preferences*
Let h be a c.d.f on $[0,1]$. A preference functional Ψ is said to characterize rank-dependent preferences, with distortion h, if for all loss and profit variable X in $\mathcal{D}$, we have

$$\Psi = \Psi_h(X) = -\int_0^1 Q_X(1-\alpha)dh\,(\alpha)\,, \tag{3.1}$$

where $Q_X(1-\alpha)$ is the $(1-\alpha)^{th}$ quantile of X.

The definition of RDU that we adopt is related to many notions in the literature of decision making under uncertainty and risk. It can be related to the Yaari functional representation (1987) of RDU. Indeed, using $h(0) = 0$ and $h(1) = 1$, and by integrating by parts, we obtain that $\Psi_h(X) = -\int h\left(S_X\left(x\right)\right)dx$, where $S_X(t) = 1 - F_X(t)$ is the survival function of the random variable X. Rank-dependent utility can also be expressed as the Choquet expectation of X with respect to a capacity [2] μ, denoted by $E_\mu X$. To see this, remark that by choosing μ, defined directly on events by $\mu(A) = -h(\mathbb{P}(A))$, for all $A \in \mathcal{F}$, $\Psi_h(X)$ can be rewritten as

$$\Psi_h(X) = \int_{-\infty}^{0} \left(\mu\left(\{X > x\}\right) - 1\right)dx + \int_0^{+\infty} \left(\mu\left(\{X > x\}\right)\right)dx = E_\mu X.$$

Parallel to the literature on decision making under uncertainty, RDU are widely used in the literature of risk assessment, mostly as distortion

2. A Choquet capacity is a normalized monotone set function. That is, for two events A and B : $A \subseteq B \Rightarrow \mu(A) \leq \mu(B)$, $\mu(\emptyset) = 0$ and $\mu(\Omega) = 1$.

risk measure (DRM) (e.g. Wang (1996), Gourieroux (2010)). The DRM with distortion function h of a loss X, that we denote $\Pi_h(X)$, is obtained as $\Pi_h(X) = \int_0^1 Q_X(1-\alpha)dh\,(\alpha)$, that is $\Psi_h(X) = -\Pi_h(X)$. The negative sign in the expression of $\Psi_h(X)$ means that a decision maker with rank-dependent preferences is better off, the lower the distortion measure of the risk he is facing.

In general, families of rank-dependent functions can be obtained by using different families of distortions. Example of distortions are summarized in Table 3.1. The power-law ($Pow(\theta)$) parametrization is given by $h(u,\theta) = 1-(1-u)^\theta$ with $\theta \in [1,+\infty[$, while the exponential ($Exp(\theta)$) parametrization is characterized by the distortion $h(u,\theta) = \frac{1-e^{-\theta u}}{1-e^{-\theta}}$. The most famous representative of the class of DRMs is the conditional value-at-risk ($CVaR$) at risk level p, with distortion function $h\,(\alpha,p) = \frac{\alpha}{p} \wedge 1$ parametrized by the risk level $p \in [0;1]$.

Example 1. *Conditional Value-at-risk ($CVaR_p(X)$)*
The CVaR at risk level p of a loss $X \in \mathcal{D}$ is defined as the average loss under the condition that X exceeds its $(1-p)^{th}$ quantile, also called the value-at-risk at risk level p and denoted $VaR_p(X)$. That is :

$$CVaR_p(X) = E\left(X|\, X \geq VaR_p(X)\right), \tag{3.2}$$

with $VaR_p(X)$ such that $P\left(X \geq VaR_p(X)\right) = p$.

Table 3.1 – Example of families of distortion

Distortion	$h(u,p)$	Parameter
Power	$1-(1-u)^p$	$p \in [1,+\infty]$
Exponential	$\frac{1-e^{-pu}}{1-e^{-p}}$	$p \in [1,+\infty]$
CVaR	$\frac{u}{p} \wedge 1$	$p \in [0,1]$

From expression (3.1), rank-dependent preferences are not linear in the probability distribution of X, which is distorted by the nonlinear function h. Therefore, for pessimistic decision makers, the expected utility framework cannot be used to derive the risk premium. In order to apply a definition of risk premium in the same spirit as the original definition of Pratt(1964), we need to find the conditions under which the rank-dependent preference can be replaced by an expected utility criterion, and derive the corresponding utility function over outcomes used in such a context.

The solution is provided by the notion of local utility function obtained by approximating the functional Ψ_h around the risk F_X using Machina's (1982) results. The specific condition needed is the smoothness of the preference functional Ψ_h. The local utility can then be used to rank risky alternatives around F_X and derive notions of local risk premium and local risk aversion.

Local utility function

Definition 3. *(Machina (1982)) Consider a real-valued preference functional Ψ on $\mathcal{D}$, continuous relative to the topology of convergence in distribution. Furthermore, suppose Ψ smooth (or differentiable) on $\mathcal{D}$ and let F be a distribution in $\mathcal{D}$. The local utility function relative to Ψ at F is the Fréchet derivative of Ψ at F, that is the function $U(\cdot;F)$ such that :*

$$\Psi(F^*) - \Psi(F) - \int U(x;F)(dF^*(x) - dF(x)) \to 0, \tag{3.3}$$

when F^* converges to F.

The notion of local utility function extends the von Neumann-Morgenstern (1944) framework to cases where the independent axiom is violated but preferences are smooth. Indeed, under the smoothness of his preference functional, a decision maker acts around a given random prospect X as an expected utility maximizer. That is, there exists a function $U(\cdot;F_X)$, such that for a differential shift F_X^* of F_X, the decision maker will prefer F_X^* to F_X, if and only if $\int U\,(x;F_X)dF_X^* \geq \int U\,(x;F_X)dF_X$.

Machina's results can be extended to cases where the smoothness assumption does not hold as considered in Safra and Segal (2002). In this study we suppose that this assumption is always satisfied.

3.3 Local utility function and risk premium for rank-dependent preferences

3.3.1 Local utility function for rank-dependent preferences

In this Section, we apply Machina's definition to derive the local utility function for rank-dependent preferences defined by Equation (3.1). The expression for the local utility in the RDU setting seems to be known in the literature (e.g. Segal (1985)). For completeness, we present this result in Proposition 3, and prove it in appendix A, given that we have been unable to find an existing proof. In the remainder, the distortion function h is supposed to satisfy assumption A.

Assumption A
The function h satisfies the following conditions

1. $h\ :\ [0;1] \mapsto R_+$
2. h is increasing and concave on $[0;1]$
3. $h'(1) < \infty$

Proposition 3. *Consider a distortion preference Ψ over the set $\mathcal{D}$, defined as in Equation (3.1).*
The local utility function $U_h(\cdot; F_X)$ relative to preference Ψ at F_X, evaluated at x is given by :

$$U_h(x; F_X) = -\int_{-\infty}^{x} h'\left(1 - F_X\left(t\right)\right) dt = -\int_{-\infty}^{x} h'\left(S_X\left(t\right)\right) dt. \tag{3.4}$$

In the sequel, if the distribution F_X is known, $U_h(x; F_X)$ will be simply denoted by $U_h(x)$.

For $X \in \mathcal{D}$, the first and the second order derivative of U_h at x are respectively $U_h'(x) = -h'(S_X(x))$ and $U_h''(x) = -F_X'(x)h''(S_X(x))$. Under Assumption A, the local utility U_h defined over losses is decreasing and concave, or equivalently, the function U_h^*, defined over returns by $U_h^*(x) = U_h(-x)$ for all $x \in \mathbb{R}$, is increasing and concave.

3.3.2 Local risk premium

Consider a decision maker with smooth preferences Ψ_h over the set $\mathcal{D}$. We define the local risk premium of the risk X, by the amount $\pi = \pi(h, F_X)$, that the decision maker is ready to pay in order to avoid the random loss X, around the distribution of F_X of X. Around the risk F_X, a smooth-preference decision maker behaves as an expected utility maximizer, using the local utility function at F_X, π can be defined as :

$$U_h(EX + \pi; F_X) = E_X\left(U_h(X; F_X)\right), \tag{3.5}$$

where $U_h(\cdot; F_X)$ is the local utility function relative to Ψ_h at F_X. The resulting risk premium $\pi(h, F_X)$ is a function of the distribution of the risk X and of the distortion h.
If the local utility function is invertible and its inverse H is known, then the expression for the risk premium $\pi_h = \pi(h, F_X)$ is :

$$\pi_h = H\left(E_X\left(U_h(X; F_X)\right)\right) - EX. \tag{3.6}$$

3.3.3 Local risk aversion

An approximation of the risk premium similar to that of Pratt (1964), can be derived and is given by (3.7). In the RDU case, the definition of the risk aversion function r relies on two approximations : at the distribution level and at the outcome level. Under suitable regularity conditions, expanding the local utility $U_h(\cdot)$ in the neighborhood of x gives :

$$U_h(EX + \pi) = U_h(EX) + \pi U_h'(EX) + o\left(\pi^2\right)$$

and

$$EU_h(X) = E\left[U_h(EX) + (X - EX)\,U_h(EX) + \frac{1}{2}\,(X - EX)^2\,U_h(EX) + O\left((X - EX)^3\right)\right].$$

This implies that :

$$\pi(h, F_X) = \frac{1}{2}\sigma_X r(h, F_X) + o(\sigma_X^2), \tag{3.7}$$

where the risk aversion function r is $r\left(h, F_X\right) = -\frac{U_h''(EX)}{U_h'(EX)}$. The function r is completely characterized by the distribution F_X of X and by the function h.

In the framework of RDU, an explicit expression for r can be derived if h is given. Using $U_h'(x) = -h'(S_X(x))$ and $U_h''(x) = F_X'(x)h''(S_X(x))$, gives :

$$r(F_X, h) = -\frac{F_X'(EX)h''(S_X(EX))}{h'(S_X(EX))}, \tag{3.8}$$

which is positive under assumption A. The risk aversion function can be used to compare any two arbitrary elements (h_1, X_1) and (h_2, X_2). Note that in the expected utility case, the loss distribution is not distorted since $h(x) = x$. This also implies, using (3.8) and $h'' \equiv 0$, that the associated local risk aversion is 0. Therefore, $r(h, F_X)$ measures the level of risk aversion only due to the pessimism of the decision maker, and can be used as a measure of the degree of pessimism.

Given a degree of pessimism, Equation (3.7) reveals that the risk premium can be approximated as a linear function of volatility σ_X. As a result, higher level of volatility will correspond to higher level of the premium due to pessimism, since $r \geq 0$.

3.3.4 Local utility function and risk premium for the CVaR

In this section, we determine the local utility function and risk premium in the conditional value-at-risk (CVaR) case. In addition, we discuss how these measures can derived in more general cases (power-law or exponential distortion).

Conditional Value-at-risk (CVaR)

Suppose the decision maker considers the $(1-p)^{th}$ quantile, $VaR_p(X)$, as a threshold, and is better-off for lower average losses above that threshold, then his preferences can be represented by the functional $\Psi_p(X) = -CVaR_p(X)$ [3]. We will refer to such preferences as CVaR preferences. The parameter p represents an acceptable loss probability level set by the decision maker, and will be referred to as the level of risk (Gourieroux and Liu (2006b)). For

3. An alternative interpretation is that such decision maker prefers higher returns when the loss $VaR_p(X)$ is exceeded since $\Psi_p(X)$ can also be written $\Psi_p(X) = E(Y|Y \leq Q_Y(p))$, where Y is the profit and loss variable.

example, for an institution p will be the percentage of loss that can be afforded on a particular position. High tolerance to risk, as expressed by large p, correspond to higher level of utility $\Psi_p(X)$ for a given risk X. Proposition 2.1 shows that, for the local utility function, such result only hold in the case of extreme losses.
In Proposition 4, we also determine the local utility function and risk premium for CVaR preferences, and analyze how they are affected by the risk level p. The derivations are based on the fact that the CVaR is a DRM with $h(\alpha, p) = \frac{\alpha}{p} \wedge 1$, $h'(\alpha, p) = \frac{1}{p} 1_{[0,p]}(\alpha)$.

Proposition 4. *Consider a decision maker with CVaR preferences at risk level p. We have the following results concerning its local utility function $U_p(\cdot; F_X)$ and risk premium $\pi(p) = \pi(p, X)$.*

1. *$U_p(x; F_X) = \frac{VaR_p(X) - x}{p} 1_{x > VaR_p(X)}$.*
2. *For two parameters p_1 and p_2 such that $p_1 < p_2$, we have $U_{p_1}(x) \geq U_{p_2}(x)$ if and only if $x \leq c(p_1, p_2)$, with $c(p_1, p_2) = \frac{p_2 VaR_{p_1}(X) - p_1 VaR_{p_2}(X)}{p_2 - p_1}$.*
3. *$\pi(p) = [pCVaR_p(X) + (1-p) VaR_p(X)] - EX$.*
4. *For a given risk X, π is decreasing with p.*

The first part of Proposition 4 reveals that, the decision maker with CVaR preferences has a null local utility when the threshold $VaR_p(X)$ is not exceeded. However, when $VaR_p(X)$ is exceeded, this local utility becomes negative, and is proportional to the extent to which $VaR_p(X)$ is exceeded. The second result of Proposition 2 implies that decision makers with higher risk level are better-off, relative to decision makers with lower risk level, only when extreme losses occur. Indeed, if we consider two decision makers respectively characterized by the risk levels p_1 and p_2 with $p_1 < p_2$. The level p_1 corresponds to higher local utility, as long as the loss magnitude does not exceed a threshold given by $c(p_1, p_2)$. The third part of the Proposition states that the risk premium is the weighted average of the CVaR and the VaR, respectively with weights p and $1 - p$, in excess of the expected loss. Since the CVaR also take into account the magnitude of the loss when it occurs, the major implication of Proposition 2.3 is that decision makers with higher tolerance to risk, care more about the magnitude than the probability of the loss when it occurs. The last part of Proposition 4 supports the intuition according to which, higher level of risk, correspond to a lower level of risk premium. These findings are illustrated in Figure 3.1.

Contrarily to the CVaR, the power-law and the exponential distortion functions described in Table (3.1), do not lead to closed-form analytical expressions for the local utility and the risk premium. However, for a given risk F_X, their value can be computed using numerical procedures or estimated from a given random sample, as discussed in Section 3.4.

FIGURE 3.1 – Local utility function and risk premium for CVaR preferences

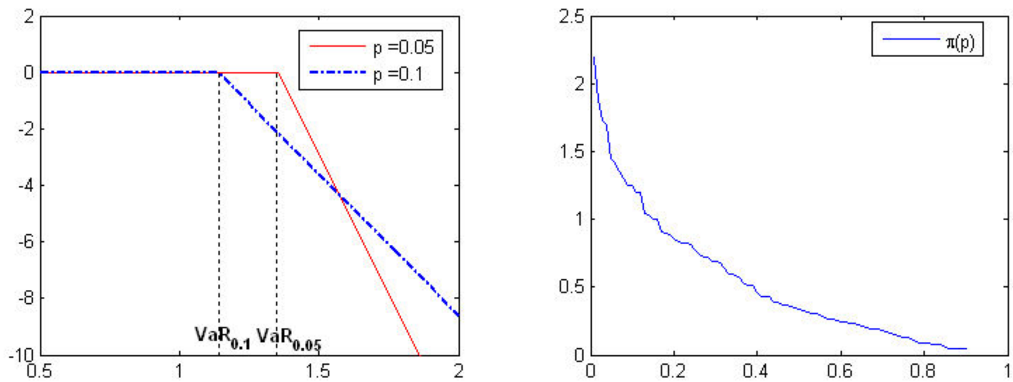


Figure 3.1 displays local utility functions and the local risk premium, computed using the results of Proposition 4 and returns distributed according to $N(\mu = 0.046; \sigma = 0.953)$. On the left panel, the decision maker with the lower risk level (p=0.05) has a higher VaR, and cares less about the magnitude of the incurred loss than the decision maker with risk level $p = 0.1$. The right panel describes the intuition according to which, for a given risk X, the higher the risk level p, the smaller the risk premium. The limiting case being that a 100% tolerance to risk, corresponds to a null risk premium.

3.4 Estimating the local utility function and the risk premium

The local risk premium generalizes the notion of risk premium, with the additional feature that the degree of pessimism is taken into account as a function of the loss distribution. In this section, we present a procedure to estimate the local risk premium using available data. We adopt the i.i.d setting for expository purpose and also because it corresponds to the most frequently used approach in practice (e.g. Gourieroux and Liu (2006), Rockafeller and Uryasev (2000)). We consider the distortion h to be given and fixed.

Suppose that we observe i.i.d one-dimensional realizations $x_1, \cdots, x_T$ of a loss variable X with distribution F_X. Our objective is to construct estimators for the local utility function and the local risk premium associated with a rank-dependent utility Ψ_h, for h known. An emphasis is put on the CVaR.

3.4.1 Estimating the local utility function

Estimating $U_h(x)$, for a given outcome x, consists in estimating the integral of the unknown function $\phi = -h' \circ S_X$ between $-\infty$ and x. The functions S_X and ϕ can respectively be estimated by $\hat{S}_X$ and $\hat{\phi} = -h'(\hat{S}_X)$, where $\hat{S}_X(x) = 1 - \hat{F}_X(x)$ is the empirical survival function. The estimated function $\hat{\phi}$ is then defined by interval, and take the values $-h'(\hat{S}_X(x_{(t)}))$ on the interval $[x_{(t)}, x_{(t+1)}[$ for $t = 1, \cdots, T$. Define $\Delta x_{(t)} = x_{(t+1)} - x_{(t)}$, we propose to estimate $U_h(x)$ by the integral of $\hat{\phi}$:

$$\hat{U}_h(x) = -\sum_{t=1}^{t_l(x)} h'\left(\hat{S}_X(x_{(t)})\right)\Delta x_{(t)}.$$

When $T \to \infty$, $\Delta x_{(t)} \to 0$ and, using $t_l(x)$ defined in Section 3.2.1, $x_{t_l(x)} \to x$ *a.s*, so that $\hat{U}_h(x)$ is expected to be consistent as we formally show below.

Assumption B

1. $x_1, \cdots, x_T$ is an i.i.d random sample from X.
2. For all $x \in R$, $\int_{-\infty}^{x} h''\left(S_X(t)\right) dt < \infty$.

Lemma 1. *Let ψ be a nonrandom function defined on R, for a fixed $x \in R$, and under Assumption B.1, we have :*

$$\sum_{t=1}^{t_l(x)} \psi(x_{(t)})\Delta x_{(t)} \to \int_{-\infty}^{x} \psi(u)du \quad a.s \quad when \;\; T \to \infty.$$

The proof of lemma 1 is a consequence of assumption B.1, and is provided in appendix A

Proposition 5. *Under assumptions A and B, for a fixed x, $\hat{U}_h(x) \to U_h(x)$ a.s, when $T \to +\infty$.*

Proposition 6. *Under assumptions A and B, for all x :*

$$\sqrt{T}\ \left[\hat{U}_h(x) - U_h(x)\right] \Rightarrow -\int_{-\infty}^{x} h'\left(S_X\left(t\right)\right) B(F_X(t))dt,$$

where $B(\cdot)$ is a standard Brownian bridge. The limiting distribution of $\hat{U}_h(x)$ is a Gaussian process with pointwise variance given by

$$V(x) = \int_{-\infty}^{x}\int_{-\infty}^{x} h''\left(S_X(u_1)\right) h''\left(S_X(u_2)\right)\left[F_X\left(u_1 \wedge u_2\right) - F_X\left(u_1\right) F_X\left(u_2\right)\right] du_1 du_2.$$

$V(x)$ can be consistently estimated by :

$$\sum_{t_1=1}^{t_l(x)}\sum_{t_2=1}^{t_l(x)} h''\left(\hat{S}_X(x_{(t_1)})\right) h''\left(\hat{S}_X(x_{(t_2)})\right)\left[\hat{F}_X\left(x_{(t_1)} \wedge x_{(t_2)}\right) - \hat{F}_X\left(x_{(t_1)}\right)\hat{F}_X\left(x_{(t_2)}\right)\right]\Delta x_{(t_1)}\ \Delta x_{(t_2)}.$$

The proofs of Proposition 5 and 6 are provided in appendix A. The consistency result of Proposition 5 is illustrated by Figure 3.2. Using Proposition 6, bootstrap procedures can be designed to approximate the limiting distribution of the local utility function.

3.4.2 Estimating the risk premium for a given distotion h

Estimating the risk premium for a given function h requires the computation of the inverse H of U_h. We first provide a procedure to estimate such a function. Under Assumption A, U_h is a concave and decreasing function, then admits an inverse that can be determined from U_h using the procedure described below and illustrated in Figure 3.3.

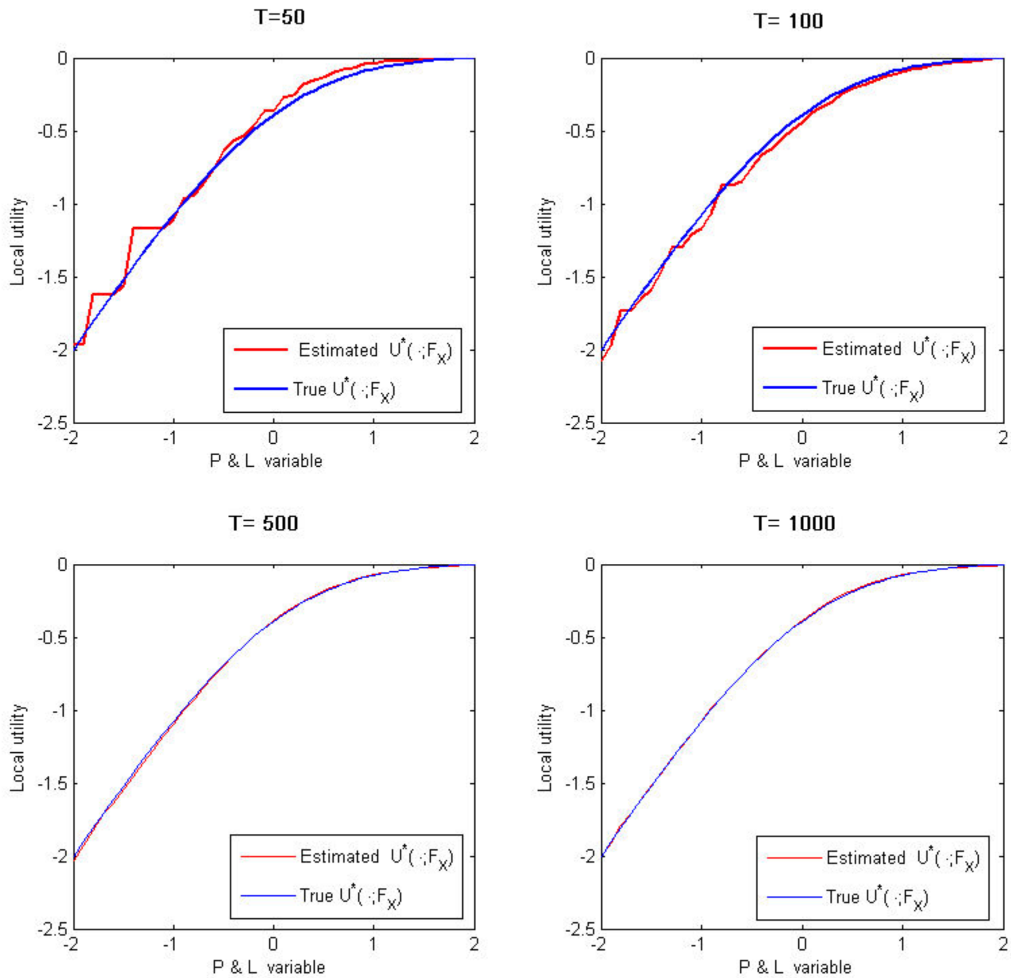
After observing the losses x_t, for a given real x in the interval $I_t = \left[x_{(t)}, x_{(t+1)}\right)$, we have $t_l(x) = t$ and $\hat{U}_h(x) = -\sum_{j=1}^{t} h'\left(\hat{S}_X(x_{(j)})\right)\Delta x_{(j)} = \lambda_t$.

This implies that, conditional on the observed losses, the estimated function $\hat{U}_h$ only takes the values in the set $\{\lambda_t\}_{t=1}^{T}$. Besides, these values are such that $\lambda_T \leq \cdots \leq \lambda_1 \leq \lambda_0$, with $\lambda_0 = 0$.

Given these remarks, we propose to estimate the reciprocal H of the function U_h, by inverting $\hat{U}_h$. Let $y \in \mathbb{R}$, there exists $i^* \in \{0, \cdots, T\}$ such that $y \in]\lambda_{i^*+1}, \lambda_{i^*}]$. As illustrated in Figure 3.3 , i^* can be expressed as

$$i^*(y) = \inf_{0 \leq i \leq T}\{i : y \leq \lambda_i\}. \tag{3.9}$$

Figure 3.2 – Estimated local utility as the sample T increases



For each value of $T \in \{50, 100, 500, 1000\}$, Figure 3.2 plots the theoretical local utility function numerically computed, and the estimated local utility function. The dataset used is calibrated to daily returns on the U.S. consumption good industry from French library ($N(\mu = 0.046; \sigma = 0.953)$). The theoretical utility function is evaluated on a grid of 1000 points on the interval $[-2; 2]$, while the local utility is only evaluated at each points of the generated samples. The distortion considered is $h(u) = 1 - (1 - u)^2$. These plots illustrate the consistency of the estimators proposed in Section 4 : the discrepancy between the theoretical and the estimated local utility curves reduces as the sample size increases.

Consequently, $y = \lambda_{i^*+1} + \alpha(\lambda_{i^*} - \lambda_{i^*+1}) = \alpha\lambda_{i^*} + (1-\alpha)\lambda_{i^*+1}$, with $\alpha = \frac{y-\lambda_{i^*+1}}{\lambda_{i^*}-\lambda_{i^*+1}}$. By linear interpolation, we have the following approximation of H :

$$\hat{H}(y) = (1-\alpha)x_{(i^*)} + \alpha x_{(i^*+1)}, \tag{3.10}$$

where i^* and λ_i are defined by (3.9).

As the final step before the estimation of π_h, we propose to estimate the expected local utility $E_X\left(U_h(X;F_X)\right)$ by its sample counterpart $\hat{E}_X\left(\hat{U}_h(X;F_X)\right) = T^{-1}\sum_{t=1}^{T}\hat{U}_h(x_t)$.

By replacing the estimates of the expected local utility and the inverse of the local utility in Equation (3.6), the estimator of π_h that we propose is :

$$\hat{\pi}_h = \hat{H}\left(T^{-1}\sum_{t=1}^{T}\hat{U}_h(x_t)\right) - T^{-1}\sum_{t=1}^{T}x_t. \tag{3.11}$$

FIGURE 3.3 – Inverting a local utility function

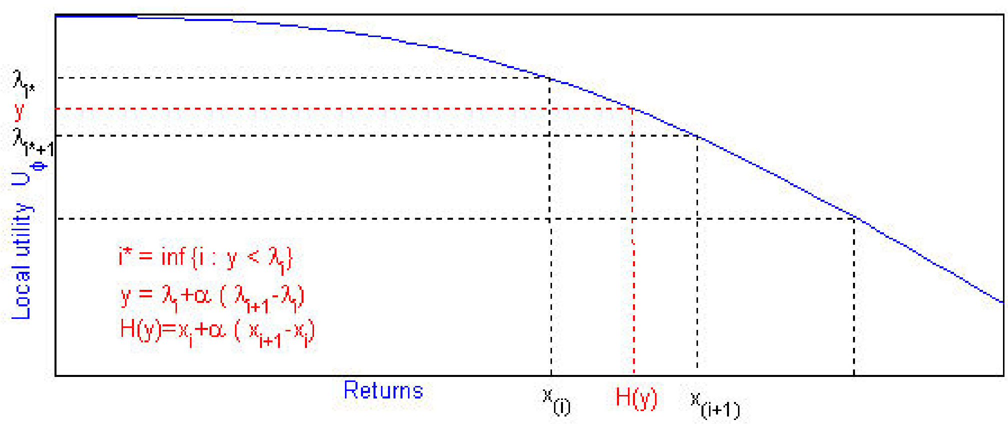


Figure 3.3 illustrates the numerical procedure used to compute the inverse H of a local utility function U_h. Losses $x_{(i)}$ on the horizontal axis represent observations arranged in increasing order, while on the vertical axis, the λ_i represent the associated level of utility.

In the next section, we show the asymptotic normality of the risk premium in the CVaR case.

3.4.3 Estimating the CVaR premium

In the Conditional value-at-risk case, the closed-form analytical expression of Proposition 4, makes it easy to construct estimators for the local utility and the risk premium. For a given p, $\pi(p)$ can be estimated by its sample counterpart $\hat{\pi}(p)$. From Proposition 4, $\hat{\pi}(p)$ is obtained by replacing the empirical estimates of $CVaR_p(X)$, $VaR_p(X)$ and $E(X)$ in $\pi(p)$:

$$\hat{\pi}(p) = \left[p\widehat{CVaR}_p(X) + (1-p)\widehat{VaR}_p(X)\right] - \bar{X}_T. \tag{3.12}$$

$\bar{X}_T$ is the sample mean, while $\widehat{VaR}_p(X)$ [4] and $\widehat{CVaR}_p(X)$, as distortion risk measures, can be estimated by L-statistics (Gourieroux and Liu (2006)) :

$$\hat{\Pi}_T(p) = \sum_{t=1}^{T} x_{(t)} \left[h\left(1 - \frac{t-1}{T}\right) - h\left(1 - \frac{t}{T}\right)\right]. \tag{3.13}$$

We have

$$\widehat{VaR}_p(X) = \sum_{t=1}^{T} x_{(t)} \left[1_{\left(\frac{t-1}{T} \leq 1-p\right)} - 1_{\left(\frac{t}{T} \leq 1-p\right)}\right] = \begin{cases} x_{(1-p)T} \; if \; (1-p)T \; is \; an \; integer \\ x_{[(1-p)T]+1}, \end{cases}$$

and $\widehat{CVaR}_p(X) = \frac{1}{Tp} \sum_{[(1-p)T]}^{T} x_{(t)}$, obtained by replacing the function $h(\alpha, p) = \frac{\alpha}{p} \wedge 1$ in (3.13) ; $[b]$ denotes the integer part of b. The convergence of $\hat{\pi}(p)$ in (3.12) results are summarized in Proposition 7.

Proposition 7. *Suppose assumption B.1 holds and denote by X a random variable with finite variance σ_X^2. Then for all real x, and $p \in [0;1]$, we have the following convergences :*

1. $\hat{\pi}(p) \to \pi(p)$ *a.s.*
2. $\sqrt{T}(\hat{\pi}(p) - \pi(p)) \to N(0, \sigma_\pi^2)$, *with* $\sigma_\pi^2 = p^2\omega_1^2 + (1-p)^2\omega_2^2 + \sigma_X^2$ *and*

$$\begin{aligned} \omega_1^2 &= \frac{V(X \,|X \geq VaR_p(X)) + (1-p)(CVaR_p(X) - VaR_p(X))}{p} \\ \omega_2^2 &= \frac{p(1-p)}{f(Q_X(1-p))}. \end{aligned}$$

4. VaR is a particular DRM with a distortion $h(\alpha, p) = 1_{\alpha \geq p}$ that is not concave. The VaR is not coherent because it does not satisfy the subadditivity property.

The function f is the probability density function of X, and ω_1 and ω_2 are the asymptotic variances of $\widehat{CVaR}_p(X)$ and $\widehat{VaR}_p(X)$ respectively.

The proof is provided in appendix C. Its main ingredient is that $\bar{X}_T$, $\widehat{VaR}_p(X)$ and $\widehat{CVaR}_p(X)$ are asymptotically uncorrelated.

In addition to providing the asymptotic distribution of the CVaR premium, Proposition 7 reveals that the volatility of the risk premium is monotonic relative to the volatility of the loss and profit variable σ_X. This fact is illustrated by our simulations (see Figure 3.4).

3.5 Monte Carlo study

In this section, we perform Monte Carlo simulations to assess the small sample properties of the proposed estimators for the local risk premium, considering four distortion functions h and three sample sizes. We use a loss distribution X which we assume to be normally distributed and that we calibrate to daily returns on the Consumption good industry (Cnsmr), over the period 1963-2010 ($\mu = -0.046; \sigma = 0.953$). The daily returns are extracted from the 5 Industry Portfolios (hereafter 5IP) compiled on Kenneth French website [5]. Using the calibrated distribution, we generate 1000 samples for different sample size $T \in \{60, 100, 200\}$. For each generated sample, the estimated local risk premium is computed using equations (3.11) and (3.12) and the four distortions from Table 3.1 : the power distortion for $\theta = 2$ and $\theta = 3$, the exponential distortion with $p = 1$ and the CVaR distortion with $p = 0.05$. We then obtained for each h and each T, a series of risk premia. The summary statistics over 1000 replications of the estimated risk premia $\hat{\pi}$ are reported in Table 3.2. The theoretical risk premia π^* are computed using the true distribution and the respective distortion h. The reported bias, based on π^* are found to be of the order $10^{-2}\pi^*$, for all h. As expected, these bias decrease with the sample size T. The series of estimated risk premia also exhibit a ratio of the standard deviation to the mean of order 10^{-2}, which corresponds to a small dispersion of the estimates around their respective true value.

The results of the simulations performed suggest that the proposed estimator are expected to be consistent and to have little estimation error, even in relatively small sample.

5. http : //mba.tuck.dartmouth.edu/pages/faculty/ken.french/.

Table 3.2 – Statistics on the estimated local risk premia throughout 1000 replications, with normal returns, for four distortions

Risk premium					
Sample size	Statistics	$Pow(2)$	$Pow(3)$	$Exp(1)$	$CVaR_{0.05}$
	π^*	0.30043	0.46072	0.16298	1.58758
	Bias	0.00547	0.00744	0.02604	0.06564
	Std	0.04586	0.05689	0.03867	0.24521
$T = 60$	q1	0.27491	0.42873	0.16238	1.48647
	median	0.30120	0.46427	0.18110	1.64010
	q3	0.32951	0.50371	0.20586	1.80270
	Bias	-0.00341	-0.00179	0.01237	0.03370
	Std	0.03142	0.04176	0.02474	0.18081
$T = 100$	q1	0.27544	0.43063	0.15857	1.49569
	median	0.29405	0.45664	0.17061	1.61413
	q3	0.31560	0.48483	0.18726	1.72706
	Bias	-0.00109	-0.00668	0.00151	0.00809
	Std	0.01946	0.02744	0.01287	0.12472
$T = 200$	q1	0.27641	0.43504	0.15611	1.51145
	median	0.28935	0.45330	0.16328	1.58993
	q3	0.30201	0.47222	0.17166	1.67808

Table 2 describes the small sample properties of the estimators proposed in Section 4. For all the values of h, we observe a small bias that decreases, when the sample T increases, and a low standard deviation. The distortion functions h are described in Table (3.1), and summarize how the decision maker distort probability distributions.

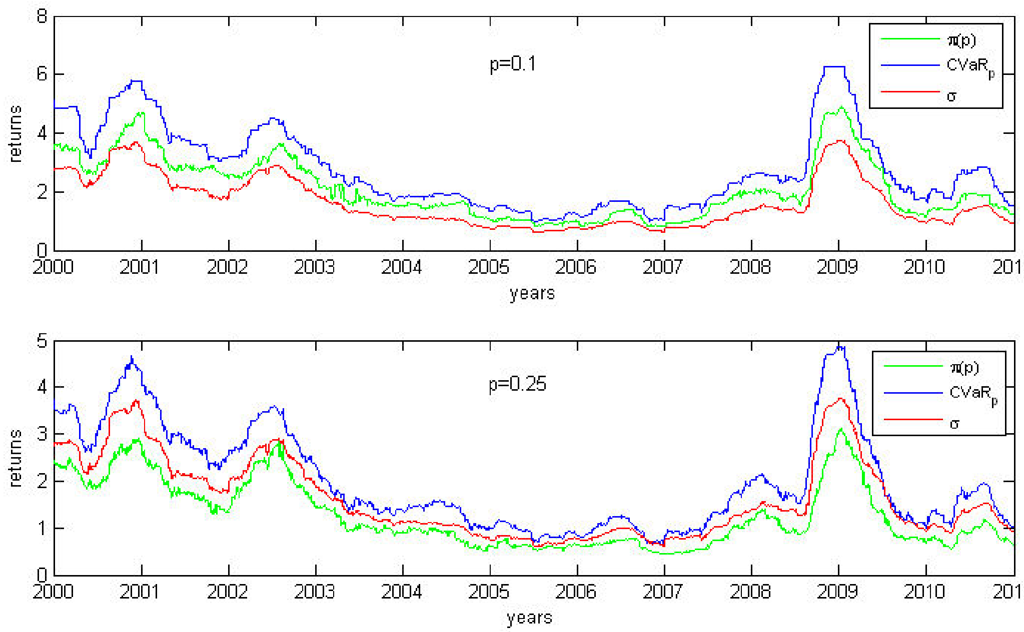
3.6 Empirical illustration

In this section we empirically illustrate the use of the proposed estimator for the local risk premium. As in the previous simulations, the loss distribution is derived from 5IP daily returns. Note however that the proposed estimators could equally be applied to a market index or to any portfolio. Decision makers are supposed to have CVaR preferences, that is, conditional to the fact the VaR is exceeded, they prefer higher returns. To reflect the structural change in the industry portfolio returns, we adopt a rolling window approach with a window size $M = 100$. At each date t, formula (3.11) is used to compute the risk premium $\hat{\pi}_t$ based on the M previous daily observations. The window is then shifted one day ahead. This step is repeated recursively until the end of the sample, and provides the time series $\{\hat{\pi}_t\}_{t=M+1}^{T}$ for each of the five industries considered.

Figure 3.5 exhibits two major periods of high level of local risk premia, 2000-2003 and 2008-2011. The latter is likely to be a reflection of the high level of volatility that U.S stock market has experienced as a consequence of the subprime crisis. The former period coincides with the aftermath of the IT bubble burst in 2000, that resulted in the decline of the U.S stock market. Figure 3.5 reveals the highest volatility in hightech, which was precisely the industry the most affected. Indeed, compared to other portfolios, investor were willing to pay more to avoid holding the hightech industry portfolio. Between the two highly volatile periods, the intermediary period (2004-2007) is characterized by a very low volatility and a local risk premium which fluctuates on average between 1 and 2 for all industries. The evidences in Figure 3.5 are supported by statistics in Table 3.3, computed for the three periods mentioned above, and also provide us with a measure of the relative severity of the crisis during these periods. This relative severity varies according to the industry, even though globally these periods are characterized by high values of risk premia. For example the premium paid during the period 2000-2003 for Cnsmr is approximately twice the premium paid in the stable period (2004-2007), while the premium paid during 2008-2011 is four times the premium paid in the stable period. For the Health industry, the maximum premium in the periods 2000-2003 and 2008-2011 is approximately the same, and is twice the premium during the 2004-2007 period.

Figure 3.4 reveals that the risk premium, the average conditional loss and the volatility as measured by standard deviation tend to move together. This can be explained by the fact that for a given p, a high level of volatility corresponds to a high level of CVaR and VaR. An additional explanation can be provided by the approximation in Equation (3.7) : high degree of pessimism as measured by the coefficient $r(h, F_X)$, and high volatility σ_X, will correspond to high risk premium $\pi(h, F_X)$. The proposed local risk premium measure can therefore be used as a measure of risk, with the additional feature that it also reflects the level of pessimism.

FIGURE 3.4 – Comparing CVaR premium with CVaR and volatility
Comparison of series of out-of-sample CVaR premia, CVaR and standard deviation computed for daily returns on the Hitech industry, using CVaR preferences with a risk level of $p \in \{0.1, 0.25\}$ and a rolling window of $M = 100$



We also considered the exponential and the power-law hazard distortions, and two different values for the rolling window, $M = 60$ and $M = 200$. The results obtained are essentially the same and are available upon request.

FIGURE 3.5 – Out-of-sample CVaR risk premia for 5IP

The Figures shows series of out-of-sample risk premia computed for daily returns from 5IP, using CVaR preferences with a risk level of $p = 0.1$ and a rolling window of $M = 100$

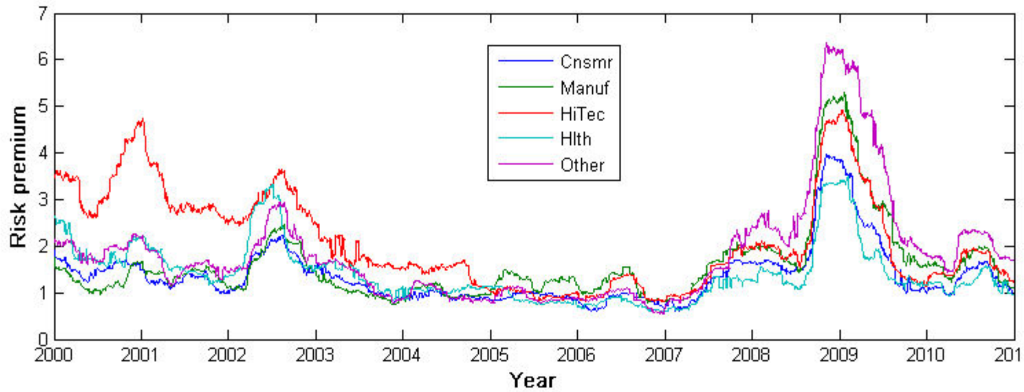


TABLE 3.3 – Statistics on Out-of-sample daily risk premia for 5IP

Table 3 displays statistics on CVaR risk premia during stressed market conditions (2000-2003 and 2008-2011), and during normal market conditions (2004-2007). Notice that the premium paid as a consequence of pessimism depends on the industry considered, and is higher during stressed market conditions. For example in the aftermath of the dot-com crisis, the HiTec industry had the highest premium. In addition, for all industries the premium is relatively higher during stressed market conditions.

Periods	Industry	Mean	Std	min	q1	median	q3	max
2000-2003	Cnsmr	2.5319	0.5442	1.5162	1.9328	2.7082	2.9974	3.4391
	Manuf	2.4281	0.5587	1.4217	2.0540	2.2243	2.9055	3.5873
	HiTec	4.7710	1.0326	3.1489	3.7573	4.7490	5.9120	6.4908
	Hlth	3.5802	1.0112	1.6109	2.6239	4.0809	4.3303	5.1993
	Other	3.4392	0.4914	1.6529	3.2022	3.4872	3.7607	4.3133
2004-2007	Cnsmr	1.4941	0.1621	0.9580	1.3860	1.5405	1.6119	1.7232
	Manuf	2.0352	0.3829	1.1416	1.7368	2.0155	2.3424	2.6826
	HiTec	1.8618	0.3593	1.2084	1.5543	1.8762	2.1021	2.9147
	Hlth	1.6187	0.2725	0.9465	1.4471	1.7268	1.7900	2.1093
	Other	1.5329	0.2391	1.1143	1.2885	1.5674	1.7036	1.9794
2008-2011	Cnsmr	3.2947	1.4640	1.5883	2.3979	2.6445	3.5417	6.8231
	Manuf	4.4095	2.0265	2.0355	3.1033	3.6885	5.3268	8.9533
	HiTec	3.9518	1.7627	1.8345	2.8841	3.0651	4.4400	8.1658
	Hlth	3.1312	1.4933	1.6869	2.0263	2.4379	4.0318	6.2901
	Other	5.2471	2.5359	2.3878	3.5664	3.8850	7.1112	11.0069

3.7 Conclusion and future work

This chapter derives the local risk premium and proposes a measure of the degree of pessimism for rank-dependent preferences. These measures are derived as a function of the loss distribution and of a distortion function characterizing the decision maker's pessimism. A statistical procedure is then provided for the nonparametric analysis of the local utility function and the local risk premium. A particular attention is devoted to the Conditional Value-at-Risk as the most commonly used rank-dependent utility. To evaluate the small sample properties of the proposed estimates, a Monte Carlo study is conducted. The proposed measures and their estimates have many possible applications, once the distortion function is known. For example, the risk premium can be used to measure the degree or riskiness of a market, an industry or an instrument, and is consistent with results obtained using standard risk measures like the variance and the CVaR. To illustrate these possibilities, we conducted an empirical study using daily returns on 5 U.S. industry portfolios. We obtained evidences of high volatility in the U.S. stock market during the periods 2000-2003 and 2008-2011. The developed statistical tools also allowed us to measure the relative level of pessimism during these two crisis periods. Throughout this chapter, we assumed that the decision maker's level of pessimism, described by a function h or a parameter p, is known. These values could be estimated in a first step, to obtain a data-dependent distortion function. Our future investigations are oriented toward deriving h or the p, based on observed risk premia. A straightforward generalization of this chapter is to consider risk averse and pessimistic decision makers. In such a framework, we plan to investigate how the total premium can be decomposed into premium due to risk aversion, and premium due to pessimism.

Bibliographie

[1] Adam, A., Houkari, M., Laurent, J. P., 2008. "Spectral risk measures and portfolio selection," *Journal of Banking & Finance*, 32, 1870-1882.

[2] Allais, M., 1953. "Le comportement de l'homme rationnel devant le risque, critique des postulats et axiomes de l'école americaine," *Econometrica*, 21, 503-546.

[3] Artzner, P., F. Delbaen, J.-M Eber and D. Heath, 1999. "Coherent measures of risk," *Mathematical Finance*, 9, 203-228.

[4] Bassett, G. W., Koenker, R., Kordas, G., 2004. "Pessimistic portfolio allocation and Choquet expected utility," *Journal of Financial Econometrics*, 2, 477-492.

[5] Gourieroux, C., J. P. Laurent, and O. Scaillet, 2000. "Sensitivity analysis of value-at-risk," *Journal of Empirical Finance*, 7, 225-245.

[6] Gourieroux, C. and W. Liu, 2006a. "Efficient portfolio analysis using distortion risk measures," *CREST Working Paper*, 2006-17.

[7] Gourieroux, C. and W. Liu, 2006b. "Sensitivity analysis of distortion risk measures," *CREST working paper*, 2006-03.

[8] Kahneman, D. and A. Tversky, 1979. "Prospect Theory : An Analysis of Decision under Risk," *Econometrica*, 47, 263-292.

[9] Kozhan, R., Schmid, W., 2009. "Asset allocation with distorted beliefs and transaction costs," *European Journal of Operational Research*, 194, 236-249.

[10] Krokhmal, P., Palmquist, J., Uryasev, S., 2002. "Portfolio optimization with conditional value-at-risk objective and constraints," *Journal of Risk*, 4, 43-68.

[11] Kusuoka, S., 2001. "On law invariant coherent risk measures," *Advances in Mathematical Economics*, 3, 83-95.

[12] Markowitz, H. M., 1952. "Portfolio selection," *The Journal of Finance*, 7, 77-91.

[13] Machina, M., 1982. "Expected utility analysis without the independence axiom," *Econometrica*, 50, 277-323.

[14] Pratt, J., 1964. "Risk aversion in the small and in the large," *Econometrica*, 32, 122-136.

[15] Quiggin, J., 1982. "A theory of anticipated utility," *Journal of Economic Behavior and Organization*, 3, 323-343.

[16] Rothschild, M. and J. E. Stiglitz, 1970. "Increasing risk : I, a definition," *Journal of Economic Theory*, 2, 225-243.

[17] Rothschild, M. and J. E. Stiglitz, 1971. "Increasing risk : II, its economic consequences," *Journal of Economic Theory*, 3, 66-84.

[18] Rockafeller, R. T. and S. Uryasev, 2000. "Optimization of conditional value-at-risk objective and constraints," *Journal of Risk*, 2, 21-41.

[19] Safra, Z. and U. Segal, 2002. "On the ecomomic meaning of Machina's Fréchet differentiability assumption," *Journal of Economic Theory*, 104, 450-461.

[20] Scaillet, O., 2004. "Nonparametric estimation and sensitivity analysis of expected shortfall," *Mathematical Finance*, 14, 115-129.

[21] Schmeidler, D., 1986. "Subjective probability and expected utility without additivity," *Econometrica*, 57, 571-587.

[22] Segal, U., 1987. "Some remarks on Quiggin's anticipated utility," *Journal of Economic Behavior and organization*, 8, 145-154.

[23] Wang, S., 1996. "Premium calculation by transforming the layer premium density," *ASTIN Bulletin*, 26, 71-92.

[24] Wang, S., 2000. "A class of distortion operators for pricing financial and insurance risks," *Journal of Risk and Insurance*, 67, 15-36.

[25] Yaari, M. E., 1987. "The dual theory of choice under risk," *Econometrica*, 55, 95-115.

Conclusion Générale

Cette thèse propose des outils statistiques pour améliorer l'implémentation des règles de choix de portefeuille, et pour mesurer le risque lorsque le preneur de décision est pessimiste.

Dans le premier chapitre, nous abordons la question de l'erreur d'estimation dans le cadre de l'analyse moyenne-variance. Nous proposons de régulariser le problème de choix de portefeuille en utilisant des techniques de régularisation tirées de la littérature des problèmes inverses. Ces techniques de régularisation à savoir le ridge, la coupure spectrale, et le Landweber-Fridman contiennent un paramètre de régularisation, dont la valeur optimale est choisie pour minimiser la perte d'utilité espérée d'un investisseur moyenne-variance. Nous montrons que cela équivaut à sélectionner le paramètre de régularisation de manière à minimiser un critère de validation croisée généralisée, corrigée du biais introduit par la régularisation.

Pour évaluer les performances de nos règles régularisées, nous faisons des simulations en utilisant un modèle à trois facteurs calibré aux données du marché boursier américain, ainsi qu'une étude empirique utilisant 48 et 100 portefeuilles d'industrie américaine. Les règles sont essentiellement comparées en fonctions de leur perte d'utilité espérée et de leur ratios de Sharpe. La principale conclusion est que, dans les cas, où le problème inverse sous-jacent est mal posé, une régularisation de la matrice de covariance améliore considérablement les performances du problème moyenne-variance, fournit souvent de meilleurs résultats que les stratégies actuelles d'allocation d'actifs et donne de meilleurs performances que le portefeuille naïf surtout dans les cas mal posé.

La méthodologie proposée dans ce premier chapitre peut être utilisée pour construire toute règle d'investissement nécessitant une estimation de la matrice de covariance et étant donné un critère de performance. Les règles d'investissement que nous proposons, ont un aspect pratique, en ce sens qu'ils sont faciles à mettre en œuvre et constituent une alternative valable aux règles en vigueur dans les cas mal posés, tel que démontré par nos simulations.

Dans le second chapitre, nous examinons dans quelle mesure l'adoption de portefeuilles optimaux domestiques conduisent à une réduction de l'exposition au risque de devise, et comment la prime de change correspondante est affectée. Nous nous limitons aux portefeuilles de minimum variance (PMV), car cette stratégie correspond au risque le plus faible parmi les portefeuilles moyenne-variance, et a été documentée avoir une meilleure performance hors-échantillon. En outre, nous considérons une famille plus générale de PMV, obtenue en appliquant la méthode de régularisation de coupure spectrale, c'est à dire en faisant varier le pourcentage retenu des composantes principales de la matrice de covariance. Plus ce pourcentage est bas, plus le portefeuille obtenu est stable. Ainsi, cette étude a également permis d'étudier les liens entre la stabilité du portefeuille et de risque de change.

Le deuxième chapitre aboutit à trois résultats principaux. Tout d'abord, nous constatons que le risque de change reste un facteur primé, et variant dans le temps pour les portefeuilles domestiques optimaux. Toutefois, pour certain niveaux de stabilité, il est possible d'éliminer complètement la composante marchés émergents du risque de change. Deuxièmement, bien que la plupart des portefeuilles d'industries sont exposés au risque de change, que ce soit directement ou indirectement, nos résultats suggèrent également que les PMV ont le potentiel de réduire significativement l'exposition au risque de change de plus de la moitié de sa taille. Troisièmement, le risque de change pour les portefeuilles optimaux étudiés est économiquement important et la contribution de la prime de change à la prime totale, reste à peu près égale à la moyenne industrielle.

L'évaluation de la pertinence du risque de change pour les portefeuilles domestiques optimaux est faite dans le cadre d'un modèle conditionnelle d'évaluation international des actifs financiers. Ainsi, notre étude fournit des évidences supplémentaires sur la tarification du risque de change, aux États-Unis. Ce travail a également des implications importantes en termes de stratégies de couverture pour les investisseurs nationaux. Précisément, l'adoption des poids stables est susceptible de produire des expositions faibles et stables aux risques de change, et donc implique des coûts de couverture inférieurs.

Le dernier chapitre se distingue des deux premiers en ce sens qu'il ne traite pas du problème du choix de portefeuille, mais propose des mesures de risques associées à un type particulier de préférences. En effet, ce chapitre dérive la prime de risque locale, propose une mesure du degré de pessimisme des préférences rang-dépendant (RDU), et finalement propose des outils statistiques pour les estimer ces mesures. Les mesures de risques sont obtenues en utilisant la notion d'utilité locale de Machina (1980), et sont fonction de la distribution de perte et d'une fonction de distorsion qui caractérise le pessimisme du décideur. Une procédure statistique est établie pour l'analyse non-paramétrique de la fonction d'utilité locale et la prime de risque local. Une attention particulière est consacrée à la mesure CVaR (valeur - à - risque conditionnelle) puisqu'elle représente la fonction d'utilité rang-dépendant la plus couramment utilisée.

Les propriétés en petit échantillon des estimations proposées sont évaluées, par une étude de Monté Carlo. Les mesures proposées et leurs estimations ont de nombreuses applications possibles, dès que la fonction de distorsion est connue. Par ailleurs, ces mesures sont conformes aux résultats obtenus pour les mesures de risque standards, comme la variance et de la VaR. Nous illustrons l'utilisations de ces mesures en utilisant des rendements quotidiens sur 5 portefeuilles de l'industrie américaine. Nous obtenons des évidences de forte volatilité pour le marché boursier américain pendant les périodes 2000-2003 et 2008-2011. Les outils statistiques développés nous ont également permis de mesurer le niveau relatif de pessimisme au cours de ces deux périodes de crise. Pour finir, dans les dérivations de ce chapitre, nous avons supposé que le niveau de pessimisme du décideur, décrite par une fonction h ou un paramètre p, qui sont connus. Ces valeurs peuvent alternativement être estimées dans un premier temps, à partir de primes de risque observées.

Annexes

Annexe A

Chapter 1

A.1 Optimality proofs

An important element of the proofs consists in orthogonalizing the regressors and showing that this will not affect our estimate. Let P be the matrix having the orthonormal eigenvectors of RR'/T as columns. Note that $P' = P^{-1}$ and hence $P'(RR'/T)P$ is a diagonal matrix with λ_j^2 on the diagonal. The model

$$1 = R\beta + u \tag{A.1}$$

can be rewritten as

$$P1 = PR\beta + Pu$$

or equivalently

$$y = X\beta + e \tag{A.2}$$

with $y = P1$, $X = PR$, and $e = Pu$. Note that $E\left(ee'\right) = E\left(uPP'u'\right) = E\left(uu'\right) = \omega^2 I$. Let x_t be the tth row of X. If one applies a regularization on $X'X$ as we did on $R'R$ in Section 3, then the resulting estimator is equal to $\hat{\beta}_\tau$. And we get $X\hat{\beta}_\tau = \tilde{M}_T\left(\tau\right)y$ where now $\tilde{M}_T$ is a diagonal matrix with $q\left(\tau, \lambda_j^2\right)$ on the diagonal. Moreover, as $\left\| X\left(\hat{\beta}_\tau - \beta\right)\right\|^2 = \left(\hat{\beta}_\tau - \beta\right)' R'P'PR\left(\hat{\beta}_\tau - \beta\right) = \left\| R\left(\hat{\beta}_\tau - \beta\right)\right\|^2$, it is indifferent to look at the problem $\left\| X\left(\hat{\beta}_\tau - \beta\right)\right\|$ in Model (A.2) or at the problem $\left\| R\left(\hat{\beta}_\tau - \beta\right)\right\|$ in Model (A.1). Most of the proof will rely on Model (A.2).

Another important element in the proof is the following. As x_t (or r_t) is random, so are the eigenvalues $\hat{\lambda}_j^2$ and the matrix $\tilde{M}_T\left(\tau\right)$. However the eigenvalues converge at a fast rate ($\sqrt{T}$) to a limit which is non random λ_j^2 (see Theorem 3 of Carrasco and Florens, 2000). It is therefore possible to replace $\hat{\lambda}_j^2$ by λ_j^2 and treat $\tilde{M}_T(\tau)$ as non random in the analysis. From now on, all results are derived for $\hat{\lambda}_j^2$ replaced by λ_j^2.

Preliminary result :

To prove Proposition 1, we need the following preliminary result.

Lemma 2. *Under Assumption A and if N goes to infinity, we have*

$$\frac{1}{T}E\left[\left\|R\left(\hat{\beta}_\tau-\beta\right)\right\|^2\right]=\frac{1}{T}E\left[\left\|R\left(\hat{\beta}_\tau-\beta_\tau\right)\right\|^2\right]+\frac{1}{T}E\left\|R\left(\beta_\tau-\beta\right)\right\|^2$$

with

$$\frac{1}{T}E\left[\left\|R\left(\hat{\beta}_\tau-\beta_\tau\right)\right\|^2\right]=O\left(\frac{1}{T\tau}\right)$$

and

$$\begin{aligned}\frac{1}{T}E\left\|R\left(\beta_\tau-\beta\right)\right\|^2 &= O\left(\tau^{\nu+1}\right) \text{ for SC, LF,}\\ &= O\left(\tau^{\min(\nu+1,2)}\right) \text{ for T.}\end{aligned}$$

In summary

$$\frac{1}{T}E\left\|R\left(\hat{\beta}_\tau-\beta\right)\right\|^2\sim\sigma^2\frac{1}{T\tau}+C\tau^{\nu+1}$$

which is minimized for $\tau=T^{-1/(\nu+2)}$ and is equivalent to $T^{-(\nu+1)/(\nu+2)}$. We can check that Li (1986)'s condition for optimality holds, namely that $inf_\tau E\left\|R\left(\hat{\beta}_\tau-\beta\right)\right\|^2\to\infty$. Note that if N is fixed, this condition is not fulfilled.

Proof of Lemma 2. As discussed earlier, it is indifferent to work on $\left\|X\left(\hat{\beta}_\tau-\beta\right)\right\|$ in Model (A.2) or on $\left\|R\left(\hat{\beta}_\tau-\beta\right)\right\|$ in Model (A.1). From now on, we work with Model (A.2) : $y=X\beta+e$. We have $X\widehat{\beta}_\tau=\tilde{M}_T(\tau)\,y$

and $\hat{E}\left(X\widehat{\beta}_\tau|X\right) \equiv X\beta_\tau = \tilde{M}_T(\tau)X\beta$ where $\hat{E}(.|X)$ denotes the linear projection on X. Hence,

$$\begin{aligned} X\left(\widehat{\beta}_\tau - \beta_\tau\right) &= \tilde{M}_T(\tau)e, \\ X(\beta_\tau - \beta) &= \left(\tilde{M}_T(\tau) - I_T\right)X\beta. \end{aligned}$$

The first equality in Lemma 2 follows from the fact that the cross-product vanishes, indeed :

$$E\left\langle X\left(\widehat{\beta}_\tau - \beta_\tau\right), X(\beta_\tau - \beta)\right\rangle = \sum_j q_j(1-q_j)E\left(e_j x_j'\beta\right) = 0,$$

where q_j denotes $q\left(\tau, \lambda_j^2\right)$.

$$\begin{aligned} E\left\|X\left(\widehat{\beta}_\tau - \beta_\tau\right)\right\|^2 &= E\left(e'\tilde{M}_T(\tau)^2 e\right) \\ &= \omega^2 \sum_j q\left(\tau, \lambda_j^2\right)^2 \\ &\leq \omega^2 \sup q\left(\tau, \lambda_j^2\right) \sum_j q\left(\tau, \lambda_j^2\right) \\ &= O\left(\frac{1}{\tau}\right) \end{aligned}$$

by Lemma 4 of Carrasco (2012). We have

$$\begin{aligned} \frac{1}{T}\|X(\beta_\tau - \beta)\|^2 &= \frac{1}{T}\left\|\left(\tilde{M}_T(\tau) - I_T\right)X\beta\right\|^2 \\ &= \sum_j (q_j - 1)^2 \frac{\langle X\beta, \hat{v}_j\rangle^2}{T} \\ &= \sum_j \lambda_j^2 (q_j - 1)^2 \left\langle \beta, \hat{\phi}_j\right\rangle^2 \\ &= \sum_j \lambda_j^{2+2\nu}(q_j - 1)^2 \frac{\left\langle \beta, \hat{\phi}_j\right\rangle^2}{\lambda_j^{2\nu}} \\ &\leq \sup \lambda_j^{2+2\nu}(q_j - 1)^2 \sum_j \frac{\left\langle \beta, \hat{\phi}_j\right\rangle^2}{\lambda_j^{2\nu}}. \end{aligned}$$

Taking the expectation, one obtains

$$\begin{aligned}\frac{1}{T}E\left\|X\left(\beta_\tau-\beta\right)\right\|^2 &\leq \sup \lambda_j^{2+2\nu}\left(q_j-1\right)^2\sum_j \frac{\langle\beta,\phi_j\rangle^2}{\lambda_j^{2\nu}} \\ &= \left\{\begin{array}{l} O\left(\tau^{(\nu+1)}\right) \text{ for SC and LF,} \\ O\left(\tau^{\min(\nu+1,2)}\right) \text{ for Ridge}\end{array}\right.\end{aligned}$$

by Assumption A and Proposition 3.11 of Carrasco, Florens, and Renault (2007).

Proof of Proposition 1.

First we analyze the elements of (1.12) :

$$\begin{aligned}&\hat{\beta}_\tau\left(\mu'\beta\right)-\beta\left(\hat{\mu}'\hat{\beta}_\tau\right)\\ =&\left(\hat{\beta}_\tau-\beta\right)\left(\mu'\beta\right)-\beta\left(\hat{\mu}'\hat{\beta}_\tau-\mu'\left(\beta-\hat{\beta}_\tau+\hat{\beta}_\tau\right)\right)\\ =&\left(\hat{\beta}_\tau-\beta\right)\left(\mu'\beta\right)-\beta\left(\left(\hat{\mu}-\mu\right)'\hat{\beta}_\tau+\mu'\left(\hat{\beta}_\tau-\beta\right)\right)\\ =&\underbrace{\left(\hat{\beta}_\tau-\beta\right)\mu'\beta}_{\text{(a1)}}-\underbrace{\beta\left(\hat{\mu}-\mu\right)'\left(\hat{\beta}_\tau-\beta\right)}_{\text{(a2)}}-\underbrace{\beta\left(\hat{\mu}-\mu\right)'\beta}_{\text{(a3)}}-\underbrace{\beta\mu'\left(\hat{\beta}_\tau-\beta\right)}_{\text{(a4)}}\end{aligned} \quad \text{(A.3)}$$

Term (a2) : $\left|\left(\hat{\mu}-\mu\right)'\left(\hat{\beta}_\tau-\beta\right)\right|^2\leq\left\|\hat{\mu}-\mu\right\|^2\left\|\hat{\beta}_\tau-\beta\right\|^2=O_p\left(\frac{N}{T}\right)\left\|\hat{\beta}_\tau-\beta\right\|^2$.

Term (a3) : $\left|\left(\hat{\mu}-\mu\right)'\beta\right|^2\leq\left\|\hat{\mu}-\mu\right\|^2\left\|\beta\right\|^2=O_p\left(\frac{N}{T}\right)$ because $\left\|\beta\right\|^2<\infty$.

So both terms (a2) and (a3) will be negligible compared to $\left\|\hat{\beta}_\tau-\beta\right\|^2$ by Assumption B(iv) and Lemma 3.

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

Since $\mu'\left(\hat{\beta}_\tau-\beta\right)=o_p\left(1\right)$, we have

$$\frac{\hat{\beta}_\tau-\beta}{\left(1-\hat{\mu}'\hat{\beta}_\tau\right)}=\frac{\hat{\beta}_\tau-\beta}{\left(1-\mu'\beta\right)}+O_p\left(\left(\hat{\beta}_\tau-\beta\right)\mu'\left(\hat{\beta}_\tau-\beta\right)\right).$$

$$\begin{aligned}
&\left(\hat{\beta}_\tau-\beta\right)'\Sigma\left(\hat{\beta}_\tau-\beta\right)\\
=\ &\left(\hat{\beta}_\tau-\beta\right)'\hat{\Sigma}\left(\hat{\beta}_\tau-\beta\right)+\left(\hat{\beta}_\tau-\beta\right)'\left(\Sigma-\hat{\Sigma}\right)\left(\hat{\beta}_\tau-\beta\right).
\end{aligned}$$

Moreover,

$$\begin{aligned}
\left(\hat{\beta}_\tau-\beta\right)'\left(\Sigma-\hat{\Sigma}\right)\left(\hat{\beta}_\tau-\beta\right) &\leq \left\|\hat{\beta}_\tau-\beta\right\|^2\left\|\Sigma-\hat{\Sigma}\right\|\\
&= O_p\left(\frac{1}{\sqrt{T}}\left\|\hat{\beta}_\tau-\beta\right\|^2\right)
\end{aligned}$$

by Theorem 4 of Carrasco and Florens (2000) and the Hilbert-Schmidt assumption of Σ.

$$\begin{aligned}
&\left(\hat{\beta}_\tau-\beta\right)'\Sigma\left(\hat{\beta}_\tau-\beta\right)\\
=\ &\left(\hat{\beta}_\tau-\beta\right)'\left[\frac{R'R}{T}-\left(\frac{R'1}{T}\right)\left(\frac{R'1}{T}\right)'\right]\left(\hat{\beta}_\tau-\beta\right)+O_p\left(\frac{\left\|\hat{\beta}_\tau-\beta\right\|^2}{\sqrt{T}}\right)\\
=\ &\frac{1}{T}\left(\hat{\beta}_\tau-\beta\right)'R'R\left(\hat{\beta}_\tau-\beta\right)-\left(\hat{\beta}_\tau-\beta\right)'\hat{\mu}\hat{\mu}'\left(\hat{\beta}_\tau-\beta\right)+O_p\left(\frac{\left\|\hat{\beta}_\tau-\beta\right\|^2}{\sqrt{T}}\right)\\
=\ &\frac{1}{T}\left\|R\left(\hat{\beta}_\tau-\beta\right)\right\|^2-\left(\hat{\mu}'\left(\hat{\beta}_\tau-\beta\right)\right)^2+O_p\left(\frac{\left\|\hat{\beta}_\tau-\beta\right\|^2}{\sqrt{T}}\right).
\end{aligned}$$

Using (1.12) and (A.3), we obtain

$$\begin{aligned}
\gamma\left(\hat{x}_\tau-x^*\right) &= \frac{\left(\hat{\beta}_\tau-\beta\right)}{(1-\mu'\beta)}+\frac{\beta\mu'\left(\hat{\beta}_\tau-\beta\right)}{(1-\mu'\beta)^2}\\
&\quad+O_p\left(\left(\hat{\beta}_\tau-\beta\right)\mu'\left(\hat{\beta}_\tau-\beta\right)\right)+O_p\left(\sqrt{\frac{N}{T}}\right).
\end{aligned}$$

$$
\begin{aligned}
& \left(\hat{x}_\tau - x^*\right)' \Sigma \left(\hat{x}_\tau - x^*\right) && \text{(A.4)} \\
= \quad & \frac{\left(\hat{\beta}_\tau - \beta\right)' \Sigma \left(\hat{\beta}_\tau - \beta\right)}{\left(1-\mu'\beta\right)^2} \\
& + \frac{\left(\hat{\beta}_\tau - \beta\right)' \mu\beta'\Sigma\beta\mu' \left(\hat{\beta}_\tau - \beta\right)}{\left(1-\mu'\beta\right)^4} \\
& + 2\frac{\left(\hat{\beta}_\tau - \beta\right)' \Sigma\beta\mu' \left(\hat{\beta}_\tau - \beta\right)}{\left(1-\mu'\beta\right)^3} \\
& + O_p\left(\left(\hat{\beta}_\tau - \beta\right)' \Sigma \left(\hat{\beta}_\tau - \beta\right) \mu' \left(\hat{\beta}_\tau - \beta\right)\right) + o_p\left(\sqrt{\frac{N}{T}} \Sigma \left(\hat{\beta}_\tau - \beta\right)\right).
\end{aligned}
$$

Replacing Σ by $R'R/T - \hat{\mu}\hat{\mu}'$, (A.4) is equal to

$$
\begin{aligned}
& \frac{\frac{1}{T}\left\| R\left(\hat{\beta}_\tau - \beta\right)\right\|^2}{\left(1-\mu'\beta\right)^2} && \text{(A.5)} \\
& - \frac{\left(\hat{\mu}'\left(\hat{\beta}_\tau - \beta\right)\right)^2}{\left(1-\mu'\beta\right)^2} && \text{(A.6)} \\
& + \frac{1}{T}\frac{\left\| R\beta\mu'\left(\hat{\beta}_\tau - \beta\right)\right\|^2}{\left(1-\mu'\beta\right)^4} && \text{(A.7)} \\
& - \frac{\left(\widehat{\mu}'\beta\mu'\left(\hat{\beta}_\tau - \beta\right)\right)^2}{\left(1-\mu'\beta\right)^4} && \text{(A.8)} \\
& + \frac{2}{T}\frac{\left(\hat{\beta}_\tau - \beta\right)' R'R\beta\mu' \left(\hat{\beta}_\tau - \beta\right)}{\left(1-\mu'\beta\right)^3} && \text{(A.9)} \\
& - 2\frac{\left(\hat{\beta}_\tau - \beta\right)' \widehat{\mu}\left(\widehat{\mu}'\beta\right)\mu' \left(\hat{\beta}_\tau - \beta\right)}{\left(1-\mu'\beta\right)^3} && \text{(A.10)} \\
& + rest\left(\tau\right). && \text{(A.11)}
\end{aligned}
$$

Note that $\widehat{\mu}'\beta$ is a scalar that can be approximated by $\mu'\beta$. The next step consists in taking the expectation of each term and writing $\hat{\beta}_\tau - \beta = \hat{\beta}_\tau - \beta_\tau + \beta_\tau - \beta$.

We turn our attention to $E\left[\left(\hat{\mu}'\left(\hat{\beta}_\tau - \beta\right)\right)^2\right]$. The variance term is given by

$$\begin{aligned}
\frac{1}{T^2}E\left[\left(\hat{\mu}'\left(\hat{\beta}_\tau - \beta_\tau\right)\right)^2\right] &= \frac{1}{T^2}E\left[\left(\frac{1_T'}{T}R\left(\hat{\beta}_\tau - \beta_\tau\right)\right)^2\right] \\
&= \frac{1}{T^2}E\left[\left(\frac{1_T'}{T}M_T(\tau)\,e\right)^2\right] \\
&= \frac{1}{T^2}E\left(e'M_T(\tau)\,\frac{1_T 1_T'}{T^2}M_T(\tau)\,e\right) \\
&= \frac{1}{T^2}E\left(\left(\sum q_j e_j\right)^2\right) \\
&= \frac{\sum q_j^2\omega^2}{T^2} + \frac{1}{T^2}\sum_j\sum_{i\neq j} q_j q_i \omega^2 \\
&\leq \frac{\sum q_j^2\omega^2}{T^2} + \frac{\omega^2}{T^2}\sum_j q_j \sum_i q_i \qquad \text{(A.12)} \\
&\leq \frac{1}{T^2}E\left\|R\left(\hat{\beta}_\tau - \beta_\tau\right)\right\|^2 + O\left(\frac{1}{\tau^2 T^2}\right) \qquad \text{(A.13)}
\end{aligned}$$

Therefore, this term is negligible with respect to (A.5). However, the bias term is not :

$$\begin{aligned}
\left(\hat{\mu}'\left(\beta_\tau - \beta\right)\right)^2 &= \left(\frac{1_T'}{T}\left(M_T(\tau) - I_T\right)R\beta\right)^2 \\
&\leq \frac{\left\|1_T\right\|^2}{T^2}\left\|\left(M_T(\tau) - I_T\right)R\beta\right\|^2 \\
&\leq \frac{1}{T}\left\|\left(M_T(\tau) - I_T\right)R\beta\right\|^2.
\end{aligned}$$

Therefore, the bias of (A.6) may be of the same order as that of (A.5).

Let us examine the term (A.7) :

The variance corresponding to (A.7) is

$$\frac{1}{T}E\left[\left(\hat{\beta}_\tau - \beta_\tau\right)'\mu\beta'R'R\beta\mu'\left(\hat{\beta}_\tau - \beta_\tau\right)\right]$$

Note that $\frac{R'R}{T}\beta = \frac{R'R}{T}E\left(\frac{R'R}{T}\right)^{-1}\mu \simeq \mu$. Replacing μ by $\hat{\mu}$, we get

$$\begin{aligned} & \frac{1}{T}E\left[\left(\hat{\beta}_\tau - \beta_\tau\right)'\mu\beta' R' R\beta\mu'\left(\hat{\beta}_\tau - \beta_\tau\right)\right] \\ \simeq \; & \beta'\mu E\left[\left(\hat{\mu}'\left(\hat{\beta}_\tau - \beta_\tau\right)\right)^2\right] \end{aligned}$$

which is negligible.

The bias corresponding to (A.7) is

$$\begin{aligned} & \frac{\left(\beta_\tau - \beta\right)'\mu\beta'\frac{R'R}{T}\beta\mu'\left(\beta_\tau - \beta\right)}{\left(1-\mu'\beta\right)^4} \\ \simeq \; & \mu'\beta\frac{\left(\mu'\left(\beta_\tau - \beta\right)\right)^2}{\left(1-\mu'\beta\right)^4}. \end{aligned}$$

Similarly, the variances of the terms (A.8) to (A.10) are negligible. The bias term corresponding to (A.8) is equal to

$$\begin{aligned} & -\frac{\left(\widehat{\mu}'\beta\mu'\left(\beta_\tau - \beta\right)\right)^2}{\left(1-\mu'\beta\right)^4} \\ \simeq \; & -\frac{\left(\mu'\beta\right)^2\left(\mu'\left(\beta_\tau - \beta\right)\right)^2}{\left(1-\mu'\beta\right)^4}. \end{aligned}$$

Combining the biases of (A.7) and (A.8), we obtain

$$\frac{\left(\mu'\beta\right)\left(\mu'\left(\beta_\tau - \beta\right)\right)^2}{\left(1-\mu'\beta\right)^3}.$$

The bias corresponding to the term (A.9) is given by

$$\begin{aligned} & 2\frac{\left(\beta_\tau - \beta\right)'\frac{R'R}{T}\beta\mu'\left(\beta_\tau - \beta\right)}{\left(1-\mu'\beta\right)^3} \\ \simeq \; & 2\frac{\left(\mu'\left(\beta_\tau - \beta\right)\right)^2}{\left(1-\mu'\beta\right)^3}. \end{aligned}$$

The bias term corresponding to (A.10) is equal to

$$\begin{aligned} & -2\frac{\left(\beta_\tau - \beta\right)'\widehat{\mu}\left(\widehat{\mu}'\beta\right)\mu'\left(\beta_\tau - \beta\right)}{\left(1-\mu'\beta\right)^3} \\ \simeq \; & -2\frac{\left(\mu'\beta\right)\left(\mu'\left(\beta_\tau - \beta\right)\right)^2}{\left(1-\mu'\beta\right)^3}. \end{aligned}$$

Combining the bias terms of (A.6) to (A.10), we obtain

$$\begin{aligned}&\frac{(\mu'\beta)(\mu'(\beta_\tau-\beta))^2}{(1-\mu'\beta)^3}+\frac{(\mu'(\beta_\tau-\beta))^2}{(1-\mu'\beta)^2}\\=\;&\frac{(\mu'(\beta_\tau-\beta))^2}{(1-\mu'\beta)^3}.\end{aligned}$$

Combining all the terms together, we obtain :

$$\begin{aligned}&\gamma^2(1-\mu'\beta)^2E\left[(\hat{x}_\tau-x^*)'\Sigma(\hat{x}_\tau-x^*)\right]\\=\;&\frac{1}{T}E\left\|R\left(\hat{\beta}_\tau-\beta\right)\right\|^2+\frac{(\mu'(\beta_\tau-\beta))^2}{(1-\mu'\beta)}+rest(\tau).\end{aligned}$$

Proof of Proposition 2.

Intermediate result : Let $\Omega=E(x_tx_t')=E(X'X)/T=E(R'R)/T.$

$$\begin{aligned}R_T^*(\tau)&=\frac{1}{T}E\left\|X\beta-X\widehat{\beta}_\tau\right\|^2+(1.15)\\&=\frac{1}{T}\sum_j(1-q_j)^2(\beta'\Omega\beta)+\frac{\omega^2}{T}\sum_j q_j^2+(1.15)\end{aligned}\qquad\text{(A.14)}$$

where $\omega^2=E\left(e_j^2\right).$

Proof of (A.14) : Let $A_T(\tau)=I-M_T(\tau)$. Following the proof of Li (1987), we have

$$\begin{aligned}&\frac{1}{T}E\left\|X\beta-X\widehat{\beta}_\tau\right\|^2\\=\;&\frac{1}{T}E\left\|X\beta-M_T(\tau)(X\beta+e)\right\|^2\\=\;&\frac{1}{T}E\left\|A_T(\tau)X\beta-M_T(\tau)e\right\|^2\\=\;&\frac{1}{T}\left\|A_T(\tau)X\beta\right\|^2+\frac{1}{T}E\left\|M_T(\tau)e\right\|^2+\frac{2}{T}E\left[\langle A_T(\tau)X\beta,M_T(\tau)e\rangle\right].\end{aligned}$$

As $M_T(\tau)$ and $A_T(\tau)$ are diagonal matrices, we have

$$\begin{aligned}\frac{1}{T}\left\|A_T(\tau)X\beta\right\|^2&=\frac{1}{T}\sum_j(1-q_j)^2E\left[\left(x_j'\beta\right)^2\right]\\&=\frac{1}{T}\sum_j(1-q_j)^2(\beta'\Omega_T\beta),\end{aligned}$$

where $\beta'\Omega_T\beta = \mu'\beta < 1$,

$$E\left\langle A_T(\tau) X\beta, M_T(\tau) e\right\rangle = \sum_j q_j (1 - q_j) E\left(e_j x_j'\beta\right) = 0$$

and $E\left(\|M_T(\tau) e\|^2\right) = E\left(\sum_j q_j^2 e_j^2\right) = \omega^2 \sum_j q_j^2$.

First we establish the optimality for a Mallows' C_L criterion where ω^2 is assumed to be known. Let

$$\hat{\tau}_C = \arg\min_{\tau} \{Crit(\tau)\}$$

where

$$Crit(\tau) = \frac{1}{T}\|(I_T - M_T(\tau)) y\| + 2\omega^2 \frac{1}{T} tr M_T(\tau) + (1.15)$$

We want to show that

$$\frac{L_T(\hat{\tau}_C)}{\inf_\tau L_T(\tau)} \xrightarrow{P} 1. \tag{A.15}$$

$Crit(\tau)$ can be rewritten as

$$\begin{aligned} & Crit(\tau) \\ = \; & \frac{1}{T}\|e\|^2 + L_T(\tau) + \frac{2}{T}\left\langle e, A_T(\tau) X\beta\right\rangle \\ & + \frac{2}{T}\left(\omega^2 tr M_T(\tau) - \left\langle e, M_T(\tau) e\right\rangle\right) \\ & + (1.15) - \frac{\left(\mu'(\beta_\tau - \beta)\right)^2}{1 - \mu'\beta} \\ & - rest(\tau). \end{aligned}$$

To establish (A.15), we need to show that in probability

$$Crit(\tau) - c_T = L_T(\tau)\left(1 + \varepsilon_T(\tau)\right)$$

where

$$\sup_\tau |\varepsilon_T(\tau)| = o_p(1).$$

Hence, to establish (A.15), it is sufficient to show that

$$\sup_{\tau} \frac{\left|\frac{1}{T}\left\langle e, A_T(\tau) X\beta\right\rangle\right|}{R_T^*(\tau)} \rightarrow 0, \tag{A.16}$$

$$\sup_{\tau} \frac{\frac{1}{T}\left|\omega^2 trM_T(\tau) - \left\langle e, M_T(\tau) e\right\rangle\right|}{R_T^*(\tau)} \rightarrow 0 \tag{A.17}$$

$$\sup_{\tau} \frac{\left|rest(\tau)\right|}{R_T^*(\tau)} \rightarrow 0, \tag{A.18}$$

$$\sup_{\tau} \left|(1.15) - \frac{\left(\mu'\left(\beta_\tau - \beta\right)\right)^2}{1-\mu'\beta}\right| / R_T^*(\tau) \rightarrow 0, \tag{A.19}$$

$$\sup_{\tau} \left|\frac{L_T(\tau)}{R_T^*(\tau)} - 1\right| \rightarrow 0 \tag{A.20}$$

Proof of the optimality of Mallows' C_L criterion for SC and LF :

Now we focus to the case of SC and LF regularizations. By a slight abuse of notation, τ denotes from now on the number of eigenvalues retained in SC or the number of iterations in LF. Note that τ is an integer number and lies in a discrete index set $H_T = \{1, 2, ..., N\}$ for SC since $N \leq T$ and $H_T = \{1, 2, ..., T\}$ for LF. We are going to check the conditions (A.16) to (A.20). We follow a proof similar to that of Li (1987).

Consider (A.16). By Chebyshev's inequality, we have

$$P\left[\sup_{\tau \in H_T} \frac{\left|\frac{1}{T}\left\langle e, A_T(\tau) X\beta\right\rangle\right|}{R_T^*(\tau)} > \delta\right] \leq \sum_{\tau \in H_T} \frac{\frac{1}{T^{2m}} E\left[\left\langle e, A_T(\tau) X\beta\right\rangle^{2m}\right]}{\delta^{2m} R_T^*(\tau)^{2m}} \tag{A.21}$$

for some integer m. Using the fact that $A_T(\tau)$ is diagonal, we have

$$\begin{aligned} E\left[\left\langle e, A_T(\tau) X\beta\right\rangle^{2m}\right] &= E\left[\left(\sum_j e_j\left(1-q_j\right) x_j'\beta\right)^{2m}\right] \\ &< C\left(\sum_j E\left(\left(e_j x_j'\beta\right)^{2m}\right)^{1/m} \left(1-q_j\right)^2\right)^m \end{aligned}$$

for some constant C by Theorem 2 of Whittle (1960) using the fact that $e_j x_j'\beta$ are independent and have mean zero (which follows from r_t, u_t i.i.d.

and $E(u_t r_t) = 0$). Moreover, Lp inequality implies that

$$\begin{aligned}\frac{1}{T}\sum_j E\left(\left(e_j x_j'\beta\right)^{2m}\right)^{1/m}(1-q_j)^2 &\leq \frac{1}{T}\sum_j E\left(\left(e_j x_j'\beta\right)^2\right)(1-q_j)^2 \\ &\leq C\frac{1}{T}\beta'\Omega_T\beta\sum_j (1-q_j)^2 \qquad \text{(A.22)} \\ &\leq CR_T^*(\tau)\end{aligned}$$

where inequality (A.22) follows from the fact that $e_t = y_t - x_t'\beta$ and $e'X\beta = \sum_t e_t x_t'\beta$. Moreover as $e_t x_t'\beta$ are independent with mean zero, we have

$$E\left(\left(e_t x_t'\beta\right)^2\right) = E\left(\frac{\left(e'X\beta\right)^2}{T}\right) = E\left(\frac{\left(u'R\beta\right)^2}{T}\right) = E\left(\left(u_t r_t'\beta\right)^2\right).$$

Now replacing u_t by $1 - r_t'\beta$, we obtain

$$E\left(\left(u_t r_t'\beta\right)^2\right) \leq E\left(\left(r_t'\beta\right)^2\right) + E\left(\left(r_t'\beta\right)^4\right) < CE\left(\left(r_t'\beta\right)^2\right)$$

where C can be taken equal to $2\big(E\left(\left(r_t'\beta\right)^2\right) + E\left(\left(r_t'\beta\right)^4\right)\big)/E\left(\left(r_t'\beta\right)^2\right)$.

Therefore, (A.21) is no greater than $C\delta^{-2m}\sum_{\tau\in H_T}\left(TR_T^*(\tau)\right)^{-m}$ which tends to zero by assumption B3.

Now, consider (A.17),

$$\omega^2 tr M_T(\tau) - \langle e, M_T(\tau)e\rangle = \sum_j \left(\omega^2 - e_j^2\right)q_j.$$

Again by Chebyshev's inequality,

$$P\left[\sup_\tau \frac{\frac{1}{T}\left|\omega^2 tr M_T(\tau) - \langle e, M_T(\tau)e\rangle\right|}{R_T^*(\tau)} > \delta\right] \leq \sum_\tau \frac{\frac{1}{T^{2m}}E\left[\left(\sum_j\left(\omega^2 - e_j^2\right)q_j\right)^{2m}\right]}{R_T^*(\tau)^{2m}}.$$

By Whittle (1960, Theorem 2), we have

$$\begin{aligned}\frac{1}{T^{2m}}E\left[\left(\sum_t\left(\omega^2 - e_t^2\right)q_t\right)^{2m}\right] &\leq C\frac{1}{T^{2m}}\left(\sum_j q_j^2 E\left(\left(\omega^2 - e_j^2\right)^{2m}\right)^{1/m}\right)^m \\ &\leq C'\frac{1}{T^{2m}}\left(\sum_j q_j^2\right)^m \\ &\leq C'\frac{1}{T^m}R_T^*(\tau).\end{aligned}$$

Hence, (A.17) holds by Assumption B3.

Consider (A.19). Using $\mu = \mu - \hat{\mu} + \hat{\mu}$, we have

$$\begin{aligned} & (1.15) - \frac{(\mu'(\beta_\tau - \beta))^2}{1 - \mu'\beta} \\ = \ & (1.15) - \frac{(\hat{\mu}'(\beta_\tau - \beta))^2}{1 - \mu'\beta} - \frac{\left((\mu - \hat{\mu})'(\beta_\tau - \beta)\right)^2}{1 - \mu'\beta} - \frac{\hat{\mu}'(\beta_\tau - \beta)(\mu - \hat{\mu})'(\beta_\tau - \beta)}{1 - \mu'\beta} \end{aligned}$$

Note that $\widehat{\mu}'(\beta_\tau - \beta) = \frac{1_T'(M_T(\tau) - I_T)R\beta}{T}$ and

$$\frac{\left(1_T'(M_T(\tau) - I_T)R\hat{\beta}\right)^2}{T^2\left(1 - \widehat{\mu}'\widehat{\beta}\right)} = \frac{\left(1_T'(M_T(\tau) - I_T)R\hat{\beta}\right)^2}{T^2(1 - \mu'\beta)} + \frac{\left(1_T'(M_T(\tau) - I_T)R\hat{\beta}\right)^2}{T^2(1 - \mu'\beta)^2}\left(\mu'\beta - \widehat{\mu}'\widehat{\beta}\right).$$

Moreover,

$$\begin{aligned} & \left|\frac{\left(1_T'(M_T(\tau) - I_T)R\hat{\beta}\right)^2}{T^2} - (\widehat{\mu}'(\beta_\tau - \beta))^2\right| \\ = \ & \left|\frac{1_T'(M_T(\tau) - I_T)R\left(\hat{\beta} - \beta\right)}{T}\frac{1_T'(M_T(\tau) - I_T)R\left(\hat{\beta} + \beta\right)}{T}\right| \\ \leq \ & \frac{\|1_T'(M_T(\tau) - I_T)\|^2}{T}\frac{\left\|R\left(\hat{\beta} - \beta\right)\right\|}{\sqrt{T}}\frac{\left\|R\left(\hat{\beta} + \beta\right)\right\|}{\sqrt{T}} \\ = \ & \sum_j \frac{(q_j - 1)^2}{T} o_p(1) \\ \leq \ & CR_T^*(\tau)\, o_p(1) \end{aligned}$$

because $\left\|R\left(\hat{\beta} + \beta\right)\right\|^2/T = \left(\hat{\beta} + \beta\right)'\left(\sum r_t r_t'/T\right)\left(\hat{\beta} + \beta\right) < \infty$ because $\beta'\Omega_T\beta < 1$.

Consider (A.20). Given $L_T = L_T^* + rest$ and (A.18) holds, it suffices to prove the result for L_T^* instead of L_T. Let $d(\tau)$ denote $(\mu'(\beta_\tau - \beta))^2/(1 -$

$\mu'\beta$).

$$\begin{aligned} L_T^*(\tau) &= \frac{1}{T}\left\|X\left(\hat{\beta}_\tau-\beta\right)\right\|^2+d(\tau) \\ &= \frac{1}{T}\|A_T(\tau)X\beta-M_T(\tau)e\|^2+d(\tau) \\ &= \frac{1}{T}\|A_T(\tau)X\beta\|^2+\frac{1}{T}\|M_T(\tau)e\|^2-\frac{2}{T}\langle A_T(\tau)X\beta,M_T(\tau)e\rangle+d(\tau). \end{aligned}$$

$$\begin{aligned} L_T^*(\tau)-R_T^*(\tau) &= \frac{1}{T}\left(\|M_T(\tau)e\|^2-E\left(\|M_T(\tau)e\|^2\right)\right) \\ &\quad -\frac{2}{T}\langle A_T(\tau)X\beta,M_T(\tau)e\rangle. \end{aligned}$$

Hence, we need to show that

$$\begin{aligned} \sup_\tau \frac{\frac{1}{T}|\langle A_T(\tau)X\beta,M_T(\tau)e\rangle|}{R_T^*(\tau)} &\to 0, &\text{(A.23)} \\ \sup_\tau \frac{\frac{1}{T}\left(\|M_T(\tau)e\|^2-E\left(\|M_T(\tau)e\|^2\right)\right)}{R_T^*(\tau)} &\to 0. &\text{(A.24)} \end{aligned}$$

Note that $|\langle A_T(\tau)X\beta,M_T(\tau)e\rangle| = \left|\sum_j q_j(1-q_j)x_j'\beta e_j\right| \leq \left|\sum_j (1-q_j)x_j'\beta e_j\right|$ because $\sup|q|\leq 1$. Similarly, $\left|\|M_T(\tau)e\|^2-E\left(\|M_T(\tau)e\|^2\right)\right| = \left|\sum_j\left(\omega^2-e_j^2\right)q_j^2\right| \leq \left|\sum_j\left(\omega^2-e_j^2\right)q_j\right|$. So the same proof as for (A.16) and (A.17) can be used to establish (A.23) and (A.24).

This completes the proof of the optimality of Mallows' C_L criterion for SC and LF.

<u>Relating the GCV to the C_L via the Nil trace estimator</u> :
Denote $\gamma_T = X\beta$ and $\hat{\gamma}_T = M_T(\tau)y_T$.

The GCV can be related to C_L via the nil-trace estimate, defined from the estimate of interest $\hat{\gamma}_T$ as

$$\bar{\gamma}_T = -\alpha y_T + (1+\alpha)\hat{\gamma}_T$$

with

$$\alpha = \frac{n^{-1}trM_T(\tau)}{1-n^{-1}trM_T(\tau)}$$

In other words, $\bar{\gamma}_T = \bar{M}_T(\tau)\, y_T$ with

$$\bar{M}_T(\tau) = -\alpha I + (1+\alpha) M_T(\tau)$$

From the previous expression, $tr\bar{M}_T(\tau) = 0$. Then selecting a model from the class $\{\hat{\gamma}_T(\tau)\ :\ \tau \in H_T\}$ using the GCV criterion, is equivalent to using the C_L to select an estimate from $\{\bar{\gamma}_T(\tau)\ :\ \tau \in H_T\}$. Following Li 87, we adopt the following assumptions :

Assumption C

A4 : $\inf\limits_{\tau \in H_T} L_T(\tau) \to 0$

A5 : $for\ any\ \{\tau_T \in H_T\}\ such\ that\ :\ \ T^{-1} tr M_T(\tau)\, M_T(\tau)' \to 0$
we $have\ \frac{\left(T^{-1} tr M_T(\tau)\right)^2}{T^{-1} tr M_T(\tau) M_T(\tau)'} \to 0$

A6 : $\sup\limits_{\tau \in H_T} T^{-1} tr M_T(\tau) \le \gamma_1 \quad for\ some\ 1 > \gamma_1 > 0$

A7 : $\sup\limits_{\tau \in H_T} \frac{\left(T^{-1} tr M_T(\tau)\right)^2}{T^{-1} tr M_T(\tau) M_T(\tau)'} \le \gamma_2 \quad for\ some\ 1 > \gamma_2 > 0$

Under assumption A1-A7, the crucial inequality (6.3 and 6.4 in Li, 87)

$$c_1 \le \frac{tr\bar{M}_T(\tau)\,\bar{M}_T(\tau)'}{tr M_T(\tau)\, M_T(\tau)'} \le c_2, \quad c_1 \le \frac{\bar{R}_T(\tau)}{R_T(\tau)} \le c_2,$$

where c_1 and c_2 are two positive constants, still hold. The justification of the replacement of the original estimate by the nil-trace estimate is then essentially the same as in Li 87, Theorem 3.2.

Let us check that the assumption C is satisfied for SC. A4 is implied by Lemma 3 in Appendix. Then note that

$$T^{-1} tr M_T(\tau)\, M_T(\tau)' = \frac{1}{T} \sum_j q_j = \frac{1}{T\tau} = \frac{N_\tau}{T}$$

where N_τ is the number of selected eigenvectors. A5 is trivially satisfied. A6 and A7 are equivalent to

$$\frac{N_\tau}{T} < 1. \tag{A.25}$$

Given that N_τ is necessarily smaller than N (the number of assets) and N is limited by assumption B(iv), condition (A.25) holds. The asymptotic optimality of SC follows. Similar proof apply to LF assuming some extra conditions.

A.2 Homotopy - LARS Algorithms for penalized least-squares

Homotopy (Continuation) is a general approach for solving a system of equation by tracking the solution of nearby system of parametrized equation. In the penalized Lasso case the Homotopy variable is the penalty term. We give below a detailed description of the Homotopy/LARS algorithm which provides the solution path to the l1-penalized least-squares objective function :

$$\tilde{x}(\tau) = \arg\min_{x} \|y - Rx\|_2^2 + \tau \|x\|_1 .$$

The solution to this minimization problem $\tilde{x}(\tau)$ is provided as a continuous piecewise function of the penalty τ satisfying the variational equations given by :

$$\begin{cases} (R'(y-Rx))_i = \frac{\tau}{2}\mathrm{sgn}(x_i) & x_i \neq 0 \\ |(R'(y-Rx))_i| \leq \frac{\tau}{2} & x_i = 0 \end{cases}$$

Meaning that the residual correlations $b_i = (R'(y-Rx))_i$ corresponding to non zero weights are equal to $\tau/2$ in absolute value, while the absolute residual correlation corresponding the zero weights must by bounded by $\tau/2$. Throughout the algorithm, it is critical to identify the set of active elements, that is the components with non zeros weights. At a given iteration k of the algorithm this set is denoted by $J_k = \left\{ i \;\; for\ which\, |b_i| = \frac{\tau^k}{2} \right\}$, and also corresponds to the set of maximal residual correlations components.

The algorithm starts with an initial solution satisfying the variational equations, for a penalty term suitably chosen. The obvious initial solution is obtained by setting all the weights to zeros. The corresponding penalty term τ_0, must then satisfy $\tau_0 \equiv 2\max_i |(R'y)_i|$. Hence we have that $\tilde{x}(\tau) = 0$ for all $\tau \geq \tau_0$. This allow us to set $J_1 = \{i^*\}$, where $i^* = arg\max_i |(R'y)_i|$.

From one iteration k to the next, the algorithm manages to update the active set J_k, which represents the support of $\tilde{x}(\tau_k)$, so that the first-order conditions remains satisfied. Hence in each iteration $k+1$, the vector b decreases at the same rate γ^{k+1} in the active set to preserve the same level of correlation for active elements.

$$(b^{k+1})_{J_{k+1}} = (b^k)_{J_{k+1}} - \gamma^{k+1}(sign(b^k))_{J_{k+1}}$$

This result is obtained by updating the optimal weights while moving along a walking direction u^{k+1} :

$$x(\tau^{k+1}) = x(\tau^k) + \gamma^{k+1} u^{k+1}$$

Denote R_J the submatrix consisting of the columns J of R, the walking direction u^{n+1} is a solution to a linear system :

$$R'_{J_{k+1}} R_{J_{k+1}} (u^{k+1})_{J_{k+1}} = (\operatorname{sgn}(b^k)_{J_{k+1}}) = \left(\operatorname{sgn}(b^k_j)_{j \in J_{k+1}}\right) = v^{k+1}$$

The remaining components of u^{k+1} are set to *zero* that is :

$$u_i^{k+1} = 0 \quad for \;\; i \notin J_{k+1}$$

The step γ_{k+1} to make in direction u^{k+1} to find $x(\tau^{k+1})$ is the minimum value such that an inactive element becomes active or the reverse.

If an inactive element i becomes active, it means that its correlation reached the maximal correlation in the descent procedure. And then is must be case that :

$$\left|b_i^k - \gamma^{k+1} r_i'^{k+1}\right| = \left|(b^k)_{J_{k+1}} - \gamma^{k+1} v_{J_{k+1}}\right| = \tau_{k+1} = \frac{\tau^k}{2} - \gamma^{k+1}$$

with r_i is the i^{th} column of R. This implies that :

$$\gamma^{k+1} = \frac{\frac{\tau^k}{2} - b_i^k}{1 - r_i'^{k+1}} \; or \; \gamma^{n+1} = \frac{\frac{\tau^k}{2} + b_i^k}{1 + r_i'^{k+1}}$$

The optimal step is then given by :

$$\gamma_+^{k+1} = \min_{i \in J^c}{}^+ \left\{ \frac{\frac{\tau^k}{2} - b_i^k}{1 - r_i'^{k+1}} \; ; \frac{\frac{\tau^k}{2} + b_i^k}{1 + r_i'^{k+1}} \right\}$$

On the other hand, if γ^{n+1} is such that an active element i reaches zero then (**??**) implies that :

$$\gamma_-^{k+1} = -\frac{x_i^k}{u_i^{k+1}}$$

The smallest step to make so that an element leaves the active set :

$$\gamma_-^{k+1} = \min_{i \in J_{k+1}}{}^+ \left(-\frac{x_i^k}{u_i^{k+1}} \right)$$

Finally the next step is given by :

$$\gamma^{k+1} = \min\left\{\gamma_+^{k+1}, \gamma_-^{k+1}\right\}$$

At the end of each stage the corresponding penalty term is $\tau^{k+1} = \tau^k - 2\gamma^{k+1}$ which is smaller than τ^k. We stop when τ^{k+1} becomes negative. After $q+1$ iterations the Algorithm provides $q+1$ breakpoints $\tau_0 > \tau_1 > \cdots > \tau_q$ and their corresponding minimizers $x(\tau_i)$. From there, the optimal solution for any τ can be deduced by linear interpolation.

Annexe B

Chapter 2

B.1 Additional tables

TABLE B.1 – Time varying industry betas relative to equity factors

The table reports the sample mean of the time-varying equity-related factor betas estimated from the system of Equations (2.19)-(2.21). The portfolios are the 48 Fama French industry portfolioS and a set of global minimum portfolio constructed using different percentage of PCs from 10% to 100%. The equity risk factors are the returns on the World market index from MSCI, the returns on small minus big firms (SMB), and returns on high minus low book-to-market value firms (HML).

Panel A : Mean of the estimated time-varying betas			
Industry Portfolios	WM	SMB	HML
1oN	0.842	0.372	-0.573
Gmin(100%)	0.495	0.061	-0.044
Gmin(90%)	0.518	-0.036	0.023
Gmin(80%)	0.559	0.006	-0.1
Gmin(70%)	0.52	0.054	-0.094
Gmin(60%)	0.556	0.048	-0.183
Gmin(50%)	0.551	0.001	-0.149
Gmin(40%)	0.593	-0.02	-0.187
Gmin(30%)	0.614	-0.065	-0.227
Gmin(20%)	0.648	-0.001	-0.236
Gmin(10%)	0.692	0.121	-0.326

Table B.1 (Continued) : Time varying portfolio betas relative to the equity factors

Industry Portfolios	WM	SMB	HML
Panel A : Mean of the estimated time-varying betas			
Agric	0.706	0.531	-0.506
Food	0.546	-0.162	-0.271
Soda	0.702	0.061	-0.206
Beer	0.617	-0.13	-0.418
Smoke	0.52	-0.066	-0.156
Toys	0.953	0.635	-0.773
Fun	1.087	0.663	-0.972
Books	0.78	0.214	-0.509
Hshld	0.673	-0.017	-0.445
Clths	0.907	0.544	-0.704
MedEq	0.714	0.409	-0.925
Drugs	0.702	-0.119	-0.789
Chems	0.929	0.133	-0.448
Rubbr	0.852	0.646	-0.662
Txtls	0.838	0.694	-0.339
BldMt	0.905	0.296	-0.524
Cnstr	1.082	0.725	-0.588
Steel	1.282	0.919	-1.011
FabPr	0.93	0.984	-0.626
Mach	1.151	0.799	-0.975
ElcEq	1.148	0.246	-1.025
Autos	1.059	0.32	-0.253
Aero	0.946	0.163	-0.359
Ships	0.715	0.415	-0.343
Guns	0.454	0.143	0.003
Gold	0.612	0.74	-0.147
Mines	0.957	0.612	-0.391
Coal	0.971	0.55	-0.44
Comps	1.278	0.892	-1.869
Chips	1.367	1.031	-1.846
LabEq	1.223	1.035	-1.448
Paper	0.797	0.125	-0.368
Boxes	0.898	0.142	-0.523
Rtail	0.862	0.357	-0.81
Meals	0.805	0.242	-0.496
Banks	0.836	0.019	-0.216
Panel B : Summary statistics of mean betas			
Average	0.883	0.411	-0.622
Standard deviation	0.223	0.353	0.430
t-Tests : H_0 : Average of all industries =0	23.387	6.885	-8.562

Annexe C

Chapter 3

C.1 Optimality proofs

Proof of proposition 3

We expand the functional Φ around F_0 to determine the local utility function relative to Φ at F_0. Let F be a differential shift of F_0. We have

$$\begin{aligned}\Phi(F) - \Phi(F_0) &= -\int h\left(1 - F\left(t\right)\right) dt + \int h\left(1 - F_0\left(t\right)\right) dt \quad \text{(C.1)}\\ &= -\int \left[h\left(1 - F\left(t\right)\right) - h\left(1 - F_0\left(t\right)\right)\right] dt. \quad \text{(C.2)}\end{aligned}$$

Using the expansion :

$$\begin{aligned}h\left(1 - F\left(t\right)\right) - h\left(1 - F_0\left(t\right)\right) &= h'\left(1 - F_0\left(t\right)\right)\left[F_0\left(t\right) - F\left(t\right)\right]\\ &+ \frac{1}{2} f''\left(1 - F_0\left(t\right)\right)\left[F_0\left(t\right) - F\left(t\right)\right]^2 + o\left(\|F_0 - F\|^2\right).\end{aligned}$$

We have that

$$\begin{aligned}\Phi(F_0) - \Phi(F) &= \int h'\left(1 - F_0\left(t\right)\right)\left[F_0\left(t\right) - F\left(t\right)\right] dt\\ &+ \frac{1}{2}\int h''\left(1 - F_0\left(t\right)\right)\left[F_0\left(t\right) - F\left(t\right)\right]^2 dt + o\left(\|F - F_0\|^2\right).\end{aligned}$$

The first term can be rewritten by integration by part :

$$\begin{aligned}\int h'\left(1-F_0\left(t\right)\right)\left[F_0\left(t\right)-F\left(t\right)\right]dt &= g(t)\left[F_0\left(t\right)-F\left(t\right)\right]\Big|_0^1 \\ &- \int_0^1 g(t)\left[F_0'\left(t\right)-F'\left(t\right)\right]dt \\ &= \int_0^1 g(t)\left[dF\left(t\right)-dF_0\left(t\right)\right]\end{aligned}$$

with $g'(t)=h'\left(1-F_0\left(t\right)\right)$. Since

$$\lim_{\|F-F_0\|\to 0}\frac{\frac{1}{2}\int_0^1 h''\left(1-F_0\left(t\right)\right)\left[F_0\left(t\right)-F\left(t\right)\right]^2 dt+o\left(\|F-F_0\|^2\right)}{\|F-F_0\|}=0$$

we have that :

$$\Phi(F)-\Phi(F_0)=-\int_0^1 g\left(t\right)\left[dF\left(t\right)-dF_0\left(t\right)\right]+o\left(\|F-F_0\|\right),$$

So that by definition :

$$U_\Psi(x;F_0)=-g(t)=-\int_{-\infty}^{x} h'\left(1-F_0\left(t\right)\right)dt. \tag{C.3}$$

Proof of proposition 4

Here we want to show that the local utility of CVaR preference at x is given by :

$$U(x;F_X)=\frac{VaR_p(X)-x}{p}1_{x>VaR_p(X)}.$$

Using the fact that $h'\left(\alpha,p\right)=\frac{1}{p}1_{[0,p]}(\alpha)$, the expression for the utility in the CVaR case is given by :

$$\begin{aligned} U(x;F) &= -\frac{1}{p}\int_{-\infty}^{x} 1_{[0,p]}\left(1-F\left(t\right)\right)dt \\ &= -\frac{1}{p}\int_{-\infty}^{x} 1_{[Q_X(1-p);+\infty[}\left(t\right)dt \\ &= -\frac{1}{p}\int_{Q_X(1-p)}^{x} dt \\ &= -\frac{x-VaR_p(X)}{p}1_{x>VaR_p(X)}. \end{aligned}$$

We can now compute the $CVaR$ premium.

On the one hand :

$$U(EX+\pi;F) = \frac{EX+\pi-VaR_p(X)}{p}1_{x>Q_X(1-p)}(EX+\pi).$$

On the other hand :

$$\begin{aligned} EU(X;F) &= E\frac{X-VaR_p(X)}{p}1_{x>VaR_p(X)} \\ &= \frac{1}{p}EX1_{X>VaR_p(X)} - VaR_p(X) \\ &= CVaR_p(X) - VaR_p(X). \end{aligned}$$

π is then the solution to :

$$\frac{EX+\pi-VaR_p(X)}{p} = CVaR_p(X) - VaR_p(X)$$

and is given by :

$$\pi(p) = [pCVaR_p(X) + (1-p)VaR_p(X)] - EX$$

$$\frac{\partial}{\partial p}\pi(p) = CVaR_p(X) - VaR_p(X) + p\frac{\partial}{\partial p}CVaR_p(X) + (1-p)\frac{\partial}{\partial p}VaR_p(X).$$

From Gourieroux and Lui (2006), p.11, we have :

$$p\frac{\partial}{\partial p}CVaR_p(X) = VaR_p(X) - CVaR_p(X).$$

Thus

$$\frac{\partial}{\partial p}\pi(p) = (1-p)\frac{\partial}{\partial p}VaR_p(X) \leq 0.$$

Proof of lemma 1

The function ψ is deterministic, then showing lemma 1, is equivalent to show that the series of random element $(x_{(t)})_{t=1,\cdots,T}$ in $\sum_{t=1}^{t_l(x)} \psi(x_{(t)})\Delta x_{(t)}$, is such that $\Delta x_{(t)} \to 0$ for all t, and that $x_{t_l(x)} \to x$ *a.s when* $T \to \infty$.

This condition is satisfied since as T becomes large, $x_{(t)}$ and $x_{(t+1)}$ become arbitrarily close . This comes from the fact that $x_{(t)}$ is the empirical quantile function of X evaluated at $\frac{t}{T}$. Thus, when $T \to \infty$, $\frac{1}{T} \to 0$ and $x_{(t)} = \hat{Q}_X\left(\frac{t}{T}\right)$ and $x_{(t+1)} = \hat{Q}_X(\frac{t}{T} + \frac{1}{T})$ become arbitrarily close. In addition, the Glivenko-Cantelli theorem leads to $x_{t_l(x)} \to x$ *a.s* when $T \to \infty$.

Proof of proposition 5

the empirical quantile function of X evaluated at $\frac{t}{T}$ by $\hat{Q}_X\left(\frac{t}{T}\right)$ or $x_{(t)}$, the order statistic of order t. The proof of proposition 5 amounts to we show that :

$$\left|\sum_{t=1}^{t_l(x)} h'\left(\hat{S}_X(x_{(t)})\right)\Delta x_{(t)} - \int_{-\infty}^{x} h'\left(S_X\left(t\right)\right)dt\right| \to 0 \quad a.s$$

We have :

$$\begin{aligned} U_h(x) - \hat{U}_h(x) &= \sum_{t=1}^{t_l(x)} h'(\hat{S}_X(x_{(t)}))\Delta x_{(t)} - \sum_{t=1}^{t_l(x)} h'\left(S_X(x_{(t)})\right)\Delta x_{(t)} \\ &+ \sum_{t=1}^{t_l(x)} h'\left(S_X(x_{(t)})\right)\Delta x_{(t)} - \int_{-\infty}^{x} h'\left(S_X\left(u\right)\right)du. \end{aligned}$$

We first show that :

$$\sum_{t=1}^{t_l(x)} h'\left(\hat{S}_X(x_{(t)})\right)\Delta x_{(t)} - \sum_{t=1}^{t_l(x)} h'\left(S_X(x_{(t)})\right)\Delta x_{(t)} \to 0. \tag{C.4}$$

The expression in the left hand side of (C.4) can be written as

$$\sum_{t=1}^{t_l(x)} \left[h'\left(\hat{S}_X(x_{(t)})\right) - h'\left(S_X(x_{(t)})\right)\right]\Delta x_{(t)}. \tag{C.5}$$

Now, by the delta method

$$h'\left(\hat{S}_X(x_{(t)})\right) - h'\left(S_X(x_{(t)})\right) \approx -h''\left(S_X(x_{(t)})\right)\left(\hat{F}_X(x_{(t)}) - F_X(x_{(t)})\right).$$

Besides using Assumption B, we have

$$\left|\sum_{t=1}^{t_l(x)}\left[h'\left(\hat{S}_X(x_{(t)})\right) - h'\left(S_X(x_{(t)})\right)\right]\Delta x_{(t)}\right| \tag{C.6}$$

$$\leq C\sup_x\left|\hat{F}_X(x) - F_X(x)\right|\left|\sum_{t=1}^{t_l(x)} h'\left(S_X(x_{(t)})\right)\Delta x_{(t)}\right| \tag{C.7}$$

where C is a positive constant. By the Glivenko-Cantelli theorem and using Lemma 1 applied function $\psi = h'' \circ S_X$, the term (C.7) converge to 0 a.s. and so does (C.6).

Second, applying Lemma1 to $\psi = h' \circ S_X$ leads to :

$$\sum_{t=1}^{t_l(x)} h'\left(S_X(x_{(t)})\right)\Delta x_{(t)} \to \int_{-\infty}^{x} h'\left(S_X(u)\right) du. \tag{C.8}$$

Combining (C.4) and (C.8) establish proposition 5.

Proof of proposition 6

Proving proposition 6 is then equivalent to proving that $\sqrt{T}\ \left[U_h(x) - \hat{U}_h(x)\right] \Rightarrow \int_{-\infty}^{x} h''\left(S\left(t\right)\right) B(F(t))dt$.

We have :

$$U_h(x) - \hat{U}_h(x) \;=\; \sum_{t=1}^{t_l(x)} h'\left(\hat{S}_X(x_{(t)})\right)\Delta x_{(t)} - \sum_{t=1}^{t_l(x)} h'\left(S_X(x_{(t)})\right)\Delta x_{(t)} \;+ o_p(1).$$

In addition, we have :

$$\sqrt{T}(\sum_{t=1}^{t_l(x)} h'(\hat{S}_X(x_{(t)}))\Delta x_{(t)} - \sum_{t=1}^{t_l(x)} h'(S_X(x_{(t)}))\Delta x_{(t)} \ ) = \sqrt{T}\sum_{t=1}^{t_l(x)} h'(\hat{S}_X(x_{(t)})) - h'\left(S_X(x_{(t)})\right)\Delta x_{(t)}$$

$$= \sum_{t=1}^{t_l(x)} h''\left(S_X(x_{(t)})\right)\sqrt{T}\left(\hat{F}_X(x_{(t)}) - F_X(x_{(t)})\right)\Delta x_{(t)}$$

$$= \sum_{t=1}^{t_l(x)} h''\left(S_X(x_{(t)})\right)G(x_{(t)})\Delta x_{(t)} \quad + o_p(1)$$

$$\to \int_{-\infty}^{x} h''\left(S\left(t\right)\right)G(t)dt$$

using lemma 1.Which completes the proof of proposition 6.

Proof of proposition 7

First notice that $\bar{X}_T$, $\widehat{VaR}_p(X)$ and $\widehat{CVaR}_p(X)$ are asymptotically uncorrelated. Indeed we have :

$$Cov\left(\bar{X}_T, \widehat{VaR}_p(X)\right) = \frac{1}{T}\sum_{t=1}^{T} Cov\left(x_{(t)}, x_{[(1-p)T]}\right) = \frac{\hat{\sigma}_X^2}{T},$$

$$Cov\left(\bar{X}_T, \widehat{CVaR}_p(X)\right) = \frac{1}{T^2 p}\sum_{t=1}^{T}\sum_{[(1-p)T]}^{T} Cov\left(x_t, x_j\right) = \frac{\hat{\sigma}_X^2}{T} \ \ and$$

$$Cov\left(\widehat{VaR}_p(X), \widehat{CVaR}_p(X)\right) = \frac{1}{T}\sum_{t=[(1-p)T]}^{T} Cov\left(x_{1-p}, x_t\right) = \frac{\hat{\sigma}_X^2}{T}$$

From Gourieroux et Lui (2006), we know that $\widehat{CVaR}_p(X)$ and $\widehat{VaR}_p(X)$ are asymptotically normal

$$\sqrt{T}\left(\widehat{CVaR}_p(X) - CVaR_p(X)\right) \to N\left(0, \omega_1^2\right) \ \ and \ \ \sqrt{T}\left(\widehat{VaR}_p(X) - VaR_p(X)\right) \to N\left(0, \omega_2^2\right)$$

with ω_1^2 and ω_2^2 defined in proposition 7. We then have that

$$\sqrt{T}\left(\begin{array}{c}\widehat{CVaR}_p(X)-CVaR_p(X)\\ \widehat{VaR}_p(X)-VaR_p(X)\\ \bar{X}_T-EX\end{array}\right)\to N\left(0,\Sigma\right)$$

where the asymptotic covariance matrix Σ is diagonal and defined by :

$$\Sigma=\left[\begin{array}{ccc}\omega_1^2 & 0 & 0\\ 0 & \omega_2^2 & 0\\ 0 & 0 & \sigma_X^2\end{array}\right]$$

Now since

$$\sqrt{T}\left(\hat{\pi}\left(p\right)-\pi\left(p\right)\right)=\left(\begin{array}{ccc}p & 1-p & -1\end{array}\right)\sqrt{T}\left(\begin{array}{c}\widehat{CVaR}_p(X)-CVaR_p(X)\\ \widehat{VaR}_p(X)-VaR_p(X)\\ \bar{X}_T-EX\end{array}\right)$$

we have

$$\sqrt{T}\left(\hat{\pi}\left(p\right)-\pi\left(p\right)\right)\to N\left(0,\sigma_\pi^2\right)$$

$$\sigma_\pi^2=\left(\begin{array}{ccc}p & 1-p & -1\end{array}\right)\Sigma\left(\begin{array}{ccc}p & 1-p & -1\end{array}\right)'=p^2\omega_1^2+(1-p)^2\omega_2^2+\sigma_X^2$$

which ends the proof.