Institut des ACTUAIRES

ENSAE IP PARIS

Mémoire présenté devant l'ENSAE Paris
pour l'obtention du diplôme de la filière Actuariat
et l'admission à l'Institut des Actuaires
le 17/11/2021

Par : **Lamia Lamrani**

Titre : **Gestion de risque de portefeuilles financiers avec des outils de matrices aléatoires**

Confidentialité : ☒ NON ☐ OUI (Durée : ☐ 1 an ☐ 2 ans)

Les signataires s'engagent à respecter la confidentialité indiquée ci-dessus

Membres présents du jury de la filière

Nom : Pr. Christian-Yann Robert

Membres présents du jury de l'Institut des Actuaires

Entreprise : Chaire d'Econophysique et de Systèmes Complexes, École Polytechnique et Capital Fund Management

Signature :

Directeur du mémoire en entreprise :

Nom : Marc Potters
Signature :

Fondation du Risque
c/o Institut Louis Bachelier
Palais Brongniart
28, Place de la Bourse
75002 PARIS

Autorisation de publication et de mise en ligne sur un site de diffusion de documents actuariels
(après expiration de l'éventuel délai de confidentialité)

Signature du responsable entreprise

Secrétariat :

Signature du candidat

Bibliothèque :

Ecole Nationale de la Statistique et de l'Administration Economique (ENSAE)
5, avenue Henry Le Chatelier - 91120 PALAISEAU, FRANCE

Résumé

Dans le cadre de l'approche moyenne/variance de Markowitz, l'estimation de la matrice de variance-covariance des rendements des actifs est cruciale pour l'investisseur qui souhaite sélectionner un portefeuille ou pour le gestionnaire d'actifs qui veut contrôler les risques de son portefeuille. La diversification des actifs fait que le nombre d'actifs présents dans un portefeuille est grand (de l'ordre de plusieurs centaines). Par exemple pour estimer la matrice de variance-covariance d'un portefeuille contenant 500 actions il faudrait de l'ordre de quatre à cinq ans de données quotidiennes. On cherche cependant à privilégier les données récentes dans la mesure où les séries financières ne sont pas stationnaires. Dans le cas où le nombre d'observations n'est pas très grand par rapport à la taille de l'objet observé, l'estimateur empirique a une efficacité limitée. Il entraîne même une sous-estimation du risque de portefeuille lorsqu'il est combiné à la théorie de Markowitz.

Le but de ce travail est donc d'étudier différents estimateurs des matrices de variance-covariance et de comparer leurs efficacités. On étudiera en particulier deux estimateurs provenant de la théorie des matrices aléatoires. Dans ce cadre, chaque cellule de la matrice à estimer est une variable aléatoire qui représente la covariance entre un couple de rendements de deux actifs financiers. On étudiera l'estimateur de Ledoit-Péché pour les matrices de variance-covariance dites "invariantes par rotation" et un estimateur récent, développé pour les matrices de crosscovariance, qui permet d'étudier les corrélations entre différentes classes d'actifs (ex : des actions et des futures). Nous étudierons également un estimateur obtenu par crossvalidation, une technique courante en machine learning. Ces estimateurs seront testés sur des données simulées puis sur des données financières françaises et américaines. Nous chercherons à montrer leur intérêt pour la gestion des risques de grands portefeuilles financiers.

Mots clés : Matrices Aléatoires, Estimation de variance-covariance, Finance, Gestion de risque.

Abstract

Within the framework of Markowitz mean/variance optimization, the estimation of the covariance matrix of the returns of the assets is of paramount importance for the investor who wants to select a portfolio or for the asset manager who wishes to control the risks of his portfolio. Investors who rely on diversification tend to hold a large number of different assets, typically hundreds of assets. For example to estimate a covariance matrix of size 500, one needs around four or five years of daily financial data. Also to have a more realistic view of the market behavior, investors tend to consider recent data, therefore the amount of available data to estimate the covariance matrix is limited : although bigger than the size of the object to be estimated it is not much bigger. In this case it is well known that the empirical estimator has limited efficiency especially when combined with Markowitz theory.

The goal of this work is to study different estimators and to compare their efficiency. We will focus in particular on two estimators coming from random matrix theory. In this framework, each entry of the matrix to estimate is seen as a random variable which represents the covariance between two financial assets. We will study Ledoit-Péché estimator for covariance matrices which are "rotationally invariant" and a recently developed estimator for crosscovariance matrices which enables to study the correlations between different types of assets(eg : stocks, bonds, futures...). Estimation by crossvalidation, a common technique of machine learning will also be considered. These estimators will be tested on simulated data and then on French and American financial data. We will try to show the interest and efficiency of these estimators for risk management of large financial portfolios.

Key words : Random Matrix, Covariance Estimation, Finance, Risk Management.

Remerciements

Je tiens tout d'abord à remercier sincèrement mon tuteur M. Marc Potters, Chief Investment Officer du Capital Fund Management, qui a supervisé mes six mois de stage. Je le remercie pour les discussions hebdomadaires, ses conseils, ses explications, les nombreuses pistes suggérées et pour sa grande disponibilité. Je pense avoir énormément appris sous sa direction.

Je remercie également M. Pierre Mergny, doctorant en matrices aléatoires de la Chaire d'Econophysique et de l'Université Paris-Saclay, pour ses conseils et son aide précieuse.

J'adresse également mes remerciements au Pr. Laure Giovangigli, Enseignant-Chercheur à l'ENSTA Paris, et au Pr. Christian-Yann Robert, Professeur à l'ENSAE Paris, pour leurs conseils et suivi tout au long du stage.

Je remercie le Pr. Michael Benzaquen, Directeur de la Chaire d'Econophysique, pour son aide dans ma recherche de stage. Je remercie également les membres de la Chaire d'Econophysique et le Capital Fund Management pour leur accueil.

Enfin, je remercie tout particulièrement le Pr. Benoît Collins, Professeur à l'Université de Kyoto, qui m'a le premier transmis le goût des matrices aléatoires en supervisant mon stage de recherche de Master 1.

Table des matières

1 Introduction

Une question importante en finance est la construction d'un portefeuille d'investissement optimal à partir d'un ensemble d'actifs (actions, obligations, matières premières, devises...). Harry Markowitz a proposé une réponse avec une approche moyenne/variance [15]. Pour construire son portefeuille optimal, l'investisseur a le choix entre maximiser son rendement espéré sous une contrainte de risque ou de manière équivalente il peut minimiser son risque sous une contrainte de rendement espéré minimal. Le problème est quadratique et la solution peut facilement être calculée. Toutefois, la théorie de Markowitz suppose la connaissance de la matrice de variance-covariance des rendements des actifs et du vecteur des rendements espérés de chaque actif. L'estimation du vecteur de l'espérance des rendements est une question de prédiction : l'investisseur va essayer de prévoir au mieux les rendements futurs en fonction de l'information dont il dispose ; tandis que l'estimation de la matrice de variance-covariance est une question de risque. Dans ce rapport, nous allons nous intéresser à la problématique de l'estimation des matrices de variance-covariance qui est fondamentale pour la gestion des risques mais aussi pour la sélection d'un portefeuille financier où il faut souvent gérer plusieurs centaines d'actifs simultanément et dans un cadre dynamique.

Une première approche pour estimer la matrice de variance-covariance pourrait être de partir des observations des rendements et d'en tirer une matrice de variance-covariance empirique. Toutefois il est connu en statistique que l'estimateur empirique est d'efficacité moindre quand le nombre d'observations n'est pas très grand par rapport à la taille de l'objet que l'on cherche à estimer. En pratique, il est rare d'avoir accès à de tels jeux de données et il est parfois peu pertinent de considérer des données trop anciennes. Si l'on considère l'exemple financier, pour estimer une matrice empirique de variance-covariance efficacement pour un portefeuille avec quelques centaines d'actifs, il faudrait des données quotidiennes sur quatre ou cinq ans. Or on sait que les séries financières ne sont pas stationnaires donc il faut privilégier l'usage de données récentes même si moins nombreuses. Pour toutes ces raisons, l'usage de l'estimateur empirique est d'une efficacité limitée et il est nécessaire de développer d'autres techniques pour estimer les covariances [11] [16].

Nous allons en particulier nous intéresser à la classe des estimateurs dits "invariants par rotation" qui ont été dérivés de la théorie des matrices aléatoires. On voit la matrice de variance-covariance comme une matrice aléatoire où chaque cellule est une variable aléatoire représentant la covariance entre deux séries temporelles (par exemple des rendements d'actifs financiers). Les matrices aléatoires ont été introduites pour la première fois par John Wishart en 1928 qui a étudié des matrices aléatoires gaussiennes [23]. C'est toutefois dans les années 1950 que les matrices aléatoires vont prendre un réel essor avec les travaux de Eugene Wigner sur les noyaux d'atomes lourds comme par exemple l'uranium [22]. Dans de tels atomes, les interactions entre particules sont très corrélées et complexes, il est donc presque impossible d'écrire analytiquement l'opérateur Hamiltonien du système. Wigner a eu l'idée de remplacer le Hamiltonien par une matrice aléatoire et d'en étudier les valeurs propres qui représentent les niveaux d'énergie. Cela a permis une meilleure compréhension des systèmes complexes.

Aujourd'hui les matrices aléatoires forment un domaine de recherche dynamique avec de nombreuses applications : physique quantique, théorie de l'information, mathématiques financières, etc. Par exemple, en théorie des nombres, il a été établi que les zéros de la fonction zêta de Riemann (intiment liée aux nombres premiers) correspondent aux valeurs propres d'un certain opérateur matriciel aléatoire [9]. Ici nous nous intéresserons particuliè-

rement aux applications financières pour l'estimation de covariance.

On considèrera l'estimateur de Ledoit et Péché [10] pour les matrices de variance-covariance. L'idée est de considérer que nous ne savons rien sur les vecteurs propres et donc de travailler uniquement sur l'estimation des valeurs propres, on passe alors d'un problème d'estimation de taille n^2 à un problème de taille n. Le spectre empirique étant plus large que celui de la "vraie matrice" de covariance, on va principalement multiplier les valeurs propres empiriques par des coefficients souvent inférieurs à 1. Il s'agit d'un shrinkage non linéaire. Nous regarderons aussi un estimateur récemment développé pour les matrices de crosscovariance [2]. Il s'agit d'une extension de l'estimateur de Ledoit et Péché aux matrices rectangulaires, par exemple dans le cas où l'on souhaiterait étudier les corrélations entre des rendements d'actions et de futures.

En parallèle, nous regarderons l'estimation par crossvalidation, une technique courante en machine learning [20][17]. La technique consiste à découper en blocs l'échantillon puis se fixer un ensemble où l'on optimise (l'échantillon privé du bloc), aussi appelé ensemble d'optimisation puis un ensemble sur lequel on valide le modèle (le bloc). Ensuite on moyenne sur les différents blocs. L'intérêt de cette technique est qu'elle ne nécessite quasiment aucune hypothèse sur les données de départ. À notre connaissance, la question du nombre de blocs optimal pour la crossvalidation est encore ouverte pour l'estimation des matrices de covariance.

Nous allons comparer ces techniques d'abord sur des données simulées, ensuite nous mènerons des études sur des données financières françaises puis sur des données américaines du S&P500. Le cœur de ce projet est donc l'estimation des matrices de variance-covariance pour mieux appréhender les corrélations entre actifs financiers et pour mieux estimer la variance d'un portefeuille financier. Une application importante de l'estimation de matrices de variance-covariance est le problème d'allocation optimale de Markowitz que l'on présentera également [15].

2 Formulation du problème et démarche

Nous allons commencer par un rappel de l'approche moyenne/variance de Markowitz. Cela permettra d'illustrer le champ d'application de l'estimation de covariance avant d'aborder le problème de l'estimation. Tout au long du rapport les noms de matrices seront en gras pour les différentier des vecteurs.

2.1 Théorie du portefeuille de Markowitz

Soit un portefeuille contenant n actifs financiers et soit $\pi \in \mathbb{R}^n$ le vecteur des positions prises sur chaque actif (la vente à découvert est autorisée). On considère le cas où tous les actifs sont risqués. L'investisseur souhaite alors déterminer les quantités optimales à investir dans chaque actif. Par portefeuille optimal on entend un portefeuille qui maximise le rendement moyen espéré sous une contrainte de risque ou de manière équivalente un portefeuille qui minimise le risque sous une contrainte de rendement espéré. Ce problème a été formalisé et résolu par Harry Markowitz en 1952 [15].

Commençons par définir quelques notations. Pour alléger le rapport, on écrira souvent "matrice de covariance" à la place de "matrice de variance-covariance".

Soient $\boldsymbol{C} \in \mathbb{R}^{n\times n}$ la matrice de covariance des rendements et $g \in \mathbb{R}^n$ le vecteur des espérances de rendement pour chaque actif. Ces deux objets sont considérés comme connus. Un portefeuille π est un vecteur à composantes réelles dont la somme des éléments est égale à 1. Soient α le niveau de risque maximal toléré par l'investisseur et $\mathcal{G}$ l'espérance minimale de rendement que souhaite l'investisseur.

Le risque du portefeuille est alors :

$$Risque(\pi) = \pi^T \boldsymbol{C} \pi$$

Et le rendement du portefeuille est :

$$Rendement(\pi) = \sum_{i=1}^{n} \pi_i g_i$$

Le problème d'optimisation a alors deux formulations équivalentes [17] [4] :

$$\text{Primal} \begin{cases} \max\limits_{\pi\in\mathbb{R}^n} \pi^T g \\ \\ \text{s.t } \pi^T \boldsymbol{C} \pi \leq \alpha \\ \quad \sum_{i=1}^n \pi_i = 1 \end{cases} \qquad \text{Dual} \begin{cases} \min\limits_{\pi\in\mathbb{R}^n} \frac{1}{2}\pi^T \boldsymbol{C} \pi \\ \\ \text{s.t } \pi^T g \geq \mathcal{G} \\ \quad \sum_{i=1}^n \pi_i = 1 \end{cases}$$

Considérons par exemple la forme duale. Elle se résout simplement grâce à un Lagrangien :

$$\mathcal{L}(\pi, \gamma) = \frac{1}{2}\pi^T \boldsymbol{C} \pi - \gamma \pi^T g$$

$$\frac{\partial \mathcal{L}}{\partial \pi} = \pi^T \boldsymbol{C} - \gamma g^T$$

D'où :

$$\frac{\partial \mathcal{L}}{\partial \pi} = 0 \Longleftrightarrow \pi_{opt} = \gamma \boldsymbol{C}^{-1} g$$

La valeur de γ telle que $\pi^T g = \mathcal{G}$ est

$$\gamma = \frac{\mathcal{G}}{g^T \boldsymbol{C}^{-1} g}$$

On a alors l'expression suivante pour le portefeuille optimal [17] [4] :

$$\pi_{opt} = \mathcal{G} \frac{g^T \boldsymbol{C^{-1}}}{g^T \boldsymbol{C^{-1}} g}$$

Comme nous l'avons évoqué dans l'introduction, la théorie de Markowitz offre un cadre quantitatif simple d'utilisation sous réserve de connaître $\boldsymbol{C}$ et g.

2.2 Problème d'estimation et estimateur empirique

Problème d'estimation de covariance

Soit $X \in \mathbb{R}^n$ un vecteur aléatoire observé au fil du temps t pour $t = 1, 2, .., T$. Dans le cas financier, $X(t)$ représentera le vecteur des rendements au temps t des actifs. On définit la matrice **X** de taille $n \times T$:

$$\mathbf{X} = (X(1), X(2), X(3), ..., X(T))$$

Nous supposons que les $X(1), ..., X(n)$ sont des observations du vecteur X indépendantes. On suppose que nos données sont centrées quitte à remplacer $X(t)$ par :

$$X(t) - \overline{X} \text{ avec } \overline{X} = \frac{1}{T}\sum_{t=1}^{T} X(t)$$

Notre question est alors la suivante :

Comment estimer la matrice de covariance des $(X(t))$ à partir des observations dont on dispose ?

(Problème de covariance)

Comme nos données sont supposées centrées, on peut négliger les produits d'espérances et la matrice de covariance empirique de X est alors :

$$\boldsymbol{C_{emp}} = \frac{1}{T}\boldsymbol{X}\boldsymbol{X^T}$$

Cet estimateur converge vers la vraie matrice de covariance pour $T \to \infty$, il est efficace en pratique si la taille de l'objet qu'on estime est négligeable devant le nombre d'observations dont on dispose.

Dans le cas où la taille de la matrice de covariance n'est pas négligeable devant le nombre d'observations T, l'estimateur empirique a une efficacité limitée. Il est également largement inefficace pour des données financières lorsque combiné à l'optimisation de Markowitz qui fait appel à l'inverse de la matrice de covariance [16] [11].

Dans le cas où X est simulé par ordinateur à l'aide d'une loi multivariée, il existe une "vraie matrice" de covariance $\boldsymbol{C}$ que l'on cherche à estimer. S'il s'agit de données financières, il n'existe probablement pas de vraie matrice de covariance mais on essaie au mieux de l'estimer sur une période et ensuite de voir si l'estimateur reste efficace en dehors de l'échantillon de temps. On cherche donc à estimer une matrice de variance-covariance à partir de l'estimateur empirique.

Problème d'estimation de crosscovariance

Nous allons maintenant présenter le problème de l'estimation de la crosscovariance. Ce problème semble au début pouvoir se résumer par un problème d'estimation de matrice de covariance.

Soient $X \in \mathbb{R}^n$ et $Y \in \mathbb{R}^p$ deux vecteurs aléatoires. Nous disposons de T observations de X supposées indépendantes et T observations de Y également indépendantes :

$$\mathbf{X} = (X(1), X(2), X(3), ..., X(T)) \text{ de taille } n \times T$$
$$\mathbf{Y} = (Y(1), Y(2), Y(3), ..., Y(T)) \text{ de taille } p \times T$$

Tout comme le problème de covariance, les données sont supposées indépendantes et centrées quitte à remplacer $X(t)$ et $Y(t)$ par :

$$X(t) - \overline{X} \text{ avec } \overline{X} = \frac{1}{T}\sum_{t=1}^{T} X(t)$$

$$Y(t) - \overline{Y} \text{ avec } \overline{Y} = \frac{1}{T}\sum_{t=1}^{T} Y(t)$$

Nous cherchons à répondre à la question suivante :

Comment estimer la crosscovariance entre les vecteurs aléatoires X et Y à partir des observations dont on dispose ?

Une première approche est de considérer que la crosscovariance est un sous-problème de l'estimation de covariance, on peut définir :

$$Z = \begin{pmatrix} X \\ Y \end{pmatrix}, \text{ un vecteur aléatoire}$$

Ensuite on estime la matrice de covariance associée aux données $\boldsymbol{Z} = (Z(1), ..., Z(T))$, on appelle $\boldsymbol{\Sigma}$ la vraie matrice de variance-covariance de Z que l'on ne connaît pas, cette matrice est de la forme :

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{C} \\ \boldsymbol{C}^T & \boldsymbol{B} \end{pmatrix}$$

où **A** est de taille $n \times n$ et est la matrice de variance-covariance de X, **B** est de taille $p \times p$ et est la matrice de variance-covariance de Y et **C** de taille $n \times p$ est la matrice de crosscovariance de X et Y.

Nous cherchons alors à estimer $\boldsymbol{C}$ dont l'estimateur empirique est :

$$\boldsymbol{C_{XY}} = \frac{1}{T}\boldsymbol{X}\boldsymbol{Y}^T$$

Ici aussi nous pouvons négliger les produits d'espérances qui apparaissent dans la covariance parce que les données sont supposées centrées.

Il se trouve que l'approche covariance devient moins efficace quand les données que l'on regarde sont de natures différentes (par exemple des actions et des températures) [2]. Il existe donc des cas où il faut développer une approche spécifique du problème de crosscovariance.

2.3 Portefeuille de Markowitz et estimateur empirique

Nous voudrions illustrer sur un exemple l'erreur d'estimation de risque de portefeuille que peut entraîner l'usage de l'estimateur empirique pour le problème d'allocation optimale de Markowitz.
Pour cela on considère $X \in \mathbb{R}^n$ un vecteur aléatoire de loi normale centrée et de matrice de variance-covariance $\mathbf{\Sigma}$. À partir de là on génère T observations de X indépendantes :$X(1), ..., X(T)$ et on pose la matrice des observations :

$$\boldsymbol{X} = (X(1), ..., X(T)) \in \mathbb{R}^{n \times T}$$

L'estimateur empirique de la matrice de variance-covariance des données est alors :

$$\boldsymbol{E} = \frac{1}{T} \boldsymbol{X} \boldsymbol{X^T}$$

On tire de manière aléatoire un vecteur $g \in \mathbb{R}^n$ qui représentera l'espérance des rendements pour chaque actif et nous nous sommes fixés un rendement minimal de portefeuille espéré $\mathcal{G} = 10\%$.

On peut alors calculer les portefeuilles suivants à l'aide de la théorie de Markowitz :

- le portefeuille empirique :
$$\pi_{emp} = \mathcal{G} \frac{g^T \boldsymbol{E^{-1}}}{g^T \boldsymbol{E^{-1}} g}$$
- le "vrai" portefeuille optimal :
$$\pi_{opt} = \mathcal{G} \frac{g^T \mathbf{\Sigma^{-1}}}{g^T \mathbf{\Sigma^{-1}} g}$$

On va considérer les risques associés à ces portefeuilles :

- le risque du portefeuille empirique :
$$R_{emp} = \pi_{emp}^T \mathbf{\Sigma} \pi_{emp}$$
- le risque du portefeuille empirique estimé avec la matrice empirique $\boldsymbol{E}$, il s'agit du portefeuille de risque minimal avec un rendement de $\mathcal{G}$ dans l'échantillon, on l'appelle le risque in-sample [4]
$$R_{in} = \pi_{emp}^T \boldsymbol{E} \pi_{emp}$$
- le vraie risque du portefeuille optimal de Markowitz :
$$R_{opt} = \pi_{opt}^T \mathbf{\Sigma} \pi_{opt}$$

On va maintenant calculer le risque de ces portefeuilles pour plusieurs couple de paramètres (n, T) où n est la taille du portefeuille et T est le nombre d'observations, nous fixons le ratio taille/observations $q = \frac{n}{T} = 0.8$. Pour chaque couple (n, T) nous effectuons 50 simulations et nous en retirons un risque moyen pour chaque type de portefeuille :

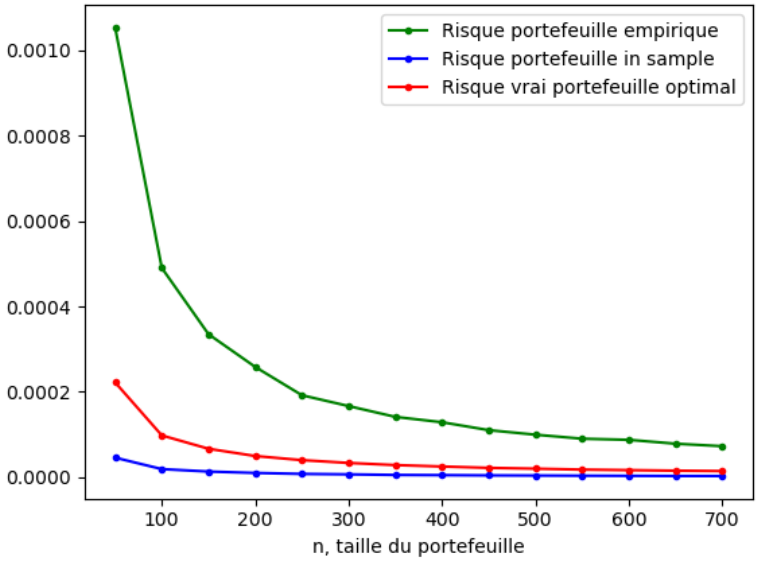

FIGURE 1 – Risque du portefeuille empirique (en vert), risque du portefeuille optimal (en bleu) et risque in-sample(en rouge) moyens pour différentes valeurs de (n, T) avec $q = \frac{n}{T} = 0.8$

On remarque que l'usage de l'estimateur empirique pour la sélection du portefeuille optimal de Markowitz entraîne une sous-estimation de risque du portefeuille. Le risque in-sample $\pi_e^T \boldsymbol{E} \pi_e$ est même plus petit que le risque véritable du portefeuille optimal de Markowitz (en rouge sur le graphe) que l'on aurait construit en connaissant la vraie matrice de variance-covariance des données $\boldsymbol{\Sigma}$. Le risque réel du portefeuille construit avec l'estimateur empirique (en vert) est bien supérieur au risque du portefeuille construit avec la connaissance de $\boldsymbol{\Sigma}$ (en bleu).

Dans la théorie de Markowitz, chaque vecteur propre de la matrice de variance-covariance représente un portefeuille de norme 1 (au sens de la norme L^2) avec pour risque la valeur propre associée. Le portefeuille optimal est alors une combinaison linéaire de ces vecteurs propres.

Nous allons regarder sur un exemple la répartition des proportions investies sur les valeurs propres empiriques pour la construction du portefeuille empirique (à partir de la matrice de variance-covariance $\boldsymbol{E}$) et nous allons la comparer à la répartition des proportions investies sur les valeurs propres de la vraie variance-covariance $\boldsymbol{\Sigma}$ pour la construction du véritable portefeuille optimal :

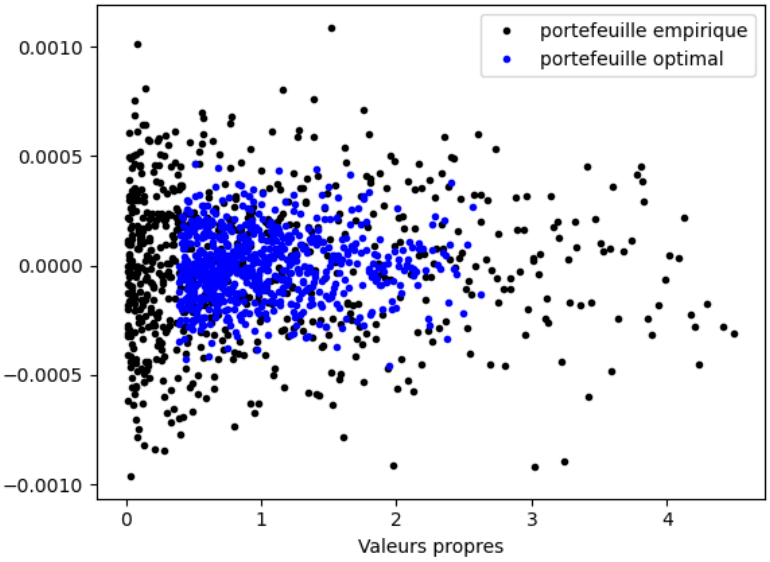

FIGURE 2 – Répartition des proportions investies par le portefeuille empirique sur les différentes valeurs propres empiriques et des proportions investies par le portefeuille optimal sur les valeurs propres de la vraie matrice de variance-covariance, $n = 700$, $q = 0.8$

Une première remarque est que le spectre empirique est beaucoup plus large que le spectre de la vraie variance-covariance des données. Le portefeuille empirique met beaucoup de poids sur les portefeuilles associés aux valeurs propres les moins risquées alors qu'elles sont plus petites que les valeurs propres minimales de la véritable matrice de variance-covariance. Le portefeuille empirique met aussi moins de poids sur les portefeuilles liés aux grandes valeurs propres qu'il surestime par rapport aux grandes valeurs propres de la vrai matrice $\mathbf{\Sigma}$. Ainsi le portefeuille construit avec l'estimateur empirique va nécessairement être plus risqué que le portefeuille construit avec la connaissance de $\mathbf{\Sigma}$.

L'objectif de ce projet va être d'étudier des techniques permettant de réduire la largeur du spectre empirique afin de se rapprocher de la vraie matrice de variance-covariance des données, d'avoir une meilleure estimation des risques et de pouvoir mener des allocations de Markowitz plus efficaces.

3 Présentation des données et de leurs propriétés

L'objectif de ce rapport est d'étudier différents estimateurs qui permettent de corriger une partie de l'erreur due à l'utilisation de l'estimateur empirique pour estimer la variance des portefeuilles et pour le problème d'allocation optimale de Markowitz. À terme, nous souhaiterions appliquer ces différents estimateurs à des données financières réelles. Nous les appliquerons aux actions du SBF120 (France) et aux actions du S&P500 (États-Unis) pour l'estimation de risques. Nous souhaiterions donc commencer par un rappel des différents risques financiers puis par une présentation de quelques caractéristiques des données financières.

3.1 Les risques financiers

La gestion des risques est devenue un enjeu capital pour les banques, les assurances et tous les autres acteurs gérant des portefeuilles d'actifs financiers. Il existe différents risques financiers, nous allons en rappeler les plus importants afin de spécifier un peu plus le cadre de notre étude [7].

Risque de marché

Il s'agit du risque de perte de la valeur d'un portefeuille financier à cause des mouvements des indices (S&P500, SBF120, FTSE,...), des devises, matières premières ou encore des taux d'emprunt. En général les acteurs "choisissent" le ou les risques auxquels ils souhaitent être exposés. Par exemple un desk de trading spécialisé dans les options sera peu atteint par les mouvements des marchés d'actions mais aura beaucoup d'exposition au risque de volatilité qui constitue le cœur de leur expertise.

Risque de Liquidité

Pour un investisseur, il peut s'agir du risque de ne pas trouver un acheteur pour ses actifs même à un prix inférieur à celui du marché. Si l'investisseur dispose d'une position de grande taille et que les volumes échangés sont bas voire inexistants, la cession de ses placements peut même se révéler impossible. Des crises de liquidité entraîne en général de grands bid-ask spreads sur les marchés.

Risque Opérationnel

Il s'agit du risque lié aux catastrophes naturelles, aux erreurs humaines ou techniques ou à la fraude. Il est difficile de couvrir de tels risques sur les marchés même s'il existe quelques actifs pour cela comme par exemple les *catastrophe bonds*(CAT). Il y a sinon également la possibilité de faire appel à un assureur pour se couvrir contre certains risques naturels.

Risque de Crédit

Il s'agit du risque de défaut de la contrepartie ou du risque qu'elle paie sa créance en dehors des délais.

Risque de business

Il s'agit du risque de changements dans des variables du business plan qui mettraient en danger sa viabilité (ex : un produit devenu désuet à cause d'une innovation chez un concurrent...).

Dans ce rapport nous nous focaliserons donc sur le risque de marché et en particulier sur les mouvements des rendements des actifs financiers puisque nous cherchons à estimer au mieux les corrélations entre actifs.

3.2 Actifs financiers et faits stylisés

Certaines propriétés générales se retrouvent sur la plupart des actifs financiers, elles sont appelées faits stylisés. Tout d'abord posons quelques notations (on considère l'absence de dividendes) :

- S_t : le prix d'un actif à un instant t ($t \in \mathbb{N}$), on peut par exemple considérer qu'entre t et $t+1$ s'écoule une journée de cotation.
- Le log-rendement :
$$R_{t+1} = ln(S_{t+1}) - ln(S_t)$$
- Le rendement arithmétique :
$$r_{t+1} = \frac{S_{t+1} - S_t}{S_t}$$

Le log-rendement a l'avantage de pouvoir se calculer très simplement sur une période donnée, par exemple sur la période allant de $t+1$ à $t+K$:

$$R_{t+1:t+K} = ln(S_{t+K}) - ln(S_t) = \sum_{k=1}^{K} ln(S_{t+k}) - ln(S_{t+k-1}) = \sum_{k=1}^{K} R_{t+k}$$

Toutefois on peut considérer les deux rendements comme équivalents sous l'hypothèse que les variations de prix sont petites à court terme (on note par $\sim$ l'équivalence) :

$$R_{t+1} = ln(1 + \frac{S_{t+1} - S_t}{S_t}) \sim \frac{S_{t+1} - S_t}{S_t} \sim r_{t+1}$$

Nous allons faire une série de remarques sur les propriétés des actifs financiers qui nous aiderons à orienter nos simulations et à interpréter celles sur des données financières [6] [7].

Absence d'auto-corrélation à court terme (c'est-à-dire de 1 jour à 3 mois)

Les rendements quotidiens sont très peu corrélés :

$$Corr(R_{t+1}, R_{t+1-\tau}) = 0, \text{ avec } \tau = 1, 2, 3, ..., 100$$

Il paraît donc difficile de prévoir un rendement à partir de son propre passé.

Queues de distribution des rendements quotidiens

Les rendements quotidiens ont des queues plus épaisses que la loi normale, il y a donc une probabilité de perte plus grande que si l'hypothèse normale était vérifiée.

Asymétries de la distribution des rendements

Il arrive que des actifs subissent de grandes baisses de prix mais les mouvements à la hausse ne sont pas aussi importants.

Corrélations entre actifs

Les corrélations semblent dépendre du temps. Les corrélations entre actions sont souvent positives et ont tendance à être plus grandes dans des marchés à fortes volatilité (ex : lors d'une crise financière).

Clusters de volatilité

On peut observer une auto-corrélation positive sur plusieurs jours avec différentes mesures de volatilité. Cela signifie que les événements à haute-volatilité ont tendance à former des clusters dans le temps. Autrement-dit si une action subit de grandes variations de prix sur une journée, il y a de fortes chances qu'il en soit encore de même le lendemain.

Il existe bien sûr une multitude d'autres faits stylisés sur les actifs financiers mais ceux-ci nous paraissent suffisants pour notre étude. Nous allons nous intéresser aux matrices de variance-covariance financières à partir de données quotidiennes et souvent sur une période d'observation supérieure à un an.

4 Théorie des matrices aléatoires

Fondamentalement une matrice aléatoire n'est qu'une matrice où chaque entrée est une variable aléatoire. Toutefois, après avoir connu un premier essor en physique quantique grâce aux travaux de E. Wigner, elles ont montré une réelle efficacité pour l'étude de systèmes complexes corrélés et plusieurs branches d'applications se sont développées [5] [9] [22]. Dans le cadre de notre étude, nous allons utiliser les matrices aléatoires pour estimer les matrices de covariance des actifs financiers. Les estimateurs de Ledoit-Péché et celui pour les matrices de crosscovariances que nous allons présenter s'appuient sur la théorie des matrices aléatoires [10] [2]. On peut également noter qu'ils sont utilisés par certaines entreprises dans le cadre de la gestion des risques de portefeuilles financiers.

4.1 Quelques éléments d'algèbre linéaire et de matrices aléatoires

Commençons déjà par rappeler la définition d'une matrice variance-covariance :

Définition 1. *Soient $X_1, X_2, ..., X_n$ des variables aléatoires à variances finies. La matrice de variance-covariance des $(X_1, X_2, ..., X_n)$ est la matrice $\mathbf{\Sigma}$ de taille $n \times n$ telle que :*

$$\Sigma_{ij} = Cov(X_i, X_j), \text{ pour } 1 \leq i, j \leq n$$

Ainsi les éléments diagonaux de $\mathbf{\Sigma}$ sont les variances des $(X_i)_{1 \leq i \leq n}$

Proposition 1. *Une matrice de variance-covariance est symétrique et semi-définie positive. Ses valeurs propres sont donc positives ou nulles.*

Le théorème spectral que l'on rappelle ici, nous permet de savoir que toute matrice de covariance est diagonalisable sur une base orthonormée :

Définition 2. *Soit $\boldsymbol{A} \in \mathbb{R}^{n \times n}$. $\boldsymbol{A}$ est dite orthogonale si $\boldsymbol{A}$ est inversible et d'inverse égal à sa transposée : $\boldsymbol{A}\boldsymbol{A}^{\boldsymbol{T}} = \boldsymbol{Id}$.*

Théorème 1. *Soit $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ une matrice symétrique (resp. hermitienne dans $\mathbb{C}^{n \times n}$). Alors il existe une matrice orthogonale $\boldsymbol{V}$ (resp. unitaire) et une matrice diagonale $\boldsymbol{D}$ telles que :*

$$\boldsymbol{A} = \boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^{-1}$$

(Théorème Spectral)

Nous allons commencer par quelques définitions. Certains objets, comme la trace normalisée ou la transformée de Stieltjes reviendront à de nombreuses reprises dans ce rapport, que ce soit dans la compréhension théorique des estimateurs que l'on étudie ou dans les applications [17][21].

Aussi, l'hypothèse d'invariance par rotation est très importante pour les estimateurs que l'on va présenter plus loin dans ce rapport [10] [2]. Nous voudrions donc définir cette notion puis l'illustrer avec les matrices de Wigner [17].

Définition 3. *Soit $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, on note τ la trace normalisée :*

$$\tau(\boldsymbol{A}) = \frac{1}{n} Tr(\boldsymbol{A})$$

L'intérêt de définir la trace normalisée est de pouvoir regarder la limite pour $n \to +\infty$ tout en gardant une trace finie. La trace normalisée représente également la moyenne arithmétique des valeurs propres d'une matrice.

Définition 4. *Une matrice Wigner à valeurs réelles est une matrice symétrique $\boldsymbol{X} \in \mathbb{R}^{n\times n}$ contenant des entrées gaussiennes centrées et indépendantes de loi $\mathcal{N}(0, \sigma_d^2)$ sur la diagonale et $\mathcal{N}(0, \sigma_{od}^2)$ en dehors [17] .*

Nous définissons m_k le moment d'ordre k ($k \in \mathbb{N}$) d'une matrice $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ de la manière suivante :

$$m_k = \mathbb{E}(\tau(\boldsymbol{A}^k))$$

Nous souhaiterions déterminer les variances des éléments diagonaux et hors diagonaux d'une matrice de Wigner afin que les moments d'ordres 1 et 2 soient indépendants de la taille de la matrice $X \in \mathbb{R}^{n\times n}$ pour $n \to +\infty$. [17]

Calculons donc les deux premiers moments d'une matrice de Wigner à l'aide de la trace normalisée :

$$\tau(\boldsymbol{X}) = \frac{1}{n}\mathbb{E}(Tr(\boldsymbol{X})) = \frac{1}{n}\mathbb{E}(\boldsymbol{X}) = 0.$$

Le moment d'ordre 1 est donc déjà indépendant de la taille de $\boldsymbol{X}$.

$$\tau(\boldsymbol{X}^2) = \frac{1}{n}\mathbb{E}(Tr(\boldsymbol{X}\boldsymbol{X}^T)) = \frac{1}{n}\mathbb{E}(\sum_{i=1}^{n}\sum_{j}^{n}\boldsymbol{X}_{ij}^2)) = \frac{1}{n}[n(n-1)\sigma_{od}^2 + n\sigma_d^2] = (n-1)\sigma_{od}^2 + \sigma_d^2$$

On choisit, par exemple, la normalisation suivante en prenant un $\sigma > 0$ et en posant [17] :

$$\sigma_{od}^2 = \frac{\sigma^2}{n}$$

et

$$\sigma_d^2 = \frac{2\sigma^2}{n}$$

De telles matrices de Wigner constituent l'ensemble GOE (*Gaussian orthogonal ensemble*). σ^2 est appelé la variance de la matrice de Wigner. Nous verrons ci-après que cette normalisation permet d'obtenir des matrices dites invariantes par rotation (d'où le choix du facteur 2 pour σ_d qui peut ne pas paraître naturel en premier lieu).

Proposition 2. *Soit $\boldsymbol{X} \in \mathbb{R}^{n\times n}$ une matrice de Wigner de variance σ^2 (c'est-à-dire que la variance de ses éléments hors diagonale est $\sigma_{od}^2 = \frac{\sigma^2}{n}$ et que celle des ses éléments diagonaux est $\sigma_d^2 = 2\sigma_{od}$). La densité spectrale de $\boldsymbol{X}$, c'est-à-dire la loi des ses valeurs propres, suit la loi du demi-cercle [17] [21] :*

$$\forall\lambda \in [-2\sigma, 2\sigma],\ \rho(\lambda) = \frac{\sqrt{4\sigma^2 - \lambda^2}}{2\pi\sigma^2}$$

Proposition 3. *Une matrice de Wigner de taille $n\times n$ de l'ensemble GOE est invariante par rotation. Cela signifie que pour $\boldsymbol{X} \in \mathbb{R}^{n\times n}$ matrice de Wigner GOE et pour toute $\boldsymbol{O}$ matrice orthogonale $\boldsymbol{O}\boldsymbol{X}\boldsymbol{O}^{\boldsymbol{T}}$, qui a les mêmes valeurs propres que $\boldsymbol{X}$, a la même probabilité d'apparaître que $\boldsymbol{X}$. On notera :*

$$\boldsymbol{O}\boldsymbol{X}\boldsymbol{O}^{\boldsymbol{T}} \overset{loi}{=} \boldsymbol{X}$$

Démonstration. [17] Il suffit de montrer qu'un vecteur gaussien centré iid reste gaussien de même loi s'il subit une rotation. Soit $v \in \mathbb{R}^n$ un vecteur gaussien iid et soit $\boldsymbol{O}$ une matrice orthogonale (c'est-à-dire que $\boldsymbol{O^T} = \boldsymbol{O^{-1}}$).
On pose $w = \boldsymbol{O}v$.
w est bien gaussien car chacun de ses éléments est une combinaison linéaire de variables aléatoires gaussiennes. On calcule ensuite la covariance de ce vecteur :

$$\mathbb{E}(w_i w_j) = \sum_{k,l} \boldsymbol{O}_{ik}\boldsymbol{O}_{jl}\mathbb{E}(v_k v_l) = \sum_{k,l} O_{ik}O_{jl}\delta_{kl} = (\boldsymbol{OO^T})_{ij} = \delta_{ij}$$

où $(\boldsymbol{OO^T})_{ij}$ est le terme d'indices (i, j) de la matrice $\boldsymbol{OO^T}$.

Ainsi w reste centré avec une covariance égale à l'Identité comme v.

Revenons maintenant aux matrices, soit $\boldsymbol{A}$ une matrice de Wigner de taille $n \times n$, on peut écrire : $\boldsymbol{A} = \boldsymbol{H} + \boldsymbol{H^T}$ (à une constante multiplicative près) avec $\boldsymbol{H}$ une matrice à entrées gaussiennes iid de variance égale à 1.
Chaque colonne de $\boldsymbol{H}$ est invariante par rotation comme montré ci-dessus, ainsi :

$$\boldsymbol{OH} \overset{loi}{=} \boldsymbol{H}$$

Les lignes de $\boldsymbol{OH}$ sont invariantes par rotation donc :

$$\boldsymbol{OHO^T} \overset{loi}{=} \boldsymbol{OH}$$

Enfin :

$$\boldsymbol{OAO^T} \overset{loi}{=} \boldsymbol{O}(\boldsymbol{H} + \boldsymbol{H^T})\boldsymbol{O^T} \overset{loi}{=} \boldsymbol{H} + \boldsymbol{H^T} = \boldsymbol{X}$$

□

Définition 5. *Soit $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ [17] :*

- *La résolvante :*

$$G_{\boldsymbol{A}}(z) = (z\boldsymbol{Id} - \boldsymbol{A})^{-1}$$

où $z \in \mathbb{C}$, la résolvante n'est pas définie sur les valeurs propres de $\boldsymbol{A}$.

- *La transformée de Stieltjes :*

$$g_n^{\boldsymbol{A}}(z) = \frac{1}{n}Tr(G_{\boldsymbol{A}}(z)) = \frac{1}{n}\sum_{k=1}^{n}\frac{1}{z - \lambda_k}$$

où $(\lambda_1, ..., \lambda_n)$ sont les valeurs propres de $\boldsymbol{A}$ et où $z \in \mathbb{C}$. Comme pour la résolvante, la transformée de Stieltjes n'est pas définie sur les valeurs propres de $\boldsymbol{A}$.

La transformée de Stieltjes est un outil très important en matrices aléatoires, notamment parce qu'elle est reliée à la densité des valeurs propres par la formule de Sokhotski-Plemelj [17] :

Proposition 4. *Soit g_μ la transformée de Stieltjes associée à la mesure μ. Par exemple, μ pourrait être la densité des valeurs propres d'une matrice aléatoire. La partie imaginaire de la transformée de Stieltjes et la mesure μ sont liées par la relation suivante :*

$$\forall z = x + i\eta \in \mathbb{C}, \mu(x) = -\frac{1}{\pi}\lim_{\eta\to 0^+}(Img_\mu(x + i\eta)dx)$$

(Formule d'inversion de la Stieltjes)

4.2 Matrice Inverse Wishart et estimateur dans le cadre Bayésien

Dans cette partie nous allons définir la notion de matrice Inverse-Wishart. Ces matrices vont revenir tout au long du rapport dans les simulations. On commence par donner la définition d'une matrice Inverse Wishart [17] .

Définition 6. *Soit $n \in \mathbb{N}$ et $p \in]0,1[$. On pose $q^* = \frac{p}{1+p}$ et $T = \lfloor \frac{n}{q^*} \rfloor$. Pour $1 \leq t \leq T$, m_t est un vecteur généré par une loi normale standard multivariée ($\mathcal{N}(0, \boldsymbol{C})$) et $\boldsymbol{C}$ la matrice de covariance. On note $\boldsymbol{M} = (m_1, ..., m_T)$ une matrice de taille $n \times T$.*
On appelle matrice Wishart de taille n et de paramètre p l'objet suivant :

$$\boldsymbol{W} = \frac{1}{T}\boldsymbol{M}\boldsymbol{M}^{\boldsymbol{T}}$$

Si $\boldsymbol{C} = \boldsymbol{Id}$ on dit que la matrice est une Wishart blanche.
Une Inverse Wishart de paramètre p est alors :

$$(1 - q^*) * \boldsymbol{W}^{-\mathbf{1}}$$

où $(1 - q^*)$ est une constante de normalisation.

Dans le problème de l'estimation de covariance, nous observons une matrice de covariance empirique $\boldsymbol{E}$. Les données, supposées centrées, sont générées à partir d'une matrice de covariance $\boldsymbol{C}$ inconnue. On considère que l'observable $\boldsymbol{E}$ est constituée de $\boldsymbol{C}$ et d'un bruit multiplicatif $\boldsymbol{W}$ (Wishart blanche)[17] :

$$\boldsymbol{E} \overset{loi}{=} \boldsymbol{C}^{\frac{\mathbf{1}}{\mathbf{2}}}\boldsymbol{W}\boldsymbol{C}^{\frac{\mathbf{1}}{\mathbf{2}}}$$

Nous cherchons à identifier $\mathbb{P}(\boldsymbol{C}|\boldsymbol{E})$, on cherche donc à estimer $\boldsymbol{C}$ à partir de la matrice empirique des données $\boldsymbol{E}$.

Considérons, dans un premier temps, l'estimation de Bayes pour des variables aléatoires. Nous écrirons ensuite un équivalent dans le cas matriciel [17] :
y est la variable observable de x et on suppose connue la loi de y sachant x. Alors :

- $P(y|x)$ est la loi de l'échantillon
- $P(x|y)$ est la loi postérieure

$$P(x|y) = \frac{P(y|x)P_0(x)}{P(y)} = \frac{1}{Z}P(y|x)P_0(x) \text{ où } Z = \int dx P(y|x)P_0(x)$$

Considérons maintenant le cas matriciel. Pour une estimation de la vraie matrice de covariance $\boldsymbol{C}$, l'estimation Bayésienne devient (on note par $\propto$ la relation de proportionnalité) [17] :

$$P(\boldsymbol{C}|\boldsymbol{E}) \propto P(\boldsymbol{E}|\boldsymbol{C})P_0(\boldsymbol{C})$$

avec :

$$P(\boldsymbol{E}|\boldsymbol{C}) \propto (det\boldsymbol{C})^{-T/2} exp[-\frac{T}{2} Tr(\boldsymbol{C^{-1}E})]$$

Pour $P_0(\boldsymbol{C})$ on peut supposer une densité, par exemple celle d'une Inverse Wishart de taille n à T^* degrés de liberté avec $T^* > n+1$ (ie : $q^* = \frac{n}{T^*}$) :

$$P_0(\boldsymbol{C}) \propto (det\boldsymbol{C})^{-(T^*+n+1)/2} exp(-\frac{T^*-n-1}{2} Tr(\boldsymbol{C^{-1}X}))$$

Avec ce choix, on a :

$$P(\boldsymbol{C}|\boldsymbol{E}) \propto (det\boldsymbol{C})^{-(T+T^*+n+1)/2} exp[-\frac{T}{2} Tr(\boldsymbol{C^{-1}E^*})]$$

Où :

$$\boldsymbol{E^*} = \boldsymbol{E} + \frac{T^*-n-1}{T}\boldsymbol{X}$$

Proposition 5. *On se place dans le cas d'une matrice Inverse Wishart $\boldsymbol{C}$ de taille n, de paramètre p. On pose $X \in \mathbb{R}^n$ un vecteur aléatoire normal centré de covariance $\boldsymbol{C}$ puis on génère T observations de X : $(X(t))_{1\leq t\leq T}$, ces observations sont supposées indépendantes. On appelle $\boldsymbol{E} = \frac{1}{T}\boldsymbol{X}\boldsymbol{X^T}$ la matrice empirique des observations. On pose : $q = \frac{n}{T}$, $q^* = \frac{n}{T^*}$ et $p = \frac{q^*}{1-q^*}$. L'estimateur optimal (c'est-à-dire ici $\mathbb{E}(\boldsymbol{C}|\boldsymbol{E})$) de $\boldsymbol{C}$ à partir de $\boldsymbol{E}$ s'obtient de la manière suivante [17] :*

$$\mathbb{E}(\boldsymbol{C}|\boldsymbol{E}) = r\boldsymbol{E} + (1-r)\boldsymbol{Id} \quad \textit{avec} \quad r = \frac{T}{T+T^*-n-1}$$

(Estimateur optimal cas Inverse Wishart)

On peut facilement donner une approximation de r, le coefficient du shrinkage linéaire, en fonction de p et q en considérant n comme grand (on néglige le -1 au dénominateur, on note $\sim$ l'équivalence) :

$$r = \frac{n}{q}\frac{1}{\frac{n}{q}+\frac{n}{q^*}-n-1} \sim \frac{1}{1+\frac{q}{q^*}-q} \sim \frac{1}{1-q+\frac{q(1+p)}{p}} \sim \frac{p}{p+q}$$

Nous avons voulu illustrer cette proposition avec $\boldsymbol{C}$ une Inverse Wishart de paramètres $n = 600$ et $p = 0.25$, on génère alors $T = 2000$ observations à partir d'une loi normale centrée de covariance $\boldsymbol{C}$, on a donc ici $q = \frac{n}{T} = 0.3$. Nous avons ensuite tracé les valeurs propres empiriques (de $\boldsymbol{E}$), les vraies valeurs propres (de $\boldsymbol{C}$) et les valeurs propres de l'estimateur optimal de $\boldsymbol{C}$ à partir de $\boldsymbol{E}$ (celles $\boldsymbol{X}$) :

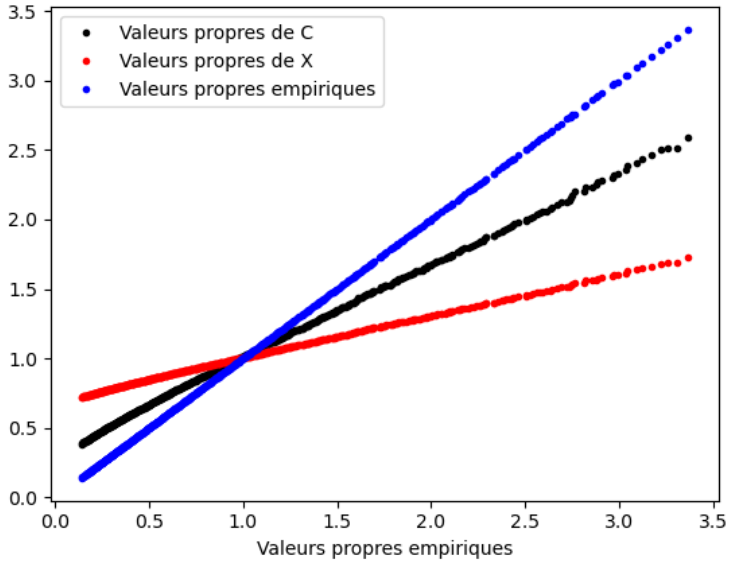

FIGURE 3 – Valeurs propres de différents estimateurs d'une matrice Inverse Wishart de paramètres $n = 600$, $p = 0.25$ et $q = 0.3$ en fonction des valeurs propres empiriques

On remarque que le spectre de la matrice empirique est plus large que le spectre de la vraie Inverse Wishart $\boldsymbol{C}$. Les valeurs propres empiriques minimales sont plus petites que les vraies valeurs propres et les valeurs propres empiriques maximales sont plus plus grandes que les valeurs propres de $\boldsymbol{C}$ maximales. Nous remarquons que la largeur du spectre de $\boldsymbol{X}$ est encore plus petite que celle de $\boldsymbol{C}$.

Pour obtenir les valeurs propres de $\boldsymbol{X}$ on a conservé les vecteurs propres empiriques et on a appliqué une même fonction affine à chaque valeur propre empirique avec un coefficient directeur strictement inférieur à 1. Cette transformation a permis de réduire la largeur du spectre empirique afin de se rapprocher du spectre de la matrice $\boldsymbol{C}$ que l'on cherche à estimer. Cette technique qui vise à réduire le bruit en multipliant les valeurs propres par un même coefficient strictement inférieur à 1 s'appelle le shrinkage linéaire [11].

Nous verrons plus tard que l'estimateur de Ledoit-Péché des matrices de covariance est un shrinkage non linéaire. Il recouvre en revanche le shrinkage linéaire lorsqu'on estime une matrice Inverse Wishart. L'idée est aussi de réduire la largeur du spectre empirique en faisant subir une transformation affine aux valeurs propres avec un coefficient directeur inférieur à 1 mais le coefficient n'est pas le même pour chaque valeur propre. [10]

Des simulations nous ont montré que l'erreur de l'estimateur du shrinkage linéaire $\boldsymbol{X}$ tel que défini précédemment par rapport à la matrice $\boldsymbol{C}$ est linéaire en n à q et p fixés. On considère l'erreur au sens de la norme de Frobenius :

$$Erreur = \| \boldsymbol{X} - \boldsymbol{C} \|_F^2$$

avec $\|.\|_F$ la norme de Frobenius : $\|\boldsymbol{M}\|_F = \sqrt{Tr(\boldsymbol{M}\boldsymbol{M}^{\boldsymbol{T}})}$

Nous avons donc voulu le vérifier théoriquement en calculant l'espérance de l'erreur du shrinkage linéaire. On pose $\boldsymbol{C}^* = r\boldsymbol{E} + (1-r)\boldsymbol{Id}$, on a :

$$\tau((\boldsymbol{C} - \boldsymbol{C}^*)^2) = \tau((\boldsymbol{C} - r\boldsymbol{E} - (1-r)\boldsymbol{Id})^2)$$

$$\tau((\boldsymbol{C}-\boldsymbol{C}^*)^2) = \tau(\boldsymbol{C^2} - r\boldsymbol{C}\boldsymbol{E} - (1-r)\boldsymbol{C} - r\boldsymbol{E}\boldsymbol{C} + r^2\boldsymbol{E^2} + r(1-r)\boldsymbol{E} - (1-r)\boldsymbol{C} + r(1-r)\boldsymbol{E} + (1-r)^2\boldsymbol{Id}$$

En utilisant le fait que $\tau(\boldsymbol{A}\boldsymbol{B}) = \tau(\boldsymbol{B}\boldsymbol{A})$ on a :

$$\tau((\boldsymbol{C} - \boldsymbol{C}^*)^2) = \tau(\boldsymbol{C^2}) - 2r\tau(\boldsymbol{C}\boldsymbol{E}) - 2(1-r)\tau(\boldsymbol{C}) + 2r(1-r)\tau(\boldsymbol{E}) + r^2\tau(\boldsymbol{E^2}) + (1-r)^2\tau(\boldsymbol{Id})$$

En parallèle, on a que [17]

- $\mathbb{E}(\tau(\boldsymbol{C})) = \mathbb{E}(\tau(\boldsymbol{E})) = \tau(\boldsymbol{Id}) = 1$
- $\mathbb{E}(\tau(\boldsymbol{C^2})) = 1 + p$
- $\mathbb{E}(\tau(\boldsymbol{C}\boldsymbol{E})) = 1 + p$
- $\mathbb{E}(\tau(\boldsymbol{E^2})) = 1 + p + q$

Après quelques simplifications, on obtient :

$$\mathbb{E}(\tau((\boldsymbol{C} - \boldsymbol{C}^*)^2)) = p(1-r)^2 + r^2 q$$

Et en utilisant l'hypothèse n grand :

$$Erreur = \mathbb{E}(\tau((\boldsymbol{C} - \boldsymbol{C}^*)^2)) \sim \frac{pq}{p+q}$$

4.3 Théorème de Marcenko-Pastur

Le théorème de Marcenko Pastur décrit le comportement asymptotique des valeurs propres d'une matrice aléatoire quand sa taille tend vers l'infini [14]. Ce théorème est d'une grande importance pour notre étude des estimateurs des matrices de covariance, tout d'abord parce que les estimateurs type Ledoit-Péché se focalisent sur le nettoyage des valeurs propres mais aussi parce que l'objectif des estimateurs va être de développer des résultats similaires en taille finie, pour des grandes matrices. Nous avons adapté à nos notations les notations de l'article *A very short proof of Marcenko-Pastur Theorem* [24].

Soit $X_{nT} \in \mathbb{R}^{n\times T}$ telles que les colonnes $\{x_{nk}\}_{k=1}^{T}$ sont les copies iid d'un même vecteur aléatoire $x_n \in \mathbb{R}^n$ pour tous $n, T \geq 1$. Tous les éléments aléatoires du problème sont définis sur le même espace de probabilité.

Le but est d'étudier la mesure empirique des valeurs propres de la matrice $\boldsymbol{E} = \frac{1}{T}\boldsymbol{X_{nT}}\boldsymbol{X_{nT}^T}$:

$$\mu = \frac{1}{n}\sum_{i=1}^{n}\delta_{\lambda_i}$$

avec $\lambda_1 \leq ... \leq \lambda_n$ valeurs propres de la matrice $\boldsymbol{E}$.

Définition 7. *Pour $q > 0$ on définit :*

$$\mu_q(x) = (1-\frac{1}{q})^+\delta_0 + \frac{\sqrt{(\lambda_+ - x)(x-\lambda_-)}}{2\pi q x}\boldsymbol{1}(x \in [\lambda_-, \lambda_+])dx$$

Avec : $x^+ = max(x, 0)$, $\lambda_- = (1-\sqrt{q})^2$, $\lambda_+ = (1+\sqrt{q})^2$ et $\boldsymbol{1}$ la fonction indicatrice.

(Loi de Marcenko-Pastur)

Voici une représentation de la densité de Marcenko-Pastur pour différentes valeurs de q :

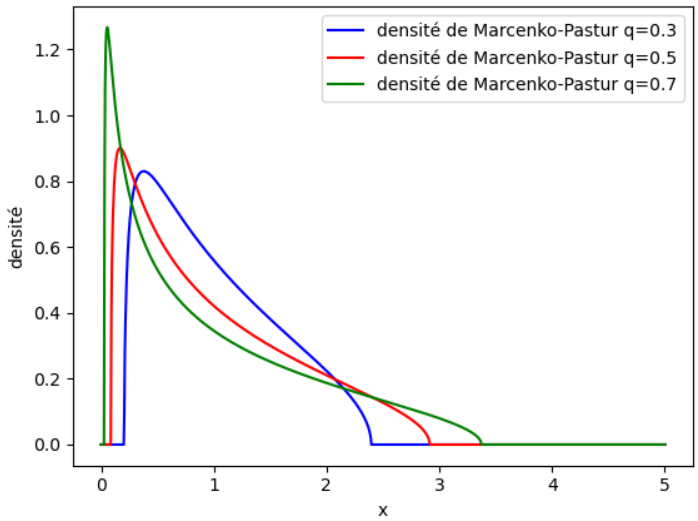

FIGURE 4 – Densité de Marcenko-Pastur pour différentes valeurs de q

Pour illustrer la convergence de la densité des valeurs propres d'une matrice aléatoire vers la loi de Marcenko-Pastur, nous avons généré des données à partir d'une normale centrée de covariance identité avec $q = \frac{n}{T} = 0.3$ fixé où T représente le nombre d'observations. Nous avons testé pour $n = 50$ et $n = 700$. Générer un grand nombre de matrices de covariance empirique pour une même taille pour établir la densité spectrale empirique revient à augmenter le nombre d'observations. On aurait donc pu obtenir des densités empiriques plutôt proches de la densité de Marcenko-Pastur même pour des petites tailles.

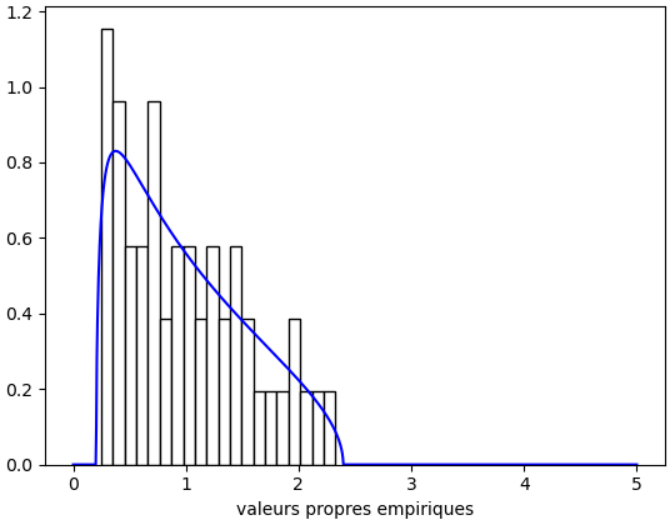

FIGURE 5 – Densité spectrale empirique pour $n = 50$, $q = 0.3$ et une covariance égale à l'identité (en bleu densité de Marcenko-Pastur)

On voit que le spectre empirique est ne correspond pas vraiment au spectre théorique de Marcenko-Pastur valable pour $n \to \infty$. La taille de la matrice de covariance est trop petite en être vraiment fidèle.

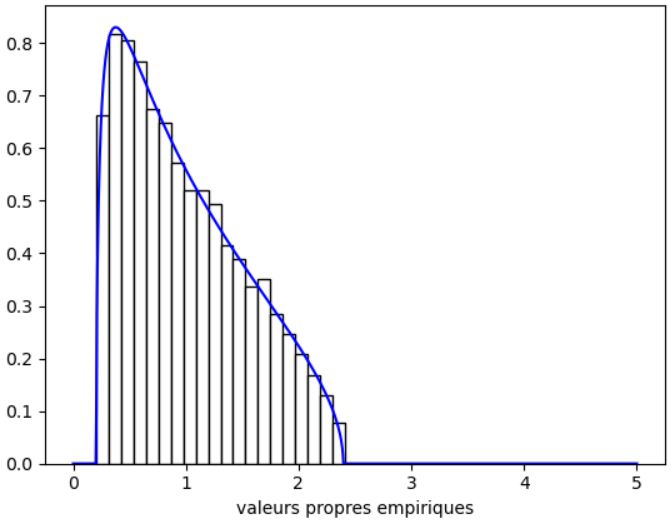

FIGURE 6 – Densité spectrale empirique pour $n = 700$, $q = 0.3$ et une covariance égale à l'identité (en bleu densité de Marcenko-Pastur)

Après avoir testé pour $n = 700$ on se rend compte qu'on est là beaucoup plus proche de

la densité de Marcenko-Pastur.

Pour que la convergence en loi vers la densité de Marcenko-Pastur, l'article pose trois hypothèses [24] :

- les vecteurs générés à partir du vecteur aléatoire x_n sont centrées
- les $\{X_{nk}\}_{k=1}^{T}$ sont indépendants
- Condition de Lindberg :

$$\forall \epsilon > 0, \lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{E}(X_{nk}^2 \mathbf{1}(|X_{nk}| > \epsilon\sqrt{n}) = 0$$

où **1** est la fonction indicatrice.

La condition de Lindberg est une condition suffisante pour pouvoir utiliser le théorème Central limite.

On rappelle que μ_n converge vers μ dans le sens de la topologie faible si et seulement si pour toute fonction f continue et bornée on a :

$$\int f d\mu_n \to \int f d\mu$$

Théorème 2. *Pour tout $T = T(n)$ avec $\frac{n}{T} \to_{n\to\infty} q > 0$, μ_{nT} converge presque-sûrement pour la convergence au sens de la topologie faible vers μ_q :*

$$\mathbb{P}(\mu_{nT} \to \mu_q, n \to \infty) = 1$$

(Théorème de Marcenko-Pastur)

Le théorème de Marcenko-Pastur est fondamental en matrices aléatoires car il établit des conditions qui impliquent la convergence de la densité des valeurs propres d'une matrice aléatoire vers une densité connue quand la taille de la matrice tend vers $+\infty$. Les estimateurs que l'on va définir ci-après vont se concentrer sur le "nettoyage" des valeurs propres empiriques pour se rapprocher de la "vraie matrice de variance-covariance" des données. L'enjeu va donc être d'avoir un analogue de la distribution de Marcenko-Pastur avec une matrice de taille finie mais grande, typiquement de l'ordre quelques centaines. Cet ordre de grandeur correspond à celui des portefeuilles financiers.

4.4 Estimateur de Ledoit-Péché pour les matrices de covariance

On dispose de données de moyenne nulle générées à partir d'une matrice de covariance $\mathbf{\Sigma}$ que l'on ne connaît pas a priori. Ledoit et Péché ont développé un estimateur de $\mathbf{\Sigma}$ en faisant l'hypothèse de l'invariance par rotation [10]. Nous allons dans un premier temps exposer le problème et ses hypothèses tels qu'ils figurent dans l'article de Ledoit et Péché puis nous présenterons une partie de la preuve dans le cas particulier où les données sont générées à partir d'une loi normale multivariée [1]. La preuve dans le cas gaussien est détaillée dans les annexes.

Commençons par poser quelques notations pour décrire le problème :

— $X \in \mathbb{R}^n$ un vecteur normal centré de matrice de covariance $\mathbf{\Sigma}$

— $\boldsymbol{X} = (X(1), ..., X(T)) \in \mathbb{R}^{n \times T}$ des observations de X iid

— $\boldsymbol{E} = \frac{1}{T}\boldsymbol{X}\boldsymbol{X}^{\boldsymbol{T}}$ la matrice de covariance empirique de valeurs propres $(\lambda_1, ...\lambda_n)$ et vecteurs propres $(u_1, ..., u_2)$.

L'objectif de l'article de Ledoit et Péché est d'étudier le comportement spectral de $\boldsymbol{E}$ dans le cas où n est grand pour ensuite estimer $\mathbf{\Sigma}$.

L'article suppose que $\boldsymbol{X} = \mathbf{\Sigma}_{\boldsymbol{n}}^{\boldsymbol{1/2}}\boldsymbol{H_n}$ où :

- $\boldsymbol{H_n}$ est une matrice de taille $n \times T$ dont les entrées sont des variables aléatoires de moyenne nulle, de variance égale à 1 et admettant un moment centré d'ordre 12 borné par une constante B indépendante de n et T (pour une variable aléatoire Y cela signifie que $\mathbb{E}((Y - \mathbb{E}(Y))^{12}) < B$).

- la matrice de covariance $\mathbf{\Sigma_n}$ de taille $n \times n$ est une matrice aléatoire symétrique définie positive indépendante de $\boldsymbol{H_n}$

- $\frac{T}{n} \to \gamma > 0$ quand n et T tendent vers l'infini (tout au long du rapport on préfère utiliser $q = \frac{n}{T}$ qui n'est donc que l'inverse de γ).

- On note $\tau = (\tau_1, ..., \tau_n)$ les valeurs propres de $\mathbf{\Sigma_n}$ et $\rho_n(\tau) = \frac{1}{n}\sum_{j=1}^{n} \mathbf{1}_{[\tau_j, +\infty[}(\tau)$ la distribution empirique du spectre de $\mathbf{\Sigma_n}$. On suppose que cette densité converge vers une limite déterministe $\rho(\tau)$ en tout point de continuité de ρ pour n tend vers l'infini. ρ définit une loi de probabilité dont le support $Supp(\rho) \subset [a_1, a_2]$ avec $0 < a_1 \leq a_2 < \infty$.

Ces hypothèses sont vérifiées lorsque X est un vecteur aléatoire gaussien centré de covariance $\mathbf{\Sigma}$. La loi normale admet des moments de tous ordres. Les moments d'ordres impairs sont nuls et par exemple le moment d'ordre 12 est égal à :

$$m_{12} = \frac{(12)!}{2^6 \times 6!}$$

La dernière hypothèse est vérifiée par le théorème de Marcenko-Pastur [14].

On se place désormais dans le cas de données générées par une loi gaussienne pour présenter la preuve. Cette preuve provient de l'article *A very short proof of Ledoit-Péché's RIE formula for covariance matrices* [1].

L'objectif est d'estimer au mieux $\mathbf{\Sigma}$ dans l'ensemble des estimateurs invariants par rotation (RIE : *Rotationnally Invariant Estimators*). Cette hypothèse d'invariance trouve une cohérence dans le cadre de l'estimation Bayésienne : comme on dispose de peu de données, les vecteurs propres ne peuvent pas être estimés avec précision. Ainsi, chaque base orthonormée de vecteurs propres est considérée comme équiprobable et on décide donc de conserver les vecteurs propres de l'estimateur empirique en l'absence d'information supplémentaire.

On obtient alors un problème d'optimisation que l'on peut formuler de la manière suivante :

$$\underset{Estimateur \in RIE}{\text{argmin}} \quad \|\boldsymbol{Estimateur} - \mathbf{\Sigma}\|_F$$

avec $\|.\|_F$ la norme de Frobenius : $\|\boldsymbol{M}\|_F = \sqrt{Tr(\boldsymbol{M}\boldsymbol{M}^{\boldsymbol{T}})}$

On obtient alors directement l'estimateur invariant par rotation optimal :

$$\boldsymbol{\xi}(\boldsymbol{E}) = \widehat{\boldsymbol{E}} = \sum_{k=1}^{n} \widehat{\lambda}_k u_k u_k^T$$

avec

$$\xi(\lambda_k) = \widehat{\lambda}_k = u_k^T \mathbf{\Sigma} u_k$$

On remarque que $(\xi(\lambda_k))$ dépend de $\mathbf{\Sigma}$ qui nous est a priori inconnue. Pour cela, on appelle ξ la fonction oracle. L'objet de la preuve va être d'approximer cette quantité à l'aide de variables connues.

Dans un premier temps, nous pouvons considérer la densité empirique des valeurs propres de $\boldsymbol{E}$ et $\widehat{\boldsymbol{E}}$:

$$\sum_{k=1}^{n} \delta_{\lambda_k} \qquad \text{et} \qquad \sum_{k=1}^{n} u_k^T \mathbf{\Sigma} u_k \delta_{\lambda_k}$$

L'article définit les quantités suivantes :

— $G(z) = \tau((z\boldsymbol{Id} - \boldsymbol{E})^{-1})$, la trace normalisée de la résolvante de $\boldsymbol{E}$

— $L(z) = \tau(\mathbf{\Sigma}(z\boldsymbol{Id} - \boldsymbol{E})^{-1})$

Les mesures empiriques sont discrètes mais grâce à la formule d'inversion de la transformée de Stieltjes qui approxime la densité des valeurs propres, nous allons pouvoir passer à des quantités continues, on rappelle cette formule [17] :

Proposition 6. *Soit g_μ la transformée de Stieltjes associée à la mesure μ. La transformée de Stieltjes et la mesure sont liées par la relation suivante :*

$$\forall z = x + i\eta \in \mathbb{C}, \mu(x) = -\frac{1}{\pi} \lim_{\eta \to 0^+} (Img_\mu(x + i\eta)dx)$$

(Formule d'inversion de la Stieltjes)

Nous allons lier ξ à la quantité L en nous rappelant que nous sommes dans le cadre Bayésien (hypothèse de l'équiprobabilité des bases de vecteurs propres) et en utilisant la formule de l'espérance totale :

$$\begin{aligned}\tau(\xi(\boldsymbol{E})(z\mathbf{Id}-\boldsymbol{E})^{-1}) &= \tau(\mathbf{E}(\boldsymbol{\Sigma}|\boldsymbol{E})(z\mathbf{1}-\boldsymbol{E})^{-1}) \\ &= \tau(\boldsymbol{\Sigma}(z\mathbf{Id}-\boldsymbol{E})^{-1}] = L(z)\end{aligned}$$

Si on écrit $L(z)$ dans sa forme continue, on obtient :

$$L(z) = \int_{\mathbb{R}} \rho_E(\lambda)\frac{\xi(\lambda)}{z-\lambda}d\lambda$$

Il nous reste ensuite à utiliser la formule d'inversion de la transformée de Stieltjes sur les quantités L et G ce qui nous donnera une nouvelle formulation de ξ :

$$\lim_{\eta\to 0^+} ImL(\lambda+i\eta) = -\pi\rho_E(\lambda)\xi(\lambda)$$

$$\lim_{\eta\to 0^+} ImG(\lambda+i\eta) = -\pi\rho_E(\lambda)$$

Soit $\epsilon > 0$ tel que $[\lambda_k-\epsilon, \lambda_k+\epsilon]\cap\{\lambda_1,...,\lambda_n\} = \{\lambda_k\}$:

$$\widehat{\lambda_k} = \lim_{\eta\to 0^+} \frac{\int_{\lambda_k-\epsilon}^{\lambda_k+\epsilon} ImL(x+i\eta)dx}{\int_{\lambda_k-\epsilon}^{\lambda_k+\epsilon} ImG(x+i\eta)dx}$$

Le résultat principal de l'article est l'approximation de $L(z)$ par des quantités connues :

Théorème 3.

$$\forall z\in\mathbb{C}\setminus\mathbb{R}, L(z) = 1 - \frac{1}{1-q+zG(z)} + o(1)$$

avec $o(1)$ une quantité qui tend vers 0 pour $n, T \to +\infty$ tel que $\frac{n}{T}\to q$, avec q une constante.

On rappelle que dans le cas où n et T tendent vers l'infini avec $\frac{n}{T} = q$, la densité des valeurs propres de $\boldsymbol{\Sigma}$ est connue par le théorème de Marcenko-Pastur [14].

Ici on cherche une sorte de développement limité de $L(z)$ pour avoir un résultat similaire dans le cas où n et T sont grands mais finis, typiquement n sera de l'ordre d'une ou plusieurs centaines.

La preuve dans le cas gaussien est présentée en détail dans les annexes, ici on se contente de donner directement la formule de Ledoit-Péché :

Théorème 4. *Pour $z = \lambda + i\eta$ avec $\eta > 0$:*

$$\xi(\lambda) = \lim_{\eta \to 0^+} \frac{\lambda}{|1 - q + q\lambda g_\mu(\lambda + i\eta)|^2}$$

(Formule de Ledoit-Péché)

4.5 Estimateur invariant par rotation pour les matrices de crosscovariance

Nous allons maintenant introduire l'estimation des matrices de crosscovariance, une généralisation de l'estimateur de Ledoit-Péché aux matrices rectangulaires. Tout ce qui est présenté dans cette partie provient de l'article *Optimal cleaning for singular values of crosscovariance matrices* de F. Benaych-George, J-P. Bouchaud et M. Potters [2]. On commencera par définir le problème d'estimation de crosscovariance :

Soit $Z = \begin{pmatrix} X \\ Y \end{pmatrix} \in \mathbb{R}^{n+p}$ un vecteur aléatoire tel que $\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}(0, \mathbf{\Sigma})$ avec $\mathbf{\Sigma} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{C} \\ \boldsymbol{C^T} & \boldsymbol{B} \end{pmatrix}$.

On génère T observations $(Z(t))_{1 \leq t \leq T}$ du vecteur aléatoire Z, à l'aide d'une normale multivariée centrée de covariance $\mathbf{\Sigma}$, on suppose les observations indépendantes. L'objectif est d'estimer $\boldsymbol{C}$ à partir de la matrice de crosscovariance empirique $\boldsymbol{C_{XY}}$ dont on peut écrire sa décomposition en valeurs singulières :

$$\boldsymbol{C_{XY}} = \frac{1}{T}\boldsymbol{X}\boldsymbol{Y^T}$$

$$\boldsymbol{C_{XY}} = \sum_{k=1}^{n} s_k u_k v_k^T = [u_1, ..., u_n] diag(s_1, s_2, ..., s_n) [v_1, v_2, ..., v_n]^T$$

On pose également :

$$\boldsymbol{C_X} = \frac{1}{T}\boldsymbol{X}\boldsymbol{X^T}, \qquad \boldsymbol{C_Y} = \frac{1}{T}\boldsymbol{Y}\boldsymbol{Y^T}$$

L'estimateur empirique de $\mathbf{\Sigma}$ la vraie matrice de covariance de Z est alors :

$$\boldsymbol{E} = \begin{pmatrix} \boldsymbol{C_X} & \boldsymbol{C_{XY}} \\ \boldsymbol{C_{XY}^T} & \boldsymbol{C_Y} \end{pmatrix}$$

Comme dans le cas de l'estimateur de Ledoit-Péché pour les matrices de covariance [10], on va chercher un estimateur invariant par rotation, on va donc conserver les vecteurs singuliers de l'estimateur empirique. On a alors un problème d'optimisation où l'on cherche à estimer les valeurs singulières $s_1^c, ..., s_n^c$:

$$argmin || [u_1, ..., u_n] diag(s_1^c, ..., s_n^c) [v_1, ..., v_n]^T - \boldsymbol{C} ||_F$$

avec $\| \ . \ \|_F$ la norme de Frobenius.

La solution est l'oracle suivant :

$$s_k^c = \xi(s_k) = u_k^T \boldsymbol{C} v_k, \text{ pour } k \in [1, n]$$

Ici les mesures empiriques importantes sont les suivantes :

$$m_{C_{XY}, C} = \frac{1}{2n} \sum_{k=1} u_k^T \boldsymbol{C} v_k (\delta_{s_k} - \delta_{-s_k})$$

$$v_{C_{XY}} = \frac{1}{2n} \sum_{k=1}^{n} (\delta_{s_k} + \delta_{s_k})$$

En utilisant la formule d'inversion de la transformée de Stieltjes, on a la formule suivante :

$$s_k^c = \lim_{\eta \to 0^+} \frac{\text{Stieltjes } m_{C_{XY},C}(x+i\eta)}{\text{Stieltjes } \nu_{C_{XY}}(x+i\eta)}$$

Ensuite on calcule les transformées de Stieltjes associées aux mesures empiriques :

$$\text{Stieltjes } m_{C_{XY},C}(z) = \int \frac{dm_{C_{XY},C}(s)}{z-s}$$

$$= \frac{1}{2n}\sum_{k=1}^{n}(\frac{1}{z-s_k} - \frac{1}{z+s_k})u_k^T \boldsymbol{C} v_k$$

$$= \frac{1}{2n}\sum_{k=1}^{n}\frac{2s_k}{z^2-s_k^2}u_k^T \boldsymbol{C} v_k = \frac{1}{n}\sum_{k=1}^{n}\frac{s_k}{z^2-s_k^2}Tr(u_k^T \boldsymbol{C} v_k)$$

$$= \frac{1}{n}\sum_{k=1}^{n}\frac{s_k}{z^2-s_k^2}Tr(\boldsymbol{C^T} u_k v_k^T)$$

$$= \frac{1}{n}Tr(\boldsymbol{C^T}\sum_{k=1}^{n}\frac{s_k}{z^2-s_k^2}u_k v_k^T) = \frac{1}{n}Tr(\boldsymbol{C^T}(z^2 - \boldsymbol{C_{XY}C_{XY}^T})^{-1}\boldsymbol{C_{XY}})$$

$$= \frac{1}{n}TrG\boldsymbol{C_{XY}C^T}$$

On fait le même calcul pour la transformée de Stieltjes de ν :

$$\text{Stieltjes } \nu_{C_{XY}} \text{ at z} = \int \frac{d\nu_{C_{XY}(s)}}{z-s}$$

$$\frac{1}{2n}\sum_{k=1}^{n}(\frac{1}{z-s_k} + \frac{1}{z+s_k})$$

$$= \frac{1}{n}\sum_{k=1}^{n}\frac{z}{z^2-s_k^2}$$

On pose ensuite les quantités suivantes :

- $\mathbf{G} = (z^2 - \boldsymbol{C_{XY}C_{XY}^T})^{-1}$
- $\widetilde{\mathbf{G}} = (z^2 - \boldsymbol{C_{XY^T}C_{XY}})^{-1}$
- $G = \frac{1}{T}Tr\mathbf{G}$
- $L = \frac{1}{T}Tr\mathbf{G}\boldsymbol{C_{XY}C^T}$

Proposition 7.

$$s_k^{cleaned} = u_k^T \boldsymbol{C} v_k \simeq \frac{ImL(z)}{Im(zG(z))} \textit{ pour } z = s_k + i\eta, \textit{ avec } \eta \ll 1$$

Tout comme pour la preuve de Ledoit-Péché, l'enjeu va maintenant être d'approximer la quantité L qui dépend encore de C qu'on ne connaît pas, pour cela l'article définit différentes quantités [2] :

$$H = \frac{1}{T}Tr\mathbf{G}C_{XY}C_{XY}^2, \quad A = \frac{1}{T}Tr\mathbf{G}C_X, \quad B = \frac{1}{T}Tr\widetilde{\mathbf{G}}C_Y, \quad \Theta = z^2\frac{AB}{1+H}$$

Théorème 5. *La fonction L vérifie :*

$$L = \frac{H - \Theta}{1 + H - \Theta} + O\left(\frac{1}{T|Imz|^5}\right)$$

On se contente, ici, d'admettre ce théorème et de signaler que la preuve complète est exposée dans l'article *Optimal cleaning for singular values of crosscovariance matrices* [2]. Cette preuve est assez similaire à la preuve de l'estimateur de Ledoit-Péché dans le cas gaussien [1] et se base également sur l'inégalité concentration de la mesure dans le cas gaussien [21] et sur la formule de Stein [1].

Le théorème d'estimation des matrices de crosscovariance est pour le moment uniquement démontré dans le cas où les données sont générées à partir d'une normale multivariée. Il y a cependant de fortes chances que ce résultat soit généralisable à d'autres types de données et il peut se révéler très intéressant pour estimer les corrélations différentes classes d'actifs financiers.

5 Crossvalidation

La crossvalidation est une technique issue du machine learning qui a l'avantage de ne nécessiter quasiment aucune hypothèse sur les données de départ. On découpe nos observations en ensembles de tailles égales appelés blocs, les observations privées du bloc sont l'ensemble d'optimisation et les observations dans le bloc forment l'ensemble de validation. La technique est appelée crossvalidation et non juste validation parce qu'on effectue cette optimisation/validation sur chaque bloc avant de moyenner [17] [20].

5.1 Fonctionnement de la crossvalidation

Cette partie a pour but d'expliquer le fonctionnement de la technique pour l'estimation des matrices de variances-covariance et pour l'estimation des matrices de crosscovariances. On commence par définir les données du problème :

— T : nombre d'observations

— $X \in \mathbb{R}^n$ un vecteur normal centré de matrice de covariance $\mathbf{\Sigma}$

— $\boldsymbol{X} = (X(1), ..., X(T)) \in \mathbb{R}^{n \times T}$, matrice des observations de X iid

— k : nombre de blocs

— taille des blocs : $\lfloor \frac{T}{k} \rfloor$

Les blocs sont définis sur des données dont les indices t sont dans un intervalle de la forme suivante :

$$\left[i \times \left\lfloor \frac{T}{k} \right\rfloor, (i+1) \times \left\lfloor \frac{T}{k} \right\rfloor \right] \text{ pour } i = 0, ..., k$$

Crossvalidation pour les matrices de variance-covariance

Prenons un des blocs sur les données $[T1, T2]$, on définit alors :

— $\boldsymbol{C_{in}}$ la matrice de covariance empirique sur les données $[T1, T2]$ de vecteurs propres $u_{in} = (u_{in}(1), ..., u_{in}(n))$

— $\boldsymbol{C_{out}}$ la matrice de covariance empirique sur $[0, T1] \cup [T2, T]$ de vecteurs propres $u_{out} = (u_{out}(1), ..., u_{out}(n))$

Sur ce bloc, l'estimateur des valeurs propres est le suivant :

$$\xi(\lambda_i) = u_{out}^T(i) \boldsymbol{C_{in}} u_{out}(i)$$

On peut alors reconstituer "une matrice de covariance" :

$$\boldsymbol{\Xi_k} = u_{out} \xi u_{out}^T$$

L'estimateur de covariance n'est alors que la moyenne de ces estimateurs sur les différents blocs :

$$\boldsymbol{\Xi} = \frac{1}{K}\sum_{k=1}^{K} \boldsymbol{\Xi_k}$$

Voici un premier exemple de crossvalidation où les données sont générées à partir d'une matrice de covariance Inverse Wishart de paramètres $n = 500$, $q = 0.5$, et $p = 0.25$. On a $T = \lfloor n/q \rfloor = 1000$ et on fait tourner la technique pour 40 blocs. En vert on a tracé les valeurs propres $\xi(\lambda_i) = u_{out}^T(i)\boldsymbol{C_{in}}u_{out}(i)$, en noir les valeurs propres de l'estimateur de crossvalidation $\boldsymbol{\Xi}$ et en rouge les valeurs propres nettoyées par l'estimateur de Ledoit-Péché en fonction des valeurs propres empiriques sur l'ensemble des données :

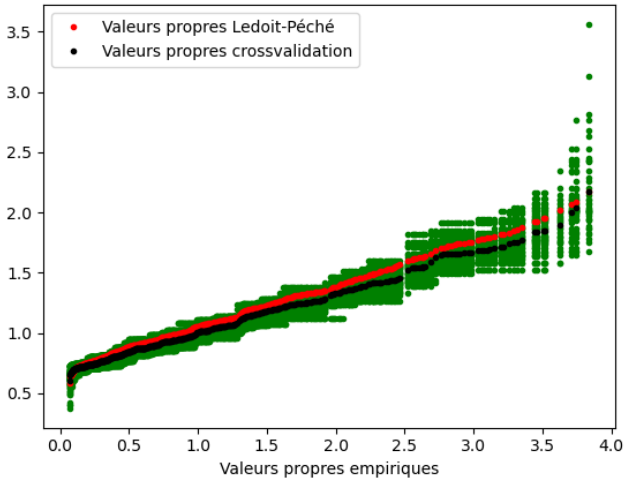

FIGURE 7 – Crossvalidation pour une matrice de covariance Inverse Wishart avec $n = 500$, $q = 0.5$ et $p = 0.25$ pour un nombre de 40 blocs

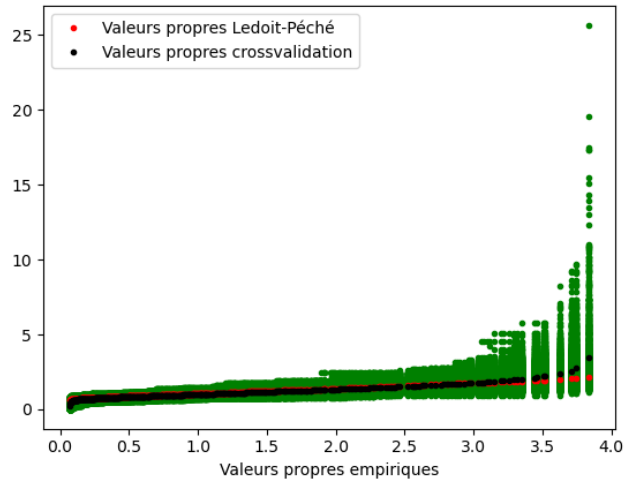

FIGURE 8 – Crossvalidation pour une matrice de covariance Inverse Wishart avec $n = 500$, $q = 0.5$ et $p = 0.25$ pour un nombre de 1000 blocs

On remarque par ces deux graphes que le nombre de blocs avec lequel on fait tourner la crossvalidation semble influer sur les valeurs propres de l'estimateur de crossvalidation et en particulier sur l'estimation des plus grandes valeurs propres.

Crossvalidation pour les matrices de crosscovariance

On peut aussi utiliser la crossvalidation sur des matrices de crosscovariance, il suffit d'adapter légèrement, prenons encore un bloc définit sur les données $[T1, T2]$:

— $(\boldsymbol{C_{XY}})_{\boldsymbol{in}}$ la matrice de crosscovariance empirique sur les données $[T1, T2]$ de vecteurs singuliers $\boldsymbol{u_{in}} = (u_{in}(1), ..., u_{in}(n))$ et $\boldsymbol{v_{in}} = (v_{in}(1), ..., v_{in}(n))$.

— $(\boldsymbol{C_{XY}})_{\boldsymbol{out}}$ la matrice de crosscovariance empirique sur $[0, T1] \cup [T2, T]$ de vecteurs propres $\boldsymbol{u_{out}} = (u_{out}(1), ..., u_{out}(n))$ et $\boldsymbol{v_{out}} = (v_{out}(1), ..., v_{out}(n))$

Sur ce bloc, l'estimateur des valeurs propres est le suivant :

$$\xi(\lambda_i) = u_{out}^T(i)\boldsymbol{C_{in}}v_{out}(i)$$

On peut alors reconstituer "une matrice de covariance" :

$$\boldsymbol{\Xi_k} = u_{out}\xi v_{out}^T$$

Comme dans le cas de la variance-covariance, l'estimateur de crosscovariance n'est alors que la moyenne de ces estimateurs sur les différents blocs :

$$\boldsymbol{\Xi} = \frac{1}{K}\sum_{k=1}^{K}\boldsymbol{\Xi_k}$$

Nous allons également prendre un exemple de crossvalidation pour l'estimation de crosscovariance. Ici, les données sont générées à partir d'une matrice de covariance Inverse Wishart de taille $n = 600$ de paramètres $q = 0.5$, et $p = 0.25$. Pour obtenir la matrice de crosscovariance à estimer, on extrait le coin en haut à droite de l'Inverse Wishart de dimensions $n_1 = 500$ et $n_2 = 100$. On a $T = \lfloor (n_1 + n_2)/q \rfloor = 1000$ et on fait tourner la technique pour 40 blocs. En vert on a tracé les valeurs singulières $\xi(\lambda_i) = u_{out}^T(i)\boldsymbol{C_{in}}v_{out}(i)$, en noir les valeurs singulières de l'estimateur de crossvalidation $\boldsymbol{\Xi}$ et en rouge les valeurs singulières nettoyées par l'algorithme de Ledoit-Péché en fonction des valeurs singulières empiriques sur l'ensemble des données :

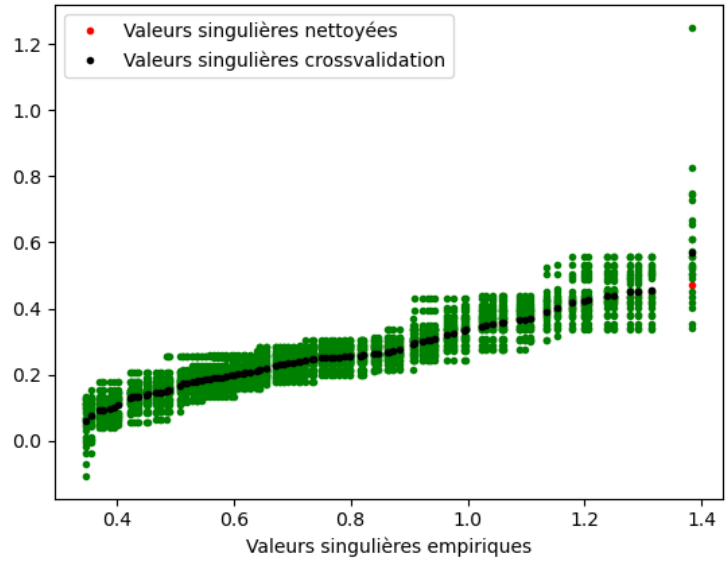

FIGURE 9 – Crossvalidation pour une matrice de crosscovariance Inverse Wishart avec pour dimensions (n_1, n_2) avec $n_1 = 500$, $n_2 = 100$, $q = 0.5$ et $p = 0.25$ pour un nombre de 25 blocs

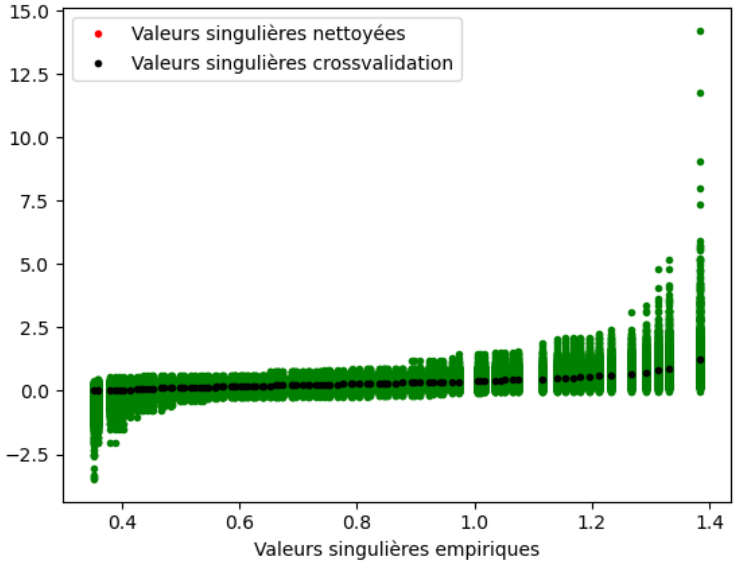

FIGURE 10 – Crossvalidation pour une matrice de crosscovariance Inverse Wishart avec pour dimensions (n_1, n_2) avec $n_1 = 500$, $n_2 = 100$, $q = 0.5$ et $p = 0.25$ pour un nombre de 1200 blocs

Comme dans le cas de l'estimation de covariance par crossvalidation, on remarque que certains nombres de blocs semblent plus avantageux que d'autres pour faire tourner l'algorithme de crossvalidation. En particulier choisir un nombre de bloc égal à T semble augmenter la largeur des spectres de crossvalidation (les valeurs singulières vertes sur les graphes). Nous regarderons plus en détail la question du nombre de blocs optimal dans les parties suivantes.

5.2 Nombre de blocs optimal pour la crossvalidation

Il semblerait qu'il existe un nombre de blocs optimal pour faire tourner la crossvalidation. Par exemple, nous avons tracé l'erreur de crossvalidation par rapport une vraie covariance $\boldsymbol{C}$ Inverse Wishart :

$$\text{Erreur de crossvalidation} = \| \boldsymbol{C_{cross}} - \boldsymbol{C} \|_F^2 \text{ , avec } \| \, . \, \|_F \text{ la norme de Frobenius}$$

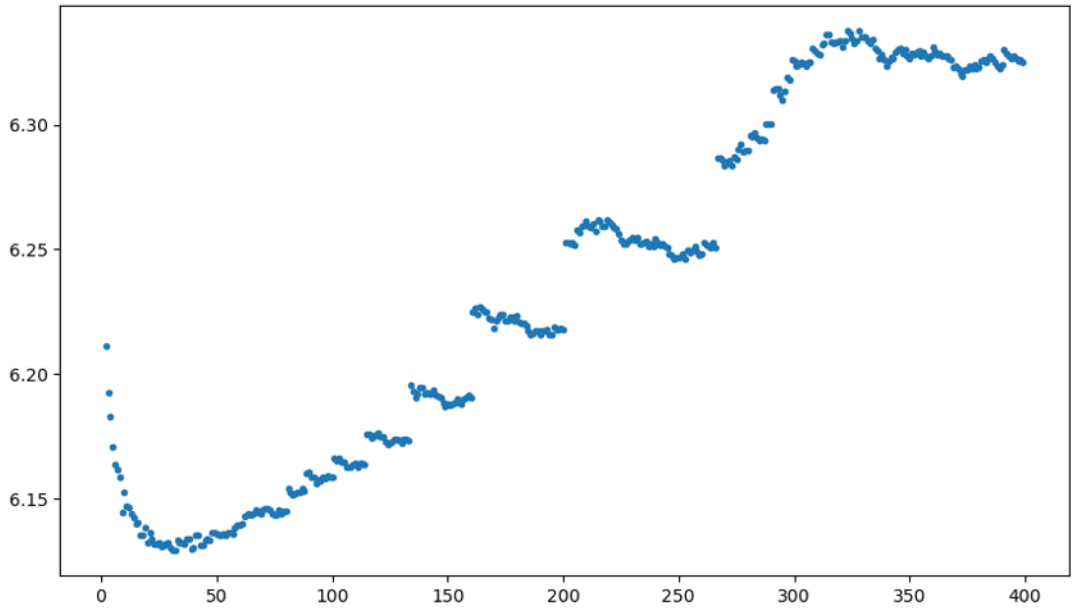

FIGURE 11 – Erreur de crossvalidation par rapport à la vraie Inverse Wishart, $n = 200$, $q = 0.5$ et $p = 1.5$, en fonction du nombre de blocs

Ce graphe nous montre une erreur en loi puissance qui admet un minimum. On observe également des paliers qui sont dûs au fait qu'on prenne une partie entière pour la taille des blocs. Par exemple, pour le graphe précédent où $T = 400$, on a que la taille des blocs utilisée pour la crossvalidation ($\lfloor T/k \rfloor$) reste constante pour $k = 50, 51, ..., 79$.

Nous pensons que le nombre de blocs pour faire tourner la crossvalidation se situe autour de $\sqrt{n}$ où n représente la taille de la matrice de covariance à estimer. Nous allons tenter d'étudier cette hypothèse empiriquement pour les matrices de covariance Inverse Wishart et dans le cas particulier Identité.

5.3 Crossvalidation pour une matrice de covariance Inverse Wishart

Ici, notre vraie matrice de covariance $\boldsymbol{C}$ est une Inverse Wishart. Ce cas est intéressant car le shrinkage linéaire est optimal et le coefficient de shrinkage a une expression analytique simple [17] [11], on rappelle son expression pour $\boldsymbol{C}$ de paramètres n, $q = \frac{n}{T}$ et p (où $p = \frac{q^*}{1-q^*}$ et $q^* = \frac{n}{T^*}$:

$$\mathbb{E}(\boldsymbol{C}|\boldsymbol{E}) = r\boldsymbol{E} + (1-r)\boldsymbol{Id} \quad \text{avec} \quad r = \frac{T}{T + T^* - n - 1}$$

Nous générons ensuite les données observées à partir d'une normale multivariée de moyenne nulle et de covariance $\boldsymbol{C}$. Nous voudrions trouver une expression du nombre de blocs optimal en fonction des paramètres de notre problème : n (la taille de la matrice de covariance), p (le paramètre de l'Inverse Wishart) et q (le quotient de n par le nombre d'observations T).

Dans un premier temps, nous nous sommes limités aux diviseurs de T comme nombre de blocs pour la crossvalidation. L'idée est que prendre des nombres de blocs qui ne divisent pas T revient à changer la valeur de q et peut créer du bruit dans nos observations. En premier lieu, on souhaite déterminer la dépendance en n, quand p et q sont fixés.

Dans ce but nous avons mené 30 simulations pour chaque valeur de n avec une même matrice inverse Wishart avec les paramètres suivants :

- $p = 0.3$
- $q = 0.5$
- n allant de 400 à 740 avec un pas de 20

Nous avons commencé par regarder le nombre de blocs optimaux pour la crossvalidation en prenant comme critère d'optimalité celui qui minimise l'erreur au sens de la norme de Frobenius par rapport à la vraie matrice de covariance. Nous avons également regardé le nombre de blocs qui minimise l'erreur par rapport à l'estimateur oracle.
On rappelle que les vecteurs propres de l'oracle sont ceux de l'estimateur empirique et que ses valeurs propres sont :

$$\xi(\lambda_k) = \widehat{\lambda}_k = u_k^T \boldsymbol{C} u_k$$

où $\boldsymbol{C}$ est la matrice de variance-covariance à estimer et $(u_1, .., u_n)$ sont les vecteurs propres de l'estimateur empirique et $(\lambda_1, .., \lambda_n)$ les valeurs propres empiriques.

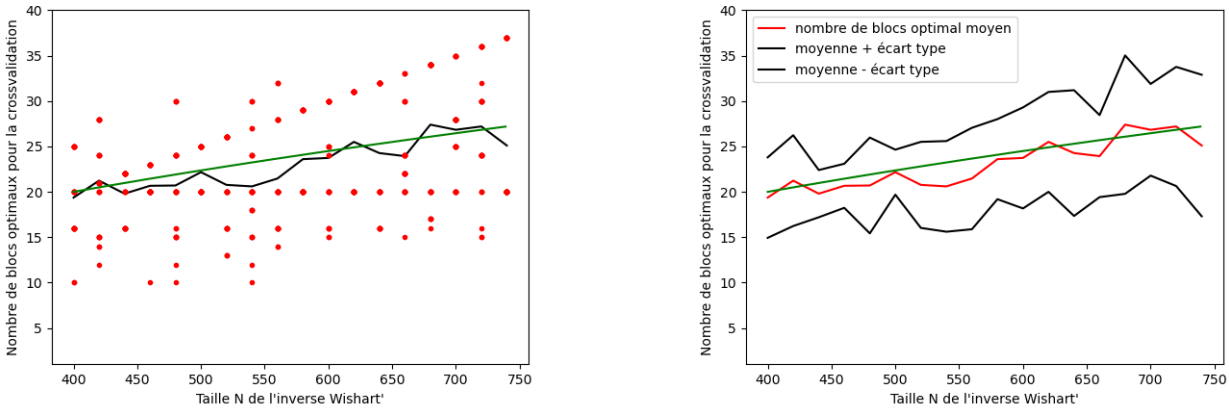

FIGURE 12 – Dépendance en n du nombre de blocs optimal pour la crossvalidation par rapport à une Inverse Wishart $p = 0.42$, $q = 0.5$

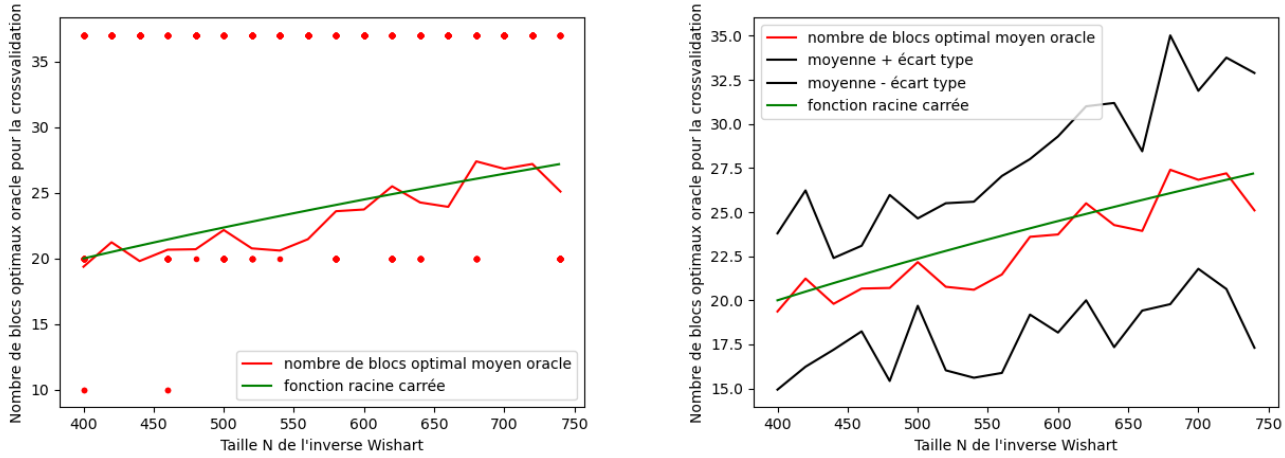

FIGURE 13 – Dépendance en n du nombre de blocs optimal pour la crossvalidation par rapport à l'estimateur oracle d'une Inverse Wishart $p = 0.42$, $q = 0.5$

Nous remarquons que le nombre optimal moyen est assez proche de $\sqrt{n}$ mais que des écarts-types importants demeurent lorsque l'on considère le nombre de blocs optimal qui minimise l'erreur par rapport à la vraie matrice de covariance ou par rapport à l'estimateur oracle. Il faut maintenant vérifier la variation d'erreur associée à ces nombres de blocs optimaux différents. Si les écart-types de l'erreur sont négligeables, on pourra considérer que $\sqrt{n}$ est une bonne approximation. On pourra également considérer que ces variations sont principalement dues aux arrondis numériques et au fait que comme on a considéré uniquement les diviseurs de T, on se prive de nombres de blocs proches de $\sqrt{n}$ ce qui accroit les écart-types puisqu'on oscille principalement sur les nombres de blocs proches de la racine de n.

Nous allons donc maintenant regarder la dépendance en n de l'erreur optimale de crossvalidation :

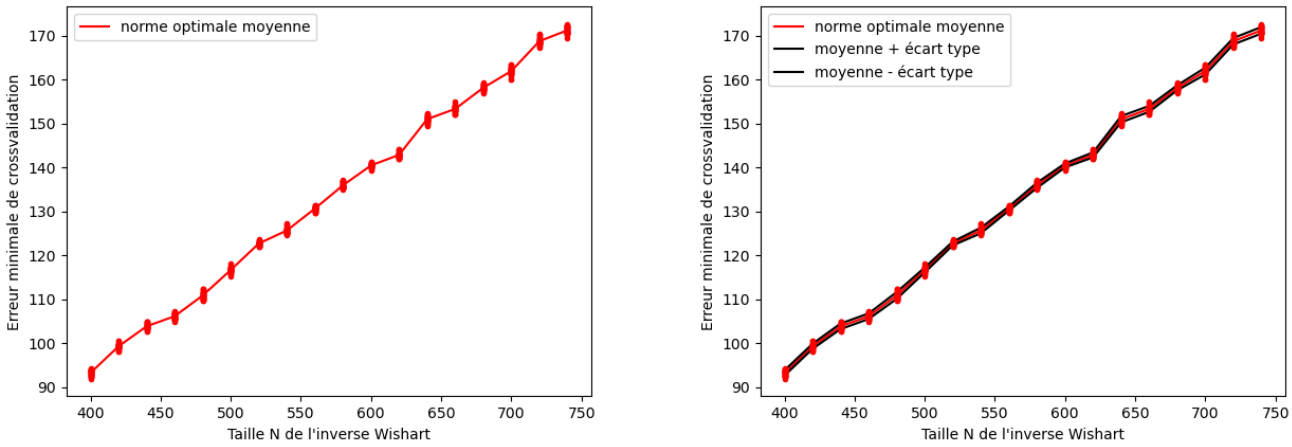

FIGURE 14 – Dépendance en n de l'erreur minimale pour la crossvalidation par rapport à une Inverse Wishart $p = 0.3$, $q = 0.5$

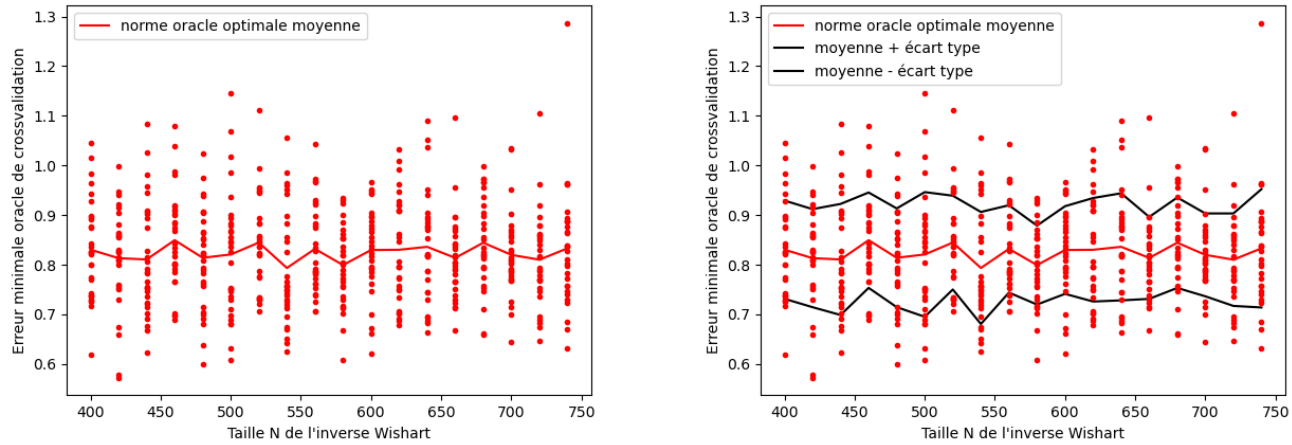

FIGURE 15 – Dépendance en n de l'erreur minimale pour la crossvalidation par rapport à l'estimateur oracle d'une Inverse Wishart $p = 0.3$, $q = 0.5$

L'erreur minimale de crossvalidation par rapport à la vraie Inverse Wishart est quasi-linéaire en n.

Nous nous sommes également intéressés à la dépendance en p du nombre de blocs optimal de crossvalidation et à la dépendance en p de l'erreur à n et q fixés.

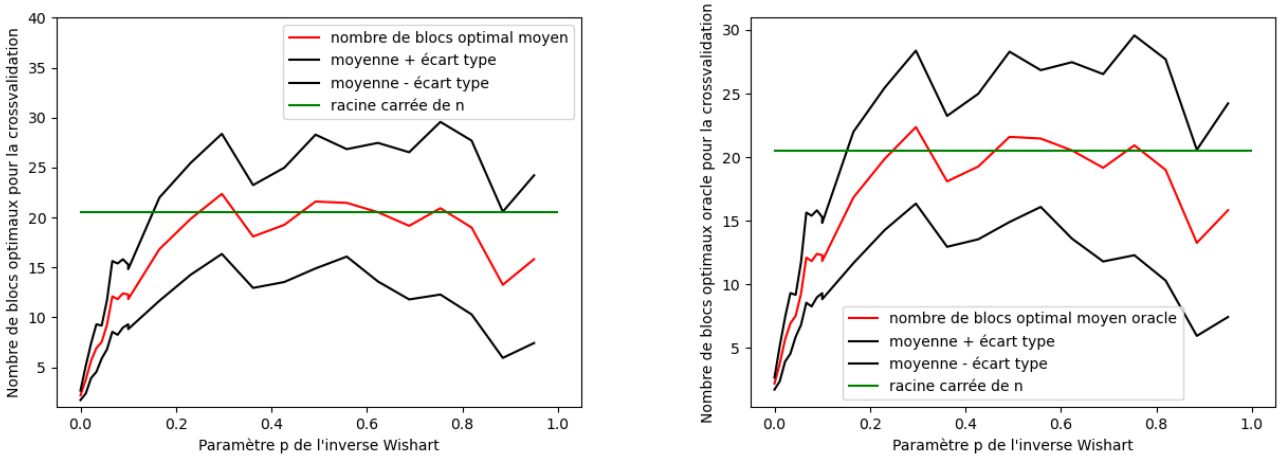

FIGURE 16 – Dépendance en p du nombre de blocs optimal pour la crossvalidation par rapport à la vraie Inverse Wishart (à gauche) et par rapport à l'oracle (à droite) $n = 420$, $q = 0.5$

Le nombre de blocs optimal pour la crossvalidation est identique que l'on considère l'erreur par rapport à la vraie matrice de variance-covariance ou par rapport à l'oracle. On remarque deux régimes :

- $p \in [0, 0.2]$ le nombre de blocs optimal croît pour atteindre $\sqrt{n}$
- $p \in [0.2, 1[$ le nombre de blocs optimal est en moyenne proche de la racine carrée de la taille de la matrice (la droite $y = \sqrt{420}$ en vert sur les graphes)

Nous avons remarqué que lorsque p est très grand (exemple : $q^* = 0.9$, $p = 9$), il arrive que le nombre de blocs optimal pour la crossvalidation soit égal à T. Les données financières ne sont pas stationnaires et les matrices de variance-covariance des actifs varient en fonction du temps [6]. Il se pourrait donc qu'on soit dans le cas de figure où le nombre optimal pour la crossvalidation est souvent égal à T pour des données financières.

5.4 Crossvalidation pour une matrice de covariance Identité

Nous générons des données gaussiennes à partir d'une matrice de covariance égale à l'identité ($C = Id$). Empiriquement nous trouvons une erreur par rapport à l'identité linéaire et croissante de K, le nombre de blocs choisi pour la crossvalidation. Ainsi nous trouvons $K = 2$ comme nombre optimal de blocs pour faire tourner la technique.

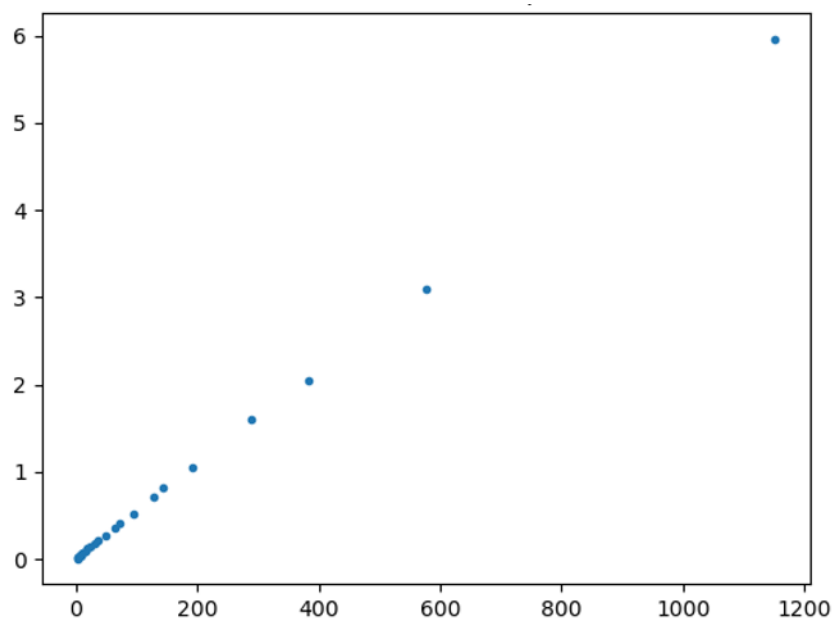

FIGURE 17 – Erreur de crossvalidation pour $\boldsymbol{C = Id}$ de taille $n = 576$, $q = 0.5$ en fonction de k nombre de blocs choisis

Nous avons tenté de prouver mathématiquement cette observation en faisant appel à la théorie des perturbations [17].

On se place dans le cas $\boldsymbol{C = Id}$. On génère donc des vecteurs à l'aide d'une loi normale multivariée de moyenne nulle et de matrice variance-covariance égale à $\boldsymbol{C}$. On note n (la taille de la matrice $\boldsymbol{C}$), T (le nombre de données générées), K (le nombre de blocs), T_B la taille de chaque bloc.

On suppose de plus que $K >> 1$ et $T_B << T$ et on définit les matrices empiriques suivantes :

- $\boldsymbol{E} = \frac{1}{T}\boldsymbol{H}\boldsymbol{H^T}$, la matrice empirique des observations sur tout l'échantillon
- $\boldsymbol{E_1} = \frac{1}{T_B}\boldsymbol{H_1}\boldsymbol{H_1^T}$, la matrice empirique des observations sur le premier bloc
- $\boldsymbol{E_{\overline{1}}} = \frac{1}{T-T_B}\boldsymbol{H_{\overline{1}}}\boldsymbol{H_{\overline{1}}^T}$, avec $(v_{\overline{1}}^l)_{1\leq l\leq N}$ une base de vecteurs propres orthonormée de $\boldsymbol{E_{\overline{1}}}$
- $\boldsymbol{E_{\overline{12}}} = \frac{1}{T-2T_B}\boldsymbol{H_{\overline{12}}}\boldsymbol{H_{\overline{12}}^T}$, la matrice empirique des observations sur l'échantillon privé des deux premiers blocs.

On note $\boldsymbol{\Xi_k}$ l'estimateur de crossvalidation sur le bloc numéro k ($1 \leq k \leq n$) :

$$\boldsymbol{\Xi_k} = \sum_{l=1}^{n} (v_{\overline{1}}^{l^T}\boldsymbol{E_1}v_{\overline{1}}^l)v_{\overline{1}}^l v_{\overline{1}}^{l^T}$$

L'estimateur final de crossvalidation n'étant que la moyenne sur l'ensemble des blocs :

$$\boldsymbol{\Xi} = \frac{1}{K}\sum_{k=1}^{K} \boldsymbol{\Xi_k}$$

À partir de là on pose :

- $\widetilde{\boldsymbol{E_1}} = \boldsymbol{E_1} - \boldsymbol{Id}$
- $\widetilde{\boldsymbol{\Xi_k}} = \sum_{l=1}^{n} (v_{\overline{1}}^{l^T} \widetilde{\boldsymbol{E_1}} v_{\overline{1}}^{l}) v_{\overline{1}}^{l} v_{\overline{1}}^{l^T}$
- $\widetilde{\boldsymbol{\Xi}} = \frac{1}{K}\sum_{k=1}^{K} \widetilde{\boldsymbol{\Xi_k}}$

Retirer l'identité permet d'avoir des matrices de covariance empiriques d'espérance nulle et cela simplifiera les calculs.

On rappelle que les les $\boldsymbol{\Xi_k}$ sont d'espérances égales, l'erreur de crossvalidation s'exprime alors comme l'espérance de la quantité suivante :

$$\frac{1}{n}Tr(\widetilde{\boldsymbol{\Xi}}^2) = \frac{1}{nK^2}\sum_{a,b} Tr(\widetilde{\boldsymbol{\Xi_1}}\widetilde{\boldsymbol{\Xi_b}}) = \frac{1}{nK}Tr(\widetilde{\boldsymbol{\Xi_1}}^2) + \frac{K^2-K}{nK}Tr(\widetilde{\boldsymbol{\Xi}}_1\widetilde{\boldsymbol{\Xi}}_2)$$

Le calcul est détaillé dans les annexes et fait appel à la théorie des perturbations, on se contente ici d'en donner les grandes lignes. Pour le premier terme $\mathbb{E}(Tr(\widetilde{\Xi}_1^2))$, on trouve :

$$\frac{1}{nK}\mathbb{E}(Tr(\widetilde{\boldsymbol{\Xi_1}}^2)) = \frac{2}{KT_B} = \frac{2}{T}$$

On va maintenant nous intéresser au second terme de l'erreur : $Tr(\widetilde{\boldsymbol{\Xi_1}}\widetilde{\boldsymbol{\Xi_2}})$.

Avec quelques manipulations simples on peut écrire :

$$\widetilde{\boldsymbol{E_1}} = \frac{T-2T_B}{T-T_B}[\boldsymbol{E_{\overline{12}}} + \frac{T_B}{T-2T_B}\boldsymbol{E_2}]$$

$$\widetilde{\boldsymbol{E_2}} = \frac{T-2T_B}{T-T_B}[\boldsymbol{E_{\overline{12}}} + \frac{T_B}{T-2T_B}\boldsymbol{E_2}]$$

$\frac{T-2T_B}{T-T_B}$ n'est qu'une constante de normalisation qui n'a pas d'influence sur les vecteurs propres. On pose $\alpha = \frac{T_B}{T-2T_B}$, la théorie des perturbations nous donne une approximation des vecteurs propres de $\boldsymbol{E_1}$ et de $\boldsymbol{E_2}$ dans la base des vecteurs propres de $\boldsymbol{E_{\overline{12}}}$ si α est suffisamment petit [17].

Après quelques calculs, on trouve :

$$Tr(\widetilde{\boldsymbol{\Xi_1}}\widetilde{\boldsymbol{\Xi_2}}) = 4\alpha^2 \sum_{l}\sum_{m_1\neq l} \frac{1}{T_B}\frac{1}{(\lambda_l^{\overline{12}} - \lambda_{m_1}^{\overline{12}})(\lambda_l^{\overline{12}} - \lambda_{m_1}^{\overline{12}})} + o(\alpha^2)$$

Toutefois ce terme n'est pas convergeant en général pour des matrices à valeurs réelles. Ce calcul nous apprend donc que l'hypothèse des petites perturbations des vecteurs propres dans le cadre de la crossvalidation n'est pas valable pour notre problème. Les vecteurs propres $(v_{\overline{1}}^{l})$ et $(v_{\overline{2}}^{l})$ sont donc fortement perturbés par rapport aux vecteurs propres $(v_{\overline{12}}^{l})$, et ce, même quand le nombre de blocs est grand.

5.5 Erreur de crossvalidation, cas Inverse Wishart

Nous pensons que l'erreur de crossvalidation par rapport à la vraie matrice de covariance pourrait être la somme de deux erreurs indépendantes : l'erreur du shrinkage optimal (linéaire dans le cas Inverse Wishart) par rapport à la vraie matrice de covariance plus une seconde erreur entre l'estimateur de crossvalidation et le shrinkage optimal.

Dans la figure qui suit nous avons effectué trente simulations pour chaque nombre de blocs et tracé l'erreur de crossvalidation par rapport au shrinkage optimal au sens de la norme de Frobenius :

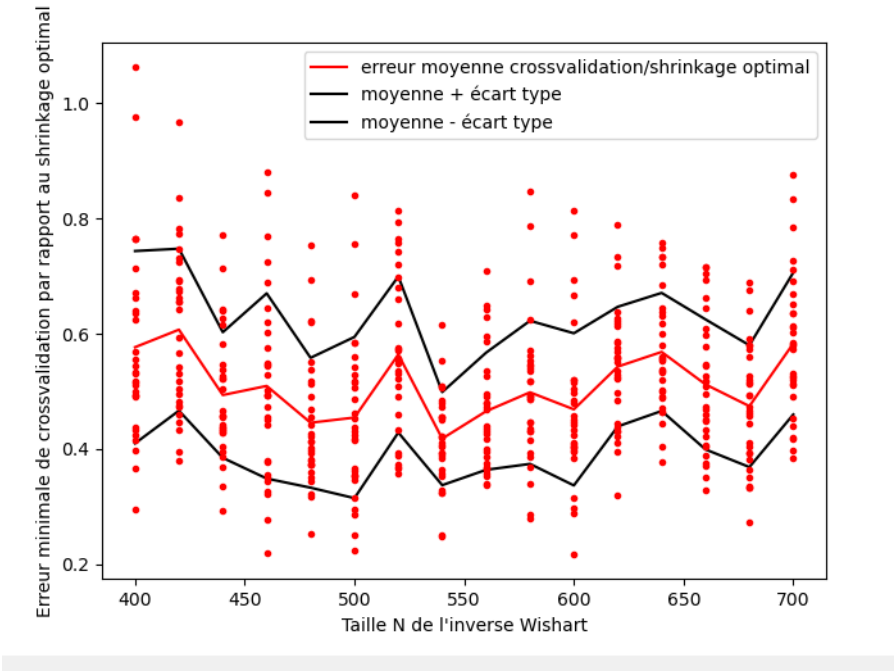

FIGURE 18 – Erreur de crossvalidation par rapport au shrinkage linéaire en focntion de n pour une Inverse Wishart avec $q = 0.5$, $p = 0.5$

De ce graphe, il semble que l'erreur moyenne de crossvalidation par rapport au shrinkage linéaire ne dépende pas de n.

Nous avons également tracé la somme des erreurs de la crossvalidation/shrinkage linéaire et shrinkage linéaire/vraie matrice pour voir si cela correspond à l'erreur de crossvalidation par rapport à la vraie matrice de covariance, ces deux erreurs semblent se superposer :

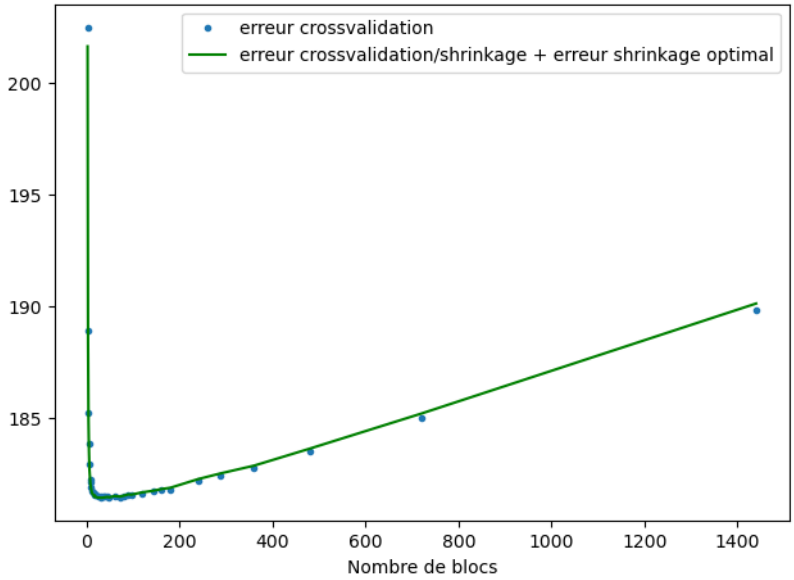

FIGURE 19 – Erreur de crossvalidation en fonction du nombre de blocs pour une Inverse Wishart $n = 720$, pour $p = 0.5$, $q = 0.5$

La forme de l'erreur de crossvalidation par rapport au shrinkage optimal semble être la somme d'un terme linéaire croissant du nombre de blocs et d'un terme en $\frac{1}{K}$, nous avons conjecturé une erreur de la forme :

$$f(K) = c_0 + \frac{c_1}{K} + c_2 K$$

où c_0 serait l'erreur du shrinkage optimal soit $c_0 = \frac{npq}{p+q}$, c_1 un terme en $n\sqrt{n}$ et c_2 en $\frac{1}{\sqrt{n}}$ pour n grand.

En dérivant, on aurait :

$$\frac{d}{dK} f(K) = -\frac{c_1}{K^2} + c_2$$

Et en résolvant $\frac{d}{dK} f(K) = 0$ on a

$$K_{opt} = \sqrt{\frac{c_1}{c_2}}$$

ce qui donnerait bien un terme en $\sqrt{n}$ pour n grand.

Nous avons tenté de calibrer cette formule sur les données, mais cette approximation n'est pas tout à fait satisfaisante étant trop pointue notamment pour les petites valeurs de n (~ 100). Aussi, il paraît étrange d'avoir des termes en $\sqrt{n}$ pour un calcul d'erreur qui correspond à une variance.

Nous pensons donc qu'une forme plus appropriée pour l'erreur serait la suivante :

$$f(K) = c_0 + \frac{c_1}{K^2} + c_2 \frac{K^2}{1 + c_3 K}$$

La division par $1+c_3K$ sert à assurer que le terme quadratique devienne linéaire par rapport au nombre de blocs lorsque celui-ci est grand.

Pour calibrer cette formule nous avons généré des données. Nous avons construit une base de données à cinq colonnes contenant les paramètres suivants :

- n
- p
- q
- k_blocs : tableau des diviseurs de T (1 exclu) qui serviront de nombre de blocs pour faire tourner la crossvalidation
- erreur_cross_shrinkage : tableau l'erreur entre l'estimateur obtenu par crossvalidation et le shrinkage linéaire optimal pour chaque nombre élément

Nous avons testé la crossvalidation pour les valeurs suivantes :

- $n \in [100, 150, 200, 250]$
- $q^* \in [0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9]$, on rappelle que $p = \frac{q^*}{1+q^*}$.
- $q \in [0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9]$

Enfin pour chaque triplet (n, p, q) on effectue 20 simulations. L'élaboration de ce tableau de résultats a demandé plus d'une cinquantaine d'heures de simulation par ordinateur. Une suite de ce projet serait d'analyser ces données.

6 Simulations avec des données synthétiques

Nous avons commencé par utiliser des données synthétiques simulées par ordinateur afin de tester nos algorithmes.

6.1 Première étape : choix de la matrice de variance-covariance

Une première question a été de générer des matrices de covariances. Au début on ne cherche pas nécessairement à être réalistes par rapport aux données financières donc la seule contrainte est d'obtenir une matrice définie positive. Une première réponse simple est de tirer aléatoirement les objets suivants :

- $\boldsymbol{X}$ une matrice orthogonale
- $\boldsymbol{\Omega}$ une matrice diagonale qui va contenir les valeurs propres de la covariance (ex : on peut les tirer uniformément entre 0.1 et 1)

La matrice de variance-covariance s'écrit alors simplement :

$$\boldsymbol{\Sigma} = \boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{\Omega}\boldsymbol{X}$$

.

Si l'on ne souhaite pas utiliser une fonction pré-codée pour tirer une matrice orthogonale, on peut tirer une matrice de Wigner $\boldsymbol{H}$ à l'aide de normales centrée réduites. Il suffit de l'ajouter à sa transposée puis de la diagonaliser. La matrice $\boldsymbol{V}$ des vecteurs propres sera alors une base orthonormée de vecteurs tirés selon la mesure de Haar. La mesure de Haar est l'équivalent de la mesure de Lebesgue dans $\mathbb{R}$ sur des espaces un peu plus généraux et est donc un équivalent de la loi uniforme. A priori nous n'avons pas de préférence sur les bases de vecteurs propres et donc tirer notre base de manière uniforme paraît pertinent.

Aussi nous allons mener des simulations sur différents cas particuliers de la matrice de covariance comme par exemple l'Identité ou l'Inverse-Wishart.

6.2 Estimateur de Ledoit-Peché pour les matrices de covariance

On dispose ici de T données à partir desquelles on calcule une matrice empirique de corrélation $\boldsymbol{E}$ des observations qui est de taille n. On note $q = \frac{n}{T}$. L'objectif est d'estimer au mieux la matrice de covariance des données que l'on ne connaît pas a priori. Voici l'algorithme de Ledoit-Péché de nettoyage des valeurs propres [3] :

Algorithm 1 Algorithme de Ledoit-Péché

Require: $\boldsymbol{E}$ matrice de corrélation de valeurs propres $(\lambda_k)_{1\leq k\leq n}$, $q = \frac{n}{T}$
for $k \in [|1, n|]$ **do**

$$z_k = \lambda_k - i\frac{1}{\sqrt{n}}$$

$$s_k(z_k) = \frac{1}{n}\sum_{j=1, j\neq k}^{n} \frac{1}{z_k - \lambda_j}$$

$$\xi_k^{RIE} = \frac{\lambda_k}{|1 - q + qz_k s_k(z_k)|^2}$$

end for
Facultatif : Régression isotonique sur les $(\xi_k^{RIE})_{1\leq k\leq n}$

On effectue une régression isotonique car on s'attend à avoir des valeurs propres nettoyées ξ qui soient monotones des valeurs propres de l'estimateur empirique. La régression isotonique se définit de la manière suivante [17] :

$$\hat{\xi}_k = \text{argmin}(\sum_{k=1}^{n}(\hat{\xi}_k - \xi_k)^2)$$

avec : $\hat{\xi}_1 \leq \hat{\xi}_2 \leq ... \leq \hat{\xi}_{n-1} \leq \hat{\xi}_n$

Voici un exemple de nettoyage des valeurs propres d'une matrice de covariance empirique tirée de manière gaussienne centrée à partir d'une Inverse Wishart :

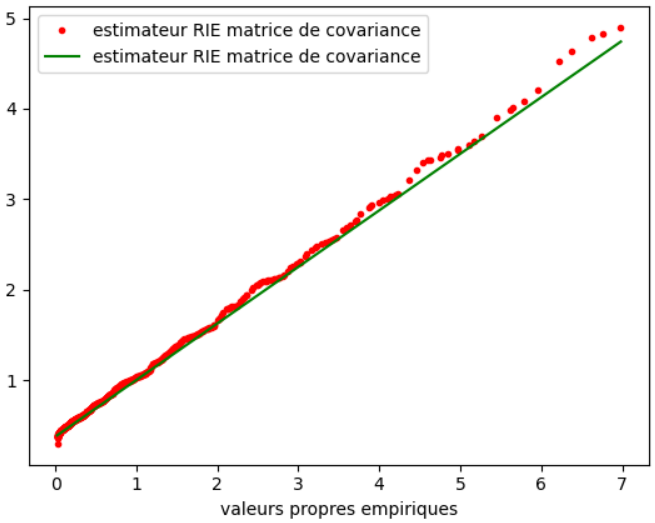

FIGURE 20 – Valeurs propres nettoyées d'une Inverse Wishart de paramètres $n = 600$, $q = 0.6$ et $p = 0.3$ par rapport aux valeurs propres empiriques

On remarque à nouveau que le spectre empirique est plus large que le spectre de la vraie Inverse Wishart et donc que la pente de notre estimateur est inférieure à 1.

Nous avons voulu observer l'erreur d'estimation de la matrice nettoyée $\boldsymbol{C}^*$ par rapport à la vraie Inverse Wishart $\boldsymbol{C}$:

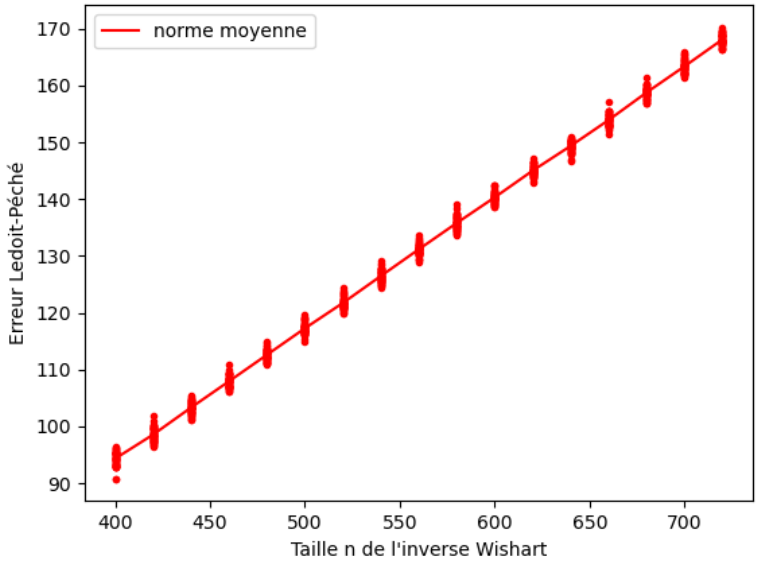

FIGURE 21 – Dépendance en n de l'erreur de l'estimateur de Ledoit-Péché par rapport à la vraie Inverse Wishart, $p = 0.42$, $q = 0.5$

Nous avions préalablement calculé la pente théorique de l'erreur qui correspond à l'espérance de la trace normalisée de l'erreur :

$$\mathbb{E}(\tau(\boldsymbol{C} - \boldsymbol{C}^*)^2) = \frac{pq}{p+q}$$

avec $r = \frac{T*}{T+T*+n-1} \simeq \frac{p}{p+q}$.

Ici, $p = 0.42$ et $q = 0.5$. La pente théorique de l'erreur est $\tau(\boldsymbol{C} - \boldsymbol{C}^*)^2 \simeq 0.2308$ tandis que la pente observée est $\tau(\boldsymbol{C} - \boldsymbol{C}^*)^2 \simeq 0.23$ ce qui est cohérent. On observe bien une erreur linéaire en n la taille de la matrice de covariance.

Nous avons également voulu comparer l'efficacité de l'estimateur de Ledoit-Péché par rapport à l'estimateur empirique pour différentes matrices Inverse Wishart de tailles n. Nous avons donc généré 50 matrices de covariance pour chaque n. On a par ailleurs fixé : $p = 0.42$ et $q = 0.5$ comme précédemment.

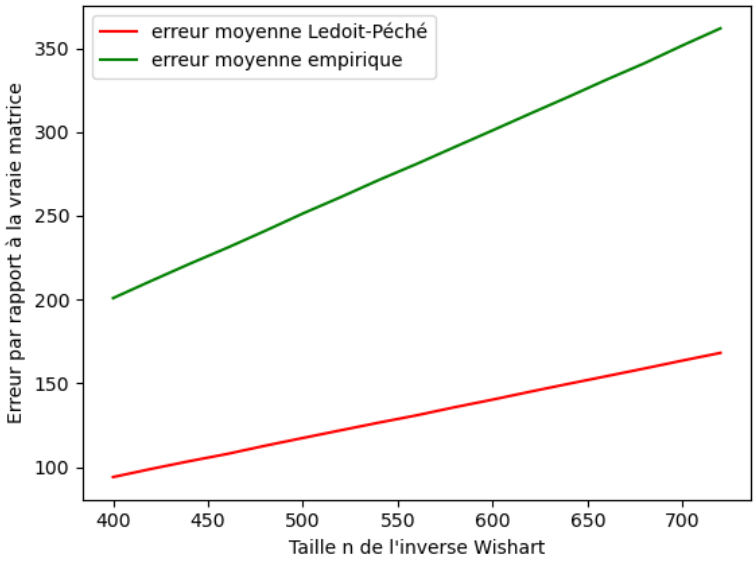

FIGURE 22 – Erreur de l'estimateur de Ledoit-Péché et de l'estimateur empirique par rapport à la vraie Inverse Wishart, $p = 0.42$, $q = 0.5$ en fonction de n

Nous avons effectué un autre exemple d'implémentation de l'estimateur de Ledoit-Péché sur une matrice de covariance où :

- la base de vecteurs propres est tirée de manière aléatoire selon la mesure de Haar
- les valeurs propres sont tirées de manière uniforme entre 0 et 3 et une valeur propre est tirée de manière uniforme entre $0.15n$ et $0.25n$ avec n la taille de la matrice de covariance. Le fait d'avoir une valeur propre bien plus grande que les autres est une caractéristique des matrices de covariance financières où la plus grande valeur propre est autour de $0.2n$ et représente le *market mode* [17] [18]. Il s'agit ici d'imiter un peu cette caractéristique des matrices financières mais nous reviendrons plus en détails sur le *market mode* dans la partie financière du rapport.

Voici un exemple des valeurs propres nettoyées pour une matrice de covariance de ce type de taille 600 :

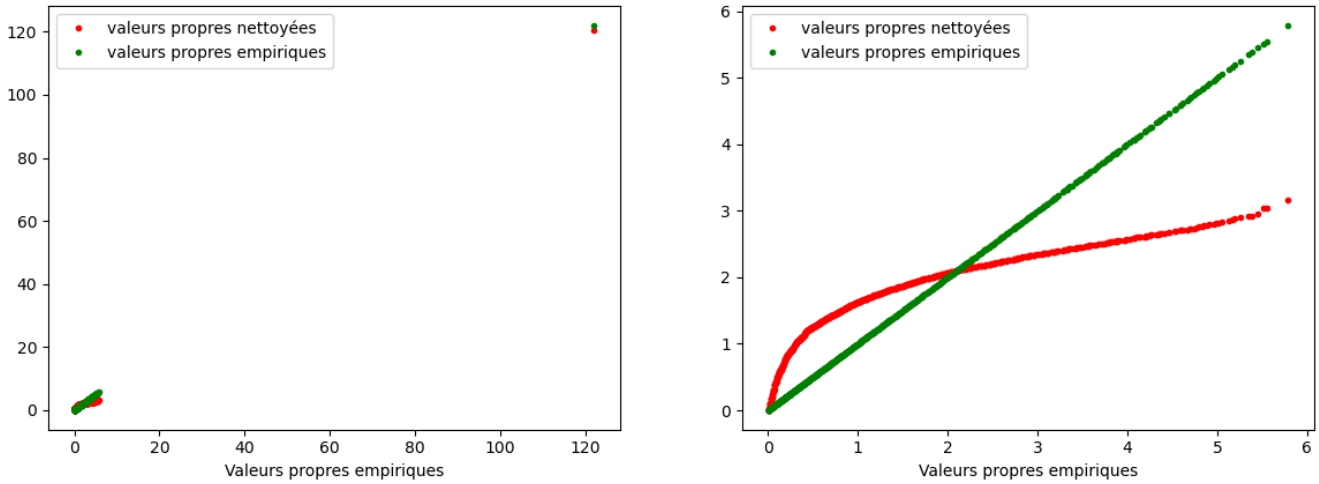

FIGURE 23 – Valeurs propres nettoyées pour une matrice de covariance de taille $n = 600$, $q = 0.5$ (dans le second graphe on a exclu la valeur propre maximale)

On remarque que le nettoyage n'est plus linéaire comme c'était le cas pour l'Inverse Wishart.

La simulation suit ensuite le même schéma que celle avec une Inverse Wishart, on fait 50 simulations pour chaque valeur de n (taille de la matrice de covariance) allant de 400 à 720 avec un pas de 20 et q fixé à 0.5. On trace donc l'erreur moyenne en fonction de n :

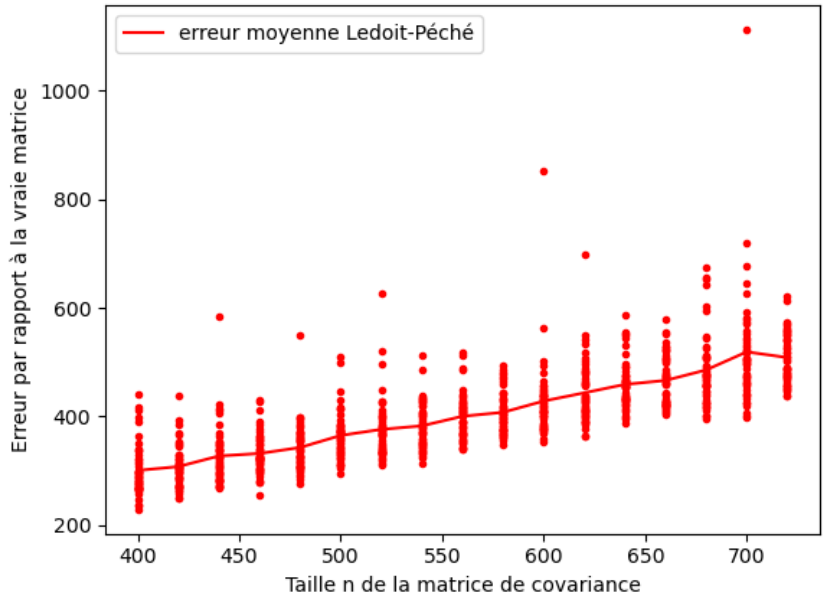

FIGURE 24 – Erreur de l'estimateur de Ledoit-Péché et de l'estimateur empirique par rapport à la vraie Inverse Wishart, $p = 0.42$, $q = 0.5$ en fonction de n

On peut également comparer à l'erreur moyenne de l'estimateur empirique :

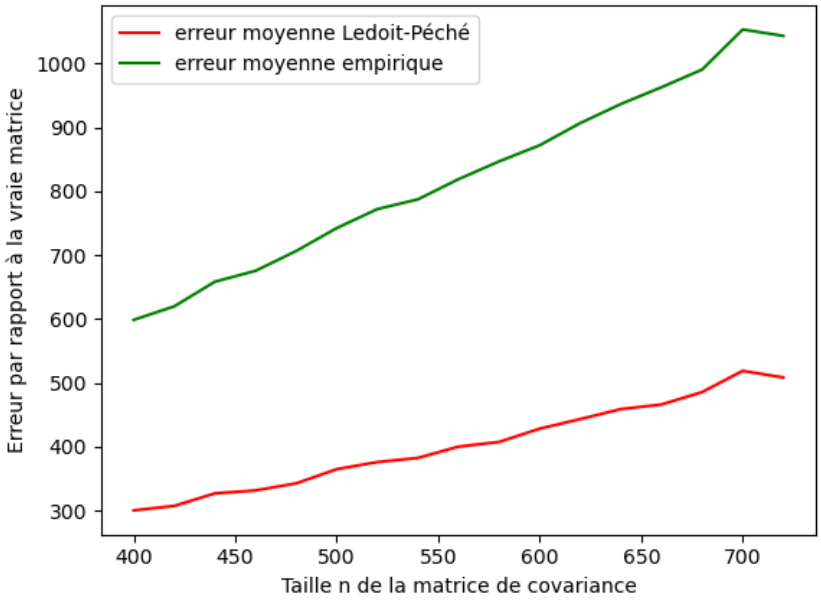

FIGURE 25 – Erreur de l'estimateur de Ledoit-Péché et de l'estimateur empirique par rapport à la vraie matrice de covariance, $q = 0.5$, en fonction de n

Nous avons présenté ici un exemple d'implémentation pour l'estimateur de Ledoit-Péché des matrices de variance-covariance. On peut voir l'implémentation de la transformée de Stieltjes comme un produit de convolution entre la densité empirique des valeurs propres la matrice à estimer et d'un noyau de Cauchy. Nous allons poser quelques notations pour illustrer ce propos [13] [17].

On rappelle que la densité empirique d'une matrice aléatoire de valeurs propres $(\lambda_k)_{1\leq k\leq n}$ est une somme de Dirac :

$$\rho_{emp}(x) = \sum_{k=1}^{n} \delta_{\lambda_k}$$

À partir de là on définit une densité lissée par le noyau ρ_S avec S pour "*smoothed*", il s'agit du produit de convolution entre la densité empirique et le noyau :

$$\rho_S(x) = \rho_{emp} \star K_\eta(x) = \frac{1}{n} \sum_{k=1}^{n} K_{\eta_k}(x - \lambda_k)$$

Avec K_η un noyau de largeur η bien choisi.

Le noyau de Cauchy est le choix le plus standard dans nos simulations mais le noyau de Wigner peut aussi se révéler intéressant.

Définition 8. *On appelle noyau de Cauchy la densité de probabilité suivante pour $u \in \mathbb{R}$ [17] :*

$$K_\eta^C(u) = \frac{1}{\pi} \frac{\eta}{u^2 + \eta^2}$$

η est appelée largeur de la densité.

Définition 9. *On appelle noyau de Wigner la densité suivante [17] :*

$$K_\eta^W(u) = \begin{cases} \frac{\sqrt{4\eta^2-u^2}}{2\pi\eta^2} \; si \; -2\eta \leq u \leq 2\eta \\ 0 \; si \; |u| > 2\eta \end{cases}$$

Les densités de Cauchy et de Wigner ont l'allure suivante :
Nous pouvons alors réécrire l'algorithme de Ledoit-Péché avec un noyau de Wigner [17] :

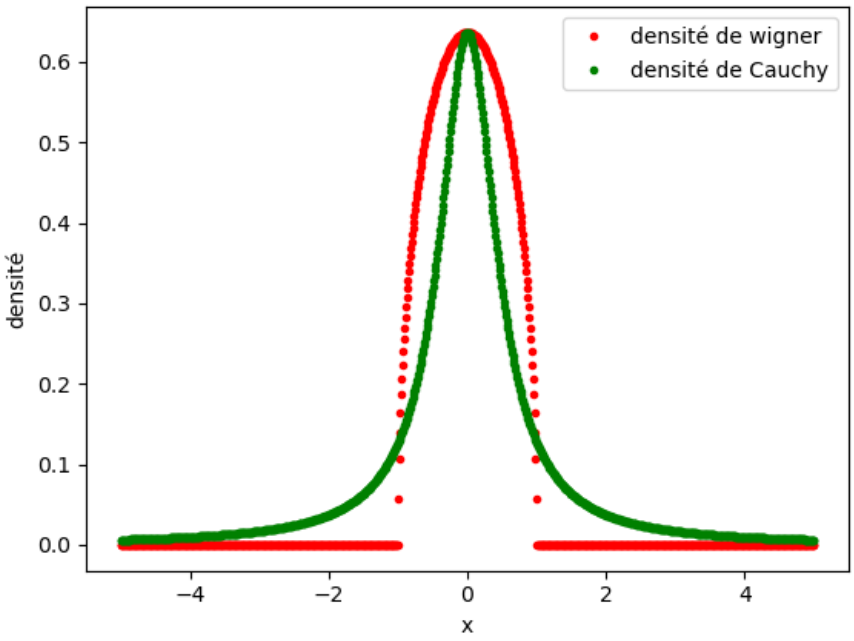

FIGURE 26 – Densités de Wigner et de Cauchy pour $\eta = 0.5$

Algorithm 2 Algorithme de Ledoit-Péché avec un noyau de Wigner

Require: $\boldsymbol{E}$ matrice de corrélations de valeurs propres $(\lambda_k)_{1\leq k\leq n}$, $q = \frac{n}{T}$
for $k \in [|1, n|]$ **do**

$$z_k = \lambda_k - i\frac{1}{\sqrt{n}}$$

$$s_k(z_k) = \frac{1}{n}\sum_{j=1, j\neq k}^{n} \frac{z-\lambda_k}{2\eta_k^2}\left(1 - \sqrt{1 - \frac{4\eta_k^2}{(z-\lambda_k)^2}}\right)$$

$$\xi_k^{RIE} = \frac{\lambda_k}{|1 - q + qz_k s_k(z_k)|^2}$$

end for
Facultatif : Régression isotonique sur les $(\xi_k^{RIE})_{1\leq k\leq n}$

Voici un exemple avec une covariance dont les vecteurs propres sont tirés selon la mesure de Haar et avec des valeurs propres tirées uniformément entre 0 et 3 et une grande

valeur propre entre $0.15n$ et $0.25n$. À q fixé à 0.5 on a tracé l'erreur d'estimation moyenne avec 30 matrices générées aléatoirement pour chaque valeur de n.

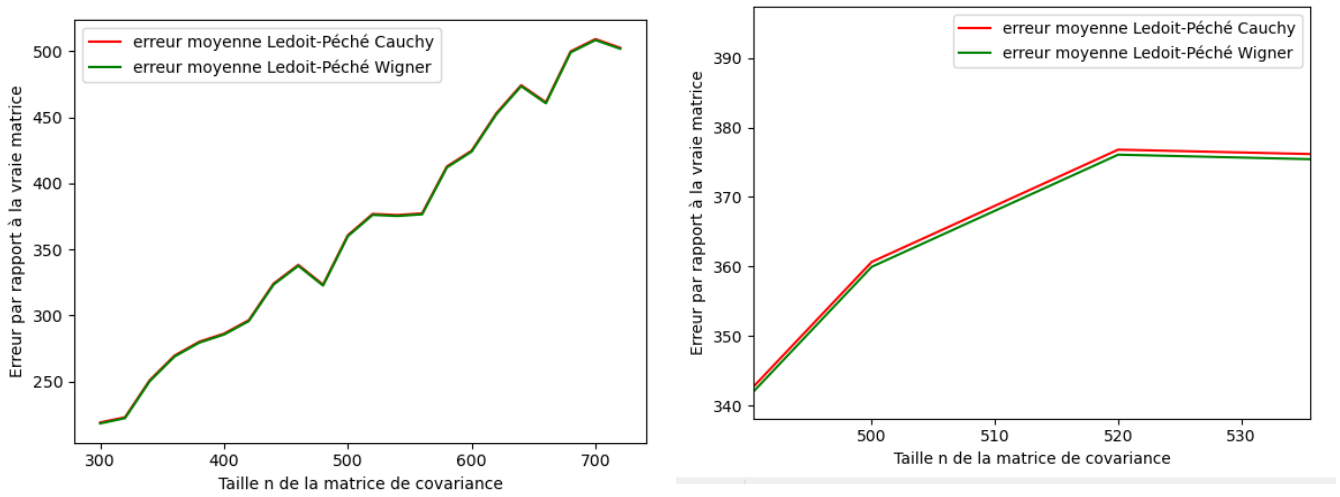

FIGURE 27 – Dépendance en n de l'erreur de nettoyage avec noyau de Cauchy et noyau de Wigner d'une matrice de covariance avec $q = 0.5$ (le graphe de droite est un zoom du premier)

On observe que pour ce type de matrice utiliser un noyau de Wigner plutôt qu'un noyau de Cauchy nous permet d'obtenir une erreur d'estimation légèrement inférieure, par exemple pour $n = 600$, on a une erreur moyenne de $427,7$ avec le noyau de Cauchy qui devient 427.0 pour le noyau de Wigner. Ces différences pourraient être plus exacerbées pour d'autres types de matrices de covariance.

Par exemple, il est connu que les queues de distribution des rendements financiers sont épaisses [6]. Il se pourrait donc que dans le cas de données financières, un noyau de Wigner permette une estimation plus efficace qu'un noyau de Cauchy.

6.3 Comparaison entre le nettoyage et la crossvalidation

Nous avons voulu appliquer le nettoyage de Ledoit-Péché et la crossvalidation simultanément afin de voir si une des deux techniques offrait des meilleurs résultats. Voici un exemple pour des matrices Inverse Wishart dont nous avons fait varier la taille pour p et q fixés. Pour l'erreur de crossvalidation nous avons retenu l'erreur minimale donnée par le nombre de blocs optimal (environ $\sqrt{n}$ car $0.2 \leq p \leq 1$).

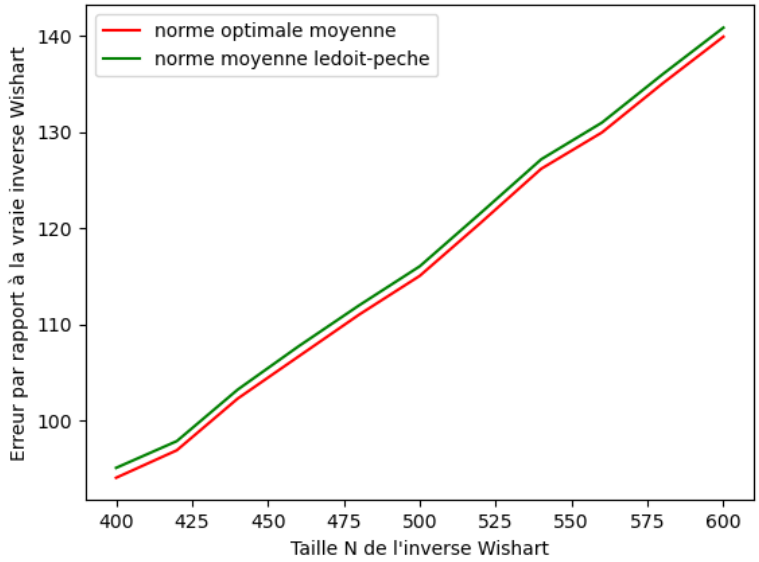

FIGURE 28 – Erreur de crossvalidation et du nettoyage de Ledoit-Péché en fonction de n, pour $p = 0.42$, $q = 0.5$

Dans ce cas, nous remarquons que l'implémentation numérique est légèrement meilleure avec la crossvalidation optimale en moyenne. Même si nous pensons que le nombre de blocs optimal est autour de $\sqrt{n}$, nous avons précédemment vu que ce nombre est assez volatile selon les simulations. Nous avons voulu avoir une idée de la largeur de la zone où le nombre de blocs permet d'avoir une erreur comparable ou meilleure que le nettoyage de Ledoit-Péché, nous avons donc pris un exemple de matrice Inverse-Wishart de taille $n = 720$ et avec $p, q = 0.5$:

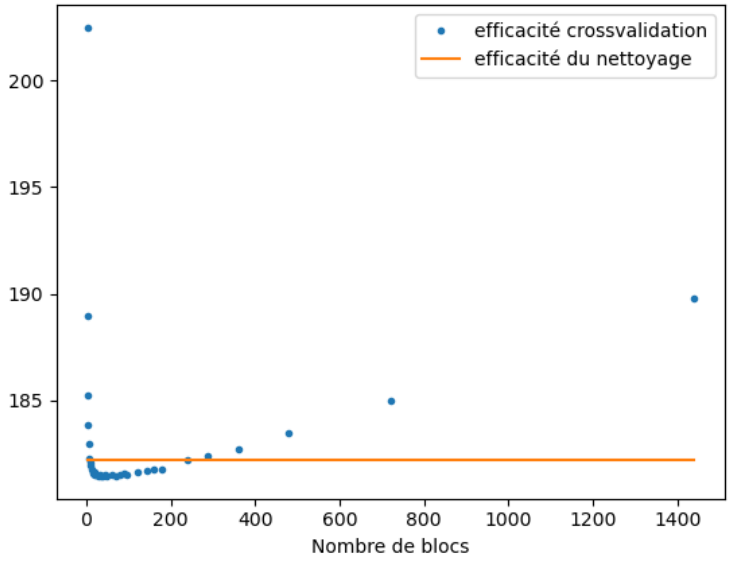

FIGURE 29 – Erreur de crossvalidation et du nettoyage de Ledoit-Péché en fonction du nombre de blocs pour une Inverse Wishart de paramètres $n = 720$, $p = 0.5$, $q = 0.5$

Nous avons tracé ce même graphe pour différentes valeurs de n, p et q et il semble que l'ensemble des blocs où la crossvalidation a une efficacité similaire au nettoyage de Ledoit-Péché est suffisamment large pour pouvoir prendre $\sqrt{n}$ blocs pour appliquer la crossvalidation. Ce nombre de blocs peut demander des adaptations pour des valeurs de p petites (inférieures à 0.2) ou des valeurs de q grandes (supérieures à 0.8).

Nous pensons que la crossvalidation permet de retrouver l'erreur théorique de l'estimateur de Ledoit-Péché. Cette conjecture serait intéressante dans le sens où la crossvalidation ne nécessite quasiment aucune hypothèse sur les données de départ.

6.4 Portefeuille de Markowitz avec une covariance Inverse Wishart

L'objectif ici est de développer un exemple d'utilisation de la théorie du portefeuille optimal de Markowitz combinée avec différents estimateurs de matrices de covariances. On souhaite comparer l'efficacité des estimateurs sur un exemple.

Nous prenons un exemple où l'on tire aléatoirement une matrice Inverse Wishart de de paramètres $q = 0.5$ et $p = 0.25$ et de taille $n \in [50, 100, 150, ..., 700]$ que l'on appelle $\mathbf{\Sigma}$. Pour chaque taille n fixée on génère T vecteurs $X(t) \in \mathbb{R}^n$ pour $t = 1, ..., T$ à l'aide d'une normale multivariée centrée et de matrice variance-covariance $\mathbf{\Sigma}$. Les $X(t)$ sont supposés indépendants.

On pose la matrice des données observées qui pourrait être en finance des rendements d'actifs observés quotidiennement (n serait le nombre d'actifs et T le nombre de jours de cotations observé) :

$$\boldsymbol{X} = (X(1), X(2), ..., X(T))$$

La matrice de variance-covariance empirique est la suivante (quitte à retrancher la moyenne, on considère que les données sont centrées) :

$$\boldsymbol{E} = \frac{1}{T}\boldsymbol{X}\boldsymbol{X}^{\boldsymbol{T}}$$

On appelle $\boldsymbol{E_{cleaned}}$ la matrice empirique nettoyée par l'algorithme de Ledoit-Péché. On a tiré de manière aléatoire un vecteur g qui représente l'espérance des rendements pour chaque actif et nous nous sommes fixés un rendement minimal de portefeuille espéré $\mathcal{G} = 10\%$.

On peut alors calculer les portefeuilles suivants à l'aide de la théorie de Markowitz :

- le portefeuille empirique :
$$\pi_{emp} = \mathcal{G}\frac{g^T\boldsymbol{E^{-1}}}{g^T\boldsymbol{E^{-1}}g}$$
- le portefeuille nettoyé :
$$\pi_{cleaned} = \mathcal{G}\frac{g^T\boldsymbol{E_{cleaned}^{-1}}}{g^T\boldsymbol{E_{cleaned}^{-1}}g}$$
- le "vrai" portefeuille optimal :
$$\pi_{opt} = \mathcal{G}\frac{g^T\mathbf{\Sigma^{-1}}}{g^T\mathbf{\Sigma^{-1}}g}$$

On va considérer les risques suivants :

- le risque du portefeuille empirique :
$$R_{emp} = \pi_{emp}^T\mathbf{\Sigma}\pi_{emp}$$
- le risque du portefeuille nettoyé :
$$R_{cleaned} = \pi_{cleaned}^T\mathbf{\Sigma}\pi_{cleaned}$$
- le risque optimal de portefeuille au sens de Markowitz :
$$R_{opt} = \pi_{opt}^T\mathbf{\Sigma}\pi_{opt}$$

Pour chaque $n \in [50, 100, 150, ..., 700]$ avec $q = 0.5$ et $p = 0.25$ nous avons généré 50 fois une matrice de variance-covariance Inverse Wishart de paramètres n et p puis généré des données et déterminé chacun des portefeuilles précédents ainsi que le risque associé. Nous avons tracé dans un graphe les risques moyens en fonction de n :

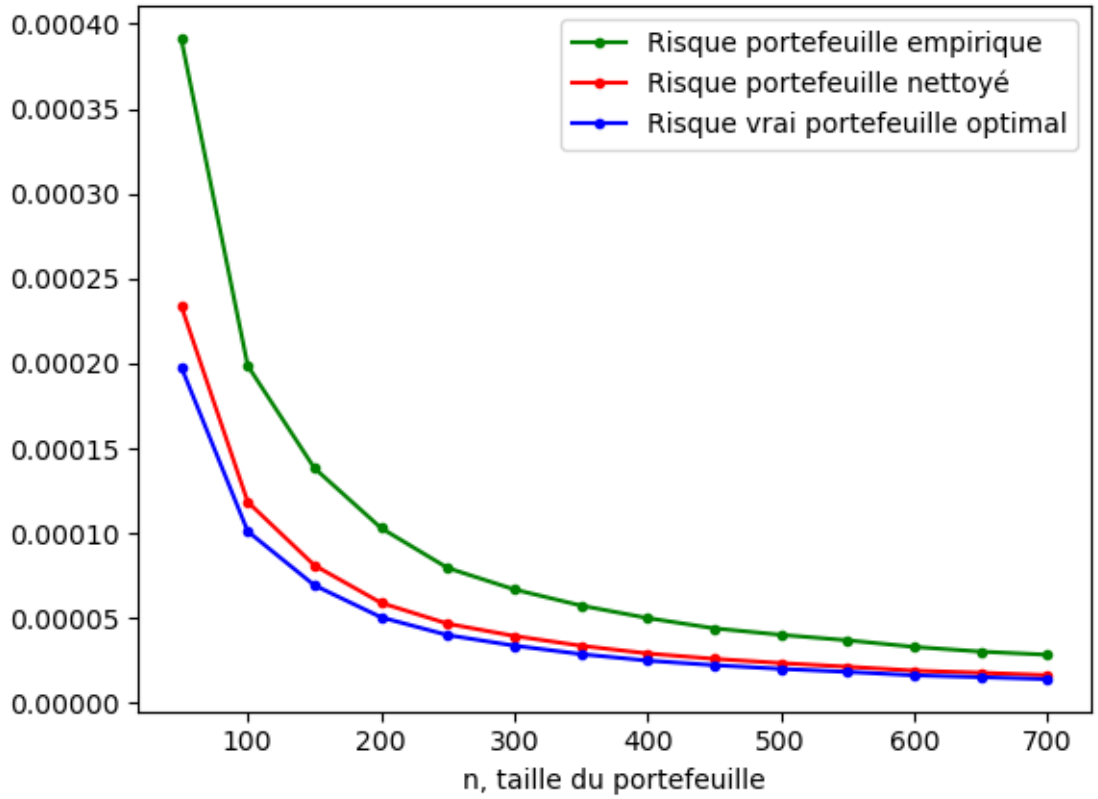

FIGURE 30 – Risque moyen des portefeuilles de Markowitz générés grâce aux estimateurs empirique, nettoyé et à la vraie matrice de variance-covariance Inverse Wishart de paramètres $p = 0.25$, $q = 0.5$ en fonction de n la taille du portefeuille

On remarque que l'estimateur empirique donne un portefeuille sous-optimal. Autrement dit dont on sous-estime le risque. Considérons la vraie matrice de variance-covariance $\mathbf{\Sigma}$ et notons $\lambda_1 \geq ... \geq \lambda_n$ ses valeurs propres et $(u_1, ..., u_n)$ ses vecteurs propres. Le vecteur u_i pour un $1 \leq i \leq n$ représente un portefeuille de norme 1 (au sens de la norme L^2) dont le risque est λ_i. Or le spectre de la matrice empirique $\boldsymbol{E}$ est plus large que le spectre de la vraie covariance $\mathbf{\Sigma}$, $\boldsymbol{E}$ a donc des valeurs propres plus petites que λ_n et d'autres plus grandes que λ_1. Ainsi en appliquant la formule du portefeuille de Markowitz avec $\boldsymbol{E}$ comme estimateur on va mettre moins de poids sur les grandes valeurs propres de $\boldsymbol{E}$ car on les croît plus risquées que ce qu'elles sont en réalité et au contraire on va mettre plus de poids que si on connaissait $\mathbf{\Sigma}$ sur les petites valeurs propres car on les croît moins risquées qu'elles ne le sont en vrai. On se retrouve donc systématiquement avec un portefeuille plus risqué que le portefeuille optimal que l'on trouverait si l'on connaissait $\mathbf{\Sigma}$.

L'algorithme de Ledoit-Péché applique un shrinkage non linéaire sur les valeurs propres de $\boldsymbol{E}$ et donc réduit la largeur du spectre empirique. On remarque que cela permet de se rapprocher significativement du risque optimal.

6.5 Estimateur de crosscovariance

Comme pour l'estimateur de Ledoit-Péché, nous donnons ici l'algorithme de l'estimateur des matrices de crosscovariance, cet algorithme provient de l'article *Optimal cleaning for singular values of crosscovariance matrices* [2] :

Algorithm 3 Algorithme de nettoyage de crosscovariance

Require: $\boldsymbol{X} \in \mathbb{R}^{n\times T}$, $\boldsymbol{Y} \in \mathbb{R}^{p\times T}$ avec $n \leq p$
Calculer : $\boldsymbol{C_{XY}} = \frac{1}{T}\boldsymbol{XY^T}$, $\boldsymbol{C_X} = \frac{1}{T}\boldsymbol{XX^T}$ et $\boldsymbol{C_Y} = \frac{1}{T}\boldsymbol{YY^T}$
$u = (u_l)_{1\leq l\leq n}$ et $v = (v_l)_{1\leq l\leq p}$ les vecteurs singuliers de $\boldsymbol{C_{XY}}$ et $s = (s_l)_{1\leq l\leq n}$ les valeurs singulières de $\boldsymbol{C_{XY}}$
for $l \in [|1, n|]$ **do**

$$Coeff_{A,l} = u_l^T \boldsymbol{C_X} u_l \text{ et } Coeff_{B,l} = v_l^T \boldsymbol{C_Y} v_l$$

$$Coeff_{B,[n+1:p]} = \sum_{l=n+1}^{p} v_l^T \boldsymbol{C_Y} v_l$$

end for
for $k \in [|1, n|]$ **do**

$$z_k = s_k + i(npT)^{-1/12}$$

Calculer H, A et B de la manière suivante :

$$H(z_k) = \frac{1}{T}\sum_{l=1}^{n} \frac{s_l^2}{z^2 - s_l^2}, A(z_k) = \frac{1}{T}\sum_{l=1}^{n} \frac{Coeff_{A,l}}{z^2 - s_l^2}$$

$$B(z_k) = \frac{1}{T}(\sum_{l=1}^{n} \frac{Coeff_{B,l}}{z^2 - s_l^2} + z^{-2} Coeff_{B,[n+1:p]})$$

$$\Theta(z_k) = z_k^2 \frac{A(z_k)B(z_k)}{1 + H(z_k)} \text{ et } L(z_k) = 1 - \frac{1}{1 + H(z_k) - \Theta(z_k)}$$

$$s_k^{cleaned} = s_k \times (\frac{ImL}{ImH})_+ \text{ où } x_+ = max(x, 0)$$

end for
Facultatif : Régression isotonique sur les $(s_k^{cleaned})_{1\leq k\leq n}$

Nous avons pris un exemple de matrice de variance-covariance $\boldsymbol{\Sigma}$ Inverse Wishart de paramètres $n = 720$, $p = 0.8$ et $q = 0.5$, d'où $T = 1440$. Rappelons les données du problème : on définit donc $Z = \begin{pmatrix} X \\ Y \end{pmatrix} \in \mathbb{R}^{n_1+n_2}$ un vecteur aléatoire tel que $\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}(0, \boldsymbol{\Sigma})$ avec $\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{C} \\ \boldsymbol{C^T} & \boldsymbol{B} \end{pmatrix}$.

On cherche à estimer le coin en haut à droite $\boldsymbol{C}$ de dimensions (n_1, n_2) avec $n_1 = 500$ et $n_2 = 100$. On appelle $\boldsymbol{C_{XY}} = \frac{1}{T}\boldsymbol{XY^T}$ la matrice empirique de crosscovariance à laquelle on va appliquer l'algorithme précédent :

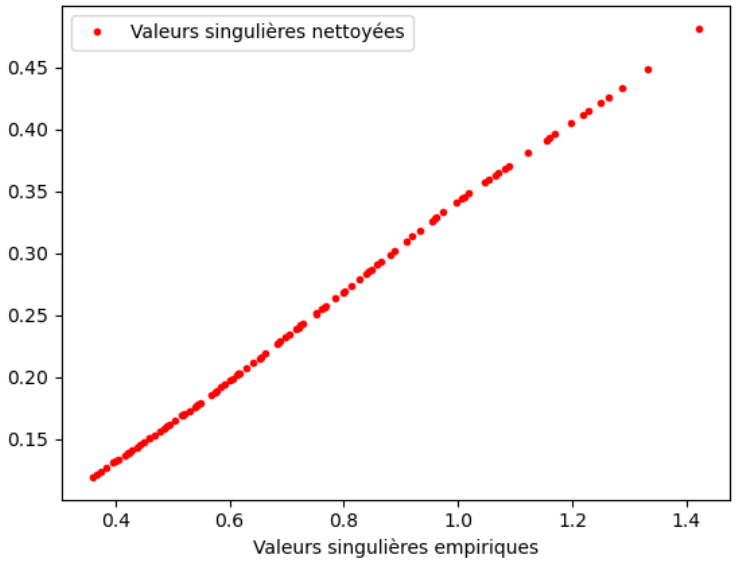

FIGURE 31 – Valeurs singulières nettoyées en fonction des valeurs singulières empiriques d'une matrice de crosscovariance Inverse Wishart de paramètres $n_1 = 500$, $n_2 = 100$, $p = 0.8$ et $q = 0.5$

Il se trouve que dans le cas où la loi de la matrice $\boldsymbol{\Sigma}$ est invariante par rotation (ce qui est le cas pour une Inverse Wishart), appliquer l'algorithme de crosscovariance à $\boldsymbol{C_{XY}}$ et appliquer l'algorithme de Ledoit-Péché à $\boldsymbol{E} = \frac{1}{T}\boldsymbol{ZZ^T}$ puis prendre le coin en haut à droite de dimension (n_1, n_2) donne le même estimateur de crosscovariance. Nous allons présenter un cas dans la partie suivante où l'estimateur de crosscovariance ne donne pas les mêmes résultats que Ledoit-Péché.

6.6 Modèle linéaire, estimation de crosscovariance

Dans cette modélisation $R = (R_1, ..., R_n)$ va jouer le rôle des rendements d'une collection de futures, matières premières, obligations, etc. On va simuler des "rendements d'actions" qui s'exprimeront comme une combinaison linéaire des rendements des futures à laquelle on ajoutera un bruit gaussien.

Même s'il ne s'agit ici que d'une modélisation avec des données générées par ordinateur par une loi normale multivariée, la recherche de facteurs explicatifs est une problématique importante en finance. On pourrait par exemple citer le modèle à facteurs de Fama et French [8].

Nous définissons donc nos rendements d'actions au temps $t \in \{0, 1, ..., T\}$ de la manière suivante :
Pour $t \in \{0, 1, ..., T\}$,

$$\forall i \in \{1, 2, ..., n\}, r_i^t = \sum_{j=0}^{p} \beta_j^i R_j^t + \epsilon_i$$

On suppose de plus que $p \geq n$, c'est-à-dire que le nombre d'actions considéré est plus grand que le nombre de futures.

Nous avons généré les données de la manière suivante :

- Les variances des futures $(R_j)_{1 \leq j \leq n}$ ont été tirées avec la valeur absolue d'une loi normale $\mathcal{N}(1, 1)$. Nous avons choisi des variances différentes parce qu'il n'y a aucune raison que de véritables futures aient la même volatilité. Les $(R_j)_{1 \leq j \leq n}$ sont indépendants, on définit $\boldsymbol{\Omega}$ la matrice diagonale contenant leurs variances.

- $\boldsymbol{\beta} = (\beta_j^i)_{1 \leq i \leq p, 1 \leq j \leq n}$, $1 \leq j \leq n$) : tirés avec une normale centrée de variance $\frac{1}{n+p}$. On note $\boldsymbol{\beta}$ la matrice des (β_j^i) de taille $p \times n$.

- ϵ=$(\epsilon_i)_{1 \leq i \leq n}$ sont des bruits gaussiens indépendants, on les suppose centrés et indépendants des $(R_j)_{1 \leq j \leq n}$. On fixe leur variance de manière à avoir autant de variance dans le bruit que dans la combinaison linéaire de futures :

$$Var(\epsilon_i) = Var(\sum_{j=0}^{n} \beta_j^i R_j^t)$$

On va désormais réécrire ce modèle sous forme matricielle. Dans ce modèle, la vraie matrice de crosscovariance est $\boldsymbol{C_{vraie}}$ telle que :

$$\forall (i, j) \in \{1, .., p\} \times \{1, ..., n\}, C_{vraie}[i, j] = \beta[i, j] * Var(R(j))$$

où $Var(.)$ représente la variance.

Ainsi la matrice de covariance de ce modèle, $\boldsymbol{\Sigma}$ est de la forme :

$$\begin{pmatrix} \boldsymbol{\Omega} & \boldsymbol{\Omega\beta} \\ \boldsymbol{\beta^T\Omega} & \boldsymbol{\beta^T\Omega\beta + "bruit"} \end{pmatrix}$$

Remarque : Passer de $\boldsymbol{\beta^T\Omega\beta}$ à $\boldsymbol{\beta^T\Omega\beta + "bruit"}$ revient juste à multiplier par deux la diagonale de la matrice $\boldsymbol{\beta^T\Omega\beta}$:

$$\boldsymbol{\beta^T\Omega\beta + "bruit"} = \boldsymbol{\beta^T\Omega\beta + diag(\beta^T\Omega\beta)}$$

Nous avons mené plusieurs simulations pour ce modèle en fixant $q = 0.5$ et en faisant varier n et p tout en gardant la contrainte $n+p = 600$. Pour chaque triplet (n, p, q) nous avons généré 30 matrices de variance-covariance $\boldsymbol{\Sigma}$ et pour chacune de ces simulations, nous avons calculé l'erreur au sens de la norme de Frobenius par rapport à la vraie matrice de crosscovariance $\boldsymbol{C_{vraie}}$ avec l'estimateur empirique, l'estimateur de Ledoit-Péché (qui consiste à appliquer l'algorithme sur la matrice de variance-covariance empirique puis à conserver le coin en haut à droite de dimensions (n, p)) et avec de l'estimateur de crosscovariance. Nous avons tracé les erreurs moyennes en fonction de p (on a alors $n = 600 - p$) :

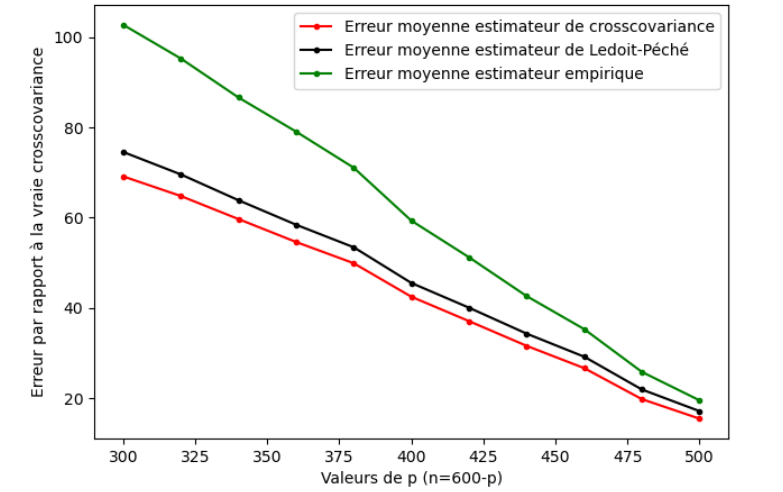

FIGURE 32 – Erreur par rapport à la vraie matrice de crosscovariance avec différents estimateurs pour $q = 0.5$ et $n + p = 600$ fixés en prenant différentes valeurs de (n, p)

Nous remarquons que nous sommes dans un cas où l'estimateur de crosscovariance est plus efficace que l'estimateur de Ledoit-Péché. Cela tient au fait qu'on ne peut pas considérer $\boldsymbol{\Sigma}$ comme invariante par rotation. Nous pensons qu'il pourrait en être de même sur les données financières dans le cas où l'on regarderait les corrélations entre des rendements d'actions et de futures.

7 Application aux données financières

7.1 Estimation de risque sur des données du SBF120

Nous avons mené une expérience avec des données publiques du site Yahoo Finance en téléchargeant les prix de clôture quotidiens des 120 entreprises du SBF120 de 2002 à Août 2021. Le SBF120 couvre les secteurs majeurs de l'économie française. Même si la composition du SBF120 a évolué, nous avons choisi de conserver les actions présentes dans le SBF120 en Août 2021. Nous ne tenons donc pas compte des sorties/entrées et des éventuelles fusions et acquisitions qui auraient pu avoir lieu sur les 20 dernières années.

À partir de ces données, on cherche à estimer une matrice de covariance de taille 120×120 sur une période de 2 ans. En effet, en retirant les week-ends et les quelques jours fériés sans cotations, on a environ 252 jours de cotations par an ce qui nous donne un paramètre $q \sim 0.2$. Au départ nous avions mené cette simulation sur des données du CAC40 mais la taille réduite 40×40 nous paraissait petite pour obtenir une bonne efficacité pour nos estimateurs. Nous avons donc préféré le SBF120 pour palier ce problème.

Nous allons donc tester l'algorithme de nettoyage de Ledoit-Péché ainsi que la technique de crossvalidation en faisant changer le nombre de blocs. Ici, nous n'avons pas accès à une "vraie matrice de covariance" donc nous prendrons comme matrice à estimer la covariance réalisée sur l'année suivante, on note cette matrice $\boldsymbol{C}$.

On note $\boldsymbol{E}$ la matrice empirique de covariance sur les deux année et $\lambda_1, ..., \lambda_n, u_1, ..., u_n$ ses valeurs propres et vecteurs propres ($n \leq 120$).

L'oracle est alors le suivant :

$$\xi(\lambda_k) = u_k^T \boldsymbol{C} u_k$$

Nous avons voulu estimer la covariance des actifs des quatre premiers mois de 2019 à partir des données de l'année 2018.

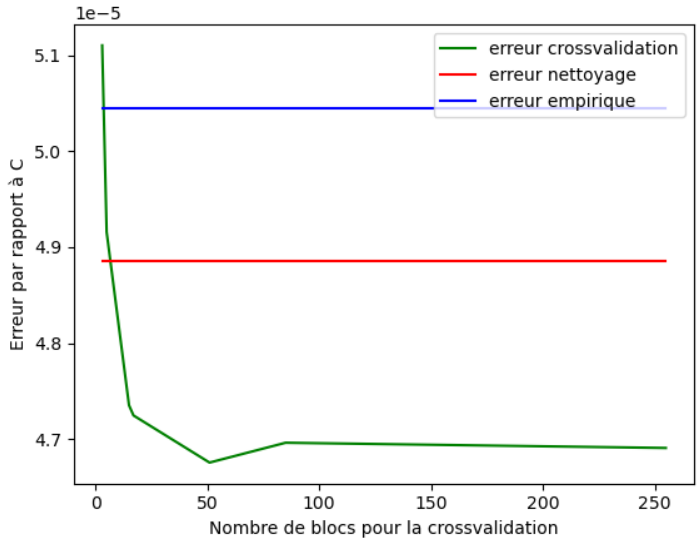

FIGURE 33 – Erreur d'estimation de la covariance sur les quatre premiers mois de 2019 à l'aide des données de 2018

Ce graphe nous permet de voir la nécessité de normaliser les rendements des actifs avant d'appliquer les différents algorithmes d'estimation. Les erreurs et les valeurs propres des estimateurs sont tellement infimes que l'on ne peut pas affirmer que les différences d'efficacité ne sont pas dues à un artifice numérique. Une autre raison qui justifie la normalisation des rendements est que les estimateurs sont construits pour des données centrées et de même variance.

Il y a plusieurs possibilités pour normaliser des rendements. Par exemple, nous pouvons normaliser les rendements par la volatilité des actifs calculée par une moyenne mobile sur une année [17] [18].

Pour cela on note :

- σ_t la volatilité d'un actif au jour t calculée par une moyenne mobile simple sur 252 jours de trading (soit environ une année) :

$$\sigma_{jour}(t) = \sqrt{252} \times \sigma_{annee}(t), \text{ avec } \sigma_{annee}(t) = \frac{\sum_{i=t-253}^{t-1}(r_i - \overline{r_i})^2}{252 - 1}$$

- $\overline{r_t}$ la moyenne des rendements d'un actif sur 252 jours de trading, où :

$$\overline{r_t} = \frac{1}{252} \sum_{i=t-253}^{t-1} r_i$$

Le rendement normalisé est alors :

$$r_{normalise}(t) = \frac{r_t - \overline{r_t}}{\sigma_t}$$

Aussi il paraît plus naturel de considérer la matrice de corrélation plutôt que la matrice de covariance afin d'avoir des matrices de norme fixée à 1.

Si on note la matrice de covariance :

$$\boldsymbol{Cov} = (C_{ij})_{1 \leq i,j \leq n}$$

Alors la matrice de corrélation n'est que :

$$\boldsymbol{Corr} = (C'_{ij})_{1 \leq i,j \leq n}, \text{ où } C'_{ij} = \frac{C_{ij}}{\sqrt{C_{ii}}\sqrt{C_{jj}}} = \frac{C_{ij}}{\sigma_i \sigma_j}$$

Autrement dit, on se contente de diviser chaque terme ij par les volatilités des actifs i et des actifs j.

Si on conserve le facteur $\sqrt{252}$ dans la volatilité, on aura besoin de transformer nos estimateurs en matrices de corrélation tandis que si on normalise directement par σ_{annee} on obtient des matrices de variance-covariance déjà proches de matrices de corrélations. Les éléments diagonaux sont déjà très proches de 1.

Nous avons commencé par une moyenne mobile simple pour la volatilité mais dans le cas de données financières, donc non stationnaires, la moyenne mobile exponentielle pourrait être plus efficace. L'avantage de la moyenne exponentielle est qu'elle met plus de poids

sur les données récentes.

Nous avons normalisé les rendements par la volatilité historique obtenue par une moyenne simple sur une année. Voici le spectre de la matrice de corrélation empirique calculée sur les années 2017 et 2018 soit $q = \frac{n}{T} \sim 0.25$:

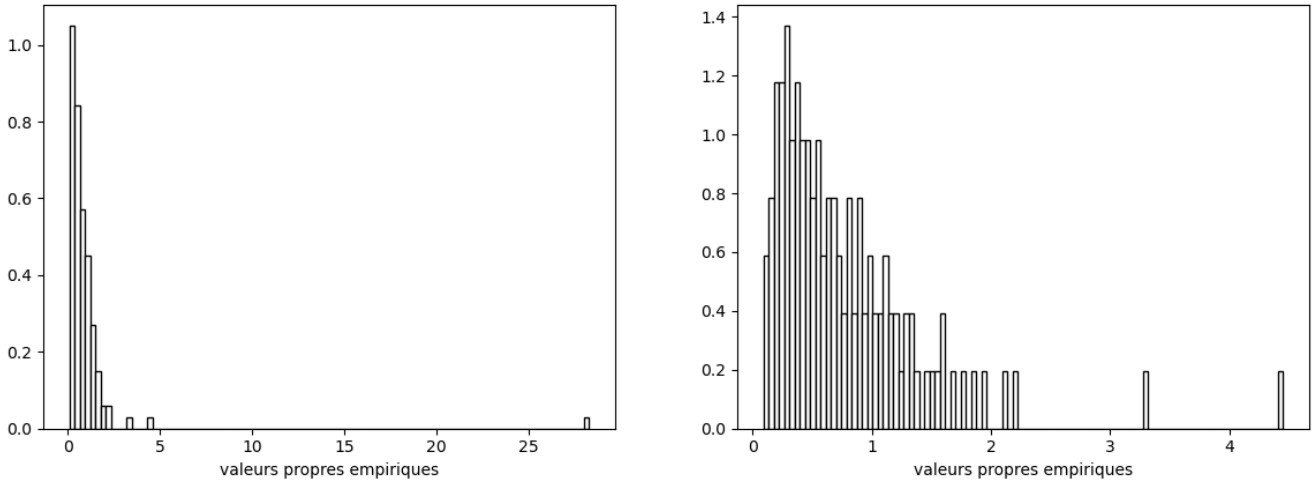

FIGURE 34 – Spectre de la matrice de corrélation empirique pour 118 actions du SBF120 sur les données 2017-2018 avec des rendements normalisés

Une première observation est que les actions sont généralement toutes positivement corrélées entre elles même s'il peut arriver que certaines actions soient temporairement négativement corrélées. On peut par exemple avoir une compagnie aérienne négativement corrélée à une compagnie pétrolière sur une période où le prix du pétrole a fortement augmenté. Une autre explication peut être liée au bruit dû au choix de l'échantillon d'observation. À partir de là on peut rappeler un théorème important de l'algèbre linéaire [17] :

Théorème 6. *Soit $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ une matrice symétrique avec des entrées strictement positives ($\forall 1 \leq i, j \leq n, a_{ij} > 0$). Alors il existe une valeur propre maximale λ , égale au rayon spectral de $\boldsymbol{A}$, et dont le vecteur propre associé v est tel que toutes ses composantes sont positives. De plus, v est le seul vecteur propre de $\boldsymbol{A}$ dont toutes les composantes sont positives à une constante positive multiplicative près.*

(Théorème de Perron-Frobenius)

La plus grande valeur propre, que l'on observe dans le graphe de la densité spectrale empirique représente le mode de marché (*'market mode'*) [17]. Dans nos études on remarque qu'elle est autour de $0.25n$ sur les données récentes mais elle peut devenir plus importante en temps de crise, par exemple la plus grande valeur propre empirique sur les données 2018-2019 était autour de 27 tandis que sur les données 2008-2009 au cœur de la crise financière des sub-primes elle était autour de 39.

Les actifs ont également tendance à être plus corrélés en temps de crise [6]. Pour illustrer cela, nous avons tracé les corrélogrammes des matrices de corrélation empiriques sur les données 2005-2007 (années boursières relativement calmes, 97 des actions actuelles du

SBF120 en faisaient partie à l'époque) et la matrice de corrélation empirique sur les années de crise financière 2008-2009 :

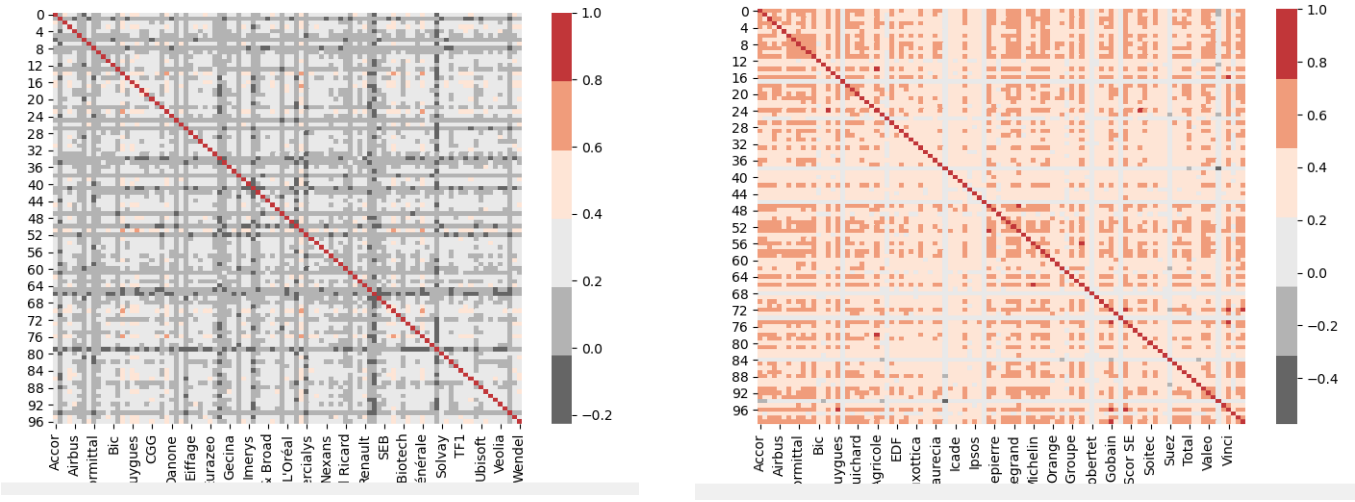

FIGURE 35 – Corrélogrammes des matrices empiriques de corrélation des actifs sur les données 2005-2007 (à gauche) et sur les données 2008-2009 (à droite)

Nous pouvons également réitérer l'observation avec l'exemple plus récent de la crise financière liée au coronavirus en considérons les données Mars 2018-Mars 2019 pour la première matrice de corrélation empirique et Mars 2020- Mars 2021 pour la seconde. Cette crise est d'une nature différente puisqu'elle est due à un choc exogène aux marchés :

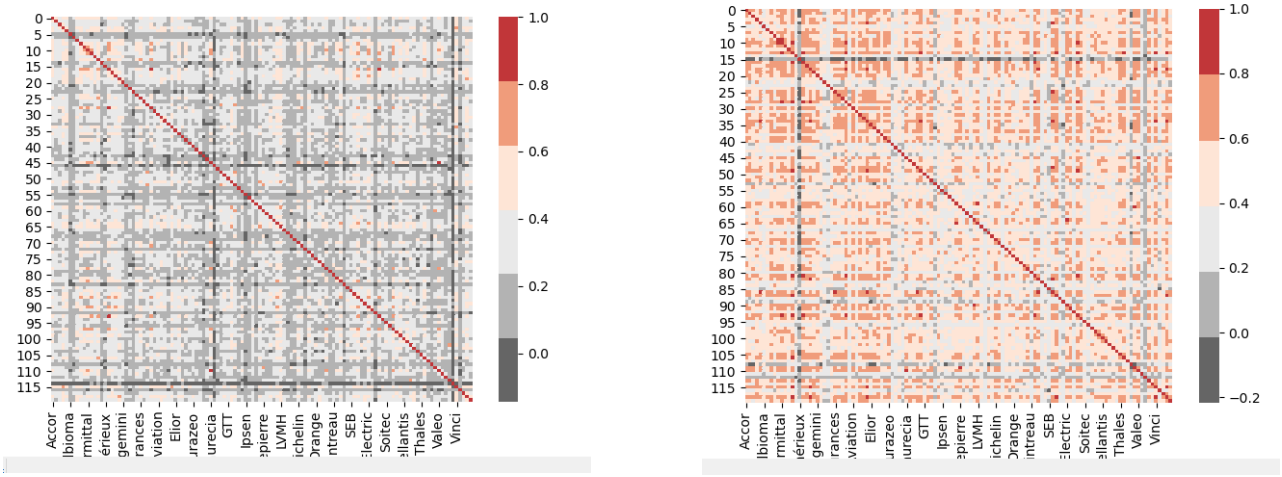

FIGURE 36 – Corrélogrammes des matrices empiriques de corrélation des actifs sur les données 2019-2020 (à gauche) et sur les données 2020-2021 (à droite)

La crise financière avait déjà commencé un peu avant Mars 2020 ce qui fait que le mode de marché (la plus grande valeur propre de la corrélation empirique) était déjà en hausse par rapport à une période normale sur les données Mars 2019-2020. Le mode de marché était autour de 35 pour les données 2019-2020, il est passé à 57 sur les données de l'année suivante en plein cœur de la crise du coronavirus.

Nous pouvons également développer une normalisation des rendements encore plus simple qui consiste juste à calculer la moyenne et l'écart type des rendements sur la période d'observation (1 ou 2 ans ici). Ensuite on retranche leur moyenne aux rendements et on divise par la volatilité sur toute la période d'observation. Cette normalisation nous suffit ici car nous sommes dans une volonté d'estimation de variance-covariance sur une période et non sur le développement une stratégie dynamique d'investissement.

Par exemple nous avons voulu estimer la matrice de variance-covariance de l'année 2018 à partir des données de l'année 2017 avec différents estimateurs : l'estimateur empirique, l'estimateur de Ledoit-Péché avec noyau de Cauchy, l'estimateur de Ledoit-Péché avec noyau de Wigner (distribution avec des queues plus épaisses que la distribution de Cauchy) et avec l'estimateur oracle :

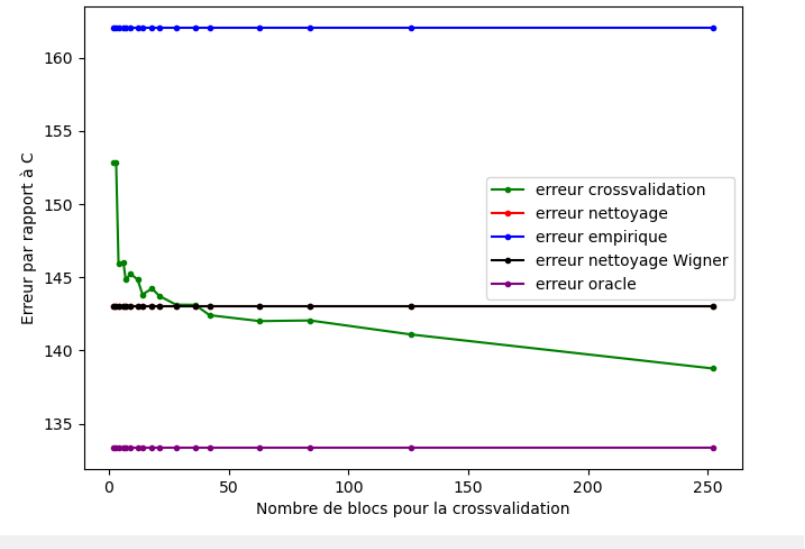

FIGURE 37 – Erreur d'estimation de la matrice de variance-covariance des actifs du SBF120 de l'année 2018 à partir des données 2017 avec différents estimateurs

Une première remarque est que tous les estimateurs sont plus efficaces que l'estimateur empirique. La crossvalidation peut être plus efficace que les algorithmes de nettoyage à condition de bien choisir le nombre de blocs. Nous avons testé la crossvalidation sur d'autres périodes, il semble que le nombre de blocs optimal soit souvent égal à T. L'algorithme de Ledoit-Péché avec un noyau de Wigner semble aussi légèrement plus efficace que celui avec un noyau de Cauchy.

Nous avions également pour projet de rajouter des matières premières, obligations et taux de change à notre ensemble d'actions afin de tester l'algorithme de crosscovariance sur les données financières. Cela aurait demandé certaines adaptations.

La plupart des futures sont côtés sur les marchés financiers américains, les heures d'ouvertures ne sont donc pas les mêmes que celles de marché français à cause du décalage horaire. La fermeture de la bourse de Paris se fait quelques heures après l'ouverture de la bourse américaine. En considérant des données quotidiennes on risque de rajouter du bruit dans nos simulations.

En effet, les prix sont influencés par l'intensité des échanges et cette intensité n'est pas constante durant la journée de trading. En particulier l'ouverture et la clôture sont des périodes où les échanges s'intensifient et les prix peuvent donc varier fortement à ces moments là. De plus, le prix des futures américains peuvent avoir une influence sur les prix des actions françaises. Il paraît donc risqué de mettre sur le même plan des prix de clôtures obtenus à plusieurs heures de décalage. Pour cela il faut mener les simulations, non plus sur des données quotidiennes mais sur des données hebdomadaires (par exemple le prix de clôture du vendredi).

Nous allons maintenant considérer le marché américain en nous intéressant aux actions du S&P500.

7.2 Application sur des données américaines du S&P 500

Nous disposons des données financières quotidiennes de clôture du 1 Janvier 1985 au 14 Juin 2021, soit 9510 jours de données dont certains, souvent fériés, ne contiennent pas de prix. Le S&P500 contient les 500 entreprises les plus liquides des États-Unis (ie : les plus grandes capitalisations boursières). Toutefois les 500 entreprises qui le composent à un instant donné ne sont pas restées les mêmes sur les 37 années de données dont nous disposons. Au total 917 entreprises sont passées ou entrées dans le S&P500. Par exemple, Tesla n'existait pas en 1985.

En premier lieu nous avons voulu observer la distribution des valeurs propres de la matrice empirique de corrélation des actifs du S&P500. Nous avons considérer trois ans de données soit $q \sim 2/3$ sur les données de début 2012 à fin 2014 qui sont des années boursières sans crise financière :

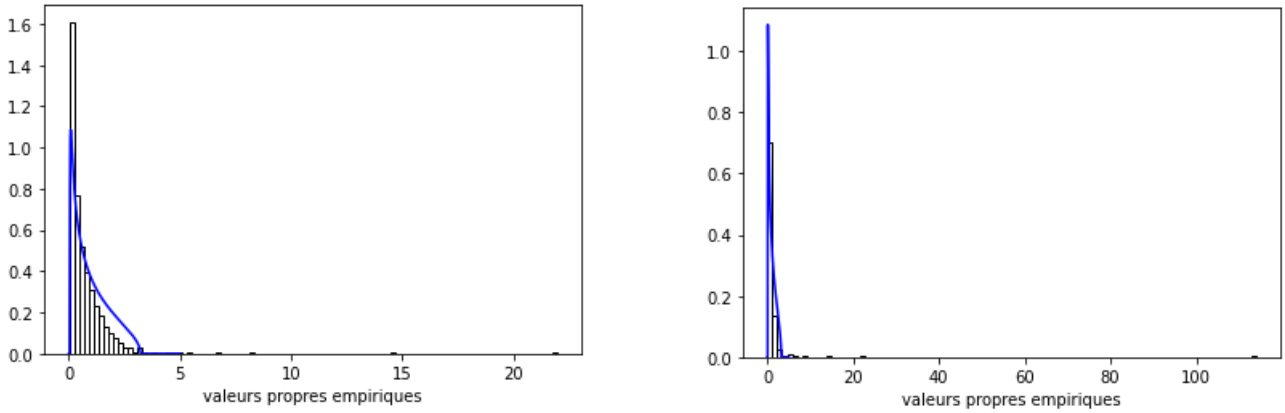

FIGURE 38 – Distribution empirique des valeurs propres empiriques de la variance-covariance des actifs du S&P500 les années 2012-2014 (le graphe de gauche est un zoom du second). En bleu nous avons tracé la distribution théorique de la loi de Marcenko-Pastur.

Nous observons que la loi de Marcenko-Pastur contient une bonne partie des valeurs propres empiriques. Ceci constitue une illustration de l'intérêt d'utiliser les matrices aléatoires pour la modélisation et la compréhension des corrélations d'actifs financiers.

Nous avons également voulu observer les corrélations entre les actifs financiers en temps de crise. Nous avons donc tracé les corrélogrammes des actions du S&P500 sur les trois années précédents la crise des subprimes (de début 2004 à fin 2006) puis le corrélogramme sur la période début 2007 à fin 2009 qui comprend la crise financière :

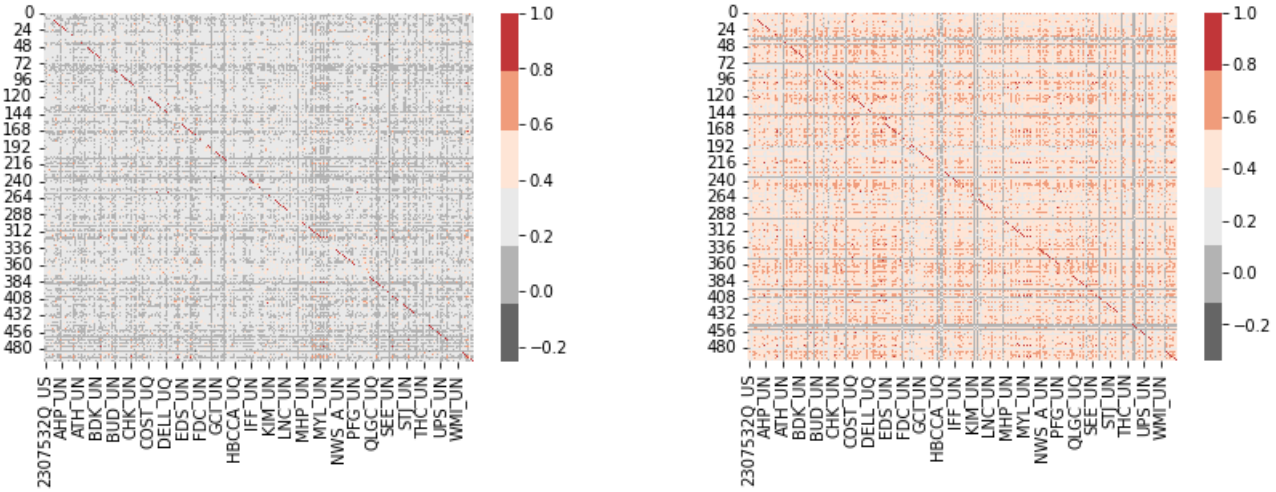

FIGURE 39 – Corrélogrammes des matrices de corrélation empirique des actifs du S&P500 sur les données des années 2007-2009 (à gauche) et sur les données 2004-2006 (à droite)

Comme pour les données françaises nous observons une forte augmentation des corrélations entre actifs qui est due que la volatilité à tendance a fortement augmenter pendant une crise et un grand nombre d'actions possèdent simultanément une forte tendance baissière. On peut également regarder le mode de marché (la plus grande valeur propre de la matrice de corrélation empirique), il passe de 113 sur la période 2004-2006 à 210 sur la période 2007-2009. On remarque que le mode de marché est autour de $0.25n$ avec n le nombre d'actifs considéré lorsque les marchés ne sont pas en crise. Nous avions le même ordre de grandeur pour les données françaises du SBF120.

Nous avons voulu tester nos estimateurs pour estimer la matrice de corrélation des actifs des deux années 2015-2017 $\boldsymbol{C}$ à partir des données début 2012 à fin 2014 dont la matrice de corrélation empirique est $\boldsymbol{E}$. Nous menons donc une estimation hors des périodes de crise. Nous regardons l'erreur d'estimation par rapport à $\boldsymbol{C}$ au sens de la norme de Frobenius comme c'est le cas dans tous le rapport. Nous avons testé l'estimateur de Ledoit-Péché avec un noyau de Cauchy et avec un noyau de Wigner ainsi que la crossvalidation pour différents blocs.

On a également considéré l'estimateur oracle dont on rappelle la définition [1]. On appelle $(\lambda_1, ..., \lambda_n), (u_1, ..., u_n)$ les valeurs propres et vecteurs propres de la matrice de corrélation $\boldsymbol{E}$, l'oracle conserve les vecteurs propres de $\boldsymbol{E}$ et a pour valeurs propres :

$$\xi(\lambda_k) = u_k^T \boldsymbol{C} u_k$$

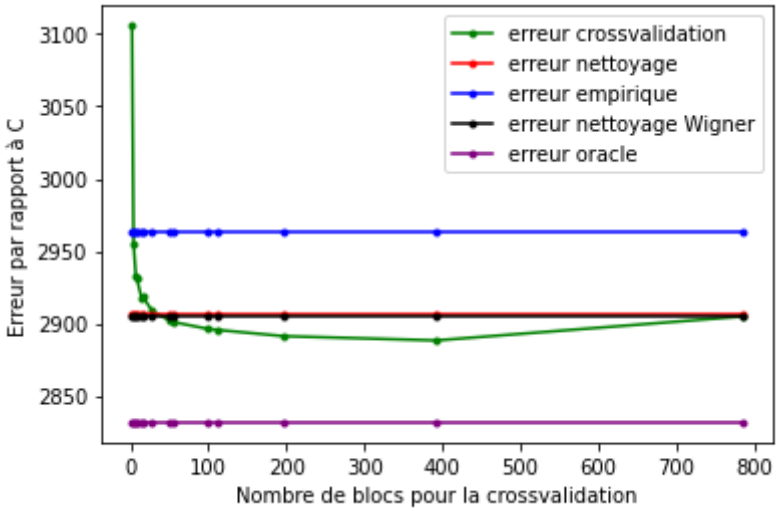

FIGURE 40 – Erreur d'estimation de la matrice de corrélation des actifs du S&P500 de la période 2015-2016 à partir des données 2012-2014 avec différents estimateurs

On rappelle que l'estimateur oracle est l'estimateur optimal sous l'hypothèse d'invariance par rotation. Un premier constat est que les trois estimateurs considérés (Ledoit-Péché avec deux noyaux différents et la crossvalidation) permettent de de rapprocher significativement de l'erreur de l'oracle par rapport à la performance de l'estimateur empirique. La crossvalidation permet d'obtenir de meilleurs résultats que le nettoyage de Ledoit-Péché mais le nombre de blocs à choisir pour que cela soit le cas est assez volatile.

Il aurait été intéressant d'observer l'évolution de ces erreurs d'estimation sur une fenêtre glissante de quelques mois [18] et aussi de tester nos estimateurs avec des moyennes mobiles exponentielles à la place de moyenne simple des observations pour la matrice de corrélation empirique [20].

Expérience avec Markowitz, comment estimer g les rendements futurs ?

Supposons que l'on souhaite constituer un portefeuille 'virtuel' à partir de la théorie de Markowitz et ensuite suivre sa performance. On rappelle l'expression du problème d'optimisation et de la forme du portefeuille optimal :

$$\begin{cases} \min\limits_{\pi\in\mathbb{R}^N} \dfrac{1}{2}\pi^T \boldsymbol{C}\pi \\ \\ \text{s.t } \pi^T g \geq \mathcal{G} \\ \quad \sum_{i=1}^n \pi_i = 1 \end{cases}$$

Le portefeuille optimal s'écrit de la manière suivante :

$$\pi_{opt} = \mathcal{G}\frac{g^T \boldsymbol{C^{-1}}}{g^T \boldsymbol{C^{-1}} g}$$

Nous avons vu comment estimer $\boldsymbol{C}$, il nous reste cependant à fixer g le vecteur des rendements futurs sur une période donnée.

Une première approche pourrait être de considérer que l'investisseur ne dispose d'aucune information pour déterminer les rendements futurs des différents actifs. On peut choisir g des manières suivantes :

- en tirant aléatoirement les rendements futurs avec une loi normale par exemple.
- en effectuant une moyenne exponentielle mobile sur une période (exemple de période d'un mois) et en essayant de prolonger la courbe.

Définition 10. *Soit $(X_t) \in \mathbb{R}^{\mathbb{N}}$ les rendements d'un actif et soit d une période d'observation. La moyenne mobile exponentielle (EMA) au temps t et de période d est alors :*

$$EMA_t = \alpha * X_t + (1 - \alpha) * EMA_{t-1}$$

avec :

$$\alpha = \frac{2}{d+2}$$

Une deuxième approche, plus réaliste, serait de considérer que l'investisseur a quelques idées des rendements futurs grâce à des analyses financières qu'il a mené et grâce à l'information dont il dispose. Dans la mesure où cette modélisation porte sur la gestion des risques et non sur la prédiction on pourrait calculer les rendements futurs réalisés et leur ajouter un bruit gaussien. On considérerait alors cela comme les rendements futurs estimés par l'investisseur :

$$g(i) = R(i) + \eta(i)$$

avec R le vecteur des rendements futurs et η un bruit gaussien indépendant de R.

Il faudrait ensuite calculer des portefeuilles optimaux de Markowitz à l'aide de nos estimateurs, suivre leurs performances et éventuellement les adapter tous les jours de trading si l'on souhaite implémenter une stratégie dynamique.

Nous allons développer un exemple avec les mêmes données que pour le précédent graphe. On cherche à estimer la matrice de corrélation $\boldsymbol{C}$ des actifs sur les données 2015-2016 à partir des données 2012-2014. On note $\boldsymbol{E}$ la matrice empirique de corrélation des actifs sur la période 2012-2014. Pour cette simulation nous fixerons $\mathcal{G} = 20\%$ le rendement de portefeuille minimal espéré et g le vecteur de l'espérance des rendements futurs sera la somme des rendements réalisés sur la période 2015-2016 et d'un bruit gaussien indépendant.

On va donc considérer les portefeuilles suivants :

- le portefeuille empirique :
$$\pi_{emp} = \mathcal{G}\frac{g^T \boldsymbol{E^{-1}}}{g^T \boldsymbol{E^{-1}} g}$$
- le portefeuille nettoyé avec un noyau de Cauchy, on note $\boldsymbol{E_{Cauchy}}$ la matrice nettoyée avec un noyau de Cauchy :
$$\pi_{Cauchy} = \mathcal{G}\frac{g^T \boldsymbol{E_{Cauchy}^{-1}}}{g^T \boldsymbol{E_{Cauchy}^{-1}} g}$$
- le portefeuille nettoyé avec un noyau de Wigner, on note $\boldsymbol{E_{Wigner}}$ la matrice nettoyée avec un noyau de Cauchy :
$$\pi_{Wigner} = \mathcal{G}\frac{g^T \boldsymbol{E_{Wigner}^{-1}}}{g^T \boldsymbol{E_{Wigner}^{-1}} g}$$

- le portefeuille généré par l'estimateur de crossvalidation avec k blocs, $\boldsymbol{E_{cross}(k)}$:

$$\pi_{cross}(k) = \mathcal{G}\frac{g^T \boldsymbol{E_{cross}(k)^{-1}}}{g^T \boldsymbol{E_{cross}(k)^{-1}} g}$$

- le portefeuille généré par l'estimateur oracle, $\boldsymbol{\xi(E)}$:

$$\pi_{oracle} = \mathcal{G}\frac{g^T \boldsymbol{\xi(E)^{-1}}}{g^T \boldsymbol{\xi(E)^{-1}} g}$$

- le portefeuille optimal (c'est-à-dire de risque minimal et de rendement $\mathcal{G}$ sur la période 2015-2016 :

$$\pi_{opt} = \mathcal{G}\frac{g^T \boldsymbol{C^{-1}}}{g^T \boldsymbol{C^{-1}} g}$$

Dans le cas des données financières il n'existe probablement pas de "vraie" matrice de variance-covariance des actifs donc nous ne pouvons pas parler de "vrai" risque de portefeuille mais plutôt de risque réalisé sur une période par un portefeuille. Nous considérons également les risques réalisés associés à ces portefeuilles :

$$Risque(\pi) = \pi^T \boldsymbol{C}\pi$$

avec π un portefeuille donné.

Nous avons donc tracé les risques réalisés par ces différents estimateurs :

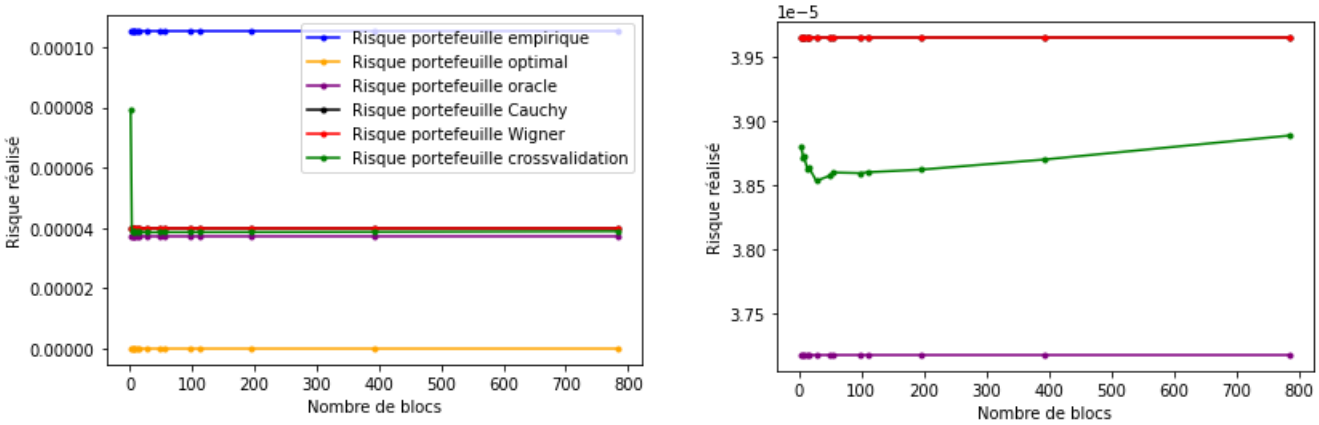

FIGURE 41 – Risques réalisés sur la période 2015-2016 pour des portefeuilles d'actifs du S&P500 construits avec différents estimateurs de la matrice de variance-covariance des actifs. Le second graphe est un zoom du premier auquel on l'on considère les portefeuilles construits avec les estimateurs de Ledoit-Péché, la crossvalidation et avec l'oracle (on a retiré la première valeur de risque du portefeuille de crossvalidation avec 2 blocs

Une première remarque est que tous les estimateurs utilisés permettent de construire des portefeuilles moins risqués que le portefeuille construit avec l'estimateur empirique. Le nettoyage avec noyau de Cauchy et avec noyau de Wigner semblent donner des résultats très similaires en termes d'efficacité. Quant à la crossvalidation, le portefeuille le moins risqué semble être construit avec un nombre de blocs en racine carré du nombre d'actifs considéré mais l'estimateur avec T blocs semble aussi efficace. Enfin nos différents estimateurs se rapprochent significativement de l'efficacité de l'oracle qui est le mieux que l'on puisse faire sous l'hypothèse d'invariance par rotation.

Il pourrait être intéressant d'appliquer ces méthodes dans le cadre d'une stratégie dynamique et de suivre les performances et les risques des différents portefeuilles sur la durée.

8 Conclusion

La gestion de risque de portefeuille est une problématique importante en finance qui se pose à tout investisseur. Nous avons ici abordé cette question sous l'angle de l'estimation des matrices de variance-covariance et de la théorie du portefeuille optimal de Markowitz. Un constat a été de voir que l'usage de l'estimateur empirique pour la matrice de variance-covariance induit une sous-estimation du risque de portefeuille car le spectre de l'estimateur empirique est plus large que celui de la "vraie" matrice de variance-covariance des données.

Nous avons donc étudié différents estimateurs pour tenter de faire mieux que l'estimateur empirique. La classe d'estimateurs que nous avons étudiée est celle des estimateurs invariants par rotation. Dans ce cas, on conserve les vecteurs propres empiriques et on se concentre sur l'estimation des valeurs propres. Nous avons considéré l'estimateur de Ledoit-Péché dérivé de la théorie des matrices aléatoires avec un noyau de Cauchy et avec un noyau de Wigner pour les matrices de variance-covariance. L'usage d'un noyau de Wigner semble améliorer légèrement la performance de l'estimateur avec noyau de Cauchy. Nous avons également étudié un estimateur de crosscovariance pour des matrices de variance-covariance connues.

Nous avons considéré la crossvalidation qui a l'avantage de ne pas nécessiter d'hypothèses lourdes sur les données de départ et nous nous sommes intéressés à la question du nombre de blocs optimal. Nous émettons plusieurs hypothèses concernant la crossvalidation. Tout d'abord nous pensons que la crossvalidation permettrait d'obtenir la même erreur théorique que l'estimateur de Ledoit-Péché. Aussi, en ce qui concerne le nombre de blocs optimal nous pensons qu'il est en racine de la taille de la matrice de variance-covariance dans la plupart des cas mais que pour les données financières, fortement non stationnaires, il semble être très volatile et peut se rapprocher de T,le nombre d'observations que l'on considère.

Ces différents estimateurs ont été testés sur des données financières françaises et américaines. Il pourrait être intéressant de tester les différents algorithmes sur des données macroéconomiques ou sur des données d'autres natures si elles sont suffisamment fréquentes. Nous pensons également que ces estimateurs peuvent être utiles à la gestion de risques pour de nombreux acteurs amenés à gérer de grands portefeuilles d'actifs financiers. On pourrait donner l'exemple des assureurs qui dans un contexte de taux bas qui semble s'inscrire dans la durée doivent de plus en plus investir dans des actifs risqués sur les marchés financiers comme dans le cas pour les contrats en unité de compte ou pour les PER (plan épargne retraite). Ces estimateurs sont utiles pour estimer la variance d'un portefeuille, les corrélations entre actifs financiers et peuvent également servir pour construire un portefeuille à l'aide de la théorie de Markowitz.

Une grande partie de ce stage a été consacrée à l'étude des matrices Inverse Wishart, nous souhaiterions approfondir la question du nombre de blocs optimal pour la crossvalidation dans ce cas là en analysant les données de crossvalidation que nous avons généré et ensuite si possible d'un point de vue plus théorique. Nous avons également pour projet d'appliquer l'algorithme de crosscovariance aux données financières.

9 Note de synthèse

Nous nous intéressons à la gestion de risques de portefeuilles financiers sous l'angle de l'estimation des matrices de variance-covariance des rendements des actifs. La théorie du portefeuille optimal de Markowitz propose un cadre théorique simple pour déterminer les proportions à investir dans chaque actifs financier dès lors qu'on détermine la matrice de variance-covariance des actifs et le vecteur de l'espérance des rendements futurs.

Dans ce problème, nous disposons des observations au fil du temps t pour $t = 1, 2, .., T$ d'un vecteur aléatoire $X \in \mathbb{R}^n$. Dans le cas financier, X_t représentera le vecteur des rendements au temps t des actifs. On définit la matrice $\boldsymbol{X}$ de taille $n \times T$:

$$\boldsymbol{X} = (X(1), X(2), X(3), ..., X(T))$$

Nous supposons que les $X(1), ..., X(n)$ sont des observations de X indépendantes et centrées (quitte à leur retrancher leur moyenne). On se demande alors comment estimer la covariance des (X_t) à partir des observations dont on dispose.

Une première réponse est d'utiliser l'estimateur empirique. Comme nos données sont supposées centrées, on peut négliger les produits d'espérances et la matrice de variance-covariance empirique de $\boldsymbol{X}$ est alors :

$$\boldsymbol{C_{emp}} = \frac{1}{T}\boldsymbol{X}\boldsymbol{X^T}$$

Nous allons maintenant introduire la notion de portefeuille optimal de Markowitz et expliquer pourquoi l'usage de l'estimateur empirique peut donner de grandes erreurs d'estimation.

Estimation de risques et théorie de Markowitz

Soit un portefeuille contenant n actifs financiers et soit $\pi \in \mathbb{R}^n$ le vecteur des positions prises sur chaque actif (la vente à découvert est autorisée). L'investisseur souhaite alors déterminer les quantités optimales à investir dans chaque actif, par portefeuille optimal on entend un portefeuille qui maximise le rendement moyen espéré sous une contrainte de risque ou de manière équivalente. Ce problème a été formalisé et résolu par Harry Markowitz en 1952 [15].

Commençons par définir quelques notations. Soient $\boldsymbol{C} \in \mathbb{R}^{n \times n}$ la matrice de variance-covariance des rendements et $g \in \mathbb{R}^n$ le vecteur des espérance de rendement pour chaque actif. Ces deux objets sont considérés comme connus. Soit $\mathcal{G}$ l'espérance minimale de rendement que souhaite l'investisseur. Le problème d'optimisation a alors pour solution [17] :

$$\pi_{opt} = \mathcal{G}\frac{g^T\boldsymbol{C^{-1}}}{g^T\boldsymbol{C^{-1}}g}$$

La théorie de Markowitz offre un cadre quantitatif simple d'utilisation sous réserve de connaître $\boldsymbol{C}$ et g. Lorsque l'on dispose d'un nombre fini d'observations des données tel que la taille de la matrice de variance-covariance n'est pas négligeable devant lui, l'estimateur empirique est largement inefficace. Dans ce la cas, l'estimateur empirique est "bruité"

et son spectre est systématiquement plus large que celui de la "vraie" matrice de variance-covariance des données. Les matrices de variance-covariance sont diagonalisables et chaque vecteur propre représente un portefeuille de norme 1 (au sens L^2) avec pour risque la valeur propre associée. Le portefeuille optimal est alors une combinaison linéaire de ces portefeuilles. Lorsque l'on combine cet estimateur à la théorie de Markowitz, on se retrouve à moins investir sur les plus grandes valeurs propres car on les croît plus grandes qu'elle ne le sont en réalité et on investit plus sur les petites valeurs propres qui sont elles plus petites que les valeurs propres minimales de la vraie matrice de variance-covariance. On se retrouve donc à sous-estimer le risque de notre portefeuille.

L'estimateur de Ledoit-Péché pour les matrices de covariance

Le spectre empirique étant plus large que le spectre de la vraie matrice de covariance. L'estimateur de Ledoit-Péché va appliquer un shrinkage non linéaire sur les valeurs propres empiriques afin de réduire la largeur du spectre. L'idée est de conserver les vecteurs propres empiriques et d'appliquer une fonction affine à chaque valeur propre avec un coefficient directeur inférieur à 1 afin de réduire la largeur du spectre.

On suppose que la matrice de covariance $\mathbf{\Sigma}$ (que l'on ne connaît pas a priori) d'un signal d'un signal Z est invariante par rotation (pour toute matrice orthogonale $\boldsymbol{O}$, $\boldsymbol{O\Sigma O^T}$ a la même loi que la matrice $\mathbf{\Sigma}$). Ledoit et Péché ont développé un estimateur de $\mathbf{\Sigma}$ à partir d'outils de la théorie des matrices aléatoires [10]. Cet estimateur nécessite quelques hypothèses sur les données que l'on listera dans le rapport [10][1].

On définit les éléments suivants :
— $X \in \mathbb{R}^n$ un vecteur aléatoire centré de matrice de covariance Σ
— $\boldsymbol{X} = (X(1), ..., X(T)) \in \mathbb{R}^{n \times T}$ des copies iid de X
— $\boldsymbol{E} = \frac{1}{T}\boldsymbol{X}\boldsymbol{X^T}$ la matrice de covariance empirique des observations, de valeurs propres $(\lambda_1, ...\lambda_n)$ et de vecteurs propres $(u_1, ..., u_n)$.

L'objectif est d'estimer au mieux $\mathbf{\Sigma}$ dans la classe des estimateurs invariants par rotation (RIE : "rotationnally invariant estimators"). Dans la mesure où l'estimation des vecteurs propres est quasi-impossible avec une taille de matrice non négligeable devant le nombre d'observations, il est cohérent dans le cadre Bayésien de considérer toutes les bases orthonormées comme équiprobables. On peut alors formuler un problème d'optimisation [1] :

$$\underset{Estimateur \in RIE}{\operatorname{argmin}} \quad \|\boldsymbol{Estimateur} - \mathbf{\Sigma}\|_F$$

avec $\|.\|_F$ la norme de Frobenius : $\|\boldsymbol{M}\|_F = \sqrt{Tr(\boldsymbol{M}\boldsymbol{M^T})}$

L'estimateur optimal, invariant par rotation, est alors :

$$\boldsymbol{\xi(E)} = \widehat{\boldsymbol{E}} = \sum_{k=1}^{n} \widehat{\lambda}_k u_k u_k^T$$

avec

$$\xi(\lambda_k) = \widehat{\lambda}_k = u_k^T \mathbf{\Sigma} u_k$$

$(\xi(\lambda_k))$ dépend de $\mathbf{\Sigma}$ qui est inconnue, c'est pourquoi on appelle Ξ la fonction oracle. L'idée est alors de chercher à approximer les mesures empiriques des matrices $\boldsymbol{E}$ et $\widehat{\boldsymbol{E}}$:

$$\sum_{k=1}^{n} \delta_{\lambda_k} \text{ and } \sum_{k=1}^{n} u_k^T \mathbf{\Sigma} u_k \delta_{\lambda_k}$$

L'estimateur de Ledoit-Péché se calcule grâce à la transformée de Stieltjes de la matrice $\boldsymbol{E}$:

Définition 11. *La transformée de Stieltjes d'une matrice $\boldsymbol{A}$ de taille $n \times n$ est [17] :*

$$g_n^{\boldsymbol{A}}(z) = \frac{1}{n} Tr((z\boldsymbol{Id} - \boldsymbol{A})^{-1}) = \frac{1}{n} \sum_{k=1}^{n} \frac{1}{z - \lambda_k}$$

où les $(\lambda_1, ..., \lambda_n)$ sont les valeurs propres de $\boldsymbol{A}$ et où $z\mathbb{C}$. La transformée de Stieltjes n'est pas définie dans le cas où z appartient au spectre de $\boldsymbol{A}$.

Théorème 7. *On note g_μ la transformée de Stieltjes de la matrice $\boldsymbol{E}$. Pour $z = \lambda + i\eta$, les valeurs propres de l'estimateur oracle sont [1] :*

$$\xi(\lambda) = \lim_{\eta \to 0^+} \frac{\lambda}{|1 - q + q\lambda g_\mu(\lambda + i\eta)|^2}$$

(Formule de Ledoit-Péché)

La crossvalidation

La crossvalidation est une technique issue du machine learning qui a l'avantage de ne nécessiter quasiment aucune hypothèse sur les données de départ, on suppose juste que nos données sont centrées. On découpe nos observations en ensembles de tailles égales appelés blocs, les observations privées du bloc sont l'ensemble d'optimisation et les observations dans le bloc forment l'ensemble de validation. La technique est appelée crossvalidation et non juste validation parce qu'on effectue cette optimisation/validation sur chaque bloc avant de moyenner [17] [20].

On commence par définir les données du problème :
— T : nombre d'observations
— k : nombre de blocs
— taille des blocs : $\lfloor \frac{T}{k} \rfloor$

Les blocs sont définis sur des données dont les indices t sont dans un intervalle de la forme suivante :

$$\left[i \times \left\lfloor \frac{T}{k} \right\rfloor, (i+1) \times \left\lfloor \frac{T}{k} \right\rfloor \right] \text{ pour } i = 0, ..., k$$

Crossvalidation pour les matrices de covariance

Prenons un des blocs sur les données $[T1, T2]$, on définit alors :

— $\boldsymbol{C_{in}}$la matrice de covariance empirique sur les données $[T1, T2]$ de vecteurs propres $u_{in} = (u_{in}(1), ..., u_{in}(n))$

— $\boldsymbol{C_{out}}$ la matrice de covariance empirique sur $[0, T1] \cup [T2, T]$ de vecteurs propres $u_{out} = (u_{out}(1), ..., u_{out}(n))$

Sur ce bloc, l'estimateur des valeurs propres est le suivant :

$$\xi(\lambda_i) = u_{out}^T(i)\boldsymbol{C_{in}}u_{out}(i)$$

On peut alors reconstituer "une matrice de covariance" :

$$\boldsymbol{\Xi_k} = u_{out}\xi u_{out}^T$$

L'estimateur de covariance n'est alors que la moyenne de ces estimateurs sur les différents blocs :

$$\boldsymbol{\Xi} = \frac{1}{K}\sum_{k=1}^{K}\boldsymbol{\Xi_k}$$

Remarque : Il semble y avoir un nombre de blocs optimal pour la crossvalidation, ce problème étant encore ouvert pour nos problèmes d'estimation de covariance, nous avons mené une étude empirique sur le sujet pour certaines matrices de variance-covariance.

Ces différents estimateurs ont été présentés, testés à l'aide du langage Python puis illustrés sur des exemples. Dans un deuxième temps, nous les avons testés sur des données financières françaises et américaines et avons pu voir qu'ils permettent de corriger une partie substantielle de l'erreur de l'estimateur empirique pour l'estimation de matrices de variances-covariance.

10 Executive Summary

Risk management of financial portfolios is an important question in finance and it will be tackled here from the point of view of covariance matrix estimation. Markowitz' optimal portfolio theory brings a simple theoretical framework to determine the proportions to be invested in each asset as soon as the covariance matrix of asset returns and the vector of expected future returns are known.

In this problem, one has access to observations at time t for $t = 1, 2, .., T$ of a random vector $X \in \mathbb{R}^n$. In the financial case, X_t will represent the vector of the financial returns of the assets at time t. Let us define matrix $\boldsymbol{X}$ of size $n \times T$:

$$\boldsymbol{X} = (X(1), X(2), X(3), ..., X(T))$$

Let us assume that $X(1), ..., X(n)$ are independant observations of X and are centered (one can always substract the mean of the data). The question is then to determine the covariance matrix of the (X_t) from the data at disposal.

A first answer could be to use the empirical estimator. As our data set is assumed to be centered, products of expectations can be neglected and the empirical covariance matrix of X is then :

$$\boldsymbol{C_{emp}} = \frac{1}{T}\boldsymbol{X}\boldsymbol{X^T}$$

We will then introduce the notion of optimal portfolio of Markowitz and explain briefly why the empirical estimator should not be used for portfolio selection purposes.

Markowitz' framework

Let us consider a portfolio built on n securities and $\pi \in \mathbb{R}^n$ the position taken on each asset. Taking long or short position are allowed. The goal of an investor is to determine the optimal proportions of each asset to buy or short-sell. However at this point, one still has to define what an optimal portfolio is. One answer could be to consider mean/variance optimization, an optimal portfolio is the one that maximizes the expected return under a desired level of risk or one that minimizes the level of risk such that it is over a minimum expected return. This problem was formulated and solved by Harry Markowitz in 1952 [15].

Let us first define several notations. The covariance matrix $C \in \mathbb{R}^{n \times n}$ of the returns of assets and $g \in \mathbb{R}^n$ the expected gain of each asset, these quantities are supposed to be known. Let us call $\mathcal{G}$ the minimum expected return the investor is wishing for. The optimal portfolio is defined by [4][17] :

$$\pi_{opt} = \mathcal{G}\frac{g^T\boldsymbol{C^{-1}}}{g^T\boldsymbol{C^{-1}}g}$$

Markowitz' theory offers a simple quantitative framework under the assumption of knowing $\boldsymbol{C}$ and g. When the size of the covariance matrix is non-negligible in front of the amount of data at disposal the use of the rmpirical estimator for portfolio selection is inefficient. In this case, the spectrum of the empirical estimator is systematically larger than the

spectrum of the real covariance matrix of the data. Covariance matrices are diagonalizable and each eigenvector represent a portfolio of norm 1 (according to the L^2 norm) with a risk equal to the associated eigenvalue. Markowitz optimal portfolio is then a linear combination of these portfolios. When the empirical estimator is combined to Markowitz' framework, less is invested on the highest eigenvalues because we believe them higher so riskier than what they are in reality and on the contrary more is invested on the smallest empirical eigenvalues that we think smaller than the smallest eigenvalues of the real covariance matrix. Therefore the risk of the selected portfolio is largely underestimated.

Ledoit-Péché's estimator for covariance matrices

As the sprectrum of the empirical covariance matrix is larger than the one of the real covariance matrix, other estimators should be considered for portfolio selection or risk mangement in general. Ledoit and Péché developed an estimator which apply a nonlinear shrinkage on the empirical eigenvalues to reduce the size of the spectrum [10]. The idea is to keep the empirical eigenvectors and to apply a linear function on the eigenvalues with a coefficient smaller than 1 to reduce the length of the spectrum.

Let us suppose that the covariance matrix $\mathbf{\Sigma}$ (which is unknown) of a signal Z is rotationnally invariant (for any orthogonal matrix $\boldsymbol{O}$, $\boldsymbol{O\Sigma O^T}$ is equally likely to appear as $\mathbf{\Sigma}$). Ledoit and Péché developed an estimator of $\mathbf{\Sigma}$ with the knowledge of the observations of the signal Z using tools from random matrix theory [10]. This estimators is based on several assumptions that are listed in the report.

We define the following elements :
— $X \in \mathbb{R}^n$ a centered random vector with covariance matrix $\mathbf{\Sigma}$ (which is unknown)
— $\boldsymbol{X} = (X(1), ..., X(T)) \in \mathbb{R}^{n \times T}$ independant and identically distributed copies of X
— $E = \frac{1}{T} XX'$ the sample covariance matrix of observations, with eigenvalues $(\lambda_1, ...\lambda_n)$ and eigenvectors $(u_1, ..., u_2)$.

The goal is to best estimate $\mathbf{\Sigma}$ inside the set of rotationnally invariant estimators. As the estimation of eigenvectors is almost impossible as the size of the covariance matrix to estimate is not negligeable in front of the number of observations we have, it is coherent in the Bayesian framework to consider all eigenvector bases are equally likely to appear. Therefore we are going to keep the eigenvectors of the empirical estimator and focus on the cleaning of the eigenvalues. The estimation can be formulated as an optimization problem [1] :

$$\underset{Estimator \in RIE}{\operatorname{argmin}} \quad \|\boldsymbol{Estimator} - \mathbf{\Sigma}\|_F$$

with $\|.\|_F$ the Frobenius norm : $\|\boldsymbol{M}\|_F = \sqrt{Tr(\boldsymbol{MM^T})}$

The optimal estimator which is rotationnally invariant is then :

$$\boldsymbol{\xi(E)} = \widehat{\boldsymbol{E}} = \sum_{k=1}^{n} \widehat{\lambda}_k u_k u_k^T$$

with

$$\xi(\lambda_k) = \widehat{\lambda}_k = u_k^T \mathbf{\Sigma} u_k$$

$(\xi(\lambda_k))$ depends on $\mathbf{\Sigma}$ which is unknown, therefore $\mathbf{\Xi}$ is called the oracle function. The idea

is then to approximate the empirical distributions of the matrices $\boldsymbol{E}$ and $\widehat{\boldsymbol{E}}$:

$$\sum_{k=1}^{n} \delta_{\lambda_k} \text{ and } \sum_{k=1}^{n} u_k^T \boldsymbol{\Sigma} u_k \delta_{\lambda_k}$$

The estimator of Ledoit-Péché can be computed with the use of the Stieltjes transform of matrix $\boldsymbol{E}$:

Définition 12. *The Stieltjes transform of a matrix $\boldsymbol{A}$ of size $n \times n$ is [17] :*

$$g_n^A(z) = \frac{1}{n} Tr((z\boldsymbol{Id} - \boldsymbol{A})^{-1}) = \frac{1}{n} \sum_{k=1}^{n} \frac{1}{z - \lambda_k}$$

where $(\lambda_1, ..., \lambda_n)$ are the eigenvalues of $\boldsymbol{A}$ and $z \in \mathbb{C}$, the Stieltjes transform is not defined on the sprectrum of $\boldsymbol{A}$.

Théorème 8. *Let us note g_μ the Stieltjes transform of matrix $\boldsymbol{E}$. For $z = \lambda + i\eta$, the eigenvalues of the oracle estimator are [1] :*

$$\xi(\lambda) = \lim_{\eta \to 0^+} \frac{\lambda}{|1 - q + q\lambda g_\mu(\lambda + i\eta)|^2}$$

(Ledoit-Péché's Formula)

Crossvalidation

Crossvalidation is a common tool in machine learning which has the advantage of almost not needing any assumption on the starting data. To run crossvalidation, the sample of observations is cut into equally sized blocs. The data outside the bloc constitute the set of optimization whereas the data inside the bloc is the set of validation. It is called crossvalidation and not validation because optimization and validation are run on every bloc before taking the mean of these estimators [17] [20].

Let us begin by introducing some notations :
— T : number of observations
— k : number of blocs
— size of the blocs : $\lfloor \frac{T}{k} \rfloor$

Blocs are defined on data of the following form :

$$\left[i \times \left\lfloor \frac{T}{k} \right\rfloor, (i+1) \times \left\lfloor \frac{T}{k} \right\rfloor \right] \text{ for } i = 0, ..., k$$

Let us consider a bloc on the data $[T1, T2]$, the empirical matrices being considered are :

— $\boldsymbol{C_{in}}$ the empirical covariance matrix on the data $[T1, T2]$ with eigenvectors $u_{in} = (u_{in}(1), ..., u_{in}(n))$

— $\boldsymbol{C_{out}}$ the empirical covariance matrix on the data $[0, T1] \cup [T2, T]$ with eigenvectors $u_{out} = (u_{out}(1), ..., u_{out}(n))$

On this bloc, the estimator of the eigenvalues is the following :

$$\xi(\lambda_i) = u_{out}^T(i)\boldsymbol{C_{in}}u_{out}(i)$$

A "real covariance matrix" can then be put together :

$$\boldsymbol{\Xi_k} = u_{out}\xi u_{out}^T$$

Finally the covariance matrix estimator is the mean of these estimators on the different blocs :

$$\boldsymbol{\Xi} = \frac{1}{K}\sum_{k=1}^{K}\boldsymbol{\Xi_k}$$

Remark : There seems to exist an optimal number of blocs to run crossvalidation and this is still an open problem for covariance optimization. We did an empirical study on this subject for certain types of covariance matrices.

These different estimators were introduced, tested using the Python langage and then illustrated on examples. After this, the estimators were also tested on French and American financial data to show that they can significantly correct a part of the error due to the use of the empirical estimator for covariance estimation.

11 Annexes

11.1 Preuve pour l'estimateur de Ledoit-Péché, cas gaussien

On se place désormais dans le cas de données générées par une loi gaussienne pour présenter la démonstrations. Cette preuve provient de l'article *A very short proof of Ledoit-Péché's RIE formula for covariance matrices* [1].

L'objectif est d'estimer au mieux $\mathbf{\Sigma}$ dans l'ensemble des estimateurs invariants par rotation (RIE : *rotationnally invariant estimator*). Cette hypothèse d'invariance trouve une cohérence dans le cadre de l'estimation bayésienne : comme on dispose de peu de données, les vecteurs propres ne peuvent pas être estimés avec précision. Ainsi, chaque base orthonormée de vecteurs propres est considérée comme équiprobable et on décide donc de conserver les vecteurs propres de l'estimateur empirique en l'absence d'informations supplémentaires.

On obtient alors un problème d'optimisation que l'on peut formuler de la manière suivante :

$$\underset{RIEestimators}{\text{argmin}} \quad \|Estimator - \mathbf{\Sigma}\|_F$$

avec $\|.\|_F$ la norme de Frobenius : $\|\boldsymbol{M}\|_F = \sqrt{Tr(\boldsymbol{M}\boldsymbol{M}^{\boldsymbol{T}})}$

On obtient alors directement l'estimateur invariant par rotation optimal :

$$\xi(E) = \widehat{E} = \sum_{k=1}^{n} \widehat{\lambda}_k u_k u_k^T$$

avec

$$\xi(\lambda_k) = \widehat{\lambda}_k = u_k^T \mathbf{\Sigma} u_k$$

On remarque que $(\xi(\lambda_k))$ dépend de $\mathbf{\Sigma}$ qui nous est a priori inconnue. Pour cela, on appelle ξ la fonction oracle. L'objet de la preuve va être d'approximer cette quantité à l'aide de variables connues.

Dans un premier temps nous pouvons considérer les valeurs propres de $\boldsymbol{E}$ et $\widehat{\boldsymbol{E}}$:

$$\sum_{k=1}^{n} \delta_{\lambda_k} \qquad \text{et} \qquad \sum_{k=1}^{n} u_k^T \mathbf{\Sigma} u_k \delta_{\lambda_k}$$

L'article définit les quantités suivantes :

— $G(z) = \tau((z - \boldsymbol{E})^{-1})$ la trace normalisée de la résolvante de $\boldsymbol{E}$

— $L(z) = \tau(\mathbf{\Sigma}(z - \boldsymbol{E})^{-1})$

Les mesures empiriques sont discrètes mais grâce à la formule d'inversion de la transformée de Stieltjes qui approxime la densité des valeurs propres, nous allons pouvoir passer à des quantités continues, on rappelle cette formule [17] :

Proposition 8. *Soit g_μ la transformée de Stieltjes associée à la mesure μ. La transformée de Stieltjes et la mesure sont liées par la relation suivante :*

$$\forall z = x + i\eta \in \mathbb{C}, \mu(x) = -\frac{1}{\pi} \lim_{\eta \to 0^+} (Img_\mu(x + i\eta)dx)$$

(Formule d'inversion de la Stieltjes)

Nous allons lier ξ à la quantité L en se rappelant que nous sommes dans le cadre Bayésien (hypothèse de l'équiprobabilité des bases de vecteurs propres) et en utilisant la formule de l'espérance totale :

$$\begin{aligned}\tau(\xi(\boldsymbol{E})(z\mathbf{Id}-\boldsymbol{E})^{-1}) &= \tau(\mathbf{E}(\boldsymbol{\Sigma}|\boldsymbol{E})(z\mathbf{Id}-\boldsymbol{E})^{-1}) \\ &= \tau(\boldsymbol{\Sigma}(z\mathbf{Id}-\boldsymbol{E})^{-1}] = L(z)\end{aligned}$$

Si on écrit $L(z)$ dans sa forme continue, on obtient :

$$L(z) = \int_{\mathbb{R}} \rho_E(\lambda)\frac{\xi(\lambda)}{z-\lambda}d\lambda$$

Il nous reste ensuite à utiliser la formule d'inversion de la transformée de Stieltjes sur les quantités L et G ce qui nous donnera une nouvelle formulation de ξ :

$$\lim_{\eta\to 0^+} ImL(\lambda+i\eta) = -\pi\rho_E(\lambda)\xi(\lambda)$$

$$\lim_{\eta\to 0^+} ImG(\lambda+i\eta) = -\pi\rho_E(\lambda)$$

Soit $\epsilon > 0$ tel que $[\lambda_k-\epsilon, \lambda_k+\epsilon]\cap\{\lambda_1,...,\lambda_n\} = \{\lambda_k\}$:

$$\widehat{\lambda_k} = \lim_{\eta\to 0^+} \frac{\int_{\lambda_k-\epsilon}^{\lambda_k+\epsilon} ImL(x+i\eta)dx}{\int_{\lambda_k-\epsilon}^{\lambda_k+\epsilon} ImG(x+i\eta)dx}$$

Le résultat principal de l'article est l'approximation de $L(z)$ par des quantités connues :

Théorème 9.

$$\forall z\in\mathbb{C}\setminus\mathbb{R}, L(z) = 1 - \frac{1}{1-q+zG(z)} + o(1)$$

avec $o(1)$ une quantité qui tend vers 0 pour $n,T\to+\infty$ tel que $\frac{n}{T}\to q$, avec q une constante.

On rappelle que dans le cas où n et T tendent vers l'infini avec $\frac{n}{T} = q$, la densité des valeurs propres de $\boldsymbol{C}$ est connue par le théorème de Marcenko-Pastur.

Ici on cherche une sorte de développement limité de $L(z)$ pour avoir un résultat similaire dans le cas où n et T sont grands mais finis, typiquement n sera de l'ordre d'une ou plusieurs centaines.

La preuve dans le cas gaussien s'appuie sur deux résultats : un corolaire de la formule de Stein et la concentration de la mesure.

Proposition 9. *Soit $X = (X_1, ..., X_d)$ un vecteur normal centré de covariance $\mathbf{\Sigma}$ et $f : \mathbb{R}^d \to \mathbb{R}$ une function de classe $\mathbf{C}^1$ (dérivable et de dérivée continue) tel que son gradient ait au plus une croissance polynomiale en $+\infty$:*

$$\forall i_0 \in [|1, d|], \mathbb{E}(X_{i_0}) f(X_1, ..., X_d) = \sum_{i=1}^{n} \Sigma_{i_0 k} \mathbb{E}(\partial_k F)(X) X)_k$$

(Formule de Stein)

Corollaire 1. *Soit $X = (X_1, ..., X_{d)}$ un vecteur normal centré de covariance $\mathbf{\Sigma}$, $F : \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^d$ de gradient avec une croissance au plus polynomiale en $+\infty$:*

$$\mathbb{E}X' F(X) X = Tr\Sigma \mathbb{E}F(X) + \sum_{i=1}^{d} (\mathbb{E}\Sigma(\partial_k F)(X) X)_k$$

Proposition 10. *Soit $X = (X_1, ..., X_{d)}$ une vecteur normal centré de covariance Σ et $f : \mathbb{R}^d \to \mathbb{R}$ une $\mathbf{C}^1$ fonction de classe $\mathbf{C}^1$ de gradient ∇f :*

$$Var(f(X) \leq \mathbb{E}\|\nabla f(X)\|^2$$

De plus, si f est k-Lipschitz, on a :

$$\forall t > 0, \mathbb{P}(|f(X) - \mathbb{E}f(X)| \geq t) \leq 2 \exp^{-\frac{t^2}{2k^2}}$$

(Concentration de la mesure pour les vecteurs gaussiens)

Définissons $\boldsymbol{H} = \frac{1}{T} Tr\boldsymbol{GE}$, on va chercher à calculer son espérance, le but sera à terme de lier cette quantité à $L(z) = \tau(\boldsymbol{G\Sigma})$:

$$\mathbb{E}Tr\boldsymbol{GE} = \mathbb{E}\frac{1}{T} \sum_t Tr\boldsymbol{GX(t)X(t)^T} = \frac{1}{T} \mathbb{E}\boldsymbol{X(t)^T GX(t)}$$

Par la formule de Stein :

$$\mathbb{E}Tr\boldsymbol{GE} = \frac{1}{T} \sum_t \mathbb{E}Tr\boldsymbol{G\Sigma} + \frac{1}{T} \sum_t \sum_{k=1}^{n} \mathbb{E}e_k^T \mathbf{\Sigma} (\frac{\partial}{\partial X(t)_k} G) X(t)$$

On calcule les dérivées partielles :

$$\frac{\partial}{\partial X(t)_k} \boldsymbol{G} = \frac{1}{T} \boldsymbol{G}(X(t) e_k^T + e_k X(t)^T) \boldsymbol{G}$$

D'où :

$$\mathbb{E}Tr\boldsymbol{GE} = \frac{1}{T} \sum_t \mathbb{E}Tr\boldsymbol{G\Sigma} + \frac{1}{T^2} \sum_t \sum_{k=1}^{n} \mathbb{E}e_k^T \mathbf{\Sigma} (G(X(t) e_k^T + e_k X(t)^T) \boldsymbol{G}) X(t)$$

$$= \frac{1}{T}\sum_t \mathbb{E}Tr\boldsymbol{G\Sigma} + \frac{1}{T^2}\sum_t\sum_{k=1}^{n}\mathbb{E}(e_k^T\boldsymbol{\Sigma G}X(t)e_k^T\boldsymbol{G}X(t) + e_k^T\boldsymbol{\Sigma G}e_kX(t)^T\boldsymbol{G}X(t))$$

$$= \frac{1}{T}\sum_t \mathbb{E}Tr\boldsymbol{G\Sigma} + \frac{1}{T^2}\sum_t\mathbb{E}(X(t)^T\boldsymbol{G\Sigma G}X(t) + tr(\boldsymbol{\Sigma G})X(t)^T\boldsymbol{G}X(t))$$

$$= \frac{1}{T}\sum_t \mathbb{E}Tr\boldsymbol{G\Sigma} + \frac{1}{T}\mathbb{E}Tr\boldsymbol{G\Sigma GE} + \frac{1}{T}\mathbb{E}Tr(\boldsymbol{\Sigma G})Tr(\boldsymbol{GE})$$

Il s'agit maintenant de lier les traces avec les quantités L ($= \tau(\boldsymbol{G\Sigma})$) et $\boldsymbol{H}$ ($= \tau(\boldsymbol{GE})$) :

- $\frac{1}{T}\sum_t \mathbb{E}Tr\boldsymbol{G\Sigma} = TL$ par définition de L(z)
- $\frac{1}{T}\mathbb{E}Tr\boldsymbol{G\Sigma GE} = o(1)$ admis ([1])
- $\frac{1}{T}\mathbb{E}Tr(\boldsymbol{\Sigma G})Tr(\boldsymbol{GE}) = TLH + o(1)$ par concentration de la mesure pour les vecteurs gaussiens

On obtient alors le développement suivant :

$$H = L + LH + o(1)$$

Dans la mesure où nous souhaitons approximer L, il nous reste à prouver que l'on peut diviser par $1 + H$. L sera alors écrite comme une fonction de H qui est une quantité connue.

Une première remarque est que :

$$\boldsymbol{G(z)}(z - \boldsymbol{E}) = \boldsymbol{Id_n} \Rightarrow z\boldsymbol{G(z)} - \boldsymbol{Id_n} = \boldsymbol{GE}$$

On définit la constante $q = \frac{n}{T}$ et on pose $z = x + i\eta$ avec $\eta > 0$.

$$H(z) = \tau(\boldsymbol{GE}) = \tau(z\boldsymbol{G}) - \frac{n}{T} = z\boldsymbol{G} - q = \frac{q}{n}\sum_{i=1}^{n}\frac{\lambda_i}{z - \lambda_i}$$

$$|Im(H(z))| = \frac{q|\eta|}{n}\sum_{i=1}^{n}\frac{\lambda_i}{(x - \lambda_i)^2 + \eta^2} \geq \frac{|\eta|Tr\boldsymbol{E}/T}{2x^2 + 2\|\boldsymbol{E}\|^2 + \eta^2}$$

On prouve ensuite que la partie imaginaire de H est non nulle. On a pour cela besoin d'un lemme que l'on va admettre :

Lemme 1. *Supposons que $n, T \to \infty$ tel que $q = \frac{n}{T}$ est borné et que $\boldsymbol{\Sigma}$ est bornée par la norme opérateur. Alors il existe une constant $C > 0$ tel qu'avec une probabilité tendant vers 1 on ait :*

$$\|\boldsymbol{E}\| \leq qC\|\boldsymbol{\Sigma}\| \text{ and } Tr\boldsymbol{E} \geq \frac{Tr\boldsymbol{\Sigma}}{T}$$

Par ce lemme, il existe $c > 0$ tel que $|1 + H| > c$ avec une probabilité qui tend vers 1. Alors :

$$L = \frac{H}{1+H} + o(1) = 1 - \frac{1}{1+H} + o(1)$$

Pour $z \to \lambda \in \mathbb{R}$:

$$\frac{ImL(z)}{ImG(z)} \sim \frac{Im\frac{-1}{1+H}}{ImG} \sim \frac{-Im\frac{1+\overline{H}}{|1+\overline{H}|^2}}{ImG} \sim \frac{\frac{\lambda Im(G)}{|1+H|^2}}{ImG} \sim \frac{\lambda}{|1+H|^2} \sim \frac{\lambda}{|1-q+\lambda G|^2}$$

En résumé, la formule de Ledoit-Péché est la suivante :

Théorème 10. *Pour $z = \lambda + i\eta$ avec $\eta > 0$:*

$$\xi(\lambda) = \lim_{\eta \to 0^+} \frac{\lambda}{|1 - q + q\lambda g_\mu(\lambda + i\eta)|^2}$$

(Formule de Ledoit-Péché)

11.2 Crossvalidation cas Identité, calcul avec la théorie des perturbations

On se place dans le cas $\boldsymbol{C} = \boldsymbol{Id}$. On génère donc des vecteurs à l'aide d'une loi normale multivariée de moyenne nulle et de matrice variance-covariance égale à $\boldsymbol{C}$. On note n (la taille de la matrice $\boldsymbol{C}$), T (le nombre de données générées), K (le nombre de blocs), T_B la taille de chaque bloc.
On suppose de plus que $K >> 1$ et $T_B << T$ et on définit les matrices empiriques suivantes :

- $\boldsymbol{E} = \frac{1}{T}\boldsymbol{H}\boldsymbol{H^T}$, la matrice empirique des observations sur tout l'échantillon
- $\boldsymbol{E_1} = \frac{1}{T_B}\boldsymbol{H_1}\boldsymbol{H_1^T}$, la matrice empirique des observations sur le premier bloc
- $\boldsymbol{E_{\overline{1}}} = \frac{1}{T-T_B}\boldsymbol{H_{\overline{1}}}\boldsymbol{H_{\overline{1}}^T}$, avec $(v_{\overline{1}}^l)_{1\leq l\leq n}$ une base de vecteurs propres orthonormée de $\boldsymbol{E_{\overline{1}}}$
- $\boldsymbol{E_{\overline{12}}} = \frac{1}{T-2T_B}\boldsymbol{H_{\overline{12}}}\boldsymbol{H_{\overline{12}}^T}$, la matrice empirique des observations sur l'échantillon privé des deux premiers blocs.

On note $\boldsymbol{\Xi_k}$ l'estimateur de crossvalidation sur le bloc numéro k ($1 \leq k \leq n$) :

$$\boldsymbol{\Xi_k} = \sum_{l=1}^{n} (v_{\overline{1}}^{l^T} \boldsymbol{E_1} v_{\overline{1}}^l) v_{\overline{1}}^l v_{\overline{1}}^{l^T}$$

L'estimateur final de crossvalidation n'étant que la moyenne sur l'ensemble des blocs :

$$\boldsymbol{\Xi} = \frac{1}{K}\sum_{k=1}^{K} \boldsymbol{\Xi_k}$$

À partir de là on pose :

- $\widetilde{\boldsymbol{E_1}} = \boldsymbol{E_1} - \boldsymbol{Id}$
- $\widetilde{\boldsymbol{\Xi_k}} = \sum_{l=1}^{n} (v_{\overline{1}}^{l^T} \widetilde{\boldsymbol{E_1}} v_{\overline{1}}^l) v_{\overline{1}}^l v_{\overline{1}}^{l^T}$
- $\widetilde{\boldsymbol{\Xi}} = \frac{1}{K}\sum_{k=1}^{K} \widetilde{\boldsymbol{\Xi_k}}$

Retirer l'identité permet d'avoir des matrices de covariance empiriques d'espérance nulle et cela simplifiera les calculs.

On rappelle que les les $\boldsymbol{\Xi_k}$ sont d'espérances égales, l'erreur de crossvalidation s'exprime alors comme l'espérance de la quantité suivante :

$$\frac{1}{n}Tr(\widetilde{\boldsymbol{\Xi}}^2) = \frac{1}{nK^2}\sum_{a,b} Tr(\widetilde{\boldsymbol{\Xi_1}}\widetilde{\boldsymbol{\Xi_b}}) = \frac{1}{nK}Tr(\widetilde{\boldsymbol{\Xi_1^2}}) + \frac{K^2-K}{nK}Tr(\widetilde{\boldsymbol{\Xi}}_1\widetilde{\boldsymbol{\Xi}}_2)$$

On commence par estimer le premier terme qui est plus simple :

$$\widetilde{\boldsymbol{\Xi}_1}^2 = \sum_{l,k} (v_{\overline{1}}^{l^T} \widetilde{\boldsymbol{E_1}} v_{\overline{1}}^l) v_{\overline{1}}^l v_{\overline{1}}^{l^T} (v_{\overline{1}}^{k^T} \widetilde{\boldsymbol{E_1}} v_{\overline{1}}^k) v_{\overline{1}}^k v_{\overline{1}}^{k^T}$$

On rappelle que les termes $(v_{\overline{1}}^{l^T} \widetilde{\boldsymbol{E_1}} v_{\overline{1}}^l)$ sont des réels, c'est l'expression de $\widetilde{\boldsymbol{E_1}}(ll)$ dans la base des $v_{\overline{1}}$. De plus cette base est orthonormée donc : $v_{\overline{1}}^{l^T} v_{\overline{1}}^l = \delta_{lk}$.

$$\widetilde{\boldsymbol{\Xi_1}}^2 = \sum_{l} (v_{\overline{1}}^{l^T} \widetilde{E}_1 v_{\overline{1}}^l)^2 v_{\overline{1}}^l v_{\overline{1}}^{l^T}$$

En passant à la trace :

$$Tr(\widetilde{\boldsymbol{\Xi_1}}^2) = \sum_{l} (v_{\overline{1}}^{l^T} \widetilde{\boldsymbol{E_1}} v_{\overline{1}}^l)^2$$

On veut désormais estimer l'espérance de cette trace. Il s'agit de la somme des carrés des éléments diagonaux de $\widetilde{\boldsymbol{E_1}}$. Les éléments diagonaux de $\boldsymbol{E_1}$ sont des variances empiriques de variables aléatoires gaussiennes d'espérance, on leur a ensuite retranché l'identité. Soit σ_l^2 la variance empirique de $\widetilde{\boldsymbol{E_1}}$, on rappelle que les données sont générées à partir d'une loi gaussienne multivariée de covariance identité :

$$\mathbb{E}((\sigma_l^2-1)^2) = \mathbb{E}(\sigma_l^4) - 2\mathbb{E}\sigma_l^2 + 1 = 2$$

En sommant, on a :

$$\frac{1}{K}\mathbb{E}(Tr(\widetilde{\boldsymbol{\Xi_1}}^{\,2})) = \frac{2n}{KT_B} = \frac{2n}{T} = 2q$$

On va maintenant nous intéresser au second terme de l'erreur : $Tr(\widetilde{\boldsymbol{\Xi_1}}\widetilde{\boldsymbol{\Xi_2}})$.

Avec quelques manipulations simples on peut écrire :

$$\widetilde{\boldsymbol{E_1}} = \frac{T-2T_B}{T-T_B}[\boldsymbol{E_{\overline{12}}} + \frac{T_B}{T-2T_B}\boldsymbol{E_2}]$$

$$\widetilde{\boldsymbol{E_2}} = \frac{T-2T_B}{T-T_B}[\boldsymbol{E_{\overline{12}}} + \frac{T_B}{T-2T_B}\boldsymbol{E_2}]$$

$\frac{T-2T_B}{T-T_B}$ n'est qu'une constante de normalisation qui n'a pas d'influence sur les vecteurs propres. On pose $\alpha = \frac{T_B}{T-2T_B}$, la théorie des perturbations nous donne une approximation des vecteurs propres de $\boldsymbol{E_1}$ et de $\boldsymbol{E_2}$ dans la base des vecteurs propres de $\boldsymbol{E_{\overline{12}}}$ si α est suffisamment petit :

$$v_{\overline{1}}^l = v_{\overline{12}}^l + \alpha \sum_{m\neq l} \frac{v_{\overline{12}}^{m^T} E_2 v_{\overline{12}}^l}{\lambda_l^{\overline{12}} - \lambda_m^{\overline{12}}} v_{\overline{12}}^m$$

$$v_{\overline{12}}^k = v_{\overline{12}}^k + \alpha \sum_{m\neq k} \frac{v_{\overline{12}}^{m^T} E_1 v_{\overline{12}}^k}{\lambda_k^{\overline{12}} - \lambda_m^{\overline{12}}} v_{\overline{12}}^m$$

Avec un peu de calculs on trouve que les termes d'ordre 0 et 1 en α dans le développement de $Tr(\boldsymbol{\Xi_1}\boldsymbol{\Xi_2})$ sont nuls. On écrit donc les termes d'ordre 2 :

$$Tr(\boldsymbol{\Xi_1}\boldsymbol{\Xi_2}) = 2\alpha^2 \sum_l \sum_{m_1\neq l} \frac{(v_{\overline{12}}^{m_1^T}\boldsymbol{E_2}v_{\overline{12}}^l)}{(\lambda_l^{\overline{12}} - \lambda_{m_1}^{\overline{12}})}(v_{\overline{12}}^{m_1^T}\boldsymbol{E_1}v_{\overline{12}}^l) \sum_{m_2\neq l} \frac{(v_{\overline{12}}^{m_2^T}\boldsymbol{E_1}v_{\overline{12}}^l)}{(\lambda_l^{\overline{12}} - \lambda_{m_2}^{\overline{12}})}(v_{\overline{12}}^{m_2^T}\boldsymbol{E_2}v_{\overline{12}}^l)$$

$$+2\alpha^2 \sum_l \sum_{m_1\neq l} \frac{(v_{\overline{12}}^{m_1^T}\boldsymbol{E_2}v_{\overline{12}}^l)}{(\lambda_l^{\overline{12}} - \lambda_{m_1}^{\overline{12}})}(v_{\overline{12}}^{l^T}\boldsymbol{E_1}v_{\overline{12}}^{m_1}) \sum_{m_2\neq l} \frac{(v_{\overline{12}}^{m_2^T}\boldsymbol{E_1}v_{\overline{12}}^l)}{(\lambda_l^{\overline{12}} - \lambda_{m_2}^{\overline{12}})}(v_{\overline{12}}^{l^T}\boldsymbol{E_2}v_{\overline{12}}^{m_2})$$

$$+2\alpha^2 \sum_l \sum_{m_1\neq l} \frac{(v_{\overline{12}}^{m_1^T}\boldsymbol{E_2}v_{\overline{12}}^l)}{(\lambda_l^{\overline{12}} - \lambda_{m_1}^{\overline{12}})}(v_{\overline{12}}^{l^T}\boldsymbol{E_1}v_{\overline{12}}^l)\frac{(v_{\overline{12}}^{m_1^T}\boldsymbol{E_1}v_{\overline{12}}^l)}{(\lambda_l^{\overline{12}} - \lambda_{m_1}^{\overline{12}})}(v_{\overline{12}}^{l^T}\boldsymbol{E_2}v_{\overline{12}}^l)$$

$$+2\alpha^2 \sum_l \sum_{m_1\neq l} \frac{(v_{\overline{12}}^{m_1^T}\boldsymbol{E_2}v_{\overline{12}}^l)}{(\lambda_l^{\overline{12}} - \lambda_{m_1}^{\overline{12}})}(v_{\overline{12}}^{l^T}\boldsymbol{E_1}v_{\overline{12}}^l)\frac{(v_{\overline{12}}^{l^T}\boldsymbol{E_1}\widetilde{v}_{\overline{12}}^{m_1})}{(\lambda_l^{\widetilde{12}} - \lambda_{m_1}^{\overline{12}})})(v_{\overline{12}}^{l^T}\boldsymbol{E_2}v_{\overline{12}}^l) + o(\alpha^2)$$

Nous devons désormais déterminer l'espérance de cette trace. Prenons le premier terme de la somme :

$$\mathbb{E}(2\alpha^2 \sum_l \sum_{m_1 \neq l} \frac{(v_{\overline{12}}^{m_1^T} \boldsymbol{E_2} v_{\overline{12}}^l)}{(\lambda_l^{\overline{12}} - \lambda_{m_1}^{\overline{12}})} (v_{\overline{12}}^{m_1^T} \boldsymbol{E_1} v_{\overline{12}}^l) \sum_{m_2 \neq l} \frac{(v_{\overline{12}}^{m_2^T} \boldsymbol{E_1} v_{\overline{12}}^l)}{(\lambda_l^{\overline{12}} - \lambda_{m_2}^{\overline{12}})} (v_{\overline{12}}^{m_2^T} \boldsymbol{E_2} v_{\overline{12}}^l)$$

$$= \alpha^2 \sum_l \sum_{m_1 \neq l, m_2 \neq l} \mathbb{E}((E_1)_{m_1 l}(E_1)_{m_2 l}(E_2)_{m_1 l}(E_2)_{m_2 l}$$

Les éléments de matrices considérés précedemment sont exprimés dans la base des vecteurs propres $(v_{\overline{12}}^l)$. On utilise que $\boldsymbol{E_1}$ et $\boldsymbol{E_2}$ sont indépendantes :

$$= \alpha^2 \sum_l \sum_{m_1 \neq l, m_2 \neq l} \mathbb{E}((E_1)_{m_1 l}(E_1)_{m_2 l}) \mathbb{E}((E_2)_{m_1 l}(E_2)_{m_2 l}) \frac{1}{(\lambda_l^{\overline{12}} - \lambda_{m_1}^{\overline{12}})(\lambda_l^{\overline{12}} - \lambda_{m_2}^{\overline{12}})}$$

Lorsque la matrice de covariance $\boldsymbol{C} = \boldsymbol{Id}$ comme ici, les matrices empiriques sont des Wishart blanches, nous rappelons la formule de Wick aussi appelée théorème d'Isserlis nécessaire pour calculer l'espérance [19] :

Théorème 11. *Soit $(X_1, X_2, ..., X_r)$ des gaussiennes multivariées indépendantes de moyenne nulle. On a :*

$$\mathbb{E}[X_1 X_2 ... X_n] = \sum_{p \in P_n^2} \prod_{\{i,j\} \in p} \mathbb{E}[X_i X_j] = \sum_{p \in P_n^2} \prod_{\{i,j\} \in p} Cov(X_i X_j)$$

(Formule de Wick)

Pour $m_1 \neq l$, $m_2 \neq l$:

$$\mathbb{E}(W_{m_1 l} W_{m_2 l}) = \frac{1}{T_B^2} \sum_{t_1, t_2} \mathbb{E}(H_{m_1 t_1} H_{l t_1} H_{m_2 t_2} H_{l t_2})$$

Où $\boldsymbol{H}$ est une matrice à entrées gaussiennes iid centrées réduites.

On commence par calculer le terme générique de la somme :

$$\mathbb{E}(H_{m_1 t_1} H_{l t_1} H_{m_2 t_2} H_{l t_2}) = \mathbb{E}(H_{m_1 t_1} H_{l t_1}) \mathbb{E}(H_{m_2 t_2} H_{l t_2}) + \mathbb{E}(H_{m_1 t_1} H_{m_2 t_2}) \mathbb{E}(H_{l t_1} H_{l t_2}) + \mathbb{E}(H_{m_1 t_1} H_{l t_2}) \mathbb{E}(H_{l t_1} H_{m_2 t_2})$$

Il suffit alors de calculer les trois termes :

- $\mathbb{E}(H_{m_1 t_1} H_{l t_1}) \mathbb{E}(H_{m_2 t_2} H_{l t_2}) = \delta_{m_1 l} \delta_{m_2 l} = 0$ car $m_1, m_2 \neq 0$
- $\mathbb{E}(H_{m_1 t_1} H_{m_2 t_2}) \mathbb{E}(H_{l t_1} H_{l t_2}) = \delta_{m_1 m_2} \delta_{t_1 t_2}$
- $\mathbb{E}(H_{m_1 t_1} H_{l t_2}) \mathbb{E}(H_{l t_1} H_{m_2 t_2}) = \delta_{m_1 l} \delta_{t_1 t_2} \delta_{m_2 l} = 0$

Donc :

$$\mathbb{E}(W_{m_1 l} W_{m_2 l}) = \frac{1}{T_B^2} \sum_{t_1, t_2} \delta_{m_1 m_2} \delta_{t_1 t_2} = \frac{1}{T_B} \delta_{m_1 m_2}$$

Nous pouvons maintenant revenir au terme en α^2 :

$$\alpha^2 \sum_l \sum_{m_1 \neq l, m_2 \neq l} \mathbb{E}((E_1)_{m_1 l}(E_1)_{m_2 l}) \mathbb{E}((E_2)_{m_1 l}(E_2)_{m_2 l}) \frac{1}{(\lambda_l^{\overline{12}} - \lambda_{m_1}^{\overline{12}})(\lambda_l^{\overline{12}} - \lambda_{m_2}^{\overline{12}})}$$

$$= \alpha^2 \sum_l \sum_{m_1 \neq l, m_2 \neq l} \frac{1}{T_B} \delta_{m_1 m_2} \frac{1}{(\lambda_l^{\overline{12}} - \lambda_{m_1}^{\overline{12}})(\lambda_l^{\overline{12}} - \lambda_{m_2}^{\overline{12}})}$$

$$= \alpha^2 \sum_l \sum_{m_1 \neq l} \frac{1}{T_B} \frac{1}{(\lambda_l^{\overline{12}} - \lambda_{m_1}^{\overline{12}})(\lambda_l^{\overline{12}} - \lambda_{m_1}^{\overline{12}})}$$

En utilisant la même méthode sur les autres termes, on a finalement :

$$Tr(\mathbf{\Xi_1} \tilde{\Xi}_2) = 4\alpha^2 \sum_l \sum_{m_1 \neq l} \frac{1}{T_B} \frac{1}{(\lambda_l^{\overline{12}} - \lambda_{m_1}^{\overline{12}})(\lambda_l^{\overline{12}} - \lambda_{m_1}^{\overline{12}})} + o(\alpha^2)$$

Toutefois ce terme n'est pas convergeant en général pour des matrices à valeurs réelles. Ce calcul nous apprend donc que l'hypothèse des petites perturbations des vecteurs propres dans le cadre de la crossvalidation n'est pas valable pour notre problème. Les vecteurs propres $(v_{\overline{1}}^l)$ et $(v_{\overline{2}}^l)$ sont donc fortement perturbés par rapport aux vecteurs propres $(v_{\overline{12}}^l)$, et ce, même quand le nombre de blocs est grand.

Table des figures

Références

[1] Benaych-Georges, F. *A very short proof of Ledoit-Péché's RIE formula for covariance matrices*. http://www.cmapx.polytechnique.fr/ benaych/Short_ proof_of_Ledoit_Peche.pdf

[2] Benaych-Georges F., Bouchaud J. P. and Potters M. *Optimal cleaning for singular values of crosscovariance matrices*. preprint, arXiv :1901.05543, 2019.

[3] Bun, J., Bouchaud, J.-P., Potters, M. *Cleaning correlation matrices*. Risk magazine, 2016.

[4] Bun, J., Bouchaud, J.-P., Potters, M. *Cleaning large correlation matrices : Tools from Random Matrix Theory*. Physics Reports Volume 666, Review article, pp. 1–109, 2017.

[5] Collins B., Nechita I. *Random matrix techniques in quantum information theory*. Journal of Mathematical Physics, 57(1), 2016.

[6] Cont, R. *Empirical Properties of Asset Returns : Stylized Facts and Statistical Issues*. Quantitative Finance, 1(2), pp.223–236, 2001.

[7] Christoffersen, P.F. *Elements of Financial Risk Management*. Academic Press, San Diego, 2003.

[8] Fama, E. F., French, K. R. *Common risk factors in the returns on stocks and bonds*. Journal of Financial Economics 33(1), pp. 3-56, 1993.

[9] Keating J., Snaith N. *Random matrix theory and* $\zeta(1/2+it)$. Communications in Mathematical Physics, 214, pp. 57–89, 2000.

[10] Ledoit, O., Péché, S. *Eigenvectors of some large sample covariance matrix ensembles*. Probability Theory and Related Fields, 151(1), pp. 233–264, 2011.

[11] Ledoit, O. and Wolf, M. *Honey, I shrunk the sample covariance matrix*. Journal of Portfolio Management, 30(4), pp. 110–119, 2004.

[12] Ledoit, O., Wolf, M. *Nonlinear shrinkage estimation of large-dimensional covariance matrices*. The Annals of Statistics, 40(2), pp. 1024-1060, 2012.

[13] Ledoit O., Wolf M. *Analytical Nonlinear Shrinkage of Large-Dimensional Covariance Matrices*. University of Zurich, Department of Economics, Working Paper No. 264, 2018.

[14] Marčenko V. A. and Pastur L. A. *Distribution of eigenvalues for some sets of random matrices*. Matematicheskii Sbornik, 114(4), pp. 507–536, 1967.

[15] Markowitz H. *Portfolio Selection*. The Journal of Finance, Vol. 7 (1), pp. 77-91, 1952.

[16] Michaud R. *The Markowitz Optimization Enigma : Is 'Optimized' Optimal ?*. Financial Analysts Journal, 5(1), pp. 31—42, 1989.

[17] Potters M., Bouchaud J-P. *A First Course in Random Matrix Theory : For Physicists, Engineers and Data Scientists*. Cambridge University Press, 2020.

[18] Sandoval L., Franca I.D.P.. *Correlation of financial markets in times of crisis*. Physica A 391, pp. 187–208, 2012.

[19] Speicher R. *Free probability Theory (Lecture notes)*. https://arxiv.org/abs/1908.08125

[20] Tan V., Zohren S. *Large Non-Stationary Noisy Covariance Matrices : A Cross-Validation Approach*. Econometrics : Mathematical Methods & Programming eJournal, 2020.

[21] Tao T. *Topics in random matrix theory*. Graduate Studies in Mathematics, 132, American Mathematical Society, 2012.

[22] Eugene Wigner. *Characteristic vectors of bordered matrices with infinite dimensions*. Annals of mathematics, vol. 62, pp. 546–564, 1955.

[23] Wishart J. *The Generalised Product Moment Distribution In Samples From A Normal Multivariate Population*. Biometrika, Volume 20A(1-2), pp. 32–52, 1928.

[24] Yaskov, P. *A short proof of the Marchenko-Pastur theorem*. C. R. Math. Acad. Sci. Paris, 354, pp. 319–322, 2016.