

# Quelques notes sur l'investissement d'un portefeuille à un actif en présence d'information complémentaire au marché

Thierry Bazier-Matte

February 2, 2017

## Table des matières

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Exposition du problème et hypothèses . . . . .	2
1.2	Dimensionnalité de l'information . . . . .	5
1.3	Risque et garanties statistiques sur la décision . . . . .	5
1.4	Interprétations . . . . .	5
1.5	Objectifs . . . . .	5
<b>2</b>	<b>Optimisation moderne de portefeuille</b>	<b>6</b>
2.1	Théorie classique du portefeuille . . . . .	6
2.2	Portefeuille universel / Papiers d'Elad Hazan . . . . .	7
2.3	Théorie de portefeuille régularisé . . . . .	7
2.4	Fama and French et suivants? . . . . .	7
2.5	Articles du NIPS . . . . .	7
2.6	Papiers de Ben Van Roy . . . . .	7
2.7	Conclusions: Notre problème par rapport à ces deux disciplines . . . . .	7
<b>3</b>	<b>Décisions non linéaires</b>	<b>8</b>
3.1	Dualité . . . . .	8
3.2	Alternate problem . . . . .	9
<b>4</b>	<b>Garanties statistiques</b>	<b>12</b>
<b>5</b>	<b>Dimensionnalité du problème</b>	<b>12</b>
<b>6</b>	<b>Distributions synthétiques</b>	<b>12</b>
<b>7</b>	<b>Conclusion</b>	<b>12</b>

# 1 Introduction

## 1.1 Exposition du problème et hypothèses

Ce mémoire vise à établir clairement et rigoureusement comment un investisseur *averse au risque* disposant d'*information complémentaire* au *marché* peut utiliser cette information pour accroître son *utilité espérée* ou, de façon équivalente, son *rendement équivalent certain*.

**Modélisation du marché** Nous entendons ici par *marché* n'importe quel type d'actif financier ou spéculatif dans lequel on peut investir une partie de sa fortune dans l'espoir de la voir fructifier au cours d'une période de temps arbitraire. Ainsi, tout au long de l'exposé théorique qui suivra, il peut être pertinent d'avoir en tête les rendements quotidiens issus des grands indices boursiers (par exemple les 500 plus grandes capitalisations américaines). Cependant, le traitement qui sera développé pourrait tout aussi bien s'appliquer à une action cotée en bourse dont on considère les rendements mensuels.<sup>[Nécessaire?]</sup> Mathématiquement, l'idée de marché peut ainsi être réduite à celle d'une variable aléatoire  $R(t)$  décrivant l'évolution du rendement de l'actif en question.

Relativement à l'idée de marché, nous ferons également l'hypothèse que l'univers a une influence sur ces rendements. Il serait par exemple raisonnable de croire que le prix du pétrole a une influence sur l'évolution du rendement du marché américain. De la même façon, l'annonce d'un scandale aura à son tour des répercussions sur la valeur du titre de la compagnie dont il est l'objet. En outre, il a été montré par Fama et French que le rendement d'une action pouvait s'expliquer comme une combinaison de quelques facteurs fondamentaux (la taille de l'entreprise, le risque de marché et le ratio cours/valeur). On peut alors considérer un vecteur d'information  $\vec{X}(t) = (X_1(t), X_2(t), \dots)$  dont chaque composante représente une information particulière, par exemple l'absence ou la présence d'un certain type de scandale, un ratio comptable, le prix d'un certain actif financier.<sup>[Rephrase]</sup> D'un point de vue probabiliste, on dira donc qu'il existe une forme de dépendance entre  $R(t)$  et  $\{\vec{X}(\tau) \mid \tau < t\}$  l'ensemble des événements antérieurs à  $t$ . Le processus joint de ces deux événements sera désormais défini comme *la distribution totale de marché*, ou simplement le marché.

**Stationarité** Bien qu'un tel modèle permette de représenter de façon très générale l'évolution d'un marché, nous formulerons l'hypothèse supplémentaire selon laquelle le marché est un processus *stationnaire*. Ceci permet notamment d'évacuer la notion temporelle afin de ne représenter qu'une distribution de causes (l'information  $X$ ) et d'effet (l'observation des rendements  $R$ ). Cette hypothèse est assez contraignante. Elle suppose d'une part que les réalisations passées n'ont aucun effet sur les réalisations futures (indépendance) et d'autre part que la distribution de marché est figée dans le temps, ce qui implique notamment l'absence de probabilité de faillite. Elle implique aussi que le marché ne peut être vu comme un environnement adversarial qui réagirait par exemple aux décisions d'un investisseur. Ceci vient notamment mettre en cause

la théorie des marchés efficients selon laquelle une brèche dans l'absence d'arbitrage serait immédiatement colmatée par des spéculateurs (effet d'autorégulation). Nous aurons toutefois l'occasion de revenir plus en détail sur les liens à faire entre cet exposé et l'effcience des marché.

**Approche mathématique et statistique** Dans ce qui suit, nous noterons par  $M$  la distribution de marché. Le vecteur aléatoire d'information sera par ailleurs formé de  $m$  composantes; pour l'instant, aucune hypothèse par rapport à la dépendance des composantes de  $X$  ne sera formulée. À ce point-ci, on a donc le modèle de marché suivant:

$$M = (R, X_1, \dots, X_m). \quad (1)$$

On fera également l'hypothèse qu'on possède un ensemble de  $n$  éléments échantillonnés à partir de  $M$ , de sorte que:

$$\{r_i, x_{i1}, \dots, x_{im}\}_{i=1}^n \sim M \quad (2)$$

représente notre ensemble d'échantillonnage (aussi appelé ensemble d'entraînement). Le domaine des rendements possibles de  $R$  sera noté  $\mathbf{R} \subseteq \mathcal{R}$  et celui du vecteur d'information  $X$  sera noté  $\mathbf{X} \subseteq \mathcal{R}^m$ . Le vecteur d'observations de rendement sera noté  $r \in \mathcal{R}^n$  et la matrice d'information par  $X \in \mathcal{R}^{n \times m}$ .

**Modélisation de la préférence** Indépendamment de la notion de marché, on a d'autre part l'aspect d'aversion au risque qui est modélisé par une fonction d'utilité  $u : \mathbf{R} \rightarrow \mathbf{U}$ , où  $\mathbf{R} \subseteq \mathcal{R}$  est le domaine (fermé ou non) des rendements considérés et  $\mathbf{U} \subseteq \mathcal{R}$  celui des *utilités*.

Bien qu'en pratique il soit plus facile de travailler sur des fonctions possédant des valeurs dans  $\mathbf{U}$ , en pratique cet espace est adimensionnel<sup>[Citation needed]</sup>, de sorte que nos résultats seront présentés dans l'espace des rendements  $\mathbf{R}$ .

**Fonction de décision** Donnés ces éléments de base, le but de ce mémoire sera alors de déterminer une fonction de décision d'investissement  $q : \mathbf{X} \rightarrow \mathbf{P} \subseteq \mathcal{R}$  maximisant l'utilité espérée de l'investissement.

Mathématiquement on a donc le problème fondamental suivant:

$$\underset{q \in \mathbf{Q}}{\text{maximiser}} \quad \mathbf{E}u(R \cdot q(X)), \quad (3)$$

où l'optimisation a lieu dans un espace de fonctions  $\mathbf{Q}$  à préciser.

Cependant, comme la distribution  $(X, R) = M$  est inconnue, il est impossible de déterminer la fonction  $q^*$  minimisant cet objectif. On dispose toutefois d'un échantillon de  $M$  dont on peut se servir pour approximer le problème (SAA, voir Shapiro<sup>[Citation needed]</sup>):

$$\underset{q \in \mathbf{Q}}{\text{maximiser}} \quad n^{-1} \sum_{i=1}^n u(r_i q(x_i)), \quad (4)$$

mais encore ici le problème est mal spécifié, puisqu'aucune contrainte n'a été posée sur l'espace  $\mathcal{Q}$ . Par exemple, il suffirait de prendre pour  $q$  un dictionnaire associant à  $x_i$  la valeur  $\alpha r_i$ , où  $\alpha > 0$ , et à toute autre valeur de  $x$  une valeur nulle pour avoir une valeur d'utilité arbitrairement grande à mesure que  $\alpha \rightarrow \infty$ .

**Risque in-échantillon et hors échantillon** une telle fonction  $q$  est qu'elle se généralise très mal. En effet pour toute observation  $x$  qui ne figurerait dans l'ensemble d'entraînement,  $q$  prescrirait alors un investissement nul. Il y a alors une énorme différence entre l'utilité observée au sein de notre échantillon et l'utilité hors échantillon.

Donnée une fonction de décision  $q \in \mathcal{Q}$  et un échantillon de  $M$ , on définit le *risque in-échantillon* ou *risque empirique* par

$$\hat{R}(q) = n^{-1} \sum_{i=1}^n \ell(r_i q(x_i)), \quad (5)$$

où  $\ell = -u$ . De la même façon, on définit le *risque hors-échantillon* ou *erreur de généralisation* par

$$R(q) = \mathbf{E} \ell(R \cdot q(X)). \quad (6)$$

On peut souhaiter d'une bonne fonction de décision qu'elle performe bien hors échantillon, aussi la quantité  $R(q) - \hat{R}(q)$  sera-t-elle primordiale et beaucoup d'attention lui sera consacrée dans les prochaines sections. Notons que le risque hors-échantillon étant théoriquement impossible à calculer, en pratique on segmentera l'ensemble d'échantillonnage en deux parties, l'une dédiée à l'apprentissage, l'autre à évaluer la performance hors échantillon.

**Régularisation** Afin de contrecarrer le risque hors échantillon, la solution est en fait de pénaliser la complexité de la fonction de décision  $q$  (rasoir d'Occam). Ainsi, on étudiera en profondeur le choix d'une fonctionnelle  $R : \mathcal{Q} \rightarrow \mathcal{R}$  permettant de quantifier la complexité de  $q$ . L'objectif serait alors

$$\underset{q \in \mathcal{Q}}{\text{maximiser}} \quad n^{-1} \sum_{i=1}^n u(r_i q(x_i)) - R(q). \quad (7)$$

Par exemple, comme les mesures sur  $x$  peuvent comporter de l'incertitude ou du bruit, il serait souhaitable que la décision  $q(x_1)$  soit proche de  $q(x_2)$ , si  $x_1$  et  $x_2$  sont eux même proches dans l'espace  $\mathbf{X}$ . Si  $R$  encodait une telle préférence, ne fonction discontinue comme le dictionnaire présenté plus haut sera alors hautement défavorisée, et une fonction plus lisse y serait préférée.

[**Todo:** Introduire la validation croisée ainsi que le paramètre  $\lambda$  dans l'objectif.]

**Espaces de décision** En pratique, ce mémoire ne considérera que des espaces de Hilbert pour  $\mathcal{Q}$ . Un des avantages des espaces de Hilbert, c'est qu'ils induisent naturellement une notion de norme  $\|\cdot\|_H$ , qu'on peut intuitivement relié au concept

de complexité. Nous nous intéresserons donc aux propriétés induites par  $R(q) = \|q\|_H^2 = \langle q, q \rangle$ . Il y a aussi moyen, sous des conditions assez techniques (théorème de la représentation) de généraliser la norme  $L_2$  de  $q$  à une norme  $L_p$  général. En particulier, nous verrons qu'une régularisation donnée par norme  $L_1$  induit certaines propriétés d'éparsité dans la solution.

**Décisions linéaires** De façon générale, la forme de décision la plus simple est celle qui combine linéairement les  $p$  observations de  $x \in \mathbf{X} \subseteq \mathcal{R}^p$ ; autrement dit lorsque qu'on contraint  $\mathbf{Q} = \mathbf{X}^*$ , *i.e.*, à l'espace dual de  $\mathbf{X}$ . En langage plus clair, à toute fonction  $q \in \mathbf{Q}$  il existe un vecteur de dimension  $p$  tel que la décision dérivée de l'observation  $x$  sera donnée par  $q(x) = \langle q, x \rangle = q^T x$ .

La régularisation  $L_2$  de  $q$  devient alors tout simplement  $R(q) = q^T q = \|q\|^2$  et la fonction optimale de décision  $q^*$  sera alors déterminée en résolvant le problème d'optimisation suivant:

$$\underset{q \in \mathcal{R}^p}{\text{maximiser}} \quad n^{-1} \sum_{i=1}^n u(r_i q^T x_i) - \lambda \|q\|^2. \quad (8)$$

## 1.2 Dimensionnalité de l'information

[**Todo:** Discussion du phénomène big data, de l'importance de  $p$ ]

## 1.3 Risque et garanties statistiques sur la décision

[**Todo:** Discussion sur les méthodes de risques hors échantillon, complexité de l'échantillonnage, mesure Rademacher, distance par rapport à la "meilleure" décision]

## 1.4 Interprétations

Interprétation géométrique dans l'espace  $X$

Interprétation statistique (avec matrix covariance)

Autre?

## 1.5 Objectifs

## 2 Optimisation moderne de portefeuille

Dans ce document, nous allons tenter de classifier et de répertorier la plupart des méthodes ayant rapport, de près ou de loin, à l'intersection des méthodes statistiques avancées et de l'apprentissage machine avec la théorie du portefeuille, en présentant pour chacune d'elle leurs avantages et leurs inconvénients.

### 2.1 Théorie classique du portefeuille

Une revue de littérature sur la théorie du portefeuille serait fondamentalement incomplète sans l'article fondateur de Markowitz, publié en 1952 [?].

Nous allons montrer que le cadre théorique développé par Markowitz peut être considéré comme un cas particulier de notre algorithme, pour autant que l'on considère un portefeuille à un seul actif.

Soit  $w \in \mathcal{R}^k$  le vecteur représentant la répartition du portefeuille de Markowitz à  $k$  actifs à optimiser. Alors un investisseur *markowitzien* souhaite résoudre le problème suivant:

$$\begin{aligned} &\text{minimiser} && w^T \Sigma w \\ &\text{tel que} && \mu^T w = \mu_0, \end{aligned} \tag{9}$$

où  $\Sigma \in \mathcal{R}^{k \times k}$  est la covariance du rendement des actifs et  $\mu \in \mathcal{R}^k$  le vecteur d'espérance. **[Todo: Montrer formellement.]** Par la théorie de l'optimisation convexe, il existe une constante  $\gamma \in \mathcal{R}$  telle que le problème énoncé est équivalent à

$$\text{maximiser} \quad \mu^T w + \gamma w^T \Sigma w. \tag{10}$$

Dans le cas où on considère un portefeuille à un seul actif, alors ce problème se réduit alors à

$$\text{maximiser} \quad \mu q - \gamma \sigma^2 q^2, \tag{11}$$

où on a posé  $\mu := \mathbf{E}R$  et  $\sigma^2 := \text{Var } R$ .

Supposons qu'un investisseur soit doté d'une utilité quadratique paramétrée par

$$u(r) = r - \frac{\gamma}{\sigma^2 + \mu^2} \sigma^2 r^2, \tag{12}$$

et que l'information factorielle intégrée à l'algorithme ne consiste uniquement qu'en les rendements eux mêmes; autrement dit, le vecteur d'information  $X$  se réduirait tout simplement à un terme constant fixé à 1, *i.e.*,  $X \sim 1$ . **[Todo: expliquer].**

Avec une utilité (12) et l'absence d'information supplémentaire, l'objectif de (??) devient aussitôt

$$\text{EU}(qR) = q\mathbf{E}R - \frac{\gamma}{\sigma^2 + \mu^2} \sigma^2 q^2 \mathbf{E}R^2. \tag{13}$$

Or, puisque  $\text{Var } R = \mathbf{E}R^2 - (\mathbf{E}R)^2$ , on a que  $\mathbf{E}R^2 = \sigma^2 + \mu^2$ , ce qui entraîne donc que (12) s'exprime par

$$\text{maximiser } \mathbf{E}U(qR) = \mu q - \gamma \sigma^2 q^2, \quad (14)$$

ce qui est tout à fait identique à (11).

Nous suggérons au lecteur intéressé par l'équivalence des diverses formulations d'optimisation de portefeuille dans un univers de Markowitz [?] et [?], tous deux publiés à l'occasion du soixantième anniversaire de [?].

## 2.2 Portefeuille universel / Papiers d'Elad Hazan

Ce mémoire sera également consacré aux garanties statistiques de performance des estimateurs  $q^*$ .

Bien que le modèle soit différent et de nature itérative, le *portefeuille universel* de [?] est à notre connaissance un des premiers modèles de gestion de portefeuille à exploiter une distribution arbitraire tout en proposant des garanties statistiques de convergence.

Voir [?, ?].

## 2.3 Théorie de portefeuille régularisé

[?]

## 2.4 Fama and French et suivants?

[?]

## 2.5 Articles du NIPS

## 2.6 Papiers de Ben Van Roy

## 2.7 Conclusions: Notre problème par rapport à ces deux disciplines

### 3 Décisions non linéaires

#### 3.1 Dualité

Let us consider the following problem, optimized over  $q \in \mathcal{R}^p$ :

$$\text{minimiser } \sum_{i=1}^n \ell(r_i q^T x_i) + n\lambda \|q\|^2, \quad (15)$$

where  $\ell = -u$ . Alternatively, this problem can be respecified using slack vector  $\xi \in \mathcal{R}^n$  as

$$\begin{aligned} &\text{minimiser } \sum_{i=1}^n \ell(\xi_i) + n\lambda \|q\|^2 \\ &\text{tel que } \xi_i = r_i q^T x_i. \end{aligned} \quad (16)$$

Let  $\alpha \in \mathcal{R}^n$ . The Lagrangian of (16) can be written as

$$\mathcal{L}(q, \xi, \alpha) = \sum_{i=1}^n \ell(\xi_i) + n\lambda \|q\|^2 + \sum_{i=1}^n \alpha_i (r_i q^T x_i - \xi_i). \quad (17)$$

Because the objective (16) is convex and its constraints are affine in  $q$  and  $\xi$ , Slater's theorem states that the duality gap of the problem is zero. In other words, solving (15) is equivalent to maximizing the Lagrange dual function  $g$  over  $\alpha$ :

$$\text{maximiser } g(\alpha) = \inf_{q, \xi} \mathcal{L}(q, \xi, \alpha). \quad (18)$$

Now, note that

$$g(\alpha) = \inf_{q, \xi} \left\{ \sum_{i=1}^n \ell(\xi_i) + n\lambda \|q\|^2 + \sum_{i=1}^n \alpha_i (r_i q^T x_i - \xi_i) \right\} \quad (19)$$

$$= \inf_{\xi} \left\{ \sum_{i=1}^n \ell(\xi_i) - \alpha^T \xi \right\} + \inf_q \left\{ \sum_{i=1}^n \alpha_i r_i q^T x_i + n\lambda \|q\|^2 \right\} \quad (20)$$

$$= -\sup_{\xi} \left\{ \alpha^T \xi - \sum_{i=1}^n \ell(\xi_i) \right\} + \inf_q \left\{ \sum_{i=1}^n \alpha_i r_i q^T x_i + n\lambda \|q\|^2 \right\} \quad (21)$$

$$= -\sum_{i=1}^n \ell^*(\alpha_i) + \inf_q \left\{ \sum_{i=1}^n \alpha_i r_i q^T x_i + n\lambda \|q\|^2 \right\}. \quad (22)$$

Where  $\ell^*$  is the convex conjugate of the loss function and is defined by

$$\ell(\alpha_i) = \sup_{\xi_i} \{ \alpha_i \xi_i - \ell(\xi_i) \}. \quad (23)$$

Note that the identity

$$f(\xi_1, \dots, \xi_n) = \sum_{i=1}^n \ell(\xi_i) \implies f^*(\xi_1, \dots, \xi_n) = \sum_{i=1}^n \ell^*(\xi_i) \quad (24)$$



was used. Consider now the second part of (22). Since the expression is differentiable, we can analytically solve for  $q$ :

$$\nabla_q \left\{ \sum_{i=1}^n \alpha_i r_i q^T x_i + n\lambda \|q\|^2 \right\} = 0 \quad (25)$$

implies that

$$q = -\frac{1}{2n\lambda} \sum_{i=1}^n \alpha_i r_i x_i \quad (26)$$

at the infimum.

Using (26), we can eliminate  $q$  from (22), so that

$$g(\alpha) = -\sum_{i=1}^n \ell^*(\alpha_i) - \frac{1}{2n\lambda} \sum_{i,j=1}^n \alpha_i \alpha_j r_i r_j x_i^T x_j + \frac{1}{4n\lambda} \sum_{i,j=1}^n \alpha_i \alpha_j r_i r_j x_i^T x_j \quad (27)$$

$$= -\sum_{i=1}^n \ell^*(\alpha_i) - \frac{1}{4n\lambda} (\alpha \circ r)^T K (\alpha \circ r). \quad (28)$$

Therefore, in its dual form, the problem (15) is equivalent to solving

$$\text{minimiser} \quad \sum_{i=1}^n \ell^*(\alpha_i) + \frac{1}{4n\lambda} (\alpha \circ r)^T K (\alpha \circ r). \quad (29)$$

**Prescribed investment** In its original form, given a feature vector  $\tilde{x}$ , the algorithm (15) suggests an investment size of  $p_0 = q^T \tilde{x}$ , where  $q$  is the trained value obtained by optimizing (15). In the dual formulation (29), with optimal value  $\alpha$ , we have from (26):

$$p_0 = q^T x_0 \quad (30)$$

$$= -\frac{1}{2n\lambda} \sum_{i=1}^n \alpha_i r_i x_i^T x_0. \quad (31)$$

[**Todo:** Insert kernel formulation with vector  $\phi$ .]

### 3.2 Alternate problem

We now consider a new problem, slightly different from (15) where a regularization based on the sum of the square of the investment sizes  $q^T x_i$  is applied:

$$\text{minimiser} \quad \sum_{i=1}^n \ell(r_i q^T x_i) + \gamma \sum_{i=1}^n (q^T x_i)^2 + n\lambda \|q\|^2. \quad (32)$$

Again, this problem can be respecified using slack vector  $\xi \in \mathcal{R}^n$  as

$$\begin{aligned} & \text{minimiser} \quad \sum_{i=1}^n \ell(\xi_i) + \gamma \sum_{i=1}^n (\xi_i/r_i)^2 + n\lambda \|q\|^2 \\ & \text{tel que} \quad \xi_i = r_i q^T x_i. \end{aligned} \quad (33)$$

The constraints in (33) are again affine, so that Slater's theorem apply.

The lagrangian of (33) is

$$\mathcal{L}(q, \xi, \alpha) = \sum_{i=1}^n \ell(\xi_i) + \gamma \sum_{i=1}^n (\xi_i/r_i)^2 + n\lambda \|q\|^2 + \sum_{i=1}^n \alpha_i (r_i q^T x_i - \xi_i), \quad (34)$$

and we seek its infimum over  $(q, \xi)$ .

$$\inf_{q, \xi} \left\{ \sum_{i=1}^n \ell(\xi_i) + \gamma \sum_{i=1}^n (\xi_i/r_i)^2 + n\lambda \|q\|^2 + \sum_{i=1}^n \alpha_i (r_i q^T x_i - \xi_i) \right\} \quad (35)$$

$$= \inf_{\xi} \left\{ \sum_{i=1}^n \ell(\xi_i) + \gamma \sum_{i=1}^n (\xi_i/r_i)^2 - \alpha^T \xi \right\} + \inf_q \left\{ \sum_{i=1}^n \alpha_i r_i q^T x_i - n\lambda \|q\|^2 \right\} \quad (36)$$

$$= -\sup_{\xi} \left\{ \alpha^T \xi - \left( \sum_{i=1}^n \ell(\xi_i) + \gamma \sum_{i=1}^n (\xi_i/r_i)^2 \right) \right\} - \frac{1}{4n\lambda} (\alpha \circ r)^T K (\alpha \circ r). \quad (37)$$

Let  $f_i(\xi_i) := h_1(\xi_i) + h_2(\xi_i) = \ell(\xi_i) + \gamma(\xi_i/r_i)^2$ . Then, using (24), the first expression of (37) can be restated as

$$-\sup_{\xi} \left\{ \alpha^T \xi - \sum_{i=1}^n f_i(\xi_i) \right\} = -\sum_{i=1}^n f_i^*(\alpha_i). \quad (38)$$

Let us introduce another identity:

$$(h_1 + h_2)^*(\alpha_i) = \inf_{\alpha'_i + \alpha''_i = \alpha_i} \{h_1^*(\alpha'_i) + h_2^*(\alpha''_i)\}. \quad (39)$$

Using (39), (38) can be written as

$$-\sum_{i=1}^n f_i^*(\alpha_i) = -\sum_{i=1}^n (h_1 + h_2)^*(\alpha_i) \quad (40)$$

$$= -\sum_{i=1}^n \inf_{\alpha'_i + \alpha''_i = \alpha_i} \{h_1^*(\alpha'_i) + h_2^*(\alpha''_i)\}. \quad (41)$$

The first conjugate function  $h_1^*$  is simply  $\ell^*$ . The second conjugate function can be derived analytically:

$$h_2^*(\alpha''_i) = \sup_{\xi_i} \{\alpha''_i \xi_i - h_2(\xi_i)\} \quad (42)$$

$$= \sup_{\xi_i} \{ \alpha_i'' \xi_i - \gamma (\xi_i / r_i)^2 \}. \quad (43)$$

The supremum occurs when

$$\xi_i = \frac{r_i^2}{2\gamma} \alpha_i''. \quad (44)$$

Therefore, (43) simplifies to

$$h_2^*(\alpha_i'') = \frac{r_i^2}{4\gamma} (\alpha_i'')^2. \quad (45)$$

Putting it all back together, the dual of (32) is

$$- \sum_{i=1}^n \inf_{\alpha_i' + \alpha_i'' = \alpha_i} \left\{ \ell^*(\alpha_i') + \frac{r_i^2}{4\gamma} (\alpha_i'')^2 \right\} - \frac{1}{4n\lambda} (\alpha \circ r)^T K(\alpha \circ r), \quad (46)$$

which is equivalent to

$$- \sum_{i=1}^n \ell^*(\alpha_i) - \frac{1}{4\gamma} \sum_{i=1}^n (r_i \beta_i)^2 - \frac{1}{4n\lambda} (r \circ (\alpha + \beta))^T K(r \circ (\alpha + \beta)), \quad (47)$$

with new optimization variables  $\alpha = \alpha', \beta = \alpha'' \in \mathcal{R}^n$ . The dual optimization problem is therefore

$$\text{minimiser } \sum_{i=1}^n \ell^*(\alpha_i) + \frac{1}{4\gamma} \|r \circ \beta\|^2 + \frac{1}{4n\lambda} (r \circ (\alpha + \beta))^T K(r \circ (\alpha + \beta)). \quad (48)$$

**Prescribed investment [Todo: ]**

## **4 Garanties statistiques**

Hello

## **5 Dimensionnalité du problème**

## **6 Distributions synthétiques**

## **7 Conclusion**