Projet recherche 2014
Compte-rendu 2
Elements de résolution et conclusions

Adrien EHRHARDT

20 avril 2014

Table des matières

Résumé

J'ai pu m'intéresser dans un premier temps à la théorie des problèmes linéaires en nombres entiers et déterminer 3 logiciels qui pourraient me servir dans ma démarche, une fois que j'aurai été en mesure de fournir un modèle de ce que je voulais optimiser. Cela étant fait, je m'intéresse dans ce compte-rendu plus particulièrement à l'application de cette théorie à travers un modèle, deux solveurs, une méthode pour affecter un poids à chaque action sélectionnée et les résultats.

Je m'explique : nous verrons tout d'abord quel est l'objectif financier que je cherche à réaliser et comment on le modélise. On verra ensuite quelles solutions logicielles et implémentations s'offrent à moi. Une fois qu'on a sélectionné des produits financiers par cette méthode, il nous faut les pondérer : savoir quelle partie de notre fortune sera dédié à quel produit. On s'intéressera donc essentiellement à cette méthode de pondération pour terminer sur les résultats qu'auraient donné ces investissements.

Première partie

Rappels financiers

Chapitre 1

Qu'est-ce qu'un fonds d'indice ?

La "bourse" désigne la structure permettant l'échange de flux financiers à travers des produits financiers. Sur le marché européen par exemple, il s'agit d'Euronext. En France, l'ensemble des sociétés cotées échangent leurs produits sur Euronext. Dans un souci de représenter la tendance globale de l'économie des grandes entreprises françaises, la Compagnie des Agents de Change (fusionnée depuis à Euronext) détermine une liste de 40 actions sélectionnées parmi les 100 sociétés dont le volume d'échange est le plus important.

La mise à jour de cette liste est faite trimestriellement par un comité d'expert dans un souci de représentativité, non seulement elle doit refléter le marché financier par la taille des entreprises présentes mais aussi par la capitalisation flottante (les produits disponibles sur le marché) et le volume de transactions.

Les valeurs entrantes se situent dans une autre liste de 20 entreprises parmi les 100 françaises dont le volume d'échange est le plus important, le CAC Next 20.

Ces listes sont appelées des indices car, après un calcul propre à l'indice et basé sur différents paramètres des valeurs qui le composent, elles représentent fidèlement l'évolution du marché financier du type d'entreprise dans l'indice. L'évolution de l'indice CAC40 (notée en points et fixé en 87 à 1000) est une fonction de différents paramètres des 40 valeurs qui le composent et est formée de telle manière qu'il reflète l'ensemble du marché des grandes entreprises françaises.

Il existe de nombreux autres indices pour permettre de refléter chaque marché et ayant chacun sa propre méthode de calcul ; un exemple courant est le S&P500 des 500 plus grandes valeurs américaines. Certains indices sont plus spécifiques et s'intéressent à des domaines précis, comme les indices spécialisés sur les entreprises engagées dans le développement durable.

Comme dit plus haut, les méthodes de calcul d'un indice diffèrent selon les indices (il n'y a pas de méthode précise), bien que la plupart se basent sur la capitalisation boursière (valeur de l'action * nombre d'actions = la valeur boursière d'une entreprise) comme le CAC40 qui se calcule ainsi :

1. Multiplication du dernier cours de clôture de chaque action par son nombre d'actions
2. Additionner les montants
3. Diviser la somme obtenue par le diviseur de 184,629,565.030481 (correspondant à la capitalisation totale du 31 décembre 1987 (1re cotation), multipliée par un facteur d'ajustement K lié aux changement arrivés depuis (euros, émission de titres))

Bien sûr, cette valeur évolue constamment et le CAC40 est mis à jour toutes les 15 secondes lorsque le marché est ouvert.

On parlera plus en détail dans une partie ultérieure de la méthode de calcul et notamment du critère de pondération (le premier point de la méthode) car elle est controversée dans les milieux financiers.

Ces indices sont devenus des produits financiers à part entière, de telle sorte qu'on peut reconstruire soi-même ces indices (la composition est connue) ou même investir directement dans un seul produit qu'on appelle "tracker". Ces trackers sont proposés sur le marché par des fonds d'investissement qui reconstruisent l'indice entièrement ou partiellement (i.e. en ne prenant en compte qu'une partie des valeurs le composant) et essayent de "coller" le plus fidèlement possible aux indices (et donc plus généralement au marché) qu'ils représentent.

Chapitre 2

Style de management

2.1 Management actif

Le management actif est une méthode d'investissement. C'est celle que l'on s'imagine généralement lorsqu'on pense à la bourse et plus particulièrement aux traders. On s'intéresse ici principalement aux inefficacités du marché ainsi qu'à ses propres opinions sur la croissance future du marché pour investir, notamment dans des valeurs qu'on pense sous-évaluée, ou en faisant des techniques short-long (vendre quelque chose à une date future qu'on ne possède pas encore).

Cette théorie correspond grosso modo à celle du "modern portfolio". Le mot "actif" veut tout simplement dire que les positions sont changées très souvent et que l'on se base sur des prédictions futures pour prendre des décisions, cela dans le but de "faire mieux que le marché". Si par exemple on investit en France et on compose notre portefeuille à partir d'actions françaises, notre point de comparaison est le CAC40. En prenant des décisions tôt et en changeant souvent de position pour prendre avantage de dysfonctionnements à court terme du marché, on espère dépasser la rentabilité du CAC40.

2.2 Management passif

Bien au contraire du management actif, le management passif consiste à investir d'après une stratégie prédéterminée, en ce sens que l'on ne se base que sur les résultats passés des valeurs, on ne s'intéresse ni aux techniques de prévision ni aux inefficacités du marché dans la mesure où la position sera "tenue" contrairement au management actif. En effet, l'intérêt de cette stratégie est de rebalancer ses positions très peu souvent. L'objectif ici n'est pas de sur-performer le marché mais de le suivre, c'est-à-dire d'obtenir la même rentabilité. Les indices entrent dans cette catégorie de management.

Chapitre 3

Avantages et inconvénients - l'essor des indices

Les indices ont connu un essor important au cours de ces vingt dernières années du fait que bien souvent les fonds ayant un management actif finissaient par sous-performer le marché. Il était d'abord réservé aux marchés d'actions puis commence à apparaître sur le marché des bonds. Certains fonds sont spécialisés dans cette approche.

Les oppositions entre management actif et passif sont liées à deux critères fondamentaux en finance de marché : le couple risque/retour sur investissement : tandis que le management passif cherche à minimiser le risque par diversification, le management actif cherche le profit comme premier critère.

De manière intrinsèque, du fait que le management actif ait pour but de sur-performer le marché contrairement au management passif qui ne consiste qu'à le suivre au mieux, il semblerait que le management actif soit préférable. Dès lors, pourquoi la management passif a connu tellement d'essor récemment ? Tout simplement parce que dans les faits le management actif ne permettait pas de remplir ses objectifs.

Premièrement, les critères de choix des actions et les informations sur lesquels se basent les managers actifs de fonds pour trouver "de bonnes opportunités" ne sont pas fondés la plupart du temps. Deuxièmement, des frais très importants sont engendrés. Supposons qu'un particulier investisse dans un mutual funds (managé activement) : il doit faire face à des frais à l'entrée (fixes + part variable) pour dissuader les petits investisseurs, des frais de management (on s'occupe de manager son investissement), des frais administratifs (le fonctionnement du mutual fund), et bien sûr les coûts liés aux investissements : on rebalance très souvent ses positions et chaque transaction engendre des faits - ces transactions sont nombreuses du fait du style de management. Ces coûts sont pris en compte dans le calcul du taux de rentabilité du fonds, ce qui conduit bien souvent à une sous-performance du marché sur lequel on réalise les investissements.

C'est ainsi que bien souvent, les indices, en plus d'avoir des frais extrêmement bas, ont un taux de rentabilité bien supérieur comme le montre l'étude de Standard&Poor, SPIVA. L'article du Market Watch, *Active fund managers continue to disappoint* montre en revanche que deux mutual funds activement managés ont largement surperformé leur indice de référence. En revanche, l'expérience montre que les années suivantes et précédentes, ce n'était pas le cas : les mutual funds qui arrivent à battre le marché changent et en moyenne sur plusieurs années, il est trop risqué d'investir dans un mutual fund en pensant que celui-ci battra le marché... On ne le sait tout simplement pas d'avance.

Une explication très parlante de la bizarrerie de ce système se trouve chez Market Watch : *A* *parable for the active investor.* Prenons l'exemple d'un chien et de son maître : les mouvements du chien en laisse sont imprévisibles ; en revanche ceux du maître sont connus (tout droit à vitesse constante par exemple) de telle sorte que les mouvements du chien sont limités par sa laisse. Par analogie, le chien représente le cours des actions, le maître la croissance économique du pays reflétée par l'indice. La laisse est le marché ; chaque entreprise, du fait des interconnections avec le marché et les autres entreprises ne peut pas s'en éloigner durablement.

Dès lors, il devient clair que les indices sont très intéressants tant pour l'absence de management qu'ils demandent que pour leur rentabilité. Depuis la création du CAC, sa rentabilité annuelle moyenne a été de 8.72 % (max : 57.4 - min : 42.68).

On va maintenant se placer dans la peau d'un investisseur privé souhaitant construire son propre indice. Il dispose d'un pool d'actions sur le marché qu'il veut refléter. Malheureusement, pour le CAC40 et, de façon plus évidente pour le S&P500, il s'agirait d'acheter 40 ou pire, 500, valeurs, sachant que les actions sont indivisibles, que certaines sont très chères, et que la méthode de poids par capitalisation donnera un pourcentage d'investissement à placer sur chaque action, la cagnotte de départ se doit d'être énorme ! De plus, il faut périodiquement rebalancer ses positions car les capitalisations boursières des actions contenues dans l'indice car leur part dans l'évolution de l'indice change ce qui engendre des frais ! Plus il y a d'actions, plus il y a de changements à faire et donc plus il y a de frais !

Ces changements de position sont faits pour booster la rentabilité du fonds. Dans la plupart des mutual funds, la turnover annuel (la part de positions changés au cours de l'année) est de 100% dans la plupart des cas. Ce constat est d'autant plus gênant que les investisseurs choisissent ce type de fonds afin de placer leur argent sur une longue période de temps, ce qui ne nécessite par conséquent pas changer constamment de positions.

Restons donc modestes : nous allons sélectionner une sous-population de ces actions de telle sorte que l'évolution de ce "sous-indice" colle statistiquement au mieux avec l'indice. On aura ainsi théoriquement autant de retour sur investissement pour moins de valeurs à gérer et moins de frais (ce faisant, on a cependant augmenté la volatilité, donc le risque en étant moins diversifié).

Chapitre 4

Modélisation

Pour rester général, supposons que le marché que nous voulons refléter se compose de n actifs. Dans la suite on désigne ce marché par "le marché n". On suppose que l'on dispose d'un moyen de mesurer la similarité entre les actions :

On note ρ_{ij} la similarité entre l'actif i et l'actif j (la matrice des "similarités" est donc symétrique).

Généralement, on choisit un système tel que ρ_{ii}=1 et -1 $\leq$ ρ_{ij} $\leq$ 1 — plus les actifs sont similaires, plus leur "similarité" est proche de 1.

On souhaite suivre cet indice en utilisant seulement q actifs (q < n). Dans la suite, on désigne ces valeurs par "l'indice q".

Pour ce faire, on choisit de modéliser le problème avec des variables de décision binaires :

— $x_{ij} = 1$ (i et j $\in$ $|[1\,;\!n]|$) si l'action i (dans le marché n) est représentée par l'action j (dans l'indice q) i.e. l'action j est l'action la plus similaire à l'action i *et* se trouvant dans le fonds. $x_{ij} = 0$ sinon.

— $y_j = 1$ si l'action j (dans le marché n) est présente dans l'indice q. $y_j = 0$ sinon.

L'objectif est de maximiser la ressemblance de l'indice q au marché n soit :

$max_{i,j} \sum_{i=1}^{n} \sum_{j=1}^{n} \rho_{ij} x_{ij}$

Abordons maintenant les contraites ; elles sont au nombre de n*n+n+1, en plus de la contrainte de binarité des variables de décision.

— Le nombre d'actifs dans l'indice doit être q, ce qui se traduit par la contrainte :
$\sum_{j=1}^{n} y_j = q$

— Pour i donné, un seul x_{ij} doit valoir 1 - il y a un seul actif j de l'indice q qui est "le plus similaire" à l'actif i dans le marché n, ce qui se traduit par les n contraintes :
$\forall$ i $\in$ $|[1\,;\!n]|$, $\sum_{j=1}^{n} x_{ij} = 1$

— Si $x_{ij} = 1$, l'actif j doit être dans l'indice q, c'est-à-dire $y_j = 1$ ce qui se traduit par les n^2 contraintes :
$\forall$ i, j $\in$ $|[1\,;\!n]|$; $x_{ij} \leq y_j$

Le modèle final de choix des q actifs dans notre fonds d'indice est donc le suivant :

$$
\begin{array}{llccl}
z = max & \sum_{i=1}^{n} \sum_{j=1}^{n} & & \rho_{ij} x_{ij} & \\
\text{subject to} & \sum_{j=1}^{n} y_j = q & = & \text{q} & \\
 & \sum_{j=1}^{n} x_{ij} & = & 1 & \forall \text{ i} \in |[1\,;\!n]| \\
 & x_{ij} & \leq & y_j & \forall \text{ i, j} \in |[1\,;\!n]|
\end{array}
$$

Une fois que l'on a sélectionné q actions, il faut investir dans ces q actions pour constituer l'indice. Pour déterminer le poids que prendra chaque action de l'indice q, on peut employer plusieurs méthodes dont on discutera plus tard. Cependant, les deux plus répandues sont la capitalisation boursière (utilisée pour la construction du CAC40 par exemple) et la valeur de l'action.

Pour un modèle "basique", on s'intéresse à la valeur V_i de chaque action i du marché n que l'on cherche à modéliser. On cherche à investir non seulement à hauteur de la valeur de l'action j se trouvant effectivement dans l'indice q mais également à hauteur de la valeur des action i du marché n que l'action j de l'indice q représente.

Vous suivez toujours ? Pour ce faire, on va tout simplement procéder au calcul suivant :

$\omega_j = \sum_{i-1}^{n} V_i x_{ij}$

où ω_j représente le poids de l'action j présent dans l'indice q. Dès lors, il devient facile de connaître la fraction de notre fortune à investir dans l'action j ; c'est :

$\frac{\omega_j}{\sum_{k=1}^{n} \omega_k}$

On se servira d'abord de ce modèle de "poids" pour les actions retenues dans notre indice avant de discuter des méthodes de pondération.

La suite logique est donc de tester ce modèle : l'implémenter sur différents logiciels puis s'intéresser au critère de pondération et terminer sur les résultats nets que l'on aurait pu obtenir en investissant dans ce fonds maison.

Deuxième partie

Utilisation des outils d'optimisation

Afin de me servir de ce modèle pour choisir q actifs à partir de n actifs et ainsi créer mon indice, il me fallait encore deux choses :

- choisir un sytème à modéliser :
  - définir un marché n que l'on veut suivre : bonds, actions "growth", penny stocks, marché américain (S&P 500), marché français (CAC40), ...
  - définir combien d'actions représentent ce marché (40 en France par exemple en général). Pour cela, on se base toujours sur un indice de référence pré-existant.
  - définir un nombre d'actions à retenir pour constituer son propre indice.
    Pour ce point précis, je n'ai mené aucune recherche : je ne sais pas quel est le nombre optimal d'actions à acheter par rapport au marché de référence pour le suivre efficacement. Peu d'articles sont disponibles à ce propos. En effet, les investisseurs disposant d'un capital important désirant suivre le CAC40 peuvent se permettre d'investir dans chaque action à hauteur de la capitalisation boursière de chaque entreprise. Par ailleurs, certains Mutual Funds vendent des fonds d'indice directement : on n'achète qu'un seul produit et le Mutual Fund s'occupe de faire la balance pour nous. Ce qu'il faut retenir est que ce critère dépend beaucoup du nombre d'actions n du marché à refléter ainsi que du capital que l'on désire investir.
  - choisir la mesure de "ressemblance" entre les actions. J'ai décidé de choisir la corrélation, comme j'en parlerai plus tard, mais on pourrait choisir une autre mesure. Cependant, la corrélation est l'outil majoritairement utilisé.
- implémenter le modèle sur un ou des logiciel(s) pour la réalisation.
  En effet, comme on l'a vu dans le précédent compte-rendu, ces problèmes linéaires en nombres entiers sont très difficiles à résoudre et il est nécessaire d'en implémenter le modèle sur ordinateur pour le résoudre (et même dans ce cas-là cela peut poser des problèmes comme on le verra par la suite).

Chapitre 5

Matlab

Je me suis tourné naturellement vers Matlab (que j'utilisais pour mon projet G1/G2), logiciel de calcul et de programmation dans de nombreux domaines scientifiques, lorsqu'il s'agissait d'implémenter le modèle vu un peu plus tôt. Je n'ai pas été surpris de trouver un nombre important de fonctions prévues pour l'optimisation ainsi qu'une toolbox destinée dont je parle dans les lignes qui suivent.

5.1 Lockbox Problem

Pour me familiariser avec Matlab et la résolution de problèmes linéaires en nombres entiers, j'ai décidé de résoudre un problème présent dans Optimization Methods in Finance, dit "Lockbox Problem".

Une entreprise de VAD reçoit des chèques sur l'échelle des Etats-Unis, découpés en 4 régions. Du fait des services postal et bancaire, le client a rempli son obligation dès lors que le chèque et fait et envoyé alors qu'il se passe plusieurs jours avant que l'entreprise ne puisse utiliser l'argent. Il y a donc de l'intérêt perdu sur ce transit. L'entreprise se pose donc la question de la rentabilité de la création de 4 lockboxes, des bureaux destinés à la réception des chèques. On dispose du volume total échangé par région, du délai par région et par lockbox de réception, et donc de l'intérêt perdu :

From	L.A.	Cincinnati	Boston	Houston
West	60	120	180	180
Midwest	48	24	60	60
East	216	180	72	180
South	126	90	108	54

Figure 5.1 – Intérêts perdus

On se demande quels lockboxes doivent être ouverts pour minimiser les pertes (coûts de fonctionnement du lockbox + intérêts perdus) soit la fonction :

$60x_{11}+120x_{12}+180x_{13}+180x_{14}+48x_{21}+21x_{22}+60x_{23}+60x_{24}+216x_{31}+180x_{32}$
$+72x_{33}+180x_{34}+126x_{41}+90x_{42}+108x_{43}+54x_{44}+90y_1+90y_2+90y_3+90y_4$

Concernant les contraintes, il est clair que chaque région des Etats-Unis devra être affectée à un seul lockbox d'où

$\sum_j x_{ij} = 1 \; \forall$ i

Il faut par ailleurs ajouter la contrainte qu'une région ne peut être assignée qu'à un lockbox ouvert :

$x_{1j}+x_{2j}+x_{3j}+x_{4j} \leq 4y_j \; \forall$ j (on a mis 4, on aurait pu mettre n'importe quoi de plus grand) Avec cette inégalité, si $y_j = 0$ aucune région ne peut être affectée au lockbox j.

5.2 Fonction bintprog

Pour résoudre le problème précédent, on utilise la fonction bintprog de Matlab qui est la fonction dédiée à la résolution des problèmes linéaires en nombres entiers.

Cette fonction fonctionne sous le modèle suivant :

$$\begin{array}{cccc} min & f^t x & & \\ \text{subject to} & Ax & \leq & \text{b} \\ & A_{eq}x & = & b_{eq} \\ & x & binaire & \end{array}$$

La fonction s'écrit " x = bintprog(f,A,b,Aeq,beq) "

Notons d'abord qu'il ne peut que minimiser, et non pas maximiser la fonction objectif. Il suffira donc de prendre l'opposé de la fonction objectif ci-dessus. Par ailleurs, remarquons que Matlab distingue les contraintes d'égalité et d'inégalité ce qui va nous compliquer la tâche.

En effet, Aeq, beq, A, b et f doivent être écrits sous forme de matrices... On verra plus tard que cela posera problème car il faut donner ces matrices en arguments. Il faut donc les avoir dans le workspace ou les calculer. Pour de "gros" calculs, ces matrices sont énormes et ralentissent énormément le programme. Prenons par exemple le cas de S&P 500. Nous sommes en présence de 500 actions, soit d'après le modèle énoncé plus haut 250500 variables et autant de contraintes. On comprend bien que des matrices carrées de cette taille sont difficiles à gérer par Matlab. De plus, ce n'est pas la notation "naturelle" qui vient à l'esprit quand on jette un oeil au modèle et à l'ordre des contraintes. Les détails des problèmes rencontrés se trouvent en chapitre 8.

5.3 Optimization toolbox et résultat

Matlab est dotée d'une Optimization Toolbox, regroupant toutes les fonctions d'optimisation dont bintprog. Une illustration de cette toolbox est disponible sur la page suivante. Celle-ci permet notamment de jouer sur les paramètres de l'algorithme utilisé pour la résolution (branch and bound, cutting planes, et leurs paramètres respectifs vus dans le premier compte-rendu) mais les résultats issus des changements de paramètres étaient trop minimes pour que j'en parle ici comme je l'ai fait dans le premier compte-rendu. En effet, la taille du problème résolu est trop faible pour que ces différences soient probantes.

En pratique, je n'ai pas pu poursuivre l'utilisation de cette toolbox avec le modèle d'indice car je devais générer les matrices par Matlab et que la toolbox ne permet que de saisir directement des matrices (fastidieux dès lors que la matrice est de taille n^2 !).

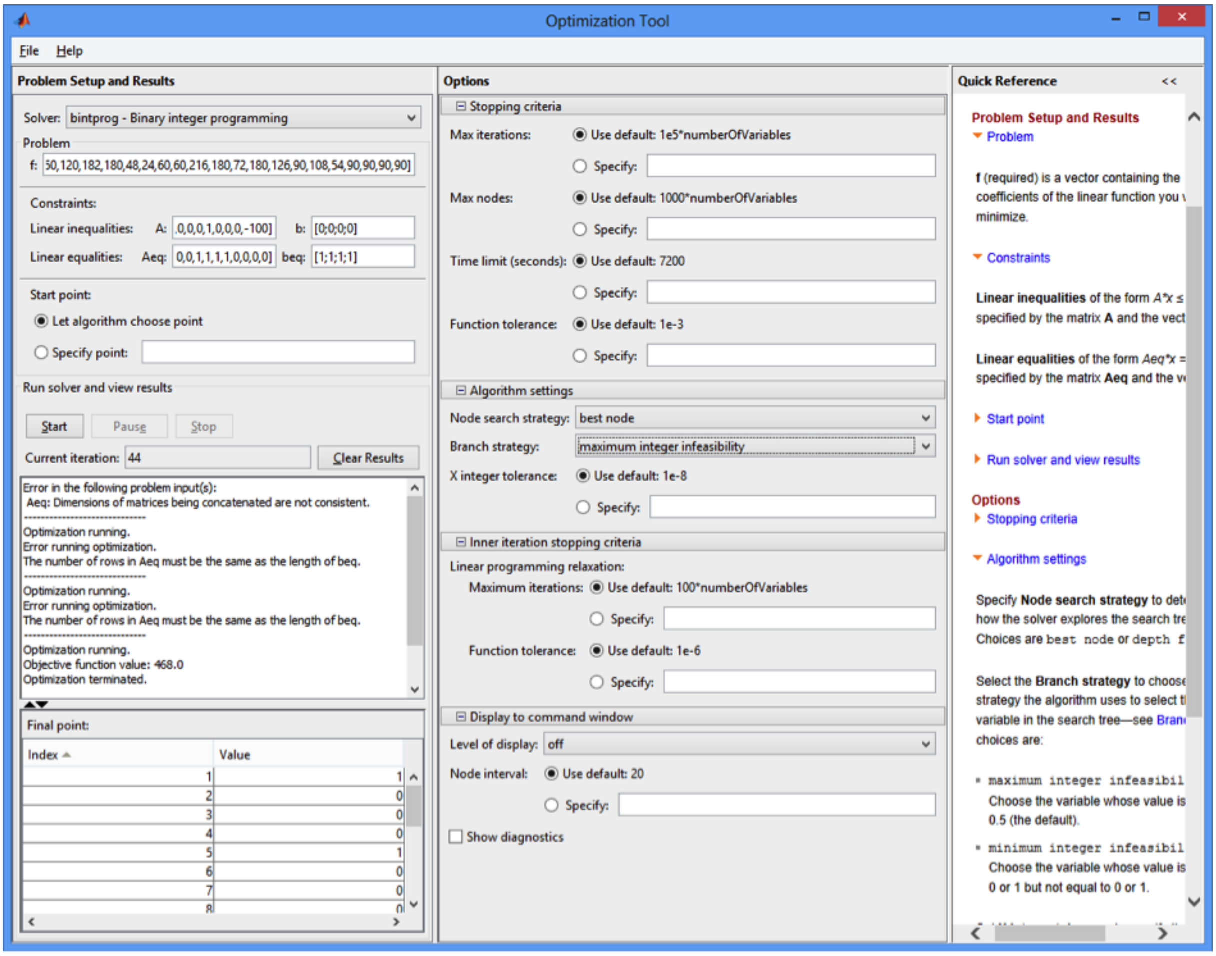


FIGURE 5.2 – Utilisation de Matlab pour la résolution des problèmes binaires

Chapitre 6

AMPL et CPLEX

6.1 AMPL - un logiciel dédié à l'optimisation

AMPL, l'acronyme de "A Mathematical Programming Language" est en langage de modélisation de problèmes d'optimisation. Il est tourné entièrement vers l'optimisation et est très rudimentaire ; il a été développé entre autres par Bell Labs et continue d'être utilisé dans de nombreuses entreprises et dans les labos de recherche.

Ce langage a permis de modéliser les données du problème linéaire en nombres entiers bien plus simplement que Matlab.

Par ailleurs, on peut disposer d'une interface graphique bien pratique.

Il faut distinguer quatre choses :

— le fichier .mod contient le modèle du problème à résoudre (variables, fonction objectif, contraintes, etc)

— le fichier .dat qui contient les données du problème à résoudre (la valeur des paramètres entre autres)

— AMPL se charge de compiler les deux fichiers précédents et d'appeler un solveur

— le solveur : il se charge de résoudre le problème. Il existe de nombreux solveurs différents dont une douzaine sont compatibles avec AMPL. Le plus connu (et le rare qui se charge très bien des problèmes linéaires en nombres entiers) est CPLEX, développé par IBM, et encore massivement utilisé aujourd'hui.

La manière dont les logiciels interagissent est déterminée par le schéma suivant :

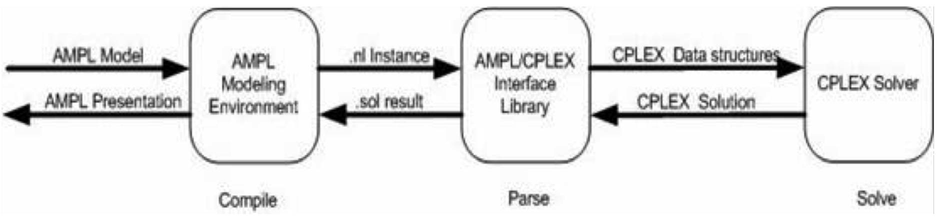


FIGURE 6.1 – Interaction entre logiciels

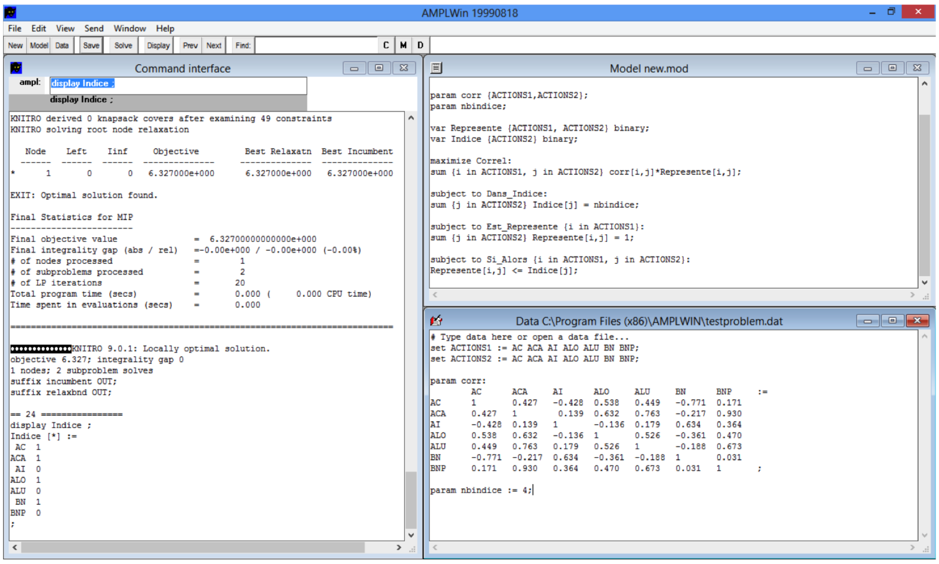


FIGURE 6.2 – Interface graphique pour le langage AMPL

6.2 Associé à un solveur de problèmes linéaires en nombres entiers : CPLEX

installation d'un solveur spécial car le solveur de base ne marche pas pour les MIP IBM CPLEX utilisation transparente dans l'interface graphique -> screenshot de résolution triviale

Le solveur CPLEX a donc été installé et fonctionne avec AMPL. Il fait partie d'une suite logicielle plus complète faite par IBM et destinée à la résolution de problèmes d'optimisation mais on lui préfère souvent le langage libre AMPL pour la modélisation.

Un problème linéaire en nombres entiers dont nous parlerons plus tard, entré sous AMPL qui appelle CPLEX est visible sur la figure suivante :

```
== 24 ================
display Indice ;
Indice [*] :=
BBBY  0
  CA  0
COST  1
DLTR  1
 KER  0
  KR  0
 MEO  1
SHLD  0
 SWY  1
 TGT  1
TSCO  1
 WMT  0
```

FIGURE 6.3 – Résolution d'un problème linéaire en nombres entiers

Chapitre 7

Préparation des données

7.1 Matlab

7.1.1 Données de corrélation

Tout d'abord, il a fallu choisir une méthode pour représenter la similarité entre les actifs. Pour ce faire, les investisseurs se basent très souvent sur la corrélation. Présentons succinctement cet outil statistique : Considérons X, variable aléatoire continu d'un espace probabilisé (Ω, A, p) , décrivant le cours et plus précisément sa valeur à la clôture. On a choisi de se baser sur la valeur à la clôture car c'est ce sur quoi se base le CAC40.

Considérons X_i les réalisations de ces cours de clôture (par jour, par mois, par an, ...). Dans mon cas, je me suis basé sur les données par jour pour plus de précision.

On définit la variance comme : $V[X] = E[X^2] - E[X]^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{X})^2$

L'écart-type s'écrit $\sigma_X = \sqrt{V[X]}$

Considérons X^1 la variable aléatoire décrivant le cours de l'action 1 et X^2 pour l'action 2. Leur covariance s'écrit : $\sigma_{X^1X^2} = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i^1 x_i^2 P(X^1 = x_i^1 et X^2 = x_i^2) - E[X^1]E[X^2]$

On définit enfin leur corrélation : $Corr(X^1, X^2) = \frac{\sigma_{X^1X^2}}{\sigma_{X^1}\sigma_{X^2}}$

Comme le montre le graphique suivant, une corrélation proche de 1 signifie que les cours évoluent systématiquement dans le même sens et d'autant. Une corrélation proche de 0 démontre une absence de relation entre les deux cours. Une corrélation proche de -1 signifie que les cours évoluent dans des directions opposées (ce qui n'est pas rare, un fait économique peut servir une industrie tout en en disqualifiant d'autres).

Pour posséder ces corrélations, il a fallu tout d'abord obtenir les cours de toutes les actions composant l'indice à modéliser : on a commencé par le CAC40. Pour ce faire, on obtient sur Yahoo Finance les cours exportables au format Excel. Après avoir concaténé les cours des actions composant le CAC, on utilise l'outil "Corrélation" présent dans l'"Analyseur de données" une fois le ruban Développeur d'Excel activé. On obtient ainsi la corrélation voulue :

Tableau que l'on charge ensuite dans Matlab par l'intermédiaire d'une variable globale .mat dont on se servira dans les fonctions suivantes.

7.1.2 Ecriture de la fonction objectif

Comme on a pu le voir dans la section dédiée aux spécificités de Matlab un peu plus haut, l'ensemble des variables sont regroupées au sein d'un vecteur x. Or les variables sont au nombre de n*n+n. On choisit la répartition suivante :

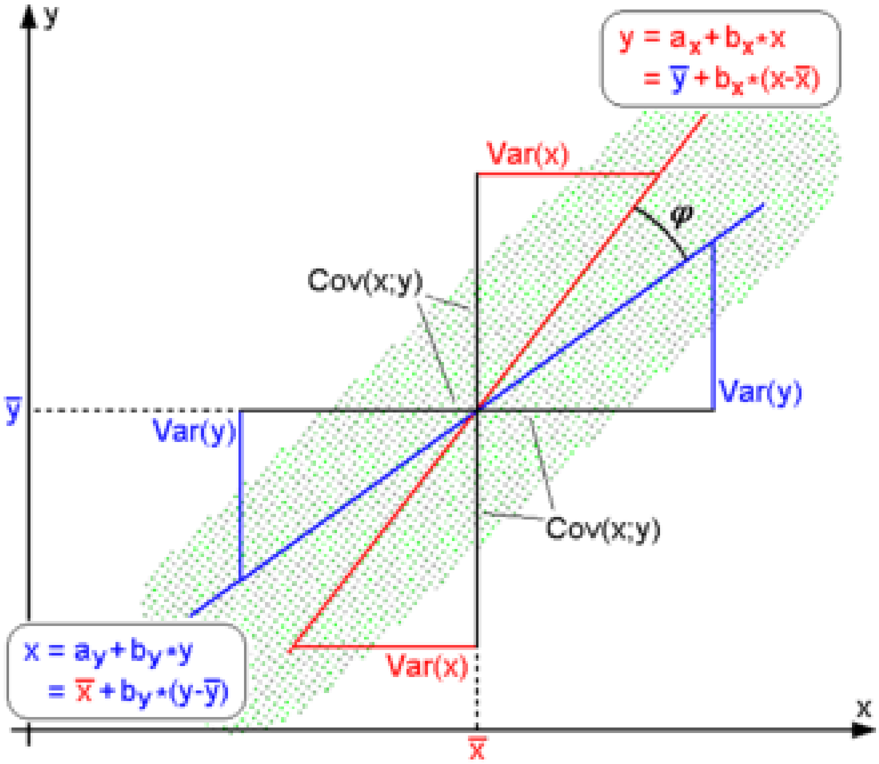


FIGURE 7.1 – Illustration des relations entre X et Y et leur corrélation

x = (x_{11}, x_{12}, ..., x_{1n}, ..., x_{nn}, y_1, ..., y_n)
Le vecteur de la fonction objectif f s'écrit donc :
f = (ρ_{11}, ρ_{12}, ..., ρ_{1n}, ..., ρ_{nn}, 0, ..., 0)
Pour faire cela, on se sert de la fonction objectif qui calcule la matrice à multiplier à la matrice de corrélation pour obtenir le vecteur f. Ce fameux vecteur s'écrit :

7.1.3 Ecriture des contraintes

Comme on l'a vu au-dessus, les contraintes devront également être écrites sous forme de matrices. Ces matrices devront, dans le cas de A_eq et A, être multipliées par x ; elles possèdent donc n*n+n colonnes. Dans le cas de b_eq et b, elles devront être constituées d'une seule colonne mais d'autant de lignes que de contraintes dans chacun des cas.

Un récapitulatif des contraintes est visible sur le tableau suivant :

Contraintes d'égalité

A_{eq} Il y a n+1 contraintes d'égalité. Les n premières contraintes sont du type
∀ i ∈ ||[1 ;n]||, $\sum_{j=1}^{n} x_{ij} = 1$
La n+1 ième est la contrainte
$\sum_{j=1}^{n} y_j = q$
Pour obtenir à la ligne i la somme des n termes consécutifs x_{i1}, x_{i2}, ..., x_{in}, il suffit de mettre un palier de "1" à la ligne i de longueur n sur les colonnes allant de n*(i-1) à n*i.

	AC	ACA	AI	ALO	ALU	BN	BNP	CA	CAP	CS	DG	EDF	EI	EN	FP	GLE	GSZ	KER	LR	MC	ML	OR	ORA	PUB	RI	RNO	SAF	SAN	SGO	SU	TEC	VIE	VIV	VK
AC	1,000																																	
ACA	0,427	1,000																																
AI	-0,428	0,139	1,000																															
ALO	0,538	0,632	-0,136	1,000																														
ALU	0,449	0,763	0,179	0,526	1,000																													
BN	-0,771	-0,217	0,634	-0,361	-0,188	1,000																												
BNP	0,171	0,930	0,364	0,470	0,673	0,031	1,000																											
CA	0,500	0,937	0,055	0,629	0,642	-0,292	0,842	1,000																										
CAP	-0,142	0,700	0,571	0,110	0,596	0,353	0,865	0,555	1,000																									
CS	-0,225	0,639	0,641	0,059	0,542	0,427	0,838	0,500	0,960	1,000																								
DG	0,054	0,712	0,572	0,183	0,750	0,192	0,825	0,568	0,912	0,885	1,000																							
EDF	0,643	0,858	-0,005	0,551	0,773	-0,425	0,708	0,892	0,459	0,392	0,593	1,000																						
EI	-0,849	-0,276	0,491	-0,522	-0,365	0,833	0,024	-0,376	0,372	0,427	0,130	-0,574	1,000																					
EN	0,671	0,841	0,002	0,650	0,820	-0,457	0,683	0,855	0,428	0,362	0,594	0,936	-0,613	1,000																				
FP	-0,068	0,560	0,683	0,158	0,461	0,163	0,722	0,480	0,774	0,784	0,808	0,417	0,166	0,454	1,000																			
GLE	0,320	0,972	0,252	0,572	0,748	-0,123	0,971	0,901	0,795	0,741	0,800	0,811	-0,149	0,800	0,680	1,000																		
GSZ	0,803	0,614	-0,213	0,730	0,585	-0,666	0,358	0,689	-0,013	-0,074	0,203	0,815	-0,848	0,833	0,162	0,519	1,000																	
KER	-0,761	0,048	0,616	-0,292	-0,155	0,835	0,322	-0,043	0,567	0,633	0,325	-0,288	0,863	-0,326	0,357	0,152	-0,643	1,000																
LR	-0,571	0,332	0,670	-0,255	0,180	0,673	0,588	0,197	0,827	0,868	0,674	0,046	0,726	-0,011	0,614	0,451	-0,423	0,867	1,000															
MC	-0,751	-0,126	0,729	-0,334	-0,175	0,764	0,173	-0,240	0,441	0,502	0,325	-0,393	0,793	-0,384	0,414	0,023	-0,615	0,835	0,748	1,000														
ML	-0,480	0,428	0,616	-0,104	0,258	0,543	0,688	0,268	0,863	0,876	0,684	0,087	0,654	0,056	0,643	0,544	-0,362	0,774	0,897	0,685	1,000													
OR	-0,742	0,052	0,596	-0,398	-0,127	0,817	0,344	-0,041	0,636	0,692	0,403	-0,256	0,911	-0,327	0,395	0,174	-0,664	0,938	0,907	0,789	0,815	1,000												
ORA	0,859	0,524	-0,337	0,702	0,540	-0,727	0,247	0,620	-0,122	-0,200	0,104	0,770	-0,911	0,801	0,031	0,418	0,966	-0,739	-0,549	-0,697	-0,465	-0,761	1,000											
PUB	-0,614	0,183	0,661	-0,395	0,084	0,699	0,471	0,057	0,787	0,829	0,614	-0,090	0,799	-0,141	0,586	0,318	-0,566	0,862	0,952	0,743	0,875	0,936	-0,665	1,000										
RI	-0,846	-0,313	0,503	-0,430	-0,484	0,773	-0,016	-0,393	0,278	0,329	0,051	-0,621	0,931	-0,627	0,225	-0,178	-0,790	0,838	0,654	0,830	0,579	0,839	-0,861	0,715	1,000									
RNO	-0,448	0,433	0,671	-0,160	0,251	0,618	0,670	0,330	0,881	0,901	0,753	0,186	0,638	0,120	0,681	0,555	-0,294	0,808	0,954	0,673	0,863	0,859	-0,416	0,923	0,575	1,000								
SAF	-0,593	0,158	0,609	-0,438	0,132	0,684	0,439	0,005	0,765	0,811	0,623	-0,082	0,783	-0,147	0,501	0,290	-0,574	0,818	0,940	0,729	0,854	0,906	-0,662	0,968	0,661	0,893	1,000							
SAN	-0,267	0,412	0,305	0,089	0,104	0,393	0,531	0,329	0,532	0,524	0,303	0,089	0,396	0,005	0,194	0,408	-0,221	0,616	0,562	0,420	0,645	0,569	-0,257	0,506	0,338	0,546	0,481	1,000						
SGO	0,358	0,764	0,398	0,551	0,898	-0,044	0,738	0,634	0,655	0,629	0,828	0,712	-0,265	0,786	0,639	0,788	0,550	-0,006	0,313	0,042	0,346	-0,017	0,464	0,183	-0,305	0,394	0,206	0,125	1,000					
SU	0,603	0,837	-0,029	0,848	0,721	-0,347	0,678	0,809	0,353	0,280	0,465	0,811	-0,529	0,821	0,299	0,780	0,811	-0,251	-0,034	-0,300	0,045	-0,286	0,754	-0,221	-0,507	0,076	-0,216	0,177	0,744	1,000				
TEC	-0,746	-0,434	0,510	-0,254	-0,443	0,634	-0,203	-0,510	0,019	0,077	-0,091	-0,647	0,693	-0,620	0,148	-0,315	-0,568	0,540	0,345	0,730	0,318	0,511	-0,648	0,363	0,790	0,261	0,333	0,116	-0,242	-0,402	1,000			
VIE	0,679	0,828	-0,061	0,807	0,733	-0,435	0,643	0,857	0,324	0,235	0,462	0,891	-0,614	0,906	0,314	0,775	0,890	-0,350	-0,105	-0,400	-0,035	-0,369	0,839	-0,275	-0,597	0,037	-0,280	0,042	0,728	0,942	-0,490	1,000		
VIV	0,515	0,893	0,103	0,545	0,691	-0,378	0,811	0,869	0,564	0,518	0,659	0,853	-0,403	0,866	0,594	0,875	0,689	-0,114	0,218	-0,184	0,278	-0,083	0,607	0,081	-0,383	0,312	0,061	0,202	0,725	0,769	-0,480	0,788	1,000	
VK	0,718	0,692	-0,108	0,844	0,687	-0,470	0,469	0,713	0,109	0,061	0,287	0,782	-0,744	0,833	0,162	0,596	0,902	-0,450	-0,291	-0,472	-0,206	-0,541	0,879	-0,458	-0,710	-0,177	-0,455	-0,006	0,684	0,901	-0,506	0,907	0,672	1,000

FIGURE 7.2 – Tableau des corrélations entre les actions du CAC40

```
function f = objectif(n,data);

f = zeros(1,n*(n+1));

for i=1:n
    for j=1:n
        f(1,n*(i-1)+j)=data(i,j)
    end
end
for k=(n*n+1):(n*(n+1))
    f(1,k)=0
end
f=transpose(f)
```

FIGURE 7.3 – Le vecteur qui, multipliée à la matrice des corrélations, donne f

Ce procédé est aussi valable pour la n+1 ième ligne car on veut sommer les y_j.
Cette matrice est donnée par la fonction suivante :

b_{eq} Comme précisé au paragraphe précédent, les n premières sommes donnent 1 alors que la dernière donne q (b_{eq} est donc une fonction à deux variables). La fonction suivante donne la matrice :

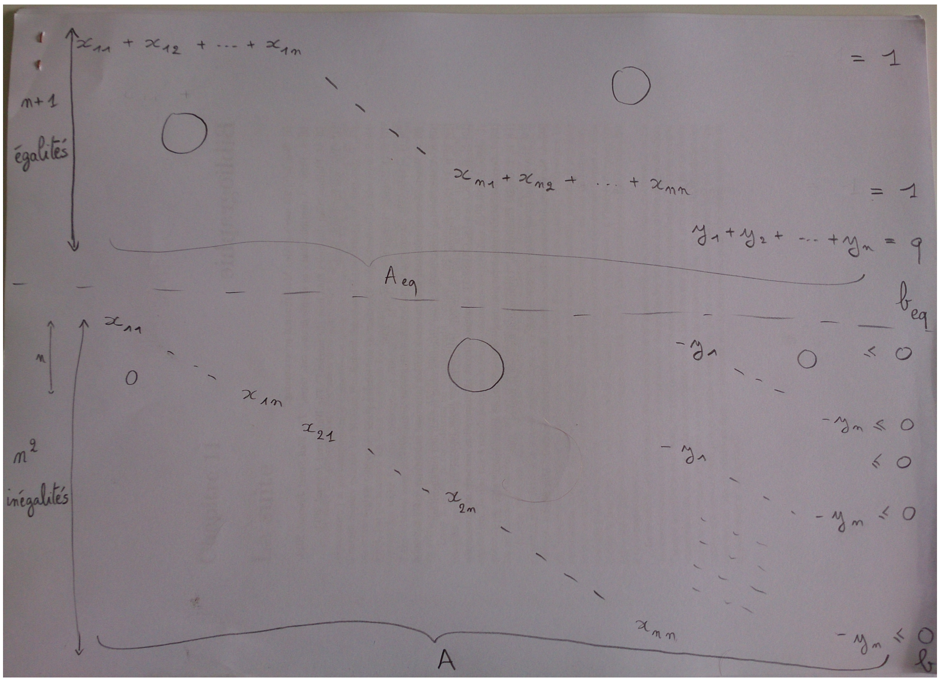


FIGURE 7.4 – Tableau des contraintes sous forme "éclatée"

```
function Aeq = egalites(n);

Aeq = zeros((n+1),n*(n+1));

for i=1:(n+1)
    for j=1:n
        Aeq(i,n*(i-1)+j)=1
    end
end
```

FIGURE 7.5 – Matrice Aeq pour la résolution avec Matlab

Contraintes d'inégalité

A Il s'agit ici de satisfaire les n*n contraintes du type ∀ i, j ∈ ||[1 ;n]|| ; $x_{ij} \leq y_j$

Il faut donc "mettre" x_{ij} et y_j du même "côté" pour, à chaque ligne de la matrice A, écrire 1 à la place x_{ij} et -1 à la place y_j.

On procède de la façon suivante : des lignes i=(k-1)*n + 1 à k*n, on remplit la diagonale de "1" afin d'obtenir sur la ligne (k-1)*n + 1 la variable x_{i1} et la variable y_1 ...jusqu'à la ligne k*n où on obtient la variable x_{in} et la varible y_n.

La fonction suivante donne la matrice A :

```
function beq = egalites2(n,q);

beq = zeros((n+1),1);

for i=1:n
    beq(i,1)=1
end
beq(n+1,1)=q
```

FIGURE 7.6 – Matrice beq pour la résolution avec Matlab

```
function A = inegalites(n);

A = zeros(n*n,n*(n+1));

for k=1:n
    for i=(k-1)*n+1:k*n

        A(i,i)=1
        A(i,n*n+i-(k-1)*n)=-1

    end
end
```

FIGURE 7.7 – Matrice A pour la résolution avec Matlab

b Comme on a "passé" y_j de l'autre côté de la contrainte, la matrice b est nulle (de taille (n*n, 1)).

7.1.4 Modèle final

Enfin, dans la mesure où la fonction bintprog de Matlab ne peut que minimiser, il a fallu chercher, pour maximiser, à minimiser l'opposé de la fonction objectif ! La fonction résolvant le problème d'optimisation en fonction de n, q ainsi que de la matrice de corrélation est la suivante :

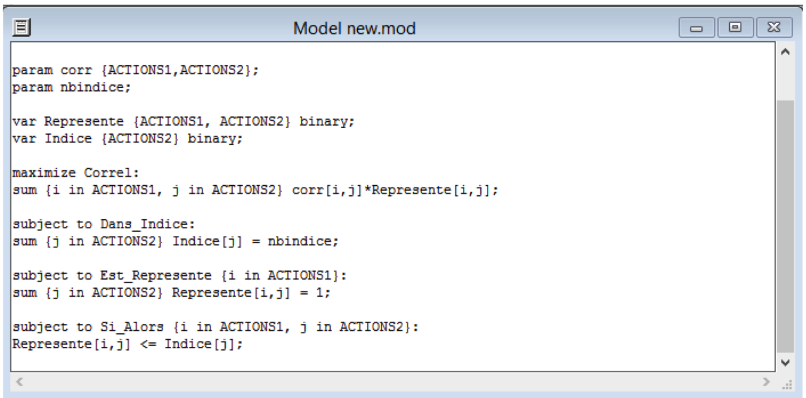
```
function x=indice(n,q,data);
global y
y=bintprog(-objectif(n,data),inegalites(n),
 zeros(n*n,1),egalites(n),egalites2(n,q))
```

FIGURE 7.8 – Fonction résolvant le modèle sous Matlab

7.2 AMPL - CPLEX

7.2.1 Fichier model

Le fichier "model" contient toutes les informations liées purement au modèle : fonction objectif, contraintes, variables de décision, ainsi que d'autres paramètres.




```
param corr {ACTIONS1,ACTIONS2};
param nbindice;

var Represente {ACTIONS1, ACTIONS2} binary;
var Indice {ACTIONS2} binary;

maximize Correl:
sum {i in ACTIONS1, j in ACTIONS2} corr[i,j]*Represente[i,j];

subject to Dans_Indice:
sum {j in ACTIONS2} Indice[j] = nbindice;

subject to Est_Represente {i in ACTIONS1}:
sum {j in ACTIONS2} Represente[i,j] = 1;

subject to Si_Alors {i in ACTIONS1, j in ACTIONS2}:
Represente[i,j] <= Indice[j];
```

Figure 7.9 – Fichier model pour AMPL

Les sets

Les "sets" dans AMPL sont des listes. On peut les définir de manière implicite, comme dans Matlab, c'est-à-dire avoir une simple numérotation de 1 à n. Mais ici, on peut lister des noms (les noms des actions par exemple) et le parcours pour j dans le "set" se fait naturellement de valeur en valeur. Il devient ainsi plus facile de faire référence aux données directement plutôt qu'à un classement de 1 à n abstrait.

Les paramètres

Les paramètres sont des données que nous fournissons dans le fichier data, qui interviennent dans le fichier model et que nous sommes susceptibles de modifier. Cela inclut par exemple le nombre d'actions que nous allons retenir pour former l'indice.

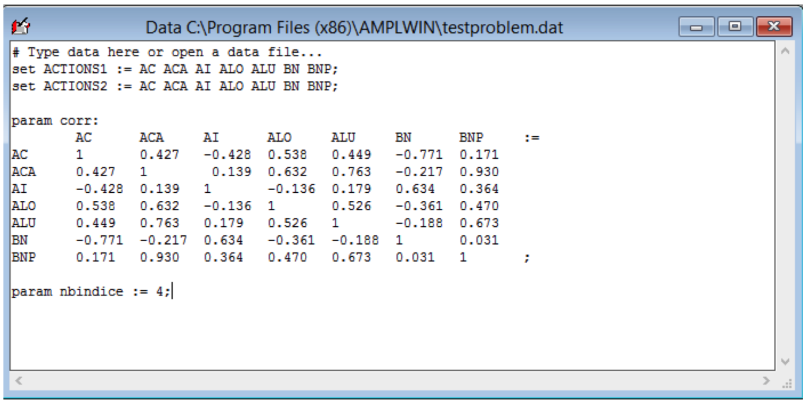
Dans le fichier model, ces paramètres sont entrés comme des "inconnus" qui seront fixés par le fichier data. Dans le fichier data, les paramètres sont entrés sous forme de nombres.

Les variables

Ce sont tout simplement les variables de décision du problème, les "inconnus" que va rechercher le solveur pour optimiser la fonction objectif.

7.2.2 Fichier data

Le fichier data sert uniquement à utiliser le modèle avec des données différentes (et aussi en changer la taille !), c'est-à-dire qu'on peut faire évoluer le fichier data et réutiliser le fichier modèle, ce qui veut dire dans notre cas changer le marché n (en changeant les sets ainsi que la matrice de corrélation) ainsi que l'indice q (en changeant le paramètre éponyme).




```
# Type data here or open a data file...
set ACTIONS1 := AC ACA AI ALO ALU BN BNP;
set ACTIONS2 := AC ACA AI ALO ALU BN BNP;

param corr:
        AC      ACA     AI      ALO     ALU     BN      BNP     :=
AC      1       0.427   -0.428  0.538   0.449   -0.771  0.171
ACA     0.427   1        0.139  0.632   0.763   -0.217  0.930
AI      -0.428  0.139   1       -0.136  0.179   0.634   0.364
ALO     0.538   0.632   -0.136  1       0.526   -0.361  0.470
ALU     0.449   0.763   0.179   0.526   1       -0.188  0.673
BN      -0.771  -0.217  0.634   -0.361  -0.188  1       0.031
BNP     0.171   0.930   0.364   0.470   0.673   0.031   1       ;

param nbindice := 4;
```

FIGURE 7.10 – Fichier data pour AMPL

Chapitre 8

Premier essai - Matlab

8.1 Résolution

Comme on l'a vu dans le premier compte-rendu, la difficulté de résolution des problèmes linéaires en nombres entiers est exponentionnelle (arbre de recherche binaire) avec le nombre de contraintes. On résout donc le problème en augmentant progressivement le nombre d'actions à partir desquels on construit l'indice.

8.1.1 Premier exemple

On sélectionne de manière arbitraire les 8 premières valeurs dans l'ordre alphabétique de l'abréviation du cours pour un premier test. On décide de ne retenir que 2 valeurs sur ces 8 pour bâtir notre indice et d'en comparer les résultats.

Pour notre exemple de 8 valeurs pour un indice de 2 actions, le calcul des pondérations a été fait en utilisant une petite fonction écrite pour l'occasion que vous trouverez en annexe. Les solutions sont :

Il faut retenir Crédit Agricole à hauteur de 87 % et Danone à hauteur de 13%.

On dispose donc d'un portefeuille composé de Crédit Agricole et Danone parmi les 8 sociétés :

1. Accor
2. Crédit Agricole
3. Air Liquide
4. Alstom
5. Alcatel-Lucent
6. Danone
7. BNP Paribas
8. Carrefour

Si on calcule le rendement des 8 premières valeurs du CAC sur l'année passée (pondéré), on obtient : 13,8 %

La moyenne arithmétique des rendements de ces 8 valeurs est : 37 %

Le rendement de notre portefeuille d'indice a été : 54,5 %

La moyenne arithmétique des rendements des actifs du portefeuille est : 27,5 %

```
w =

          0
   294.4100
          0
          0
          0
    42.7800
          0
          0
```

FIGURE 8.1 – Fichier poids pour Matlab

8.1.2 Temps de calcul

Matlab n'est malheureusement pas ralenti par le temps d'exécution de la fonction bintprog en elle-même, chargée de déterminer la solution, mais plutôt par le calcul des fonctions annexes présentées plus haut chargées de stocker des matrices de taille très importante (pour peu que le nombre d'actifs dans le marché n soit grand).

Dans mon cas, au-delà de 10 actions dans le marché n d'origine, Matlab a été incapable de me fournir une réponse en moins d'1/2h, d'où la nécessité de passer à un langage orienté entièrement vers l'optimisation et qui n'aurait pas besoin de calculer de lourdes matrices. Je parle donc de mes résolutions avec AMPL et CPLEX dans la prochaine partie de ce compte-rendu.

Troisième partie

Modélisation complète CPLEX

Chapitre 9

La grande distribution

Dans cette partie, le modèle présenté 7.2.1 n'a pas évolué. Dans la mesure où dans un premier temps j'ai été limité à 300 contraintes et/ou variables de décision pour la version étudiante de mes solveurs, je ne pouvais refléter qu'un marché de taille 16 ! J'ai donc décidé de refléter un marché particulier, et non un marché censé décrire une économie comme les plus connus qui comportent beaucoup d'actifs CAC40, S&P, etc.

Concernant le marché de la grande distribution, j'ai pu m'appuyer sur des sites répertoriant les cours "sectoriels", c'est-à-dire les cours des grandes entreprises de chaque secteur. Je me suis ensuite focalisé sur les titres les plus importants (les plus grandes capitalisations & ceux côtés sur les plus gros marchés) pour aboutir à une liste de 12 enseignes :

1. BBBY - Bed Bath and Beyond
2. CA - Carrefour
3. COST - Costco Wholesale Corporation
4. DLTR - Dollar Tree, Inc.
5. KER - Kering
6. KR - The Kroger Company
7. MEO - Métro
8. SHLD - Sears Holding Corporation
9. SWY - Safeway
10. TGT- Target Corporation
11. TSCO - Tesco
12. WMT - Wal-Mart

Le fichier data se présente donc de la manière suivante :

Le solveur, "habitué" à fonctionner sur des données bien plus larges (en dehors des versions d'essai...) résout le problème instantanément. Deux essais ont été menés : le premier modélise le marché de 12 actions par 8 actions tandis que le deuxième se contente de 6 actions.

Les résultats issus de l'exploitation de ces indices sont visibles dans la partie suivante.

```
# Type data here or open a data file...
set ACTIONS1 := MMM AXP T BA CAT CVX CSCO DD XOM GE GS HD INTC IBM JNJ JPM MCD MRK MSFT NKE PFE PG KO TRV UTX UNH VZ V WMT DIS;
set ACTIONS2 := MMM AXP T BA CAT CVX CSCO DD XOM GE GS HD INTC IBM JNJ JPM MCD MRK MSFT NKE PFE PG KO TRV UTX UNH VZ V WMT DIS;

param corr:
     MMM AXP T BA CAT CVX CSCO DD XOM GE GS HD INTC IBM JNJ JPM MCD MRK MSFT NKE PFE PG KO TRV UTX UNH VZ V WMT DIS :=
MMM 1 0.294 0.171 0.118 0.043 0.468 -0.082 -0.041 0.151 0.118 -0.252 0.319 -0.054 0.079 0.103 0.055 0.148 0.152 -0.012 0.074 0.08 0.024 0.159 0.241 0.226 0.541 -0.061 0.905 0.561 0.187
AXP 0.294 1 0.09 0.174 0.079 0.354 0.504 0.012 0.322 0.426 -0.07 0.445 0.331 0.108 0.229 0.702 0.234 0.118 0.201 0.036 0.19 0.241 0.155 -0.13 0.429 0.136 0.118 0.941 0.399 -0.008
T 0.171 0.09 1 -0.113 -0.404 -0.397 0.657 0.421 0.014 0.604 -0.027 -0.087 0.268 0.068 0.032 0.187 0.158 0.578 0.488 -0.501 0.509 0.132 0.307 0.238 -0.314 -0.004 0.665 0.669 -0.192 0.451
BA 0.118 0.174 -0.113 1 0.298 0.462 -0.201 -0.013 0.297 -0.038 0.734 0.342 -0.075 0.089 -0.127 0.228 0.195 -0.279 -0.207 0.493 -0.105 -0.101 -0.119 0.335 0.591 -0.079 -0.081 0.893 0.338 -0.119
CAT 0.043 0.079 -0.404 0.298 1 0.447 -0.261 -0.256 0.375 -0.365 0.352 0.325 -0.035 -0.26 0.009 0.082 0.262 -0.417 -0.266 0.585 -0.199 0.014 -0.191 0.357 0.479 0.031 -0.28 0.498 0.067 -0.139
CVX 0.468 0.354 -0.397 0.462 0.447 1 -0.098 -0.357 0.434 -0.165 0.123 0.62 0.038 -0.126 0.089 0.186 0.177 -0.295 -0.139 0.537 -0.067 -0.066 -0.118 0.388 0.567 0.117 -0.432 0.831 0.63 -0.11
CSCO -0.082 0.504 0.657 -0.201 -0.261 -0.098 1 0.504 0.209 0.831 -0.158 0.467 0.783 -0.041 0.291 0.606 -0.044 0.53 0.785 -0.251 0.69 0.35 0.39 0.112 0.128 0.052 0.634 0.246 -0.152 0.289
DD -0.041 0.012 0.421 -0.013 -0.256 -0.357 0.504 1 0.289 0.453 0.064 0.074 0.288 0.894 0.552 0.312 0.134 0.319 0.383 -0.136 0.024 0.44 0.718 0.065 -0.419 0.11 0.634 0.742 -0.578 0.197
XOM 0.151 0.322 0.014 0.297 0.375 0.434 0.209 0.289 1 0.044 0.493 0.483 0.23 0.329 0.495 0.426 0.342 -0.049 0.06 0.438 -0.078 0.05 0.362 0.272 0.258 -0.14 -0.174 0.736 0.07 0.061
GE 0.118 0.426 0.604 -0.038 -0.365 -0.165 0.831 0.453 0.044 1 -0.294 0.246 0.674 0.387 0.098 0.451 -0.044 0.46 0.765 -0.308 0.686 0.232 0.343 0.116 -0.091 0.076 0.515 0.517 -0.155 0.314
GS -0.252 -0.07 -0.027 0.734 0.352 0.123 -0.158 0.064 0.493 -0.294 1 -0.056 -0.293 0.225 0.002 0.123 0.434 -0.324 -0.436 0.316 -0.434 -0.37 0.035 0.509 0.157 -0.404 -0.264 0.216 0.023 0.427
HD 0.319 0.445 -0.087 0.342 0.325 0.62 0.467 0.074 0.483 0.246 -0.056 1 0.474 0.227 0.504 0.46 0.012 -0.091 0.445 0.491 0.252 0.247 0.034 0.475 0.513 0.267 -0.09 0.976 0.341 0.169
INTC -0.054 0.331 0.268 -0.075 -0.035 0.038 0.783 0.288 0.23 0.674 -0.293 0.474 1 -0.078 0.121 0.602 -0.138 0.214 0.772 0.059 0.649 0.299 0.161 0.176 0.202 -0.002 0.273 0.593 -0.191 0.188
IBM 0.079 0.108 0.068 0.089 -0.26 -0.126 -0.041 0.894 0.329 0.387 0.225 0.227 -0.078 1 0.578 0.187 0.329 0.208 -0.1 0.135 -0.203 0.29 0.741 0.126 -0.312 -0.137 0.09 0.746 -0.284 0.241
JNJ 0.103 0.229 0.032 -0.127 0.009 0.089 0.291 0.552 0.495 0.098 0.002 0.504 0.121 0.578 1 0.175 0.08 0.303 0.237 0.128 0.006 0.461 0.527 -0.054 -0.14 0.174 -0.047 0.929 -0.124 0.364
JPM 0.055 0.702 0.187 0.228 0.082 0.186 0.606 0.312 0.426 0.451 0.123 0.46 0.602 0.187 0.175 1 0.156 0.149 0.47 0.074 0.354 0.358 0.278 0.18 0.492 0.037 0.307 0.756 0.017 0.087
MCD 0.148 0.234 0.158 0.195 0.262 0.177 -0.044 0.134 0.342 -0.044 0.434 0.012 -0.138 0.329 0.08 0.156 1 0.14 -0.167 -0.086 -0.099 -0.071 0.244 0.426 0.107 -0.194 0.096 0.787 0.258 0.173
MRK 0.152 0.118 0.578 -0.279 -0.417 -0.295 0.53 0.319 -0.049 0.46 -0.324 -0.091 0.214 0.208 0.303 0.149 0.14 1 0.533 -0.587 0.513 0.348 0.32 0.015 -0.255 0.086 0.577 0.914 -0.065 0.528
MSFT -0.012 0.201 0.488 -0.207 -0.266 -0.139 0.785 0.383 0.06 0.765 -0.436 0.445 0.772 -0.1 0.237 0.47 -0.167 0.533 1 -0.192 0.785 0.449 0.286 0.244 0.092 0.035 0.552 0.83 -0.264 0.502
NKE 0.074 0.036 -0.501 0.493 0.585 0.537 -0.251 -0.136 0.438 -0.308 0.316 0.491 0.059 0.135 0.128 0.074 -0.086 -0.587 -0.192 1 -0.226 -0.064 0.072 0.186 0.537 -0.101 -0.56 0.189 0.193 -0.212
PFE 0.08 0.19 0.509 -0.105 -0.199 -0.067 0.69 0.024 -0.078 0.686 -0.434 0.252 0.649 -0.203 0.006 0.354 -0.099 0.513 0.785 -0.226 1 0.285 0.048 0.172 0.17 0.003 0.507 0.955 -0.001 0.394
PG 0.024 0.241 0.132 -0.101 0.014 -0.066 0.35 0.44 0.05 0.232 -0.37 0.247 0.299 0.29 0.461 0.358 -0.071 0.348 0.449 -0.064 0.285 1 0.309 -0.017 -0.016 0.108 0.253 0.856 -0.199 0.277
KO 0.159 0.155 0.307 -0.119 -0.191 -0.118 0.39 0.718 0.362 0.343 0.035 0.034 0.161 0.741 0.527 0.278 0.244 0.32 0.286 0.072 0.048 0.309 1 0.054 -0.238 -0.081 0.382 -0.454 -0.257 0.419
TRV 0.241 -0.13 0.238 0.335 0.357 0.388 0.112 0.065 0.272 0.116 0.509 0.475 0.176 0.126 -0.054 0.18 0.426 0.015 0.244 0.186 0.172 -0.017 0.054 1 0.095 0.038 0.245 0.948 -0.037 0.495
UTX 0.226 0.429 -0.314 0.591 0.479 0.567 0.128 -0.419 0.258 -0.091 0.157 0.513 0.202 -0.312 -0.14 0.492 0.107 -0.255 0.092 0.537 0.17 -0.016 -0.238 0.095 1 0.17 -0.23 0.882 0.52 -0.247
UNH 0.541 0.136 -0.004 -0.079 0.031 0.117 0.052 0.11 -0.14 0.076 -0.404 0.267 -0.002 -0.137 0.174 0.037 -0.194 0.086 0.035 -0.101 0.003 0.108 -0.081 0.038 0.17 1 0.079 0.916 0.176 -0.139
VZ -0.061 0.118 0.665 -0.081 -0.28 -0.432 0.634 0.634 -0.174 0.515 -0.264 -0.09 0.273 0.09 -0.047 0.307 0.096 0.577 0.552 -0.56 0.507 0.253 0.382 0.245 -0.23 0.079 1 0.898 -0.362 0.595
V 0.905 0.941 0.669 0.893 0.498 0.831 0.246 0.742 0.736 0.517 0.216 0.976 0.593 0.746 0.929 0.756 0.787 0.914 0.83 0.189 0.955 0.856 -0.454 0.948 0.882 0.916 0.898 1 0.923 0.962
WMT 0.561 0.399 -0.192 0.338 0.067 0.63 -0.152 -0.578 0.07 -0.155 0.023 0.341 -0.191 -0.284 -0.124 0.017 0.258 -0.065 -0.264 0.193 -0.001 -0.199 -0.257 -0.037 0.52 0.176 -0.362 0.923 1 -0.107
DIS 0.187 -0.008 0.451 -0.119 -0.139 -0.11 0.289 0.197 0.061 0.314 0.427 0.169 0.188 0.241 0.364 0.087 0.173 0.528 0.502 -0.212 0.394 0.277 0.419 0.495 -0.247 -0.139 0.595 0.962 -0.107 1 ;

param nbindice := 4;
```

FIGURE 9.1 – Fichier data pour la grande distribution

```
== 24 =================
display Indice ;
Indice [*] :=
BBBY  0
  CA  0
COST  1
DLTR  1
 KER  0
  KR  0
 MEO  1
SHLD  0
 SWY  1
 TGT  1
TSCO  1
 WMT  0
```

FIGURE 9.2 – L'indice à 6 actions

```
== 29 ================
display Indice ;
Indice [*] :=
BBBY  0
  CA  1
COST  1
DLTR  1
 KER  1
  KR  0
 MEO  0
SHLD  1
 SWY  1
 TGT  1
TSCO  1
 WMT  0
;
```

FIGURE 9.3 – L'indice à 8 actions

Chapitre 10

Dow Jones Industrial Average

Cet exemple aurait été le plus intéressant du fait qu'il est exclusivement construit sur le prix de l'action. J'aurais ainsi pu comparer les méthodes de pondération sur cet exemple et comparer réellement le cours de l'indice aux cours que j'obtiens pour mon propre indice (les autres indices, tels que le CAC40, ont un cours trop complexe à évaluer dans le cadre de ce travail de projet recherche).

Cependant CPLEX a cessé de fonctionner, me laissant avec le solveur Knitro qui est limité à 300 contraintes. Après avoir compilé les cours des 30 valeurs du Dow Jones Industrial Average, avoir calculé les différents ROI ainsi que les corrélations, je n'ai pu m'en servir pour en sortir un indice du fait des caprices d'AMPL.

C'est sans doute là ma plus grande déception de ce projet recherche, d'autant plus frustrant qu'il n'est qu'une limitation informatique.

Quatrième partie

Critère de pondération

Pour pondérer les résultats, on se sert en premier lieu de la méthode vue en ainsi que dans le premier compte-rendu sur le prix de l'action. Je me suis également intéressé à la capitalisation boursière puisque c'est un exemple de pondération classique pour la plupart des indices boursiers. Enfin, devant les récents papiers financiers disant que la méthode de capitalisation boursière était obsolète (puisque d'autres méthodes arrivent à un rendement supérieur), je me suis intéressé à la simple méthode de la moyenne arithmétique.

Il ne faut pas oublier également que ces méthodes se font non seulement en regard des actifs retenus dans l'indice et présentés dans la partie précédente à travers les figures indice6 et indice8 mais aussi par *les actifs qu'ils représentent* ! C'est là le point crucial de l'indice : si sur les 6 actions retenus pour le marché 12, chacune en représente 2, ou si au contraire, une action en représente 7 et les 5 autres se représentent eux-mêmes, cela change entièrement la composition de notre portefeuille (conséquence du comportement du marché).

Pour faire les calculs de pondération, il faut donc également se procurer la matrice "Représente" qui, dans chaque colonne non nulle, présente des 1 aux lignes dont les actions sont représentés par la colonne en question.

Cette matrice se présente de la façon suivante :

```
== 21 =================
display Represente ;
Represente [*,*]
:     BBBY  CA COST DLTR KER  KR MEO SHLD SWY TGT TSCO WMT    :=
BBBY   0     0   1    0    0   0   0   0    0   0   0    0
CA     0     0   0    0    0   0   1   0    0   0   0    0
COST   0     0   1    0    0   0   0   0    0   0   0    0
DLTR   0     0   0    1    0   0   0   0    0   0   0    0
KER    0     0   0    0    0   0   0   0    0   1   0    0
KR     0     0   1    0    0   0   0   0    0   0   0    0
MEO    0     0   0    0    0   0   1   0    0   0   0    0
SHLD   0     0   0    0    0   0   1   0    0   0   0    0
SWY    0     0   0    0    0   0   0   0    1   0   0    0
TGT    0     0   0    0    0   0   0   0    0   1   0    0
TSCO   0     0   0    0    0   0   0   0    0   0   1    0
WMT    0     0   1    0    0   0   0   0    0   0   0    0
;
```

FIGURE 10.1 – Matrice des représentants du marché

Chapitre 11

Absurdité du prix de l'action

Certains indices se basent sur le prix de l'action pour pondérer les actifs dans l'indice. C'est par exemple le cas du Dow Jones Industrial Average. Certains investisseurs privilégient encore cette technique ainsi que celle de la capitalisation boursière, bien que comme on le verra plus tard, les fonds d'investissement recherchent de plus en plus des investissements alternatifs, afin d'éviter les pièges des deux techniques classiques mais aussi les surperformer ...

En effet, le prix de l'action est bien souvent arbitraire : il relève d'un choix de la société émettrice pour un couple (nombre d'actions, volume). On peut décider de n'avoir que des investisseurs fortunés et donc une action forte pour une volume faible. Le prix ne traduit donc en rien le dynamisme, la volatilité d'une action, etc. De plus, les prix des actions d'un marché peuvent fortement varier : il s'échange des actions de l'ordre de 500 \$ ainsi que d'autres de l'ordre du \$ sur le marché sans que l'on puisse conclure sur la proéminence de l'une ou l'autre sur le marché.

Ce critère, en plus de paraître trop arbitraire, montre des résultats médiocres en termes de retours sur investissement comme le montre le graphique suivant. C'est pour cette raison que certains fonds d'investissement vont même jusqu'à pondérer sur l'inverse du prix de l'action !

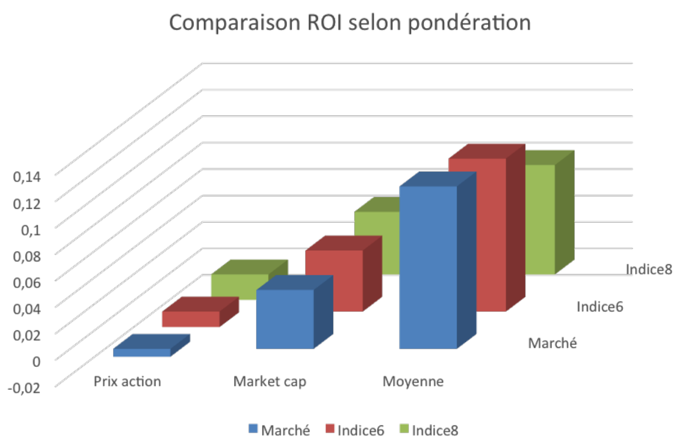


FIGURE 11.1 – Fichier de résultats pour la grande distribution

Chapitre 12

Capitalisation boursière

Comme on a pu le remarquer sur la figure des résultats, c'est la méthode de pondération avec la capitalisation boursière totale des entreprises représentées par chaque action dans l'indice. En effet, le retour sur investissement est quasiment identique pour les indices 6 et 8 que pour le marché 12, ce qui est remarquable ! Même si je ne peux l'expliquer, il semblerait que le fait que la plupart des indices boursiers se basent sur la capitalisation boursière pour le choix des portefeuilles soit plus que fondé, puisque l'objectif reste non pas de surperformer le marché mais bien de l'imiter au mieux.

Le seul bémol à émettre quant à ce résultat ne concerne pas le reflet du marché mais bien le retour sur investissement qu'il en donne puisque l'on voit bien sur la figure suivante qu'une simple méthode de moyenne arithmétique (où chacune des 6, 8 ou 12 actions est prise en quantité égale) surperforme l'indice. Cette critique se trouve également sur l'article anglais de wikipédia *Stock market index*. On commentera ce fait dans le chapitre suivant.

Une comparaison des deux méthodes précédentes est disponibles ici :

Le lecteur souhaitant approfondire la question trouvera des informations utiles au lien suivant : $http : \backslash\backslash www.schwab.com\backslash public\backslash schwab\backslash resource\_center\backslash fundamentally\_weighted\_indexes.html$

Comparing index characteristics

	F FUNDAMENTAL	M MARKET-CAP
Portfolio weighting	Economic factors	Cap weighting
Portfolio construction	Value tilt	Larger-cap tilt
Portfolio turnover	Reconstitution and rebalancing	Reconstitution
Tax efficiency	Typically	Typically
Cost structure	Low cost	Lowest cost

FIGURE 12.1 – Comparaison prix de l'action - capitalisation boursière

Chapitre 13

Moyenne arithmétique

Cette méthode de pondération soulève à la fois beaucoup d'espoir mais aussi de nombreuses problématiques.

Commençons bien sûr par les côtés positifs. Comme j'ai pu le lire dans de nombreux articles financiers et comme j'ai pu le découvrir moi même, une simple moyenne arithmétique surperforme bien souvent le marché. En effet, en prenant une part égale de chaque actif de l'indice, on obtient un rendement nettement supérieur à celui obtenu pour la capitalisation boursière ou le prix de l'action. Cela tend à confirmer qu'en termes de retour sur investissement la capitalisation n'est pas la méthode la plus indiquée. En revanche, cela n'enlève en rien le fait que, probablement, c'est la méthode de capitalisation boursière qui rend le mieux compte de l'évolution *réelle* de l'économie.

On peut donc chercher à surperformer un indice basé sur la capitalisation boursière en batissant à partir de celui-ci un indice pondéré par une simple moyenne ! Ce qui se traduirait par surperformer l'évolution réelle de l'économie. Cela induit bien sûr un risque plus important (les "grosses" capitalisations sont considérées moins volatiles).

En revanche, cette méthode de pondération pose également un certain nombre de questions. En effet, on remarque que même si les résultats des indices 6 et 8 sont bien plus proches du marché que ce n'était pas le cas pour la méthode du prix de l'action, l'indice 8 est plus loin du résultat obtenu pour le marché que l'indice 6. Ceci contredit le fait que plus on augmente le nombre d'actifs dans l'indice plus on se rapproche du marché. Nous ne pouvons donc conclure sur la précision de ce critère de pondération par rapport au marché.

Ensuite, le fait que nous prenons chaque actif à part égale dans l'indice est problématique dans le sens où nous ne prenons pas du tout en compte le fait que certaines actions en représentent d'autres ! Par exemple, une action qui en représenterait 7 autres dans l'indice serait compté au même titre que les 5 autres actifs de l'indice qui se représentent eux-mêmes, ce qui semble illogique et éloigné du but initial de refléter au mieux le marché. C'est peut-être bien ici que se situe l'erreur de précision dont nous parlions au paragraphe précédent.

A l'inverse, on ne peut pas affecter un poids égal au nombre d'actions représentées dans le marché puisque cela voudrait dire que, toujours dans l'exemple précédent, une action représente 7/12 ème du marché à elle-seule, donc 7/12 de notre investissement total. Encore une fois, nous nous éloignons du but initial qui était de minimiser les risques spécifiques (ou risques intrinsèques) : les risques liés à un titre en particulier, à la mauvaise gestion d'une entreprise en particulier, et n'affectant pas le marché, afin de minimiser le risque du portefeuille au seul risque de marché (le marché que l'on veut modéliser).

Le lecteur souhaitant approfondir la question trouvera des informations utiles au lien suivant :

$http : \backslash\backslash online.wsj.com\backslash news\backslash articles\backslash SB10001424052748703555804576101812395730494$

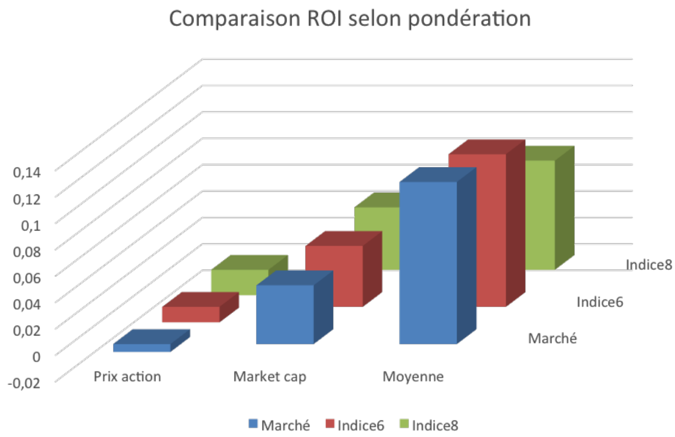


Figure 13.1 – Fichier de résultats pour la grande distribution

Chapitre 14

Conclusion

Concernant les objectifs de ce projet recherche, je les considère atteints dans la mesure où je me suis intéressé d'une part à l'algorithmie pure des solveurs utilisés dans le cas de la programmation linéaire en nombres entiers. Ce travail a fait l'objet d'un premier compte-rendu. Comme prévu au début du projet recherche, je n'ai pas eu le temps de concevoir et tester mon propre modèle de résolution. Ce temps supplémentaire m'a permis d'aller plus loin dans les concepts financiers et les résultats "simples" de mes modèles qui ont fait l'objet de ce deuxième compte-rendu.

Concernant ces résultats plus particulièrement, j'ai pu vérifier la plupart des choses très empiriques que j'ai pues lire dans des magazines financiers, à savoir la méthode de construction des indices, la méthode de pondération, l'absurdité de la méthode du prix de l'action, la fiabilité de la méthode par capitalisation boursière qui reste donc la plus adaptée au reflet d'une économie, ... J'ai aussi pu confirmer qu'une simple moyenne arithmétique pouvait parfois permettre d'obtenir un bon rendement mais j'ai aussi exhibé les problèmes liés à cette méthode, là où les articles que j'ai pus lire ne sont pas allés.

On comprend dès lors tout l'intérêt des mathématiques, de l'informatique, et en particulier de l'optimisation en finance, qui paraissent nettement plus solides que la "théorie" qui encadre l'analyse graphique des courbes de laquelle on dégage des tendances (dans le cas du management actif). On comprend donc d'autant plus l'arrivée massive d'ingénieurs dans le milieu de la finance de marché et en particulier la structuration des produits, l'analyse quantitative, ...

Si je m'intéresse plus particulièrement aux compétences que j'ai pu acquérir, je dirais d'une part que du côté de l'optimisation j'ai pu affrondir le cours d'ANO par tout ce qui concerne la programmation en nombres entiers mais aussi faire des lectures sur l'état actuelle de la recherche dans ces domaines et tout ce qui permet d'accélérer les algorithmes tels que Branch&Bound. Un compte-rendu entier (la moitié de mon travail en projet recherche) y est dédié. Par ailleurs, j'ai pu approfondir l'utilisation de Matlab et ses toolbox, de même qu'apprendre un nouveau langage, AMPL, et plusieurs de ses solveurs, ce qui n'est pas négligeable pour mon projet professionnel.

D'autre part, concernant la finance, j'ai pu approfondir également mes connaissances existantes et limitées dans le domaine. C'est, pour le coup, un plus indéniable pour mon CV, puisque, souhaitant travailler dans le monde de la finance de marché, ce n'est pas les mathématiques qui me manquent pour y entrer, mais bien les connaissances purement financières qui ne sont pas enseignées à l'école. Le seul problème que j'y ai rencontré et qui est gênant pour un ingénieur dans le milieu est le caractère très empirique de ce que j'ai pu lire à ce sujet, mais qui dans les faits ce sont toujours vérifiés.

En conclusion, le projet recherche a été une bonne expérience que je recommanderai à tous les centraliens souhaitant travailler en autonomie. Il faut cependant avoir une idée de départ, même vague, et être certain que l'on peut maintenir pour ce sujet une motivation nécessaire à la rédaction des compte-rendus, aux recherches parfois fastidieuses mais aussi aux échecs (la programmation dans mon cas).