

# Scénario: Exploration, classification des encours boursiers parisiens

## Résumé

Scénario d'analyse d'un jeu de données : l'ensemble des séries des encours boursier à Paris. Description et classification de ces courbes. Utilisation des différentes techniques descriptives uni, bi et [multidimensionnelles](#) : [ACP](#), [k-means](#), [CAH](#).

## 1 Objectif

Le travail proposé s'intéresse aux cours des actifs boursiers sur la place de Paris de 2000 à 2005. Seules 349 entreprises sont considérées, celles qui ont été régulièrement cotées sur la période considérée. Les autres, présentant trop de données manquantes car introduites ou exclues au cours de la période, par exemple à la suite d'une fusion, ont été éliminées. Le travail ne se veut pas exhaustif mais illustratif.

Remarque : Il serait tout à fait intéressant de compléter les données en considérant les cours de 2006 à 2008.

L'objectif principal et la réalisation d'une classification des entreprises au regard du comportement de leur titre au cours de la période et plus particulièrement autour des difficultés rencontrés en 2002. Il s'agit donc de déterminer des groupes ou classes homogènes quant à ce comportement. Ce travail est relativement original car absent des analyses financières classiques que se focalisent sur le comportement d'un titre avec des indicateurs très sophistiqués associés à la série chronologique. L'originalité vient de ce que ce sont toutes les séries qui sont simultanément étudiées. Ce pourrait être le travail préalable d'un analyste qui, voulant constituer un portefeuille "équilibré", chercherait des classes de comportement homogènes afin de compléter une information plus spécifique sur chaque entreprise et leur secteur d'activité. Il ne s'agit donc pas d'une étude sur les qualités prédictives de modèles mais simplement exploratoire. Après un descriptif de l'origine des données, la première partie concerne le traite-

ment préalable, la deuxième une description multidimensionnelle par analyse en composantes principales, la troisième, la recherche d'une classification, de sa représentation et de son interprétation.

## 2 Données

Beaucoup de sites, dont les principaux moteurs de recherche, proposent des historiques des grandes places boursières. Ces historiques fournissent pour une action (définie par son code), pour une période donnée et pour chaque jour : le cours d'ouverture, le maximum, le minimum, le cours de fermeture et le nombre de titres échangés. La première tâche est un travail de compilation afin de récupérer l'ensemble des titres gérées, ici par la bourse de Paris. La deuxième est un filtrage afin de ne conserver que les titres cotés sur toute une période (2000 à 2005). La troisième nécessite un choix de "granularité". En effet, nous nous intéressons au comportement global sur la période en négligeant les fluctuations du titre à court terme ainsi que la volatilité (la variance) associée. Ceci nécessite un lissage des données qui peut être opéré de façon très diverses (splines, ondelettes...). Le choix fait est celui de la simplicité : on s'intéresse aux moyennes mensuelles des cours. Cet indicateur étant relativement "grossier", il n'est plus important de savoir s'il s'agit de la moyenne du cours à l'ouverture, la fermeture, le min ou le max, c'est le montant à l'ouverture qui a été choisi.

Les données sont disponibles dans le répertoire "data" : fichier `paris2005.dat`.

## 3 Travail à réaliser

Différents outils sont disponibles pour répondre à la question posée, il s'agit de les mettre en œuvre afin d'y répondre de façon pertinente. Trois phases seront considérées ci-dessous qui abordent les principales questions. Les initiatives personnelles, les commentaires et interprétations complémentaires seront bienvenues.

### 3.1 Description élémentaire des séries

- Tracer simultanément tous les cours en fonction du temps puis étudier la distribution (histogramme, diagramme boîte) de la variable statistique

regroupant tous les cours pour toute la période. Commenter son profil, quelle transformation nécessite ces données.

- Retracer les trajectoires des cours, une information parasite et inutile pour les objectifs de l'étude est présente dans les données ; Quelle est-elle ? par quelle transformation y remédier ?
- Refaire le même graphe, commenter l'allure générale des trajectoires.

## 3.2 Analyse en composantes principales

Calculer les analyses en composantes principales centrées réduites et non réduites.

- Commenter l'intérêt respectif de ces deux analyses. Regarder en particulier les "biplots" des deux premiers axes. Laquelle semble-t-il opportun de privilégier.
- Tracer l'éboulis des valeurs propres, les diagrammes boîtes des composantes principales. Discuter le choix de la dimension.
- Tracer les coordonnées des dix premiers vecteurs propres comme fonction du temps. En quoi cette représentation a-t-elle un sens ? À quelle interprétation des premiers axes factoriels conduit cette représentation ?
- Interprétation du premier plan factoriel. Remarquer la répartition particulière des vecteurs CM1 à CM72. Identifier et spécifier les caractéristiques des entreprises les plus remarquables en marge du nuage.

## 3.3 Classification

Il s'agit de regrouper les entreprises en considérant les profils de leurs cours. Rechercher différentes classification. Discuter le nombre de classes. Tester différentes distances entre trajectoires (euclidienne, 1-corrélation, 1-corrélation<sup>2</sup>), différents types de saut.

- Ascendante hiérarchique,
- par réallocation dynamique (kmeans),
- par réallocation dynamique initialisée à partir des barycentres des classes de la classification ascendante hiérarchique.
- Comparer les classifications obtenues par des tableaux croisés,

Représenter la classification qui vous semble la plus pertinente dans les axes de l'ACP ou d'un MDS. Commenter les positions respectives de ces classes dans les axes.

Enfin, tracer un graphique superposant les trajectoires d'une même classe, un graphique par classe. Commenter les profils généraux de ces classes.

## Annexe : quelques outils de programmation

### Traiter avec R la matrice des cours

```
# prise en compte des données
paris=read.table("paris2005.dat")
# Quelques cours remarquables après transformations
ts.plot(t(lparis0[c("Degrenne", "Fala",
                  "Cegedim", "Telefonica"),]),gpars=list(lty=c(1:4)))
# les vecteurs propres associés aux 6
premières valeurs propres de la fonction princomp
plot.ts(pca$loadings[,1:6])
```