

Contrôle des connaissances "Machine learning"

PIRES - KESADRI- ASLIMI

2 juin 2014

Chapitre 1

Description des données.

1.1 Introduction

Tout d'abord, nous avons utilisé pour ce projet, deux logiciels qui sont R et Rstudio. En effet, les graphiques issus de l'ACP sont difficilement interprétables avec Rstudio, tandis que les graphiques d'évolution des cours le sont. Il a donc fallu alterner entre ces deux interfaces. Concernant les packages, nous avons utilisé ici deux packages à savoir FactoMineR et Tseries. FactoMineR permet de réaliser l'analyse des composantes principales et Tseries permet de travailler sur la matrice contenant tous les cours.

1.2 Description des données

Sur les 100 entreprises composant le FTSE (Financial Times Stock Exchange), nous avons choisi de prendre 10 capitalisations parmi les 15 plus importantes de cette place financière. Il s'agit ici, de HSBC, Vodafone, British American Tobacco, Diageo, Imperial Tobacco, British Petroleum, GSK, AstraZeneca, BS Group et Tesco. On remarque que ces cours concernent des domaines divers et variés comme l'industrie pétrolière, bancaire, agro-alimentaire, pharmaceutique ou encore celle du tabac. L'objectif étant de mesurer l'impact de ces valeurs décisives sur le reste du marché.

Au départ, nous voulions prendre les dix plus grosses capitalisations du FTSE. Cependant, nous avons rencontré des problèmes sur Yahoo Finance quant à l'interprétation des données. En effet, il arrive que certains titres ne soient pas cotés en bourse certains jours (Krash, Fusion acquisition etc...), d'où un nombre inégal de valeurs téléchargées. Il a donc fallu trouver une solution, nous avons alors pris dix des quinze plus grosses capitalisations du FTSE et nous avons également changé la période d'étude des cours. Nous avons pris des cours cotés de 2005 à 2009 pour

étudier l'impact de la crise sur ces valeurs. Notre but d'étudier l'intérêt de l'ACP pour ce domaine d'étude.

Chapitre 2

Description de la méthode

2.1 Les données

1er Etape On importe les cours d'ouvertures (SERIESO) et de fermetures (SERIESC), puis on fait une moyenne de ces cours (SERIESmean). En effet, on ne se contente pas de d'utiliser les cours d'ouverture ou de fermeture car en début de séance le marché est extrêmement volatile alors qu'en fin de séance, tous les acteurs cherchent à se couvrir (Hedger). C'est pourquoi il nous a sembler judicieux de prendre une moyenne de ces deux cours pour qu'elle reflète au mieux la réalité du marché.

2nd Etape - Summary On utilise la fonction Summary, celle-ci nous donne un détail des cours et permet une analyse statistique de base sur chacun des cours. On trace alors, l'ensemble des cours sur un même graphique.

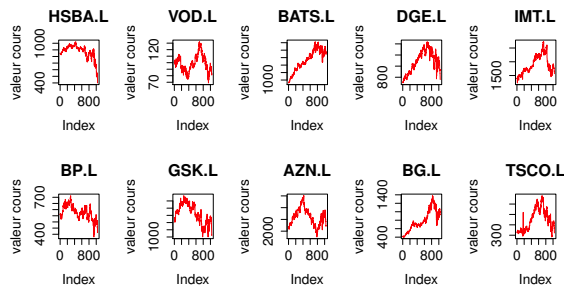


FIGURE 2.1 – 10cours

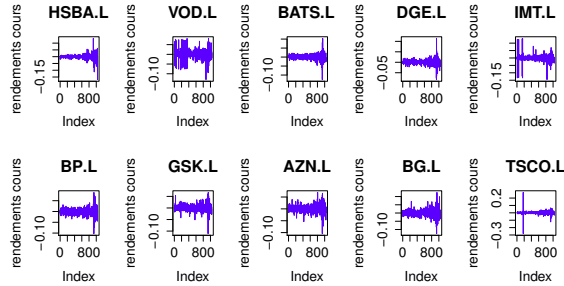


FIGURE 2.2 – 10rendement

3ème Etape - Calcul des log rendement On calcul les rendements sur les cours moyens à l'aide de la fonction `diff(log(SERIESmean))`. Puis on utilise à nouveau la fonction `Summary` et on trace les fonctions associées aux Logrendements.

2.2 Analyse des Composantes Principales (ACP)

Méthode On utilise la fonction `PCA(rSERIES)` pour réaliser une ACP. On obtient alors deux graphiques. Le premier graphique Variable Factor Map (ci-dessous) qui permet de voir comment les individus (cours) sont repartit par rapport au deux premières dimensions.

Quant au deuxième (ci-dessous), il sert a voir comment les variables (les entreprises) sont repartit par rapport aux deux premières dimensions.

2.3 III-Reconstitution des données)

Méthode On reconstitue alors, les cours de bourse à partir de la première composante principal. K correspond, ici, à la contribution de chaque cours à la composante principales.

Graph plus summary

On se sert alors des rendements obtenues pour tracer les rendements reconstitués.

2.4 Commentaire sur le code

Vous trouverez, en pièce jointe, le code R associé à ce rapport.

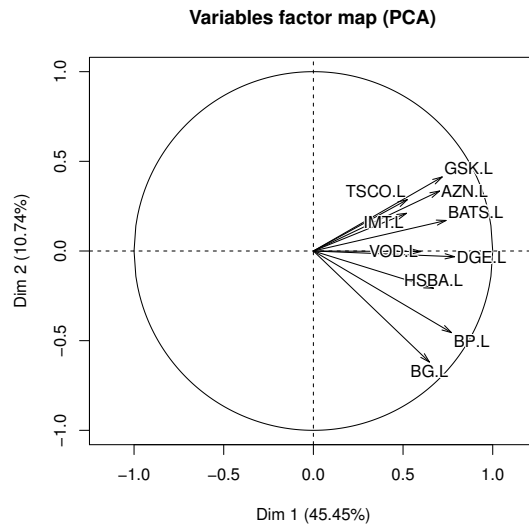


FIGURE 2.3 – camem

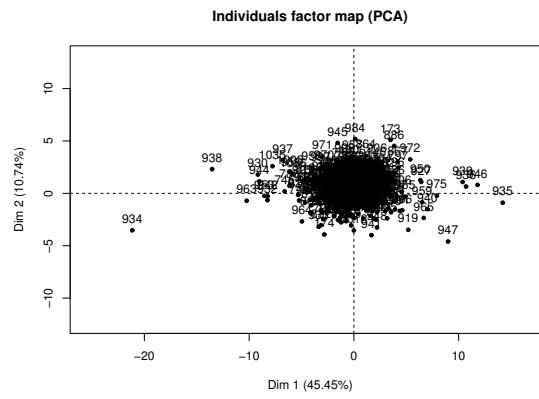


FIGURE 2.4 – nuagePTS

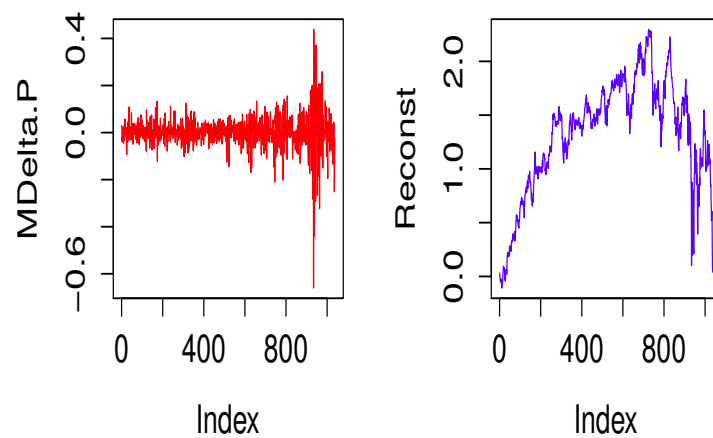


FIGURE 2.5 – coursRdtRecon

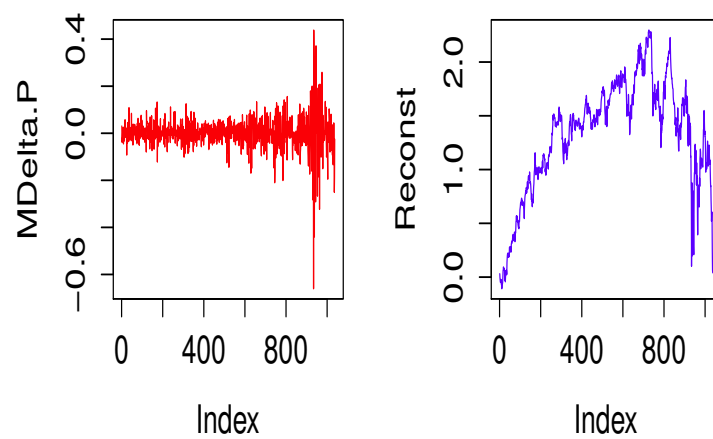


FIGURE 2.6 – coursRdtRecon

Chapitre 3

La description des résultats

3.1 Interprétation des données

Graph cour plus rendement Summary

Interprétation :

On remarque que le summary effectué confirme les graphiques obtenues.

3.2 Interprétation ACP

Nuage de pt :

Interprétation : Tout les indinivdu sont centré autour d'une

Camembert :

En dim1 :Tout les cours sont corrélée, ici entre 0,5 et 1

En dim 2 : On distingue ici que 3 groupes se demarque tres nettement. blabla
tab etc...

En dim 2, on explique 55 pourcent des données

> summary(D.PCA)

Call : PCA(rSERIES)

Eigenvalues Dim.1 Dim.2 Dim.3 Dim.4 Dim.5 Dim.6 Dim.7 Dim.8 Variance
4.545 1.074 0.845 0.762 0.715 0.590 0.490 0.416 Cumulative

D.PCAvarcoord summary 2 lignes : confirme notre interprétation plus valeur
précise

ie tabac 2 vrai valeur

3.3 Reconstitution de l'indice domestique

Graph final plus blabla

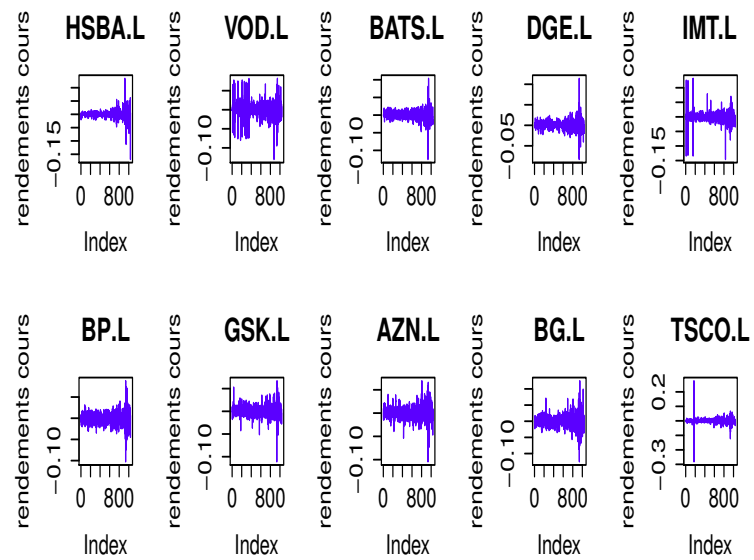


FIGURE 3.1 – 10rendement

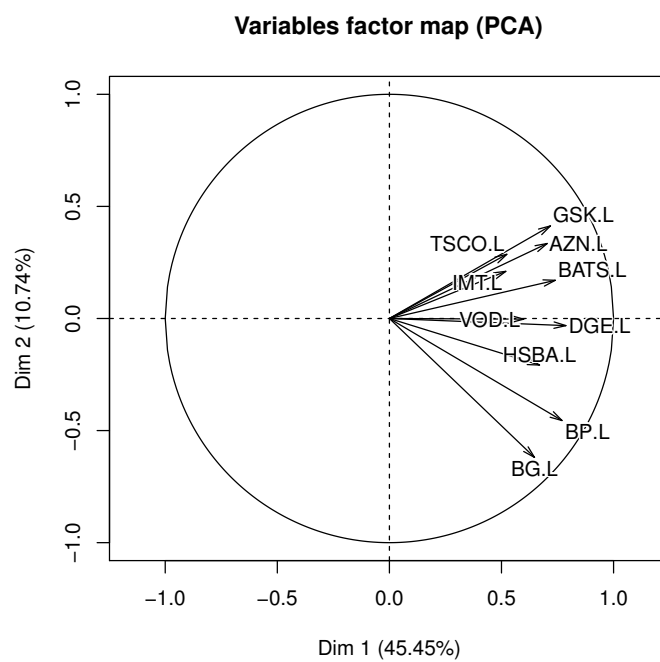


FIGURE 3.2 – camem

Chapitre 4

Conclusion

L'analyse des composantes principales s'est révélée très efficace. En effet, avec 47 % des données, nous avons pu expliquer l'indice domestique. En effet, avec seulement 10 cours, correspondant à 10 des 15 plus grosses capitalisations du FTSE, nous avons dégager la même tendance générale que celle des 100 cours de cette indice. Nous avons ainsi, mis en évidence l'intérêt de l'ACP pour l'étude des cours financiers.

Chapitre 5

Sources