

Stylized Facts of Financial Market Using the Machine Learning

*Henri Makika**

May 24, 2019

0. Faits stylisés sur les rendements des marchés financiers

1. Faits stylisés pour les séries univariées

Avant de passer à la modélisation des risques des marchés financiers, il est utile d'examiner et de revoir les caractéristiques typiques des données des marchés financiers. Celles-ci sont résumées dans la littérature sous forme de «faits stylisés» (nous renvoyons le lecteur de lire Campbell et al. 1997; McNeil et al. 2005). Ces propriétés observées ont des implications importantes pour déterminer si le modèle de risque choisi est approprié ou non. En d'autres termes, un modèle de risque qui ne prend pas correctement en compte les caractéristiques de la série chronologique des données des marchés financiers ne sera pas non plus utile pour dériver des mesures de risque. Pour les données observées sur les marchés financiers, les faits stylisés suivants peuvent être énoncés:

- i. Les données chronologiques sur les retours, en particulier les résultats journaliers, ne sont généralement pas indépendantes et sont distribuées de manière identique (*iid*). Ce fait n'est pas compromis par les faibles valeurs absolues du coefficient d'autocorrélation de premier ordre.
- ii. La volatilité des processus de retour n'est pas constante dans le temps.
- iii. Les rendements absolus ou carrés sont hautement autocorrélés.
- iv. La distribution des rendements des marchés financiers est leptokurtic¹. La survenue d'événements extrêmes est plus probable par rapport à la distribution normale.
- v. Les rendements extrêmes sont observés de près dans le temps (clustering de volatilité²).
- vi. La distribution empirique des rendements est biaisée à gauche; les rendements négatifs sont plus susceptibles de se produire que les rendements positifs.

Notre objectif est donc de montrer comment analyser le marché financier en utilisant les techniques à la frontière technologique (Machine Learning).

Les packages pour cette analyse :

```
library(timeSeries) # Pour l'analyse de times series
library(fBasics)    # Pour l'analyse basique sur la finance
library(evir)       # Le package où se trouve la base de données
```

*Master's degree in economics (University of Campinas, São Paulo). E-mail : hd.makika@gmail.com

¹Voir statistique descriptive

²Terme approprié à la finance

Détails

Il s'agit donc de l'ensemble de données des déclarations quotidiennes, qui est contenu dans le package *evir*. Cette série commence en janvier 1973 et se termine en juillet 1996 et comprend 6146 observations. nous allons vérifier si les faits stylisés (cités ci-dessus) s'appliquent au rendement des actions de Siemens.

Les packages *fBasics* (Würtz et al. 2014) et *evir* sont chargés en premier avant de commencer l'analyse. Les fonctions contenues dans ces packages sont utilisées pour produire certains graphiques. Ensuite, le jeu de données *Siemens* est chargé et converti en un objet de la classe *timeSeries* avec la fonction du même nom. La représentation chronologique des pourcentages de rendements, une représentation en boîte de ces dernières, ainsi que la fonction de l'autocorrélation et l'autocorrélation partielle, sont ensuite produites et illustrées ci-bas. Comme on peut le déduire du graphique de la série chronologique, il existe un regroupement de la volatilité.

Téléchargement de la base de données

```
data(siemens)
```

Analyse des données

```
SieDates = as.character(format(as.POSIXct(attr(siemens, "times")),
                                "%Y-%m-%d"))

SieRet = timeSeries(siemens*100, charvec = SieDates)
colnames(SieRet) = "SieRet"
```

Stylised facts I

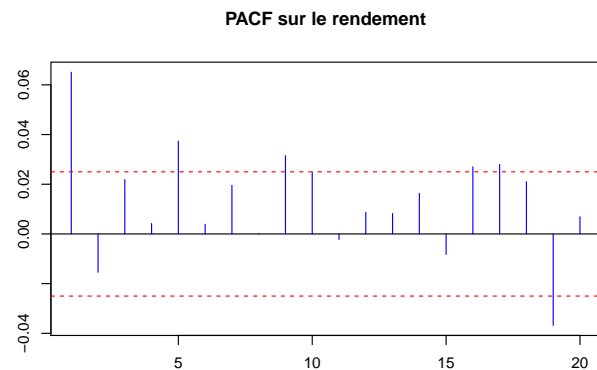
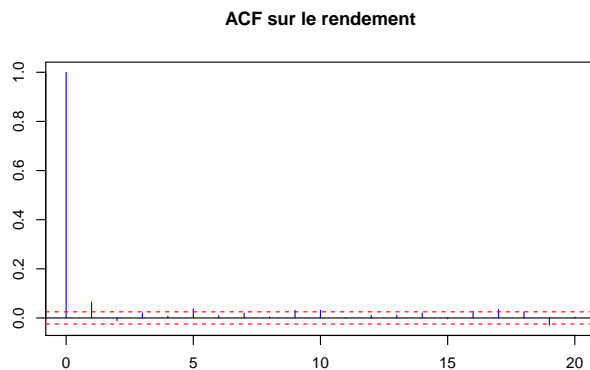
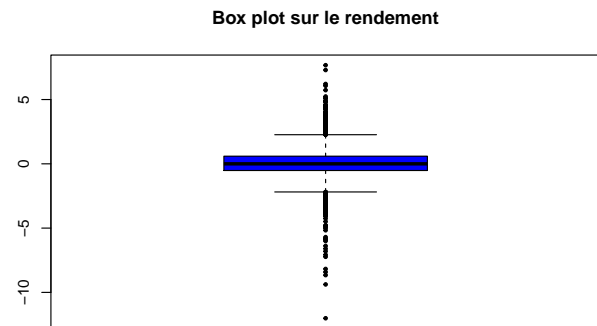
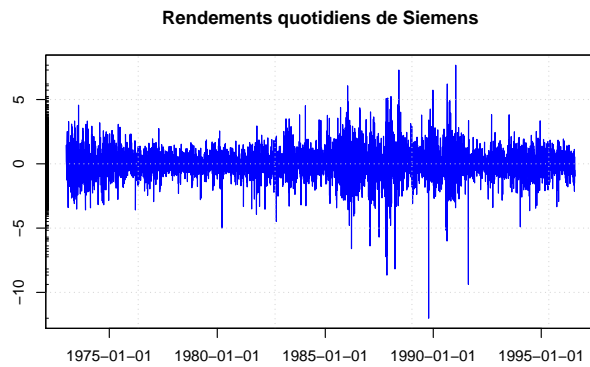
```
par(mfrow = c(2, 2))

seriesPlot(SieRet, title = FALSE, main = "Rendements quotidiens de Siemens",
           col = "blue")

boxplot(SieRet, title = FALSE, main = "Box plot sur le rendement",
        col = "blue", cex = 0.5, pch = 19)

acf(SieRet, main = "ACF sur le rendement", lag.max = 20,
    ylab = "", xlab = "", col = "blue", ci.col = "red")

pacf(SieRet, main = "PACF sur le rendement", lag.max = 20,
    ylab = "", xlab = "", col = "blue", ci.col = "red")
```



Les rendements sont biaisés à gauche et les queues épaisses sont évidentes, comme le montre le graphique en boîte (panneau supérieur droit). La perte la plus importante est survenue le 16 octobre 1989, à -12.01%. Le rendement le plus élevé de 7.67% a été enregistré le 17 janvier 1991. L'asymétrie est de -0,52 et l'excès de kurtosis de 7.74, ce qui indique clairement une queue épaisse. La fonction d'autocorrélation (ACF) et la fonction d'autocorrélation partielle (PACF) suggèrent une légère autocorrélation de premier ordre. Incidemment, la série montre une variation systématique de la fréquence hebdomadaire; Bien que significatif, il est beaucoup moins prononcé que l'autocorrélation quotidienne.

```
summary(SieRet)
```

```
##      SieRet
##  Min.   :-12.01116
##  1st Qu.: -0.51991
##  Median :  0.00000
##  Mean   :  0.02131
##  3rd Qu.:  0.59380
##  Max.    :  7.67287
```

Stylised facts II

```
SieRetAbs = abs(SieRet)

SieRet100 <- tail(sort(abs(series(SieRet))), 100)[1]

idx <- which(series(SieRetAbs) > SieRet100, arr.ind = TRUE)

SieRetAbs100 = timeSeries(rep(0, length(SieRet)),
```

```

charvec = time(SieRet))
SieRetAbs100[idx, 1] = SieRetAbs[idx]

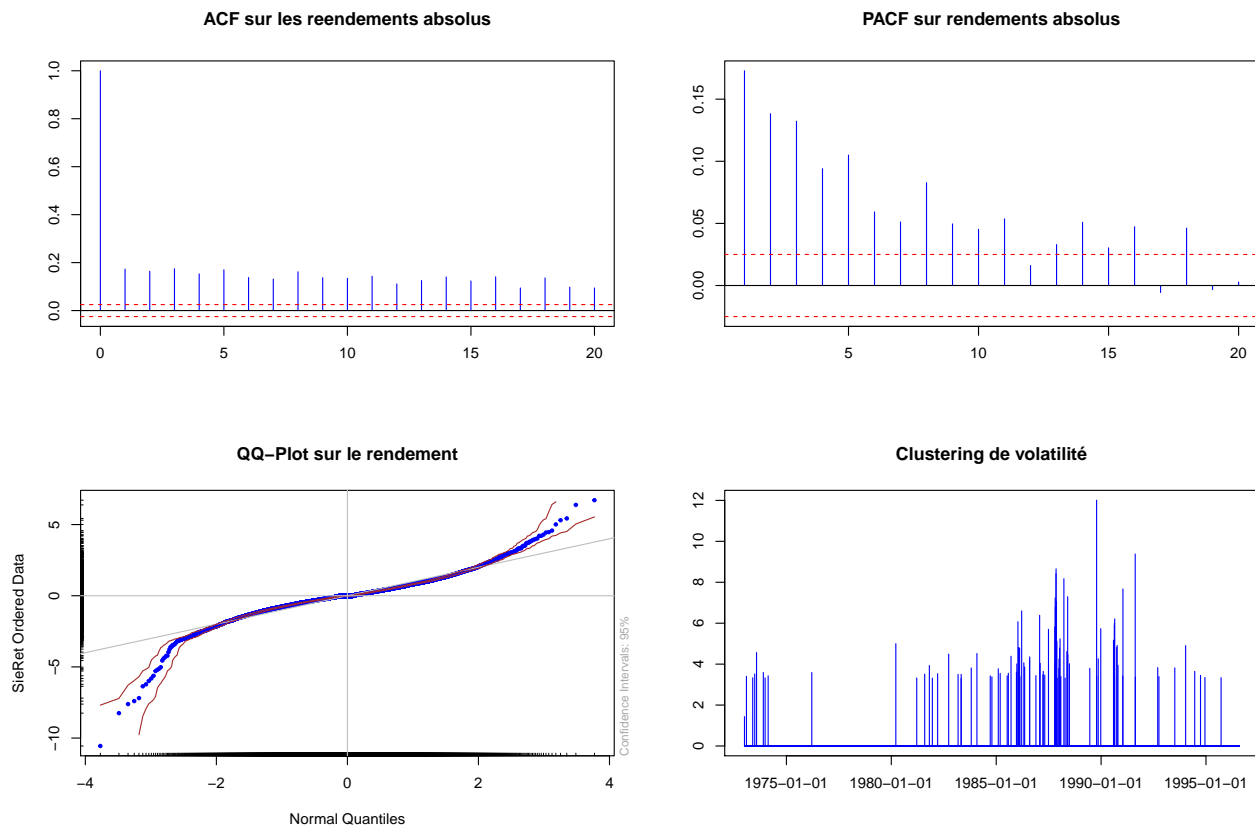
par(mfrow=c(2,2))
acf(SieRetAbs, main = "ACF sur les reendements absolus", lag.max = 20,
    ylab = "", xlab = "", col = "blue", ci.col = "red")

pacf(SieRetAbs, main = "PACF sur rendements absolus", lag.max = 20,
    ylab = "", xlab = "", col = "blue", ci.col = "red")

qqnormPlot(SieRet, main = "QQ-Plot sur le rendement", title = FALSE,
    col = "blue", cex = 0.5, pch = 19)

plot(SieRetAbs100, type = "h", main = "Clustering de volatilité",
    ylab = "", xlab = "", col = "blue")

```



Dans le tableau supérieur de nos graphiques, la fonction d'autocorrélation et de l'autocorrélation partielle des rendements absolus sont tracées. On voit que, ces différences sont significativement différentes de zéro et ne s'effacent que lentement. Dans le même tableau (graphique inférieur gauche), le graphique quantile-quantile (QQ) est produit par rapport à la distribution normale. Les asymétries négatives et les queues lourdes se reflètent dans leurs valeurs quantitatives sur ce graphique. Enfin, dans le tableau (inférieur droit) les 100 plus grands rendements absolus ont été extraits de l'objet *SieRend*. Ces valeurs sont affichées dans le graphique inférieur droit. Ce graphique chronologique confirme plus clairement ce qui pouvait déjà être déduit du graphique supérieur gauche de notre tableau : premièrement, l'existence d'un regroupement de la volatilité (clustering volatility); et deuxièmement, que les rendements deviennent plus volatils dans la seconde moitié de la période considérée. Bien que ces faits stylisés ne soient illustrés que par les rendements boursiers de

Siemens, ils sont non seulement valables pour la quasi-totalité des rendements boursiers pour cette période, mais s'appliquent également à d'autres classes d'actifs, telles que les obligations, les devises et les contrats à terme standardisés.

2. Faits stylisés pour des séries multivariées

Le point (1) présentait les faits stylisés pour les rendements univariés des marchés financiers. Du point de vue du portefeuille, les caractéristiques des séries à rendement multivarié présentent un intérêt aussi considérable. Voici donc les faits stylisés pour les séries multivariées :

- i. La valeur absolue des corrélations croisées entre les séries de rendements est moins prononcée et les corrélations contemporaines sont généralement les plus fortes.
- ii. En revanche, les rendements absolus révèlent des corrélations croisées élevées. Cette conclusion empirique est similaire au cas univarié.
- iii. Les corrélations contemporaines ne sont pas constantes dans le temps.
- iv. Les observations extrêmes d'une série de rendements s'accompagnent souvent d'extrêmes dans l'autre série de rendements.

Pour ce point (2), les données boursières européennes quotidiennes pour la France, l'Allemagne et le Royaume-Uni sont utilisées. Cet ensemble de données est d'abord chargé dans R de manière automatique et converti à partir d'un objet avec l'attribut de classe *mts* en un objet *zoo*. Le package *zoo* (voir Zeileis et Grothendieck, 2005) est une alternative au package *timeSeries* qui était utilisé dans le cas univarié. Ensuite, un graphique de série chronologique des données d'index est produit. La corrélation entre ces trois marchés boursiers européens est assez évidente dans le graphique ci-dessous.

Packages

```
library(zoo)
```

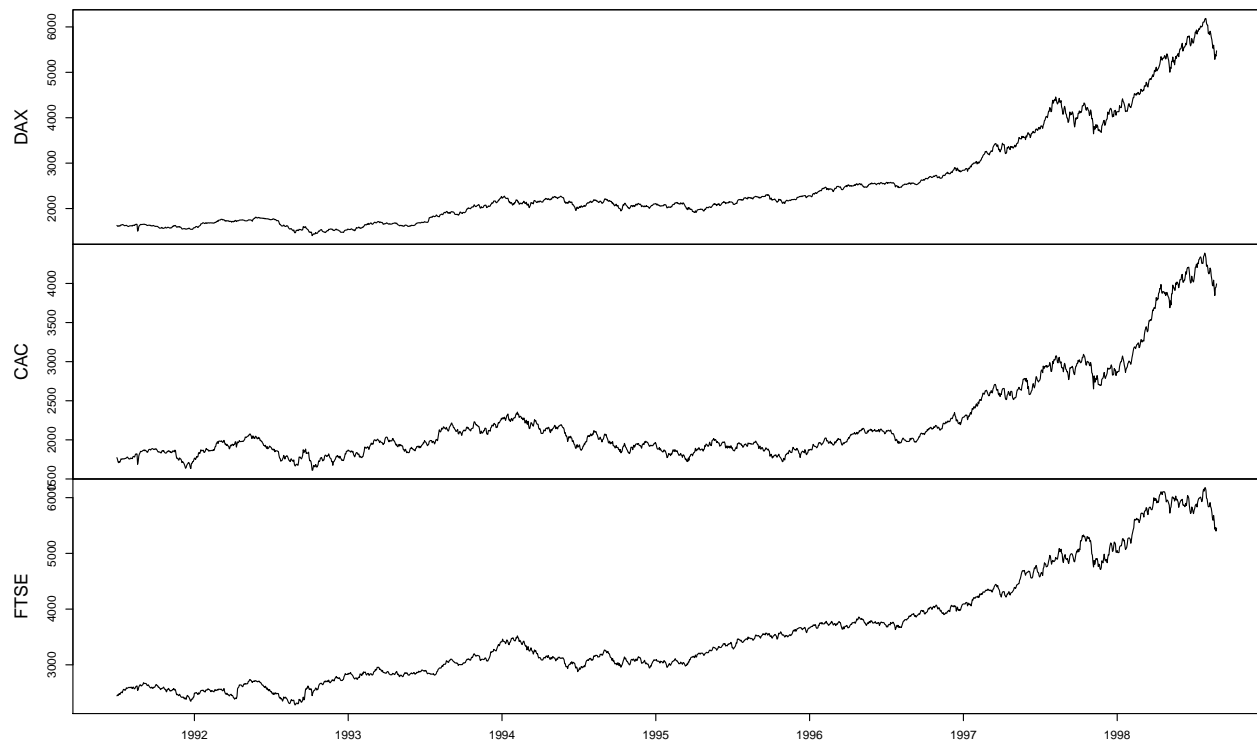
Lecture des données

```
data("EuStockMarkets")
head(EuStockMarkets[,-2])

##           DAX      CAC    FTSE
## [1,] 1628.75 1772.8 2443.6
## [2,] 1613.63 1750.5 2460.2
## [3,] 1606.51 1718.0 2448.2
## [4,] 1621.04 1708.1 2470.4
## [5,] 1618.16 1723.1 2484.7
## [6,] 1610.61 1714.3 2466.8
```

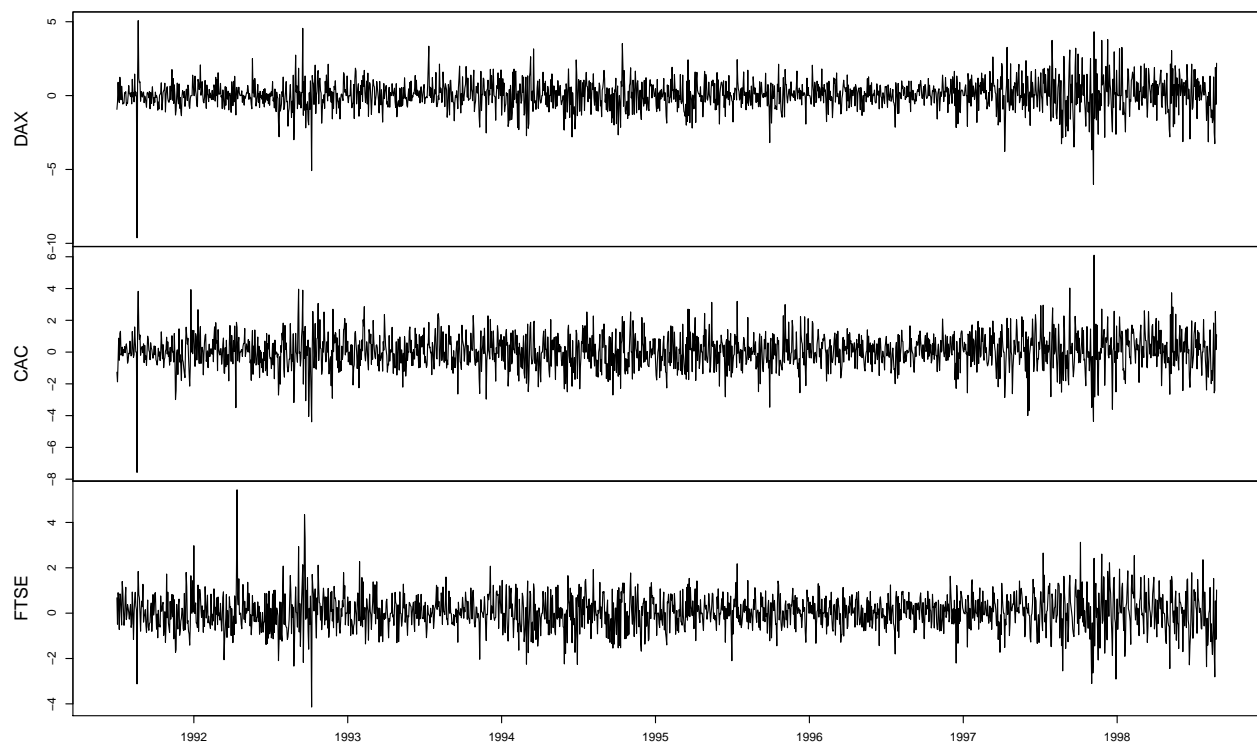
Graphiques Times series

```
EuroStock = as.zoo(EuStockMarkets[, c("DAX", "CAC", "FTSE")])
plot(EuroStock, xlab = "", main = "")
```



Pourcentage des rendements

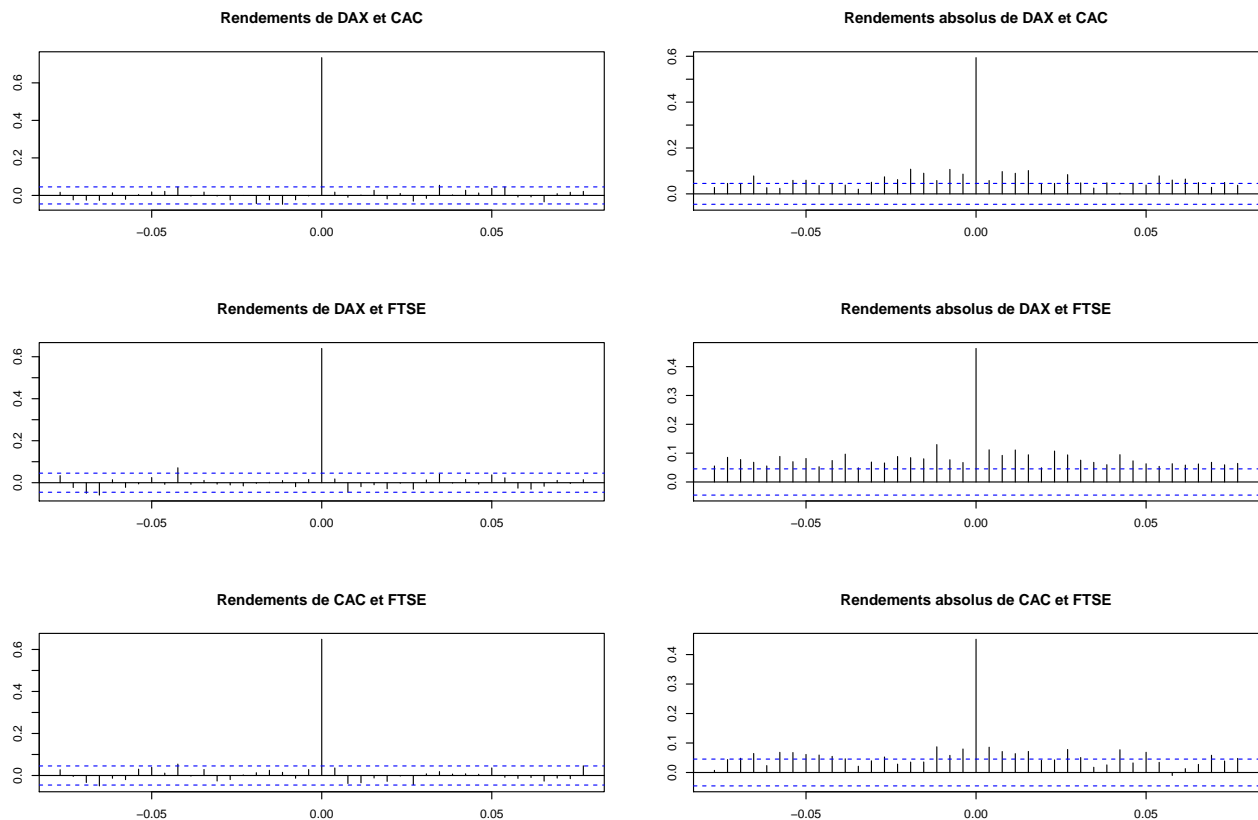
```
EuroRend = diff(log(EuroStock))*100
plot(EuroRend, xlab = "", main = "")
```



Corrélations croisées

```
layout(matrix(1:6, nrow = 3, ncol = 2, byrow = TRUE))

ccf(EuroRend[,1], EuroRend[,2], ylab = "", xlab = "", lag.max = 20,
    main = "Rendements de DAX et CAC")
ccf(abs(EuroRend[,1], abs(EuroRend[,2], ylab = "", xlab = "",
    lag.max = 20, main = "Rendements absolus de DAX et CAC")
ccf(EuroRend[,1], EuroRend[,3], ylab = "", xlab = "", lag.max = 20,
    main = "Rendements de DAX et FTSE")
ccf(abs(EuroRend[,1], abs(EuroRend[,3], ylab = "", xlab = "",
    lag.max = 20, main = "Rendements absolus de DAX et FTSE")
ccf(EuroRend[,2], EuroRend[,3], ylab = "", xlab = "", lag.max = 20,
    main = "Rendements de CAC et FTSE")
ccf(abs(EuroRend[,2], abs(EuroRend[,3], ylab = "", xlab = "",
    lag.max = 20, main = "Rendements absolus de CAC et FTSE")
```



Comme indiqué précédemment, les rendements eux-mêmes sont à peine corrélés de manière croisée entre les différents marchés (France, Allemagne et Royaume-Uni) et diminuent assez rapidement (graphiques à gauche). Cependant, des corrélations croisées significatives sont évidentes pour leur contrepartie absolue. Ces *facts* sont les plus prononcés pour les Allemands vis-à-vis du marché boursier britannique.

Corrélations roulantes

```
Coroul = function(x){
  dimension = ncol(x)
```

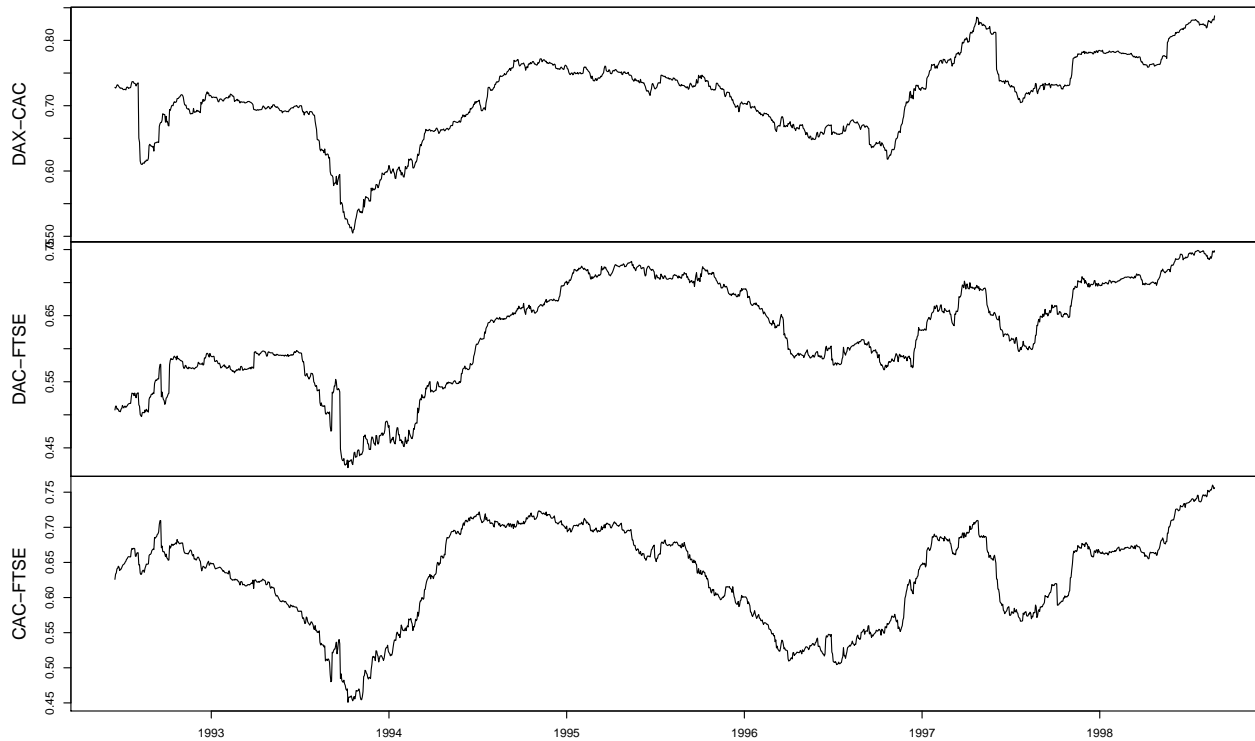
```

rccorrel = cor(x)[lower.tri(diag(dimension), diag = FALSE)]
return(rccorrel)
}

rccorrel = rollapply(EuroRend, width = 250, Coroul, align = "right",
                    by.column = FALSE)

colnames(rccorrel) = c("DAX-CAC", "DAC-FTSE", "CAC-FTSE")
plot(rccorrel, main = "", xlab = "")

```



Les trois graphiques des corrélations croisées sont assez similaires pour toutes les paires représentées, ainsi que leurs plages. Pour le DAX/CAC, la corrélation se situe entre 0.55 et 0.84, pour le DAX/FTSE, les valeurs sont comprises entre 0.44 et 0.75, et enfin pour le CAC/FTSE, la corrélation se situe dans l'intervalle compris entre 0.47 et 0.78.

3. Optimisation de portefeuille robuste

Nous présentons des méthodes d'optimisation de portefeuille dans lesquelles une sorte de mesure du risque et son niveau sont directement intégrés dans la pondération des instruments financiers.

Portefeuilles à risque optimal

Dans cette section, un outil graphique, à savoir les diagrammes de surface de risque, permet de contrôler toutes les répartitions d'actifs réalisables en ce qui concerne la diversification c'est-à-dire le risque implicite. Les fonctions permettant de créer ce type de tracé sont contenues dans le package *fPortfolio*.

Conclusion

Merci!