16/12/2019

Analyse multifactorielle du fonds Fidelity European Growth Fund

Cette étude est une analyse multifactorielle de type Fama-French du fonds luxembourgeois Fidelity.

« This paper argues that many of the CAPM average-return anomalies are related, and they are captured by the three [or five] – factor model in Fama and French (1993). » - Multifactor Explanations of Asset Pricing Anomalies

– Fama & French

Wiem Ghazouani
Louis Grunenwald
Mamadou Missiliou Haidara
Anna-Salima Rakass

Table des matières

Introduction

Notre étude porte sur l'analyse de performance de *Fidelity Funds - European Growth Fund*, qui s'agit des fonds principalement investis en actions de sociétés cotées européennes. Ces fonds construisent notre variable d'intérêt pour laquelle nous recherchons des variables qui peuvent contribuer à son explication. Par ailleurs, nous étudions les relations éventuelles existantes entre ces variables explicatives et leurs effets sur le comportement du *European Growth Fund*.

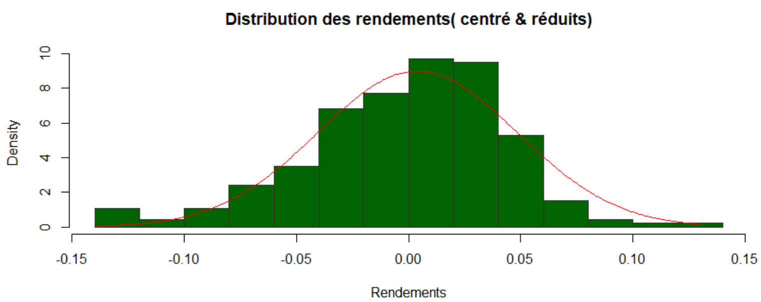
Pour ce fait, nous mettons en place une description brève des *Fidelity Funds - European Growth Fund A-DIST-EUR,* puis nous étudions les divers volets des performances.
Nous abordons à la suite la partie pratique de collecte de variables et leur étude, et nous finissons avec la partie modélisation.

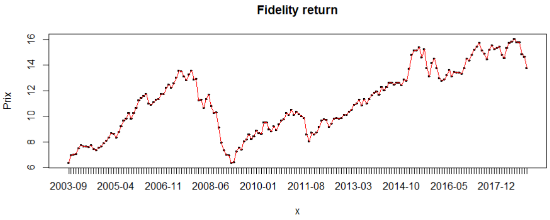
I) **<u>Présentation générale</u>**

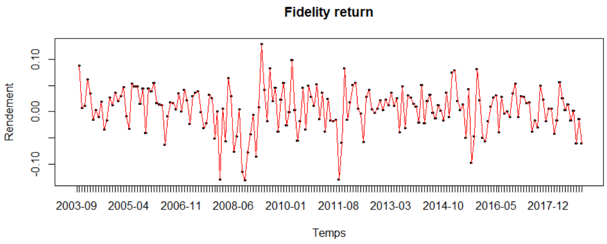
Fidelity Funds - European Growth Fund A-DIST-EUR est une SICAV lancée le 01/10/1990 par la société Fidelity (FIL Inv Mgmt (Lux) S.A.). Son code ISIN est LU0048578792. Les gérants sont *Matthew Siddle* depuis le 01/07/2012 et *Hélène Powell* depuis le 01/07/2019.

Sur 5 ans, la SICAV a reçu une note de 3 étoiles sur 5 par Morningstar.
La SICAV Fidelity Funds - European Growth Fund A-DIST-EUR a pour indice de référence 100 % FTSE World Europe TR EUR. Elle est classée dans la catégorie Sicavonline "*Actions Europe Grandes Caps. Mixte*" et dans la catégorie AMF "*Fonds étrangers*". Ces fonds sont toutefois libellés en EUR.

Au 03/12/2019, la SICAV avait une capitalisation de 6 435,31 M. EUR. Elle n'est pas éligible au PEA. Elle n'est pas non plus éligible au PEA PME. La SICAV peut être souscrite au sein du Compte titres proposé par Sicavonline.







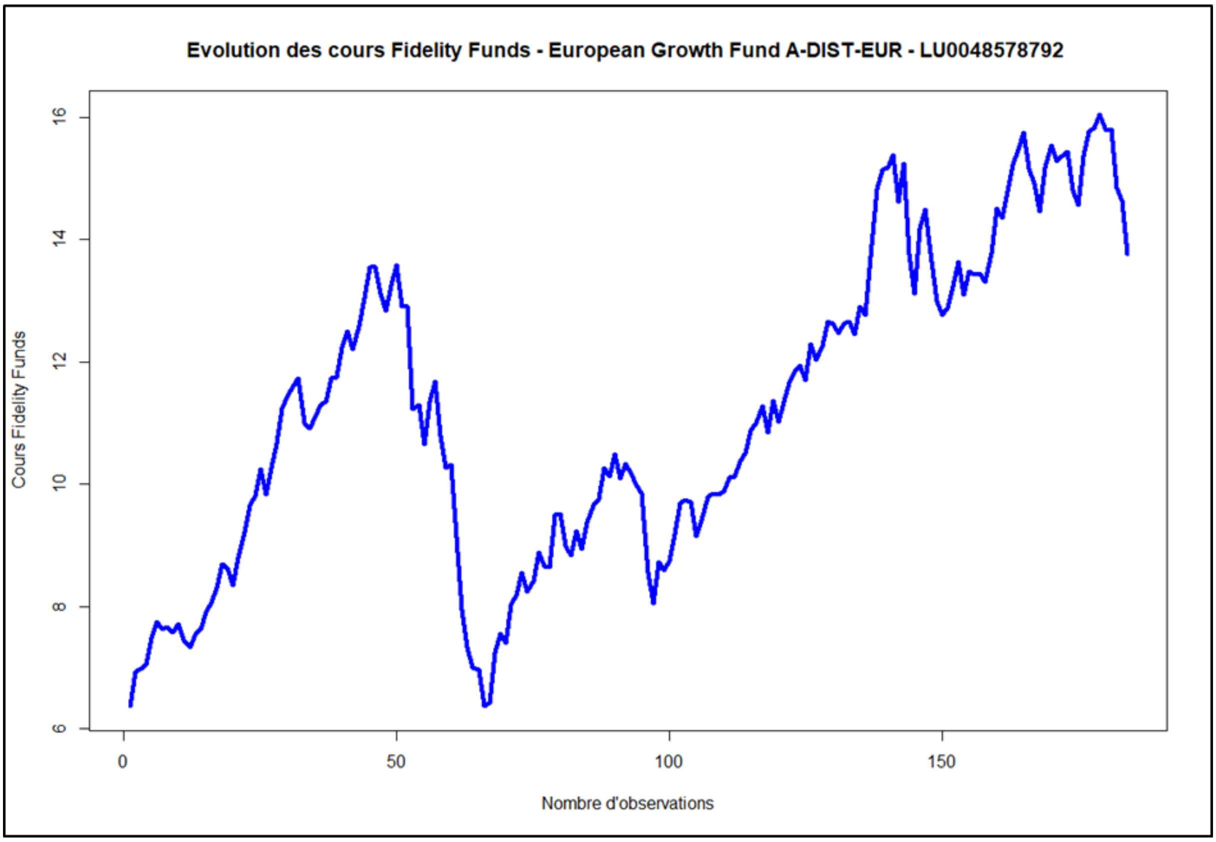

```
> summary(table$Prix)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  5.116   8.078  10.370  10.759  13.175  16.270
```

```
> summary(table$Variation)
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-0.139581 -0.017811  0.009404  0.003940  0.034284  0.129208
```

II) Analyse de performance

Les données qu'on se dispose sur Fidelity Funds - European Growth Fund A-DIST-EUR s'agissent des cours historiques mensuels, allons de 28/09/2003 jusqu'au 31/12/2018
En termes de nombre d'observations nous disposons de 184 observations.

A) Lecture de graphique



Visuellement, la variable d'intérêt suit un mouvement haussier tout au long de la période étudiée, mais pour certaines périodes, Il a eu des chutes importantes, surtout à partir de la 50ième observation qui coïncide avec la date 31/10/2007, ceci peut s'expliquer par diverses tensions économiques et conjoncturelles, surtout que cette période a témoigné la crise des Subprime, qui a affecté l'économie mondiale et dont les conséquences ont duré pour longtemps.

B) Statistiques descriptives

Tableau 1 : Résumé statistique de Fidelity Funds - European Growth Fund

Mesures	Valeurs
Minimum	6.361000
Maximum	16.040000
1. Quartile	9.099250
3. Quartile	13.235000
Mean	11.190397
Median	11.060000

Variance	6.800147
Stdev	2.607709
Skewness	0.097294
Kurtosis	-1.085722

De 28/09/2003 jusqu'au 31/12/2018 le cours du fond vaut au moyenne 11.2€. Il atteint un prix maximal de 16.04€ et un prix minimal de 6.361€. Le cours risqué est d'ordre 6.8€.

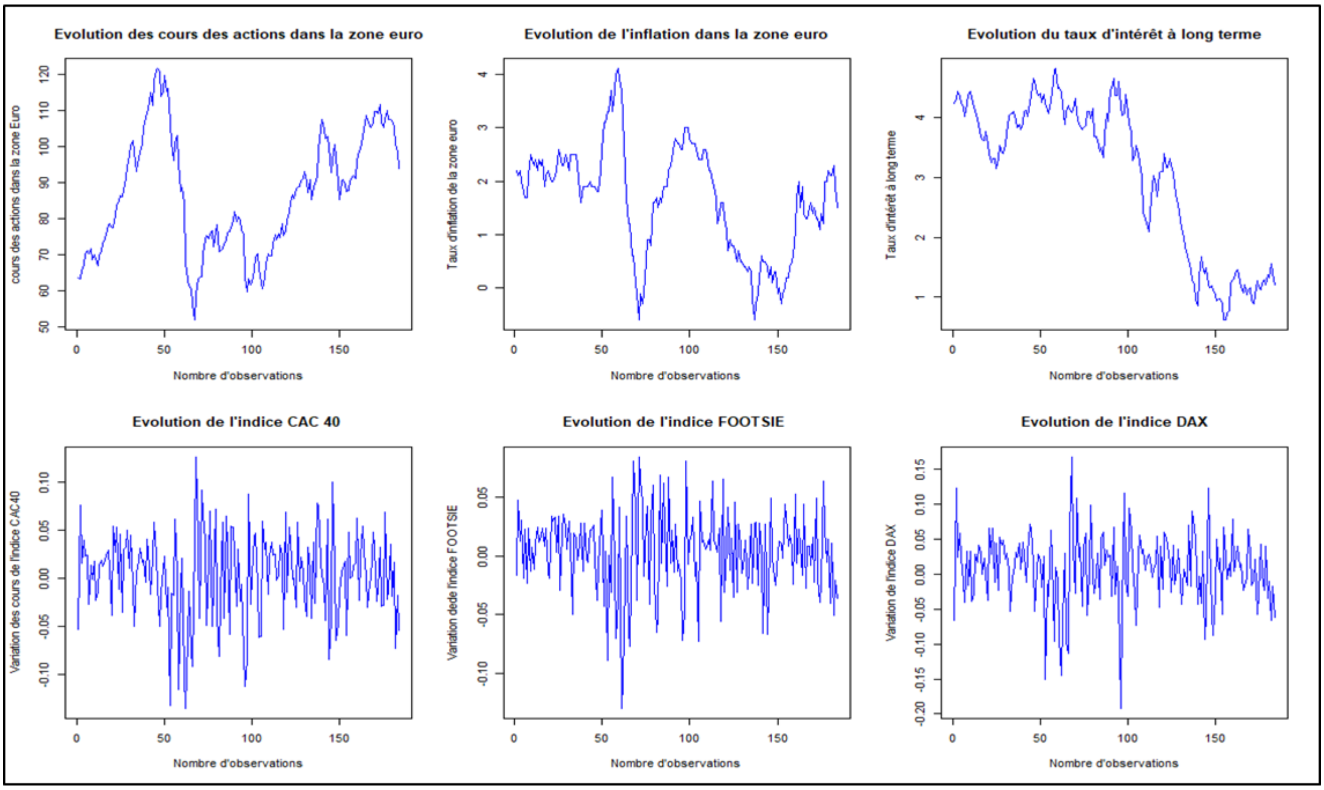
C) Variables explicatives

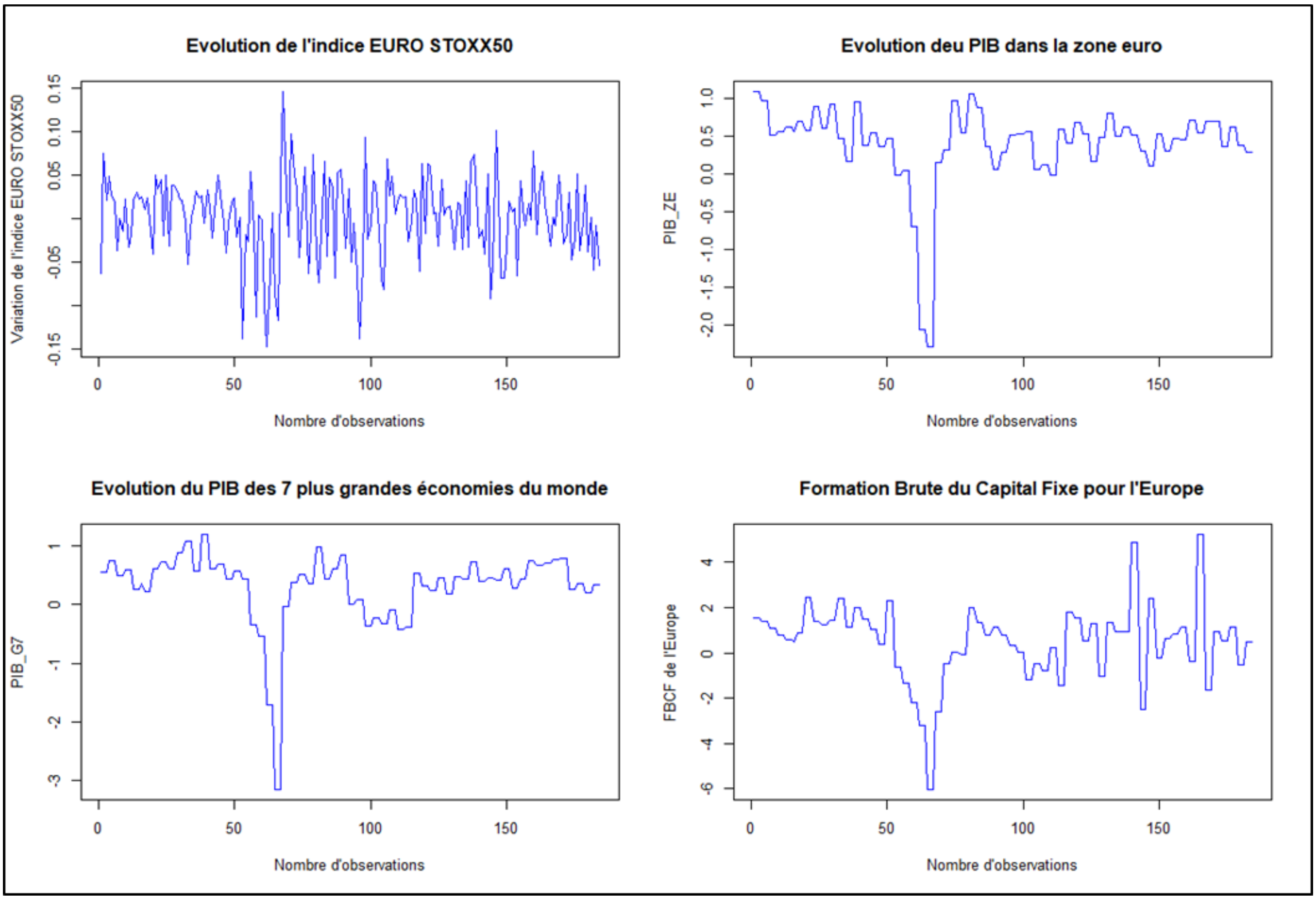
Afin d'expliquer le comportement évolutif du fond en question nous avons collecté plusieurs variables à partir du site de l'OCDE. Nous présentons l'ensemble des variables récoltées dans le dictionnaire suivant:

Tableau 2 : Dictionnaire des variables

Libellés	Explication
Cours_actions_ZE	Cours des actions dans la zone euro
Inflation_ZE	Inflation de la zone euro
TX_LT	Taux d'intérêt à long terme
Var_CAC40	Variation de l'indice boursier CAC 40 de la bourse de Paris
Var_FTSE100	Variation de l'indice boursier footsie des cent entreprises britanniques
Var_DAX	Variation de l'indice boursier allemand
Var_EUST50	Variation de l'indice boursier EURO STOXX 50 au niveau de la zone euro
TX_change_ZE	Taux de change dans la zone euro
PIB_ZE	PIB dans la zone euro
PIB_G7	PIB des 7 plus grandes économies du monde
FBCF_EU	Formation Brute du Capital Fixe pour l'Europe

D) Visualisation des variables explicatives





E) <u>Statistiques des variables explicatives</u>

> basicStats(base)

	Cours_action_ZE	Inflation_ZE	TX_LT	Var_CAC40	Var_FTSE100	Var_DAX
nobs	184.000000	184.000000	184.000000	184.000000	184.000000	184.000000
NAs	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Minimum	51.828740	-0.600000	0.613300	-0.135200	-0.130200	-0.191900
Maximum	121.662100	4.100000	4.814500	0.125600	0.084500	0.167600
1. Quartile	72.881450	0.800000	1.517925	-0.023400	-0.019000	-0.022850
3. Quartile	100.611075	2.400000	4.103950	0.032025	0.026900	0.038625
Mean	86.701529	1.660870	3.043943	0.002968	0.003276	0.007329
Median	87.029060	1.900000	3.508650	0.008400	0.008000	0.013200
Sum	15953.081390	305.600000	560.085600	0.546200	0.602700	1.348500
SE Mean	1.225935	0.074602	0.094740	0.003330	0.002666	0.003726
LCL Mean	84.282744	1.513680	2.857019	-0.003602	-0.001985	-0.000022
UCL Mean	89.120314	1.808059	3.230868	0.009539	0.008536	0.014679
Variance	276.536759	1.024034	1.651540	0.002040	0.001308	0.002554
Stdev	16.629394	1.011946	1.285123	0.045170	0.036164	0.050535
Skewness	0.150146	-0.191406	-0.542233	-0.439775	-0.514277	-0.489065
Kurtosis	-1.022081	-0.528194	-1.284727	0.337587	0.719302	1.601056

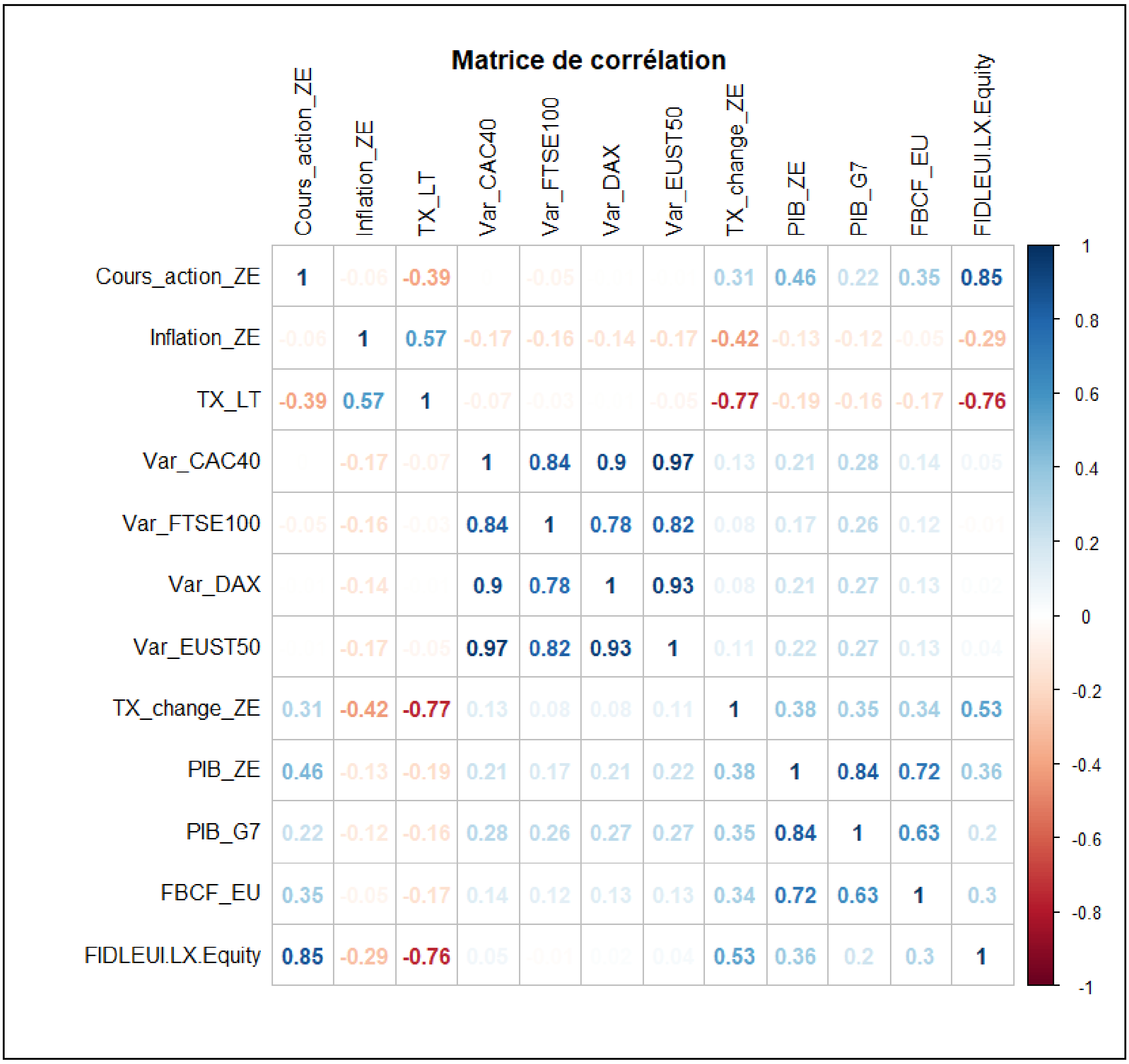
	Var_EUST50	TX_change_ZE	PIB_ZE	PIB_G7	FBCF_EU	FIDLEUI.LX.Equity
nobs	184.000000	184.000000	184.000000	184.000000	184.000000	184.000000
NAs	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Minimum	-0.146900	0.682675	-3.157149	-2.279426	-6.041325	6.361000
Maximum	0.146900	0.904035	1.193740	1.092960	5.213289	16.040000
1. Quartile	-0.026500	0.730638	0.216647	0.302741	-0.394237	9.099250
3. Quartile	0.031700	0.847186	0.618300	0.617882	1.358247	13.235000
Mean	0.001990	0.791412	0.310048	0.392333	0.494737	11.190397
Median	0.007150	0.778294	0.444112	0.507213	0.798211	11.060000
Sum	0.366100	145.619852	57.048749	72.189192	91.031696	2059.033000
SE Mean	0.003469	0.005021	0.047236	0.041371	0.126686	0.192243
LCL Mean	-0.004854	0.781505	0.216850	0.310706	0.244784	10.811099
UCL Mean	0.008834	0.801319	0.403245	0.473959	0.744691	11.569694
Variance	0.002214	0.004639	0.410547	0.314934	2.953081	6.800147
Stdev	0.047053	0.068110	0.640740	0.561190	1.718453	2.607709
Skewness	-0.461294	0.351493	-3.071163	-3.094245	-0.671571	0.097294
Kurtosis	0.669896	-1.052830	12.912804	11.723583	3.047298	-1.085722

>

F) Etude de la corrélation

L'une des règles primordiales dans le domaine de la Data Science est d'étudier les relations de dépendance entre les variables déployées.

Deux variables liées portent généralement la même information au niveau de l'explication d'un phénomène précis. Corriger ce problème nous mène à éviter la redondance des informations.



Nous considérons une forte corrélation à partir d'un seuil de 60%. Il ressort de la matrice de corrélation ci-dessus que:

- La variable Var_Cac40 est fortement corrélé avec Var_FTSE100 (r=0.84), avec Var_DAX (r=0.9) et avec Var_EUST50 (r=0.97).
- La variable Var_FTSE100 est fortement corrélé avec Var_DAX (r=0.78) et avec Var_EUST50 (r=0.93)
- La variable PIB_ZE est fortement corrélé avec PIB_G7 (r=0.84) et avec FBCF_EU (r=0.72).
- La variable PIB_G7 est fortement corrélée FBCF_EU (r=0.63)
- En terme de valeur absolu, la variable TX_change_ZE est fortement corrélé avec TX_LT (r=0.77).

A priori, nous possédons une idée sur les variables corrélées, nous tendons à choisir certaines parmi elles pour la partie modélisation et non pas le déploie de toutes les variables.
Par ailleurs

- La variable d'intérêt est fortement et positivement corrélée avec la variable Cours_action_ZE (r=0.85). Aussi, elle est fortement et négativement corrélée avec la variable TX_LT (r=-0.76).

III) Analyses théories classiques

La Généralisation du CAPM peut être basée sur les modèles multifactoriels (Fama-French : 3 Facteurs et 5 Facteurs). Ce sont aussi des modèles linéaires, mais ils ne posent aucunes hypothèses sur l'aversion au risque des investisseurs.
Le premier modèle est introduit par Ross (1976) est basé une évaluation de l'actif par arbitrage. On l'appelle l'APT (Arbitrage Pricing Theory).
Ce modèle ne suppose pas la normalité des rendements, mais seulement que les investisseurs sont averses au risque, sans spécifier de fonction d'utilité particulière.
Quand il n'existe qu'un seul facteur correspondant au rendement de marché, ce modèle est le CAPM.
Le problème est donc de bien identifier les facteurs. De nombreuses études empiriques ont étudiés la détermination de ces facteurs macroéconomiques ou financiers.
Par exemple, le modèle à 3 facteurs de Fama-French prend en compte notamment le « book-to-market ratio » (valeur comptable de l'action d'une société à sa valeur de marché), et la taille de l'entreprise mesurée par sa capitalisation boursière.
Fama & French suppose donc que le marché est efficient mais que plus d'un facteur est nécessaire pour expliquer le rendement d'un actif, et que ces facteurs sont les mêmes pour tous les investisseurs.
Le modèle d'évaluation par arbitrage (modèle MEA ou *APT*) est un modèle financier d'évaluation des actifs d'un portefeuille qui s'appuie sur l'observation des anomalies du MEDAF et considère les variables propres aux firmes susceptibles d'améliorer davantage le pouvoir prédictif du modèle d'évaluation. Pour lutter contre l'instabilité du β (qui indique le risque spécifique par type de facteur), le modèle MEA introduit des facteurs macroéconomiques et spécifiques. Cependant selon le principe d'Absence d'Opportunité d'Arbitrage (AOA) : les portefeuilles ou les actifs présentant les mêmes risques doivent s'échanger au même prix. Ce modèle n'intègre aucun facteur relatif aux préférences des investisseurs. La Méthode d'Evaluation par Arbitrage (MEA ou *APT*) peut se décrire ainsi :

$$E(R) = Rf + \beta 1(R1 - Rf) + \beta 2(R2 - Rf) + \ldots + \beta n(Rn - Rf),$$

Le MEDAF ou CAPM (Capital Asset Pricing Model), suppose qu'il existe un actif sans risque avec un rendement Rf. Dans ce modèle chaque portefeuille efficient est une combinaison de l'actif sans risque et du portefeuille de marché M, qui correspond au point de tangence entre les deux frontières efficientes (avec et sans l'actif sans risque).
Ce modèle a été aussi amélioré par le modèle Fama-French à trois facteurs. Fama est souvent perçu comme le père de l'hypothèse d'efficience des marchés, débutée avec sa thèse de doctorat dans un article de mai 1970 du "*Journal of Finance*", nommé *Efficient Capital Markets : A Review of Theory and Empirical Work*. Fama propose deux concepts fondamentaux qui définissent les marchés comme efficients. Le premier est le niveau d'efficience des marchés : (les modèles utilisés peuvent changer en fonction du niveau d'efficience considéré) :

- L'efficience peut être forte ;

(« *Stock Prices reflect all information (public and private) about a firm* »);

- L'efficience peut être semi-forte

(« *Stock Prices reflect all publicly available information about a firm.* »);

- L'efficience peut être faible

(« *Stock Prices reflect all past market price and volume information* »).

Le deuxième est de démontrer l'incompatibilité d'un équilibre des marchés et de son efficience. Eugène Fama souhaite donc améliorer le modèle d'évaluation des actifs MEDAF, et propose de rajouter au modèle d'évaluation par arbitrage (modèle MEA ou *APT*) des facteurs complémentaires.
Nous essayerons donc ici dans un premier temps d'évaluer la qualité de prévision des rendements grâce aux modèles financiers dits classiques. D'une part le modèle du MEDAF et d'autres part les modèles multifactoriels de Fama & French (à 3 et 5 facteurs)

<u>Le ratio de Sharp</u>

Ce ratio est basé sur la Capital Market Line (CML). Pour chaque portefeuille efficient on retrouve cette égalité :

$$\frac{\overline{R}_P - R_f}{\sigma(R_P)} = \frac{\overline{R}_M - R_f}{\sigma(R_M)}.$$

Ce ratio est donc la pente de la CML. Comme proposé par Sharp ce ratio peut être considéré comme une mesure de performance on peut prendre le numérateur comme la prime de risque (positive ou négative) et le dénominateur comme un indicateur de risque. Autrement dit, c'est le rapport entre l'excès de rendement moyen du portefeuille et la mesure totale du risque du portefeuille.

Pour exemplifier, le gérant d'un fonds peut regarder si son excès de rendement moyen est suffisant pour compenser un risque plus élevé que celui du portefeuille de marché. Si un portefeuille est bien diversifié, son ratio de Sharp est proche de celui du portefeuille de marché.

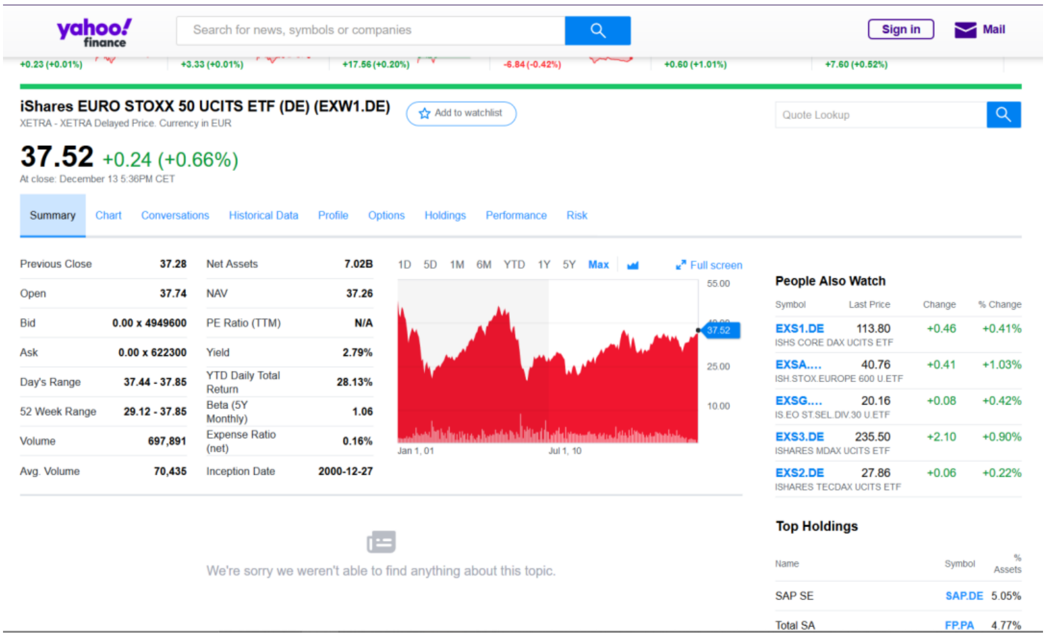
Plus le ratio de Sharp est élevé mieux c'est.

A) <u>Le MEDAF</u>

<u>Régression selon le MEDAF</u>

$$E(R_x) = R_f + \beta(R_M - R_f)$$

Pour ce faire, nous utilisons les rendements du fonds Fidelity en tant que rendement du portefeuille ainsi que celui du marché, à travers un fonds ETF répliquant bien l'indice de marché (r_m) de référence du fonds (en Europe).



Nous calculons dans un premier temps le bêta de ce modèle par la formule suivante :

$$\beta_{portfolio} = cov(R_p, R_m)/\sigma_m$$

Un fois ce bêta calculé, nous l'appliquons par la suite dans l'équation afin de déterminer R_x. Nous obtenons dès lors, une estimation du rendement du portefeuille.

Afin de connaître la qualité de l'estimation nous avons calculé le R^2 comme suit :

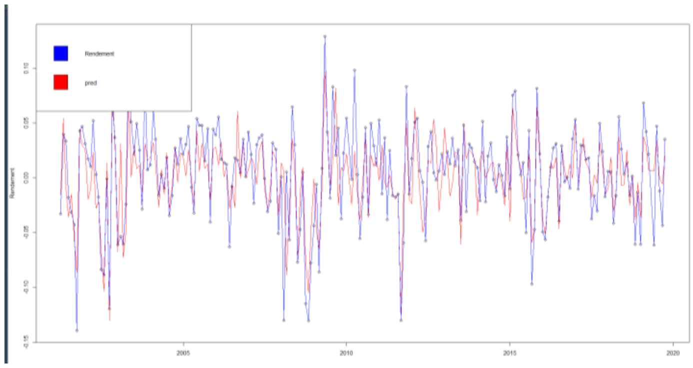
$$R^2 = \frac{SCReg}{SCT}$$

Nous obtenons alors un R^2 de 62,6% ce qui n'est pas mauvais mais pas excellent.
Nous en avons aussi profité pour calculer le ratio de Sharp par la formule qui suit :

$$S_P = \frac{R_P - R_F}{\sigma_P}$$

A la suite de ce calcul, nous avons déterminé que le Sharp ratio se situait autour de 0 à 0.065 précisément. Nous pouvons donc émettre un jugement de valeur en disant le fond n'est pas très efficace.

```r
a$pred <- a$Rf + a$beta * a$`Rm-Rf`
```



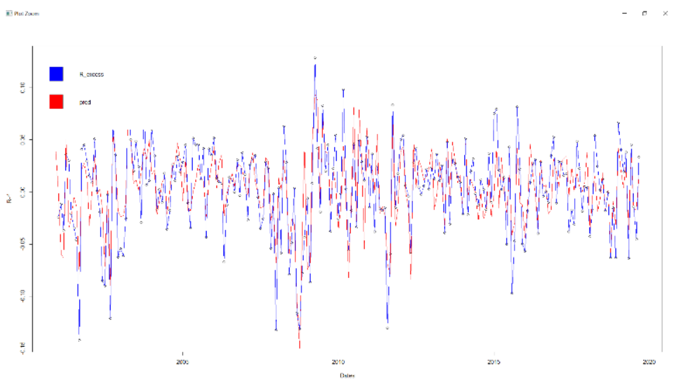
B) Fama & French

$$r_{Pt} - r_{ft} = \alpha_P + \beta_P\left(R_{Mt} - R_{ft}\right) + b_S SMB_t + b_H HML_t + \varepsilon_{Pt}$$

Ce modèle prend donc en compte les petites valeurs (*Small Caps*) (SMB), les actions ayant un *book-to-market* ratio (cours/valeur comptable d'une action, HML) élevé et la différence entre le portefeuille de marché et le taux sans risque.

Modèle Fama & French à 3 facteurs

```
Call:
lm(formula = R_excess ~ `Mkt-RF` + SMB + HML, data = FFF)

Residuals:
      Min        1Q    Median        3Q       Max
-0.088390 -0.017723 -0.000132  0.016295  0.073082

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.259e-05  1.847e-03   0.039    0.969
`Mkt-RF`     6.807e-01  3.627e-02  18.768   <2e-16 ***
SMB         -2.529e-02  9.559e-02  -0.265    0.792
HML         -1.113e-01  8.225e-02  -1.353    0.177
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02733 on 221 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.6276,	Adjusted R-squared:  0.6226
F-statistic: 124.2 on 3 and 221 DF,  p-value: < 2.2e-16
```



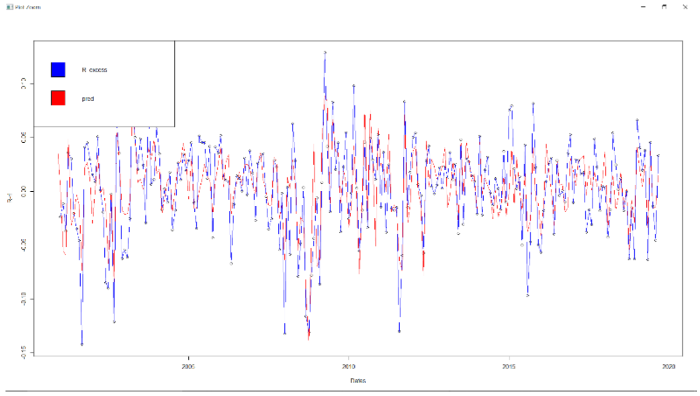
$$R_{it} - R_{Ft} = a_i + b_i(R_{Mt} - R_{Ft}) + s_i SMB_t + h_i HML_t + r_i RMW_t + c_i CMA_t + e_{it}$$

Pour le modèle à 5 facteurs, il suffit simplement de rajouter 2 nouveaux termes :

- RMW (=Robust Minus Weak) : explique la différence de rendement entre les entreprises à rentabilité robuste et faible
- CMA (=Conservative Minus Agressive) : représente la différence de rendement entre les entreprises ayant des politiques d'investissement faibles et élevées

Modèle Fama & French à 5 facteurs

```
Call:
lm(formula = R_excess ~ `Mkt-RF` + SMB + HML + RMW + CMA, data = FFF)

Residuals:
      Min        1Q    Median        3Q       Max
-0.080196 -0.018213 -0.000801  0.016318  0.073161

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.0007383  0.0019887   0.371    0.711
`Mkt-RF`     0.6439174  0.0438266  14.692   <2e-16 ***
SMB         -0.0624098  0.0981636  -0.636    0.526
HML         -0.0099926  0.1226407  -0.081    0.935
RMW         -0.0225132  0.1451158  -0.155    0.877
CMA         -0.2334534  0.1518956  -1.537    0.126
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0273 on 219 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.6319,    Adjusted R-squared:  0.6235
F-statistic: 75.18 on 5 and 219 DF,  p-value: < 2.2e-16
```



	Three factor Model	Five factor Model
AIC	-975,4382	-974,0206
BIC	-958,3577	-950,1079

IV) Régressions Econométriques

A) Régression simple

Pour voir la qualité de nos variables et pour avoir un bon référentiel, nous avons réalisé une régression simple sans traitement sur les variables sélectionnées. Les résultats étant assez bons (R^2=0.95), cela nous a conforté dans le choix d'utiliser ces variables.

```
Call:
lm(formula = FIDLEUI.LX.Equity ~ ., data = table)

Residuals:
     Min       1Q   Median       3Q      Max
-1.16418 -0.42227 -0.04238  0.40358  1.37550

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     10.376818   1.166677   8.894 7.88e-16 ***
Cours_action_ZE  0.101753   0.003699  27.507  < 2e-16 ***
Inflation_ZE     0.140343   0.057629   2.435   0.0159 *
TX_LT           -1.300540   0.068980 -18.854  < 2e-16 ***
Var_CAC40        0.025675   0.048700   0.527   0.5987
Var_FTSE100     -0.023495   0.023442  -1.002   0.3176
Var_DAX          0.017065   0.025782   0.662   0.5089
Var_EUST50      -0.007063   0.053083  -0.133   0.8943
TX_change_ZE    -5.437055   1.160539  -4.685 5.67e-06 ***
PIB_ZE          -0.140636   0.171661  -0.819   0.4138
PIB_G7           0.085921   0.164912   0.521   0.6030
FBCF_EU          0.038320   0.038660   0.991   0.3230
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6151 on 172 degrees of freedom
Multiple R-squared:  0.9477,    Adjusted R-squared:  0.9444
F-statistic: 283.4 on 11 and 172 DF,  p-value: < 2.2e-16
```

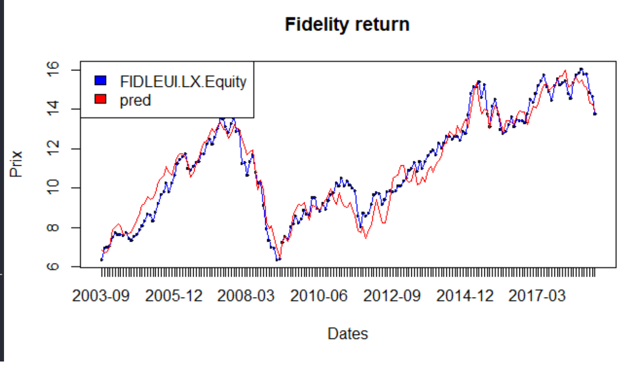
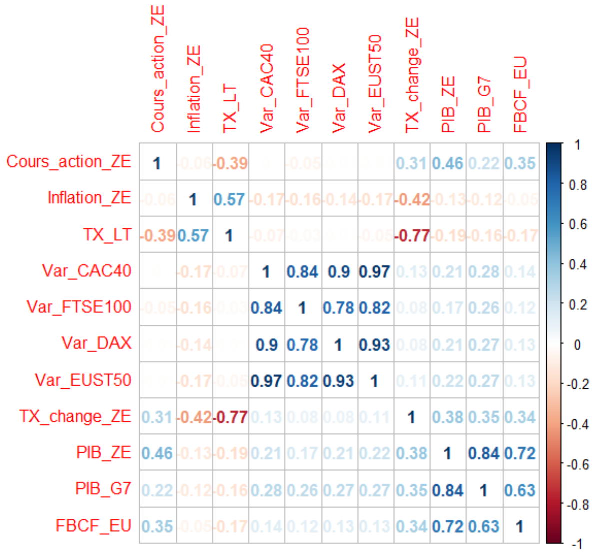


table2['pred']= 10.376818 + 0.101753 * table$Cours_action_ZE + 0.140343 * table$Inflation_ZE + (-1.300540) * table$TX_LT + (-5.437055) * table$TX_change_ZE

B) Régression après traitement



Matrice de corrélation :

Nous considérons une forte corrélation à partir d'un seuil de 60%. Il ressort de la matrice de corrélation ci-dessus que :

- La variable Var_Cac40 est fortement corrélée avec Var_FTSE100 (r=0.84), avec Var_DAX (r=0.9) et avec Var_EUST50 (r=0.97).
- La variable Var_FTSE100 est fortement corrélée avec Var_DAX (r=0.78) et avec Var_EUST50 (r=0.93)
- La variable PIB_ZE est fortement corrélée avec PIB_G7 (r=0.84) et avec FBCF_EU (r=0.72).
- La variable PIB_G7 est fortement corrélée FBCF_EU (r=0.63)
- En termes de valeur absolu, la variable TX_change_ZE est fortement corrélée avec TX_LT (r=0.77).

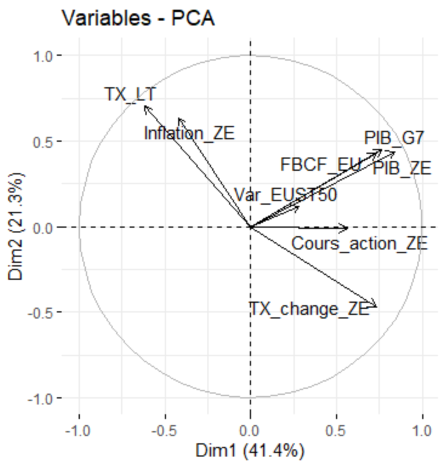
A priori, nous possédons une idée sur les variables corrélées, nous tendons à choisir certaines parmi elles pour la partie modélisation et non pas le déploie de toutes les variables.
Par ailleurs

- La variable d'intérêt est fortement et positivement corrélée avec la variable Cours_action_ZE (r=0.85). Aussi, elle est fortement et négativement corrélée avec la variable TX_LT (r=-0.76).

1) Première approche : suppression des variables corrélées

Dans cette première approche, nous décidons de supprimer quelques variables trop fortement corrélées (Var_Cac40, Var_FTSE100, Var_DAX)

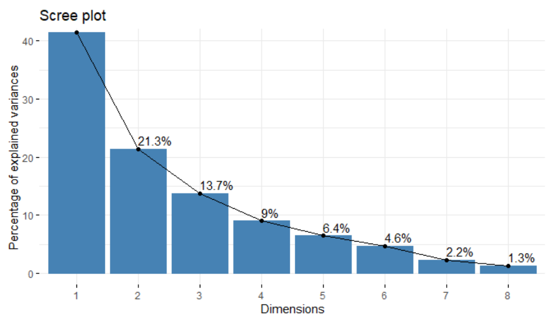
Dans le cadre d'optimisation, nous réalisons une Analyse des Composantes Principales (ACP) afin de déterminer les composantes principales de nos données à travers leurs valeurs propres :



	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Cours_action_ZE	0.5622844	-0.01186713	-0.53756058	0.58367384	-0.2071293
Inflation_ZE	-0.4221028	0.63393871	-0.34098478	0.18857714	0.5080786
TX_LT	-0.6200773	0.70276157	0.10645983	-0.04166271	-0.1402261
Var_EUST50	0.2842217	0.11900937	0.79831652	0.50031750	0.1153751
TX_change_ZE	0.7308124	-0.47086835	-0.04459298	-0.10823411	0.4108637

Ci-dessus, nous pouvons voir les valeurs propres de chacune des variables sur les dimensions choisies dans le cercle de corrélation. Nous voyons que certaines, semblent colinéaire dans ce cercle, mais ne le sont pas parfaitement car des variables colinéaires ont une corrélation proche de 1. Elles peuvent aussi être colinéaires sur ces dimensions et non les autres. Une erreur fréquente est de confondre *multi-colinéarité* et *corrélation*. Si des variables colinéaires sont *de facto* fortement corrélées entre elles, deux variables corrélées ne sont pas forcément colinéaires. En termes non statistiques, il y a colinéarité lorsque deux ou plusieurs variables mesurent la même chose. C'est donc pour cela qu'on aborde le cercle de corrélation avec beaucoup de recul.

Ci- dessous, nous pouvons observer un histogramme de la variance expliquée par chaque dimension en fonction de sa valeur propre :



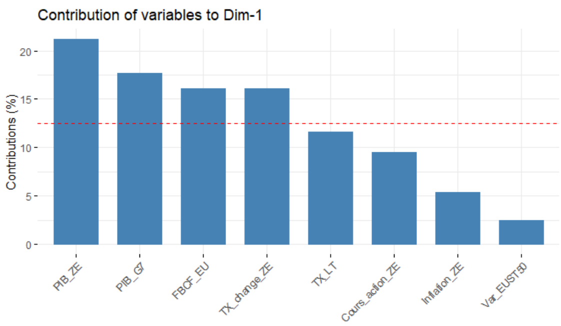
	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	3.3137884	41.422355	41.42235
Dim.2	1.7053475	21.316844	62.73920
Dim.3	1.0954503	13.693129	76.43233
Dim.4	0.7187968	8.984960	85.41729
Dim.5	0.5156065	6.445081	91.86237
Dim.6	0.3714545	4.643181	96.50555
Dim.7	0.1770953	2.213691	98.71924
Dim.8	0.1024607	1.280759	100.00000

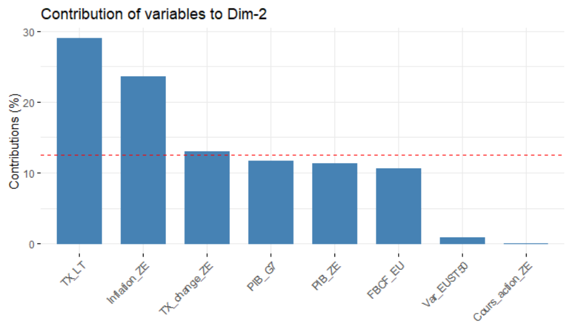
Figure : Inertie totale expliquée par chaque composante principale

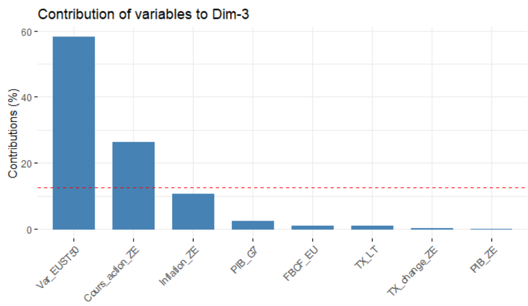
Le nombre d'axes est déterminé par le point, au-delà duquel les valeurs propres restantes sont toutes relativement petites et de tailles comparables. Ce graphique indique l'inertie expliquée par nombre de dimension choisie. En effet, on voit ici que la première dimension apporte plus de 40% d'informations du modèle. Dans ce cas-là, il nous sera nécessaire de sélectionner les variables caractérisant cette première dimension. Nous décidons de ne pas perdre trop d'information et sélectionnons donc les 3 premières dimensions qui expliqueront autour de 76% de variabilité du modèle (cf : tableau de droite avec les variances cumulées). En effet, il existe bien une inflexion entre la troisième et la quatrième valeur propre, même si finalement nous n'aurons pas une belle explication de l'inertie (près de 25 % restera inexpliquée).

Maintenant que les dimensions ont été choisies, il nous faut déterminer les variables caractérisant ces dimensions

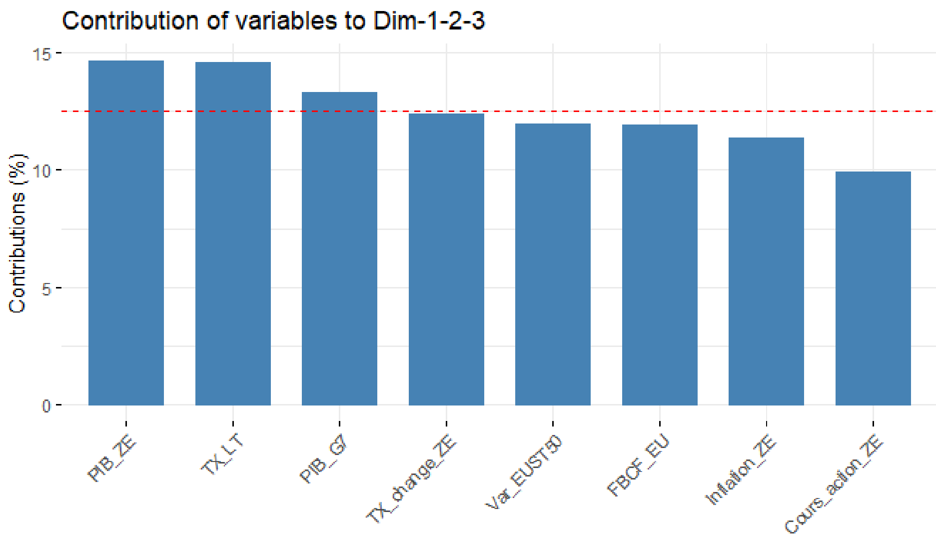
Ci-dessous l'importance de chaque variable pour les dimensions choisies :







On voit bien que les dimensions ne font pas ressortir les mêmes variables dans leurs compositions.

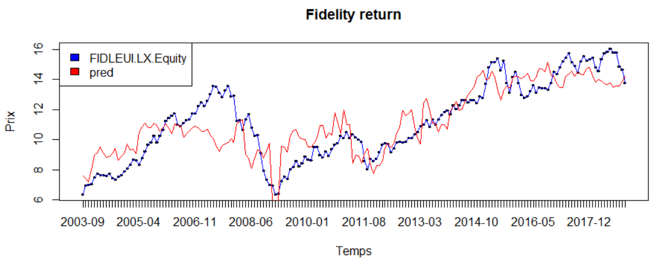


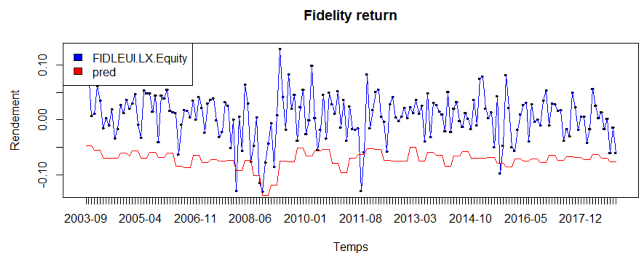
Nous décidons de sélectionner les variables majeurs de ces 3 dimensions, à savoir le PIB_ZE, le TX_LT, le PIB_G7 et le TX_change_ZE. Avec ces 4 variables nous allons réaliser notre régression.

<u>Résultat de la régression</u> :

```
Call:
lm(formula = FIDLEUI.LX.Equity ~ ., data = table)

Residuals:
    Min      1Q  Median      3Q     Max
-2.7589 -1.1713 -0.1525  0.9550  4.0291

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   24.7304     2.4713  10.007  < 2e-16 ***
TX_LT         -1.8387     0.1360 -13.520  < 2e-16 ***
TX_change_ZE -10.1995     2.7311  -3.735 0.000253 ***
PIB_ZE         2.2822     0.3221   7.085 3.08e-11 ***
PIB_G7        -1.4752     0.3632  -4.062 7.27e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.485 on 179 degrees of freedom
Multiple R-squared:  0.6827,    Adjusted R-squared:  0.6756
F-statistic: 96.28 on 4 and 179 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = Variation ~ ., data = table)

Residuals:
     Min       1Q   Median       3Q      Max
-0.14273 -0.02253  0.00213  0.02554  0.13002

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.082432   0.067386  -1.223   0.2228
TX_LT         0.003611   0.003697   0.977   0.3300
TX_change_ZE  0.083390   0.074345   1.122   0.2635
PIB_ZE       -0.019473   0.008579  -2.270   0.0244 *
PIB_G7        0.042810   0.009660   4.432 1.63e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03927 on 178 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.1487,    Adjusted R-squared:  0.1296
F-statistic: 7.772 on 4 and 178 DF,  p-value: 8.539e-06
```



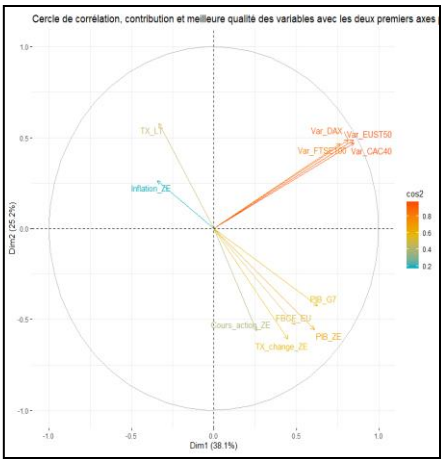


A gauche nous avons notre régression effectuée sur les prix du fond Fidelity et à droite les celle sur les rendements.

2) <u>2ème approche : non suppression des variables corrélées</u>

Il est toujours litigieux de supprimer des variables non parfaitement corrélées car on risque toujours la perte d'information. En effet si cette variable n'est pas colinéaire, nous n'avons aucune raison de la supprimer car elle contient une info propre à elle.

<u>Réalisation de l'ACP</u> :



	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Cours_action_ZE	0.2588638	-0.5631194	0.06366975	0.736768539	-0.238177217
Inflation_ZE	-0.3435153	0.2608874	0.65464775	0.347644143	0.503175504
TX_LT	-0.3332497	0.5782036	0.66029710	-0.098639924	-0.141265319
Var_CAC40	0.8504850	0.4701966	-0.07097303	0.079073058	0.030489869
Var_FTSE100	0.7659483	0.4679324	-0.05368413	0.008991446	0.009205230
Var_DAX	0.8145561	0.4852725	-0.01916288	0.083481631	0.007133484
Var_EUST50	0.8460357	0.4859464	-0.05336806	0.082337099	0.010941772

L'objectif de l'Analyse en Composantes Principales (ACP) est de revenir à un espace de dimension réduite (par exemple 2) en déformant le moins possible la réalité. Il s'agit donc d'obtenir le résumé le plus pertinent possible des données initiales.

L'ACP est une méthode factorielle qui consiste à projeter le nuage de points sur un sous-espace, en perdant le moins d'information possible. Elle est principalement utilisée pour 2 raisons :

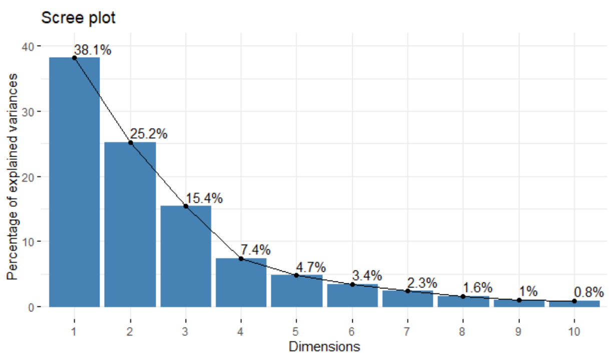
- D'une part évaluer l'importance des variables dans l'explication de la variabilité du modèle
- Et d'autre part, pouvoir diminuer la dimension de la matrice de base à l'aide d'axes majeurs calculés par l'ACP. Cette réduction dimensionnelle permet, dans le cas de données importantes, d'optimiser le temps de calcul.

Dans l'analyse en composantes principales, les variables sont souvent normalisées. Ceci est particulièrement recommandé lorsque les variables sont mesurées dans différentes unités, sinon, le résultat de l'ACP obtenue sera fortement affecté.

Techniquement, l'approche consiste à transformer les données en soustrayant à chaque valeur la moyenne de la variable et en la divisant par l'écart type. A l'issue de cette transformation les données obtenues sont dites données centrées-réduites. L'ACP appliquée à ces données transformées est appelée ACP normée. Par défaut, sur le logiciel r, la fonction **PCA**() [dans le package *FactoMineR*], normalise les données automatiquement. On peut nous même le faire par l'ajout de l'argument scale.unit qui prend la valeur TRUE. Soit le format simplifié : PCA(X, scale.unit = TRUE, graph = FALSE).

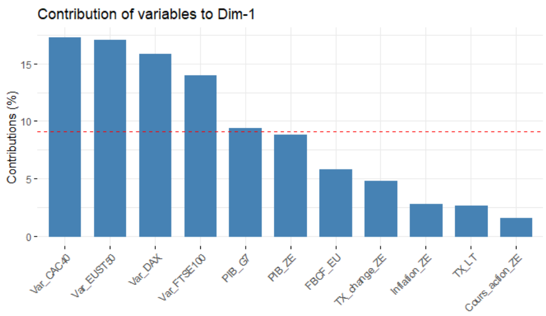
La lecture graphique montre que :

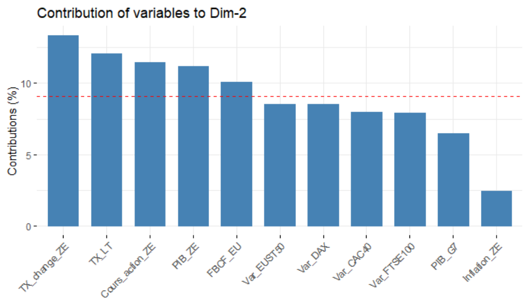
- Var_CAC40, Var_DAX, Var_EUST50 et Var_FTSE100 sont fortement et positivement corrélées avec l'axe 1. Ces variables sont les mieux présentées et les plus contributives à la construction du premier axe.
- PIB_ZE, PIB_G7, FBCF_EU et Cours_action_ZE sont faiblement corrélés avec l'axe 1.
- Inflation_ZE et TX_LT sont faiblement et négativement corrélés avec l'axe 1

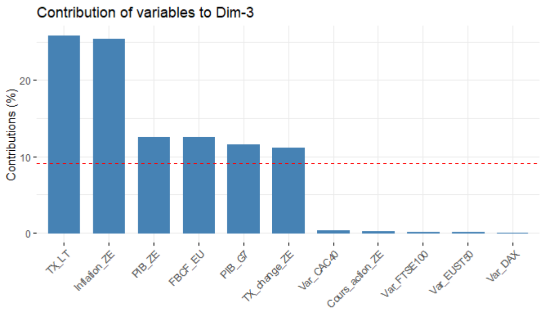


	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	4.19365804	38.1241640	38.12416
Dim.2	2.76701092	25.1546447	63.27881
Dim.3	1.68975928	15.3614480	78.64026
Dim.4	0.80874967	7.3522697	85.99253
Dim.5	0.52105112	4.7368284	90.72935
Dim.6	0.37329881	3.3936255	94.12298
Dim.7	0.25470851	2.3155319	96.43851
Dim.8	0.17470566	1.5882333	98.02675
Dim.9	0.10498894	0.9544449	98.98119
Dim.10	0.09232855	0.8393504	99.82054
Dim.11	0.01974050	0.1794591	100.00000

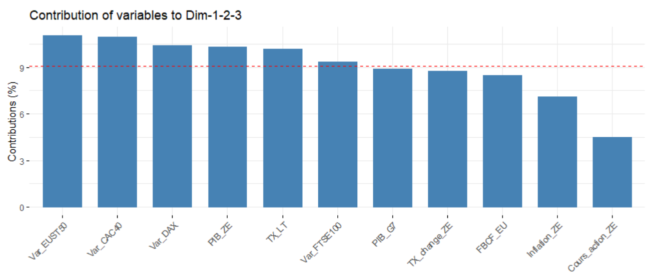
Nous voyons, suite à l'ACP réalisée ci-dessus, que la variabilité est un peu mieux expliquée par 2 dimensions. Nous choisissons encore une fois de conserver les 3 dimensions qui nous permettent de conserver encore une fois une bonne explication de la variabilité de notre modèle.







Nous observons sur les graphiques ci-dessus les variables expliquant le mieux chacune des dimensions choisies. C es dimensions étant les plus importantes dans l'explication de la variabilité du modèle, nous pouvons par conséquent affirmer que les variables qui en ressortent sont les plus importantes dans la détermination de nos résultats.



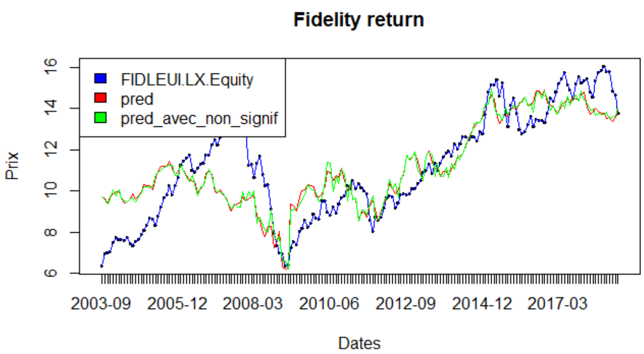
Enfin, suite à ce graph agrégé, nous décidons de conserver les 6 variables expliquant le mieux la variabilité de nos données. Nous effectuerons donc la régression sur ces variables.

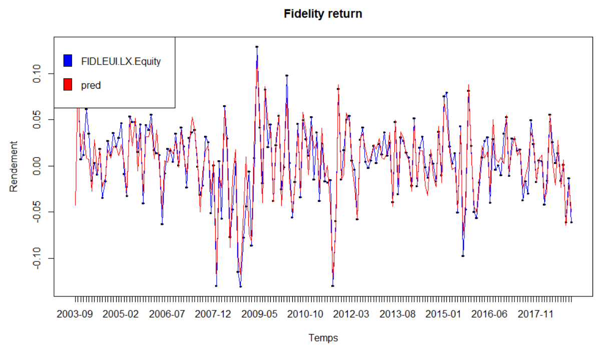
Elément qui est assez intéressant à remarquer, nous avons des variables d'importance ici, que nous avions précédemment retiré dans la première approche car trop corrélée. Cependant nous nous rendons compte que celles-ci ramènent néanmoins beaucoup d'informations. Nous préfèreront donc, à terme, conserver ces données, plutôt que de les supprimer arbitrairement.

Résultat régression :

```
Call:
lm(formula = FIDLEUI.LX.Equity ~ ., data = table)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.3334 -1.2878 -0.1054  0.9848  4.3899 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 15.30715    0.33003  46.381  < 2e-16 ***
TX_LT       -1.45026    0.09554 -15.180  < 2e-16 ***
Var_CAC40    0.02522    0.12678   0.199    0.843    
Var_FTSE100 -0.08042    0.06079  -1.323    0.188    
Var_DAX      0.03316    0.06734   0.492    0.623    
Var_EUST50  -0.03276    0.13787  -0.238    0.812    
PIB_ZE       0.96379    0.19413   4.965 1.61e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.613 on 177 degrees of freedom
Multiple R-squared:  0.6298,	Adjusted R-squared:  0.6172 
F-statistic: 50.18 on 6 and 177 DF,  p-value: < 2.2e-16

>
```

```
Call:
lm(formula = Variation ~ ., data = table)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.051976 -0.008039 -0.001134  0.007962  0.070577 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -7.801e-04  3.220e-03  -0.242 0.808822    
TX_LT        3.792e-05  9.337e-04   0.041 0.967649    
Var_CAC40    6.167e-03  1.235e-03   4.992 1.43e-06 ***
Var_FTSE100  2.331e-03  5.936e-04   3.928 0.000123 ***
Var_DAX      4.624e-03  6.567e-04   7.041 4.12e-11 ***
Var_EUST50  -3.986e-03  1.344e-03  -2.966 0.003434 ** 
PIB_ZE       1.058e-03  1.896e-03   0.558 0.577563    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01572 on 176 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.8652,	Adjusted R-squared:  0.8606 
F-statistic: 188.2 on 6 and 176 DF,  p-value: < 2.2e-16
```





Nous obtenons ici des résultats supérieurs à notre première approche, ce modèle, plus juste, est préférable à la régression sans les variables corrélées. Il nous permet ici potentiellement d'estimer assez bien des rendements futurs. Cependant nous restons encore une fois avec des résultats plus faibles que la régression avec toutes les variables car nous avons utilisé une ACP qui par son approche nous fait nécessairement perdre de l'information. Ainsi suite à cette ACP nous avons des résultats assez équivalents aux modèles classiques du MEDAF ou de Fama & French.

Cependant même si les résultats sont relativement satisfaisant, nous nous interrogeons tout de même sur la pertinence du modèle linéaire. Nous décidons donc de tester la linéarité du modèle dans la partie suivante.

C) Test d'évaluation de la significativité du lien linéaire entre les variables explicatives et la variable d'intérêt

L'objectif de ces tests est de s'assurer que les variables explicatives sont linéairement indépendantes c'est-à-dire que la matrice de variance covariance est de rang complet (singulière) par conséquent que l'estimateur des MCO est sans biais.

1) <u>Modèle linéaire</u>

Le test d'évaluation de la significativité du lien linéaire entre les variables explicatives est la variable d'intérêt est valide, si les résidus sont :

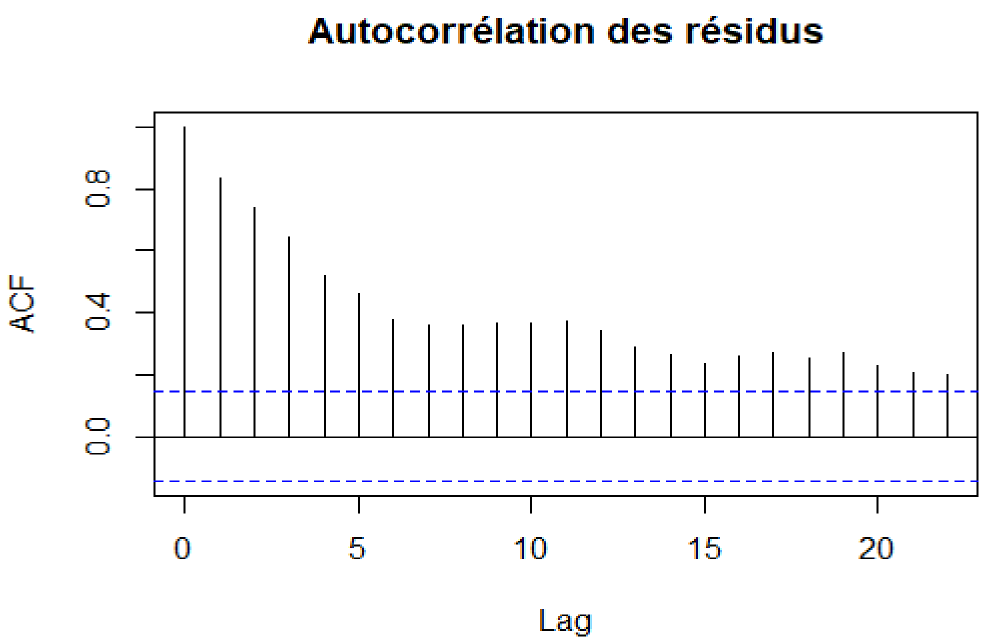
- Indépendants
- Distribués selon une loi normale de moyenne 0
- Distribués de façon homogène, c'est-à-dire avec une variance constante

2) <u>Evaluation de l'hypothèse d'indépendance des résidus</u>

Il y'a autocorrélation des résidus lorsque le résidu d'un point quelconque est lié à celui du point suivant dans le tableau de donnée

Les pointillées horizontaux, sont les intervalles de confiance du coefficient de corrélation égal é 0. Ici, le plot nous montre qu'une autocorrélation significative est présente pour tous les lags.

.



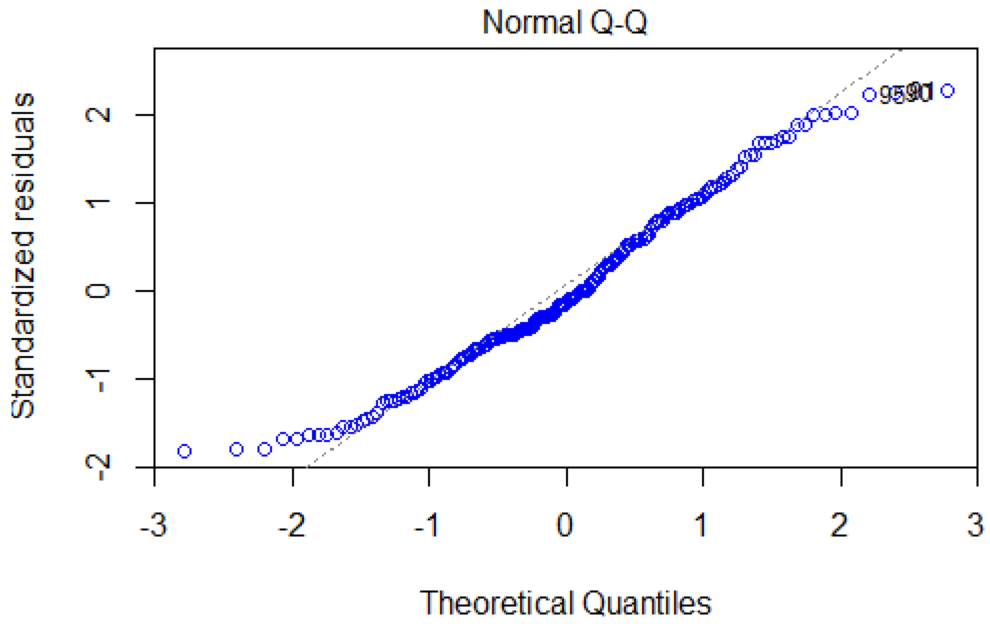
Pour avoir une confirmation de nos résultats, nous effectuer le test de Durbin Watson. Ce test est employé pour évaluer la présence d'une autocorrélation pour un lag de valeur 1.

L'hypothèse d'indépendance des résidus est rejetée lorsque la p-value du test est inférieure à 0.05.

```
lag Autocorrelation D-W Statistic p-value
  1       0.8364767     0.3129584       0
Alternative hypothesis: rho != 0
```

Les résultats ci-dessus montrent que le test de Durbin Watson rejette l'hypothèse d'indépendance des résidus car p-value est inférieur à 0.05

3) Analyse de l'hypothèse de normalité des résidus



Cette hypothèse peut s'évaluer graphiquement à l'aide d'un QQplot. (Ci-dessus)

Si les résidus sont bien distribués le long de la droite figurant sur le plot, alors l'hypothèse de normalité est acceptée. A l'inverse, s'ils s'en écartent, alors l'hypothèse de normalité est rejetée. Dans notre cas les résidus semblent être bien distribués selon la loi normale. Néanmoins, nous allons effectuer le test de Shapiro pour avoir la confirmation.

Le test de Shapiro-Wilk

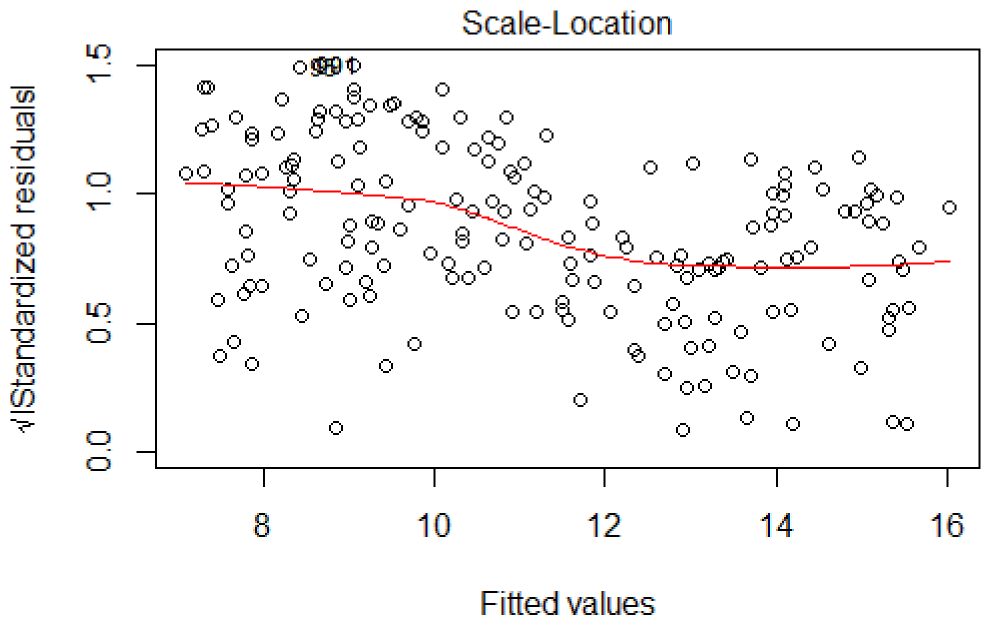
Le test de Shapiro-Wilk peut également être employé pour évaluer la normalité des résidus.

```
        Shapiro-Wilk normality test

data:  residus
W = 0.97754, p-value = 0.004607
```

Ici nous rejetons l'hypothèse de normalité car le p-value est inférieur à 0.05.

4) Evaluation de l'hypothèse d'homogénéité des résidus

L'hypothèse d'homogénéité peut se vérifier de façon visuelle, pour cela il faut réaliser un "residuals vs fitted plot". Les "fitted" correspondent aux réponses prédites par le modèle, pour les valeurs observées de la variable prédictive. Plus précisément, on utilise pour ce plot la racine carrée des résidus standardisés.



On observe la courbe rouge, qui est aussi une courbe de régression locale, si elle est plate on conclut que les résidus sont répartis de façon homogène tout le long du gradient des valeurs prédites. Et donc que l'hypothèse d'homogénéité des résidus est acceptée.

Mais dans le graphique ci-dessus, on voit que la courbe est descendante, alors cela signifie que la dispersion des résidus n'est pas homogène. Dans ce cas, l'hypothèse d'homogénéité des résidus est rejetée.

D) Modèle Log-Niveau

Les différents tests réalisés précédemment n'ayant pas été concluant, nous allons proposer un autre modèle pour expliquer notre fond. Pour cela nous avons introduit de la non-linéarité en transformant les prix du fond en valeurs logarithmique puis, nous avons refait la régression, les résultats obtenus sont les suivants :

```
Call:
lm(formula = log_fond ~ Var_FTSE100 + Cours_action_ZE + TX_LT +
    Var_CAC40 + Var_DAX + Var_EUST50 + PIB_ZE + PIB_G7, data = base)

Residuals:
      Min        1Q    Median        3Q       Max
-0.196483 -0.044286 -0.001192  0.041802  0.148280

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      1.7955461  0.0432713  41.495   <2e-16 ***
Var_FTSE100     -0.1808604  0.2735150  -0.661    0.509
Cours_action_ZE  0.0098545  0.0004129  23.869   <2e-16 ***
TX_LT           -0.0867392  0.0045868 -18.910   <2e-16 ***
Var_CAC40        0.0785874  0.5671239   0.139    0.890
Var_DAX          0.1901829  0.3014394   0.631    0.529
Var_EUST50      -0.0290468  0.6183636  -0.047    0.963
PIB_ZE          -0.0112751  0.0180968  -0.623    0.534
PIB_G7           0.0091112  0.0191658   0.475    0.635
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.072 on 175 degrees of freedom
Multiple R-squared:  0.9143,    Adjusted R-squared:  0.9104
F-statistic: 233.5 on 8 and 175 DF,  p-value: < 2.2e-16
```

Dans ce modèle les résultats très satisfaisants :

Le R-carré (R-Squared) (R-carré) est égal à 93.68% ce qui veut dire que le modèle explique seulement **91.43%** des variations du fond. C'est un indicateur important de la qualité du modèle.

Le **p-value : < 2.2e-16** Cette valeur est le résultat du test de Fisher (test de significativité globale du modèle). Et puisqu'elle est inférieure à 0.05 alors le modèle est globalement significatif.

Deux variables sont significatives, il s'agit respectivement du cours des actins dans la zone euro et du taux d'intérêt de long terme.

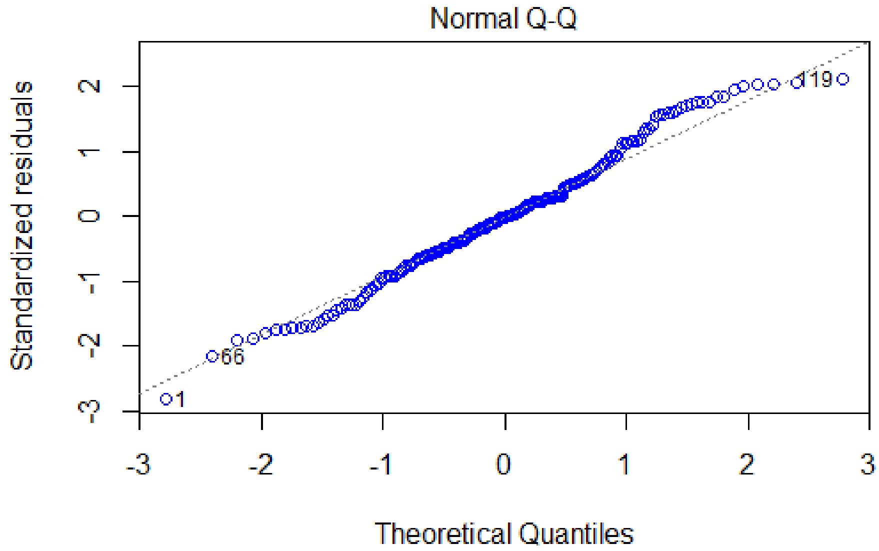
Interprétation des coefficients :

- **0.0098545**= cela signifie que lorsque le cours des actions de la zone euro augmente d'une unité cela entraine, toutes choses étant égale par ailleurs, une augmentation de la valeur du fond de **0.98%**. On observe que la probabilité marginale de rejet (pr(>|t|) de cette variable est inférieure à 0.05 alors le test de significativité partielle de Student est vérifié, la variable a bien un effet sur la valeur du fond.

- **-0.0867392**= C'est le coefficient estimé de la variable tu taux d'intérêt. On s'aperçoit qu'il y'a une relation inverse entre le taux d'intérêt et la valeur du fond. Une augmentation de la valeur du taux d'intérêt d'une unité entraine une diminution de la valeur du fond de **8.67%**, toutes choses étant égales par ailleurs. Pour cette variable aussi la probabilité marginale de rejet (pr(>|t|) est inférieure à 0.05 donc le test est significatif. Le taux d'intérêt a une influence significative sur l'appréciation du fond.

Test d’évaluation de la normalité des résidus

Pour ce modèle aussi, nous avons réalisé un test de normalité des résidus pour voir si ces derniers sont distribués selon une loi normale Contrairement au modèle précédent on peut observer sur le QQplot. (Ci-dessous) que la distribution des résidus semble s’aligné parfaitement sur la droite. On accepte l’hypothèse de normalité



Le test de Shapiro réalisé sur ce modèle donne les résultats suivants :

```
	Shapiro-Wilk normality test

data:  residus
W = 0.98669, p-value = 0.07995
```

Les résultats viennent confirmer l’intuition observée sur le graphique, nous acceptons l'hypothèse de normalité car la p-value est supérieur à 0.05.

Pour aller loin dans la comparaison des modèles, nous allons utiliser l’**AIC** (Critère d'information d'Akaike) et le **BIC** (Critère d'information bayésien), les résultats sont les suivant pour les suivants :

	Linéaire	Log-Niveau
AIC	377.651	-435.2985
BIC	409.8004	-403.1492

On peut observer que le modèle log Niveau est meilleur que le modèle linéaire de départ au regard des valeurs des critères de l’AIC et du BIC.