L'analyse des motifs (patterns) dynamiques : rappels

- La normalité d'un nombre au sens de Borel implique l'uniformité de la distribution des motifs
- Les fréquences conditionnelles sont toutes égales à $\frac{1}{2}$
- On peut également montrer que les hausses et les baisses des cours ne sont pas corrélées.
- On note I_t, la variable qui prend la valeur 1 en cas de hausse à la date t et 0 sinon
  - *I_t variable indicatrice de hausse à la date t*
  - *Il y aura hausse à la date* t *et hausse à la date* $t-k$, *où* $k > 0$ *est un retard (lag) si* $I_t \times I_{t-k} = 1$
  - ***On peut vérifier*** *(exercice à faire)* ***que le coefficient de corrélation linéaire entre*** $\mathbf{I_t}$ ***et*** $\mathbf{I_{t-k}}$ ***est nul pour tout*** $\mathbf{k > 0}$

L'analyse des motifs (patterns) dynamiques : rappels

- Hors, la nullité des autocorrélations est aussi une propriété du jeu de pile ou face, donc des suites aléatoires !
- Tests « portemanteau » (Box et Pierce (1970), Ljung et Box (1978))
  - *Utilisés pour savoir si l'on est en présence d'un « bruit blanc fort »*
- La statistique de Ljung et Box (1978) est calculée comme $Q = n(n+2)\sum_{k=1}^{h}\frac{\hat{\rho}_k^2}{n-k}$, où $\hat{\rho}_k$ est l'autocorrélation pour un décalage de k
- La suite C_2 est déterministe, ses autocorrélations sont nulles et pourtant elle vérifie de le test de Ljung et Box.

L'analyse des motifs (patterns) dynamiques : rappels

- La normalité d'un nombre au sens de Borel implique l'uniformité de la distribution des motifs
- Les fréquences conditionnelles sont toutes égales à $\frac{1}{2}$
- Hausses et baisses ne sont pas corrélées
- Borel (1909) a montré que la probabilité qu'un nombre réel dans [0,1] soit normal est égale à 1 !
- Il existe potentiellement beaucoup de suites déterministes qui se comportent comme des suites aléatoires

Emile Borel

Exercice: L'analyse des motifs (patterns) dynamiques

- Comme précédemment, on note I_t l'indicatrice de hausse à la date t.
  - $I_t = 1$ *si hausse des prix (H) entre* $t-1$ *et* t
  - $I_t = 0$ *si baisse des prix (B) entre* $t-1$ *et* t
- On note $\pi_{11} = P(I_t = 1 | I_{t-1} = 1)$ la probabilité qu'une hausse soit suivie d'une hausse
    - On suppose que cette probabilité ne dépend de t.
  - *D'où* $\pi_{10} = P(I_t = 0 | I_{t-1} = 1) = 1 - \pi_{11}$
  - π_{10} *probabilité qu'une hausse soit suivie d'une baisse*
- De même, on note $\pi_{00} = P(I_t = 0 | I_{t-1} = 0)$: probabilité qu'une baisse soit suivie d'une baisse.
  - $\pi_{01} = P(I_t = 1 | I_{t-1} = 0) = 1 - \pi_{00}$
  - π_{01} *probabilité qu'une baisse soit suivie d'une hausse*

L'analyse des motifs (patterns) dynamiques

- Exercice : utiliser la loi des probabilités totales pour calculer la probabilité non conditionnelle de hausse

L'analyse des motifs (patterns) dynamiques

- Exercice : utiliser la loi des probabilités totales pour calculer la probabilité non conditionnelle de hausse
  - *La probabilité non conditionnelle de hausse* $P(I_t = 1)$ *s'écrit comme* $P(I_t = 1 | I_{t-1} = 1) \times P(I_{t-1} = 1) + P(I_t = 1 | I_{t-1} = 0) \times P(I_{t-1} = 0)$
- En supposant les probabilités indépendantes de la date, écrire $P(I_t = 1)$ en fonction des probabilités conditionnelles

L'analyse des motifs (patterns) dynamiques

- Exercice : utiliser la loi des probabilités totales pour calculer la probabilité non conditionnelle de hausse
  - *La probabilité non conditionnelle de hausse* $P(I_t = 1)$ *s'écrit comme* $P(I_t = 1 | I_{t-1} = 1) \times P(I_{t-1} = 1) + P(I_t = 1 | I_{t-1} = 0) \times P(I_{t-1} = 0)$
- En supposant les probabilités indépendantes de la date, écrire $P(I_t = 1)$ en fonction des probabilités conditionnelles
  - $P(I_t = 1) = \frac{\pi_{01}}{\pi_{01}+\pi_{10}}$
- En déduire une condition pour que $P(I_t = 1) = \frac{1}{2}$

L'analyse des motifs (patterns) dynamiques

- Exercice : utiliser la loi des probabilités totales pour calculer la probabilité non conditionnelle de hausse
  - *La probabilité non conditionnelle de hausse* $P(I_t = 1)$ *s'écrit comme* $P(I_t = 1 | I_{t-1} = 1) \times P(I_{t-1} = 1) + P(I_t = 1 | I_{t-1} = 0) \times P(I_{t-1} = 0)$
- En supposant les probabilités indépendantes de la date, écrire $P(I_t = 1)$ en fonction des probabilités conditionnelles
  - $P(I_t = 1) = \frac{\pi_{01}}{\pi_{01}+\pi_{10}}$
- En déduire une condition pour que $P(I_t = 1) = \frac{1}{2}$
  - $\pi_{01} = \pi_{10}$
- On supposera par la suite que $P(I_t = 1 | I_{t-1}, I_{t-1}, \dots) = P(I_t = 1 | I_{t-1})$ (propriété dite de Markov)

L'analyse des motifs (patterns) dynamiques

- On se place dans les conditions précédentes.
- Calculer l'espérance de gain $E[G]$ à chaque date pour une stratégie momentum, en supposant que l'on gagne 1 euro en cas de succès et que l'on perd 1 euro en cas d'échec.

L'analyse des motifs (patterns) dynamiques

- On se place dans les conditions précédentes.
- Calculer l'espérance de gain $E[G]$ à chaque date pour une stratégie momentum, en supposant que l'on gagne 1 euro en cas de succès et que l'on perd 1 euro en cas d'échec.
  - $E[G|I_{t-1}=0] = P(I_t=0|I_{t-1}=0) - P(I_t=1|I_{t-1}=0) = 1 - 2\pi_{10}$ *où* $E[G|I_{t-1}=0]$ = *espérance de gain sachant* $I_{t-1}=0$
  - $E[G|I_{t-1}=0] = 1 - 2\pi_{10}$
  - *Comme* $\pi_{01} = \pi_{10}$, $E[G] = 1 - 2\pi_{10}$
- Ecrire une condition pour qu'une stratégie momentum (respectivement contrarian) soit profitable en moyenne

L'analyse des motifs (patterns) dynamiques

- On se place dans les conditions précédentes.
- Calculer l'espérance de gain $E[G]$ à chaque date pour une stratégie momentum, en supposant que l'on gagne 1 euro en cas de succès et que l'on perd 1 euro en cas d'échec.
  - $E[G|I_{t-1}=0] = P(I_t=0|I_{t-1}=0) - P(I_t=1|I_{t-1}=0) = 1 - 2\pi_{10}$ *où* $E[G|I_{t-1}=0]$ = *espérance de gain sachant* $I_{t-1}=0$
  - $E[G|I_{t-1}=0] = 1 - 2\pi_{10}$
  - *Comme* $\pi_{01} = \pi_{10}$, $E[G] = 1 - 2\pi_{10}$
- Ecrire une condition pour qu'une stratégie momentum (respectivement contrarian) soit profitable en moyenne
  - *Stratégie momentum profitable si* $\pi_{01} = \pi_{10} < \frac{1}{2}$
  - *Stratégie contrarian si* $\pi_{01} = \pi_{10} > \frac{1}{2}$

L'analyse des motifs (patterns) dynamiques

- On veut tester si $\pi_{01} = \pi_{10} = \frac{1}{2}$, c'est-à-dire si les motifs HH, HB, BH et BB sont équirépartis.
  - *Remarque : Avec l'hypothèse de Markov, cela implique que l'on est en présence d'une suite binaire indifférente (Ville).*
  - *On considère un échantillon de taille* n
  - n_{ij} *représente le nombre de* i *suivis d'un* j
  - *Estimateurs (du maximum de vraisemblance) des probabilités de transition*
    - $\hat{\pi}_{01} = \frac{n_{01}}{n_{01}+n_{00}}$, $\hat{\pi}_{00} = \frac{n_{00}}{n_{01}+n_{00}} = 1 - \hat{\pi}_{01}$
    - $\hat{\pi}_{10} = \frac{n_{10}}{n_{10}+n_{11}}$, $\hat{\pi}_{11} = \frac{n_{11}}{n_{10}+n_{11}} = 1 - \hat{\pi}_{10}$
  - *Est-ce que* $\hat{\pi}_{01}, \hat{\pi}_{10}$ *(et donc* $\hat{\pi}_{01}$, $\hat{\pi}_{10}$*) sont proches de* $\frac{1}{2}$ *?*

L'analyse des motifs (patterns) dynamiques

- Test d'adéquation du khi-deux de Pearson
  - *Statistique de test :* $4 \times \left[\left(\frac{n_{00}}{n} - \frac{1}{4}\right)^2 + \left(\frac{n_{01}}{n} - \frac{1}{4}\right)^2 + \left(\frac{n_{10}}{n} - \frac{1}{4}\right)^2 + \right.$ $\left. + \left(\frac{n_{11}}{n} - \frac{1}{4}\right)^2 \right]$
  - *La quantité précédente suit approximativement une loi du χ^2 à 3 degrés de liberté.*
- Christoffersen : test du ratio de maximum de vraisemblance
  - *On note $\alpha = \frac{1}{2}$*
  - *Statistique de test :* $2 \log \left[\left(\frac{\hat{\pi}_{00}}{\alpha}\right)^{n_{00}} \left(\frac{\hat{\pi}_{01}}{\alpha}\right)^{n_{01}} \left(\frac{\hat{\pi}_{10}}{\alpha}\right)^{n_{10}} \left(\frac{\hat{\pi}_{11}}{\alpha}\right)^{n_{11}} \right]$
  - *Pour n grand, la statistique suit approximativement une loi du khi-deux à 2 degrés de liberté*

L'analyse des motifs (patterns) dynamiques

- Exemple numérique avec Excel :
  - *B1=ALEA() simule une variable uniforme dans* $[0,1]$ *avec le générateur de nombres pseudo-aléatoires d'Excel*
  - *Recopier 252 fois dans la colonne B*
  - *C1= =SI(B1<0,5;1;0)*
  - *C2=SI(C1=1;SI(B2<0,55;1;0);SI(B2<0,45;1;0))*
  - *Recopier 251 fois dans la colonne C*
  - *La colonne C contient les indicatrices de hausse et de baisse*
  - *Ici,* $\pi_{01} = \pi_{10} = 0{,}55 > \frac{1}{2}$
  - $E[G] = 2\pi_{10} - 1 = 0{,}1$
  - *Sur une année :* 252 *jours de trading*
  - *Espérance de gain sur une année* $=$ 25,2 *euros*

L'analyse des motifs (patterns) dynamiques

- Le gain quotidien G peut s'écrire comme $2B - 1$ où B est une variable de Bernoulli de paramètre $p = 0{,}55$
  - *L'écart-type de B est …*
  - *L'écart-type de G est …*

L'analyse des motifs (patterns) dynamiques

- Le gain quotidien G peut s'écrire comme $2B - 1$ où B est une variable de Bernoulli de paramètre $p = 0{,}55$
  - *L'écart-type de B est* $\sqrt{p(1-p)}$
  - *L'écart-type de G est* $2\sqrt{p(1-p)}$

L'analyse des motifs (patterns) dynamiques

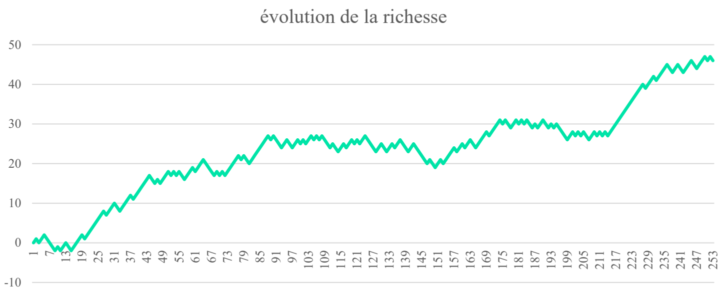
- Le gain quotidien G peut s'écrire comme $2B - 1$ où B est une variable de Bernoulli de paramètre $p = 0{,}55$
  - *L'écart-type de* B *est* $\sqrt{p(1-p)}$
  - *L'écart-type de* G *est* $2\sqrt{p(1-p)}$
- Gain annuel : somme de gains quotidiens indépendants
  - *Ecart-type du gain annuel est …*
  - *Ratio : espérance de gain / écart-type* $= \cdots$
  - *Horizon de* 2 *ans, espérance de gain* $= \cdots$
  - *Écart-type :* $2\sqrt{2 \times 252 \times p(1-p)} = \cdots$
  - *Ratio : espérance de gain / écart-type* $= \cdots$

L'analyse des motifs (patterns) dynamiques

- Le gain quotidien G peut s'écrire comme $2B - 1$ où B est une variable de Bernoulli de paramètre $p = 0{,}55$
  - *L'écart-type de* B *est* $\sqrt{p(1-p)}$
  - *L'écart-type de* G *est* $2\sqrt{p(1-p)}$
- Gain annuel : somme de gains quotidiens indépendants
  - *Ecart-type du gain annuel est* $2\sqrt{252 \times p(1-p)} = 15{,}8$ *euros*
  - *Ratio : espérance de gain / écart-type* $= \frac{25{,}2}{15{,}8} = 1{,}6 < 2$
  - *Horizon de* 2 *ans, espérance de gain* $= 2 \times 25{,}2 = 50{,}4$ *euros*
  - *Écart-type :* $2\sqrt{2 \times 252 \times p(1-p)} = 22{,}3$ *euros*
  - *Ratio : espérance de gain / écart-type* $= \frac{50{,}4}{22{,}3} = 2{,}25 > 2$
  - *Pour un horizon de deux ans, gain significatif*

L'analyse des motifs (patterns) dynamiques

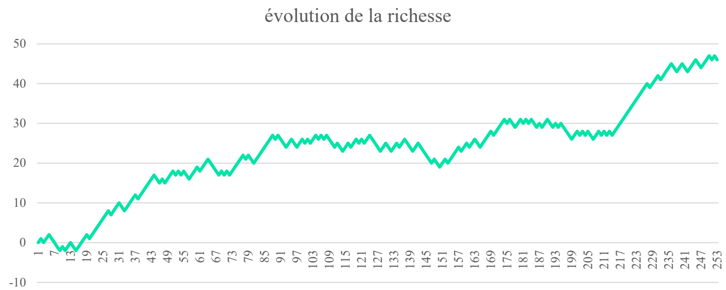
- Exemple numérique avec Excel (suite) :
  - *Une trajectoire de la richesse sur une année*



  - *Statistique de Christoffersen pour cette trajectoire* $= 1{,}19$.
  - *La valeur critique au seuil de* 5% *pour la loi du* χ^2 *à* 2 *degrés de liberté est égale à* $5{,}99$
  - *On ne rejette pas l'hypothèse nulle* $\pi_{01} = \pi_{10} = \frac{1}{2}$ …

L'analyse des motifs (patterns) dynamiques

- (Autre) exemple numérique avec Excel (suite) :



  - *Statistique de Christoffersen pour cette trajectoire* $= 3{,}77$.
  - *On ne rejette pas non plus l'hypothèse nulle* $\pi_{01} = \pi_{10} = \frac{1}{2}$
  - *Puissance du test :* $P(\text{rejet } H_0 | H_1)$ *est faible*
  - *On a du mal à détecter une déviation d'une suite aléatoire*

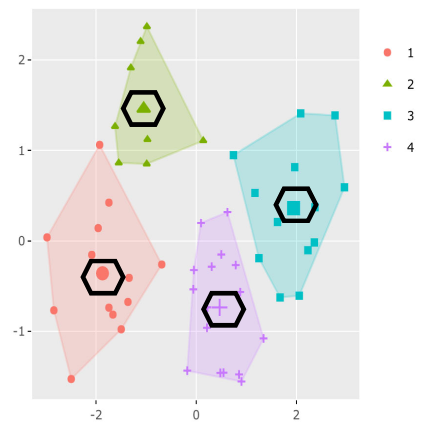
Exercice : classification et indices boursiers

- Préambule : on peut chercher à regrouper des actions en classes (clusters) en fonction de certaines caractéristiques
  - *Taille (capitalisation boursière) : blue chips, small caps*
  - *Market to Book (rapport entre capitalisation boursière et valeur des fonds propres comptables)*
    - Critères utilisés dans le modèle à trois facteurs de Fama et French
  - *Secteur d'activité : industrie, finance, etc.*
  - *Zone géographique d'activité : Europe, Amérique du Nord, Asie, …*
    - Critères utilisés dans l'analyse du risque de crédit
  - *Caractéristiques statistiques : corrélation, similitude des rentabilités passées (facteur « momentum » de Cahart)*
    - Des actions dont les rentabilités passées présentent des motifs similaires vont appartenir au même groupe

Exercice : classification et indices boursiers

- Une classification consiste à affecter chaque action à un groupe d'appartenance ou classe.
- Les actions d'un même groupe ou classe sont supposées avoir des caractéristiques similaires et distinctes des actions des autres groupes
- En statistique, une méthode courante de classification cherche à former des groupes (clusters, agrégats), tels que la variance intra-groupe soit minimale et la variance inter-groupe maximale
  - *La similarité entre deux membres d'une même classe est élevée*
  - *La dissemblance entre deux membres de classes différentes est élevée*

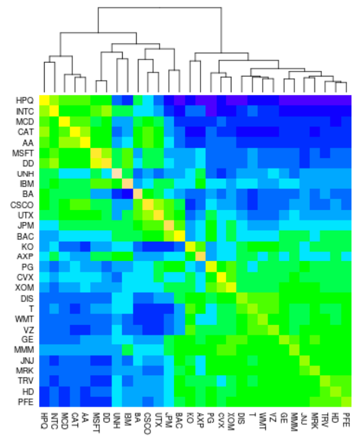
Exercice : classification et indices boursiers



- Un exemple avec quatre clusters
- les distances intragroupes sont minimales
- Les distances intergroupes maximales
- Chaque hexagone ⬡ encadre le barycentre de chaque classe
  - *Barycentre : moyenne*
- Distances intergroupes calculées à partir des distances entre les barycentres de chaque classe
- Est-ce que l'individu moyen est représentatif de sa classe ?

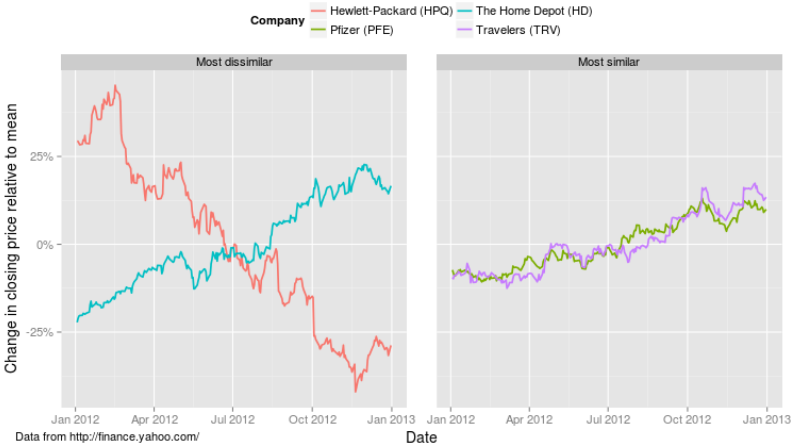
Exercice : classification et indices boursiers

- Classification des actions du Dow Jones
- A partir des évolutions des prix au cours de l'année 2012
- « Heatmap » des corrélations
- En jaune/vert, corrélations positives
- En bleu corrélations négatives
- Au dessus du graphique : regroupements des titres
- Classification hiérarchique ascendante
  - *Codes sources R*
    - http://www.alastairsanderson.com/projects/Dow-Jones-Industrial-Average-stock-clustering-analysis/



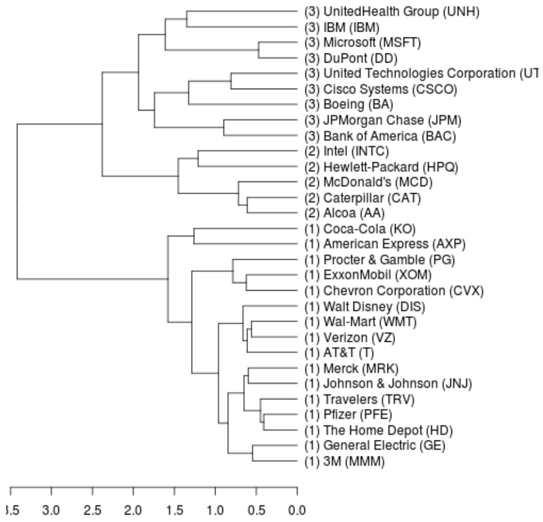
Exercice : classification et indices boursiers

- Regroupement des titres en fonction de leur proximité
  - *Les prix des deux paires les plus semblables et les plus dissemblables*



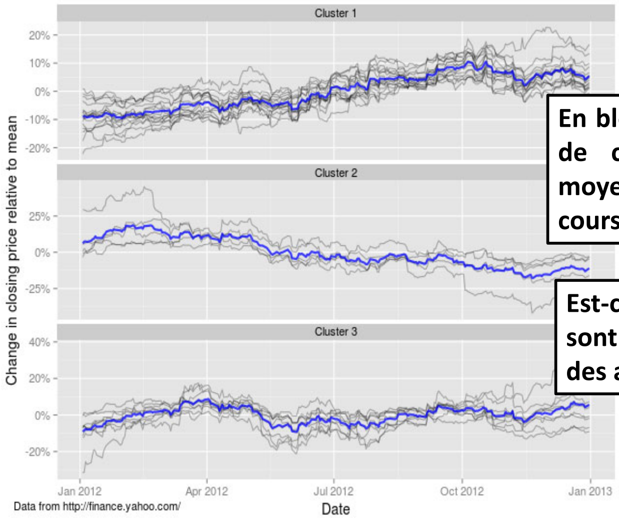
Exercice : classification et indices boursiers

- Regroupements des titres en clusters en fonction de leur proximités
- Classification hiérarchique ascendante
- On va considérer les trois premiers clusters



Exercice : classification et indices boursiers

- Evolution des prix des actions des trois premiers secteurs
  - *Après recentrage et mise à l'échelle*



En bleu, on représente le barycentre de chaque classe, c'est-à-dire la moyenne des évolutions de chaque cours boursiers dans sa classe.

Est-ce que les évolutions en bleu sont représentatives des évolutions des actions dans chaque classe ?

Exercice : classification et indices boursiers

- Est-ce que l'individu moyen est représentatif de sa classe ?
- Formalisation : on considère $i = 1, \dots, n$ actions appartenant à une même classe (ou indice boursier)
  - $r_1, \dots, r_n$ *: rentabilités (supposées centrées)*
  - $\sigma_1, \dots, \sigma_n$ *: écarts-types des rentabilités (supposés positifs)*
  - ρ_{ij} *: coefficient de corrélation linéaire entre* r_i *et* r_j
  - $\omega_1, \dots, \omega_n$ *: poids des actions dans l'indice.*
  - *Ces poids sont supposés positifs ou nuls (barycentre)*
  - $r = \omega_1 r_1 + \cdots + \omega_n r_n$ *: rentabilité de l'indice (ou barycentre de la classe)*
  - *On notera* σ*, l'écart-type de la rentabilité de l'indice*

Exercice : classification et indices boursiers

- Question 1 : Donner l'expression du Bêta du titre i par rapport à l'indice de référence.

Exercice : classification et indices boursiers

- Question 1 : Donner l'expression du Bêta du titre i, β_i, par rapport à l'indice de référence.
  - $\beta_i = \frac{cov(r_i,r)}{cov(r,r)} = \frac{\omega_1 cov(r_i,r_1)+\cdots+\omega_n cov(r_i,r_n)}{\sigma^2}$
  - $\beta_i = \frac{\sigma_i}{\sigma^2} \times (\omega_1\rho_{i1}\sigma_1 + \cdots + \omega_n\rho_{in}\sigma_n)$
- Question 2 : on suppose que les $\rho_{ij} \geq 0$. Montrer que les β_i sont > 0
  - *L'hypothèse est cohérente avec l'homogénéité intra-groupe, les corrélations deux à deux devant être élevées*
- Question 3 : à quoi est égal $\omega_1\beta_1 + \cdots + \omega_n\beta_n$?
  - *Moyenne des Bêtas ?*

Exercice : classification et indices boursiers

- Questions 1 &2 : Donner l'expression du Bêta du titre i, β_i, par rapport à l'indice de référence.
  - $\beta_i = \frac{\sigma_i}{\sigma^2} \times (\omega_1\rho_{i1}\sigma_1 + \cdots + \omega_n\rho_{in}\sigma_n) > 0$
- Question 3 : à quoi est égal $\omega_1\beta_1 + \cdots + \omega_n\beta_n$?
  - $cov(r,r) = cov(r, \omega_1 r_1 + \cdots + \omega_n r_n) = \omega_1 cov(r,r_1) + \cdots + \omega_1 cov(r,r_n)$
  - *En divisant à gauche et à droite par* $cov(r,r)$
  - $1 = \omega_1\beta_1 + \cdots + \omega_n\beta_n$ *(linéarité des bêtas)*
  - *Remarque :* $1 = \frac{cov(r,r)}{cov(r,r)}$ *: Bêta de l'indice*
- Question 4 : On définit le risque idiosyncratique comme $\varepsilon_i = r_i - \beta_i r$. Quel est le coefficient de corrélation linéaire entre ε_i et r ?

Exercice : classification et indices boursiers

- Question 1 : Donner l'expression du Bêta du titre i, β_i, par rapport à l'indice de référence.
  - $\beta_i = \frac{\sigma_i}{\sigma^2} \times (\omega_1\rho_{i1}\sigma_1 + \cdots + \omega_n\rho_{in}\sigma_n) > 0$
- Question 3 : $\omega_1\beta_1 + \cdots + \omega_n\beta_n = 1$
- Question 4 : On définit le risque idiosyncratique comme $\varepsilon_i = r_i - \beta_i r$. Quel est le coefficient de corrélation linéaire entre ε_i et r ?
  - *Numérateur du coefficient de corrélation linéaire :* $cov(\varepsilon_i, r)$
  - $cov(\varepsilon_i, r) = cov(r_i - \beta_i r, r) = cov(r_i, r) - \beta_i cov(r,r) = 0$
  - *Puisque* $\beta_i = cov(r_i, r)/cov(r,r)$
- Question 5 : décomposer le risque total du titre i

Exercice : classification et indices boursiers

- Question 3 : $\omega_1\beta_1 + \cdots + \omega_n\beta_n = 1$
- Question 4 : On définit le risque idiosyncratique comme $\varepsilon_i = r_i - \beta_i r$ (les rentabilités sont centrées)
  - *Coefficient de corrélation linéaire entre ε_i et $r = 0$*
- Question 5 : décomposer le risque total du titre i
  - $r_i = \beta_i r + \varepsilon_i$
  - $\sigma_i^2 = \beta_i^2\sigma^2 + \sigma_{\varepsilon_i}^2$ *(car ε_i et r non corrélés)*
  - *$\beta_i\sigma$ est le « risque de marché » du titre i*
- Question 6 : décomposer le risque de l'indice en fonction des risques de marché

Exercice : classification et indices boursiers

- Question 3 : $\omega_1\beta_1 + \cdots + \omega_n\beta_n = 1$
- Question 4 : On définit le risque idiosyncratique comme $\varepsilon_i = r_i - \beta_i r$ (les rentabilités sont centrées)
  - *Coefficient de corrélation linéaire entre ε_i et $r = 0$*
- Question 5 : décomposer le risque total du titre i
  - *$\beta_i\sigma$ est le risque de marché du titre i*
- Question 6 : décomposer le risque de l'indice en fonction des risques de marché
  - $\omega_1\beta_1 + \cdots + \omega_n\beta_n = 1 \Rightarrow \sigma = \omega_1(\beta_1\sigma) + \cdots + \omega_n(\beta_n\sigma)$
  - *Le risque de l'indice est la moyenne pondérée des risques de marché des titres qui le composent*
- Question 7 : en déduire une inégalité

Exercice : classification et indices boursiers

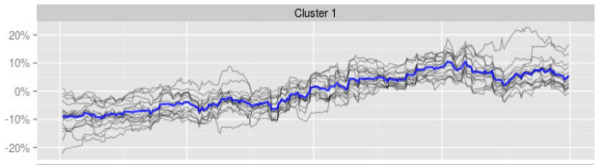
- Question 3 : $\omega_1\beta_1 + \cdots + \omega_n\beta_n = 1$
- Question 5 : $\beta_i\sigma$ est le risque de marché du titre i
- Question 6 : $\sigma = \omega_1(\beta_1\sigma) + \cdots + \omega_n(\beta_n\sigma)$
- Question 7 : en déduire une inégalité
  - $\sigma_i^2 = \beta_i^2\sigma^2 + \sigma_{\varepsilon_i}^2 \Rightarrow \sigma_i \geq \beta_i\sigma$
  - *L'écart entre σ_i et $\beta_i\sigma$ est d'autant plus élevé que le risque idiosyncratique σ_{ε_i} est élevé*
  - $\omega_1\sigma_1 + \cdots + \omega_n\sigma_n \geq \sigma$
  - *Le risque du barycentre (indice), σ est inférieur à la moyenne des risques individuels, σ_i des actions composant la classe*
- Question 8 : donner une signification au ratio $\beta_i\sigma/\sigma_i$

Exercice : classification et indices boursiers

- Question 3 : $\omega_1\beta_1 + \cdots + \omega_n\beta_n = 1$
- Question 5 : $\beta_i\sigma$ est le risque de marché du titre i
- Question 6 : $\sigma = \omega_1(\beta_1\sigma) + \cdots + \omega_n(\beta_n\sigma)$
- Question 7 : en déduire une inégalité
  - $\sigma_i^2 = \beta_i^2\sigma^2 + \sigma_{\varepsilon_i}^2 \Rightarrow \sigma_i \geq \beta_i\sigma$
  - $\omega_1\sigma_1 + \cdots + \omega_n\sigma_n \geq \sigma$
- Question 8 : donner une signification au ratio $\beta_i\sigma/\sigma_i$
  - *Notons ρ_i le coefficient de corrélation linéaire entre r_i et r*
  - $\rho_i = \frac{cov(r_i,r)}{\sigma_i\sigma} = \frac{cov(\beta_i r+\varepsilon_i,r)}{\sigma_i\sigma} = \frac{\beta_i cov(r,r)+cov(\varepsilon_i,r)}{\beta_i} = \frac{\beta_i\sigma^2}{\beta_i}$
  - $\rho_i = \beta_i\sigma/\sigma_i$

Exercice : classification et indices boursiers

- Remplacer un élément de la classe (action) par un indice (barycentre) amène à des trajectoires trop lissées



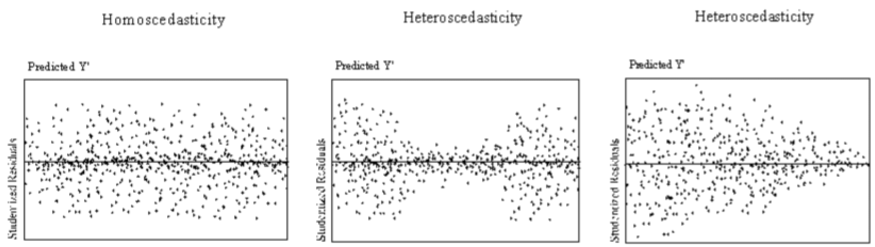
- Le barycentre n'est en général pas représentatif de la classe en termes de risque
  - *Sauf les corrélations intra-classes sont très élevées (proches de un)*
- Le phénomène est d'autant plus marqué que les corrélations intra-classes sont faibles

Remarque : corrélation entre les risques idiosyncratiques

- Suivant une tradition initiée par Sharpe, le modèle de marché est interprété comme un modèle à un seul facteur
  - $r_i = \beta_i r + \varepsilon_i$
    - Comme précédemment, on considère des rentabilités centrées
- Ceci ressemble à la décomposition précédente. Mais on rajoute l 'hypothèse : $r, \varepsilon_1, \dots, \varepsilon_n$ sont indépendantes
  - *L'indépendance entre r et ε_i implique la nullité de la corrélation mais la réciproque n'est pas vraie*
  - *Elle permet d'assurer que l'on a bien une relation linéaire où ε_i joue le rôle d'un « bruit ».*
  - *Indépendance ⇒ variance de ε_i ne dépend pas r, par exemple qu'elle n'est pas plus élevée pour des valeurs extrêmes de r*
  - *On parle alors d'un modèle homoscédastique*
  - *On peut aussi ajouter l'hypothèse que : $r, \varepsilon_1, \dots, \varepsilon_n$ sont gaussiennes*

Remarque : corrélation entre les risques idiosyncratiques

- Distribution des résidus pour des modèles linéaires homoscédastiques et hétéroscédastiques



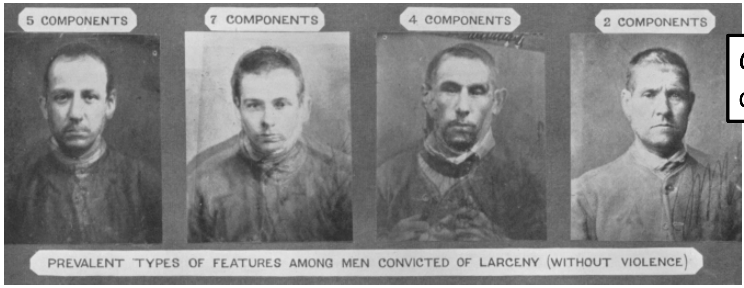
- Sous les conditions précédentes, la méthode des moindres carrés ordinaires permet une estimation optimale des β_i
- Dans le cas de la finance, il y a un petit problème avec les hypothèses précédentes : l'indépendance des $\varepsilon_1, \dots, \varepsilon_n$

Remarque : corrélation entre les risques idiosyncratiques

- Risques idiosyncratiques et modèle à un facteur
  - $r_i = \beta_i r + \varepsilon_i,\ i = 1, \dots, n$
  - $r = \omega_1 r_1 + \cdots + \omega_n r_n = (\omega_1\beta_1 + \cdots + \omega_n\beta_n) r + \omega_1\varepsilon_1 + \cdots + \omega_n\varepsilon_n = r + \omega_1\varepsilon_1 + \cdots + \omega_n\varepsilon_n$ *car* $\omega_1\beta_1 + \cdots + \omega_n\beta_n = 1$
  - *Par indépendance de $r, \varepsilon_1, \dots, \varepsilon_n$, la variance de $r + \omega_1\varepsilon_1 + \cdots + \omega_n\varepsilon_n$ est égale à $\sigma^2 + \omega_1^2\sigma_{\varepsilon_1}^2 + \cdots + \omega_n^2\sigma_{\varepsilon_n}^2$.*
  - *On obtient donc $\omega_1^2\sigma_{\varepsilon_1}^2 + \cdots + \omega_n^2\sigma_{\varepsilon_n}^2 = 0$*
  - *Comme les poids $\omega_1, \dots, \omega_n$ sont positifs, $\sigma_{\varepsilon_1} = \cdots = \sigma_{\varepsilon_n} = 0$ et comme les variables sont centrées $\varepsilon_1 = \cdots = \varepsilon_n = 0$.*
  - *Le modèle à un facteur est dégénéré si le facteur est un indice*
  - *L'équation rouge signifie $\omega_1\varepsilon_1 + \cdots + \omega_n\varepsilon_n = 0$, c'est-à-dire pas de risque idiosyncratique dans l'indice*
  - $\omega_1\varepsilon_1 + \cdots + \omega_n\varepsilon_n = 0$ *incompatible avec $\varepsilon_1, \dots, \varepsilon_n$ indépendants*

Exercice : classification et indices boursiers

- L’individu moyen est-il représentatif de sa classe ?
- Problème est ancien dans l’histoire de la statistique
  - *Francis Galton l’un des pères de la statistique : coefficient de corrélation, retour à la moyenne, etc. s’est par exemple intéressé*
  - *En superposant des photos de voleurs à la tire (après recentrage sur les yeux et mise à l’échelle), il a obtenu des visages moyens*



Composite portraits de voleurs à la tire

Exercice : classification et indices boursiers

- L’individu moyen est-il représentatif de sa classe ?
  - Galton (1879). Composite portraits, made by combining those of many different persons into a single resultant figure. *The Journal of the Anthropological Institute of Great Britain and Ireland*.

Images composites (moyennes) de criminels : à gauche, à partir de 9 individus, à droite à partir de 5, au milieu la moyenne des images de droite et de gauche

  - *Un criminel « moyen » ressemble à un anglais moyen pauvre …*
  - *Gommer les « traits pathogènes » rend l’image « sympathique »*

Exercice : classification et indices boursiers

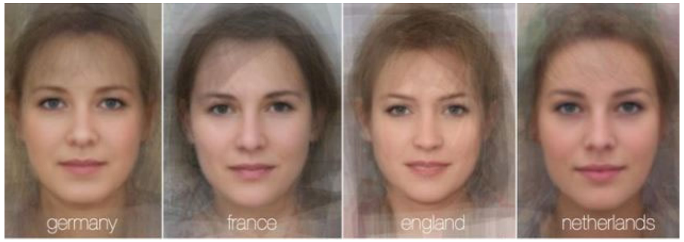
- Adolphe Quételet avait utilisé le concept *d’homme moyen physique*, un homme dont les caractéristiques physiques seraient les moyennes de tous les hommes
  - *Si la population se compose d’hommes de 1,70 m pour 50% et de 1,90 m pour les autres 50%, il n’y a pas d’homme moyen*
- Galton savait que les moyennes effacent les différences
  - *« Tout ce qui est commun reste, tout ce qui est individuel tend à disparaître »*
    - Galton (1883) Inquiries into Human Faculty and its Development
    - Morini (2010). Francis Galton ou comment photographier une moyenne. *Mathématiques et sciences humaines.*
    - Brogowski (2003). De l'idéal (dé) tourné en Witz. La photographie composite de Francis Galton et ses résonances. *Revue d'esthétique.*

Exercice : classification et indices boursiers

- Une moyenne de 60 visages fait apparaître un effet de halo et ne ressemble pas encore à un « vrai visage »
  - *Le risque spécifique n’est pas complètement éliminé*

Exercice : classification et indices boursiers

- Ci-dessous des images obtenues par moyennage à partir de photographies de femmes de différents pays
  - Source Face Research Lab, Institute of Neuroscience and Psychology at the University of Glasgow, http://facelab.org/

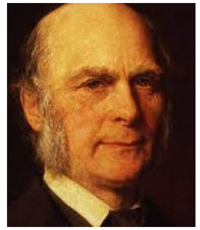


- *Selon certains, les visages symétriques, arrondis, à la peau homogène, aux contours réguliers, avec des courbures fines sont jugés plus attractifs.*
- *Or, c'est ce que fait le moyennage.*

Exercice : classification et indices boursiers

- Kant (*critique du jugement, du jugement esthétique, analytique du beau*) fait appel à la notion d'image moyenne comme élément de référence (à tort ?)
  - *Bien plus, quand l'esprit veut instituer des comparaisons, l'imagination (...) rappelle les images les unes sur les autres, et par cet assemblage de plusieurs images de la même espèce, fournit une moyenne qui sert de mesure commune.*
  - *Si maintenant on cherche (...) pour cet homme moyen la tête moyenne, pour celle-ci le nez moyen , etc., cette figure donnera l'idée normale du bel homme dans le pays où se fait la comparaison.*

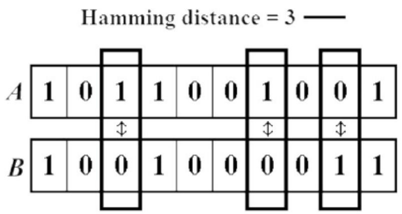
+



= ?

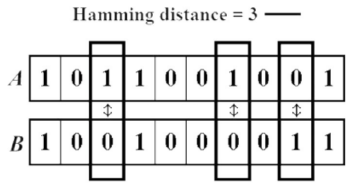
Exercice : classification et indices boursiers

- Analyse et la classification des motifs boursiers
  - *d : distance de Hamming entre deux motifs*



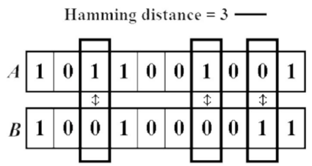
- *Définissons la classe* C_A *des éléments proches du motif A comme l'ensemble des motifs de distance à A* ≤ 3
- $C_A = \{M, d(A, M) \leq 3\}$
- *Par construction, A devrait être représentatif de cette classe*

Exercice : classification et indices boursiers



- *Supposons que* C_A *se résume à* $\{A, B\}$
- *De par sa* ***centralité****, A est dit* ***prototype*** *de la classe* C_A
- *Remarquons que* $\frac{A+B}{2}$ *n'est même pas un motif : 0,5 ne correspond ni à une hausse, ni à une baisse.*
- *Il ne faut pas confondre centralité et moyenne, cela dépend de la définition pertinente des proximités entre motifs.*
- *On retrouve la problématique du sens à donner à la moyenne arithmétique.*

Exercice : classification et indices boursiers



- *Définissons maintenant C comme l'ensemble des motifs qui prennent les mêmes valeurs que A en position* 1,2,4,5,6,8,10.
- *Ce qui caractérise C, c'est un ensemble de traits communs*
- *Supposons que C se résume à $\{A, B\}$*
- *Ni A, ni B n'est privilégié : ils sont représentatifs de C*
- L'exemple précédent illustre deux approches de la classification et par conséquent de la représentativité
- *Théorie du prototype* et *approche par conditions nécessaires et suffisantes*

- Classe C définie par conditions nécessaires et suffisantes : identité des valeurs de A en position 1,2,4,5,6,8,10
  - *A et B ont le même statut*
- C_A est constituée selon l'approche prototypique : A est le prototype de C_A

	CNS (Conditions nécessaires et suffisantes)	Prototype (version standard)
Structure des catégories horizontales	Homogène : les membres d'une catégorie ont un statut équivalent. Ex : dans la catégorie **oiseau**, *moineau, aigle, autruche, manchot* ont le même statut. Les catégories ont des limites nettes, elles sont rigides.	Prototypique : il existe une entité centrale qui représente le « meilleur exemplaire », les autres membres sont de « plus ou moins bons » exemplaires. Ex : *aigle* et *moineau* sont « plus oiseau » que *manchot* ou *autruche*. Les limites sont floues.
Appartenance à une catégorie	De type vrai/faux : un exemplaire appartient ou n'appartient pas à une catégorie, tout exemplaire est également représentatif. La détermination de l'appartenance est analytique. Les cas marginaux ne sont pas gérés.	Le degré de représentativité d'un exemplaire correspond à son degré d'appartenance à la catégorie. La détermination de l'appartenance est plus globale et s'effectue sur la base du degré de similarité avec le prototype. Les cas marginaux sont pris en compte (ex : chaise à 3 pieds...)
Partage des propriétés	Tous les membres d'une catégorie possèdent obligatoirement un même ensemble de propriétés, lequel constitue la condition nécessaire d'appartenance.	Chaque membre d'une catégorie possède au moins une propriété commune avec le prototype. C'est une « ressemblance de famille » qui les regroupe.

Tableau extrait de Kleiber : La sémantique du prototype (1990)