Société de Calcul Mathématique SA
Outils d'aide à la décision
depuis 1995

Analyse critique de la réplication d'indices

en Ingénierie Financière

I. Méthodologie

Rapport rédigé par la

Société de Calcul Mathématique SA

Mai 2018

rédaction : Bernard Beauzamy, Adrien Schmitt

Siège social et bureaux : 111, Faubourg Saint Honoré, 75008 Paris. Tel : 01 42 89 10 89. Fax : 01 42 89 10 69. www.scmsa.eu
Société Anonyme au capital de 56 200 Euros. RCS : Paris B 399 991 041. SIRET : 399 991 041 00035. APE : 7219Z

Résumé Opérationnel

La "réplication d'indices" cherche à reproduire, aussi bien que possible, les résultats d'un "hedge fund" en se servant des valeurs d'un certain nombre d'indices dont on pense qu'ils peuvent expliquer ses performances.

Les arguments en faveur d'une telle approche sont : moins de frais, plus de disponibilité des sommes investies ("liquidité"), éventuellement moins de risques, etc. De ce fait, la réplication peut intéresser certains investisseurs, qui ne peuvent investir directement dans les fonds répliqués (pour des raisons réglementaires, de faible périodicité de valorisation, de risque d'illiquidité, de ticket d'entrée trop élevé …). Elle rend le produit plus accessible.

A l'inverse, une objection possible est que l'objectif même de la réplication d'indices est factice et ne répond pas aux besoins d'un investisseur, lequel, par définition, veut gagner de l'argent. Cette réplication, "parent pauvre" d'une politique de gestion, sera tout au plus capable d'assurer un suivi de mauvaise qualité.

Nous ne cherchons pas ici à trancher ce débat de fond ; nous faisons simplement une analyse des méthodes mathématiques, présentées dans la littérature spécialisée, pour réaliser la réplication.

Notre conclusion est simple :

Ces méthodes mathématiques sont entièrement dépourvues de valeur scientifique. Il s'agit de formules "factices", auxquelles on attribue une validité qu'elles n'ont pas, en jouant sur les mots et sur les définitions. Les formules empruntées aux statistiques, à l'origine, sont simples, mais on leur prête un sens qu'elles n'ont pas, et un pouvoir explicatif qu'elles n'ont pas.

Notre méthode de travail :

Entre 2009 et 2011, nous avons réalisé six études pour AXA Private Equity (maintenant Ardian Investments). Nous suivrons ici la même méthode : nous lisons en détail tous les documents utiles et nous jugeons "sur pièces", comme nous lirions une démonstration mathématique :

- Les hypothèses sont-elles clairement énoncées ; sont-elles réalistes ?
- Les raisonnements sont-ils clairement détaillés ?
- Les conclusions apparaissent-elles solidement établies ?

Un élément essentiel de la recherche scientifique, comme le souligne Richard Feynman, est que tous les éléments de jugement doivent être disponibles, et pas seulement ceux qui vont dans un sens favorable. Un modèle de prévision donne de bons résultats pour l'année 2011, soit. Mais :

- Quels résultats obtient-il les autres années ?
- Repose-t-il sur des arguments valides, ou bien le résultat est-il le fait du hasard ?
- Qu'ont donné les autres modèles, cette année-là et les autres années ?

I. L'objectif de la réplication

Il s'agit de reproduire, aussi bien que possible, les résultats d'un "hedge fund" en se servant des valeurs d'un certain nombre d'indices dont on pense qu'ils peuvent expliquer ses performances.

L'objectif de la réplication est de reproduire le comportement boursier d'une stratégie de Hedge Fund. Les raisons invoquées sont les suivantes :

1) ce type de produits n'est pas facilement accessible (ticket d'entrée élevé, fréquence de valorisation trop faible, liquidité du produit qui ne satisfait pas les investisseurs) ;

2) le comportement de ces stratégies se différencie des autres produits financiers et donc suscite un intérêt pour diversifier les portefeuilles (gestion du risque global).

Prenons un exemple concret, pour clarifier le vocabulaire et exposer la méthode.

Partons de l'indice S&P 500, indice Standard and Poor constitué des 500 plus grosses entreprises américaines. On peut être tenté de "répliquer" cet indice (on dit parfois "cloner") de la manière suivante : on choisira une dizaine d'entreprises sur les 500, on observera les variations de leur cours de bourse et, à partir de ces variations-là seulement, on va constituer une "combinaison linéaire" (voir plus loin) de ces 10 facteurs ; si les dix entreprises ont été convenablement choisies, on peut espérer que le résultat sera proche de l'indice S&P 500.

Le travail à faire sera :

– Faire un choix judicieux des 10 facteurs parmi les 500 ;
– Déterminer une méthode appropriée pour les combiner en un indice unique.

En admettant que ceci soit fait correctement, les avantages supposés sont :

– C'est plus facile à réaliser, puisque dix informations seulement sont utiles ;
– Ce peut être totalement automatisé ;
– Cela coûtera moins cher en frais de gestion : pour être performant, un gérant de fonds doit en permanence acheter et vendre, et ces transactions sont rémunérées. Ici, la stabilité de la méthode fera faire des économies ;
– C'est plus "transparent", en ce sens que la méthode peut être rendue publique, alors que le gérant d'un fonds ne dit pas comment il fait.

Mais, inversement :

– On a certainement perdu de l'information par rapport à l'indice de départ, et on ne sait pas laquelle ;
– On remplace la gestion intelligente, faite par un spécialiste qui sait quand acheter ou vendre, par une politique "passive", copiée sur les aléas antérieurs du marché, mais incapable de s'adapter à une évolution ultérieure, si elle n'a jamais été rencontrée.

C'est donc, par principe, une tentative d'automatisation, à des fins de pure simplification. On cherche à se débarrasser de l'intelligence humaine (c'est très souvent assumé, et même mis en avant), en la remplaçant par des mécanismes automatiques ; le résultat, dans un univers d'une très grande complexité, ne peut être que le "parent pauvre" de la politique de gestion conduite par un spécialiste. Du reste, les responsables des stratégies de réplication le disent clairement : nous ne cherchons pas à gagner plus d'argent que ne le fait le fonds d'origine, mais seulement à suivre celui-ci. Cela facilite l'accès : plus d'investisseurs peuvent acheter ce type de produits. Cela fait plus d'encours à gérer et donc une rémunération plus importante pour le responsable.

L'idée d'analyser les différents facteurs, pour voir lesquels concourent le plus aux résultats, peut être intéressante pour le gérant d'un hedge fund lui-même : il peut ainsi constater que certains éléments ne concourent pas aux résultats, et vouloir s'en débarrasser. C'est le même principe que nous rencontrons dans d'autres contextes : lorsqu'on dispose d'un réseau de capteurs, on peut vouloir simplifier ce réseau en supprimant les capteurs redondants, mais on perd alors en robustesse.

Mais cette simplification de l'information, par un opérateur compétent, n'est pas une politique systématique et, plus tard, il pourra vouloir réintégrer les facteurs supprimés. Soyons clairs : nous n'avons aucune objection, bien au contraire, à l'encontre de l'analyse fine de la contribution de chaque facteur aux résultats ; c'est l'idée d'adopter un mécanisme aveugle et permanent qui nous paraît malsaine.

Dans la mesure où un spécialiste est rémunéré "au résultat", comme c'est généralement le cas, il nous semble que la politique d'économie visant à remplacer le spécialiste par un robot médiocre n'est pas justifiée pour un investisseur, sauf guerre des prix alignée sur la médiocrité.

II. Les méthodes de la réplication

La plupart des approches sont des modèles linéaires (combinaisons linéaires de facteurs). Très grossièrement, la littérature fait apparaître trois ensembles de méthodes :

1. Les facteurs sont choisis au sein d'un ensemble prédéfini pour leur cohérence économique attendue à la création de cette performance (des indices, des variables macro-économiques,...). Il y a donc là un choix réfléchi.

2. Les facteurs sont déterminés par une méthode purement automatique d'ajustement linéaire au sens des moindres carrés. Cette méthode, purement aveugle, est la plus répandue.

3. Les performances des facteurs sont construites à partir des sensibilités à ces facteurs pré-calculées par une méthode exogène (souvent des z-scores), approche appelée "cross-sectional regression" (opposée aux time-series regression).

Nous nous limiterons à l'analyse de la seconde méthode. Comme nous le verrons plus loin, la littérature sur ces questions est extrêmement confuse. Le vocabulaire lui-même est mal défini ; les gens ne parlent pas de la même chose. Les méthodes, les algorithmes, ne sont jamais convenablement présentés ; les résultats qui apparaissent dans les articles sont systématiquement contredits par des notes en bas de page qui annoncent "ces résultats sont hypothétiques". En bref, pour réaliser son analyse, le mathématicien professionnel ne dispose pas des éléments habituellement disponibles pour la recherche scientifique, que ce soit en biologie ou en astronomie, ou dans tout autre discipline. Nous avons donc choisi de prendre les choses à leur début.

III. L'utilisation d'outils mathématiques

Même lorsqu'ils sont anciens, les outils mathématiques doivent être utilisés avec discernement ; ceci est vrai pour tous les outils : il existe différents types de marteaux et de ciseaux. L'utilisation "en aveugle", comme elle est faite ici, conduit à des erreurs.

Voyons ceci de plus près sur un exemple simple, très pertinent ici : celui de l'ajustement linéaire (malheureusement, le vocabulaire utilisé, surtout pour les modèles financiers, est celui de "régression" ; ce vocabulaire vient de Francis Galton, 1886, qui a étudié la variation de la taille d'individus ; par la suite, les mathématiciens ont été incapables d'imposer un vocabulaire plus pertinent).

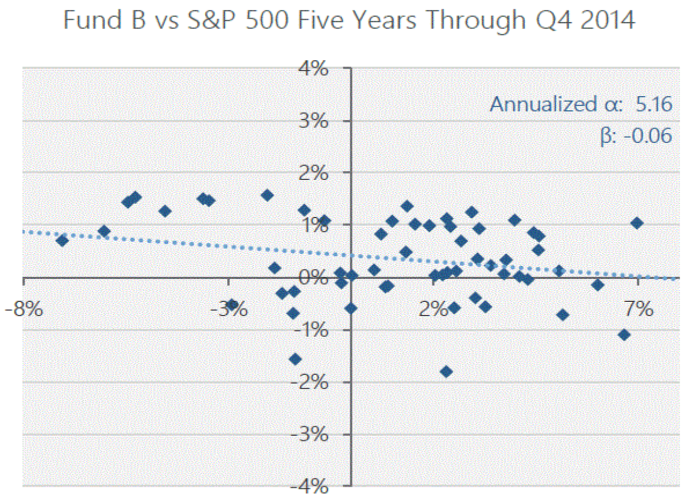


Fig. Un exemple d'ajustement linéaire

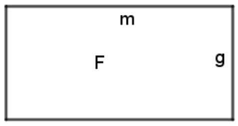
Pour réaliser un ajustement, on part d'un nuage de points (ensemble quelconque de points dans le plan, ou dans l'espace) et on cherche à faire passer une droite qui les approche au mieux. Voici un exemple (figure ci-dessus) ; dans ce cas, les points sont très dispersés et l'ajustement n'est pas fameux. Mais enfin, on a le droit de le réaliser.

Mathématiquement parlant, étant donné des points (x_i, y_i) dans le plan $(i = 1,\dots,N)$, on cherche une droite d'équation $y = ax + b$ qui minimise l'ensemble des distances aux points ; on cherche à minimiser $\sqrt{\sum_{i=1}^{N}\left(y_i - (ax_i + b)^2\right)}$, ce qui se calcule facilement. Tout le monde fait cela depuis plus d'un siècle, et bien des logiciels ont automatisé cette fonction (dont Excel, bien sûr).

Malheureusement, il existe deux "pièges", auxquels personne ne pense jamais, tant la méthode est devenue habituelle ; ils sont décrits en Annexe I. Disons que ce sont des choses importantes pour un mathématicien professionnel, mais qui relèvent plutôt du purisme, par comparaison avec les énormes bourdes que nous verrons bientôt.

IV. Des abus dans le vocabulaire

Prenons une formule mathématique quelconque ; elle ne se transpose pas nécessairement au monde réel. On a parfaitement le droit de dessiner un rectangle, d'appeler m le grand côté, g le petit côté et F la superficie, et on obtient la formule (tout à fait exacte) $F = mg,$ mais on n'a pas pour autant démontré la formule fondamentale de la dynamique !



C'est dans la conversion aux modèles financiers de formules mathématiques simples, élaborées dans d'autres cadres, que se situent les erreurs majeures, comme nous allons le voir.

A. Approche probabiliste

Il est tout à fait légitime de considérer que les résultats à venir d'un indice boursier (par exemple le S&P 500, pour fixer les idées) peuvent être considérés comme des réalisations d'une variable aléatoire. Non pas qu'ils soient dus au hasard, mais parce qu'ils résultent d'influences que nous ne pouvons pas, ou ne voulons pas, connaître (voir notre livre [MPPR2] pour des explications détaillées). L'approche probabiliste des outils financiers est donc justifiée.

Le problème est celui de la loi que l'on va attribuer à cette variable aléatoire (en abrégé v.a.). On la construit traditionnellement sur la base d'un historique plus ou moins long. Mais cela n'a de sens que si la loi est "stationnaire", c'est-à-dire qu'elle ne varie pas dans le temps.

"Stationnaire" ne signifie pas que le processus est constant, mais que sa loi est toujours la même. Par exemple, si on joue à pile ou face, le processus n'est pas constant (on tire l'un ou l'autre selon les cas), mais la probabilité de chaque résultat reste 1/2 dans chaque cas, tout le temps.

Dans le monde réel, très peu de processus ont réellement une loi stationnaire. On peut penser à la température un jour donné (par exemple le 15 mai) : elle varie d'une année sur l'autre, mais toujours selon la même loi (si l'on excepte les hypothèses de réchauffement climatique). En d'autres termes, la probabilité d'avoir une température $\geq 12°C$ serait la même tous les ans le 15 mai. Sur des tranches de dix mille ans, la proportion d'années où l'on enregistre le 15 mai une température $\geq 12°C$ serait à peu près constante (ceci est invérifiable !).

Si l'on veut comprendre quelque chose aux probabilités, il faut toujours se référer à la "loi empirique des grands nombres" ; par exemple dire "tel événement a une probabilité de 1/3" signifie que, sur une répétition de l'expérience faite 10 000 fois, à peu près un tiers verront l'événement réalisé.

Le S&P 500 n'a certainement pas une loi stationnaire, parce que son évolution dépend de quantité de facteurs qui, pour certains d'entre eux, ne se sont jamais rencontrés. Transposer les résultats généraux de la théorie des probabilités aux modèles financiers est donc par essence incorrect, sauf si on le fait sur de brèves périodes, suffisamment calmes, ou bien, à la rigueur, si l'on recherche un résultat grossier, pour lequel la robustesse paraîtra satisfaisante.

Ceci est déjà une première critique, assez fondamentale, à l'égard des méthodes de réplication, puisqu'elles font (sans le savoir et sans le dire) une hypothèse de stationnarité.

B. Application à la réplication

L'objectif de la réplication statistique d'un indice sera donc de se rapprocher de cet indice en utilisant seulement un certain nombre de variables explicatives.

La réplication, mathématiquement parlant, fonctionne de la manière suivante ; c'est une variante simple de l'ajustement, vu précédemment.

On dispose d'une variable d'intérêt Y, par exemple le résultat d'un indice. On dispose en outre d'un certain nombre de facteurs $X_1,...,X_N$, appelés facteurs explicatifs. Ces facteurs peuvent être endogènes à Y (comme par exemple une entreprise qui intervient dans le S&P 500) ou être complètement exogènes (ils n'ont, a priori, rien à voir : la consommation des ménages au Japon).

On écrit une décomposition de la forme :

$$Y = \alpha + \beta_1 X_1 + \cdots + \beta_N X_N + reste \qquad (1)$$

Les coefficients $\alpha, \beta_1,..., \beta_N$ sont calculés comme expliqué précédemment (ajustement au sens des moindres carrés) et le reste, inconnu, est simplement la différence :

$$reste = Y - \left(\alpha + \beta_1 X_1 + \cdots + \beta_N X_N\right) \qquad (2)$$

La partie "linéaire" :

$$Y_0 = \alpha + \beta_1 X_1 + \cdots + \beta_N X_N \qquad (3)$$

est la réplication recherchée.

Attention : on a réalisé une combinaison linéaire de facteurs, mais, sauf cas exceptionnel, les facteurs eux-mêmes ne sont pas linéaires. Ce sont des séries temporelles quelconques.

Il est bien évident qu'une décomposition du type (1) est toujours possible, avec n'importe quelles variables $Y, X_1,..., X_N$; tout est dans le "reste". Or ce reste disparaît systématiquement dans la littérature relative aux modèles financiers : dans les articles académiques, on lui prête des vertus qu'il n'a pas (généralement gaussien de moyenne nulle), dans les brochures commerciales, il n'est tout simplement pas mentionné.

C. *In hoc signo vinces*

A partir de ces formules très simples, et largement surévaluées comme nous l'avons vu plus haut, le monde de la gestion d'actifs financiers croit avoir découvert l'outil universel, le signe par lequel il vaincra. Il va accorder à tous les éléments de la formule des significations qu'ils n'ont pas dans la réalité.

– Le coefficient "alpha" est attribué aux capacités intrinsèques du gestionnaire, à son habileté. Mais ceci est tout à fait absurde : des ajustements faits sur des séries temporelles gérées par le même gestionnaire, mais sur des périodes différentes, donnent des coefficients "alpha" différents.

– Les coefficients "beta" sont simplement les parts prises par chaque facteur dans la composition d'ensemble ; on les associe à des "pentes" (ce qui n'est légitime que dans le cas d'un ajustement linéaire sur un nuage de points). Ils indiquent en réalité la "sensibilité" du résultat (on dirait plus simplement sa "dépendance") par rapport à chaque facteur. Si par exemple le coefficient β_1 est grand, le résultat Y dépendra plus du facteur X_1 que si β_1 était petit. Il faut bien noter que ces coefficients peuvent être négatifs.

– Le "risque", ou encore "volatilité", est défini comme l'écart-type σ de la série temporelle considérée.

Cette définition du "risque" est très surprenante ; il n'est pas légitime de l'associer à l'écart-type de la série temporelle. En effet, revenons à l'exemple des températures vu plus haut. On peut s'interroger sur la variabilité des températures le 15 mai ; pour cela, on utilisera l'historique des températures du 15 mai (enregistrées à Paris sur 150 ans environ), peut-être aussi l'ensemble des températures du mois de mai, mais on exclura certainement l'été et l'hiver, non représentatifs pour cette question. On voit donc bien sur cet exemple que l'écart-type de l'ensemble de la série temporelle ne caractérise pas convenablement le "risque".

Il faut bien voir que, dans la littérature sur ces sujets, il n'y a pas de consensus sur les définitions, ce qui ajoute à la confusion. Beaucoup d'articles discutent à perte de vue sur la signification "financière" du alpha et des betas. Comme nous le verrons plus loin, un article, qui constate que les définitions existantes ne sont pas satisfaisantes, propose d'en ajouter une troisième, encore moins claire.

On est ici en présence d'un dogme de nature religieuse. La corporation des tenants de la réplication a découvert une formule, qui est sa Bible. Elle est publiée, enseignée dans des Ecoles prestigieuses, fait l'objet de publications, de thèses, de communications dans des congrès, etc. Il ne servirait à rien de la discuter, tout comme il aurait été inutile de vouloir, avant Galilée, convaincre un prélat de l'Eglise catholique que la Terre tournait autour du Soleil, et non l'inverse. L'Eglise, après Galilée, a mis environ 200 ans pour l'admettre, et il faudra bien le même laps de temps aux responsables de modèles financiers.

Cette situation, insistons-y une seconde, n'est nullement limitée : on rencontre exactement la même avec les spécialistes d'épidémiologie (santé des populations) qui voient des dangers partout, à partir de modèles mathématiques incorrects.

Nous passons maintenant en revue, de manière plus détaillée, les erreurs commises.

V. Erreurs commises

A. Erreurs fondamentales

On lit un peu partout dans les articles analysés une assertion relative à l'équilibre des marchés. Les marchés, naturellement, tendraient vers un certain équilibre. Or ceci est absolument faux : dans quelque domaine que ce soit, la Nature ne recherche pas un équilibre, mais procède par des oscillations dont les amplitudes (positives ou négatives) vont en croissant. Nous n'insisterons pas davantage sur ce point ici.

De la même façon, les articles académiques reposent tous sur une "fonction d'utilité" (par exemple l'aversion au risque). Cette fonction d'utilité est purement académique ; elle n'a aucun sens en pratique. Elle va conditionner le résultat : selon qu'elle sera convexe ou concave, croissante ou décroissante, les résultats seront différents. Il faudrait au minimum faire une "analyse de sensibilité" : changer cette fonction, pour voir en quoi les résultats sont modifiés, mais aucun article ne fait jamais cette analyse.

Il faut rejeter par principe tout article qui repose sur une hypothèse d'équilibre et tout article qui met en œuvre une fonction d'utilité : ceci élimine la quasi-totalité des modèles économiques actuels, lesquels sont entièrement dépourvus de valeur effective.

Nous n'insisterons pas davantage sur ces deux points fondamentaux, parce que la réplication n'en dépend pas réellement : ils sont mentionnés dans la plupart des articles, comme éléments de base pour la réflexion.

B. Erreurs techniques

1. Choix de la période de référence

On travaille, par définition, sur des séries temporelles. La question est donc de savoir quand elles commencent et quand elles finissent. Si on veut prévoir un indice au moyen d'un ensemble de facteurs, on analysera les valeurs de toutes ces variables sur une période, pour réaliser l'ajustement, et on s'en servira pour prévoir la suite. Mais si on le fait, par exemple, sur la période 2000-2018 (allant jusqu'à l'époque actuelle), on réalise un ajustement qui n'aura aucune valeur prédictive, puisqu'on utilise les données actuelles ; c'est ce qu'un article (voir plus loin) appelle "look-ahead bias". Mais ce n'est pas un biais, c'est une malhonnêteté ! Par principe, il faut utiliser les données 2000-2015 pour prévoir 2016, etc.

Il est licite d'utiliser tout l'historique disponible : 1995 à 2015, si les données existent, pour prévoir 2016. Il n'y a pas de raison valable de "tronquer" ; bien au contraire, un bon gestionnaire a une mémoire bien au-delà de deux ans. La stratégie est en fait le résultat d'un compromis entre avoir suffisamment de données pour faire des estimations et pas trop non plus sinon on perd la stationnarité.

2. Valeurs réelles ou pourcentages

Les notations sont ici incroyablement confuses. Il semble que, dans bien des cas, les séries temporelles soient constituées de pourcentages d'un mois sur l'autre, et l'étude statistique est alors totalement incorrecte. Il faut travailler sur les valeurs réelles, éventuellement ramenées "base 100" à une date de référence.

Voici un exemple qui semble incorrect ; il est reproduit de
https://www.abcbourse.com/apprendre/19_variance_covariance.html

Exemple concret sur le marché

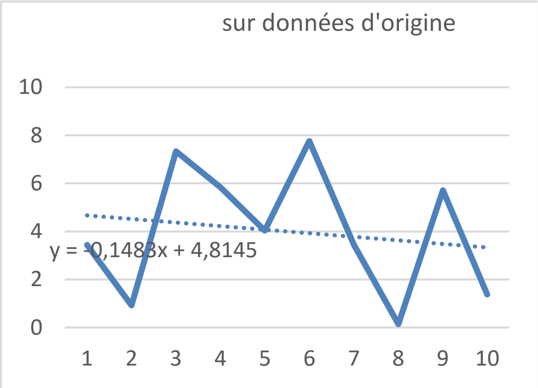
Prenons un exemple pour illustrer le calcul d'une covariance en finance. Il s'agira ici de comparer deux actions : Air France et Société Générale, dont nous avons récupéré les variations historiques mensuelles.

Mois	Société Générale	Air France
Juin 2009	-5,17%	-18,97%
Juillet 2009	15,87%	-2,97%
Août 2009	24,77%	20,61%
Septembre 2009	-2,15%	16,62%
Octobre 2009	-17,49%	-15,78%
Novembre 2009	3,33%	0,96%
Décembre 2009	4,39%	4,17%
Janvier 2010	-13,77%	7,55%
Février 2010	-4,29%	-17,08%
Mars 2010	15,25%	19,27%
Avril 2010	-13,23%	1,71%
Mai 2010	-13,24%	-17,06%
Moyenne	-0,4775%	-0,0808%

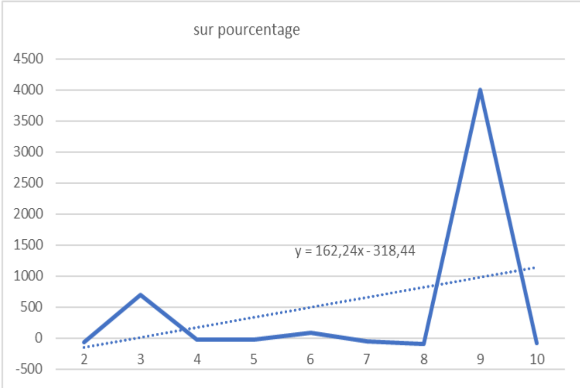
On n'obtient pas les mêmes résultats, selon que l'on fait l'ajustement sur des valeurs réelles ou sur des pourcentages d'accroissement d'une fois sur l'autre. Voici un exemple, que nous avons élaboré pour la circonstance. Il s'agit de dix valeurs ; à gauche les valeurs et à droite les accroissements :

3,43502462	
0,92459559	-73,0832906
7,33710825	693,547828
5,83089352	-20,5287245
4,030568	-30,8756369
7,76345372	92,6143838
3,44406903	-55,6374115
0,13940573	-95,9522958
5,71684778	4000,87009
1,36916518	-76,0503474

Ajustement sur données d'origine



Ajustement sur pourcentage



Valeur prévue dans le premier cas, pour le temps 11 : 3,18 , dans le second : 21,44. Les résultats sont complètement différents.

3. Corrélation et indépendance

Lorsque l'ajustement est réalisé, on calcule le coefficient de corrélation, qui permet, selon les auteurs, de mesurer la qualité de l'ajustement. Ceci est totalement faux, comme nous allons le voir, et l'ensemble de la construction liée à cette corrélation est totalement incorrect.

On lit, par exemple sur ABC Bourse :
https://www.abcbourse.com/apprendre/19_variance_covariance.html
"Plus la covariance est faible et plus les séries sont indépendantes et inversement plus elle est élevée et plus les séries sont liées". Ceci est radicalement faux : la covariance ne mesure que la dépendance linéaire. Deux séries peuvent être absolument liées (par exemple X et X^2, ou bien X et $\sqrt{X}$, etc. : quand on connaît l'une, on connaît l'autre) et avoir une covariance nulle. Voir notre livre MPPR2, page 128, pour un exemple probabiliste très simple.

On trouvera en Annexe II un exemple frappant de deux séries temporelles totalement liées l'une à l'autre : quand on connaît x_k on peut calculer y_k et inversement ; pourtant leur corrélation linéaire est nulle.

VI. Références

FEVRIER 2017 LES ETF : CARACTERISTIQUES, ETAT DES LIEUX ET ANALYSE DES RISQUES - LE CAS DU MARCHE FRANÇAIS
document AMF
http://www.amf-france.org/technique/multimedia?docId=workspace://SpacesStore/2d61ede7-b0be-40fa-8654-fe438a33ad00_fr_1.0_rendition

Etude sur la réplication statistique des fonds de couverture, Fady Sallit, HEC Montréal, 2008.
http://biblos.hec.ca/biblio/memoires/m2008no102.pdf

Barclay Hedge : SG CTA Index (formerly Newedge CTA Index)
https://www.barclayhedge.com/research/indices/calyon/

SG Prime Services Indices
https://cib.societegenerale.com/en/prime-services-indices/

Wikipedia : Gestion alternative
https://fr.wikipedia.org/wiki/Gestion_alternative

C. Livres

Rémy Estran, Etienne Harb, Iryna Veryzhenko : Gestion de portefeuille, Dunod, 2017

Bernard Beauzamy : Introduction to Banach spaces and their geometry. Second edition. North-Holland Mathematics Studies, 68. Notas de Matemática, 86. North-Holland Publishing Co., Amsterdam, 1985.

Bernard Beauzamy : Méthodes Probabilistes pour l'étude des phénomènes réels. Ouvrage édité et commercialisé par la Société de Calcul Mathématique SA, ISBN 2-9521458-0-6, ISSN 1767-1175. Mars 2004. Seconde Edition, 2016.

Bernard Beauzamy : Nouvelles Méthodes Probabilistes pour l'évaluation des risques. Ouvrage édité et commercialisé par la Société de Calcul Mathématique SA. ISBN 978-2-9521458-4-8. ISSN 1767-1175, avril 2010.

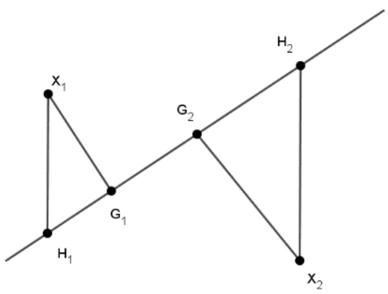
Olga Zeydina et Bernard Beauzamy : Probabilistic Information Transfer. Ouvrage édité et commercialisé par la Société de Calcul Mathématique SA. ISBN: 978-2-9521458-6-2, ISSN : 1767-1175. Relié, 208 pages, mai 2013.

Annexe I

Deux remarques préliminaires à propos de l'ajustement linéaire

Remarquons d'abord que l'on utilise la distance euclidienne (avec des carrés). Pourquoi celle-là ? Il en existe bien d'autres (voir le livre [BB_Banach]). La distance euclidienne est très commode, mais elle accentue les grandes valeurs au détriment des petites (du fait du carré). Par exemple, si on cherche la distance de l'origine au point de coordonnées $(5, 0.1)$ on trouve $d = \sqrt{25 + 0.01} \approx 5.00099$. Mais supposons que l'on se soit trompé d'un facteur 100% sur la seconde coordonnée, qui est en réalité 0.02 ; la distance ainsi corrigée sera $d = \sqrt{25 + 0.04} \approx 5.0039$, soit une variation de $0.06\,\%$: c'est infime. Une erreur majeure sur la plus petite des deux coordonnées ne se voit pas.

Plus grave, la distance est calculée en parallèle avec l'axe Oy. Sur le graphe ci-contre, on minimise la somme des longueurs X_1H_1 et X_2H_2 (au carré) et non pas la vraie distance à la droite, qui est X_1G_1 et X_2G_2 respectivement. Cela fait peu de différence si la droite obtenue est quasiment horizontale, mais cela fait une énorme différence si la droite a une pente très forte. En d'autres termes, si le nuage de points est "très incliné", l'ajustement au sens des moindres carrés est tout simplement faux. On ne trouvera pas un seul article, dans les disciplines appliquées, où cette particularité est mentionnée !



Annexe II

Coefficient de corrélation linéaire et indépendance

Voici un exemple très simple de deux séries temporelles, complètement dépendantes l'une de l'autre (à chaque instant je puis calculer x_k à partir de y_k et réciproquement), mais dont le coefficient de corrélation linéaire est nul.

Considérons, sur l'intervalle $\left[0,\frac{\pi}{2}\right]$, les deux fonctions :

$$f(x)=x-\alpha$$

$$g(x)=\cos(x)$$

On a :

$$\int_0^{\pi/2} f(x)dx=\frac{\pi^2}{8}-\frac{\alpha\pi}{2}$$

$$\int_0^{\pi/2} g(x)dx=1$$

$$\int_0^{\pi/2} f(x)g(x)dx=\frac{\pi}{2}-\alpha-1$$

Par conséquent, si l'on choisit pour α la valeur $\alpha=\frac{1}{4}\frac{\pi^2-4\pi+8}{\pi-2}\approx 1.161...$ on aura :

$$\int_0^{\pi/2} f(x)g(x)dx=\left(\int_0^{\pi/2} f(x)dx\right)\times\left(\int_0^{\pi/2} g(x)dx\right)$$

Fabriquons maintenant des séries temporelles en discrétisant les fonctions ; posons :

$$x_k=\frac{\pi}{2}f\left(\frac{k}{1000}\frac{\pi}{2}\right)$$

$$y_k=\frac{\pi}{2}g\left(\frac{k}{1000}\frac{\pi}{2}\right)$$

$$z_k = \frac{\pi}{2} f\left(\frac{k}{1000}\frac{\pi}{2}\right) g\left(\frac{k}{1000}\frac{\pi}{2}\right)$$

On a :

$$\frac{1}{1000}\sum_{k=1}^{1000} x_k \approx \int_0^{\pi/2} f(x)\,dx$$
$$\frac{1}{1000}\sum_{k=1}^{1000} y_k \approx \int_0^{\pi/2} g(x)\,dx$$

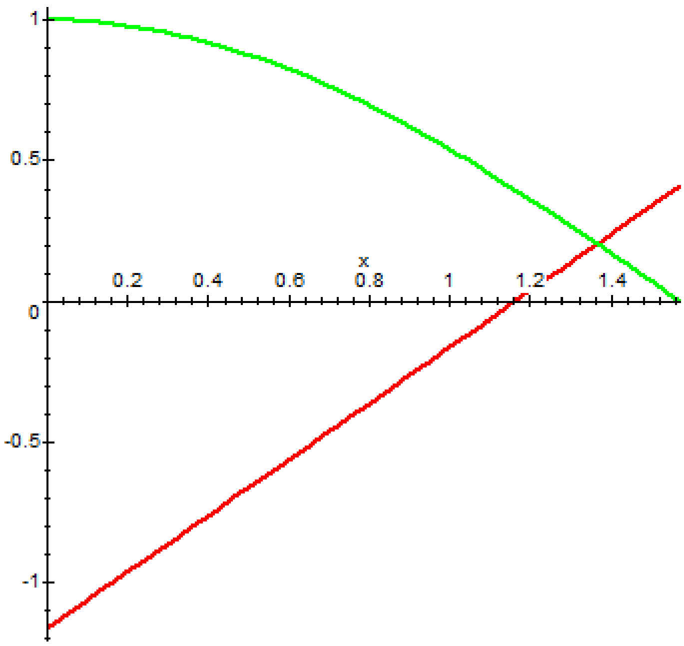
$$\frac{1}{1000}\sum_{k=1}^{1000} z_k \approx \int_0^{\pi/2} f(x)\,g(x)\,dx$$

Considérant les x_k comme des réalisations d'une variable aléatoire X, les y_k de Y et les z_k de Z , on a donc :

$$E(XY) = \frac{1}{1000}\sum_{k=1}^{1000} z_k \approx \left(\frac{1}{1000}\sum_{k=1}^{1000} x_k\right)\left(\frac{1}{1000}\sum_{k=1}^{1000} y_k\right) = E(X)E(Y)$$

ce qui prouve que le coefficient de corrélation linéaire est nul.

Or les deux fonctions f, g sont évidemment dépendantes : si je connais $y = f(x)$, je connais $x = y + \alpha$ et je connais $g(x) = \cos(y+\alpha)$, et inversement, puisque le cosinus est inversible sur le segment $\left[0, \frac{\pi}{2}\right]$. Les deux séries temporelles représentent la même chose, et pourtant leur corrélation linéaire est nulle.



Graphe des deux fonctions