

CONCENTRATION ET DISPERSION : ANALYSE DES SERIES CHRONOLOGIQUES A FORTE VARIABILITE*

Anne-Valère Haenni Amo[†]

Claude Tricot[‡]

Résumé

Le point de départ de cette recherche est la théorie des prix de Benoît Mandelbrot qui a détrôné la théorie des fluctuations régies par le hasard de Bachelier basée sur le mouvement brownien, en faveur de la classe des processus α -stables. En effet, si l'on considère des trajectoires boursières, la loi estimant la densité d'échantillon s'avère ne pas être normale mais stable sans variance. Dans cette optique, la plupart des mesures de dispersion n'ont plus de sens parce qu'elles ont en commun de se fonder sur une notion de distance et dépendent trop de ce fait des caractéristiques de l'échantillon. Nous introduisons, à cet effet, un indice de concentration emprunté à la géographie (Tricot (1971), Tricot & Raffestin (1979)) pour le convertir en mesure de dispersion en jouant sur la complémentarité potentielle entre concentration et dispersion. Cet indice présente l'avantage d'être obtenu à partir d'une modélisation convenable et de se fonder sur des probabilités.

Mots clés : loi α -stable, variance infinie, chroniques épaisses, volatilité, mesure de dispersion, indice de concentration

Introduction

En qualifiant la loi normale de "reine des statistiques", Fuchs (1996) résume sa place dans la littérature scientifique. Il serait difficile de la détrôner puisque la plupart des outils statistiques sont développés sur la base d'une hypothèse de normalité. Elle présente l'atout majeur d'offrir un formalisme extrêmement puissant, de bénéficier de la propriété de stabilité (pour décrire un phénomène, toutes les fonctions de distributions sont de même type pour toutes les tailles d'intervalle de temps) et enfin, l'argument souvent avancé pour justifier l'utilisation des statistiques gaussiennes est le théorème central limite. Ce résultat fondamental de la théorie des probabilités postule la normalité pour une variable dès que celle-ci résulte d'un grand nombre de petits phénomènes additifs indépendants.

Pour certains théoriciens dont l'instigateur fut Bachelier (1900), il a paru naturel de supposer que les rendements des titres financiers étaient des variables normales dans la mesure où l'on pouvait considérer les variations de prix comme le fruit de flux continus d'informations et d'ordres d'achat et de vente survenant de façon aléatoire sur le marché. Cette propriété de

* Cet article est basé sur une thèse de doctorat (Haenni Amo (1998)) soutenue au Département d'Econométrie de la Faculté des sciences économiques et sociales de l'Université de Genève, Suisse

[†] Anne-Valère Haenni Amo, docteur ès sciences économiques et sociales, mention économétrie et statistique, occupe actuellement les fonctions d'analyste quantitatif à l'Union Bancaire Privée.

[‡] Claude Tricot est Professeur Honoraire de Statistique à la Faculté des sciences économiques et sociales de l'Université de Genève

non-prévisibilité du comportement des marchés est ce que l'on appelle *efficience des marchés*. Cette formalisation, connu sous le nom de "*mouvement brownien*", de "*processus de Wiener*" ou de "*processus de Wiener-Lévy*", est fondée sur trois hypothèses : la normalité, l'indépendance et la stationnarité. Ainsi, initialement, la loi normale fut, de façon légitime, pressentie comme le candidat idéal pour modéliser le comportement des changements de cours boursiers et nous offrait la chance de progresser en terrain connu et conquis.

Jusqu'ici aucun modèle théorique basé sur la loi normale ne s'est vu empiriquement vérifié a posteriori (Mandelbrot (1963b), Mandelbrot (1967), Fama (1965)). En effet, les queues de distribution d'une loi normale ne rendent pas compte de façon suffisamment réaliste de la survenance de variations de grandes amplitudes sur les marchés financiers. En d'autres termes, les fluctuations boursières sont plus sévères que celles décrites par la distribution de Gauss. Cette dernière attribue une probabilité dérisoire aux "chocs" et aux "catastrophes" puisque ses queues décroissent selon une loi exponentielle. On peut donc la considérer comme une première approche de modélisation. La formalisation brownienne de la théorie des fluctuations de Bachelier n'est qu'une forme bénigne du hasard. Selon Benoît Mandelbrot, la notion du hasard est multiforme. Dans la nature, on ne rencontre pas seulement le hasard bénin, en l'occurrence, le hasard financier lui, est *sauvage* ou *malin*. Mandelbrot définit un tel hasard en faisant référence à deux récits bibliques. Un hasard sauvage se diagnostique par les symptômes suivants : lorsqu'il manifeste un *Effet de Noé*, c'est-à-dire la présence non négligeable de valeurs exceptionnelles et un *Effet de Joseph*, c'est-à-dire lorsqu'il incorpore une interdépendance statistique à caractère pseudo-périodique.

Face à ces symptômes de non-normalité, l'une des solutions de rechange a été de recourir aux lois dites α -*stables*, ou de façon équivalente, de *Pareto-Lévy*. On doit cette nouvelle interprétation du comportement des marchés à Mandelbrot (1963b), inspiré d'une part, par la loi sur les revenus personnels développée par Pareto en 1897 et d'autre part, par la stabilité des variables aléatoires telle que définie par Lévy (Lévy (1954)). En effet, ces distributions permettent de formaliser les effets de leptokurticité ("queues de distribution épaisses"), et de conserver les propriétés de stabilité par addition et de convergence de sommes normées vers une loi limite. Elles rendent fidèlement compte des risques extrêmes. Cette classe de distribution est d'autant plus intéressante que la loi normale en est un cas particulier et qu'elle offre une version généralisée du théorème central limite. Les queues décroissent selon une loi de puissance, elles décroissent si lentement que la variance, mesurant l'amplitude des fluctuations, est infinie. Cette divergence est à l'origine de la réticence des financiers à recourir à ces lois et explique pourquoi la proposition de Mandelbrot a eu du mal à s'imposer. En effet, la variance empirique est l'instrument fétiche de bon nombre d'économistes pour exprimer la notion de volatilité.

On soulignera que d'autres approches ont été proposées pour modéliser les fluctuations financières. Nous citerons tout d'abord le t-student (Blattberg & Gonedes (1974)) qui est une distribution leptokurtique. On peut également imaginer un processus défini à partir d'un processus à trajectoires continues et d'un processus de Poisson. La présence d'un processus de Poisson, appelé processus de saut, permet, par ses trajectoires discontinues, de modéliser la survenance d'événements qualifiés de "*rare*s" au cours du temps. Un tel processus permet ainsi d'incorporer des variations brutales de cours boursiers (Roger (1991)). Une approche économétrique du problème a également permis d'exprimer l'idée d'une variance variable dans le temps. Il s'agit des modèles ARCH (Autoregressive Conditional Heteroskedasticity) de Engel (1982) et GARCH (Generalized ARCH), une généralisation due à Bollerslev (Bollerslev (1986a)),

Bollerslev (1986b)). Ces modèles conditionnellement gaussiens produisent des queues épaisses. Nous mentionnerons aussi les mélanges discrets ou infinis de lois normales (Jewel (1982)) et enfin, le modèle de DuMouchel (1973) basé sur le mélange d'une distribution normale et d'une distribution stable symétrique dans les proportions p et $1-p$. Le modèle stable permet d'engendrer les queues leptokurtiques alors que la bosse centrale serait le résultat du modèle gaussien.

La plupart de ces modélisations alternatives permettent de reproduire l'effet de leptokurticité mais présentent toutefois le désavantage de ne pas être additivement stables (invariantes par addition), une propriété indispensable à la théorie financière moderne, les rendements étant additifs par changement d'échelle.

Très peu de travaux se sont cependant penchés sur la question de la mesure de la volatilité financière pouvant se substituer à la variance empirique (Copeland & Weston (1983)). Notre recherche a pour but d'analyser le comportement des chroniques financières épaisses en l'absence de variance à l'aide des lois stables et d'introduire une nouvelle mesure de dispersion évaluée non pas à partir de l'échantillon mais du modèle estimé.

A cet effet, le papier s'articulera en deux parties :

Les lois stables non-gaussiennes ont des distributions de probabilité dont certains moments n'existent pas toujours, formalisant de façon réaliste la variabilité extrême des marchés financiers. Son utilisation soulève donc un problème non négligeable, à savoir la caractérisation de la volatilité. Le but de la section 1 est de proposer une mesure de dispersion pouvant se substituer aux mesures usuelles, dont la variance, dans le cas notamment de distributions à queues épaisses. Dans la section 2, les différentes notions théoriques introduites à la section précédente sont appliquées à un portefeuille de titres dont la diversification nous permettra par la suite d'effectuer une étude comparative de la volatilité des titres.

1 Dispersion, concentration et risque financier

Les lois α -stables non-gaussiennes décrivent de façon plus réaliste les fluctuations de grande amplitude sur les marchés financiers que ne le fait une loi normale, en leur affectant une probabilité plus élevée. Par ailleurs, un recours aux lois stables non-gaussiennes permet de conserver l'une des vertus de la distribution de Laplace-Gauss, à savoir la stabilité des rentabilités boursières : à un changement d'échelle près, les fluctuations sont toutes décrites par la même loi (Mandelbrot (1963b)).

Cependant, à partir du moment où l'on renonce à la loi normale, on renonce simultanément à un formalisme puissant. En effet, comme les lois stables non-gaussiennes décroissent lentement pour rendre compte du phénomène de "queues épaisses" observé sur les marchés, l'écart-type qui mesure l'amplitude des fluctuations, devient infini. Cette divergence est perturbante lorsque l'on sait que l'écart-type est la notion fondamentale pour exprimer la propension d'une valeur boursière à accomplir des mouvements de cours plus amples que les variations moyennes du marché, ce que l'on appelle la *volatilité*. Une volatilité infinie représente donc un obstacle intellectuel.

Ainsi, comme l'utilisation de modèles à variance infinie rend la plupart des mesures de dispersion non pertinentes pour traduire la notion de volatilité, on introduit un indice de concentration emprunté à la géographie (Tricot (1971), Tricot & Raffestin (1979), Tricot (1987)).

1.1 Un indice de concentration de la géographie

La concentration n'est pas une grandeur à mesurer mais à repérer au moyen d'un indice qui permet des comparaisons. Si l'on admet la complémentarité entre les notions de concentration et de dispersion, une concentration minimale devrait correspondre à une dispersion maximale, alors qu'une concentration maximale devrait correspondre à une dispersion minimale.

1.1.1 Définition

Les données en géographie sont des données par régions. L'indice que nous avons emprunté à Tricot (1971) permet de mesurer la concentration d'une population sur un territoire E en recouvrant le territoire par une grille de mailles égales. Une première approche de la concentration s'exprime par la proportion de mailles vides. Tricot & Raffestin (1979) montre que cette concentration est donnée pratiquement par l'indice

$$T = \sum_{\substack{i=1 \\ d_i \leq d}}^K \left(\frac{s_i}{S} - \frac{n_i}{N} \right), \quad (1)$$

où S exprime la surface du territoire E comprenant N habitants. d désigne la *densité moyenne*, $d = \frac{N}{S}$. Si l'on suppose que le territoire est découpé en K régions,

- s_i est la surface de la région i ,
- n_i est le nombre d'habitants dans la région i ,
- d_i est la densité de la région i , $d_i = \frac{n_i}{s_i}$,

où $i = 1, \dots, K$.

Par conséquent, les relations suivantes sont évidentes :

$$\sum_{i=1}^K s_i = S \quad \text{et} \quad \sum_{i=1}^K n_i = N.$$

On distinguera trois types de régions :

1. si $d_i = d$, la région sera dite de concentration normale,
2. si $d_i > d$, la région sera dite de forte concentration,
3. si $d_i < d$, la région sera dite de faible concentration.

On verra facilement que l'indice T peut se définir de façon équivalente

$$T = \frac{1}{2} \sum_{i=1}^K \left| \frac{s_i}{S} - \frac{n_i}{N} \right|.$$

Les propriétés suivantes sont établies à partir de cette définition alternative de l'indice.

1.1.2 Propriétés de l'indice

Propriété 1 L'indice de concentration T est compris entre 0 et 1.

Propriété 2 $T = 1$ si et seulement si $\left| \frac{n_i}{N} - \frac{s_i}{S} \right| = \frac{n_i}{N} + \frac{s_i}{S}$ pour tout i

La condition $\left| \frac{n_i}{N} - \frac{s_i}{S} \right| = \frac{n_i}{N} + \frac{s_i}{S}$ pour tout i peut s'exprimer de façon alternative. En effet, on peut réarranger l'ordre des termes en valeur absolue que l'on additionne pour définir l'indice T de sorte que :

- pour $i = 1, \dots, k$, $\frac{n_i}{N} \geq \frac{s_i}{S}$. Alors,

$$\left| \frac{n_i}{N} - \frac{s_i}{S} \right| = \frac{n_i}{N} - \frac{s_i}{S} = \frac{n_i}{N} + \frac{s_i}{S},$$

ce qui est le cas si et seulement si

$$s_i = 0, \quad n_i \neq 0, \quad i = 1, \dots, k;$$

- pour $i = k + 1, \dots, K$, $\frac{n_i}{N} < \frac{s_i}{S}$. Alors,

$$\left| \frac{n_i}{N} - \frac{s_i}{S} \right| = \frac{s_i}{S} - \frac{n_i}{N} = \frac{s_i}{S} + \frac{n_i}{N},$$

si et seulement si

$$s_i \neq 0, \quad n_i = 0, \quad i = k + 1, \dots, K.$$

Ainsi, cette condition exprime le fait que si une région contient des habitants ($n_i \neq 0$), sa surface doit être nulle ($s_i = 0$). Par contre, si sa surface n'est pas nulle ($s_i \neq 0$), elle ne doit pas contenir d'habitants ($n_i = 0$).

En d'autres termes, cette propriété exprime le fait que l'indice T prend sa valeur maximale $T = 1$ si et seulement si la population est concentrée en quelques points de surface nulle qui sont des régions.

Propriété 3 $T = 0$ si et seulement si $\frac{n_i}{N} = \frac{s_i}{S}$ pour tout i . Le nombre d'habitants de chaque région est proportionnel à la surface de la région. On parlera de population à répartition homogène dans la contrée.

1.1.3 Généralisation

Une généralisation est faite par Tricot & Raffestin (1979) en remarquant que s_i est une mesure de la région i et n_i une autre mesure de la région i . Ainsi, l'indice est généralisable à tout ensemble, par exemple un ensemble probabilisé, ce en choisissant deux mesures.

Soit $\mu(\cdot)$ une mesure de probabilité. Soit $s(\cdot)$ la mesure de longueur ordinaire. On considère ces deux mesures sur un ensemble E . On suppose l'existence d'une densité, $\frac{d\mu}{ds}$. La densité

moyenne sur E sera notée $\frac{\mu(E)}{s(E)}$. On appellera E_1 le sous-ensemble de E tel que $\frac{d\mu}{ds} \leq \frac{\mu(E)}{s(E)}$.

En généralisant (1), on peut donc écrire que la concentration de la distribution de probabilité sur E est donnée par :

$$T = \int_{E_1} \left(\frac{ds}{s(E)} - \frac{d\mu}{\mu(E)} \right) = \frac{s(E_1)}{s(E)} - \frac{\mu(E_1)}{\mu(E)}. \quad (2)$$

Le tableau suivant résume cette adaptation de l'indice de la géographie à la statistique :

Géographie	Statistique
Territoire E	Ensemble E
N	$\mu(E)$
S	$s(E)$
$d = \frac{N}{S}$	$\frac{\mu(E)}{s(E)}$
n_i	$d\mu$
s_i	ds
$d_i = \frac{n_i}{s_i}$	$\frac{d\mu}{ds}$
$E_1 : d_i \leq d$	$E_1 : \frac{d\mu}{ds} \leq \frac{\mu(E)}{s(E)}$
$T = \sum_{d_i \leq d} \left(\frac{s_i}{S} - \frac{n_i}{N} \right)$	$T = \int_{E_1} \left(\frac{ds}{s(E)} - \frac{d\mu}{\mu(E)} \right)$

TAB. 1 - L'indice T comme indice statistique

E est l'ensemble fondamental, à savoir l'ensemble des valeurs admissibles pour une variable aléatoire univariée X . Il s'agit d'un sous-ensemble de \mathbb{R} , à savoir d'un intervalle $[A, B]$ tel que $A < B$. $\mu([A, x])$ n'est rien d'autre que la fonction de répartition de notre variable aléatoire, que l'on note habituellement $F_X(x)$. Ainsi, $\mu(E)$ est toujours unitaire, $\mu(\cdot)$ étant une mesure de probabilité. $s(\cdot)$ est la mesure de Lebesgue, à savoir la longueur de l'intervalle considéré : $s(E) = |B - A|$. Enfin, $\frac{d\mu}{ds}$ est la fonction de densité de X , notée habituellement $f_X(x)$. Par conséquent, E_1 correspond aux valeurs prises par X dont les fréquences relatives sont inférieures à la borne $\frac{1}{|B - A|}$. En résumé,

$$\begin{aligned}
 E &= [A, B] \\
 \mu(E) &= 1 \\
 s(E) &= |B - A| \\
 E_1 &= \left\{ x \in [A, B] \mid f_X(x) \leq \frac{1}{|B - A|} \right\} \\
 T &= \frac{s(E_1)}{|B - A|} - \mu(E_1)
 \end{aligned}$$

1.1.4 Un indice de dispersion

Dispersion et concentration sont complémentaires dans le cas d'une distribution unimodale. On donnera ici une explication quelque peu intuitive.

Lorsque la distribution de probabilité considérée est unimodale, on ne distingue qu'un seul pôle de concentration. Cette concentration des probabilités se situant dans un voisinage du mode, on s'imagine aisément qu'en s'en écartant, la concentration décroît en faveur de la dispersion.

A présent, si l'on considère une distribution de probabilité bimodale, on est en présence de deux pôles de concentration. Plus ces pôles sont éloignés l'un de l'autre plus la dispersion est forte, alors que la concentration reste la même.

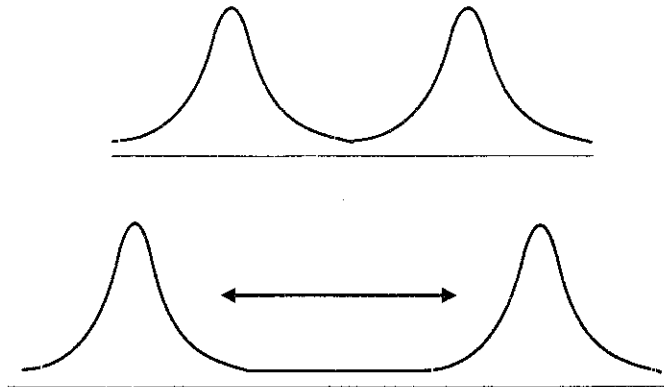


FIG 1 – Concentration et dispersion dans le cas d'une distribution bimodale

Par conséquent, dans le cas d'une distribution unimodale, l'indice $1 - T$ nous permet de mesurer la dispersion d'une variable aléatoire univariée X sur un intervalle $[A, B]$.

1.2 Calcul de l'indice T dans le cas d'une loi α -stable symétrique

1.2.1 Estimation de la densité d'une loi stable

Par la suite, on cherchera à évaluer l'indice de dispersion $1 - T$ dans le contexte des marchés financiers. A cet effet, la distribution de probabilité considérée est une loi stable non-gaussienne.

qui présente l'inconvénient de ne pas toujours avoir de forme analytique. Ce manque de formalisme rend toute manipulation plus complexe d'autant plus que l'évaluation numérique de ces fonctions de densité reste une question ouverte (Ripley & Venables (1994)) Cependant, la connaissance de quatre coefficients α, β, C et δ ¹ permet de les paramétrer de façon à les approcher numériquement au moyen de l'outil informatique (Haenni Amo (1998)). En effet, on en obtient une estimation numérique par application directe du théorème d'inversion de sa fonction caractéristique correspondante, notée $\varphi_X(t)$:

$$f_X(x) = \begin{cases} \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp(-itx) \varphi_X(t) dt \\ \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp(-itx - C^\alpha |t|^\alpha) dt \quad \text{si } \beta = \delta = 0. \end{cases}$$

La densité stable "théorique" est estimée par la série convergente suivante, en se basant sur les sommes de Riemann :

$$\hat{f}_X(\hat{\alpha}, \hat{\beta}, \hat{C}, \hat{\delta}; x_i) = \frac{1}{2\pi} \sum_{t=-L}^{+L} e^{-itx_i} \varphi_X(\hat{\alpha}, \hat{\beta}, \hat{C}, \hat{\delta}; t) \Delta t, \quad (3)$$

où $\hat{\alpha}, \hat{\beta}, \hat{C}$ et $\hat{\delta}$ sont les paramètres estimés à partir d'un échantillon donné. On obtient des résultats très concluants à partir de $L = 500$ et $\Delta t = 1$.

1.2.2 Calcul de l'indice

Cette section a pour but d'adapter l'indice de concentration T au cas d'une distribution stable symétrique $\beta = 0$ ². Ainsi, on suppose que $X \sim S_{\alpha,0}(C, \delta)$.

Supposer une variable aléatoire stable implique que l'ensemble de ses valeurs admissibles soit \mathbb{R} tout entier et non pas un intervalle. Il faut donc au préalable définir l'ensemble E sur lequel est calculé l'indice.

¹Les lois α -stables univariées sont spécifiées à partir de quatre paramètres :

- α est l'*exposant caractéristique* ou l'*indice de stabilité* qui n'est rien d'autre qu'un coefficient d'aplatissement (kurtosis), aussi bien pour le mode que pour les queues :

$$0 < \alpha \leq 2,$$

- β est le *paramètre d'asymétrie*, mesurant le degré d'asymétrie (skewness) de la distribution :

$$-1 \leq \beta \leq 1,$$

($\beta = -1$, $\beta = 0$ et $\beta = 1$ expriment respectivement une distribution totalement asymétrique à gauche, symétrique et totalement asymétrique à droite),

- C est le *paramètre d'échelle* auquel on fait jouer le rôle d'une mesure de dispersion quelquefois, $C \geq 0$,
- δ est le *paramètre de localisation* ou de *translation*, traduisant la position de la moyenne lorsque $\alpha > 1$, $\delta \in \mathbb{R}$.

On notera $X \sim S_{\alpha,\beta}(C, \delta)$ une variable aléatoire X distribuée selon une loi stable de paramètres α, β, C et δ (Samorodnitsky & Taqqu (1994))

Lorsque $\alpha = 2$, $\beta = 0$, $C = \sigma/\sqrt{2}$ et $\delta = \mu$, la loi α -stable n'est rien d'autre que la loi normale $N(\mu, \sigma^2)$.

²**Propriété 4** $X \sim S_{\alpha,\beta}(C, \delta)$ est symétrique si et seulement si $\beta = 0$ et $\delta = 0$. X est symétrique par rapport à δ si et seulement si $\beta = 0$ (Samorodnitsky & Taqqu (1994)).

Dans un premier temps, on construit un intervalle $[A, B]$ contenant les réalisations de notre variable aléatoire X . On notera cet échantillon donné (x_1, \dots, x_n) . A ce stade, $\mu(E) = F_X(B) - F_X(A)$ est presque unitaire. En d'autres termes, l'aire sous la densité de probabilité est proche de 1 sur l'intervalle $[A, B]$, les queues étant négligées.

On opéra pour $A \leq \min_i \{x_i\}$ et $B \geq \max_i \{x_i\}$, sachant que l'échantillon est concentré sur l'intervalle $[x_{\min}, x_{\max}]$

Ainsi, $E = [A, B] \supseteq [x_{\min}, x_{\max}]$

Dans un deuxième temps, on estime les paramètres α , C et δ à partir de l'échantillon considéré en ayant recours à la méthode proposée par McCulloch (1986).

La densité de probabilité étant considérée exclusivement sur $[A, B]$ et $\mu(\cdot)$ devant être unitaire, on effectue la modification suivante sur la fonction de densité :

$$f_X(\hat{\alpha}, 0, \hat{C}, \hat{\delta}; x) \longrightarrow \frac{f_X(\hat{\alpha}, 0, \hat{C}, \hat{\delta}; x)}{F_X(B) - F_X(A)} = f(x),$$

où $x \in [A, B]$ et $f(x)$ représente la densité tronquée sur l'intervalle $[A, B]$.

Par conséquent, la fonction de répartition correspondante s'écrit :

$$F(x) = \int_A^x \frac{f_X(\hat{\alpha}, 0, \hat{C}, \hat{\delta}; t)}{F_X(B) - F_X(A)} dt = \frac{F_X(x) - F_X(A)}{F_X(B) - F_X(A)}$$

A présent, $F(A) = 0$ et $F(B) = \mu(E) = \mu([A, B]) = 1$.

A ce stade, la forme en cloche des distributions α -stables symétriques nous permet aisément de définir l'ensemble E_1 comme l'union de deux intervalles :

$$E_1 = \left\{ x \in [A, B] \mid f(x) \leq \frac{1}{|B - A|} \right\} = [A, c] \cup [d, B] \quad \text{où } c, d \in \mathbb{R}$$

Quant aux points x sur l'intervalle $[A, B]$ dont la densité α -stable est égale à la quantité $\frac{1}{|B - A|}$, ils se comptent forcément au nombre de 0, 1 ou 2. En effet, l'ensemble

$$\left\{ x \in [A, B] \mid f(x) = \frac{1}{|B - A|} \right\}$$

s'identifie respectivement

- à l'ensemble vide lorsque $\max_i \{f(x_i)\} < \frac{1}{|B - A|}$,
- au mode lorsque $\max_i \{f(x_i)\} = \frac{1}{|B - A|}$,
- à deux points c et d tels que $A \leq c < d \leq B$, lorsque $\max_i \{f(x_i)\} > \frac{1}{|B - A|}$

Dans les deux premiers cas, $E_1 = E = [A, B]$ et l'indice de concentration est nul. Dans le troisième cas, $E_1 = [A, c] \cup [d, B]$. Il nous reste alors à calculer $\mu(E_1)$ pour connaître l'indice dans ce dernier cas de figure.

A partir des paramètres estimés $\hat{\alpha}$, \hat{C} et $\hat{\delta}$, on peut obtenir la valeur de $\mu(E_1)$ correspondante en utilisant la fonction de répartition d'une loi α -stable tronquée, notée F . On évalue cette dernière aux points c et d :

$$\begin{aligned}
\mu(E_1) &= F(c) + (1 - F(d)) \\
&= \frac{F_X(c) - F_X(A)}{F_X(B) - F_X(A)} + \left(1 - \frac{F_X(d) - F_X(A)}{F_X(B) - F_X(A)}\right) \\
&= 1 - \frac{F_X(d) - F_X(c)}{F_X(B) - F_X(A)}
\end{aligned}$$

$\mu(E_1)$ peut être aisément évalué en recourant à une table de quantiles stables centrés et réduits (Nolan (1997)).

Enfin, disposant à ce stade de tous les éléments nécessaires à l'évaluation de notre indice de concentration, on peut écrire

$$T = \frac{|c - A| + |B - d|}{|B - A|} - \mu(E_1).$$

Illustration

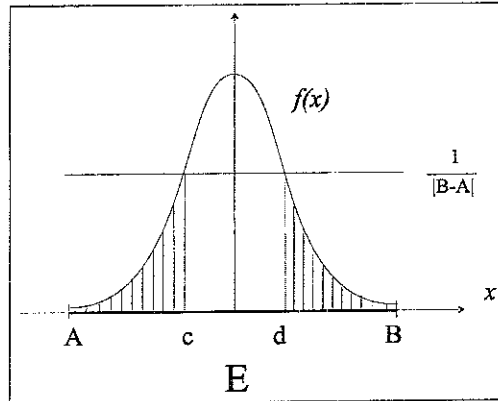


FIG. 2 – Calcul de l'indice pour une loi α -stable symétrique.

$$\text{où } E = [A, B]$$

$$E_1 = [A, c] \cup [d, B]$$

$$s(E_1) = |c - A| + |B - d|$$

La zone hachurée correspond à $\mu(E_1)$, alors que $\mu(E) = 1$.

2 Applications

On cherche ici à effectuer une étude comparative de la volatilité de différents titres sur la base de différentes mesures de dispersion, notamment l'indice $1 - T$.

2.1 La base de données

La base de données est composée des prix de clôture journaliers $(P_t)_{t=0..M}$ de l'or et de différentes actions suisses, françaises, anglaises, américaines et de Hong Kong entre le 16 avril 1991 et le 18 février 1998³ en prenant soin d'évacuer les jours fériés dans chaque pays. A partir de ces cours, on déduit les chroniques de rendements logarithmiques $(R_{t-1,t}) = (\log P_t - \log P_{t-1}) = (X_i)_{t=i=1..M}$.

Nous disposons au total de dix échantillons dont les caractéristiques sont réunies dans le tableau suivant :

CODE	N	Société	Secteur Principal
ABB	1720	Abb AG	Electricité
NESTLE	1720	Nestle SA	Aliments et Boissons
ROCHE	1720	Roche Holding AG	Médicaments, Cosmétiques et Produits de Santé
OR	1733	-	-
BARCLAYS	1733	Barclays PLC	Finance
DANONE	1712	Group Danone SA	Aliments et Boissons
BOEING	1731	The Boeing Company	Aérospatial
MERL	1731	Merrill Lynch & Co., Inc.	Finance
NIKE	1731	Nike, Inc.	Habillement et Textiles
HUTCHISON	1698	Hutchison Whampoa Ltd.	Divers

TAB. 2 – Description de la base de données

2.2 Mesure de la dispersion

Dans un premier temps, pour l'ensemble des rentabilités boursières, on évalue numériquement l'indice de dispersion $1 - T$ ainsi que différentes mesures de dispersion usuelles telles que l'étendue, l'intervalle semi-interquartile, la semi-variance, l'écart moyen et la variance que l'on notera respectivement $R, ISI, sVar, EM$ et Var . En ce qui concerne l'intervalle semi-interquartile, il serait intéressant de comparer les valeurs numériques de cette mesure obtenues à partir de l'échantillon avec celles du modèle proposé ($X \sim S_X(\alpha, \beta, C, \delta)$), ce que l'on notera ISI_M . Dans un deuxième temps, on classe ces valeurs boursières dans l'ordre décroissant de la volatilité selon chaque mesure de dispersion.

³SOURCE : DATASTREAM.

1 Calcul de l'indice de dispersion $1 - T$ (On choisit $A = -30$ et $B = 30$.)

TITRES	$\min x_i$	$\max x_i$	$d = -c$	$\mu(E_1)$	$1 - T$
ABB	-15.68	6.12	3.22	5.74	0.165
NESTLE	-6.29	4.73	2.62	3.12	0.118
ROCHE	-8.34	8.11	2.62	4.34	0.131
OR	-3.36	3.30	1.81	3.59	0.096
BARCLAYS	-9.95	14.38	3.62	5.07	0.172
DANONE	-8.55	7.76	3.02	2.88	0.130
BOEING	-11.91	11.00	3.42	3.95	0.154
MERL	-12.30	10.38	4.23	4.43	0.185
NIKE	-14.47	11.43	4.03	5.22	0.187
HUTCHISON	-12.56	21.63	4.03	6.54	0.200

TAB 3 – Valeurs numériques de l'indice $1 - T$

2. Les mesures de dispersion

TITRES	\sqrt{Var}	EM	\sqrt{sVar}	ISI	ISI_M	R
ABB	1.55	1.07	1.03	0.765	0.764	21.80
NESTLE	1.00	0.75	0.61	0.565	0.565	11.01
ROCHE	1.09	0.72	0.69	0.480	0.479	16.45
OR	0.58	0.40	0.37	0.280	0.277	6.66
BARCLAYS	1.64	1.19	0.99	0.885	0.884	24.33
DANONE	1.29	0.98	0.76	0.780	0.778	16.31
BOEING	1.58	1.15	0.93	0.885	0.882	22.91
MERL	2.03	1.57	1.19	1.240	1.236	22.68
NIKE	2.01	1.47	1.21	1.110	1.107	25.89
HUTCHISON	2.23	1.54	1.35	1.100	1.097	34.19

TAB 4 – Les autres mesures de dispersion.

3 Classification des titres dans l'ordre décroissant de la volatilité

$1 - T$	\sqrt{Var}	EM	\sqrt{sVar}	ISI	ISI_M	R
HUICHISON	HUICHISON	MERL	HUICHISON	MERL	MERL	HUICHISON
NIKE	MERL	HUICHISON	NIKE	NIKE	NIKE	NIKE
MERL	NIKE	NIKE	MERL	HUICHISON	HUICHISON	BARCLAYS
BARCLAYS	BARCLAYS	BARCLAYS	ABB	BARCLAYS	BARCLAYS	BOEING
ABB	BOEING	BOEING	BARCLAYS	BOEING	BOEING	MERL
BOEING	ABB	ABB	BOEING	DANONE	DANONE	ABB
ROCHE	DANONE	DANONE	DANONE	ABB	ABB	ROCHE
DANONE	ROCHE	NESTLE	ROCHE	NESTLE	NESTLE	DANONE
NESTLE	NESTLE	ROCHE	NESTLE	ROCHE	ROCHE	NESTLE
OR	OR	OR	OR	OR	OR	OR

TAB. 5 – Classification des valeurs boursières dans l'ordre décroissant de la variabilité

En premier lieu, on ne manquera pas de remarquer dans le tableau 4 la proximité des valeurs numériques de l'intervalle semi-interquartile obtenues à partir de l'échantillon et du modèle α -stable, ce qui nous conforte dans l'hypothèse de modélisation.

Dans l'optique où l'on s'est placé, c'est-à-dire en proposant un modèle à variance infinie, les calculs de la variance, de la semi-variance, de l'écart moyen et de l'étendue n'ont pas de sens. Ces mesures ont un sens dans un contexte de statistique descriptive, elles seront en effet toujours définies à partir de l'échantillon. Par contre, dans le cadre qui nous intéresse, on peut remettre en doute leur qualité d'estimateur de la variabilité des cours.

Les deux seules mesures compatibles avec un modèle produisant des queues de distribution épaisses sont l'indice $1 - T$ et l'intervalle semi-interquartile. Mais, comment expliquer la différence dans leurs classifications respectives des titres ? L'indice $1 - T$ tient compte à la fois de la structure centrale et de l'épaisseur des queues. En effet, il peut se redéfinir comme la somme de deux quantités dont la première est représentative de la structure centrale (la distance $c - d$) et la seconde des queues ($\mu(E_1)$) :

$$1 - T = 1 - \frac{|c - A| + |B - d|}{|B - A|} + \mu(E_1) = \frac{|d - c|}{|B - A|} + \mu(E_1).$$

On peut alors imaginer deux cas de figure pour une dispersion élevée : tout d'abord, une distribution leptokurtique à la fois au niveau des queues et du mode (la distance $c-d$ est faible, alors que la probabilité concentrée dans les queues $\mu(E_1)$ est grande) puis, dans un deuxième temps, une distribution plus étalée (la distance $c-d$ est plus grande cette fois). On pourrait ainsi obtenir le même indice de deux manières différentes : à partir d'une distribution de queues épaisses à laquelle correspondrait une quantité $\mu(E_1)$ élevée ou, à partir d'une distribution plus étalée autour du mode à laquelle correspondrait cette fois une distance $c-d$ plus grande.

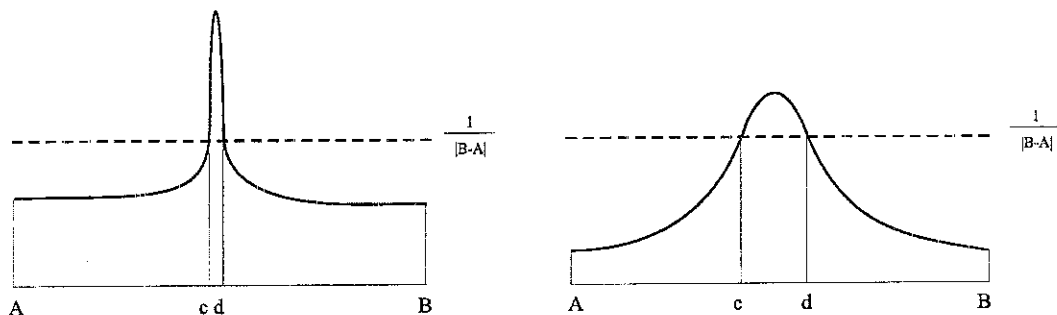


FIG. 3 - Indice $1 - T$

L'indice $1 - T$ tient donc compte à la fois du facteur "mode" et du facteur "queues" ce qui n'est pas le cas de l'intervalle semi-interquartile comme le montre l'exemple suivant. En effet, dans l'exemple caricatural ci-dessous, on peut remarquer que l'intervalle semi-interquartile, ne tient pas du tout compte de la région centrale.

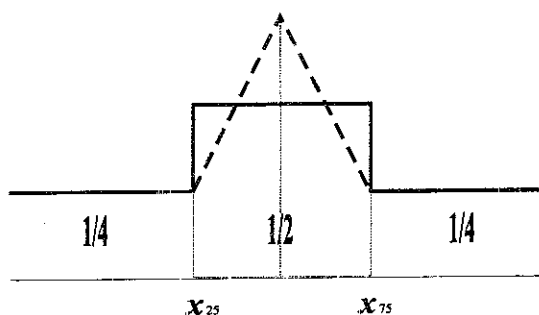


FIG. 4 - Intervalle semi-interquartile

Pour des queues identiques, les deux distributions ci-dessus représentent une dispersion différente au niveau du mode. Lorsque la bosse centrale prend la forme d'un triangle, on peut aisément imaginer que la dispersion est plus faible que pour un mode rectangulaire. Or, l'intervalle semi-interquartile ne permet pas de faire ressortir cette nuance et traduirait dans ce cas une variabilité identique.

Ainsi, si l'on admet que l'indice $1 - T$ fait foi, à partir du tableau 5, on remarquera que les titres ayant réalisé des performances diamétralement opposées sont HUTCHISON, coté sur la Bourse de Hong Kong (caractérisée, on le sait, par une forte volatilité depuis le krach de 1997), et l'or

3 Conclusion

Notre choix de modélisation impliquant l'inexistence du moment d'ordre 2, la variance empirique n'est plus pertinente pour mesurer la notion de volatilité, c'est-à-dire la propension d'un marché financier à la variabilité. En effet, la variance du modèle étant infinie, la variance d'échantillon considérée comme un estimateur ne représente rien. Le caractère erratique du deuxième moment empirique permet de douter à juste titre de la validité de toutes les mesures de dispersion dérivées comme la semi-variance d'échantillon. Il en est de même de l'écart moyen dans la mesure où le modèle n'admet pas non plus de moment d'ordre 1 si $\alpha \leq 1$, et plus généralement de l'étendue. On a introduit, à cet effet, un indice de concentration emprunté à la géographie que l'on a converti en indice de dispersion en jouant sur la complémentarité potentielle entre concentration et dispersion.

En résumé, le but général de ce travail de recherche était de traduire la notion de volatilité sur les marchés financiers. Or, si l'on considère une série chronologique prise comme un échantillon, la loi qui estime la densité d'échantillon s'avère ne pas être normale mais stable, sans variance. Dans ce contexte, si l'on se base uniquement sur un échantillon, les mesures de dispersion usuelles ne représentent rien car elles n'estiment rien. Elles ont en commun de se fonder sur une notion de distance et par conséquent, elles dépendent trop des caractéristiques de l'échantillon. L'indice $1 - T$ et l' ISI_M sont, en revanche, obtenus à partir d'une modélisation convenable et se fondent sur des probabilités. A travers des études empiriques, ces deux instruments de rechange se sont révélés efficaces. En effet, chacun offre une mesure pertinente de la notion de volatilité d'une valeur boursière, en ayant l'avantage, par rapport aux autres mesures de dispersion, d'être compatible avec une modélisation α -stable non-gaussienne. On mesurera donc la volatilité des cours boursiers par $1 - T$ ou l'intervalle semi-interquartile mesuré, non pas à partir de l'échantillon mais du modèle. On soulignera toutefois que le calcul de l'indice $1 - T$ fait intervenir à la fois la structure centrale et l'épaisseur des queues de la distribution théorique et se trouve donc être préférable à l' ISI_M .

Il reste à ajouter que le paramètre d'échelle C , pouvant parfois faire figure de mesure de dispersion, n'a pas été sélectionné pour classer les chroniques pour les raisons suivantes : C est proportionnel à l'écart-type dans le cas de la loi normale, or ici nous n'avons pas de variance ; C s'identifie à l'intervalle semi-interquartile pour la loi de Cauchy⁴, lequel s'avère moins bon que l'indice $1 - T$.

Enfin, la modélisation stable nous a servi d'une part, à mettre en évidence le caractère erratique du moment d'ordre 2 empirique et d'autre part, à calculer une nouvelle mesure de dispersion basée sur des probabilités aisément adaptable à n'importe quel domaine d'étude pour appréhender le problème de l'estimation du degré de variabilité d'un phénomène. Il a l'avantage de toujours être défini à partir du moment où l'on a opté pour une loi de probabilité spécifique.

Pour terminer, on soulignera qu'envisager le calcul de l'indice $1 - T$ dans le cas multivarié, dans le contexte de la théorie du portefeuille par exemple, semble une perspective de recherche intéressante. La théorie du portefeuille dans le cas "stable", notamment la mesure du risque d'un portefeuille stable, a, en effet, déjà fait l'objet de recherches fructueuses (Belkacem (1996), Belkacem & Lévy Véhel & Walter (1996)).

⁴La loi de Cauchy, comme la loi normale, est une loi α -stable particulière. Dans ce cas, $\alpha = 1$, $\beta = 0$, le paramètre δ s'identifie à la médiane et C à l'intervalle semi-interquartile

Références

- BACHELIER, L. (1900) Théorie de la spéculation. *Annales Scientifiques de l'Ecole Normale Supérieure*, III-17
- BELKACEM, L. (1996) *Processus stables et applications en finance*. Thèse de doctorat à l'Université Paris IX Dauphine.
- BELKACEM, L., LÉVY VÉHEL, J., WALTER, C. (1996) *CAPM, Risk and Portfolio selection in "Stable" Markets* Rapport de recherche 2776, INRIA
- BLATTBERG, R., GONNEDES, N. J. (1974) A Comparison of the Stable and Student Distributions as Statistical Models for Stock Prices. *Journal of Business*, 47 : 244-280
- BOLLERSLEV, T. (1986a) A Conditional Heteroskedasticity Time Series Model for Speculative Prices and Rates of Return. *Revue of Economics and Statistics*, 69 : 542-547.
- BOLLERSLEV, T. (1986b) Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 7 : 307-327
- COPELAND, T. E., WESTON, J. F. (1983) *Financial Theory and Corporate Policy*. Addison & Wesley, New York.
- DUMOUCHEL, W. H. (1973) Stable Distributions in Statistical Inference : 1. Symmetric Stable Distributions Compared to Other Symmetric Long-tailed Distributions. *Journal of the American Statistical Association*, 68.
- ENGEL, D. D. (1982) The Multiple Stochastic Integral. *Mem Amer Math Soc*, 28 : 209-245
- FAMA, E. F. (1965) The Behavior of Stock Market Prices. *Journal of Business*, 38.
- FUCHS, A. (1996) La loi normale, reine des statistiques. *Dossier pour la science*, hors série, avril.
- HAENNI AMO, A.-V. (1998) *Mesure de la volatilité des chroniques fractales*. Thèse de doctorat à l'Université de Genève (Département d'Econométrie).
- JEWEL, N. (1982) Mixtures of exponential distributions. *Annuaire Statistique*, 10 : 479-484.
- LÉVY, P. (1954) *Théorie de l'addition des variables aléatoires*. Gauthier-Villars, 2nd edn.
- MANDELBROT, B. B. (1963b) The Variation of Certain Speculative Prices. *Journal of Business*, 36.
- MANDELBROT, B. B. (1967) The Variation of Some Other Speculative Prices. *Journal of Business*, 40.
- MCCULLOCH, J. H. (1986) Simple Consistent Estimators of Stable Distribution Parameters. *Communications in Statistics-Computation and Simulation*, 15 : 1109-1136
- NOLAN, J. P. (1997) Numerical Calculation of Stable Densities and Distribution Functions. *Stochastic Models*, 13 : 759-774.
- RIPLEY, B. D., VENABLES, W. N. (1994) *Modern Applied Statistics with S-plus*. New York : Springer-Verlag
- ROGER, P. (1991) *Les outils de la modélisation financière*. Presses universitaires de France.
- SAMORODNITSKY, G., TAQQU, M. S. (1994) *Stable Non-Gaussian Random Processes : Stochastic Models with Infinite Variance*. New York : Chapman & Hall.
- TRICOT, C. (1971) Concentration d'un ensemble de points. *Cahiers du Département d'Econométrie, Faculté des SES, Université de Genève*.
- TRICOT, C., RAFFESTIN, C. (1979) Concentration et connexité ou Pour une approche du concept de territorialité, dans G. Gaudard, ed., colloques économiques 8/1 *Modèles et politiques de l'espace économique*. Fribourg : Editions Universitaires
- TRICOT, C. (1987) *Statistique spatiale*. Avignon : Coll. Brouillons Dupont, ed. Dupont