

INTRODUCTION GÉNÉRALE

Dans cette introduction, nous faisons un panorama de l'industrie de la gestion d'actifs. Après avoir présenté les différents acteurs (sociétés de gestion et investisseurs), nous nous intéressons à la réglementation de la gestion d'actifs, et plus particulièrement de la gestion collective. Nous abordons ensuite les différents styles de gestion, que nous pouvons caractériser par l'univers d'investissement et par la technique de gestion utilisée. La quatrième section est consacrée à la place de la gestion quantitative en asset management. Enfin, dans la dernière section, nous présentons le plan du livre.

L'industrie de la gestion d'actifs

La gestion d'actifs (en anglais *asset management*) représente au 31 décembre 2009 environ 52,6 billions de dollars¹ d'actifs sous gestion². Elle a largement bénéficié du rebond du marché, puisque le montant d'AuM avait fortement baissé en 2008 pour s'établir à 47 billions de dollars³, mais on est encore loin des 57 billions de dollars de l'année 2007. La

¹Nous rappelons qu'en français un billion est mille milliards (et correspond à un *trillion* en américain).

²Le terme anglais correspondant est *assets under management* (ou AuM).

³Boston Consulting Group, « In Search of Stable Growth », Global Asset Management Report, juillet 2010, www.bcg.com.

répartition des actifs sous gestion est la suivante. La gestion collective sous forme de fonds représente environ 23 \$bln, alors que les hedge funds et les ETF gèrent respectivement 1,6 \$bln et 1 \$bln. Les 27 \$bln restants correspondent à des mandats de gestion et à la gestion structurée.

Dans cette section, nous présentons respectivement les sociétés d'investissement et leurs clients. En particulier, nous faisons la distinction entre la gestion traditionnelle et la gestion hedge funds. Nous consacrons aussi un paragraphe aux agences de notation.

La gestion traditionnelle

La gestion traditionnelle est dominée par de très grands acteurs (*global players*). À côté des géants que sont BlackRock ou Pimco, on trouve une myriade de petites sociétés de gestion. Contrairement à ces géants qui couvrent toutes les classes d'actifs, ces boutiques sont spécialisées dans quelques stratégies.

Les grands acteurs

Sur le tableau 1, nous reportons le classement du journal *Pensions & Investments* concernant les plus grosses sociétés de gestion. On voit donc que les 5 plus grosses sociétés de gestion représentent environ 15% des actifs gérés et que les 20 premières sociétés gèrent environ 45% des actifs. Ce phénomène de concentration s'est encore accentué avec la crise. Ainsi, la fusion entre BlackRock et Barclays Global Investors (BGI) a donné naissance en 2009 à un mastodonte qui gère 3,36 billions de dollars au 31 mars 2010⁴. En France, l'année 2009⁵ a vu aussi la création d'Amundi, issue de la fusion de CAAM⁶ et SGAM⁷. Cette course au plus gros est dictée par une concurrence de plus en plus forte dans la gestion traditionnelle. Cela se traduit par une baisse générale des frais de gestion, ce qui a entraîné une diminution de la marge depuis 2006⁸. Afin de rétablir une rentabilité plus importante, les sociétés de gestion se sont lancées dans une vaste opération de réduction des coûts et d'industrialisation. Une première solution est d'automatiser la chaîne de traitement front-middle-back. Cela passe par une automatisation informatique de l'ensemble des processus. La deuxième solution est de gérer plus d'actifs afin de bénéficier des effets d'échelle.

⁴Source : www.blackrock.com.

⁵Cette année a été aussi l'occasion pour BNP Paribas Asset Management d'absorber Fortis AM.

⁶Crédit Agricole Asset Management.

⁷Société Générale Asset Management.

⁸Boston Consulting Group, « Conquering the Crisis », Global Asset Management Report, juillet 2009, www.bcg.com.

Tableau 1. *Les plus gros gérants au 31 décembre 2008 (en milliards de dollars)*

Rang	Nom	Pays	AuM
1	Barclays Global Investors	GB	1 516
2	Allianz Group	DE	1 462
3	State Street Global	US	1 444
4	Fidelity Investments	US	1 389
5	AXA Group	FR	1 383
6	BlackRock	US	1 307
7	Deutsche Bank	DE	1 150
8	Vanguard Group	US	1 145
9	J.P. Morgan Chase	US	1 136
10	Capital Group	US	975
11	Bank of New York Mellon	US	928
12	UBS	CH	821
13	BNP Paribas	FR	810
14	Goldman Sachs Group	US	798
15	ING Group	NL	777
16	Crédit Agricole	FR	776
17	HSBC Holdings	GB	735
18	Legg Mason	US	698
19	Natixis	FR	630
20	Wells Fargo	US	574

Source : The P&I/Watson Wyatt 500 (www.pionline.com)

Les boutiques

À côté de ces grosses sociétés de gestion, on trouve un certain nombre de boutiques. Ces boutiques sont généralement créées par un ancien gérant qui monte sa propre société de gestion. La logique de ces boutiques est très différente des précédentes. Elles sont généralement spécialisées sur un style de gestion, et leur succès dépend très fortement de la performance financière des fonds gérés. Sur le tableau 2, nous reportons le nombre de sociétés de gestion dans quelques pays européens⁹.

Nous remarquons qu'il existe de grandes différences entre les pays. Ainsi, il peut apparaître curieux que le nombre de sociétés de gestion en France soit beaucoup plus grand qu'au Royaume-Uni ou en Allemagne. En fait, les boutiques sont une spécialité très française. Parmi les plus célèbres, on peut citer La Française des Placements, Oddo AM, etc. Ces boutiques gèrent généralement moins de vingt milliards d'euros. Certaines sont ce-

⁹European Fund and Asset Management Association, « Asset Management in Europe – Facts and Figures », *EFAMA's Third Annual Review*, avril 2010, www.efama.org.

Nous reportons les statistiques d'actifs sous gestion (en milliards de dollars), du nombre de fonds ainsi que l'AuM moyen par fonds recueillies par l'Investment Company Institute³³ sur le tableau 4. Nous remarquons que, sur les 23 \$bln d'actifs collectés par les OPCVM dans le monde, près de la moitié sont aux États-Unis. Nous notons aussi une très grande différence entre l'AuM moyen d'un fonds aux États-Unis et dans le reste du monde.

Tableau 4. AuM et nombre d'OPCVM dans le monde

Région	AuM	(en %)	Nombre	(en %)	AuM moyen
Monde	22 964		65 735		349
Amérique	12 597	54,9%	16 982	25,8%	742
Europe	7 546	32,9%	33 054	50,3%	228
Asie et Pacifique	2 715	11,8%	14 795	22,5%	184
Afrique	106	0,5%	904	1,4%	118
États-Unis	11 121	48,4%	7 691	11,7%	1 446
Luxembourg	2 294	10,0%	9 017	13,7%	254
France	1 806	7,9%	7 982	12,1%	226
Australie	1 199	5,2%			
Irlande	861	3,7%	2 721	4,1%	316
Brésil	784	3,4%	4 744	7,2%	165
Royaume-Uni	729	3,2%	2 266	3,4%	322
Japon	661	2,9%	3 656	5,6%	181
Canada	565	2,5%	2 075	3,2%	272
Chine	381	1,7%	547	0,8%	697
Allemagne	318	1,4%	2 067	3,1%	154
Italie	279	1,2%	675	1,0%	414
Espagne	270	1,2%	2 588	3,9%	104
Corée du Sud	265	1,2%	8 703	13,2%	30

Source : Investment Company Institute, 2010.

Les mandats discrétionnaires

La gestion sous mandat contractualise une relation bilatérale entre le gérant et l'investisseur. Elle se distingue de la gestion collective puisque c'est une gestion non publique et concerne donc principalement les investisseurs institutionnels. L'exemple typique est le fonds de pension qui cherche à répliquer un indice d'actions ou d'obligations. Celui-ci va lancer un appel d'offre (*request for proposal* ou RFP). Aidé d'un consultant, le

³³Investment Company Institute, *Investment Company Fact Book — A Review of Trends and Activity in the Investment Company Industry*, 50th edition, 2010, www.icifactbook.org.

fonds de pension effectue une première sélection de gérants (*short list*). Dans un deuxième temps, une étude approfondie des gérants sélectionnés conduira le fonds de pension à choisir un gérant (ou plusieurs si le notionnel est gros).

Par définition, il est très difficile d'obtenir des statistiques sur ces mandats discrétionnaires puisqu'ils ne sont pas publics. Néanmoins, on estime qu'en Europe les mandats discrétionnaires représentent environ 50% des actifs sous gestion fin 2008 (EFAMA, 2010). Ces mandats peuvent aussi être complétés par des services de conseil. Ainsi, BlackRock Solutions, filiale de BlackRock spécialisée dans la gestion des risques, l'allocation stratégique, etc., a une clientèle représentant 9 billions de dollars.

Les ETF

Les trackers ou exchange traded funds (ETF) sont des fonds qui répliquent la performance d'un indice et qui sont négociables en bourse. Contrairement aux OPCVM classiques qui offrent une liquidité journalière, les ETF sont négociés en continu. Pour cela, la société de gestion a recours à un (ou plusieurs) market maker qui va assurer la liquidité intra-journalière. Avant l'ouverture de la bourse, le market maker reçoit l'inventaire du portefeuille, et s'engage ensuite à maintenir des prix d'achat et de vente avec un écart maximum et une quantité minimum dans des conditions normales de marché.

Les ETF sont des produits relativement récents. Le premier ETF a été lancé par State Street en 1993. Ce tracker appelé Standard and Poor's Deposit Receipt (SPDR que l'on prononce "spider") réplique la performance du S&P 500 et il est aujourd'hui le plus gros ETF avec 68 milliards de dollars d'actifs sous gestion au 1er juillet 2010. En France, il faut attendre le 15 janvier 2001 pour que Lyxor Asset Management lance le premier tracker sur la bourse de Paris. Le marché des ETF est beaucoup plus petit que celui des fonds traditionnels, puisqu'il représente environ un billion de dollars en 2010³⁴. Néanmoins, c'est un marché qui connaît une très forte croissance avec des encours qui augmentent de 30% en moyenne par an. Outre les trackers classiques sur les indices d'actions ou d'obligations, on trouve désormais un vaste univers d'investissement allant des indices de matières premières à des indices de style en passant par des indices sectoriels. Depuis quelques années, on voit aussi apparaître des ETF sur indices de stratégie³⁵ et des ETF actifs³⁶.

³⁴On estime que fin 2009 il y a un peu plus de 1900 ETF dans le monde. *Source* : BlackRock.

³⁵Par exemple, des stratégies optionnelles de type *covered call*.

³⁶Ce sont des ETF d'allocation dynamique entre classes d'actifs ou des ETF structurés intégrant une protection de capital, des effets de levier ou donnant la performance inverse de l'indice.

Les véhicules d'investissement de gestion alternative

De par la nature des stratégies mises en place, les réglementations européennes et américaines sont peu adaptées aux hedge funds. Ceci explique que les hedge funds vont chercher des véhicules dans des pays beaucoup moins réglementés afin de limiter les impacts réglementaires sur la gestion du fonds. À côté des zones *offshore* traditionnelles, les hedge funds utilisent aussi des véhicules des zones *onshore* qui présentent une fiscalité alléger. De par leur flexibilité absolue, les fonds Caïmans sont les véhicules les plus répandus. « *Ils ne connaissent aucune restriction quant aux investissements sous-jacents. Tout type de placement est permis, y compris des placements immobiliers et illiquides* » (Lecocq, 2008)³⁷. Ces fonds peuvent être des fonds communs de placement ou structurés sous la forme de sociétés. Dans tous les cas, une dizaine de jours suffit pour obtenir l'agrément et créer le fonds. En Europe, Guernesey, Jersey et Malte offrent des véhicules proches de ceux des îles Caïmans. Depuis quelques années, ces véhicules d'investissement sont concurrencés par des véhicules plus réglementés. C'est notamment le cas avec les fonds irlandais, luxembourgeois et suisses. Par exemple, l'Irlande offre deux types de véhicules. Pour les fonds d'investisseurs professionnels (*professional investor fund* ou PIF), la souscription minimum est de 125 000 euros alors que celle-ci est relevée à 250 000 euros pour les fonds d'investisseurs qualifiés (*qualifying investor fund* ou QIF). Ces derniers plus souples ont la préférence des hedge funds. En 2007, le Luxembourg a créé les fonds d'investissement spécialisés (SIF). Depuis quelque temps, des *hedge funds* de très grande notoriété (BlueCrest, Man AHL, Winton) proposent d'investir dans leurs stratégies via des véhicules UCITS³⁸.

Remarque 2. *La frontière entre gestion traditionnelle et gestion alternative est moins importante aujourd'hui que par le passé. Tout d'abord, on note une convergence entre ces deux gestions du point de vue des techniques de gestion. À partir de 2005 sont apparus des fonds long only actions gérés par des hedge funds en utilisant leur moteur de stock picking implémenté dans leur stratégie long/short. En 2006, les sociétés de gestion traditionnelle ont lancé des fonds actions appelés 130/30. Ces fonds sont construits sur une exposition acheteuse de 130% couverte en partie par une exposition vendeuse de 30%. On peut voir ces fonds comme un fonds actions tilté par une stratégie equity market neutral. Ensuite, on note que les hedge funds se tournent de plus en plus vers des formats contrôlés (plateformes de comptes gérés, newcits, etc.).*

³⁷D. Lecocq, « Quelle juridiction pour les hedge funds ? », *Banque & Finance*, mai-juin 2008, p. 46-47.

³⁸On utilise le terme Newcits pour désigner ce type de fonds.

liste des lauréats du prix Graham et Dodd qui récompense chaque année les meilleurs articles du FAJ⁷⁸. On y trouve bien sûr Fisher Black et ses co-auteurs (parmi lesquels Emmanuel Derman et Robert Litterman) mais aussi Michael Brennan, John Campbell, Hayne Leland, Andrew Lo, Robert Merton, Richard Roll, Stephen Ross, Mark Rubinstein, William Sharpe, etc. On retrouve ce même souci d'éducation professionnelle dans la revue *Journal of Portfolio Management*. Celle-ci est créée en 1974 par Peter Bernstein dont l'idée originelle est de fournir aux professionnels l'état de l'art des méthodes financières. Édité aujourd'hui par Frank Fabozzi, *Journal of Portfolio Management* continue à diffuser au sein de la profession les outils et méthodes indispensables pour faire de la gestion d'actif. C'est notamment dans cette revue que seront publiés la régression de style de Sharpe, la loi fondamentale de la gestion active de Grinold ou encore les stratégies de rebalancement de Mark Rubinstein⁷⁹.

Comme nous venons de le voir, la frontière entre le monde académique et le monde professionnel est moins prononcée qu'en Europe. C'est donc tout naturel que les professeurs de finance des universités américaines aient une activité professionnelle. Tous les noms prestigieux ont été à un moment de leur vie impliqués dans la création d'un fonds, dans l'élaboration d'une stratégie ou même dans la construction d'un business. Ainsi, Fisher Black et Myron Scholes sont à l'origine du fonds Stagecoach Fund chez Wells Fargo au début des années soixante-dix, Hayne Leland et Mark Rubinstein mettront en place l'assurance de portefeuille dans la société de gestion LOR, William Sharpe fait du conseil dans diverses sociétés de gestion dont Wells Fargo Investment Advisors et Frank Russell Company, Andrew Lo crée sa propre société de gestion AlphaSimplex, etc. Parmi ces différentes expériences, certains parcours ont changé l'histoire de l'asset management.

Au début des années soixante-dix, Barr Rosenberg est Professeur de finance à l'Université de Californie à Berkeley. C'est un jeune économètre qui a fait sa thèse en 1968 à Harvard University. Son sujet porte sur l'estimation des paramètres d'un modèle linéaire lorsque ceux-ci sont aléatoires. Celui-ci a déjà des publications prestigieuses dans *Journal of the American Statistical Association* ou *Biometrika*⁸⁰ lorsqu'il s'intéresse à la prévision des risques communs et spécifiques des actions⁸¹. En 1974, il crée la société de conseil Barr Rosenberg Associates (BARRA) qui a pour objectif de développer la gestion de portefeuille auprès des institutionnels. En par-

⁷⁸Celle-ci est disponible à l'adresse www.cfapubs.org/page/AllpastGrahamAndDoddWinners?journalCode=faj.

⁷⁹P.L. Bernstein et F.J. Fabozzi, *Streetwise: The Best of the Journal of Portfolio Management*, Princeton University Press, 1998.

⁸⁰B. Rosenberg, « Linear Regression with Randomly Dispersed Parameters », *Biometrika*, 60(1), 1973, p. 65-72.

⁸¹B. Rosenberg et W. McKibben, « The Prediction of Systematic and Specific Risk in Common Stocks », *Journal of Financial and Quantitative Analysis*, 8, 1973, p. 317-333.

ticulier, il développe le modèle multi-factoriel BARRA⁸² qui est une alternative au modèle APT. Alors que ce dernier s'appuie sur des facteurs statistiques, le modèle BARRA utilise des facteurs économiques. BARRA est un succès énorme très utilisé par les fonds de pension et les investisseurs institutionnels. Rosenberg quitte BARRA en 1985⁸³ pour créer RIEM (Rosenberg Institutional Equity Management). La philosophie de gestion consiste à combiner les modèles de portefeuille à la BARRA avec des modèles statistiques de valorisation des actions. En 1999, RIEM est racheté par AXA IM et devient AXA Rosenberg. Cette société gère 41 milliards de dollars au 31 mai 2010.

Barr Rosenberg est l'un des pionniers à quitter le monde académique pour se lancer dans la gestion quantitative⁸⁴. Il est vite suivi par d'autres universitaires. L'histoire de BGI est racontée dans le chapitre 10 de Bernstein (2007). Celle-ci débute avec le lancement du premier fonds indicel par Wells Fargo Investment Advisors. D'autres innovations suivront (comme le fonds Stagecoach Fund), mais la rentabilité n'est pas au rendez-vous. En 1990, Wells Fargo vend finalement sa société de gestion à la banque Barclays qui la renomme Barclays Global Investors. Il faut encore quelques années pour que ces solutions innovantes soient acceptées et que BGI devienne la première société de gestion au monde. Bernstein (2007) rapporte les propos suivants de Blake Grossman, le CEO de BGI⁸⁵ :

« *It was an evangelical undertaking* Grossman declared as he looked back over the whole story from the struggles of the 1970s onward. *Indexing and computer-driven strategies based on theoretical models are still the core of our work. They color everything we do.* » (Bernstein, 2007, page 147).

Dans cette entreprise d'évangélisation, Grossman est aidé par Richard Grinold et Ronald Kahn. Grinold rejoint BGI en 1993 pour être responsable de la recherche. Auparavant, il a été Professeur à l'Université de Californie à Berkeley et a travaillé chez BARRA, tout comme Kahn, qui a un Ph.D. en physique de Harvard et qui rejoint BGI en 1998. En 1995, Grinold et Kahn publient un livre intitulé *Active Portfolio Management: Quantitative Theory and Applications*, qui synthétise leurs travaux de recherche chez BARRA. On remarque que la première référence du livre est un article sur le problème des données et l'estimation robuste d'un modèle paru

⁸²B. Rosenberg, « Extra-Market Components of Covariance in Security Returns », *Journal of Financial and Quantitative Analysis*, 9, 1974, p. 263-274.

⁸³En 2004, MSCI rachète BARRA pour offrir une solution intégrée avec d'une part les indices et d'autre part les modèles factoriels d'optimisation.

⁸⁴Le lecteur qui désire connaître de façon plus détaillée l'histoire de Barr Rosenberg peut consulter le chapitre 3 de Mintz *et al.* (1998).

⁸⁵Il est intéressant de noter que Grossman a été l'assistant de recherche de William Sharpe à Stanford. Le monde de la finance américaine a décidément été façonné par quelques personnalités que sont Harry Markowitz, William Sharpe, Fisher Black, etc.

Le ratio Kappa défini par Kaplan et Knowles (2004) est¹⁵ :

$$\text{Kappa}_m(\theta) = \frac{\mu - \theta}{\sqrt[n]{\text{lp}_m(\theta)}}$$

On remarque que $\text{Sortino}(\theta) = \text{Kappa}_2(\theta)$. Une autre mesure de performance a été proposée par Keating et Shadwick (2002)¹⁶ :

$$\text{Omega}(\theta) = \frac{\int_{\theta}^{\infty} (1 - \mathbf{F}(x)) \, dx}{\int_{-\infty}^{\theta} \mathbf{F}(x) \, dx}$$

Kaplan et Knowles (2004) ont montré que nous avons la relation suivante $\text{Omega}(\theta) = \text{Kappa}_1(\theta) + 1$. La grande difficulté avec ces ratios est de définir la bonne valeur du seuil θ , qui est perçu comme le rendement minimum acceptable de l'investisseur (*minimum acceptable return*). C'est donc un paramètre endogène propre à chaque investisseur. C'est pourquoi on utilise généralement deux mesures universelles de seuil : $\theta = 0$ et $\theta = r$.

Remarque 7. *Les ratios précédents sont très utilisés dans l'industrie des hedge funds à cause du caractère non gaussien des rendements des stratégies alternatives. Dans la pratique, ces ratios donnent des résultats relativement proches du ratio de Sharpe¹⁷.*

5.4. Un exemple

À titre d'illustration, nous considérons la stratégie (S) de covered call BXM qui est étudiée à la page 479. Nous supposons que le benchmark (B) de cette stratégie est l'indice S&P 500. Nous reportons sur la page suivante à quoi pourrait ressembler un reporting¹⁸. Sur le tableau 9, nous indiquons les performances mensuelles (en %) de la stratégie. Nous indiquons aussi le rendement et la volatilité (en %) de la stratégie et du benchmark pour les différentes années calendaires. Nous pouvons bien sûr enrichir ce reporting en incluant une attribution de performance, une analyse temporelle de la corrélation, etc.

¹⁵P.D. Kaplan et J.A. Knowles, « Kappa: A Generalized Downside Risk-Adjusted Performance Measure », *Journal of Performance Measurement*, 8(3), 2004, p. 42-54.

¹⁶C. Keating et W.F. Shadwick, « A Universal Performance Measure », *Journal of Performance Measurement*, 6(3), 2002, p. 59-84.

¹⁷M. Eling et F. Schuhmacher, « Does the Choice of Performance Measure Influence the Evaluation of Hedge Funds », *Journal of Banking and Finance*, 31(9), 2007, p. 2632-2647.

¹⁸BM et WM correspondent aux mois qui présentent le meilleur rendement (*best month*) et le plus mauvais rendement (*worst month*).

Tableau 8. Statistiques (01/01/1990 – 31/12/2009)

	S		B	
Rendement	9,1%		8,2%	
Volatilité	13,0%		18,6%	
Sharpe	0,37		0,21	
TE	8,7%			
IR	0,09			
Alpha	0,8%			

	1%		5%	
	S	B	S	B
1D	-9,0%	-9,0%	Max	-40,1%
1W	-18,0%	-18,3%	Début	26/12/07
1M	-24,5%	-29,7%	Fin	09/03/09
6M	-33,1%	-40,7%		
1Y	-35,5%	-46,3%		

	1%		5%	
	S	B	S	B
1D	2,4%	3,1%	1,2%	1,7%
1W	5,4%	6,4%	2,4%	3,6%
1M	10,9%	13,3%	5,0%	6,9%
6M	19,9%	24,3%	8,1%	11,9%
1Y	29,4%	39,0%	16,1%	25,0%

	S	B
Skewness	-0,32	0,00
Kurtosis	26,71	12,72
BM	8,2%	11,4%
	(Mars 09)	(Déc. 91)
WM	-15,0%	-16,8%
	(Oct. 08)	(Oct. 08)

ou encore :

$$\min_{q \in \mathbb{R}} \sum_i \chi_\alpha(y_i - q)$$

avec $\chi_\alpha(u) = u \times (\alpha - \mathbf{1}\{u < 0\})$. Si on considère maintenant le modèle linéaire $Y = X\beta + U$ et que l'on applique la définition précédente à la variable aléatoire U , l'estimateur $\hat{\beta}_\alpha$ de la régression quantile pour le niveau α est :

$$\hat{\beta}_\alpha = \arg \min \sum_i \chi_\alpha(y_i - x_i^\top \beta)$$

Notons que pour $\alpha = 50\%$, la régression quantile est connue sous le nom de régression médiane et nous avons :

$$\hat{\beta}_{50\%} = \arg \min \sum_i |y_i - x_i^\top \beta|$$

Cela revient à minimiser la valeur absolue des résidus. Cette estimation est aussi appelée régression linéaire robuste.

1.3.2. Écriture du problème sous forme LP

Nous avons :

$$\begin{aligned} y_i &= x_i^\top \beta + u_i \\ &= \sum_{k=1}^K x_{i,k} (\beta_k^+ - \beta_k^-) + u_i^+ - u_i^- \end{aligned}$$

Nous obtenons le programme d'optimisation LP :

$$\begin{aligned} z^* &= \arg \min c^\top z \\ \text{s.c.} &\begin{cases} Az = b \\ z \geq 0 \end{cases} \end{aligned}$$

avec Y le vecteur des endogènes, X la matrice des exogènes, z le vecteur $(\beta^+ \quad \beta^- \quad u^+ \quad u^-)^\top$, $A = (X \quad -X \quad I \quad -I)$, $b = Y$ et $c = (0 \quad 0 \quad \alpha \mathbf{1} \quad (1 - \alpha) \mathbf{1}^-)^\top$. Pour résoudre ce problème LP, Portnoy et Koenker (1997) proposent d'utiliser la méthode des points intérieurs².

1.3.3. Extension à l'estimation non paramétrique

Dans le cas d'une seule variable exogène, nous pouvons modifier la méthode précédente pour construire une estimation non paramétrique de la fonction quantile conditionnelle $Q(\alpha | X = x)$. Pour cela, nous introduisons les poids w_i dans la fonction d'optimisation :

$$\min_{q \in \mathbb{R}} \sum_i w_i \chi_\alpha(y_i - (\beta_0(x) + \beta_1(x)(x_i - x)))$$

²S. Portnoy et R. Koenker, « The Gaussian Hare and the Laplacian Tortoise: Computability of Squared-Error versus Absolute-Error Estimators », *Statistical Science*, 12, 1997, p. 279-296.

avec :

$$w_i = \mathcal{K} \left(\frac{x - x_i}{h} \right)$$

et \mathcal{K} un noyau non paramétrique. Nous avons :

$$\hat{Q}(\alpha | X = x) = \hat{\beta}_0$$

$\hat{\beta}_0$ est l'estimation non paramétrique du modèle linéaire local. De même, nous pouvons utiliser un modèle quadratique :

$$\min_{q \in \mathbb{R}} \sum_i w_i \chi_q \left(y_i - \left(\beta_0(x) + \beta_1(x)(x_i - x) + \beta_2(x)(x_i - x)^2 \right) \right)$$

et définir $\hat{\beta}_0$ comme l'estimation non paramétrique du modèle quadratique local³.

1.3.4. Application à l'estimation du skew bêta

On considère le modèle linéaire :

$$R_t^S = \beta R_t^B + u_t$$

où R_t^S et R_t^B sont les rendements de la stratégie et du benchmark. On cherche à estimer la fonction $\beta(r) = Q(\alpha | R_t^B = r)$. $\beta(r)$ est le β -rendement de la stratégie expliqué par le benchmark. On considère les indices investissables HFRX de hedge funds calculés par HFR⁴. On calcule la fonction $\beta(r)$ en utilisant le modèle quadratique local de régression linéaire avec l'indice S&P 500 comme benchmark. Cette fonction est représentée sur le graphique 1 pour les stratégies equity hedge (HFRX EH) et equity market neutral (HFRX EMN) et pour l'indice global (HFRX). On remarque que le bêta n'est pas constant et qu'il est plus élevé lorsque les rendements du S&P 500 sont très négatifs dans le cas de la stratégie equity hedge et de l'indice global. L'estimation d'un bêta non conditionnel peut donc être biaisée. Par contre, on vérifie que la stratégie equity market neutral ne présente pas de composante directionnelle significative.

2. La programmation quadratique

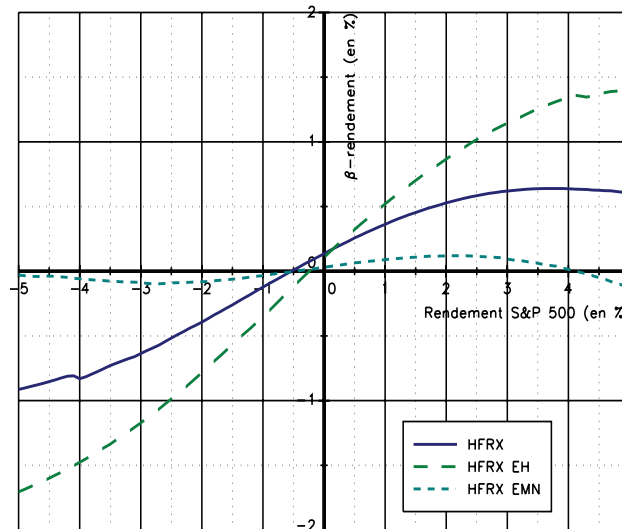
2.1. Spécification d'un programme quadratique

2.1.1. Définition

Un programme quadratique (QP) est un problème d'optimisation dont la fonction objectif est quadratique et dont les contraintes sont linéaires.

³Cette partie est développée à la page 297.

⁴La société Hedge Fund Research, Inc. est spécialisée dans les bases de données de hedge funds. Tous les mois, elle publie des indices non investissables qui reflètent la performance globale de l'industrie de la gestion alternative funds. Contrairement à ces indices appelés HFRI, les indices HFRX sont des indices journaliers et investissables. Source : www.hedgefundresearch.com.

Graphique 1. Estimation de la fonction $\beta(r)$ 

On cherche donc la solution x^* telle que :

$$x^* = \arg \min \frac{1}{2} x^\top Q x - x^\top R$$

sous les contraintes :

$$Sx \leq T$$

On remarque que ce système de contraintes permet de traiter aussi le cas de contraintes linéaires d'égalité $Ax = B$ puisque cela revient à considérer deux inéquations $Ax \geq B$ et $Ax \leq B$:

$$\begin{bmatrix} -A \\ A \end{bmatrix} x \leq \begin{bmatrix} -B \\ B \end{bmatrix}$$

De la même façon, nous pouvons considérer un espace borné $x^- \leq x \leq x^+$ puisque nous avons :

$$\begin{bmatrix} -I \\ I \end{bmatrix} x \leq \begin{bmatrix} -x^- \\ x^+ \end{bmatrix}$$

C'est pourquoi la plupart des logiciels numériques préfèrent traiter le système suivant plus facile à appréhender pour l'utilisateur⁵ :

$$\begin{aligned} Ax &= B \\ Cx &\leq D \\ x^- &\leq x \leq x^+ \end{aligned}$$

⁵C'est par exemple le cas de la procédure `Qprog` du logiciel Gauss.

particulier, on note que les coefficients des facteurs growth valent respectivement $-94,4\%$ (indice large) et $92,6\%$ (indice smid). Si on somme ces coefficients, on obtient donc une exposition globale au facteur growth très faible, ce qui indique qu'il y a sûrement un problème de multi-colinéarité dans la régression MCO.

2.3. Application au portefeuille de variance minimale

2.3.1. Résultats théoriques

Considérons un portefeuille composé de n actifs. Nous notons Σ la matrice de covariance des rendements. Le problème du portefeuille de variance minimale est le suivant :

$$\begin{aligned} x^* &= \arg \min x^\top \Sigma x \\ \text{s.c. } & \mathbf{1}^\top x = 1 \text{ et } \mathbf{0} \leq x \leq \mathbf{1} \end{aligned}$$

On cherche donc le portefeuille qui présente la variance la plus faible sous la contrainte que les poids soient positifs et que la somme des poids soit égale à 1.

Considérons tout d'abord le cas où l'on n'impose pas la contrainte $\mathbf{0} \leq x \leq \mathbf{1}$. Le lagrangien du problème d'optimisation est¹² :

$$f(x; \lambda_0) = \frac{1}{2} x^\top \Sigma x - \lambda_0 (\mathbf{1}^\top x - 1)$$

La solution x^* vérifie le système suivant des conditions de premier ordre :

$$\begin{cases} \partial_x f(x; \lambda_0) = \Sigma x - \lambda_0 \mathbf{1} = 0 \\ \partial_{\lambda_0} f(x; \lambda_0) = \mathbf{1}^\top x - 1 = 0 \end{cases}$$

Nous obtenons $x = \lambda_0 \Sigma^{-1} \mathbf{1}$. Comme $\mathbf{1}^\top x - 1 = 0$, nous avons $\lambda_0 = 1 / (\mathbf{1}^\top \Sigma^{-1} \mathbf{1})$. La solution est donc :

$$x^* = \frac{1}{\mathbf{1}^\top \Sigma^{-1} \mathbf{1}} \Sigma^{-1} \mathbf{1}$$

Nous remarquons que la contribution marginale en risque est :

$$\frac{\partial \sigma(x)}{\partial x} = \frac{\Sigma x}{\sqrt{x^\top \Sigma x}} = \frac{\lambda_0 \mathbf{1}}{\sqrt{x^\top \Sigma x}}$$

La solution du problème de variance minimale sans contraintes de *short-selling* vérifie donc les conditions suivantes :

$$\frac{\partial \sigma(x)}{\partial x_i} = \frac{\partial \sigma(x)}{\partial x_j} \quad \forall i, j$$

¹²Pour des raisons d'écriture, nous considérons la fonction avec un coefficient d'échelle de $1/2$, mais cela ne change pas la solution du problème.

Introduisons maintenant la contrainte de *no short-selling*. Le lagrangien devient :

$$f(x; \lambda_0) = \frac{1}{2}x^\top \Sigma x - \lambda_0 (\mathbf{1}^\top x - 1) - \lambda^\top x$$

La solution x^* vérifie le nouveau système des conditions de premier ordre :

$$\begin{cases} \partial_x f(x; \lambda_0) = \Sigma x - \lambda_0 \mathbf{1} - \lambda = 0 \\ \partial_{\lambda_0} f(x; \lambda_0) = \mathbf{1}^\top x - 1 = 0 \end{cases}$$

ainsi que les conditions de Kuhn-Tucker :

$$\min(\lambda_i, x_i) = 0$$

Nous avons donc $x = \lambda_0 \Sigma^{-1} \mathbf{1} + \Sigma^{-1} \lambda$. Il n'est pas possible de définir la solution analytique. Néanmoins, nous avons :

$$\frac{\partial \sigma(x)}{\partial x} = \frac{\Sigma x}{\sqrt{x^\top \Sigma x}} = \frac{\lambda_0 \mathbf{1} + \lambda}{\sqrt{x^\top \Sigma x}}$$

À l'optimum, si $x_i = 0$ alors $\lambda_i > 0$ et si $x_i > 0$ alors $\lambda_i = 0$. La solution du problème de variance minimale avec des contraintes de *no short-selling* vérifie donc les conditions suivantes :

$$\frac{\partial \sigma(x)}{\partial x_i} = \frac{\partial \sigma(x)}{\partial x_j} = \frac{\lambda_0}{\sqrt{x^\top \Sigma x}} \quad \text{si } x_i \neq 0 \text{ et } x_j \neq 0$$

et :

$$\frac{\partial \sigma(x)}{\partial x_i} = \frac{\lambda_0 + \lambda_i}{\sqrt{x^\top \Sigma x}} \quad \text{si } x_i = 0$$

À l'optimum, les contributions marginales en risque sont donc égales pour tous les actifs qui composent le portefeuille de variance minimale (c'est-à-dire ceux qui ont un poids strictement positif).

Nous pouvons dériver l'expression analytique de la solution optimale dans le cas où toutes les corrélations sont égales. Notons $\Sigma = \sigma \sigma^\top \odot R$ la matrice de covariance avec $R = C_n(\rho)$ la matrice de corrélation constante¹³. Nous avons $\Sigma^{-1} = \Gamma \odot R^{-1}$ avec $\Gamma_{i,j} = (\sigma_i \sigma_j)^{-1}$ et :

$$R^{-1} = \frac{\rho \mathbf{1} \mathbf{1}^\top - ((n-1)\rho + 1)I}{(n-1)\rho^2 - (n-2)\rho - 1}$$

Comme $x = (\Sigma^{-1} \mathbf{1}) / (\mathbf{1}^\top \Sigma^{-1} \mathbf{1})$, nous en déduisons que¹⁴ :

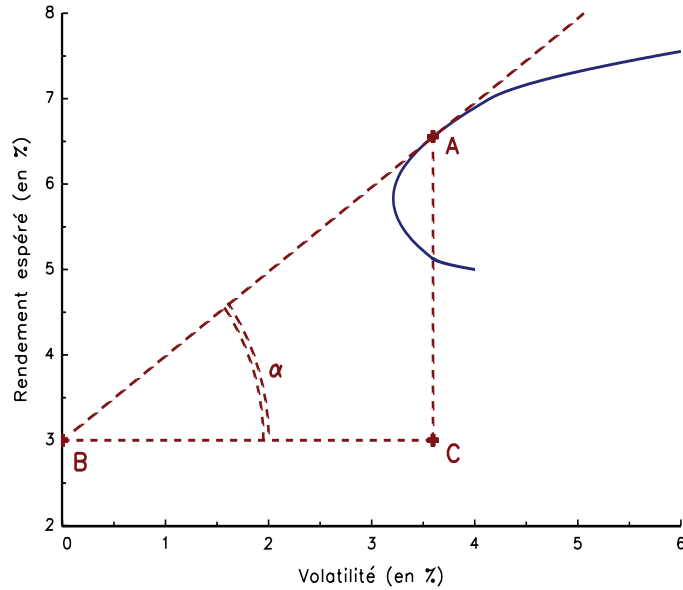
$$x_i = \frac{-((n-1)\rho + 1)\sigma_i^{-2} + \rho \sum_{j=1}^n (\sigma_i \sigma_j)^{-1}}{\sum_{k=1}^n \left(-((n-1)\rho + 1)\sigma_k^{-2} + \rho \sum_{j=1}^n (\sigma_k \sigma_j)^{-1} \right)}$$

¹³Nous avons donc $R_{i,j} = \rho$ si $i \neq j$ et $R_{i,i} = 1$.

¹⁴Nous utilisons la propriété $\text{tr}(AB) = \text{tr}(BA)$.

2.4.4. Portefeuille de marché et ratio de Sharpe

Graphique 9. Portefeuille de marché et ratio de Sharpe



Sur le graphique 9, nous reportons la frontière efficiente précédente ainsi que le taux sans risque qui est égal à 3%. Tobin a montré qu'il existe un seul portefeuille optimal lorsqu'il existe un actif sans risque. Sur le graphique, celui-ci correspond au point A. Dans ce cas, on voit que toute combinaison linéaire de ce portefeuille et de l'actif sans risque produit des portefeuilles qui dominent ceux appartenant à la frontière efficiente de Markowitz. Le portefeuille optimal (ou portefeuille de marché) est le portefeuille qui présente le plus grand ratio de Sharpe :

$$x^M = \arg \max \text{sh}(x | r)$$

D'un point de vue graphique, le ratio de Sharpe s'interprète comme la tangente de l'angle \widehat{ABC} :

$$\text{sh}(x | r) = \frac{\mu(x) - r}{\sigma(x)} = \frac{\overline{AC}}{\overline{BC}} = \tan \alpha$$

On remarque donc que les portefeuilles qui se situent sur la droite (A, B) ont le même ratio de Sharpe que le portefeuille de marché. Cette droite est appelée la droite de marché (*capital market line* ou CML).

Pour trouver le portefeuille de marché, on considère donc le problème suivant :

$$\phi^M = \arg \max \frac{\mu(x^*(\phi)) - r}{\sigma(x^*(\phi))}$$

Connaissant ϕ^M , nous en déduisons immédiatement que :

$$x^M = x^*(\phi^M)$$

La recherche du maximum est plus compliquée que résoudre les μ - et σ -problèmes d'allocation. Théoriquement, la fonction $\text{sh}(x^*(\phi) | r)$ est croissante jusqu'à ϕ^M , puis décroissante. Il semble donc très facile de trouver la solution en utilisant l'algorithme de Newton-Raphson. Néanmoins, il arrive que la décroissance (respectivement que la croissance) ne soit pas très forte notamment lorsque le taux sans risque r est élevé (respectivement faible). De plus, la fonction n'est pas forcément de classe C^2 (voir le graphique 10). Pour éviter ces problèmes, il est important de partir d'un point initial qui ne corresponde pas au portefeuille de variance minimale ou au portefeuille d'espérance maximale de rendement.

Reprenons l'exemple numérique précédent. Nous avons représenté le portefeuille de marché lorsque le taux sans risque r est égal à 3% sur le graphique 9. Les relations entre le ratio de Sharpe $\text{sh}(x^* | r)$ d'une part et les paramètres $\sigma(x^*)$, $\mu(x^*)$ et $\phi(x^*)$ d'autre part sont reportées sur le graphique 10 lorsque le taux sans risque vaut respectivement 1%, 3% et 6%. Les compositions du portefeuille de marché sont données dans le tableau 8.

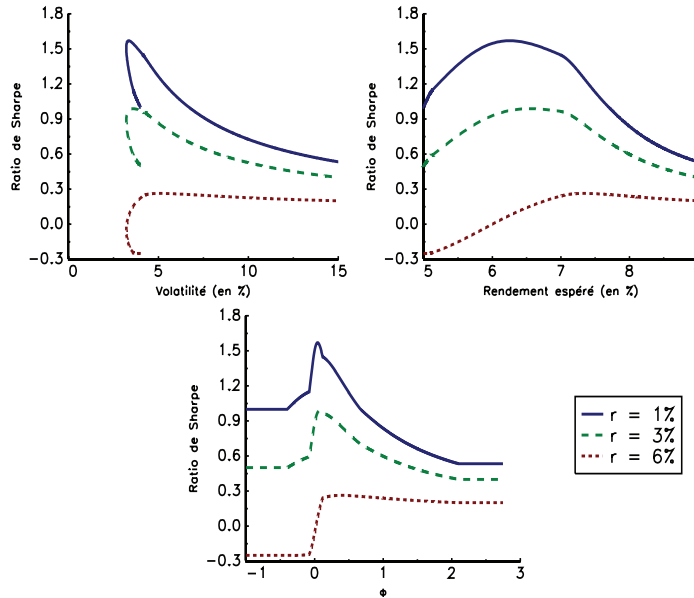
Tableau 8. Portefeuilles de marché

r	2,00	3,00	4,00	5,00	6,00
x_1^M	28,00	20,10	2,02		
x_2^M	5,63	6,70	9,14	12,10	21,82
x_3^M	48,77	54,99	69,22	70,22	68,07
x_4^M	17,61	18,22	19,62	17,69	10,11
ϕ^M	0,05	0,07	0,11	0,17	0,38
$\mu(x^M)$	6,38	6,55	6,95	7,07	7,34
$\sigma(x^M)$	3,44	3,59	4,07	4,26	5,07
$\text{sh}(x^M)$	1,27	0,99	0,72	0,48	0,26

Remarque 9. La condition de premier ordre du programme d'optimisation $x^* = \arg \max \text{sh}(x | r)$ est $\partial_x \text{sh}(x | r) = 0$. Le portefeuille optimal vérifie donc la relation suivante :

$$\frac{\partial_x \mu(x) - r}{\partial_x \sigma(x)} = \frac{\mu(x) - r}{\sigma(x)}$$

On peut montrer que cette condition est suffisante (Scherer, 2007).

Graphique 10. Ratio de Sharpe en fonction de $\sigma(x^*)$, $\mu(x^*)$ et $\phi(x^*)$ 

2.5. Construction d'un portefeuille long/short avec contrôle de volatilité

Un portefeuille long/short consiste en un ensemble de positions acheteuses financées par des positions vendeuses. Dans les problèmes d'allocation précédents, nous imposons que la somme des poids $\sum_{i=1}^n x_i$ soit égale à 1. Dans le cas d'un portefeuille long/short, cette condition est donc remplacée par la condition suivante :

$$\sum_{i=1}^n x_i = 0$$

Nous supposons aussi que nous disposons pour chaque stratégie i d'un score s_i et que notre objectif est de maximiser le score linéaire suivant :

$$x^* = \arg \max \sum_{i=1}^n s_i x_i$$

Le score du portefeuille x est noté $s(x) = \sum_{i=1}^n s_i x_i$. Nous verrons dans le chapitre sur les outils de scoring des méthodes statistiques pour construire ces scores. Mais ces scores peuvent aussi être très simples :

- s_i peut par exemple être égal au rendement espéré μ_i de la stratégie (calculé par un modèle de prévision économétrique ou correspondant à une vue discrétionnaire) ;

Implémentation du modèle

D'un point de vue pratique, voici donc les différentes étapes du modèle de Black-Litterman :

1. on calcule la matrice Σ de variance-covariance des rendements des actifs en utilisant les données historiques ;
2. on en déduit le vecteur d'équilibre π qui est la solution du problème inverse de Markowitz ;
3. le gérant spécifie ses vues de marché et la conviction dans ses vues (cela revient à spécifier les matrices P , Q et Ω) ;
4. à partir d'une matrice de poids Γ , on calcule l'espérance conditionnelle μ_{cond} ;
5. on utilise cette information pour résoudre le problème d'optimisation classique :

$$\begin{aligned} x^{\text{BL}} &= \arg \min x^\top \Sigma x - \phi x^\top \mu_{\text{cond}} \\ \text{s.c. } & \mathbf{1}^\top x = 1 \text{ et } \mathbf{0} \leq x \leq \mathbf{1} \end{aligned}$$

La difficulté majeure avec l'algorithme précédent est la spécification de la matrice Γ . Généralement, on pose :

$$\Gamma = \tau \Sigma$$

Le paramètre τ est un paramètre d'ajustement qui permet de calibrer une volatilité cible de tracking error (Meucci, 2005). De nouveau, nous avons un problème inverse. Si nous paramétrons la solution x^{BL} en τ :

$$\begin{aligned} x^{\text{BL}}(\tau) &= \arg \min x^\top \Sigma x - \phi x^\top \mu_{\text{cond}}(\tau) \\ \text{s.c. } & \begin{cases} \mathbf{1}^\top x = 1 \\ \mathbf{0} \leq x \leq \mathbf{1} \\ \mu_{\text{cond}}(\tau) = \pi + \tau \Sigma P^\top (P \tau \Sigma P + \Omega)^{-1} (Q - P \pi) \end{cases} \end{aligned}$$

alors le problème inverse est défini par :

$$\tau = \left\{ \tau > 0 : \sigma(x^{\text{BL}}(\tau) | x^*) = \sqrt{(x^{\text{BL}}(\tau) - x^*)^\top \Sigma (x^{\text{BL}}(\tau) - x^*)} = \eta \right\}$$

avec η la volatilité cible de tracking error ex-ante. Le modèle de Black-Litterman nécessite donc la résolution de deux problèmes inverses : le premier pour déterminer le vecteur π que nous calculons analytiquement et le second pour définir le scalaire τ que nous calibrons numériquement. Pour cela, nous utilisons la méthode de la bi-section.

Un exemple

Reprenons l'exemple précédent. Nous avons calculé π page 137. On suppose que le gérant a une vue absolue sur le premier actif et une vue relative sur les deuxième et troisième actifs :

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \end{pmatrix}$$

$$Q = \begin{pmatrix} q_1 \\ q_{2-3} \end{pmatrix}$$

On suppose que q_1 et q_{2-3} sont égaux à 5%. Cela veut dire que le gérant pense que l'espérance de rendement du premier actif est 5% et que l'espérance de sur-performance du second actif par rapport au troisième actif est aussi de 5%. Il nous faut maintenant spécifier l'incertitude des vues du gérant. On a :

$$\Omega = \begin{pmatrix} \varpi_1^2 & 0 \\ 0 & \varpi_{2-3}^2 \end{pmatrix}$$

On suppose que $\varpi_1 = 5\%$ et $\varpi_{2-3} = 10\%$. Cela indique que le gérant a plus confiance dans sa première vue que dans la seconde. Posons $\tau = 1$. L'ensemble de ces paramètres définit notre cas générique (1). Les autres cas sont des variantes de ce cas générique. Dans le cas (2), on a $q_1 = 10\%$ et $q_{2-3} = 0\%$. Le cas (3) correspond à $\varpi_1 = \varpi_{2-3} = 20\%$. Enfin, τ est égal à 10% dans le cas (4). Nous reportons les solutions x^{BL} dans le tableau ainsi que la volatilité de la tracking error $\sigma(x^{\text{BL}}(\tau) | x^*)$ par rapport au portefeuille stratégique x^* donné par le cas (0).

Tableau 19. *Portefeuilles Black-Litterman*

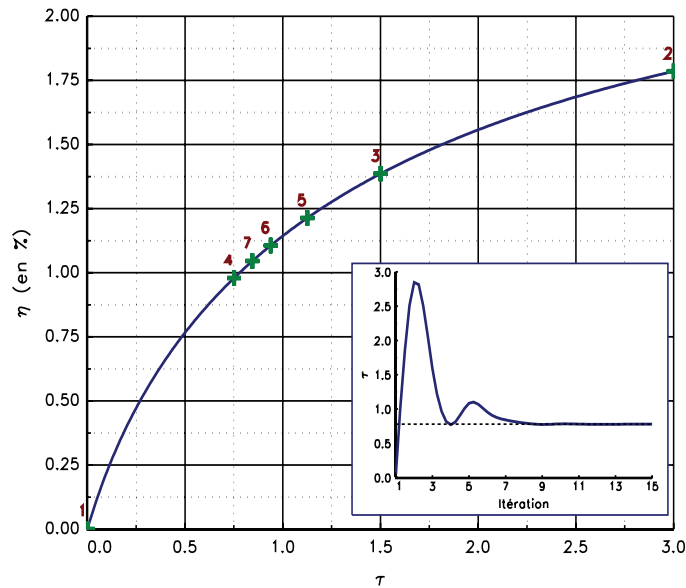
	(0)	(1)	(2)	(3)	(4)
x_1^{BL}	10,00	28,78	54,40	11,86	12,90
x_2^{BL}	20,00	15,16	7,27	17,76	18,77
x_3^{BL}	50,00	42,30	33,10	51,00	49,29
x_4^{BL}	20,00	13,76	5,24	19,38	19,03
$\sigma(x^{\text{BL}} x^*)$	0,00	1,14	2,81	0,34	0,22

Considérons le jeu de paramètres (1) et supposons que le gérant s'impose une contrainte de volatilité de tracking error égale à 1%. Nous représentons sur le graphique 17 le chemin emprunté par l'algorithme de la bi-section pour trouver la solution :

$$x^{\text{BL}} = \begin{pmatrix} 26,0\% \\ 15,6\% \\ 43,7\% \\ 14,7\% \end{pmatrix}$$

Nous remarquons que le modèle de Black-Litterman a sur-pondéré le premier actif (26% au lieu de 10%). Il aurait dû aussi sur-pondérer le troisième actif et sous-pondérer le second actif puisque le gérant pense que la différence de rendement est seulement de 5% alors que le marché évalue cette différence à 8%. Néanmoins, il n'en tient pas compte car l'incertitude du gérant est deux fois plus importante pour la deuxième vue. Il préfère donc jouer pleinement la première vue d'autant plus que la contrainte de volatilité de tracking error est forte.

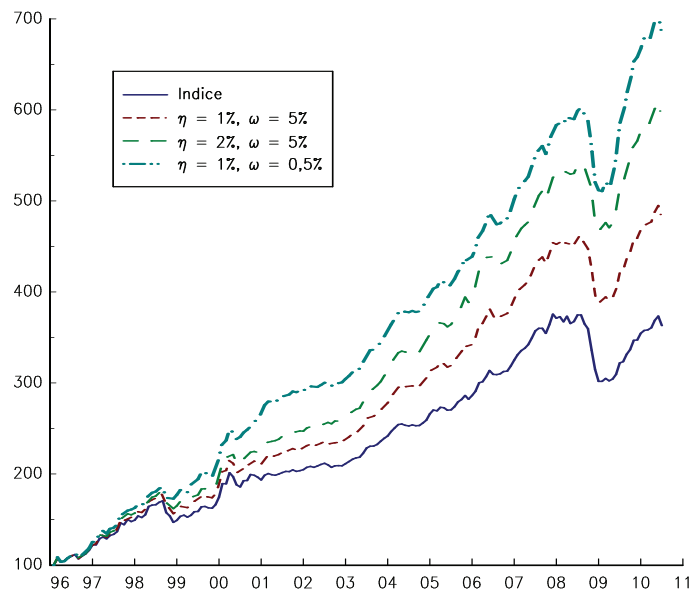
Graphique 17. Solution du problème inverse de Black-Litterman



Remarque 16. Le paramètre η de tracking error permet plus ou moins de tilter le portefeuille stratégique. Ce paramètre joue néanmoins le même rôle que la matrice d'incertitude Ω . Pour s'en convaincre, nous considérons un exemple avec l'indice hedge funds de Credit Suisse Tremont. Celui-ci est composé de dix sous-indices correspondant chacun à une stratégie de gestion alternative. Supposons que nous connaissons chaque mois la performance moyenne des sous-indices pour les douze prochains mois. Nous cherchons alors à tilter l'indice en incorporant ces vues dans le modèle de Black-Litterman. Nous supposons que la matrice Ω est une matrice diagonale d'éléments $\Omega_{i,i} = \omega^2$. La matrice Σ est la matrice de covariance empirique des vingt-quatre derniers mois. Le portefeuille de marché correspond à la composition mensuelle de l'indice. Le graphique 18 présente

les résultats du backtest⁴⁶ pour différentes valeurs de η et ω . Une augmentation de η a donc un effet similaire à une diminution de ω . Une plus grande tracking error est équivalente à une plus grande certitude des vues tactiques.

Graphique 18. Incertitude et tracking error dans le modèle de Black-Litterman



4. La programmation dynamique

Les principes de la programmation dynamique ont été énoncés par Richard Bellman dans les années cinquante. Ils permettent de résoudre des problèmes de contrôle optimal. Dans les années soixante-dix, Robert Merton a utilisé intensivement ces techniques pour résoudre de nombreux problèmes de finance tels que la valorisation d'option américaine, l'optimisation inter-temporelle entre la consommation et l'épargne, et plus généralement les problèmes de choix d'investissement en avenir incertain.

⁴⁶Nous faisons l'hypothèse $sh^* = 0,50$ pour calculer les primes de risque d'équilibre.

4.1. L'approche de Bellman

4.1.1. Le principe d'optimalité de Bellman

Considérons un problème d'optimisation dynamique entre les dates 0 et T . La programmation dynamique repose sur le principe d'optimalité de Bellman : si le problème général peut être décomposé en sous-problèmes, alors une sous-trajectoire de la trajectoire optimale est elle-même optimale pour la fonction objectif restreinte aux trajectoires ayant pour origine celle de cette sous-trajectoire. Ce principe permet de résoudre le problème général de façon descendante. On détermine la solution optimale en partant de la date T et on résout successivement le sous-problème entre les dates t et $t - 1$.

Considérons le problème d'optimisation dynamique discrète suivant⁴⁷. On note t les dates de décision entre 0 et T . Soient $e(t)$ et $v(t)$ les variables d'état et de contrôle⁴⁸. On définit $f(t, e(t), v(t))$ la fonction de gain et $e(t + 1) = g(t, e(t), v(t))$ la fonction de transfert. Connaissant l'état initial du système e_0 , on cherche à déterminer la suite des contrôles optimaux $v^*(t)$ qui permet de maximiser la fonction objectif définie par :

$$J(v) = \sum_{t=0}^{T-1} f(t, e(t), v(t)) + h(T, e(T))$$

On impose aussi que les variables d'état et de contrôle satisfont certaines contraintes de la forme $e(t) \in \mathcal{E}(t)$ et $v(t) \in \mathcal{V}(t)$. Le problème d'optimisation est donc :

$$\begin{aligned} \max \quad & \sum_{t=0}^{T-1} f(t, e(t), v(t)) + h(T, e(T)) \\ \text{s.c.} \quad & \begin{cases} e(t+1) = g(t, e(t), v(t)) \\ e(t) \in \mathcal{E}(t) \\ v(t) \in \mathcal{V}(t) \\ e(0) = e_0 \end{cases} \end{aligned}$$

Pour résoudre ce problème, on considère le sous-problème suivant entre $t - 1$ et t :

$$h(t, e(t)) = \max_{v(t)} f(t, e(t), v(t)) + h(t + 1, g(t, e(t), v(t)))$$

Sous certaines conditions, la suite des solutions de ces sous-problèmes est la solution du problème général.

⁴⁷Cette partie s'inspire très largement du chapitre 4 Partie B de Demange et Rochet (1997).

⁴⁸En automatique, la variable de contrôle est appelée la variable de commande.

4.3. Quelques extensions du modèle de Merton

Le modèle de Merton a donné lieu à de nombreuses extensions, en particulier dans le domaine de la gestion dite de long terme. C'est notamment le cas de la gestion d'un fonds de pension, qui est un mécanisme de système de retraite. Durant la vie active, l'employé cotise au fonds de pension, le fonds de pension gère les sommes reçues en investissant dans des actifs, et l'employé reçoit au moment de la retraite un capital ou une pension.

Dans le cas d'un fonds de pension *defined benefit* (DB), la pension est définie ex-ante. Dans ce cas, le risque est transféré au fonds de pension, et celui-ci peut être (virtuellement) en faillite si l'actif est inférieur au passif. L'approche *liability-driven investment* (LDI) est une méthode de gestion actif-passif adaptée pour les fonds de pension dans cette situation⁵⁷.

Dans le cas d'un fonds de pension *defined contribution* (DC), la pension dépend de la performance du portefeuille de l'employé. Le risque est ainsi entièrement supporté par l'employé. Celui-ci doit donc faire des choix d'investissement alors qu'il a généralement une connaissance très faible de la théorie financière. Ces choix sont d'autant plus compliqués qu'ils peuvent dépendre du cycle de vie de l'investisseur⁵⁸. Les sociétés de gestion proposent alors deux solutions simples à l'investisseur. La première solution consiste à investir dans un fonds profilé, c'est-à-dire un fonds diversifié correspondant à une certaine aversion pour le risque. On distingue généralement trois types d'investisseur : l'investisseur conservateur (C), l'investisseur modéré (M) et l'investisseur agressif (A). À chacun de ces types d'investisseur correspond un fonds profilé. On a ainsi des fonds prudents (P), des fonds équilibrés (E) et des fonds dynamiques (D). Ces fonds sont aussi appelés des fonds "*lifestyle*", c'est-à-dire des fonds qui dépendent du style de vie de l'investisseur qui est appréhendé par son aversion pour le risque. La deuxième solution consiste à investir dans des fonds générationnels. L'idée sous-jacente est de prendre en compte le cycle de vie de l'investisseur. Dans ce cas, la part de l'actif risqué est grande lorsque l'investisseur est jeune, elle diminue avec l'âge pour atteindre une valeur faible au moment de la retraite. Ces fonds générationnels sont appelés des fonds "*lifecyle*".

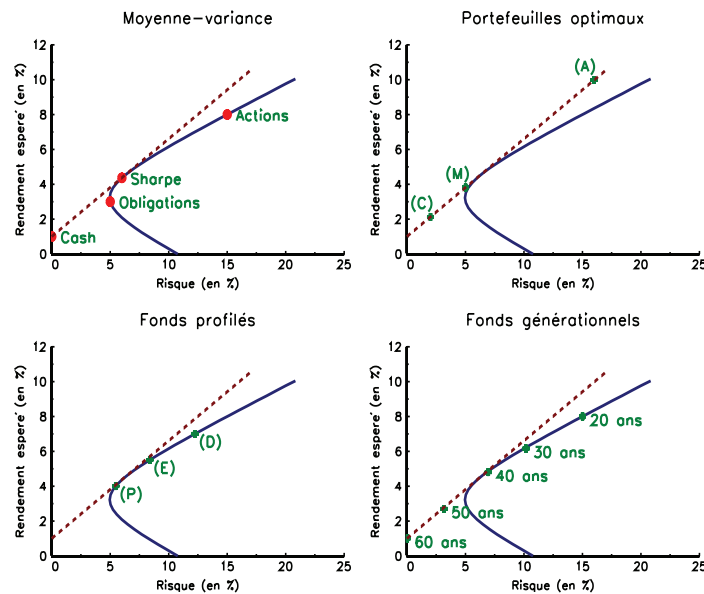
⁵⁷B. Bruder, G. Jamet et G. Lasserre, « Beyond Liability-Driven Investment: New Perspectives on Defined Benefit Pension Fund Management », *Lyxor White Paper Series*, 2, 2010, www.lyxor.com.

L. Martellini et V. Milhau, « Measuring the Benefits of Dynamic Asset Allocation Strategies in the Presence of Liability Constraints », *EDHEC Risk Working Paper*, 2009.

⁵⁸P.A. Samuelson, « Lifetime Portfolio Selection by Dynamic Stochastic Programming », *Review of Economics and Statistics*, 51(3), 1969, p. 239-246.

Sur le graphique 21, nous reportons les différentes solutions⁵⁹. Sur le premier graphe, on illustre l'approche moyenne-variance et la détermination du portefeuille de marché. Sur le deuxième graphe, on indique le couple rendement-risque des portefeuilles optimaux des trois investisseurs. Ces portefeuilles sont sur la droite de marché. Ce qui les distingue, ce sont les proportions relatives entre le portefeuille de marché et l'actif sans risque. Ainsi, un investisseur conservateur (C) détient plus d'actif sans risque qu'un investisseur agressif (A). Mais la proportion entre les actions et les obligations est la même. Celle-ci est donnée par le portefeuille de marché. Les fonds profilés sont illustrés sur le troisième graphe. La proportion entre actions et obligations varie en fonction du profil. Ainsi, le fonds prudent (P) détient plus d'obligations que le fonds dynamique (D). Enfin, les fonds générationnels sont reportés sur le quatrième graphique. Selon ce principe, un investisseur jeune (20 ans) détient quasiment toute sa richesse en actions. Avec l'âge, celui-ci va réduire la part en actions au bénéfice des obligations et de l'actif sans risque. Au moment de la retraite, le portefeuille est presque entièrement investi dans l'actif sans risque.

Graphique 21. Illustration des fonds profilés et générationnels



⁵⁹Le taux sans risque est fixé à 1%, les rendements espérés des obligations et des actions sont égaux à 3% et 8% et les volatilités correspondantes sont de 5% et 15%. La corrélation entre les obligations et les actions est de 20%.

L'existence des fonds profilés et générationnels est donc contradictoire avec le théorème de séparation de Tobin (Campbell et Viciera, 2002). De même, la théorie financière classique ne permet pas de résoudre le problème des fonds de pension *defined benefit*. Dans les paragraphes suivants, nous voyons comment on peut réconcilier ces pratiques de marché et la théorie financière en adaptant le cadre d'analyse de Merton.

4.3.1. L'approche liability-driven investment

Nous reprenons la méthodologie présentée dans Bruder *et al.* (2010). On considère un fonds de pension DB qui doit faire face à un ensemble de paiements futurs. On calcule d'abord la valeur présente de la structure du passif du fonds de pension, ainsi que les sensibilités aux différents paramètres de marché (taux d'intérêt, inflation, etc.). Ceci permet de construire le portefeuille de couverture du passif (*liability hedging portfolio* ou LHP). Si la valeur de l'actif est supérieure ou égale à la valeur du portefeuille LHP, alors le fonds de pension a tout intérêt à acheter le portefeuille LHP. Dans ce cas, le passif est entièrement couvert et se présente sous la forme d'un portefeuille obligataire composé de zéro-coupons, d'obligations indexées sur l'inflation, etc. Si la valeur de l'actif ne lui permet pas d'acheter le portefeuille LHP, alors le fonds de pension n'est pas solvable. Pour le redevenir, celui-ci doit investir dans des actifs risqués. L'approche LDI permet alors de définir la stratégie dynamique optimale d'exposition aux actifs risqués.

Notons $A(t)$ la valeur de l'actif du fonds et $L(t)$ la valeur présente du passif. On définit le ratio de financement (*funding ratio*) de la façon suivante :

$$F(t) = \frac{A(t)}{L(t)}$$

Le programme d'optimisation du fonds de pension est alors :

$$\max \mathbb{E}_{\mathbb{P}} [U(F(T))]$$

sous la contrainte de budget :

$$A_0 = \mathbb{E}_{\mathbb{Q}} \left[e^{-\int_0^T r_t dt} A(T) \right]$$

Comme nous l'avons vu précédemment, la contrainte de budget traduit le fait que le fonds de pension achète un actif contingent $A(T)$ dont le prix initial $\mathbb{E}_{\mathbb{Q}} \left[\exp \left(-\int_0^T r_t dt \right) A(T) \right]$ est égal à la richesse initiale du fonds de pension, qui est la valeur du passif A_0 . On en déduit que la solution optimale $A^*(T)$ vérifie la relation suivante :

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \frac{U'(F^*(T))}{\lambda_L}$$

avec λ_L le lagrangien associé à la contrainte de budget. Bruder *et al.* (2010) montrent alors que la résolution du problème précédent revient à

par le fait que les valeurs propres de Σ sont :

$$\lambda_i(\Sigma) = \begin{cases} \lambda_i(A_1^2) & \text{si } \lambda_i(A_1^2) > 0 \\ -\lambda_i(A_2^2) & \text{sinon} \end{cases}$$

Ce résultat se déduit de la décomposition $A = A_1 + iA_2$. Une autre façon de trouver la matrice de covariance la plus proche consiste donc à faire une décomposition en valeurs propres de Σ , à mettre les valeurs propres négatives à zéro, puis à recomposer la matrice avec ces nouvelles valeurs propres.

2.3.2. Calcul de la matrice de corrélation la plus proche

On définit la matrice de corrélation la plus proche d'une matrice carrée arbitraire A de la façon suivante :

$$\gamma(A) = \min \{ \|A - X\| : X \text{ est une matrice de corrélation} \}$$

Pour résoudre ce problème, Higham (2002) propose d'utiliser la méthode de projections alternées qui est une méthode itérative²⁰ :

$$A \longleftarrow P_U(P_S(A))$$

où P_U et P_S sont des projections sur les espaces S et U qui sont définis par :

$$S = \{X = X^\top : X \geq 0\}$$

et :

$$U = \{X = X^\top : \text{diag}(X) = \mathbf{1}\}$$

L'algorithme pour calculer la matrice de corrélation la plus proche de A correspond alors aux étapes suivantes :

1. on pose $\Delta S_0 = 0$ et $X_0 = A$;
2. à l'itération k , on calcule $R_k = X_{k-1} - \Delta S_{k-1}$, $Y_k = P_S(R_k)$, $\Delta S_k = Y_k - R_k$ et $X_k = P_U(Y_k)$;
3. on recommence l'étape 2 jusqu'à la convergence de l'algorithme ; la solution est alors X_k .

On peut considérer différentes classes de projection pour l'implémentation numérique de l'algorithme. Dans ce livre, nous utilisons les projections suivantes :

$$P_U(A) = (P_{m,n}) \quad \text{avec } P_{m,n} = \begin{cases} 1 & \text{si } m = n \\ A_{m,n} & \text{sinon} \end{cases}$$

et :

$$P_S(A) = TS^+T^*$$

²⁰N.J. Higham, « Computing the Nearest Correlation Matrix — A Problem from Finance », *IMA Journal of Numerical Analysis*, 22(3), 2002, p. 329-343.

où TST^* est la décomposition Schur de A et $S^+ = (S_{m,n}^+)$ avec $S_{m,n}^+ = \max(S_{m,n}, 0)$.

Si on applique cet algorithme à une matrice de corrélation uniforme $C_n(\rho)$ avec une valeur ρ quelconque, nous avons :

$$P_U(P_S(C_n(\rho))) = C_n(\rho^*)$$

avec $\rho^* = \max(\rho, -1/(n-1))$. Dans le cas où $\rho < -1/(n-1)$, nous obtenons donc la matrice de corrélation constante de dimension n la plus petite, qui est la solution la plus naturelle du problème. Considérons maintenant la matrice suivante :

$$A = \begin{pmatrix} 1,0 & & & & & \\ 0,9 & 1,0 & & & & \\ 0,5 & 0,6 & 1,0 & & & \\ 0,2 & 0,9 & 0,0 & 1,0 & & \\ 0,9 & 0,0 & 0,9 & 0,0 & 1,0 & \end{pmatrix}$$

En appliquant l'algorithme précédent, nous obtenons la matrice de corrélation suivante :

$$\rho(A) = \begin{pmatrix} 1,0000 & & & & & \\ 0,6933 & 1,0000 & & & & \\ 0,6147 & 0,4571 & 1,0000 & & & \\ 0,2920 & 0,7853 & 0,0636 & 1,0000 & & \\ 0,7376 & 0,2025 & 0,7876 & -0,0901 & 1,0000 & \end{pmatrix}$$

2.4. L'intégration numérique

Les méthodes d'intégration numérique permettent de calculer les intégrales de la forme :

$$I = \int_a^b f(x) dx$$

On distingue généralement les algorithmes numériques à pas réguliers de ceux à pas irréguliers.

2.4.1. Les méthodes des trapèzes et de Simpson

Une première approche est d'approcher l'intégrale I par une succession de trapèzes. Soient $x_i = a + ih$ avec $h = n^{-1}(b-a)$ et n le nombre de nœuds. Nous avons :

$$\begin{aligned} \int_a^b f(x) dx &\simeq \sum_{i=1}^n \frac{h}{2} (f(x_{i-1}) + f(x_i)) \\ &= h \left(\frac{f(a) + f(b)}{2} + \sum_{i=1}^{n-1} f(x_i) \right) \end{aligned}$$

tes valeurs de x_T . L'utilisation de l'équation de Fokker-Planck est notamment intéressante lorsque nous ne disposons pas de la solution analytique de $f(x_T)$.

Reprenons l'exemple du processus de Ornstein-Uhlenbeck :

$$dx(t) = a(b - x(t)) dt + \sigma dW(t)$$

Nous cherchons à déterminer la densité f associée à la distribution $F(x_T) = \Pr\{x(T) \leq x_T \mid x(0) = x_0\}$. La première façon de procéder est de résoudre l'équation de Feynman-Kac :

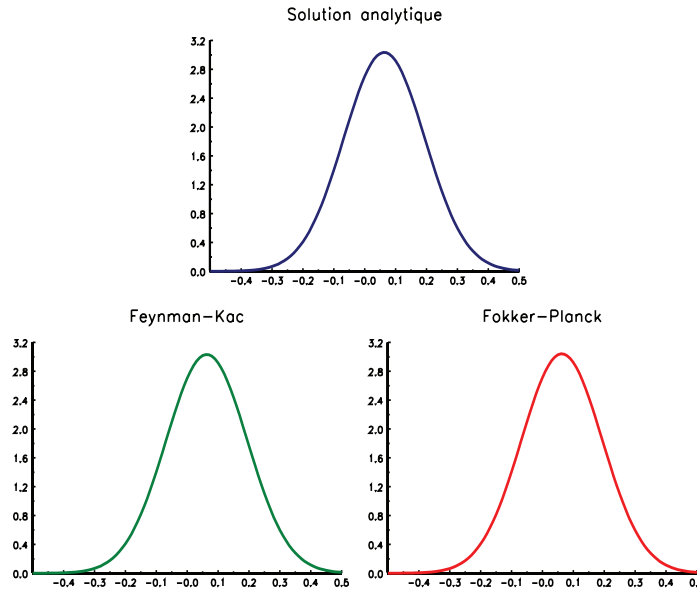
$$\begin{cases} \frac{1}{2}\sigma^2(t, x) \partial_x^2 u(t, x) + (a(b - x(t))) \partial_x u(t, x) + \partial_t u(t, x) = 0 \\ u(T, x) = \phi(T, x) \end{cases}$$

L'autre méthode consiste à résoudre l'équation de Fokker-Planck :

$$\begin{cases} \frac{1}{2}\sigma^2 \partial_x^2 u(t, x) - a(b - x) \partial_x u(t, x) = \partial_t u(t, x) - au(t, x) \\ u(0, x) = \mathbf{1}\{x = x_0\} \end{cases}$$

Sur le graphique 21, nous représentons les deux solutions numériques ainsi que la solution exacte pour les valeurs $x_0 = 0$, $a = 1$, $b = 10\%$, $\sigma = 20\%$ et $T = 1$.

Graphique 21. Densité du processus de Ornstein-Uhlenbeck



Dans le cas du mouvement brownien géométrique :

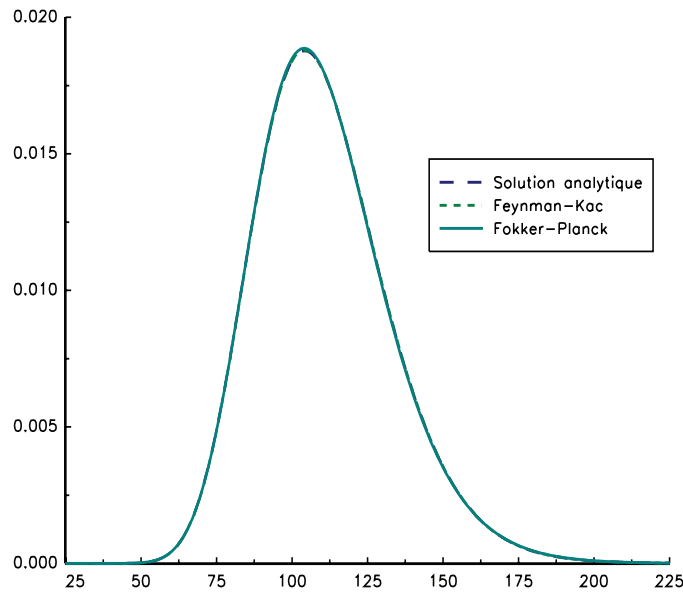
$$dx(t) = \mu x(t) dt + \sigma x(t) dW(t)$$

l'équation de Fokker-Planck s'écrit :

$$\begin{cases} \frac{1}{2}\sigma^2 x^2 \partial_x^2 u(t, x) - (\mu x - 2\sigma^2 x) \partial_x u(t, x) = \partial_t u(t, x) + (\mu - \sigma^2) u(t, x) \\ u(0, x) = \mathbf{1}\{x = x_0\} \end{cases}$$

Nous représentons la solution analytique ainsi que les solutions numériques de Feynman-Kac et de Fokker-Planck sur le graphique 22 pour les valeurs $x_0 = 100$, $\mu = 10\%$, $\sigma = 20\%$ et $T = 1$.

Graphique 22. Densité du mouvement brownien géométrique



L'équation de Fokker-Planck se généralise dans le cas de processus de diffusion multi-dimensionnels. Par exemple, dans le cas bivarié :

$$\begin{cases} dx_1(t) = \mu_1(t, x_1, x_2) dt + \sigma_1(t, x_1, x_2) dW_1(t) \\ dx_2(t) = \mu_2(t, x_1, x_2) dt + \sigma_2(t, x_1, x_2) dW_2(t) \\ \mathbb{E}[W_1(t) W_2(t)] = \rho dt \end{cases}$$

nous avons :

$$\begin{cases} \frac{1}{2}\partial_{x_1}^2 [\sigma_1^2(\cdot) u(\cdot)] + \frac{1}{2}\partial_{x_2}^2 [\sigma_2^2(\cdot) u(\cdot)] + \rho \partial_{x_1, x_2}^2 [\sigma_1(\cdot) \sigma_2(\cdot) u(\cdot)] \\ -\partial_{x_1} [\mu_1(\cdot) u(\cdot)] - \partial_{x_2} [\mu_2(\cdot) u(\cdot)] = \partial_t u(\cdot) \\ u(0, x_{1,0}, x_{2,0}) = \mathbf{1}\{x_1 = x_{1,0}, x_2 = x_{2,0}\} \end{cases}$$

Cette équation a pour solution $u(T, x_{1,T}, x_{2,T}) = f(x_{1,T}, x_{2,T}) dx_1 dx_2$ avec f la densité associée à la distribution :

$$F(x_1, x_2) = \Pr\{x_1(T) \leq x_{1,T}, x_2(T) \leq x_{2,T} \mid x_1(0) = x_{1,0}, x_2(0) = x_{2,0}\}$$

Considérons par exemple le processus de diffusion suivant :

$$\begin{cases} dx_1(t) = \mu x_1(t) dt + x_1(t) \sqrt{x_2(t)} dW_1(t) \\ dx_2(t) = a(b - x_2(t)) dt + \sigma \sqrt{x_2(t)} dW_2(t) \\ \mathbb{E}[W_1(t)W_2(t)] = \rho t \end{cases}$$

Ce processus a été introduit en finance par Heston (1993) pour modéliser la volatilité stochastique⁴³. L'équation de Fokker-Planck associée est :

$$\begin{cases} \frac{1}{2} \partial_{x_1}^2 [x_1^2 x_2 u(t, x_1, x_2)] + \frac{1}{2} \partial_{x_2}^2 [\sigma^2 x_2 u(t, x_1, x_2)] + \\ \rho \partial_{x_1, x_2}^2 [\sigma x_1 x_2 u(t, x_1, x_2)] \\ - \partial_{x_1} [\mu x_1 u(t, x_1, x_2)] - \partial_{x_2} [a(b - x_2) u(t, x_1, x_2)] = \partial_t u(t, x_1, x_2) \end{cases}$$

En développant, nous obtenons :

$$\begin{cases} \frac{1}{2} x_1^2 x_2 \partial_{x_1}^2 u + \frac{1}{2} \sigma^2 x_2 \partial_{x_2}^2 u + \rho \sigma x_1 x_2 \partial_{x_1, x_2}^2 u + \\ (2x_2 + \rho \sigma - \mu) x_1 \partial_{x_1} u + (\sigma^2 + \rho \sigma x_2 - a(b - x_2)) \partial_{x_2} [u(t, x_1, x_2)] = \\ \partial_t u(t, x_1, x_2) + (\mu - a - x_2 - \rho \sigma) u(t, x_1, x_2) \end{cases}$$

Un autre modèle de volatilité stochastique (appelé SABR) est celui développé par Hagan *et al.* (2002)⁴⁴ :

$$\begin{cases} dx_1(t) = x_2(t) x_1^\beta(t) dW_1(t) \\ dx_2(t) = \nu x_2(t) dW_2(t) \\ \mathbb{E}[W_1(t)W_2(t)] = \rho dt \end{cases}$$

On montre que l'équation de Fokker-Planck est :

$$\begin{cases} \frac{1}{2} x_1^{2\beta} x_2^2 \partial_{x_1}^2 u + \frac{1}{2} \nu^2 x_2^2 \partial_{x_2}^2 u + \rho \nu x_1^\beta x_2^2 \partial_{x_1, x_2}^2 u + \\ \left(2\beta x_1^{2\beta-1} x_2^2 + 2\rho \nu x_1^\beta x_2 \right) \partial_{x_1} u + \left(2\nu^2 x_2 + \beta \rho \nu x_1^{\beta-1} x_2^2 \right) \partial_{x_2} u = \\ \partial_t u(t, x_1, x_2) - \left(\beta(2\beta - 1) x_1^{2\beta-2} x_2^2 + \nu^2 + 2\beta \rho \nu x_1^{\beta-1} x_2 \right) u \end{cases}$$

En particulier, pour $\beta = 1$, on obtient :

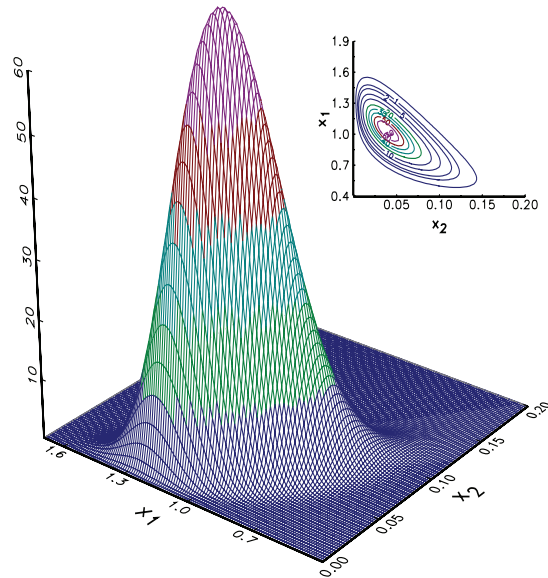
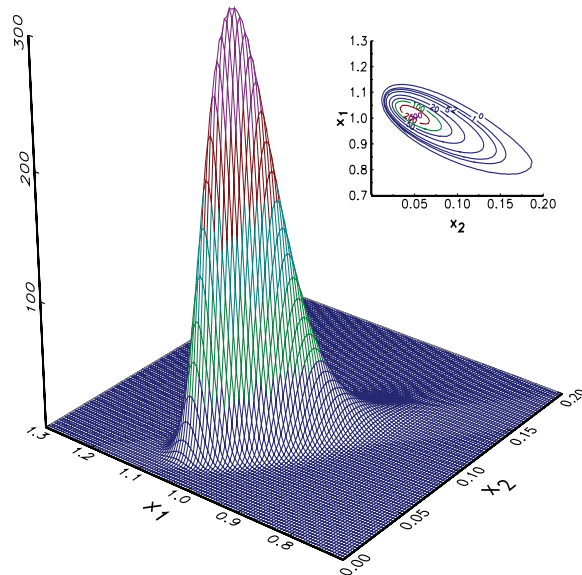
$$\begin{cases} \frac{1}{2} x_1^2 x_2^2 \partial_{x_1}^2 u + \frac{1}{2} \nu^2 x_2^2 \partial_{x_2}^2 u + \rho \nu x_1 x_2^2 \partial_{x_1, x_2}^2 u + \\ (2x_1 x_2^2 + 2\rho \nu x_1 x_2) \partial_{x_1} u + (2\nu^2 x_2 + \rho \nu x_2^2) \partial_{x_2} u = \\ \partial_t u(t, x_1, x_2) - (x_2^2 + \nu^2 + 2\rho \nu x_2) u \end{cases}$$

À titre d'illustration, nous avons représenté les densités bivariées des modèles⁴⁵ de Heston et SABR pour $T = 0,5$ sur les graphiques 23 et 24 en utilisant les méthodes numériques de résolution des edp bivariées de Kurpiel et Roncalli (1999).

⁴³S.L. Heston, « A Closed-form Solution for Options with Stochastic Volatility with Application to Bond and Currency Options », *Review of Financial Studies*, 6(2), 1993, p. 327-343.

⁴⁴P. Hagan, D. Kumar, A. Lesniewski et D. Woodward, « Managing Smile Risk », *Wilmott Magazine*, September 2002, p. 84-108.

⁴⁵Les paramètres pour le modèle de Heston sont $\mu = 0$, $a = 2$, $b = 4\%$, $\sigma = 20\%$ et $\rho = -75\%$ alors que nous utilisons $\beta = 0,75$, $\nu = 0,5$ et $\rho = -75\%$ pour le modèle SABR. Dans les deux cas, la position initiale du système est $x_1(0) = 1$ et $x_2(0) = 0,06$.

Graphique 23. Densité du modèle de Heston**Graphique 24.** Densité du modèle SABR

nous avons besoin que (\tilde{Y}, \tilde{X}) forment un vecteur gaussien pour montrer que la solution du problème d'espérance conditionnelle $\mathbb{E}[\tilde{Y} \mid \tilde{X} = x] = f(x)$ est bien :

$$\begin{aligned} f(x) &= x^\top \hat{\beta} \\ &= x^\top (X^\top X)^{-1} X^\top Y \end{aligned}$$

Si nous faisons l'hypothèse que (\tilde{Y}, \tilde{X}) forment un vecteur gaussien, cela revient à faire l'hypothèse que \tilde{U} est une variable aléatoire gaussienne puisque la projection gaussienne est linéaire. Remarquons que nous avons aussi :

$$\begin{aligned} \hat{\beta} &= (X^\top X)^{-1} X^\top Y \\ &= \beta + (X^\top X)^{-1} X^\top \tilde{U} \end{aligned}$$

En imposant l'orthogonalité entre X et \tilde{U} , nous montrons que l'estimateur $\hat{\beta}$ est sans biais :

$$\begin{aligned} \mathbb{E}[\hat{\beta}] &= \beta + (X^\top X)^{-1} \mathbb{E}[X^\top \tilde{U}] \\ &= \beta \end{aligned}$$

Si les résidus sont homoscedastiques², la matrice de covariance de $\hat{\beta}$ est :

$$\begin{aligned} \text{var}(\hat{\beta}) &= \mathbb{E}\left[\left(\hat{\beta} - \beta\right)\left(\hat{\beta} - \beta\right)^\top\right] \\ &= \sigma^2 (X^\top X)^{-1} \end{aligned}$$

1.1.2. Application au calcul de l'alpha et du bêta

Nous avons vu précédemment qu'il existe un portefeuille de marché x^M (ou portefeuille optimal) qui est préférable à tous les autres portefeuilles car il maximise le ratio de Sharpe $sh(x \mid r)$. Dans ce cas, toute allocation $x^M(\theta)$ entre le portefeuille de marché et l'actif sans risque domine la frontière efficiente de Markowitz. Le rendement aléatoire R_t du portefeuille $x^M(\theta)$ est :

$$R_t = \theta R_t^M + (1 - \theta)r$$

avec R_t^M le rendement aléatoire du portefeuille de marché et r le taux sans risque. Nous avons :

$$\mu(x^M(\theta)) = \theta\mu(x^M) + (1 - \theta)r$$

et :

$$\sigma(x^M(\theta)) = \theta\sigma(x^M)$$

²Cela veut dire que les résidus ont la même variance ou encore $\tilde{U} \sim \mathcal{N}(0, \sigma^2 I)$.

Nous vérifions que le ratio de Sharpe du portefeuille $x^M(\theta)$ est le même que celui du portefeuille de marché x^M :

$$\begin{aligned} \text{sh}(x^M(\theta) | r) &= \frac{\theta\mu(x^M) + (1-\theta)r - r}{\theta\sigma(x^M)} \\ &= \frac{\mu(x^M) - r}{\sigma(x^M)} \\ &= \text{sh}(x^M | r) \end{aligned}$$

L'ensemble des points $\{(\sigma(x^M(\theta)), \mu(x^M(\theta))), \theta \in \mathbb{R}_+\}$ est appelé la droite de marché. Le modèle CAPM est un prolongement de l'analyse précédente. Dans ce cas, le rendement R_t^i d'un titre i est lié au rendement du portefeuille de marché par la relation linéaire suivante³ :

$$R_t^i = r + \beta(R_t^M - r) + \varepsilon_i$$

avec :

$$\beta = \frac{\text{cov}(R_t^M, R_t^i)}{\sigma^2(R_t^M)}$$

Dans la pratique, on estime le bêta d'un titre à partir d'un historique de rendements et on approxime le portefeuille de marché par un benchmark représentatif (par exemple un indice). Jensen (1968) utilise ce modèle pour estimer la sur-performance d'un gérant⁴. Pour cela, il considère le modèle suivant :

$$R_t^i - r_t = \alpha + \beta(R_t^M - r_t) + \varepsilon_i$$

Les estimations $\hat{\alpha}$ et $\hat{\beta}$ de cette régression linéaire sont appelés l'alpha et le bêta du gérant. Certains styles de gestion sont basés sur ce modèle d'évaluation des actifs. Par exemple, un gérant peut privilégier des titres agressifs ($\hat{\beta} > 1$) ou des titres défensifs ($\hat{\beta} < 1$) dans la construction de son portefeuille. Certaines gestions présentent beaucoup de bêta et peu d'alpha. C'est le cas des gestions indicielles. D'autres stratégies, comme les stratégies *absolute return*, se veulent être des gestions avec peu de bêta et beaucoup d'alpha. Cette distinction entre stratégie d'alpha et stratégie de bêta s'est encore renforcée ces dernières années avec la crise financière de 2008.

1.1.3. Extension aux moindres carrés pondérés

Dans le cas des moindres carrés pondérés (*weighted least squares* ou WLS), nous avons :

$$\hat{\beta} = \arg \min \sum_{i=1}^n w_i u_i^2$$

³W.F. Sharpe, « Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk », *Journal of Finance*, 19(3), 1964, p. 425-442.

⁴M.C. Jensen, « The Performance of Mutual Funds in the Period 1945-1964 », *Journal of Finance*, 23(2), 1968, p. 389-416.

Nous pouvons montrer que l'estimateur analytique est :

$$\hat{\beta} = (X^T W X)^{-1} X^T W Y$$

où W est une matrice diagonale dont les éléments sont $W_{i,i} = w_i$.

1.1.4. Extension à la régression robuste

Reprenons le problème précédent des moindres carrés ordinaires. Nous avons :

$$\hat{\beta} = \arg \min \sum_{i=1}^n \rho(u_i) = \arg \min \sum_{i=1}^n \rho(y_i - x_i^T \beta) \quad (4.1)$$

avec $\rho(u) = u^2$. Huber (1964) propose de généraliser cette méthode en considérant d'autres fonctions $\rho(u)$ ⁵. Dans cette approche appelée *M*-estimation, la fonction ρ doit posséder les propriétés suivantes : $\rho(u) \geq 0$, $\rho(0) = 0$, $\rho(u) = \rho(-u)$ et $\rho(u_1) \geq \rho(u_2)$ si $|u_1| \geq |u_2|$. Soit $\psi(u) = \rho'(u)$. Les conditions de premier ordre du problème (4.1) forment un système de K équations :

$$\sum_{i=1}^n \psi(y_i - x_i^T \beta) x_{i,k} = 0$$

Nous pouvons écrire ce système de la façon suivante :

$$\sum_{i=1}^n \frac{\psi(y_i - x_i^T \beta)}{y_i - x_i^T \beta} (y_i - x_i^T \beta) x_{i,k} = 0$$

En posant $w(u) = \psi(u)/u$, nous obtenons finalement :

$$\sum_{i=1}^n w_i \cdot (y_i - x_i^T \beta) \cdot x_{i,k} = 0$$

Nous remarquons que nous obtenons le système de premier ordre du problème des moindres carrés pondérés. La seule différence est l'endogénéité des poids w_i qui dépendent de u_i . Pour résoudre ce système, nous utilisons l'algorithme itératif suivant :

1. on choisit une valeur initiale $\beta^{(0)}$;
2. on calcule la matrice $W^{(j-1)}$ associée aux poids $w_i = y_i - x_i^T \beta^{(j-1)}$;
3. à chaque itération j , nous calculons l'estimateur des moindres carrés :

$$\beta^{(j)} = \left(X^T W^{(j-1)} X \right)^{-1} X^T W^{(j-1)} Y$$

⁵P.J. Huber, « Robust Estimation of a Location Parameter », *Annals of Mathematical Statistics*, 35(1), 1964, p. 73-101.

4. on répète les étapes 2 et 3 jusqu'à la convergence de l'algorithme :
 $|\beta^{(j)} - \beta^{(j-1)}| \leq \varepsilon$.

L'estimateur le plus célèbre est celui obtenu avec $\rho(u) = |u|$ et $\psi(u) = \text{sign}(u)$. Il est connu sous le nom d'estimateur LAD (least absolute deviation). Une variante de l'estimateur LAD est celle proposée par Huber (1964) :

$$\rho(u) = \begin{cases} u^2 & \text{si } |u| \leq c \\ c|u| & \text{si } |u| > c \end{cases}$$

Ces deux estimateurs sont moins sensibles aux points aberrants que l'estimateur OLS d'où le nom de régression robuste. On remarque aussi que la régression quantile que nous avons présentée au paragraphe 1.3. à la page 76 est un cas particulier de régression robuste avec $\rho(u) = u \times (\alpha - \mathbf{1}\{u < 0\})$ et $\psi(u) = \alpha - \mathbf{1}\{u < 0\}$. Notons que si $\alpha = 0,5$ nous obtenons la régression médiane et $\rho(u) = u \times (0,5 - \mathbf{1}\{u < 0\}) = 0,5 \times |u|$. L'estimateur de la régression médiane correspond donc à l'estimateur LAD. Nous remarquons aussi que :

$$\begin{aligned} \rho(u) &= u \times (\alpha - \mathbf{1}\{u < 0\}) \\ &= \begin{cases} (1 - \alpha)|u| & \text{si } u \leq 0 \\ \alpha|u| & \text{si } u > 0 \end{cases} \end{aligned}$$

Posons $u = y - q$ avec $q = x^\top \beta$. Nous avons :

$$\sum_{i=1}^n \rho(u_i) = (1 - \alpha) \sum_{y_i < q_i} |y_i - q_i| + \alpha \sum_{y_i > q_i} |y_i - q_i|$$

$q_i = x_i^\top \hat{\beta}$ est bien l'estimateur de Q_i tel que $\Pr\{Y_i \leq Q_i \mid X_i = x_i\} = \alpha$ d'où le nom de régression quantile⁶.

Nous cherchons à estimer le bêta hebdomadaire de certaines actions par rapport à celui du CAC 40 pour la période allant du 1^{er} janvier 1990 au 31 décembre 2009. Nous reportons sur le graphique 1 les estimateurs OLS et LAD ainsi que l'estimateur de la régression quantile pour différentes valeurs de α .

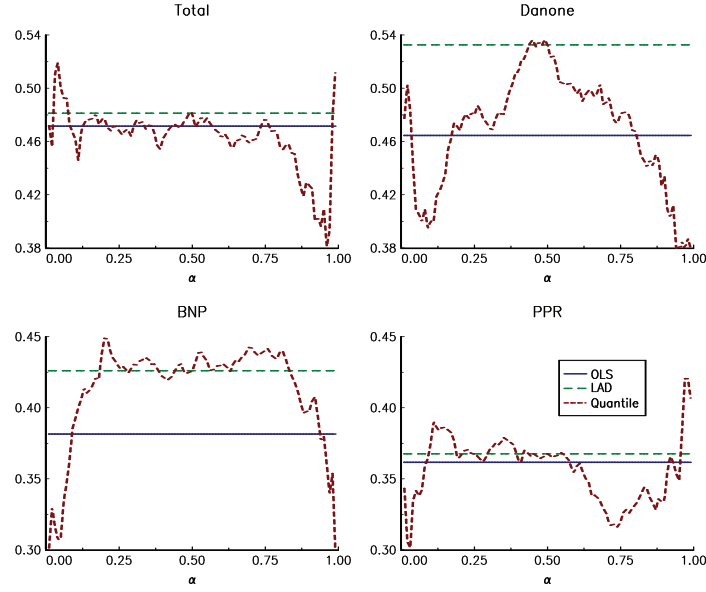
1.2. Le maximum de vraisemblance

1.2.1. Définition de l'estimateur

La méthode du maximum de vraisemblance (*maximum likelihood* ou ML) est une technique très populaire. On rappelle ici quelques résultats que l'on peut trouver dans tous les livres traitant de l'estimation statistique (Davidson et MacKinnon, 1995). Soit θ un vecteur $K \times 1$ de

⁶R. Koenker et G. Basset, « Regression Quantiles », *Econometrica*, 46(1), 1978, p. 33-50.

Graphique 1. Estimation robuste du bêta



paramètres à estimer et Θ l'espace des paramètres. Nous notons $L_i(\theta)$ la vraisemblance de l'observation i , c'est-à-dire la fonction de densité de l'observation i . $\ell_i(\theta) \equiv \ln L_i(\theta)$ est la log-vraisemblance de $L_i(\theta)$. Considérons n observations *iid*, on définit la fonction de log-vraisemblance de l'échantillon comme l'expression $\ell(\theta) = \sum_{i=1}^n \ell_i(\theta)$. $\hat{\theta}_n$ est l'estimateur du maximum de vraisemblance s'il vérifie $\ell(\hat{\theta}_n) \geq \ell(\theta) \forall \theta \in \Theta$. Nous pouvons montrer que $\hat{\theta}_n$ est sans biais, asymptotiquement efficace et possède la propriété de normalité asymptotique :

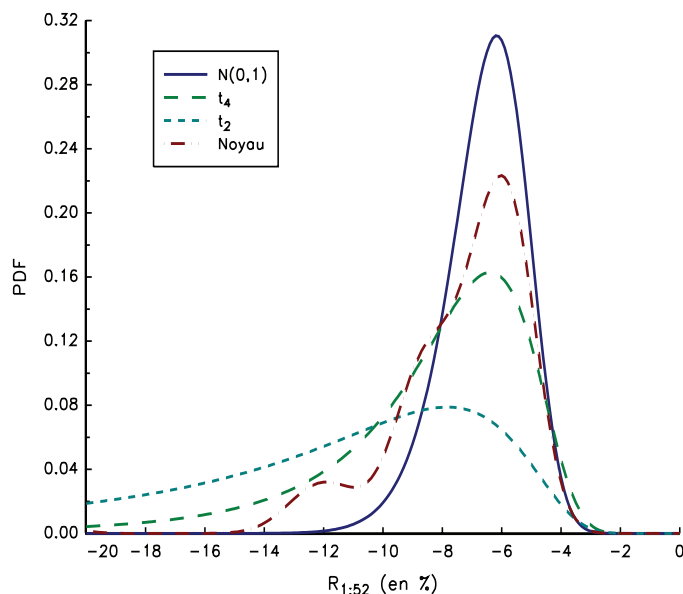
$$\sqrt{n}(\hat{\theta}_n - \theta_0) \rightarrow \mathcal{N}(\mathbf{0}, \mathcal{J}^{-1}(\theta_0))$$

avec $\mathcal{J}(\theta_0)$ la matrice d'information de Fisher définie par :

$$\mathcal{J}(\theta_0) = \mathbb{E} \left[-\frac{\partial^2 \ell_i(\theta_0)}{\partial \theta \partial \theta^\top} \right]$$

et θ_0 la vraie valeur du vecteur des paramètres. Soit $h(\theta)$ une fonction du vecteur des paramètres θ . $h(\hat{\theta}_n)$ converge alors presque sûrement vers $h(\theta_0)$ et nous avons :

$$\sqrt{n}(h(\hat{\theta}_n) - h(\theta_0)) \rightarrow \mathcal{N} \left(\mathbf{0}, \frac{\partial h(\theta_0)}{\partial \theta^\top} \mathcal{J}^{-1}(\theta_0) \frac{\partial h(\theta_0)}{\partial \theta} \right)$$

Graphique 5. Densité de $R_{1:52}$ 

1.5.2. La régression non paramétrique

Considérons maintenant le modèle $y = f(x) + u$. Si $f(x) = x^\top \beta$ (resp. $f(x) = f(x, \beta)$), nous obtenons le modèle paramétrique linéaire (resp. non linéaire) et nous avons $\hat{y} = x^\top \hat{\beta}_{ML}$ (resp. $\hat{y} = f(x, \hat{\beta}_{ML})$). La régression non paramétrique consiste à estimer directement f et nous avons $\hat{y} = \hat{f}(x)$. Par exemple, nous avons présenté le lissage spline à la page 194. Cette technique est une régression non paramétrique. Une autre méthode non paramétrique très connue est la régression polynômiale locale (*local regression* ou *loess*⁵⁵). Dans le cas d'une seule variable exogène, nous avons :

$$y_i = \beta_0(x) + \sum_{j=1}^p \beta_j(x) (x_i - x)^j + u_i$$

Pour une valeur x donnée, nous estimons les paramètres $\hat{\beta}_j(x)$ par moindres carrés pondérés en utilisant les poids⁵⁶ :

$$w_i = \mathcal{K} \left(\frac{x - x_i}{h} \right)$$

⁵⁵W.S. Cleveland et S.J. Devlin, « Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting », *Journal of the American Statistical Association*, 83(403), 1988, p. 596-610.

⁵⁶W. Härdle, *Applied Nonparametric Regression*, Cambridge University Press, 1992.

Nous en déduisons que l'estimation $\hat{y} = \hat{f}(x)$ est :

$$\begin{aligned}\hat{y} &= \hat{f}(x) \\ &= \mathbb{E} \left[\hat{\beta}_0(x) + \sum_{j=1}^p \hat{\beta}_j(x) (x-x)^j + u_i \right] \\ &= \hat{\beta}_0(x)\end{aligned}$$

Contrairement au modèle linéaire classique, la régression non paramétrique nécessite la résolution d'un problème de moindres carrés pour chaque valeur de x . Elle est donc plus gourmande en temps de calcul. La méthode précédente s'applique aussi à la régression quantile⁵⁷.

Considérons le modèle non linéaire :

$$y = (\cos(2\pi x - \pi) + 1) \varepsilon$$

avec $\varepsilon \sim \mathcal{U}_{[0,1]}$. Nous avons :

$$m(x) = \mathbb{E}[y|x] = \frac{1}{2} (\cos(2\pi x - \pi) + 1)$$

et :

$$q(x; \alpha) = F_{y|x}^{-1}(\alpha) = \alpha (\cos(2\pi x - \pi) + 1)$$

Nous pouvons estimer la fonction $m(x)$ par régression non paramétrique et la fonction $q(x; \alpha)$ par régression quantile non paramétrique. Sur le graphique 6, nous présentons un exemple avec 1 000 points (x, y) simulés. Nous comparons $m(x)$ et $q(x; \alpha)$ avec les fonctions estimées $\hat{m}(x)$ et $\hat{q}(x; \alpha)$ pour différentes valeurs d'ordre p de la régression polynomiale⁵⁸.

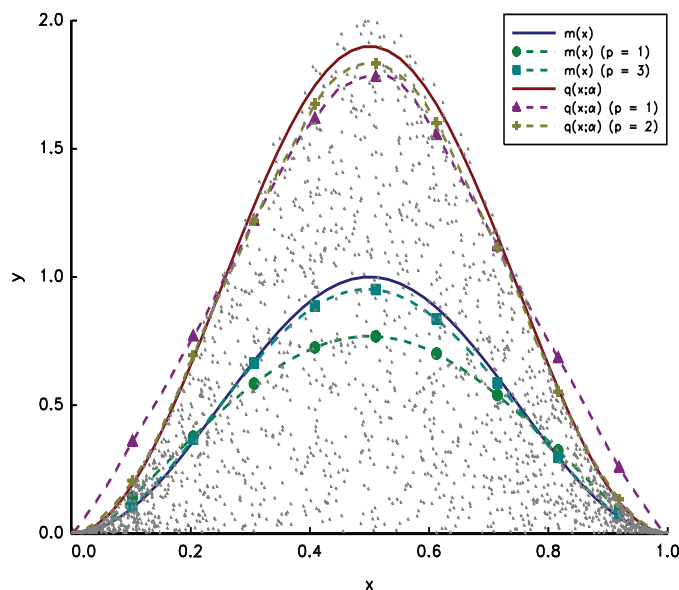
2. Modélisation de la dépendance statistique

Dans cette section, nous abordons le problème de la dépendance entre plusieurs variables aléatoires. Il existe principalement deux cadres d'analyse pour résoudre ce type de problème. Le premier considère une approche gaussienne et consiste à estimer une matrice de covariance ou de corrélation. Le second considère une approche plus générale et utilise les fonctions copules. Celles-ci sont utilisées en finance depuis peu longtemps et définissent la statistique exhaustive minimale de dépendance au sein d'un vecteur aléatoire.

⁵⁷Voir le paragraphe consacré à la régression quantile non paramétrique à la page 77.

⁵⁸ α est fixé à 95%.

Graphique 6. Régression non paramétrique



2.1. Modélisation des matrices de covariance et de corrélation

2.1.1. L'estimateur du maximum de vraisemblance

On considère un vecteur gaussien $X \sim \mathcal{N}(\mu, \Sigma)$ de dimension n . La log-vraisemblance de l'échantillon $\{x_1, \dots, x_T\}$ est égale à :

$$\ell = -\frac{nT}{2} \ln 2\pi - \frac{T}{2} \ln |\Sigma| - \frac{1}{2} \sum_{t=1}^T (x_t - \mu)^\top \Sigma^{-1} (x_t - \mu)$$

L'estimateur ML de μ vérifie $\partial_\mu \ell = \mathbf{0}$ et nous obtenons $\hat{\mu} = \bar{x}$. En utilisant les propriétés de la trace matricielle, nous avons :

$$\begin{aligned} \ell &= -\frac{nT}{2} \ln 2\pi - \frac{T}{2} \ln |\Sigma| - \frac{1}{2} \sum_{t=1}^T \text{tr} \left((x_t - \bar{x})^\top \Sigma^{-1} (x_t - \bar{x}) \right) \\ &= -\frac{nT}{2} \ln 2\pi - \frac{T}{2} \ln |\Sigma| - \frac{1}{2} \sum_{t=1}^T \text{tr} \left(\Sigma^{-1} (x_t - \bar{x})^\top (x_t - \bar{x}) \right) \\ &= -\frac{nT}{2} \ln 2\pi - \frac{T}{2} \ln |\Sigma| - \frac{1}{2} \text{tr} (\Sigma^{-1} S) \end{aligned}$$

avec :

$$S = \sum_{t=1}^T (x_t - \bar{x})^\top (x_t - \bar{x})$$

Nous en déduisons que :

$$\partial_{\Sigma^{-1}} \ell = \frac{T}{2} \Sigma - \frac{1}{2} S = 0$$

et nous obtenons finalement :

$$\hat{\Sigma} = \frac{1}{T} S$$

L'estimateur du maximum de vraisemblance de Σ est donc la matrice de covariance empirique.

2.1.2. La prise en compte d'une structure de corrélation

L'estimateur du maximum de vraisemblance suggère que nous pouvons estimer la matrice de corrélation de $X \sim \mathcal{N}(\mu, \Sigma)$ à partir des données centrées $\tilde{x}_{i,t} = \hat{\sigma}_i^{-1}(x_{i,t} - \hat{\mu}_i)$. Dans ce cas, nous notons $C = (\rho_{i,j})$ la matrice de corrélation. Nous obtenons :

$$\begin{aligned} \hat{C} &= \arg \max \ell(C) \\ \text{s.c. } & C \text{ est une matrice de corrélation} \end{aligned}$$

avec :

$$\ell = -\frac{nT}{2} \ln 2\pi - \frac{T}{2} \ln |C| - \frac{1}{2} \sum_{t=1}^T \tilde{x}_t^\top C^{-1} \tilde{x}_t$$

Il peut être pertinent de prendre en compte une structure particulière de corrélation. Par exemple, nous pouvons imposer :

$$C = C_n(\rho) = \begin{pmatrix} 1 & & \rho \\ & \ddots & \\ \rho & & 1 \end{pmatrix}$$

Dans ce cas, on parle de modèle à un seul facteur. Pour comprendre cette structure de corrélation, nous considérons le modèle suivant :

$$X_i = \sqrt{\rho} X + \sqrt{1-\rho} \varepsilon_i \quad i = 1, \dots, n$$

où X et ε_i sont des variables aléatoires gaussiennes indépendantes centrées et réduites. Nous interprétons X comme le facteur commun et ε_i comme le facteur idiosyncratique. Nous en déduisons que :

$$\begin{aligned} \mathbb{E}[X_i X_j] &= \rho \mathbb{E}[X^2] + (1-\rho) \mathbb{E}[\varepsilon_i \varepsilon_j] + \\ &\quad (\mathbb{E}[X \varepsilon_i] + \mathbb{E}[X \varepsilon_j]) \sqrt{\rho(1-\rho)} \\ &= \rho \end{aligned}$$

et $\mathbb{E}[X_i X_i] = 1$. La corrélation entre X_i et X_j est donc égale à ρ . Considérons maintenant le modèle suivant :

$$X_i = \sqrt{\rho} X + \sqrt{\rho_{(f(i))} - \rho} X_{(f(i))} + \sqrt{1 - \rho_{(f(i))}} \varepsilon_i \quad i = 1, \dots, n$$

3.1.3. Le lissage

Soit une date t^* donnée. Nous notons :

$$a_{t|t^*} = E_{t^*} [\alpha_t]$$

et :

$$P_{t|t^*} = \mathbb{E}_{t^*} \left[(a_t - \alpha_t) (a_t - \alpha_t)^\top \right]$$

pour $t \leq t^*$. Nous avons bien sûr $a_{t^*|t^*} = a_{t^*}$ et $P_{t^*|t^*} = P_{t^*}$. Le lissage est alors donné par le système récursif suivant :

$$\begin{aligned} P_t^* &= P_t T_{t+1}^\top P_{t+1|t}^{-1} \\ a_{t|t^*} &= a_t + P_t^* (a_{t+1|t^*} - a_{t+1|t}) \\ P_{t|t^*} &= P_t + P_t^* (P_{t+1|t^*} - P_{t+1|t}) P_t^{*\top} \end{aligned}$$

3.2. Quelques applications

3.2.1. Les moindres carrés récursifs

Considérons le modèle linéaire classique :

$$y_t = x_t^\top \beta + \varepsilon_t$$

avec x_t un vecteur de K variables exogènes et $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$. Ce modèle est un cas particulier d'un modèle espace-état. L'équation de mesure est en fait :

$$y_t = x_t^\top \beta_t + \varepsilon_t$$

tandis que l'équation de transition du vecteur d'état β_t est :

$$\beta_t = \beta_{t-1}$$

En utilisant le filtre de Kalman, nous en déduisons que :

$$\begin{cases} \hat{\beta}_{t|t-1} = \hat{\beta}_{t-1} \\ P_{t|t-1} = P_{t-1} \\ y_{t|t-1} = x_t^\top \hat{\beta}_{t|t-1} \\ v_t = y_t - y_{t|t-1} \\ F_t = x_t^\top P_{t|t-1} x_t + \sigma^2 \\ \hat{\beta}_t = \hat{\beta}_{t|t-1} + P_{t|t-1} x_t F_t^{-1} v_t \\ P_t = (I_K - P_{t|t-1} x_t F_t^{-1} x_t^\top) P_{t|t-1} \end{cases}$$

Considérons le vecteur $Y_t = (y_1, \dots, y_t)$ et la matrice $X_t = (x_1^\top, \dots, x_t^\top)$. Nous avons :

$$X_t^\top X_t = X_{t-1}^\top X_{t-1} + x_t x_t^\top$$

et :

$$X_t^\top Y_t = X_{t-1}^\top Y_{t-1} + x_t y_t$$

Supposons que $P_t = \sigma^2 (X_t^\top X_t)^{-1}$, nous en déduisons que :

$$F_t = \sigma^2 \left(1 + x_t^\top (X_{t-1}^\top X_{t-1})^{-1} x_t \right)$$

Posons $\lambda_t = (X_{t-1}^\top X_{t-1})^{-1} x_t$. En utilisant la formule de l'inversion de Sherman-Morrison-Woodbury présentée à la page 121, nous obtenons :

$$(X_t^\top X_t)^{-1} = (X_{t-1}^\top X_{t-1})^{-1} - \frac{\lambda_t \lambda_t^\top}{1 + \lambda_t^\top x_t}$$

Cette équation correspond exactement à l'équation du filtre de Kalman :

$$\begin{aligned} P_t &= (I_K - P_{t|t-1} x_t F_t^{-1} x_t^\top) P_{t|t-1} \\ &= \sigma^2 (X_{t-1}^\top X_{t-1})^{-1} - \frac{\sigma^2 (X_{t-1}^\top X_{t-1})^{-1} x_t x_t^\top (X_{t-1}^\top X_{t-1})^{-1} \sigma^2}{\sigma^2 \left(1 + x_t^\top (X_{t-1}^\top X_{t-1})^{-1} x_t \right)} \\ &= \sigma^2 \left((X_{t-1}^\top X_{t-1})^{-1} - \frac{\lambda_t \lambda_t^\top}{1 + \lambda_t^\top x_t} \right) \end{aligned}$$

ce qui prouve que l'hypothèse $P_t = \sigma^2 (X_t^\top X_t)^{-1}$ est vérifiée. Nous en déduisons que :

$$\hat{\beta}_t = \hat{\beta}_{t|t-1} + (X_{t-1}^\top X_{t-1})^{-1} x_t \frac{v_t}{1 + \lambda_t^\top x_t}$$

Le filtre de Kalman est donc finalement :

$$\begin{cases} v_t = y_t - x_t^\top \hat{\beta}_{t-1} \\ \lambda_t = (X_{t-1}^\top X_{t-1})^{-1} x_t \\ F_t = \sigma^2 (1 + \lambda_t^\top x_t) \\ \hat{\beta}_t = \hat{\beta}_{t-1} + P_{t-1} x_t F_t^{-1} v_t \\ P_t = P_{t-1} - (1 + \lambda_t^\top x_t)^{-1} \lambda_t \lambda_t^\top \end{cases}$$

Ce sont aussi les équations récursives des moindres carrés ordinaires¹¹. Notons au passage que ce système récursif évite l'inversion des T matrices $X_t^\top X_t$ dans le système $\hat{\beta}_t = (X_t^\top X_t)^{-1} X_t^\top Y_t$ pour $t = 1, \dots, T$.

Nous vérifions qu'à la date terminale de l'échantillon, nous avons :

$$\hat{\beta}_T = \hat{\beta}_{OLS}$$

Considérons maintenant le filtre de lissage. Nous avons :

$$\begin{aligned} P_t^* &= P_t P_{t+1|t}^{-1} \\ &= I \end{aligned}$$

¹¹A. Spanos, *Statistical Foundations of Econometric Modelling*, Cambridge University Press, 1986.

d'où :

$$\begin{aligned}\hat{\beta}_{t|t^*} &= \hat{\beta}_t + P_t^* (\hat{\beta}_{t+1|t^*} - \hat{\beta}_t) \\ &= \hat{\beta}_{t+1|t^*}\end{aligned}$$

et :

$$\begin{aligned}P_{t|t^*} &= P_t + (P_{t+1|t^*} - P_{t+1|t}) \\ &= P_{t+1|t^*}\end{aligned}$$

En particulier, nous avons :

$$\hat{\beta}_{t|T} = \hat{\beta}_{\text{OLS}}$$

et :

$$P_{t|T} = \sigma^2 (X_T^\top X_T)^{-1} = \text{cov}(\hat{\beta}_{\text{OLS}})$$

Afin d'illustrer l'intérêt des moindres carrés récursifs, nous simulons le modèle suivant :

$$\begin{cases} y_t = 2 + t + 3x_t + \varepsilon_t & t = 1, \dots, 100 \\ y_t = 10 + t + 3x_t + \varepsilon_t & t = 101, \dots, 200 \end{cases}$$

avec $\varepsilon_t \sim \mathcal{N}(0, 1)$. La variable exogène x_t est simulée à partir de la distribution $\mathcal{U}_{[0,5]}$. Sur le graphique 2, nous reportons la valeur estimée de la constante c_t . On voit qu'à partir de $t > 100$, on a un changement de comportement de l'estimateur. Si on considère l'évolution de l'innovation v_t ainsi que les intervalles de confiance à 99%, on vérifie le changement structurel du modèle. On identifie généralement les ruptures de tendance des estimateurs par les tests CUSUM et CUSUMSQ¹². Notons $w_t = (1 + \lambda_t^\top x_t)^{-1/2} v_t$ l'innovation standardisée. Sous l'hypothèse $\mathcal{H}_0 : \beta_t = \beta_{t-1}$, la statistique CUSUM définie par $W_t = s^{-1} \sum_{i=1}^t w_i$ avec $s^2 = (T - K)^{-1} \sum_{t=1}^T (y_t - x_t^\top \hat{\beta}_t)^2$ suit la loi $\mathcal{N}(0, t - K)$. La statistique CUSUMSQ correspond à $V_t = \sum_{i=1}^t w_i^2 / \sum_{i=1}^T w_i^2$. Sous \mathcal{H}_0 , V_t suit une distribution Bêta de paramètres $(T - t)/2$ et $(t - K)/2$.

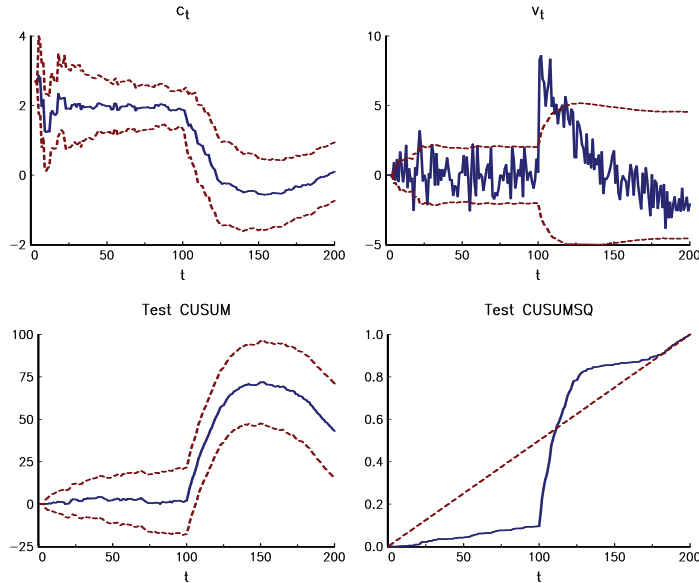
3.2.2. L'estimation des composantes inobservables

Le filtre de Kalman permet d'estimer très facilement les composantes inobservables du vecteur d'état α_t . Prenons l'exemple d'un modèle de tendance déterministe :

$$\begin{cases} y_t = \mu_t + \varepsilon_t \\ \mu_t = \beta t \end{cases}$$

¹²R.L. Brown, J. Durbin et J.M. Evans, « Techniques for Testing the Constancy of Regression Relationships over Time », *Journal of the Royal Statistical Society B*, 37(2), 1975, p. 149-192.

Graphique 2. Moindres carrés récursifs et test CUSUM



avec $\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2)$. Pour estimer la tendance μ_t , il suffit d'estimer le paramètre β par moindres carrés ordinaires et nous avons $\hat{\mu}_t = \hat{\beta}t$. Remarquons que le modèle précédent s'écrit aussi :

$$\begin{cases} y_t = \mu_t + \varepsilon_t \\ \mu_t = \mu_{t-1} + \beta \end{cases}$$

Une façon de modéliser la tendance stochastique est donc de perturber l'équation $\mu_t = \mu_{t-1} + \beta$ par un bruit $\eta_t \sim \mathcal{N}(0, \sigma_\eta^2)$. Nous obtenons alors un modèle espace-état :

$$\begin{cases} y_t = \mu_t + \varepsilon_t \\ \mu_t = \mu_{t-1} + \beta + \eta_t \end{cases}$$

Pour obtenir la correspondance avec notre modèle canonique de la page 349, il suffit de poser $Z_t = 1$, $d_t = 1$, $H_t = \sigma_\varepsilon^2$, $T_t = 1$, $c_t = \beta$ et $Q_t = \sigma_\eta^2$. On peut donc estimer $\hat{\mu}_t$ et $\hat{\mu}_{t|t-1}$ en employant le filtre de Kalman. Supposons maintenant que la pente de la tendance est aussi stochastique :

$$\begin{cases} \mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t \\ \beta_t = \beta_{t-1} + \zeta_t \end{cases}$$

$\{\alpha_{t-1}^{(1)}, \dots, \alpha_{t-1}^{(N_s)}\}$ et des poids $\{w_{t-1}^{(1)}, \dots, w_{t-1}^{(N_s)}\}$ représentant une approximation de $f(\alpha_{t-1} | \theta, \mathcal{I}_{t-1})$ avec N_s le nombre de particules simulées. On peut alors écrire :

$$f(\alpha_t | \theta, \mathcal{I}_t) \propto f(y_t | \alpha_t, \theta) \sum_{i=1}^{N_s} w_{t-1}^{(i)} f(\alpha_t | \alpha_{t-1}^{(i)}, \theta)$$

Le problème se résume alors à déterminer l'échantillon $\{\alpha_t^{(1)}, \dots, \alpha_t^{(N_s)}\}$ ainsi que les poids $\{w_t^{(1)}, \dots, w_t^{(N_s)}\}$. Ces poids peuvent être choisis par la méthode dite d'échantillonnage préférentiel (*Importance Sampling*).

4.1. Échantillonnage préférentiel

Supposons ainsi qu'il existe une distribution *cible* $p(x)$ dont il est difficile de tirer un échantillon, mais qu'il est possible d'évaluer (éventuellement à une constante de normalisation près). Supposons maintenant que l'on connaisse une distribution *instrumentale* $q(x)$ dont il est possible de générer un échantillon $\{x^{(1)}, \dots, x^{(N_s)}\}$. On suppose que la distribution q possède un support englobant le support de p tel que le ratio $q(x)/p(x)$ est borné pour tout x du support. La distribution $p(x)$ peut alors être approchée par :

$$p(x) \approx \sum_{i=1}^{N_s} w^{(i)} \delta(x - x^{(i)})$$

où δ est la fonction de Dirac. Le poids normalisé $w^{(i)}$ associé à $x^{(i)}$ est calculé en deux étapes, en posant d'abord :

$$w^{(i)} = \frac{p(x^{(i)})}{q(x^{(i)})}$$

puis en normalisant chaque poids par $\sum_{i=1}^{N_s} w^{(i)}$.

4.2. Calcul des poids pour les techniques SMC

Dans le cas du calcul des poids d'un filtre particulière, la technique d'échantillonnage préférentiel est appliquée à la distribution jointe de $\alpha_{0:t} = (\alpha_0, \dots, \alpha_t)$:

$$\begin{aligned} f(\alpha_{0:t} | y_t, \mathcal{I}_{t-1}) &= f(\alpha_t | y_t, \alpha_{0:t-1}, \mathcal{I}_{t-1}) f(\alpha_{0:t-1} | \mathcal{I}_{t-1}) \\ &= \frac{f(y_t | \alpha_t) f(\alpha_t | \alpha_{0:t-1})}{f(y_t | \mathcal{I}_{t-1})} f(\alpha_{0:t-1} | \mathcal{I}_{t-1}) \\ &\propto f(y_t | \alpha_t) f(\alpha_t | \alpha_{0:t-1}) f(\alpha_{0:t-1} | \mathcal{I}_{t-1}) \end{aligned}$$

où la dépendance par rapport à θ est implicite. On en déduit que :

$$\begin{aligned}
 w_t^{(i)} &\propto \frac{f(y_t | \alpha_t^{(i)}) f(\alpha_t^{(i)} | \alpha_{0:t-1}^{(i)}) f(\alpha_{0:t-1}^{(i)} | \mathcal{I}_{t-1})}{q(\alpha_t^{(i)}, \alpha_{0:t-1}^{(i)} | y_t, \mathcal{I}_{t-1})} \\
 &= \frac{f(y_t | \alpha_t^{(i)}) f(\alpha_t^{(i)} | \alpha_{0:t-1}^{(i)})}{q(\alpha_t^{(i)} | y_t, \alpha_{0:t-1}^{(i)}, \mathcal{I}_{t-1})} \times \frac{f(\alpha_{0:t-1}^{(i)} | \mathcal{I}_{t-1})}{q(\alpha_{0:t-1}^{(i)} | y_t, \mathcal{I}_{t-1})} \\
 &= w_{t-1}^{(i)} \frac{f(y_t | \alpha_t^{(i)}) f(\alpha_t^{(i)} | \alpha_{0:t-1}^{(i)})}{q(\alpha_t^{(i)} | y_t, \alpha_{0:t-1}^{(i)}, \mathcal{I}_{t-1})} \quad (5.10)
 \end{aligned}$$

où l'on a partitionné la densité instrumentale de la manière suivante :

$$q(\alpha_{0:t} | y_t, \mathcal{I}_{t-1}) = q(\alpha_t, \alpha_{0:t-1} | y_t, \mathcal{I}_{t-1}) \times q(\alpha_{0:t-1} | \mathcal{I}_{t-1})$$

On voit donc naturellement apparaître l'algorithme générique suivant à chaque pas de temps t :

1. on simule $\alpha_t^{(i)} \sim q(\alpha_t^{(i)} | y_t, \alpha_{0:t-1}^{(i)}, \mathcal{I}_{t-1})$;
2. on met à jour le poids $w_t^{(i)}$ en utilisant l'équation (5.10) ;
3. on répète N_s fois les étapes 1 et 2 pour produire un échantillon de taille N_s ;
4. on normalise chaque poids $w_t^{(i)}$ en divisant par $\sum_{i=1}^{N_s} w_t^{(i)}$;
5. le cas échéant, on peut échantillonner de nouveau $\{\alpha_t^{(1)}, \dots, \alpha_t^{(N_s)}\}$ pondérés par $\{w_t^{(1)}, \dots, w_t^{(N_s)}\}$ pour produire un nouvel échantillon avec des poids plus équilibrés.

Un des principaux problèmes des filtres particulières est la dégénérescence des échantillons $\{\alpha_t^{(i)}, w_t^{(i)}\}$ avec le temps. Ce phénomène apparaît lorsque la variance des poids $\{w_t^{(1)}, \dots, w_t^{(N_s)}\}$ est trop grande, avec pour résultat de concentrer l'échantillon autour de quelques particules avec un large poids. À mesure que ce phénomène s'accroît, l'approximation de la densité $f(\alpha_t | \theta, \mathcal{I}_t)$ se dégrade. Pour combattre ce phénomène, il est souvent nécessaire de *remplir* l'échantillon avec de nouvelles particules, par exemple en échantillonnant à nouveau pour ré-équilibrer les poids comme dans l'algorithme précédent²⁶. C'est pourquoi il existe de nom-

²⁶Certains algorithmes plus sophistiqués introduisent des variables auxiliaires pour éviter ce phénomène.

M. Pitt et N. Shephard, « Filtering via Simulation: Auxiliary Particle Filters », *Journal of the American Statistical Association*, 94(446), 1999, p. 590-599.

breuses versions de l'algorithme générique précédent²⁷. Les plus connus sont ceux présentés par Arulampalam *et al.* (2002) : algorithme SIS (*Sequential Importance Sampling*), algorithme GPF (*Generic Particle Filter*), algorithme SIR (*Sampling Importance Resampling*) et algorithme RPF (*Regularized Particle Filter*).

4.3. Quelques exemples

4.3.1. La répliation d'allocation dynamique

Roncalli et Weisang (2008) considèrent un gérant actions qui a une allocation dynamique entre l'indice MSCI USA et l'indice MSCI EMU et cherchent à estimer les positions prises par ce gérant en ne connaissant que la valeur du fonds²⁸. Ils supposent que le gérant change les poids à une fréquence mensuelle. Sur le graphique 8, nous représentons les deux indices, les poids de l'indice MSCI USA du gérant ainsi que la valeur du fonds correspondant. La performance mensuelle R_t du fonds est égale à :

$$R_t = w_t^{(\text{USA})} R_t^{(\text{USA})} + (1 - w_t^{(\text{USA})}) R_t^{(\text{EMU})}$$

avec $R_t^{(\text{USA})}$ et $R_t^{(\text{EMU})}$ les performances mensuelles des indices MSCI USA et MSCI EMU. Comme c'est un fonds actions *long only*, nous avons la contrainte $w_t^{(\text{USA})} + w_t^{(\text{EMU})} = 1$ d'où $w_t^{(\text{EMU})} = 1 - w_t^{(\text{USA})}$. Pour estimer les poids, nous pouvons procéder comme dans le cas du paragraphe sur le bêta alternatif. Nous spécifions :

$$\begin{cases} R_t = \beta_t R_t^{(\text{USA})} + (1 - \beta_t) R_t^{(\text{EMU})} + \varepsilon_t \\ \beta_t = \beta_{t-1} + \eta_t \end{cases}$$

Nous pouvons donc estimer la position $\beta_t = w_t^{(\text{USA})}$ en utilisant le filtre de Kalman avec le modèle espace-état suivant :

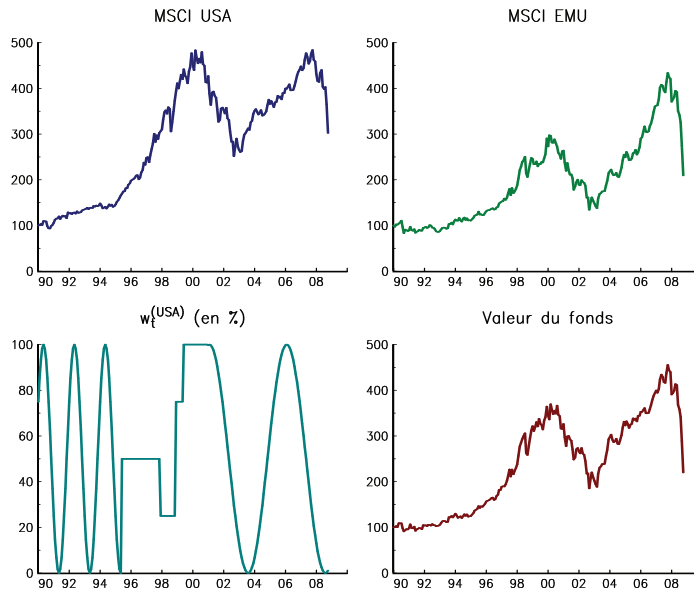
$$\begin{cases} R_t - R_t^{(\text{EMU})} = \beta_t (R_t^{(\text{USA})} - R_t^{(\text{EMU})}) + \varepsilon_t \\ \beta_t = \beta_{t-1} + \eta_t \end{cases}$$

Sous la forme générique, la variable d'observation y_t est égale à $R_t - R_t^{(\text{EMU})}$ et nous avons $Z_t = R_t^{(\text{USA})} - R_t^{(\text{EMU})}$. Nous avons donc seulement deux paramètres à estimer par maximum de vraisemblance σ_ε et σ_η . Nous estimons ensuite la variable d'état β_t par filtrage de Kalman ou filtrage particulière. Nous obtenons les résultats du graphique 9. Nous notons que les deux filtres donnent de très bons résultats.

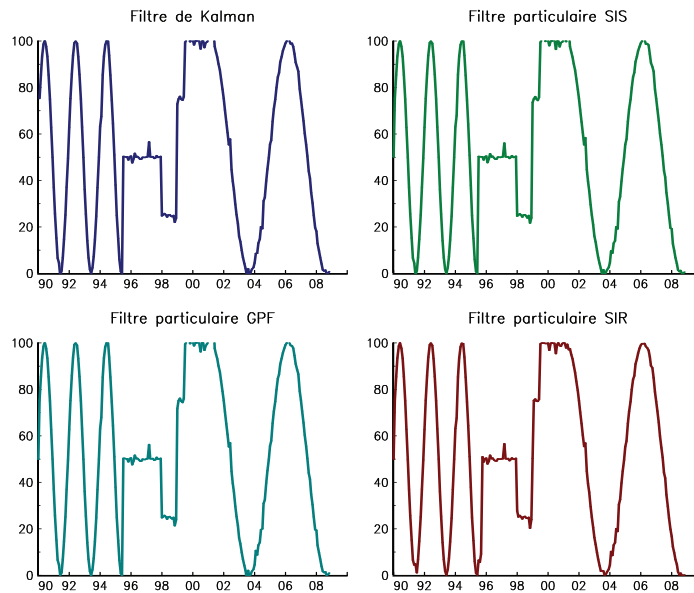
²⁷S. Arulampalam, S. Maskell, N. Gordon et T. Clapp, « A Tutorial on Particle Filters for Online Nonlinear/non-Gaussian Bayesian tracking », *IEEE Transaction on Signal Processing*, 50(2), 2002, p. 174-188.

²⁸T. Roncalli et G. Weisang, « Tracking Problems, Hedge Fund Replication and Alternative Beta », *Working paper*, 2008, ssrn.com/abstract=1325190.

Graphique 8. Un exemple d'allocation dynamique



Graphique 9. Estimation de l'allocation par filtrage



4.3.2. Un modèle non linéaire

On considère l'exemple suivant :

$$\begin{cases} f(\alpha_t | \alpha_{t-1}) = f_{\mathcal{N}}(\alpha_t; T_t(\alpha_{t-1}), Q_t) \\ f(y_t | \alpha_t) = f_{\mathcal{N}}(y_t; \kappa \alpha_t^2, H_t) \end{cases}$$

Cela veut dire que la représentation espace-état est :

$$\begin{cases} \alpha_t = T_t(\alpha_{t-1}) + \eta_t \\ y_t = \kappa \alpha_t^2 + \varepsilon_t \end{cases} \quad (5.11)$$

avec $\eta_t \sim \mathcal{N}(0, Q_t)$ et $\varepsilon_t \sim \mathcal{N}(0, H_t)$. Nous remarquons que l'équation de mesure est non linéaire ainsi que l'équation d'état si la fonction $T_t(\alpha)$ est non linéaire. L'exemple précédent a souvent été étudié avec la spécification suivante²⁹ :

$$T_t(\alpha) = \frac{\alpha}{2} + \frac{25\alpha}{1 + \alpha^2} + 8 \cos(1.2 \times t)$$

En considérant les valeurs numériques suivantes $\kappa = 1/20$, $Q_t = 1$ et $R_t = 10$, nous simulons le modèle (5.11) et nous estimons α_t en considérant les algorithmes des filtres particuliers. Les densités de vraisemblance et de transition *a priori* sont données par $f(y_t | \alpha_t)$ et $f(\alpha_t | \alpha_{t-1})$. Nous supposons que la densité *a posteriori* $f(\alpha_t | \alpha_{t-1}, y_t)$ est égale à $f(\alpha_t | \alpha_{t-1})$, ce qui veut dire que la connaissance de y_t ne permet pas d'avoir plus d'information sur l'état α_t . Pour l'échantillonnage préférentiel, on simule les particules $\alpha_t^{(i)}$ de la façon suivante :

$$\alpha_t^{(i)} = T_t(\alpha_{t-1}^{(i)}) + \eta_t^{(i)}$$

avec $\eta_t^{(i)} \sim \mathcal{N}(0, Q_t)$. Ce schéma est utilisé pour simuler la densité *a priori* $q(\alpha_t | \alpha_{t-1})$ et la densité *a posteriori* $q(\alpha_t | \alpha_{t-1}, y_t)$. Sur le graphique 10, nous reportons l'exemple d'une simulation avec 1 000 particules. Les estimations $\hat{\alpha}_t$ sont égales à :

$$\hat{\alpha}_t = \sum_{i=1}^{N_s} w_t^{(i)} \alpha_t^{(i)}$$

avec $\alpha_t^{(i)}$ la valeur simulée de α_t pour la i -ième particule et $w_t^{(i)}$ le poids de $\alpha_t^{(i)}$. Pour chaque simulation, nous pouvons calculer la statistique RMSE :

$$\text{RMSE} = \sqrt{\frac{1}{T} \sum_{t=1}^T (\alpha_t - \hat{\alpha}_t)^2}$$

²⁹B.P. Carlin, N.G. Polson et D.S. Stoffer, « A Monte Carlo Approach to Nonnormal and Nonlinear State-Space Modeling », *Journal of the American Statistical Association*, 87(418), 1992, p. 493-500.

G. Kitagawa, « Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models », *Journal of Computational and Graphical Statistics*, 5(1), 1996, p. 1-25.

modèle de la régression linéaire devient :

$$y_t = x_t^\top \beta + \varepsilon_t$$

avec :

$$\begin{aligned} \varepsilon_t &= \exp\left(\frac{1}{2}h_t\right) \epsilon_t \\ h_t &= \gamma_0 + \gamma_1 \epsilon_{t-1} + \dots + \gamma_q \epsilon_q + \phi_1 h_{t-1} + \dots + \phi_p h_{t-p} + v_t \end{aligned}$$

Notons que les processus ϵ_t et v_t ne sont pas forcément gaussiens, et que l'on peut utiliser des distributions avec des queues épaisses, plus adaptées aux séries financières. L'estimation des paramètres de ce modèle par maximum de vraisemblance est considérée comme un problème difficile en général car la fonction de vraisemblance est un mélange des distributions des variances conditionnelles h_t . Dans son expression la plus simple, le modèle à *volatilité stochastique canonique* est :

$$\begin{cases} y_t = \exp\left(\frac{1}{2}h_t\right) \epsilon_t \\ h_t = \gamma_0 + \phi_1 h_{t-1} + v_t \end{cases} \quad (5.14)$$

On peut toutefois l'estimer par quasi-maximum de vraisemblance en utilisant le filtre de Kalman³⁵. Pour cela, il faut linéariser le modèle. Notons tout d'abord que nous pouvons paramétrer le modèle (5.14) de la façon suivante :

$$\begin{cases} y_t = \sigma_0 \exp\left(\frac{1}{2}h_t\right) \epsilon_t \\ h_t = \phi_1 h_{t-1} + v_t \end{cases} \quad (5.15)$$

avec $\sigma_0 = \exp(\gamma_0/2)$. Harvey *et al.* (1994) proposent de transformer les observations $y_t = r_t$ de la manière suivante :

$$\ln r_t^2 = \kappa + h_t + \xi_t$$

avec $\kappa = \gamma_0 + \mathbb{E}[\ln \epsilon_t^2]$ et $\xi_t = \ln \epsilon_t^2 - \mathbb{E}[\ln \epsilon_t^2]$. Nous obtenons alors un modèle espace-état :

$$\begin{cases} \ln r_t^2 = \kappa + h_t + \xi_t \\ h_t = \phi_1 h_{t-1} + v_t \end{cases} \quad (5.16)$$

où la variable d'observation est $\ln r_t^2$ et la variable d'état est h_t .

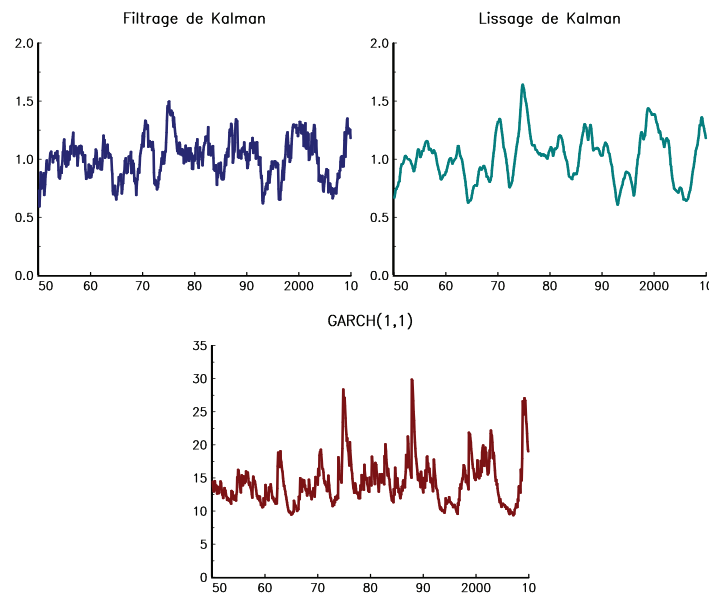
Reprenons l'exemple des rendements mensuels de l'indice S&P 500. Nous estimons les paramètres du modèle (5.16) par maximum de vraisemblance en employant le filtre de Kalman. Nous obtenons :

$$\begin{cases} \ln r_t^2 = -7,766 + h_t + \xi_t \\ h_t = 0,926 \times h_{t-1} + v_t \end{cases}$$

³⁵A. Harvey, E. Ruiz et N. Shephard, « Multivariate Stochastic Variances Models », *Review of Economic Studies*, 61(2), 1994, p. 247-264.

avec $\sigma_v = 0,229$ et $\sigma_\xi = 2,118$. Tous les coefficients sont significatifs au seuil de confiance de 99%. Nous reportons sur le graphique 16 la volatilité estimée par le filtre de Kalman, avant et après lissage, que nous comparons à celle obtenue avec le modèle GARCH(1,1) de la section précédente. Nous notons que les profils de volatilité sont similaires. Le graphique 17 présente des diagnostics sur les résidus standardisés $\epsilon_t^* = r_t \exp(-\frac{1}{2}h_t)$. Nous pouvons noter la présence d'auto-corrélation partielle marginalement significative, ce qui suggère un modèle mal spécifié. Nous pouvons aussi vérifier que les résidus standardisés ϵ_t^* ne sont pas distribués selon une loi normale.

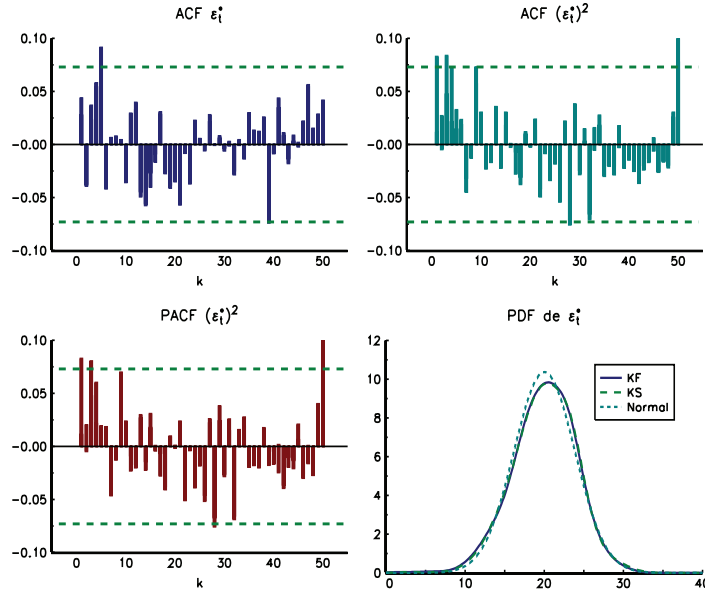
Graphique 16. Volatilité stochastique estimée pour l'indice S&P 500



Remarque 44. Le modèle canonique de volatilité stochastique a donné lieu à de nombreuses extensions, comme la prise en compte d'asymétrie³⁶. Notons aussi que l'on peut employer d'autres méthodes que celle du filtre de Kalman pour estimer ce modèle³⁷.

³⁶A.C. Harvey et N. Shephard, « Estimation of an Asymmetric Stochastic Volatility Model for Asset Returns », *Journal of Business and Economic Statistics*, 14(4), 1996, p. 429-434.

³⁷C. Broto et E. Ruiz, « Estimation Methods for Stochastic Volatility Models: A Survey », *Journal of Economic Surveys*, 18(5), 2004, p. 613-649.

Graphique 17. Diagnostics des résidus standardisés ϵ_t^* 

5.2.2. Estimation du modèle à volatilité stochastique canonique par MCMC

Supposons maintenant que l'on désire estimer le modèle à volatilité stochastique suivant en utilisant la méthode MCMC :

$$r_t = \beta + \varepsilon_t \quad (5.17)$$

avec :

$$\begin{aligned} \varepsilon_t &= \exp\left(\frac{1}{2}h_t\right) \epsilon_t \\ h_t &= \gamma + \phi h_{t-1} + \sigma_v v_t \end{aligned}$$

où ϵ_t et η_t sont deux bruits blancs gaussiens de moyenne nulle et de variance unitaire tels que $\mathbb{E}[\epsilon_t \eta_{t+l}] = 0$ pour tout l et $\mathbb{E}[\epsilon_t \epsilon_{t+l}] = 0$ pour tout $l \neq 0$. Les paramètres de ce modèle sont la constante β , les coefficients γ et ϕ , la volatilité σ_v . Nous devons aussi estimer la volatilité stochastique h_t pour $t = 1, \dots, T$ et la position initiale h_0 . L'information connue *a priori* sur le modèle, ainsi que l'incertitude associée, est donnée par les distributions *a priori* :

– l'incertitude associée à h_0 est traduite par :

$$h_0 \sim \mathcal{N}(m_0, C_0)$$

- la distribution *a priori* de β est donnée par une loi normale :

$$\beta \sim \mathcal{N}(\beta_0, A_0)$$

- Notons $\omega = (\gamma, \phi)^\top$. La distribution *a priori* de (ω, σ_v^2) est donnée par une loi normale-inverse-gamma $(\omega, \sigma_v^2) \sim \mathcal{NIG}(\omega_0, \Omega_0, \nu_0, s_0^2)$ ou encore :

$$\begin{aligned}\omega | \sigma_v^2 &\sim \mathcal{N}(\omega_0, \sigma_v^2 \Omega_0) \\ \sigma_v^2 &\sim \mathcal{IG}(\nu_0/2, \nu_0 s_0^2/2)\end{aligned}$$

Les hyper-paramètres associés sont donc $m_0, C_0, \beta_0, A_0, \omega_0, \Omega_0, \nu_0$ et s_0^2 . Notons $\mathbf{h} = (h_1, \dots, h_T)^\top$ et $\mathbf{r} = (r_1, \dots, r_T)^\top$. On désire estimer la distribution *a posteriori* $f(\beta, \omega, \sigma_v^2, \mathbf{h} | \mathbf{r})$ des paramètres du modèle³⁸. On fait l'hypothèse que cette distribution peut être décomposée de la manière suivante :

$$f(\beta, \omega, \sigma_v^2, \mathbf{h} | \mathbf{r}) = f(\mathbf{h} | \beta, \omega, \sigma_v^2, \mathbf{r}) \times f(\omega, \sigma_v^2 | \beta, \mathbf{h}, \mathbf{r}) \times f(\beta | \mathbf{h}, \mathbf{r})$$

On peut donc appliquer l'échantillonnage de Gibbs. Cela revient à simuler à l'itération i les paramètres suivants :

1. le vecteur \mathbf{h} des volatilités stochastiques :

$$\mathbf{h}_i \sim f(\mathbf{h} | \beta_{i-1}, \omega_{i-1}, \sigma_{v,i-1}^2, \mathbf{r})$$

2. le vecteur (ω, σ_v^2) des paramètres :

$$(\omega_i, \sigma_{v,i}^2) \sim f(\omega, \sigma_v^2 | \beta_{i-1}, \mathbf{h}_i, \mathbf{r})$$

3. la constante β :

$$\beta_i \sim f(\beta | \mathbf{h}_i, \mathbf{r})$$

Pour les distributions *a priori* sélectionnées ci-dessus pour les paramètres ω, σ_v^2 et β , il existe des distributions *a posteriori* dites *conjuguées*³⁹ :

$$- f(\omega, \sigma_v^2 | \beta, \mathbf{h}, \mathbf{r})$$

Posons $X = (\mathbf{1}, (h_0, \dots, h_{T-1})^\top)$. Il est possible de démontrer que la distribution *a posteriori* $f(\omega, \sigma_v^2 | \beta, \mathbf{h}, \mathbf{r})$ est une loi normale-inverse-

³⁸Notons que le vecteur \mathbf{h} des volatilités stochastiques est ici considéré comme un paramètre du modèle.

³⁹Dans le paradigme bayésien, si la distribution *a posteriori* appartient à la même famille de distributions que la distribution *a priori*, les distributions sont dites *conjuguées*, la distribution *a priori* est appelée distribution *a priori conjuguée* de la fonction de vraisemblance.

$S(z_t) = z_t - z_{t-1}$ est appelée la *forme stationnaire* de z_t et nous avons :

$$f_{S(z)}(\lambda) = (1 - 2\theta_1 \cos \lambda + \theta_1^2) \frac{\sigma_u^2}{2\pi} + 2(1 - \cos \lambda) \frac{\sigma_v^2}{2\pi}$$

Dans le cas général, si la forme stationnaire d'un processus X_t existe, il doit exister un polynôme $P(L)$ tel que :

$$S(X_t) = P(L) X_t$$

Les modèles structurels

Les modèles structurels sont des modèles à composantes inobservables (Harvey, 1990). Ils admettent donc généralement une représentation espace-état. Ils sont la plupart du temps non stationnaires, c'est pourquoi il est nécessaire de trouver la forme stationnaire du processus pour calculer la densité spectrale associée. Le modèle *Local Level* (LL) est défini par :

$$\begin{cases} y_t = \mu_t + \varepsilon_t \\ \mu_t = \mu_{t-1} + \eta_t \end{cases}$$

avec $\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ et $\eta_t \sim \mathcal{N}(0, \sigma_\eta^2)$. La forme stationnaire de y_t est :

$$\begin{aligned} S(y_t) &= y_t - y_{t-1} \\ &= \eta_t + (1 - L) \varepsilon_t \end{aligned}$$

Nous en déduisons que :

$$2\pi f_{S(y)}(\lambda) = \sigma_\eta^2 + 2(1 - \cos \lambda) \sigma_\varepsilon^2$$

Le modèle *Local Linear Trend* (LLT), que nous avons vu à la page 355, peut être mis sous la forme suivante :

$$\begin{cases} y_t = \mu_t + \varepsilon_t \\ \mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t \\ \beta_t = \beta_{t-1} + \zeta_t \end{cases}$$

avec $\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2)$, $\eta_t \sim \mathcal{N}(0, \sigma_\eta^2)$ et $\zeta_t \sim \mathcal{N}(0, \sigma_\zeta^2)$. La forme stationnaire de y_t est :

$$\begin{aligned} S(y_t) &= (1 - L)^2 y_t \\ &= \zeta_{t-1} + (1 - L) \eta_t + (1 - L)^2 \varepsilon_t \end{aligned}$$

Nous en déduisons que :

$$2\pi f_{S(y)}(\lambda) = \sigma_\zeta^2 + 2(1 - \cos \lambda) \sigma_\eta^2 + 4(1 - \cos \lambda)^2 \sigma_\varepsilon^2$$

Le modèle *Basic Structural Model* (BSM) a pour expression :

$$\begin{cases} y_t = \mu_t + \gamma_t + \varepsilon_t \\ \mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t \\ \beta_t = \beta_{t-1} + \zeta_t \\ \sum_{i=0}^{s-1} \gamma_{t-i} = \omega_t \end{cases}$$

avec $\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2)$, $\eta_t \sim \mathcal{N}(0, \sigma_\eta^2)$, $\zeta_t \sim \mathcal{N}(0, \sigma_\zeta^2)$ et $\omega_t \sim \mathcal{N}(0, \sigma_\omega^2)$. La forme stationnaire de y_t est :

$$\begin{aligned} S(y_t) &= (1-L)(1-L^s)y_t \\ &= (1-L)(1-L^s)\varepsilon_t + (1-L^s)\eta_t + \\ &\quad \sum_{i=1}^s L^s \zeta_{t-i} + (1-2L+L^2)\omega_t \end{aligned}$$

Nous en déduisons que :

$$\begin{aligned} 2\pi f_{S(y)}(\lambda) &= g_\lambda((1-L)(1-L^s))\sigma_\varepsilon^2 + g_\lambda(1-L^s)\sigma_\eta^2 + \\ &\quad g_\lambda\left(\sum_{i=1}^s L^s\right)\sigma_\zeta^2 + g_\lambda(1-2L+L^2)\sigma_\omega^2 \end{aligned}$$

avec :

$$\begin{aligned} g_\lambda(1-L^s) &= 2(1-\cos s\lambda) \\ g_\lambda(1-2L+L^2) &= 6-8\cos\lambda+2\cos 2\lambda \\ g_\lambda\left(\sum_{i=1}^s L^s\right) &= s+2\sum_{j=1}^{s-1}(s-j)\cos j\lambda \\ g_\lambda((1-L)(1-L^s)) &= 4(1-\cos s\lambda)(1-\cos\lambda) \end{aligned}$$

Une variante du modèle BSM a été utilisée par Lehman Brothers pour développer le modèle Cassandra de prévision de la volatilité implicite des cours de change. Dans ce modèle, la volatilité implicite Σ_t se décompose de la façon suivante :

$$\Sigma_t = \mu_t + \beta_t + \gamma_t + \varepsilon_t$$

avec μ_t l'équilibre de long terme, β_t l'effet de retour à la moyenne et γ_t la composante saisonnière. Plus formellement, nous avons :

$$\begin{cases} \mu_t = \mu_{t-1} + \eta_t \\ \beta_t = \phi\beta_{t-1} + \zeta_t \\ \sum_{i=0}^{s-1} \gamma_{t-i} = \omega_t \end{cases}$$

avec η_t , ζ_t et ω_t trois bruits blancs indépendants de variance σ_η^2 , σ_ζ^2 et σ_ω^2 . μ_t est donc une marche aléatoire, β_t est un processus AR(1) et γ_t est un processus saisonnier stochastique puisque nous avons $\gamma_t = \gamma_{t-s} + \omega_t - \omega_{t-1}$. Notons que si $\sigma_\omega^2 = 0$, alors la composante saisonnière est déterministe ($\gamma_t = \gamma_{t-s}$). Pour estimer ce modèle par la méthode du maximum de vraisemblance dans le domaine des fréquences, il nous faut trouver la densité spectrale de la forme stationnaire. Nous avons :

$$\begin{aligned} S(y_t) &= (1-L)(1-L^s)y_t \\ &= (1-L)(1-L^s)\varepsilon_t + \\ &\quad (1-L^s)\eta_t + \\ &\quad \frac{1-L-L^s+L^{s+1}}{1-\phi L}\zeta_t + \\ &\quad (1-2L+L^2)\omega_t \end{aligned}$$

Nous en déduisons que :

$$2\pi f_{S(y)}(\lambda) = g_\lambda ((1-L)(1-L^s)) \sigma_\varepsilon^2 + g_\lambda (1-L^s) \sigma_\eta^2 + \\ g_\lambda \left(\frac{1-L-L^s+L^{s+1}}{1-\phi L} \right) \sigma_\zeta^2 + g_\lambda (1-2L+L^2) \sigma_\omega^2$$

avec :

$$g_\lambda \left(\frac{1-L-L^s+L^{s+1}}{1-\phi L} \right) = \frac{g_\lambda (1-L-L^s+L^{s+1})}{g_\lambda (1-\phi L)} \\ g_\lambda (1-\phi L) = 1-2\phi \cos \lambda + \phi^2 \\ g_\lambda (1-L-L^s+L^{s+1}) = 4-4\cos \lambda -4\cos s\lambda + \\ 2\cos(s-1)\lambda + 2\cos(s+1)\lambda$$

Les autres expressions ont déjà été données dans le cas du modèle BSM.

Le modèle *Cycle Model* (CM) est défini par :

$$\begin{cases} y_t = \psi_t \\ \begin{bmatrix} \psi_t \\ \psi_t^* \end{bmatrix} = \rho \begin{bmatrix} \cos \lambda_c & \sin \lambda_c \\ -\sin \lambda_c & \cos \lambda_c \end{bmatrix} \begin{bmatrix} \psi_{t-1} \\ \psi_{t-1}^* \end{bmatrix} + \begin{bmatrix} \kappa_t \\ \kappa_t^* \end{bmatrix} \end{cases}$$

avec $\kappa_t \sim \mathcal{N}(0, \sigma^2)$ et $\kappa_t^* \sim \mathcal{N}(0, \sigma^2)$. Harvey (1990) montre que ce processus s'écrit aussi :

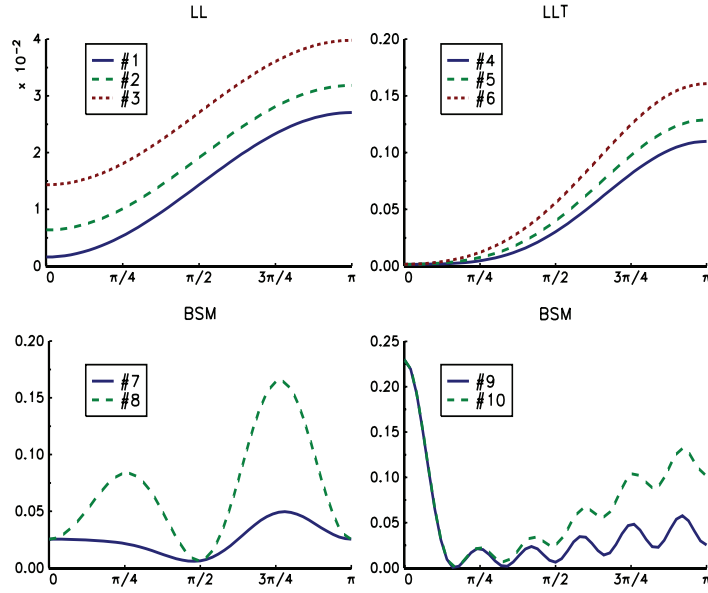
$$y_t = (1-2\rho \cos \lambda_c L + \rho^2 L^2)^{-1} \\ \times ((1-\rho \cos \lambda_c L) \kappa_t + (\rho \sin \lambda_c L) \kappa_t^*)$$

Nous en déduisons que :

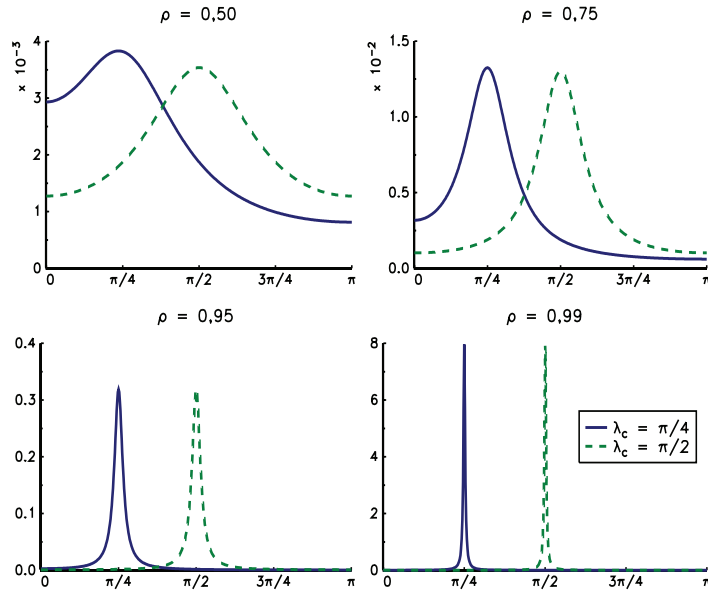
$$2\pi f_y(\lambda) = \frac{1 + \rho^2 - 2\rho \cos \lambda_c \cos \lambda}{1 + \rho^4 + 4\rho^2 \cos^2 \lambda_c - 4\rho(1 + \rho^2) \cos \lambda_c \cos \lambda + 2\rho^2 \cos 2\lambda} \sigma^2$$

Nous avons représenté des différentes densités spectrales sur les graphiques 26 et 27. Les jeux de paramètres utilisés sont les suivants. Pour le modèle LL, $\sigma_\varepsilon = 0,20$ et $\sigma_\eta = 0,10$ pour (#1) – respectivement $\sigma_\eta = 0,20$ pour (#2) et $\sigma_\eta = 0,30$ pour (#3). Pour le modèle LLT, $\sigma_\varepsilon = 0,20$, $\sigma_\zeta = 0,10$ et $\sigma_\eta = 0,10$ pour (#4) – respectivement $\sigma_\eta = 0,20$ pour (#5) et $\sigma_\eta = 0,30$ pour (#6). Pour le modèle BSM, $\sigma_\varepsilon = 0,10$, $\sigma_\eta = 0,10$, $\sigma_\zeta = 0,10$, $\sigma_\omega = 0,10$ et $s = 4$ pour (#7) – respectivement $s = 12$ pour (#9). Pour (#8), nous avons $\sigma_\varepsilon = 0,20$, $\sigma_\eta = 0,30$, $\sigma_\zeta = 0,10$, $\sigma_\omega = 0,10$ et $s = 4$ tandis que pour (#10) nous avons $\sigma_\varepsilon = 0,10$, $\sigma_\eta = 0,10$, $\sigma_\zeta = 0,10$, $\sigma_\omega = 0,20$ et $s = 12$. Il est intéressant de noter l'influence du caractère stochastique de la saisonnalité sur le processus. Pour le modèle cyclique, nous avons pris $\sigma = 0,10$. Nous vérifions que si $\rho \rightarrow 1$, nous obtenons la densité spectrale d'un cycle déterministe, qui est parfaitement localisée à la fréquence λ_c . Plus ρ est faible, moins le processus est bien localisé.

Graphique 26. Densité spectrale des processus LL, LLT et BSM



Graphique 27. Densité spectrale du modèle de cycle stochastique



6.4. L'estimation dans le domaine spectral

6.4.1. Le périodogramme

La transformée de Fourier discrète⁴⁶ d'une série temporelle $\{x_t\}$ avec $t = 1, \dots, n$ est définie par :

$$d(\lambda_j) = \sum_{t=1}^n x_t e^{-i\lambda_j t}$$

pour $\lambda_j = 2\pi j/n$ avec $j \in \{0, 1, \dots, n-1\}$. On construit le périodogramme $I(\lambda_j)$ de la façon suivante :

$$I(\lambda_j) = \frac{|d(\lambda_j)|^2}{2\pi n}$$

Sous certaines conditions⁴⁷, on montre que :

$$\lim_{n \rightarrow \infty} I(\lambda) = f_x(\lambda)$$

Cet estimateur a cependant une variance très importante. C'est pourquoi on préfère utiliser dans la pratique un périodogramme lissé. Celui-ci est défini par :

$$I^*(\lambda_j) = \sum_{s=-m}^{s=m} w_m(s) I(\lambda_j)$$

avec $w_m(s)$ une fenêtre de lissage et m la largeur de la fenêtre. La fonction $w_m(s)$ est égale à $W(s/m)$ avec $W(u)$ une fonction normalisée. Voici les fonctions de lissage les plus utilisées :

$$W(u) = 1 - |u| \quad (\text{Bartlett})$$

$$W(u) = \left(1 - 6|u|^2 + 6|u|^3\right) \cdot \mathbf{1}\left\{|u| < \frac{1}{2}\right\} +$$

$$2(1 - |u|)^3 \cdot \mathbf{1}\left\{|u| \geq \frac{1}{2}\right\} \quad (\text{Parzen})$$

$$W(u) = 1 - 2a + 2a \cos(\pi u) \quad (\text{Tukey})$$

$$W(u) = 1 \quad (\text{Rectangular})$$

$$W(u) = (\pi u)^{-1} \sin(\pi u) \quad (\text{Daniell})$$

$$W(u) = 3 \left((\pi u)^{-3} \sin(\pi u) - (\pi u)^{-1} \cos(\pi u) \right) \quad (\text{Priestley})$$

Sur le graphique 28, nous considérons un processus AR(1) avec $\phi = 0,5$ et $\sigma = 20\%$. Nous représentons sa densité spectrale $f_x(\lambda)$, puis nous

⁴⁶En anglais, cette fonction est appelée dft ou *discrete fourier transform*.

⁴⁷En particulier, le processus doit être stationnaire.

simulons ce processus pour n égal à 1 000 observations. Nous calculons ensuite le périodogramme $I(\lambda)$. Nous remarquons que $I(\lambda)$ est fortement bruité. Nous estimons ensuite la densité spectrale en utilisant la technique de lissage. Nous calculons ainsi $I^*(\lambda_j)$ en considérant la fenêtre de Tukey avec $a = 0,25$ et $m = 50$. Nous obtenons un estimateur beaucoup moins bruité. Afin de voir l'impact du lissage, nous construisons la fonction d'autocorrélation associée. Pour cela, il suffit d'appliquer la transformée de Fourier discrète inverse à $I(\lambda_j)$ et $I^*(\lambda_j)$ pour estimer la fonction d'autocovariance :

$$\hat{\gamma}(k) = \frac{1}{n} \sum_{j=1}^n I(\lambda_j) e^{i\lambda_j k}$$

et de normaliser :

$$\hat{\rho}(k) = \frac{\hat{\gamma}(k)}{\hat{\gamma}(1)}$$

Dans le cas d'un processus AR(1), la fonction d'autocorrélation théorique $\rho(k)$ est égale à ϕ^k . Sur le graphique 29, nous représentons cette fonction ainsi que la fonction d'autocorrélation empirique. Nous reportons aussi la fonction $\hat{\rho}(k)$ estimée à partir du périodogramme. Nous vérifions que celle-ci est exactement égale à la fonction empirique. Nous remarquons que ces deux fonctions ne convergent pas vers 0 lorsque le retard k est relativement grand, ce qui n'est pas le cas de celle estimée à partir du périodogramme lissé avec la fenêtre de Tukey.

6.4.2. La méthode d'estimation de Whittle

Whittle (1953) propose une méthode d'estimation des processus gaussiens stationnaires dans le domaine des fréquences⁴⁸. Considérons le processus gaussien $x = \{x_t, t = 1, \dots, n\}$ avec $x \sim \mathcal{N}(0, \Sigma)$. La fonction de log-vraisemblance du processus x est alors :

$$\ell = -\frac{n}{2} \ln 2\pi - \frac{1}{2} \ln |\Sigma| - \frac{1}{2} x^\top \Sigma^{-1} x \quad (5.19)$$

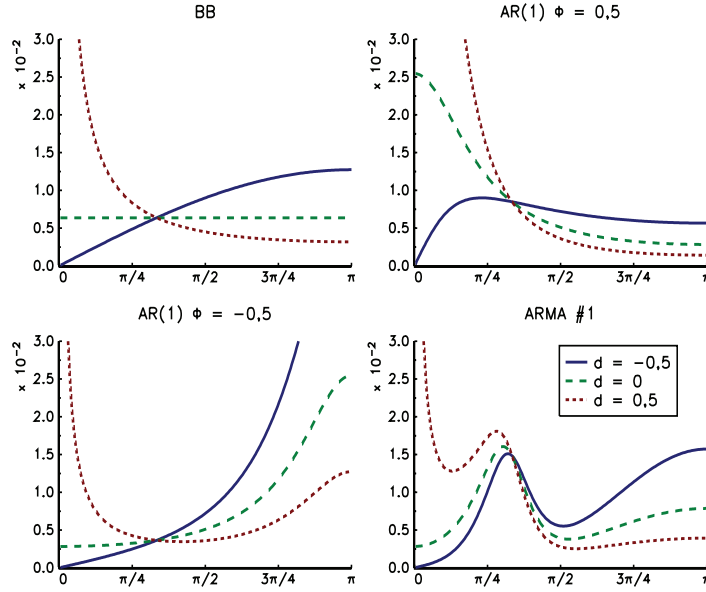
Whittle montre que :

$$\begin{aligned} \ln |\Sigma| &\simeq \sum_{j=0}^{n-1} \ln f(\lambda_j) \\ x^\top \Sigma^{-1} x &\simeq \sum_{j=0}^{n-1} \frac{I(\lambda_j)}{f(\lambda_j)} \end{aligned}$$

⁴⁸P. Whittle, « The Analysis of Multiple Stationary Time Series », *Journal of the Royal Statistical Society B*, 15(1), 1953, p. 125-139.

Y. Pawitan et F. O'Sullivan, « Nonparametric Spectral Density Estimation using Penalized Whittle Likelihood », *Journal of the American Statistical Association*, 89(426), 1994, p. 600-610.

Graphique 35. Densité spectrale de quelques processus ARFIMA



le domaine spectral. Dans le cas du processus ARFIMA avec $d > 0$, on observe le même phénomène que dans le cas du processus fractionnaire pur, c'est-à-dire que les composantes de très basse fréquence dominent très largement la densité spectrale. Pour $\lambda \rightarrow 0$, nous avons donc :

$$\ln f(\lambda) \simeq (\ln \sigma^2 - \ln 2\pi) - d \ln \left(4 \sin^2 \frac{\lambda}{2} \right)$$

Geweke et Porter-Hudak (1983) utilisent ce résultat et proposent d'estimer d par la régression linéaire suivante⁵² :

$$\ln I(\lambda_j) = c - d \ln \left(4 \sin^2 \frac{\lambda_j}{2} \right) + u_j$$

avec $\lambda \leq \lambda_{\min}$. La détermination de l'intervalle $]0, \lambda_{\min}]$ a donné lieu à une abondante littérature sur le sujet⁵³. En général, λ_{\min} est fixé à $2\pi\sqrt{T}$ où T est la longueur de l'historique utilisé pour estimer le périodogramme.

⁵²J. Geweke et S. Porter-Hudak, « The Estimation and Application of Long Memory Time Series Models », *Journal of Time Series Analysis*, 4(4), 1983, p. 221-238.

⁵³C.M. Hurvich et K.I. Beltrao, « Automatic Semiparametric Estimation of the Memory Parameter of a Long-Memory Time Series », *Journal of Time Series*, 15(3), 1994, p. 285-302.

R.T. Baillie, « Long Memory Processes and Fractional Integration in Econometrics », *Journal of Econometrics*, 73(1), 1996, p. 5-59.

Relation avec l'exposant de Hurst

Notons H l'exposant de Hurst. Soit c une constante. Il existe plusieurs façons de caractériser H .

1. La distribution du processus y à la date ct est égale à celle du processus y à la date t multipliée par c^H :

$$y_{ct} \stackrel{d}{=} c^H y_t$$

On dit dans ce cas que le processus est auto-similaire (*self-similar*).

2. Si on considère la distribution asymptotique de ρ_k , on vérifie que :

$$\rho_k \sim ck^{2H-2}$$

3. Du point de vue de la représentation spectrale, on obtient :

$$f(\lambda) \sim c|\lambda|^{-2H-1}$$

4. Nous pouvons aussi définir l'exposant de Hurst par rapport au paramètre fractionnaire puisque nous avons :

$$H = d + \frac{1}{2}$$

Pour estimer ce paramètre, on utilise généralement la statistique R/S (ou *rescaled range*). Soit S_T l'écart-type empirique de $\{y_1, \dots, y_T\}$ et R_T la statistique définie par⁵⁴ :

$$R_T = \max_t \sum_{i=1}^t (y_i - \bar{y}) - \min_t \sum_{i=1}^t (y_i - \bar{y})$$

La statistique R/S que l'on note Q est donc égale à :

$$Q_T = \frac{R_T}{S_T}$$

On montre que $Q_T \sim cT^H$ pour $T \rightarrow \infty$. On peut donc estimer le coefficient H en considérant la régression linéaire suivante :

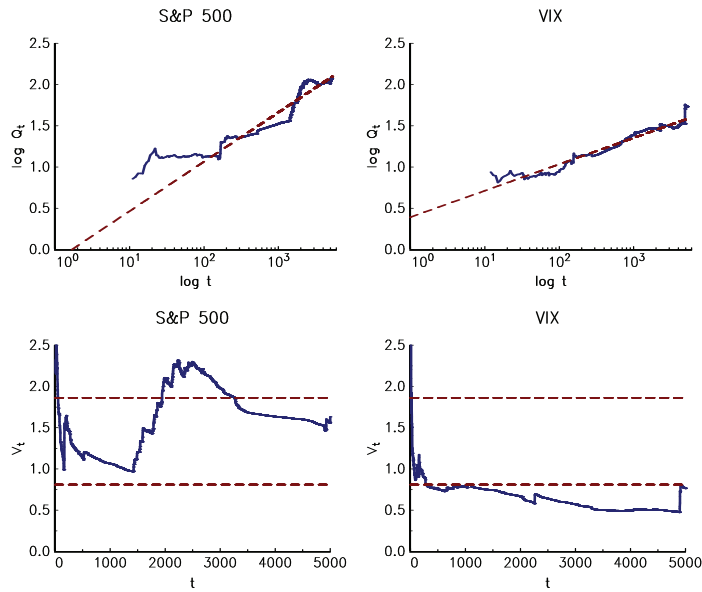
$$\ln Q_t = a + H \ln t + u_t$$

Remarque 48. Notons que l'on peut tester l'hypothèse $H = 0$ en utilisant la statistique $V_t = Q_t/\sqrt{t}$ dont les valeurs critiques sont données dans le tableau 2 de Lo (1991). Par exemple, au seuil de confiance de 95%, les valeurs critiques sont 0,809 et 1,862.

⁵⁴A.W. Lo, « Long-Term Memory in Stock Market Prices », *Econometrica*, 59(5), 1991, p. 1279-1313.

Considérons le rendement journalier du S&P 500 et la variation journalière du VIX de janvier 1990 à décembre 2009. Sur le graphique 36, nous reportons la relation estimée $\ln Q_t = \hat{a} + \hat{H} \ln t$ pour ces deux séries. Nous obtenons $\hat{H} = 0,60$ pour l'indice S&P 500 et $\hat{H} = 0,32$ pour l'indice VIX. En utilisant la statistique V_t , nous ne refusons pas l'hypothèse $H = 0,50$ pour l'indice S&P 500 au seuil de confiance de 95%. Pour l'indice VIX, celle-ci est refusée. Si nous calculons le paramètre fractionnaire implicite, nous obtenons donc $\hat{d} = 0,10$ pour l'indice S&P 500 et $\hat{d} = -0,18$ pour l'indice VIX⁵⁵. Ce dernier présente donc un caractère *mean reverting*.

Graphique 36. Analyse R/S des indices S&P 500 et VIX



6.6.3. La détection et l'identification des cycles

Soit λ_c la fréquence du cycle. Nous avons vu que $f(\lambda_c)$ doit être beaucoup plus élevée que $f(\lambda)$ pour $\lambda \neq \lambda_c$. En 1929, Fisher considère alors la statistique g suivante⁵⁶ :

$$g = \frac{I(\lambda_c)}{\sum_{j=1}^q I(\lambda_j)}$$

⁵⁵L'estimation directe du paramètre fractionnaire par maximum de vraisemblance dans le domaine des fréquences donne respectivement $\hat{d} = -0,05$ et $\hat{d} = -0,15$.

⁵⁶voir Priestley (1994).

avec $q = \lfloor n/2 \rfloor$. Sous l'hypothèse que le processus possède bien un cycle de fréquence λ_c , alors Fisher montre que la distribution de g est :

$$\Pr \{g \geq a\} = 1 - \sum_{j=1}^q (-1)^j \binom{q}{j} (1 - jx)_+^{q-1}$$

Si cette probabilité est plus petite que le niveau α , alors on rejette l'hypothèse d'absence de cycle. Dans la pratique, la fréquence λ_c n'est pas connue, on choisira donc la fréquence correspondant à la plus grande valeur du périodogramme :

$$\lambda_c = \left\{ \lambda : I(\lambda) = \max_{1 \leq j \leq q} I(\lambda_j) \right\}$$

Dans le cas où le nombre d'observations n est grand, Priestley (1994) a montré que la statistique $g^* = 2ng$ a pour distribution $\Pr \{g^* \geq a\} = ne^{-a/2}$.

Nous considérons l'exemple du nombre annuel de peaux de lynx récoltées au Canada entre 1820 et 1940⁵⁷. Nous représentons sur le graphique 37 la série temporelle $x_t = \log(y_t)$ avec y_t le nombre de peaux. Si nous calculons le périodogramme de x_t , nous observons un pic à la fréquence $\lambda_c = 0,6614$. Le test de Fisher prend la valeur $g = 0,59674$, ce qui donne une p -value proche de 0. La présence d'un cycle est donc confirmée et la période du cycle est égale à $p = 2\pi/\lambda_c$ soit 9,5 années. Sur le graphique 37, nous avons représenté ce cycle (quasi décennal) c_t ainsi que la partie résiduelle $x_t - c_t$. Finalement, nous obtenons :

$$\begin{aligned} x_t &= c_t + \varepsilon_t \\ &= 2,90 + 0,607 \times \sin\left(\frac{2\pi}{p}t - 1,13\right) + \varepsilon_t \end{aligned}$$

6.6.4. L'extraction de composantes fréquentielles déterminées

Dans le paragraphe précédent, nous avons volontairement omis d'expliquer comment nous avons calculé la composante c_t . Nous rappelons que la transformée de Fourier du logarithme du nombre de peaux de lynx x_t est :

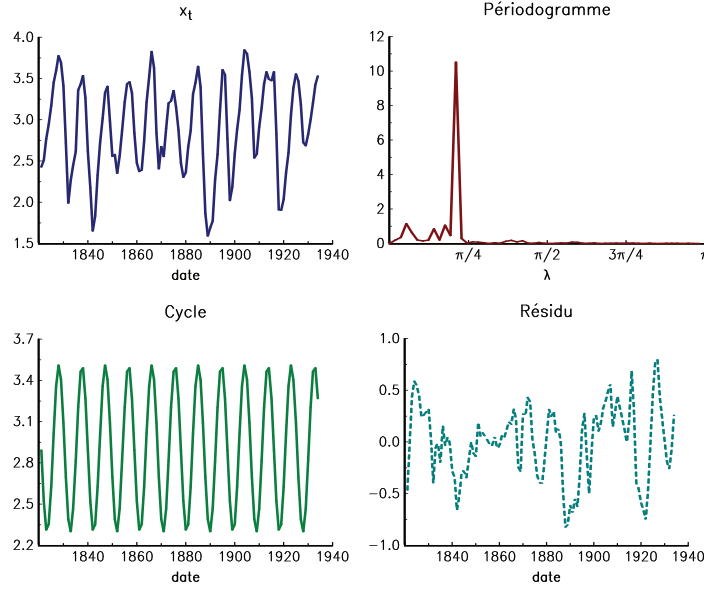
$$d(\lambda_j) = \sum_{t=1}^n x_t e^{-i\lambda_j t}$$

La transformée inverse est donnée par :

$$x_t = \frac{1}{n} \sum_{j=1}^n d(\lambda_j) e^{i\lambda_j t}$$

⁵⁷H. Tong, *Non-Linear Time Series*, Oxford University Press, 1993.

Graphique 37. Détection du cycle des peaux de lynx



Si nous définissons $d_c(\lambda)$ de la façon suivante :

$$d_c(\lambda) = \begin{cases} d(\lambda_c) & \text{si } \lambda = \lambda_c \text{ ou } \lambda = 2\pi - \lambda_c \\ 0 & \text{sinon} \end{cases}$$

alors la composante cyclique est égale à :

$$c_t = \bar{x} + \frac{1}{n} \sum_{j=1}^n d_c(\lambda_j) e^{i\lambda_j t}$$

Nous avons donc estimé la composante cyclique en considérant la transformée de Fourier inverse des coefficients de Fourier correspondant au cycle de fréquence λ_c .

La méthode précédente se généralise facilement. Considérons une partition de $\Lambda = \{\lambda_1, \dots, \lambda_n\}$:

$$\Lambda = \bigcup_{k=1}^m \Lambda_k$$

avec $\Lambda_k \cap \Lambda_{k'} = \emptyset$ et construisons la composante x_t^k en prenant la transformée de Fourier inverse des coefficients $d^k(\lambda)$ définis par :

$$d^k(\lambda) = \begin{cases} d(\lambda) & \text{si } \lambda \in \Lambda_k \\ 0 & \text{sinon} \end{cases}$$

Alors nous avons réussi à décomposer le signal x_t en m sous-signaux :

$$x_t = \sum_{k=1}^m x_t^k$$

Si un processus est composé de plusieurs cycles de fréquences différentes, il est alors très facile de reconstruire ces cycles par l'analyse spectrale. Cette méthode permet donc d'extraire des composantes fréquentielles déterminées. Elle est liée au théorème suivant.

Théorème 2. Soit la série temporelle $\{x_t, t = 1, \dots, n\}$ et $d(\lambda_j)$ la transformée de Fourier avec $\lambda_j = 2\pi j/n$. On passe de la représentation temporelle à la représentation fréquentielle de la façon suivante :

$$\begin{cases} d(\lambda_j) = \sum_{t=1}^n x_t e^{-i\lambda_j t} \\ x_t = n^{-1} \sum_{j=1}^n d(\lambda_j) e^{i\lambda_j t} \end{cases}$$

Et nous avons l'identité suivante concernant la répartition énergétique :

$$\sum_{t=1}^n |x_t|^2 = n^{-1} \sum_{j=1}^n |d(\lambda_j)|^2$$

Ce résultat est connu sous le nom de décomposition de Parseval.

Considérons une série temporelle $x_t = c_t + u_t$ composée d'une composante cyclique de long terme et d'une composante résiduelle. La composante de long terme est la somme de trois cycles c_t^i élémentaires dont les périodes sont supérieures à 5 ans. La composante résiduelle est la somme de 5 cycles de très court terme (inférieurs à 1 an) et d'un bruit gaussien. Nous représentons x_t ainsi que les composantes c_t^i , c_t et u_t sur le graphique 38 pour les 2500 dates. Nous considérons le signal reconstruit avec seulement les m fréquences les plus significatives. Nous avons donc :

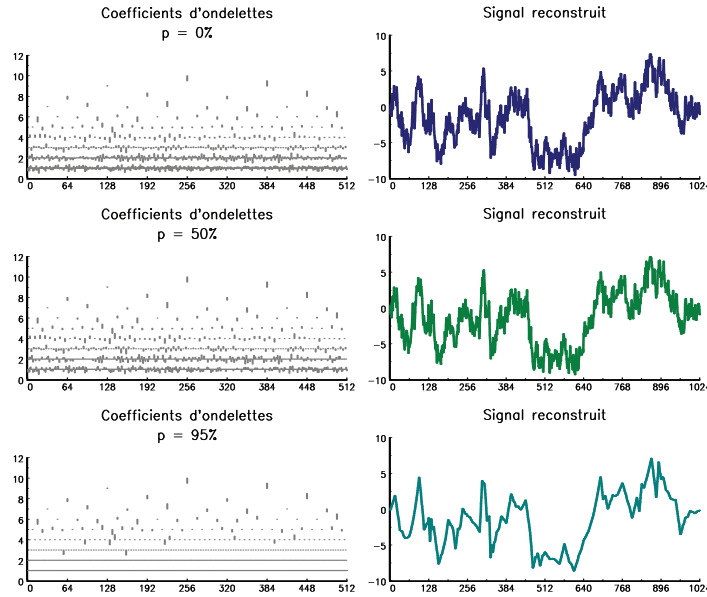
$$x_t^m = n^{-1} \sum_{j=1}^n d^m(\lambda_j) e^{i\lambda_j t}$$

avec $d^m(\lambda_j) = 0$ pour les 2500 fréquences à l'exception des m fréquences qui présentent les coefficients $|d(\lambda_j)|$ les plus élevés. Le graphique 39 présente le signal reconstruit x_t^m pour différentes valeurs de m . On voit donc que nous captions une grande partie de la dynamique de x_t avec peu de coefficients. Une autre façon d'illustrer ceci est d'utiliser le théorème de Parseval. Nous avons :

$$\kappa_m = \frac{\sum_{t=1}^n |x_t^m|^2}{\sum_{t=1}^n |x_t|^2} = \frac{n^{-1} \sum_{k=n-m+1}^n |d_{k:n}|^2}{\sum_{t=1}^n |x_t|^2}$$

avec $d_{k:n}$ la k -ième statistique d'ordre parmi n de $d(\lambda_j)$. Si nous considérons les résultats du tableau 7, nous voyons que nous expliquons 35% de la variance avec seulement une fréquence, 70% de la variance avec seulement deux fréquences, etc. Avec 50 fréquences sur 2500 (soit seulement 2%), nous arrivons à expliquer plus de 96% de la variance.

Graphique 45. Le filtrage par la méthode des seuils



signaux (voir le graphique 46). Le premier signal correspond à la sous-bande W^9 , le deuxième signal est la somme des 2 sous-bandes suivantes, etc. Le quatrième signal est celui défini par les coefficients d'ondelettes $c_{(0)}$, $d_{(0)}$, $d_{(1)}$ et $d_{(2)}$. La somme des 4 signaux correspond bien au signal d'origine.

7.3.2. Le débruitage

On utilise généralement la méthode des seuils pour éliminer le bruit d'un signal. Soient $W = \{w_n\}$ et $\tilde{W} = \{\tilde{w}_n\}$ les vecteurs des coefficients d'ondelettes du signal avant et après le débruitage. Les différentes techniques⁶⁶ de *shrinkage* sont :

- *Hard shrinkage*

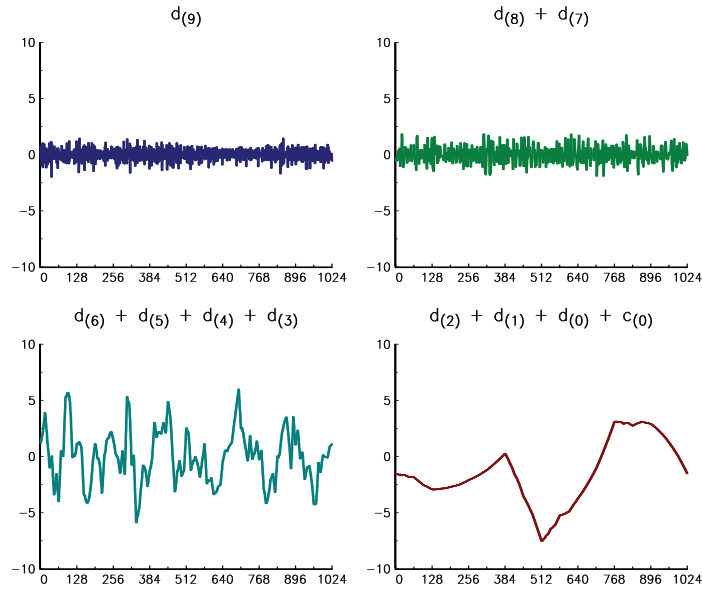
Soit w^+ un scalaire. Nous avons :

$$\tilde{w}_n = w_n \cdot \mathbf{1}\{|w_n| > w^+\}$$

⁶⁶D.L. Donoho et I.M. Johnstone, « Ideal Spatial Adaptation via Wavelet Shrinkage », *Biometrika*, 81(3), 1994, p. 425-455.

D.L. Donoho et I.M. Johnstone, « Adapting to Unknown Smoothness via Wavelet Shrinkage », *Journal of the American Statistical Association*, 90(432), 1995, p. 1200-1224.

Graphique 46. Le codage en sous-bandes



– *Soft shrinkage*

Dans cette méthode, les coefficients \tilde{w}_n sont définis par :

$$\tilde{w}_n = \text{sgn}(w_n) \cdot (|w_n| - w^+)_+$$

– *Quantile shrinkage*

Cette méthode correspond à la méthode *Hard shrinkage* avec w^+ le p -ième quantile des coefficients $|w_n|$.

– *Semi-soft shrinkage*

Soient w_1 et w_2 deux scalaires tels que $w_1 < w_2$. Nous avons :

$$\tilde{w}_n = \begin{cases} 0 & \text{si } |w_n| \leq w_1 \\ \text{sgn}(w_n) (w_2 - w_1)^{-1} w_2 (|w_n| - w_1) & \text{si } w_1 < |w_n| \leq w_2 \\ w_n & \text{si } |w_n| > w_2 \end{cases}$$

En pratique, les coefficients w_n sont réduits (en divisant par l'écart-type) avant d'être transformés.

Considérons le processus x_t défini par :

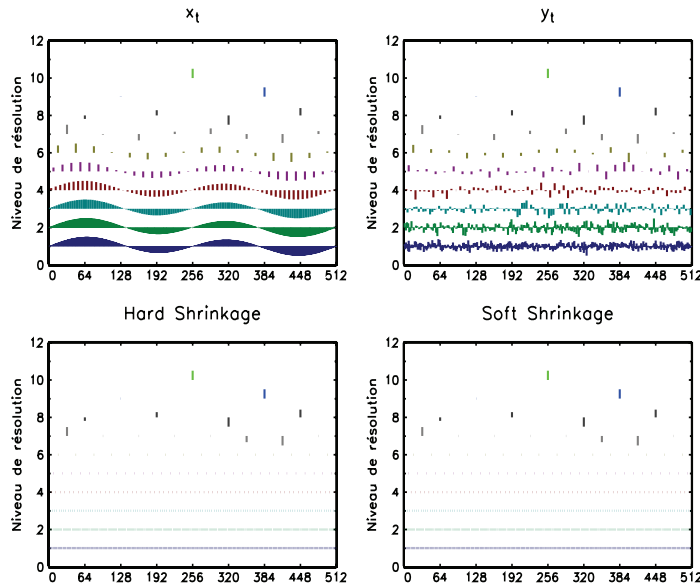
$$x_t = \sin(t) + \sin(2t)$$

Nous construisons le processus bruité y_t de la façon suivante :

$$y_t = x_t + e_t$$

avec $e_t \sim \mathcal{N}(0, 0.4^2)$. Sur le graphique 48, nous représentons y_t ainsi que le processus débruité avec les méthodes *Hard shrinkage* et *Soft shrinkage*. À titre d'information, nous reportons aussi les coefficients \tilde{w}_n utilisés par ces méthodes et nous les comparons aux coefficients w_n sur le graphique 47. Donoho et Johnson (1994) suggèrent d'appliquer les méthodes de débruitage sur les coefficients des premières sous-bandes (et non sur l'ensemble des coefficients d'ondelettes) en considérant $w^+ = \sqrt{2 \ln n_j}$ avec n_j le nombre de coefficients de la j -ième sous-bande. À titre d'illustration, nous reproduisons l'exemple de Donoho et Johnson (1994) sur le graphique 49.

Graphique 47. Coefficients d'ondelettes utilisés pour le débruitage



7.3.3. Les processus de mémoire longue

Une des applications les plus célèbres des ondelettes est l'estimation et la détection des processus de mémoire longue. Dans un article de 1992, Wornell et Oppenheim montrent que les ondelettes sont parfaitement adaptées à ce problème⁶⁷. Considérons un processus y_t défini par $(1 - L)^d y_t = \varepsilon_t$

⁶⁷G.W. Wornell et A. Oppenheim, « Estimation of Fractal Signals from Noisy Measurements using Wavelets », *IEEE Transactions on Signal Processing*, 40(3), 1992, p. 611-623.

avec $\hat{\mu}$ et $\hat{\sigma}$ la tendance et la volatilité réalisée. Nous pouvons montrer que C_t est une fonction croissante puis décroissante du multiplicateur.

3. La valeur du coussin est négativement corrélée à la volatilité réalisée. Supposons que :

$$dS_t = \mu S_t dt + \sigma(t) S_t dW_t$$

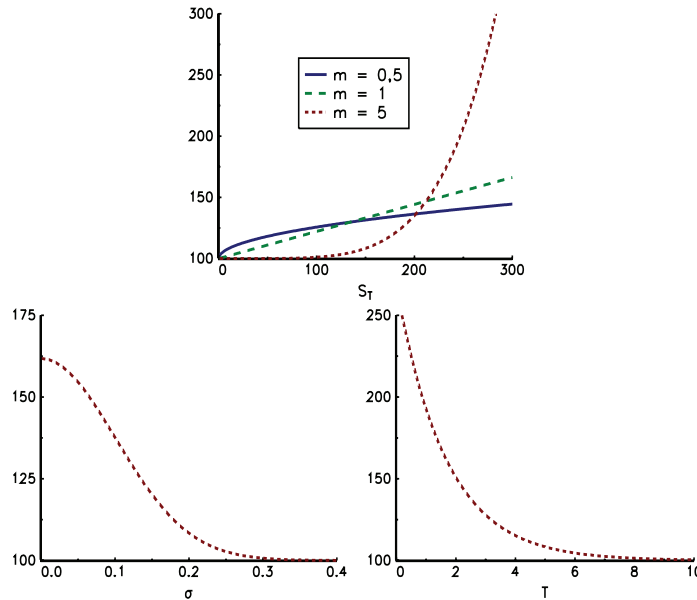
Nous pouvons montrer que :

$$C_t = C_0 \left(\frac{S_t}{S_0} \right)^m \exp \left(-(m-1)rt - \frac{1}{2}(m^2-m) \int_0^t \sigma^2(\theta) d\theta \right)$$

Comme $m > 1$, C_t est une fonction décroissante de la volatilité réalisée $\int_0^t \sigma^2(\theta) d\theta$.

Le graphique 4 illustre ces différentes propriétés.

Graphique 4. Valeur de C_T en fonction de S_T , σ et T



1.1.4. Optimalité de la gestion CPPI

On considère le cas d'un multiplicateur qui varie dans le temps. Nous avons :

$$dC(t) = (m(t)(\mu - r) + r)C(t) dt + m(t)\sigma C(t) dW(t)$$

Posons $e(t) = m(t)C(t)$ la quantité investie à la date t . En reprenant le cadre d'analyse du chapitre 2, nous en déduisons que l'équation HJB devient :

$$\frac{\partial J}{\partial t}(t, c) + \max_{c \in \mathbb{R}} \left[(e(\mu - r) + rc) \frac{\partial J}{\partial c}(t, c) + \frac{1}{2} e^2 \sigma^2 \frac{\partial^2 J}{\partial c^2}(t, c) \right] = 0$$

avec la contrainte $c > 0$. Nous obtenons finalement :

$$e^* = -\frac{(\mu - r)}{\sigma^2} \frac{\partial_c J(t, c)}{\partial_c^2 J(t, c)}$$

En utilisant le même raisonnement que celui mené dans le modèle de Merton au chapitre 2, nous pouvons montrer que $m^*(t)$ est constant dans le cas d'une fonction d'utilité HARA⁵ :

$$U(x) = \frac{(x - g)^\gamma}{\gamma}$$

Nous en déduisons alors que la gestion CPPI est optimale lorsque l'investisseur adopte un programme de maximisation d'utilité HARA. Nous rappelons que, dans le cas d'une fonction d'utilité CRRA, c'est la stratégie *constant mix* qui est optimale.

Remarque 53. *L'équation HJB précédente reste valide si on introduit des contraintes supplémentaires, par exemple une exposition maximale (Black et Perold, 1992). Dans le cas d'une garantie américaine, l'approche par martingale est privilégiée⁶.*

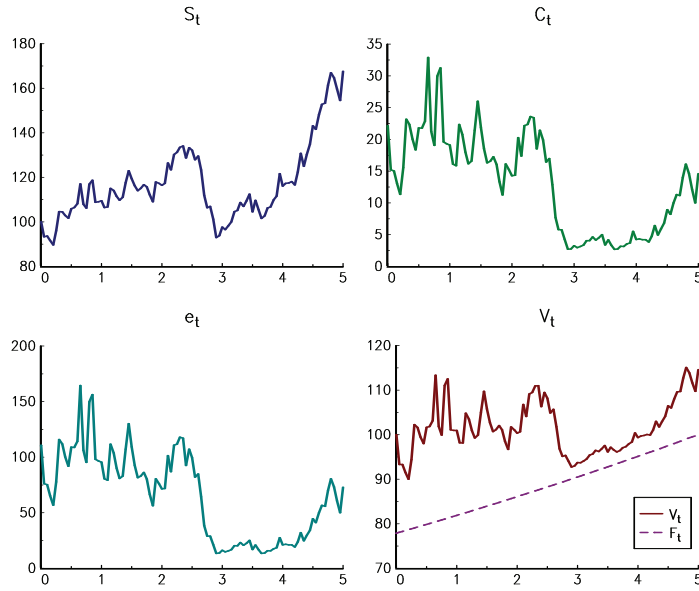
1.1.5. La méthode CPPI en pratique

L'analyse précédente suppose que la trajectoire du sous-jacent est continue. Dans ces conditions, nous avons $\Pr\{C_t < 0\} = 0$. Nous sommes donc dans une situation d'assurance parfaite. Reprenons notre exemple et simulons une trajectoire du processus. Sur le graphique 5, nous représentons la trajectoire S_t ainsi que la valeur du coussin C_t , l'exposition e_t et la valeur de la stratégie CPPI V_t pour $m = 5$. Supposons que l'on utilise un multiplicateur plus élevé, par exemple $m = 7$. Dans ce cas, nous obtenons les résultats du graphique 6. Nous remarquons que lorsque t est égal à 3 ans, la valeur du coussin est très faible. L'exposition est quasi nulle. Néanmoins, nous arrivons à partir de la quatrième année à nous réexposer à l'actif sans risque de façon significative grâce à la forte hausse de S_t . Malheureusement cette situation n'est pas réaliste, car elle suppose

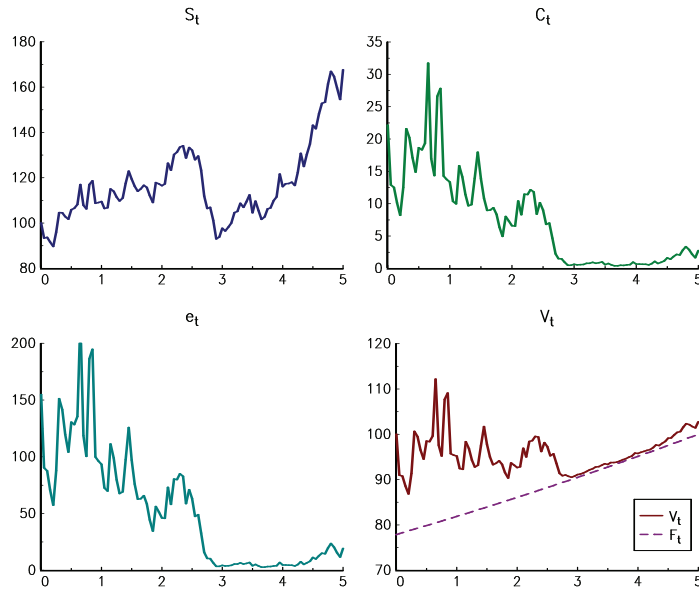
⁵F. Black et A. Perold, « Theory of Constant Proportion Portfolio Insurance », *Journal of Economic Dynamic and Control*, 16(3-4), 1992, p. 403-426.

⁶N. El Karoui, M. Jeanblanc et V. Lacoste, « Optimal Portfolio Management with American Capital Guarantee », *Journal of Economic Dynamics and Control*, 29(3), 2005, p. 449-468.

Graphique 5. Simulation de la stratégie CPPI ($m = 5$)



Graphique 6. Simulation de la stratégie CPPI ($m = 7$)



une gestion parfaitement continue (le gérant réagit instantanément aux fluctuations du marché) et une absence de sauts. Dans la pratique, nous aurions sûrement monétarisé le fonds la troisième année, car il y a un trop grand risque de ne pas se réajuster à temps. Afin de se rapprocher de la réalité, nous considérerons le modèle précédent en temps discret.

En temps discret, lorsqu'on considère une période de rebalancement dt , la valeur du portefeuille évolue de la manière suivante :

$$V_{t+dt} - V_t = (V_t - e_t) \frac{B_{t+dt} - B_t}{B_t} + e_t \frac{S_{t+dt} - S_t}{S_t}$$

avec $B_{t+dt} = (1 + r dt) B_t$ et $F_{t+dt} = (1 + r dt) F_t$. On se donne aussi les valeurs initiales pour les paramètres V_0, B_0 et S_0 . Le gérant garantit un capital G à maturité T . Nous avons donc : $F_0 = G(1 + r dt)^{-T/dt}$.

Reprenons notre exemple avec un niveau de garantie égal à 95 afin de bien comprendre l'influence des différents paramètres m, σ et dt . Nous supposons que l'actif risqué évolue selon un mouvement brownien géométrique⁷. Nous simulons un ensemble de trajectoires. Pour chaque trajectoire simulée, nous appliquons la stratégie du coussin en temps discret pour rebalancer le portefeuille entre l'actif risqué et l'actif sans risque. Avec la méthode Monte-Carlo, nous pouvons alors calculer la probabilité $\Pr\{C_T < 0\}$ d'avoir un coussin négatif avant la maturité. Nous pouvons aussi évaluer le coût de la garantie comme l'espérance des coussins négatifs à maturité T , c'est-à-dire $\mathbb{E}[C_T | C_T < 0]$. Cette quantité (exprimée en % de l'investissement initial) est appelée le risque de gap (*gap risk*). Les graphiques 7 et 8 représentent l'évolution de ces quantités en fonction des paramètres m, σ et dt . Les valeurs par défaut de m et σ sont égales à 5 et 20%. Pour le graphique 7, nous utilisons une période de rebalancement de 30 jours calendaires. Nous vérifions que la probabilité et le coût de la garantie sont des fonctions croissantes de m, σ et dt .

Dans la pratique, nous devons choisir un multiplicateur optimal qui permet de maximiser l'utilité de l'investisseur. Dans le cas d'une fonction d'utilité linéaire, nous avons :

$$m^* = \arg \max \mathbb{E}[C_t]$$

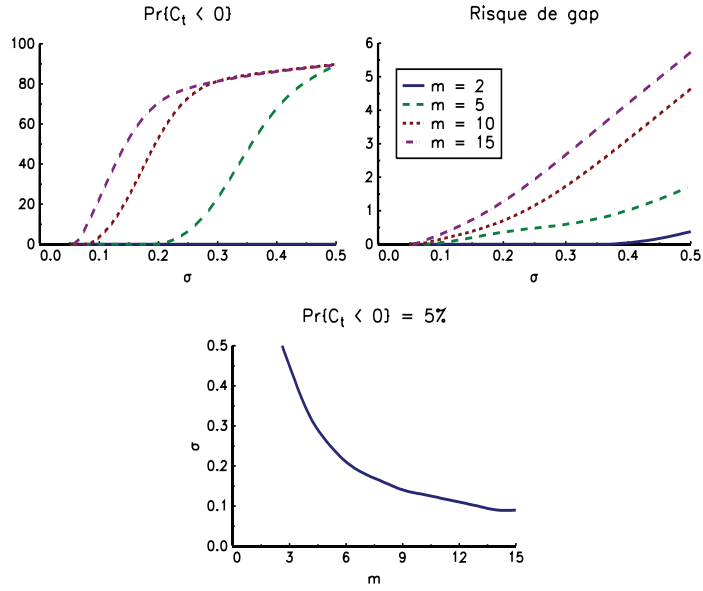
Nous obtenons :

$$\begin{aligned} \mathbb{E}[C_t] &= \exp\left(\left(r + m(\mu - r) - \frac{1}{2}m^2\sigma^2\right)t + \frac{1}{2}m^2\sigma^2t\right) \\ &= \exp((r + m(\mu - r))t) \end{aligned}$$

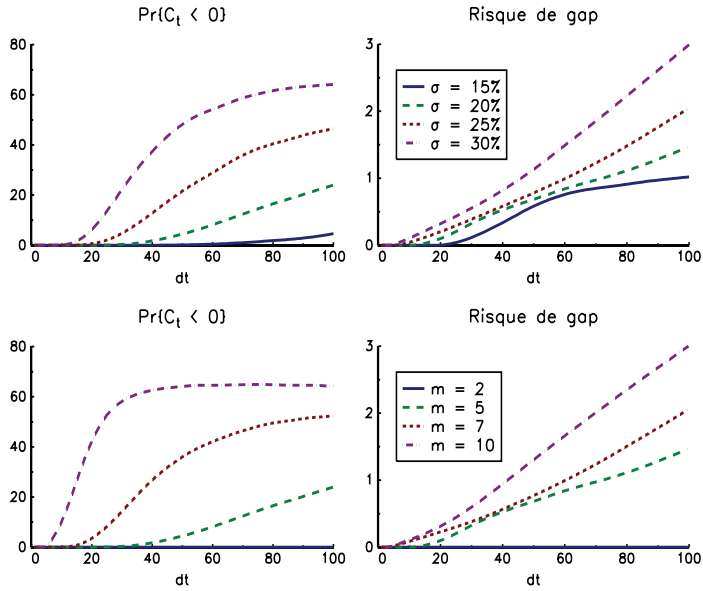
Le multiplicateur optimal m^* est donc $+\infty$. Cette solution n'est pas acceptable, car elle permet de gagner beaucoup avec une probabilité très

⁷Afin d'éliminer l'effet tendance, nous posons $\mu = 0$.

Graphique 7. Influence du paramètre σ ($dt = 30$ jours)



Graphique 8. Influence de la période de rebalancement dt



faible. Si on utilise une fonction d'utilité logarithmique, cela revient à maximiser l'espérance de rendement de V_t ou encore :

$$m^* = \arg \max \mathbb{E} \left[\ln \frac{C_t}{C_0} \right]$$

Nous avons :

$$\mathbb{E} \left[\ln \frac{C_t}{C_0} \right] = \left(r + m(\mu - r) - \frac{1}{2} m^2 \sigma^2 \right) t$$

La condition de premier ordre s'écrit :

$$\frac{\partial}{\partial m} \mathbb{E} \left[\ln \frac{C_t}{C_0} \right] = (\mu - r) t - m \sigma^2 t = 0$$

Nous en déduisons que le multiplicateur optimal est :

$$m^* = \frac{(\mu - r)}{\sigma^2}$$

Ce multiplicateur optimal est donc le ratio de Sharpe de l'actif sans risque divisé par la volatilité. Pour tenir compte de la fréquence de rebalancement, nous constatons que :

$$V_{t+dt} = V_t + (V_t - e_t) r_{t,t+dt} + e_t R_{t,t+dt}$$

où $r_{t,t+dt}$ et $R_{t,t+dt}$ sont les rendements sans risque et risqué entre t et $t + dt$. Nous en déduisons que :

$$C_{t+dt} = C_t (1 + (1 - m) r_{t,t+dt} + m R_{t,t+dt})$$

Nous devons vérifier que le coussin est positif à tout instant :

$$\begin{aligned} C_{t+dt} > 0 &\Leftrightarrow 1 + (1 - m) r_{t,t+dt} + m R_{t,t+dt} > 0 \\ &\Leftrightarrow R_{t,t+dt} > -\frac{1}{m} (1 + (1 - m) r_{t,t+dt}) \end{aligned}$$

En pratique $r_{t,t+dt} \simeq 0$, et donc $R_{t,t+dt} > -m^{-1}$ et $m < -R_{t,t+dt}^{-1}$. Comme ceci doit être vérifié à tout instant, le multiplicateur doit être inférieur à l'inverse du drawdown d de fréquence dt :

$$m < \frac{1}{d}$$

Le multiplicateur $m^+ = d^{-1}$ est appelé le multiplicateur maximum⁸. Au delà de ce multiplicateur, il y a un risque d'avoir un coussin négatif, et donc d'être appelé en garantie. Sur le tableau 2, nous reportons les valeurs

⁸P. Bertrand et J.L. Prigent, « Portfolio Insurance: The Extreme Value Approach to the CPPI Method », *Finance*, 23(2), 2002, p. 68-86.

prises par m^* et m^+ pour différentes valeurs de σ et d . Supposons que le ratio de Sharpe soit égal à 0,25, alors le multiplicateur optimal est égal à 5 si $\sigma = 5\%$. Si le ratio de Sharpe est égal à 1, celui-ci est égal à 20! Considérons un sous-jacent de type actions développées, ce qui correspond à une volatilité de long terme de l'ordre de 20%. Dans ce cas, $m^* = 1,25$ si $sh = 0,25$ et $m^* = 5$ si $sh = 1$. Pour cette classe d'actifs, nous avons des drawdowns journaliers de l'ordre de 10%, ce qui donne un multiplicateur maximum égal à 10. Considérons maintenant un sous-jacent de type hedge funds. Pour une volatilité égale à 10% et un ratio de Sharpe égal à 1, le multiplicateur optimal est égal à 10. Néanmoins, comme la liquidité de ce fonds est mensuelle et qu'il existe des risques extrêmes, nous considérons que le drawdown est égal à 20%, ce qui donne $m^+ = 5$. Dans ce dernier exemple, nous sommes donc obligés d'utiliser un multiplicateur plus petit que celui optimal du fait du risque de gap.

Tableau 2. Calcul des multiplicateurs m^* et m^+

	σ	5%	10%	15%	20%	30%	50%
sh = 0,25	m^*	5	2	1,67	1,25	0,83	0,5
sh = 1,00	m^*	20	10	6,67	5	3,33	2
	d	2%	5%	10%	20%	30%	50%
	m^+	50	20	10	5	3,33	2

1.2. L'approche cœur-satellite

Amenc *et al.* (2004) proposent d'utiliser la méthode CPPI pour construire des stratégies cœur-satellite⁹. Notons B_t , X_t et S_t les processus qui décrivent l'actif sans risque, le cœur et le sous-jacent du satellite. Nous supposons que :

$$\begin{cases} dB_t &= rB_t dt \\ dX_t &= \mu_X X_t dt + \sigma_X X_t dW_t^X \\ dS_t &= \mu_S S_t dt + \sigma_S S_t dW_t^S \end{cases}$$

avec $\mathbb{E}[W_t^X W_t^S] = \rho t$. Nous investissons une proportion α de la richesse initiale dans le cœur et le reste dans une stratégie CPPI portant sur l'actif S_t . Nous avons donc :

$$V_t = \alpha V_0 \frac{X_t}{X_0} + (1 - \alpha) V_0 \frac{C_t}{C_0}$$

La valeur initiale du coussin est $C_0 = (1 - \alpha) V_0$. Par définition, nous avons :

$$V_t \geq \alpha V_0 \frac{X_t}{X_0}$$

⁹N. Amenc, P. Malaise et L. Martellini, « Revisiting core-satellite investing », *Journal of Portfolio Management*, 31(1), 2004, p. 64-75.

Cette garantie est très différente de celle donnée en assurance de portefeuille, puisqu'elle est relative à la valeur du cœur. Considérons un CPPI classique. Nous avons :

$$dC_t = mC_t \frac{dS_t}{S_t} - (m-1)rC_t dt$$

Nous en déduisons que :

$$V_t = \alpha V_0 \frac{X_t}{X_0} + (1-\alpha) V_0 \left(\frac{S_t}{S_0} \right)^m \exp \left(- \left((m-1)r + \frac{1}{2} (m^2 - m) \sigma_S^2 \right) t \right) \quad (6.1)$$

Amenc *et al.* (2004) proposent de financer le levier à partir du cœur et non à partir de l'actif sans risque. Dans cette version du CPPI Long/Short, nous avons :

$$\begin{aligned} dC_t^{LS} &= mC_t^{LS} \frac{dS_t}{S_t} - (m-1)C_t^{LS} \frac{dX_t}{X_t} \\ &= (m\mu_S - (m-1)\mu_X) C_t^{LS} dt + m\sigma_S C_t^{LS} dW_t^S - \\ &\quad (m-1)\sigma_X C_t^{LS} dW_t^X \end{aligned}$$

Nous en déduisons que :

$$C_t^{LS} = C_0^{LS} \exp \left(\mu t + m\sigma_S W_t^S - (m-1)\sigma_X W_t^X \right)$$

avec :

$$\mu = m\mu_S - (m-1)\mu_X - \frac{1}{2}m^2\sigma_S^2 - \frac{1}{2}(m-1)^2\sigma_X^2 + m(m-1)\rho\sigma_S\sigma_X$$

Nous obtenons finalement :

$$V_t^{LS} = \alpha V_0 \frac{X_t}{X_0} + (1-\alpha) V_0 \left(\frac{S_t}{S_0} \right)^m \left(\frac{X_t}{X_0} \right)^{1-m} \exp \left(-\frac{1}{2} (m^2 - m) \sigma^2 t \right) \quad (6.2)$$

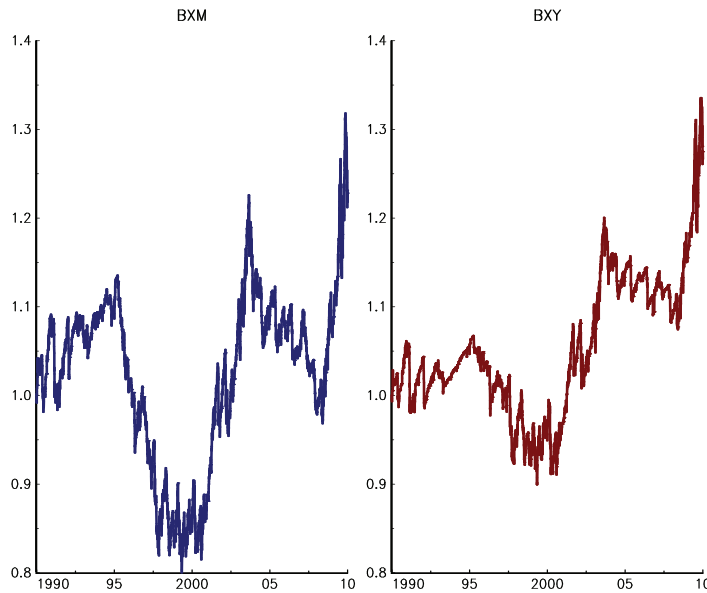
avec :

$$\sigma = \sqrt{\sigma_S^2 + \sigma_X^2 - 2\rho\sigma_S\sigma_X}$$

Si nous comparons les équations (6.1) et (6.2), nous remarquons des effets différents .

1. Dans le cas du CPPI classique, le coussin ne dépend pas de la performance du cœur. Pour le CPPI Long/Short, la valeur du coussin est négativement corrélée à la performance du cœur. Il y a donc un effet amortisseur sur la valeur de V_t^{LS} .
2. La corrélation ρ influence la valeur de V_t dans le cas du CPPI à travers la dynamique jointe de (X_t, S_t) mais n'a aucun effet sur le coussin. Ce n'est pas vrai pour le CPPI Long/Short. Les coûts de gamma sont d'autant plus importants que σ est élevé. Il est donc préférable d'utiliser le CPPI Long/Short lorsque le satellite est positivement corrélé au cœur.

Graphique 28. Comparaison des indices *BXM* et *BXY* avec l'indice *S&P 500*



est respectivement de 9,0% et 9,5% alors qu'elle est seulement de 8,1% pour le S&P 500. De plus, la volatilité de ces indices est beaucoup plus faible. Elle vaut respectivement 13,2% et 15,0%, alors qu'elle est égale à 18,9% pour le benchmark. Ces indices Covered Call présentent donc des ratios de Sharpe long terme bien supérieurs à celui du S&P 500, ce qui en fait de bons candidats pour une allocation stratégique (surtout l'indice BXM).

2.3.2. La stratégie Bull Spread

Description de la stratégie

Nous notons $C(K_c)$ le prix de l'option d'achat de prix d'exercice K_c . $P(K_p)$ est le prix de l'option de vente de prix d'exercice K_p . La stratégie Bull Spread est composée de trois positions :

1. une position longue dans l'actif risqué S_t ;
2. une position courte sur l'option d'achat $C(K_c)$;
3. et une position longue sur l'option de vente $P(K_p)$ avec $K_p < K_c$.

Nous considérons une période avec deux dates 0 et T . Le PnL de cette stratégie est :

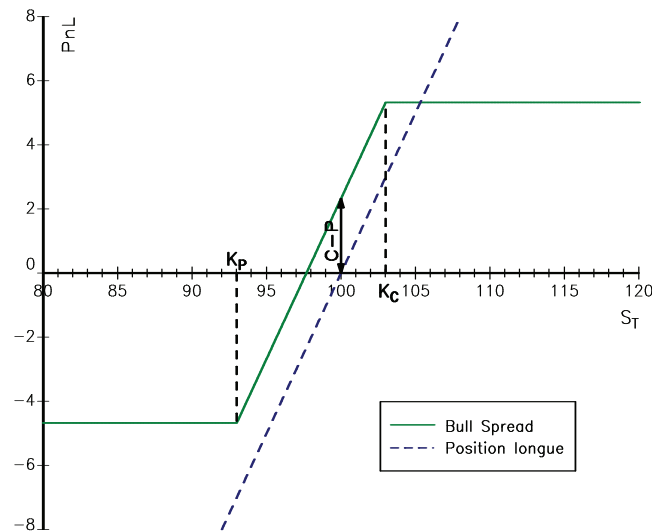
$$\begin{aligned} \text{PnL} = & (S_T - S_0) + \\ & C(K_c) - \max(0, S_T - K_c) + \\ & \max(0, K_p - S_T) - P(K_p) \end{aligned}$$

Nous obtenons trois cas :

$$\text{PnL} = \begin{cases} (K_p - S_0) + C(K_c) - P(K_p) & \text{si } S_T \leq K_p \\ (S_T - S_0) + C(K_c) - P(K_p) & \text{si } K_p < S_T < K_c \\ (K_c - S_0) + C(K_c) - P(K_p) & \text{si } S_T \geq K_c \end{cases}$$

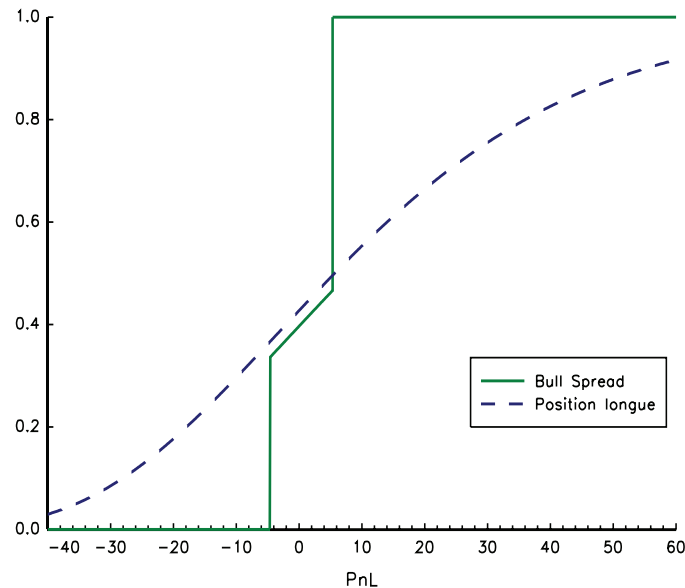
Nous représentons la fonction PnL¹⁷ sur le graphique 29. Notons aussi que les fonctions de distribution des stratégies long only et Bull Spread sont difficilement comparables, puisque les payoffs sont très différents (voir le graphique 30).

Graphique 29. *Payoff de la stratégie Bull Spread*



¹⁷Nous supposons que $S_0 = 100$. Les prix d'exercice des options sont respectivement $K_c = 103$ et $K_p = 93$. La maturité T des options est fixée à un an. Nous considérons une volatilité implicite Σ égale à 30%. Nous supposons aussi que le taux de dividende et le taux sans risque sont nuls.

Graphique 30. Fonction de distribution du PnL



Considérons le cas particulier lorsque nous finançons complètement l'achat de l'option de vente par la vente de l'option d'achat. Cela veut dire que $C(K_c) = P(K_p)$. Sachant un prix d'exercice K_c pour le call, nous pouvons trouver un prix d'exercice K_p qui vérifie $C(K_c) = P(K_p)$, et vice versa. Pour cela, nous devons résoudre une équation linéaire par la méthode de Newton-Raphson ou la méthode de la bi-section. Soient $C_{BS}(K, \Sigma, T)$ et $P_{BS}(K, \Sigma, T)$ les prix du call et du put en fonction des paramètres K , Σ et T . Nous posons $K_p = \kappa(K_c)$:

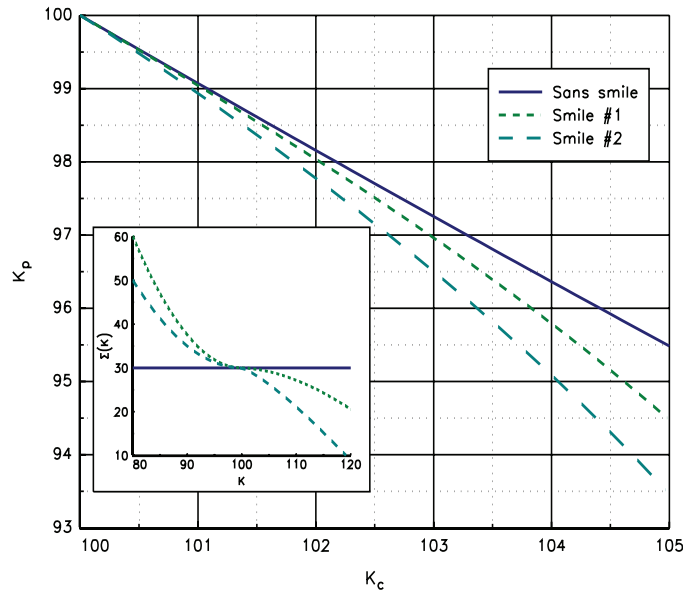
$$\begin{aligned} K_p &= \kappa(K_c) \\ &= \{K : P_{BS}(K, \Sigma, T) = C_{BS}(K_c, \Sigma, T)\} \end{aligned}$$

Sur le graphique 31, nous représentons cette fonction pour l'ensemble précédent des paramètres (avec une volatilité fixée à un mois). Notons que si nous tenons compte du smile, le profil des prix d'exercice est moins favorable à la stratégie Bull Spread. Ainsi, le prix d'exercice implicite du put est plus petit que celui calculé sans l'effet smile.

Rationalité de la stratégie

Cette stratégie est très célèbre depuis la fraude Madoff¹⁸. En effet, elle était présentée comme la principale source de performance des hedge funds

¹⁸P. Clauss, T. Roncalli et G. Weisang, « Risk Management Lessons from Madoff Fraud », in J.J. Choi et M. Papaioannou (eds), *Credit, Currency or Derivatives Ins-*

Graphique 31. Prix d'exercice implicite $K_P = \kappa(K_C)$ 

gérés par Bernard Madoff. Depuis, on sait que cette performance était obtenue par un schéma de Ponzi.

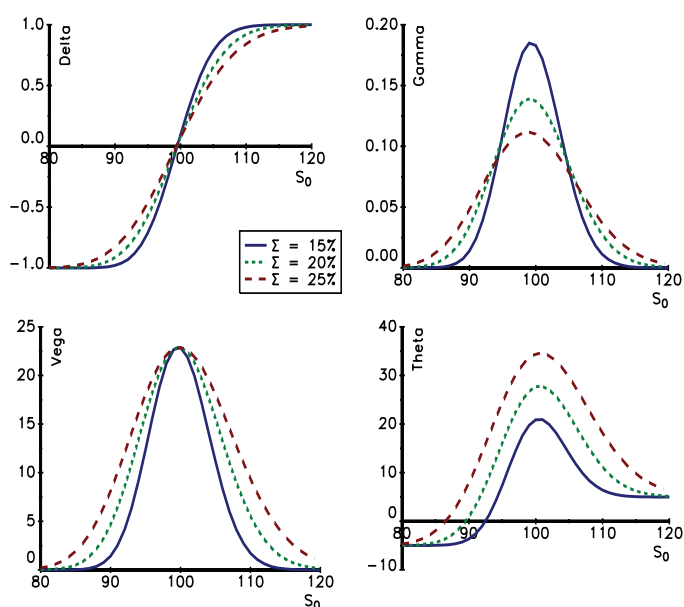
L'idée sous-jacente de la stratégie Bull Spread est d'éliminer partiellement les effets de la queue de distribution de l'actif risqué. Dans le paragraphe suivant, nous présentons un backtest avec une maturité et des prix d'exercice fixes. En réalité, ce type de stratégie est activement géré. Par exemple, le gérant peut choisir une maturité pour le put beaucoup plus lointaine que celle du call afin de minimiser la prime à payer et l'effet du smile. Le gérant pourra aussi revendre le put si le *mark-to-market* de celui-ci est très important, et racheter un put avec un prix d'exercice beaucoup plus bas que précédemment. De même, il pourra racheter le call si le *mark-to-market* est très faible, et revendre un call avec plus de prime. On peut aussi imaginer que le gérant pourra annuler complètement sa position Bull Spread si la valeur *mark-to-market* de la position globale est fortement positive.

Voyons un exemple. Nous supposons que $S_0 = 100$. Les prix d'exercice des options sont respectivement $K_c = 105$ et $K_p = 97$. La maturité T des options est fixée à 3 mois. Nous considérons une volatilité implicite Σ

truments of Global Financial Stability or Crisis ?, International Finance Review, 10, Emerald Group Publishing Limited, 2009.

Une hausse de la volatilité implicite est favorable à la stratégie. Disposant d'un modèle de dynamique de la volatilité implicite, il est alors possible de construire des signaux d'achat et de vente de straddle. La construction de ces modèles étant relativement difficile, nous pouvons nous contenter de modéliser la dynamique de la volatilité historique et considérer que celle-ci est un bon proxy de Σ_t .

Graphique 34. Sensibilités de l'option straddle



3.1.2. Application à l'arbitrage court terme de la volatilité de change

Dans ce paragraphe, nous modélisons la volatilité instantanée des cours de change. Nous construisons ensuite des signaux de volatilité qui sont utilisés dans une stratégie d'achat et de vente d'options straddle de maturité très courte.

On note σ_t la volatilité historique (Σ_t étant la volatilité implicite). Les taux d'intérêt domestique et étranger sont respectivement r et r^* . À la date t , le prix du straddle ATM est égal à W_t . Dans le cas d'une position vendeuse, le PnL est égal à :

$$\text{PnL} = W_t - |S_T - S_0|$$

On propose de calculer le signal comme la probabilité d'avoir un PnL positif :

$$\begin{aligned}\mathcal{SC}_t &= \Pr \{ \text{PnL} \geq 0 \} \\ &= \Pr \{ S^-(\Sigma_t) \leq S_T \leq S^+(\Sigma_t) \}\end{aligned}$$

avec $S^\pm(\Sigma_t) = S_t \pm W_t$. Si on suppose que les rendements du sous-jacent sont indépendants et gaussiens $\ln(S_{t+1}/S_t) \sim \mathcal{N}(\mu_t, \sigma_t^2)$, on a :

$$\mathcal{SC}_t = \Phi\left(\frac{\ln S^+(\Sigma_t) - m_t}{\sigma_t \sqrt{T-t}}\right) - \Phi\left(\frac{\ln S^-(\Sigma_t) - m_t}{\sigma_t \sqrt{T-t}}\right)$$

avec $m_t = \ln S_t + \mu_t(T-t)$. On suppose généralement que les rendements sont centrés — μ_t est égal à 0 dans ce cas-là. Si on se place dans le cas d'un mouvement brownien géométrique, on peut aussi choisir $\mu_t = r - r^* - \sigma_t^2/2$.

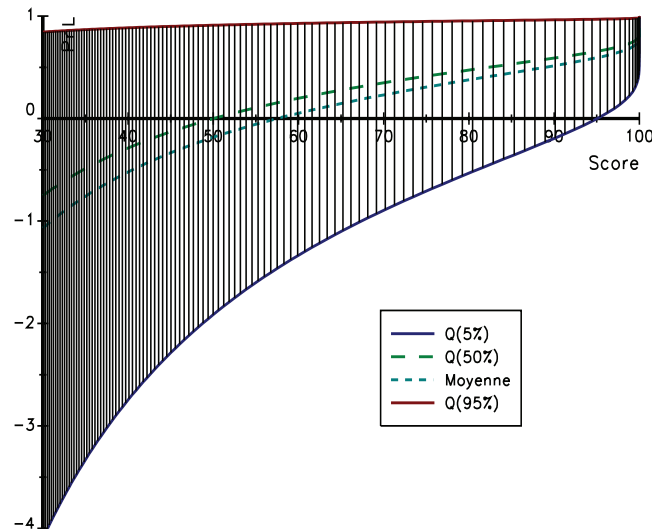
On se fixe ensuite des seuils s^+ et s^- avec $s^+ > s^-$. On vend le straddle si $\mathcal{SC}_t > s^+$ et on l'achète si $\mathcal{SC}_t < s^-$. Le rationnel de cette stratégie est simple. À l'instant t , le marché anticipe une volatilité Σ_t du cours de change entre t et T . Notre estimation σ_t suggère une autre valeur pour la volatilité sur cette même période. On peut par exemple utiliser un modèle GARCH pour estimer cette volatilité. On voit que l'approximation faite par notre modèle, i.e. assimiler la volatilité instantanée GARCH à la volatilité historique réalisée, ne peut marcher que pour des périodes très courtes ($T-t \simeq 0$). Comme on ne couvre pas les positions, il est nécessaire de ne pas laisser dériver la position trop longtemps à cause du décalage du spot.

Considérons une stratégie seulement vendeuse de straddle. Posons $r = r^* = 5\%$. Prenons une volatilité implicite égale à 10% et une maturité égale à un jour. Nous pouvons alors représenter la distribution du PnL pour chaque valeur du score \mathcal{SC}_t . Sur le graphique 35, nous reportons les quantiles 5% et 95%, la médiane et la moyenne du PnL normalisé (c'est-à-dire le PnL rapporté à la valeur de la prime du straddle). Nous simulons cette stratégie de vente de straddle de maturité un jour pour la période 1995-2009 en considérant 8 devises par rapport au dollar. Ces devises sont le dollar australien (AUD), le dollar canadien (CAD), l'euro (EUR), la livre sterling (GBP), le yen (JPY), la couronne norvégienne (NOK), le dollar néo-zélandais (NZD) et la couronne suédoise (SEK). Le graphique 36 présente les PnLs journaliers normalisés en fonction du score \mathcal{SC}_t , alors que le graphique 37 correspond au backtest de la stratégie consistant à vendre un dollar chaque fois que le score \mathcal{SC}_t est supérieur au seuil s^+ . Le cas $s^+ = 0$ correspond à une vente systématique quel que soit le score. Ce type de stratégie vise à capter le spread positif entre la volatilité implicite et ma volatilité réalisée constatée historiquement. On note cependant deux périodes. La période de 1998 à 2004 est très favorable à cette stratégie. Depuis 2004, la rentabilité de cette stratégie est beaucoup plus faible. On

note que fixer un seuil s^+ égal à 50% permet d'éliminer certaines positions perdantes, notamment depuis 2004.

Remarque 57. Une des difficultés dans la construction de ces backtests est la prise en compte des coûts de frottement. Le backtest précédent est obtenu avec les valeurs mid des volatilités implicites ATM sans intégrer de coûts bid-ask. L'intégration de ces coûts pourrait changer complètement la pertinence de la stratégie. De même, le changement de sources de données pourrait affecter les résultats du backtest. Il faut donc toujours être prudent et modeste avec des backtests de stratégie optionnelle. Nous aborderons ce problème plus en détail dans le dernier chapitre.

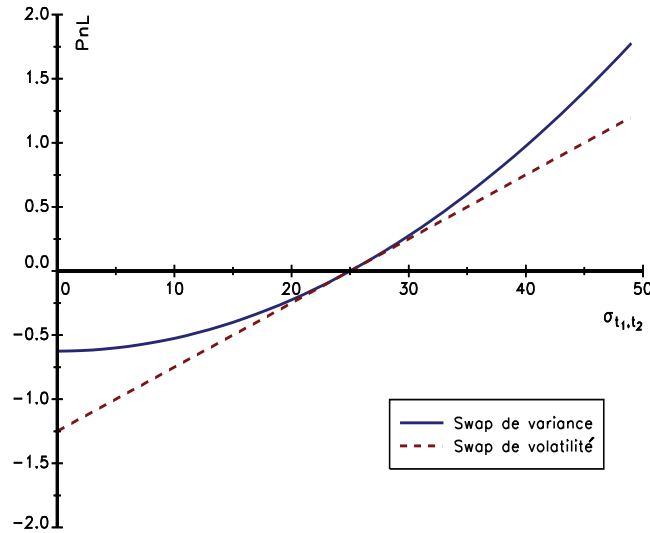
Graphique 35. Relation théorique entre la valeur du score et le PnL



3.2. Les swaps de variance

Un swap de variance est un contrat financier qui permet de s'exposer à la variance réalisée future. En quelques années, c'est devenu un instrument indispensable pour la construction de stratégies de volatilité.

Graphique 38. PnL du swap de variance



position sur le rendement journalier :

$$\text{PnL}(t-1, t) \geq 0 \Leftrightarrow \begin{cases} S_t \geq S_{t-1} e^{K/(100\sqrt{252})} \\ S_t \leq S_{t-1} e^{-K/(100\sqrt{252})} \end{cases}$$

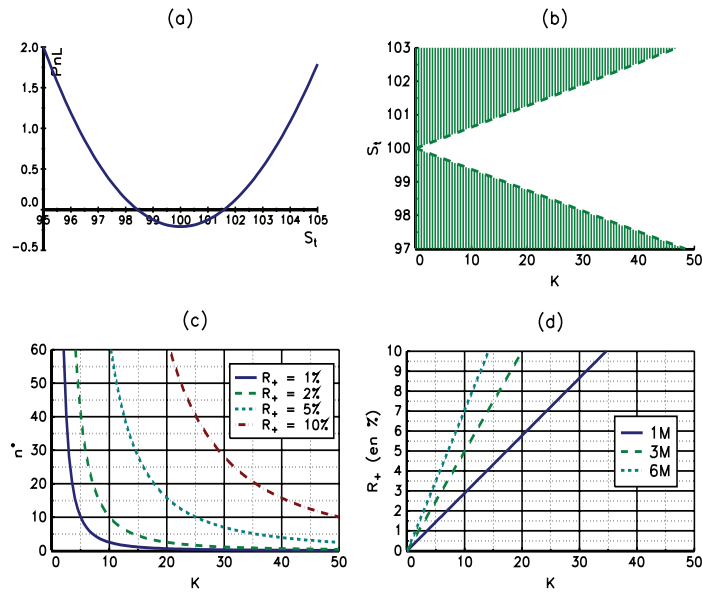
Le PnL journalier est positif si la valeur absolue du rendement journalier excède une certaine quantité de volatilité implicite. Nous représentons sur le graphique 39(b) le cône au-delà duquel le PnL journalier est positif. Par exemple, si le prix d'exercice est égal à 25 et la valeur du sous-jacent est égale à 100, le PnL journalier est positif si le rendement est supérieur à 1,62% ou inférieur à -1,60%. Considérons maintenant le cas où les rendements sont nuls à l'exception d'une journée de trading où nous observons une forte variation du sous-jacent. Notons R_+ et n^* le nombre de journées de trading couvertes par cette variation R_+ . Nous obtenons :

$$n^* \leq \frac{100^2 \times 252 \times R_+^2}{K^2}$$

La statistique n^* est reproduite sur le graphique 39(c). Par exemple, si nous considérons un swap de variance de prix d'exercice 25%, alors une seule variation journalière de $\pm 10\%$ permet d'avoir un PnL positif si la maturité du contrat est inférieure à 41 jours de trading. Une nouvelle écriture de la condition précédente est $R_+ \geq \sqrt{n}K / (100 \times \sqrt{252})$. Cette fonction correspond à celle donnée par le graphique 39(d). Elle indique

pour chaque maturité la variation journalière qui permet d'être sûr que le PnL est positif sur toute la période. Par exemple, si on observe un rendement de $\pm 10\%$ sur une seule journée, on est sûr que le PnL d'un swap de variance de maturité 3 mois et de prix d'exercice 15% est positif.

Graphique 39. Comportement du swap de variance



3.2.2. Valorisation du swap de variance

Considérons une option d'achat de maturité T . Nous rappelons que les expressions du véga et du gamma sont :

$$v = \frac{\partial C}{\partial \sigma} = e^{(b-r)T} \phi(d_1) S_0 \sqrt{T}$$

$$\Gamma = \frac{\partial^2 C}{\partial S^2} = e^{(b-r)T} \phi(d_1) \frac{1}{S_0 \sigma \sqrt{T}}$$

Nous en déduisons que la sensibilité \mathcal{V} de l'option à la variance est :

$$\begin{aligned} \mathcal{V} &= \frac{\partial^2 C}{\partial \sigma^2} \\ &= \frac{1}{2\sigma} \frac{\partial C}{\partial \sigma} \\ &= \frac{1}{2} T S_0^2 \Gamma \end{aligned}$$

Rappelons que la formule de robustesse de Black-Scholes présentée à la page 651 montre que le PnL d'une position optionnelle couverte en delta est sensible à la volatilité réalisée. Nous avons :

$$\begin{aligned} \text{PnL} &= e^{-rT} \int_0^T \frac{1}{2} \Gamma(t) (\Sigma^2(T, K) - \sigma^2(t)) S^2(t) dt \\ &= e^{-rT} \int_0^T \frac{\mathcal{V}(t)}{T-t} (\Sigma^2(T, K) - \sigma^2(t)) dt \end{aligned}$$

Si la position optionnelle est un call, le PnL est aussi sensible à la position du sous-jacent. L'idée de Demeterfi *et al.* (1999) est de construire un portefeuille d'options d'achat qui est beaucoup moins sensible à la variation du sous-jacent, c'est-à-dire qui présente une sensibilité $\mathcal{V}(t)$ constante²². Soient $\mathcal{V}_j(t)$ et $\mathcal{V}(t)$ les sensibilités de la j -ième option et du portefeuille. En notant $w_j(t)$ le poids de la j -ième option dans le portefeuille, nous obtenons :

$$\mathcal{V}(t) = \sum_{j=1}^m w_j(t) \mathcal{V}_j(t)$$

Sur le graphique 40, nous représentons la sensibilité $\mathcal{V}(t)$ du portefeuille d'options pour différentes valeurs de m lorsque les poids sont proportionnels à $1/K^2$. Nous remarquons que lorsque $m \rightarrow \infty$, nous avons :

$$\mathcal{V}(t) = \int_0^\infty \frac{1}{K^2} \frac{\partial^2 C(K)}{\partial \sigma^2} = c$$

Une exposition à la variance revient donc à détenir un portefeuille d'options avec tous les prix d'exercice pondérées en $1/K^2$. Ce principe est la base de la valorisation et de la réplique statique du swap de variance.

Nous considérons un swap de variance de maturité T dont le payoff $N \times (V - K_{\text{var}})$ dépend de la différence entre la variance réalisée V et le prix d'exercice K_{var} . Nous supposons que la dynamique du sous-jacent est :

$$dS(t) = \mu(t) dt + \sigma(t) dW(t)$$

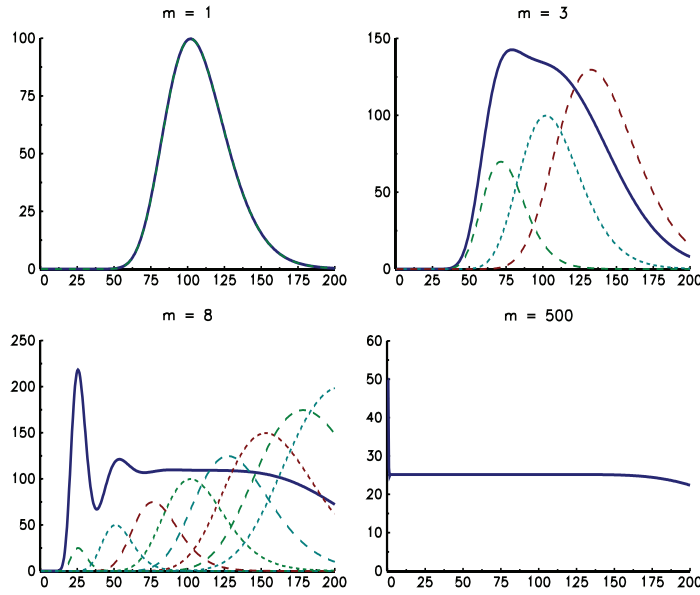
La variance réalisée est définie par :

$$V = \frac{1}{T} \int_0^T \sigma^2(t) dt$$

Le swap de variance a un prix nul si :

$$K_{\text{var}} = \mathbb{E}[V] = \frac{1}{T} \mathbb{E} \left[\int_0^T \sigma^2(t) dt \right]$$

²²K. Demeterfi, E. Derman, M. Kamal et J. Zou, « More Than You Ever Wanted To Know About Volatility Swaps », *Quantitative Strategies Research Notes*, Goldman Sachs, 1999.

Graphique 40. Sensibilité \mathcal{V} du portefeuille d'options et des calls

Dans le cas d'un mouvement brownien géométrique, nous avons aussi :

$$K_{\text{var}} = \frac{2}{T} \mathbb{E} \left[\int_0^T \frac{dS(t)}{S(t)} - \ln \frac{S(T)}{S(0)} \right]$$

Demeterfi *et al.* (1999) montrent alors que²³ :

$$K_{\text{var}} = \frac{2}{T} \left(e^{rT} \int_0^{F_0} \frac{1}{K^2} P(K) dK + \int_{F_0}^{\infty} \frac{1}{K^2} C(K) dK \right) + \frac{2}{T} \left(rT - \ln \frac{F_0}{S_0} - \left(\frac{S_0}{F_0} e^{rT} - 1 \right) \right) \quad (6.5)$$

où F_0 est le prix *forward* du sous-jacent, $C(K)$ et $P(K)$ sont les prix des calls et puts de prix d'exercice K et de maturité T .

Remarque 59. L'équation (6.5) est très importante car elle indique qu'il est possible de répliquer un swap de variance avec un ensemble de calls et puts. Comme nous l'avons vu précédemment, cette réplique statique

²³Le lecteur peut aussi consulter la référence suivante qui propose une démonstration beaucoup plus claire de ce résultat.

P. Carr et R. Lee, « Volatility Derivatives », *Annual Review of Financial Economics*, 1, 2009, p. 1-21.

est très particulière puisque les poids sont inversement proportionnels au carré des strikes des options – $w(K) \propto K^{-2}$.

Le résultat précédent permet à Demeterfi *et al.* (1999) de dériver des formules approchées pour le prix d'exercice du swap de variance. En particulier, si le smile est linéaire, c'est-à-dire que $\Sigma(K) = \Sigma(F) - s(K - F)/F$, ces auteurs montrent que²⁴ :

$$K_{\text{var}} = \Sigma^2(F) (1 + 3s^2T + \dots)$$

Pour valoriser le swap de variance en *mark-to-market*, nous devons tout d'abord remarquer que la variance telle que définie dans ce swap est additive. Cela veut dire que si $t_1 < t_2 < t_3$, nous avons²⁵ :

$$v_{t_1, t_3} = \frac{n_{1,2}}{n_{1,3}} v_{t_1, t_2} + \frac{n_{2,3}}{n_{1,3}} v_{t_2, t_3}$$

Considérons un swap de variance à la date t_1 pour la période $[t_1, t_3]$ et dont le prix d'exercice est K_{t_1, t_3} . À la date t_2 , nous en déduisons que :

$$\begin{aligned} \text{PnL} &= N_{\text{var}} \times (v_{t_1, t_3} - K_{t_1, t_3}^2) \\ &= N_{\text{var}} \times \left(\frac{n_{1,2}}{n_{1,3}} v_{t_1, t_2} + \frac{n_{2,3}}{n_{1,3}} v_{t_2, t_3} - \frac{n_{1,2} + n_{2,3}}{n_{1,3}} K_{t_1, t_3}^2 \right) \\ &= N_{\text{var}} \times \frac{n_{1,2}}{n_{1,3}} (v_{t_1, t_2} - K_{t_1, t_3}^2) + N_{\text{var}} \times \frac{n_{2,3}}{n_{1,3}} (v_{t_2, t_3} - K_{t_1, t_3}^2) \end{aligned}$$

Le premier terme du PnL est connu à la date t_2 et correspond à la volatilité réalisée sur la période $[t_1, t_2]$. Le second terme correspond à un swap de variance pour la période $[t_2, t_3]$. Notons K_{t_2, t_3} son prix d'exercice, nous avons :

$$\begin{aligned} \text{PnL} &= N_{\text{var}} \times \frac{n_{1,2}}{n_{1,3}} (v_{t_1, t_2} - K_{t_1, t_3}^2) + \\ &N_{\text{var}} \times \frac{n_{2,3}}{n_{1,3}} (v_{t_2, t_3} - K_{t_2, t_3}^2 + K_{t_2, t_3}^2 - K_{t_1, t_3}^2) \end{aligned}$$

²⁴Cette formule d'approximation est très utilisée par les quants pour produire des backtests.

²⁵En effet, nous avons :

$$\begin{aligned} &\frac{n_{1,2}}{n_{1,3}} \left(\frac{1}{n_{1,2}} \sum_{t=t_1}^{t_2} \ln^2 \frac{S_t}{S_{t-1}} \right) + \frac{n_{2,3}}{n_{1,3}} \left(\frac{1}{n_{2,3}} \sum_{t=t_2}^{t_3} \ln^2 \frac{S_t}{S_{t-1}} \right) \\ &= \frac{1}{n_{1,3}} \sum_{t=t_1}^{t_3} \ln^2 \frac{S_t}{S_{t-1}} \end{aligned}$$

(resp. sur le taux actuariel). Les résultats sont présentés sur le tableau 7 dans le cas de l'obligation de maturité 5 ans. Nous vérifions que le calcul basé sur la sensibilité donne de bons résultats.

Tableau 6. Prix, taux de rendement actuariel et sensibilité d'une obligation de maturité T et de coupon 5%

T	$R(0, T)$	$B(0, T)$	$P(0)$	R^*	S
1	0,52	99,48	104,45	0,52	-103,91
2	0,99	98,04	107,92	0,98	-208,83
3	1,42	95,86	110,53	1,39	-312,51
4	1,80	93,10	112,43	1,76	-413,32
5	2,15	89,92	113,74	2,08	-510,09

Tableau 7. Impact d'une variation des taux sur le prix de l'obligation 5 ans

ΔR^* (en pbs)	$\check{P}(0)$	ΔP	$\hat{P}(0)$	ΔP	$S \times \Delta R^*$
-50	116,32	2,58	116,33	2,59	2,55
-30	115,28	1,54	115,28	1,54	1,53
-10	114,25	0,51	114,25	0,51	0,51
0	113,74	0,00	113,74	0,00	0,00
10	113,23	-0,51	113,23	-0,51	-0,51
30	112,23	-1,51	112,22	-1,52	-1,53
50	111,23	-2,51	111,22	-2,52	-2,55

4.2.2. La stratégie de roll-down

Considérons un investisseur qui a un horizon de placement d'un an. La façon la plus naturelle d'investir est d'acheter une obligation de maturité un an. Le prix actuel de l'obligation est $P = C(1) \times B(0, 1) + N \times B(0, 1)$. Au bout d'un an, il reçoit le coupon $C(1)$ et la valeur de l'obligation est N . On en déduit que le taux de rendement de cet investissement est égal au taux zéro-coupon de maturité 1 an :

$$\begin{aligned}
 R &= \frac{N + C(1)}{C(1)B(0, 1) + NB(0, 1)} - 1 \\
 &= \frac{1}{B(0, 1)} - 1 \\
 &= R(0, 1)
 \end{aligned}$$

L'investisseur peut aussi acheter une obligation de maturité T et la revendre un an plus tard. Dans ce cas, le rendement de l'investissement est :

$$R = \frac{P(1) + C(1)}{P(0)} - 1$$

Le prix $P(1)$ n'est pas connu aujourd'hui puisque c'est le prix de revente. Si on suppose néanmoins que la courbe des taux reste inchangée, nous pouvons calculer R . Considérons la courbe des taux précédente. Soit une obligation de maturité 5 ans dont le coupon annuel est de 5%. Le prix $P(0)$ de l'obligation est égale aujourd'hui à 113,74. Si on calcule la valeur de cette obligation dans un an avec la courbe des taux inchangée, nous obtenons $P(1) = 112,43$. Nous en déduisons que :

$$R = \frac{112,43 + 5}{113,74} - 1 = 3,24\%$$

Ce rendement est bien supérieur au taux zéro coupon $R(0,1)$ qui est égal à 0,52%. Cette stratégie qui consiste à considérer que la courbe des taux ne va pas bouger et à investir sur une maturité plus grande que l'horizon d'investissement est appelée une stratégie de *roll-down*⁴³. Nous représentons sur le graphique 51 l'excès de rendement $R - R(0,1)$ de cette stratégie. Néanmoins, il existe une grande différence entre ces deux rendements puisque $R(0,1)$ est certain alors que R n'est pas du tout certain. Il existe donc un risque en cas de montée des taux. Reprenons l'exemple en supposant que les taux augmentent de 2% dans un an. Dans ce cas, la valeur de l'obligation devient $P(1) = 104,55$ et nous obtenons :

$$R = \frac{104,55 + 5}{113,74} - 1 = -3,68\%$$

Remarque 62. L'excès de rendement $\bar{R} = R - R(0,1)$ d'une stratégie *roll-down* comprend deux composantes : le *carry* désigne le rendement provenant du portage de coupon tandis que le *roll-down* désigne le rendement provenant de la déformation (ou non) de la courbe des taux. Nous avons :

$$\bar{R} = \underbrace{\frac{C(1)}{P(0)} - R(0,1)}_{\text{carry}} + \underbrace{\frac{P(1)}{P(0)} - 1}_{\text{roll-down}}$$

Ce *carry* est lui-même décomposable en une partie coupon et une partie financement. Sur le tableau 8, nous reportons l'excès de rendement \bar{R} ainsi que la décomposition entre *carry* et *roll-down* pour les obligations de maturité 1 an à 5 ans. Dans le cas d'un coupon de 5%, le *roll-down* est négatif car le prix de l'obligation est bien au-dessus du pair. Si on considère un zéro-coupon, on a un *carry* négatif car il n'y a pas de coupon et un *roll-down* positif. Dans le cas d'une obligation avec un coupon de 2%, le *carry* et le *roll-down* sont tous deux positifs pour une maturité supérieure à deux ans.

On appelle *breakeven* \mathfrak{R} la variation de taux au-delà de laquelle la stratégie n'est plus rentable. Celui-ci est égal à 76 pbs dans notre exemple.

⁴³On la trouve aussi dans la littérature sous le nom de "Riding the Yield Curve".

Tableau 8. Décomposition entre carry et roll-down

T	P(0)	Carry			Roll-down	\bar{R}
		Coupon	Financement	Total		
Coupon = 5%						
1	104,45	4,79	-0,52	4,26	-4,26	0,00
2	107,92	4,63	-0,52	4,11	-3,21	0,90
3	110,53	4,52	-0,52	4,00	-2,36	1,64
4	112,43	4,45	-0,52	3,92	-1,69	2,24
5	113,74	4,40	-0,52	3,87	-1,15	2,72
Zéro Coupon						
1	99,48	0,00	-0,52	-0,52	0,52	0,00
2	98,04	0,00	-0,52	-0,52	1,47	0,94
3	95,86	0,00	-0,52	-0,52	2,27	1,75
4	93,10	0,00	-0,52	-0,52	2,96	2,44
5	89,92	0,00	-0,52	-0,52	3,54	3,02
Coupon = 2%						
1	101,47	1,97	-0,52	1,45	-1,45	0,00
2	101,99	1,96	-0,52	1,44	-0,51	0,93
3	101,73	1,97	-0,52	1,44	0,26	1,70
4	100,83	1,98	-0,52	1,46	0,89	2,35
5	99,45	2,01	-0,52	1,49	1,39	2,88

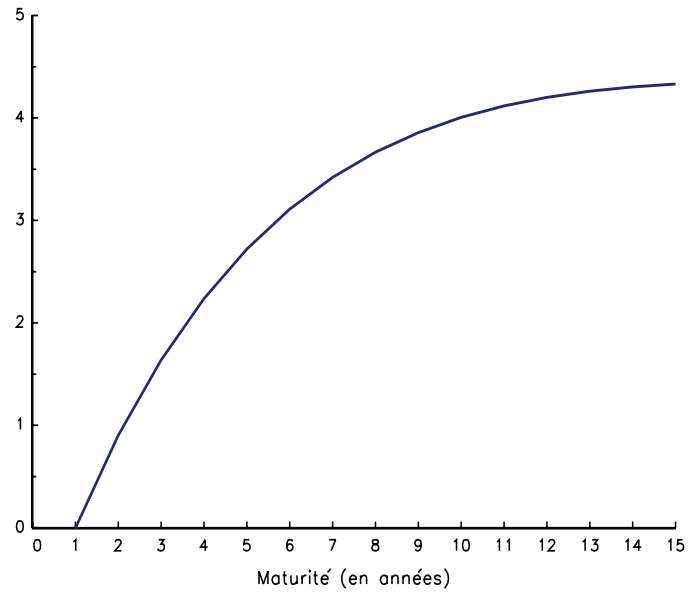
En considérant ce mouvement de taux, nous vérifions que le rendement de la stratégie est égal au taux zéro coupon $R(0, 1)$:

$$R = \frac{109,33 + 5}{113,74} - 1 = 0,52\%$$

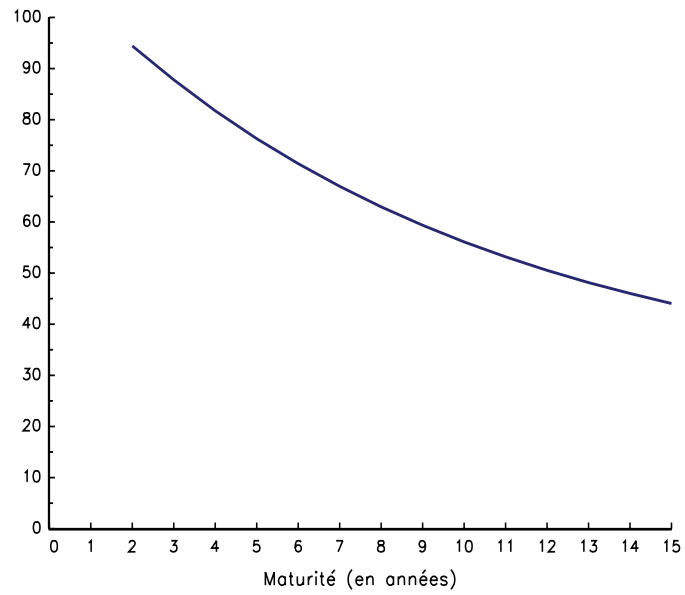
À titre d'illustration, nous représentons les valeurs de *breakeven* de la stratégie de *roll-down* sur le graphique 52.

Remarque 63. *Le breakeven est une meilleure mesure du carry et du roll-down. C'est en effet une mesure homogène au risque de taux. Il correspond en première approximation au carry et au roll-down en euros ajustés par la sensibilité de l'obligation aux mouvements des taux. Or le carry et le roll-down s'interprètent souvent en terme de protection qu'ils offrent contre le risque d'une hausse de taux. Dans notre exemple, nous remarquons que le carry et le roll-down en absolu sont bien plus importants sur les maturités longues. Cette perception n'est pas la bonne, car si on pondère par la sensibilité au risque de taux, le carry et le roll-down sont meilleurs sur les maturités courtes. Dans notre exemple, cette protection est de 95 pbs par an pour l'obligation 2 ans et de 76 pbs par an pour l'obligation 5 ans. Autrement dit, pour la même exposition au risque d'un mouvement parallèle de la courbe de taux (c'est-à-dire la même durée), une obligation 2 ans offre un meilleur carry et roll-down qu'une obligation 5 ans. Nous devons aussi noter qu'une stratégie de roll-down est justifiée*

Graphique 51. Excès de rendement $\bar{R} = R - R(0,1)$ (en %) de la stratégie de roll-down



Graphique 52. Breakeven \mathfrak{R} (en pbs) de la stratégie de roll-down



dans un environnement de faible volatilité. Idéalement, il faudrait que rien ne bouge, c'est-à-dire que la volatilité soit nulle. C'est pourquoi il est important de comparer le breakeven de la stratégie à la volatilité associée. Généralement, les taux courts présentent une volatilité plus importante que les taux longs. Dans notre exemple, il faudrait que la volatilité du taux 2 ans soit 25% plus forte que la volatilité du taux 5 ans pour justifier le différentiel de carry et de roll-down observé⁴⁴.

Remarque 64. Au premier ordre, nous pouvons calculer le breakeven en comparant les rendements forward aux rendements spot futurs. Nous avons :

$$\mathfrak{R} \simeq R_{t,1Y}^* - R_{t+1,0}^*$$

où $R_{t,1Y}^*$ est le yield-to-maturity 1 an forward et $R_{t+1,0}^*$ est le yield-to-maturity spot dans un an en supposant la courbe des taux inchangée. Dans notre exemple, le taux zéro 4 ans dans 1 an est égal à 1,80% et le taux forward 1 an de maturité résiduelle 4 ans est égal à 2,55%. Nous avons $\mathfrak{R} \simeq 75$ pbs par an. Si on réajuste par la sensibilité et la convexité, nous trouvons un carry et un roll-down sur le zéro-coupon 5 ans égal à 3,05% (à comparer avec le chiffre de 3,02% calculé avec la méthode exacte).

4.2.3. Les stratégies de barbell

Dans le cas de la stratégie précédente portant sur une seule maturité, on parle de portefeuille *bullet*. Plus généralement, on cherche à jouer un scénario de mouvement de courbe de taux. Pour cela, on peut utiliser un portefeuille *barbell* construit à partir de plusieurs maturités. L'exemple typique est un portefeuille portant sur les maturités 2 ans, 5 ans et 10 ans. On impose généralement que la sensibilité du portefeuille soit nulle afin de tirer parti des mouvements de pentification et de courbure de la courbe des taux. Notons n_1 , n_2 et n_3 le nombre d'obligations de maturité T_1 , T_2 et T_3 avec $T_1 < T_2 < T_3$. La sensibilité du portefeuille S est nulle si la relation suivante est vérifiée :

$$n_1 S_1 + n_2 S_2 + n_3 S_3 = 0$$

Par convention, nous fixons $n_2 = -1$. Pour déterminer le portefeuille, nous avons donc deux variables n_1 et n_3 à estimer. Pour cela, nous devons imposer une deuxième relation. Martinelli *et al.* (2003) proposent 4 méthodes pour construire un portefeuille barbell⁴⁵.

1. Dans un portefeuille 50/50, la sensibilité de chaque jambe longue est la même $-n_1 S_1 = n_3 S_3$. Nous avons donc :

$$\begin{cases} n_1 = -n_2 \frac{S_2}{S_1} \\ n_3 = -n_2 \frac{S_2}{S_3} \end{cases}$$

⁴⁴Le chiffre de 25% correspond à $95/76 - 1$.

⁴⁵Notons néanmoins que plus de 90% des stratégies implémentées sont des portefeuilles barbell 50/50.

- Dans le cas de plusieurs actifs, nous pouvons fixer une allocation cible w^* et *tilter* cette exposition :

$$w(t) = f_e(s(t), w^*)$$

Par exemple, nous pouvons spécifier :

$$w_i(t) = \frac{(1 + s_i(t)) \times w_i^*}{\sum_{i=1}^n (1 + s_i(t)) \times w_i^*}$$

Ce type de fonction d'exposition est utilisé dans le cas d'une gestion *benchmarkée*.

- Nous pouvons aussi construire une exposition sans référence à un *benchmark*. La fonction d'exposition pourra être de la forme :

$$w_i(t) = \eta \times f_e(s_i(t))$$

avec $f_e'(x) \geq 0$.

- Si $f_e(x) \geq 0$ et $\eta = (\sum_{i=1}^n f_e(s_i(t)))^{-1}$ (les poids sont positifs et la somme des poids est égale à 100%), nous obtenons une stratégie *total return*.
- Si $f_e(x) \in \mathbb{R}$, c'est-à-dire que nous pouvons avoir des positions *short*, la stratégie est dite *absolute return*.

On peut aussi utiliser les modèles d'allocation présentés dans le chapitre 2. Le modèle dit *Trend following Markowitz* est notamment très populaire. Il correspond au modèle de Markowitz avec un objectif de volatilité où les primes de risque sont remplacées par les scores.

5.1.2. Un exemple

Nous reprenons l'univers d'investissement de la page 93 en ne considérant que quatre classes d'actifs : les actions américaines, les actions européennes, les obligations américaines et les obligations européennes. Nous simulons une stratégie *trend following* en rebalançant le portefeuille tous les mois avec des poids obtenus à partir du programme d'optimisation suivant :

$$\begin{aligned} w_t^* &= \arg \max s_t^\top w_t \\ \text{s.c.} & \begin{cases} \sigma(w_t) \leq \sigma^* \\ w_t \in \Omega \end{cases} \end{aligned}$$

À la date t , le vecteur s_t des scores correspond aux rendements observés sur la dernière année des quatre classes d'actifs. Nous utilisons la matrice de covariance empirique des rendements journaliers observés sur la dernière année pour mesurer le risque $\sigma(w_t)$ du portefeuille. C'est donc un programme d'allocation de type Markowitz avec une contrainte de volatilité σ^* et des contraintes supplémentaires $w_t \in \Omega$.

1. Dans le cas d'une stratégie *benchmarkée*, nous définissons le benchmark w^* comme étant le portefeuille équilibré. La volatilité cible

σ^* à la date t est égale à la volatilité historique du benchmark calculée sur la dernière année et Ω correspond à :

$$\begin{cases} 0,7 \cdot w^* \leq w_t \leq 1,3 \cdot w^* \\ \mathbf{1}^\top w_t = 1 \end{cases}$$

Dans cette stratégie (notée #1), la déviation par rapport au benchmark est contrôlée par des contraintes sur les poids. Nous pouvons aussi construire une autre stratégie (notée #2) où la contrainte porte sur la volatilité de l'erreur de réplication :

$$\begin{cases} \sqrt{(w - w^*)^\top \Sigma (w - w^*)} \leq 2\% \\ \mathbf{1}^\top w_t = 1 \end{cases}$$

La simulation de ces deux stratégies et du benchmark est représentée sur le graphique 56.

2. Pour la stratégie *total return*, nous avons :

$$\begin{cases} \sigma^* = 6\% \\ \mathbf{0} \leq w_t \leq \mathbf{1} \\ \mathbf{1}^\top w_t = 1 \end{cases}$$

Contrairement à la stratégie précédente, il n'y a pas de contrainte de déviation des poids par rapport à un benchmark. Nous remarquons d'ailleurs que la simulation du graphique 57 présente un ratio de Sharpe meilleur que la stratégie précédente.

3. Pour la stratégie *absolute return*, nous imposons :

$$\begin{cases} \sigma^* = 4\% \\ -0,5 \cdot \mathbf{1} \leq w_t \leq 0,5 \cdot \mathbf{1} \\ \mathbf{1}^\top w_t = 0 \end{cases}$$

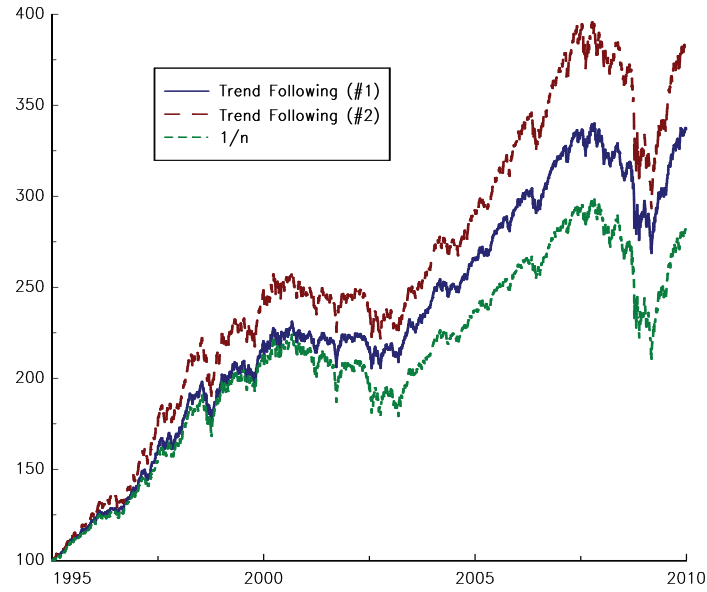
Les poids peuvent être maintenant négatifs, ce qui veut dire que certaines positions sont *short*. Nous imposons aussi que la somme des poids est nulle. La stratégie n'a donc pas de biais structurellement *long* ou *short*. La simulation de la stratégie est présentée sur le graphique 58.

Les résultats de ces simulations sont proches de ceux obtenus par certains travaux académiques⁷¹.

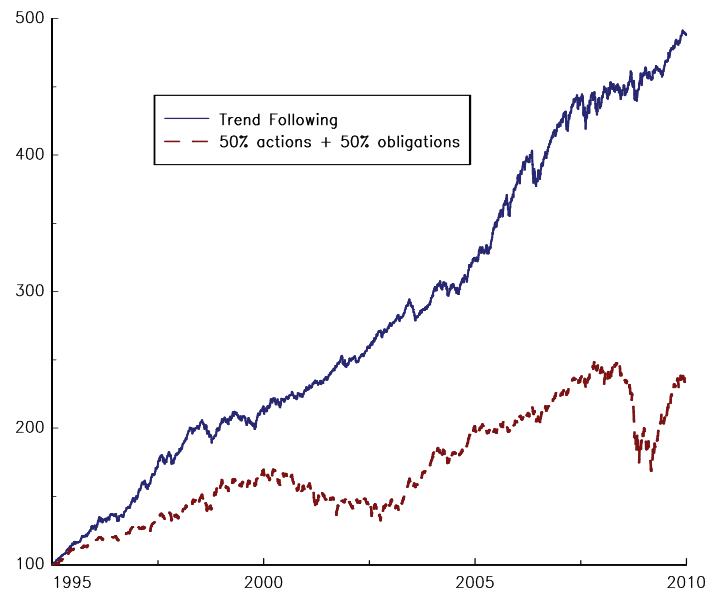
⁷¹M.T. Faber, « A Quantitative Approach to Tactical Asset Allocation », *Journal of Wealth Management*, 9(4), 2007, p. 69-79.

O. ap Gwilym, A. Clare, J. Seaton et S. Thomas, « Price and Momentum as Robust Tactical Approaches to Global Equity Investing », *Journal of Investing*, 19(3), 2010, p. 80-91.

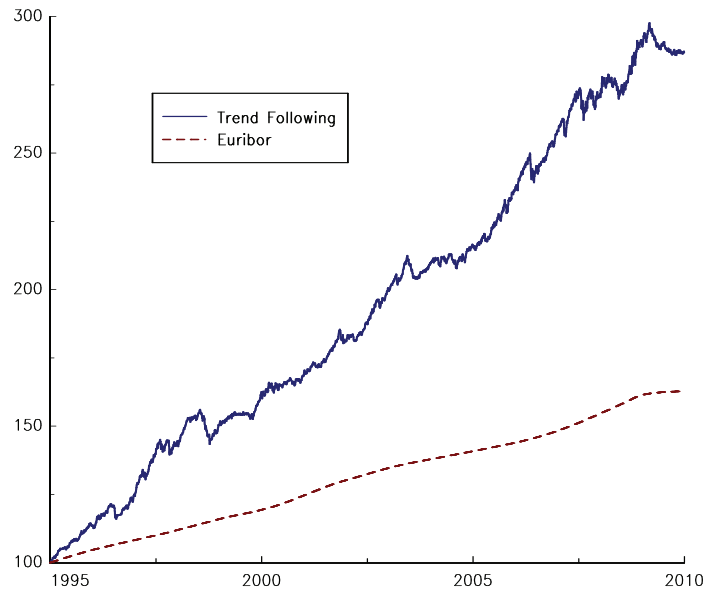
Graphique 56. Backtest de la stratégie trend following benchmarkée



Graphique 57. Backtest de la stratégie trend following total return



Graphique 58. Backtest de la stratégie trend following absolute return



5.1.3. Profil optionnel d'une stratégie trend following

Les stratégies *trend following* sont intensivement utilisés par les CTA. Comme nous l'avons vu précédemment, on peut caractériser les stratégies dynamiques de trading par un profil optionnel équivalent. Ainsi, Fund et Hsieh (2001) analysent les stratégies CTA comme des profils optionnels de type *lookback straddle*⁷².

5.2. La stratégie mean reverting

La stratégie de retour à la moyenne est très utilisée dans le cas de l'arbitrage de volatilité ou du *pair trading*. Deux types de modélisation sont couramment employés. Le premier repose sur le processus de Ornstein-Uhlenbeck alors que le second utilise la cointégration.

5.2.1. Le modèle de Ornstein-Uhlenbeck

Notons S_t la variable d'intérêt. Cela peut être par exemple une volatilité, un spread de volatilité, une différence de deux prix d'actions ou un spread de taux d'intérêt. Nous considérons un processus de Ornstein-

⁷²W. Fung et D.A. Hsieh, « The Risk in Hedge Fund Strategies: Theory and Evidence from Trend Followers », *Review of Financial Studies*, 14(2), 2001, p. 313-341.

Uhlenbeck pour modéliser la dynamique de S_t . Nous avons :

$$dS_t = a(b - S_t) dt + \sigma dW_t$$

avec $a \geq 0$. La solution de ce processus est donnée à la page 238. Nous rappelons que $S_t | S_0$ est une variable aléatoire gaussienne avec :

$$\mathbb{E}[S_t | S_0] = S_0 e^{-a(t-t_0)} + b \left(1 - e^{-a(t-t_0)}\right)$$

et :

$$\sigma^2 [S_t | S_0] = \sigma^2 \left(\frac{1 - e^{-2a(t-t_0)}}{2a} \right)$$

Trois paramètres interviennent dans la définition du processus. Nous avons :

$$\lim_{t \rightarrow \infty} \mathbb{E}[S_t | S_0] = b$$

b est donc la moyenne de long terme de S_t . Nous remarquons que :

$$\mathbb{E}[S_t | S_0] - b = (S_0 - b) e^{-a(t-t_0)}$$

Plus le paramètre a est élevé, plus la déviation par rapport à b sera en moyenne petite au cours du temps. a est donc une mesure de la vitesse pour retourner à la moyenne b . Enfin, σ est un paramètre de volatilité. Nous montrons l'influence de ces paramètres sur le graphique 59.

Si nous discrétisons le processus de Ornstein-Uhlenbeck, nous obtenons :

$$S_t - S_{t-1} = a(b - S_{t-1}) + \eta_t$$

avec $\eta_t \sim \mathcal{N}(0, \sigma^2)$. Nous pouvons construire un modèle de retour à la moyenne plus général :

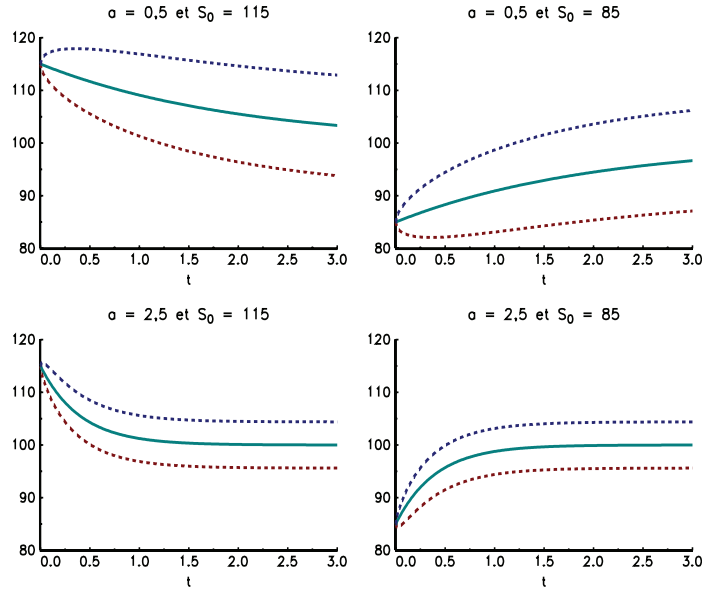
$$\begin{cases} S_t - S_{t-1} = \sum_{i=1}^p \alpha_i (b_{t-i} - S_{t-i}) + \eta_t \\ b_t = b_{t-1} + \varepsilon_t \end{cases}$$

avec $\eta_t \sim \mathcal{N}(0, \sigma^2)$ et $\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2)$. C'est un modèle AR(p) qui intègre une moyenne stochastique. Pour estimer ce modèle, nous utilisons le filtre de Kalman en utilisant la représentation espace-état suivante :

$$\begin{cases} S_t = Z \alpha_t \\ \alpha_t = T \alpha_{t-1} + R \begin{pmatrix} \eta_t \\ \varepsilon_t \end{pmatrix} \end{cases}$$

avec $\alpha_t = (S_t, \dots, S_{t-p+1}, b_t, \dots, b_{t-p+1})^\top$, $R = [\mathbf{e}_1 \quad \mathbf{e}_p]$ et :

$$T = \begin{pmatrix} 1 - \alpha_1 & -\alpha_2 & \cdots & -\alpha_p & \alpha_1 & \cdots & \cdots & \alpha_p \\ & & & 0 & 0 & \cdots & \cdots & 0 \\ & I_{p-1} & & \vdots & \vdots & & & \vdots \\ & & & 0 & 0 & \cdots & \cdots & 0 \\ & & & & 1 & 0 & \cdots & 0 \\ & 0_p & & & & & & \vdots \\ & & & & & I_{p-1} & & \vdots \\ & & & & & & & 0 \end{pmatrix}$$

Graphique 59. Illustration de l'effet de retour à la moyenne ($b = 100$)

Remarque 65. Si $p = 1$, ce modèle correspond à la version discrétisée du processus bi-dimensionnel suivant :

$$\begin{cases} dS_t = a(b_t - S_t) dt + \sigma dW_t \\ db_t = dZ_t \end{cases}$$

avec Z_t un mouvement brownien indépendant de W_t . Nous pouvons considérer d'autres processus pour modéliser la moyenne stochastique b_t . Par exemple, nous pouvons supposer que b_t est aussi un processus de Ornstein-Uhlenbeck :

$$\begin{cases} dS_t = a(b_t - S_t) dt + \sigma dW_t \\ db_t = \kappa(\theta - b_t) dt + \xi dZ_t \end{cases}$$

La discrétisation de ce processus est immédiate et l'estimation de b_t se fait par le filtre de Kalman.

5.2.2. Les modèles à correction d'erreurs

Nous supposons que S_t est la différence⁷³ entre deux processus X_t et Y_t :

$$S_t = X_t - Y_t$$

Si X_t et Y_t sont $I(1)$ et que S_t est $I(0)$, alors nous savons que (X_t, Y_t) admet une représentation sous forme de correction d'erreurs (VECM). La

⁷³ou une combinaison linéaire de X_t et Y_t .

des actifs financiers, nous choisissons les actifs financiers i qui vérifient la condition $s_i \geq (1 - \alpha)n$. Dans le cas d'un portefeuille long/short, les actifs financiers de la position acheteuse \mathcal{L} vérifient $s_i \geq (1 - \alpha)n$ alors que ceux de la position vendeuse \mathcal{S} vérifient $s_i \leq \alpha n$. Notons que ces sélections s'appellent des "top k lists" et font l'objet d'études statistiques² depuis le début des années 2000 et le développement des moteurs de recherche internet.

Les exemples de sélection d'actifs financiers basés sur la méthode des rangs ne manquent pas. Par exemple, de nombreux gérants actions utilisent cette approche pour sélectionner les titres de leur portefeuille à partir de facteurs comptables comme les ratios *Price-to-Book*³ (P/B) ou *Price-Earnings*⁴ (P/E).

1.1.2. La méthode des probabilités

Nous associons maintenant à chaque actif financier i une variable aléatoire X_i . Nous supposons que cette variable aléatoire est positivement corrélée à notre signal d'investissement. Nous pouvons alors construire le score s_i de la façon suivante :

$$s_i = \Pr \{X_i \geq x_i\}$$

Si l'actif financier i a une probabilité importante que X_i prenne de fortes valeurs, alors le score est élevé indiquant un signal d'investissement fort. Dans le cas d'une relation négative avec le signal d'investissement, nous avons :

$$s_i = \Pr \{X_i \leq x_i\}$$

Par définition, S est une distribution uniforme et prend ses valeurs dans l'intervalle $[0, 1]$.

Considérons un processus de Ornstein-Uhlenbeck $Y_i(t)$. Pour tout actif financier i , nous associons les paramètres a_i , b_i et σ_i de $Y_i(t)$. À la date t , la position de $Y_i(t)$ est connue. Nous cherchons à connaître la position de $Y_i(t)$ à la date $t + \theta$. La variable aléatoire X_i est donc $Y_i(t + \theta)$. Si la

²R. Fahin, R. Kumar et D. Sivakumar, « Comparing Top K Lists », *SIAM Journal on Discrete Mathematics*, 17(1), 2003, p. 134-160.

³Le ratio *Price-to-Book* est le rapport entre le cours de marché du titre et la book value. Celle-ci est la valeur comptable nette de l'entreprise. Ce ratio mesure donc le rapport entre la valeur de marché des capitaux propres et leur valeur comptable. Généralement, on préfère les actions qui ont un faible ratio P/B.

⁴Le ratio *Price-Earnings* est le rapport entre le cours de marché du titre et le bénéfice par action. Ce ratio est utilisé pour mesurer la cherté présente d'une action.

relation avec le signal d'investissement est négative, nous avons :

$$\begin{aligned}
 s_i &= \Pr \{X_i \leq x_i\} \\
 &= \Pr \{Y_i(t+\theta) \leq x_i\} \\
 &= \Phi \left(\sqrt{\frac{2a_i}{1-e^{-2a_i\theta}}} \left(\frac{x_i - (Y_i(t) e^{-a_i\theta} + b_i(1-e^{-a_i\theta}))}{\sigma_i} \right) \right)
 \end{aligned} \tag{7.1}$$

Voici le type de score que nous pouvons utiliser pour une stratégie *mean reverting*. Supposons que $Y_i(t)$ représente le spread de volatilité historique de maturité θ entre deux actions i_1 et i_2 :

$$Y_i(t) = \sigma_{t-\theta,t}^{(i_1)} - \sigma_{t-\theta,t}^{(i_2)}$$

À la date t , les volatilités implicites de maturité θ sont cotées respectivement $K_{t,t+\theta}^{(i_1)}$ et $K_{t,t+\theta}^{(i_2)}$. Le score de la paire i est donc :

$$s_i = \Pr \left\{ \sigma_{t,t+\theta}^{(i_1)} - \sigma_{t,t+\theta}^{(i_2)} \leq K_{t,t+\theta}^{(i_1)} - K_{t,t+\theta}^{(i_2)} \right\}$$

Ce score mesure la probabilité que le spread de volatilité réalisée entre t et $t+\theta$ soit inférieur au spread coté par le marché. Un score élevé indique qu'il faut vendre la volatilité du premier actif et acheter celle du second actif. Pour calculer ce score, nous devons tout d'abord estimer les paramètres a_i , b_i et σ_i par la méthode du maximum de vraisemblance en considérant l'historique $\{Y_i(s), s \leq t\}$. Ensuite, il suffit de calculer l'expression (7.1) pour chaque paire de trading. Par exemple, pour $a_i = 0,7$, $b_i = -3,0$ et $\sigma_i = 3,8$, la valeur du score est 97,6% si le spread actuel et le spread implicite sont égaux à -1 et $3,5$ (voir le graphique 1). Dans ce scoring appliqué au spread de volatilité, notons que le gérant est intéressé par les paires présentant des scores très élevés ou très faibles.

1.1.3. La méthode des z-scores

Un z-score est une normalisation de la variable aléatoire X . Nous avons :

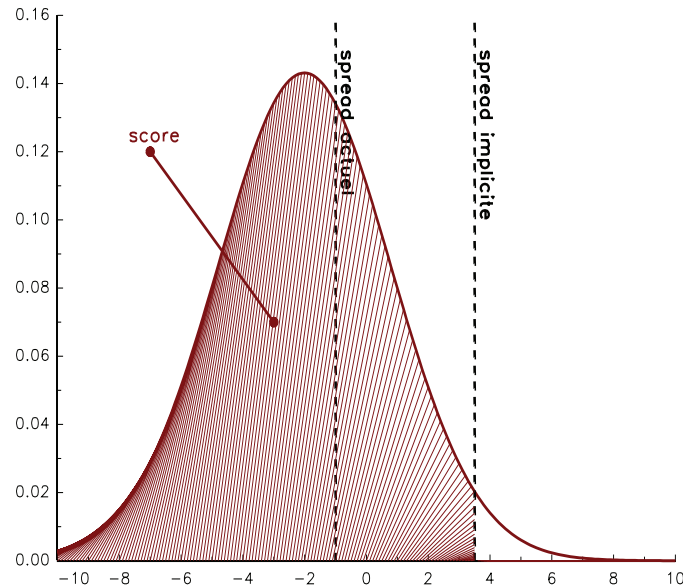
$$z_i = \frac{x_i - m}{\sigma}$$

avec :

$$\begin{aligned}
 m &= \frac{1}{n} \sum_{i=1}^n x_i \\
 \sigma &= \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - m_i)^2}
 \end{aligned}$$

Les paramètres de normalisation m et σ sont donc la moyenne et l'écart-type des valeurs prises par l'ensemble Ω des actifs financiers. Dans certains

Graphique 1. Illustration du score de probabilité



cas, la normalisation ne se fait pas entre les scores des différents actifs financiers à la date t , mais porte sur l'historique $[t - p, t]$ de la variable $x_i(t)$. Dans ce cas, nous avons :

$$z_i(t; p) = \frac{x_i(t) - m_i(t; p)}{\sigma_i(t; p)}$$

avec :

$$m_i(t; p) = \frac{1}{p+1} \sum_{\theta=0}^p x_i(t-\theta)$$

$$\sigma_i(t; p) = \sqrt{\frac{1}{p+1} \sum_{\theta=0}^p (x_i(t-\theta) - m_i(t; p))^2}$$

Les paramètres de normalisation $m_i(t; p)$ et $\sigma_i(t; p)$ sont donc la moyenne et l'écart-type de la variable sur la période glissante $[t - p, t]$. Dans ce dernier cas, on parle de z-score temporel alors qu'on réserve le nom de z-score spatial pour le premier cas. Supposons que $x_i(t)$ corresponde au P/E du titre i , le z-score $z_i(t; p)$ normalise le P/E du titre i par rapport aux p dernières observations du P/E du titre i . En revanche, le z-score z_i normalise le P/E du titre i par rapport au P/E de l'ensemble des titres.

La distribution d'un z-score est complexe puisqu'elle dépend de la distribution de X . Néanmoins, sous certaines conditions, on considère qu'elle

doit être proche d'une distribution normale centrée et réduite $\mathcal{N}(0, 1)$. D'ailleurs, dans certaines applications⁵, les z-scores sont bornés à gauche par -3 et à droite par $+3$. Le choix de l'intervalle $[-3, +3]$ correspond au fait que la probabilité $\Pr\{|\mathcal{N}(0, 1)| \geq 3\}$ est très faible⁶.

1.1.4. La normalisation d'un score

Souvent, il est plus pratique de normaliser un score afin de changer de support. Prenons l'exemple d'un score S qui prend ses valeurs dans $[0, 1]$. Dans le cas d'un portefeuille long/short, le score $S^* = 2 \times S - 1$ peut être plus pertinent. En effet, il prend ses valeurs dans $[-1, 1]$ et on distingue plus facilement les candidats *short* (ceux qui ont un score négatif) des candidats *long* (ceux qui ont un score positif).

Plus généralement, la normalisation d'un score consiste à appliquer une transformation de la forme $S^* = g(S)$. On peut spécifier la fonction g de différentes façons.

1. La méthode min-max.

La méthode min-max⁷ consiste à normaliser le score afin qu'il prenne ses valeurs dans $[0, 1]$. Nous avons :

$$s_i^* = \frac{s_i - \min_i s_i}{\max_i s_i - \min_i s_i}$$

2. La méthode PIT (*probability integral transform*).

Notons \mathbf{F}_S la distribution du score S et \mathbf{F}_{S^*} la distribution désirée pour le score S^* . La fonction g est définie de la façon suivante :

$$S^* = \mathbf{F}_{S^*}^{-1}(\mathbf{F}_S(S))$$

Reprenons notre exemple précédent. Nous avons $S \sim \mathcal{U}_{[0,1]}$ et nous voulons que $S^* \sim \mathcal{U}_{[-1,1]}$. Comme $\mathbf{F}_S(x) = x$ et $\mathbf{F}_{S^*}(x) = (x + 1)/2$, nous en déduisons que :

$$g(x) = \mathbf{F}_{S^*}^{-1}(\mathbf{F}_S(x)) = 2x - 1$$

3. La méthode PIT empirique.

La distribution théorique du score S n'étant pas toujours connue, on peut utiliser sa distribution empirique $\hat{\mathbf{F}}_S$. Nous avons donc :

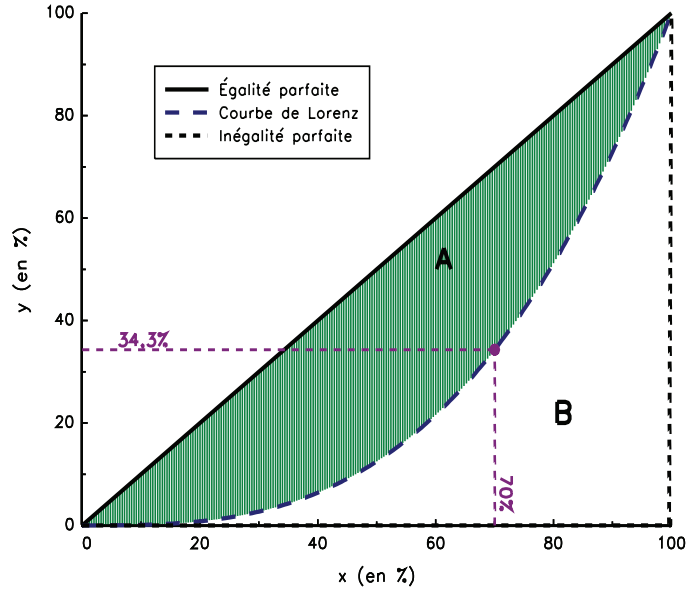
$$S^* = \mathbf{F}_{S^*}^{-1}(\hat{\mathbf{F}}_S(S))$$

⁵C'est notamment le cas lorsque le z-score est normalisé pour prendre des valeurs entières. Dans ce cas, les valeurs du z-score vont généralement de -3 à $+3$ avec un pas de 1.

⁶Elle vaut exactement 0,26%.

⁷A. Jaina, K. Nandakumara et A. Ross, « Score Normalization in Multimodal Biometric Systems », *Pattern Recognition*, 38(12), 2005, p. 2270-2285.

Graphique 9. Courbe de Lorenz



$\Pr\{S \leq s\}$ $\mathbf{F}_0(s) = \Pr\{S < s \mid Y = 0\}$ et $\mathbf{F}_1(s) = \Pr\{S < s \mid Y = 1\}$, nous obtenons :

$$\begin{cases} x(s) = 1 - \mathbf{F}(s) \\ y(s) = 1 - \mathbf{F}_0(s) \end{cases}$$

Cette courbe de Lorenz mesure la performance du score à ne pas sélectionner les mauvais paris. Nous pouvons aussi construire la courbe de Lorenz qui mesure la performance du score à sélectionner les bons paris. Dans ce cas, nous avons :

$$\begin{cases} x(s) = \Pr\{S > s\} \\ y(s) = \Pr\{S > s \mid Y = 1\} \end{cases}$$

ou encore :

$$\begin{cases} x(s) = 1 - \mathbf{F}(s) \\ y(s) = 1 - \mathbf{F}_1(s) \end{cases}$$

Cette courbe s'appelle la courbe de précision. Enfin, la courbe ROC (Receiver Operating Characteristic) est définie par :

$$\begin{cases} x(s) = \Pr\{S > s \mid Y = 0\} = 1 - \mathbf{F}_0(s) \\ y(s) = \Pr\{S > s \mid Y = 1\} = 1 - \mathbf{F}_1(s) \end{cases}$$

Pour chacune de ces courbes, nous pouvons calculer un coefficient de Gini G ainsi qu'un coefficient de Gini normalisé G^* :

$$G^* = \frac{G(\mathcal{L})}{G(\mathcal{L}^*)}$$

avec \mathcal{L}^* la courbe de Lorenz correspondant au score parfait.

À titre d'illustration, nous considérons l'exemple du graphique 8. Nous représentons la courbe de sélection, la courbe ROC et la courbe de précision sur le graphique 10. Les coefficients de Gini correspondants sont donnés dans le tableau 3. Il est intéressant de remarquer que les coefficients de Gini normalisé sont très proches quelle que soit la courbe de performance. Si nous reprenons notre exemple initial long/short, le coefficient de Gini vaut $-0,32$ pour le premier score et $0,64$ pour le deuxième score. Il ne fait plus aucun doute que ce dernier est le plus performant.

Graphique 10. Courbes de sélection, ROC et de précision

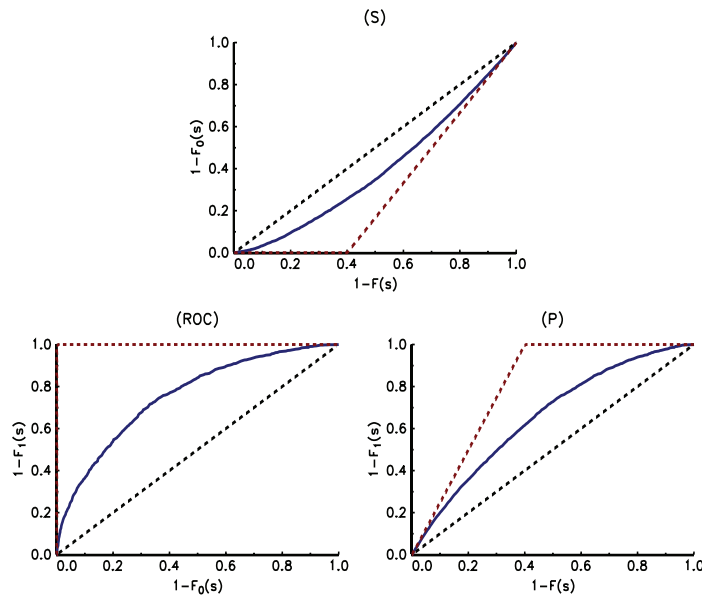


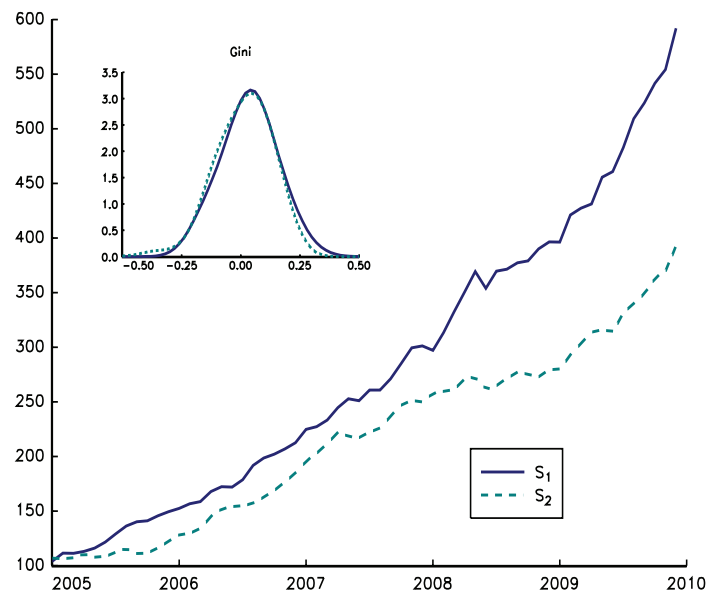
Tableau 3. Coefficients de Gini

Courbe	$G(\mathcal{L})$	$G(\mathcal{L}^*)$	G^*
Sélection	0,20	0,40	0,51
ROC	0,51	1,00	0,51
Précision	0,31	0,60	0,51

Remarque 71. Le graphique 11 présente un backtest long/short sur 100 actions. Tous les mois, on détermine les 50 positions acheteuses et les 50 positions vendeuses en fonction d'un score. Si on considère la performance financière, le score S_1 est largement meilleur que le score S_2 . Pourtant,

si on compare la densité du coefficient de Gini estimée à partir des 60 valeurs mensuelles du coefficient de Gini, il n'est pas évident qu'un score domine l'autre. Une étude plus approfondie du backtest révèle que la très bonne performance financière du score S_1 peut être expliquée par quelques rendements aberrants.

Graphique 11. Performance financière et performance du score



3. Les méthodes d'apprentissage statistique

L'apprentissage statistique (*statistical learning*) désigne de façon générale tout modèle statistique qui permet de prédire une variable Y de sortie à partir d'un nombre fini de variables d'entrée X . Par exemple, la régression linéaire est un modèle d'apprentissage statistique. Dans un sens plus restreint, elle désigne les modèles qui sont estimés sur une population test et qui sont ensuite validés sur une autre population. Ce sont donc des algorithmes qui comprennent deux phases : une phase d'apprentissage et une phase de validation croisée.

Parmi les différentes méthodes d'apprentissage statistique, les méthodes d'ensemble désignent les algorithmes qui permettent d'améliorer la performance de prévision d'un système de classification ou de scoring. L'idée

est de combiner plusieurs estimations du système plutôt que d'en utiliser une seule ou d'agréger plusieurs modèles, tout en évitant un surajustement (*overfitting*). Notons L un algorithme d'apprentissage (*learning algorithm*) et M le nombre de combinaisons (ou d'itérations). Pour chaque itération m , on utilise ensuite une méthode \mathcal{M} qui permet de définir une hypothèse de sortie $h_m : X \rightarrow Y$. La méthode d'ensemble définit alors le système de scoring comme une combinaison de ces différentes hypothèses de sortie :

$$H(x) = f(w_m, h_m)$$

où w_m sont les poids optimaux estimés par l'algorithme d'ensemble. Par exemple, dans le cas d'un système de classification où les sorties sont 1 (bien classé) et -1 (mal classé), $H(x)$ peut prendre la forme suivante :

$$H(x) = \text{sign} \left(\sum_{m=1}^M w_m h_m(x) \right)$$

On classe les algorithmes d'ensemble selon la méthode \mathcal{M} . On distingue en particulier les algorithmes de *boosting* qui sont basés sur des méthodes de construction adaptative des algorithmes de *bagging* qui utilisent une méthode aléatoire.

3.1. Les méthodes de boosting

Les méthodes de boosting trouvent leur origine dans les articles fondateurs de Shapire (1990)²² et Freund (1995)²³. En 1996, ces deux auteurs proposent l'algorithme AdaBoost pour les systèmes de classification²⁴. Friedman *et al.* (2000) résument cette procédure de la façon suivante²⁵ :

1. les poids des observations sont fixés à $\pi_i = 1/n$ pour $i = 1, \dots, n$;
2. on répète les séquences suivantes pour $m = 1, \dots, M$:
 - (a) on estime le système de classification $h_m(x) \in \{-1, 1\}$ en utilisant les poids π_i sur la population d'apprentissage ;
 - (b) on calcule l'erreur e_m ainsi que le poids w_m définis par :

$$e_m = \mathbb{E}_\pi [\mathbf{1}\{y \neq h_m(x)\}]$$

et :

$$w_m = \log \left(\frac{1 - e_m}{e_m} \right)$$

²²R.E. Shapire, « The Strength of Weak Learnability », *Machine Learning*, 5(2), juin 1990, p. 197-227.

²³Y. Freund, « Boosting a Weak Learning Algorithm by Majority », *Information and Computation*, 121(2), 1995, p. 256-285.

²⁴Y. Freund et R.E. Shapire, « Experiments with a New Boosting Algorithm », in *Machine Learning: Proceedings of the Thirteenth International Conference*, Morgan Kaufman, 1996.

²⁵J. Friedman, T. Hastie et R. Tibshirani, « Additive Logistic Regression: A Statistical View of Boosting », *Annals of Statistics*, 28(2), 2000, p. 337-407.

Pour que $\Delta_\mu > 0$, on doit vérifier la condition suivante :

$$\phi \geq \frac{\mu - r}{\sigma}$$

Pour que la règle stop loss soit efficace dans un marché trend following, le paramètre ϕ doit être supérieur au ratio de Sharpe de la stratégie. On obtient donc un résultat paradoxal. Plus la stratégie présente un ratio de Sharpe élevé, plus la règle de stop loss est inefficace.

Le message de ces résultats est très clair. Un stop loss est efficace si la stratégie présente un rendement ajusté du risque peu attractif. Un stop loss peut donc se justifier dans le cas d'une stratégie long only sur des classes d'actifs traditionnelles. D'ailleurs, un backtest sur les indices actions montre que certaines règles stop loss apportent de la valeur. En revanche, si la stratégie présente un très bon rendement ajusté du risque, alors il ne faut surtout pas mettre en place un stop loss⁵.

Malgré ces résultats théoriques, beaucoup de praticiens continuent d'utiliser ces règles de stop loss. Toute la difficulté est alors de calibrer les paramètres n , n' , ℓ et g . On peut calibrer ces paramètres sur la trajectoire historique, mais cela revient à optimiser le backtest. La méthode la plus couramment utilisée est de définir ces paramètres en fonction de la volatilité de la stratégie⁶. Plaçons-nous dans le cas d'un mouvement brownien géométrique classique. Nous avons :

$$\Pr \{R(t; n) \leq -\ell\} = \Phi \left(\frac{-\ell - \left(\mu - \frac{1}{2}\sigma^2\right) \frac{n}{260}}{\sigma \sqrt{\frac{n}{260}}} \right)$$

On calibre la valeur de ℓ en se donnant une période d'observation n (exprimée en jours de trading) et une probabilité d'activer le stop $p_\ell = \Pr \{R(t; n) \leq -\ell\}$. Nous en déduisons que :

$$\ell = \Phi^{-1}(1 - p_\ell) \sigma \sqrt{\frac{n}{260}} - \left(\mu - \frac{1}{2}\sigma^2\right) \frac{n}{260}$$

Si on néglige l'effet tendance, l'expression de ℓ devient :

$$\ell = \Phi^{-1}(1 - p_\ell) \sigma \sqrt{\frac{n}{260}}$$

Par exemple, si n correspond à deux semaines, le seuil ℓ est égal à 13,7% dans le cas où $\sigma = 30\%$ et $p_\ell = 1\%$.

⁵Dit autrement, un stop loss est efficace si la stratégie sous-jacente est mauvaise.

⁶D.L. Schalow, « Setting Stops with Standard Deviations », *Journal of Portfolio Management*, 22(4), 1996, p. 58-61.

Tableau 1. *Statistiques des backtests de stop loss (volatilité 1 an)*

	S&P 500	$p = 5\%$	$p = 20\%$	$n = 60$
Performance (en %)	8,20	9,30	6,84	7,85
Volatilité (en %)	18,59	12,71	11,84	13,32
Ratio de Sharpe	0,21	0,39	0,22	0,27

Tableau 2. *Statistiques des backtests de stop loss (volatilité 3 mois)*

	S&P 500	$p = 5\%$	$p = 20\%$	$n = 60$
Performance (en %)	8,20	8,03	4,89	6,51
Volatilité (en %)	18,59	12,73	11,94	17,37
Ratio de Sharpe	0,21	0,29	0,05	0,13

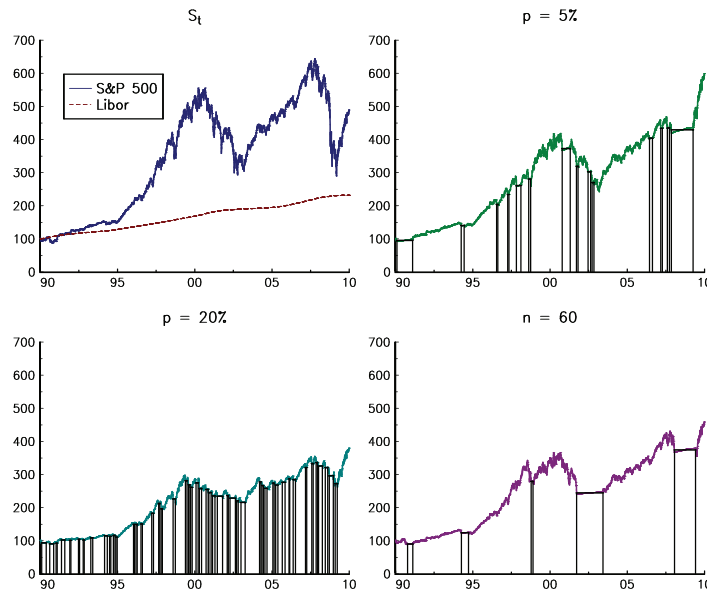
Sur le graphique 1, nous présentons les résultats de la stratégie stop loss appliquée à l'indice S&P 500. Nous supposons que $n = n'$ et $p_\ell = p_g$. Les valeurs par défaut sont n égal à 20 jours de trading et p égal à 5%. La volatilité σ est calibrée à chaque date t en prenant la volatilité historique sur la dernière année. Les statistiques de performance sont présentées dans le tableau 2. À première vue, on obtient des résultats plutôt satisfaisants. Nous remarquons néanmoins que ces résultats sont sensibles aux choix des paramètres. En particulier, si on utilise la volatilité historique 3 mois pour calibrer le paramètre ℓ , nous obtenons des résultats beaucoup moins favorables.

Remarque 72. *La mise en place de prise de profit ressemble à première vue à une règle de stop loss dans le cas d'un marché mean reversion. En fait, ces deux règles sont très différentes. Dans une règle stop loss, on essaye de limiter une perte qui a déjà été réalisée en partie. Dans une règle take profit, on est dans une configuration où la stratégie gagne de l'argent. Toute la difficulté est alors de s'arrêter avant le retournement. Mettre en place une prise de profit revient à croire que l'on est capable de vendre au plus haut et d'acheter au plus bas !*

1.2. La gestion de l'exposition

Le modèle précédent pose une difficulté majeure. C'est un modèle d'allocation binaire. Dans la pratique, on préfère gérer l'exposition de façon dynamique plutôt que d'utiliser une règle d'allocation $w_t \in \{0, 1\}$.

L'exemple le plus simple est la méthode de contrôle de volatilité. Nous avons vu précédemment qu'il y a une relation négative entre la volatilité et le rendement. Une faible volatilité est généralement associée à un marché haussier tandis qu'une forte volatilité est le signe d'un marché baissier.

Graphique 1. Simulation historique de règles stop loss sur l'indice S&P 500

Considérons la règle suivante :

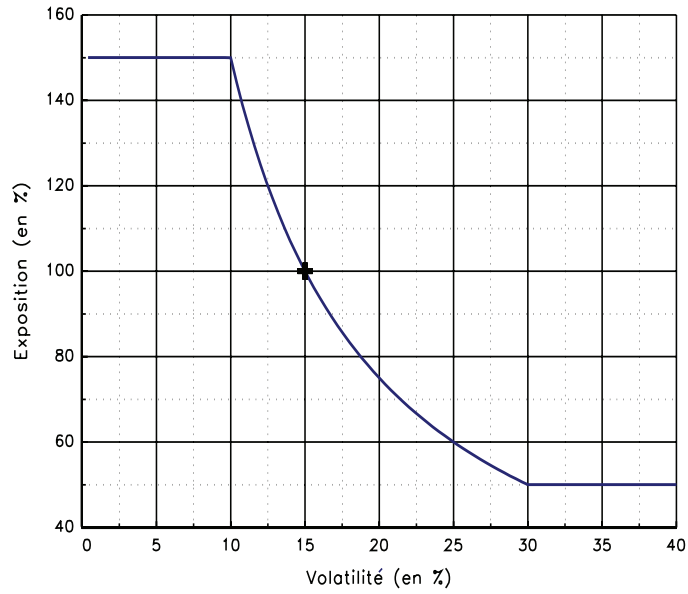
$$w_t = \max \left(\min \left(\frac{\sigma^*}{\sigma_{t-1}}, w_+ \right), w_- \right)$$

w_t est donc une fonction décroissante de la volatilité σ_{t-1} . Si $\sigma_{t-1} < \sigma^*$, alors $w_t > 1$. Nous avons aussi $w_t \in (w_-, w_+)$. À titre d'illustration, le graphique 2 présente la fonction d'exposition w_t pour les paramètres $w_- = 50\%$, $w_+ = 150\%$ et $\sigma^* = 15\%$. Avec ce mécanisme, nous réduisons donc l'exposition dans un environnement à forte volatilité.

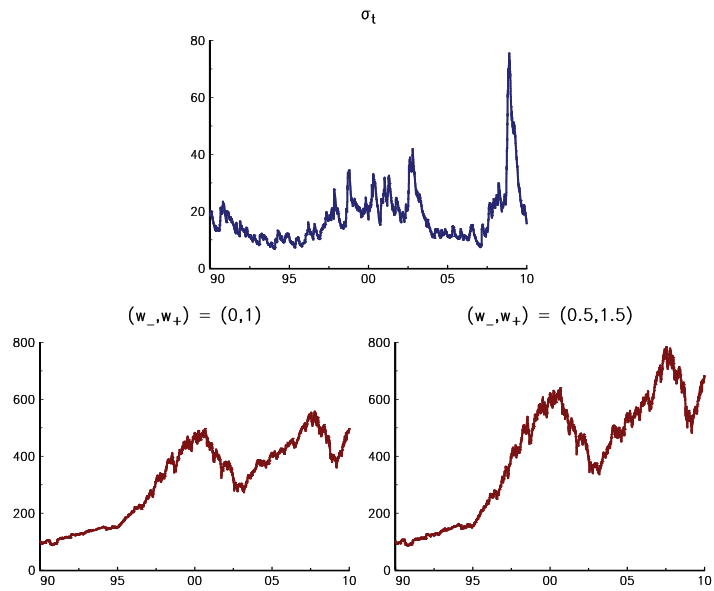
Considérons l'exemple avec l'indice S&P 500. Nous utilisons un modèle GARCH(1,1) pour mesurer la volatilité σ_t . La volatilité cible σ^* est fixée à 20%. Dans le cas $(w_-, w_+) = (0, 1)$, nous obtenons une performance légèrement supérieure à celle de l'indice S&P 500, mais avec une volatilité et un drawdown moindres. Dans le cas $(w_-, w_+) = (50\%, 150\%)$, nous obtenons des résultats similaires en terme de mesure de risque, et la performance est améliorée.

Remarque 73. La gestion de l'exposition en fonction de la volatilité est une stratégie très répandue parmi les hedge funds et le CTA en particulier. Par rapport à la règle d'allocation 0/1, elle permet de ne pas définir une règle de désactivation du stop ad-hoc. Néanmoins, elle n'a de sens que s'il existe une relation entre les régimes de volatilité et de performance.

Graphique 2. Exemple de fonction d'exposition w_t



Graphique 3. Simulation historique du contrôle de volatilité avec l'indice S&P 500



5.1. Topologie des marchés de matières premières

Les marchés financiers des matières premières sont des marchés fortement régulés³⁴. Ceci s'explique par le caractère relativement complexe des contrats, par la nature des transactions et par l'enjeu économique sous-jacent. Contrairement aux actifs financiers traditionnels (par exemple les marchés des actions ou des obligations), les prix des matières premières sont déterminés par l'équilibre entre l'offre et la demande sur le marché physique et ces prix ont une influence importante sur les futures de matières premières. Une mauvaise récolte, une catastrophe naturelle ou un changement climatique vont donc impacter l'évolution des futures, tout comme un embargo ou une suspension d'exportation.

On distingue généralement deux grandes catégories d'acteurs : ceux qui cherchent à se couvrir et ceux qui spéculent. Parmi la première catégorie, on trouve les producteurs, les négociants, les consommateurs et les entreprises agro-alimentaires. Dans la seconde catégorie, on trouve les hedge funds et les opérateurs de marché. Ce sont ces derniers qui assurent la liquidité de ces marchés à terme. La part des opérations de trading (ou des positions spéculatives) est difficile à mesurer³⁵. Plusieurs études montrent cependant que cette part augmente depuis quelques années. Ceci peut avoir un impact non négligeable sur le prix des matières premières physiques. Ainsi, certaines études expliquent la crise alimentaire de 2007-2008 par la part croissante de ces opérations de spéculation³⁶.

La plupart des marchés se trouvent aux États-Unis, à l'exception du LME qui est spécialisé dans les métaux :

- CME (Chicago Mercantile Exchange) ;
- LME (London Metal Exchange) ;
- CBOT (Chicago Board of Trade) ;
- NYBOT (New York Board of Trade) ;
- NYMEX (New York Mercantile Exchange).

³⁴Aux États-Unis, l'agence de régulation est la commission CFTC (*commodity futures trading commission*). De nombreuses données concernant l'activité des marchés de futures sont disponibles sur son site web : www.cftc.gov.

³⁵L'article de Working (1953) est la référence classique concernant la part de la spéculation sur le marché des futures. Pour une revue récente de ce sujet appliqué aux matières premières agricoles, vous pouvez consulter Sanders *et al.* (2010).

H. Working, « Futures Trading and Hedging », *American Economic Review*, 43(3), 1953, p. 314-343.

D.R. Sanders, S.H. Irwin et R.P. Merrin, « The Adequacy of Speculation in Agricultural Futures Markets: Too Much of a Good Thing ? », *Applied Economic Perspectives and Policy*, 32(1), 2010, p. 77-94.

³⁶M. Robles, M. Torero et J. von Braun, « When Speculation matters », *International Food Policy Research Institute*, 57, ageconsearch.umn.edu, février 2009.

On classe généralement les matières premières en trois grandes catégories : l'énergie, les métaux et l'agriculture. Le tableau 4 présente les futures les plus liquides pour chaque catégorie³⁷.

Tableau 4. *Principaux futures de matières premières*

Énergie	Métaux	Agriculture
Fuel	Argent	Cacao
Pétrole	Or	Soja
Gasoil	Platine	Maïs
Gaz naturel	Palladium	Coton
Sans plomb	Zinc	Blé
	Plomb	Sucre
	Aluminium	Bétail

Outre les contrats futures, on peut investir dans les matières premières en utilisant des indices, comme l'indice GSCI (Goldman Sachs Commodity Index), l'indice RICI (Rogers International Commodity Index) ou l'indice CRB (Reuters CRB Futures Index). La composition de ces indices est très différente que ce soit en termes de matières premières ou de poids. Certains sont très focalisés sur le secteur de l'énergie comme l'indice GSCI alors que d'autres sont plus équilibrés (comme l'indice RICI qui contient 35 matières premières). On peut investir dans ces différents indices via des fonds de réplcation, des ETFs ou des produits structurés. Ceci explique l'impact énorme que peuvent avoir ces indices sur le marché des futures³⁸.

5.2. Spécificité des futures de matières premières

Par rapport aux contrats portant sur les actifs financiers traditionnels, les futures de matières premières comportent deux spécificités :

1. la valorisation des futures est beaucoup plus complexe ;
2. la courbe *forward* de structure par terme présente des caractéristiques originales.

³⁷Ces contrats sont dénommés en anglais : *heating oil*, *crude oil* (coté à Chicago) ou *brent* (coté à Londres), *gas oil*, *natural gas* et *rbob gasoline* pour l'énergie, *silver*, *gold*, *platinum*, *palladium*, *zinc*, *lead* et *aluminium* pour les métaux, *cocoa*, *soybeans*, *corn*, *cotton*, *wheat*, *sugar* et *live cattle* pour les matières agricoles, auxquels il faut ajouter le café (*coffee*) et l'huile de soja (*soybean oil*).

³⁸Dans le cas des matières premières agricoles, les positions doivent être déclarées à la CFTC. Pendant très longtemps, la CFTC distinguait les positions long et short, et les positions commerciales (à titre de couverture) et non commerciales (à titre de spéculation). Depuis peu, elle a créé une nouvelle catégorie appelée *index trackers* pour contrôler l'impact des indices sur ces matières premières.

Il n'existe pas un modèle théorique consensuel pour valoriser les futures de matières premières. Tout d'abord, il convient de prendre en compte le coût de stockage s de la matière première. Soit $S(t)$ le prix spot. Le prix forward est donc de la forme :

$$F(t, T) = S(t) e^{(r+s)(T-t)}$$

La détention physique de la matière première entraîne certes un coût, mais elle peut constituer un avantage dans certaines situations. C'est notamment le cas lorsque les niveaux de stocks sont bas dans un contexte de demande forte. On peut modéliser cet avantage comme une prime c (appelée *convenience yield*), que l'on assimile à un paramètre de liquidité de la matière première. Nous avons donc :

$$F(t, T) = S(t) e^{(r-c+s)(T-t)}$$

Ceci explique que la valeur du futures peut être inférieure à la valeur spot si $c > r + s$. Ce modèle ne permet cependant pas d'expliquer toute la dynamique des futures de matières premières³⁹. Les prix à terme obéissent plus à un équilibre intertemporel de l'offre et la demande qu'à une relation d'arbitrage mathématique⁴⁰.

L'autre spécificité des matières premières est le caractère contango et backwardation de la courbe forward. Prenons l'exemple du futures *Crude Oil*. Nous représentons la structure par terme de ce futures à différentes dates sur le graphique 4. Selon les périodes, cette structure par terme peut être croissante ou décroissante⁴¹. Le terme contango désigne une courbe *forward* croissante alors que le terme backwardation correspond à une courbe forward décroissante. Ce phénomène va forcément avoir une incidence sur la méthode de roulage des positions de futures. Généralement, le contrat le plus liquide est le premier contrat. Au moment de l'échéance de ce contrat, on peut rouler sa position sur le contrat de maturité la plus proche (ou contrat *nearby*). Cette façon de rouler la position va avoir une incidence sur le rendement de la stratégie. En effet, si la courbe est en backwardation, le roulage de la position va conduire à un gain alors que nous avons une perte si la courbe est en contango (voir le graphique 5). Une stratégie basée sur les futures de matières premières comporte donc un rendement lié à la méthode de roulage. Ce rendement est appelé *roll return*⁴². D'autres méthodes de *roll de futures* sont possibles, par

³⁹E.S. Schwartz, « The Stochastic Behavior of Commodity Prices: Implications for Valuation and Hedging », *Journal of Finance*, 52(3), 1997, p. 923-973.

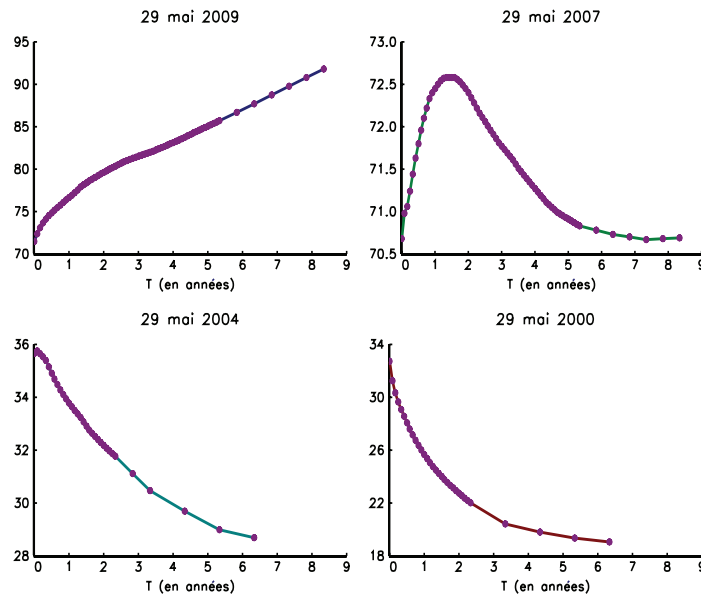
⁴⁰Notons de plus que cette relation ne peut s'appliquer aux matières premières non stockables comme l'électricité.

⁴¹Ce phénomène s'apparente à ce que l'on trouve dans les structures par terme de taux d'intérêt.

⁴²C. Harvey et C. Erb, « The Strategic and Tactical Value of Commodity Futures », *Financial Analysts Journal*, mars 2006, p. 69-97.

exemple rouler la position sur le contrat le plus liquide (c'est-à-dire celui qui présente le plus grand volume) ou rouler le contrat sur le deuxième contrat. Sur le graphique 6, nous comparons ces méthodes de roll dans le cas du café. Nous vérifions que la composante *roll return* est importante.

Graphique 4. Structure par terme du futures Crude Oil



Remarque 77. *Mis à part les hedge funds, l'investissement dans les matières premières est encore marginalisé et peu fréquent parmi les investisseurs institutionnels. Celui-ci est encore perçu comme très risqué. Pourtant, les matières premières peuvent être des actifs très diversifiants dans une allocation stratégique⁴³.*

6. La valorisation des produits optionnels

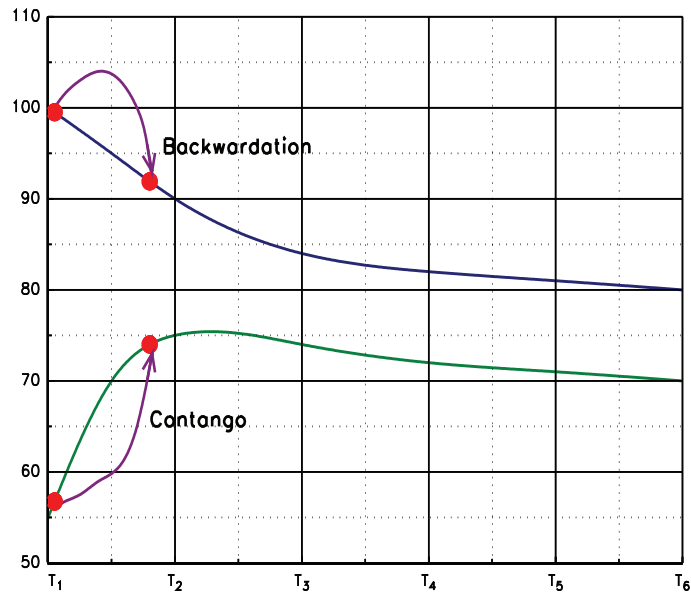
Dans cette section⁴⁴, nous rappelons les principaux concepts concernant la valorisation des produits optionnels et dérivés. Pour un traitement plus exhaustif, le lecteur peut consulter Gatheral (2006)⁴⁵, Hull (2008), Portait

⁴³G. Gorton et K.G. Rouwenhorst, « Facts and Fantasies about Commodity Futures », *Financial Analysts Journal*, 62(2), 2006, p. 47-68.

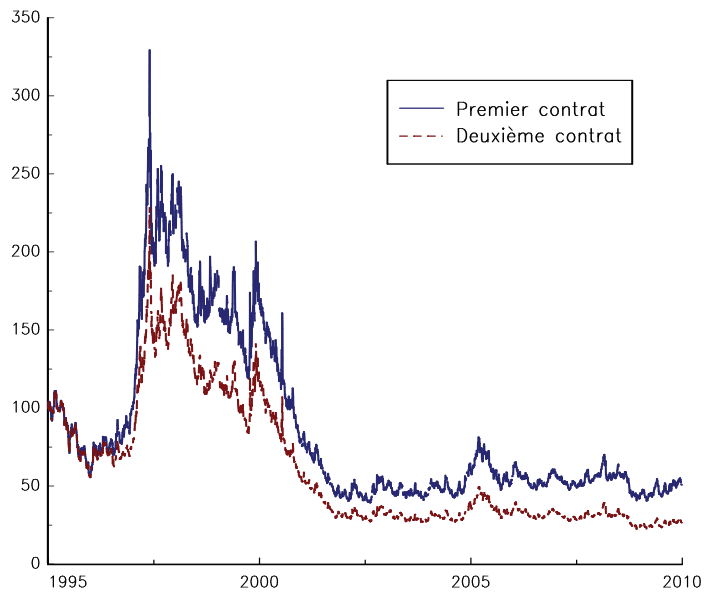
⁴⁴Cette section est tirée en partie de Roncalli (2009).

⁴⁵J. Gatheral, *The Volatility Surface: A Practitioner's Guide*, John Wiley & Sons, 2006.

Graphique 5. Les phénomènes contango et backwardation



Graphique 6. Comparaisons des méthodes de roll dans le cas du café

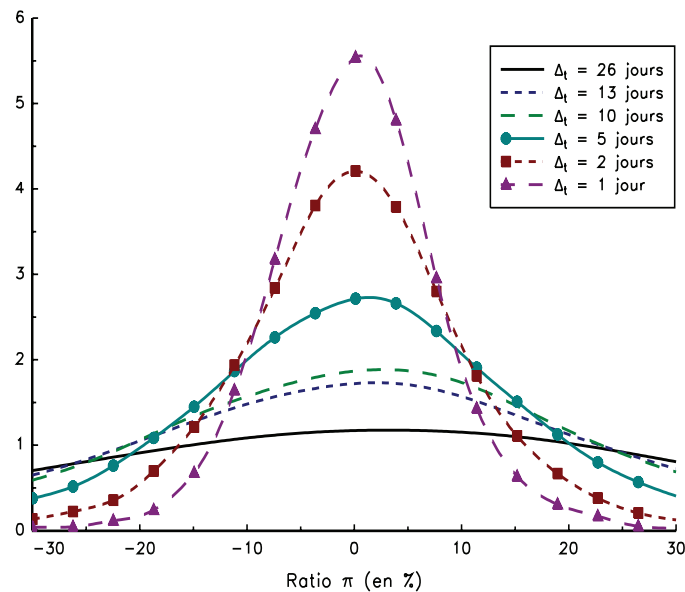


Pour mesurer l'efficacité d'une couverture, on étudie le ratio suivant :

$$\pi = \frac{\text{PnL}(T)}{C(0)}$$

Dans le cas d'une couverture parfaite, ce ratio π est nul. On utilise souvent la volatilité $\sigma(\pi)$ comme mesure d'efficacité. On reprend l'exemple de l'option précédente, mais avec une maturité de 130 jours ou 6 mois (et on suppose qu'une année correspond à 260 jours de trading). On simule 10 000 trajectoires du mouvement brownien géométrique et on estime la densité du ratio π qui est représentée sur le graphique 8 pour différents pas de rebalancement Δ_t . On vérifie que la couverture est d'autant plus efficace que la fréquence de rebalancement est élevée. Ainsi, l'efficacité de la couverture mesurée par $\sigma(\pi)$ est égale respectivement à 8% pour une fréquence journalière alors qu'elle vaut 37% pour une fréquence de rebalancement de 26 jours (voir le graphique 9).

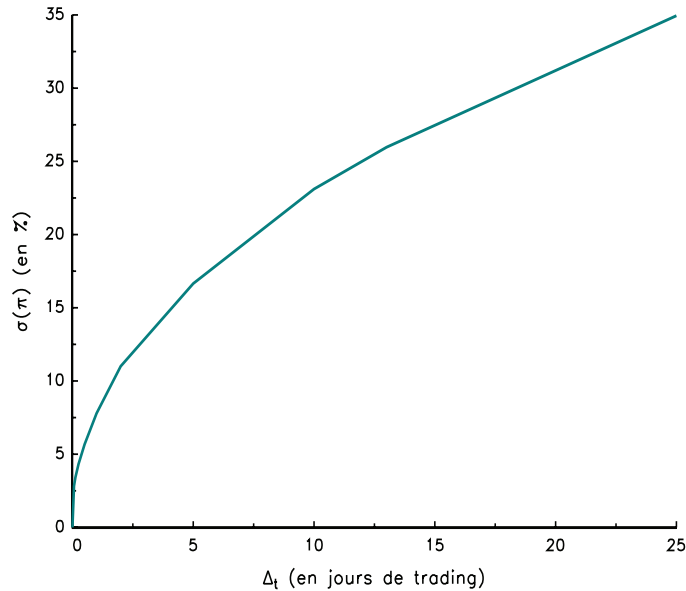
Graphique 8. Densité du ratio de PnL



6.3. La gestion des options

Nous voyons dans un premier temps comment valoriser une position optionnelle dans un fonds d'investissement. Le deuxième paragraphe est consacré aux coefficients de sensibilité qui sont utilisés pour la couverture dynamique et pour comprendre les risques de la position optionnelle. Nous

Graphique 9. Relation entre l'efficacité $\sigma(\pi)$ et la fréquence Δ_t de la couverture



études ensuite la volatilité implicite qui est une paramétrisation de la distribution risque-neutre.

6.3.1. Valorisation des stratégies optionnelles en mark-to-market

Supposons que nous vendons un produit optionnel dans une stratégie à la date t_0 , par exemple un call d'échéance T et de prix d'exercice K . À la date t_0 , nous recevons la prime de l'option C_0 . À la date d'échéance T , le PnL de cette vente de call est égal à :

$$\text{PnL}(T) = C_0 - (S(T) - K)^+$$

Ce PnL peut être positif ou négatif en fonction de l'évolution du sous-jacent. Le cas le plus favorable est lorsque le sous-jacent est inférieur au prix d'exercice et, dans ce cas, nous gagnons la prime de l'option. Que vaut maintenant le PnL pour $t < T$, par exemple pour $t = t_0$? On pourrait être tenté de répondre que le PnL à la date t_0 est égal à la prime du call. Dans ce cas, nous ne valorisons qu'une seule jambe de la stratégie. Avec un tel raisonnement, la valeur de la stratégie serait égale à C_0 jusqu'à $t < T$. Nous observerions alors un possible saut du PnL à la date T si l'option était exercée. Ceci pose un problème dans le cas d'un fonds ouvert. Si nous valorisons les produits optionnels de cette façon, les clients ont tout

intérêt à sortir du fonds dès que la prime de l'option est incluse dans la NAV du fonds.

Dans la pratique, toute stratégie doit être valorisée en mark-to-market pour éviter les incohérences précédentes. Dans l'exemple de vente d'un call, nous avons :

$$\text{PnL}(t) = C_0 - C(t)$$

avec $C(t)$ le prix de marché de l'option à la date t . De cette façon, nous n'aurons pas un saut de PnL à la date d'exercice. Remarquons qu'à l'échéance, le mark-to-market $C(T)$ de l'option est le payoff $(S(T) - K)^+$ et nous retrouvons le PnL précédent. Que vaut maintenant le PnL à l'origine ? La réponse est zéro, puisque le mark-to-market de l'option est justement la prime reçue. En fait, dans la pratique, il est très légèrement négatif. En effet, si on décide de racheter l'option sur le marché, elle ne sera pas exactement égale à C_0 mais légèrement inférieure à cause du spread bid-ask. Le PnL à la date t_0 est en fait le bid-ask de l'option. Pour être plus précis, il faudrait donc valoriser la stratégie en tenant compte de cette asymétrie entre le prix acheteur et le prix vendeur, sachant qu'à l'échéance on converge toujours vers le même PnL $C_0 - (S(T) - K)^+$.

On considère un exemple de vente de call à la monnaie d'échéance 50 jours de trading. Pour avoir des valeurs entières de t , on suppose que l'année comporte 250 jours de trading. La valeur actuelle S_0 du sous-jacent est égale à 100. On considère un taux d'intérêt constant et égal à 5%. Le tableau 7 présente les résultats à différentes valeurs de t pour trois trajectoires simulées $S(t)$ du sous-jacent avec l'hypothèse d'une volatilité implicite constante et égale à 20%. La prime de l'option C_0 est égale à 4,0689. À l'échéance, on vérifie que le PnL est bien égal à la prime moins le payoff à maturité. Ainsi, pour la première trajectoire, le PnL final est égal à la prime puisque le sous-jacent finit à 93,641 et donc l'option n'est pas exercée. Pour la seconde trajectoire, on obtient un PnL très négatif puisque l'option finit très en dedans de la monnaie. Le payoff à maturité est égal à 12,838 et on vérifie que :

$$\text{PnL}(T) = 4,0689 - 12,838 = -8,77$$

Remarquons que le PnL était encore plus défavorable cinq jours de trading avant l'échéance. Le tableau 8 reprend les mêmes trajectoires simulées avec des volatilités implicites qui ne sont plus constantes, ce qui produit des trajectoires différentes de PnL mais un PnL final qui reste le même.

Remarque 80. *Nous utilisons ce principe de valorisation en mark-to-market avec toutes les stratégies optionnelles présentées dans ce livre (covered call, bull spread, variance swap, etc.). Une légère différence peut apparaître cependant concernant la prime reçue ou versée. Dans certains cas, la prime reçue est capitalisée alors que dans d'autres cas elle est*

Tableau 7. Valorisation en MtM avec une volatilité implicite constante

t	$S_1(t)$	$\text{PnL}_1(t)$	$S_2(t)$	$\text{PnL}_3(t)$	$S_3(t)$	$\text{PnL}_3(t)$
0,00	100,000	0,00	100,000	0,00	100,000	0,00
0,02	98,513	1,01	98,530	1,00	104,350	-2,61
0,04	103,921	-2,06	98,362	1,32	103,206	-1,55
0,06	103,069	-1,20	100,815	0,27	103,477	-1,49
0,08	104,012	-1,63	105,871	-3,10	98,183	1,91
0,10	102,891	-0,53	112,434	-8,93	98,992	1,81
0,12	97,277	2,82	112,517	-8,88	99,155	2,04
0,14	95,471	3,58	112,605	-8,85	103,859	-0,65
0,16	94,868	3,88	110,901	-7,04	102,751	0,57
0,18	95,453	4,01	114,888	-10,92	105,594	-1,65
0,20	93,641	4,07	112,838	-8,77	104,209	-0,14

Tableau 8. Valorisation en MtM avec une volatilité implicite non constante

t	$\Sigma_1(t)$	$\text{PnL}_1(t)$	$\Sigma_2(t)$	$\text{PnL}_3(t)$	$\Sigma_3(t)$	$\text{PnL}_3(t)$
0,00	0,200	0,00	0,200	0,00	0,200	0,00
0,02	0,170	1,51	0,341	-1,35	0,209	-2,74
0,04	0,147	-1,38	0,400	-1,82	0,257	-2,39
0,06	0,140	-0,44	0,445	-3,35	0,211	-1,64
0,08	0,120	-0,84	0,483	-6,49	0,202	1,88
0,10	0,116	0,31	0,516	-11,13	0,248	1,22
0,12	0,114	3,64	0,546	-10,85	0,181	2,26
0,14	0,097	4,04	0,574	-10,43	0,234	-0,90
0,16	0,092	4,07	0,600	-8,35	0,202	0,55
0,18	0,078	4,07	0,624	-11,15	0,209	-1,66
0,20	0,051	4,07	0,647	-8,77	0,200	-0,14

réinvestie immédiatement dans la stratégie⁵⁵. De même, la prime versée peut être financée par la stratégie ou en empruntant.

6.3.2. Les coefficients de sensibilité

Nous avons vu précédemment que le delta d'une option est défini par :

$$\Delta = \frac{\partial C}{\partial S}$$

C'est la sensibilité du prix de l'option à l'actif sous-jacent. Dans le cas d'un portefeuille d'option portant sur le même sous-jacent, le delta d'un portefeuille est égal à la somme des deltas individuels. Le delta est une quantité importante pour le gérant de l'option qui va construire le portefeuille de

⁵⁵Dans l'exemple utilisé dans cette section, nous n'avons ni capitalisé ni réinvesti la prime.

Considérons la vente d'une option de maturité T et de strike K . Nous supposons que le vendeur valorise l'option avec une volatilité implicite $\Sigma(T, K)$. Le PnL du vendeur dépendra de la différence entre la variance qui a été vendue $v(T) = \Sigma^2(T, K) \times T$ et la variance $\hat{v}(T)$ subie par la couverture dynamique. Plus précisément, El Karoui *et al.* (1998) ont montré que⁵⁸ :

$$\text{PnL} = e^{-rT} \int_0^T \frac{1}{2} \Gamma(t) (\Sigma^2(T, K) - \sigma^2(t)) S^2(t) dt$$

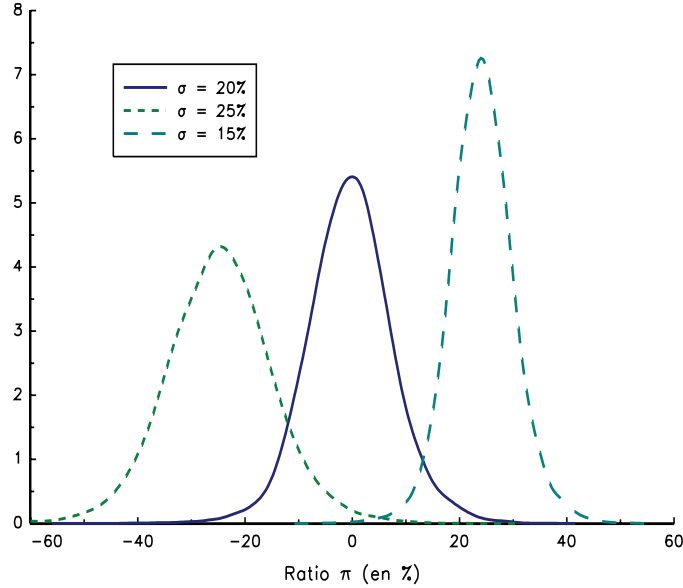
Cette formule est connue sous le nom de formule de robustesse de Black-Scholes. Dans le cas d'une option européenne d'achat, $\Gamma(t)$ est positif. On peut donc dégager un PnL positif si la variance réalisée est inférieure à la variance qui a été vendue. Reprenons l'exemple du graphique 8 à la page 643. Nous reprenons les mêmes paramètres et nous estimons la densité du ratio d'efficacité π de la couverture dynamique pour un pas de rebalancement Δ_t égal à 1 jour de trading. Nous supposons que le trader a vendu l'option avec une volatilité implicite Σ égale à 20% et qu'il utilise cette même volatilité pour gérer la couverture dynamique. Si nous supposons que la dynamique du sous-jacent est un mouvement brownien géométrique dont le coefficient de diffusion est σ , nous obtenons le graphique 13. Dans le cas où $\sigma = \Sigma$, nous retrouvons les résultats précédents, c'est-à-dire que la distribution de π est centrée sur 0. Si la volatilité réalisée σ est très inférieure (resp. supérieure) à la volatilité implicite Σ , alors il est fort probable que le trader ait un PnL positif (resp. négatif). Dans notre exemple, nous avons ainsi :

$$\begin{aligned} \Pr \{ \pi > 0 \mid \Sigma = 20\%, \sigma = 15\% \} &= 99,94\% \\ \Pr \{ \pi > 0 \mid \Sigma = 20\%, \sigma = 25\% \} &= 1,05\% \end{aligned}$$

Remarque 81. *Pour certains produits (options cliquets par exemple), le coefficient $\Gamma(t)$ change de signe au cours du temps. Dans ce cas, il n'est pas possible de définir un niveau de volatilité implicite conservateur. Ce problème a donné lieu au développement du modèle à volatilité incertaine⁵⁹ (uncertain volatility model ou UVM). L'idée sous-jacente est de considérer que la volatilité est comprise entre une valeur minimum et une valeur maximum - $\sigma(t) \in [\sigma_-, \sigma_+]$ - et de couvrir l'option avec la volatilité σ_- lorsque le gamma de l'option est négatif et avec la volatilité σ_+ lorsque le gamma devient positif. On calcule alors le prix de l'option par edp.*

⁵⁸N. El Karoui, M. Jeanblanc et S.E. Shreve, « On the Robustness of the Black-Scholes Equation », *Mathematical Finance*, 8(2), 1998, p. 93-126.

⁵⁹M. Avellaneda, A. Levy et A. Paras, « Pricing and Hedging Derivative Securities in Markets with Uncertain Volatilities », *Applied Mathematical Finance*, 2(2), 1995, p. 73-88.

Graphique 13. Densité du ratio π en fonction de la volatilité réalisée*Relation entre la volatilité implicite et la densité risque-neutre*

Breeden et Litzenberger (1978) ont montré que le smile et la densité risque-neutre sont complètement liés⁶⁰. Nous avons :

$$\Pr \{S(T) \leq K\} = 1 + e^{r(T-t)} \cdot \partial_K C_t(T, K)$$

où $C_t(T, K)$ est le prix à l'instant t de l'option européenne d'achat d'échéance T et de strike K . Soient $\Sigma_t(T, K)$ la surface de la volatilité implicite et $C_t^*(T, K, \Sigma)$ le prix BS de l'option de volatilité implicite Σ . Nous avons :

$$\begin{aligned} \Pr \{S(T) \leq K\} &= 1 + e^{r(T-t)} \cdot \partial_K C_t^*(T, K, \Sigma_t(T, K)) \\ &= 1 + e^{r(T-t)} \cdot \partial_K C_t^*(T, K, \Sigma_t) + \\ &\quad e^{r(T-t)} \cdot \partial_\Sigma C_t^*(T, K, \Sigma_t) \cdot \partial_K \Sigma_t(T, K) \end{aligned}$$

avec :

$$\begin{aligned} \partial_K C^*(T, K, \Sigma) &= -e^{-r(T-t)} \Phi(d_2) \\ \partial_\Sigma C_t^*(T, K, \Sigma) &= S_0 e^{(b-r)(T-t)} \sqrt{T-t} \phi \left(d_2 + \frac{1}{2} \Sigma \sqrt{T-t} \right) \end{aligned}$$

Si on s'intéresse à la fonction de densité, nous avons :

$$\partial_K \Pr \{S(T) \leq K\} = e^{r(T-t)} \cdot \partial_K^2 C_t(T, K)$$

⁶⁰D. Breeden et R. Litzenberger, « State Contingent Prices Implicit in Option Prices », *Journal of Business*, 51, 1978, p. 621-651.

INDEX

1/n (portefeuille), 130, 537
130/30 (stratégie), 25, 109

A

ACP (analyse en composantes principales), 173–177, 305, 309

Action

- Equity swap, 626
- Gestion active, 26
- Gestion indiciaire, 103–114
- Indice, 627
- Screening, 581
- Valorisation, 624

AdaBoost (algorithme), 578

Adams-Bashforth (algorithme de), 209

Adams-Moulton (algorithme de), 209

AIFM (directive), 21

Algèbre linéaire, 170

Allocation

- Stratégique, 17, 32, 73, 128, 138, 357, 480, 634

- Tactique, 32, 138, 357, 551

Alpha (de Jensen), 16, 34, 106, 263, 361, 459

AMF, 17

Antithétique (variable), 251

Apprentissage statistique, 316, 577

APT

- Logiciel, 307
- Modèle, 35, 191

Arbitrage

- Convertible, 31, 527
- Relative value, 507–533
- Statistique, 31, 507
- Volatilité, 498, 525

Arbre binomial, 653

ARCH (modèle), 281, 378–383

ARFIMA (processus), 412, 434

ARMA (modèle), 339–341, 355, 399, 430, 541

Assurance de portefeuille, 443

AuM (assets under management), 9

Aversion au risque, 124, 154, 155, 160, 506

B

- Backfilling, 595
 Backtest, 47–71, 484, 488
 Backwardation, 315, 633
 Bagging (algorithme), 580
 Bande (matrice), 187, 196, 226
 Barbell (stratégie), 522
 Bayes-Stein (estimateur de), 308
 Bellman, voir Programmation dynamique
 Benchmark, 20, 51, 86, 360, 506, 537, 614, 627
 Bêta, 111, 263, 308, 357, 528, 626
 BFGS (algorithme), 123
 Bi-section (algorithme), 97, 119, 140, 184, 213, 482, 616, 649
 Biais du survivant, 595
 Bid-ask (spread), 114, 489, 601, 645
 Bilinéaire (interpolation), 82
 Black-Litterman (modèle de), 138–143, 315
 Black-Scholes (formule), 207, 461, 637
 Blend (style), 26
 Bond picking, 28, 583
 Boosting (algorithme), 578
 Bootstrap, 580
 Bottom-up (approche), 29
 Box-Muller (algorithme de), 237, 258
 Brownien géométrique (mouvement), 62, 153, 196, 207, 229, 237, 271, 281, 286, 588, 636, 643
 Broyden (algorithme de), 121, 213, 226
 Budget de risque, 128
 Bull Spread (stratégie), 480–485
 Buy and hold (stratégie), 49, 67
- C**
- CARA (utilité), 124
 Carry trade, 27, 101, 509
 Cauchy (problème de), 208
 CDO (collateralized debt obligation), 508, 528
 CDS (credit default swap), 527, 619
 CESR, 17
 Change (taux de), 509, 629
 Cholesky (décomposition de), 171, 237
 CIR (processus), 272, 286, 657
 Coûts de transaction, 114, 474
 Cœur-satellite (stratégie), 457–460
 Cointégration, 342, 542
 Commodities, voir Matières premières
 Constant mix (stratégie), 49, 160, 452, 599
 Contango, 315, 633
 Contribution en risque, 90, 128, 163, 531
 Contrôle (variable de), 253
 Convenience yield, voir Matières premières
 Convertible (arbitrage de), 31, 527
 Copule
 - Archimédienne, 314
 - Définition, 312
 - Normale, 314, 560
 - t de Student, 314
 Correction d'erreurs (modèle à), voir VECM
 Corrélation
 - Constante, 91, 131, 310
 - Estimation, 199, 299–312
 - Factorielle, 300, 532
 - Implicite, 501
 - Simulation, 244
 - Trading, 501
 Couverture de change, 55
 Covariance (matrice de)
 - Analyse factorielle, 303
 - Estimation, 197, 299–312
 - Ledoit-Wolf, 310
 - Maximum de vraisemblance, 299
 Covered call (stratégie), 24, 474–480

CPPI (méthode), 32, 448–457
 Crank-Nicholson (schéma), 224
 Crédit (stratégie de), 526
 Creuse (matrice), 187, 226
 CRRRA (utilité), 124, 154, 161, 452
 CSSF, 18
 CTA (stratégie), 30, 465, 540, 630
 Currency swap, 630

D

Data snooping, 597
 DCF (discounted cash flows), 625
 Delta (coefficient de sensibilité),
 461, 486, 495, 526, 527,
 637, 639, 646
 Dense (matrice), 187
 Densité spectrale, 395
 DFP (algorithme), 123
 Différences (méthode des), 201,
 228
 Différences finies (algorithme des),
 221–232, 638
 Dirichlet (conditions de), 225
 Discrépance faible (séquence à),
 255
 Discrimination (courbe de), 569
 Dispersion, 133, 501
 Distressed securities, 31
 Diversification, 93, 130
 Dividende, 625
 Dominance stochastique, 572
 Drawdown, 64, 472
 Duration, 27, 517, 617

E

Échantillonnage d'importance, 253,
 292, 389
 Échantillonnage d'un portefeuille,
 108, 627
 EDO (équation différentielle or-
 dinaire), 147, 208–220
 EDP (équation aux dérivées par-
 tielles), 147, 189, 221, 636
 EDS (équation différentielle sto-
 chastique), 149, 237
 Efficience des marchés, 546, 625

EM (algorithme), 273
 EMM (méthode d'estimation), 287
 Ensemble (méthode d'), 577
 Eonia, 614
 Equity hedge, 30, 78
 Equity market neutral, 25, 31, 78,
 111, 508, 528–533
 Equity swap, 626
 ERC (portefeuille), 132, 530
 Espace-état (modèle), 341, 349–
 370, 384, 541
 ETF (Exchange Traded Fund),
 24, 632
 Euler
 · Algorithme, 209, 222, 238,
 272
 · Décomposition, 128
 Euribor, 614
 Eurodollar, 624
 Event driven, 31
 Explicite (schéma), 223
 Exponentielle (matricielle), 220

F

Facteurs (modèle à), 131, 173, 303,
 358
 Faure (générateur de), 256
 FCP, voir OPCVM
 Feynman-Kac (théorème de), 227
 Filtrage
 · Kalman, 349, 541
 · Particulaire, 370–377, 391
 · Spectral, 422
 · Temps-fréquence, 430
 Fixed income, 613
 Flechter-Reeves (algorithme de),
 123
 FLS (flexible least squares), 189
 Fokker-Planck (équation de), 227,
 272
 Fonds
 · Flexible, 163
 · Lifestyle, 155
 · Profilé, 160, 357
 · Target date, 165
 Fonds de pension, 155

- Forward
- Contrat, 611
 - Prix, 620, 621
 - Taux, 55, 515, 522
- Fractionnaire (processus), voir AR-FIMA
- Frais de gestion, 13, 51–55, 614
- Frontière efficiente, 95
- Fund picking, 29
- Futures
- Contrat, 506, 612, 623, 630
 - Prix, 620, 622, 630
 - Taux d'intérêt, 623
- ### G
- Gamma (coefficient de sensibilité), 486, 494, 647
- GARCH (modèle), 378–383, 488, 590
- Garp (style), voir Blend
- Gear (algorithme de), 209
- Générateur congruentiel linéaire, 234
- Générateur markovien, 181
- Gestion
- Active, 28
 - Alternative, 12, 30
 - Collective, 18
 - Discrétionnaire, 29
 - Diversifiée, 29, 73
 - Indicielle, 28, 103–114
 - Overlay, 31
 - Passive, 28, 103
 - Profilée, 29
 - Quantitative, 29
 - Sous mandat, 23
 - Structurée, 32, 443–460
 - Tiltée, 103
 - Traditionnelle, 10
- Gibbs (échantillonnage de), 289, 389
- Gini (coefficient de), 196, 574
- Glide path, 165
- Global macro (stratégie), 30, 357, 546–552
- GMM (méthode généralisée des moments), 275–283
- Gordon-Shapiro (formule de), 625
- Gradient conjugué (algorithme du), 122
- Griddy Gibbs (échantillonnage), 290, 389
- Growth
- Facteur, 87, 626
 - Style, 26, 135
- ### H
- Halton (générateur de), 256
- Hamilton-Jacobi-Bellman (équation de), 147, 452
- Hammersley (séquence de), 256
- HARA (utilité), 124, 452
- Hedge funds, 12, 30, 78, 142, 315, 357, 528
- Hermite (quadrature de), 202
- Hermitienne (matrice), 170
- Heston (modèle de), 231, 286, 657
- Histogramme, 293
- Hit rate, 566
- Hopscotch (algorithme), 226
- Hurst (exposant de), 16, 416
- ### I
- IGARCH (modèle), 380
- Implicite (schéma), 224
- Importance sampling, 371
- In-sample (simulation), 595
- Indice, 627
- Inférence indirecte, 287
- Information (ratio d'), 65, 104
- Instrumentale (variable), 279
- Intégration numérique, 152, 200–208, 248, 545, 617, 638
- Interpolation
- Bilinéaire, 82
 - Quadratique, 84
 - Spline cubique, 189, 195
- Inversion (méthode de l'), 235
- ### J
- Jacobi (matrice de), 205

K

Kalman (filtre de), 189, 286, 341,
349–357, 362, 383, 541
Kappa (mesure), 69
Kendall (tau de), 313, 559
Kohonen (carte de), 334
Kolmogorov-Smirnov (statistique
de), 411, 572

L

Laguerre (quadrature de), 202
LAPACK, 170
Legendre (quadrature de), 202
Levier (effet de), 13, 19, 51, 56,
528
LHP (liability hedging portfolio),
157
LHS (liability hedging swap), 160
Liability-driven investment (LDI),
157
Libor, 614
Liquidité, 601, 604
Lissage

- Kalman, 351
- Ondelettes, 432
- Spectral, 405
- Spline cubique, 65, 195

LOG (utilité), 125, 456
Logit (modèle), 122, 270
LogitBoost (algorithme), 579
Long/short (stratégie), 13, 30, 100–
103, 109, 117, 357, 458,
468, 508, 528, 554, 570
Lorenz (courbe de), voir Gini
LU (décomposition), 172
Lucas (modèle de), 279, 552, 626

M

Marché

- Comptant, 610
- Dérivé, 610
- Organisé, 610

Mark-to-market (valorisation), 477,
497, 622, 626, 639, 644
Market maker, 24
Market neutral, 528

Markov (chaîne de), 179, 288, 565
Markowitz (modèle de), 94
MARS (Multivariate Adaptive Re-
gression Splines), 336, 554
Matières premières, 12, 315, 630–
634
Matrice aléatoire, 309
Maximum de vraisemblance, 122,
266–275, 277, 286, 347,
381, 384, 406, 409
MCMC (algorithme), 288–293, 386
MDP (portefeuille), 133
Mean reverting, 49, 412, 467, 468,
500, 540–545, 556, 587
Mémoire longue (processus de),
434
Merger arbitrage, 31
Metropolis-Hastings (algorithme
de), 292
Milstein (schéma de), 242
Min-max (score), 558
Moindres carrés

- Généralisés, 356
- Ordinaires, 261, 346, 381
- Pondérés, 264
- Récursifs, 351

Moments (méthode des), 276
Momentum, voir Trend following
et Mean reverting
Monétaire dynamique (stratégie),
508
Monte Carlo (méthode de), 152,
233–258, 283–293, 638
Mortgage (arbitrage de), 527
Multigestion, 29
Mutual fund, voir OPCVM

N

NAV (Net Asset Value), 51, 645
Neumann (conditions de), 225
Newcits, 25
Newton-Raphson (algorithme de),
97, 120, 236, 482
Non funded (stratégie), 50
Non paramétrique

- Estimation, 293–299

- Régression, 297
 - Statistique, 572
- Noyau (méthode du), 78, 293

O

- Obligation
- Convertible, 527
 - Gestion active, 27
 - Valorisation, 615
- OBPI (méthode), 444
- Oméga (mesure), 69
- Ondelettes (analyse en), 424–439
- OPCVM, 22
- Optimalité de Bellman, voir Programmation dynamique
- Optimisation
- Linéaire, 74
 - Non linéaire, 119–143
 - Quadratique, 78
 - Sous contraintes, 125, 126
- Option
- Call, 461, 468, 496, 613
 - Contrat, 613
 - Couverture dynamique, 639
 - Put, 468, 496, 613
 - Spread option, 207, 253
 - Straddle, 485, 492
 - Strangle, 465
 - Stratégie, 447, 461–485
- Ordre (statistique d'), 295
- Ornstein-Uhlenbeck (processus de), 161, 227, 229, 238, 271, 282, 540, 555
- Orthogonale (matrice), 170
- Out-of-sample (simulation), voir In-sample

P

- P/E (ratio), 555, 581
- Pair trading, 31, 530, 553
- Parseval (décomposition de), 420
- Particulaire (filtrage), 370, 391
- Performance (courbe de), 568
- Périodogramme, 405, 409, 430
- PIT (probability integral transform), 558

- Polak-Ribiere (algorithme de), 123
- Pont brownien, 242
- Ponzi (schéma de), 218, 483
- Portage, 507, 616, 647
- Portefeuille de marché, 98
- PPR (Projection Pursuit Regression), voir MARS
- Prédiction-correction (algorithme de), 209
- Price-to-book (ratio), 555, 582
- Pricing, voir Valorisation
- Private equity, 12, 21, 27
- Probit (modèle), 122, 270
- Problème inverse, 136
- Programmation
- Dynamique, 143–167
 - Linéaire, 74
 - Quadratique, 78–118, 262

Q

- QR (décomposition), 121, 184
- QuadPack, 205
- Quadratique (interpolation), 84
- Quadratures (méthode des), 201, 250
- Quasi linéaire (relation), 185
- Quasi Monte Carlo (méthode de), 255
- Quasi-Newton (algorithme de), 122

R

- Racine carrée (matrice), 171, 197
- Racine unité (test de), 345
- Rang (statistique de), 554
- Real estate, 21
- Rebalancement (d'un portefeuille), 47, 596
- Réduction de variance, 250
- Régression (modèle de)
- De style, 16, 35, 84, 358
 - Linéaire, 84, 261–267
 - Loess, 297
 - Non paramétrique, 78, 297
 - Quantile, 76, 266, 298, 315
 - Robuste, 265
- Relative value, voir Arbitrage

- Rendement actuariel, 516
 Rentabilité (calcul de), 59
 Réplication (technique de), 28, 108, 358, 365, 373
 Reporting, 58
 Réseau de neurones, 316–335, 554
 Rhô de Spearman, 572
 Risk budgeting, voir Budget de risque
 Risque (mesure de), 61
 ROC (Receiver Operating Characteristic), 575
 Roll-down (stratégie), 518
 Roll return, 633
 Rotation du portefeuille, voir Turnover
 Rotation sectorielle (stratégie), 547
 Runge-Kutta (algorithme de), 209
- ### S
- SABR (modèle), 231, 658–661
 Schur (décomposition), 176, 200
 Scoring, 100, 553
 Screening, 553, 580–583
 SEC, 18
 Sector neutral, 31, 528
 Sélection (courbe de), 568
 Semi-variance, 67
 Shannon (entropie de), 439, 563
 Sharpe (ratio de), 65, 98
 Sherman-Morrison-Woodbury (formule de), 121, 352
 Short bias (stratégie), 30
 Shrinkage (méthode de), 308, 432
 SICAV, voir OPCVM
 Signal (d'achat ou de vente), 488
 Simplexe
 - Algorithme, 74
 - Approximation, 191
 Simpson (méthode de), 200
 Simulation, voir Monte Carlo
 Skew bêta, 78, 337
 Small cap (facteur), 26, 35, 87, 365, 626
 SMC (Sequential Monte Carlo), 371, 386
 SMM (méthode simulée des moments), 283
 Sobol (générateur de), 256
 SOM (self-organizing map), voir Kohonen
 Sortino (ratio de), 68
 Spearman (rhô de), 313
 Spectrale (analyse), 283, 395–424, 535
 Spline (fonction), 194
 Spot
 - Marché, 610
 - Prix, 610, 620, 630, 633
 - Taux, 514, 610
 SQP (algorithme), 126, 193
 Stationnaire (forme), 400
 Stock picking, 25, 28, 31, 87, 530
 Stop loss, 466, 506, 586
 Structure par terme, voir Taux d'intérêt
- ### T
- Swap
 - Action, 626
 - Change, 630
 - Contrat, 612
 Swap de variance, 489–509
 Swaption, 526
 Symétrique (matrice), 170
- ### T
- Take profit, 589
 Tau de Kendall, 572
 Taux d'intérêt
 - Carry, 519
 - Eonia, 614
 - Euribor, 614
 - FRA, 620
 - Instantané, 614
 - Libor, 614
 - Monétaire, 613
 - Parité couverte (CIP), 509, 630
 - Parité non couverte (UIP), 509
 - Stratégie, 514
 - Structure par terme, 173, 217, 227, 514, 615, 618

- Taux de change (stratégie), 487, 509
- Taux de succès, voir Hit rate
- Thêta (coefficient de sensibilité), 486, 647
- Thêta-schéma, 223
- Tobit (modèle), 271
- Top-down (approche), 29, 547
- Tracker, voir ETF
- Tracking error, 28, 103, 140, 538, 627
- Trading de dispersion, 501
- Transformations (méthode des), 236
- Transformée de Fourier, 405, 426
- Transformée en ondelettes, 425
- Trapèzes (méthode des), 200
- Trend following, 31, 101, 135, 412, 462, 467, 534–540, 587
- Tridiagonal (algorithme), 188, 195, 225
- Turnover, 115, 597
- U**
- UCITS (directive), 17
- Unitaire (matrice), 170
- Utilité (fonction d'), 16, 123
- V**
- Valeur actuelle nette (VAN), 609
- Valeurs propres (décomposition en), 170, 198, 205
- Valeurs singulières (décomposition en), 172
- Valorisation, 609
- Value
- Facteur, 87, 626
 - Style, 26
- Value-at-Risk, 20, 64, 198
- VAR (processus), 339, 356
- Variance minimale (portefeuille de), 90, 130
- Vasicek (modèle de), 161, 227
- VECM (modèle), 342–349, 542
- Véga (coefficient de sensibilité), 486, 494, 647
- VIX (indice de volatilité), 417, 504
- Volatilité
- Arbitrage, 498, 525, 540
 - Contrôle, 102, 531, 589
 - Estimation, 62, 130, 376
 - Historique, 471
 - Implicite, 471, 527, 649, 658
 - Locale, 655
 - Réalisée, 489, 495, 650
 - Smile, 497, 649
 - Stochastique, 376, 383–394, 657
 - Stratégie, 484
- Vovol, 286, 383, 658
- W**
- Whittle (estimation de), 406
- Y**
- Yield book, 40
- Yield curve, voir Taux d'intérêt
- Yield-to-maturity, 517, 522
- Z**
- Z-score, 556
- Zéro-coupon, 151, 157, 174, 227, 446, 514, 615

Table des matières

Remerciements	5
Préface	7
Introduction générale	9
L'industrie de la gestion d'actifs	9
La réglementation	17
Les différents styles de gestion.....	26
La gestion quantitative	32
Plan du livre	41

Première partie

Les outils mathématiques

Chapitre 1 – La construction d'un backtest	47
1. Calcul de la trajectoire d'un panier de stratégies.....	47
1.1. Les stratégies financées	47
1.2. Les stratégies non financées	50
2. Prise en compte des frais de gestion	51
3. Couverture d'une position de change.....	55
4. Les effets de levier.....	56
5. Le reporting d'un backtest.....	58
5.1. Les mesures de rentabilité	59

5.2.	Les mesures de risque	61
5.3.	Les mesures de performance ajustée du risque	65
5.4.	Un exemple.....	69
Chapitre 2 – Les méthodes d'optimisation		73
1.	La programmation linéaire	74
1.1.	L'algorithme du simplexe	74
1.2.	La méthode des points intérieurs	76
1.3.	Application à la régression quantile.....	76
2.	La programmation quadratique	78
2.1.	Spécification d'un programme quadratique.....	78
2.2.	Application à la régression de style	84
2.3.	Application au portefeuille de variance minimale....	90
2.4.	Problème d'allocation de Markowitz.....	94
2.5.	Construction d'un portefeuille long/short avec contrôle de volatilité.....	100
2.6.	La gestion indicielle actions	103
2.7.	Prise en compte des coûts de transaction	114
3.	L'optimisation non linéaire	119
3.1.	Résolution d'équations non linéaires.....	119
3.2.	Les algorithmes numériques d'optimisation non linéaire	122
3.3.	La prise en compte de contraintes linéaires d'égalité	125
3.4.	Le principe de la programmation quadratique séquentielle	126
3.5.	Allocation stratégique sous contraintes de budget de risque	128
3.6.	La construction de portefeuilles diversifiés.....	130
3.7.	Exemples de problèmes inverses.....	136
4.	La programmation dynamique	143
4.1.	L'approche de Bellman	144
4.2.	L'optimisation dynamique de portefeuille.....	153
4.3.	Quelques extensions du modèle de Merton	155
Chapitre 3 – Les méthodes numériques		169
1.	L'algèbre linéaire	170
1.1.	Les méthodes de décomposition.....	170
1.2.	Les matrices bandes et creuses	187
2.	Les méthodes d'approximation	191
2.1.	Approximation d'un simplexe.....	191
2.2.	Les fonctions splines cubiques	194
2.3.	Approximation d'une matrice définie positive	196
2.4.	L'intégration numérique	200
2.5.	La résolution d'équations différentielles ordinaires ..	208
2.6.	La méthode des différences finies	221
3.	Les méthodes de simulation et de Monte Carlo	233

3.1.	Simulation de nombres aléatoires	233
3.2.	Simulation des processus de diffusion	237
3.3.	Simulation d'une matrice de corrélation	244
3.4.	La méthode de Monte Carlo	247

Deuxième partie

Les outils économétriques

Chapitre 4 – Les outils statistiques		261
1.	Les différentes méthodes d'estimation	261
1.1.	La régression linéaire	261
1.2.	Le maximum de vraisemblance	266
1.3.	La méthode généralisée des moments	275
1.4.	Les méthodes d'estimation basées sur les simulations	283
1.5.	L'estimation non paramétrique	293
2.	Modélisation de la dépendance statistique	298
2.1.	Modélisation des matrices de covariance et de corrélation	299
2.2.	Les fonctions copules	312
3.	Les réseaux de neurones artificiels et autres modèles sta- tistiques d'apprentissage	316
3.1.	Le perceptron et les réseaux de neurones multi- couches	318
3.2.	Les cartes de Kohonen	334
3.3.	Le modèle MARS	336
Chapitre 5 – La modélisation des séries temporelles		339
1.	Les modèles ARMA	339
1.1.	Le cas du modèle VAR(1)	339
1.2.	Extension aux modèles ARMA	340
2.	Les modèles à correction d'erreurs	342
2.1.	La notion de cointégration	342
2.2.	Les mécanismes à correction d'erreurs	344
2.3.	Tests et estimation des relations de cointégration ...	344
3.	Les modèles espace-état	349
3.1.	Spécification et estimation d'un modèle espace-état	349
3.2.	Quelques applications	351
4.	Les filtres particuliers	370
4.1.	Échantillonnage préférentiel	371
4.2.	Calcul des poids pour les techniques SMC	371
4.3.	Quelques exemples	373
5.	Les modèles à volatilité conditionnelle ou stochastique	376
5.1.	Les modèles ARCH et GARCH	378
5.2.	Les modèles à volatilité stochastique	383
6.	L'analyse spectrale	395

6.1.	Définition de la densité spectrale	395
6.2.	Localisation dans le domaine des fréquences	396
6.3.	Quelques propriétés de la densité spectrale.....	398
6.4.	L'estimation dans le domaine spectral	405
6.5.	Extension au cas multi-dimensionnel	409
6.6.	Quelques applications	411
7.	L'analyse en ondelettes	424
7.1.	La représentation temps-fréquence	424
7.2.	La transformée en ondelettes	425
7.3.	Quelques applications	430
7.4.	L'analyse en paquets d'ondelettes	438

Troisième partie

La gestion quantitative

Chapitre 6 – Les stratégies quantitatives	443
1. La gestion structurée	443
1.1. L'assurance de portefeuille.....	443
1.2. L'approche cœur-satellite	457
2. Les stratégies optionnelles	461
2.1. Dualité avec les stratégies de gestion	461
2.2. Les stratégies de call et de put	468
2.3. Les stratégies plus complexes	474
3. Les stratégies de volatilité	484
3.1. Relation avec les stratégies optionnelles	485
3.2. Les swaps de variance	489
4. Les stratégies d'arbitrage et de portage	507
4.1. Le carry trade	509
4.2. Les stratégies de taux d'intérêt.....	514
4.3. Les stratégies de crédit	526
4.4. La stratégie <i>equity market neutral</i>	528
5. Les stratégies de momentum	533
5.1. La stratégie <i>trend following</i>	534
5.2. La stratégie <i>mean reverting</i>	540
5.3. Les stratégies de momentum et l'hypothèse d'efficience des marchés.....	546
6. Les stratégies <i>global macro</i>	546
6.1. La rotation sectorielle	547
6.2. L'allocation tactique.....	551
Chapitre 7 – Les outils de scoring	553
1. La construction des scores	553
1.1. Les méthodes d'élaboration des scores	553
1.2. L'agrégation de scores.....	559
2. L'évaluation des scores.....	562

2.1.	L'entropie de Shannon	563
2.2.	Les outils graphiques	567
2.3.	Les mesures statistiques de performance	572
3.	Les méthodes d'apprentissage statistique	577
3.1.	Les méthodes de boosting.....	578
3.2.	Les méthodes de bagging.....	580
4.	Les outils de screening	580
4.1.	Le screening d'actions.....	581
4.2.	Le screening d'autres univers	583
Chapitre 8 – La gestion des risques		585
1.	Le risque de perte	586
1.1.	La définition de <i>stop loss</i>	586
1.2.	La gestion de l'exposition	589
2.	Le risque de simulation	592
2.1.	Les données	592
2.2.	L'estimation historique des paramètres.....	595
2.3.	Le problème du rebalancement du portefeuille	596
3.	Le risque d'investissement	597
3.1.	La rotation du portefeuille.....	597
3.2.	La prise en compte de la liquidité	601
Conclusion générale		607
Annexe – Les instruments financiers		609
1.	Description générale des différents types de contrats.....	610
1.1.	Les titres au comptant	610
1.2.	Les produits dérivés	610
2.	Les produits de taux d'intérêt	613
2.1.	Le marché monétaire	613
2.2.	Les obligations.....	615
2.3.	Les instruments à terme.....	620
3.	Les actions.....	624
3.1.	Modèles de valorisation	625
3.2.	Les contrats swaps sur actions.....	626
3.3.	Les indices d'actions	627
4.	Les devises.....	629
4.1.	Les forward de taux de change	630
4.2.	Les swaps de change	630
5.	Les matières premières.....	630
5.1.	Topologie des marchés de matières premières.....	631
5.2.	Spécificité des futures de matières premières.....	632
6.	La valorisation des produits optionnels	634
6.1.	Le modèle de Black et Scholes.....	636
6.2.	La couverture dynamique en delta des produits optionnels	639
6.3.	La gestion des options.....	643

6.4. Les autres modèles de valorisation	653
Bibliographie générale	663
Index	667