

HEC MONTRÉAL
École Affiliée à l'Université de Montréal

Gauthier WEBANCK

**Le modèle de chaînes de Markov cachées dans une stratégie de
réplication de fonds de couvertures**

Mémoire présenté à HEC Montréal en vue de l'obtention du grade de
Maîtrise ès Science de la Gestion (M.Sc.) en Ingénierie Financière

Avril 2011

©Gauthier Webanck, 2011

Remerciements

Je tiens d'abord à remercier mes co-directeurs de mémoire, Bruno Rémillard et Nicolas Papageorgiou, qui ont été déterminants tout au long de mon travail. Je tiens aussi à les remercier de m'avoir donné la problématique sur laquelle j'ai travaillé car je l'ai trouvée très pertinente, pratique et elle m'a permis d'approfondir mes connaissances en ingénierie financière.

Je tiens également à remercier Alexandre Hocquard, Hugues Langlois et Nicolas Ponce à côté de qui j'ai pu travailler durant ce projet de recherche. Leurs expériences et leur compréhension des différents modèles présentés dans ce mémoire, ainsi que leurs connaissances en méthodes numériques m'ont permis d'avancer et d'avoir une meilleure compréhension des problèmes auxquels j'ai été confrontés. Je tiens aussi à exprimer ma reconnaissance envers le Groupe de Recherche en Placements Alternatifs DGIA-HEC qui nous a permis de travailler en groupe durant toute une année, et ce, dans d'excellentes conditions.

Les dernières personnes que je tiens à remercier, et non les moindres, sont mes parents, ma soeur, mon frère et Marie, pour leur soutien inconditionnel depuis le début de mes études. Leur soutien, moral et financier m'a permis de m'épanouir dans un domaine qui m'a toujours passionné.

Table des matières

1	Introduction	1
2	Revue de littérature	3
2.1	La réplication des fonds de couverture	3
2.1.1	Intérêts et objectifs de la réplication	3
2.1.2	Les différentes approches de réplication	4
2.2	Le modèle de chaînes de Markov cachées	8
3	La réplication de fonds de couverture en étapes	13
3.1	La modélisation des rendements	14
3.1.1	Définition d'un vecteur bivarié d'un mélange de gaussiennes	15
3.1.2	Estimation de la copule $\mathcal{C}_{\mathcal{P},\mathcal{R}}$	16
3.2	Le calcul de la fonction de <i>payoff</i> g	17
3.3	La stratégie optimale de réplication dynamique	18
3.3.1	La couverture localement optimale	19
3.4	Hypothèses de travail, problématiques et enjeux	22
4	Estimation par chaînes de Markov avec changements de régimes	23
4.1	Définition du maximum de vraisemblance	23

4.2	L'algorithme EM et la fonction Q	24
4.3	La théorie derrière un modèle de chaînes de Markov avec changement de régimes	26
4.4	Formule d'estimation basée sur la fonction Q	27
5	Implantation du modèle	32
5.1	Algorithme du modèle dans un cas gaussien	32
5.1.1	Calcul des points de départ	34
5.1.2	Étape E : Calcul des probabilités conditionnelles des régimes	35
5.1.3	Étape M : Estimation des paramètres de notre modèle avec changements de régimes gaussiens	36
5.1.4	Distribution conditionnelle et transformée de Rosenblatt dans le cas bivarié	37
5.2	Test d'adéquation basé sur la transformée de Rosenblatt	38
5.2.1	Cas univarié	38
5.2.2	Cas bivarié	39
5.3	Adaptation de la stratégie optimale de répliation dans le cas d'une estimation par chaînes de Markov cachées	41
5.3.1	Algorithme de la stratégie de couverture optimale	43
5.3.2	Algorithme dans le cas d'une loi normale univariée	45
5.3.3	Algorithme dans le cas d'une loi normale multivariée	46
5.3.4	Validation de l'algorithme par une tarification d'option européenne univariée	47
6	Analyse des résultats	51
6.1	Estimations sur des séries temporelles bivariées	51
6.1.1	Estimations pour la période de 1997 à 2006	53

6.1.2	Estimations pour la période de 1997 à 2001	57
6.1.3	Estimations pour la période de 2002 à 2006	60
6.2	Réplication d'une distribution normale	63
6.2.1	Réplication avec les données de la période entre 1997 et 2006 .	63
6.2.2	Réplication avec les données de la période entre 1997 et 2001 .	65
6.2.3	Réplication avec les données de la période entre 2002 et 2006 .	66
6.2.4	Comparaison des résultats avec notre article de référence . . .	67
6.3	Réplication d'une distribution de Johnson	69
6.3.1	Réplication avec les données de la période entre 1997 et 2006 .	70
6.3.2	Réplication avec les données de la période entre 1997 et 2001 .	71
6.3.3	Réplication avec les données de la période entre 2002 et 2006 .	72
6.4	Synthèse des résultats et problèmes rencontrés	74
7	Conclusion	75

Chapitre 1

Introduction

La forte croissance qu'ont connu les fonds de couvertures ces dernières années a récemment éveillé l'intérêt des chercheurs. Après s'être intéressée à différentes méthodes d'évaluations de leurs performances afin de pouvoir les classer, la recherche s'est rapidement penchée sur des méthodes qui permettraient de répliquer les propriétés des fonds de couvertures.

Dans ce mémoire, nous nous baserons essentiellement sur le travail réalisé par Papageorgiou et al. (2008) concernant une stratégie optimale de couverture afin de répliquer les propriétés des rendements des fonds de couvertures. Plus précisément, nous nous intéresserons à l'estimation des lois sous-jacentes des différents actifs à utiliser dans la stratégie de réplification. Pour cela, nous utiliserons une méthode alternative, le modèle de chaînes de Markov avec changements de régimes, plus communément appelée chaînes de Markov "cachées". Ce modèle, contrairement à celui utilisé par les auteurs (estimation par mélange de gaussiennes), permet de capturer l'auto-corrélation des rendements. Ainsi, en améliorant nos estimations, nous espérons améliorer la stratégie de couverture optimale et la réplification des propriétés des fonds de couvertures.

Dans un premier temps, nous énoncerons les différentes étapes de la réplification ainsi que le modèle de chaînes de Markov avec changements de régimes. Par la suite,

nous détaillerons l'algorithme d'estimation et nous énoncerons comment modifier la stratégie optimale de couverture proposé par Papageorgiou et al. (2008) afin que cette dernière prenne en compte nos nouvelles estimations. Ainsi, nous serons en mesure de tester et comparer nos estimations face à un modèle basé sur les mélanges de Gaussiennes. Enfin, nous tenterons de répliquer certaines distributions données afin de valider notre modèle de chaînes de Markov cachées dans une stratégie de réplication de fonds de couvertures.

Chapitre 2

Revue de littérature

Avant de débiter tout travail de recherche, il est primordial de parcourir les différents articles et publications englobant notre domaine d'étude afin de mieux discerner les frontières à partir desquelles nous travaillons sur quelque chose de nouveau. Nous allons donc couvrir l'ensemble de notre sujet à partir de deux grands axes : dans un premier temps nous parlerons de la répliation de fonds de couvertures puis des chaînes de Markov dites "cachées" et leurs applications en finance.

2.1 La répliation des fonds de couverture

2.1.1 Intérêts et objectifs de la répliation

Les fonds de couvertures sont devenus de plus en plus populaires ces dernières décennies, depuis leur apparition en 1949. Nous pouvons en compter aujourd'hui plus de 8 000, représentant au total plus de 1 000 milliards de dollars investis. Leur popularité est due à différents facteurs. D'abord, ils sont connus pour avoir des rendements élevés en comparaison à d'autres fonds d'investissements standards. En effet, les fonds de couvertures sont contraints à beaucoup moins de régulations au niveau de leur stratégie d'investissement. En plus d'être des produits d'investissement à performance supérieure, ils se sont révélés être un excellent outil de diversification.

En effet, nous avons pu observer que leur performance était assez indépendante de la performance des marchés. Ainsi cette faible corrélation avec les portefeuilles classiques est un avantage nouveau pour les gestionnaires de portefeuilles souhaitant atténuer les fluctuations de leurs portefeuilles en période de récession. Cette nouvelle tendance est surtout apparue quand les fonds de couvertures ont commencé à afficher de moins bonnes performances. Liang (1999) fut l'un des premiers à s'intéresser à ces fonds. Dans son article, il conclut que les frais de gestion sont assez importants. De plus, les rendements étaient positivement corrélés au volume d'actifs sous gestion ainsi qu'à la période où l'investissement est bloqué et non retirable. Le point le plus important fut certainement de mentionner leurs faibles corrélations avec les actifs traditionnels.

Néanmoins, le fait d'être contraints à beaucoup moins de régulations permettait aux gestionnaires de fonds plus de discrétion quant à leur stratégie d'investissement. Le manque de transparence faisait apparaître ces fonds comme des boîtes noires. De plus, d'autres contraintes comme leurs faibles liquidités ou des frais de gestion excessifs ont commencé à susciter, dans le monde académique, beaucoup d'intérêt dans la fin des années 90. La recherche s'est d'abord interrogée sur des mesures de performances plus appropriées à ce type d'investissements dans le but de pouvoir les comparer et les classer entre eux. Néanmoins, l'attention se tourna rapidement sur la création de stratégies tentant de répliquer les fonds de couvertures.

2.1.2 Les différentes approches de répliation

Sharpe (1992) fut l'un des premiers à proposer un modèle qui permettait de répliquer un fond mutuel. Il s'agissait d'un modèle de répliation par facteurs. Dans son article l'auteur tente de répliquer les rendements mensuels de fonds mutuels à l'aide d'un modèle régressif multifactoriel. Néanmoins, cette méthode semblait moins bien

adaptée aux fonds mutuels car il était nécessaire d'identifier les sources de risque afin de bien répliquer les rendements. Or, dans le cas des fonds de couvertures, le manque de transparence rendait cette tâche très difficile. En effet, il était très difficile d'identifier les stratégies des gestionnaires, et donc de savoir dans quels types d'actifs ils transigeaient. De plus, pour les fonds basant leur stratégie sur l'arbitrage, la vente à découvert ou l'utilisation d'un levier très important, il était impossible de répliquer car il n'y avait pas d'actif correspondant disponible.

Ce sont Fung and Hsieh (1997) qui mettront cela de l'avant dans leur article de 1997. En effet, avec le modèle de Sharpe, ils ne parvenaient qu'à expliquer environ 25% des variations dans les rendements mensuels des fonds de couverture. Néanmoins, ils vont proposer une modification à apporter à ce modèle grâce à la découverte d'un fait de taille : les rendements des fonds de couverture ne sont pas linéaires mais plutôt proches de ceux que peuvent procurer les options. Cela viendrait donc expliquer les coefficients de déterminations non significatifs des techniques utilisées par Sharpe à l'aide d'actifs financiers classiques. Ainsi, il paraissait donc évident qu'un modèle linéaire à facteurs n'était pas vraiment approprié à ce type d'investissement. Dans Fung and Hsieh (2001), les auteurs viennent approfondir leur travail initial en utilisant des stratégies optionnelles de type "Lookback Straddle" comme étant la stratégie optimale pour répliquer la répartition des profits de fonds de couvertures. La non linéarité des rendements de fonds de couvertures s'expliquait en partie par le fait qu'ils investissaient dans des produits à rendements non linéaires, telles que les options, mais aussi à cause de l'utilisation de techniques de trading dynamique et l'imputation de frais de gestion asymétrique.

Mitchell and Pulvino (2001), Agarwal and Naik (2004) ainsi que Diez de los Rios and Garcia (2006), se sont également penchés sur cette non linéarité des rendements afin de tenter de trouver des comportements similaires dans les stratégies d'option. Mitchell and Pulvino (2001) vont trouver que les fonds de couverture utilisant l'ar-

bitrage ont un rendement très peu corrélé avec le marché mais très proche d'une stratégie de vente d'option de vente non couverte sur indice. Agarwal and Naik (2004), quant à eux, vont appliquer la méthode par régression multi-factorielle utilisant des options à la monnaie ou hors de la monnaie. Enfin, Diez de los Rios and Garcia (2006) vont finalement améliorer cette technique en utilisant différents degrés de hors jeu afin d'avoir un R^2 plus significatif.

Ce modèle de régression multifactorielle commençait à atteindre ses limites au niveau de son perfectionnement par les chercheurs. Peut être fallait-il se lancer sur une piste bien différente afin d'obtenir de meilleurs résultats. Amin and Kat (2005) vont se pencher essentiellement sur les travaux de Dybvig (1988a,b) et de Glosten and Jagannathan (1994). En effet, ils vont utiliser leurs travaux sur les mesures de performance de fonds de couverture. De plus, ils vont se servir du modèle introduit par Dybvig de tarification de la distribution d'un flux monétaire à échéance, mais cette fois en le modifiant pour le rendre en temps continu. A partir de cela, ils vont arriver à cette conclusion : lors de la répliation, ce qui importe est de répliquer les propriétés statistiques d'un fond, ainsi que son comportement face au portefeuille de l'investisseur initial. Cela fut un tournant pour la recherche en ce qui concerne la répliation de fonds de couvertures. A partir de cela, Amin et Kat ont développé une stratégie de trading dynamique en se servant de contrats à terme. Le but était donc de répliquer la distribution marginale des fonds de couvertures. Néanmoins, à ce stade, répliquer la dépendance entre les fonds et le portefeuille de l'investisseur n'était que très peu efficace.

C'est en retravaillant sur la question et en réutilisant le travail des nombreux auteurs que nous avons cités, que Kat and Palaro (2005) vont décider d'utiliser les copules afin de mieux répliquer la dépendance. Une copule est une fonction de répartition dont les marges sont uniformes. Elles sont maintenant beaucoup utilisées en finance depuis qu'elles ont été introduites par Sklar (1959). Il montra qu'une

copule peut permettre la décomposition d'une distribution jointe à n-dimension en n-distributions marginales. Ce domaine est très vaste et ne concerne que de façon moindre notre futur sujet de recherche, ainsi nous n'élaborerons pas ce sujet.

Ainsi, Kat et Palaro vont utiliser les copules afin de dériver une stratégie de trading dynamique qui génère des rendements avec des propriétés statistiques similaires aux fonds de couvertures. Ainsi, leur technique peut être vue comme la création d'un fond synthétique qui va tenter d'avoir un certain comportement statistique donné. Dans leur article suivant, Kat and Palaro (2006) testent leur méthode hors échantillon afin d'observer les convergences des moments, à savoir, la moyenne, l'écart type, l'asymétrie (skewness) et l'aplatissement (kurtosis) ainsi que la dépendance avec un portefeuille donné. Leurs résultats sont concluants, au niveau de la convergence du 2ème, 3ème et 4ème moment, ainsi que pour la corrélation, et ce, même dans un cas où elle est négative.

Néanmoins, certains problèmes persistent dans leur méthode et beaucoup de choses restent à faire évoluer. Ainsi, Papageorgiou et al. (2008), reprenant le travail de Dybvig (1988b) et de Kat and Palaro (2005), vont remédier à certaines difficultés tout en perfectionnant leurs approches. Un des problèmes majeurs, qui est d'ailleurs mis en évidence par Kat et Palaro eux-mêmes, est le fait que leur mesure d'efficience est calculée dans un monde de Black Scholes (Black and Scholes, 1973). Or, nous savons que les rendements des fonds de couverture ainsi que ceux des contrats à termes utilisés ne sont pas normaux. Ainsi, il y a clairement une contradiction qui aurait nécessairement pour conséquence une mauvaise répliation des moments et de la corrélation. Ainsi, afin de remédier à ce problème, Papageorgiou et al. (2008) proposent d'utiliser une méthodologie basée sur la tarification d'options américaines. Ceci leur permet d'implanter une stratégie de hedging optimale qui a pour but de minimiser l'erreur au carré de répliation. La méthode utilisée est inspirée du travail de Schweizer (1995). Cette méthode permet de déterminer la stratégie à variance op-

timale qui minimisera la perte nette des stratégies, c'est-à-dire la différence entre le contingent et les gains cumulés à partir des transactions de la stratégie de couverture.

Un autre problème majeur concerne les lois mensuelles qu'ils utilisent. Provenant des données des fonds, qui ne sont disponibles que de manières mensuelles, on ne peut que construire des lois mensuelles. Néanmoins, le problème dans l'article de Kat et Palaro est qu'ils assument qu'elles ne sont pas infiniment divisibles. Ainsi, la loi des rendements journaliers peut ne pas être compatible avec la loi des rendements mensuels. Ceci pose problème car il est essentiel de disposer des lois journalières afin de déterminer la stratégie de couverture optimale à mettre en place pour répliquer notre fond. Papageorgiou et al. (2008) proposent alors d'utiliser plusieurs mélanges de lois normales puis de résoudre pour la loi appropriée des rendements mensuels. En effet, une somme de lois gaussiennes reste une loi gaussienne. Enfin, concernant l'estimation de la copule, ils recommandent de ne pas utiliser la méthode IFM mais plutôt une méthode basée sur les rangs normalisés et la transformée de Rosenblatt énoncée dans l'article de Genest et al. (2009).

2.2 Le modèle de chaînes de Markov cachées

Initialement introduits et étudiés à la fin des années 60 et début 70, les modèles de chaînes de Markov cachées (HMM) sont devenus plus populaires à partir des années 80. Poritz (1982), Rabiner and Juang (1985), et Rabiner (1989), furent les premiers à parler de chaînes de Markov cachées dans leurs travaux qui n'étaient d'ailleurs nullement reliés à l'économétrie ou à la finance mais plutôt à la reconnaissance de signaux sonores.

Ce modèle utilise une chaîne de Markov afin de modéliser les changements de caractéristiques statistiques qui peuvent seulement être probabilistiquement manifestés

à travers les observations. Un HMM est donc un procédé doublement stochastique dans lequel se trouve une chaîne de Markov non observable (d'où le terme "hidden") pour laquelle chaque état est associé à une fonction de densité (distribution de probabilités). Une chaîne de Markov est définie par une matrice de transition entre états. La fonction de probabilité peut, elle, être une représentation soit paramétrique, soit non paramétrique. L'observation est donc une fonction probable de l'état. L'échantillon, lui, ne peut être observable qu'à travers un ensemble de procédés stochastiques qui produit la séquence des observations.

Dans Rabiner and Juang (1985), les auteurs se concentrent essentiellement sur des densités de mélanges de gaussiennes autorégressives. Poritz (1982) a été le premier à montrer comment l'idée de l'analyse de prédiction linéaire peut être incorporée aux HMM. Néanmoins, dans son modèle, il a seulement considéré un seul mélange de gaussienne autorégressive par états de la chaîne de Markov. Dans leur article, Juang et Rabiner, quant à eux, tentent d'incorporer un mélange de densités gaussiennes autorégressives. Enfin, Rabiner (1989) va tenter de mieux expliquer la théorie des HMM, du concept le plus simple (avec chaîne de Markov discrète) à des modèles plus sophistiqués (modèle avec des densités continues).

Les chaînes de Markov avec changement de régression ont été initialement introduites par Goldfeld and Quandt (1982a). Dans leur article, ils traitent une problématique où, pour un intervalle de temps, ils utilisent une certaine régression, et pour un autre, ils utilisent une autre régression. Dans cet article, ils approfondissent la possibilité de pouvoir avoir plusieurs changements. Il s'agit donc de changements d'états de manière récursive. Il introduit l'utilisation de chaîne de Markov, avec une matrice de transition pour ces deux états avec leurs probabilités associées. Cosslett and Lee (1985) travailleront sur le même sujet mais seront reconnus comme étant les premiers à calculer correctement la fonction de vraisemblance.

Hamilton (1989) reprendra ces articles afin d'étudier la possibilité que les rendements du PIB soient sujets à des sauts auto corrélés en temps discret. Pour cela il utilise un algorithme utilisant un filtre itératif non linéaire. Ce filtre permet également d'estimer les paramètres d'une population grâce à la méthode du maximum de vraisemblance et propose les bases afin de prévoir les valeurs futures des séries. Il remarque que de tels sauts sont observables mais qu'ils le sont surtout entre une phase de récession et de croissance. Il en conclut que le passage d'une phase de croissance à une phase de récession est associé à une chute de 3% dans la valeur présente du futur PIB. De plus, ce changement annonce similairement une chute de 3% dans la prévision à long terme du PIB.

Par la suite, il va tenter d'améliorer son approche dans Hamilton (1990), en introduisant à un algorithme qui utilise le maximum de vraisemblance afin d'obtenir des paramètres pour des processus sujets à des sauts (en temps discret) et ce dans les paramètres de l'auto régression. Les sauts sont eux-mêmes modélisés comme étant la résultante d'un processus de Markov en temps discret. En effet, nous avons pu remarquer que les mouvements importants dans les prix des actifs sont souvent dus à des événements identifiables. Un exemple peut être la décision par la Réserve Fédérale Américaine de changer son taux directeur. Il commence donc à parler clairement de séries temporelles qui sont sujettes à des changements de régimes.

Dans Hamilton (1990), l'auteur utilise l'algorithme "Expectation Maximisation (EM)" de Dempster et al. (1977) car celui-ci est numériquement très robuste. Il s'agit d'un algorithme applicable pour trouver les estimateurs par maximum de vraisemblance d'une série de données incomplètes. Il permet également de trouver les estimateurs quand la fonction de vraisemblance peut être simplifiée en assumant l'existence de valeurs pour des paramètres qui seraient manquants (pour le cas des HMM). Son écrit propose une caractérisation analytique de la dérivée de la fonction de log-vraisemblance de l'échantillon.

L'article de Bilmes (1998) nous sera très utile car il explique de façon très détaillée le problème d'estimation de paramètres pour le maximum de vraisemblance et comment l'algorithme EM peut être utilisé pour trouver sa solution. Il explique clairement les algorithmes afin d'estimer les paramètres pour un mélange de densité gaussienne ainsi que pour les paramètres d'un HMM. De plus, il améliore le travail de Dempster et al. (1977) pour assurer la convergence, rendant cette méthode encore plus robuste.

Nous nous intéresserons aussi au travail de Qian and Titterton (1991), car ceux-ci mettent également en évidence les difficultés rencontrées lors de l'estimation des paramètres pour les HMM et proposent certaines méthodes pour y remédier. Plusieurs de ces méthodes reposent sur la simulation Monte Carlo. Ils proposent comme alternative à l'algorithme EM proposé par Hamilton (1990), l'algorithme SRM basé sur les pseudo fonctions de vraisemblances, pour des modèles de mélanges et des modèles avec structure de dépendance.

Ce n'est qu'en 1996, que Hamilton (1996) parle de changements de séries temporelles sous forme de chaînes de Markov, ce qui peut donc clairement être assimilé à un HMM. Le but de cet article est de vérifier si l'approche donne des résultats proches des données réelles. Pour cela il utilise plusieurs tests. Il utilise une technique se servant du multiplicateur de Lagrange, initialement utilisé par Breusch (1974), Godfrey and Wickens (1981) et Engle (1982a,b). Il utilise également l'approche générale pour le test de spécification développée par Newey (1985), Tauschen (1985) et White (1985). Il fait aussi une analyse Monte Carlo de ces tests. Il en conclut que le biais n'est pas très significatif pour des échantillons au dessus de 100. Néanmoins, en ce qui concerne les échantillons plus petits que 50, le biais est assez sévère.

De nombreux articles en économétrie évoquent alors les HMM. Nous citerons ce-

lui de Asbrink and T. (1998) qui est assez significatif pour notre sujet d'étude. Dans cet article, les auteurs montrent que lorsqu'il faut modéliser une série de rendements présentant une dépendance d'ordre supérieur, le HMM permet de bien les modéliser. Ils utilisent à cet effet une série de rendements du S&P 500. Néanmoins, ils précisent qu'il est cependant plus difficile de modéliser la fonction d'auto corrélation faiblement diminutive pour les rendements absolus. Leur étude diffère des autres utilisant le HMM car le nombre de régimes n'est pas préalablement défini. Également, Giudici and Vandekerkhove (2000) soulignent que les HMM sont une classe de modèles pour modéliser des phénomènes aléatoires faiblement dépendants. Un modèle à chaînes de Markov cachées peut être vu comme une extension des modèles de mélanges. Le but de cet article est de tester le rapport de vraisemblance pour les modèles de HMM. Sous des conditions appropriées, on montre que la théorie asymptotique standard des tests du rapport de vraisemblance est valide.

Enfin, Remillard et al. (2010) montrent comment modéliser des rendements d'actifs à l'aide de chaînes de Markov cachées. Ils énoncent des propriétés de prévisions une fois le modèle estimé ainsi qu'un test d'adéquation permettant de déterminer le nombre optimal d'états pour la matrice Markovienne, et ce pour des séries temporelles multivariées. Enfin, ils décrivent comment adapter ce modèle dans une stratégie de couverture. Cet article nous sera donc très utile afin d'implanter notre stratégie de couverture dans le but de répliquer une distribution donnée, et ce, tout en conservant les propriétés de notre estimation par chaînes de Markov cachées.

Ainsi, nous retiendrons que les chaînes de Markov avec changements de régimes sont bien adaptées à de petits échantillons qui présentent de l'autocorrélation. Cette méthode nous permettrait, dans le cadre de la réplification de fonds de couverture, de mieux estimer nos lois concernant les rendements journaliers, pour le portefeuille et les actifs transigés dans la stratégie de réplification. Une fois l'autocorrélation captée dans notre estimation, nous pourrions répliquer notre fond en la prenant en compte.

Chapitre 3

La répliation de fonds de couverture en étapes

Rappelons l'idée de base que nous définissons par répliation : à partir de deux titres risqués S_1 et S_2 , il est possible de «reproduire» à la fois le rendement d'un troisième titre risqué S_3 , mais également sa dépendance avec le titre S_1 . L'important n'est pas alors de reproduire la valeur du titre S_3 mais ses propriétés statistiques. Il existe, pour une période de temps T fixée, une fonction g telle que la loi de probabilité jointe de $S_1(T)$ et $g(S_1(T), S_2(T))$ soit la même que la loi jointe de $S_1(T)$ et $S_3(T)$. Ainsi, il faut déterminer une méthode pour générer le flux $g(S_1(T), S_2(T))$ sans investir dans le titre S_3 . Nous utiliserons donc un portefeuille composé d'un titre non risqué S_0 et des titres S_1 et S_2 rebalancés périodiquement à $t = 1, \dots, T - 1$.

Le problème de répliation peut se diviser en trois grandes étapes. Dans un premier temps, il s'agit de modéliser au mieux les fonctions de distribution ainsi que les fonctions des copules. Ensuite, vient se poser le problème du calcul de la fonction de *payoff*. Enfin, il faut trouver une stratégie de trading dynamique qui nous procurera la meilleure approximation possible de la fonction de *payoff*. Ainsi, nous allons tenter de détailler ces étapes en se servant de la méthodologie de Papageorgiou et al. (2008).

Précisons la notation :

- \mathcal{P} représente le portefeuille initial de l'investisseur.

- \mathcal{R} représente l'actif de réserve, utilisé dans la replication.
- \mathcal{I} représente le fond de couverture à répliquer.

- $F_{\mathcal{P}}$ représente la fonction de distribution marginale de \mathcal{P}_T .
- $F_{\mathcal{R}}$ représente la fonction de distribution marginale de \mathcal{R}_T .
- $F_{\mathcal{I}}$ représente la fonction de distribution marginale de \mathcal{I}_T .

- $R_{0,T}^{\mathcal{P}}$ représente les rendements de \mathcal{P}_T .
- $R_{0,T}^{\mathcal{R}}$ représente les rendements de \mathcal{R}_T .
- $R_{0,T}^{\mathcal{I}}$ représente les rendements de \mathcal{I}_T .

3.1 La modélisation des rendements

Le défi majeur de cette étape est la modélisation bivariée des rendements journaliers de la réserve \mathcal{R} et du portefeuille \mathcal{P} . Nous avons effectivement besoin des lois journalières car la stratégie de réplification sera elle-même journalière. Néanmoins, il faut nous assurer qu'avec cette stratégie nous retombons sur la distribution mensuelle souhaitée correspondant à celle du fond de couverture. Le fait d'utiliser un mélange de gaussiennes pour modéliser notre distribution bivariée nous procure deux caractéristiques intéressantes : une distribution bivariée est infiniment divisible et un mélange de gaussiennes permet de capter de hauts niveaux d'asymétrie (skewness). Ainsi, nous devrions obtenir notre loi mensuelle et aurons plus de facilité à répliquer l'asymétrie importante que connaissent les fonds de couvertures. Il faut également estimer la loi mensuelle univariée et mensuelle du fond de couverture. Pour cela, Papageorgiou et al. (2008) utilisent un test d'adéquation que nous citerons plus tard. Concernant le choix des copules, nous nous baserons essentiellement sur l'article de Genest et al. (2009).

3.1.1 Définition d'un vecteur bivarié d'un mélange de gaussiennes

Un vecteur aléatoire bivarié X est un mélange de gaussiennes avec m régimes et de paramètres $(\pi_k)_{k=1}^m$, $(\mu_k)_{k=1}^m$ et $(A_k)_{k=1}^m$, si sa densité est donnée par

$$f(x) = \sum_{k=1}^m \pi_k \phi_2(x; \mu_k, A_k)$$

où $\phi_2(x; \mu, A) = \frac{e^{-\frac{1}{2}(x-\mu)^\top A^{-1}(x-\mu)}}{2\pi\sigma_1\sigma_2(1-\rho^2)^{1/2}}$ est la densité d'un vecteur de Gaussiennes bivariées

de moyenne $\mu = (\mu_1, \mu_2)^\top$ et comme matrice de covariance $A = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$.

Sa fonction de répartition est

$$F(x_1, x_2) = \sum_{k=1}^m \pi_k \Phi_2\left(\frac{x_1 - \mu_{k1}}{\sigma_{k1}}, \frac{x_2 - \mu_{k2}}{\sigma_{k2}}; \rho_k\right),$$

où $\Phi_2(\cdot, \cdot; \rho)$ est la fonction de répartition bivariée standard gaussienne de corrélation ρ .

Quelques propriétés intéressantes concernant les mélanges de gaussiennes bivariées

Une propriété qui est très importante que Papageorgiou et al. (2008) démontrent dans leur article, est le fait qu'une somme de mélanges gaussiennes indépendantes est toujours un mélange de gaussiennes.

Considérons la distribution conditionnelle d'un mélange de gaussiennes bivariées $X = (X^{(1)}, X^{(2)})$. Posons $\beta_k = \rho_k \frac{\sigma_{k2}}{\sigma_{k1}}$ et $\alpha_k = \mu_{k2} - \beta_k \mu_{k1}$, $k = 1, \dots, m$. Ainsi, il est facile de vérifier que la distribution conditionnelle de $X^{(2)}$ sachant $X^{(1)} = x_1$ est un

mélange de gaussiennes de paramètres $\{\tilde{\pi}_k(x_1)\}_{k=1}^m$, $\{\tilde{\mu}_k(x_1)\}_{k=1}^m$, $\{\tilde{\sigma}_k^2\}_{k=1}^m$, où

$$\tilde{\pi}_k(x_1) = \frac{\pi_k \phi(x_1; \mu_{k1}, \sigma_{k1}^2)}{\sum_{j=1}^m \pi_j \phi(x_1; \mu_{j1}, \sigma_{j1}^2)} \quad (3.1)$$

et

$$\tilde{\mu}_k(x_1) = \alpha_k + \beta_k x_1, \quad \tilde{\sigma}_k^2 = \sigma_k^2 (1 - \rho_k^2). \quad (3.2)$$

3.1.2 Estimation de la copule $\mathcal{C}_{\mathcal{P}, \mathcal{R}}$

Il n'y a pas de restriction concernant le choix de la copule $\mathcal{C}_{\mathcal{P}, \mathcal{R}}$, entre les rendements mensuels du fond de couvertures et les portefeuilles de l'investisseur. Supposons que nous ayons les rendements mensuels $(Y_1, Z_1), \dots, (Y_n, Z_n)$ qui appartiennent à une famille de copule C_θ . Pour estimer θ , nous utiliserons la méthode IFM. Néanmoins, nous ne l'utiliserons pas comme paramètre de la copule qui repose sur la distribution marginale estimée. En effet, une mauvaise spécification des distributions marginales va créer un biais dans le choix de la copule. Pour que notre estimation soit plus robuste, il est préférable d'utiliser les rangs normalisés, i.e. si R_{i1} représente le rang de Y_i dans Y_1, \dots, Y_n et si R_{i2} représente le rang de Z_i dans Z_1, \dots, Z_n , avec $R_{ij} = 1$ pour les plus petites observations, nous posons

$$U_i = \frac{R_{i1}}{n+1}, \quad V_i = \frac{R_{i2}}{n+1}, \quad i = 1, \dots, n.$$

Pour estimer θ , nous maximisons la fonction de pseudo-log-vraisemblance

$$\sum_{i=1}^n \log c_\theta(U_i, V_i),$$

comme suggéré dans l'article de Genest et al. (1995). Par exemple, si la copule est la copule gaussienne de corrélation ρ , l'estimateur de pseudo-vraisemblance pour ρ

défini le coefficient de Van Der Waerden comme étant la corrélation entre les paires $\{\Phi^{-1}(U_i), \Phi^{-1}(V_i); i = 1, \dots, n\}$. Pour d'autres familles qui peuvent être indexées par le Tau de Kentall, e.g., les familles de Clayton, Frank et Gumbel, nous pouvons estimer le paramètre par inversion du Tau de Kendall de l'échantillon. Voir, e.g., Genest et al. (2006).

Enfin, pour le test d'adéquation, nous utiliserons une statistique de type Cramér-von Mises pour la copule empirique ou pour la transformée de Rosenblatt. Ce dernier pourrait être le meilleur choix sachant que $\frac{\partial}{\partial u} \mathcal{C}_{1,3}(u, v)$ avait besoin d'être calculé pour l'évaluation de la fonction de *payoff*. Ces tests sont décrits dans Genest et al. (2009).

3.2 Le calcul de la fonction de *payoff* g

Après avoir estimé les distributions nécessaires ainsi que les fonction des copules, nous devons calculer la fonction de payoff g . On la calcule telle que la loi de probabilité jointe de $\mathcal{P}(T)$ et $g(\mathcal{P}(T), \mathcal{R}(T))$ soit la même que la loi jointe de $\mathcal{P}(T)$ et $\mathcal{I}(T)$. On cherche alors la fonction de profit la moins chère qui sera capable de transformer la distribution jointe de l'actif de réserve et du portefeuille de l'investisseur en distribution jointe du fond et du portefeuille de l'investisseur. Cette dernière, telle que mentionnée par Papageorgiou et al. (2008), a été trouvée par Kat and Palaro (2005) :

$$g(x, y) = Q \{x, P (R_{0,T}^{\mathcal{R}} \leq y | R_{0,T}^{\mathcal{P}} = x)\},$$

où R représente les rendements, $Q(x, \alpha)$ est le quantile d'ordre α de la loi conditionnelle de $R_{0,T}^{\mathcal{I}}$ sachant $R_{0,T}^{\mathcal{P}} = x$, i.e., pour tout $\alpha \in (0, 1)$, $q(x, \alpha)$ qui satisfait

$$P \{R_{0,T}^{\mathcal{I}} \leq Q(x, \alpha) | R_{0,T}^{\mathcal{P}} = x\} = \alpha.$$

Utilisant les propriétés des copules, e.g. Nelsen (1999), les distributions condition-

nelles peuvent être exprimées en terme de marges et des copules associées.

$$P(R_{0,T}^{\mathcal{R}} \leq y | R_{0,T}^{\mathcal{P}} = x) = \frac{\partial}{\partial u} \mathcal{C}_{\mathcal{P},\mathcal{R}}(u, v) \Big|_{u=F_{\mathcal{P}}(x), v=F_{\mathcal{R}}(y)}.$$

Notons que $\frac{\partial}{\partial u} \mathcal{C}_{\mathcal{P},\mathcal{R}}(u, v) = P\{F_2(R_{0,T}^{\mathcal{R}}) \leq v | F_{\mathcal{P}}(R_{0,T}^{\mathcal{P}}) = u\}$. De plus, si $\mathcal{Q}(u, \alpha)$ est le quantile d'ordre α de la fonction de distribution $\frac{\partial}{\partial u} \mathcal{C}_{\mathcal{P},\mathcal{I}}(u, v)$, nous obtenons donc

$$Q(x, \alpha) = F_{\mathcal{R}}^{-1} \circ \mathcal{Q}(F_{\mathcal{P}}(x), \alpha).$$

Dans cette méthodologie, étant donné que les rendements mensuels $(R_{0,T}^{\mathcal{P}}, R_{0,T}^{\mathcal{R}})$ sont modélisés par un mélange de gaussiennes de paramètres $(\pi_k)_{k=1}^m$, $(\mu_k)_{k=1}^m$ et $(A_k)_{k=1}^m$, les distributions conditionnelles peuvent être exprimées ainsi

$$P(R_{0,T}^{\mathcal{R}} \leq y | R_{0,T}^{\mathcal{P}} = x) = \sum_{k=1}^m \tilde{\pi}_k(x) \phi\{y; \tilde{\mu}_k(x), \tilde{\sigma}^2\}$$

où $\tilde{\pi}_k(x)$, $\tilde{\mu}_k(x)$ et $\tilde{\sigma}^2$ sont donnés par (3.1) et (3.2).

3.3 La stratégie optimale de réplication dynamique

Papageorgiou et al. (2008) s'inspirent du travail de Schweizer (1995) pour répliquer la fonction g . Pour cela, nous sélectionnons le portefeuille (V_0, φ) qui minimise l'espérance de l'erreur de réplication

$$E[\beta_T^2 \{V_T(V_0, \varphi) - C_T\}^2],$$

où β_T est le facteur d'actualisation et $C_T = 100 \exp\left\{g\left(R_{0,T}^{(1)}, R_{0,T}^{(2)}\right)\right\}$ est le flux monétaire à échéance.

Afin d'y arriver, Papageorgiou et al. (2008) développent une extension des ré-

sultats de Schweizer (1995)). Il est important de noter qu'il n'y a pas d'évaluation risque neutre dans leur évaluation et que tous les calculs sont fait sous mesure de probabilité objective.

Ainsi, ils montrent que si la réplication dynamique fonctionne, c'est-à-dire, $V_T = C_T$, alors le rendement sur investissement peut être décomposé ainsi :

$$\log(V_T/V_0) = \log(100/V_0) + g\left(R_{0,T}^{(1)}, R_{0,T}^{(2)}\right).$$

Ainsi, tel que proposé par Kat and Palaro (2005), nous pouvons interpréter $\alpha = \log(100/V_0)$ comme une mesure de performance.

3.3.1 La couverture localement optimale

Supposons que (Ω, P, \mathcal{F}) est un espace probabilisable de filtration $\mathbb{F} = \{\mathcal{F}_0, \dots, \mathcal{F}_T\}$ sous lequel les processus stochastiques sont définis. Ainsi, dans notre cas, supposons que le processus de prix S_t est bi-dimensionnel, c'est-à-dire $S_t = (S_t^{(1)}, S_t^{(2)})$.

Définissons β_t comme étant le facteur d'actualisation : β_t est la valeur, à la période 0, à investir dans l'actif sans risque pour qu'elle aie une valeur de 1\$ à la période t . Par définition, $\beta_0 = 1$. Nous assumons également que le processus β est prévisible, c'est-à-dire que β_t est \mathcal{F}_{t-1} -mesurable pour tout $t = 1, \dots, T$.

Une stratégie dynamique de réplication peut être décrite par une valeur initiale V_0 et une séquence de vecteurs de poids $\varphi = (\varphi_t)_{t=0}^T$, où pour tout $j = 1$ et $j = 2$, $\varphi_t^{(j)}$ indique le nombre d'actifs (S^1, S^2) investis durant la période $(t-1, t]$. Puisque φ_t peut dépendre seulement des valeurs S_0, \dots, S_{t-1} , le processus stochastique φ_t est supposé prévisible. Initialement, $\varphi_0 = \varphi_1$, et la valeur initiale du portefeuille est V_0 . Ainsi, le montant investi initialement dans l'actif sans risque est :

$$V_0 - \sum_{j=1}^d \varphi_1^{(j)} S_0^{(j)} = V_0 - \varphi_1^\top S_0.$$

Puisque la stratégie de couverture doit être auto-financée, pour tout $t = 1, \dots, T$,

$$\beta_t V_t(V_0, \varphi) - \beta_{t-1} V_{t-1}(V_0, \varphi) = \varphi_t^\top (\beta_t S_t - \beta_{t-1} S_{t-1}). \quad (3.3)$$

En utilisant la condition d'auto-financement, (3.3),

$$\beta_T V_T = \beta_T V_T(V_0, \varphi) = V_0 + \sum_{t=1}^T \varphi_t^\top (\beta_t S_t - \beta_{t-1} S_{t-1}). \quad (3.4)$$

Le problème de stratégie de réplication pour un flux monétaire C est donc équivalent à trouver un stratégie (V_0, φ) qui fait en sorte que l'erreur

$$G_T(V_0, \varphi) = \beta_T V_T(V_0, \varphi) - \beta_T C \quad (3.5)$$

soit la plus petite possible. Papageorgiou et al. (2008) choisissent l'erreur espérée au carré comme mesure de qualité pour la réplication. Il est également naturel de supposer que les prix (S_t^1, S_t^2) sont de carrés intégrables. Ils assument également que la stratégie de couverture φ satisfait la même propriété, surtout que pour tout $t = 1, \dots, T$, $\varphi_t^\top (\beta_t S_t - \beta_{t-1} S_{t-1})$ a des moments d'ordre supérieurs finis.

A des fins de simplicité, définissons :

$$\Delta_t = S_t - E(S_t | \mathcal{F}_{t-1}), \quad t = 1, \dots, T.$$

Sous la condition suivante, la matrice de covariance conditionnelle Σ_t of Δ_t existe et est définie comme telle :

$$\Sigma_t = E \{ \Delta_t \Delta_t^\top | \mathcal{F}_{t-1} \}, \quad 1 \leq t \leq T.$$

Papageorgiou et al. (2008)) traitent le cas d -dimensionnel même si pour des fins de réplifications, nous n'avons besoin que du cas bi-dimensionnel. Ainsi, ils supposent que Σ_t est inversible pour tout $t = 1, \dots, T$. Si Σ_t n'est pas inversible pour certains t , il existerait une $\varphi_t \in \mathcal{F}_{t-1}$ qui fait en sorte que $\varphi_t^\top S_t = \varphi_t^\top E(S_t | \mathcal{F}_{t-1})$, c'est-à-dire

que $\varphi_t^\top S_t$ est prévisible. Leurs hypothèses peuvent être interprétées comme disant que la dimension véritable des actifs est d . Ainsi, nous pouvons énoncer les résultats majeurs de leur article :

Theorem 1 *Supposons que Σ_t est inversible pour tout $t = 1, \dots, T$. Ainsi, le risque $E\{G^2(V_0, \varphi)\}$ est localement minimisé en choisissant récursivement $\varphi_T, \dots, \varphi_1$ satisfaisant*

$$\varphi_t = (\Sigma_t)^{-1} E(\{S_t - E(S_t|\mathcal{F}_{t-1})\} C_t | \mathcal{F}_{t-1}), \quad t = T, \dots, 1, \quad (3.6)$$

où C_T, \dots, C_0 sont définis récursivement en autant que $C_T = C$ et

$$\beta_{t-1} C_{t-1} = \beta_t E(C_t | \mathcal{F}_{t-1}) - \varphi_t^\top E(\beta_t S_t - \beta_{t-1} S_{t-1} | \mathcal{F}_{t-1}), \quad (3.7)$$

pour $t = T, \dots, 1$.

De plus, la valeur optimale de V_0 est C_0 , et

$$E(G^2) = \sum_{t=1}^T E(\beta_t^2 G_t^2),$$

où $G_t = \varphi_t^\top \{S_t - E(S_t|\mathcal{F}_{t-1})\} - \{C_t - E(C_t|\mathcal{F}_{t-1})\}$, $1 \leq t \leq T$.

Maintenant que nous avons trouvé la stratégie optimale de couverture, d'après le critère de la moyenne des erreurs au carré, nous pouvons nous demander quel est le lien entre le prix donné par C_0 et le prix suggéré par la méthode de la mesure martingale. La réponse étant donnée par le résultat suivant.

Corollary 1 *Pour tout $t = 1 \dots, T$, définissons*

$$U_t = 1 - \Delta_t^\top (\Sigma_t)^{-1} E(S_t - \beta_{t-1} S_{t-1} / \beta_t | \mathcal{F}_{t-1}). \quad (3.8)$$

De plus, définissons $M_0 = 1$ et $M_t = U_t M_{t-1}$, $1 \leq k \leq n$. Ainsi $(M_t, \mathcal{F}_t)_{t=0}^T$ est

une martingale et

$$\beta_{t-1}C_{t-1} = E(\beta_t C_t U_t | \mathcal{F}_{t-1}).$$

En particulier $\beta C_t M_t$ est une martingale et $C_0 = E(\beta_T C_T M_T | \mathcal{F}_0)$. De plus, $E(\beta_t S_t U_t | \mathcal{F}_{t-1}) = \beta_{t-1} S_{t-1}$, donc $\beta_t S_t M_t$ est une martingale.¹

3.4 Hypothèses de travail, problématiques et enjeux

Nous avons vu que Papageorgiou et al. (2008) amélioreraient la méthode de Kat and Palaro (2005) au niveau de l'estimation des lois en utilisant des mélanges de gaussiennes. En effet, nous avons vu que ces dernières ont la particularité d'être infiniment divisibles. Néanmoins, cette méthode suppose que les rendements soient indépendants les uns des autres, ainsi elle néglige la possibilité qu'il y ait présence d'autocorrélation dans les séries de rendements journaliers des différents actifs. Néanmoins, si on utilise la loi journalière sur une période d'un mois, nous pouvons constater une différence avec la loi mensuelle, alors que celles-ci devraient être par définition semblables. La différence provient en partie de l'erreur d'estimation des lois journalières et mensuelles, mais nous pensons que la présence d'autocorrélation dans nos séries de rendements est la source majeure de cette différence.

Ainsi, nous utiliserons le modèle de chaînes de Markov avec changements de régimes afin d'introduire de la dépendance entre les rendements de nos séries. Notre travail portera donc essentiellement sur l'estimation de la distribution de la loi jointe des rendements $(R_{0,T}^P, R_{0,T}^P)$ utilisés par la suite dans la fonction de *payoff* g et dans la stratégie optimale de couverture.

1. Quand le marché est complet, il existe une unique mesure martingale Q et le risque associé à la stratégie optimale est nul. Ainsi M_t est positive, et nous avons donc dans notre cas une représentation explicite de la densité Q cohérent avec P .

Chapitre 4

Estimation par chaînes de Markov avec changements de régimes

Afin de mieux comprendre la théorie derrière un modèle de chaînes de Markov avec changement de régimes, nous commencerons par clarifier ce qu'est un maximum de vraisemblance. Ensuite, nous pourrions parler de l'algorithme EM tel qu'énoncé dans Bilmes (1998) qui est utilisé dans l'estimation des paramètres d'un mélange de gaussiennes ainsi que dans notre modèle.

4.1 Définition du maximum de vraisemblance

Le maximum de vraisemblance est grandement utilisé en ingénierie financière afin d'estimer les paramètres de nombreux modèles.

Si nous avons une fonction de densité $p(x|\Theta)$ ayant pour paramètres Θ , pouvant être, dans le cas d'une gaussienne, sa moyenne et sa variance. Si nous avons également une série de données de taille N , provenant, par supposition, de cette distribution, c'est-à-dire $\mathcal{X} = \{x_1, \dots, x_N\}$. Ainsi, nous assumons que ce vecteur de données est indépendant et identiquement distribué avec la distribution p . La densité

de l'échantillon est donc

$$p(\mathcal{X}|\Theta) = \prod_{i=1}^N p(x_i|\Theta) = \mathcal{L}(\Theta|\mathcal{X}).$$

Cette fonction $\mathcal{L}(\Theta|\mathcal{X})$ est appelée la fonction de vraisemblance. Dans le cas du maximum de vraisemblance, notre but est de trouver les paramètres Θ qui maximisent \mathcal{L} . Ainsi nous cherchons Θ^* où

$$\Theta^* = \operatorname{argmax}_{\Theta} \mathcal{L}(\Theta|\mathcal{X})$$

Nous avons souvent tendance à maximiser $\log(\mathcal{L}(\Theta|\mathcal{X}))$ car cela est plus facile analytiquement. Ainsi, nous pouvons voir que cette estimation peut être plus ou moins facile dépendamment de la forme de la fonction de densité $p(x|\Theta)$. Dans le cas d'une simple gaussienne, les estimateurs sont trouvés facilement de manière analytique. Néanmoins, pour notre modèle, nous allons avoir besoin de techniques plus élaborées.

4.2 L'algorithme EM et la fonction Q

Cet algorithme est une méthode générale pour trouver le maximum de vraisemblance de paramètres d'une distribution provenant d'un échantillon comportant des données manquantes. Celles-ci peuvent être de plusieurs formes. Dans notre cas, nous pouvons simplifier la fonction de vraisemblance en assumant l'existence ainsi que la valeur de paramètres manquants ou plutôt «cachés».

Ainsi, nous supposons que l'échantillon \mathcal{X} est observé et généré par une distribution quelconque. Nous appelons \mathcal{X} l'échantillon incomplet. Nous supposons également qu'il existe un échantillon complet $\mathcal{Z} = (\mathcal{X}, \mathcal{Y})$ et une fonction de densité jointe

$$p(z|\Theta) = p(x, y|\Theta) = p(y|x, \Theta)p(x|\Theta)$$

Cette densité fait surface à partir d'une fonction de densité marginale $p(x|\Theta)$, de variables cachées et de valeurs de paramètres «devinés».

Nous pouvons donc définir la nouvelle fonction de vraisemblance comme étant $\mathcal{L}(\Theta|\mathcal{Z}) = \mathcal{L}(\Theta|\mathcal{X}, \mathcal{Y}) = p(\mathcal{X}, \mathcal{Y}|\Theta)$ comprenant toutes les données. Nous l'appellerons la fonction de vraisemblance complète. Cette fonction est aléatoire car l'information \mathcal{Y} est inconnue, aléatoire et supposément gouvernée par une distribution sous-jacente.

L'algorithme EM trouve d'abord la valeur espérée de la fonction de log-vraisemblance complète $\log p(\mathcal{X}, \mathcal{Y}|\Theta)$ en respectant les données manquantes \mathcal{Y} étant donné les données observées \mathcal{X} et la valeur estimée actuelle des paramètres. Ainsi nous définissons la fonction Q comme étant

$$Q(\Theta, \Theta^{(i-1)}) = E [\log p(\mathcal{X}, \mathcal{Y}|\Theta) | \mathcal{X}, \Theta^{(i-1)}]$$

où $\Theta^{(i-1)}$ sont les valeurs estimées actuelles des paramètres que nous utilisons pour évaluer et Θ sont les nouveaux paramètres que nous optimisons pour augmenter Q . Il est important de comprendre que \mathcal{X} et $\Theta^{(i-1)}$ sont des constantes, Θ est une variable normale que nous cherchons à ajuster et \mathcal{Y} est une variable aléatoire gouvernée par une distribution $f(y|\mathcal{X}, \Theta^{(i-1)})$. Ainsi, nous pouvons à nouveau écrire l'équation précédente comme étant

$$E [\log p(\mathcal{X}, \mathcal{Y}|\Theta) | \mathcal{X}, \Theta^{(i-1)}] = \int_{\mathbf{y} \in \Upsilon} \log p(\mathcal{X}, y|\Theta) f(y|\mathcal{X}, \Theta^{(i-1)}) d\mathbf{y}.$$

Notons que $f(y|\mathcal{X}, \Theta^{(i-1)})$ est la distribution marginale des données inobservables et est dépendante des observations \mathcal{X} et des paramètres actuels. De plus, Υ est l'espace des valeurs que peut prendre y .

Ainsi, cet algorithme se fait en deux étapes. La première, l'étape E, est l'évaluation de cette espérance. L'argument Θ correspond aux paramètres qui vont être optimisés dans le but de maximiser la vraisemblance. Le second argument $\Theta^{(i-1)}$ correspond aux paramètres utilisés pour évaluer cet espérance. La deuxième, l'étape M de l'algorithme, est la maximisation de cette espérance précédemment calculée. Ainsi nous trouvons

$$\Theta^{(i)} = \operatorname{argmax}_{\Theta} Q(\Theta, \Theta^{(i-1)}).$$

Ces deux étapes sont répétées autant que nécessaire. Chaque itération est garantie d'augmenter le log-vraisemblance et l'algorithme est garanti de converger vers un maximum local de la fonction de vraisemblance.

4.3 La théorie derrière un modèle de chaînes de Markov avec changement de régimes

Nous avons maintenant les outils nécessaires pour pouvoir nous pencher sur le modèle nous concernant. Ainsi, un modèle à chaînes de Markov dites «cachées» est un modèle de probabilités jointes d'une collection de variables aléatoires $\{O_1, \dots, O_T, Q_1, \dots, Q_T\}$. Les variables O_t sont soit des observations continues soit discrètes et les variables Q_t sont dites «cachées» et discrètes. Dans un modèle à chaînes de Markov cachées, il y a deux hypothèses conditionnelles émises à propos de ces variables qui font en sorte que son algorithme est facilement implémentable. Ces dernières sont 1), la t^{me} variable cachée, sachant la variable cachée au temps $(t-1)$ est indépendante des variables précédentes, donc

$$P(Q_t | Q_{t-1}, O_{t-1}, \dots, Q_1, O_1) = P(Q_t | Q_{t-1})$$

et 2), l'observation au temps t , sachant l'observation cachée au temps t , est indépendante des autres variables donc

$$P(O_t \mid Q_T, O_T, Q_{T-1}, O_{T-1}, \dots, Q_{t+1}, O_{t+1}, Q_t, Q_{t-1}, O_{t-1}, \dots, Q_1, O_1) = P(O_t \mid Q_t)$$

Dans cette section, nous expliquerons comment, à partir de l'algorithme EM, nous trouverons les estimateurs du maximum de vraisemblance des paramètres d'un modèle de chaînes de Markov cachées à partir de vecteurs d'observations caractéristiques. Cet algorithme est connu sous le nom d'algorithme de Baum-Welch.

Q_t est une variable aléatoire discrète avec N valeurs possibles $\{1, \dots, N\}$. Nous supposons que la chaîne de Markov cachée sous-jacente définie par $P(Q_t \mid Q_{t-1})$ est indépendante du temps t . Ainsi, nous pouvons présenter $P(Q_t \mid Q_{t-1})$ comme une matrice de transition stochastique et indépendante du temps $A = \{a_{i,j}\} = p(Q_t = j \mid Q_{t-1} = i)$. Concernant le cas $t = 1$, il est décrit par la distribution de l'état initial, $\pi_i = p(Q_1 = i)$. Nous disons que nous sommes dans l'état j au temps t si $Q_t = j$. Une séquence particulière d'états est décrite par $q = (q_1, \dots, q_T)$ où $q_t \in \{1, \dots, N\}$ est l'état au temps t .

Une séquence particulière d'observations O est décrite par $O = (O_1 = o_1, \dots, O_T = o_T)$. La probabilité d'un vecteur d'observation au temps t à l'état j est décrite par $b_j(o_t) = p(O_t = o_t \mid Q_t = j)$. L'ensemble des paramètres pour toutes les distributions observées est présenté par $B = \{b_j(\cdot)\}$.

4.4 Formule d'estimation basée sur la fonction Q

Considérons $O = (o_1, \dots, o_T)$ comme étant les données observées et $q = (q_1, \dots, q_T)$ comme étant les données inobservables ou «cachées». La fonction de vraisemblance

comprenant des données manquantes est donnée par $P(O | \Delta)$ alors que la fonction de vraisemblance complète est $P(O, q | \Delta)$. La fonction Q est donc

$$Q(\Delta, \Delta') = \sum_{q \in \mathcal{Q}} \log P(O, q | \Delta) P(O, q | \Delta')$$

où Δ' sont nos estimés initiaux des paramètres, pouvant correspondre aux estimés précédents ou à une simple supposition, et \mathcal{Q} est l'espace de toutes les séquences d'états possibles de longueur T . Sachant une séquence d'état q , représenter $P(O, q | \Delta')$ est simple.

$$P(O, q | \Delta) = \pi_{q_0} \prod_{t=1}^T a_{q_{t-1}q_t} b_{q_t}(o_t)$$

La fonction Q devient donc

$$\begin{aligned} Q(\Delta, \Delta') &= \sum_{q \in \mathcal{Q}} \log \pi_{q_0} P(O, q | \Delta') + \sum_{q \in \mathcal{Q}} \left(\sum_{t=1}^T \log a_{q_{t-1}q_t} \right) P(O, q | \Delta') \\ &\quad + \sum_{q \in \mathcal{Q}} \left(\sum_{t=1}^T \log b_{q_t}(o_t) \right) P(O, q | \Delta') \end{aligned} \quad (4.1)$$

Etant donné que les paramètres que nous souhaitons optimiser sont maintenant séparés en trois de manière indépendante dans la somme, nous pouvons optimiser chaque terme de manière individuelle.

La première partie de l'équation (4.1) devient

$$\sum_{q \in \mathcal{Q}} \log \pi_{q_0} P(O, q | \Delta') = \sum_{i=1}^N \log \pi_i p(O, q_0 = i | \Delta')$$

sachant qu'en sélectionnant tous les $q \in \mathcal{Q}$, nous ne faisons que sélectionner à répétition les valeurs de q_0 , la partie de droite est simplement l'expression marginale pour le temps $t = 0$. En ajoutant le multiplicateur de Lagrange γ , en utilisant la

contrainte $\sum_i \pi_i = 1$ et en annulant la dérivé, nous obtenons

$$\frac{\partial}{\partial \pi_i} \left(\sum_{i=1}^N \log \pi_i p(O, q_0 = i | \Delta') + \gamma \left(\sum_{i=1}^N \pi_i - 1 \right) \right) = 0$$

En prenant la dérivé, en sommant les i pour avoir γ et en résolvant pour avoir les π_i , nous obtenons

$$\pi_i = \frac{P(O, q_0 = i | \Delta')}{P(O | \Delta')}$$

La deuxième partie de l'équation (4.1) devient

$$\sum_{q \in \mathcal{Q}} \left(\sum_{t=1}^T \log a_{q_{t-1} q_t} \right) p(O, q | \Delta') = \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^T \log a_{ij} P(O, q_{t-1} = i, q_t = j | \Delta')$$

A cause de ce terme, nous regardons pour chaque temps t toutes les transitions de i à j et lui allouant sa probabilité correspondante. La partie de droite est simplement la somme des marginales jointes pour les temps $t - 1$ et t . De façon similaire, nous pouvons utiliser le multiplicateur de Lagrange avec la contrainte $\sum_{j=1}^N a_{ij} = 1$ pour avoir

$$a_{ij} = \frac{\sum_{t=1}^T P(O, q_{t-1} = i, q_t = j | \Delta')}{\sum_{t=1}^T P(O, q_{t-1} = i | \Delta')}$$

Le troisième terme de l'équation (4.1) devient

$$\sum_{q \in \mathcal{Q}} \left(\sum_{t=1}^T \log b_{q_t}(o_t) \right) P(O, q | \Delta') = \sum_{i=1}^N \sum_{t=1}^T \log b_i(o_t) p(O, q_t = i | \Delta')$$

Car pour ce terme nous observons, pour chaque temps t , l'émission d'un état et nous pondérons chaque émission par sa probabilité respective. La partie de droite est simplement la somme des marginales au temps t .

Pour des distributions discrètes, nous pouvons également utiliser un multiplica-

teur de Lagrange mais cette fois-ci avec la contrainte $\sum_{j=1}^L b_i(j) = 1$. Seules les observations qui sont égales à v_k contribuent à la k^{me} valeurs de probabilité, donc nous obtenons

$$b_i(k) = \frac{\sum_{t=1}^T P(O, q_t = i \mid \Delta') \delta_{o_t, v_k}}{\sum_{t=1}^T P(O, q_t = i \mid \Delta')}$$

Pour des mélanges de gaussiennes, la forme de la fonction Q est un peu différente, les variables «cachées» doivent inclure non seulement la séquence des états, mais aussi une variable qui indique les caractéristiques du mélange pour chaque état et à chaque temps. Ainsi, nous pouvons écrire Q comme cela

$$Q(\Delta, \Delta') = \sum_{q \in \mathcal{Q}} \sum_{m \in \mathcal{M}} \log P(O, q, m \mid \Delta) P(O, q, m \mid \Delta')$$

où m est un vecteur $m = \{m_{q_11}, m_{q_22}, \dots, m_{q_T T}\}$ qui indique les caractéristiques du mélange pour chaque temps. Si nous faisons de même avec l'équation (4.1), le premier et le deuxième terme resteraient inchangés car les paramètres sont indépendants de m qui est marginalisé par la somme. Le troisième terme de l'équation (4.1) deviendrait

$$\sum_{q \in \mathcal{Q}} \sum_{m \in \mathcal{M}} \left(\sum_{t=1}^T \log b_{q_t}(o_t, m_{q_t t}) \right) P(O, q, m \mid \Delta') = \sum_{i=1}^N \sum_{l=1}^M \sum_{t=1}^T \log \left(c_{il} b_{il}(o_t) \right) p(O, q_t = i, m_{q_t t} = l \mid \Delta')$$

Bilmes (1995) montre que nous pouvons optimiser cette dernière équation afin d'obtenir

$$c_{il} = \frac{\sum_{t=1}^T P(q_t = i, m_{q_t t} = l \mid O, \Delta')}{\sum_{t=1}^T T \sum_{l=1}^M P(q_t = i, m_{q_t t} = l \mid O, \Delta')}$$

$$\mu_{il} = \frac{\sum_{t=1}^T o_t P(q_t = i, m_{q_t t} = l \mid O, \Delta')}{\sum_{t=1}^T P(q_t = i, m_{q_t t} = l \mid O, \Delta')}$$

et

$$\Sigma_{il} = \frac{\sum_{t=1}^T (o_t - \mu_{il})(o_t - \mu_{il})^T P(q_t = i, m_{q_t t} = l \mid O, \Delta')}{\sum_{t=1}^T P(q_t = i, m_{q_t t} = l \mid O, \Delta')}$$

Si nous souhaitons un modèle avec changements de régimes gaussiens au lieu de mélange, nous avons simplement à définir $l=1$.

Chapitre 5

Implantation du modèle

Dans ce chapitre, nous détaillerons clairement l'algorithme afin d'estimer notre modèle de chaînes de Markov avec changements de régimes gaussiens pour une série chronologique univariée et bivariée. Nous énoncerons également un test d'adéquation, basé sur la transformée de Rosenblatt, afin de pouvoir choisir le nombre de régimes optimal dans notre estimation. Enfin, nous détaillerons l'algorithme de la stratégie de couverture optimale, énoncé par Remillard et al. (2010), afin de pouvoir répliquer une distribution donnée tout en conservant l'autocorrélation capturée lors de notre estimation par chaînes de Markov avec changements de régimes gaussiens.

5.1 Algorithme du modèle dans un cas gaussien

Nous énoncerons l'algorithme tel qu'énoncé dans l'article de Remillard et al. (2010).

Afin d'exprimer la densité jointe $f_{1:k}$ de X_1, \dots, X_k , définissons, pour tout $i \in \{1, \dots, r\}$ et tout $k \geq 1$,

$$\gamma_k(i) = E \left\{ \mathbb{I}(\tau_k = i) \prod_{t=1}^k g_t(\tau_t) \right\}$$

Si nous définissons ν comme étant la loi stationnaire, notons que $\gamma_0(j) = \nu_j$ semble

être un choix logique. Ainsi

$$Z_k = \sum_{i=1}^r \gamma_k(i) = E \left\{ \prod_{t=1}^k g_t(\tau_t) \right\} = f_{1:k}(x_1, \dots, x_k).$$

Grâce au travail de Del Moral (2004), nous pouvons trouver une formule récursive pour les γ_k . Plus précisément, grâce au travail de Remillard et al. (2010), nous avons, pour tout $k \geq 1$ et tout $i \in \{1, \dots, r\}$

$$\gamma_1(i) = \nu_i g_1(i),$$

$$\gamma_k(i) = g_k(i) \sum_{\alpha=1}^r \gamma_{k-1}(\alpha) Q_{\alpha i}, k \geq 2.$$

Ensuite, pour tout $k \geq 2$, la densité conditionnelle de X_1, \dots, X_{k-1} notée par $f_{k:1}$, peut être exprimée comme un mélange.

$$\begin{aligned} f_{k:1}(x_k | x_1, \dots, x_{k-1}) &= f_{1:k}(x_1, \dots, x_k) / f_{1:k-1}(x_1, \dots, x_{k-1}) \\ &= \frac{\sum_{i=1}^r \sum_{j=1}^r \gamma_{k-1}(i) Q_{ij} f_j(x_k)}{\sum_{i=1}^r \gamma_{k-1}(i)} \\ &= \sum_{j=1}^r W_{j,k-1} f_j(x_k), \end{aligned}$$

où

$$W_{j,k-1} = \frac{\sum_{i=1}^r \gamma_{k-1}(i) Q_{ij}}{\sum_{i=1}^r \gamma_{k-1}(i)}, j \in \{1, \dots, r\}$$

Étant donné que $\gamma_0(j) = \nu_j$, $W_{j,0} = \nu_j$, pour tout $j \in \{1, \dots, r\}$.

Ensuite, provenant de la formule de Kallianpur-Sriberl (Rémillard, 2007, Chapitre 6) nous avons pour tout $k \geq 1$ et pour tout $i \in \{1, \dots, r\}$,

$$\eta_k(i) = P(\tau_k = i | X_1, \dots, X_k) = \frac{\gamma_k(i)}{Z_k}$$

5.1.1 Calcul des points de départ

Les points de départ sont cruciaux dans ce genre d'algorithme afin qu'il converge bien vers un maximum absolu et non un maximum local. Ainsi, nous utiliserons comme point de départ les paramètres d'une estimation de mélange pour le même nombre de régimes. Nous détaillerons ainsi cet algorithme pour un échantillon bivarié, tel qu'énoncé dans Papageorgiou et al. (2008).

Soit y_1, \dots, y_n un échantillon aléatoire d'un mélange de gaussiennes bivariées de paramètres $\pi = (\pi_k)_{k=1}^m$, $\mu = (\mu_k)_{k=1}^m$ et $A = (A_k)_{k=1}^m$. Commençons avec un estimateurs initial $\theta^{(0)}$. Prenant un estimateur $\theta^{(\ell)} = (\pi^{(\ell)}, \mu^{(\ell)}, A^{(\ell)})$ de paramètre $\theta = (\pi, \mu, A)$, posons

$$\pi_k(y_i, \theta^{(\ell)}) = \frac{\pi_k^{(\ell)} \phi_2(y_i; \mu_k^{(\ell)}, A_k^{(\ell)})}{\sum_{j=1}^m \pi_j^{(\ell)} \phi_2(y_i; \mu_j^{(\ell)}, A_j^{(\ell)})}, \quad i = 1, \dots, n,$$

et définissons le nouvel estimateur $\theta^{(\ell+1)} = (\pi^{(\ell+1)}, \mu^{(\ell+1)}, A^{(\ell+1)})$.

$$\pi_k^{(\ell+1)} = \frac{1}{n} \sum_{i=1}^n \pi_k(y_i, \theta^{(\ell)}),$$

$$\mu_k^{(\ell+1)} = \frac{1}{n} \sum_{i=1}^n y_i \pi_k(y_i, \theta^{(\ell)}) / \pi_k^{(\ell+1)},$$

et

$$A_k^{(\ell+1)} = \frac{1}{n} \sum_{i=1}^n (y_i - \mu_k^{(\ell+1)}) (y_i - \mu_k^{(\ell+1)})^\top \pi_k(y_i, \theta^{(\ell)}) / \pi_k^{(\ell+1)},$$

pour $k = 1, \dots, m$. Au fur et à mesure que ℓ augmente, les nombres $\{\pi_k(y_i, \theta^{(\ell)}); k = 1, \dots, i = 1, \dots, n\}$ se stabilisent et les estimateurs convergent. En ce qui concerne la matrice de transition Q , nous prendrons comme point de départ la matrice indépendante tirée des pondérations du mélange.

5.1.2 Étape E : Calcul des probabilités conditionnelles des régimes

Il est primordial de déterminer les probabilités

$$\lambda_k(i) = P(\tau_k = i \mid X_1, \dots, X_n),$$

ainsi que,

$$\Lambda_k(i, j) = P(\tau_k = i, \tau_{k+1} = j \mid X_1, \dots, X_n)$$

pour tout $1 \leq k \leq n$ et tout $i, j \in \{1, \dots, r\}$

Pour cela, définissons

$$\tilde{\gamma}_n(i) = 1$$

$$\tilde{\gamma}_k(i) = \sum_{\beta=1}^r \tilde{\gamma}_{k+1}(\beta) Q_{i\beta g_{k+1}(\beta)}, \quad 1 \leq k \leq n-1$$

Ainsi pour tout $i, j \in \{1, \dots, r\}$, nous pouvons vérifier que

$$\lambda_k(i) = \frac{\gamma_k(i) \tilde{\gamma}_k(i)}{\sum_{\alpha=1}^r \gamma_k(\alpha) \tilde{\gamma}_k(\alpha)} \quad k = 1, \dots, n$$

$$\Lambda_k(i, j) = \frac{Q_{ij} \gamma_k(i) \tilde{\gamma}_{k+1}(j) g_{k+1}(\beta)}{\sum_{\alpha=1}^r \sum_{\beta=1}^r Q_{\alpha\beta} \gamma_k(\alpha) \tilde{\gamma}_{k+1}(\beta) g_{k+1}(\beta)} \quad k = 1, \dots, n-1,$$

et $\Lambda_n(i, j) = \lambda_n(i) Q_{ij}$

Dans leur article article, Remillard et al. (2010) montrent que ces formules sont correctes pour tout $1 \leq k \leq n-1$ car

$$\sum_{j=1}^r \Lambda_k(i, j) = \sum_{j=1}^r \frac{Q_{ij} \gamma_k(i) \tilde{\gamma}_{k+1}(j) g_{k+1}(j)}{\sum_{\alpha=1}^r \sum_{\beta=1}^r Q_{\alpha\beta} \gamma_k(\alpha) \tilde{\gamma}_{k+1}(\beta) g_{k+1}(\beta)} = \frac{\gamma_k(i) \tilde{\gamma}_k(i)}{\sum_{\alpha=1}^r \gamma_k(\alpha) \tilde{\gamma}_k(\alpha)} = \lambda_k(i)$$

Utilisant la définition de $\tilde{\gamma}_k$, $\sum_{j=1}^r \Lambda_n(i, j) = \sum_{j=1}^r \lambda_n(i) Q_{ij} = \lambda_n(i)$.

De la même façon, pour tout $1 \leq k \leq n-1$,

$$\sum_{i=1}^n \Lambda_k(i, j) = \sum_{i=1}^r \frac{Q_{ij} \gamma_k(i) \tilde{\gamma}_{k+1}(j) g_{k+1}(j)}{\sum_{\alpha=1}^r \sum_{\beta=1}^r Q_{\alpha\beta} \gamma_k(\alpha) \tilde{\gamma}_{k+1}(\beta) g_{k+1}(\beta)} = \frac{\gamma_{k+1}(i) \tilde{\gamma}_{k+1}(j)}{\sum_{\alpha=1}^r \gamma_{k+1}(\alpha) \tilde{\gamma}_{k+1}(\alpha)} = \lambda_{k+1}(i)$$

utilisant la définition de γ_{k+1} .

5.1.3 Étape M : Estimation des paramètres de notre modèle avec changements de régimes gaussiens

Dans notre cas, les densités f_1, \dots, f_r sont celles d'une distribution gaussienne de moyenne $(\mu_k)_{k=1}^r$ et de matrice de covariance $(A_k)_{k=1}^r$. L'étape de maximisation (M-Step) consiste en un renouvellement récursif des paramètres $(\nu_k)_{k=1}^r, (\mu_k)_{k=1}^r, (A_k)_{k=1}^r$ et de la matrice de transition Q en définissant, pour tout $i, j \in \{1, \dots, r\}$,

$$\nu'_i = \sum_{k=1}^n \lambda_k(i) / n$$

$$\mu'_i = \sum_{k=1}^n x_k w_k(i)$$

$$A'_i = \sum_{k=1}^n (x_k - \mu'_i)(x_k - \mu'_i)^T w_k(i)$$

$$Q'_{ij} = \sum_{k=1}^n \Lambda_k(i, j) \text{Big} / \sum_{k=1}^n \lambda_k(i) = \frac{1}{n} \sum_{k=1}^n \Lambda_k(i, j) \text{Big} / \nu'_i$$

$$\text{où } w_k(i) = \lambda_k(i) \text{Big} / \sum_{l=1}^n \lambda_l(i)$$

Dans Remillard et al. (2010), on précise que ν' ne correspond pas exactement à la loi stationnaire de Q' mais qu'elle s'en rapproche plus n est grand.

En effet, pour tout $j \in \{1, \dots, r\}$,

$$\sum_{i=1}^r \nu'_i Q'_{ij} = \frac{1}{n} \sum_{k=1}^n \sum_{i=1}^r \Lambda_k(i, j) = \frac{1}{n} \sum_{k=2}^{n+1} \delta_k(j) = \nu'_j + \frac{\delta_{n+1}(j) - \delta_1(j)}{n} \neq \nu'_j$$

Mais,

$$\max_{1 \leq j \leq r} \left| \sum_{i=1}^r \nu'_i Q'_{ij} - \nu'_j \right| \leq 1/n$$

5.1.4 Distribution conditionnelle et transformée de Rosenblatt dans le cas bivarié

Nous allons montrer comment trouver la transformée de Rosenblatt Ψ_k correspondant à la densité (5.1).

$$\Psi_k^{(1)}(z_1) = \Psi_k^{(j)}(x_1, \dots, x_{k-1}, z_1) = \sum_{\alpha=1}^r W_{\alpha, k-1} F_{\alpha, 1}(z_1)$$

et

$$\Psi_k^{(2)}(z_1, z_2) = \Psi_k^{(2)}(x_1, \dots, x_{k-1}, z_1, z_2) = \frac{\sum_{\alpha=1}^r W_{\alpha, k-1} f_{\alpha, 1}(z_1) F_{\alpha, 2}(z_2)}{\sum_{\alpha=1}^r W_{\alpha, k-1} f_{\alpha, 1}(z_1)}$$

où f et F sont les fonctions de densité et de répartition d'une loi normale bivariée de moyenne μ_k et de matrice de covariance $\Sigma_k = \begin{pmatrix} v_k^{(1)} & \rho_k \sqrt{v_k^{(1)} v_k^{(2)}} \\ \rho_k \sqrt{v_k^{(1)} v_k^{(2)}} & v_k^{(2)} \end{pmatrix}$.

Ainsi, $f_{k,2}$ est la fonction de densité d'une distribution gaussienne de moyenne $\mu_k^{(2)} + \beta_k(y_k^{(1)} - \mu_k^{(1)})$ et de variance $v_k^{(2)}(1 - \rho_k^{(2)})$ où $\beta_k = \rho_k \sqrt{v_k^{(2)}/v_k^{(1)}}$.

Ainsi, nous définissons que $U_1 = \Psi_1(X_1)$ et $U_2 = \Psi_2(X_1, X_2)$ sont indépendants et uniformément distribués sur $[0, 1]^2$.

5.2 Test d'adéquation basé sur la transformée de Rosenblatt

Afin de déterminer le nombre optimal de régimes pour l'estimation de notre modèle de chaînes de Markov avec changements de régimes gaussiens, nous nous baserons sur l'article de Durbin (1973) qui est aussi utilisé par Papageorgiou et al. (2008). Nous avons vu précédemment comment trouver la transformée de Rosenblatt, nous détaillerons ici comment l'utiliser dans un test d'adéquation.

5.2.1 Cas univarié

Soit X_1, \dots, X_n un échantillon de taille n provenant d'une distribution continue F dans R . Supposons que les hypothèses à tester soient

$$\mathcal{H}_0 : F \in \mathcal{F} = \{F_\theta; \theta \in \Theta\} \quad \text{vs} \quad \mathcal{H}_1 : F \notin \mathcal{F}$$

Par exemple, la famille paramétrique \mathcal{F} pourrait être la famille des mélanges de gaussiennes univariées avec m régimes.

Le test statistique proposé est basé sur Durbin (1973). Soit $\theta_n = T_n(X_1, \dots, X_n)$ un estimateur de θ , selon Genest et al. (2009) et définissons

$$D_n(u) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(U_i \leq u), \quad u \in [0, 1],$$

où $U_i = F_{\theta_n}(X_i)$, $i = 1, \dots, n$. Pour tester \mathcal{H}_0 face à \mathcal{H}_1 , nous pouvons utiliser le test statistique de Cramér-von Mises

$$\begin{aligned} S_n &= n \int_0^1 \{D_n(u) - u\}^2 du \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \left\{ \frac{U_i^2 + U_j^2 - 2 \max(U_i, U_j)}{2} + \frac{1}{3} \right\}. \end{aligned}$$

Comme les U_i sont «presque distribués uniformément dans $[0, 1]$ » sous l'hypothèse nulle, de grandes valeurs de S_n devraient entraîner le rejet de l'hypothèse nulle. Néanmoins, la distribution de S_n dépend en général du paramètre θ . Pour calculer la P -value de S_n , nous pouvons utiliser la technique de rééchantillonnage paramétrique (parametric bootstrap) comme nous la décrivons ici.

- a) Calculer θ_n et S_n .
- b) Pour des N grand (par exemple 1000), répéter les étapes suivantes pour chaque $k \in \{1, \dots, N\}$:
 - (i) Générer un échantillon aléatoire $X_{1,k}, \dots, X_{n,k}$ provenant de la distribution F_{θ_n} .
 - (ii) Calculer

$$\begin{aligned}\theta_{n,k} &= T_n(X_{1,k}, \dots, X_{n,k}), \\ U_{i,k} &= F_{\theta_{n,k}}(X_{i,k}), \quad i = 1, \dots, n, \\ S_{n,k} &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \left\{ \frac{U_{i,k}^2 + U_{j,k}^2 - 2 \max(U_{i,k}, U_{j,k})}{2} + \frac{1}{3} \right\}.\end{aligned}$$

Une P -value approximative pour le test basé sur la statistique S_n de Cramér–von Mises est donc donnée par

$$\frac{1}{N} \sum_{k=1}^N \mathbb{I}(S_{n,k} > S_n).$$

5.2.2 Cas bivarié

Nous utiliserons ici le travail de Rémillard et al. (2010).

Soit $(X_1, Y_1), \dots, (X_n, Y_n)$ un échantillon de taille n provenant d'une distribution continue F dans \mathbb{R}^2 . Supposons que les hypothèses à tester soient

$$\mathcal{H}_0 : F \in \mathcal{F} = \{F_\theta; \theta \in \Theta\} \quad \text{vs} \quad \mathcal{H}_1 : F \notin \mathcal{F}$$

Par exemple, la famille paramétrique \mathcal{F} peut être des familles de mélanges de gaussiennes bivariées avec m régimes. Notons G_θ la fonction de distribution de X_i et H_θ la fonction de distribution conditionnelle de Y_i sachant X_i , i.e., $H_\theta(x, y) = P(Y_i \leq y | X_i = x)$.

Le test statistique proposé est basé sur l'article de Durbin (1973) et la transformée de Rosenblatt.

Supposons que $\theta_n = T_n(X_1, Y_1, \dots, X_n, Y_n)$ est un estimateur régulier de θ , selon Genest et al. (2009), nous posons

$$D_n(u, v) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(U_i \leq u, V_i \leq v), \quad u, v \in [0, 1],$$

où $U_i = G_{\theta_n}(X_i)$, $V_i = H_{\theta_n}(X_i, Y_i)$, $i = 1, \dots, n$. pour tester \mathcal{H}_0 face à \mathcal{H}_1 , nous pouvons utiliser la statistique de Cramér-von Mises

$$\begin{aligned} S_n &= n \int_0^1 \int_0^1 \{D_n(u, v) - uv\}^2 dudv \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \left[\frac{1}{9} - \frac{1}{4}(1 - U_i^2)(1 - V_i^2) - \frac{1}{4}(1 - U_j^2)(1 - V_j^2) \right. \\ &\quad \left. + \{1 - \max(U_i, U_j)\} \{1 - \max(V_i, V_j)\} \right]. \end{aligned}$$

Étant donné que la paire (U_i, V_i) est «presque distribuée uniformément sur $[0, 1]^2$ » sous l'hypothèse nulle, de grandes valeurs de S_n devraient entraîner le rejet de l'hypothèse nulle. Néanmoins, la distribution limitée de S_n dépend du paramètre non-connu θ . Pour calculer la P -value du S_n , nous pouvons utiliser l'approche de rééchantillonnage paramétrique décrite ici.

- a) Calculons θ_n et S_n .
- b) Pour des entiers N grands (par exemple 1000), répéter les étapes suivantes pour chaque $k \in \{1, \dots, N\}$:
 - (i) Générer un échantillon aléatoire $(X_{1,k}, Y_{1,k}), \dots, (X_{n,k}, Y_{n,k})$ provenant de la distribution de F_{θ_n} .

(ii) Calculer

$$\begin{aligned} \theta_{n,k}^* &= T_n(X_{1,k}, Y_{1,k}, \dots, X_{n,k}, Y_{n,k}), \\ U_{i,k} &= G_{\theta_{n,k}}(X_{i,k}), \quad V_{i,k} = H_{\theta_{n,k}}(X_{i,k}, Y_{i,k}), \quad i = 1, \dots, n \\ S_{n,k} &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \left[\frac{1}{9} - \frac{1}{4}(1 - U_{i,k}^2)(1 - V_{i,k}^2) - \frac{1}{4}(1 - U_{j,k}^2)(1 - V_{j,k}^2) \right. \\ &\quad \left. + \{1 - \max(U_{i,k}, U_{j,k})\} \{1 - \max(V_{i,k}, V_{j,k})\} \right]. \end{aligned}$$

Une P -value approximative pour le test basé sur la statistique de Cramér–von Mises S_n est donnée par

$$\frac{1}{N} \sum_{k=1}^N \mathbb{I}(S_{n,k} > S_n).$$

Nous avons donc vu comment implanter le modèle d'estimation par chaînes de Markov avec changements de gaussiennes pour des séries univariées et bivariées de série temporelle. De plus, nous avons vu comment choisir le nombre optimal de regimes avec le test d'adéquation de Rémillard et al. (2010). Nous avons également vu une alternative à ce test pour des séries univariées avec le test de log de vraisemblance. Nous allons maintenant montrer comment adapter l'algorithme de couverture optimale de Papageorgiou et al. (2008) que nous avons énoncé dans le chapitre 3.

5.3 Adaptation de la stratégie optimale de réplification dans le cas d'une estimation par chaînes de Markov cachées

Afin de conserver la dépendance sérielle captée lors de notre estimation, il est nécessaire de modifier l'approche standard de la stratégie optimale de couverture utilisée par Papageorgiou et al. (2008) où l'estimation se faisait avec des mélanges de gaussiennes.

Pour cela nous allons utiliser le travail de Remillard et al. (2010) concernant les propriétés de prévisions sérielles d'un modèle à chaînes de Markov cachées. Ainsi, les auteurs montrent que pour une fonction de payoff g , il est facile d'estimer son espérance au temps $k + 1$ sachant la filtration au temps k . Ainsi

$$E\{g(X_{k+1})|\mathcal{F}_k\} = \sum_{i=1}^r E\{g(X_{k+1})|X_k, \tau_k = i\}\eta_k(i) = \sum_{i=1}^r \sum_{j=1}^r \eta_k(i)Q_{ij} \int g(x)f_j(x)dx \quad (5.1)$$

La formule (5.1) entraîne que la loi conditionnelle de X_{k+1} sachant X_1, \dots, X_k à comme densité

$$f_{k+1:k}(x) = \sum_{i=1}^r \sum_{j=1}^r \eta_k(i)Q_{ij}f_j(x), \quad (5.2)$$

qui est un mélange avec les mêmes densités $(f_j)_{j=1}^r$ mais seulement avec des poids différents qui sont donné par $\sum_{i=1}^r \eta_k(i)Q_{ij}$ pour le régime $j, j \in \{1, \dots, r\}$. Ainsi, après avoir estimé notre modèle jusqu'au temps k , la prédiction concernant X_{k+1} est

$$\sum_{i=1}^r \sum_{j=1}^r \eta_k(i)\mu_j Q_{ij}.$$

Des intervalles de confiance peuvent être calculés en utilisant les quantiles des densités $f_{k+1:k}$ donnés par l'équation (5.2).

En utilisant les propriétés des chaînes de Markov, nous pouvons faire ces prévisions pour n'importe quel pas de temps $l \geq 1$. Ainsi,

$$E\{g(X_{k+l})|\mathcal{F}_k\} = \sum_{i=1}^r \sum_{j=1}^r \eta_k(i)(Q^l)_{ij} \int g(x)f_j(x)dx.$$

La loi conditionnelle de X_{k+l} sachant X_1, \dots, X_k a pour densité

$$f_{k+l:k}(x) = \sum_{i=1}^r \sum_{j=1}^r \eta_k(i)(Q^l)_{ij}f_j(x). \quad (5.3)$$

qui est un mélange avec les mêmes densités $(f_j)_{j=1}^r$ et poids $\sum_{i=1}^r \eta_k(i)(Q^l)_{ij}$ pour les régimes $j, j \in \{1, \dots, r\}$.

Enfin, Remillard et al. (2010) nous montrent que si la chaîne de Markov de matrice de transition Q est ergodique, alors la loi conditionnelle de X_{k+l} sachant X_1, \dots, X_k converge vers la distribution stationnaire

$$f(x) = \sum_{i=1}^r \nu_i f_i(x).$$

Ainsi, dans le cas de longues prévisions, la variable est supposée avoir un comportement indépendant du passé.

5.3.1 Algorithme de la stratégie de couverture optimale

Rappelons que d'après Papageorgiou et al. (2008), le risque $E(\beta_n^2 G^2 | \mathcal{F}_0)$ est minimisé en choisissant de manière récursive $\phi_n, \phi_{n-1}, \dots, \phi_1$ satisfaisant

$$\varphi_t = (\Sigma_t)^{-1} E(\{S_t - E(S_t | \mathcal{F}_{t-1})\} C_t | \mathcal{F}_{t-1}), \quad t = T, \dots, 1, \quad (5.4)$$

où C_T, \dots, C_0 sont définis récursivement en autant que $C_T = C$ et

$$\beta_{t-1} C_{t-1} = \beta_t E(C_t | \mathcal{F}_{t-1}) - \varphi_t^\top E(\beta_t S_t - \beta_{t-1} S_{t-1} | \mathcal{F}_{t-1}), \quad (5.5)$$

pour $t = T, \dots, 1$.

De plus, la valeur optimale de V_0 est C_0 , et

$$E(G^2) = \sum_{t=1}^T E(\beta_t^2 G_t^2), \quad (5.6)$$

où $G_t = \varphi_t^\top \{S_t - E(S_t | \mathcal{F}_{t-1})\} - \{C_t - E(C_t | \mathcal{F}_{t-1})\}$, $1 \leq t \leq T$.

On pose aussi

$$U_t = 1 - \Delta_t^\top (\Sigma_t)^{-1} E(S_t - \beta_{t-1} S_{t-1} / \beta_t | \mathcal{F}_{t-1}). \quad (5.7)$$

On obtient alors

$$\beta_{t-1}C_{t-1} = E(\beta_t C_t U_t | \mathcal{F}_{t-1}).$$

Le problème principal avec ces expressions est le choix de la filtration $F = \{\mathcal{F}_k; k = 0, \dots, n\}$. En effet, à cause de l'expression (5.1), les espérances conditionnelles qui respectent $\mathcal{F}_k = \sigma\{S_1, \dots, S_k\}$, apparaissant dans les équations (5.6) et (5.7), vont dépendre de la trajectoire (S_1, \dots, S_k) car $\eta_k(i)$ en dépend aussi. Cela est donc impossible à générer. Une alternative est d'utiliser $\mathcal{F}_k = \sigma\{(S_1, \tau_1), \dots, (S_k, \tau_k)\}$. Parce que (S_k, τ_k) est un processus de Markov, V_k va seulement dépendre de (S_k, τ_k) si C est une fonction de S_n . Sachant que τ_k ne peut que prendre des valeurs dans $\{1, \dots, r\}$, nous n'avons qu'à produire chaque V_k et chaque ϕ_k pour les points (s, i) , où s fait parti de la grille des S_k et $1 \leq i < +r$. Quand s n'est pas sur la grille, il suffit d'interpoler $V_k(\cdot, i)$ au point s . Néanmoins, les régimes n'étant pas observables, ils doivent être prédits. Ainsi, considérons n_0 valeurs passées de S , jusqu'au temps présent $t = 0$ et estimons τ_t par

$$\hat{\tau}_t = \operatorname{argmax}_i \eta_{n_0+t}(i), t = 0, \dots, T - 1$$

Cette dernière équation signifie que le régime prédit est le régime ayant la plus grande probabilité sachant l'information sur les prix jusqu'au temps $n_0 + t$. Ainsi, pour trouver les poids ϕ_{t+1} pour les périodes $(t, t+1]$, il suffit d'utiliser $\phi_{t+1}(S_t, \hat{\tau}_t)$, $t \in \{0, \dots, T - 1\}$. Dans les cas d'un mélange, il est important de noter que lorsque $Q_{ij} = \pi_j$, V_k et ϕ_{k+1} ne dépendent pas de τ_k . Ainsi, cette algorithme inclue le cas des mélanges même si la prédiction de τ_k n'y est pas nécessaire.

Enfin, dans Remillard et al. (2010), les auteurs précisent que l'utilisation de simulation par échantillonnage stratifié, contrairement à l'échantillonnage simple, semble meilleure en terme de variabilité. Cela consiste à générer $N_j \approx Q_{ij}N$ observations

$X_{k,j}$ de la densité f_j pour tout $j \in \{1, \dots, r\}$, et ainsi calculer la moyenne pondérée

$$\sum_{j=1}^r Q_{ij} \frac{1}{N_j} \sum_{k=1}^{N_j} g(X_{k,j}, j).$$

5.3.2 Algorithme dans le cas d'une loi normale univariée

Pour chaque $(t, j) \in \{1, \dots, T\} \times \{1, \dots, r\}$, générons $\xi_{\alpha,t,j} \sim \phi_j, \alpha = \{1, \dots, N\}$.
Puis définissons

$$\hat{L}_{1tj} = \frac{1}{N} \sum_{\alpha=1}^N e^{\xi_{\alpha,t,j}}, \hat{L}_{2tj} = \frac{1}{N} \sum_{\alpha=1}^N e^{2\xi_{\alpha,t,j}}$$

et

$$\hat{A}_{ti} = \sum_{j=1}^r Q_{ij} \hat{L}_{2tj} - \left(\sum_{j=1}^r Q_{ij} \hat{L}_{1tj} \right)^2$$

De plus, définissons $\hat{V}_{ti} = \sum_{j=1}^r Q_{ij} \hat{L}_{1tj}, i \in \{1, \dots, r\}$.

Si $C_T = G(S_T)$, alors définissons $\hat{f}_T(s, i) = G(s)$ et, pour chaque $t = T, \dots, 1$ et chaque $i \in \{1, \dots, r\}$, définissons

$$\hat{f}_{t-1}(s, i) = \frac{\beta_t}{\beta_{t-1}} \sum_{k=1}^r Q_{ik} \left[\frac{1}{N} \sum_{\alpha=1}^N \hat{f}_t(se^{\xi_{\alpha,t,k}}, k) \{1 - (e^{\xi_{\alpha,t,k}} - \hat{V}_{ti})(\hat{V}_{ti} - \beta_{t-1}/\beta_t)/\hat{A}_{ti}\} \right]$$

Enfin, $\hat{\phi}_t = \hat{\psi}_t(S_{t-1}, \hat{\tau}_{t-1})$, où

$$\hat{\psi}_t(s, i) = \frac{1}{s \hat{A}_{ti}} \sum_{k=1}^r Q_{ik} \left\{ \frac{1}{N} \sum_{\alpha=1}^N \hat{f}_t(se^{\xi_{\alpha,t,k}}, k) (e^{\xi_{\alpha,t,k}} - \hat{V}_{ti}) \right\}$$

Notons que si $G(s) = s$, alors $\hat{f}_t(s, i) = s$ et $\hat{\psi}_t(s, i) = 1$ pour tout s et pour tout $i \in \{1, \dots, r\}$.

5.3.3 Algorithme dans le cas d'une loi normale multivariée

Pour chaque $(t, j) \in \{1, \dots, T\} \times \{1, \dots, r\}$, générons

$$\xi_{\alpha,t,j} \sim N_d(\mu_j, A_j), \alpha = \{1, \dots, N\}$$

Puis définissons, pour chaque $l, m \in \{1, \dots, d\}$

$$(\hat{L}_{1tj})_l = \frac{1}{N} \sum_{\alpha=1}^N e^{(\xi_{\alpha,t,j})_l}, (\hat{L}_{2tj})_{lm} = \frac{1}{N} \sum_{\alpha=1}^N e^{(\xi_{\alpha,t,j})_l + (\xi_{\alpha,t,j})_m}$$

et

$$\hat{A}_{ti} = \sum_{j=1}^r Q_{ij} \hat{L}_{2tj} - \hat{V}_{ti} (\hat{V}_{ti})^T$$

où $\hat{V}_{ti} = \sum_{j=1}^r Q_{ij} \hat{L}_{1tj}$, $i \in \{1, \dots, r\}$. De plus, définissons $(X_{\alpha,t,j})_l = e^{(\xi_{\alpha,t,j})_l}$ et D_s la matrice diagonale avec $(D_s)_u = s_u$, $l \in \{1, \dots, d\}$.

Si $C_T = G(S_T)$, alors définissons $\hat{f}_T(s, i) = G(s)$ et, pour chaque $t = T, \dots, 1$ et chaque $i \in \{1, \dots, r\}$, définissons

$$\hat{f}_{t-1}(s, i) = \frac{\beta_t}{\beta_{t-1}} \sum_{k=1}^r Q_{ik} \left[\frac{1}{N} \sum_{\alpha=1}^N \hat{f}_t(D_s X_{\alpha,t,k}, k) \{1 - (X_{\alpha,t,k} - \hat{V}_{ti})^T (\hat{A}_{ti})^{-1} (\hat{V}_{ti} - \beta_{t-1} \beta_t \mathbf{1})\} \right]$$

où $(\mathbf{1})_l = 1$ pour tout $l = 1, \dots, m$.

Enfin, $\hat{\phi}_t = \hat{\psi}_t(S_{t-1}, \hat{\tau}_{t-1})$, où

$$\hat{\psi}_t(s, i) = D_s^{-1} (\hat{A}_{ti})^{-1} \sum_{k=1}^r Q_{ik} \left\{ \frac{1}{N} \sum_{\alpha=1}^N \hat{f}_t(D_s X_{\alpha,t,k}, k) (X_{\alpha,t,k} - \hat{V}_{ti}) \right\}.$$

Notons que si $G(s) = \lambda^T s$, alors $\hat{f}_t(s, i) = \lambda^T s$ et $\hat{\psi}_t(s, i) = \lambda$ pour tout s et pour tout $i \in \{1, \dots, r\}$.

Ainsi, dans le chapitre suivant nous testerons ce modèle. Néanmoins, afin de valider cet algorithme, nous le testerons sur une série de rendements corrélés en tarifant une option européenne univariée. Ainsi, une fois l'algorithme validé, nous pourrons le

tester sur des séries de rendements bivariés afin de répliquer certaines distributions données.

5.3.4 Validation de l'algorithme par une tarification d'option européenne univariée

Dans cette partie, nous allons tenter de valider notre stratégie de couverture en tarifant une option de type européenne univariée et en estimant nos paramètres sur des données corrélées.

Dans un premier temps, nous définissons les paramètres pour la simulation des données corrélées :

$$Q = \begin{pmatrix} 0.9 & 0.1 \\ 0.3 & 0.7 \end{pmatrix}$$
$$Moyenne = \begin{pmatrix} 0.0015 & -0.0007 \end{pmatrix}$$
$$Variance = \begin{pmatrix} 0.0005 & 0.0001 \end{pmatrix}$$

Ainsi, nous simulons 10000 rendements journaliers à partir desquels nous pourrions estimer notre modèle. Nos paramètres estimés pour notre modèle de chaînes de Markov cachées sont présentés dans le tableau 5.1.

Comme nous pouvons le constater, nous sommes très proches, en ce qui concerne les estimateurs du modèle par chaînes de Markov cachées, des paramètres de la simulation.

A partir de cette estimation, nous tarifons une option européenne ayant pour échéance 66 jours, un prix d'exercice de 100 sachant que le prix actuel $S_0 = 100$. Pour cela, nous re-simulons 1000 trajectoires de 66 rendements avec la même loi que précédemment.

TABLE 5.1 – Tableau représentant les estimés journaliers par modèle de chaînes de Markov cachées (HMM) pour 5000 rendements et 2 régimes

Estimateurs	HMM	
Moyenne R1	0.0015	
Moyenne R2	-0.0010	
Variance R1	0.0004879	
Variance R2	0.0000751	
Loi Stationnaire	0.7531	
	0.2469	
Q	0.8997	0.1003
	0.3060	0.6940

Ainsi, nous trouvons les prix suivants

	<u>Regime1</u>	<u>Regime2</u>	<u>BlackScholes</u>
<u>Prix</u>	6.7833	6.7109	6.8012
<u>Delta</u>	0.5607	0.5610	0.5546

Nous constatons que nos prix ainsi que nos deltas sont très proches de ceux donnés par le modèle Black & Scholes.

De plus, afin de calculer l'erreur de réplication, nous allons calculer sur, 10000 nouvelles trajectoires, la différence entre la stratégie de couverture et le flux monétaire à échéance. Ces nouvelles trajectoires doivent nécessairement être de la même loi que celle utilisée pour la toute première simulation, sur laquelle nous avons estimé nos deux modèles. A partir de ces nouvelles trajectoires, nous pourrons donc calculer la moyenne de ces erreurs ME ainsi que la racine de la moyenne des erreurs aux carrés $RMEC$.

	<u>HMM</u>
<u>ME</u>	0.0114
<u>RMEC</u>	0.8369

Nous pouvons constater, grâce à ces résultats ainsi qu'au graphique 5.1 (qui est

un histogramme des erreurs de couverture), que nos erreurs sont relativement faibles et centrées en zero. Ainsi, ces derniers nous permette de valider que notre estimation et notre tarification par chaînes de Markov cachées fonctionne dans le cas de données corrélées. Le RMEC peut paraître élevé surtout par rapport à l'ordre de grandeur du prix de l'option. Néanmoins, nous sommes en train de tarifer un produit dérivé dont le flux monétaire à échéance est non-linéaire et donc difficilement répliable. Dans le cadre de cette étude, notre flux monétaire à échéance est la fonction g , qui est une certaine distribution. Cela nous donne un profil de flux monétaire à échéance plus lisse et donc plus facile à répliquer. Nous nous attendons donc à avoir des erreurs moins élevées.

Ainsi, notre algorithme étant validé, nous pourrons dans le prochain chapitre le tester dans le cas bivarié. Pour cela nous utiliserons les données de l'article de Papageorgiou et al. (2008) afin de les estimer par chaînes de Markov cachées. Nous pourrons ainsi dans un premier temps comparer nos estimations, mais aussi répliquer certaines distributions données.

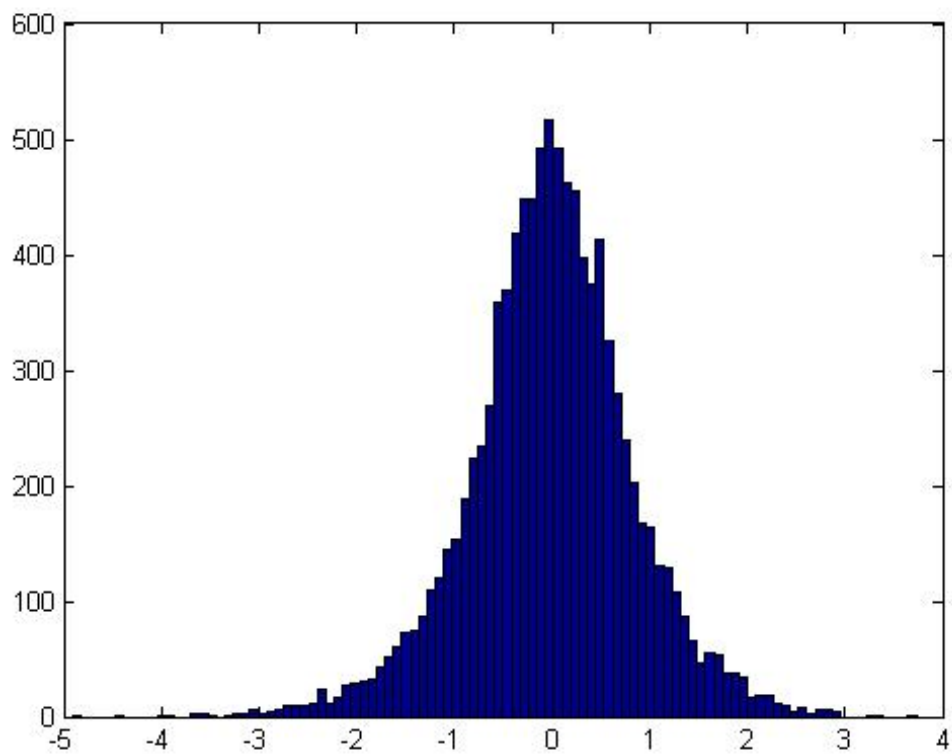


FIGURE 5.1 – Histogramme des erreurs de couverture pour la réplication d'une option européenne avec un modèle de HMM

Chapitre 6

Analyse des résultats

Dans ce chapitre, nous allons tester l'estimation et la répliation par chaînes de Markov avec changements de régimes gaussiens. Sachant que dans le modèle de répliation de Papageorgiou et al. (2008), les estimations sur les séries temporelles se font avec des mélanges de gaussiennes, nous comparerons notre nouvelle estimation à cette méthode. Dans un premier temps, nous estimerons sur des séries temporelles bivariées correspondant exactement à celle utilisé dans l'article de Papageorgiou et al. (2008). Enfin, nous tenterons, à partir de ces même données, de répliquer certaines distributions données.

6.1 Estimations sur des séries temporelles bivariées

Nous allons, dans cette partie, nous limiter aux séries de rendements utilisées par Papageorgiou et al. (2008). Ces rendements proviennent de la base de donnée CRB Trader.

Rappelons que l'estimation est faite sur une série bivariée correspondant aux rendements du portefeuille de l'investisseur et à la réserve utilisée dans la répliation. Dorénavant, la première dimension représentera le portefeuille et la deuxième dimension représentera la réserve. L'article estime pour deux types de réserves différentes,

mais nous utiliserons seulement la première. Ainsi, voici le détail de la composition de nos deux séries :

- **Portefeuille** : 50% SP500 et 50% Obligation gouvernementale Américaine 30 ans (Treasury Bonds).
- **Réserve 1** : 50% PowerShares Dynamic Small Cap Value Portfolio, 25% iShares Lehman 20 Year Treasury Bond Fund et 25% Citigroup Treasury 10 Year Bond Fund

Les estimations sont faites sur la période entre le 30 Janvier 1997 et le 29 Décembre 2006 (120 mois), ainsi que sur les deux sous-périodes entre le 30 Janvier 1997 et le 29 Décembre 2001 (59 mois) et entre le 30 Décembre 2001 et le 29 Décembre 2006 (61 mois). Voici le résumé présenté dans l'article de Papageorgiou et al. (2008)).

TABLE 6.1 – Résumé des statistiques pour les rendements journaliers du portefeuille ainsi que pour ceux de la réserve, sur les différentes périodes

Actif	Statistiques	Période 1 (97-06)	Période 2 (97-01)	Période 3 (02-06)
Portefeuille	Mean	0.0035	0.0047	0.0024
	S.Dev	0.0244	0.0289	0.0192
	Skew	-0.2150	-0.2697	-0.2482
	R. Sk	-0.0813	-0.2665	-0.1097
	Kurt	3.2109	2.6942	3.6637
	R. Kurt	3.2467	2.7483	3.6386
Réserve 1	Mean	0.0094	0.0095	0.0093
	S.Dev	0.0225	0.0260	0.0187
	Skew	0.3006	0.5346	-0.3480
	R. Sk	0.0362	0.0552	0.0159
	Kurt	5.0025	5.0399	3.2161
	R. Kurt	3.2419	4.0561	2.9244
	Corr. avec PF	0.6749	0.7054	0.6206

Afin de déterminer le nombre optimal de régimes dans notre estimation par chaînes de Markov cachées, nous utiliserons cette fois-ci le test basé sur la transformée de Rosenblatt présentée dans la partie 5.3.2. Ainsi, nous choisissons le nombre de régimes pour laquelle la P-value est supérieure à 0.05 en commençant à estimer

pour deux régimes.

Une fois que nous avons notre nombre de régimes, et les estimateurs associés, nous sommes en mesure de simuler à partir de notre modèle. Ainsi, nous simulons 500000 séquences de 22 jours afin d'avoir nos rendements mensuels. Sur ces 500000 rendements mensuels, nous pouvons estimer notre loi mensuelle. Nous l'utiliserons plus tard pour la fonction g . Sachant que nous ne nous intéressons pas à la dépendance entre les rendements mensuels, nous estimerons donc notre loi mensuelle avec un mélange de gaussiennes.

Afin de comparer notre estimation par chaînes de Markov cachées, nous estimerons sur nos rendements réels journaliers, comme le faisait Papageorgiou et al. (2008), un mélange de gaussiennes. Ainsi, à partir de celle-ci, nous re-simulerons des séquences de 22 jours afin de pouvoir comparer les moments mensuels des données ainsi que ceux des données re-simulées avec nos deux modèles.

Ainsi, pour les trois périodes suivantes, nous indiquerons les résultats du test basé sur la transformée de Rosenblatt ainsi que nos estimateurs par chaînes de Markov cachées pour le nombre de régimes choisis. De plus, nous présenterons les moments mensuels de nos simulations faites à partir de nos deux modèles afin de les comparer aux moments mensuels des données originales.

6.1.1 Estimations pour la période de 1997 à 2006

Nous allons maintenant estimer notre modèle sur la longue période, c'est-à-dire entre 1997 et 2006. Dans un premier temps, nous devons déterminer le nombre optimal de régimes à utiliser dans notre estimation de chaînes de Markov cachées comme nous l'avons mentionné plus tôt. Ainsi, comme nous pouvons le voir dans le tableau 6.2, nous utiliserons trois régimes pour cette période présentés dans le tableau 6.3 :

TABLE 6.2 – Tableau présentant les résultats du test basé sur la transformée de Rosenblatt pour la période 97-06

Régimes	Pvalue
2	0%
3	9%
4	10%
5	26%

TABLE 6.3 – Tableau présentant les estimés journaliers par modèle de chaînes de Markov cachées (HMM) pour la période 97-06, 3 régimes

Estimateurs	HMM Portefeuille	HMM Réserve	
Moyenne R1	0.182055e-003	0.348456e-003	
Moyenne R2	-0.011116e-003	0.829117e-003	
Moyenne R3	0.216623e-003	0.411299e-003	
Covariance R1	0.372465e-004	0.186967e-004	
	0.186967e-004	0.165847e-004	
Covariance R2	0.969680e-004	0.539561e-004	
	0.539561e-004	0.610620e-004	
Covariance R2	0.168346e-004	0.169078e-004	
	0.169078e-004	0.254400e-004	
Q	0.9730	0.0262	0.0007
	0.0998	0.9001	0.0000
	0.0000	0.0120	0.98800
Loi Stationnaire	0.5401		
	0.1422		
	0.3172		

En ce qui concerne les moments mensuels, nous pouvons comparer notre estimations à celles trouvés en utilisant un mélange de cinq gaussiennes, comme le faisait Papageorgiou et al. (2008) dans leur article pour cette période. Les résultats sont présentés dans le tableau 6.4

Nous voyons que notre estimation par chaînes de Markov cachées semble plus

TABLE 6.4 – Moyenne, variance, asymétrie, aplatissement et corrélation mensuels des données réelles et resimulées par nos deux modèles pour la période 97-06

Type	Portefeuille	Réserve
Données	0.00352781	0.00940346
	0.02440561	0.02248843
	-0.21226604	0.29679262
	3.15263311	4.87047165
	0.6748	
HMM	0.00347050	0.00911116
	0.02849735	0.02318492
	-0.03311449	0.14560444
	3.70115098	3.56784288
	0.7167	
Mélanges	0.00318538	0.00894803
	0.02874008	0.02321657
	-0.02659318	0.00360342
	3.05167940	3.05720402
	0.7314	

proche des moments mensuels. Cela se reflète également sur les deux graphiques présentés dans la figure 6.1

A partir de nos rendements mensuels simulés grâce à notre modèle de chaînes de Markov cachées, nous avons estimé un mélange de gaussiennes à deux régimes que nous utiliserons pour la fonction g .

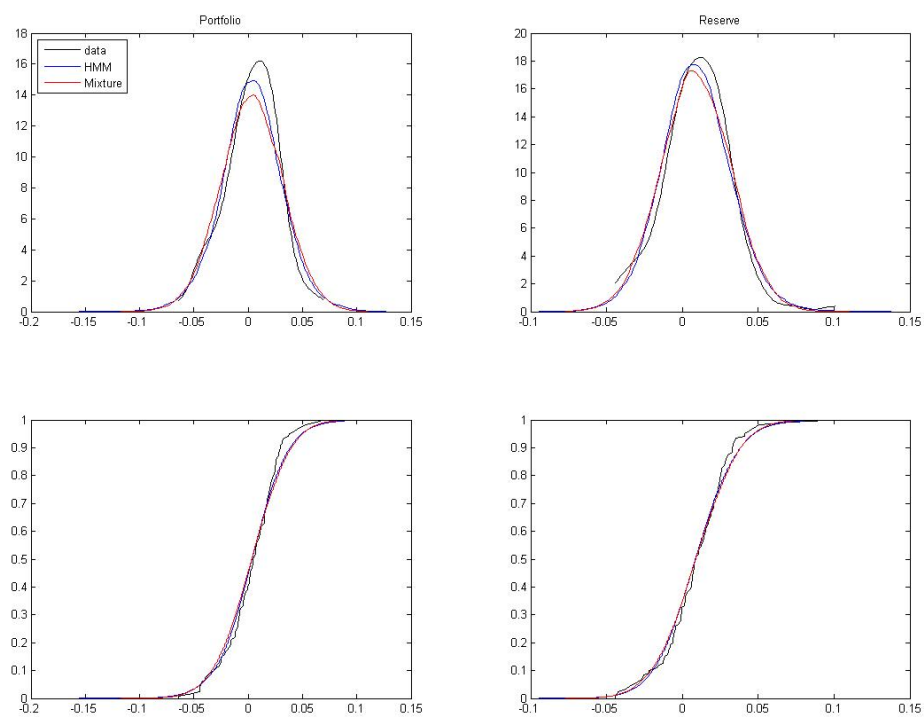


FIGURE 6.1 – Estimateurs à noyaux et fonctions de répartition des rendements mensuels réels, des rendements mensuels générés par chaînes de Markov cachées et des rendements mensuels générés par un mélange pour le période de 1997 à 2006

6.1.2 Estimations pour la période de 1997 à 2001

En ce qui concerne la première sous-période, nous devons également déterminer le nombre optimal de régimes à utiliser dans notre estimation de chaînes de Markov cachées. Ainsi, comme nous pouvons le voir dans le tableau 6.5, nous utiliserons trois régimes pour cette période présentés dans le tableau 6.6 :

TABLE 6.5 – Tableau présentant les résultats du test basé sur la transformée de Rosenblatt pour la période 97-01

Régimes	Pvalue
2	0%
3	12%
4	26%
5	89%

TABLE 6.6 – Tableau présentant les estimés journaliers par modèle de chaînes de Markov cachées (HMM) pour la période 97-01, 3 régimes

Estimateurs	HMM Portefeuille	HMM Réserve	
Moyenne R1	0.268344e-003	0.433973e-003	
Moyenne R2	0.656980e-003	-0.770651e-003	
Moyenne R3	-0.324313e-003	1.369893e-003	
Covariance R1	0.032303e-003	0.015123e-003	
	0.015123e-003	0.013146e-003	
Covariance R2	0.131856e-003	0.056458e-003	
	0.056458e-003	0.034229e-003	
Covariance R2	0.063827e-003	0.040000e-003	
	0.040000e-003	0.057323e-003	
Q	0.9125	0.0767	0.0106
	0.4009	0.4505	0.1484
	0.0112	0.1291	0.8596
Loi Stationnaire	0.6627		
	0.1391		
	0.1973		

En ce qui concerne les moments mensuels, nous pouvons également comparer

notre estimation à celle trouvée en utilisant un mélange de cinq gaussiennes, comme le faisait Papageorgiou et al. (2008) pour cette période. Les résultats sont présentés dans le tableau 6.7.

TABLE 6.7 – Moyenne, variances, asymétrie et aplatissement mensuels des données réelles et resimulées par nos deux modèles pour la période 97-01

Type	Portefeuille	Réserve
Données	0.00474203	0.00948747
	0.02893752	0.02600776
	-0.26282371	0.52092645
	2.61955045	4.77105705
	0.7054	
HMM	0.00450650	0.00997590
	0.03405242	0.02388291
	-0.06056747	0.28329971
	3.19372443	3.87216396
	0.6816	
Mélanges	0.00450596	0.01003163
	0.03392574	0.02363823
	0.00137632	0.05365754
	3.06094793	3.08601780
	0.7063	

A partir de nos rendements mensuels simulés grâce à notre modèle de chaînes de Markov cachées, nous avons estimé un mélange de gaussiennes à 2 régimes que nous utiliserons pour la fonction g .

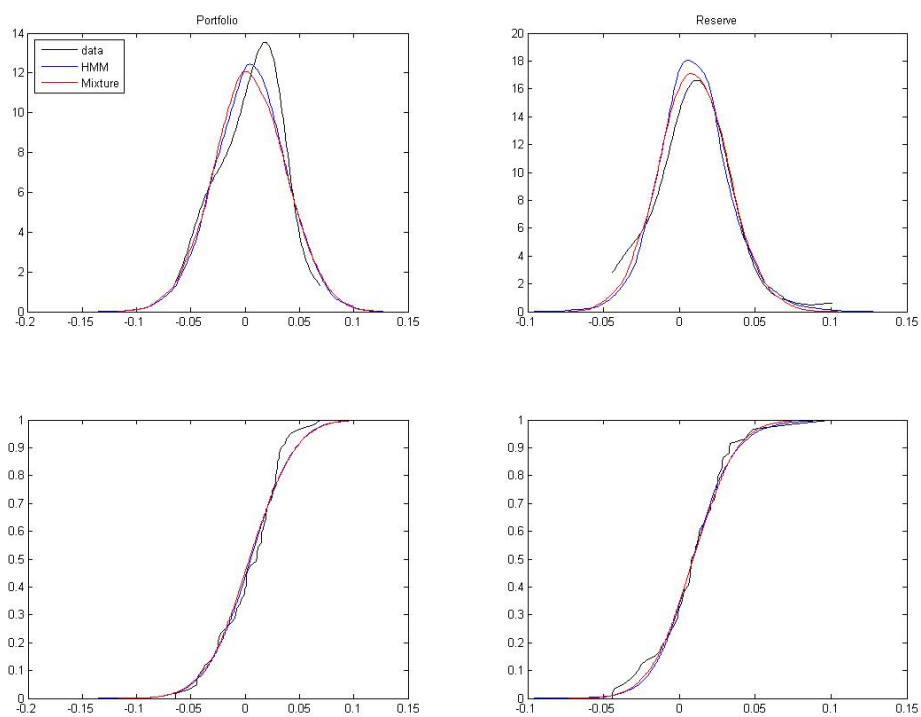


FIGURE 6.2 – Estimateurs à noyaux et fonctions de répartition des rendements mensuels réels, des rendements mensuels générés par chaînes de Markov cachées et des rendements mensuels générés par un mélange pour la période de 1997 à 2001

6.1.3 Estimations pour la période de 2002 à 2006

Pour la deuxième sous-période, nous devons également déterminer le nombre optimal de régimes à utiliser dans notre estimation de chaînes de Markov cachées. Ainsi, comme nous pouvons le voir dans le tableau 6.8, nous utiliserons trois régimes pour cette période présentés dans le tableau 6.9 :

TABLE 6.8 – Tableau présentant les résultats du test basé sur la transformée de Rosenblatt pour la période 02-06

Régimes	Pvalue
2	2%
3	19%
4	84%
5	92%

TABLE 6.9 – Tableau présentant les estimés journaliers par modèle de chaînes de Markov cachées (HMM) pour la période 02-06

Estimateurs	HMM Portefeuille	HMM Réserve	
Moyenne R1	-0.017683e-003	0.449915e-003	
Moyenne R2	0.968509e-003	0.819136e-003	
Moyenne R3	-0.054477e-003	0.261454e-003	
Covariance R1	0.419070e-004	0.269371e-004	
	0.269371e-004	0.277002e-004	
Covariance R2	0.043414e-004	0.041737e-004	
	0.041737e-004	0.068259e-004	
Covariance R2	0.209142e-004	0.213127e-004	
	0.213127e-004	0.320248e-004	
Q	0.9979	0.0021	0.0000
	0.0000	0.2263	0.7737
	0.0000	0.2713	0.7287
Loi Stationnaire	0.3743		
	0.1630		
	0.4620		

En ce qui concerne les moments mensuels, nous pouvons également comparer

notre estimation à celle trouvée en utilisant un mélange de trois gaussiennes, comme le faisait Papageorgiou et al. (2008) pour cette période. Les résultats sont présentés dans le tableau 6.10

TABLE 6.10 – Moyenne, variance, asymétrie et aplatissement mensuels des données réelles et re-simulées par nos deux modèles pour la période 02-06, 4 régimes

Type	Portefeuille	Réserve
Données	0.00235340	0.00932220
	0.01920992	0.01868581
	-0.23312017	-0.33940337
	3.47952199	3.10199961
	0.6163	
HMM	0.00258810	0.00915473
	0.02399412	0.02403299
	-0.12790635	0.00193373
	3.63263161	3.00297951
	0.7799	
Mélanges	0.00263594	0.00924638
	0.02389148	0.02404151
	-0.07014468	-0.03275267
	3.05418411	3.02612533
	0.7885	

A partir de nos rendements mensuels simulés grâce notre modèle de chaînes de Markov cachées, nous avons estimé un mélange de gaussiennes à 2 régimes que nous utiliserons pour la fonction g .

Maintenant que nous avons estimé nos paramètres pour ces différentes périodes, nous allons pouvoir les utiliser afin de déterminer si nous pouvons répliquer certaines distributions. Dans le cadre de ce mémoire, nous ne répliquerons pas de vrais fonds de couverture mais une distribution donnée. Si nous pouvons valider notre nouvelle stratégie sur ces distributions, alors nous le pourrons pour de vrais fonds.

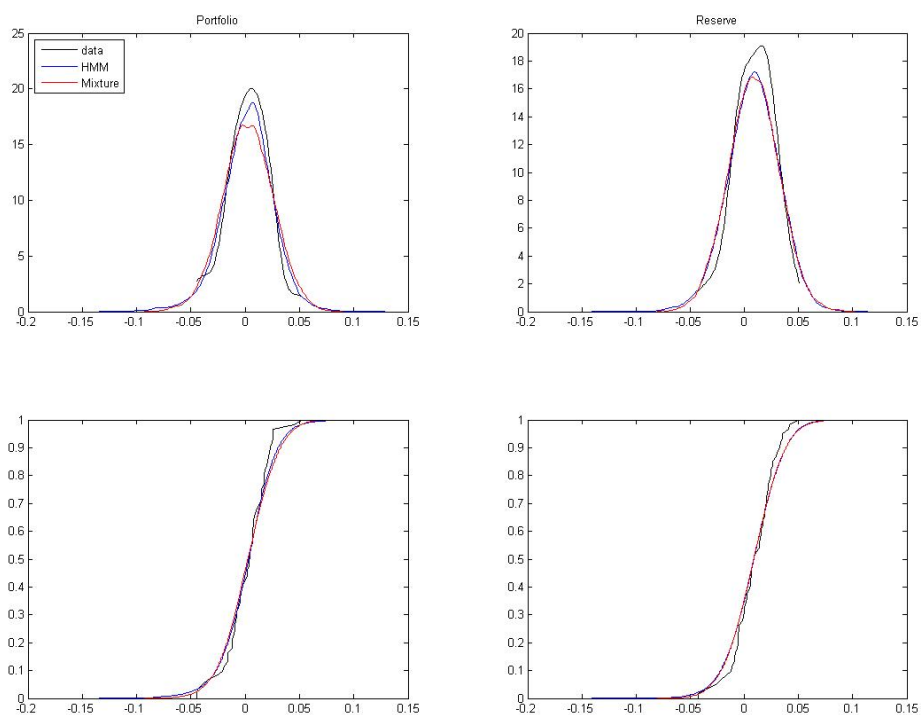


FIGURE 6.3 – Estimateurs à noyaux et fonctions de répartition des rendements mensuels réels, des rendements mensuels générés par chaînes de Markov cachées et des rendements mensuels générés par un mélange pour la période de 2002 à 2006

6.2 Réplication d'une distribution normale

Maintenant que nous avons nos estimations et donc nos lois journalières et mensuelles pour différentes périodes, nous allons pouvoir tenter de répliquer une distribution donnée. Ainsi, nous allons répliquer la même distribution que celle utilisée dans Papageorgiou et al. (2008). Il s'agit d'une distribution normale centrée d'écart-type 12% et de corrélation 30% avec le portefeuille. Comme dans leur article, nous re-simulons 10000 valeurs de $g(R_{0,T}^{(1)}, R_{0,T}^{(2)}), \log(V_T/100)$. De plus, nous normalisons la valeur de notre portefeuille ainsi que la réserve au temps $t = 1$:

- Portefeuille initial : $S_1(0) = 100$.
- Actif de réserve : $S_2(0) = 100$.

6.2.1 Réplication avec les données de la période entre 1997 et 2006

Le tableau 6.11 expose les résultats de la réplication, c'est-à-dire les caractéristiques de la loi à répliquer ainsi que celles de la fonction g et de la stratégie optimale de couverture. Le tableau 6.12 nous montre l'erreur de réplication et le tableau 6.13, la valeur de V_0 et des $Delta$ initiaux pour le portefeuille et la réserve, ainsi que leurs intervalles de confiance.

TABLE 6.11 – Résultats de la réplication basée sur 10 000 trajectoires pour $g(R_{0,T}^{(1)}, R_{0,T}^{(2)}) = \log(C_T/100)$ et $\log(V_T/100)$ avec la stratégie optimale de couverture. 97/06

Paramètres	Valeurs réelles	g	Stratégie de couverture Optimale
Moyenne	0	6.0520E-05	8.0255E-05
Écart type	0.034641	0.034726	0.034698
Asymétrie	0	-0.0138	-0.0159
Aplatissement	0	0.0034	0.0041
ρ	0.3	0.3019	0.3021

Nous pouvons constater dans le tableau 6.11 que la fonction g est très proche de la loi à répliquer et qu'il en est de même pour la stratégie optimale de couverture. La corrélation est également très proche de la cible.

TABLE 6.12 – Résultats de la réplication basée sur 10 000 trajectoires pour le payoff \tilde{g} et $\log(V_T/100)$ avec la stratégie de couverture optimale. 97/06

Paramètres	Couverture Optimale
Moyenne de l'erreur de couverture	0.001878
Racine carré MEC^2	0.018969

Grâce au tableau 6.12, nous constatons que notre erreur moyenne est centrée sur zéro. La racine carrée de la moyenne des erreurs au carré est relativement faible en comparaison à notre article de référence où elle est également de l'ordre de 10^{-2} . En ce qui concerne le critère de Kat et Palaro que nous appelons $V0$, qui nous indique si nous sur-performons la distribution à répliquer ou non, ainsi que les Deltas initiaux du portefeuille et de la réserve, nous trouvons les valeurs présentées dans le tableau 6.13.

TABLE 6.13 – Critère de performance $V0$ et Delta initiaux de la stratégie ainsi que leurs intervalles pour un niveau de confiance de 95% sur 1000 itérations. 97/06

Paramètres	Valeurs moyennes	Intervalles de Confiance
$V0$	98.6799	3.1282E-03
Delta - Portefeuille	-1.9185	2.3729E-03
Delta - Reserve	2.45293	1.6221E-03

Nous constatons que nous sur-performons ce que nous cherchons à répliquer. Concernant les intervalles de confiance, nous concluons qu'ils sont relativement petits pour notre $V0$ ainsi que pour nos *Delta*.

6.2.2 Réplication avec les données de la période entre 1997 et 2001

Nous présenterons les mêmes tableaux que précédemment, mais cette fois-ci avec les estimateurs de la période entre 1997 et 2001.

TABLE 6.14 – Résultats de la réplication basée sur 10 000 trajectoires pour $g(R_{0,T}^{(1)}, R_{0,T}^{(2)}) = \log(C_T/100)$ et $\log(V_T/100)$ avec la stratégie optimale de couverture. 97/01

Paramètres	Valeurs réelles	g	Stratégie de couverture Optimale
Moyenne	0	1.0670E-04	-1.1819E-4
Écart type	0.034641	0.034654	0.034674
Asymétrie	0	0.0469	0.0712
Aplatissement	0	0.0199	0.1683
ρ	0.3	0.2764	0.2760

Nous remarquons certaines divergences pour la fonction g comparativement à l'estimation de la période précédente. En effet, notre corrélation est plus éloignée de la cible de 30% et l'asymétrie et l'aplatissement sont également un plus éloignées. Nous constatons également que cela était bien reflété dans le tableau 6.7 où nous trouvons un écart entre nos rendements réels et ceux resimulés par chaînes de Markov cachées au niveau de ces deux moments.

TABLE 6.15 – Résultats de la réplication basée sur 10 000 trajectoires pour le payoff \tilde{g} et $\log(V_T/100)$ avec la stratégie de couverture optimale. 97/01

Paramètres	Couverture Optimale
Moyenne de l'erreur de couverture	-0.001063
Racine carrée MEC^2	0.132984

En ce qui concerne les erreurs de réplifications ainsi que les intervalles de confiance sur V_0 , Δ_P et Δ_R , nous trouvons des résultats assez similaires qu'à la période précédente sauf pour la racine carrée de la moyenne de l'erreur de couverture.

TABLE 6.16 – Critères de performance $V0$ et Delta initiaux de la stratégie ainsi que leurs intervalles de confiance pour un niveau de confiance de 95% sur 1000 itérations. 97/01

Paramètres	Valeurs moyennes	Intervalles de Confiance
$V0$	98.6684	3.0719E-03
Delta - Portfeuille	-0.7106	1.2785E-03
Delta - Reserve	2.2088	1.5182E-04

6.2.3 Réplication avec les données de la période entre 2002 et 2006

Enfin, voici les résultats pour le période entre 2002 et 2006. Nous ne voyons pas, dans le tableau 6.17, d'écart significatif vis-a-vis de notre moyenne ou notre écart type cible comme précédemment. La corrélation est également très proche de notre valeur cible de 30%. L'asymétrie ainsi que l'aplatissement sont également très satisfaisants.

TABLE 6.17 – Résultats de la réplication basée sur 10 000 trajectoires pour $g(R_{0,T}^{(1)}, R_{0,T}^{(2)}) = \log(C_T/100)$ et $\log(V_T/100)$ avec la stratégie optimale de couverture. 02/06

Paramètres	Valeurs réelles	g	Stratégie de couverture Optimale
Moyenne	0	5.9576E-05	7.4517E-5
Écart type	0.034641	0.034354	0.034346
Asymétrie	0	-0.0488	-0.0497
Aplatissement	0	0.0037	0.0049
ρ	0.3	0.3030	0.3030

Nous observons que nous sur-performons notre cible car nous avons un $V0$ inférieur à 100.

De manière générale, nous étions très proches de notre cible pour les trois périodes. Néanmoins, nous avons pu constater que les estimations étaient importantes

TABLE 6.18 – Résultats de la réplication basée sur 10 000 trajectoires pour le payoff \tilde{g} et $\log(V_T/100)$ avec la stratégie de couverture optimale. 02/06

Paramètres	Couverture Optimale
Moyenne de l'erreur de couverture	0.001468
Racine carrée MEC^2	0.009446

TABLE 6.19 – Critère de performance V_0 et Delta initiaux de la stratégie ainsi que leurs intervalles pour un niveau de confiance de 95% sur 1000 itérations - 02/06

Paramètre	Valeur moyenne	Intervalles de Confiance
V_0	98.6899	3.1845E-03
Delta - Portefeuille	-1.9298	2.1416E-03
Delta - Reserve	2.4451	1.0541E-03

et toutes divergences au niveau des moments mensuels dans notre estimation se reflétaient dans la fonction de payoff g . De plus, nous ne semblons pas constater que nous performons mieux lorsque notre estimation est sur une plus longue période. Les intervalles de confiance sur nos estimateurs sont tous relativement petits, ainsi que la racine carrée de la moyenne des erreurs de couvertures.

6.2.4 Comparaison des résultats avec notre article de référence

Il est intéressant de comparer nos résultats à ceux de Papageorgiou et al. (2008) dans lequel ils tentent également de répliquer une distribution normale. Pour cela, ils utilisent un mélange de gaussiennes à quatre régimes afin de modéliser les rendements journaliers du portefeuille et de la réserve. Les tableaux 6.20 et 6.21 présentent leurs résultats.

Ainsi, afin de comparer notre modèle au leur, nous allons utiliser une chaîne de Markov cachées à quatre régimes pour modéliser les rendements journaliers du por-

TABLE 6.20 – Résultats de notre article de référence - réplication basée sur 10 000 trajectoires pour $g\left(R_{0,T}^{(1)}, R_{0,T}^{(2)}\right) = \log(C_T/100)$ et $\log(V_T/100)$ avec la stratégie optimale de couverture.

Paramètres	Valeurs réelles	g	Stratégie de couverture Optimale
Moyenne	0	3.957E-07	3.574E-7
Écart type	0.034641	0.034957	0.034961
Asymétrie	0	-0.0589	-0.0640
Aplatissement	0	0.0299	0.0324
ρ	0.3	0.3028	0.3028

TABLE 6.21 – Résultats de notre article de référence - réplication basée sur 10 000 trajectoires pour le payoff \tilde{g} et $\log(V_T/100)$ avec la stratégie de couverture optimale.

Paramètres	Couverture Optimale
Moyenne de l'erreur de couverture	0.000004009
Racine carrée MEC^2	0.017861376

tefeuille et de la réserve. Ces régimes gaussiens ont les mêmes caractéristiques que ceux du mélange utilisée par les auteurs dans leur article. Nous utiliserons également la même loi stationnaire. Néanmoins, nous devons introduire une matrice de transition entre nos régimes qui sera la suivante :

$$Q = \begin{pmatrix} 0.30 & 0.40 & 0.25 & 0.05 \\ 0.10 & 0.60 & 0.10 & 0.20 \\ 0.07 & 0.33 & 0.50 & 0.10 \\ 0.05 & 0.10 & 0.10 & 0.75 \end{pmatrix}$$

Les tableau 6.22 et 6.23 présente les résultats que nous obtenons avec notre modèle.

Ainsi, notre réplication semble bien fonctionner dans ce cas ci. Nous remarquons que nos erreurs sont supérieures à celles trouvées dans notre article de référence

TABLE 6.22 – Résultats de la réplication basée sur 10 000 trajectoires pour $g(R_{0,T}^{(1)}, R_{0,T}^{(2)}) = \log(C_T/100)$ et $\log(V_T/100)$ avec la stratégie optimale de couverture.

Paramètres	Valeurs réelles	g	Stratégie de couverture Optimale
Moyenne	0	3.8629E-05	4.3562E-5
Ecart-type	0.034641	0.034899	0.034961
Assymétrie	0	0.0100	-0.0061
Aplatissement	0	-0.0296	0.0106
ρ	0.3	0.3007	0.3005

TABLE 6.23 – Résultats de la réplication basée sur 10 000 trajectoires pour le payoff \tilde{g} et $\log(V_T/100)$ avec la stratégie de couverture optimale.

Paramètres	Couverture Optimale
Moyenne de l'erreur de couverture	0.00069659
Racine carrée MEC^2	0.06550304

mais demeurent relativement faibles. De plus, le nombre optimal de régimes devrait être de trois et non quatre comme nous le faisons ici, ce qui augmente notre erreur quadratique. Néanmoins, en ce qui concerne notre stratégie de couverture optimale, nos résultats semblent plus proches de notre cible et donc plus satisfaisants. Nous pouvons donc conclure que notre méthode par chaînes de Markov cachées est aussi efficace que celle par mélanges de gaussiennes.

6.3 Réplication d'une distribution de Johnson

Nous allons maintenant tenter de répliquer une distribution de Johnson. Le choix de cette distribution vient du fait qu'elle peut présenter de l'asymétrie et avoir un aplatissement différent de 3. Cela nous intéresse car la distribution des rendements des fonds de couverture peuvent présenter de telles caractéristiques. Ainsi, nous tenterons de répliquer une distribution présentant une asymétrie positive et un aplatis-

sement plus forte. Cela peut être interprété comme une distribution présentant plus de rendements positifs, mais également plus de rendements extrêmes, qu'ils soient positifs ou négatifs.

6.3.1 Réplication avec les données de la période entre 1997 et 2006

Voici les résultats de la réplication pour une estimation sur la période entre 1997 et 2006. Nous pouvons constater dans le tableau 6.24 que notre fonction g , ainsi que notre stratégie de couverture optimale, présentent une asymétrie positive ainsi qu'un aplatissement en excès. L'asymétrie semble bien captée par rapport à la cible alors que l'aplatissement semble plus faible.

TABLE 6.24 – Résultats de la réplication d'une Johnson basée sur 10 000 trajectoires pour $g(R_{0,T}^{(1)}, R_{0,T}^{(2)}) = \log(C_T/100)$ et $\log(V_T/100)$ avec la stratégie optimale de couverture. 97/06

Paramètres	Valeurs réelles	g	Stratégie de couverture Optimale
Moyenne	0	-6.4141E-05	-8.6382E-05
Écart type	0.034641	0.034337	0.034382
Asymétrie	0.7838	0.7496	0.7140
Aplatissement	1.2358	0.8979	0.8121
ρ	0.3	0.2839	0.2837

En ce qui concerne le V_0 présenté dans le tableau 6.26, nous arrivons à surperformer notre cible car elle est significativement inférieure à 100. Néanmoins, nous voyons qu'elle est en moyenne faiblement supérieure à celle trouvée lors de la réplication d'une loi normale. Cela peut être interprété comme le coût associé à l'asymétrie positive demandée.

En ce qui concerne l'erreur de réplication présentée dans le tableau 6.25, nous re-

TABLE 6.25 – Résultats de la réplication basée sur 10 000 trajectoires pour le payoff \tilde{g} et $\log(V_T/100)$ avec la stratégie de couverture optimale. 97/06

Paramètres	Couverture Optimale
Moyenne de l'erreur de couverture	-0.002094
Racine carrée MEC	0.142863

TABLE 6.26 – Critère de performance V_0 et Delta initiaux de la stratégie ainsi que leurs intervalles pour un niveau de confiance de 95% sur 1000 itérations - 97/06

Paramètres	Valeurs moyennes	Intervalles de Confiance
V_0	98.7471	3.0206E-03
Delta - Portefeuille	-1.7532	2.7328E-03
Delta - Reserve	2.2413	2.4061E-03

marquons qu'elle est bien centrée sur 0. Néanmoins, la racine carrée de la moyenne des erreurs au carré est plus élevée. Ainsi, cela laisse penser qu'il est plus difficile de répliquer une distribution présentant plus de aplatissement, c'est à dire plus de rendements extrêmes.

6.3.2 Réplication avec les données de la période entre 1997 et 2001

Voici les résultats pour une estimation de nos lois sur la période entre 1997 et 2001. Les résultats diffèrent de ceux de la période précédente car les troisième et le quatrième moments de la stratégie de couverture optimale semblent bien plus proche de la cible.

Les erreurs de couverture sont également supérieures à celles pour une loi normale et le V_0 significativement supérieur. Cela nous permet de tirer les même conclusions qu'à la période précédente.

TABLE 6.27 – Résultats de la réplication d'une Johnson basée sur 10 000 trajectoires pour $g\left(R_{0,T}^{(1)}, R_{0,T}^{(2)}\right) = \log(C_T/100)$ et $\log(V_T/100)$ avec la stratégie optimale de couverture. 97/01

Paramètres	Valeurs réelles	g	Stratégie de couverture Optimale
Moyenne	0	-7.2957E-04	-5.0343E-04
Écart type	0.034641	0.033941	0.033670
Asymétrie	0.7838	0.7934	0.7267
Aplatissement	1.1119	1.2225	1.1889
ρ	0.3	0.2625	0.2680

TABLE 6.28 – Résultats de la réplication basée sur 10 000 trajectoires pour le payoff \tilde{g} et $\log(V_T/100)$ avec la stratégie de couverture optimale. 97/01

Paramètres	Couverture Optimale
Moyenne de l'erreur de couverture	0.021643
Racine carrée MEC	0.201787

TABLE 6.29 – Critère de performance V_0 et Delta initiaux de la stratégie ainsi que leurs intervalles pour un niveau de confiance de 95% sur 1000 itérations - 97/01

Paramètres	Valeurs moyennes	Intervalles de Confiance
V_0	98.7066	2.7879E-03
Delta - Portefeuille	-0.6335	1.1011E-04
Delta - Reserve	1.9779	1.5571E-03

6.3.3 Réplication avec les données de la période entre 2002 et 2006

Pour la période entre 2002 et 2006, nous trouvons des résultats similaires à la période allant de 1997 et 2006. En effet, nous n'avons pas pu bien capter l'asymétrie ainsi que la aplatissement en excès. Notre erreur, comme le montre le tableau 6.31, est également plus grande que lorsque notre cible était une loi normale et nos intervalles

de confiance sont relativement petits. Cette période, comme dans le cas d'une loi normale, semble la plus efficace pour la réplication de cette loi.

TABLE 6.30 – Résultats de la réplication d'une Johnson basée sur 10 000 trajectoires pour $g\left(R_{0,T}^{(1)}, R_{0,T}^{(2)}\right) = \log(C_T/100)$ et $\log(V_T/100)$ avec la stratégie optimale de couverture. 02/06

Paramètres	Valeurs réelles	g	Stratégie de couverture Optimale
Moyenne	0	1.0621E-04	-6.8335E-05
Écart type	0.03416410	0.034349	0.034333
Asymétrie	0.7838	0.7477	0.7014
Aplatissement	1.1119	0.8853	0.7718
ρ	0.3	0.2859	0.2843

TABLE 6.31 – Résultats de la réplication basée sur 10 000 trajectoires pour le payoff \tilde{g} et $\log(V_T/100)$ avec la stratégie de couverture optimale. 02/06

Paramètres	Couverture Optimale
Moyenne de l'erreur de couverture	-0.0017552
Racine carrée MEC	0.158296

TABLE 6.32 – Critères de performance V_0 et Delta initiaux de la stratégie ainsi que leurs intervalles pour un niveau de confiance de 95% sur 1000 itérations - 02/06

Paramètres	Valeurs moyennes	Intervalles de Confiance
V_0	98.7525	2.9558E-03
Delta - Portefeuille	-1.7667	2.9107E-03
Delta - Reserve	2.2375	2.5908E-03

Nous avons réussi à bien répliquer cette distribution pour des estimations faites pour les période entre 1997 et 2006 et entre 2002 et 2006. Les erreurs étaient relativement faibles et les moments de la fonction g et de la stratégie de couverture assez proches de notre cible. De plus, nous avons pu constater des valeurs de V_0

inférieures à 100 mais significativement supérieures à une réplification de loi normale.

6.4 Synthèse des résultats et problèmes rencontrés

Les résultats concernant les différentes réplifications de lois sont de manière générale satisfaisants. Ils nous permettent de valider cette technique qui utilise les chaînes de Markov cachées pour les estimations ainsi que dans la stratégie optimale de couverture. En effet, nous avons réussi à répliquer différentes lois données à l'aide d'estimation sur des données réelles. Ainsi, il serait intéressant de tester cette stratégie de réplification sans resimuler de nouvelles données mais en utilisant les vrais rendements futurs réalisés.

Nous avons également vu que l'estimation par chaînes de Markov cachées pouvait être plus intéressante que celle donnée par des mélanges de Gaussiennes, ce qui s'expliquerait pas la présence d'auto-corrélation dans nos données journalières.

Un des problèmes rencontrés a été, dans certains cas, notre estimation. En effet, celle-ci nous donnait une matrice de transition avec un état final, c'est-à-dire non transitoire. Ainsi, nous ne pouvions utiliser cette estimation avec ce nombre de régimes car notre modèle de prédiction nous donnerait alors seulement un régime, malgré qu'il en ait estimé d'autres dans le passé. De plus, certaines de nos estimations nous donnaient un nombre optimal de régimes où l'un d'entre eux possédait une variance nulle ce qui annulait ce régime.

Chapitre 7

Conclusion

Nous avons vu toutes les étapes qui nous permettent de construire une stratégie qui permet d'obtenir une certaine distribution mensuelle donnée. Pour cela, nous avons vu comment estimer un modèle de chaînes de Markov avec des changements de régimes de type gaussien, et ce, en univarié et en bivarié. Nous avons également expliqué comment, à partir de cette estimation, nous pouvions modifier la stratégie optimale de couverture donnée par Papageorgiou et al. (2008) afin de prendre en compte la dépendance sérielle estimée par notre modèle. Ainsi, la valeur ajoutée de ce modèle réside dans le fait qu'il prend en compte la dépendance journalière des rendements de notre portefeuille, ainsi que dans notre actif de réserve, utilisé dans la stratégie optimale de réplication.

Nous avons vu également que notre réplication pouvait être plus ou moins bonne selon les périodes sur lesquelles nous l'avons estimée. Ainsi, nous pouvons penser qu'en l'absence d'auto-corrélation un modèle utilisant les mélanges de gaussiennes serait préférable. Il serait donc intéressant dans une prochaine étude de s'intéresser à des méthodes avancées pour savoir dans quel cas il y a présence d'auto-corrélation journalière et dans quel cas il n'y en a pas. Ainsi, nous saurions pour quels actifs financiers estimer avec des chaînes de Markov cachées.

De plus, nous avons travaillé sur des données qui représentaient un mélange de

différents actifs, que ce soit au niveau du portefeuille de marché ou de l'actif de réserve. Il est plus difficile de détecter de l'auto-corrélation sur un tel mélange d'actifs. Néanmoins, dans le cas d'une option sur un actif précis, l'auto-corrélation peut être plus fréquente. Nous avons vu que notre stratégie de couverture correspondait à une tarification d'option américaine, il pourrait donc être intéressant d'appliquer cette méthode afin de tarifer des options américaines sur certains actifs présentant de l'auto-corrélation, voir même d'autres types de produits dérivés.

Bibliographie

- Agarwal, V. and Naik, N. Y. (2004). Risks and Portfolio Decisions Involving Hedge Funds. *The Review of Financial Studies*, 17(1) :63–98.
- Amin, G. and Kat, H. M. (2005). Hedge Fund Performance 1990-2000 : Do the Money Machines Really Add Value ? *Journal of Financial and Quantitative Analysis*, 38 :1–24.
- Asbrink, S., R. T. and T., T. (1998). Stylized facts of daily returns series and the hidden markov model. *Journal of Applied Econometrics*, 13 :217–244.
- Bilmes, J. (1998). A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussain Mixture and Hidden Markov Models. *Technical Report, International Computer Science Institute, Berkley*.
- Black, F. and Scholes, M. (1973). The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, 81 :637–654.
- Box, G. and G.M., J. (1976). *Time Series Analysis : Forecasting and Control*, volume second edition. Holden Day.
- Breusch, T. S., P. T. (1974). The lagrange multiplier test and its applications to model specification in econometrics. *Review of Economics Studies*, 47 :239–253.
- Cosslett, S. R. and Lee, L. F. (1985). Serial correlation in discrete variable models. *Journal of Econometrics*, 27 :79–97.
- Dempster, A. P., Laird, N., and Rubin, D. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Royal Statistical Society*, 39(1) :1–38.
- Diez de los Rios, A. and Garcia, R. (2006). Assessing and Valuing the Nonlinear Structure of Hedge Fund Returns. *Technical Report, CIRANO et CIREQ*.
- Durbin, J. (1973). Weak convergence of the sample distribution function when parameters are estimated. *Ann. Statist.*, 1(2) :279–290.

- Dybvig, P. (1988a). Distributional Analysis of Portfolio Choice. *Journal of Business*, 61 :369–393.
- Dybvig, P. (1988b). Inefficient Dynamic Portfolio Strategies or How to Throw Away a Million Dollars in the Stock Market. *Review of Financial Studies*, 1 :67–88.
- Engle, R. F. (1982a). A General Approach to Lagrange Multiplier model Diagnostic. *Journal of Econometrics*, 20 :83–104.
- Engle, R. F. (1982b). Wald, Likelihood Ratio, and Lagrange Multiplier Tests in econometrics. *Handbook of Econometrics*, 2.
- Fung, W. and Hsieh, D. A. (1997). Empirical Characteristics of Dynamic Trading Strategies : The Case of Hedge Fund. *The Review of Financial Studies*, 10(2) :274–302.
- Fung, W. and Hsieh, D. A. (2001). The Risk in Hedge Fund Strategies : Theory and Evidence from Trend Followers. *The Review of Financial Studies*, 14(2) :313–341.
- Genest, C., Ghoudi, K., and Rivest, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82 :543–552.
- Genest, C., Quessy, J.-F., and Rémillard, B. (2006). Goodness-of-fit procedures for copula models based on the integral probability transformation. *Scand. J. Statist.*, 33 :337–366.
- Genest, C. and Rémillard, B. (2008). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Ann. Inst. H. Poincaré Sect. B*, 44 :1096–1127.
- Genest, C., Rémillard, B., and Beaudoin, D. (2009). Omnibus goodness-of-fit tests for copulas : A review and a power study. *Insurance Math. Econom.*, 44 :199–213.
- Giudici, P., R. T. and Vandekerckhove (2000). Stylized facts of daily returns series and the hidden markov model. *Journal of Applied Econometrics*, 56(3) :742–747.
- Glosten, L. and Jagannathan, R. (1994). A Contingent Claim Approach to Performance Evaluation. *Journal of Empirical Finance*, 1 :133–160.
- Godfrey, L. G. and Wickens, M. (1981). Testing linear and log-linear regressions for functional form. *Review of Economics Studies*, 48 :487–496.
- Goldfeld, S. M. and Quandt, R. M. (1982a). A Markov Model for Switching Regression. *Journal of Econometrics*, 1 :3–16.

- Hamilton, J. D. (1989). A New Approach to the Economic of Nonstationary Time Series and the Business Cycle. *Econometrica*, 57 :357–384.
- Hamilton, J. D. (1990). Analysis of Times Series Subject to Changes in Regime. *Journal of Econometrics*, 45 :39–70.
- Hamilton, J. D. (1996). Specification testing in Markov-Switching Time-series Models. *Journal of Econometrics*, 70 :127–157.
- Hocquard, A., Papageorgiou, N., and Rémillard, B. (2007). Optimal hedging strategies with an application to hedge fund replication. *Wilmott Magazine*, (Jan-Feb) :62–66.
- Kat, H. M. and Palaro, H. P. (2005). Who Needs Hedge Funds? a Copula-Based Approach to Hedge Fund Return Replication. *Technical Report, Cass Business School, City University*.
- Kat, H. M. and Palaro, H. P. (2006). Tell Me What You Want, What You Really Really Want? *Technical Report, Cass Business School, City University*.
- Liang, B. (1999). On the Performance of Hedge Funds. *Financial Analyst Journal*, 65 :72–85.
- Mitchell, M. and Pulvino, T. (2001). Characteristics of Risk and Return in Risk Arbitrage. *The Journal of Finance*, 56(6) :2135–2175.
- Nelsen, R. B. (1999). *An Introduction to Copula*, volume 139. Springer-Verlag, New York.
- Newey, W. (1985). Maximum Likelihood Specification testing and Conditional Moment Tests. *Econometrica*, 53 :1047–1070.
- Papageorgiou, N., Rémillard, B., and Hocquard, A. (2008). Replicating the properties of hedge fund returns. *Journal of Alternative Investments*, 11 :8–38.
- Poritz, A. B. (1982). Linear predictive hidden markov models and the speech signal. *Journal of Econometrics*, pages 1291–1294.
- Qian, W. and Titterton, D. (1991). Estimation of parameters in hidden markov models. *Philosophical Transactions : Physical Sciences and Engineering*, 337 :407–428.
- Rabiner, L. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Review of Economics Studies*.

- Rabiner, L. and Juang, H. (1985). A probabilistic distance measure for hidden markov models.
- Rémillard, B. (2008). Modelling asset returns with markov regime-switching models. *Technical Report, HEC Montréal.*
- Remillard, B., Hocquard, A., and Papageorgiou, N. A. (2010). Option Pricing and Dynamic Discrete Time Hedging for Regime-Switching Geometric Random Walks Models. Technical report.
- Rémillard, B., Papageorgiou, N., and Soustra, F. (2010). Dynamic Copulas. *Gerad.*
- Rémillard, B., Papageorgiou, N., and Soustra, F. (2010). Dynamic copulas. Technical Report G-2010-18, Gerad.
- Schweiser, M. (1995). Variance-Optimal Hedging in Discrete Time. *Mathematics of Operations Research*, 20(1) :10–32.
- Sharpe, W. (1992). Asset Allocation : Management Style and Performance Measurement. *Journal of Portfolio Management*, pages 7–19.
- Sklar, A. (1959). Fonctions de Répartition à n Dimensions et leurs Marges. *Publ. Inst. Université de Paris*, 8 :229–231.
- Tauschen, C. M. (1985). Diagnostic testing and evaluation of maximum likelihood models. *Journal of Econometrics*, 30 :415–443.
- White, H. (1985). Specification testing in dynamic models, in : Truman f. bewley, ed. *Advances in Econometrics, Fifth World Congress, Cambridge University press*, 2.