



HAL
open science

Contribution à l'allocation stratégique de portefeuille à l'aide de probabilités de retournement de tendances

Quentin Gai Gianetto

► **To cite this version:**

Quentin Gai Gianetto. Contribution à l'allocation stratégique de portefeuille à l'aide de probabilités de retournement de tendances. Statistiques [stat]. Télécom Bretagne, Université de Rennes 1, 2012. Français. NNT : 2012TELB0186 . tel-03973666

HAL Id: tel-03973666

<https://hal.science/tel-03973666>

Submitted on 4 Feb 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Sous le sceau de l'Université européenne de Bretagne

Télécom Bretagne

En habilitation conjointe avec l'Université de Rennes 1

Ecole Doctorale – MATISSE

Contribution à l'allocation stratégique de portefeuille à l'aide de probabilités de retournement de tendances

Thèse de Doctorat

Mention : Traitement du signal et télécommunications

Présentée par **Quentin Gai Gianetto**

Département : Image et traitement de l'information

Laboratoire : Lab-STICC Pôle : Connaissance Information et Décision (CID)

Directeur de thèse : M. Jean-Marc Le Caillec

Soutenue le 30 Janvier 2012

Jury :

- M. Alain Hillion, Directeur de recherche CNRS, Lab-STICC, Institut Télécom – *Président*
- M. Yves Rakotondratsimba, Professeur, École Centrale d'Électronique de Paris – *Rapporteur*
- M. Jean-Marc Vesin, Professeur, École Polytechnique Fédérale de Lausanne (Suisse) – *Rapporteur*
- M. Patrick Hénaff, Maître de conférences, Institut d'Administration des Entreprises de Paris – *Examineur*
- M. Erwan Marrec, Responsable gestion quantitative à Fédéral Finance Gestion, Brest – *Examineur*
- M. Jean-Marc Le Caillec, Professeur, Télécom Bretagne – *Directeur de thèse*

**Contribution à l'allocation stratégique de portefeuille à l'aide de
probabilités de retournement de tendances**

Quentin Gai Gianetto

Résumé

Cette thèse porte sur l'allocation stratégique d'un portefeuille d'investissement. L'enjeu est de définir un univers d'investissement composé de plusieurs classes d'actifs puis de mettre en place une stratégie d'allocation au sein de cet univers qui permette de maximiser l'espérance de rendement du portefeuille. Généralement, on suppose que les marchés répondent à une hypothèse d'efficience, que les investisseurs sont rationnels, que les coûts de transactions sont négligeables, etc. Dans ce cadre, la valeur d'un portefeuille correspondrait à la somme des valeurs de N actifs risqués, évoluant de manière totalement aléatoires, et d'un actif sans risque dont la valeur évolue de manière déterministe avec le temps. Néanmoins, cette approche, classique dans le domaine de la modélisation probabiliste d'un portefeuille, repose sur des hypothèses simplistes éloignées de la réalité. Dans cette thèse, on aborde l'allocation stratégique de portefeuille d'une manière différente. On considère que les marchés ne sont pas efficaces mais sont sensibles à des préoccupations d'ordre macroéconomique ou liées aux croyances des investisseurs. L'enjeu principal consiste à pouvoir mesurer de manière pertinente l'information macroéconomique et l'information liée aux croyances des investisseurs en temps réel. Il faut ensuite pouvoir incorporer cette information au sein de l'allocation du portefeuille d'investissement. Nous avons donc proposé des outils portant sur l'estimation et la prédiction de tendances pour des séries non-stationnaires, la synthèse de l'information disponible aux investisseurs et le lien qui peut être fait avec l'allocation stratégique du portefeuille.

Mots clés : Séries chronologiques, Représentation parcimonieuse, Analyse quasi-temps réel, Prédiction non-linéaire, Chaînes de Markov cachées, Allocation stratégique de portefeuille

Abstract

This thesis concerns the strategic asset allocation of a portfolio composed of several asset classes. The stake is to select a universe of investment and, next, to set up an allocation within this universe which allows to maximize the returns of the investment portfolio. Generally, it is supposed an efficient-market hypothesis, rational investors, negligible transaction costs, etc. In this ideal frame, the value of a portfolio would correspond to the sum of the values of N risky assets, evolving in a totally random way, and of the value of a risk-free asset which evolves in a deterministic way. Nevertheless, this approach, classic in the domain of the probabilistic modeling of a portfolio, bases on simplistic hypotheses which are remote from the reality. In this thesis, we approach the strategic asset allocation of a multi-asset portfolio in a different way. We consider that markets are not efficient but are sensitive to macroeconomic concerns or are connected to the beliefs of the investors. The main objective consists in being able to measure pertinently the macroeconomic information and the information connected to the real-time beliefs of the investors. It is then necessary to be able to incorporate this information within the portfolio allocation. We thus proposed tools concerning the estimation and the prediction of trends for non-stationary time series, the determination and the fusion of different kinds of information, and the link which can be made with the portfolio allocation.

Keywords : Time Series Analysis, Parsimonious Representations, Quasi-Real Time Analysis, Nonlinear Prediction, Hidden Markov Models, Strategic Asset Allocation

« Comprendre le monde pour un homme, c'est le réduire à l'humain »

Albert Camus
Le mythe de Sisyphe

Remerciements

Cette thèse CIFRE est le fruit d'une collaboration entre Télécom Bretagne et la société Fédéral Finance Gestion, filiale du Crédit Mutuel Arkéa. Je remercie tout d'abord mon directeur de thèse, M. Jean-Marc Le Caillec, de m'avoir proposé ce sujet de thèse suite à mon stage de Master 2. Je le remercie d'avoir suivi mon travail avec attention et de m'avoir toujours soutenu à mes moments de doutes. Je lui suis également reconnaissant pour sa disponibilité et sa gentillesse. Je tiens également à remercier M. Erwan Marrec, responsable de la gestion quantitative à Fédéral Finance Gestion, pour avoir encadré mon travail au sein de l'entreprise. Je le remercie pour son professionnalisme et pour m'avoir initié à la finance quantitative. Ses remarques sur les aspects pratiques de la finance de marché m'ont été utiles pour que cette thèse puisse correspondre aux attentes de l'entreprise.

Mes remerciements s'adressent également aux professeurs Yves Rakotondratsimba et Jean-Marc Vesin pour avoir accepté de rapporter ma thèse et pour l'intérêt qu'ils ont porté à mes travaux. Je vous remercie tout particulièrement pour l'ensemble des commentaires très constructifs que vous avez fait concernant ma thèse. Je remercie M. Yves Rakotondratsimba d'avoir mis en avant l'importance essentielle de la tendance en finance de marché, alors que celle-ci est généralement négligée dans les mathématiques financières, et M. Jean-Marc Vesin pour avoir souligné le peu d'utilisation de techniques issues du traitement du signal dans ce domaine. Ces deux remarques donnent tout son sens à ma thèse. Un grand merci également à M. Alain Hillion d'avoir accepté de présider le jury de ma soutenance ainsi qu'à M. Patrick Hénaff d'avoir accepté de faire partie de mon jury. Les commentaires que vous avez émis lors de ma soutenance ont permis des précisions essentielles et ont ouvert des pistes de recherche très intéressantes.

J'adresse également mes remerciements aux membres et aux thésards du laboratoire Image et Traitement de l'Information de Télécom Bretagne pour leur accueil et leur sympathie. Un merci spécial à Corinne Le Lann et à Frédéric Maussang pour leurs gentilleses. Je remercie Corinne pour s'être occupée des formalités administratives et Frédéric pour avoir accepté de partager son bureau avec moi : merci d'avoir pris le temps d'écouter mes pérégrinations.

J'associe à ces remerciements mes anciens collègues de travail à Fédéral Finance Gestion et en particulier ceux de la gestion quantitative : Erwan bien sûr mais aussi Cyril, Caroline, Pierre-Yves, Franck et Nicolas. Un merci spécial à Cyril et Caroline pour le temps que vous m'avez accordé et pour votre professionnalisme : cela a été un plaisir de travailler avec vous. Merci également à Jamal, thésard au sein de Fédéral Finance Gestion, avec qui j'ai pu échanger sur différents sujets de recherche durant ma thèse.

Je remercie également les membres du laboratoire IRMAR-INSA de Rennes qui m'ont permis d'avoir un environnement calme et serein pour terminer la rédaction de ma thèse. Sans cela, j'aurais sûrement eu plus de difficultés à la terminer.

Un grand merci aussi à mes parents, Isabelle et Dominique, pour m'avoir soutenu et réconforté lorsque je doutais, ainsi que ma soeur Ellen et mon frère Germain, pour l'affection qu'ils m'ont portés pendant ce travail. Je remercie également Riwal, Laetitia et Angelina pour leur amitié et leur hospitalité. Merci aussi à mes amis brestois : Camille, Mike et Anne-laure, Cindy et Mickael, Ific, Lolo et Sophie ; ainsi qu'à mes amis rennais : Salah, Baptiste, Jean-Louis, Clément, Leila, Gael, Jo, Yoann et Basile pour tous ces bons moments passés ensemble durant cette thèse.

Enfin, je remercie ma femme, Ilham, pour son soutien et ses conseils qui m'ont permis de terminer cette thèse.

Table des matières

Introduction générale	9
1 Introduction à la gestion de portefeuille	13
1.1 Investir dans un titre financier négociable	14
1.1.1 La formation du prix de marché	14
1.1.2 La modélisation probabiliste de l'incertitude liée à l'évolution d'un investissement	15
1.2 Sur l'hypothèse d'efficience des marchés	16
1.2.1 L'hypothèse classique d'efficience des marchés	16
1.2.2 Déterminer le prix d'équilibre d'un titre financier	18
1.2.3 Existence d'anomalies à l'efficience informationnelle des marchés actions	21
1.2.4 Biais comportementaux des investisseurs	23
1.3 La dynamique des prix de marché	26
1.3.1 Persistances et tendances dans les prix de marché	26
1.3.2 Répartition inconditionnelle des rendements	31
1.3.3 La dynamique des dépendances entre actifs	32
1.4 Univers d'investissement et gestion de portefeuille en pratique	34
1.4.1 Différentes classes d'actifs	34
1.4.2 L'analyse des prix de marché	36
1.4.3 Le processus d'allocation d'actifs	41
1.5 Conclusion	43
2 Optimisation de portefeuille et estimation de tendances	45
2.1 L'optimisation classique « moyenne-variance »	45
2.1.1 Optimisation « moyenne-variance »	46
2.1.2 Hypothèse d'ergodicité et propriétés des estimateurs en échantillons finis	49
2.2 Formalisation d'une stratégie d'investissement en temps discret	49
2.3 Définition de la stratégie optimale	52
2.4 Approximation de la stratégie optimale	55
2.5 Conclusion	58
3 Estimation non-paramétrique de tendances en gestion de portefeuille	59
3.1 Estimer une tendance dans le domaine temporel	60
3.1.1 Filtrage linéaire dans le domaine temporel	61

3.1.2	Filtrage dans le domaine fréquentiel	62
3.1.3	Filtrage linéaire optimum et nécessité de méthodes locales et adaptatives . . .	64
3.2	Problématiques de régression non-paramétrique	66
3.2.1	Le régressogramme	66
3.2.2	L'hypothèse de régularité de la tendance	67
3.2.3	La régression sur base de fonctions orthonormales	67
3.2.4	Estimation du paramètre de lissage	69
3.3	Méthodes non-paramétriques d'estimation de tendances en pratique	71
3.3.1	Techniques de filtrage non-paramétrique classiques en économie et finance . . .	72
3.3.2	La régression non-paramétrique polynomiale locale	74
3.3.3	La régression à l'aide de fonctions splines	76
3.3.4	La régression en utilisant des bases d'ondelettes	78
3.3.5	La régression en utilisant une décomposition modale empirique	81
3.4	Comparaison des différentes méthodes	83
3.4.1	Critères de comparaison	84
3.4.2	Modèles utilisés pour la génération des données	85
3.4.3	Méthodes d'estimation de tendances utilisées	86
3.4.4	Interprétation des résultats	86
3.5	Conclusion	91
4	Prédiction non-linéaire du risque lié au marché action en utilisant différents indicateurs	95
4.1	Présentation d'indicateurs du risque lié au marché action	96
4.1.1	Indicateurs liés à la macroéconomie	96
4.1.2	Indicateurs liés aux croyances des investisseurs	100
4.1.3	Étude de la dépendance avec l'évolution du marché action	106
4.2	Extraire une composante chaotique sous-jacente à l'évolution des indicateurs de risque	108
4.2.1	Sur la prédictibilité d'une série temporelle	108
4.2.2	Exposant maximal de Lyapunov et horizon de prédiction	113
4.2.3	Tester l'imprévisibilité d'une série temporelle	114
4.2.4	Estimation non-linéaire de la composante chaotique	121
4.3	Prédiction non-linéaire des composantes chaotiques	123
4.3.1	Prédiction d'une seule composante chaotique	123
4.3.2	Prédiction de plusieurs composantes chaotiques	130
4.4	Correction des effets de bord	133
4.4.1	Formulation du problème	133
4.4.2	Prédiction de la révision de l'estimation à l'aide d'un processus ARMA(p, q) . .	135
4.5	Application à des indicateurs liés aux croyances des investisseurs	136
4.6	Conclusion	137
5	Construire un indicateur synthétique de l'augmentation de facteurs de risque à partir d'une chaîne de Markov cachée	141
5.1	Coder qualitativement les indicateurs	142
5.2	Fusionner les indicateurs à l'aide d'une chaîne de Markov cachée	143
5.2.1	L'approche de Grégoir et Lenglar	143
5.2.2	Présentation du modèle de chaîne de Markov cachée	144

5.2.3	Présentation de l'algorithme d'Espérance-Maximisation	145
5.2.4	Étape d'espérance et algorithme forward-backward	146
5.2.5	Étape de maximisation	147
5.2.6	Insertion d'une étape Stochastique	148
5.2.7	Construction d'un indicateur synthétique d'augmentation des indicateurs . . .	149
5.3	Applications et résultats	150
5.3.1	Construction d'une horloge d'investissement	150
5.3.2	Construction d'un indicateur de retournement du risque lié aux croyances des investisseurs	154
5.4	Conclusion	154
6	Allocation stratégique de portefeuille à partir de stratégies momentums réglées par une horloge d'investissement	159
6.1	Sélectionner un univers d'investissement diversifié	160
6.1.1	La diversification	160
6.1.2	Classification ascendante hiérarchique des actifs	162
6.1.3	Détermination d'un portefeuille composé d'un nombre minimal de classes nécessaires à sa diversification	164
6.2	Stratégies « momentums » réglées par une horloge d'investissement	170
6.2.1	Présentation d'une stratégie momentum classique	170
6.2.2	Allocation en fonction d'une horloge d'investissement	174
6.2.3	Allocation par le « momentum »	175
6.2.4	Backtesting de l'allocation stratégique proposée	178
6.3	Conclusion	180
	Conclusion générale	181

Introduction générale

Motivations et contexte

Le sujet de cette thèse concerne l'allocation stratégique d'un portefeuille d'investissement. L'enjeu est de définir un univers d'investissement composé de plusieurs classes d'actifs puis de mettre en place une stratégie d'allocation au sein de cet univers qui permette de maximiser l'espérance de rendement du portefeuille. Généralement, il est supposé que les marchés sont efficients. Dans ce cadre, la valeur d'un portefeuille correspond à la somme des valeurs de N actifs risqués, évoluant de manière totalement aléatoires, et d'un actif sans risque dont la valeur évolue de manière déterministe avec le temps. L'approche classique consiste à définir l'allocation du portefeuille en maximisant son espérance de rendement tout en minimisant sa variance ou, tout du moins, une mesure du risque liée à un investissement dans ce portefeuille. Dans cette thèse, nous abordons l'allocation stratégique de portefeuille d'une manière différente. Les marchés ne sont pas considérés efficients mais sensibles à des préoccupations d'ordre macroéconomique ou liées aux croyances des investisseurs. La problématique consiste alors à définir des « outils » adéquats afin de mesurer de manière pertinente l'information macroéconomique et l'information liée aux croyances des investisseurs en temps réel. Ces outils doivent permettre d'incorporer cette information au sein de l'allocation du portefeuille d'investissement.

Objectif et contributions

L'objectif de cette thèse est de fournir des outils adéquats permettant de prédire et synthétiser l'information disponible aux investisseurs en temps réel afin qu'ils puissent optimiser leur stratégie d'investissement. La première étape a été de sélectionner un certain nombre d'indicateurs susceptibles de fournir une information utile pour les investisseurs. Nous avons considéré que cette information était utile lorsque l'indicateur avait une signification propre et qu'il possédait une corrélation significative avec l'évolution des marchés actions. Une fois que ces indicateurs ont été choisis, nous nous sommes intéressés à la prédiction de leurs directions futures.

Pour cela, il faut pouvoir extraire une tendance au sein de l'évolution de ces indicateurs. Nous avons également mis en évidence que l'estimation d'une composante prévisible au sein des rendements des investissements du portefeuille est essentielle pour définir la stratégie optimale à adopter. L'estimation de tendances est donc un enjeu crucial en gestion de portefeuille.

Dans ce but, nous avons testé différentes méthodes non-paramétriques d'estimation de tendances sur des séries générées par des modèles proches de ceux qu'il est possible d'inférer d'après des séries

financières réelles. Il est apparu que la décomposition modale empirique est un outil intéressant pour ce type d'estimation. Néanmoins, la tendance ainsi estimée semble parfois « trop proche » des données : elle semble imprévisible. Afin de s'assurer que la tendance estimée soit prévisible, nous avons développé un test statistique de l'hypothèse d'imprévisibilité d'une série temporelle. Ce test peut alors servir de seuil lors de l'estimation non-linéaire de la tendance. Concernant l'étape de prédiction proprement dite, nous avons exploité le comportement chaotique de la tendance estimée pour pouvoir la prédire en utilisant une régression non-paramétrique polynomiale locale. Nous avons plus particulièrement étendu cette méthode de prédiction au cas multivarié afin de prédire des ensembles de variables. Ainsi, il n'est pas nécessaire de supposer que ces variables évoluent de manière indépendante entre elles lors de la prédiction. Une autre contribution de cette thèse a été le développement d'une méthode originale afin de construire un indicateur synthétique d'augmentation « globale » d'un ensemble de facteurs de risque. Cet indicateur permet de distinguer des phases de hausses et de baisses généralisées de ces facteurs. Cette méthode se base sur des chaînes de Markov cachées multivariées qualitatives.

Nous nous sommes également intéressé au concept de diversification de portefeuille. Nous avons proposé une méthode originale basée sur une classification ascendante hiérarchique de différentes classes d'actifs. Cette méthode se fonde sur l'évolution plus ou moins corrélée de la tendance des différents actifs considérés. La méthode de diversification est ensuite construite en deux étapes : détermination d'un nombre de classes permettant de réduire suffisamment la partie diversifiable de la variance du portefeuille puis sélection des actifs parmi les classes retenues en minimisant la partie non-diversifiable de la variance du portefeuille.

Enfin, nous avons proposé une allocation stratégique de portefeuille prenant en compte les moments des différentes classes d'actifs mais également le contexte économique synthétisé à partir de nos indicateurs.

Organisation de la thèse

Le plan de cette thèse est le suivant. Le premier chapitre est une introduction à la gestion de portefeuille. Nous revenons sur l'hypothèse d'efficience des marchés et nous présentons un certain nombre de faits contraires à cette hypothèse.

Dans le second chapitre, nous nous intéressons à la stratégie d'investissement optimale à adopter entre deux instants de rebalancement du portefeuille lorsqu'un investisseur est soumis à un certain nombre de contraintes. Nous montrons que l'estimation de tendances déterministes localisées dans le temps est essentielle pour approximer la stratégie optimale.

Dans le troisième chapitre, nous exposons un certain nombre de méthodes d'estimations non-paramétriques de tendances au sein de séries temporelles que nous comparons sur des séries générées par des modèles économétriques proches de ceux qu'il est possible d'inférer à partir de séries de prix ou de rendements de prix.

Dans le quatrième chapitre, nous nous intéressons à des indicateurs macroéconomiques et à des indicateurs liés aux croyances des investisseurs. Ces indicateurs sont généralement qualifiés de « facteurs de risque ». Nous proposons une manière originale d'estimer la composante « prévisible » au sein de ces facteurs de risque. Nous présentons également une méthode de prédiction non-linéaire applicable sur un ensemble de tels indicateurs.

Dans le cinquième chapitre, nous proposons de synthétiser l'ensemble des prédictions réalisées au sein d'un même indicateur en utilisant une méthode originale basée sur des chaînes de Markov

cachées qualitatives.

Le sixième et dernier chapitre se consacre à l'allocation stratégique de portefeuille proprement dite. Dans un premier temps, nous réfléchissons au choix d'un univers d'investissement suffisamment diversifié. Nous proposons une méthode de sélection basée sur une classification ascendante hiérarchique et sur l'indépendance de l'évolution des tendances des rendements. Nous nous intéressons ensuite à la construction de stratégies d'allocation basées sur des facteurs de risques. Nous proposons d'allouer stratégiquement un portefeuille en prenant en compte le cycle macroéconomique tel que mesuré par nos indicateurs tout en utilisant les momentums des différentes classes d'actifs.

Chapitre 1

Introduction à la gestion de portefeuille

Sommaire

1.1	Investir dans un titre financier négociable	14
1.1.1	La formation du prix de marché	14
1.1.2	La modélisation probabiliste de l'incertitude liée à l'évolution d'un investissement	15
1.2	Sur l'hypothèse d'efficience des marchés	16
1.2.1	L'hypothèse classique d'efficience des marchés	16
1.2.2	Déterminer le prix d'équilibre d'un titre financier	18
1.2.3	Existence d'anomalies à l'efficience informationnelle des marchés actions	21
1.2.4	Biais comportementaux des investisseurs	23
1.3	La dynamique des prix de marché	26
1.3.1	Persistances et tendances dans les prix de marché	26
1.3.2	Répartition inconditionnelle des rendements	31
1.3.3	La dynamique des dépendances entre actifs	32
1.4	Univers d'investissement et gestion de portefeuille en pratique	34
1.4.1	Différentes classes d'actifs	34
1.4.2	L'analyse des prix de marché	36
1.4.3	Le processus d'allocation d'actifs	41
1.5	Conclusion	43

Introduction

L'objectif de ce chapitre est de proposer une introduction aux problématiques liées à la gestion de portefeuilles composés de plusieurs titres financiers négociables. Le chapitre est composé de cinq parties. La première partie explique comment se forme un prix de marché puis comment modéliser, à l'aide d'une hypothèse probabiliste, l'incertitude issue d'un investissement dans un titre. Dans une seconde partie, nous nous intéressons à la dynamique d'un investissement dans un portefeuille de titres. Afin de mieux cerner cette dynamique, nous introduisons la théorie néo-classique d'efficience des marchés. Nous exposons ensuite diverses anomalies à cette théorie ainsi que différentes critiques qui lui sont souvent faites dans la littérature. L'exposé de cette théorie et de ses critiques permet

de mieux comprendre la complexité de la modélisation du processus d'investissement dans un titre financier négociable. Après cet aspect théorique, la troisième partie du chapitre se concentre sur les études statistiques de la dynamique des prix. Nous y abordons les différents types de dépendances constatées au sein de l'évolution des prix et discutons de l'existence d'un cycle de marché. La quatrième partie du chapitre traite de la gestion de portefeuille en pratique. Différentes classes d'actifs, souvent présentes dans les portefeuilles de gérants de fonds, sont alors introduites. L'analyse de ces actifs s'opère le plus souvent de deux façons distinctes : soit par une analyse fondamentale, soit par une analyse technique. Ces deux sortes d'analyse des titres sont discutées puis comparées. Une allocation des titres dans le portefeuille est généralement effectuée suite à cette phase d'analyse. Enfin, la dernière partie du chapitre correspond à notre conclusion.

1.1 Investir dans un titre financier négociable

Lorsqu'on investit dans un titre financier négociable, on ne connaît pas la valeur future de cet investissement. Cette valeur dépend à la fois des flux futurs que l'actif versera (dividendes, intérêts,...) et de la plus ou moins-value que l'investisseur pourra en tirer. Les flux futurs versés par l'actif peuvent être précisés ou non lors de l'émission du titre tandis que la plus ou moins-value dépend de l'évolution du prix de marché. Ce prix étant issu d'un équilibre entre offre et demande, il est très difficile de l'anticiper de manière exacte au vu de la multiplicité des agents et des causes pouvant influencer sur cet équilibre. Afin de modéliser la dynamique d'un investissement, on suppose généralement que sa valeur correspond à la réalisation d'une variable aléatoire, l'évolution de cette valeur correspondant à un processus stochastique.

Dans ce premier paragraphe, nous introduisons différents concepts permettant d'utiliser l'hypothèse probabiliste dans la gestion d'un portefeuille¹ composé de titres financiers négociables. En premier lieu, nous expliquons de quelle manière se forme le prix de marché d'un tel titre. La modélisation probabiliste permettant de quantifier l'incertitude liée aux rendements d'un investissement dans ce type d'actif est ensuite présentée.

1.1.1 La formation du prix de marché

Un titre financier négociable est un contrat offrant des garanties de gains futurs suivant certaines conditions définies au préalable. Ainsi, un investisseur détenant une obligation à taux fixe² touchera des intérêts définis lors de l'émission de cette obligation. Mais ces actifs financiers possèdent également une valeur intrinsèque variant avec le temps. Cette valeur, généralement appelée valeur « fondamentale », correspond à l'actualisation des flux futurs versés par le titre. Chaque investisseur détenant ou souhaitant détenir un titre effectue sa propre actualisation de cette valeur fondamentale. S'il décide d'acquérir une position dans un titre suite à cette actualisation, il passe un ordre de bourse dans lequel il spécifie, entre autres, le prix auquel il souhaite acheter une certaine quantité de cet actif financier. Cet ordre est alors répertorié dans un carnet d'ordre spécifique à l'actif dans lequel l'ensemble de l'offre et de la demande concernant le titre sont confrontés (voir Figure 1.1). Ce processus dynamique d'interaction entre ordres d'achat et ordres de vente entraîne l'évolution du prix de marché du titre.

1. Un portefeuille est une collection d'investissements financiers

2. Voir paragraphe 4.1.2.

BUY ORDERS		SELL ORDERS	
SHARES	PRICE	SHARES	PRICE
100	37.3600	300	37.4500
100	37.3500	400	37.5000
600	37.2300	1100	37.5100
1000	37.2000	1000	37.5200
100	37.0600	100	37.5400
1000	36.8000	200	37.6100
190	36.7500	100	37.7300
507	36.5100	5	37.7600
3000	36.4600	350	37.7900
1000	36.4500	1000	37.8000
72 Buy Orders		118 Sell Orders	

FIGURE 1.1 – Ensemble des ordres à cours limité présents dans le carnet d’ordres d’INET pour le titre EBAY, côté sur le NASDAQ. Les deux premières (resp. dernières) colonnes représentent les ordres d’achat (resp. de vente) à cours limité. À la date d’observation, un investisseur souhaitant vendre (resp. acheter) 100 titres peut soumettre indifféremment un ordre au marché ou un ordre à la meilleure limite. Cela déclenchera une transaction avec le meilleur ordre d’achat (resp. de vente) présent dans le carnet. On l’identifie sur la première ligne : le prix de vente (resp. d’achat) correspondant sera ainsi de 37,36\$ (resp. 37,45\$). (Exemple issu de [Moi(2008)])

Le prix de marché d’un titre est donc issu du principe économique fondamental de l’offre et de la demande. Ce principe établit que le prix d’un titre et la quantité échangée sont déterminés par l’intersection des courbes d’offre et de demande, où la courbe de demande représente l’ensemble des quantités désirées par les acheteurs pour différents prix et la courbe d’offre représente l’ensemble des quantités que les vendeurs sont enclin à offrir en fonction de ces prix. L’intersection des deux courbes détermine un équilibre, un couple (*quantité, prix*), satisfaisant à la fois les acheteurs et les vendeurs simultanément. La courbe de demande correspond donc à l’agrégation de différents désirs individuels d’acheteurs, chacun ayant optimisé son choix par rapport à une contrainte de budget dépendant des prix mais aussi d’autres facteurs qui peuvent être d’ordres économiques, monétaires, psychologiques, etc... De la même façon, la courbe d’offre est l’agrégation de ce que chaque vendeur du titre est enclin à offrir, chaque vendeur étant soumis à des contraintes de ressources dépendant des prix mais également d’autres facteurs. Le point d’équilibre entre offre et demande au sein du carnet d’ordre spécifique au titre détermine le prix de marché, également appelé prix d’équilibre, du titre financier.

1.1.2 La modélisation probabiliste de l’incertitude liée à l’évolution d’un investissement

À un instant donné, chaque titre financier se distingue à l’aide de deux caractéristiques essentielles : la durée d’investissement souhaitée et le devenir supposé du capital investi dans ce titre sur cette période. Dans cette optique, on définit le rendement d’un investissement financier entre deux instants t et $t + \Delta t$ ($\Delta t > 0$ constant), noté $R_{t,t+\Delta t}$, comme le pourcentage d’évolution entre la valeur de l’investissement (ou capital investi) à l’instant t , W_t , et la valeur de l’investissement à l’instant $t + \Delta t$, $W_{t+\Delta t}$,

$$W_{t+\Delta t} = (1 + R_{t,t+\Delta t})W_t \Leftrightarrow R_{t,t+\Delta t} = \frac{W_{t+\Delta t}}{W_t} - 1 \quad \forall t \in I \subseteq \mathbb{R}$$

Nous appellerons Δt l’horizon d’investissement.

En général, le gestionnaire de portefeuille considère que la meilleure façon d'estimer le rendement à venir est de déterminer la valeur la plus probable de ce rendement. La dynamique du rendement est alors modélisée sous une hypothèse probabiliste. On définit ainsi un espace de probabilité complet $(\Omega, \mathcal{F}, \mathbb{P})$ dans lequel Ω représente l'ensemble de tous les historiques de rendements possibles et où \mathbb{P} est une mesure de probabilité liée à l'observation historique des rendements et de leurs fréquences. Pour Δt fixé, les rendements observés sont alors assimilés à la réalisation d'une variable aléatoire $R_{t,t+\Delta t}$ telle que :

$$R_{t,t+\Delta t}(\omega) : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

$\mathcal{B}(\mathbb{R})$ représentant la tribu Borélienne de \mathbb{R} . \mathcal{F} est une σ -algèbre représentant l'ensemble de l'information disponible aux investisseurs.

Cette dernière joue un rôle prépondérant dans la quantification de l'incertitude liée aux rendements futurs. Elle représente la capacité de cet investisseur à faire une distinction entre tous les historiques de rendements possibles. Dans un cadre probabiliste, on peut considérer que cette information est contenue dans la tribu des rendements observables jusqu'à l'instant présent t . Elle est alors modélisée en munissant l'espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ d'une filtration $(\mathcal{F}_t)_{t \in I}$. Cette dernière permet de formaliser l'évolution de l'information disponible au fur et à mesure du temps. En supposant que le rendement entre t et $t + \Delta t$ ne dépend que du passé et non du futur, il est logique de supposer que $R_{t,t+\Delta t}$ est \mathcal{F}_t -mesurable. On dit alors que $(R_{t,t+\Delta t})_{t \in I}$ est adapté à la filtration $(\mathcal{F}_t)_{t \in I}$.

La famille $(R_{t,t+\Delta t}, \mathcal{F}_t)_{t \in I}$ est appelée processus stochastique. Lorsque la référence à $(\mathcal{F}_t)_{t \in I}$ est claire, on simplifie cette notation en notant ce processus $(R_{t,t+\Delta t})_{t \in I}$. On peut considérer que $(R_{t,t+\Delta t}, \mathcal{F}_t)_{t \in I}$ est une fonction de deux variables :

$$R : \Omega \times I \rightarrow \mathbb{R}, (\omega, t) \mapsto R_{t,t+\Delta t}(\omega)$$

En conséquence, on considérera l'évènement $\{R_{t,t+\Delta t}(\omega) \in A\}$ pour $A \in \mathcal{B}(\mathbb{R})$ comme presque sûr si $\mathbb{P}(R_{t,t+\Delta t}(\omega) \in A) = 1$ et comme impossible si $\mathbb{P}(R_{t,t+\Delta t}(\omega) \in A) = 0$.

Jusqu'au début des années 80, de nombreuses études empiriques ont laissé supposer que le processus stochastique des prix était proche d'une marche aléatoire, le processus des rendements de prix étant donc assimilé à un bruit indéterministe. Ces constatations donnèrent naissance à la théorie d'efficience des marchés.

1.2 Sur l'hypothèse d'efficience des marchés

Dans ce second paragraphe, nous présentons l'hypothèse classique d'efficience des marchés. Nous présentons différents « faits stylisés » remettant en cause cette hypothèse ainsi que différents biais comportementaux des investisseurs contraires à l'hypothèse de rationalité.

1.2.1 L'hypothèse classique d'efficience des marchés

Sur les marchés financiers, on distingue trois types d'efficience : l'efficience allocationnelle, l'efficience opérationnelle et l'efficience informationnelle. L'efficience *allocationnelle* correspond à la capacité du marché à allouer de manière optimale les ressources financières. L'efficience *opérationnelle* désigne une situation où les intermédiaires financiers mettent en relation des acheteurs et des vendeurs de titres de manière optimale, c'est-à-dire pour un coût le plus faible possible. Toutefois, ces

deux types d'efficience dépendent d'une troisième sorte d'efficience : l'efficience *informationnelle* du marché.

Les origines de cette efficience informationnelle peuvent être attribuées à [Sam(1965)] dans un article nommé « *Proof that Properly Anticipated Prices Fluctuate Randomly* » ainsi qu'à Fama ([Fam1(1965)], [Fam2(1965)], [Fam(1970)]) : lorsque les rendements de prix sont correctement anticipés, c'est-à-dire lorsqu'ils incorporent l'information et les espérances de tous les participants du marché de manière instantanée, ils sont imprévisibles. La théorie d'efficience des marchés de titres se base donc essentiellement sur une dimension informationnelle puisque l'allocation optimale des ressources ne peut se faire que si les prix intègrent toute l'information disponible tandis que la réalisation de transactions au meilleur coût dépend de la disponibilité de cette information.

Cette hypothèse a connu un succès à la fois empirique et théorique important jusqu'au milieu des années 80, avant qu'un certain nombre d'études ne viennent la remettre en cause. Elle repose sur l'idée que les participants actifs du marché espèrent accumuler des gains à partir de l'information qu'ils perçoivent. Pour cela, ils vont profiter des plus petits avantages procurés par cette information. En incorporant de manière précipitée celle-ci dans les prix du marché, ils vont rapidement éliminer les opportunités d'investissement qui ont motivées leurs échanges. Si ceci arrive instantanément, ce qui peut être le cas dans un monde idéalisé de marchés sans coûts de transactions, alors les prix reflèteraient toujours et entièrement toute l'information disponible aux investisseurs. Par conséquent, les rendements de prix seraient entièrement aléatoires et imprévisibles. Il serait donc impossible d'accumuler des profits anormaux puisqu'il n'existerait pas de possibilités d'arbitrage. Dans ce contexte, seule une stratégie auto-financée (ou *buy-and-hold*), consistant à investir un certain capital et à ne pas rebalancer les positions détenues, pourrait être optimale. En 1970, Fama fait évoluer cette hypothèse en en déclinant trois formes [Fam(1970)]. La forme forte de l'efficience de marché consiste à affirmer que le prix reflète toute l'information disponible qu'elle soit publique³ ou privée⁴, la forme semi-forte que les prix reflètent toute l'information publique disponible puis, la forme faible, que les prix reflètent l'historique des prix passés.

L'hypothèse d'efficience du marché suppose donc que les investisseurs sont « rationnels », dans le sens où leur bien-être, représenté par une fonction d'utilité⁵, est optimisé selon les opportunités d'investissement détectées. En d'autres termes, l'individu cherche à atteindre des objectifs prédéterminés très spécifiques de la manière la plus sûre avec le moins possible de coûts. Certains, comme [Shl(2000)] et [Mal(2003)], affirment qu'il n'est pas nécessaire que tous les participants du marché soient rationnels pour que le marché soit efficient. Dans le cas où certains investisseurs auraient des comportements irrationnels dans leur perception de l'information et les décisions qui en découlent, les marchés pourraient tout de même tendre vers cette efficience. En effet, on pourrait imaginer que les décisions irrationnelles des investisseurs sont inversement corrélées, auquel cas une décision d'un investisseur irrationnel compenserait alors la décision d'un autre. L'ensemble des investisseurs irrationnels du marché n'aurait ainsi pas d'impact sur l'équilibre des prix. Si l'irrationalité devient commune à un grand groupe d'investisseurs agissant de manière corrélée et étant ainsi capables d'écarter le prix de la valeur fondamentale du titre⁶, on pourrait alors supposer l'existence d'« arbitrageurs » rationnels

3. c'est-à-dire disponible à l'ensemble des investisseurs.

4. c'est-à-dire disponible à un investisseur mais pas aux autres.

5. Voir chapitre 2.

6. Voir paragraphe 4.2.1.

qui repèreraient rapidement l'écart du prix et agiraient en conséquence. En vendant un actif sur-évalué sur un marché et en achetant un actif similaire sous-évalué sur un autre marché, ils créeraient une dynamique faisant revenir les prix de marché vers leurs valeurs fondamentales. L'hypothèse supposerait donc simplement qu'il existe de nombreux arbitrageurs rationnels qui agissent de manière rapide et sans aucune contrainte.

L'hypothèse d'efficience des marchés a été étendue dans plusieurs autres directions, incluant par exemple de l'information asymétrique⁷ [Rah(2001)] ou bien encore des coûts de transactions [Jen(1978)]. Mais les grandes lignes restent les mêmes : des investisseurs individuels forment des espérances de manière rationnelle, les marchés agrègent l'information de manière efficiente et l'équilibre des prix incorpore toute l'information disponible.

En pratique, il est difficile de tester l'hypothèse d'efficience du marché puisqu'on ne connaît pas la valeur fondamentale du titre. En effet, supposons que le prix d'équilibre d'un titre à l'instant t , noté P_t , s'écrit comme l'attente rationnelle d'une valeur fondamentale V^* conditionnellement à l'information disponible en t :

$$P_t = \mathbb{E}[V^* | \mathcal{F}_t]$$

En $t + \Delta t$, on aurait $P_{t+\Delta t} = \mathbb{E}[V^* | \mathcal{F}_{t+\Delta t}]$. Par conséquent, l'espérance du changement de prix entre t et $t + \Delta t$ s'écrirait comme :

$$\mathbb{E}[P_{t+\Delta t} - P_t | \mathcal{F}_t] = \mathbb{E}[\mathbb{E}[V^* | \mathcal{F}_{t+\Delta t}] - \mathbb{E}[V^* | \mathcal{F}_t] | \mathcal{F}_t] = \mathbb{E}[V^* | \mathcal{F}_t] - \mathbb{E}[V^* | \mathcal{F}_t] = 0$$

Nous avons ici supposé que la valeur fondamentale du titre était constante. On aurait tout aussi bien pu supposer qu'elle fut dépendante du temps, auquel cas nous aurions eu :

$$\mathbb{E}[P_{t+\Delta t} - P_t | \mathcal{F}_t] = \mathbb{E}[V_{t+\Delta t}^* | \mathcal{F}_t] - \mathbb{E}[V_t^* | \mathcal{F}_t]$$

ce qui n'est pas forcément nul ou constant. L'espérance conditionnelle des variations de prix peut ainsi correspondre à celle des variations de la valeur fondamentale du titre. On voit donc qu'un test de l'hypothèse d'efficience basé sur les prix nécessite de choisir un modèle de détermination du prix d'équilibre P_t .

Dans le paragraphe suivant, nous présentons brièvement une méthode de détermination du prix d'équilibre fréquemment utilisée et consistant à évaluer les flux futurs versés par un actif. Nous présentons également deux célèbres modèles permettant de caractériser l'évolution du prix d'équilibre : le « MEDAF » et l'« APT ».

1.2.2 Déterminer le prix d'équilibre d'un titre financier

Méthode d'actualisation des flux futurs versés par un titre

Cette méthode d'évaluation des titres repose sur l'idée qu'il est possible d'anticiper les flux futurs que l'actif nous rapportera. Supposons que le rendement d'un investissement dans un titre financier s'écrive formellement comme :

$$R_{t,t+\Delta t} = \frac{P_{t+\Delta t} + F_{t,t+\Delta t}}{P_t} - 1$$

7. L'information asymétrique désigne une situation où certains investisseurs ont une information pertinente et d'autres non.

avec P_t le prix de marché à l'instant t et $F_{t,t+\Delta t}$ l'ensemble des flux, autres que le prix de revente de cet actif, versés par le titre sur l'intervalle $]t, t + \Delta t]$. Ainsi, il semble logique d'écrire :

$$\widehat{P}_t = \mathbb{E} \left[\frac{P_{t+\Delta t} + F_{t,t+\Delta t}}{R_{t,t+\Delta t} + 1} \middle| \mathcal{F}_t \right]$$

avec \widehat{P}_t le prix actualisé à l'instant t . On suppose couramment $\mathbb{E}[R_{t,t+\Delta t} | \mathcal{F}_t]$ égale à une constante k , également appelée taux d'actualisation. Auquel cas, sur $n \times \Delta t$ périodes de temps futurs,

$$\widehat{P}_t = \mathbb{E} \left[\frac{P_{t+n \times \Delta t}}{(1+k)^{n \times \Delta t}} \middle| \mathcal{F}_t \right] + \mathbb{E} \left[\frac{F_{t,t+n \times \Delta t}}{(1+k)^{n \times \Delta t}} \middle| \mathcal{F}_t \right]$$

En général, il est raisonnable de supposer une condition de transversalité :

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\frac{P_{t+n \times \Delta t}}{(1+k)^{n \times \Delta t}} \middle| \mathcal{F}_t \right] = 0$$

La valeur fondamentale d'un titre serait ainsi égale à la valeur actualisée des flux futurs qu'il versera :

$$\widehat{P}_t = \lim_{n \rightarrow \infty} \frac{\mathbb{E}[F_{t,t+n \times \Delta t} | \mathcal{F}_t]}{(1+k)^{n \times \Delta t}}$$

Cette dernière relation est difficilement applicable en pratique puisqu'elle requiert une prévision infinie des flux versés. Néanmoins, elle met en avant l'importance de la dynamique des flux versés par l'actif, autres que son prix de revente, dans l'évaluation de sa valeur fondamentale. Pour le marché action, ces flux correspondent généralement aux dividendes ou aux "free cash flow" (voir par exemple [Mye(1994)]). Plusieurs modèles mathématiques d'évaluation de la valeur fondamentale d'une action ont ainsi été proposés (modèle de [Gor(1956)], de Bates, de Molodovski,...). Pour le marché obligataire, ces flux correspondent généralement à des intérêts perçus (voir par exemple [Kal(1993)]).

Les modèles d'actualisation des flux futurs sont les plus utilisés mais il existe d'autres méthodes d'évaluation du prix d'équilibre. Ainsi, on peut supposer que la valeur fondamentale d'une action dépend uniquement de ce que l'entreprise possède à l'instant présent. Suivant cette approche, on peut se contenter d'analyser différentes quantités issues du bilan comptable d'une entreprise afin d'évaluer la valeur fondamentale de l'action correspondante. Une autre méthode consiste à procéder par une approche comparative. On suppose, par exemple, que l'évolution d'une action d'une entreprise appartenant à un secteur particulier est similaire à celles d'autres entreprises appartenant au même secteur.

Les deux prochains modèles présentés ne traitent pas de l'évaluation du prix d'équilibre mais plutôt de l'évaluation de son évolution.

Modèle d'évaluation des actifs financiers

Afin de caractériser l'évolution du prix d'équilibre, [Sha(1964)], [Lin(1965)] et [Mos(1966)] ont proposés un « modèle d'évaluation des actifs financiers » (le MEDAF ou *CAPM* en anglais). Celui-ci repose sur plusieurs hypothèses assez restrictives :

1. Les investisseurs sont averses au risque et évaluent un portefeuille de titres dans un cadre espérance-variance des rendements. Ils maximisent l'utilité de leur richesse de fin de période⁸.
2. Les actifs présents sur les marchés sont parfaitement divisibles.
3. Les marchés sont sans frictions (il n'existe pas de coûts de transaction).
4. Les marchés de titres sont parfaits⁹.
5. Tous les investisseurs sont rationnels, il n'existe que de "petits" investisseurs (aucun investisseur ne possède un portefeuille dont la composition affecte à tout instant le prix de marché) et les investisseurs sont libres de leurs décisions.
6. Les investisseurs ont accès aux mêmes opportunités d'investissement et les anticipations de rendements faites par les investisseurs sont identiques pour tous (homogènes).

Sous ces hypothèses, l'espérance du rendement en excès d'un actif financier A peut s'exprimer comme :

$$\mathbb{E}[R_{t,t+\Delta t}^A] - r_F = \beta_A \times (\mathbb{E}[R_{t,t+\Delta t}^M] - r_F)$$

où $R_{t,t+\Delta t}^A$ est le rendement du prix de marché de l'actif A entre t et $t+\Delta t$, r_F le rendement sans risque, $R_{t,t+\Delta t}^M$ le rendement du portefeuille de marché¹⁰ et $\beta_A = \frac{\text{Cov}[R_{t,t+\Delta t}^A, R_{t,t+\Delta t}^M]}{\mathbb{V}[R_{t,t+\Delta t}^M]}$; $\text{Cov}[R_{t,t+\Delta t}^A, R_{t,t+\Delta t}^M]$ symbolisant la covariance entre $R_{t,t+\Delta t}^A$ et $R_{t,t+\Delta t}^M$ et $\mathbb{V}[R_{t,t+\Delta t}^M]$ la variance de $R_{t,t+\Delta t}^M$.

Si ce modèle possède l'avantage d'être relativement simple, on ne peut le tester empiriquement puisque les rendements du portefeuille de marché sont inobservables dans la réalité. Il existe cependant de nombreuses études empiriques utilisant des approximations de ces rendements du marché (avec des indices de marché par exemple). Les résultats concluent généralement sur son incapacité à expliquer de manière fiable la dynamique des rendements de prix [Fam(2003)]. Ces mauvais résultats peuvent s'expliquer soit par les nombreuses hypothèses faites par le modèle, soit par des approximations imparfaites du portefeuille de marché.

L'« Arbitrage Pricing Theory » (ou APT)

Un autre type de modèle offrant une alternative au MEDAF et appelé *Arbitrage Pricing Theory* (ou APT) a pris une place conséquente dans la littérature financière. L'APT se base notamment sur une absence d'opportunité d'arbitrage, ce qui correspond à une exigence plus faible que la maximisation de fonctions d'utilité. Il fut introduit par [Ros(1976)] puis présenté comme une alternative au MEDAF pouvant être testée empiriquement (voir [Rol(1977)] et [Rol(1980)]). Ce modèle repose sur les hypothèses 2. à 6. précédemment évoquées pour le MEDAF, additionnées de celles-ci :

1. Le rendement d'un actif correspond à une combinaison linéaire de facteurs de risque systématique décorrélés, inconnus et dépendant du temps, additionnée d'un résidu.
2. Il n'existe pas d'opportunité d'arbitrage.

8. Voir chapitre 2.

9. Un marché parfait est un marché où l'on peut acheter et vendre sans restriction de quantité et où il n'existe pas de taxes.

10. Portefeuille équipondéré de l'ensemble des actifs risqués disponibles sur le marché.

L'absence d'opportunité d'arbitrage signifie qu'il ne peut exister de configurations de prix telles qu'un investisseur puisse acquérir une richesse certaine sans mise de départ. Formellement, ce modèle exprime le rendement entre t et $t + \Delta t$ de l'actif A comme :

$$R_{t,t+\Delta t}^A = \mu^A + \beta^A f_t + \epsilon_t^A$$

avec μ^A une constante, β^A un vecteur de dimension $(1 \times k)$, f_t le vecteur de dimension $(k \times 1)$ des k facteurs de risque à l'instant t et ϵ_t^A un résidu, également appelé risque idiosyncratique de l'actif A , indépendant des facteurs de risque. L'APT suppose également que :

- Le nombre d'actifs présents sur le marché est supérieur à k .
- $\forall A, \forall t, \quad \mathbb{E}[\epsilon_t^A] = 0$.
- $\forall A, \forall B, \forall t, \quad \mathbb{E}[\epsilon_t^A \epsilon_t^B] = 0$.
- $\forall t, \forall i, \quad \mathbb{E}[f_{t,i}] = 0$.
- $\forall t, \forall i, \forall j, \quad \mathbb{E}[f_{t,i} f_{t,j}] = 0$.

Plus particulièrement, on a $\mathbb{E}[R_{t,t+\Delta t}^A] = \mu^A$. Dans ce cadre, le MEDAF peut être vu comme un cas particulier du modèle APT ayant un seul facteur de risque et un risque idiosyncratique nul.

Le principal reproche fait à ce modèle est qu'il se présente comme un modèle pouvant être testé empiriquement mais il ne spécifie explicitement aucun des facteurs de risque impactant la dynamique du prix d'équilibre (voir [Dhr(1984)], [Rol(1984)]). Il existe ainsi plusieurs manières de définir ces facteurs, soit par des méthodes statistiques appropriées de décomposition du signal (analyse en composantes principales, analyse en composantes indépendantes,...), soit par la sélection de facteurs de risque exogènes vérifiant les conditions du modèle (variables macroéconomiques, variables financières,...). Les tests empiriques de ce modèle sont donc très variés.

Pour le marché action, les méthodes statistiques concluent généralement sur un nombre de facteurs de risque compris entre 1 et 10 avec un nombre de facteurs significatifs augmentant lorsque le nombre d'actifs présents sur le marché augmente. Plus le nombre de facteurs augmente et plus les méthodes statistiques concluent sur la prédominance d'une composante (la « composante de marché ») [Trz(1986)].

Il existe de nombreux autres modèles de détermination des prix de marché, notamment des modèles multifactoriels ne respectant pas toutes les conditions de l'APT (indépendances entre les différents facteurs, etc...), par exemple en utilisant des facteurs de risque macroéconomiques [Che(1986)], des facteurs de risque plus spécifiques au contexte financier : facteurs taille et book-to-market de [Fam(1993)], facteur momentum de [Car(1997)] ou d'autres facteurs [Avr(2006)].

Le MEDAF et l'APT respectent l'hypothèse d'efficacité informationnelle du marché. Néanmoins, plusieurs faits ont été constatés remettant en cause cette théorie. Dans ce qui suit, on présente certaines observations souvent mises en avant comme étant des « anomalies » à la théorie économique néo-classique d'efficacité du marché.

1.2.3 Existence d'anomalies à l'efficacité informationnelle des marchés actions

Comme nous l'avons précisé précédemment, l'hypothèse d'un marché informationnellement efficace suppose que l'ensemble de l'information disponible aux investisseurs s'incorpore de manière efficace dans les prix d'équilibre. L'existence d'opportunités d'arbitrage n'est donc pas possible.

Or, un certain nombre d'observations, remettant en cause cette absence d'opportunités d'arbitrage, peuvent être constatées sur le marché action.

Biais saisonniers (Calendar effects)

Plusieurs études ont mis en avant l'existence de certains phénomènes revenant de manière régulière dans les rendements des prix de marché. Il est donc possible d'anticiper ces phénomènes du fait de leur saisonnalité. En les regroupant par fréquence d'apparition, ces phénomènes, ou biais saisonniers, sont les suivants :

- L'effet « week-end » (ou *day-of-the-week effect*). L'effet week-end décrit la tendance des prix à décroître le Lundi : les prix à la cloture du marché le Lundi ont tendance à être plus faible que les prix à la fermeture le Vendredi. Ce rendement entre le Vendredi et le Lundi tend également à être significativement plus faible que les rendements quotidiens entre les autres jours de la semaine ([Fre(1980)], [Gib(1981)], [Kei(1984)]). Cet effet serait notamment lié à une activité plus importante des investisseurs institutionnels sur les marchés actions le Lundi [Bru(2004)].
- L'effet « début de mois » (ou *week-of-the-month effect*). Cet effet décrit la tendance des rendements hebdomadaires à être significativement positifs la première semaine du mois tandis qu'ils ne sont pas significativement différents de 0 les autres semaines du mois. ([Koh(1992)], [Ari(1987)])
- L'effet « Janvier » ou « fin d'année » (ou *month-of-the-year effects*, ou *turn-of-the-year effect*). Cet effet se traduit par le fait que de nombreuses actions, et plus particulièrement celles des petites entreprises ayant faiblement performées l'année précédente, ont une tendance à avoir un rendement mensuel plus élevés entre fin Décembre et fin Janvier qu'entre les autres mois de l'année. ([Roz(1976)], [Rei(1983)], [Hau(1988)], [Lak(1984)], [Rit(1989)])
- L'effet « fin de mois » (ou *turn-of-the-month effect*). Cet effet est à rapprocher de l'effet de « début de mois ». Il se traduit par une tendance des rendements de prix quotidiens à être significativement positifs le dernier jour du mois précédant et les quatre premiers jours du mois suivant. ([Hen(1996)])
- L'effet « vacances » (ou *holiday effect*). Cet effet se traduit par un rendement anormalement élevé entre l'avant dernier et le dernier jour précédent des vacances scolaires [Kim(1994)].

Effets Small Cap/Large Cap, Growth/Value, Momentum de court-terme et Renversement de long terme

Dans ce sous-paragraphe, nous présentons différentes observations non saisonnières mais contredisant une absence d'opportunités d'arbitrage et largement documentés dans la littérature.

- [Ban(1981)] a mis en évidence que les actions d'entreprises avec une petite capitalisation ont des rendements moyens significativement plus élevés que celles ayant de grosses capitalisations. Ce phénomène serait notamment dû à une prime de risque plus élevée pour les entreprises avec une petite capitalisation. On parle d'effet « small cap/large cap ».
- Un autre effet prépondérant dans la littérature financière est l'effet « Growth/Value ». [Fam(1992)] ont identifié une prime dite « value » dans les rendements des actions aux États-Unis. En d'autres termes, les actions ayant un rapport « valeur comptable sur valeur de marché » (également appelé « book-to-market ») élevé ont des rendements moyens plus élevés que celles avec un faible book-to-market. Ces actions avec un « book-to-market » élevé sont appelées actions « value » tandis que les autres sont dites « growth ».

- L'effet « momentum de court terme » a été mis en avant par [Jeg(1993)]. Ces derniers ont remarqué que les actions ayant eu la meilleure (resp. la plus mauvaise) performance sur les 3 à 12 mois précédents ont tendance à continuer de bien (resp. mal) performer les 3 à 12 mois suivant. Les actions offrant la meilleure performance apparaissent ne pas être plus risquées que celle offrant les pires performances. Toutefois, les rendements anormaux constatés ont tendance à s'effriter durant les deux années suivantes.
- Enfin, l'effet « renversement de long terme » a été mis en avant par [DeB(1985)]. Ces derniers remarquent qu'un portefeuille composé d'actions ayant eu la pire performance sur les trois à cinq années passées a tendance à surperformer, sur les trois à cinq années suivantes, un portefeuille composé d'actions ayant eu la meilleure performance sur ces trois à cinq années passées.

Autres biais

D'autres biais sont régulièrement évoqués dans la littérature. Parmi ces biais, on trouve par exemple l'effet "stock split". Ce dernier intervient lors de l'annonce d'un partitionnement d'actions. Un tel partitionnement augmente le nombre d'actions en suspens et fait baisser la valeur de chaque action tout en préservant la capitalisation de marché de l'entreprise. Néanmoins, on constate une augmentation du prix de l'action avant et après cette annonce. De nombreuses entreprises annoncent le partitionnement de leurs actions quand leur action a augmenté à un prix trop élevé pour l'investisseur moyen. Le partitionnement des titres est donc souvent vu par les investisseurs comme un signal que l'action de l'entreprise continuera à augmenter [Lam(1987)].

Un autre biais souvent évoqué correspond à l'arbitrage sur fusion/acquisition. Quand les entreprises annoncent une fusion ou une acquisition, le prix de marché de l'action de l'entreprise acquise tend à augmenter tandis que la valeur de celle qui l'acquiert tend à baisser. L'arbitrage lors d'une fusion-acquisition entraîne un mispricing potentiel après une telle annonce. En fait, l'offre soumise pour une acquisition peut ne pas être un reflet précis de la valeur fondamentale de l'entreprise acquise. Les arbitrageurs aspirent à exploiter ce type de biais [Mit(2004)].

1.2.4 Biais comportementaux des investisseurs

Les critiques les plus dures à l'encontre de l'efficacité du marché résident dans les choix et les comportements des participants du marché. L'approche standard pour modéliser les choix est de supposer que les investisseurs optimisent l'espérance d'une fonction d'utilité dite « additivement séparable par rapport au temps »¹¹, par rapport à certaines familles de paramètres, par exemple l'aversion au risque relatif constant. Dans ce cadre, on parle de « rationalité » des investisseurs. Néanmoins, des psychologues et chercheurs en économie ont mis en lumière un certain nombre de biais comportementaux spécifiques qui sont omniprésents aux prises de décision humaines dans des situations d'incertitude, plusieurs de ces biais menant à des résultats indésirables pour la richesse économique individuelle. On relate brièvement par la suite certains de ces biais.

La sur-confiance (ou présomptuosité)

Le biais de sur-confiance provient de la mauvaise calibration de probabilités subjectives. Un individu estime des probabilités liées à certaines hypothèses de manière bien calibrée si, sur le long terme,

11. c.-à-d. $F(x_1, \dots, x_t, \dots, x_T) = \sum_{t=1}^T f_t(x_t)$

la proportion de réalisation de ces hypothèses est égale aux probabilités qu'il leur a préalablement assignées. Le degré de calibrage d'une probabilité traduit donc la manière dont la confiance en un évènement correspond à la probabilité d'occurrence effective. On trouve à l'origine de la sur-confiance différentes causes :

- la surestimation des connaissances : les individus sont présomptueux sur la précision de leurs connaissances ([Alp(1982)], [Lic(1982)]).
- l'effet mieux que la moyenne : les individus pensent que leur jugement est meilleur que la moyenne ([Sve(1981)], [Tay(1988)]).
- l'illusion du contrôle ou l'optimisme irréal : les individus sont démesurément optimistes sur les évènements futurs qui leur sont personnellement favorables ([Wol(1975)], [Mil(1975)], [Wei(1980)]).

Toutes ces causes entraînent une espérance de la probabilité du succès personnel plus élevée que la probabilité objective.

L'heuristique de représentativité

Un individu soumis à l'heuristique de représentativité tend à privilégier les informations stéréotypiques au détriment d'autres informations sur les probabilités d'évènements [Kah(1982)]. Ainsi, les investisseurs sous un biais d'heuristique de représentativité accordent plus d'importance à des informations récentes allant dans le même sens qu'à des informations structurelles dans la révision de leurs croyances [Bar(1998)]. Ces jugements expliqueraient la sur-réaction¹² du prix de marché d'une action à une série d'informations mauvaises (resp. bonnes) concernant une entreprise.

Le biais d'ancrage

Un individu soumis au biais d'ancrage n'arrive pas à ajuster aux nouvelles données l'estimation qu'il a faite de la probabilité d'un évènement. D'après [Tve(1974)], ce biais entraînerait les individus à surestimer la probabilité liée à la réalisation d'évènements conjoints et à sous-estimer celle liée à la réalisation d'évènements disjoints. Il serait notamment à l'origine de la sous-réaction¹³ des cours des actions aux annonces de bénéfices des entreprises puisque les investisseurs ne réviseraient pas suffisamment leurs croyances [Bar(1998)].

La comptabilité mentale (ou psychologique)

D'après l'hypothèse de l'utilité espérée, un investisseur doit évaluer un actif risqué d'après la valeur avec laquelle cet actif contribue à sa richesse économique individuelle totale, symbolisée par le portefeuille de l'investisseur. Or, [Tha(1981)] mettent en lumière qu'un investisseur évalue en général un actif risqué directement par rapport aux gains que celui-ci peut délivrer, indépendamment des autres actifs. Cette opération cognitive est connue sous le nom de « *narrow framing* » (mot-à-mot « cadrage étroit ») : les individus évaluent des actifs risqués ou des tickets de loteries directement par les gains que ces actifs vont engendrer sans intégration de ces gains dans un compte de résultats plus global. Ce biais fait partie de ce que [Tha(1999)] définit comme les biais de comptabilité mentale : « *l'ensemble des opérations cognitives utilisées par les individus et ménages pour organiser, évaluer et suivre des activités financières* ». Le biais de « *narrow framing* » est en accord

12. Une réaction démesurée par rapport à une réaction dite « rationnelle ».

13. Une réaction plus lente qu'une réaction rationnelle.

avec la théorie des perspectives¹⁴ de [Kah(1979)]. D'après cette théorie, les investisseurs maximisent une fonction similaire à une fonction d'utilité standard à l'exception qu'elle est définie sur les gains et pertes de chaque actifs plutôt que sur le niveau de richesse totale.

L'aversion à la perte et l'aversion myope à la perte

L'aversion à la perte se définit comme la propension psychologique à percevoir les pertes d'une certaine amplitude plus grandes que les gains d'une même amplitude par rapport à un point de référence [Kah(1979)], [Tve(1991)]. Ce point de référence peut correspondre au prix d'achat d'un titre. Si le prix de marché de ce titre descend en dessous de ce point de référence, l'investisseur averse à la perte aura alors tendance à percevoir cette perte d'une manière plus importante que si le prix avait augmenté d'une même amplitude. [Ben(1995)] combine ce concept avec la comptabilité mentale afin d'expliquer la prime de risque des marchés actions¹⁵. On parle alors d'investisseurs ayant une aversion « myope » à la perte.

L'instinct grégaire (ou comportement moutonnier)

Les pensées, actions et sentiments d'un individu sont influencés par son environnement social, environnement au sein duquel d'autres individus interagissent. L'influence de ces derniers sur un individu s'opère soit par un échange verbal ou écrit, soit par l'observation d'actions (par exemple l'observation des offres et demandes concernant un titre), soit par l'observation de conséquences d'actions (par exemple l'observation des gains propres à chaque titre, ou des prix de marché). De cette façon, le processus d'interaction social entre investisseurs peut entraîner des convergences de comportement. Ainsi, des comportements moutonniers ou encore de « suiveurs de tendance »¹⁶ ont fréquemment été observés sur les marchés financiers. [Hub(2001)] mettent en avant ce comportement des investisseurs sur le marché action en montrant que les cours d'un secteur action particulier (biotechnologies) montent suite à la diffusion d'un article du New York Times concernant des avancées technologiques possibles mais pas encore découvertes dans ce secteur. De nombreuses autres études ont observé ce comportement moutonnier [Bik(2000)] : par exemple dans la sélection de titres des gérants de fonds [Gri(1995)], ou bien encore au niveau des prédictions faites par les analystes financiers [Kri(2006)].

L'escompte hyperbolique du futur

L'escompte hyperbolique du futur s'oppose à un taux d'escompte exponentiel du futur. Dans le cas exponentiel, les préférences d'un individu sont constantes au cours du temps, peu importe le moment où la décision est prise. Dans le cas hyperbolique, les préférences d'un individu dépendent de la situation temporelle dans laquelle se trouve celui-ci. Avec un taux d'escompte hyperbolique, les préférences d'un individu peuvent donc changer sans intervention d'un élément déclencheur extérieur, mais uniquement du fait de la structure même du raisonnement de l'individu. Du fait même de ses préférences, l'individu n'exécute pas la décision qu'il avait prise antérieurement. On pourra se reporter à [Lai(1997)] ou [O'D(1999)] pour plus de détails sur ce biais comportemental.

14. « Prospect Theory » en anglais.

15. Différence entre les rendements des actions (equity) et ceux d'un actif sans risque (treasury bills).

16. « Trend followers » en anglais.

L'aversion au regret

Le regret est une peine ressentie à la suite d'une décision prise antérieurement et qui nous déplaît a posteriori. Concrètement, l'aversion au regret d'un investisseur peut ainsi le conduire au choix de conserver un portefeuille benchmark plutôt que de profiter d'opportunités d'investissements. En raisonnant de cette manière, le regret serait quantifiable par les écarts de performances entre un investissement et son portefeuille benchmark. Un individu averse au regret devrait donc compenser cet écart par des rendements sur investissement attendus suffisamment élevés [Bel(1982)], [Cla(1994)].

Toutes ces constatations tendent à faire penser que les investisseurs sont souvent irrationnels - si ce n'est toujours - en ayant des comportements prévisibles et financièrement ruineux.

1.3 La dynamique des prix de marché

Supposons que le prix à un instant t incorpore instantanément l'ensemble de l'information disponible. En conséquence, les variations de ce prix seraient imprévisibles. Il semble alors logique de modéliser le processus des prix comme un processus semblable à une marche aléatoire :

$$P_{t+1} = P_t + \epsilon_{t+1}$$

où P_t est une variable aléatoire représentant le prix à l'instant t et ϵ_{t+1} est un bruit aléatoire, centré sur 0 et de répartition symétrique, représentant l'évolution indéterministe entre t et $t + 1$.

Si le prix de marché possède ce type de comportement, aucune stratégie d'investissement ne peut battre régulièrement une stratégie auto-financée puisque l'évolution des prix serait entièrement aléatoire. Or, de nombreuses études statistiques ont mis en avant des observations contraires à cette hypothèse de marche aléatoire, notamment à travers l'existence de persistances et tendances au sein des prix.

1.3.1 Persistances et tendances dans les prix de marché

La notion de tendance est différente de celle de persistance, la tendance désignant la direction prise par les rendements des prix et la persistance désignant une dépendance sérielle au sein de ces rendements. La notion de tendance dans les prix de marché est notamment à la base de l'analyse technique¹⁷.

À travers la littérature, la dépendance au sein de l'évolution des prix est présente sous trois formes distinctes : l'indépendance des rendements de prix, la prédictibilité de la direction des prix (se traduisant par une dépendance significative dans les signes des rendements) et la dépendance de long terme de la volatilité des rendements.

L'indépendance des rendements de prix

Il est en général admis que les rendements de prix, calculés sur des horizons d'investissement allant de un jour à un an, possèdent des autocorrélations significativement nulles [Wor(1934)],

17. Voir paragraphe 4.2.2.

[Ken(1953)], [Rob(1959)]. En conséquence, on suppose qu'ils possèdent une espérance conditionnelle $E[R_{t,t+\Delta t}|\mathcal{F}_d]$ ne dépendant pas de la filtration \mathcal{F}_d (c.-à-d. $\forall \mathcal{F}_d, E[R_{t,t+\Delta t}|\mathcal{F}_d] = E[R_{t,t+\Delta t}]$). Ils sont donc indépendants au sens de cette espérance conditionnelle. Néanmoins, certaines études ont mis en avant une dépendance significative des rendements sur des horizons soit de très court terme, soit de très long terme. Ainsi, les rendements à ultra haute fréquence (avec un horizon d'approximativement 20 minutes) ont des autocorrélations faibles mais significatives dues à des effets de microstructures [Lo(1999)]. D'autres études ont montrées une dépendance des rendements sur des horizons de très longs terme (2 à 10 ans) [Cam(1988)], [Fam(1988)]. L'indépendance dans l'évolution des prix est donc une hypothèse dont la véracité dépend de l'horizon du rendement.

Il faut également remarquer que cette dépendance évolue avec le temps. Suivant la fenêtre de temps utilisée pour estimer les autocorrélations des rendements de prix, on peut obtenir des résultats différents (voir figure 1.2).

On peut toutefois se questionner sur la pertinence des autocorrélations pour mesurer une dépendance au sein de l'évolution des prix puisque celles-ci ne mesurent qu'une dépendance linéaire. Il semblerait plus judicieux d'étudier cette dépendance au moyen d'un critère d'information mutuelle par exemple.

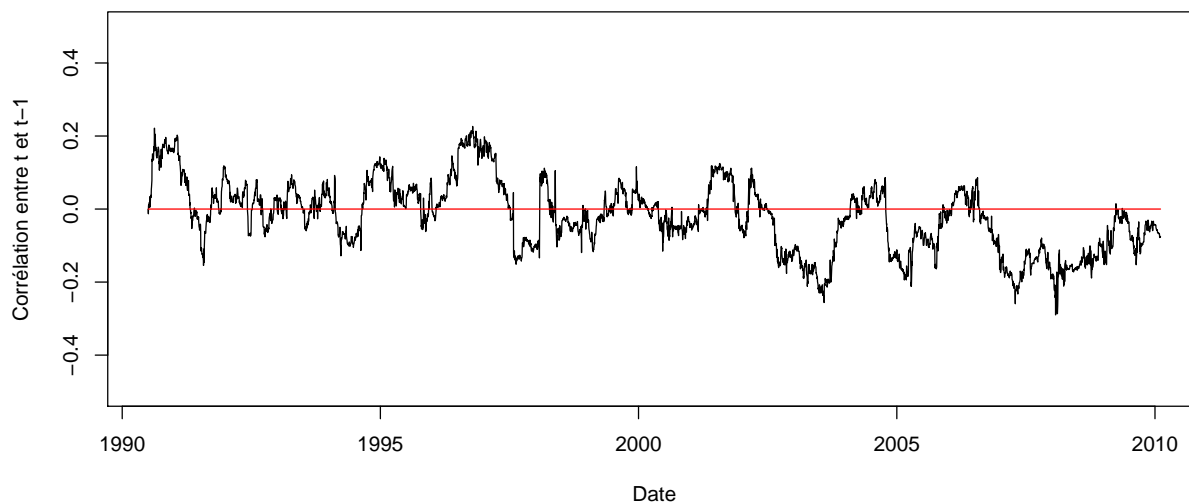


FIGURE 1.2 – Évolution du coefficient d'autocorrélation des rendements de l'indice de prix S&P 500 entre l'instant $t - 1$ et l'instant t évalué sur une fenêtre glissante de 6 mois.

Dépendance de long-terme de la volatilité des rendements

Les autocorrélations des rendements ne mesurent qu'une dépendance linéaire dans l'évolution des prix. Lorsqu'on observe de plus près la dynamique des prix, on constate que de grandes variations dans les prix sont généralement suivies par de grandes variations, de même que de petites variations sont suivies par des petites variations. On dit alors que la volatilité des prix expose une dépendance de long-terme. Concrètement, cette dépendance peut s'observer en étudiant les autocorrélations des rendements au carré. On constate généralement que cette décroissance est lente [Tay(1986)]. Lorsque

1.3. LA DYNAMIQUE DES PRIX DE MARCHÉ

la décroissance de ces autocorrélations est de type hyperbolique, les rendements au carré peuvent être modélisés à l'aide d'un processus à mémoire longue [Ban(2005)] :

$$(1 - L)^d R_{t,t+\Delta t}^2 = \epsilon_t$$

où L est l'opérateur retard et ϵ_t un bruit blanc. $0 < d < 0.5$ représente alors le paramètre de mémoire du processus à mémoire longue. Plusieurs études ont montré que les rendements au carré des prix de marché des actions peuvent être modélisés par un tel processus [Din(1993)] [Bol(1996)] [Lob(1998)].

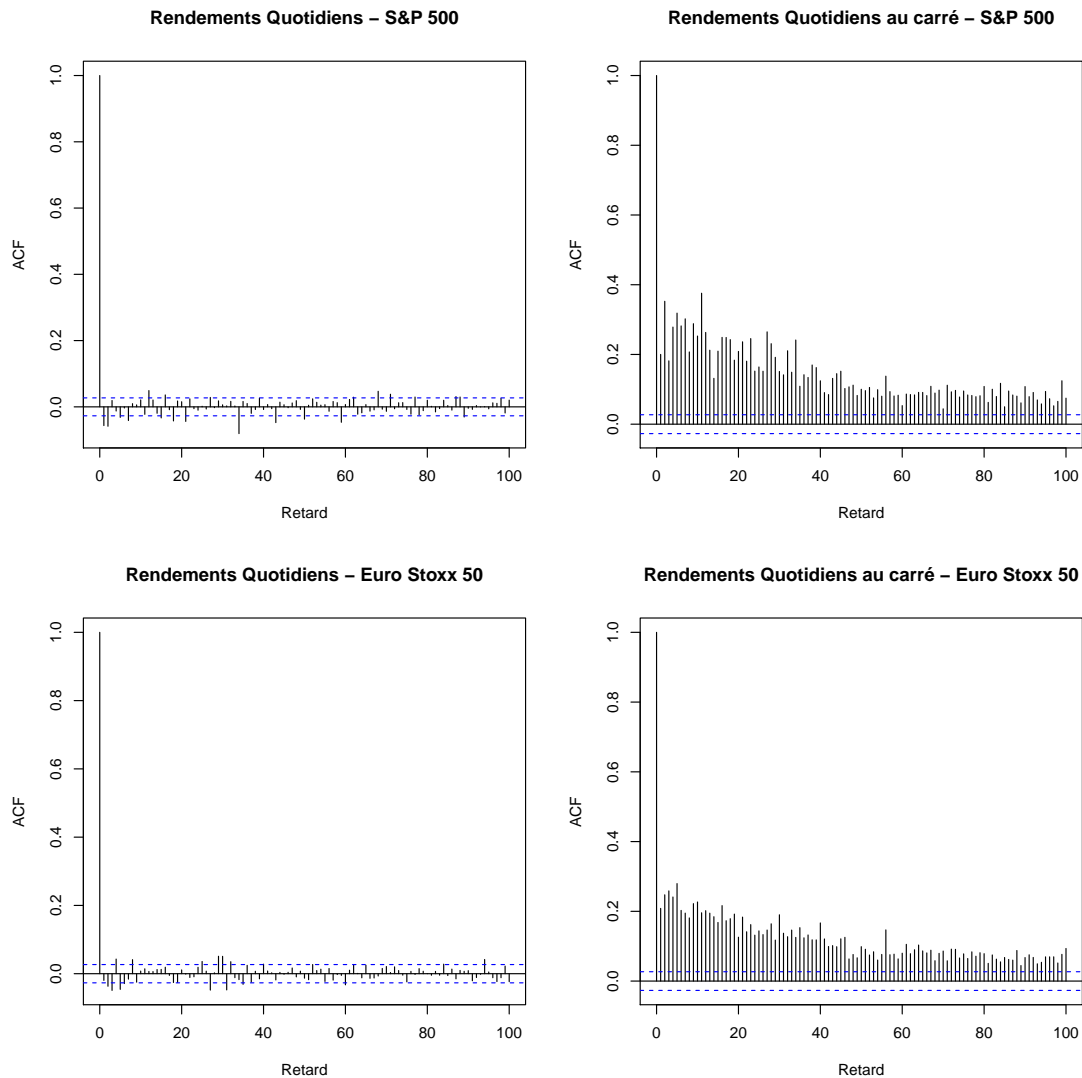


FIGURE 1.3 – Fonctions d'autocorrélation pour les rendements et les rendements au carré des indices de prix S&P500 et Euro Stoxx 50 évaluées sur la période 01/01/1990-26/10/2010.

La prédictibilité de la direction des prix et l'existence d'un « cycle de marché »

En macroéconomie, il est courant de modéliser certaines variables en considérant qu'il existe différents états du monde ou régimes. Ainsi, le NBER¹⁸ définit le cycle économique comme une succession de périodes d'expansion, de stabilité, de contraction et de récession. Chaque période se caractérise alors par un comportement particulier des variables macroéconomiques mesurées. Similairement à la macroéconomie, en finance de marché, l'idée d'un cycle appelé « cycle financier », ou « cycle de marché », a été développée. En effet, plusieurs constatations statistiques mettent en lumière l'existence de différents régimes de marché, notamment au niveau de la volatilité du marché (voir paragraphe 1.3.1 (sous-paragraphe intitulé « Relations entre ces trois types de dépendance »)). La prédictibilité de la direction des prix serait notamment plus faible en période de forte volatilité du marché [LeB(1992)].

Cette prédictibilité se traduit par une dépendance significative dans l'évolution des signes des rendements. Plusieurs études ont conclu sur cette dépendance significative des signes des rendements concernant certains indices de prix du marché action (voir par exemple [Bre(1989)], [Pes(1995)], [Whi(2000)], [Pes(2001)]). Dans ces études, ainsi qu'à travers la littérature, certaines variables économiques et financières semblent avoir un pouvoir prédictif significatif sur l'évolution des rendements en excès de prix du marché. Elles sont mesurées à travers les prix de marché de certains actifs et donnent un point de vue global sur l'état des conditions financières. Elles permettent donc de caractériser le cycle de marché et représentent des approximations d'anticipations ou de prime de risque faites par les participants actifs du marché. Parmi ces variables caractérisant le cycle financier, on trouve par exemple :

- le rendement des bons du Trésors de maturité 3 mois. Cette variable est considérée comme une approximation de l'inflation anticipée ([Fam(1977)], [Fam(1981)]).
- la différence entre le rendement moyen des bons du Trésor de plus de 10 ans de maturité et ceux de moins de 3 mois. Cette variable est considérée comme une approximation de l'anticipation de l'activité économique à long terme ([Fam(1988)]).
- la différence entre les rendements en dividendes et le rendement sans risque. Cette variable est considérée comme une approximation de la prime de risque liée au marché action ([Kei(1986)], [Cam(1989)], [Fam(1988)]).
- la différence entre le rendement moyen des obligations d'entreprises dites "investment grade" bien notées (Aaa par Moody's ou AAA par S&P) et le rendement moyen de celles ayant une note plus faible (notées Baa par Moody's ou BBB par S&P). Cette variable sert d'approximation à la prime de risque de défaut des entreprises ([Fam(1988)]).

Ces quatre variables sont les plus utilisées dans la littérature pour décrire le cycle de marché. Néanmoins, cette liste est non exhaustive et d'autres variables ont montré leur pertinence dans la description de ce cycle. Par exemple :

- la différence entre le rendement moyen d'un portefeuille d'actions à forte capitalisation et le rendement moyen d'un portefeuille composé d'actions à faible capitalisation ([Fam(1993)]). Cette variable prend en compte l'effet Small Cap/Large Cap.
- la différence entre le rendement moyen d'un portefeuille d'actions avec un book-to-market élevé et le rendement moyen d'un portefeuille composé d'actions à faible book-to-market ([Fam(1993)]). Cette variable prend en compte l'effet Growth/Value.
- la différence entre le rendement moyen d'un portefeuille composé d'actions ayant eu une bonne

18. The National Bureau of Economic Research est un organisme de recherche privé à but non lucratif.

performance sur les 11 derniers mois et le rendement moyen d'un portefeuille composé d'actions ayant eu une mauvaise performance sur la même période ([Car(1997)]). Cette variable prend en compte l'effet Momentum.

Ces variables sont d'un grand intérêt pour un gérant actif de portefeuille puisque l'état du cycle financier pourra lui indiquer quelles décisions prendre dans son allocation de portefeuille. Nous discuterons plus amplement de toutes ces variables, ainsi que d'autres facteurs de risque, dans notre troisième chapitre.

Relations entre ces trois types de dépendance

Afin d'illustrer les relations entre ces trois types de dépendance, supposons, dans un premier temps, que $R_{t,t+\Delta t}|\mathcal{F}_t$ soit une variable aléatoire suivant une loi normale de moyenne conditionnelle $\mu_{t+\Delta t}|\mathcal{F}_t$ et de variance conditionnelle finie $\sigma_{t+\Delta t}^2|\mathcal{F}_t$. La probabilité que le rendement entre t et $t+\Delta t$ soit positif est donc de :

$$\begin{aligned} \mathbb{P}(R_{t,t+\Delta t} > 0|\mathcal{F}_t) &= 1 - \mathbb{P}(R_{t,t+\Delta t} \leq 0|\mathcal{F}_t) = 1 - \mathbb{P}\left(\frac{R_{t,t+\Delta t} - \mu_{t+\Delta t}}{\sigma_{t+\Delta t}} \leq \frac{-\mu_{t+\Delta t}}{\sigma_{t+\Delta t}} \middle| \mathcal{F}_t\right) \\ &= 1 - \Phi_{\mathcal{N}(0,1)}\left(\frac{-\mu_{t+\Delta t}}{\sigma_{t+\Delta t}} \middle| \mathcal{F}_t\right) \end{aligned}$$

où $\Phi_{\mathcal{N}(0,1)}(\cdot)$ est la fonction de répartition de la loi $\mathcal{N}(0,1)$.

Dans le cas où les rendements ne suivent pas une loi normale, on peut remplacer $\Phi_{\mathcal{N}(0,1)}$ par la fonction de répartition de la variable aléatoire $Z_{t,t+\Delta t} = \frac{R_{t,t+\Delta t} - \mu_{t+\Delta t}}{\sigma_{t+\Delta t}}$. En conséquence,

$$\mathbb{E}[\mathbb{P}(R_{t,t+\Delta t} > 0)|\mathcal{F}_t] = 1 - \Phi_{\mathcal{N}(0,1)}\left(\mathbb{E}\left[\frac{-\mu_{t+\Delta t}}{\sigma_{t+\Delta t}} \middle| \mathcal{F}_t\right]\right)$$

En pratique, $\mu_{t+\Delta t}$ peut être approximée par une fonction du temps suffisamment régulière évoluant dans un voisinage de 0 comme le suggère les tests statistiques sur les rendements. D'un autre côté, $\sigma_{t+\Delta t}$ peut être approximé par un processus à mémoire longue tel qu'un de ceux proposés dans [Ban(2005)] par exemple.

Supposons dorénavant que $\mathbb{P}(Z_{t,t+\Delta t} \leq 0|\mathcal{F}_t)$ soit égale à $m \in [0, 1]$. Dans le cas où l'espérance de la tendance des rendements est nulle,

$$\mathbb{E}[\mathbb{P}(R_{t,t+\Delta t} > 0)|\mathcal{F}_t] = m$$

Dans le cas gaussien, $m = \frac{1}{2}$ et le signe du rendement à venir est donc imprévisible. Si l'espérance de la tendance est non nulle et ne dépend pas de la filtration \mathcal{F}_t , elle est égale à une constante $\mu \neq 0$. Dans ce cas, on dit que les rendements sont indépendants au sens de leur espérance conditionnelle. La prévisibilité du signe des rendements évoluerait donc en fonction de la volatilité des rendements. Plus particulièrement,

- Dans le cas où $\mu > 0$, $\mathbb{E}[\mathbb{P}(R_{t,t+\Delta t} > 0)|\mathcal{F}_t] > m$.
- Dans le cas où $\mu < 0$, $\mathbb{E}[\mathbb{P}(R_{t,t+\Delta t} > 0)|\mathcal{F}_t] < m$.
- Plus l'espérance de la volatilité conditionnellement à \mathcal{F}_t sera forte, plus $\mathbb{E}[\mathbb{P}(R_{t,t+\Delta t} > 0)|\mathcal{F}_t]$ tendra vers m . Dans le cas gaussien, le signe sera donc d'autant plus imprévisible que la volatilité sera forte.

Concernant le marché action, de nombreuses études ont montré que la volatilité des rendements quotidiens tendait à baisser (resp. monter) lorsque les prix montaient (resp. baissaient). Ce phénomène est également appelé « effet de levier » [Bla(1976)] et se traduit par une corrélation significativement négative entre les rendements de prix passés et leur volatilité future [Bou(2001)]. Cette observation statistique pourrait également se traduire par la relation suivante :

$$\text{sign}(\mu_{t+\Delta t}) \neq \text{sign}(\sigma_{t+\tau+\Delta t} - \sigma_{t+\Delta t}) \quad \text{pour un } \tau > 0 \text{ fixé.}$$

Ainsi, si les prix sont en phase baissière ($\mu_{t+\Delta t} < 0$), la volatilité à venir serait en hausse ($\sigma_{t+\tau+\Delta t} > \sigma_{t+\Delta t}$). Dans le cas gaussien, on aurait donc une prévisibilité du signe telle que $\mathbb{E}[\mathbb{P}(R_{t,t+\Delta t} > 0) | \mathcal{F}_t] < \frac{1}{2}$ mais se rapprochant de l'imprévisibilité puisque la volatilité serait en augmentation.

D'autres études proposent de définir des « régimes de marché » en se focalisant sur le niveau de volatilité. Ainsi, les périodes de faibles et fortes volatilité caractériseraient un « cycle » intrinsèque aux séries de prix. Plus particulièrement, LeBaron [LeB(1992)] montre que les autocorrélations des rendements de prix d'actions sont liées à leur niveau de volatilité. Les autocorrélations tendent à être plus grandes durant les périodes de faible volatilité et plus petites durant les périodes de forte volatilité. On parle également de « cluster » de volatilité. À partir de cette observation, LeBaron propose de modéliser les séries de prix avec un processus du type (l'espérance conditionnelle des rendements étant supposée constamment nulle) :

$$R_{t-\Delta t,t} = f(\sigma_t^2) R_{t-2\Delta t,t-\Delta t} + \epsilon_t$$

$$\sigma_t^2 = \sum_{i=1}^N R_{t-(i+1)\Delta t,t-i\Delta t}^2$$

f étant une fonction décroissance de la variance conditionnelle σ_t^2 , ce qui se traduit par de meilleures prédictions locales en période de faible volatilité des rendements.

1.3.2 Répartition inconditionnelle des rendements

Dans le sous-paragraphe précédent, nous avons discuté de la dynamique des prix, et plus particulièrement de leurs rendements. Cette dynamique induit nécessairement une répartition inconditionnelle des rendements. Il existe beaucoup de publications sur cette répartition pour les rendements des prix du marché action et quelques particularités, également appelées « faits stylisés », les caractérisent. Parmi ces particularités, on trouve principalement :

- l'existence de queues épaisses au niveau de la répartition inconditionnelle des rendements. On parle également de leptokurticité.
- une asymétrie entre les pertes et les gains. On observe plus particulièrement des valeurs extrêmes négatives plus élevées que celles positives.

En conséquence, la répartition des rendements semble en général différente de celle d'une loi normale. De nombreuses publications se sont intéressées aux skewness et kurtosis de ces rendements. Toutes concluent que ces rendements sont leptokurtiques et asymétriques. [Gem(2002)] montre que la répartition des rendements sur des horizons de très court terme (1 à 15 minutes) est plus proche de celle d'une loi hyperbolique généralisée, suggérant ainsi de remplacer le mouvement Brownien par un mouvement hyperbolique généralisé dans les processus « classiques » de modélisation de l'évolution des prix comme ceux de [Bla(1973)] et [Mer(1976)].

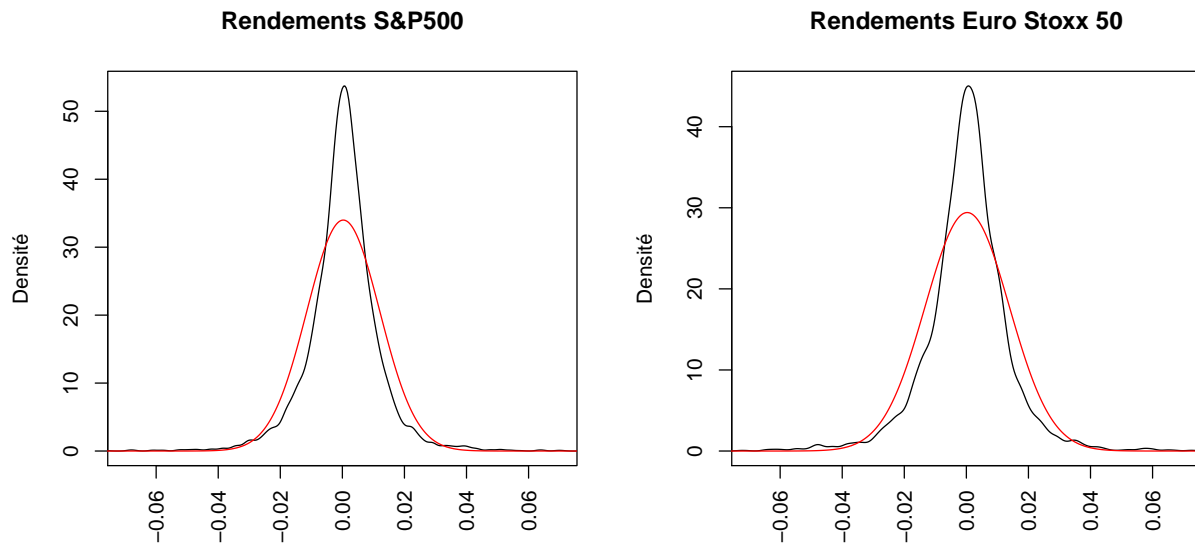


FIGURE 1.4 – Densité estimée par des noyaux gaussiens (en noir) et densité d’une loi normale ayant même moyenne et même variance que celle des rendements (en rouge) évaluées pour les rendements quotidiens des indices de prix S&P500 et Euro Stoxx 50 sur la période 01/01/1990-26/10/2010.

1.3.3 La dynamique des dépendances entre actifs

Lorsque les marchés sont en situation de stress, les propriétés de dépendance entre les actifs changent. Ainsi, des actifs dont les prix de marché étaient censés évoluer de manière indépendante peuvent se mettre à baisser de manière très corrélée. Ceci pose notamment problème pour un gérant de portefeuille optimisant son portefeuille sur une seule période de temps par exemple. Cette décision, censée être optimale, deviendrait obsolète si la configuration du marché évolue. Les propriétés de dépendance entre les actifs varient donc avec le temps [Cam(1997)].

Pour mesurer de manière empirique l’évolution des dépendances entre plusieurs classes d’actifs, on peut utiliser un coefficient de corrélation intra-classes d’actifs, noté ρ_{intra} . Il s’estime par

$$\rho_{intra}(t) = \frac{1}{mNs^2} \sum_{n=1}^N \left[\sum_{i=1}^{m-1} \sum_{j>i}^m (R_{t-\Delta t,t}^{n,i} - \bar{R})(R_{t-\Delta t,t}^{n,j} - \bar{R}) \right]$$

où :

- N représente le nombre de classes d’actifs et m le nombre d’actifs.
- $R_{t-\Delta t,t}^{n,i}$ représente le rendement entre $t - \Delta t$ et t de l’actif i issu de la $n^{\text{ème}}$ classe d’actifs.
- $\bar{R} = \frac{1}{mN} \sum_{n=1}^N \left(\sum_{j=1}^m R_{t-\Delta t,t}^{n,j} \right)$
- $s^2 = \frac{1}{mN} \sum_{j=1}^m \sum_{n=1}^N (R_{t-\Delta t,t}^{n,j} - \bar{R})^2$

1.3. LA DYNAMIQUE DES PRIX DE MARCHÉ

Plus ce coefficient de corrélation intra-classes d'actifs est élevé et plus la corrélation entre les classes d'actifs considérées est forte. La figure 1.5 représente l'évolution de ce coefficient durant la période 01/01/2000-30/03/2010 pour les classes d'actifs suivantes :

- Actions.

Les indices de prix actions considérés sont les suivants : Euro Stoxx Health Care, Euro Stoxx Oil & Gas, Euro Stoxx Technology, Euro Stoxx Telecom, Euro Stoxx Utilities, Euro Stoxx 50, Euro Stoxx Small, Euro Stoxx MID, S&P 500 Composite, S&P 500/CITIGROUP Growth, S&P 500/CITIGROUP Value, S&P 400 Midcap, S&P 600 Small cap, MSCI US Small cap Growth, MSCI US Small cap Value, MSCI US Small Cap 1750, FTSE/NAREIT Equity REITS, NIKKEI 225 Stock Average, TOPIX, MSCI Japan (Growth), MSCI Japan (Value), DJTM World Emerging, MSCI EM Asia, MSCI EM, MSCI EM Europe, MSCI EM Latin America, MSCI Malaysia (EM).

- Obligations.

Les indices obligataires considérés sont les suivants : US T.I.P.S., Taux à 2 ans US, Taux à 10 ans US, Barclays US High Yield Composite, Barclays US Corporate Investment Grade, Taux à 2 ans Allemagne, Taux à 10 ans Allemagne, Barclays PAN-EUR Aggregate High Yield, Barclays PAN-EUR Aggregate AAA.

- Matières premières.

Les indices de prix des matières premières considérés sont les suivants : S&P GSCI Commodity Excess Return, S&P GSCI Agriculture Excess Return, S&P GSCI Crude Oil Excess Return, S&P GSCI Gold Excess Return, S&P GSCI Industrial Metals Exc. Ret., S&P GSCI Livestock Excess Return, S&P GSCI Energy Excess Return.

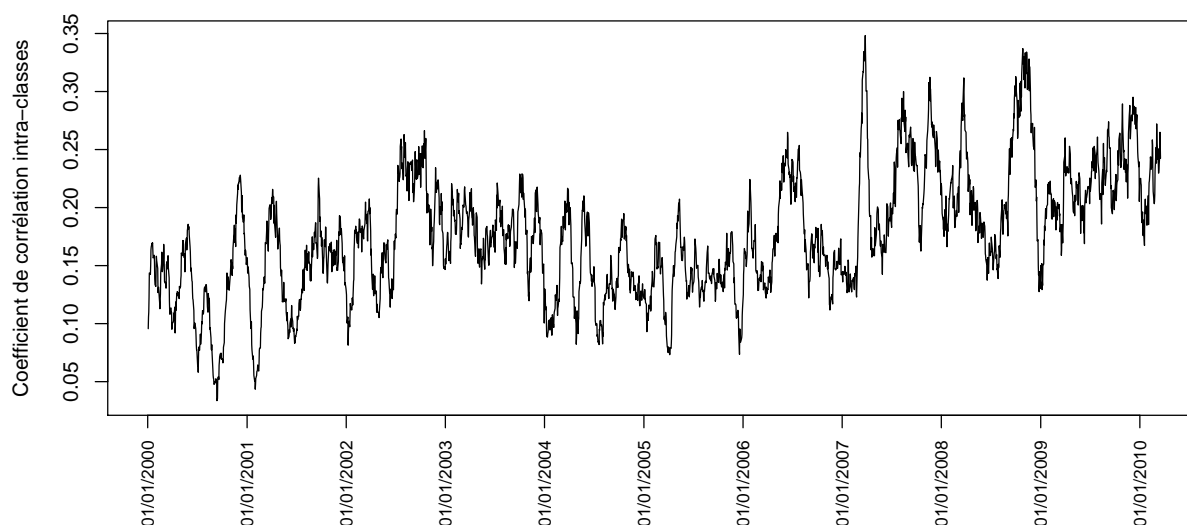


FIGURE 1.5 – Évolution d'un coefficient intra-classes d'actifs pour les actions, les obligations et les matières premières. On constate une forte augmentation de ce coefficient lors du crash d'Octobre 2008 par exemple.

Il existe différentes mesures de dépendance entre deux actifs. Les principales sont les suivantes :

- Le coefficient de corrélation. C'est la mesure la plus populaire. Néanmoins, un coefficient de corrélation nul n'est pas synonyme d'indépendance mais seulement de décorrélation linéaire.
- Le tau de Kendall. C'est une mesure de corrélation des rangs. Il correspond à la différence entre la probabilité de concordance et celle de discordance.
- Le rho de Spearman. Comme le tau de Kendall, c'est une mesure de corrélation des rangs. Il est proportionnel à la différence entre la probabilité de concordance et celle de discordance.
- L'information mutuelle moyenne. Cette mesure permet d'estimer la dépendance, linéaire ou non, entre deux séries. Elle nécessite néanmoins d'estimer la densité de la loi de probabilité associée à chaque série ainsi que celle de leur loi jointe.
- La dépendance de queue. Cette dépendance décrit la dépendance entre des valeurs extrêmes. Elle peut notamment se mesurer à l'aide de copules [Emb(2003)].

Afin de prendre en compte cette dynamique conditionnelle de la dépendance entre différents investissements, plusieurs modèles ont été proposés. On citera par exemple le modèle DCC-GARCH¹⁹ [Eng(2002)] permettant de prendre également en compte la décroissance lente des autocorrélations des rendements.

1.4 Univers d'investissement et gestion de portefeuille en pratique

Pour maximiser la valeur de son portefeuille, un gérant va tenter d'anticiper l'évolution de la valeur d'un investissement dans chacun des titres présents dans son univers d'investissement. Ces anticipations lui permettront d'arbitrer dans cet univers. Pour cela, deux sortes d'analyse des titres sont généralement pratiquées : une analyse fondamentale et/ou une analyse technique. Chacune possède ses avantages et ses inconvénients.

1.4.1 Différentes classes d'actifs

Une classe d'actifs représente une catégorie spécifique de titres financiers. Dans la suite de notre étude et afin d'avoir un univers d'investissement clair et précis, nous utiliserons des classes d'actifs généralement présentes dans les portefeuilles des gérants de fonds : actions, obligations, taux de change, futures sur matières premières et liquidités. Dans ce paragraphe, nous présentons succinctement chacune de ces classes.

Les actions

Une action est un titre émis par une société de capitaux, représentatif d'une partie du capital social de cette société. Elle confère à son acquéreur (l'actionnaire) la qualité d'associé au sein de la société. Les actionnaires disposent de droits fondamentaux : des droits pécuniaires ainsi que des droits de participation à la vie sociale de l'entreprise. Parmi les droits pécuniaires se trouvent le droit aux bénéfices²⁰, le droit au remboursement et au boni de liquidation²¹, le droit préférentiel

19. *Dynamical Conditional Correlation - Generalized Auto-Regressive Conditionally Heteroskedastic*

20. Le montant du dividende est fonction de l'évolution des résultats de la société et de l'affectation qui leur est donnée, ce qui fait de l'action une valeur à revenu variable. Le risque de la perte du capital pris par l'actionnaire est compensé par une possibilité de revalorisation de l'action : on parle alors de plus-value.

21. Sommes que se partagent les actionnaires d'une société dissoute après que les actifs aient été réalisés, que les créanciers et le personnel aient été payés et les apports aient été repris, le boni représente les bénéfices qui n'ont pas

de souscription à une augmentation de capital²² ou bien encore le droit d'attribution en cas de distribution gratuite²³. Parmi les droits de participation à la vie sociale de l'entreprise se trouvent le droit de vote dans les assemblées ainsi que le droit à l'information²⁴.

Différentes sortes d'actions sont souvent évoquées pour diversifier un portefeuille. Parmi ces types d'actions, on trouve le plus souvent :

- Des actions regroupées par secteurs d'activité. Les actions sont souvent regroupées par secteurs du fait de l'évolution supposée corrélée entre des actions d'entreprises appartenant à un même secteur industriel. La classification ICB²⁵ utilisée par Dow Jones, FTSE ou encore Datastream distingue, par exemple, 10 types de secteurs industriels : pétrole et gaz, matériaux de base, industries, biens de consommation, santé, services aux consommateurs, télécommunications, services aux collectivités, sociétés financières et technologies. Chaque secteur est divisé en un certain nombre de sous-secteurs.
- Les actions de type *Growth* et *Value*. Les actions répertoriées *Growth* sont des titres où, d'après la réalisation de leurs bénéfices attendus, une valorisation de leur prix de marché est anticipée. Par exemple, pour l'indice *S&P500/Citigroup Growth Index*, les facteurs utilisés pour déterminer quelles actions sont de style *Growth* sont les suivants :
 - Le taux de croissance du bénéfice par action²⁶ sur 5 ans
 - Le taux de croissance du ratio *sales-per-share*²⁷ sur 5 ans
 - Le taux de croissance interne²⁸ sur 5 ansLes actions répertoriées *Value* sont des titres ayant une valeur considérée décotée. On s'attend donc à une réévaluation de ces titres en fonction de facteurs de risque fondamentaux. Par exemple, pour l'indice *S&P500/Citigroup Value Index*, les facteurs utilisés pour déterminer quelles actions sont de style *Value* sont :
 - Le ratio *book-value-to-price*²⁹.
 - Le ratio *cash-flow-to-price*³⁰.
 - Le ratio *sales-to-price*³¹.
 - Le rendement en dividendes³².
- Les actions Small cap (resp. Mid cap et Large cap) sont des titres ayant une capitalisation de marché considérée faible (resp. moyenne et forte).
- Les actions dites émergentes correspondent à des actions d'entreprises situées dans des pays

été distribués au cours de la vie de l'entreprise.

22. Droit de l'actionnaire lui permettant de souscrire de nouvelles actions lors d'une augmentation de capital de l'entreprise.

23. Les attributions d'actions gratuites sont des opérations réservées aux anciens actionnaires d'une société. Le capital est ainsi réparti sur un plus grand nombre d'actions.

24. Droit destiné à permettre à l'actionnaire de contrôler la gestion des affaires sociales, à assurer une information financière légale, à connaître la stratégie de l'entreprise, les risques de l'activité de l'entreprise et perspectives financières de l'entreprise ; tout ceci de manière exacte, précise et sincère.

25. *Industry Classification Benchmark*.

26. Bénéfice net d'une société divisé par le nombre d'actions qui composent son capital.

27. Ratio entre les revenus totaux net gagnés dans une année fiscale et la moyenne du nombre d'actions en cours pour cette année fiscale.

28. Taux de croissance des capitaux propres.

29. Ratio entre la valeur comptable des capitaux propres et leur valeur de marché.

30. Ratio entre le cash-flow par action (capacité d'autofinancement par action) et le prix de marché de l'action.

31. Ratio entre le chiffre d'affaire par action sur les 12 mois précédent et le prix de marché de l'action.

32. Ratio entre le dividende par action par an et le prix de marché de l'action.

ditions émergentes (par exemple certains pays d'Amérique Latine, de l'Asie ou encore de l'Europe).

- Les actions de sociétés immobilières correspondent à des actions de sociétés dont une majeure partie des revenus est généré par des activités immobilières. Par exemple, les actions comprises dans l'indice FTSE EPRA/NAREIT Eurozone correspondent à des actions de sociétés dont 75% au minimum de l'EBITDA³³ est généré par des activités immobilières, tout en ayant une capitalisation minimum (≥ 50 millions d'euros) et un seuil de liquidité par rapport aux capitaux échangés minimum (≥ 25 millions d'euros par an).

Les obligations

Une obligation est un titre de créance représentant la part d'un emprunt émis par un État, une collectivité publique, une entreprise nationale ou une société privée. Elle donne le droit à son titulaire de percevoir une rémunération appelée « intérêt ». Si cet intérêt est à taux fixe, l'émetteur s'engage à verser un revenu constant pendant toute la durée de l'emprunt. Le rendement d'une obligation se compose de la combinaison des taux d'intérêts et de la plus ou moins value issue de l'achat de l'obligation.

Les taux de change

Le taux de change correspond à la valeur d'échange d'une monnaie contre une autre.

Les futures sur matières premières

Cette classe d'actifs regroupe des contrats à terme (ou futures) concernant des livraisons de marchandises. Ces contrats consistent à échanger un certain type de marchandise à une date future. Le contrat spécifie donc la qualité et la quantité de marchandises désirées ainsi que l'heure et le lieu de livraison. Les producteurs peuvent ainsi se couvrir contre les risques de fluctuations des prix. Ce type de contrat leur garanti un prix prédéterminé d'achat ou de vente de leurs marchandises.

Par exemple, les indices *S&P GSCI Excess Return Index* correspondent à des investissements dans des contrats à termes concernant certains types de marchandises. Ces contrats sont « rollés » au début de chaque mois car certains des contrats sous-jacents expirent le mois suivant. En effet, avant l'expiration du contrat, l'indice doit décaler son exposition sur un autre contrat future en vendant le contrat qui arrive à échéance et en achetant le contrat suivant correspondant.

Les liquidités

Le terme de liquidités désigne l'ensemble des fonds dont le détenteur peut disposer sans délai. On suppose que les liquidités sont stockés sur une sorte de compte bancaire, le taux d'intérêt de ce compte étant généralement assimilé à un taux d'intérêt sans risque.

1.4.2 L'analyse des prix de marché

Afin de sélectionner un marché, un secteur ou un titre en particulier, une phase d'analyse des prix de marché est nécessaire. Les gérants utilisent en général deux types d'analyse. Une analyse dite fondamentale et une analyse dite technique.

33. Earnings Before Interest, Taxes, Depreciation and Amortization (revenus avant intérêts, impôts, dotations aux amortissements et provisions sur immobilisations).

L'analyse fondamentale des prix de marché

L'analyse fondamentale se base sur la théorie d'évaluation des entreprises (*firm foundation theory*) qui fut développée dans les années 30 et finalement mise au point par John Burr Williams (*theory of investment value*). Elle a été popularisée par Benjamin Graham et David Dodd dans leur livre « *Security Analysis* » (1934) et par le livre de Graham « *Intelligent Investor* » [Gra(1949)]. L'un de ses plus fervents adeptes est le célèbre milliardaire *Warren Buffet*.

Le but de l'analyse fondamentale d'un titre est de trouver et d'explorer toutes les variables économiques qui influencent les gains futurs de cet actif financier. Ces variables fondamentales sont liées à certaines caractéristiques économiques, comme, par exemple, des circonstances macroéconomiques (inflation, taux d'intérêts, prix du pétrole, croissance du PIB, chômage, etc...), des caractéristiques spécifiques à certains secteurs d'activité (état de la concurrence, adéquation offre/demande, découvertes technologiques,...) ou encore des caractéristiques spécifiques aux entreprises (taux de croissance, dividendes, bénéfices, procès, grèves,...). En se basant sur ces fondamentaux économiques, un analyste fondamental essaye de calculer la vraie valeur sous-jacente, aussi appelée la valeur fondamentale, d'un actif financier.

L'analyse fondamentale considère que la valeur d'un titre lié à une entreprise est essentiellement celle des flux futurs actualisés que cet actif reversera. La valeur fondamentale d'un titre se base donc sur des données passées et sur les espérances de développements futurs que ces données engendrent. Des changements dans la valeur fondamentale de l'actif ne peuvent être dus qu'à des informations nouvelles concernant les variables économiques censées avoir un impact sur l'actif. Si la valeur fondamentale est évaluée plus élevée que le prix du marché, l'analyste fondamental conclue que le marché sous-évalue l'actif, il a alors affaire avec une opportunité d'investissement. La philosophie sous-jacente de l'analyse fondamentale est donc que le processus d'offre/demande concernant un titre fait converger sa valeur de marché vers sa valeur fondamentale : lorsque assez de gérants réalisent que le marché n'évalue pas correctement le prix de l'actif, le mécanisme d'offre/demande du marché forcera le prix de l'actif à converger vers sa valeur fondamentale. Si certains analystes fondamentaux ont un meilleur accès à l'information et possèdent un système plus sophistiqué pour interpréter et pondérer l'influence de l'information sur les gains futurs, ils pourront ainsi gagner plus que les autres.

Néanmoins, les gérants se basant sur des fondamentaux ont à étudier de nombreuses variables économiques différentes et ils doivent connaître les effets précis de ces variables sur les flux futurs de l'actif. On peut donc se demander si l'analyse fondamentale complète d'un titre est réalisable en pratique. En plus de ce problème, il peut arriver que le prix d'un actif dévie de manière persistante de sa valeur fondamentale, par exemple si de nombreux gérants surréagissent à une nouvelle information. Dans ce cas, une stratégie court-termiste basée sur une analyse fondamentale ne serait pas profitable, il faudrait donc appliquer une telle stratégie à plus long terme. Si un gérant n'a pas la patience et/ou la richesse suffisante pour attendre que la convergence vers la valeur fondamentale arrive, on peut se demander si une stratégie de long terme serait même réellement profitable. Enfin, on peut remettre en question la supposition que les variables fondamentales impactent les marchés financiers. En effet, on peut envisager, à l'inverse, que ce sont les marchés financiers qui impactent les valeurs fondamentales.

Au final, il apparaît qu'évaluer la valeur fondamentale d'un titre est une tâche délicate et qu'une stratégie basée sur la convergence du prix de marché vers cette valeur est assez hasardeuse.

Une manière d'éviter cette problématique est de considérer que les espérances des investisseurs se reflètent dans le prix de marché plus ou moins graduellement grâce au mécanisme de formation du

cours de Bourse, basé sur un équilibre entre offre et demande. Un gérant pensant de cette manière quitte le domaine de l'analyse fondamentale pour l'analyse technique.

L'analyse technique des prix de marché

Dans son livre « *The stock market barometer* » [Ham(1922)], William Peter Hamilton a posé les fondations de la théorie de Dow, la première théorie d'analyse graphique des cours. La théorie est basée sur des éditoriaux de Charles H. Dow quand il était éditeur du *Wall Street Journal* durant la période 1889-1902. Robert Rhea a popularisé cette idée dans les années 30, notamment dans son livre « *The Dow Theory* » [Rhe(1932)]. Charles Dow pensait notamment que les attentes des investisseurs concernant l'économie nationale à venir impactaient les ordres de Bourse d'une manière globale et causaient ainsi des augmentations ou des chutes de prix sur des horizons de long terme, tout ceci en avance par rapport aux développements économiques de l'instant présent. Pour quantifier cette théorie, Charles Dow a commencé à calculer des moyennes pour mesurer les mouvements de marché. Ceci a conduit à l'existence du Dow-Jones Industrial Average en mai 1896 et du Dow-Jones Railroad Average en Septembre 1896. La théorie de Dow suppose que toute l'information est contenue dans les moyennes, par conséquent, qu'aucune autre information n'est nécessaire pour pouvoir prendre des décisions de gestion. La théorie distingue trois types de mouvements de marché : les mouvements primaires (major movements), les mouvements secondaires (intermediate movements) et les mouvements tertiaires (minor movements). Ces mouvements sont également nommés « tendances ». Le but de la théorie est de détecter des changements de tendances primaires le plus tôt possible. Les tendances tertiaires, plus difficiles à identifier, tendent à être influencées par des informations de moindre importance et intervenant de manière aléatoire. Selon la théorie de Dow, les marchés baissiers et haussiers correspondent à des tendances primaires allant, respectivement, de haut en bas et de bas en haut et sont divisibles en différentes étapes reflétant l'état d'esprit des investisseurs.

Après le travail d'Hamilton et de Rhea, la littérature concernant l'analyse technique fut étendue et affinée par des pionniers comme Richard Schabacker, Robert Edwards, John Magee, et plus tard par Welles Wilder et John Murphy. L'analyse technique a ainsi été développée comme un outil standard utilisé par de nombreux praticiens pour prédire les mouvements de prix futurs de n'importe quelle classe d'actif (stocks, bonds, futures, options). Aujourd'hui l'analyse technique est devenue une industrie en elle-même. De nombreux programmes d'analyse technique sont vendus sur le marché et les newsletters et journaux d'analyse technique fleurissent.

L'analyse technique essaye de prédire les mouvements de prix futurs en se basant sur les mouvements de prix passés. De façon générale, les analystes techniques considèrent qu'il est inutile d'observer les fondamentaux car tout ce qui se passe à travers le monde se retrouve incorporé dans les prix. Ce dernier, issu du mécanisme entre l'offre et la demande, reflète les rêves, espérances, croyances, caprices et cauchemars de l'ensemble des investisseurs du marché. Leur philosophie est donc que l'information est graduellement incorporée dans le prix d'un actif. Ainsi, il n'existe pas de mouvements de prix abruptes qui incorporent immédiatement l'ensemble de l'information disponible, excepté lors d'un crash du marché, évènement très rare. Les mouvements de prix possèdent donc des tendances déterminées par les attitudes et les sentiments des investisseurs. Les changements d'attitudes peuvent être de nature psychologique, politique, monétaire, économique... L'art de l'analyse technique est d'identifier un renversement de tendance assez tôt pour pouvoir profiter de cette tendance jusqu'à ce qu'on trouve une preuve que la tendance va ou s'est renversée. Elle tente ainsi de détecter des changements dans les prix le plus rapidement possible.

Les règles de gestion basée sur l'analyse technique les plus populaires sont les techniques de « trend following » (suivi de tendance) telles que les moyennes mobiles et les filtres. Parmi les indicateurs techniques les plus populaires, on trouve par exemple :

- le « Moving Average Convergence Divergence » (*MACD*). Cet indicateur mesure la direction de la tendance des cours. Il correspond à la différence de deux moyennes mobiles exponentielles calculées sur des horizons de longueurs différentes. La formule générale d'une moyenne mobile exponentielle est la suivante :

$$MME(t, l, \alpha) = \frac{\sum_{i=0}^l \alpha(1 - \alpha)^i P_{t-i}}{\sum_{i=0}^l \alpha(1 - \alpha)^i} \quad (1.1)$$

avec P_t le prix de marché du titre à l'instant t et $\alpha \in]0, \dots, 1[$ le paramètre permettant de pondérer plus ou moins fortement les derniers prix de marché disponibles. Lorsque α tend vers 0, ce filtre tend vers une moyenne mobile équipondérée. Le *MACD* est en général évalué sur des prix quotidiens comme :

$$MACD_t = MME(t, 12, 0.1) - MME(t, 26, 0.1)$$

Lorsque la *MACD* est positive, la *MME* de court terme est supérieure à la *MME* de long terme, le titre est alors supposé suivre une tendance haussière.

- le « Relative Strength Index » (*RSI*). Cet indicateur mesure la régularité avec laquelle les cours montent ou baissent sur une période donnée. Sa formule générale est la suivante :

$$RSI_t(n) = 1 - \frac{1}{1 + \frac{H_t(n)}{B_t(n)}}$$

avec $H_t(n)$ (resp. $B_t(n)$) la moyenne des hausses (resp. baisses) sur les n derniers jours. Lorsque le *RSI* est environ au-dessus de 70 (en dessous de 30), on parle de « sur-achat » (de « sur-vente »).

- le « Momentum ». Cet indicateur mesure la cyclicité des hausses et des baisses. Sa formule générale est la suivante :

$$MOM_t(n) = P_t - P_{t-n}$$

Afin de bien cerner la problématique de l'analyse technique, il est intéressant d'étudier l'exemple des filtres d'Alexander. [Ale(1961)] proposa de définir des filtres pour extraire de possibles tendances dans les prix pouvant être masquées par des mouvements du marché de faibles amplitudes. Une stratégie de filtre achète quand le prix augmente de $x\%$ par rapport à un plus bas récent et vend lorsque le prix décline de $x\%$ par rapport à un plus haut récent. Il applique plusieurs filtres au DJIA durant la période 1897-1929 et au *S&P* Industrials durant la période 1929-1959. Alexander conclue que, dans les marchés spéculatifs, un changement de prix, une fois initié, tend à persister. Par conséquent, son étude conforte la philosophie sous-jacente à l'analyse technique selon laquelle les prix évoluent en tendance. Néanmoins, il remarque que les différents coûts de transactions peuvent réduire les résultats trouvés. Deux ans plus tard, [Man(1963)] remarque qu'il y a un défaut dans les calculs d'Alexander puisque ce dernier, dans [Ale(1961)], suppose que le gérant peut acheter exactement lorsque le prix augmente de $x\%$ et vendre lorsque le prix baisse de $x\%$. Or, ceci paraît peu probable dans la réalité. En 1964, [Ale(1964)] corrige les erreurs de calculs remarquées par Mandelbrot et prend en compte les coûts de transactions. La règle du filtre permet d'engranger des profits considérables,

supérieurs à une stratégie auto-financée, mais les coûts de transactions suppriment tous les profits. On voit donc que ces coûts de transactions sont essentiels dans l'efficacité d'une telle règle de gestion.

Une des principales critiques de l'analyse technique est qu'elle ne détecte une tendance que lorsque celle-ci est déjà établie, puisqu'elle se base sur les mouvements de prix passés. Le temps qu'une tendance soit signalée, elle peut déjà être terminée. Les analystes techniques ne réagiraient ainsi que trop tard. D'autant que, comme noté par [Osl(1995)], les livres d'analyse technique échouent à documenter la validité de leurs affirmations. Les auteurs n'hésitent pas à caractériser un modèle comme fréquent et fiable sans essayer de quantifier ces affirmations. Le manque d'analyse statistique robuste provient de la difficulté à programmer des techniques de reconnaissance de figures issue de l'analyse technique. De nombreuses figures d'analyse technique semblent ainsi de vagues suppositions sans justification précise. Une autre critique est que l'analyse technique a une nature très subjective. Il y a probablement autant de méthodes pour combiner et interpréter plusieurs indicateurs techniques qu'il y a d'analystes techniques. Ainsi, il arrive bien souvent que l'on compare l'analyse fondamentale à l'analyse technique comme l'astronomie est comparée à l'astrologie. L'analyse technique serait une sorte de finance « voodoo ». L'attitude des académistes envers l'analyse technique est bien décrite par [Mal(1996)] (p.139) : « Obviously, I'm biased against the chartist. This is not only a personal predilection but a professional one as well. Technical analysis is anathema to the academic world. We love to pick on it. Our bullying tactics are prompted by two considerations : (1) after paying transaction costs, the method does not do better than a buy-and-hold strategy for investors, and (2) it's easy to pick on. And while it may seem a bit unfair to pick on such a sorry target, just remember : it's your money we are trying to save ».

Néanmoins, les analystes techniques ont conscience que leurs méthodes ne sont en aucun cas irréfutables. Par exemple, [Pri(1998)] (p.5) remarque sur l'analyse technique « it can help in identifying the direction of a trend, but there is no known method of consistently forecasting its magnitude ». [Edw(2001)] (p.12) remarquent que « chart analysis is certainly neither easy nor foolproof ». Finalement, [Ach(1995)] (p.6) remarque « ..., I caution you not to let the software lull you into believing markets are as logical and predictable as the computer you use to analyse them ». Par conséquent, même les analystes techniques avertissent contre les décisions d'investissement basées sur leurs seules analyses.

Comparaison entre analyse fondamentale et technique

L'analyse technique possède un avantage important sur l'analyse fondamentale : elle peut être appliquée assez facilement, pour un coût assez faible et pour tout type de titres. Elle nécessite néanmoins une certaine pratique pour reconnaître certaines figures, mais, en principe, tout le monde peut l'appliquer. On pourrait objecter que l'analyse fondamentale peut également être pratiquée d'une manière assez simple. Par exemple, le nombre de voitures garées au centre commercial peut être une indication de la confiance des consommateurs dans l'économie nationale. Néanmoins, ce genre de conclusion est assez hâtive et peut être facilement erronée. L'analyse fondamentale d'un titre nécessite d'avoir accès à de nombreuses statistiques économiques fiables et de prendre le temps de les étudier. Ce qui la rend finalement bien plus couteuse que l'analyse technique.

Un autre avantage de l'analyse technique est la facilité avec laquelle on peut tester sa puissance de prédiction alors qu'il est plus difficile de le faire pour des analyses fondamentales. En effet, afin de tester des règles de gestion basées sur l'analyse technique, on a juste besoin des prix, volumes ou dividendes, ce qui peut s'obtenir facilement.

L'analyse fondamentale et l'analyse technique utilisent toutes deux des historiques de données mais d'une manière différente. L'analyste technique soutient que toute l'information est graduellement incorporer dans les prix, elle utilise donc essentiellement des historiques de prix. Tandis que l'analyste fondamentale tente d'utiliser directement toute l'information disponible représentée par des variables économiques. Elle utilise donc des historiques de quantités économiques afin de calculer la valeur fondamentale du titre.

Une autre différence importante entre ces deux types d'analyse est que l'analyste technique ne publiera jamais un prix précis. En effet, il ne se base que sur les signaux d'achats et de ventes que sa stratégie génère. A l'inverse, l'analyste fondamental a pour but de publier un prix précis correspondant à une valeur fondamentale calculée.

Pour conclure, ni l'analyse fondamentale, ni l'analyse technique ne permettent d'enregistrer des profits de façon certaine.

1.4.3 Le processus d'allocation d'actifs

Le processus d'allocation d'actifs correspond au processus consistant à investir de l'argent dans différents actifs financiers. Il n'existe pas de méthodologie clairement définie pour ce processus. Néanmoins, certains concepts sont régulièrement évoqués pour caractériser un processus d'allocation d'actifs. On distingue ainsi trois principales étapes à un processus d'investissement : l'allocation stratégique, l'analyse des titres et l'allocation tactique.

Allocation stratégique, analyse des titres et allocation tactique

Dans un premier temps, l'*allocation stratégique* a pour but de définir une stratégie d'allocation de long terme en spécifiant une philosophie, un univers et un horizon d'investissement. Le gestionnaire y définit les différentes classes d'actifs (« l'univers » d'investissement) et leurs poids approximatifs respectifs dans le portefeuille. Le point de départ de l'allocation est en général le niveau de tolérance de l'investisseur à l'égard du risque encouru. Ce niveau de tolérance permet de définir l'objectif de l'allocation d'actifs du portefeuille. Ainsi, un gestionnaire peut décider de créer un portefeuille composé d'environ 75% d'obligations et de 25% d'actions, il prendra alors moins de risques qu'un investisseur détenant environ 75% d'actions et 25% d'obligations. Ce gestionnaire peut également décider d'investir essentiellement dans des actions européennes ayant une faible capitalisation lorsqu'il investi en actions. Ce choix correspond alors à une certaine « philosophie » d'investissement. En pratique, l'allocation stratégique prend relativement peu de temps comparé à l'ensemble du processus d'allocation d'actifs.

Après cette allocation stratégique, vient, la plupart du temps, une phase d'analyse des titres afin de sélectionner les titres dans lesquels l'investisseur aura la possibilité d'investir ou non. Cette phase d'analyse des titres est en général soit fondamentale, soit technique comme nous l'avons vu précédemment. Pour le marché action, on parle également de « stock-picking ». Le stock-picking repose sur l'idée que ce sont les titres en eux-mêmes, et non l'allocation des poids accordés à chacun, qui apportent le plus de valeur ajoutée dans un portefeuille. En général, les gestionnaires tentent d'effectuer une analyse fondamentale de chaque titre la plus complète possible. Ils consacrent donc beaucoup de temps à cette phase du processus d'investissement.

Après la phase d'analyse des titres, l'univers définitif d'investissement est déterminé. Il ne reste plus qu'à définir les poids du portefeuille accordés à chaque titre. Cette étape est appelée *allocation tactique*. Elle consiste à modéliser les opportunités d'investissements, à définir des mesures précises

de performances et de risques, à implémenter des stratégies d'investissement. Elle tire parti des mouvements de marché à court terme, ce qui peut entraîner un écart, sur le court terme, par rapport à l'allocation stratégique initiale. L'allocation tactique se déroule en trois étapes. Dans un premier temps, les futurs rendements d'un investissement dans chacun des titres sont prédits. Grâce à ces prédictions, l'investisseur peut ensuite déterminer les poids associés à chaque titre dans son portefeuille, en maximisant une fonction d'utilité par exemple. Enfin, les performances de l'allocation sont testées. Suivant les résultats obtenus, l'allocation des poids de chaque actif dans le portefeuille peut être entièrement revue et modifiée.

Un exemple classique d'allocation tactique : le modèle de Markowitz

Le modèle de [Mar(1952)] suppose que l'investisseur souhaite optimiser son portefeuille de telle façon que l'espérance du rendement à venir soit maximale et que le risque associé, représenté par la variance du rendement à venir, soit minimale. Formellement, l'espérance et la variance des rendements du portefeuille peuvent s'écrire de la façon suivante :

$$\begin{aligned}\mathbb{E}[R_{t,t+\Delta t}^P] &= \sum_{i=1}^N \rho^{a_i}(t) \mathbb{E}[R_{t,t+\Delta t}^{a_i}] \\ \mathbb{V}[R_{t,t+\Delta t}^P] &= \sum_{i=1}^N (\rho^{a_i}(t))^2 \mathbb{V}[R_{t,t+\Delta t}^{a_i}] + 2 \sum_{\substack{i,j=1 \\ i>j}}^N \rho^{a_i}(t) \rho^{a_j}(t) \text{cov}(R_{t,t+\Delta t}^{a_i}, R_{t,t+\Delta t}^{a_j})\end{aligned}$$

avec $\rho^{a_i}(t)$ le poids de l'actif a_i dans le portefeuille à l'instant t et $R_{t,t+\Delta t}^{a_i}$ le rendement de l'actif a_i entre t et $t + \Delta t$. On doit donc minimiser $\mathbb{V}[R_{t,t+\Delta t}^P]$ et maximiser $\mathbb{E}[R_{t,t+\Delta t}^P]$ par rapport aux poids $\rho^{a_i}(t)$. Pour se faire, la détermination de l'ensemble des portefeuilles qui minimisent la variance pour une espérance de rentabilité fixée passe par la résolution du programme d'optimisation quadratique en ω suivant :

$$\min_{\omega} \omega^T \Sigma \omega$$

sous les contraintes $\omega^T e = 1$ et $\omega^T R = \mathbb{E}[R_{t,t+\Delta t}^P]$. Avec ω le vecteur des poids, R le vecteur composé des $\mathbb{E}[R_{t,t+\Delta t}^{a_i}]$ et Σ la matrice de variance-covariance entre les rendements de tous les actifs.

Si le modèle de Markowitz reste populaire du fait de sa simplicité, il apparaît néanmoins inapproprié en pratique. En effet, cette manière de procéder possède trois principales lacunes :

- elle n'optimise le portefeuille que sur une seule période de temps.
- elle ne prend pas en compte les caractères asymétriques et leptokurtiques de la répartition des rendements constatés en pratique.
- elle ne tient pas compte des coûts de transactions.

Les caractères asymétriques et leptokurtiques des rendements peuvent être pris en compte en prenant une autre mesure de risque que la variance. Par exemple, on pourrait tenter d'estimer le risque en utilisant la VaR ou la CVaR [Kro(2002)]. Plusieurs études ont également proposées d'incorporer les coûts de transactions dans cette optimisation de portefeuille (voir par exemple [Lob(2006)], [Dyb(2005)]). L'optimisation sur plusieurs périodes de temps peut être effectuée par une approche d'optimisation dynamique. Afin d'obtenir des solutions optimales, le principe d'optimalité de Bellman et celui du minimum de Pontryagin sont souvent utilisés [Fle(1975)]. Plusieurs publications sont parues sur ce sujet. On citera par exemple [Sam(1969)], [Dan(1993)]. Quand les variables de contrôles

(par exemple les poids) sont contraintes, les problèmes d'optimisation dynamique deviennent très difficiles à résoudre. Une technique pour résoudre de tels problèmes est le « contrôle prédictif de modèle » (voir [Her(2005)]).

Outre ces problématiques, le modèle de Markowitz ne se base que sur des historiques de rendements passés. Il repose donc sur une hypothèse d'efficience des marchés qui reste discutable (voir paragraphe 1.2.3). Si on suppose que les marchés ne sont pas efficaces, nous avons vu que certaines études parvenaient à mettre en avant un certain nombre de facteurs de risque ayant un pouvoir prédictif significatif sur les rendements à venir. Il faut alors relier ces différents facteurs aux rendements à venir en utilisant un modèle mathématique permettant d'obtenir des prédictions fiables et utilisables en temps réel au sein de l'allocation.

1.5 Conclusion

L'hypothèse d'efficience informationnelle des marchés est difficilement vérifiable et se confronte à un certain nombre d'anomalies constatées en pratique. La cause de ces anomalies pourrait notamment être le comportement parfois irrationnel des investisseurs actifs du marché. Dans ce contexte, de nombreux facteurs, plus ou moins mesurables en temps réel, impactent les prix de marché et rendent leur dynamique complexe à modéliser. Néanmoins, un gérant de portefeuille se doit d'optimiser l'allocation de son portefeuille et donc d'anticiper cette dynamique s'il veut créer de la performance. Pour cela, nous avons montré que la littérature a mis en avant un certain nombre de variables économiques et financières ayant un pouvoir prédictif sur la direction des prix agrégés de marché.

Pour optimiser l'allocation d'un portefeuille, trois étapes essentielles et distinctes ont été exposées : l'allocation stratégique, l'analyse des titres et l'allocation tactique. L'allocation stratégique est fondamentale car elle a pour but de tirer profit des tendances de long terme au sein de différentes classes d'actifs. Elle accorde un certain poids dans le portefeuille à chaque classe d'actifs. Afin de déterminer ces poids, il semble nécessaire de définir en quoi consiste la stratégie optimale que doit adopter un investisseur souhaitant optimiser son portefeuille. Dans le prochain chapitre, nous mettons en avant un certain nombre d'évènements nécessaires à la définition de cette stratégie optimale à adopter.

1.5. CONCLUSION

Optimisation de portefeuille et estimation de tendances

Sommaire

2.1	L'optimisation classique « moyenne-variance »	45
2.1.1	Optimisation « moyenne-variance »	46
2.1.2	Hypothèse d'ergodicité et propriétés des estimateurs en échantillons finis . . .	49
2.2	Formalisation d'une stratégie d'investissement en temps discret	49
2.3	Définition de la stratégie optimale	52
2.4	Approximation de la stratégie optimale	55
2.5	Conclusion	58

Introduction

L'objectif de ce chapitre est de montrer en quoi l'estimation d'une tendance au sein de l'évolution de la valeur d'un investissement est importante pour optimiser l'allocation d'un portefeuille de titres. Dans un premier temps, nous présentons l'approche classique « moyenne-variance » souvent utilisée pour optimiser un tel portefeuille de titres et les hypothèses sur lesquelles elle se base. On présente ensuite une modélisation en temps discret de l'évolution d'un portefeuille de titres. Cette modélisation nous permet de mettre en avant un certain nombre d'évènements à partir desquels un investisseur peut déduire la stratégie optimale à adopter (c'est-à-dire celle lui rapportant le plus). L'estimation d'une tendance au sein des rendements des investissements permet alors de quantifier les probabilités de réalisation de ces évènements.

2.1 L'optimisation classique « moyenne-variance »

Dans ce paragraphe, nous relatons l'approche classique moyenne-variance utilisée dans de nombreux articles (voir par exemple [Mer(1971)], [Zhou(1999)], [Baj(1998)], [Che(2005)]) et la formulation du problème d'optimisation de portefeuille qu'elle implique.

2.1.1 Optimisation « moyenne-variance »

Supposons qu'un gérant de portefeuille ait le choix d'investir dans un portefeuille composé de $N + 1$ actifs : N actifs dits « risqués » et un actif dit « sans risque ». On suppose que le prix de l'actif sans risque $S_0(t)$ suit la dynamique suivante :

$$\begin{aligned} dS_0(t) &= r(t)S_0(t)dt \\ S_0(0) &= s_0(0) > 0 \end{aligned} \quad (2.1)$$

avec $r(t) > 0$ une fonction déterministe du temps modélisant le taux d'intérêt sans risques. Dans ce cadre, $S_0(t)$ est donc une fonction déterministe du temps. Chaque actif risqué A_i possède un prix $S_i(t)$ correspondant à une variable aléatoire suivant une dynamique définie par l'équation différentielle stochastique suivante :

$$\begin{aligned} dS_i(t) &= \lambda_i(t)S_i(t)dt + S_i(t) \sum_{j=1}^N \sigma_{ij}(t)dW_j(t) \\ S_i(0) &= s_i(0) > 0 \end{aligned} \quad (2.2)$$

où les $(W_j(t))_{N \geq j \geq 1}$ sont des mouvements Browniens indépendants définis sur un espace de probabilité filtré complet $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$, $\lambda_i(t)$ un paramètre de tendance (ou *drift*) déterministe et $\sigma(t) = (\sigma_{ij}(t))_{N \times N}$ une matrice de variance-covariance déterministe. On précise que la mesure de probabilité \mathbb{P} considérée ici est une mesure de probabilité historique.

Dans ce cadre, $\sum_{j=1}^N \sigma_{ij}(t)dW_j(t)$ représente la partie aléatoire de l'évolution du prix et $\lambda_i(t)dt$ sa partie déterministe. Plus particulièrement, les paramètres $(\sigma_{ij}(t))_{i \in [1, \dots, N], j \in [1, \dots, N]}$ permettent de corrélérer l'évolution de certains prix.

Soit V_t la richesse totale du portefeuille à l'instant $t \geq 0$. Supposons que le gérant possède $N_i(t)$ parts du $i^{\text{ème}}$ titre à l'instant t . Dans ce cas,

$$V_t = \sum_{i=0}^N N_i(t)S_i(t) \quad \forall t \geq 0 \quad (2.3)$$

Plaçons-nous maintenant dans le cas où les coûts de transaction n'existent pas et où les titres sont négociés de manière continue. En utilisant le lemme d'Itô, on obtient

$$dV_t = \sum_{i=0}^N S_i(t)dN_i(t) + N_i(t)dS_i(t) + \sum_{i=0}^N d \langle N_i, S_i \rangle_t \quad (2.4)$$

où $\langle N_i, S_i \rangle_t$ correspond au processus de variation quadratique entre $N_i(t)$ et $S_i(t)$. On peut raisonnablement supposer que $N_i(t)$ est une fonction déterministe du temps suivant la dynamique suivante :

$$dN_i(t) = \mu_i(t)N_i(t)dt \quad (2.5)$$

avec $\mu_i(t)$ un paramètre de tendance déterministe. Par conséquent, $d \langle N_i, S_i \rangle_t = 0$ pour tout $i \in [0, \dots, N]$ et

$$dV_t = N_0(t)S_0(t)(\mu_0(t) + r(t))dt + \sum_{i=1}^N (\lambda_i(t) + \mu_i(t))N_i(t)S_i(t)dt + \sum_{i=1}^N \sum_{j=1}^N N_i(t)S_i(t)\sigma_{ij}(t)dW_j(t) \quad (2.6)$$

Il s'en suit que la richesse finale de l'investisseur correspond à :

$$V_T = V_0 + \int_0^T (\mu_0(t) + r(t)) N_0(t) S_0(t) dt + \sum_{i=1}^N \int_0^T (\lambda_i(t) + \mu_i(t)) N_i(t) S_i(t) dt + \sum_{i=1}^N \sum_{j=1}^N \int_0^T N_i(t) S_i(t) \sigma_{ij}(t) dW_j(t) \quad (2.7)$$

où $V_0 = v_0 > 0$.

Lorsque l'investisseur se trouve en $t = 0$, il va chercher à maximiser cette richesse finale. Pour cela, il va maximiser sa richesse finale espérée, représentée par $\mathbb{E}[V_T | \mathcal{F}_0]$, tout en minimisant l'incertitude liée à cette richesse finale représentée par la variance $\mathbb{V}[V_T | \mathcal{F}_0]$. Cela correspond à un problème d'optimisation multi-objectifs. Le problème d'optimisation de portefeuille moyenne-variance est alors défini par :

$$\left\{ \begin{array}{l} \text{Minimiser } u(-\mathbb{E}[V_T | \mathcal{F}_0], \mathbb{V}[V_T | \mathcal{F}_0]) \\ \text{avec } V_t \text{ suivant la dynamique précédemment évoquée} \\ \text{et } u \text{ une fonction d'utilité concave} \end{array} \right.$$

Afin d'illustrer ce problème d'optimisation, nous allons supposer que $N_i(t) = N_i$, $\lambda_i(t) = \lambda_i$ et $\sigma_{ij}(t) = \sigma_{ij}$ sont des fonctions constantes du temps afin de simplifier les calculs. Le lemme d'Itô nous permet d'exprimer dV_t et dV_t^2 :

$$\begin{aligned} dV_t &= N_0 S_0(t) r(t) dt + \sum_{i=1}^N \lambda_i N_i S_i(t) dt + \sum_{i=1}^N \sum_{j=1}^N N_i \sigma_{ij} S_i(t) dW_j(t) \\ &= [r(t) V_t + \sum_{i=1}^N (\lambda_i - r(t)) N_i S_i(t)] dt + \sum_{i=1}^N \sum_{j=1}^N N_i \sigma_{ij} S_i(t) dW_j(t) \end{aligned} \quad (2.8)$$

$$\begin{aligned} dV_t^2 &= 2V_t dV_t + d \langle V, V \rangle_t \\ &= [2r(t) V_t^2 + 2V_t \sum_{i=1}^N (\lambda_i - r(t)) N_i S_i(t) + \sum_{i=1}^N \sum_{j=1}^N N_i^2 \sigma_{ij}^2 S_i^2(t)] dt + 2V_t \sum_{i=1}^N \sum_{j=1}^N N_i \sigma_{ij} S_i(t) dW_j(t) \end{aligned} \quad (2.9)$$

On peut ensuite en déduire les expressions suivantes :

$$d\mathbb{E}[V_t | \mathcal{F}_0] = (r(t) \mathbb{E}[V_t | \mathcal{F}_0] + \sum_{i=1}^N (\lambda_i - r(t)) N_i \mathbb{E}[S_i(t) | \mathcal{F}_0]) dt \quad (2.10)$$

$$d\mathbb{E}[V_t^2 | \mathcal{F}_0] = (2r(t) \mathbb{E}[V_t^2 | \mathcal{F}_0] + 2 \sum_{i=1}^N (\lambda_i - r(t)) N_i \mathbb{E}[V_t S_i(t) | \mathcal{F}_0] + \sum_{i=1}^N \sum_{j=1}^N N_i^2 \sigma_{ij}^2 \mathbb{E}[S_i^2(t) | \mathcal{F}_0]) dt \quad (2.11)$$

avec pour conditions initiales $\mathbb{E}[V_0 | \mathcal{F}_0] = v_0 > 0$ et $\mathbb{E}[V_0^2 | \mathcal{F}_0] = v_0^2$, et où

$$\mathbb{E}[S_i(t) | \mathcal{F}_0] = s_i(0) e^{\lambda_i t} \quad (2.12)$$

$$\mathbb{E}[S_i(t) S_j(t) | \mathcal{F}_0] = s_i(0) s_j(0) e^{(\lambda_i + \lambda_j + \sum_{k=1}^N \sigma_{ik} \sigma_{jk}) t} \quad (2.13)$$

$$\mathbb{E}[V_t S_i(t) | \mathcal{F}_0] = \sum_{j=0}^N N_j \mathbb{E}[S_i(t) S_j(t) | \mathcal{F}_0] \quad (2.14)$$

La résolution de ces équations différentielles ordinaires permet de déduire les expressions de $\mathbb{E}[V_t|\mathcal{F}_0]$ et de $\mathbb{E}[V_t^2|\mathcal{F}_0]$ qui pourront ainsi, par la suite, permettre de définir $\mathbb{V}[V_t|\mathcal{F}_0] = \mathbb{E}[V_t^2|\mathcal{F}_0] - \mathbb{E}[V_t|\mathcal{F}_0]^2$ pour tout $t \in [0, T]$:

$$\mathbb{E}[V_t|\mathcal{F}_0] = v_0 e^{\int_0^t r(s)ds} + \sum_{i=1}^N N_i s_i(0) \int_0^t (\lambda_i - r(s)) e^{\lambda_i s + \int_s^t r(u)du} ds \quad (2.15)$$

$$\begin{aligned} \mathbb{E}[V_t^2|\mathcal{F}_0] = & v_0^2 e^{2\int_0^t r(s)ds} + 2 \sum_{i=1}^N \sum_{j=1}^N N_i N_j s_i(0) s_j(0) \int_0^t (\lambda_i - r(s)) e^{(\lambda_i + \lambda_j + \sum_{k=1}^N \sigma_{ik} \sigma_{jk})s + 2\int_s^t r(u)du} ds \\ & + \sum_{i=1}^N \sum_{j=1}^N N_i^2 s_i^2(0) \sigma_{ij}^2 \int_0^t e^{(2\lambda_i + \sum_{k=1}^N \sigma_{ik}^2)s + 2\int_s^t r(u)du} ds \end{aligned} \quad (2.16)$$

Le problème d'optimisation consiste alors à minimiser une fonction d'utilité $u(-\mathbb{E}[V_T|\mathcal{F}_0], \mathbb{V}[V_T|\mathcal{F}_0])$ par rapport aux parts détenues $(N_i)_{1 \leq i \leq N}$. On utilise généralement $u(x, y) = x + \alpha y$ où $\alpha \in \mathbb{R}^{+,*}$ correspond à un paramètre d'aversion au risque de l'investisseur.

Cette approche, régulièrement évoquée dans la littérature, repose sur les axiomes suivants :

1. L'existence d'un marché efficient et complet.
2. Pas de coûts de transactions ou de taxes.
3. Tous les participants du marché peuvent prêter et emprunter de l'argent au même taux d'intérêt sans risque.
4. Les titres sont négociés de manière continue.
5. L'évolution de la valeur du portefeuille dépend essentiellement du prix de marché des titres.

On suppose généralement que ces axiomes sont « approximativement » vérifiés. Néanmoins, dans la réalité, chacun de ces axiomes peut être remis en cause.

Nous avons ainsi vu, dans le premier chapitre, que les investisseurs sont soumis à des biais comportementaux et qu'il existe des anomalies à l'hypothèse d'efficience du marché. Cependant, certains, comme Shleifer [Shl(2000)], affirment qu'il n'est pas nécessaire que tous les participants du marché soient rationnels pour que le marché tende vers l'efficience. Il en est de même pour chacun des axiomes : bien qu'ils ne soient pas vérifiés en pratique, il est coutume de les supposer approximativement vérifiés.

Les coûts de transactions sont ainsi généralement considérés négligeables [Mer(1971)] bien que tous les intervenants du marché y soient néanmoins soumis. De même, bien qu'il n'existe pas de produits financiers ayant un taux sans risque dans la réalité, celui-ci est souvent approximé par des taux d'obligations d'états « bien notées » alors qu'il existe un risque de crédit [Wei(1989)]. C'est également le cas pour l'hypothèse de continuité du temps. En pratique, un investisseur devra toujours opéré en temps discret. Il a cependant été montré que, si le pas entre deux instants discrets où il est possible d'effectuer des transactions est suffisamment faible, l'approche en temps continu est une approximation asymptotique de la réalité acceptable [Mer(1974)]. Enfin, l'évolution de la valeur d'un portefeuille ne dépend pas seulement de l'évolution des prix des actifs qu'il détient mais également

de l'ensemble des flux versés par ces titres (dividendes, intérêts, etc.) ainsi que de flux exogènes (tels que des frais de gestion) auquel il est soumis.

Cette approche de l'optimisation de portefeuille est donc « discutable » car elle fait de nombreuses approximations sujettes à controverse. D'autant qu'une fois le modèle de l'évolution de richesse défini, il faut pouvoir l'inférer statistiquement d'après les séries de prix observés. Ceci pose le problème bien connu de l'inférence statistique d'un modèle probabiliste à partir d'un échantillon fini de données.

2.1.2 Hypothèse d'ergodicité et propriétés des estimateurs en échantillons finis

Supposons que l'on observe une série de rendements de prix $(R_{t,t+\Delta t})_{t \in [t_0, \dots, t_n]}$ pour $\Delta t \in \mathbb{N}$ fixé. L'optimisation « moyenne-variance » classique suppose qu'il existe un modèle probabiliste du type (2.2) générant ces rendements observés. Il faut alors estimer ce modèle le plus précisément possible.

Lorsqu'on tente d'inférer les paramètres du modèle, il est nécessaire de supposer que leur estimation empirique d'après un échantillon suffisamment grand de données converge vers la « vraie » valeur du paramètre. En d'autres termes, on suppose que la moyenne temporelle d'une fonction g des rendements converge vers son espérance mathématique :

$$\lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{t=t_0}^{t_n} g(R_{t,t+\Delta t}) = \int_{\mathbb{R}} g(x) f_{R_{t,t+\Delta t}}(x) dx \quad (2.17)$$

où $g : \mathbb{R} \rightarrow \mathbb{R}$ est une fonction appartenant à $L^1(\mathbb{R})$ et $f_{R_{t,t+\Delta t}}$ correspond à la densité de probabilité de $R_{t,t+\Delta t}$. Cette propriété est appelée « propriété d'ergodicité ».

Néanmoins, en pratique, n est toujours fini. Ainsi, même en supposant une hypothèse d'ergodicité, on n'est pas assuré de pouvoir mesurer exactement les paramètres du modèle. Il existe une certaine erreur statistique liée à la taille de l'échantillon observé. On peut alors se baser sur la théorie probabiliste pour estimer un intervalle de confiance lié à l'estimateur, ce qui nécessite de connaître sa loi de probabilité.

Avec cette approche, on suppose que l'estimateur varie suivant l'échantillon et qu'il correspond à la réalisation d'une variable aléatoire. Lorsque le nombre d'estimations faites sur différents échantillons tend vers l'infini, on peut utiliser le théorème centrale-limite en supposant une hypothèse d'indépendance et de distribution identique entre ces différentes estimations. On en déduit que la loi de l'estimateur est supposée converger asymptotiquement vers celle d'une loi normale, ce qui permet d'obtenir l'intervalle de confiance recherché.

Dans le paragraphe suivant, afin de se détacher de ces approches utilisant des processus de diffusion en temps continu, on décrit l'évolution de la valeur d'un investissement dans un portefeuille de titres en essayant de s'affranchir de trop d'hypothèses simplificatrices. On ne suppose aucune dynamique précise concernant l'évolution de la valeur d'un investissement dans chacun des titres sous-jacents. Nous mettons simplement en avant la relation qu'il existe entre ces valeurs et celle de l'ensemble du portefeuille en temps discret.

2.2 Formalisation d'une stratégie d'investissement en temps discret

Dans ce paragraphe, on s'attache à décrire l'évolution de la valeur d'un investissement dans un portefeuille de titres en temps discret. Nous partons du principe que l'évolution de cette valeur dépend d'une stratégie d'investissement prenant place au sein d'un univers d'investissement.

Définition 1. Un univers d'investissement U_{N+1} correspond à un ensemble fini de $N + 1$ différents actifs A_k dans lesquels l'investisseur a la possibilité d'investir : $U_{N+1} \equiv \{A_k | k \in [0, \dots, N]\}$.

Par la suite, on se placera dans le cas d'un petit investisseur¹ ayant la possibilité d'investir dans un univers d'investissement composé de $N + 1$ actifs risqués ($N \in \mathbb{N}$). L'actif ayant le risque le plus faible sera noté A_0 .

Définition 2. Le processus stochastique à temps discret de la valeur d'un investissement dans l'univers d'investissement U_{N+1} , noté $(V_t)_{t \in I}$, est tel que :

$$V_t = S_0(t) + \sum_{k=1}^N S_k(t) \quad (2.18)$$

où $S_k(t)$ correspond au montant investi dans chaque actif risqué A_k pour $k \in [1, \dots, N]$ et $S_0(t)$ correspond au montant investi dans l'actif ayant le risque le plus faible A_0 .

On remarque que $S_k(t)$ ne correspond pas à un prix de marché mais à un montant investi dans l'actif A_k .

Soit $I = [t_0, t_1, \dots, t_n]$ avec $n \in \mathbb{N}$ un ensemble d'instantes où l'investisseur a la possibilité de rebalancer son portefeuille et tels que $\forall i \in [1, \dots, n]$, $t_i \leq T$ où T correspond à l'horizon d'investissement. Si l'investisseur n'effectue aucun rebalancement de son portefeuille entre deux instantes t_i et t_{i+1} appartenant à I , on aura alors :

$$V_{t_{i+1}} = (1 + r_{t_i, t_{i+1}})S_0(t_i) + \sum_{k=1}^N (1 + R_{t_i, t_{i+1}}^k)S_k(t_i) \quad (2.19)$$

Avec :

- $r_{t_i, t_{i+1}}$ le rendement du montant investi dans l'actif ayant le risque le plus faible entre t_i et t_{i+1} .
- $R_{t_i, t_{i+1}}^k$ le rendement du montant investi dans l'actif risqué A_k entre t_i et t_{i+1} .

Supposons dorénavant que la stratégie d'investissement du gérant de portefeuille l'entraîne à rebalancer son portefeuille aux instantes $I^* = [t_1^*, \dots, t_m^*]$, où $m \in \mathbb{N}$ et tels que $I^* \subseteq I$, en achetant $a_t^k \geq 0$ et en vendant $-v_t^k \leq 0$ de l'actif A_k , pour tout $t \in I^*$ et tout $k \in [1, \dots, N]$ (c'est-à-dire pour les actifs risqués autres que l'actif ayant le risque le plus faible). Cette activité de rebalancement engendre des flux monétaires d'entrée et de sortie au niveau de la liquidité du portefeuille². En d'autres termes, lorsqu'on achète un titre, on soustrait un certain montant au niveau de cette liquidité et lorsqu'on en vend un, on ajoute un certain montant à cette liquidité. La liquidité du portefeuille est supposée être l'actif ayant le risque le plus faible. Si aucune transaction n'est effectuée entre deux instantes t_i et t_{i+1} , le rendement de la liquidité correspond au rendement $r_{t_i, t_{i+1}}$. $S_0(t_i)$ correspond donc au montant de cette liquidité en t_i .

On note t_i^{*-} et t_{i+1}^{*-} appartenant à I , les instantes qui précèdent, respectivement, t_i^* et t_{i+1}^* appartenant à I^* . En d'autres termes, on suppose qu'entre la prise de décision de l'investisseur et son action de rebalancer le portefeuille, il se déroule un laps de temps de longueur $t_i^* - t_i^{*-} > 0$. Entre

1. Un investisseur dont la composition du portefeuille n'affecte pas à tout instant les prix du marché
 2. Ensemble des fonds dont le détenteur peut disposer sans délai. Voir chapitre 1 paragraphe 4.1.5.

l'instant de la prise décision $t_i^{*, -}$ et la prochaine prise de décision $t_{i+1}^{*, -}$, la dynamique des positions détenues dans chaque actif k sera alors de :

$$\begin{cases} S_k(t_{i+1}^{*, -}) = (1 + R_{t_i^*, t_{i+1}^*}^k) \left((1 + R_{t_i^*, -}^k) S_k(t_i^{*, -}) + a_{t_i^*}^k - v_{t_i^*}^k \right) & \forall k \in [1, \dots, N] \\ S_0(t_{i+1}^{*, -}) = (1 + r_{t_i^*, t_{i+1}^*}) \left((1 + r_{t_i^*, -}) S_0(t_i^{*, -}) - \sum_{k=1}^N (1 + c_a^k(t_i^*)) a_{t_i^*}^k + \sum_{k=1}^N (1 - c_v^k(t_i^*)) v_{t_i^*}^k + C(t_i^*) \right) \end{cases} \quad (2.20)$$

Avec :

- $c_a^k(t_i^*) \in]0, 1[$ et $c_v^k(t_i^*) \in]0, 1[$ des coûts de transactions proportionnels à l'achat et à la vente en t_i^* de l'actif A_k .
- $\sum_{k=1}^N [(1 + c_a^k(t_i^*)) a_{t_i^*}^k - (1 - c_v^k(t_i^*)) v_{t_i^*}^k]$ correspond à l'ensemble des flux stockés au niveau de la liquidité du portefeuille et générés par les transactions effectuées par le gérant sur les actifs risqués, autres que la liquidité, en t_i^* .
- $C(t_i^*) \in \mathbb{R}$ des flux monétaires exogènes (frais de gestion, etc.).

À l'issu de la période d'investissement, c'est-à-dire $\forall t \in I$ tel que $T \geq t > t_m^*$, et en supposant qu'aucune transaction n'a eu lieu sur l'intervalle $]t_m^*, \dots, t]$, le montant du capital investi dans le portefeuille sera de :

$$V_t = (1 + r_{t_m^*, t}) S_0(t_m^*) + \sum_{k=1}^N (1 + R_{t_m^*, t}^k) S_k(t_m^*) \quad (2.21)$$

Connaissant la dynamique d'un portefeuille dont on a rebalancé les positions à différents instants, on est en mesure de définir le concept de stratégie d'investissement :

Définition 3. Une stratégie d'investissement Ψ peut se représenter sous la forme d'un processus tel que :

$$\Psi_t = \begin{pmatrix} a_t^1 & v_t^1 \\ \vdots & \vdots \\ a_t^N & v_t^N \end{pmatrix} = (\mathbf{a}_t \quad \mathbf{v}_t)$$

avec $\mathbf{a}_t \in \mathbb{R}^N$, $\mathbf{v}_t \in \mathbb{R}^N$ et $(a_t^k)_{t \in I}$ et $(v_t^k)_{t \in I}$ des processus représentant les montants achetés ou vendus de l'actif A_k aux instants $t \in I$. Plus particulièrement, $a_t^k = v_t^k = 0$ pour tout $k \in [1, \dots, N]$ si $t \notin I^*$.

Définition 4. Une stratégie d'investissement Ψ est dite auto-financée, ou « buy-and-hold », si $I^* \equiv \emptyset$. Dans ce cas, $\Psi_t = \mathbf{0}_{N \times 2} \quad \forall t \in I$.

Le montant du capital investi dans le portefeuille est donc fonction de la stratégie d'investissement choisie par l'investisseur. Ce dernier va chercher à maximiser la richesse finale de son portefeuille par rapport à cette stratégie :

$$\Psi^{Opt} = \arg \max_{\Psi^* \in \mathcal{A}} V_t(\Psi^*) \quad \text{avec } t > t_m^* \quad (2.22)$$

$\mathcal{A} \subseteq \mathbb{R}^{N \times 2}$ représentant l'ensemble des stratégies possibles. Ces stratégies doivent généralement répondre à un certain nombre de contraintes qui peuvent avoir plusieurs causes : réglementations, limite de liquidités disponibles, etc...

2.3 Définition de la stratégie optimale

Dans ce paragraphe, on définit le concept de « stratégie optimale » à adopter entre deux instants de décision de l'investisseur. La définition d'une stratégie optimale dépend des contraintes d'investissement auquel est soumis l'investisseur. Par exemple, dans le cas où un investisseur ne possède pas un niveau de liquidité suffisant pour satisfaire certaines réglementations, il peut être contraint de vendre un certain nombre de ses actifs. Cependant, cette vente d'actifs peut elle-même être contrainte par le nombre d'actifs que possède l'investisseur si celui-ci ne peut pas vendre à découvert. De même, si le niveau de liquidité est trop faible, l'investisseur peut être contraint dans ses achats. Nous allons donc considérer les contraintes suivantes :

$$(C1) \quad a_{t_i^*}^{k,-} \leq a_{t_i^*}^k \leq a_{t_i^*}^{k,+}$$

$$(C2) \quad v_{t_i^*}^{k,-} \leq v_{t_i^*}^k \leq v_{t_i^*}^{k,+}$$

où $a_{t_i^*}^{k,-}$, $a_{t_i^*}^{k,+}$, $v_{t_i^*}^{k,-}$ et $v_{t_i^*}^{k,+}$ appartiennent à \mathbb{R} et sont connus par l'investisseur à l'instant de sa prise de décision en $t_i^{*, -}$.

Un investisseur suivant une certaine stratégie d'investissement peut également être contraint par le montant des transactions concernant les achats et ventes de titres. Par exemple, il peut être souhaitable que les transactions ne dépassent pas un certain montant afin que la stratégie permette de générer des gains :

$$(C3) \quad \sum_{k=1}^N [(1 + c_a^k(t_i^*))a_{t_i^*}^k - (1 - c_v^k(t_i^*))v_{t_i^*}^k] \leq M$$

où $M \in \mathbb{R}$ représente un montant auquel l'ensemble des flux générés par les transactions d'achats et de ventes en t_i^* doit être inférieur.

Considérons l'évolution de la valeur du portefeuille :

$$V_{t_{i+1}^*} - V_{t_i^*} = S_0(t_{i+1}^*) - S_0(t_i^*) + \sum_{k=1}^N S_k(t_{i+1}^*) - S_k(t_i^*) \quad (2.23)$$

On sait que (voir équation (2.20)) :

$$S_0(t_{i+1}^*) = (1 + r_{t_i^*, t_{i+1}^*}^{*, -})S_0(t_i^*) + (1 + r_{t_i^*, t_{i+1}^*}^{*, -})C(t_i^*) - (1 + r_{t_i^*, t_{i+1}^*}^{*, -}) \sum_{k=1}^N [(1 + c_a^k(t_i^*))a_{t_i^*}^k - (1 - c_v^k(t_i^*))v_{t_i^*}^k] \quad (2.24)$$

$$S_k(t_{i+1}^*) = (1 + R_{t_i^*, t_{i+1}^*}^k)S_k(t_i^*) + (1 + R_{t_i^*, t_{i+1}^*}^k)(a_{t_i^*}^k - v_{t_i^*}^k) \quad (2.25)$$

Afin d'alléger les notations, nous noterons désormais :

- $r_{t_i^*, t_{i+1}^*}^{*, -}$ comme $r_{i+1; c}$, et $r_{t_i^*, t_{i+1}^*}^{*, -}$ comme $r_{i+1; nc}$
- $R_{t_i^*, t_{i+1}^*}^k$ comme $R_{i+1; c}^k$, et $R_{t_i^*, t_{i+1}^*}^k$ comme $R_{i+1; nc}^k$.
- $a_{t_i^*}^k$ comme a_i^k , et $v_{t_i^*}^k$ comme v_i^k .

2.3. DÉFINITION DE LA STRATÉGIE OPTIMALE

On peut donc réécrire l'évolution de richesse du portefeuille comme :

$$\begin{aligned}
V_{t_{i+1}^{*,-}} - V_{t_i^{*,-}} &= r_{i+1;c} S_0(t_i^{*,-}) + (1 + r_{i+1;nc}) C(t_i^*) + \sum_{k=1}^N R_{i+1;c}^k S_k(t_i^{*,-}) \\
&+ \sum_{k=1}^N a_i^k [(1 + R_{i+1;nc}^k) - (1 + r_{i+1;nc})(1 + c_a^k(t_i^*))] \\
&+ \sum_{k=1}^N v_i^k [(1 + r_{i+1;nc})(1 - c_v^k(t_i^*)) - (1 + R_{i+1;nc}^k)] \quad (2.26)
\end{aligned}$$

Puis, sous forme matricielle :

$$V_{t_{i+1}^{*,-}} - V_{t_i^{*,-}} = r_{i+1;c} S_0(t_i^{*,-}) + (1 + r_{i+1;nc}) C(t_i^*) + \mathbf{R}_{i+1;c}^{Tr} \mathbf{S}(t_i^{*,-}) + \mathbf{a}_i^{Tr} \mathbf{D}_{i,1} + \mathbf{v}_i^{Tr} \mathbf{D}_{i,2} \quad (2.27)$$

où :

- $\mathbf{R}_{i+1;c}$ le vecteur composé des $(R_{i+1;c}^k)_{k \in [1, \dots, N]}$ et $\mathbf{S}(t_i^{*,-})$ le composé des $(S_k(t_i^{*,-}))_{k \in [1, \dots, N]}$.
- \mathbf{a}_i le vecteur composé des $(a_i^k)_{k \in [1, \dots, N]}$ et \mathbf{v}_i le vecteur composé des $(v_i^k)_{k \in [1, \dots, N]}$.
- $\mathbf{D}_i = \begin{pmatrix} (1 + R_{i+1;nc}^1) - (1 + r_{i+1;nc})(1 + c_a^1(t_i^*)) & (1 + r_{i+1;nc})(1 - c_v^1(t_i^*)) - (1 + R_{i+1;nc}^1) \\ \vdots & \vdots \\ (1 + R_{i+1;nc}^k) - (1 + r_{i+1;nc})(1 + c_a^k(t_i^*)) & (1 + r_{i+1;nc})(1 - c_v^k(t_i^*)) - (1 + R_{i+1;nc}^k) \\ \vdots & \vdots \\ (1 + R_{i+1;nc}^N) - (1 + r_{i+1;nc})(1 + c_a^N(t_i^*)) & (1 + r_{i+1;nc})(1 - c_v^N(t_i^*)) - (1 + R_{i+1;nc}^N) \end{pmatrix}$

Pour maximiser cette évolution de richesse, l'investisseur ne peut jouer que sur \mathbf{a}_i et \mathbf{v}_i . Lorsqu'on dérive cette expression par rapport à \mathbf{a}_i (resp. \mathbf{v}_i), on obtient $\mathbf{D}_{i,1}$ (resp. $\mathbf{D}_{i,2}$). Dans le cas où

$$(1 + R_{i+1;nc}^k) - (1 + r_{i+1;nc})(1 + c_a^k(t_i^*)) = 0 \Leftrightarrow R_{i+1;nc}^k = r_{i+1;nc} + c_a^k(t_i^*)(1 + r_{i+1;nc})$$

le montant d'actif A_k acheté par l'investisseur n'a aucun impact sur l'évolution de richesse du portefeuille. Similairement, dans le cas où

$$(1 + r_{i+1;nc})(1 - c_v^k(t_i^*)) - (1 + R_{i+1;nc}^k) = 0 \Leftrightarrow R_{i+1;nc}^k = r_{i+1;nc} - c_v^k(t_i^*)(1 + r_{i+1;nc})$$

le montant d'actif A_k vendu par l'investisseur n'a aucun impact sur l'évolution de richesse du portefeuille.

Si $(1 + R_{i+1;nc}^k) - (1 + r_{i+1;nc})(1 + c_a^k(t_i^*)) \neq 0$ et $(1 + r_{i+1;nc})(1 - c_v^k(t_i^*)) - (1 + R_{i+1;nc}^k) \neq 0$ pour tout $k \in [1, \dots, N]$, la stratégie optimale d'achat/vente dépendra des signes des coordonnées des vecteurs $\mathbf{D}_{i,1}$ et $\mathbf{D}_{i,2}$ ainsi que des contraintes auxquelles est soumis l'investisseur. Par exemple, dans le cas où la $k^{\text{ème}}$ coordonnée du vecteur $\mathbf{D}_{i,1}$ est positive, il faut que a_i^k soit le plus élevé possible afin de maximiser l'évolution de richesse. On peut ainsi donner une définition précise de la stratégie optimale que l'investisseur aurait dû adopter à l'instant $t_i^{*,-}$, sous un certain nombre de contraintes, s'il avait eu connaissance de l'évolution future des investissements et des coûts de transactions.

Notons :

- $G_i(x) = r_{i+1;nc} + x \times (1 + r_{i+1;nc})$
- $C_M(j) = M + \sum_{k=1}^N (1 - c_v^k(t_i^*)) v_i^{k, Opt} - \sum_{l \in \mathcal{B}_j} (1 + c_a^l(t_i^*)) a_{t_i^*}^{l,+} - \sum_{l \in \mathcal{B}_j^c \setminus \{j\}} (1 + c_a^l(t_i^*)) a_{t_i^*}^{l,-}$
- $\mathcal{B}_j = \left\{ k \in [1, \dots, N] \setminus \{j\} \mid R_{i+1;nc}^k > \frac{a_{t_i^*}^{j,+}}{a_{t_i^*}^{k,+}} (R_{i+1;nc}^j - G_i(c_a^j(t_i^*))) + G_i(c_a^k(t_i^*)) \right\}$ et \mathcal{B}_j^c le complémentaire de cet ensemble dans $[1, \dots, N]$.

2.3. DÉFINITION DE LA STRATÉGIE OPTIMALE

On a alors la définition suivante :

Définition 5. La stratégie optimale d'un investisseur soumis aux contraintes (C1), (C2) et (C3) à l'instant t_i^* s'exprime comme :

$$\Psi_{t_i^*}^{Opt} = \begin{pmatrix} a_i^{1,Opt} & v_i^{1,Opt} \\ \vdots & \vdots \\ a_i^{N,Opt} & v_i^{N,Opt} \end{pmatrix}$$

où

$$\forall k \in [1, \dots, N], \quad v_i^{k,Opt} = \begin{cases} v_{t_i^*}^{k,+} & \text{si } R_{i+1;nc}^k < G_i(-c_v^k(t_i^*)) \\ v_{t_i^*}^{k,-} & \text{si } R_{i+1;nc}^k > G_i(-c_v^k(t_i^*)) \end{cases} \quad (2.28)$$

$$\forall k \in [1, \dots, N], \quad a_i^{k,Opt} = \begin{cases} H(k) & \text{si } R_{i+1;nc}^k > G_i(c_a^k(t_i^*)) \\ a_{t_i^*}^{k,-} & \text{si } R_{i+1;nc}^k < G_i(c_a^k(t_i^*)) \end{cases} \quad (2.29)$$

avec

$$H(k) = \begin{cases} a_{t_i^*}^{k,+} & \text{si } C_M(k) \geq (1 + c_a^k(t_i^*))a_{t_i^*}^{k,+} \\ C_M(k) & \text{si } (1 + c_a^k(t_i^*))a_{t_i^*}^{k,+} > C_M(k) > (1 + c_a^k(t_i^*))a_{t_i^*}^{k,-} \\ a_{t_i^*}^{k,-} & \text{si } C_M(k) \leq (1 + c_a^k(t_i^*))a_{t_i^*}^{k,-} \end{cases} \quad (2.30)$$

Preuve. On constate que la $k^{\text{ème}}$ coordonnée du vecteur $\mathbf{D}_{i,2}$, notée $\mathbf{D}_{i,2}(k)$, est

- positive lorsque $R_{i+1;nc}^k < r_{i+1;nc} - c_v^k(t_i^*)(1 + r_{i+1;nc})$.
- négative lorsque $R_{i+1;nc}^k > r_{i+1;nc} - c_v^k(t_i^*)(1 + r_{i+1;nc})$.

En conséquence, la stratégie optimale à adopter concernant les ventes d'actifs est la suivante :

$$\forall k \in [1, \dots, N], \quad v_i^{k,Opt} = \begin{cases} v_{t_i^*}^{k,+} & \text{si } R_{i+1;nc}^k < r_{i+1;nc} - c_v^k(t_i^*)(1 + r_{i+1;nc}) \\ v_{t_i^*}^{k,-} & \text{si } R_{i+1;nc}^k > r_{i+1;nc} - c_v^k(t_i^*)(1 + r_{i+1;nc}) \end{cases} \quad (2.31)$$

Concernant la stratégie optimale à adopter pour l'achat des titres, les montants pouvant être alloués doivent vérifier l'inégalité suivante, due à la contrainte (C3) :

$$M + \sum_{k=1}^N (1 - c_v^k(t_i^*))v_i^{k,Opt} \geq \sum_{k=1}^N (1 + c_a^k(t_i^*))a_i^{k,Opt} \quad (2.32)$$

Afin de maximiser l'évolution de richesse, il faut investir un maximum dans les actifs A_k tels que $D_{i,1}(k)$ est positif et acheter le moins possible dans le cas où cette coordonnée est négative. Par conséquent, il faut acheter les actifs pour lesquels le produit $a_{t_i^*}^{k,+} D_{i,1}(k)$ est le plus élevé tout en respectant la contrainte (C3). Pour se faire, nous allons créer un ensemble \mathcal{B}_j répertoriant l'ensemble des actifs A_k ayant un produit $a_{t_i^*}^{k,+} D_{i,1}(k)$ plus élevé que celui associé à l'actif A_j , c'est-à-dire $a_{t_i^*}^{j,+} D_{i,1}(j)$:

$$\mathcal{B}_j = \left\{ k \in [1, \dots, N] \setminus \{j\} \mid R_{i+1;nc}^k > \frac{a_{t_i^*}^{j,+}}{a_{t_i^*}^{k,+}} (R_{i+1;nc}^j - r_{i+1;nc} - c_a^j(t_i^*)(1 + r_{i+1;nc})) + r_{i+1;nc} + c_a^k(t_i^*)(1 + r_{i+1;nc}) \right\} \quad (2.33)$$

On peut remarquer que dans le cas où $a_{t_i^*}^{j,+} = a_{t_i^*}^{k,+}$ et où $c_a^j(t_i^*) = c_a^k(t_i^*)$, cet ensemble devient $\mathcal{B}_j = \left\{ k \in [1, \dots, N] \setminus \{j\} \mid R_{i+1;nc}^k > R_{i+1;nc}^j \right\}$.

Le montant investissable dans l'actif A_j , noté $C_M(j)$, dépend du montant maximal que l'on peut acheter, c.-à-d. $M + \sum_{k=1}^N (1 - c_v^k(t_i^*)) v_i^{k,Opt}$, auquel on soustrait le montant investi dans les actifs ayant un produit $a_{t_i^*}^{k,+} D_{i,1}(k)$ plus élevé que celui associé à l'actif A_j , c.-à-d. $\sum_{l \in \mathcal{B}_j} (1 + c_a^l(t_i^*)) a_{t_i^*}^{l,+}$, ainsi que le montant investi dans les actifs ayant un produit $a_{t_i^*}^{k,+} D_{i,1}(k)$ plus faible que celui associé à l'actif A_j (dans lesquels on investira un minimum), c.-à-d. $\sum_{l \in \mathcal{B}_j^c \setminus \{j\}} (1 + c_a^l(t_i^*)) a_{t_i^*}^{l,-}$ où \mathcal{B}_j^c correspond à l'ensemble complémentaire de \mathcal{B}_j dans $[1, \dots, N]$:

$$C_M(j) = M + \sum_{k=1}^N (1 - c_v^k(t_i^*)) v_i^{k,Opt} - \sum_{l \in \mathcal{B}_j} (1 + c_a^l(t_i^*)) a_{t_i^*}^{l,+} - \sum_{l \in \mathcal{B}_j^c \setminus \{j\}} (1 + c_a^l(t_i^*)) a_{t_i^*}^{l,-} \quad (2.34)$$

Lorsque $\mathbf{D}_{i,1}(k)$ est négatif, on investit le moins possible. Lorsque $\mathbf{D}_{i,1}(k)$ est positif, on distingue trois cas. Dans le premier cas, le montant investissable dans l'actif A_k , noté $C_M(k)$, permet d'investir un maximum dans cet actif. Dans le second cas, le montant investissable est inférieur au maximum que l'on peut investir dans cet actif. Dans le troisième cas, le montant investissable ne permet d'investir que le montant minimum imposé par la contrainte (C1). Ceci peut se réécrire de la manière suivante :

$$a_i^{k,Opt} = \begin{cases} H(k) & \text{si } R_{i+1;nc}^k > r_{i+1;nc} + c_a^k(t_i^*)(1 + r_{i+1;nc}) \\ a_{t_i^*}^{k,-} & \text{si } R_{i+1;nc}^k < r_{i+1;nc} + c_a^k(t_i^*)(1 + r_{i+1;nc}) \end{cases} \quad (2.35)$$

où

$$H(k) = \begin{cases} a_{t_i^*}^{k,+} & \text{si } C_M(k) \geq (1 + c_a^k(t_i^*)) a_{t_i^*}^{k,+} \\ C_M(k) & \text{si } (1 + c_a^k(t_i^*)) a_{t_i^*}^{k,+} > C_M(k) > (1 + c_a^k(t_i^*)) a_{t_i^*}^{k,-} \\ a_{t_i^*}^{k,-} & \text{si } C_M(k) \leq (1 + c_a^k(t_i^*)) a_{t_i^*}^{k,-} \end{cases} \quad (2.36)$$

Ce qui nous permet de définir explicitement la stratégie optimale à adopter. \square

2.4 Approximation de la stratégie optimale

Dans le paragraphe précédent, nous avons présenté la stratégie d'achat/vente qu'il aurait fallu adopter pour maximiser les rendements du portefeuille si le futur des investissements était connu. Dans ce paragraphe, on introduit de l'incertitude sur le futur des investissements à travers la théorie des probabilités. L'estimation de tendances au sein des valeurs des investissements permet alors d'approximer la stratégie optimale telle que définie précédemment.

Un investisseur soumis aux contraintes (C1), (C2) et (C3) souhaitant approximer une stratégie optimale entre deux instants de décisions $t_i^{*,-}$ et $t_{i+1}^{*,-}$ fait face à un certain nombre de règles de décision. Pour chaque actif $k \in [1, \dots, N]$, il doit estimer si les événements suivants peuvent se réaliser :

- $\ll R_{i+1;nc}^k \leq G_i(-c_v^k(t_i^*)) \gg$
- $\ll R_{i+1;nc}^k \leq G_i(c_a^k(t_i^*)) \gg$

Afin d'estimer l'ensemble \mathcal{B}_k , il doit également estimer si les événements suivants peuvent se réaliser :

- $\ll R_{i+1;nc}^j > \frac{a_{t_i^*}^{k,+}}{a_{t_i^*}^{j,+}} (R_{i+1;nc}^k - G_i(c_a^k(t_i^*))) + G_i(c_a^j(t_i^*)) \gg$ pour tout $j \in [1, \dots, N] \setminus \{k\}$.

Supposons que :

- $\forall k \in [1, \dots, N]$, $R_{i+1;nc}^k$ et $r_{i+1;nc}^k$ soient des variables aléatoires définies sur un espace de probabilité complet $(\Omega, \mathcal{F}, \mathbb{P})$.
- les coûts proportionnels de transactions $c_a^k(t_i^*)$ et $c_v^k(t_i^*)$ sont des fonctions déterministes du temps pour tout $k \in [1, \dots, N]$.

Notons $P_j(x) : \mathbb{R} \rightarrow [0, 1]$ et $Q_j(k) : [1, \dots, N] \rightarrow [0, 1]$ deux fonctions définies par :

- $P_k(x) = \mathbb{P}(R_{i+1;nc}^k \leq G_i(x))$
- $Q_k(j) = \mathbb{P}(R_{i+1;nc}^j > \frac{a_{t_i^*}^{k,+}}{a_{t_i^*}^{j,+}} (R_{i+1;nc}^k - G_i(c_a^k(t_i^*))) + G_i(c_a^j(t_i^*)))$

où \mathbb{P} est une mesure de probabilité définie sur l'espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$.

Pour tout $k \in [1, \dots, N]$, $\widehat{P}_k(c_a^k(t_i^*))$, $\widehat{P}_k(-c_v^k(t_i^*))$ et $(\widehat{Q}_k(j))_{j \in [1, \dots, N] \setminus \{k\}}$ quantifient donc les probabilités de réalisation des événements cités plus haut. Puisque ces événements permettent de définir complètement la stratégie optimale à adopter dans le cadre des conditions (C1), (C2) et (C3), l'approximation de cette stratégie va essentiellement dépendre de l'estimation de ces $(N + 1) \times N$ probabilités.

Afin d'estimer ces probabilités, le théorème de décomposition de Cramér [Cra(1961)] [Cra(1962)] nous permet de supposer qu'il existe une unique composante prévisible au sein de la série de rendements :

Théorème 1. *Soit y_t un processus aléatoire centré ($\mathbb{E}[y_t] = 0$) dont les moments du second ordre sont finis ($\mathbb{E}[y_t^2] < \infty$). Alors il existe une unique décomposition orthogonale*

$$y_t = \mu_t + \epsilon_t$$

telle que ϵ_t est une composante parfaitement non-déterministe³ (c.-à-d. que $\epsilon_t = y_t - \mathbb{E}[y_t | \mathcal{F}_t]$) tandis que μ_t est déterministe⁴ (c.-à-d. que $\mu_t = \mathbb{E}[y_t | \mathcal{F}_t]$).

Ce théorème justifie l'estimation d'une unique « tendance » déterministe sur un échantillon fini de données. Cette tendance se définit donc comme étant une composante prévisible sous-jacente à la série d'observations. En effet, supposons que l'on possède un échantillon fini de rendements $(r_{i+1})_{i \in [1, \dots, n]}$ et $(R_{t_i, t_{i+1}}^k)_{i \in [1, \dots, n]}$ pour tout $k \in [1, \dots, N]$. En pratique, on peut généralement se ramener au cas où les conditions du théorème de Cramér sont statistiquement acceptées et on peut donc supposer que :

$$r_{i+1;nc} - m_0 = \mu_{0;i+1} + \epsilon_{0;i+1} \quad \text{et} \quad R_{i+1;nc}^k - m_k = \mu_{k;i+1} + \epsilon_{k;i+1} \quad \text{pour tout } k \in [1, \dots, N]$$

où :

- $m_0 = \frac{1}{n} \sum_{i=1}^n r_{i+1;nc}$ et $m_k = \frac{1}{n} \sum_{i=1}^n R_{i+1;nc}^k$

3. « Le passé infiniment éloigné du processus x_t ne nous apprend rien sur le développement futur du processus. Toute information contenue dans ce processus doit donc être entrée comme une innovation à un certain moment. » [Cra(1962)]

4. « Toute information relativement au processus est déjà contenue dans son passé infiniment éloigné » [Cra(1962)]

2.4. APPROXIMATION DE LA STRATÉGIE OPTIMALE

- $\epsilon_{0;i+1}$ et $\epsilon_{k;i+1}$ sont des composantes parfaitement non-déterministes que l'on peut supposer être des variables aléatoires indépendamment distribuées ($\epsilon_{k;i+1}$ et $\epsilon_{k;j+1}$ sont indépendants $\forall i \neq j$) et centrées ($\mathbb{E}[\epsilon_{k;i+1}] = 0$).
- $\mu_{0;i+1}$ et $\mu_{k;i+1}$ sont des composantes déterministes donc entièrement prévisibles.

Dans ce cadre, on a alors les égalités suivantes :

$$\widehat{P}_j(x) = \widehat{\mathbb{P}}(R_{i+1;nc}^k \leq G_i(x)) = \widehat{\mathbb{P}}(\epsilon_{k;i+1} - (1+x)\epsilon_{0;i+1} \leq (1+x)m_0 - m_k + (1+x)\mu_{0;i+1} - \mu_{k;i+1}) \quad (2.37)$$

pour tout $k \in [1, \dots, N]$, et

$$\begin{aligned} \widehat{Q}_j(k) = \widehat{\mathbb{P}}\left(\epsilon_{j;i+1} - \frac{a_{t_i^*}^{k,+}}{a_{t_i^*}^{j,+}} \epsilon_{k;i+1} + \epsilon_{0;i+1} \left((1 + c_a^k(t_i^*)) \frac{a_{t_i^*}^{k,+}}{a_{t_i^*}^{j,+}} - (1 + c_a^j(t_i^*)) \right) > \right. \\ \left. \frac{a_{t_i^*}^{k,+}}{a_{t_i^*}^{j,+}} (m_k + \mu_{k;i+1} - (1 + c_a^k(t_i^*)) (m_0 + \mu_{0;i+1}) - c_a^k(t_i^*)) + (1 + c_a^j(t_i^*)) (m_0 + \mu_{0;i+1}) + c_a^j(t_i^*) \right) \end{aligned} \quad (2.38)$$

pour tout $k \in [1, \dots, N]$ et tout $j \in [1, \dots, N]$.

Afin d'estimer ces probabilités, il faut estimer les lois de probabilité générant les variables aléatoires $\epsilon_{0;i+1}$ et $\epsilon_{k;i+1}$ pour tout $k \in [1, \dots, N]$.

Cette estimation passe par l'estimation de la composante prévisible sur l'échantillon fini de données, l'estimation des réalisations du bruit correspondant alors à la différence entre les observations et la composante prévisible estimée.

Exemple d'une approximation de la stratégie optimale Nous allons considérer qu'un investisseur peut investir dans deux actifs ($k \in [1, 2]$) et est soumis aux contraintes suivantes :

- (C1) $0 \leq a_{t_i^*}^k \leq 10$ pour $k \in [1, 2]$.
- (C2) $0 \leq v_{t_i^*}^k \leq 10$ pour $k \in [1, 2]$.
- (C3) le montant des transactions concernant les achats et ventes de titres ne doit pas dépasser 20 (c.-à-d. $M = 20$).

Nous allons également supposer que les coûts proportionnels de transaction en t_i^* sont connus et égaux à 1%. Une approximation de la stratégie optimale concernant les ventes d'actifs détenus en t_i^* peut alors se définir par :

$$\forall k \in [1, 2], \quad v_i^{k,Opt} = \begin{cases} 10 & \text{si } \widehat{P}_k(-0.01) > 0.5 \\ 0 & \text{si } 1 - \widehat{P}_k(-0.01) \geq 0.5 \end{cases} \quad (2.39)$$

On vendra donc un maximum si la probabilité associée à l'évènement $R_{i+1;nc}^k < G_i(-c_v^k(t_i^*))$ est estimée supérieure à 0.5. Concernant les achats, il faut déterminer une approximation de l'ensemble \mathcal{B}_k pour $k \in [1, 2]$. On peut considérer l'ensemble :

$$\widehat{\mathcal{B}}_k = \left\{ l \in [1, 2] \setminus \{k\} \mid \widehat{Q}_k(l) > 0.5 \right\}$$

2.5. CONCLUSION

Dans notre exemple, $\widehat{Q}_k(l) = \widehat{\mathbb{P}}(R_{i+1;nc}^l > R_{i+1;nc}^k)$, c'est donc la probabilité estimée que chaque actif est un rendement supérieur à l'autre. Si $\widehat{Q}_k(l) \neq 0.5$, $\widehat{\mathcal{B}}_k$ sera donc l'ensemble vide pour un des actifs. Cet ensemble nous permet de déterminer la quantité $C_M(k)$ pour $k \in [1, 2]$, dépendant de la contrainte (C3) (c.-à-d. $M = 20$) :

$$C_M(k) = 20 + \sum_{l=1}^2 0.99 \times v_i^{l, Opt} - \sum_{l \in \widehat{\mathcal{B}}_k} 1.01 \times 10$$

Puis d'en déduire une approximation de la stratégie optimale concernant les achats de titres en t_i^* :

$$a_i^{k, Opt} = \begin{cases} H(k) & \text{si } 1 - P_k(0.01) > 0.5 \\ 0 & \text{si } P_k(0.01) > 0.5 \end{cases} \quad (2.40)$$

où

$$H(k) = \begin{cases} 10 & \text{si } C_M(k) \geq (1.01) \times 10 \\ C_M(k) & \text{si } (1.01) \times 10 > C_M(k) > 0 \\ 0 & \text{si } C_M(k) \leq 0 \end{cases} \quad (2.41)$$

2.5 Conclusion

Dans ce chapitre, nous avons mis en avant un certain nombre d'évènements nécessaires à la définition de la stratégie optimale à adopter pour optimiser un portefeuille de titres. Une approximation de cette stratégie optimale passe par l'estimation des probabilités de réalisation de ces évènements. Puisque ces probabilités dépendent des loi de probabilités des composantes aléatoires au sein des rendements des investissements, elles dépendent d'une estimation de leur composante prévisible (ou « tendance »), les composantes aléatoires (ou « bruits ») étant définies comme la différence entre les rendements et leur tendance. Dans le prochain chapitre, nous nous intéressons à différentes méthodes statistiques permettant d'estimer une telle tendance.

Puisque la dynamique de ces investissements est complexe et peut être impactée par des facteurs qui ne sont pas toujours mesurables en temps réel, il faudrait estimer cette composante prévisible d'une manière assez « flexible » en faisant peu d'hypothèses sur la modélisation de l'évolution de l'investissement et en considérant que cette modélisation peut évoluer avec le temps. Nous proposons donc, dans notre prochain chapitre, des solutions à cette problématique d'estimation de tendances. Nous étudions notamment l'apport de la régression non-paramétrique pour estimer une tendance. Ce type de régression ne fait aucune hypothèse sur la distribution initiale des investissements et semble donc tout à fait appropriée.

Estimation non-paramétrique de tendances en gestion de portefeuille

Sommaire

3.1	Estimer une tendance dans le domaine temporel	60
3.1.1	Filtrage linéaire dans le domaine temporel	61
3.1.2	Filtrage dans le domaine fréquentiel	62
3.1.3	Filtrage linéaire optimum et nécessité de méthodes locales et adaptatives	64
3.2	Problématiques de régression non-paramétrique	66
3.2.1	Le régressogramme	66
3.2.2	L'hypothèse de régularité de la tendance	67
3.2.3	La régression sur base de fonctions orthonormales	67
3.2.4	Estimation du paramètre de lissage	69
3.3	Méthodes non-paramétriques d'estimation de tendances en pratique	71
3.3.1	Techniques de filtrage non-paramétrique classiques en économie et finance	72
3.3.2	La régression non-paramétrique polynomiale locale	74
3.3.3	La régression à l'aide de fonctions splines	76
3.3.4	La régression en utilisant des bases d'ondelettes	78
3.3.5	La régression en utilisant une décomposition modale empirique	81
3.4	Comparaison des différentes méthodes	83
3.4.1	Critères de comparaison	84
3.4.2	Modèles utilisés pour la génération des données	85
3.4.3	Méthodes d'estimation de tendances utilisées	86
3.4.4	Interprétation des résultats	86
3.5	Conclusion	91

Introduction

Anticiper l'évolution d'un investissement dans un actif financier est une tâche délicate lorsqu'on ne connaît pas les flux futurs versés par cet actif. Cette anticipation peut s'avérer inexacte, notamment

si des évènements futurs incertains impactent l'évolution de l'investissement. L'actif sera donc dit « risqué » lorsqu'il existe une incertitude quantifiable concernant les gains futurs qu'il procurera. Mesurer ce risque est essentiel pour un investisseur souhaitant limiter l'incertitude de ses gains. Il n'existe malheureusement pas de définition universelle de ce risque bien qu'une certaine littérature tente d'en définir les contours [Art(1999)]. Il peut donc se mesurer de nombreuses façons. En général, on définit un point de référence où le risque est considéré nul puis on quantifie l'éloignement par rapport à ce référentiel. L'allocation optimale d'un portefeuille consisterait alors à minimiser une telle mesure de risque tout en maximisant une mesure du gain attendu [Mar(1952)].

Afin de mesurer ce gain attendu, il faut prédire le rendement à venir de l'investissement. En général, ces prédictions sont réalisées à partir d'une estimation de l'espérance conditionnelle des rendements de prix sur un horizon d'investissement prédéfini. Lorsque les modèles sont linéaires avec un bruit additif gaussien, les méthodes « classiques » des moindres carrés et du maximum de vraisemblance sont alors idéales. Néanmoins, comme nous l'avons spécifié dans notre premier chapitre, la dynamique de la valeur d'un investissement est différente de celle des prix et apparaît bien plus complexe que celle d'un simple modèle linéaire gaussien. En effet, de nombreux facteurs, plus ou moins mesurables en temps réel, ont un impact évolutif sur cette dynamique, ce qui la rend difficilement prévisible. Anticiper la dynamique de la valeur d'un investissement passe donc par une identification des facteurs l'impactant mais également par une anticipation de l'évolution de ces différents facteurs. La problématique essentielle réside alors dans la méthode de prédiction de toutes ces dynamiques.

Dans ce chapitre, nous proposons d'estimer ces dynamiques d'une manière adaptative et localisée dans le temps. Nous partons du principe que la dynamique de la série d'observations étudiée est la résultante de deux composantes. La première décrit l'évolution prévisible de cette série (la « tendance ») tandis que la seconde décrit son évolution imprévisible (le « bruit »). La tendance se définit alors comme une fonction du temps caractérisée par un certain degré de régularité (on peut par exemple supposer que la tendance est une fonction deux fois dérivable) tandis que le bruit évolue dans un voisinage de cette tendance : il est censé être la réalisation d'une variable aléatoire centrée sur 0 et indépendamment distribuée. Afin d'estimer cette dynamique d'après les observations, nous ne faisons aucune hypothèse sur la densité de probabilité de la série initiale d'observations : on parle d'estimation « non-paramétrique ». Dans ce cadre, la mesure du gain attendu correspond à une estimation non-paramétrique de la tendance future de l'évolution de l'investissement.

Ce chapitre concerne donc le filtrage de séries temporelles dans le but d'estimer une tendance de manière non-paramétrique d'après un échantillon fini en temps discret. Après une rapide introduction au filtrage temporel des séries chronologiques, nous montrons la nécessité d'utiliser des méthodes de filtrage locales et adaptatives. Nous présentons ensuite différentes méthodes de régression « non-paramétriques » afin d'estimer une tendance parmi lesquelles la régression non-paramétrique polynomiale locale, les splines de lissage, la décomposition sur base d'ondelettes et la décomposition modale empirique. Nous discutons plus particulièrement des méthodes existantes pour estimer le paramètre de lissage définissant la tendance.

3.1 Estimer une tendance dans le domaine temporel

Soit $(y_t)_{t \in \mathbb{R}}$ la variable dont on souhaite estimer la tendance. On suppose que

$$y_t = f(t) + \epsilon_t \tag{3.1}$$

avec f le paramètre de *tendance* et $(\epsilon_t)_{t \in \mathbb{R}}$ un paramètre de *bruit* oscillant dans un voisinage de cette tendance. ϵ_t est supposé posséder une bande de fréquence plus élevée que celle de la tendance. En

pratique, l'estimation de la tendance consiste à supposer une certaine régularité pour la fonction f puis à l'estimer d'après les observations $(y_t)_{t \in \mathbb{R}}$. Il est coutume de supposer ϵ_t aléatoire, stationnaire au second ordre et d'espérance inconditionnelle nulle afin de faciliter l'inférence de cette tendance.

3.1.1 Filtrage linéaire dans le domaine temporel

Un filtre *linéaire* consiste à transformer une série d'entrée x_t en une série de sortie, également appelée *réponse* et notée y_t , au moyen d'une transformation linéaire. Plus spécifiquement, la sortie y_t du filtre linéaire est le résultat d'une convolution de l'entrée x_t avec un vecteur de coefficients $w = (\dots, w_{-2}, w_{-1}, w_0, w_1, w_2, \dots)$:

$$y_t = w * x_t = \sum_{i=-\infty}^{\infty} w_i x_{t-i} \quad (3.2)$$

Dans le domaine temporel, les filtres linéaires sont classifiés selon leur réponse à un signal d'impulsion *unitaire*, également appelé *impulsion de Dirac*, et défini par $x_t = \mathbf{1}_{t=0}$ où $\mathbf{1}$ est une fonction indicatrice.

Sachant que la série d'entrée est une impulsion de Dirac, la série de sortie du filtre est appelée la réponse *impulsionnelle* du filtre. Plus particulièrement, $y_t = w_t$. Dans ce cas, si y_t est à support borné¹, le filtre est dit à réponse impulsionnelle finie (en anglais *Finite Impulse Response* ou FIR). D'un autre côté, si le support de la réponse impulsionnelle du filtre est non borné, le filtre correspondant est appelé filtre à réponse impulsionnelle infinie (en anglais *Infinite Impulse Response* ou IIR).

Filtres à réponse impulsionnelle infinie Les filtres linéaires peuvent être vus comme un cas particulier d'équations différentielles linéaires à coefficients constants :

$$y_t = \sum_{i=1}^L a_i y_{t-i} + \sum_{i=0}^M w_i x_{t-i} \quad (3.3)$$

où $L \in \mathbb{N}$ représente le nombre de valeurs retardées de la sortie y_t et $M \in \mathbb{N}$, le nombre de valeurs retardées de l'entrée x_t . La sortie du filtre dépend alors de toutes ces valeurs retardées ainsi que de la valeur de l'entrée à l'instant présent.

Supposons, par exemple, que le filtre soit défini par l'équation suivante :

$$y_t = a y_{t-1} + x_t \quad (3.4)$$

La sortie y_t ne peut être déterminée, même si x_t et a sont disponibles, sans qu'une information sur la sortie du filtre à un instant passé t_0 ($t_0 \leq t$) ne soit précisée. Supposons que l'on connaisse y_0 et que $x_t = 0$ pour $t < 0$. Une solution à l'équation peut alors être facilement obtenue en raisonnant par récurrence :

$$y_t = a^t y_0 + \sum_{i=0}^{t-1} a^i x_{t-i} \quad (3.5)$$

Dans ce cas, le filtre a une réponse impulsionnelle non nulle quelque soit t et c'est donc un filtre à réponse impulsionnelle infinie. En particulier, si $|a| > 1$, l'équation différentielle est instable et la

1. $\exists t_0 \in \mathbb{R}, \quad \forall |t| > t_0, \quad y_t = w_t = 0$

réponse impulsionnelle du filtre divergera avec le temps. Dans le cas où $|a| < 1$, le filtre est alors dit *stable* car la réponse impulsionnelle tendra asymptotiquement vers 0 avec le temps.

Plus généralement, l'équation (3.3) aura toujours une réponse impulsionnelle à support non borné. La stabilité du filtre dépendra alors des coefficients a_i . Pour la stabilité de telles équations, le lecteur intéressé pourra se référer à Hamilton [Ham(1994)].

Filtres à réponse impulsionnelle finie Un filtre à réponse impulsionnelle finie peut s'écrire sous la forme :

$$y_t = \sum_{i=-N}^M w_i x_{t-i} \quad (3.6)$$

La sortie du filtre dépend donc de N valeurs futures, de la valeur à l'instant présent, ainsi que de M valeurs passées de la série d'entrée (N et M étant des entiers naturels). Dans le cas où $N > 0$, le filtre est dit *non causal*. En économie et finance, les moyennes mobiles sont une technique de filtrage assez répandue. Une moyenne mobile centrée et équipondérée est un exemple de filtre FIR non causal tandis qu'une moyenne mobile exponentielle (voir Chapitre 1 équation 1.1) est un filtre FIR causal ($N \leq 0$).

3.1.2 Filtrage dans le domaine fréquentiel

Dans le domaine temporel, la réponse impulsionnelle d'un filtre permet de décrire et de classifier des filtres linéaires. Il est également possible de décrire ces filtres dans le domaine fréquentiel à l'aide d'une réponse fréquentielle, également appelée *fonction de transfert*. En temps discret, elle se définit par

$$H(f) = \sum_{k=-\infty}^{\infty} w_k e^{-2i\pi f k}$$

où f est la fréquence du signal d'entrée (que l'on peut apparenter à une fréquence d'oscillation²) et w_k la réponse impulsionnelle du filtre. La fonction de transfert correspond donc à la transformée de Fourier de la réponse impulsionnelle du filtre.

Dans le but d'introduire l'analyse de filtres dans le domaine fréquentiel, nous allons considérer un signal x_t ayant une fréquence connue f :

$$x_t = e^{2i\pi f t}$$

Si l'entrée correspond à cette forme d'exponentiel complexe et si le filtre est un filtre FIR :

$$y_t = \sum_{k=-\infty}^{\infty} w_k x_{t-k} = \sum_{k=-\infty}^{\infty} w_k e^{2i\pi f (t-k)} = e^{2i\pi f t} \left(\sum_{k=-\infty}^{\infty} w_k e^{-2i\pi f k} \right)$$

où le terme en parenthèse correspond à la fonction de transfert précédemment définie. On peut ainsi écrire que

$$y_t = H(f) e^{2i\pi f t}$$

Une fonction de transfert peut s'exprimer en coordonnées polaires de la façon suivante :

$$H(f) = G(f) e^{i\theta f}$$

2. inverse de la période nécessaire à une oscillation du signal d'entrée.

$G(f)$ est appelé *fonction de gain* du filtre et correspond à la norme de la fonction de transfert $|H(f)|$. $e^{i\theta f}$ est appelée *fonction de phase*. Cette décomposition nous permet de caractériser une fonction de transfert d'un filtre par sa fonction de gain $G(f)$ et l'angle θ de sa phase.

Filtres passe-haut et passe-bas Lorsque le gain est élevé pour de faibles fréquences, et faible pour des fréquences plus élevées, le filtre est dit *passe-bas*. Dans ce cas, les faibles fréquences de l'entrée sont préservées par le filtre, tandis que les composantes de hautes fréquences sont atténuées. À l'opposé, si le gain est faible pour de faibles fréquences et élevé pour de hautes fréquences, le filtre est dit *passe-haut*. Dans ce cas, seules les hautes fréquences de l'entrée sont préservées au travers du filtre. Ainsi, lorsque l'on cherche à estimer une « tendance », c'est-à-dire une série ayant une fréquence plus faible que celle des données observées, on utilise un filtre passe-bas.

On appelle filtre *idéal* un filtre possédant une fonction de transfert défini par :

$$H(f) = \mathbf{1}_{f_l \leq f \leq f_u}$$

où f_l est la fréquence de coupure la plus faible du filtre et f_u est la fréquence de coupure la plus élevée du filtre. Pour un filtre passe-bas idéal, on doit ainsi avoir $f_l = 0$ et $f_u < 0.5$. Un filtre passe-haut idéal aura des fréquences de coupure vérifiant $f_l > 0$ et $f_u = 0.5$. Néanmoins, en pratique, un filtre idéal n'est pas réalisable car cela nécessite d'estimer une infinité de coefficients ω_k . Ce filtre est donc approximé.

Considérons par exemple une simple moyenne mobile :

$$y_t = \frac{1}{2}(x_t + x_{t-1})$$

En substituant x_t par l'exponentielle complexe on obtient

$$y_t = \frac{1}{2}(e^{2i\pi f t} + e^{2i\pi f(t-1)}) = \left(\frac{1}{2} + \frac{1}{2}e^{-2i\pi f}\right)e^{2i\pi f t}$$

La fonction de transfert de cette moyenne mobile se définit donc comme

$$H(f) = \frac{1}{2}(1 + e^{-2i\pi f})$$

Elle est égale à 1 si $f = 0$. Lorsque f tend vers 0, la fonction de transfert évolue dans un voisinage de 1. C'est donc un filtre passe-bas. Lorsqu'on exprime cette fonction de transfert en coordonnées polaires,

$$H(f) = \frac{1}{2}(e^{-i\pi f} + e^{i\pi f})e^{-i\pi f} = \cos(\pi f)e^{-i\pi f}$$

La fonction de gain correspond donc à $\cos(\pi f)$ et l'angle de la phase à π . Puisque l'angle de la phase du filtre est non nul, la sortie du filtre possèdera une fonction de phase différente de celle de la série d'entrée. Ce phénomène est appelé *changement de phase* du filtre.

Ce changement de phase est important dans l'étude de séries économiques et financières. Si un filtre possède une phase non nulle, la série de sortie du filtre aura des points de retournement ayant lieu à des instants différents du signal d'entrée. Il est ainsi intéressant de sélectionner un filtre de phase nulle pour préserver les propriétés de phase de la série d'entrée.

Par exemple, une moyenne mobile centrée permet de conserver ces propriétés de phase. L'angle de la phase sera nul pour toutes les moyennes mobiles centrées symétriques car leur fonction de transfert est donnée par

$$H(f) = \omega_0 + 2 \sum_{k=1}^N w_k \cos(2\pi f k) = G(f)$$

où $N \in \mathbb{N}$ le nombre de valeurs passées et futures de l'entrée prises en compte dans le filtre.

Un filtre causal à réponse impulsionnelle finie peut également être un filtre passe-haut. Par exemple,

$$y_t = \frac{1}{2}(x_t - x_{t-1})$$

a la réponse fréquentielle suivante :

$$H(f) = \sin(\pi f) i e^{-i\pi f}$$

où le gain est nul lorsque la fréquence est nulle et augmente lorsque f augmente, atteignant son maximum à la fréquence $f = \frac{1}{2}$.

3.1.3 Filtrage linéaire optimum et nécessité de méthodes locales et adaptatives

Dans cette partie, on étudie la détermination de coefficients optimaux, au sens où il minimise l'erreur quadratique moyenne, pour un filtre à réponse impulsionnelle finie. Supposons que y_t soit un signal stationnaire au second ordre se décomposant de la manière suivante :

$$y_t = f(t) + \epsilon_t \tag{3.7}$$

où $(y_t)_{t \in \mathbb{N}}$ est le signal observé, f est une fonction déterministe du temps et $(\epsilon_t)_{t \in \mathbb{N}}$ est un bruit décorrélé et distribué avec une moyenne nulle et de variance constante. Le filtrage consiste à estimer $(f(t))_{t \in \mathbb{N}}$ d'après $(y_t)_{t \in \mathbb{N}}$. Une estimée linéaire de $f(t)$ peut être obtenue en assignant un poids w_i à chaque observation et en calculant une moyenne pondérée des observations via

$$\hat{f}(t) = \sum_{i=-N}^M w_i f(t-i) + \sum_{i=-N}^M w_i \epsilon_{t-i} \tag{3.8}$$

où N et M sont des entiers naturels. La performance de cet estimateur est évaluée en utilisant l'erreur d'estimation définie par la différence entre le signal $f(t)$ et l'estimée $\hat{f}(t)$. Si la valeur espérée de cette erreur est nulle, l'estimateur correspondant est dit *non biaisé*. On peut également évaluer l'adéquation aux données à l'aide de l'erreur quadratique moyenne qui correspond à l'espérance de cette erreur au carré. Elle peut s'exprimer comme la somme de la variance et du biais de l'estimateur au carré

$$EQM(\hat{f}(t)) = \mathbb{E}[(f(t) - \hat{f}(t))^2] = \mathbb{V}(\hat{f}(t)) + (\mathbb{E}[f(t) - \mathbb{E}[\hat{f}(t)]])^2 \tag{3.9}$$

Un estimateur sans biais est dit plus *efficace* qu'un autre estimateur sans biais s'il possède une variance plus faible [Gri(1993)].

On souhaite donc obtenir les coefficients $\mathbf{W} = (w_{-N} \dots w_M)^{Tr}$ d'un filtre permettant de minimiser l'erreur quadratique moyenne :

$$\min_{\mathbf{W}} \mathbb{E}[(f(t) - \hat{f}(t))^2] \tag{3.10}$$

On notera $e_t^2 = (f(t) - \hat{f}(t))^2$. En différenciant par rapport à w_j , on obtient :

$$\frac{\partial \mathbb{E}[e_t^2]}{\partial w_j} = -2\mathbb{E}[f(t) - \sum_{i=-N}^M w_i y_{t-i}] y_{t-j} = 0 \quad \forall j \in [-N, \dots, M] \quad (3.11)$$

Ce qui peut se réécrire sous la forme de l'équation dite de *Wiener-Hopf* [Wie(1964)] :

$$\sum_{i=-N}^M w_i \mathbb{E}[y_{t-i} y_{t-j}] = \mathbb{E}[f(t) y_{t-j}] \quad \forall j \in [-N, \dots, M] \quad (3.12)$$

Ainsi, un filtre dont les poids vérifient la relation (3.12) est censé être optimum. Cette équation peut également se réécrire comme

$$\mathbb{E}[e_t y_{t-j}] = 0 \quad \forall j \in [-N, \dots, M] \quad (3.13)$$

Cette condition est également appelée « condition d'orthogonalité » en théorie de l'estimation : l'erreur d'estimation du signal $f(t)$ est décorrélée de l'ensemble des observations $(y_k)_{k \in [-N, \dots, M]}$.

L'équation de *Wiener-Hopf* peut se réécrire sous forme matricielle :

$$\Gamma_{yy} \mathbf{W} = \Gamma_{fy} \quad (3.14)$$

avec

- Γ_{yy} la matrice de variance-covariance des observations

$$\Gamma_{yy} = \begin{pmatrix} \mathbb{E}[y_{t+N}^2] & \mathbb{E}[y_{t+N} y_{t+N-1}] & \dots & \mathbb{E}[y_{t+N} y_{t-M}] \\ \mathbb{E}[y_{t+N-1} y_{t+N}] & \mathbb{E}[y_{t+N-1}^2] & \dots & \mathbb{E}[y_{t+N-1} y_{t-M}] \\ \vdots & \vdots & \dots & \vdots \\ \mathbb{E}[y_{t-M+1} y_{t+N}] & \mathbb{E}[y_{t-M+1} y_{t+N-1}] & \dots & \mathbb{E}[y_{t-M+1} y_{t-M}] \\ \mathbb{E}[y_{t-M} y_{t+N}] & \mathbb{E}[y_{t-M} y_{t+N-1}] & \dots & \mathbb{E}[y_{t-M}^2] \end{pmatrix}$$

- Γ_{fy} le vecteur de covariance entre le paramètre de tendance $f(t)$ et le signal observé :

$$\Gamma_{fy} = (\mathbb{E}[f(t) y_{t+N}] \quad \mathbb{E}[f(t) y_{t+N-1}] \quad \dots \quad \mathbb{E}[f(t) y_{t-M}])^{Tr}$$

Si on connaît Γ_{yy} et Γ_{fy} et que la matrice de variance-covariance des observations est inversible, on peut déterminer les poids optimaux par :

$$\mathbf{W} = \Gamma_{yy}^{-1} \Gamma_{fy} \quad (3.15)$$

Lorsqu'on a obtenu les poids, on peut en déduire les erreurs d'estimation :

$$\mathbb{E}[e_t^2] = \mathbb{E}[f(t)^2] - \mathbb{E}[(\sum_{i=-N}^M w_i y_{t-i})(f(t) - e_t)] \quad (3.16)$$

où $f(t) - e_t = \hat{f}(t)$. En utilisant la condition d'orthogonalité, on a alors

$$\mathbb{E}[e_t^2] = \mathbb{E}[f(t)^2] - \mathbb{E}[(\sum_{i=-N}^M w_i y_{t-i}) f(t)] = \mathbb{E}[f(t)^2] - \sum_{i=-N}^M w_i \mathbb{E}[f(t) y_{t-i}] = \mathbb{E}[f(t)^2] - \Gamma_{fy}^{Tr} \mathbf{W} \quad (3.17)$$

Ce filtrage théorique est connu sous le nom de *filtre de Wiener* [Wie(1964)]. Il permet d'obtenir des poids optimums, au sens où ils minimisent l'erreur quadratique moyenne, lorsque la sortie du filtre y_t est stationnaire au second ordre et lorsque N , M , Γ_{yy} et Γ_{fy} sont connus. Malheureusement, en pratique, l'hypothèse de stationnarité au second ordre est souvent rejetée dans le cadre des séries temporelles de données financières. Il faut alors supposer que Γ_{yy} et Γ_{fy} dépendent du temps. La détermination d'un filtrage optimal passe donc par l'estimation d'un aspect local, au sens où N et M permettent de définir un voisinage temporel de t à partir duquel estimer $f(t)$, ainsi que d'un aspect adaptatif en faisant dépendre Γ_{yy} et Γ_{fy} du temps.

La régression non-paramétrique est un domaine assez vaste qui a pour but de déterminer un filtrage optimal en supposant simplement une hypothèse de régularité sur la fonction f . Cette hypothèse de régularité se trouve à l'origine de l'aspect local et adaptatif de la tendance estimée. Dans le prochain paragraphe, nous présentons ses différentes problématiques.

3.2 Problématiques de régression non-paramétrique

On se place dans le cadre suivant :

On observe des variables aléatoires indépendantes $(t_i, y_i)_{i=1, \dots, n}$ où le vecteur de variables explicatives $t_i \in [0, 1]$ peut être vu comme une variable représentant le temps et déterminant la distribution de $y_i \in \mathbb{R}$, qui représente l'observation à l'instant t_i . On suppose que la distribution de chaque y_i est déterminée par un paramètre f_i qui peut dépendre de l'instant t_i , $f_i = f(t_i)$. Plus particulièrement, il est généralement supposé que $f(t_i) = \mathbb{E}[Y_i | t_i]$. On considère alors le modèle suivant :

$$y_i = f(t_i) + \epsilon_i \quad \forall i \in [1, \dots, n], n \in \mathbb{N} \quad (3.18)$$

avec $y_i \in \mathbb{R}$, $t_i = \frac{i - 0.5}{n}$, $f : [0, 1] \rightarrow \mathbb{R}$ et les ϵ_i étant indépendants et identiquement distribués tels que

- $\forall i \in [1, \dots, n], \mathbb{E}[\epsilon_i] = 0$
- $\forall i \in [1, \dots, n], \forall j \in [1, \dots, n], \mathbb{E}[\epsilon_i \epsilon_j] = \sigma^2 \mathbf{1}_{i=j}$, avec $\sigma^2 \in \mathbb{R}^{+, *}$.

3.2.1 Le régressogramme

Afin d'estimer la fonction f , le régressogramme est une des premières méthodes de régression non-paramétrique. Elle fut introduite par Tuckey [Tuk(1960)]. On suppose que l'on peut décomposer l'intervalle $[0, 1]$ en différents sous-intervalles $P_j = [\frac{j-1}{\lambda}, \frac{j}{\lambda}]$ pour tout $j \in [1, \dots, \lambda-1]$ et $P_\lambda = [\frac{\lambda-1}{\lambda}, 1]$ où $\lambda \in \mathbb{N}$. L'estimateur proposé est le suivant :

$$\forall j \in [1, \dots, \lambda], \quad \forall t_i \in P_j, \quad \hat{f}(t_i) = \frac{\sum_{i=1}^n y_i \mathbf{1}_{t_i \in P_j}}{\sum_{i=1}^n \mathbf{1}_{t_i \in P_j}} \quad (3.19)$$

\hat{f} est donc une fonction continue par morceaux et constante sur chaque intervalle P_j . Plus particulièrement, cette estimation correspond à la moyenne des données comprises dans l'intervalle P_j .

λ peut être vu comme un paramètre de lissage. Il nécessite d'être choisi de manière à ce qu'il ne soit ni trop petit ni trop grand. Par exemple, si $\lambda = 1$, $\hat{f}(t_i) = \frac{1}{n} \sum_{i=1}^n y_i$, la tendance n'est pas

définie de manière locale puisqu'elle correspond à la moyenne de l'ensemble de la série d'observations $(y_i)_{i \in [1, \dots, n]}$. A l'opposé, si $\lambda = n$, $\hat{f}(t_i) = y_i$ et la tendance est pleinement informative.

On se confronte donc à une problématique essentielle en régression non-paramétrique : comment déterminer le lissage de la tendance ? Il doit permettre à la tendance estimée d'avoir un certain degré de lissage, tout en minimisant l'inadéquation de cette tendance aux données.

3.2.2 L'hypothèse de régularité de la tendance

Le point de départ définissant le lissage de la tendance est une hypothèse de régularité faite sur la fonction f de l'équation (3.18). La régularité de cette fonction peut être caractérisée de différentes façons : il est courant de supposer $f \in L^2([0, 1])$ ou $f \in W^{2,2}([0, 1])$. Une manière générale de caractériser cette régularité est d'utiliser un exposant local de Hölder. Une fonction est dite continue h -Höldérienne au point t_0 si et seulement s'il existe un polynôme de degré inférieur à h , noté P , tel que

$$|f(t) - P(t - t_0)| \leq K_{t_0} |t - t_0|^h \quad (3.20)$$

dans un voisinage de t_0 et avec K_{t_0} une constante réelle. Notons $C^h(t_0)$ l'espace des fonctions à valeurs réelles qui vérifient la propriété précédente en t_0 . Une fonction f possède un exposant local de Hölder α si

$$\forall h < \alpha, \quad f \in C^h(t_0) \quad \text{et} \quad \forall h > \alpha, \quad f \notin C^h(t_0) \quad (3.21)$$

Soit $h_f(t)$ l'exposant local de Hölder de f en t . Si $h_f(t_0) \geq 1$, f est différentiable en t_0 , tandis qu'une discontinuité de f en t_0 implique que $h_f(t_0) = 0$. Plus généralement, plus la valeur de $h_f(t_0)$ est élevée, plus f est régulière en t_0 .

Le but de la régression est donc de trouver un estimateur \hat{f} dans un ensemble de fonctions suffisamment régulières \mathcal{G}_δ où δ est un paramètre de régularité (par exemple un exposant local de Hölder minimum), tel que $\mathbb{E}[(\hat{f} - f)^2]$ est aussi petite que possible. L'hypothèse de régularité est en général formulée de façon à ce que f puisse se décomposer sur une base de fonctions orthonormales prédéfinie.

3.2.3 La régression sur base de fonctions orthonormales

Définissons préalablement la notion de base orthonormale.

Définition 6. $\{x_j\}_{j=1}^\infty$ est dite une base orthonormale de $L^2[0, 1]$ si elle satisfait :

- $\langle x_i, x_j \rangle = \mathbf{1}_{i=j}$
- $\|x_j\|_{L^2[0,1]} = 1 \quad \forall j \in [1, \infty[$
- Si $f \in L^2[0, 1]$ et $\forall j \in [1, \infty[$, $f \perp x_j$ alors $f \equiv 0$

Supposons que $f \in L^2[0, 1]$ dans l'équation (3.18). En conséquence, f peut s'exprimer comme une combinaison linéaire de la base orthonormale $\{x_j\}_{j=1}^\infty$:

$$f = \sum_{j=1}^{\infty} \beta_j x_j \quad \text{où} \quad \beta_j = \langle f, x_j \rangle \quad (3.22)$$

L'équation précédente peut se réécrire comme $f = \beta^{Tr} X$ avec $\beta^{Tr} = (\beta_1, \beta_2, \dots)$ et $X = (x_1, x_2, \dots)^{Tr}$. Néanmoins, dans ce cas X ne peut être de rang plein puisqu'elle possède une infinité de colonnes. Il existe donc une combinaison linéaire nulle des colonnes, ce qui implique la non-unicité du vecteur β .

3.2. PROBLÉMATIQUES DE RÉGRESSION NON-PARAMÉTRIQUE

Afin de remédier à ce problème, la régression sur bases de fonctions orthonormales consiste à estimer le vecteur β en présupposant une base de fonctions orthonormales $\{x_j\}_{j=1}^{\infty}$ et un « paramètre de lissage » λ permettant de rendre X de rang plein. Concrètement, on exprime f comme :

$$f = b_{\lambda}^{Tr} X_{\lambda} + u_{\lambda} \quad (3.23)$$

avec $b_{\lambda}^{Tr} = (b_1, b_2, \dots, b_{\lambda})$ et $X = (x_1, x_2, \dots, x_{\lambda})^{Tr}$ et u_{λ} un vecteur de « bruit » modélisant l'adéquation aux données tel que $u_{\lambda} = \beta^{Tr} X - b_{\lambda}^{Tr} X_{\lambda}$. L'estimateur de f s'exprime alors comme

$$\hat{f}_{\lambda} = \sum_{j=1}^{\lambda} \hat{b}_j x_j \quad (3.24)$$

où $\hat{b} = (\hat{b}_1, \dots, \hat{b}_{\lambda})^{Tr}$ est obtenu en minimisant une mesure de l'inadéquation aux données, par exemple la somme des erreurs aux carrés :

$$\hat{b} = \arg \min_b \sum_{t=1}^n (y_t - \sum_{j=1}^{\lambda} \hat{b}_j x_j(t))^2 \quad (3.25)$$

λ peut être considéré comme un paramètre de lissage car les fonctions x_j pour $j < k$ ont généralement des bandes de fréquences inférieures aux fonctions x_j pour $j \geq k$.

Remarque 1. *L'avantage important de cette approche sur bases de fonctions orthogonales est que tous les β_j sont indépendants. Par conséquent, si certains sont considérés significativement nuls, on peut les supprimer sans pour autant réestimer tous les autres coefficients.*

Choix de la base de fonctions orthonormales La base de fonctions orthonormales est déterminée d'après l'hypothèse faite sur la régularité de la fonction f . En pratique, les bases de cosinus et les bases de polynômes sont assez populaires (voir par exemple [Ost(2007)] [Art(2002)]) :

- **Base de cosinus**

On suppose ici que f est une fonction réelle stationnaire au second ordre. Par conséquent, on peut la décomposer en séries de Fourier. En d'autres termes,

$$f(t) = \beta_0 + \sum_{k=1}^{\infty} \beta_k e_k(t) \quad \forall t \in [0, 1] \quad (3.26)$$

avec β_k la transformée de Fourier de f : $\beta_k = \int_0^1 e_k(t) f(t) dt$ et

$$e_k(t) = \sqrt{2} \cos(\pi k t) \quad \forall k > 0 \quad \forall t \in [0, 1]$$

- **Base de polynômes**

La décomposition sur base de polynômes orthogonaux provient du lemme suivant :

Lemme 1. *Soit $p_i(x)_{i \in [1, \dots, n]}$ une suite de polynômes orthogonaux tels que $p_i(x)$ soit un polynôme de degré i . N'importe quel polynôme de degré n peut alors s'écrire comme une combinaison linéaire des polynômes $p_0(x), \dots, p_n(x)$.*

Si on peut approximer f par un polynôme de degré n alors on peut l'approximer par une combinaison linéaire de polynômes orthogonaux. Il existe trois principales bases de polynômes orthogonaux : les polynômes de Jacobi, ceux de Laguerre et ceux de Hermite [Vil(2005)]. On utilise souvent une base de polynômes de Legendre [Ost(2007)], qui sont en réalité un cas particulier des polynômes de Jacobi [Gab(1939)]. Si on suppose que f est une fonction continue, elle peut s'approximer par un polynôme (Théorème d'approximation de Weierstrass), cette approximation peut alors se décomposer sur une base orthonormale de polynômes de Legendre :

$$f(t) \approx \sum_{j=0}^{\infty} \beta_j P_j(t) \quad (3.27)$$

$$\text{avec } P_j(t) = \frac{Q_j(t)}{\|Q_j(t)\|} \text{ et } Q_j(t) = \frac{1}{2^j} \sum_{k=0}^j (C_j^k)^2 (t-1)^{j-k} (t+1)^k.$$

Il existe d'autres bases orthonormales de $L^2[0, 1]$, telles que celles formées par les ondelettes par exemple (voir paragraphe 3.3.4).

Remarque 2. Une autre approche serait d'essayer d'estimer empiriquement la base de fonctions sur laquelle se décompose f sans présupposer une forme explicite de fonction. Cette approche correspond à celle développée récemment par la décomposition modale empirique (voir paragraphe 3.3.5).

3.2.4 Estimation du paramètre de lissage

La régression sur bases de fonctions orthonormales dépend essentiellement de deux ingrédients clés : le choix de la base de fonctions et le choix du paramètre de lissage λ . Ce sous-paragraphe est une présentation des principales méthodes existantes pour estimer ce paramètre.

Approches classiques : C_P de Mallows et Validation croisée Une première idée serait de choisir le paramètre de lissage de manière à ce qu'il minimise une fonction mesurant l'inadéquation aux données, par exemple la somme des résidus aux carrés :

$$\lambda = \arg \min_{\lambda^*} \|\mathbf{Y} - \hat{\mathbf{f}}_{\lambda^*}\|^2 \quad (3.28)$$

avec $\mathbf{Y} = (y_1 \ y_2 \ \dots \ y_n)^{tr}$ et $\hat{\mathbf{f}}_{\lambda^*}$. Malheureusement, il existe un biais entre la somme des résidus aux carrés et l'erreur du modèle $\|\mathbf{f} - \hat{\mathbf{f}}_{\lambda}\|^2$. Ce biais est dû au fait que l'on utilise le même échantillon de données pour l'estimation du modèle et pour l'évaluation du modèle. En effet,

$$\begin{aligned} \frac{1}{n} \mathbb{E}[\|\mathbf{Y} - \hat{\mathbf{f}}_{\lambda}\|^2] &= \frac{1}{n} \mathbb{E}[\|\mathbf{Y} - \mathbf{f}\|^2 + 2(\mathbf{Y} - \mathbf{f})^{Tr}(\mathbf{f} - \hat{\mathbf{f}}_{\lambda}) + \|\mathbf{f} - \hat{\mathbf{f}}_{\lambda}\|^2] \\ &= \sigma^2(1 - \frac{2}{n} tr(\mathbf{H})) + EQM(\hat{\mathbf{f}}_{\lambda}) \end{aligned} \quad (3.29)$$

avec $EQM(\hat{\mathbf{f}}_{\lambda})$ l'erreur quadratique moyenne telle que définie par l'équation (3.9) et \mathbf{H} la matrice telle que $\hat{\mathbf{f}}_{\lambda} = \mathbf{H}\mathbf{Y}$.

C'est la raison pour laquelle le C_P de Mallows a été proposé pour estimer λ [Mal(1973)]. Il se base sur l'équation (3.29) et s'écrit comme :

$$C_P(\lambda) = \frac{1}{n} \|\mathbf{Y} - \hat{\mathbf{f}}_{\lambda}\|^2 + \frac{2}{n} \sigma^2 tr(\mathbf{H}) \quad (3.30)$$

3.2. PROBLÉMATIQUES DE RÉGRESSION NON-PARAMÉTRIQUE

Une autre manière de faire est d'évaluer le modèle sur des échantillons différents de celui qui a servi à son estimation. C'est le principe de la validation croisée. On estime le modèle sur différents sous-échantillons issus de l'échantillon initial et on minimise la somme des erreurs d'estimation sur ces sous-échantillons.

La validation croisée ordinaire consiste à sélectionner n sous-échantillons de l'échantillon initial en retirant la donnée n° i pour le sous-échantillon n° i . Notons $\widehat{\mathbf{f}}_\lambda^{(-i)}$ l'estimateur de \mathbf{f} en retirant la $i^{\text{ème}}$ donnée. Le critère de validation croisée ordinaire se définit ainsi par :

$$OCV(\lambda) = \frac{1}{n} \sum_{i=1}^n (y_i - \widehat{\mathbf{f}}_\lambda^{(-i)}(t_i))^2 \quad (3.31)$$

Néanmoins, le calcul de ce critère est assez fastidieux puisqu'il faut estimer le modèle n fois. En régression linéaire, il est possible de réduire ces calculs en estimant seulement une fois le modèle et en estimant les coefficients de la diagonale de \mathbf{H} :

$$OCV(\lambda) = \frac{1}{n} \sum_{i=1}^n \left(\frac{y_i - \widehat{f}_\lambda(t_i)}{1 - H_{ii}} \right)^2 \quad (3.32)$$

Il est parfois difficile d'estimer les coefficients H_{ii} , notamment pour des questions d'inversion de matrices, et plus facile d'estimer la trace de \mathbf{H} . Une méthode souvent utilisées est celle proposée par [Hut(1990)] qui a prouvé le lemme suivant, permettant d'estimer la trace d'une matrice à l'aide de méthodes de Monte-Carlo :

Lemme 2. Soit A une matrice $n \times n$ avec $\text{tr}(A) \neq 0$. Soit z un vecteur aléatoire dont les coordonnées sont *i.i.d.* et suivent une loi de Rademacher (c.-à-d. $z_i \in \{-1, 1\}$ et $\mathbb{P}(z_i = 1) = \mathbb{P}(z_i = -1) = \frac{1}{2}$). Alors :

- $\mathbb{E}[z^T A z] = \text{tr}(A)$
- $\mathbb{V}[z^T A z] = 2(\|A\|_F^2 - \sum_{i=1}^n A_{ii}^2)$

C'est la raison pour laquelle le critère de validation croisée généralisée, consistant à remplacer les coefficients H_{ii} par leur moyenne, égale à $\text{tr}(\mathbf{H})/n$, est souvent utilisé en pratique. Il s'écrit comme :

$$GCV(\lambda) = \frac{1}{n} \frac{\sum_{i=1}^n (y_i - \widehat{f}_\lambda(t_i))^2}{\left(1 - \frac{\text{tr}(\mathbf{H})}{n}\right)^2} = \frac{n \|\mathbf{Y} - \widehat{\mathbf{f}}_\lambda\|^2}{(n - \text{tr}(\mathbf{H}))^2} \quad (3.33)$$

En procédant de manière plus générale, on peut également extraire V sous-intervalles de $[1, \dots, n]$ formant une partition de $[1, \dots, n]$. On retire ensuite, l'un après l'autre, chacun de ces sous-intervalles de l'intervalle initial, puis on calcule le critère de validation croisée sur les données restantes. C'est ce qu'on appelle une validation croisée « V -fold » [Gei(1975)] [Zha(1993)]. Si ces sous-intervalles sont de cardinaux égaux à $d = \frac{n}{V}$, on parle également de validation croisée « d -delete ».

Méthodes de jackknife et de bootstrap La méthode de jackknife « d -delete » se rapproche de la validation croisée d -delete et consiste à sélectionner aléatoirement s échantillons de taille d parmi n données [Efr(1983)], [Wu(1986)]. On a ainsi C_n^d combinaisons possibles. On supprime alors chacun de ces échantillons dans l'échantillon original l'un après l'autre puis on calcule les estimateurs correspondant $\widehat{f}_\lambda^{(1)}, \dots, \widehat{f}_\lambda^{(s)}$ sur les $n - d$ données restantes. Pour chaque estimateur on peut ensuite calculer un critère de validation croisée :

$$CV_d^k(\lambda) = \frac{1}{n} \sum_{i=1}^n (y_i - \widehat{f}_\lambda^{(k)}(i))^2 \quad (3.34)$$

On peut ainsi estimer un critère de validation croisée « jackknife d -delete », noté $CV^{J(d)}(\lambda)$:

$$CV^{J(d)}(\lambda) = \frac{1}{s} \sum_{k=1}^s CV_d^k \quad (3.35)$$

La méthode du bootstrap consiste à rééchantillonner les erreurs de prédictions [Wu(1986)], [Sti(1985)]. On commence par calculer l'estimateur \hat{f}_λ sur l'ensemble des données puis les erreurs de prédictions correspondantes $(\hat{\epsilon}_i)_{i \in [1, \dots, n]}$. On génère alors s échantillons de taille n dont chaque valeur $\epsilon_i^{(g)}$, pour $g \in [1, \dots, s]$ et $i \in [1, \dots, n]$, est générée suivant les probabilités $\mathbb{P}(\epsilon_i^{(g)} = \hat{\epsilon}_i) = \frac{1}{n}$. On crée ensuite s nouveaux échantillons de données de taille n en ajoutant à l'estimateur initial \hat{f}_λ chaque vecteur $\epsilon^{(g)}$ généré. Sur chaque nouvel échantillon généré on peut alors estimer un nouvel estimateur $\hat{f}_\lambda^{(g)}$ pour lequel on peut calculer un critère de validation croisé :

$$CV^g(\lambda) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{f}_\lambda^{(g)}(i))^2 \quad (3.36)$$

Ce qui permet une autre approximation de l'estimateur de validation croisée, notée $CV^B(\lambda)$:

$$CV^B(\lambda) = \frac{1}{s} \sum_{g=1}^s CV^g(\lambda) \quad (3.37)$$

Méthodes de plug-in Les méthodes de plug-in tentent d'estimer un paramètre de lissage de manière « globale ». Pour cela, elles tentent de minimiser une mesure de l'erreur « globale » de l'estimation du modèle, appelée erreur quadratique moyenne intégrée et notée « $MISE$ » :

$$MISE(\lambda) = \mathbb{E} \left[\int_{-\infty}^{\infty} (\mathbf{f}(x) - \hat{\mathbf{f}}_\lambda(x))^2 dx \right] = \mathbb{E} \left[\int_{-\infty}^{\infty} \hat{\mathbf{f}}_\lambda(x)^2 dx \right] + \mathbb{E} \left[\int_{-\infty}^{\infty} \mathbf{f}(x)^2 dx \right] - 2\mathbb{E} \left[\int_{-\infty}^{\infty} \hat{\mathbf{f}}_\lambda(x) \mathbf{f}(x) dx \right] \quad (3.38)$$

Le premier terme de l'équation (3.38) peut être évalué numériquement tandis que le second ne dépend pas du paramètre de lissage. Il faut alors approximer le troisième terme de cette équation. Ceci peut être fait en minimisant un critère de validation croisée [Loa(1999)] ou le critère C_p de Mallows [Rup(1995)]. En dérivant, par rapport au paramètre de lissage, l'approximation de l'erreur quadratique moyenne intégrée ainsi obtenue, on peut alors obtenir une expression explicite pour le paramètre de lissage optimal (c.-à-d. minimisant une approximation du critère $MISE$).

3.3 Méthodes non-paramétriques d'estimation de tendances en pratique

Dans ce paragraphe, nous présentons différentes méthodes de régression non-paramétrique permettant d'estimer une tendance déterministe au sein de séries d'observations et utilisables en temps réel. Nous comparons ces différentes méthodes et discutons de leurs performances sur des échantillons finis de données.

3.3.1 Techniques de filtrage non-paramétrique classiques en économie et finance

Nous introduisons ici quelques techniques de filtrage non-paramétrique souvent utilisée dans la littérature pour estimer la tendance de séries macroéconomiques ou financières. Les méthodes présentées sont les suivantes : le filtre d'Hodrick-Prescott [Hod(1981)], le filtre passe-bas de Baxter-King [Bax(1999)], celui de Butterworth [Gom(2001)] [Har(2003)] ainsi que celui de Christiano et Fitzgerald [Chr(2003)].

Le filtre d'Hodrick-Prescott Le filtre d'Hodrick-Prescott [Hod(1981)] consiste à minimiser la fonction suivante par rapport à $\hat{f} = (\hat{f}(t_i))_{i \in [1, \dots, n]}$:

$$\sum_{i=1}^n (y_i - \hat{f}(t_i))^2 + \lambda \sum_{i=1}^n (\hat{f}(t_i) - 2\hat{f}(t_i - 1) + \hat{f}(t_i - 2))^2 \quad (3.39)$$

où $\lambda \in \mathbb{R}$ correspond au paramètre de lissage. Ce paramètre peut soit être fixé par l'utilisateur, ce qui est assez subjectif, soit être estimé à l'aide d'une des méthodes présentées dans le paragraphe 3.2.4. La solution à cette minimisation est donnée par [Pai(2010)] :

$$\hat{f} = (I + \lambda \Delta_2^{Tr} \Delta_2)^{-1} \mathbf{Y} \quad (3.40)$$

où $\mathbf{Y} = (y_i)_{i \in [1, \dots, n]}$ et où Δ_2 est une matrice $(n-2) \times n$ définie par :

$$\Delta_2(i, j) = \begin{cases} 1 & \text{si } j = i \quad \forall i \in [1, \dots, n-2] \\ -2 & \text{si } j = i + 1 \quad \forall i \in [1, \dots, n-2] \\ 1 & \text{si } j = i + 2 \quad \forall i \in [1, \dots, n-2] \\ 0 & \text{sinon} \end{cases} \quad (3.41)$$

Ainsi $\Delta_2 \hat{f}$ correspond à une discrétisation de la dérivée seconde de \hat{f} .

Cette technique d'estimation de tendance n'est en fait qu'un cas particulier de la régression non-paramétrique par splines de lissage [Pai(2010)](voir le paragraphe sur les splines de lissage 3.3.3).

Le filtre de Baxter-King Le filtre passe-bas de Baxter-King [Bax(1999)] est une approximation d'un filtre passe-bas idéal par une moyenne mobile centrée symétrique. Un filtre passe-bande idéal symétrique peut s'exprimer de la manière suivante :

$$f(t_i) = \sum_{k=-\infty}^{+\infty} w_k y_{t_i+k}$$

où $w_k = \frac{1}{\pi k} (\sin(\frac{2\pi k}{p_{min}}) - \sin(\frac{2\pi k}{p_{max}})) \mathbf{1}_{k \neq 0} + (\frac{2}{p_{min}} - \frac{2}{p_{max}}) \mathbf{1}_{k=0}$, p_{min} et p_{max} étant les périodes minimum et maximum de la composante que l'on souhaite extraire ($\frac{1}{p_{min}}$ et $\frac{1}{p_{max}}$ correspondant aux fréquences maximum et minimum). Pour un filtre passe-bas idéal, on a $+\infty > p_{min} > 2$ et $p_{max} \rightarrow +\infty$.

Baxter et King ajustent le filtre passe-bas avec la contrainte que le gain s'annule lorsque la fréquence est nulle. Cette contrainte implique que la somme des coefficients de la moyenne mobile soit nulle. Il s'écrit donc comme :

$$f(t_i) = \hat{\omega}_0 y_{t_i} + \sum_{j=1}^N \hat{\omega}_j (y_{t_i-j} + y_{t_i+j})$$

avec $\hat{w}_j = w_j - \frac{1}{2N+1} \sum_{j=-N}^N w_j$.

Le filtre de Butterworth Le filtre passe-bas de Butterworth dépend d'un paramètre positif q et d'un entier positif m [Har(2003)]. Il peut s'exprimer à l'aide de l'opérateur retard L comme :

$$B_m(L) = \frac{1}{1 + q^{-1}|1 - L|^{2m}} \quad m \in \mathbb{N}^*$$

On a alors

$$B_m(L)y_t = \sum_{j=-N}^N w_j y_{t+j}$$

Les poids w_j sont déterminés par le filtre et donc dépendant de m et q . En particulier, pour $m = 1$, on peut montrer que ces poids décroissent à un taux exponentiel de manière symétrique (voir [Har(2003)]). La forme de ce filtre est motivée par ses propriétés dans le domaine fréquentiel. En effet, le gain de ce filtre peut s'exprimer comme :

$$B_m(e^{-2i\pi f}) = \frac{1}{1 + q^{-1}2^{2m} \sin^{2m}(\pi f)}$$

En particulier, si on pose $q = (2 \sin(\pi f_u))^{2m}$,

$$B_m(e^{-2i\pi f}) = \frac{1}{1 + \left(\frac{\sin(\pi f)}{\sin(\pi f_u)}\right)^{2m}}$$

f_u correspond alors à une fréquence à laquelle le gain égale 1/2. Le filtre passe-bas laisse alors passer les fluctuations dont la fréquence est inférieure à f_u . Lorsque m tend vers l'infini, le filtre converge vers un filtre passe-bas idéal.

Le filtre de Christiano-Fitzgerald Le filtre passe-bas de Christiano et Fitzgerald [Chr(2003)] est une approximation d'un filtre passe-bas idéal minimisant l'erreur quadratique entre la sortie du filtre et le filtre passe-bas idéal. Ils approximent un filtre passe-bande idéal par

$$\hat{B}(L)y_t = \sum_{k=-N}^M \hat{w}_{t,k} y_{t+k}$$

dont les poids $w_{t,k}$ sont approximés en minimisant l'erreur quadratique suivante :

$$\hat{w}_{t,k} = \arg \min \mathbb{E}[(B(L)y_t - \hat{B}(L)y_t)^2]$$

où $B(L)$ est le filtre passe-bande symétrique idéal défini précédemment (voir filtre de Baxter-King).

Les méthodes précédentes donnent un aperçu de ce qu'il est coutume d'utiliser en finance et en économie pour extraire la tendance d'une série temporelle de façon non-paramétrique. Néanmoins, elles reposent toutes sur un paramètre qui doit être déterminé : le paramètre de lissage pour le filtre d'Hodrick-Prescott et la fréquence maximale de coupure du filtre passe-bas idéal que tentent d'approximer les autres filtres (Baxter-King, Butterworth et Christiano-Fitzgerald). Une mauvaise

spécification de ce paramètre peut impliquer une tendance estimée éloignée de l'unique composante prévisible telle que définie dans le théorème de Cramér [Cra(1961)] [Cra(1962)].

Il faut remarquer que les filtres passe-bas de Baxter-King, Butterworth et Christiano-Fitzgerald nécessitent de préciser cette fréquence maximale de coupure mais également un horizon temporel (paramètre N) pour estimer la tendance.

Dans les prochains paragraphes, nous introduisons quatre techniques de régression non-paramétrique peu utilisées dans la littérature économique et financière.

3.3.2 La régression non-paramétrique polynomiale locale

On considère le modèle suivant :

$$y_i = f(t_i) + \epsilon_i \quad \forall i \in [1, \dots, n], n \in \mathbb{N} \quad (3.42)$$

avec $y_i \in \mathbb{R}$, $t_i = \frac{i - 0.5}{n}$, $f : [0, 1] \rightarrow \mathbb{R}$ et les ϵ_i étant indépendants et identiquement distribués tels que

- $\forall i \in [1, \dots, n], \mathbb{E}[\epsilon_i] = 0$
- $\forall i \in [1, \dots, n], \forall j \in [1, \dots, n], \mathbb{E}[\epsilon_i \epsilon_j] = \sigma^2 \mathbf{1}_{i=j}$, avec $\sigma^2 \in \mathbb{R}^{+,*}$.

Supposons que f soit une fonction $p + 1$ différentiable sur \mathbb{R} . On peut donc appliquer la formule de Taylor dans un voisinage de t_i (nous noterons ce voisinage $V(t_i)$) :

$$\forall t \in V(t_i), \quad f(t) = f(t_i) + (t - t_i)f'(t_i) + \frac{(t - t_i)^2}{2}f''(t_i) + \dots + \frac{(t - t_i)^p}{p!}f^{(p)}(t_i) + o((t - t_i)^p) \quad (3.43)$$

Ce qui se réécrit plus simplement sous forme vectorielle comme :

$$\forall t \in V(t_i), \quad f(t) \approx \delta_{t,t_i}^{tr} \mathbf{B}_{t_i} \quad (3.44)$$

où : $\delta_{t,t_i}^{tr} = (1 \quad (t - t_i) \quad (t - t_i)^2 \quad \dots \quad (t - t_i)^p)$ et $\mathbf{B}_{t_i} = \begin{pmatrix} f(t_i) & f'(t_i) & \frac{f''(t_i)}{2} & \dots & \frac{f^{(p)}(t_i)}{p!} \end{pmatrix}$. Pour estimer \mathbf{B}_{t_i} , on cherche alors à minimiser de manière locale la somme des carrés des erreurs d'estimation :

$$\widehat{\mathbf{B}}_{t_i} = \arg \min_{\mathbf{B}_*} \sum_{j=1}^n (y_j - \delta_{t_j,t_i}^{tr} \mathbf{B}_*)^2 K_h(\|t_j - t_i\|)$$

avec $K_h(\|t_j - t_i\|)$ une fonction « noyau » faisant correspondre un poids faible pour les données éloignées de t_i et fort pour celles proches de t_i . h est la « fenêtre » associée à cette fonction et permet de lisser plus ou moins la tendance. Cette minimisation peut également se réécrire de façon matricielle comme :

$$\widehat{\mathbf{B}}_{t_i} = \arg \min_{\mathbf{B}_*} (\mathbf{Y} - \mathbf{D}_{t_i} \mathbf{B}_*)^{tr} \mathbf{K}_{h,t_i} (\mathbf{Y} - \mathbf{D}_{t_i} \mathbf{B}_*)$$

avec :

- \mathbf{B}_* une matrice de taille $(p + 1) \times 1$.
- $\mathbf{Y} = (y_1 \quad y_2 \quad \dots \quad y_n)^{tr}$.
- $\mathbf{K}_{h,t_i} = (K_h(\|t_1 - t_i\|) \quad K_h(\|t_2 - t_i\|) \quad \dots \quad K_h(\|t_n - t_i\|)) \mathbf{1}_{n \times n}$ avec $\mathbf{1}_{n \times n}$ la matrice identité de taille $n \times n$.
- $\mathbf{D}_{t_i} = (\delta_{t_1,t_i}^{tr} \quad \delta_{t_2,t_i}^{tr} \quad \dots \quad \delta_{t_n,t_i}^{tr})^{tr}$.

Après annulation de la dérivée de l'équation par rapport à \mathbf{B}_* , on obtient l'estimateur suivant :

$$\widehat{\mathbf{B}}_{t_i} = (\mathbf{D}_{t_i}^{tr} \mathbf{K}_{h,t_i} \mathbf{D}_{t_i})^{-1} \mathbf{D}_{t_i}^{tr} \mathbf{K}_{h,t_i} \mathbf{Y}$$

On peut ensuite déterminer le biais et la variance de cet estimateur (voir [Tsy(2004)]) :

- $\mathbb{E}[\widehat{\mathbf{B}}_{t_i} | t_1, \dots, t_n] - \mathbf{B}_{t_i} = (\mathbf{D}_{t_i}^{tr} \mathbf{K}_{h,t_i} \mathbf{D}_{t_i})^{-1} \mathbf{D}_{t_i}^{tr} \mathbf{K}_{h,t_i} \xi$ avec $\xi = (\epsilon_1 \dots \epsilon_n)^{tr}$
- $\mathbb{V}[\widehat{\mathbf{B}}_{t_i} | t_1, \dots, t_n] = \sigma^2 \text{diag}((\mathbf{D}_{t_i}^{tr} \mathbf{K}_{h,t_i} \mathbf{D}_{t_i})^{-1})$ où $\text{diag}(M)$ correspond à la diagonale de la matrice M .

Au final, la régression non-paramétrique polynomiale locale dépend de trois choix importants : celui de l'ordre p du polynôme d'approximation, le choix de la fonction « noyau » $K_h(\cdot)$ et le choix de la fenêtre h .

Choix de l'ordre du polynôme d'approximation. En pratique, l'ordre du polynôme d'approximation, noté p , doit être choisi de manière à ce qu'il ne soit pas trop élevé. En effet, augmenter l'ordre du polynôme permet de minimiser l'erreur quadratique du modèle mais diminue le lissage de la tendance et de ses dérivées. Ce lissage peut ensuite être compensé à l'aide de la fenêtre h . Mais il faut également que l'estimation de la tendance et de ses dérivées ne prenne pas un temps de calcul trop important (notamment au niveau de l'inversion de la matrice $\mathbf{D}_{t_i}^{tr} \mathbf{K}_{h,t_i} \mathbf{D}_{t_i}$ qui est de taille $p \times p$).

Au niveau asymptotique, il a été montré qu'un estimateur polynomial local de $\mu^{(k)}(t_i)$, avec k un entier correspondant à l'ordre de la dérivée, possède des propriétés intéressantes aux bords de l'échantillon si $p-k$ est impair (notamment au niveau de la correction des effets de bord [Che(1997)]).

Lorsque l'on travaille sur des échantillons finis, on peut choisir raisonnablement p entre 1 et 3 pour estimer une tendance.

Choix de la fonction « poids ». Fan et Gijbels [Fan(1996)] ont montré que la fonction asymptotiquement optimale afin d'estimer $\mu^{(k)}(t_i)$ est la fonction de Epanechnikov définie par

$$K_h(x) = \frac{3}{4} \left(1 - \frac{x^2}{h^2}\right) \mathbf{1}_{]-1,1[}\left(\frac{x}{h}\right)$$

Néanmoins, Seifert et Gasser [Sei(1996)] ont montré que la densité normale standard $\mathcal{N}(0, 1)$ possède des propriétés intéressantes en échantillon fini :

$$K_h(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2h^2}}$$

Choix de la fenêtre. La sélection d'une fenêtre est cruciale car elle détermine le niveau du lissage de la courbe. Dans la littérature, de nombreuses études se sont penchées sur cette question [Mül(1987)], [Vie(1991)], [Bro(1993)]. Les méthodes couramment utilisées, comme la validation croisée, ont été présentées dans le paragraphe 3.2.4. En pratique, une méthode *plug-in* donnant de bons résultats est celle proposée par Ruppert, Sheather et Wand [Rup(1995)] combinée avec une densité normale standard comme fonction « noyau ».

3.3.3 La régression à l'aide de fonctions splines

Reprenons le modèle (3.42). Nous avons vu que la régression polynomiale locale utilise un développement de Taylor à l'ordre p . Elle suppose alors que $o((t - t_i)^p)$ est négligeable dans l'équation (3.43) pour t appartenant à un voisinage de t_i . Les splines de régression se basent quant à elles sur la formule de Taylor avec un reste intégral. En effet, dans ce cas, en supposant que f est $p + 1$ différentiable et que sa $p + 1^{\text{ème}}$ dérivée est intégrable sur $[t_i, t]$:

$$\forall t > t_i, \quad f(t) = f(t_i) + (t - t_i)f'(t_i) + \dots + \frac{(t - t_i)^p}{p!}f^{(p)}(t_i) + \int_{t_i}^t \frac{(t - x)^p}{p!}f^{(p+1)}(x)dx \quad (3.45)$$

Il faut remarquer que, dans cette équation, t n'appartient pas forcément à un voisinage de t_i . Le reste intégral nécessite donc d'être approximé. L'idée des splines de régression est de sélectionner un certain nombre de points, aussi appelés noeuds et notés $\xi = (\xi_1, \dots, \xi_k)$, appartenant à $[t_i, t]$ puis à estimer l'intégrale par :

$$\int_{t_i}^t \frac{(t - x)^p}{p!}f^{(p+1)}(x)dx \approx \sum_{j=1}^{k-1} \frac{(t - \xi_j)^p}{p!}f^{(p+1)}(\xi_j)(\xi_{j+1} - \xi_j)\mathbf{1}_{\xi_j \in [t_i, t]} \quad (3.46)$$

L'équation (3.45) peut alors se réécrire sous forme d'un produit scalaire :

$$\forall t > t_i, \quad f(t) \approx \delta_{t, t_i, \xi}^{Tr} \mathbf{B}_{t_i, \xi} \quad (3.47)$$

avec :

- $\delta_{t, t_i, \xi}^{Tr} = (1 \quad (t - t_i) \quad (t - t_i)^2 \quad \dots \quad (t - t_i)^p \quad (t - \xi_1)_+^p \quad \dots \quad (t - \xi_{k-1})_+^p)$
- $\mathbf{B}_{t_i, \xi}^{Tr} = \left(f(t_i) \quad f'(t_i) \quad \frac{f''(t_i)}{2} \quad \dots \quad \frac{f^{(p)}(t_i)}{p!} \quad \frac{f^{(p+1)}(\xi_1)}{p!}(\xi_2 - \xi_1) \quad \dots \quad \frac{f^{(p+1)}(\xi_{k-1})}{p!}(\xi_k - \xi_{k-1}) \right)$
- $(t - \xi_i)_+^p$ signifiant $(t - \xi_i)^p \mathbf{1}_{\xi_i \in [t_i, t]}$

Dans ce cas, f est une fonction spline (c.-à-d. définie par morceaux par des polynômes). En pratique, il est courant de supposer $t_i = 0$. La principale problématique concerne alors la sélection du vecteur des noeuds ξ . Pour un niveau de lissage équivalent, l'estimateur utilisant des noeuds mal placés aura un biais plus important qu'un autre estimateur.

Une idée serait de minimiser les erreurs d'estimation par rapport aux vecteurs ξ et $\mathbf{B}_{t_i, \xi}$:

$$\widehat{\mathbf{B}}_{t_i, \xi} = \arg \min_{\xi, \mathbf{B}_*} \sum_{j=1}^n (y_j - \delta_{t_j, t_i, \xi}^{Tr} \mathbf{B}_*)^2 \quad (3.48)$$

Néanmoins, cette approche nécessite tout de même de fixer la taille du vecteur ξ (c.-à-d. le nombre des noeuds). Dans ce cadre, la fonction à minimiser n'est pas linéaire par rapport à ξ . Il faut utiliser des méthodes d'optimisation numérique du type Gauss-Newton modifié [Gal(1973)] ou des algorithmes d'évolution différentielle [Pri(1997)] pour cette minimisation. Jupp [Jup(1978)] a cependant montré qu'il n'y a pas unicité de la solution au problème de minimisation de l'équation (3.48).

D'une manière générale, l'utilisation des splines de régression est assez fastidieuse à mettre en oeuvre en pratique. Le problème du nombre et de l'emplacement des noeuds est complexe à résoudre et ne plaide pas en sa faveur. En revanche, l'utilisation des splines de lissage est une autre manière d'utiliser les fonctions splines pour estimer une tendance qui s'avère plus facile à mettre en oeuvre.

Les splines de lissage Les splines de lissage consistent à minimiser la fonction suivante :

$$\sum_{i=1}^n (y_i - \hat{f}(t_i))^2 + \lambda \int_{x_1}^{x_n} \hat{f}^{(2)}(x)^2 dx \quad (3.49)$$

λ permet de donner plus ou moins d'importance au lissage de la tendance. Lorsqu'il tend vers 0, il n'y a pas de lissage et le signal obtenu est pleinement informatif. Lorsque λ tend vers l'infini, l'estimée obtenue tend vers celle obtenue par les moindres carrés ordinaires. Ce modèle fait parti des modèles dits « pénalisés » dans le sens où la fonction à minimiser dépend d'un terme représentant l'inadéquation aux données et d'un paramètre permettant de donner plus ou moins d'importance à la complexité du modèle (pénalisant la réduction de l'inadéquation aux données).

L'idée est alors d'approximer $\hat{f}(x)$ par une fonction d'interpolation passant par les points $\hat{f}(t_1), \dots, \hat{f}(t_n)$. Dans ce cas,

$$\hat{f}(x) = \sum_{i=1}^n \hat{f}(t_i) B_i(x) \quad (3.50)$$

où $(B_i(x))_{i \in [1, \dots, n]}$ forme une base de fonctions splines. En conséquence, on peut exprimer l'intégrale de la dérivée seconde au carré de la tendance comme :

$$\int \hat{f}^{(2)}(x)^2 dx = \hat{f}^{Tr} A \hat{f} \quad (3.51)$$

avec $A_{ij} = \int B_i^{(2)}(x) B_j^{(2)}(x) dx$ pour $i \in [1, \dots, n]$ et $j \in [1, \dots, n]$. Ce qui permet de réécrire la fonction que l'on cherche à minimiser comme

$$(Y - \hat{f})^{Tr} (Y - \hat{f}) + \lambda \hat{f}^{Tr} A \hat{f} \quad (3.52)$$

Lorsqu'on dérive cette fonction par rapport à \hat{f} , on obtient

$$\hat{f} = (I + \lambda A)^{-1} Y \quad (3.53)$$

Pour évaluer la matrice A , les fonctions B -splines sont généralement utilisées car elles permettent des calculs simples et rapides. Une fonction B -spline d'ordre 1 se définit par

$$B_i^0(x) = \mathbf{1}_{x \in [t_i, t_{i+1}[} \quad (3.54)$$

et les B -splines d'ordre $k \geq 1$, par la relation de récurrence

$$B_i^k(x) = \frac{x - t_i}{t_{i+k-1} - t_i} B_i^{k-1}(x) + \frac{t_{i+k} - x}{t_{i+k} - t_{i+1}} B_{i+1}^{k-1}(x) \quad (3.55)$$

avec $i \in [0, \dots, T - k - 1]$. Il est courant d'utiliser des B -splines cubiques ou quadratiques, ce qui correspond à $k = 2$ ou 3 . On peut déduire de cette définition la dérivée d'une fonction B -splines en raisonnant par récurrence :

$$\frac{d}{dx} B_i^k(x) = (k - 1) \left\{ \frac{B_i^{k-1}(x)}{t_{i+k-1} - t_i} - \frac{B_{i+1}^{k-1}(x)}{t_{i+k} - t_{i+1}} \right\} \quad (3.56)$$

Ces formules permettent de calculer la matrice A et donc d'obtenir \hat{f} .

Remarque 3. *Lien avec le filtre d'Hodrick-Prescott [Hod(1981)]. Paige et Trindade [Pai(2010)] ont montré que le filtre d'Hodrick-Prescott est un cas spécial des splines de lissage où les $(t_i)_{i \in [1, \dots, N]}$ sont équirépartis, les $(\epsilon_i)_{i \in [1, \dots, N]}$ sont décorrélés et le degré des splines est de 1.*

3.3.4 La régression en utilisant des bases d'ondelettes

L'idée essentielle de la méthodologie des ondelettes est d'observer une série temporelle à différentes échelles de temps. Son cadre est similaire à l'analyse spectrale dans laquelle chaque fonction du domaine temporel peut être exprimée comme une somme pondérée de fonctions trigonométriques ayant différentes fréquences. Néanmoins, la fonction trigonométrique continue pendant une période de temps infinie dans le domaine spectral tandis que la fonction de base de l'analyse en ondelettes existe pendant un temps limité.

Analyse multirésolution orthonormée et approximation de fonctions L'idée d'une analyse multirésolution orthonormée de $L^2(\mathbb{R})$ est d'approximer l'espace $L^2(\mathbb{R})$ par une suite croissante d'espaces fermés $(V_j)_{j \in \mathbb{Z}}$. Chaque V_j représente un « espace d'approximation » dans lequel peut se définir l'image d'une fonction approximant une fonction f appartenant à $L^2(\mathbb{R})$. Lorsqu'on incrémente la suite $(V_j)_{j \in \mathbb{Z}}$, l'« espace d'approximation » devient alors de plus en plus proche de $L^2(\mathbb{R})$. L'approximation d'une fonction f de $L^2(\mathbb{R})$ est alors de plus en plus fine. Plus rigoureusement, une analyse multirésolution se définit de la façon suivante :

Définition 7. Une analyse multirésolution orthonormée de $L^2(\mathbb{R})$ est une suite de sous-espaces fermés $(V_j)_{j \in \mathbb{Z}}$ vérifiant les propriétés suivantes :

- $\forall j \in \mathbb{Z}, V_j \subset V_{j+1}$
- $\bigcap_{j \in \mathbb{Z}} V_j = \{0\}$
- $\bigcup_{j \in \mathbb{Z}} V_j$ est dense dans $L^2(\mathbb{R})$
- $\forall (j, k) \in \mathbb{Z}^2, f \in V_j \Leftrightarrow f(\cdot - 2^{-j}k) \in V_j$
- Soit $V_k \subset V_l$ ($k < l$). Si $f \in V_k$, il existe $g \in V_l$ telle que $g(x) = f(2^{l-k}x) \forall x \in \mathbb{R}$.
- Il existe $\varphi \in V_0$ telle que la famille $\{\varphi(\cdot - k), k \in \mathbb{Z}\}$ forme une base orthonormée de V_0 . Cette fonction φ est appelée fonction d'échelle de l'analyse multirésolution.

Ainsi, la famille de fonctions $\{\varphi_{jk} : x \rightarrow 2^{\frac{j}{2}}\varphi(2^jx - k)\}_{k \in \mathbb{Z}}$ forme une base orthonormée de V_j pour la norme $L^2(\mathbb{R})$. Plus généralement, une analyse multirésolution peut se définir à partir de bases de Riesz (dans ce cas $\{\varphi(\cdot - k), k \in \mathbb{Z}\}$ est une base de Riesz). Nous ne traiterons cependant pas de ce cas. Le lecteur intéressé par celui-ci pourra se référer à [Dau(1992)] [Mey(1992)] [Mal(1999)].

Passer d'une résolution j à une résolution $j+1$ revient à ajouter des « détails ». Plus formellement, puisque $V_j \subset V_{j+1}$ pour tout $j \in \mathbb{Z}$, il existe un espace complémentaire à V_j dans V_{j+1} , que l'on notera W_j , définit par :

$$V_{j+1} = V_j \oplus W_j$$

Puisque les espaces V_j sont emboîtés, nous pouvons écrire que l'espace $V_{j_0} \oplus (\bigoplus_{j=j_0}^{+\infty} W_j)$ est un espace dense dans $L^2(\mathbb{R})$. Similairement à ce que nous avons écrit précédemment pour φ , il existe une fonction ψ telle que $\{x \rightarrow \psi(x - k)\}_{k \in \mathbb{Z}}$ soit une base orthonormée de W_0 . Cette fonction ψ est appelée ondelette. La famille $\{\psi_{jk} : x \rightarrow 2^{\frac{j}{2}}\psi(2^jx - k)\}_{k \in \mathbb{Z}}$ forme alors une base orthonormée de l'espace W_j . Puisque $V_{j_0} \oplus (\bigoplus_{j=j_0}^{+\infty} W_j)$ est dense dans $L^2(\mathbb{R})$, on déduit de ce qui précède la décomposition suivante pour toute fonction f appartenant à $L^2(\mathbb{R})$:

$$f = \sum_{k \in \mathbb{Z}} \alpha_{j_0 k} \varphi_{j_0 k} + \sum_{j=j_0}^{+\infty} \sum_{k \in \mathbb{Z}} \beta_{jk} \psi_{jk} \quad (3.57)$$

avec $\alpha_{j_0k} = \int_{\mathbb{R}} f(t)\varphi_{j_0k}(t)dt$ et $\beta_{jk} = \int_{\mathbb{R}} f(t)\psi_{jk}(t)dt$.

Si f est définie sur un intervalle $[0, 1]$, il est possible de construire une base d'ondelettes sur $L^2[0, 1]$ en périodisant les fonctions φ et ψ . Cette périodisation restreint l'analyse multirésolution à $[0, 1]$ mais entraîne des problèmes de discontinuités aux bords de l'intervalle. Soient $\tilde{\varphi}$ et $\tilde{\psi}$ des versions périodiques de φ et ψ sur $[0, 1]$. Toute fonction f de $L^2([0, 1])$ peut alors se décomposer sous la forme :

$$f = \sum_{k \in \mathbf{I}(j_0)} \alpha_{j_0k} \tilde{\varphi}_{j_0k} + \sum_{j \geq j_0} \sum_{k \in \mathbf{I}(j)} \beta_{jk} \tilde{\psi}_{jk} \quad (3.58)$$

avec $\alpha_{j_0k} = \int_0^1 f(t)\tilde{\varphi}_{j_0k}(t)dt$ et $\beta_{jk} = \int_0^1 f(t)\tilde{\psi}_{jk}(t)dt$.

Régression non paramétrique à l'aide d'ondelettes On s'intéresse ici à l'estimation par ondelettes d'une fonction f définie sur $[0, 1]$ et à valeurs réelles :

$$y_i = f(t_i) + \epsilon_i \quad i = 1, \dots, n \quad \text{avec } n = 2^J, J \in \mathbb{N} \quad (3.59)$$

f peut être estimée de deux manières différentes. La première façon est d'estimer f par une estimation par noyau, le noyau étant élaboré à l'aide d'ondelettes. On parle d'estimation « linéaire ». La seconde façon est d'estimer f en ne prenant pas en compte les coefficients d'ondelettes de moindre importance (relativement à l'information sur f) par des techniques de seuillage. On parle d'estimation « non-linéaire ».

Estimation linéaire

Soient $(t_i, y_i)_{i \in [1, n]}$ des observations. Soit une base d'ondelettes orthogonales générée par le couple (φ, ψ) . L'estimateur linéaire de f correspond donc à une estimation de la projection de f sur un espace V_λ de la façon suivante :

$$\hat{f} = \sum_{k=0}^{2^{j_0}-1} \hat{\alpha}_{j_0,k} \varphi_{j_0,k} + \sum_{j=j_0}^{\lambda} \sum_{k=0}^{2^j-1} \hat{\beta}_{j,k} \psi_{j,k} \quad (3.60)$$

où les coefficients $\hat{\alpha}_{j_0,k}$ et $\hat{\beta}_{j,k}$ sont des coefficients empiriques d'échelle et d'ondelette respectivement :

$$\hat{\alpha}_{j_0,k} = \frac{1}{n} \sum_{i=1}^n y_i \varphi_{j_0k}(t_i) \quad \text{et} \quad \hat{\beta}_{j,k} = \frac{1}{n} \sum_{i=1}^n y_i \psi_{jk}(t_i) \quad (3.61)$$

$\hat{\alpha}_{j_0,k}$ et $\hat{\beta}_{j,k}$ étant des estimateurs de $\alpha_{j_0k} = \int_0^1 f(t)\tilde{\varphi}_{j_0k}(t)dt$ et $\beta_{jk} = \int_0^1 f(t)\tilde{\psi}_{jk}(t)dt$.

λ est donc un paramètre de lissage. Plus il est grand et plus le nombre de coefficients de l'estimation est grand. Cette dernière est donc plus fine. L'estimateur en (3.60) peut se réécrire sous la forme

$$\forall x \in [0, 1], \quad \hat{f}_n(x) = \sum_{i=1}^n K_{j_0,\lambda}(x, t_i) \mathbf{Y}_i \quad (3.62)$$

où $K_{j_0,\lambda}$ est une fonction « noyau » définie par

$$\forall (u, v) \in [0, 1]^2, \quad K_{j_0,\lambda}(u, v) = \frac{1}{n} \sum_{k=0}^{2^{j_0}-1} \varphi_{j_0,k}(u)\varphi_{j_0,k}(v) + \frac{1}{n} \sum_{j=j_0}^{\lambda} \sum_{k=0}^{2^j-1} \psi_{jk}(u)\psi_{jk}(v) \quad (3.63)$$

Ainsi, le choix de λ dépend de la régularité de la fonction f que l'on cherche à estimer. On se confronte donc encore aux problèmes rencontrés précédemment sur la sélection du paramètre de lissage de la tendance (voir paragraphe 3.2.4). Ceci invite à utiliser des techniques non-linéaires d'estimation, dites de seuillage, car elles sélectionnent cette tendance, non pas en présupposant un degré de régularité précis sur la tendance, mais en ne retenant que l'« information essentielle » de la série temporelle. Dans ce cadre, il faut donc définir ce qui constitue cette « information essentielle ».

Estimation non-linéaire

Lorsqu'on écrit les coefficients en ondelettes, on remarque qu'ils se décomposent comme

$$\forall(j, k), \quad \widehat{\beta}_{jk} = \beta_{jk} + \frac{1}{n} \sum_{i=1}^n \epsilon_i \psi_{jk}(t_i) \quad (3.64)$$

où β_{jk} représente le niveau de détail. Si le niveau de bruit est élevé comparativement à β_{jk} , on peut supposer que le terme $\frac{1}{n} \sum_{i=1}^n \epsilon_i \psi_{jk}(t_i)$ est prédominant dans l'équation (3.64). L'idée est alors de ne sélectionner que les coefficients β_{jk} suffisamment significatifs pour estimer f . Ceux-ci seront supposés contenir l'essentiel de l'information sur f . Pour cela, on utilise une fonction seuil permettant d'annuler les coefficients en ondelettes pas assez significatifs. L'estimateur seuillé de f se définit alors comme

$$\widehat{f} = \sum_{k=0}^{2^{j_0}-1} \widehat{\alpha}_{j_0 k} \varphi_{j_0 k} + \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \gamma(\widehat{\beta}_{j_0 k}, \lambda_j) \psi_{j_0, k} \quad (3.65)$$

La fonction $\gamma(\cdot, \lambda_j)$ désigne la fonction de seuillage. Elle dépend du seuil λ_j . Les fonctions de seuillage les plus connues sont les suivantes :

- Seuillage fort : $\gamma_H(u, \lambda) = u \mathbf{1}_{|u| > \lambda}$
- Seuillage doux : $\gamma_S(u, \lambda) = \text{sign}(u)(|u| - \lambda)_+$

La procédure SCAD [Fan(2001)] [Ant(2001)] réalise un compromis entre seuillage doux et seuillage fort. Elle se définit comme

$$\gamma_{SCAD}(u, \lambda) = \begin{cases} \text{sign}(u)(|u| - \lambda)_+ & \text{si } |u| \leq 2\lambda \\ \frac{1}{a-2}((a-1)u - a\lambda \text{sign}(u)) & \text{si } 2\lambda \leq |u| \leq a\lambda \\ \text{sign}(u)(|u| - \lambda)_+ & \text{si } |u| > a\lambda \end{cases} \quad (3.66)$$

avec $a = 3.7$ d'après [Fan(2001)].

Il faut cependant noter que ces façons de seuiller les coefficients ne constituent pas une liste exhaustive et que d'autres manières sont envisageables.

Choix du seuil Dans l'équation (3.65), on définit un seuil λ_j à chaque niveau de résolution j . On peut néanmoins choisir un seuil unique sur l'ensemble des niveaux de résolution si les résidus $(\epsilon_i)_{i \in [1, \dots, n]}$ sont décorrélés. Le seuil utilisé est, la plupart du temps, un seuil dit « universel » et définit par [Don(1995)] :

$$\lambda = \sigma_\epsilon \sqrt{2 \log(n)}$$

où σ_ϵ représente l'écart-type des $(\epsilon_i)_{i \in [1, \dots, n]}$. Mais celui-ci entraîne souvent un surlissage de la fonction, on le corrige alors de la façon suivante :

$$\lambda = \sigma_\epsilon \sqrt{2 \log(n) - \log(\log(n))}$$

Pour estimer σ_ϵ , Donoho [Don(1995)] propose d'utiliser un estimateur *MAD* (*Mean Absolute Deviation*) appliqué aux coefficients d'ondelettes de résolution maximale.

3.3.5 La régression en utilisant une décomposition modale empirique

La régression utilisant une décomposition modale empirique est une manière différente de réaliser une estimation non-linéaire de la tendance. C'est une technique assez récente (article fondateur datant de 1998 [Hua(1998)]) et plutôt révolutionnaire dans le sens où elle permet de décomposer un signal sur une base de fonctions sans expliciter de manière formelle ces fonctions. Celles-ci sont directement estimées d'après les observations en utilisant un algorithme entièrement piloté par les données. Elle met en avant les caractéristiques locales d'un signal et peut s'appliquer sur des signaux non-stationnaires ou sur des données générées par des systèmes non-linéaires.

Il faut cependant noter que cette approche est assez intuitive et qu'il n'existe pas, aujourd'hui, de formalisme mathématique clair pour cette technique.

Fonctions modales intrinsèques La décomposition modale empirique repose sur un concept particulier, celui de fonction modale intrinsèque. Ce concept permet de fixer un cadre pour estimer la base de fonction sur laquelle décomposer le signal. Une fonction modale intrinsèque f vérifie les propriétés suivantes :

- $f : \mathbb{R} \rightarrow \mathbb{R}$
- f possède des maxima locaux strictement positifs et des minima locaux strictement négatifs.
- $\forall I \subset \mathbb{R}$, le nombre de maxima locaux et le nombre de minima locaux de f dans I diffèrent, au plus, de 1.
- La moyenne entre une fonction d'interpolation passant par les maxima locaux de f (enveloppe supérieure) et une fonction d'interpolation passant par les minima locaux de f (enveloppe inférieure) est nulle.

Les propriétés des fonctions modales intrinsèques sont motivées par le concept de fréquence instantanée [Hua(1998)] [Cex(2005)]. Traditionnellement, la fréquence instantanée d'un signal est définie à partir de sa transformée de Hilbert. Si y_t correspond au signal, sa transformée de Hilbert se définit par

$$y_t^H = \frac{1}{\pi} v.p. \left\{ \int_{\mathbb{R}} \frac{y_s}{t-s} ds \right\}$$

où *v.p.* désigne la valeur principale de Cauchy. Soit $z_t = y_t + iy_t^H = a(t)e^{i\theta_t}$ où $a(t) = |z_t|$ correspond à l'amplitude instantanée du signal et θ_t à sa phase instantanée. La fréquence instantanée à l'instant t se définit alors par $\omega_t = \theta'_t$.

De manière formelle, on peut ainsi écrire qu'une fonction modale intrinsèque correspond à une approximation d'une composante instantanée du signal :

$$f_j(t) \approx \mathcal{R}(a_j(t)e^{i\theta_j(t)})$$

où $\mathcal{R}(x)$ représente la partie réelle de x , $a_j(t)$ l'amplitude instantanée et $\phi_j(t)$ la phase de l'enveloppe de la $j^{\text{ème}}$ fonction modale intrinsèque.

La décomposition modale empirique consiste à exprimer n'importe quel type de signal (on ne fait donc aucune hypothèse sur la répartition du signal) comme une combinaison linéaire de fonctions modales intrinsèques. Ceci est possible grâce à un algorithme de tamisage.

Algorithme de tamisage Cet algorithme permet d'estimer ces fonctions modales intrinsèques. Notons $(y_t)_{t \in I_T \subset \mathbb{N}}$ la série observée. L'algorithme de tamisage est présenté dans le tableau 3.1.

Algorithme de tamisage	
Étape 1 :	Soit $g_t = y_t$ pour tout $t \in I_T$ et $i = 0$.
Étape 2 :	Détermination des maxima locaux et des minima locaux de $(g_t)_{t \in I_T}$. S'il existe au moins deux maxima, ils sont interpolés. La fonction d'interpolation est appelée enveloppe supérieure et notée g_t^\uparrow . S'il existe au moins deux minima, ils sont interpolés. La fonction d'interpolation est appelée enveloppe inférieure et notée g_t^\downarrow . S'il y a moins de deux minima ou moins de deux maxima, on pose $r_{i+1}(y_t) = g_t$ et on stoppe l'algorithme.
Étape 3 :	On pose $d_t = (g_t^\uparrow + g_t^\downarrow)/2$ et on met à jour $g_t : g_t = g_t - d_t$ pour tout $t \in I_T$. On détermine les maxima et minima locaux de $(g_t)_{t \in I_T}$. Si $(g_t)_{t \in I_T}$ n'a pas des maxima locaux strictement positifs et des minima locaux strictement négatifs, on retourne à l'étape 2.
Étape 4 :	$(g_t)_{t \in I_T}$ est une fonction modale intrinsèque. On pose $i = i + 1$, $f_i(t) = g_t$, $y_t = y_t - f_i(t)$ pour tout $t \in I_T$ et on retourne à l'étape 1.

TABLEAU 3.1 – Algorithme de tamisage permettant d'extraire les fonctions modales intrinsèques à partir d'une série d'observations.

Lorsque l'algorithme stoppe après N itérations, $(y_t)_{t \in I_T}$ peut alors s'exprimer comme :

$$\forall t \in I_T, \quad y_t = \sum_{i=1}^N f_i(t) + r_{N+1}(y_t) \tag{3.67}$$

où f_i correspond à la $i^{\text{ème}}$ fonction modale intrinsèque et $r_{N+1}(y_t)$ correspond à un résidu. Plus particulièrement, à chaque itération de l'algorithme, f_{i+1} aura une bande de fréquence plus faible que f_i .

Remarque sur l'interpolation des extremas Il existe trois familles de méthodes pour approximer les enveloppes supérieures et inférieures :

- la collocation : la fonction passe par les points à interpoler (ex : splines d'interpolation).
- l'osculation : la fonction passe par les points à interpoler et coïncide avec les n premières dérivées en ces points (ex : interpolation de Hermite).
- l'approximation : la fonction passe dans un voisinage des points à interpoler (ex : méthodes des moindres carrés, méthodes min-max).

En pratique, l'interpolation des extremas dans le cadre de la décomposition modale empirique se fait le plus souvent à partir de splines d'interpolation cubiques. Néanmoins, ce genre de méthode entraîne des distorsions du signal aux bornes de l'intervalle d'estimation I_T , ce qui pose problème

lorsque ce sont les dernières données qui importent le plus, comme c'est le cas en gestion de portefeuille.

Régression non paramétrique et décomposition modale empirique Supposons que les données $(y_t)_{t \in I_T \subset \mathbb{N}}$ soient générées par un modèle du même type que celui défini en (3.42). On a alors :

$$\forall t \in I_T, \quad y_t = f(t) + \epsilon_t = \sum_{i=1}^N f_i(t) + r_{N+1}(y_t) \quad (3.68)$$

Où f est un paramètre de tendance et ϵ un paramètre de bruit. L'idée est alors de fixer un seuil λ tel que

$$y_t - \sum_{i=1}^{\lambda} f_i(t) \approx f(t) \quad (3.69)$$

En effet, la première fonction modale intrinsèque aura une bande de fréquence plus élevée que toutes celles des autres fonctions modales intrinsèques. Si cette fréquence est très élevée, la fonction modale intrinsèque oscillera de façon très rapide et, de cette façon, sera quasiment imprévisible. En conséquence, la somme des premières fonctions modales intrinsèques qui peuvent être considérées comme imprévisibles peut s'assimiler à une variable aléatoire indépendamment distribuée.

Approches classiques du seuillage L'approche la plus courante du seuillage est celle proposée par Huang et al. [Hua(1998)]. Elle consiste à définir un critère d'arrêt dans l'algorithme de tamisage limitant l'écart-type entre deux étapes consécutives de tamisage. Ce critère s'écrit comme :

$$SD(i) = \sum_{t \in I_T} \left(\frac{f_i(t) - f_{i-1}(t)}{f_{i-1}(t)} \right)^2 \quad (3.70)$$

Typiquement, lorsque ce critère est inférieur à un certain seuil (Huang et al. [Hua(1998)] suggèrent 0.2 ou 0.3), l'algorithme est stoppé. Cette approche est néanmoins subjective puisque ce seuil est fixé par l'utilisateur.

Une autre manière de déterminer ce seuil est de procéder par validation croisée « *d-delete* » telle que définie dans le paragraphe 3.2.4. On sélectionne alors le niveau de lissage minimisant le critère de validation croisée. Néanmoins, ce genre de méthode nécessite un certain temps de calcul puisqu'il faut opérer la décomposition modale empirique sur un nombre important d'échantillons.

3.4 Comparaison des différentes méthodes

Nous avons précédemment présenté différentes méthodes d'estimation de tendances. Une question vient alors naturellement : Existe-t-il une méthode qui estime mieux que les autres la tendance ? Qu'apporte une méthode par rapport à une autre ? Dans ce paragraphe, on tente d'apporter des réponses à ces deux questions.

La meilleure manière de comparer ces méthodes semble d'utiliser des échantillons finis de données générées sur lesquels on estime une tendance à l'aide de chacune des différentes méthodes évoquées. On peut ensuite comparer les résultats à travers différents critères judicieusement sélectionnés.

3.4.1 Critères de comparaison

Les critères qui nous ont semblés les plus pertinents pour comparer ces méthodes d'estimation sont les suivants :

- l'écart-type du bruit résultant de l'estimation de la tendance et le rapport signal-à-bruit. Ces deux critères nous permettent de caractériser l'adéquation entre tendance estimée et données générées. Ils se définissent par :

$$\sigma_{\hat{\epsilon}} = \sqrt{\frac{1}{T-1} \sum_{t=1}^T (\hat{\epsilon}_t - \bar{\hat{\epsilon}})^2} \quad \text{et} \quad SNR = \frac{\sigma_{\hat{f}}^2}{\sigma_{\hat{\epsilon}}^2}$$

où $\hat{f}(t)$ est la tendance estimée, $\hat{\epsilon}_t = y_t - \hat{f}(t)$ et $\bar{\hat{\epsilon}}$ est la moyenne empirique des résidus.

- le nombre de points de retournement de la tendance estimée. Ce critère nous permet de savoir si la tendance estimée change de directions à de nombreuses reprises ou non. On définit un point de retournement de la tendance à l'horizon p comme :

$$Turn_p(t) = \mathbf{1}_{\text{sign}(\Delta_p \hat{f}(t+p)) \neq \text{sign}(\Delta_p \hat{f}(t))}$$

où $\Delta_p = 1 - L^p$ avec L l'opérateur retard. Le critère est donc défini par

$$NT_p = \sum_{t=p+1}^{T-p} Turn_p(t)$$

Dans notre étude, nous avons choisi NT_1 .

- le test de blancheur de Lyung-Box (statistique et p-value) appliqué au bruit résultant ainsi qu'au carré du bruit. Ces critères nous permettent de savoir si on peut accepter l'hypothèse selon laquelle les résidus, ainsi que leurs carrés, évoluent de manière indépendante.
- l'estimation du paramètre de mémoire des résidus et des résidus au carré. Ces critères permettent de quantifier la dépendance au sein du bruit. Pour évaluer ce critère, nous nous sommes inspirés de la méthode de Geweke et Porter-Hudak [Gew(1983)], reprise par [Rob(1995)]. On suppose que le processus générant les résidus s'écrit comme :

$$(1 - L)^d \epsilon_t = \delta_t$$

avec d un paramètre de mémoire, δ_t un bruit blanc de variance σ_{δ}^2 . L'idée est alors d'approximer la densité spectrale de puissance de ce processus afin d'en déduire d . La densité spectrale de ce processus peut notamment s'exprimer comme³

$$f(\lambda) = \frac{\sigma_{\delta}^2}{2\pi} [4 \sin^2(\frac{\lambda}{2})]^{-d}$$

3. $f(\lambda) = \frac{g_y(e^{-i\lambda})}{2\pi}$ avec $g_y(z)$ la fonction génératrice des auto-covariances se définissant comme $g_y(z) = \sigma_{\delta}^2 [(1 - z)(1 - z^{-1})]^{-d}$.

Geweke et Porter-Hudak [Gew(1983)] suggèrent alors d'exprimer le logarithme de la densité spectrale afin d'isoler d puis d'estimer $f(\lambda)$ à l'aide d'un périodogramme. Concrètement, on aboutit à

$$\log(I_k) = \log\left(\frac{\sigma_\delta^2}{2\pi}\right) - d \log\left(4 \sin^2\left(\frac{\pi k}{T}\right)\right) + \log\left(\frac{I_k}{f\left(\frac{2\pi k}{T}\right)}\right)$$

où $I_k = \frac{1}{2\pi T} \left| \sum_{t=0}^{T-1} y_t e^{2i\pi k \frac{t}{T}} \right|^2$. Le paramètre de mémoire d peut ensuite s'estimer à l'aide de la régression linéaire suivante :

$$\log(I_k) = \beta_1 + \beta_2 \log\left(4 \sin^2\left(\frac{\pi k}{T}\right)\right) + \epsilon_k \quad \forall k \in [1, \dots, m]$$

Ce qui donne l'estimateur suivant pour le paramètre de mémoire :

$$\hat{d} = -\hat{\beta}_2 = -\frac{\sum_{k=1}^m (X_k - \bar{X}) \log(I_k)}{\sum_{k=1}^m (X_k - \bar{X})^2}$$

où $X_k = \log\left(4 \sin^2\left(\frac{\pi k}{T}\right)\right)$ et $\bar{X} = \frac{1}{m} \sum_{k=1}^m X_k$. On peut également en déduire une estimation pour σ_δ^2 :

$$\widehat{\sigma_\delta^2} = 2\pi \times e^{\widehat{\beta}_1}$$

Il est à noter que nous n'avons pas retenu le critère d'erreur quadratique moyenne. D'une part, tous les modèles utilisés pour générer les données ne possèdent pas de tendance et, lorsqu'il n'en possède pas, ce critère est incalculable. D'autre part, la génération de données à l'aide d'un bruit évoluant autour d'une tendance prédéfinie peut localement modifier cette tendance. Il n'est ainsi pas rare que quelques données censées être générées indépendamment les unes des autres exhibent tout de même quelques dépendances localement. Afin de comparer les modèles, il nous semble donc plus judicieux de s'intéresser au comportement du bruit résultant de l'estimation de la tendance qu'à la différence entre tendance estimée et tendance ayant servi à générer les données.

3.4.2 Modèles utilisés pour la génération des données

Dans un premier temps, il peut être intéressant de savoir quelles méthodes arrivent à détecter qu'une série temporelle est générée par un modèle complètement déterministe, où aucune composante aléatoire n'est présente. Pour se faire, on utilise un premier modèle qui correspond à une suite chaotique de Feigenbaum :

- Modèle 1 :

$$y_t = 3.57 y_{t-1} (1 - y_{t-1})$$

avec $y_0 = 0.7$.

Dans un second temps, on s'intéresse à savoir si les tendances estimées possèdent un nombre de points de retournement similaires. Ceci nous a conduit à générer des données à l'aide des modèles suivant :

- Modèle 2 :

$$y_t = m_2(t) + \epsilon_2(t)$$

avec $m_2(t) = 10 - \left(\frac{25}{4} \left(t - \frac{1}{2}\right)\right)^2 + \cos(100t)$ et ϵ_2 un bruit blanc distribué suivant la loi $\mathcal{N}(0, 1)$.

- Modèle 3 :

$$y_t = m_3(t) + \epsilon_3(t)$$

avec $m_3(t) = 10 - 10t + (7t)^4 \mathbf{1}_{t \leq 0.25} + (7t - 3.5)^4 \mathbf{1}_{0.25 < t \leq 0.75} + (7t - 7)^4 \mathbf{1}_{t > 0.75}$ et ϵ_3 un bruit blanc distribué suivant la loi $\mathcal{N}(0, 1)$.

Enfin, on s'intéresse à comparer les modèles sur des séries générées proches de celles parfois rencontrées en finance. Dans ce but, nous utilisons trois modèles :

- Modèle 4 :

$$y_t = \sigma_t \epsilon(t) \tag{3.71}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \tag{3.72}$$

avec $\alpha_0 = 4.10^{-6}$, $\alpha_1 = 0.08$, $\beta_1 = 0.9$ et $\epsilon(t)$ un bruit blanc distribué suivant la loi $\mathcal{N}(0, 1)$.

- Modèle 5 :

$$(1 - L)^{0.15} y_t = \epsilon_t$$

L représentant l'opérateur retard et $\epsilon(t)$ un bruit blanc distribué suivant la loi $\mathcal{N}(0, 1)$.

- Modèle 6 :

$$y_t = (1 + r_t) y_{t-1}$$

avec r_t suivant le Modèle 4 et $y_0 = 100$.

Le modèle 4 est un modèle *GARCH*(1,1) proche de celui qu'il est possible d'inférer à partir de rendements quotidiens d'indices du marché action, le second est un processus stochastique à mémoire longue dont le paramètre de mémoire est proche de celui qu'il est possible d'estimer sur des séries de taux de change, enfin, le modèle 6 correspond à une série de prix dont les rendements sont générés à l'aide du modèle 4.

La figure 3.2 représente six séries générées à l'aide de chacun de ces modèles.

3.4.3 Méthodes d'estimation de tendances utilisées

Les méthodes utilisées pour l'estimation des tendances sont présentées dans le tableau 3.2.

3.4.4 Interprétation des résultats

On génère une série de longueur 1000 à l'aide du modèle 1 (déterministe) puis 100 séries de 1000 observations à l'aide des modèles 2 à 6 (aléatoires). On applique ensuite les méthodes de régression du tableau 3.2 puis on évalue les différents critères présentés. Pour les modèles 2 à 6, on estime une moyenne et un écart-type pour chacun des critères. Les résultats sont présentés dans les tableaux 3.3 à 3.8.

Concernant le premier modèle, entièrement déterministe, on constate que seule la régression par décomposition modale empirique et un critère de validation croisée (*EMD1*) ordinaire est en mesure de détecter que le bruit aléatoire n'existe pas. Ceci est dû à l'hypothèse de régularité faite par les autres méthodes qui est incompatible avec le comportement de la série temporelle générée par le premier modèle : on aboutit alors à des résultats incohérents.

Pour les modèles 2 et 3, on constate que la méthode *EMD1* est celle qui est le plus en adéquation avec les données : elle possède une variance moyenne du bruit inférieure aux autres méthodes (0.8954

3.4. COMPARAISON DES DIFFÉRENTES MÉTHODES

RNP1	Régression non-paramétrique polynomiale locale Fonction noyau : Loi normale standard, Fenêtre : critère GCV, Ordre polynomial de 1
RNP2	Régression non-paramétrique polynomiale locale Fonction noyau : Loi normale standard, Fenêtre : critère GCV, Ordre polynomial de 2
RNP3	Régression non-paramétrique polynomiale locale Fonction noyau : Loi normale standard, Fenêtre : critère GCV, Ordre polynomial de 3
RNP4	Régression non-paramétrique polynomiale locale Fonction noyau : Loi normale standard, Fenêtre : critère RSW, Ordre polynomial de 1
RNP5	Régression non-paramétrique polynomiale locale Fonction noyau : Loi normale standard, Fenêtre : critère RSW, Ordre polynomial de 2
RNP6	Régression non-paramétrique polynomiale locale Fonction noyau : Loi normale standard, Fenêtre : critère RSW, Ordre polynomial de 3
SPL1	Régression par splines de lissage Paramètre de lissage : critère de validation croisée ordinaire, Splines cubiques
SPL2	Régression par splines de lissage Paramètre de lissage : critère de validation croisée ordinaire, Splines quadratiques
OND1	Régression sur base d'ondelettes Ondelette de Haar, Transformée en ondelette discrète Procédure SCAD et seuillage universel
OND2	Régression sur base d'ondelettes Ondelette de Haar, Transformée en ondelette discrète (Maximum Overlap) Procédure SCAD et seuillage universel
OND3	Régression sur base d'ondelettes Ondelette Least Asymmetric 8, Transformée en ondelette discrète Procédure SCAD et seuillage universel
OND4	Régression sur base d'ondelettes Ondelette Least Asymmetric 8, Transformée en ondelette discrète (Maximum Overlap) Procédure SCAD et seuillage universel
EMD1	Régression par décomposition modale empirique Critère de Validation croisée
EMD2	Régression par décomposition modale empirique Critère de validation croisée et diminution du niveau de détail de 1
EMD3	Régression par décomposition modale empirique Critère de validation croisée et diminution du niveau de détail de 2

TABLEAU 3.2 – Méthodes de régression non-paramétrique utilisées pour estimer une tendance sur les séries générées.

pour le modèle 2 et 0.905 pour le modèle 3) ainsi qu'un rapport signal-à-bruit moyen plus élevé (11.64 pour le modèle 2 et 18.09 pour le modèle 3). Cette bonne adéquation aux données s'accompagne d'une fluctuation importante de la tendance estimée : elle se retourne en moyenne 163 fois pour le modèle 2 et 165 fois pour le modèle 3. De plus, le bruit résultant de l'estimation de tendance avec cette méthode n'est pas à mémoire courte ($d = 0$) comme on pourrait si attendre mais plutôt antipersistant ($-\frac{1}{2} < d < 0$) : ses autocorrélations décroissent hyperboliquement mais sa densité spectrale de puissance est dominée par des composantes de fréquences élevées. La méthode *EMD3* estime en

Séries générées à l'aide des 6 modèles

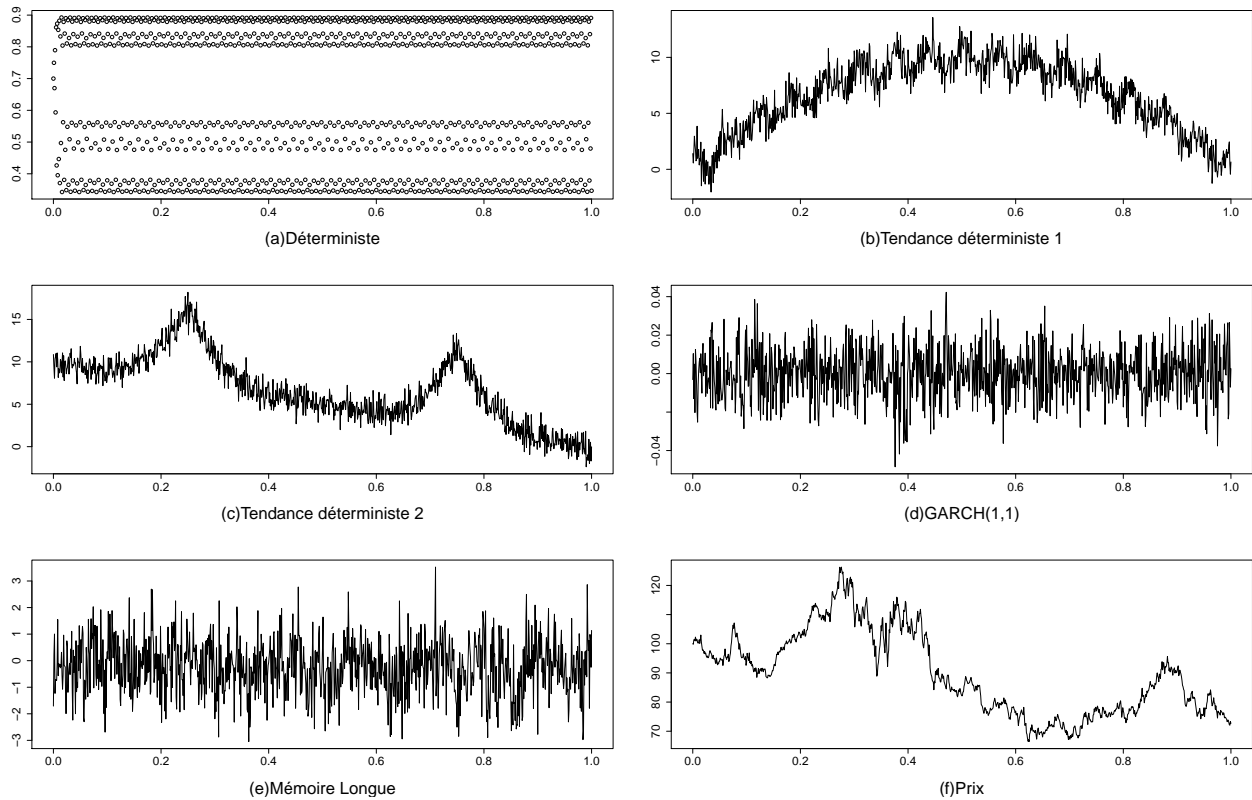


FIGURE 3.1 – (a) Suite de points générée à l’aide du modèle de Feigenbaum. (b) Série générée à l’aide du modèle 2. (c) Série générée à l’aide du modèle 3. (d) Série générée par le modèle 4. (e) Série générée à l’aide du modèle 5. (f) Série générée à l’aide du modèle 6.

moyenne un bruit proche de celui généré par un processus à mémoire courte et indépendamment distribué pour ces deux modèles. Le paramètre de mémoire du bruit est en moyenne proche de 0 (-0.03 pour le modèle 2 et -0.08 pour le modèle 3) et le test de Ljung-Box ne rejette pas l’hypothèse d’un bruit indépendamment distribué ($p.value$ de 0.274 pour le modèle 2 et de 0.365 pour le modèle 3).

En revanche, les autres méthodes ont des résultats moins évidents sur l’ensemble des modèles 2 et 3. Pour le modèle 2, les méthodes $RNP2$, $RNP3$, $OND1$, $OND2$, $OND3$ et $OND4$ ont en moyenne un bruit ayant un paramètre de mémoire proche de celui généré par un processus à mémoire courte mais le test de Ljung-Box rejette l’hypothèse d’un bruit indépendamment distribué ($p.values$ proches de 0). D’autres méthodes ont un bruit plutôt antipersistant mais ne rejettent pas l’hypothèse d’un bruit indépendamment distribué ($RNP1$, $RNP5$, $RNP6$) pour ce deuxième modèle. Pour le troisième modèle, plusieurs méthodes possèdent un bruit qui pourrait être généré de manière indépendante par un processus à mémoire courte (méthodes $RNP1$, $RNP4$, $OND3$ et $OND4$). La méthode $RNP4$ ne se retourne notamment qu’en moyenne 9 fois.

Le modèle 4 est un modèle GARCH(1,1) tel qu’on peut l’estimer d’après des rendements quotidiens du marché action. La méthode $EMD1$ possède la meilleure adéquation aux données : le bruit

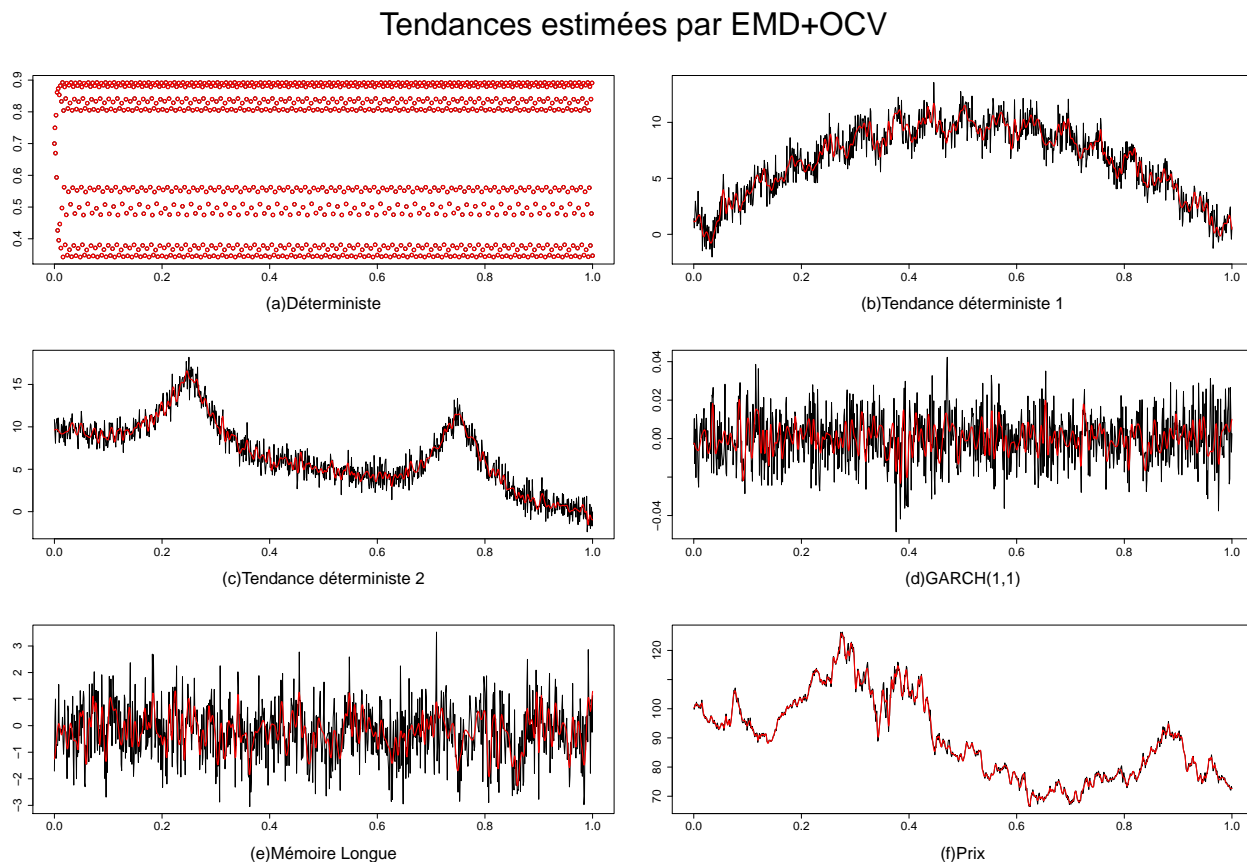


FIGURE 3.2 – En noir : série initiale d’observations. En rouge : tendance extraite par la méthode EMD1. (a) Suite de points générée à l’aide du modèle de Feigenbaum. (b) Série générée à l’aide du modèle 2. (c) Série générée à l’aide du modèle 3. (d) Série générée par le modèle 4. (e) Série générée à l’aide du modèle 5. (f) Série générée à l’aide du modèle 6.

résultant est en moyenne plus faible qu’avec les autres méthodes et le rapport signal-à-bruit plus élevé. Les méthodes *EMD2* et *EMD3* possèdent également une meilleure adéquation aux données que les autres méthodes : leur rapport signal-à-bruit est plus élevé et la variance de leur résidus est inférieure ou égale aux autres méthodes. On peut également remarquer que les méthodes basées sur la décomposition modale empirique estiment une tendance qui se retourne un grand nombre de fois. Par exemple, la méthode *RNP1* estime une tendance qui ne se retourne en moyenne qu’une ou deux fois et possède un rapport signal-à-bruit extrêmement faible (0.0007). La variance de la tendance estimée par cette méthode est donc très faible. En revanche, la tendance estimée par la méthode *EMD1* se retourne en moyenne 177 fois : sa variance est donc plus élevée et son rapport signal-à-bruit également. La méthode *EMD1* capte donc une tendance de plus court terme que celle estimée par la méthode *RNP1*.

Le bruit résultant des méthodes *EMD1*, *EMD2* et *EMD3* est estimé comme étant antipersistant en moyenne. Ce résultat contraste avec les autres méthodes pour lesquelles le bruit est estimé comme étant à mémoire courte ($d(\epsilon)$ proche de 0) et indépendamment distribué ($p.values$ élevées). Il est intéressant de remarquer que toutes les méthodes ont un paramètre de mémoire des résidus au carré semblable à celui d’un processus à mémoire longue ($0 < d < \frac{1}{2}$). Ceci est dû à la faible variance des

tendances estimées par rapport à celle du bruit et à la structure de dépendance au sein de la volatilité de la série temporelle établie par le modèle GARCH(1,1).

Pour le modèle 5, on constate également une bonne adéquation par rapport aux données des méthodes basées sur la décomposition modale empirique. On remarque également que les méthodes basées sur les splines de lissage (*SPL1* et *SPL2*) ont une adéquation aux données assez bonne par rapport aux autres méthodes. Cette meilleure adéquation se traduit également par des retournements de la tendance estimée plus nombreux. On remarque que le bruit résultant de l'estimation de la tendance par les méthodes de splines possède un paramètre de mémoire suggérant qu'il n'est pas stationnaire : le paramètre $d(\epsilon) = -0.701$ pour *SPL1* par exemple. Le bruit pourrait donc être modélisé par un modèle du type :

$$(1 - L)^{0.3}\epsilon_t = (1 - L)\delta_t$$

où δ_t est un bruit blanc.

Le modèle 6 représente une série de prix de marché quotidiens tels que leurs rendements suivent le modèle 4. On remarque que la régression par splines de lissage est la méthode exposant la meilleure adéquation aux données, devant les méthodes basées sur la décomposition modale empirique. En revanche, la tendance estimée par les splines de lissage se retourne un grand nombre de fois : 259 fois pour *SPL1* alors qu'une série comprend 1000 données, ce qui fait environ un retournement de tendance toutes les 4 données. Ces méthodes captent donc des tendances à très court terme. De plus, le bruit résultant apparaît être non-stationnaire ($d \notin [-0.5, 0.5]$). Il pourrait notamment se modéliser de la façon suivante :

$$(1 - L)^d\epsilon_t = (1 - L)^p\delta_t$$

où δ_t est un bruit blanc, $d \approx -0.18$ et $p = 2$ pour *SPL1* et $d \approx -0.096$ et $p = 3$ pour *SPL2*. Les méthodes *RNP1*, *RNP2*, *RNP4* ont également des paramètres de mémoire suggérant que le bruit résultant n'est pas stationnaire.

La méthode *EMD1* tire une fois de plus son épingle du jeu puisqu'elle possède une très bonne adéquation aux données et un bruit qui semble, en moyenne, être stationnaire au second ordre. De plus, le test de Ljung-Box ne rejette pas l'hypothèse d'un bruit indépendamment distribué.

Conclusion sur la comparaison des résultats Parmi les méthodes d'estimation de tendance utilisées, une méthode semble se dégager : celle basée sur la décomposition modale empirique et un critère de validation croisée ordinaire, notée *EMD1*. En effet, elle possède la meilleure adéquation aux données pour les modèles 1 à 5 (variance du bruit la plus faible et rapport signal-à-bruit le plus élevé). Pour le modèle 6, la régression par splines de lissage offre la meilleure adéquation aux données mais le bruit résultant est non-stationnaire tandis que la méthode *EMD1* possède une très bonne adéquation aux données et expose un bruit stationnaire au second ordre. L'inconvénient de la méthode *EMD1* réside dans le calcul du critère de validation croisée ordinaire qui prend du temps en pratique. De plus, la tendance estimée par cette méthode fluctue beaucoup par rapport aux autres méthodes. En d'autres termes, cette méthode permet d'obtenir une tendance très proche des données mais peut-être parfois « trop » proche des données. Ceci peut-être problématique pour une utilisation concrète car on souhaite obtenir une tendance qui soit prévisible. Si elle possède des fluctuations erratiques, elle sera difficilement prévisible. Nous proposons, dans le prochain chapitre, un critère différent de la validation croisée ordinaire pour la décomposition modale empirique permettant d'obtenir une tendance qui ne soit pas imprévisible.

3.5 Conclusion

Dans ce chapitre, nous avons présenté différentes méthodes non-paramétriques d'estimation de tendances puis nous les avons comparées. Il est apparu que la régression non-paramétrique par décomposition modale empirique associée à un critère de validation croisée ordinaire semble donner les meilleurs résultats, notamment pour des séries générées par un modèle GARCH(1,1) ou sur des séries de prix dont les rendements suivent un modèle GARCH(1,1), dans le sens où la tendance estimée est « proche » des données et où le bruit résultant paraît stationnaire au second ordre. Néanmoins, la tendance estimée par cette méthode paraît parfois « trop » proche des données, dans le sens où elle fluctue énormément, ce qui semble la rendre presque imprévisible. Or, on souhaite obtenir une tendance qui soit prévisible. On présente, dans le prochain chapitre, une méthode permettant de s'assurer de cette prédictibilité.

	RNP1	RNP2	RNP3	RNP4	RNP5	RNP6	SPL1	SPL2
σ_ϵ	0.2162	0.2162	0.2162	0.2162	0.2162	0.2162	0.2162	0.2162
SNR	$1.79e^{-06}$	$1.14e^{-05}$	$2.58e^{-05}$	$1.12e^{-05}$	$2.39e^{-05}$	$4.41e^{-05}$	$1.30e^{-05}$	$1.59e^{-05}$
NT_1	0	0	2	0	0	2	0	1
$L.B.(\epsilon)$	828.93[0]	828.95[0]	828.99[0]	828.96[0]	828.99[0]	829.05[0]	828.94[0]	828.95[0]
$L.B.(\epsilon^2)$	0.0999[0.752]	0.0917[0.762]	0.0922[0.761]	0.0914[0.762]	0.0940[0.759]	0.0964[0.756]	0.0937[0.759]	0.0948[0.758]
$d(\epsilon)$	-0.6426	-0.6232	-0.5938	-0.6279	-0.6007	-0.5474	-0.6447	-0.6116
$d(\epsilon^2)$	-0.1334	-0.1356	-0.1399	-0.1358	-0.1398	-0.1461	-0.1340	-0.1368

	OND1	OND2	OND3	OND4	EMD1	EMD2	EMD3
σ_ϵ	0.2162	0.2162	0.2162	0.2162	0	0.2065	0.2159
SNR	$4.59e^{-06}$	$1.08e^{-05}$	$1.67e^{-07}$	$4.66e^{-06}$	NA	0.0945	0.00409
NT_1	0	6	708	1	0	498	248
$L.B.(\epsilon)$	828.92[0]	828.93[0]	828.92[0]	828.93[0]	NA	996.78[0]	836.66[0]
$L.B.(\epsilon^2)$	0.0735[0.786]	0.0844[0.771]	0.1006[0.751]	0.0844[0.771]	NA	528.69[0]	0.9137[0.339]
$d(\epsilon)$	-0.5252	-0.6211	-0.6412	-0.6281	NA	-2.2835	-1.6431
$d(\epsilon^2)$	-0.1855	-0.1371	-0.1328	-0.1352	NA	2.4138	-0.1027

TABLEAU 3.3 – Résultats pour le Modèle 1

	RNP1	RNP2	RNP3	RNP4	RNP5	RNP6	SPL1	SPL2
σ_ϵ	1.0245	1.2209	1.2283	1.0686	0.9959	0.9945	0.9639	0.9709
SNR	8.3652	5.6668	5.6762	7.5672	9.073	9.107	9.7135	9.5986
NT_1	30.615	1.4615	1	28	29.8461	30.4615	31.1538	31.3076
$L.B.(\epsilon)$	2.96[0.245]	107.89[0]	112.99[0]	19.23[0.013]	7.84[0.261]	7.60[0.243]	6.11[0.078]	4.29[0.148]
$L.B.(\epsilon^2)$	1.663[0.393]	3.280[0.385]	3.227[0.428]	1.444[0.481]	1.234[0.448]	1.2190[0.454]	1.167[0.438]	1.167[0.437]
$d(\epsilon)$	-0.211	-0.004	-0.043	-0.153	-0.317	-0.301	-0.481	-0.516
$d(\epsilon^2)$	-0.005	0.010	0.016	-0.008	-0.005	-0.005	0.0006	-0.002

	OND1	OND2	OND3	OND4	EMD1	EMD2	EMD3
σ_ϵ	1.1606	1.1606	1.0854	1.1102	0.8958	0.9485	0.9998
SNR	6.38	6.46	7.18	6.96	11.64	10.26	9.09
NT_1	0	34.07	74.84	23.46	163.538	64.384	32.769
$L.B.(\epsilon)$	62.94[0]	66.70[0]	24.95[8.7e ⁻⁰⁶]	36.25[1.9e ⁻⁰⁷]	48.32[3.7e ⁻⁰⁷]	12.56[0.025]	4.09[0.274]
$L.B.(\epsilon^2)$	1.382[0.486]	1.950[0.497]	0.955[0.567]	1.498[0.491]	15.255[1.22e ⁻⁰²]	2.200[0.368]	1.479[0.374]
$d(\epsilon)$	0.055	0.035	-0.037	-0.031	-0.313	-0.203	-0.030
$d(\epsilon^2)$	0.0018	0.0099	-0.0028	-0.0138	0.092	0.0245	0.021

TABLEAU 3.4 – Résultats moyens pour le Modèle 2

	RNP1	RNP2	RNP3	RNP4	RNP5	RNP6	SPL1	SPL2
σ_ϵ	0.991	0.989	0.987	0.999	0.977	0.977	0.981	0.984
SNR	14.57	14.70	14.94	14.21	15.29	15.30	15.12	15.07
NT_1	13.30	23.38	14.38	8.92	18	19	13.53	11.38
$L.B.(\epsilon)$	0.905[0.504]	1.078[0.471]	1.244[0.435]	0.471[0.589]	3.020[0.170]	3.106[0.162]	2.131[0.268]	1.714[0.343]
$L.B.(\epsilon^2)$	1.712[0.362]	1.812[0.345]	2.111[0.320]	1.910[0.344]	2.309[0.304]	2.354[0.294]	2.128[0.315]	2.185[0.306]
$d(\epsilon)$	-0.076	-0.109	-0.208	-0.038	-0.315	-0.308	-0.268	-0.383
$d(\epsilon^2)$	0.011	0.0096	0.005	0.026	0.006	0.006	0.0057	0.005

	OND1	OND2	OND3	OND4	EMD1	EMD2	EMD3
σ_ϵ	1.088	1.044	0.994	1.001	0.905	0.958	0.985
SNR	12.19	13.41	14.24	14.46	18.09	15.99	15.07
NT_1	0.15	21.69	20.84	7.69	165.07	69.76	24.76
$L.B.(\epsilon)$	16.684[0.002]	6.792[0.030]	0.524[0.623]	0.404[0.654]	41.659[0]	8.858[0.019]	1.539[0.365]
$L.B.(\epsilon^2)$	1.581[0.325]	1.295[0.437]	1.790[0.319]	1.541[0.336]	22.725[0.0074]	5.289[0.249]	2.724[0.284]
$d(\epsilon)$	0.089	0.062	-0.018	-0.031	-0.275	-0.183	-0.08
$d(\epsilon^2)$	0.025	0.019	0.014	0.007	0.101	0.034	0.011

TABLEAU 3.5 – Résultats moyens pour le Modèle 3

	RNP1	RNP2	RNP3	RNP4	RNP5	RNP6	SPL1	SPL2
σ_ϵ	0.01406	0.01405	0.01403	0.01401	0.0139	0.0141	0.0140	0.0140
SNR	0.0007	0.0015	0.0036	0.0053	0.0091	0.0099	0.001	0.0034
NT_1	1.61	3.07	2.92	5	7.76	8.46	0.76	2.15
$L.B.(\epsilon)$	1.035[0.491]	1.007[0.490]	1.046[0.479]	0.943[0.510]	0.943[0.514]	0.948[0.513]	1.048[0.489]	1.046[0.483]
$L.B.(\epsilon^2)$	32.536[0.010]	32.20[0.009]	31.537[0.011]	31.578[0.011]	30.777[0.012]	30.531[0.013]	32.364[0.010]	31.529[0.010]
$d(\epsilon)$	-0.003	-0.010	-0.025	-0.048	-0.084	-0.085	-0.003	-0.021
$d(\epsilon^2)$	0.121	0.121	0.122	0.123	0.122	0.122	0.121	0.121

	OND1	OND2	OND3	OND4	EMD1	EMD2	EMD3
σ_ϵ	0.0137	0.0137	0.0136	0.0136	0.0126	0.0134	0.0137
SNR	0.0244	0.0224	0.0196	0.0159	0.295	0.137	0.07
NT_1	4.15	26.76	58.07	42.07	177	92	49.54
$L.B.(\epsilon)$	1.412[0.427]	1.216[0.422]	1.029[0.464]	1.067[0.460]	45.30[0]	10.06[0.031]	2.789[0.287]
$L.B.(\epsilon^2)$	16.578[0.008]	17.281[0.015]	12.628[0.017]	14.458[0.016]	61.588[0]	36.936[0.005]	36.279[0.008]
$d(\epsilon)$	0.0042	0.0013	-0.0005	0.0001	-0.319	-0.211	-0.125
$d(\epsilon^2)$	0.098	0.101	0.091	0.096	0.169	0.127	0.128

TABLEAU 3.6 – Résultats moyens pour le Modèle 4

	RNP1	RNP2	RNP3	RNP4	RNP5	RNP6	SPL1	SPL2
σ_ϵ	0.991	0.987	0.988	0.991	0.983	0.982	0.909	0.947
SNR	0.044	0.051	0.057	0.037	0.059	0.062	0.182	0.135
NT_1	15.07	23.92	14.46	13.31	16.62	17	68.84	37.61
$L.B.(\epsilon)$	20.91[0.06]	18.69[0.05]	19.61[0.035]	19.95[0.031]	16.45[0.051]	16.12[0.046]	5.54[0.146]	9.97[0.091]
$L.B.(e^2)$	0.922[0.523]	1.078[0.502]	1.123[0.479]	0.773[0.654]	0.841[0.571]	0.846[0.572]	1.173[0.507]	1.413[0.535]
$d(\epsilon)$	0.006	-0.004	-0.026	0.032	-0.076	-0.076	-0.701	-0.510
$d(e^2)$	0.004	0.007	0.009	0.003	0.002	0.002	-0.004	-0.0008

	OND1	OND2	OND3	OND4	EMD1	EMD2	EMD3
σ_ϵ	1.014	1.009	1.012	1.010	0.846	0.916	0.957
SNR	0.007	0.012	0.004	0.008	0.529	0.294	0.170
NT_1	0.15	22.77	44.31	25.08	167.92	89.46	46.92
$L.B.(e)$	31.34[0]	28.64[0]	30.25[0]	29.28[0]	24.57[0.0015]	1.49[0.536]	6.63[0.072]
$L.B.(e^2)$	0.889[0.551]	0.928[0.531]	0.799[0.566]	0.755[0.606]	9.586[0.057]	3.036[0.398]	0.663[0.608]
$d(\epsilon)$	0.131	0.127	0.127	0.127	-0.248	-0.117	-0.034
$d(e^2)$	0.009	0.008	0.007	0.004	0.082	0.034	0.013

TABLEAU 3.7 – Résultats moyens pour le Modèle 5

	RNP1	RNP2	RNP3	RNP4	RNP5	RNP6	SPL1	SPL2
σ_ϵ	2.560	2.438	1.890	2.399	1.965	1.956	0.539	0.623
SNR	60.268	65.126	107.438	69.99	104.69	106.16	1528.26	1129.24
NT_1	21.86	30.11	42.57	22.85	32.71	33.28	259.28	223.14
$L.B.(e)$	668.40[0]	645.49[0]	466.49[0]	631.13[0]	489.73[0]	486.29[0]	252.99[0]	228.14[0]
$L.B.(e^2)$	492.91[0]	433.27[0]	272.31[0]	417.26[0]	283.62[0]	280.61[0]	104.90[0]	92.30[0]
$d(\epsilon)$	0.581	0.559	0.152	0.568	0.221	0.234	-2.179	-3.096
$d(e^2)$	0.508	0.501	0.359	0.489	0.380	0.380	0.219	0.209

	OND1	OND2	OND3	OND4	EMD1	EMD2	EMD3
σ_ϵ	1.499	1.244	1.187	1.067	0.833	1.338	2.044
SNR	182.72	273.54	306.17	363.54	573.25	210.11	93.73
NT_1	1.14	110.71	112.71	89.29	206.34	90.57	40.42
$L.B.(e)$	157.04[0]	164.20[0]	158.64[0]	69.83[0]	15.25[0.280]	190.38[0]	524.47[0]
$L.B.(e^2)$	30.67[0]	21.91[2e ⁻⁰⁴]	23.57[0.001]	1.72[0.304]	176.08[0]	150.81[0]	355.01[0]
$d(\epsilon)$	0.191	0.139	0.155	-0.098	-0.214	-0.134	-0.484
$d(e^2)$	0.103	0.096	0.098	0.018	0.268	0.224	0.401

TABLEAU 3.8 – Résultats moyens pour le Modèle 6

Prédiction non-linéaire du risque lié au marché action en utilisant différents indicateurs

Sommaire

4.1	Présentation d'indicateurs du risque lié au marché action	96
4.1.1	Indicateurs liés à la macroéconomie	96
4.1.2	Indicateurs liés aux croyances des investisseurs	100
4.1.3	Étude de la dépendance avec l'évolution du marché action	106
4.2	Extraire une composante chaotique sous-jacente à l'évolution des indicateurs de risque	108
4.2.1	Sur la prédictibilité d'une série temporelle	108
4.2.2	Exposant maximal de Lyapunov et horizon de prédiction	113
4.2.3	Tester l'imprévisibilité d'une série temporelle	114
4.2.4	Estimation non-linéaire de la composante chaotique	121
4.3	Prédiction non-linéaire des composantes chaotiques	123
4.3.1	Prédiction d'une seule composante chaotique	123
4.3.2	Prédiction de plusieurs composantes chaotiques	130
4.4	Correction des effets de bord	133
4.4.1	Formulation du problème	133
4.4.2	Prédiction de la révision de l'estimation à l'aide d'un processus ARMA(p, q) .	135
4.5	Application à des indicateurs liés aux croyances des investisseurs	136
4.6	Conclusion	137

Introduction

Nous avons vu, dans le premier chapitre, qu'il existe un certain nombre d'études ayant mis en lumière la prédictibilité des rendements en excès en utilisant des variables économiques et financières retardées (voir par exemple [Fam(1988)], [Avr(2006)]). Parmi celles-ci nombreuses sont celles qui s'appuient sur une régression linéaire afin d'expliquer les rendements en excès du marché action à l'aide de ces variables. Dans la plupart des cas, les régresseurs significatifs peuvent être interprétés comme des estimées de ce que les participants actifs du marché souhaitent gagner pour contrebalancer

le risque lié à un investissement dans ce marché. Par exemple, il a été mis en évidence à plusieurs reprises que l'écart entre le rendement moyen en dividendes du marché action et le taux d'une obligation émise par un état « bien noté » constitue un facteur de risque significatif du marché action (voir [Kei(1986)][Fam(1988)][Cam(1989)]). Le rendement moyen en dividendes du marché action est alors considéré comme une approximation de la rémunération liée à un investissement dans le marché action tandis que le taux de l'obligation est considéré comme une approximation de la rémunération dans un actif sans risque. La différence entre les deux permet donc de mesurer une sorte de « prime »¹ globale qui peut être associée au risque du marché.

Il existe d'autres manières d'estimer de telles primes de risque. L'intérêt de ces variables réside dans le fait qu'elles caractérisent les croyances des investisseurs dans l'état des conditions financières en temps réel. Elles sont donc d'une importance cruciale pour un petit investisseur souhaitant optimiser la valeur de son portefeuille. Si ces variables s'accordent de manière uniforme pour dire que l'état des conditions financières n'est pas propice à un investissement dans le marché action alors peut-être faut-il désinvestir ce marché et se tourner vers d'autres actifs.

La prédiction de l'évolution de ces variables peut donc être un instrument décisif dans la prise de bonnes décisions d'investissement. Lorsque ces variables passent d'une phase baissière à une phase haussière, ou l'inverse, cela constitue un événement majeur : le risque lié au marché s'accroît (resp. décroît). Dans ce chapitre, on se concentre donc sur la prédiction en temps réel de telles variables et, plus particulièrement, sur celle de leurs tendances de plus ou moins long terme.

Afin d'extraire cette tendance, que nous supposons être une composante sous-jacente à la série d'observations et ayant un comportement chaotique, nous adoptons une approche originale basée sur une décomposition modale empirique et sur un test statistique de l'imprévisibilité d'une série chronologique construit à partir de la théorie du Chaos. En temps réel, cette estimation est soumise à des « effets de bords ». En d'autres termes, l'arrivée de nouvelles observations va modifier l'estimation de la tendance aux bords de l'intervalle de temps considéré pour l'estimation. Nous considérons également ce problème et proposons des solutions pour limiter ces effets.

4.1 Présentation d'indicateurs du risque lié au marché action

Dans ce paragraphe, nous mettons en avant un certain nombre de variables constituant des indicateurs du risque lié à un investissement dans le marché action. On peut les classer en deux catégories : les indicateurs liés à la théorie macroéconomique et les indicateurs liés aux croyances des investisseurs.

4.1.1 Indicateurs liés à la macroéconomie

Dans ce paragraphe, nous allons présenter deux indicateurs : le premier rend compte du contexte économique mondial et le second de l'inflation anticipée par les investisseurs.

Indicateur du contexte économique lié à une zone géographique L'idée derrière ce type d'indicateur est de supposer que les marchés actions sont affectés d'une manière globale par des nouvelles économiques concernant certains pays. Ce type d'indicateur synthétise donc l'information économique d'une zone géographique à partir d'un nombre restreint d'indicateurs nationaux afin de

1. différence de rémunération

4.1. PRÉSENTATION D'INDICATEURS DU RISQUE LIÉ AU MARCHÉ ACTION

savoir si le contexte économique est propice, ou non, à une hausse des marchés actions de cette zone. L'indicateur repose donc sur deux points essentiels :

- La sélection des indicateurs nationaux qui seront fusionnés.
- La méthodologie de fusion de ces indicateurs nationaux.

La Commission Européenne publie mensuellement un tel indicateur concernant la zone Euro sur son site internet² (voir figure 4.1). Les variables économiques prises en compte dans cet indice sont les suivantes :

- Indice de la production industrielle.
- Valeur ajoutée brute du secteur tertiaire privé.
- Indice des prix à la consommation.
- Indice de la production de la construction, du bâtiment et du génie civil.

Ces indicateurs sont calculés et disponibles via Eurostat. Lorsque l'activité économique est en expansion (resp. en contraction), ces variables économiques sont censées augmenter (resp. baisser). La méthodologie de construction de l'indicateur se base alors sur celle développée par Grégoir et Lengart [Gré(2000)] en utilisant des chaînes de Markov cachées qualitatives multivariées. Nous présentons plus amplement cette méthodologie dans le paragraphe du chapitre .

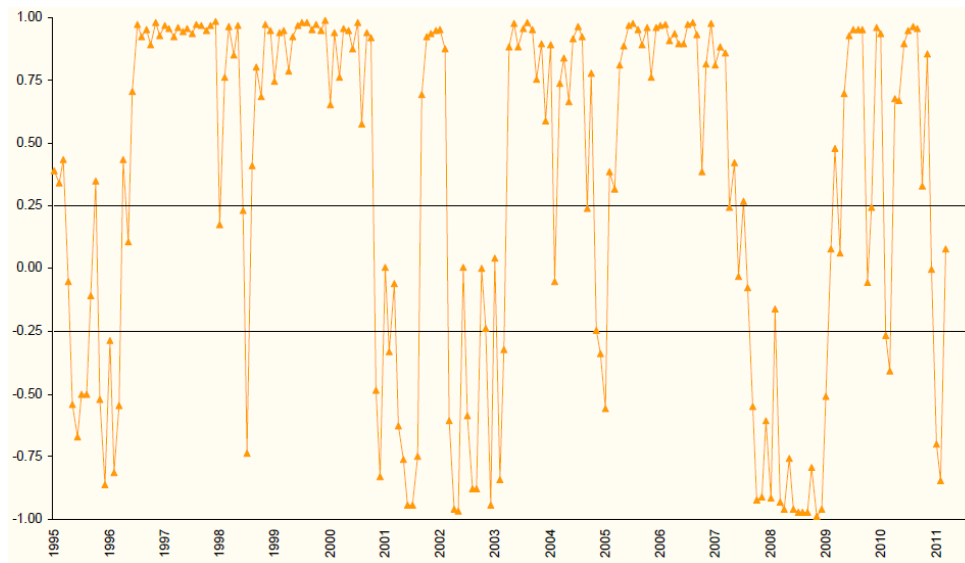


FIGURE 4.1 – Indicateur de point de retournement pour la zone Euro calculé par Eurostat à l'aide de la méthodologie de Grégoir et Lengart [Gré(2000)]. L'interprétation de cet indicateur est la suivante : lorsque l'indicateur se trouve entre -0.25 et 0.25, on se trouve dans une période d'incertitude concernant l'évolution des indicateurs. Lorsqu'il est supérieur à 0.25, on est dans une période de croissance des indicateurs macroéconomiques considérés. Lorsqu'il est inférieur à -0.25, on est dans une période de baisse des indicateurs macroéconomiques considérés.

Indicateur d'inflation anticipée Afin de caractériser la période du cycle économique dans laquelle on se trouve, il peut être intéressant d'associer à l'indicateur du contexte économique d'une

2. http://ec.europa.eu/economy_finance/publications/cycle_indicators/

4.1. PRÉSENTATION D'INDICATEURS DU RISQUE LIÉ AU MARCHÉ ACTION

zone géographique particulière, un indicateur de l'inflation anticipée au niveau de cette zone. Cette approche peut se justifier de la manière suivante.

Lors d'une expansion de l'activité économique combinée à un faible taux d'inflation, l'activité économique croît et les bénéfices des entreprises augmentent. Les investisseurs sont confiants : les prix de marché des actions augmentent et les taux des obligations d'entreprises sont plus faibles. Cet accroissement de l'activité entraîne un besoin de matières premières. Leurs prix augmentent du fait de l'augmentation de la demande. On est en période de « reprise » économique.

Il arrive néanmoins un moment où l'accélération de la croissance atteint les limites des capacités de production. On atteint alors une période de surchauffe de l'économie. Cette limitation des capacités de production entraîne une demande insatisfaite et donc une hausse de l'inflation. Cette dernière va « rogner » les profits des entreprises (par exemple si le prix du pétrole augmente).

Lorsque l'activité économique finit par se contracter tandis que l'inflation s'accroît, les dépenses de consommation des ménages augmentent ce qui entraîne une baisse de la consommation. Cette baisse de la consommation va finir par faire baisser l'inflation. L'activité des entreprises va alors devenir plus faible et leurs résultats diminuer. On entre dans une période de « stagflation ». Le risque de défaut va augmenter et provoquer une hausse des taux des obligations d'entreprises.

Lorsque l'inflation devient suffisamment faible, la consommation va finir par s'arrêter de baisser. Au bout d'un certain temps, on rentre dans une période de « relance » économique : les prix des actions vont repartir à la hausse.

Anticiper l'inflation permet donc de mieux mesurer dans quel état du cycle économique on se trouve et dans quels actifs il peut être intéressant d'investir. La figure 4.2 représente une « horloge d'investissement » trouvant sa justification dans ce qui a été dit précédemment. Ce type d'horloge est à la base de stratégies de gestion utilisées par certaines sociétés de gestion.

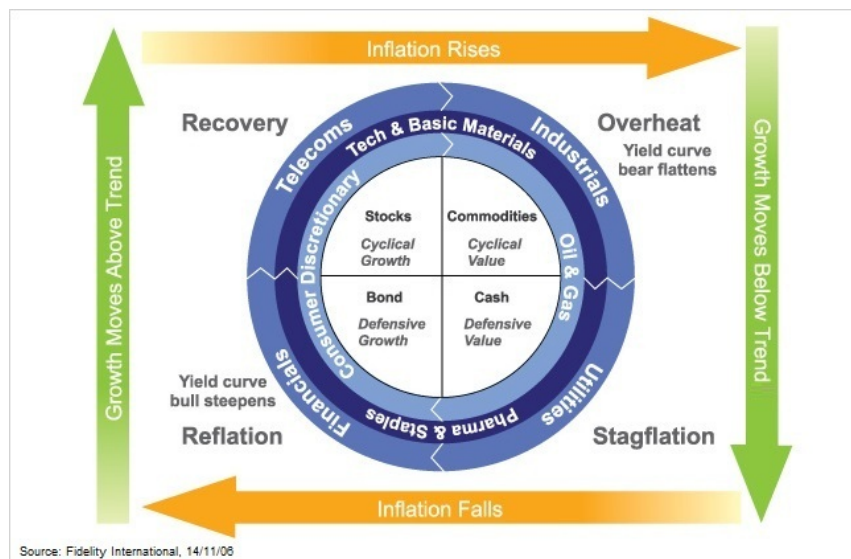


FIGURE 4.2 – Exemple d'une « horloge d'investissement » utilisée par la société de gestion Fidelity International.

Afin de mesurer l'inflation, il existe les indices de prix à la consommation (Consumer Price

4.1. PRÉSENTATION D'INDICATEURS DU RISQUE LIÉ AU MARCHÉ ACTION

Index) fournis par les instituts statistiques nationaux. Il est alors possible de construire un indice global d'inflation concernant une zone géographique en utilisant les indices de prix à la consommation de certains pays et en les fusionnant à la manière de Grégoir et Lengart [Gré(2000)]. Néanmoins, ces indicateurs sont souvent basés sur des constatations des prix à la consommation les mois passés. Ils donnent donc un point de vue retardé sur l'inflation.

Une autre manière intéressante de mesurer l'inflation est d'utiliser le point mort d'inflation. Celui-ci correspond à la différence entre les taux des obligations d'état indexées sur l'inflation et celles non indexées. Dans ce cadre, plus le point mort d'inflation est élevé, plus l'inflation anticipée est forte. Néanmoins, puisqu'il se base sur des taux obligataires, le point mort d'inflation est sensible au risque de crédit. Il faut donc faire attention à son interprétation.

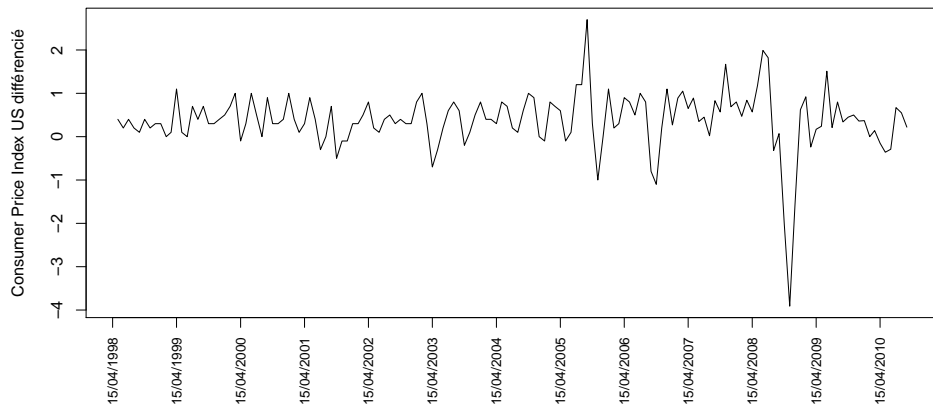


FIGURE 4.3 – Différenciation de l'indice des prix à la consommation aux États-Unis.

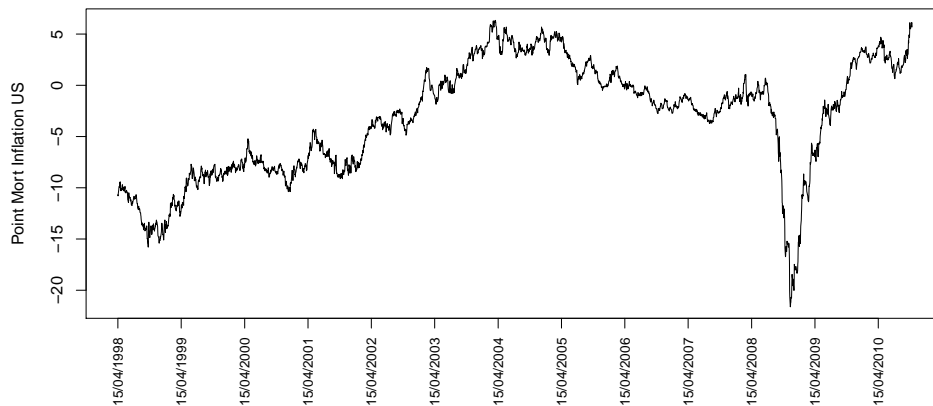


FIGURE 4.4 – Évolution du point mort d'inflation aux États-Unis.

4.1.2 Indicateurs liés aux croyances des investisseurs

Dans ce paragraphe, on présente différents indicateurs reflétant la perception qu'ont les investisseurs, d'une manière globale, de différentes sortes de risques liés au risque du marché action.

Approximation de la prime de risque du marché action La prime de risque liée à un marché est une rémunération censée compenser une perte possible du capital investi sur ce marché. Elle correspond, en quelques sortes, au prix du danger. Approximer la prime de risque d'un marché permet donc de mesurer le risque associé à ce marché. En général, la valeur actualisée de la prime de risque pour un horizon d'investissement fixé de $n\Delta t$ se définit comme (voir par exemple [Meh(1985)], [Meh(2003)]) :

$$Pr_t(t + n\Delta t) = \mathbb{E}[R_{t,t+n\Delta t}|\mathcal{F}_t] - r_{t,t+n\Delta t} \quad (4.1)$$

où $r_{t,t+n\Delta t}$ est le rendement d'un actif à faible risque que l'on considère comme étant une fonction déterministe du temps et $R_{t,t+n\Delta t}$ une variable aléatoire symbolisant le rendement de la valeur de l'investissement.

Afin d'approximer la prime de risque du marché action, certaines études suggèrent d'utiliser la différence entre le rendement moyen en dividendes et une approximation du taux « sans risque » ([Kei(1986)] [Fam(1988)] [Cam(1989)]). En effet, le rendement moyen en dividendes correspond à une approximation de la rémunération liée à un investissement dans le marché action tandis que le taux d'une obligation d'état bien notée correspond à une approximation de la rémunération dans un actif à risque très faible. La différence entre les deux permet donc de mesurer, de manière globale, la « prime » (différence de rémunération) qui peut être associée au risque du marché action.

Une seconde manière d'évaluer cette prime de risque est d'utiliser le « price-earning ratio ». Le « price-earning ratio »(ou *PER*) correspond au rapport entre le cours de bourse d'une action d'une entreprise sur le bénéfice net réalisé par cette dernière divisé par le nombre d'actions émises (bénéfice par action). Il correspond donc à un ratio prix/bénéfice par action. Nous avons vu, dans notre premier chapitre, qu'en supposant un certain nombre d'hypothèses, on peut écrire que la valeur actualisée d'un investissement sur le marché vaut

$$\widehat{V}_t = \lim_{n \rightarrow \infty} \mathbb{E}\left[\frac{F_{t,t+n\Delta t}}{(1+k)^{n \times \Delta t}}|\mathcal{F}_t\right] \quad (4.2)$$

où $k = \mathbb{E}[R_{t,t+\Delta t}|\mathcal{F}_t]$ est un taux d'actualisation constant et $F_{t,t+n\Delta t}$ l'ensemble des flux versés par le titre sur l'intervalle $]t, t + n\Delta t]$. Par souci de simplification, on pose $\Delta t = 1$. Concernant le marché action, les flux versés entre deux instants t et $t + 1$ peuvent être vu comme un dividende versé en $t + 1$ et noté D_{t+1} . Gordon [Gor(1962)] considère que ces dividendes évoluent à un taux de croissance constant g entre deux instants t et $t + 1$. Selon cette approche, on a alors :

$$\mathbb{E}\left[\frac{F_{t,t+n}}{(1+k)^n}|\mathcal{F}_t\right] = \sum_{l=1}^n \frac{D_{t+l}}{(1+k)^l} = \sum_{l=1}^n \frac{(1+g)^l}{(1+k)^l} D_t = \frac{1+g}{1+k} \frac{1 - \left(\frac{1+g}{1+k}\right)^n}{1 - \frac{1+g}{1+k}} D_t \xrightarrow{n \rightarrow +\infty} (1+g) \frac{D_t}{k-g} \quad (4.3)$$

à la condition que $g < k$. De ce fait, $\widehat{V}_t \approx (1+g) \frac{D_t}{k-g}$, ce qui permet d'en déduire k :

$$k \approx (1+g) \frac{D_t}{\widehat{V}_t} + g = \frac{D_{t+1}}{\widehat{V}_t} + g \quad (4.4)$$

D'après ce raisonnement, la prime de risque du marché peut donc être approximée par :

$$Pr_t(t + 1) \approx \frac{D_{t+1}}{\widehat{V}_t} + g - r_{t,t+1}$$

4.1. PRÉSENTATION D'INDICATEURS DU RISQUE LIÉ AU MARCHÉ ACTION

Le bénéfice par action peut être considéré comme étant une approximation grossière de l'espérance des dividendes futurs versés pour une action (voir [Cam(1988)]). Il faut cependant noter que le bénéfice net est souvent celui de l'année passée et que celui-ci peut être réinvesti dans l'entreprise plutôt que reversé sous forme de dividendes. L'utilisation du *PER* peut alors se faire pour capturer la dynamique de cette prime de risque en approximant $\frac{D_{t+1}}{\hat{v}_t}$ par $\frac{1}{PER}$:

$$Pr_t(t+1) \approx \frac{1}{PER} - r_{t,t+1}$$

On notera toutefois que cette approximation reste sujette à discussion du fait de ses nombreuses hypothèses.

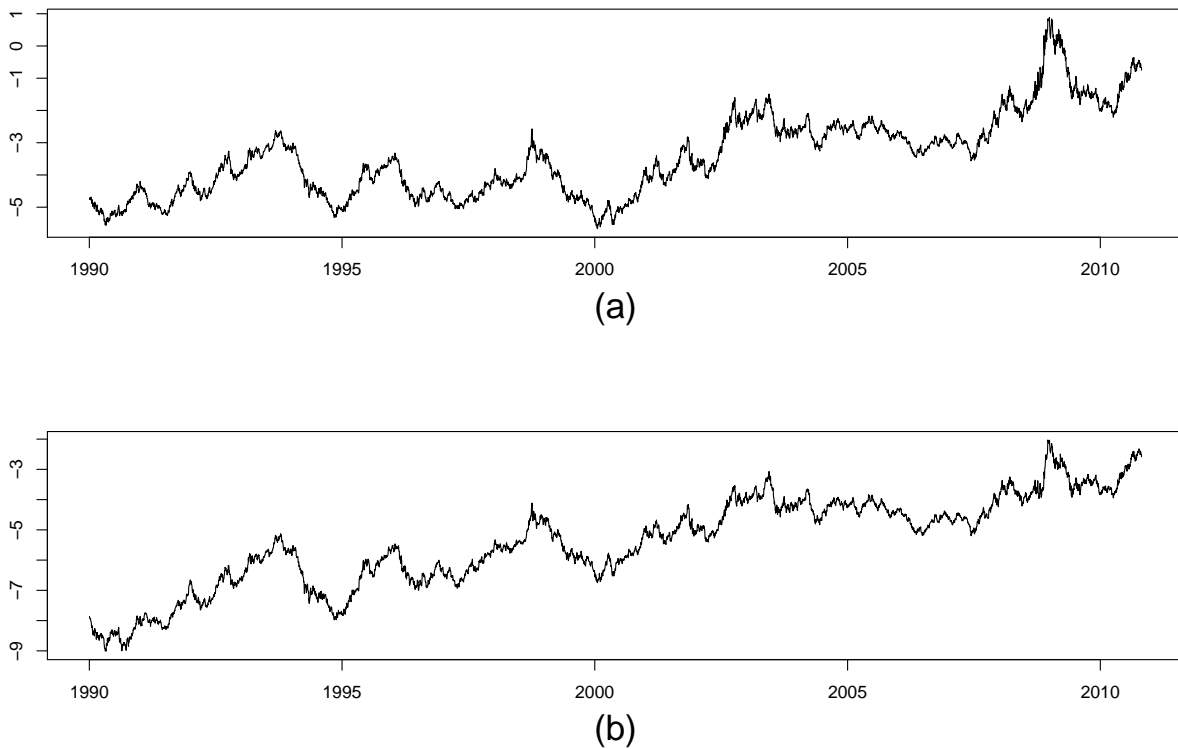


FIGURE 4.5 – (a) Approximation de la prime de risque du marché action US en utilisant la différence entre les rendements moyen en dividendes et les taux d'obligations d'état US à 10 ans. (b) Approximation de la prime de risque du marché action US en utilisant la différence entre l'inverse du PER et les taux d'obligations d'état US à 10 ans.

Approximation de la prime de risque de défaut Certaines études ont également montré la pertinence de l'utilisation de la différence entre le taux moyen des obligations d'entreprises dites « investment grade » ayant une note moyenne et le taux moyen de celles ayant une note plus élevée (voir [Fam(1988)]). Cette variable sert d'approximation globale à la prime de risque de défaut des

4.1. PRÉSENTATION D'INDICATEURS DU RISQUE LIÉ AU MARCHÉ ACTION

entreprises puisque la rémunération liée à une telle obligation sera d'autant plus faible que cette entreprise sera apte à rembourser ses créanciers (et aura donc une note élevée).

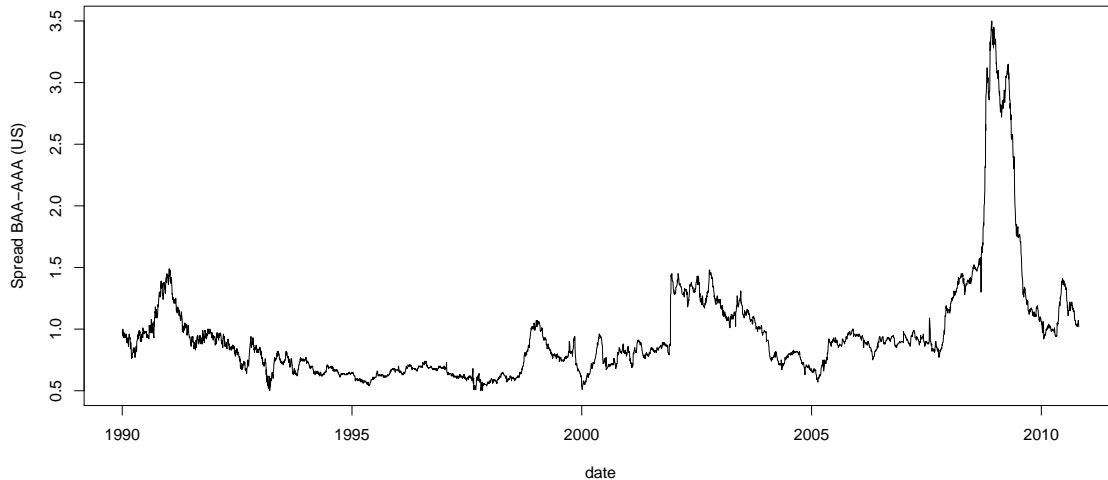


FIGURE 4.6 – Approximation de la prime de risque de défaut des entreprises US en utilisant la différence entre les taux moyens des obligations d’entreprises US notées BAA et celles notées AAA par l’agence de notation Moody’s.

Approximation du risque lié à la volatilité implicite et de l’évolution de la prime de risque liée à la volatilité du marché La volatilité implicite d’une option reflète l’anticipation faite par un investisseur concernant l’évolution future de la volatilité du sous-jacent. L’indice de volatilité implicite VIX est calculé d’après des options ayant pour sous-jacent l’indice *S&P500*. Cet indice mesure la volatilité implicite à 30 jours, espérée par les investisseurs d’une manière globale, d’après une approximation des prix d’option à-la-monnaie. Bollerslev, Tauchen et Zhou suggèrent d’utiliser cet indice afin de mesurer la variance risque-neutre pour une maturité de 30 jours [Bol(2008)]. Ils proposent une approximation de prime de risque liée à la volatilité du marché en évaluant la différence entre cet indice de volatilité implicite et la volatilité réalisée. Si cette différence est positive, les anticipations des investisseurs concernant la volatilité du marché suggèrent une augmentation de la volatilité (volatilité implicite plus élevée que celle qui s’est réalisée). Si on considère la volatilité comme une mesure de risque, ils envisagent donc une augmentation du risque. La principale difficulté de cette approximation réside dans la manière d’estimer la volatilité historique. Bollerslev, Tauchen et Zhou mesurent cette volatilité en utilisant les rendements des cours intraday à haute fréquence sur les 22 jours ouvrés passés [Bol(2008)]. La figure 4.7 représente cette approximation de prime de risque en utilisant une volatilité historique calculée sur les 22 jours ouvrés passés à l’aide de rendements quotidiens.

Indicateur d’aversion au risque des investisseurs Cet indicateur essaye de mesurer l’aversion au risque moyenne de l’ensemble des investisseurs. On suppose donc que l’ensemble des actifs est affecté par un sentiment homogène d’aversion au risque. Lorsque les investisseurs sont peu averses

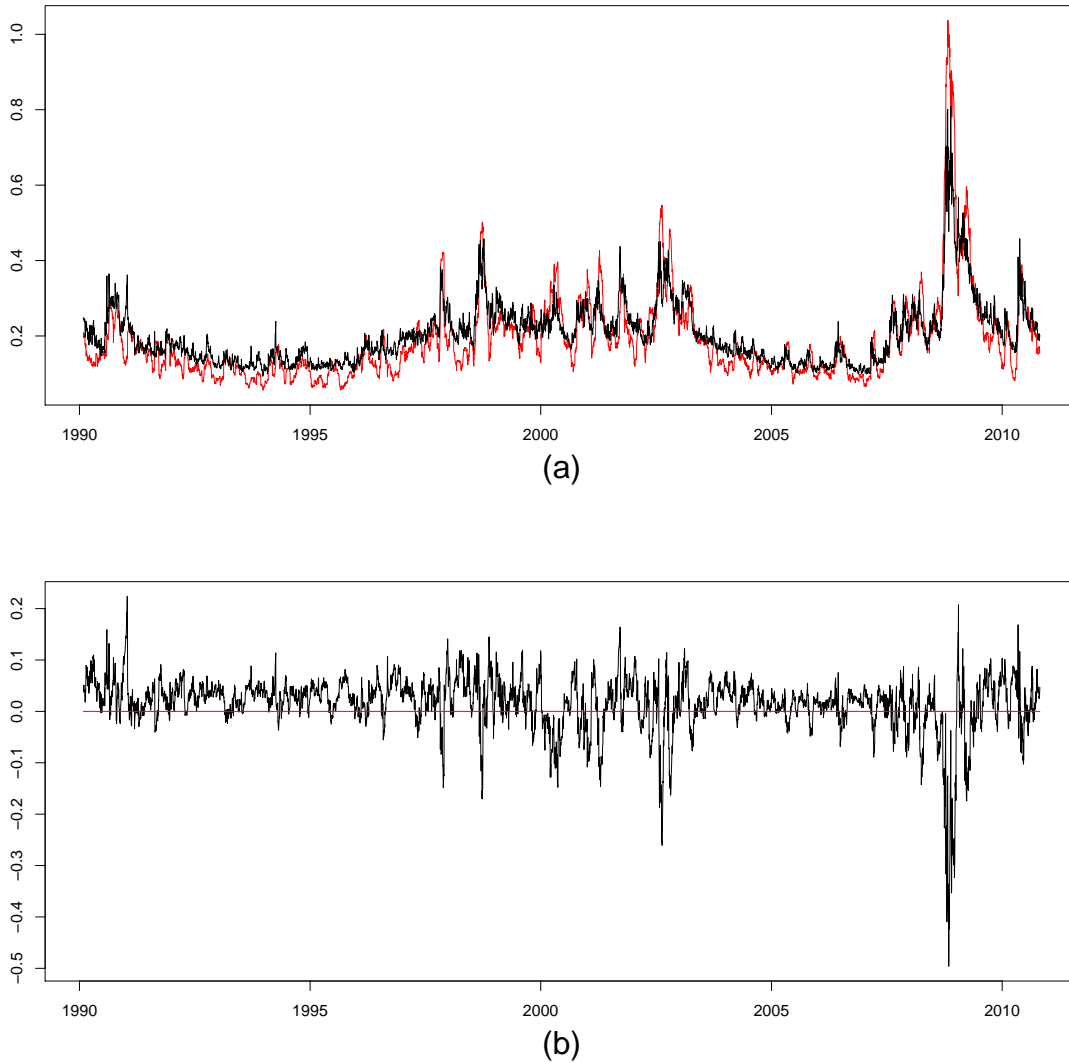


FIGURE 4.7 – (a) Évolution de l'indice VIX (en noir) et celle de la volatilité annualisée réalisée sur les 22 jours précédents mesurée à l'aide de rendements quotidiens du *S&P500* (en rouge). (b) Approximation de l'évolution de la prime de risque liée à la volatilité du marché action US (différence entre volatilité implicite et volatilité réalisée). Lorsque cette différence est positive (resp. négative) cela signifie que les investisseurs envisagent une hausse (resp. baisse) de la volatilité du marché (et donc de la prime de risque).

au risque, ils investissent dans les actifs les plus risqués. Lorsqu'ils sont très averses au risque, ils investissent dans les actifs les moins risqués.

On sélectionne un certain nombre d'actifs ayant des profils rendement/risque différents. L'idée est alors d'effectuer la régression linéaire suivante :

$$\mathbb{E}[R_{t,t+1}^i] = \alpha_{t+1} + \beta_{t+1}\rho(R_{t,t+1}^i) + \epsilon_{t+1}^i \quad (4.5)$$

4.1. PRÉSENTATION D'INDICATEURS DU RISQUE LIÉ AU MARCHÉ ACTION

où $R_{t,t+1}^i$ correspond au rendement du prix de l'actif i entre l'instant t et l'instant $t + 1$ et $\rho(R_{t,t+1}^i)$ est une mesure de risque d'autant plus élevée que le risque est élevé pour ce rendement. Le coefficient β_{t+1} mesure la relation qu'il existe entre performance et risque sur l'ensemble des actifs. S'il est négatif, cela signifie que les actifs les moins risqués offrent la meilleure performance : les investisseurs sont considérés comme averses au risque.

Une manière de faire peut être d'estimer l'espérance par la moyenne des rendements sur une période passée, et la mesure de risque par une estimation de la variance des rendements sur cette même période passée.

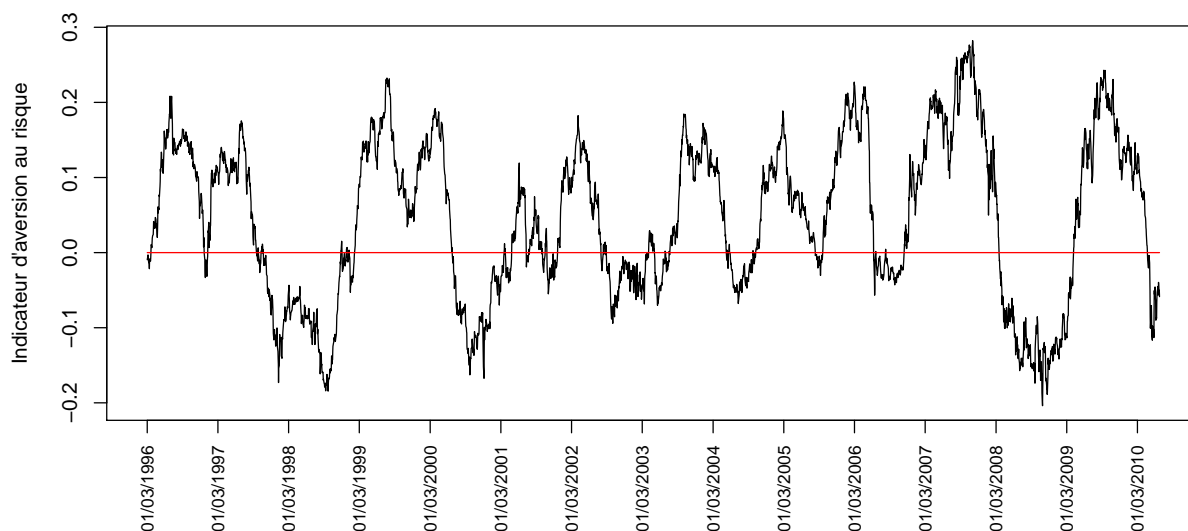


FIGURE 4.8 – Indicateur d'aversion au risque calculé à l'aide des indices actions suivants : S&P 500 Composite, FTSE 100, SBF 120, S&P/TSX Composite Index, SHANGHAI SE Composite, DAX 30 Performance, HANG SENG, India BSE (100) National, ISE National 100, Swiss Market Index, Portugal PSI-20, BEL 20, TOPIX 500, MILAN COMIT 30, Russia RTS Index, IBEX 35, Brazil BOVESPA, Mexico IPC (BOLSA), FTSE Colombia, OMX Stockholm 30 (OMXS30), Korea SE Composite (KOSPI), DJTM Indonesia, AEX INDEX (AEX), Warsaw General Index, Argentina Merval, Prague SE PX, DJTM Singapore, Oslo SE OBX, Taiwan SE Weighted, Morocco SE CFG 25, FTSE W Pakistan, ATX - Austrian Traded Index. Pour chaque indice, l'espérance et la variance des rendements sont calculées sur les 6 mois passés.

Le Put/Call Ratio Si un investisseur anticipe une baisse du prix d'un actif, il achète une option de vente, ayant pour sous-jacent cet actif et lui permettant de couvrir ses positions à terme. Plus il y a d'investisseurs anticipant une baisse des cours, plus ces investisseurs se couvrent : le nombre d'options de vente (ou « put ») échangées devient alors plus important que le nombre d'options d'achat (ou « call ») échangées. En ce sens, le rapport entre le nombre de put échangés et le nombre de call échangés permet de mesurer la croyance des investisseurs concernant l'évolution futur du marché. Si ce ratio est supérieur à 1, les investisseurs envisageraient donc une baisse du cours du sous-jacent.

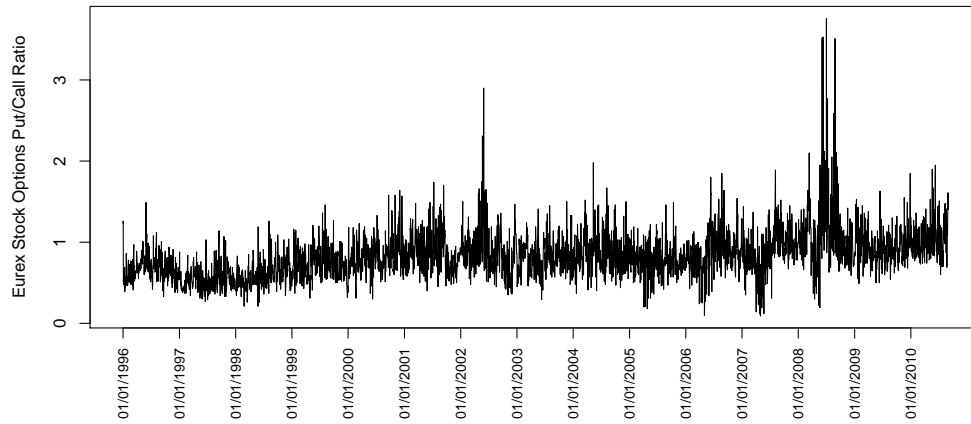


FIGURE 4.9 – Le Put/Call Ratio quotidien calculé sur les options du marché EUREX à maturité 1 mois.

Skew de volatilité 90-110 Le skew de volatilité 90-110 est la différence de volatilité implicite entre une option de vente ayant pour prix d'exercice 90 et une option d'achat ayant pour prix d'exercice 110. Un skew élevé (resp. faible) implique une prédominance des acheteurs de puts (resp. de call) et des vendeurs de calls (resp. de puts). Un skew élevé signifie donc que les investisseurs envisagent une baisse du sous-jacent.

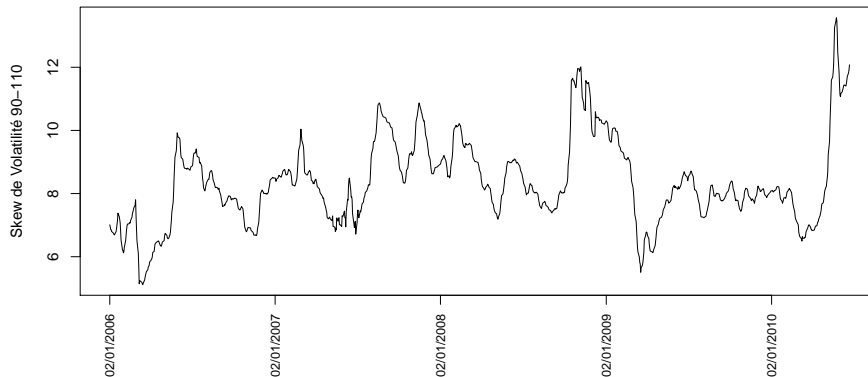


FIGURE 4.10 – Skew de volatilité évalué sur des options de maturité 3 mois ayant pour sous-jacent le DJ Eurostoxx 50.

Indicateur ARMS (Statistique TRIN) L'indicateur ARMS correspond au rapport entre le volume de transactions par titre en déclin et le volume de transactions par titre en augmentation. Si l'indicateur est supérieur à 1, les investisseurs ont tendance à vendre des titres d'une manière plus importante lorsqu'ils sont en baisse qu'ils ne les achètent lorsqu'ils sont en augmentation. Dans ce cadre, ils peuvent être considérés comme pessimistes sur l'évolution futur du marché.

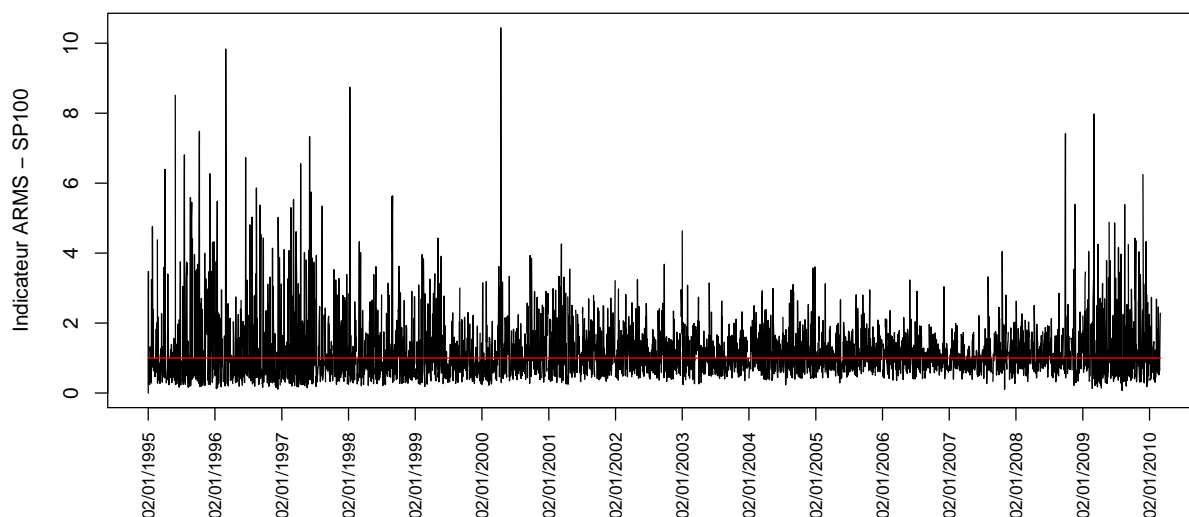


FIGURE 4.11 – Indicateur ARMS évalué sur les titres du S&P 100.

4.1.3 Étude de la dépendance avec l'évolution du marché action

Dans ce paragraphe, on étudie la dépendance entre les indicateurs précédemment présentés et l'évolution du marché action d'une manière globale. Pour cela, on étudie les corrélations de Pearson entre les différents indicateurs et les rendements quotidiens, hebdomadaires et mensuels d'indices de prix agrégés du marché action.

D'une manière générale, la valeur des différents indicateurs et la valeur des rendements quotidiens, hebdomadaires et mensuels à venir est très faible (voir tableau 4.1). Néanmoins, lorsqu'on observe la corrélation entre les indicateurs et ces rendements au carré, les corrélations sont beaucoup plus significatives (voir tableau 4.2).

Nous noterons les indicateurs de la manière suivante :

- PRUS : Différence entre les rendements moyens en dividendes du marché action US et les taux d'obligations d'état US à maturité 10 ans.
- PRDEF : Différence entre les taux d'obligations d'entreprises US notées BAA par Moody's et celles notées AAA.
- PMINFL : Point mort d'inflation US : différence entre le taux des TIPS et le taux des obligations d'état US non indexées sur l'inflation.
- SKEW : Skew de volatilité évalué sur des options de maturité 3 mois ayant pour sous-jacent le DJ Eurostoxx 50.
- VIX : Indice de volatilité implicite VIX.
- PUTC : Put/Call ratio calculé sur les options du marché EUREX à maturité 1 mois.
- ARMS : Indicateur ARMS calculé sur les titres du S&P 100.
- AVERS : Indicateur d'aversion au risque calculé sur une fenêtre de 6 mois d'après le portefeuille précédemment présenté.

Seuls deux indicateurs semblent décorrélés des rendements des indices de prix et de ces rendements

4.1. PRÉSENTATION D'INDICATEURS DU RISQUE LIÉ AU MARCHÉ ACTION

au carré : ce sont les indicateurs ARMS et PUTC. Ces indicateurs ont été calculés quotidiennement et sont donc très volatiles. Lorsqu'on calcule une moyenne mobile de ces indicateurs sur 1 semaine et 1 mois, on obtient des corrélations beaucoup plus significatives avec les indices de prix.

Indicateur	Rdts Quotidiens			Rdts Hebdo			Rdts Mensuels		
	MSCI World	S&P 500	Euro Stoxx 50	MSCI World	S&P 500	Euro Stoxx 50	MSCI World	S&P 500	Euro Stoxx 50
PRUS	0.0142	0.032	-0.0098	0.0129	0.0159	-0.0048	0.0388	0.0279	0.0056
PRDEF	0.0074	0.0132	-0.0068	-0.0247	-0.0320	-0.0253	-0.0228	-0.0419	-0.0368
PMINFL	-0.0009	-0.0072	0.0061	0.0776	0.0885	0.0638	0.1318	0.1521	0.1278
SKEW	-0.0479	-0.0366	-0.0433	-0.1567	-0.1533	-0.1429	-0.1599	-0.165	-0.1608
VIX	0.0100	0.0535	-0.0198	-0.0367	-0.0238	-0.0336	-0.0415	-0.0469	-0.0456
PUTC	-0.0049	0.0343	0.0007	-0.0251	0.0010	-0.0272	-0.0692	-0.0606	-0.0375
ARMS	-0.0024	0.0178	-0.0242	0.0174	0.0185	0.0278	0.0261	0.0272	0.0197
AVERS	0.0244	0.0116	0.0345	0.0969	0.1015	0.0953	0.2034	0.2244	0.1992

TABLEAU 4.1 – Corrélations entre les indicateurs et les rendements de différents indices de prix du marché action.

Indicateur	Rdts Quotidiens			Rdts Hebdo			Rdts Mensuels		
	MSCI World	S&P 500	Euro Stoxx 50	MSCI World	S&P 500	Euro Stoxx 50	MSCI World	S&P 500	Euro Stoxx 50
PRUS	0.2699	0.2827	0.2676	0.2751	0.3010	0.3218	0.3013	0.3323	0.368
PRDEF	0.3318	0.3476	0.2967	0.3192	0.3441	0.3457	0.298	0.3288	0.3545
PMINFL	-0.3768	-0.3981	-0.3235	-0.3552	-0.3783	-0.3522	-0.2814	-0.3139	-0.2905
SKEW	0.1944	0.1992	0.1560	0.1788	0.1730	0.1779	0.0126	0.0328	0.0178
VIX	0.4633	0.4793	0.4142	0.4275	0.4431	0.4245	0.2998	0.3339	0.3239
PUTC	0.0314	0.0373	0.0412	0.031	0.023	0.0486	0.0694	0.0644	0.0605
ARMS	0.0242	0.0138	0.0743	-0.0033	-0.0067	0.0039	0.0226	0.0197	0.0424
AVERS	-0.2708	-0.2797	-0.2323	-0.2487	-0.2538	-0.2563	-0.332	-0.3434	-0.3101

TABLEAU 4.2 – Corrélations entre les indicateurs et les rendements au carré de différents indices de prix du marché action.

		Rdts Quotidiens			Rdts Hebdo			Rdts Mensuels		
		MSCI World	S&P 500	Euro Stoxx 50	MSCI World	S&P 500	Euro Stoxx 50	MSCI World	S&P 500	Euro Stoxx 50
PUTC	MM 1s	-0.0022	0.0085	-0.0011	-0.0212	-0.0136	-0.0159	-0.1108	-0.1077	-0.0632
	MM 1m	-0.0235	-0.0169	-0.0079	-0.0698	-0.0659	-0.0383	-0.1568	-0.1558	-0.0994
ARMS	MM 1s	0.0212	0.0244	0.0309	-0.0120	-0.0097	0.0169	0.0424	0.0368	0.0473
	MM 1m	-0.0257	-0.0296	0.0105	-0.0836	-0.0965	-0.0115	0.2833	0.2832	0.3892

TABLEAU 4.3 – Corrélations entre les moyennes mobiles sur 1 semaine (MM 1s) et 1 mois (MM 1m) des indicateurs PUTC et ARMS et les rendements de différents indices de prix du marché action.

4.2. EXTRAIRE UNE COMPOSANTE CHAOTIQUE SOUS-JACENTE À L'ÉVOLUTION DES INDICATEURS DE RISQUE

		Rdts Quotidiens			Rdts Hebdo			Rdts Mensuels		
		MSCI World	S&P 500	Euro Stoxx 50	MSCI World	S&P 500	Euro Stoxx 50	MSCI World	S&P 500	Euro Stoxx 50
PUTC	MM 1s	0.0389	0.0393	0.0474	0.0422	0.0335	0.0631	0.1299	0.1212	0.1078
	MM 1m	0.0812	0.0784	0.0754	0.0929	0.0821	0.097	0.2639	0.2547	0.2039
ARMS	MM 1s	0.0798	0.0638	0.1556	0.1749	0.196	0.1682	0.1369	0.1672	0.1749
	MM 1m	0.2785	0.2557	0.2616	0.1562	0.1581	0.1967	0.2448	0.2502	0.2924

TABEAU 4.4 – Corrélations entre les moyennes mobiles sur 1 semaine (MM 1s) et 1 mois (MM 1m) des indicateurs PUTC et ARMS et les rendements au carré de différents indices de prix du marché action.

Parmi les indicateurs certains ont des corrélations similaires avec le marché action, c'est le cas des indicateurs PRUS, PRDEF, VIX et ARMS (MM 1m) qui sont positivement corrélés avec les rendements quotidiens, hebdomadaires et mensuels au carré des indices de prix. Les indicateurs AVERS et PMINFL sont inversement corrélés avec ces rendements, ce qui suggère de prendre leurs opposés pour obtenir des indicateurs corrélés significativement avec les rendements quotidiens, hebdomadaires et mensuels au carré.

Les indicateurs SKEW et PUTC (MM 1m) apparaissent plus complexes à interpréter. L'indicateur SKEW semble négativement corrélé avec les rendements hebdomadaires et mensuels des indices de prix et positivement corrélé avec les rendements au carré quotidiens et hebdomadaires. Dans les autres cas, les corrélations sont faibles. L'indicateur PUTC (MM 1m) semble négativement corrélé avec les rendements mensuels des indices de prix et positivement avec les rendements mensuels au carré. Dans les autres cas, ses corrélations avec le marché action sont également faibles.

4.2 Extraire une composante chaotique sous-jacente à l'évolution des indicateurs de risque

Dans ce paragraphe, nous présentons une méthode originale permettant d'extraire une composante chaotique, et donc prévisible à plus ou moins long terme, au sein d'une série temporelle qui apparait être aléatoire.

4.2.1 Sur la prédictibilité d'une série temporelle

Lorsqu'on souhaite étudier l'évolution des indicateurs de risque précédemment présentés, on les différencie une fois. Les séries temporelles issues de cette différenciation paraissent parfois indépendamment distribuées. Si c'est le cas, on ne peut pas prédire exactement la valeur à venir de la série. On peut seulement estimer un intervalle de confiance où sera censée se réaliser la prochaine valeur de la série. Néanmoins, rien ne nous interdit de supposer qu'il existe une composante sous-jacente à la série temporelle qui soit prévisible.

Pour séparer le prévisible de l'imprévisible, il faut se poser la question de la quantification de la prédictibilité d'une série temporelle. Une série temporelle est prévisible lorsqu'il est possible d'anticiper son évolution à partir du présent d'après ce que l'on connaît du passé. Dans ce paragraphe, nous discutons de cette notion de prédictibilité en se basant sur la théorie du Chaos et nous proposons un test statistique pour tester l'hypothèse d'imprévisibilité d'une série temporelle. Nous discutons

4.2. EXTRAIRE UNE COMPOSANTE CHAOTIQUE SOUS-JACENTE À L'ÉVOLUTION DES INDICATEURS DE RISQUE

également de l'implémentation de ce test au travers de la reconstruction de l'espace des phases ainsi que de la recherche de plus proches voisins. Ces aspects peuvent s'avérer problématiques en pratique car ils peuvent être coûteux en temps de calcul.

Théorème de plongement de Takens

Le théorème de plongement de Takens [Tak(1981)] et ses généralisations (voir par exemple [Sau(1991)]) permet de reconstruire un système dynamique non linéaire à partir de l'analyse statistique d'une série temporelle. Notre but n'est pas d'étudier la preuve de ce théorème en détails, le lecteur intéressé pourra alors se référer à l'article de J.P. Huke [Huk(1993)]. Nous présentons seulement de manière explicite le théorème ainsi que ce qu'il implique au niveau statistique.

Supposons qu'il existe un système dynamique inconnu, noté ϕ , générant une série de $T + 1$ points y_t pour $0 \leq t \leq T$:

$$y_{t+\tau} = \phi(y_t)$$

Ces points ne peuvent être directement observés dans la réalité. On suppose alors qu'il existe une fonction f représentant la mesure de ces points. On peut par exemple supposer que y_t représente le niveau de risque du marché action dans l'absolu et que $f(y_t)$ correspond à une mesure de ce risque.

On va supposer que ϕ est un C^k -difféomorphisme et que le système agit dans un espace qui peut se représenter sous la forme d'une variété M de dimension m . En d'autres termes, $\phi : M \rightarrow M$. f est supposée être une fonction suffisamment lisse, c.-à-d. que $f \in C^k(M, \mathbb{R})$ où $C^k(M, \mathbb{R})$ est l'espace des C^k -difféomorphismes allant de M à \mathbb{R} . Il existe alors y_t tel que $y_t \in M$ et

$$f(\phi(y_t)) = x_{t+\tau} \quad \forall 0 \leq t < T$$

où $(x_t)_{t \in [1, \dots, T]}$ correspond à la série de mesures observée. Le théorème de Taken prouve qu'un vecteur $z_t \in \mathbb{R}^d$ tel que :

$$z_t = (f(y_t), f(\phi(y_t)), \dots, f(\phi^{d-1}(y_t))) = (x_t, x_{t+\tau}, \dots, x_{t+(d-1)\tau})$$

ne peut être associé qu'à un unique état du système, noté y_t , si $d > 2m$. Notons $D^k(M)$ l'espace des C^k -difféomorphismes allant de M à M et $C^k(M, \mathbb{R})$ l'espace des C^k -difféomorphismes allant de M à \mathbb{R} . Le théorème de Taken est le suivant :

Théorème 2. *Soit M une m -variété compacte. Il existe un sous-espace U ouvert et dense dans $D^k(M) \times C^k(M, \mathbb{R})$ vérifiant la propriété suivante :*

$\forall (\phi, f) \in U, \forall d > 2m$, la fonction $\Psi_{(\phi, f)} : M \rightarrow \mathbb{R}^d$ donnée par $\Psi_{(\phi, f)}(y) = (f(y), f(\phi(y)), \dots, f(\phi^{d-1}(y)))$ est un morphisme injectif (ou plongement) de M dans \mathbb{R}^d .

Par le théorème de Takens, si deux vecteurs observés z_{t_1} et z_{t_2} sont égaux, on peut donc conclure qu'ils correspondent à un même état du système : $y_{t_1} = y_{t_2}$.

Lorsque l'on souhaite identifier les différents états du système dynamique linéaire ou non-linéaire générant la série d'observations, le théorème de Takens nous dit qu'il suffit de s'intéresser aux vecteurs z_t . L'image de l'ensemble des différents états du système, représenté par la variété M , par la fonction $\Psi_{(\phi, f)}$ est également appelé « espace des phases ». Statistiquement, il faut donc déduire les paramètres τ et d permettant de reconstruire l'ensemble des vecteurs z_t contenus dans la série d'observations.

4.2. EXTRAIRE UNE COMPOSANTE CHAOTIQUE SOUS-JACENTE À L'ÉVOLUTION DES INDICATEURS DE RISQUE

Lorsque l'on se trouve à un instant t et que l'on souhaite prédire un état du système, on sait que $(x_{t-(d-1)\tau}, \dots, x_{t-\tau}, x_t)$ est associé à un état du système $y_{t-(d-1)\tau}$. Si on arrive à extraire un vecteur d'observations passées $(x_{t-(d-1)\tau-\delta}, \dots, x_{t-\tau-\delta}, x_{t-\delta})$ ($\delta > \tau$) égale à $(x_{t-(d-1)\tau}, \dots, x_{t-\tau}, x_t)$ alors on sait, par ce théorème, que $y_{t-(d-1)\tau} = y_{t-(d-1)\tau-\delta}$. En conséquence, $\phi(y_{t-(d-1)\tau}) = \phi(y_{t-(d-1)\tau-\delta}) = y_{t-(d-1)\tau-\delta+\tau}$ et on a déduit l'état du système à venir.

Remarque 4. *Il faut noter que ϕ est un C^k -difféomorphisme allant de M à M . En conséquence, deux états y_{t_1} et y_{t_2} tels que $y_{t_1} \approx y_{t_2}$ auront des images par ϕ également proches : $\phi(y_{t_1}) \approx \phi(y_{t_2})$. Puisque f est également un C^k -difféomorphisme, on aura $f(\phi(y_{t_1})) \approx f(\phi(y_{t_2}))$. C'est la raison pour laquelle deux vecteurs d'observations $z_{t_1} \in \mathbb{R}^d$ et $z_{t_2} \in \mathbb{R}^d$ approximativement égaux peuvent être associés à deux états du système y_{t_1} et y_{t_2} approximativement égaux. Ce théorème justifie donc la recherche de plus proches voisins au sens d'une distance prédéfinie (par exemple euclidienne) lorsque l'on souhaite prédire une série temporelle générée par de tels ϕ et f .*

Reconstruction de l'espace des phases

D'après le théorème de Taken, la dynamique d'une série temporelle chaotique est entièrement caractérisée par l'espace des phases de dimension d défini par les vecteurs retards $Y(t, \tau, d)$:

$$\forall t \in [(d-1)\tau + 1, \dots, T], \quad Y(t, \tau, d) = (y(t), y(t-\tau), \dots, y(t-(d-1)\tau))^{Tr}$$

où $\tau \in \mathbb{N}$ est le temps de retard, $d \in \mathbb{N}$ est une dimension de plongement. En pratique, ce temps de retard et cette dimension de plongement nécessitent d'être estimés de manière appropriée.

Estimation du temps de retard Si τ est inférieur à la valeur qu'il aurait fallu sélectionner, les données du vecteur $Y(t, d, \tau)$ ne seront pas indépendantes. S'il est trop élevé, alors il n'y a pas de corrélation dynamique entre les vecteurs d'état et cela entraîne une perte d'information sur le système original [Jay(1994)]. τ est souvent choisi à l'aide de la fonction d'autocorrélation ou de l'information mutuelle [Fra(1986)] [Moo(1995)]. Dans le cas de l'information mutuelle, τ est estimé par le premier minimum local de l'information mutuelle moyenne.

Soient X_t et $X_{t-\tau}$ deux variables aléatoires. L'information mutuelle moyenne se définit comme :

$$\bar{I}_{X_t, X_{t-\tau}} = \int_{\mathbb{R}} \int_{\mathbb{R}} f_{X_t, X_{t-\tau}}(x_t, x_{t-\tau}) \log\left(\frac{f_{X_t, X_{t-\tau}}(x_t, x_{t-\tau})}{f_{X_t}(x_t)f_{X_{t-\tau}}(x_{t-\tau})}\right) dx_t dx_{t-\tau}$$

où $f_{X_t, X_{t-\tau}}$ correspond à la densité de la loi de $(X_t, X_{t-\tau})$, f_{X_t} correspond à la densité de la loi de X_t et $f_{X_{t-\tau}}$ correspond à la densité de la loi de $X_{t-\tau}$. En pratique, il faut estimer ces lois afin d'en déduire une estimation de l'information mutuelle moyenne. Fraser et Swinney [Fra(1986)] estiment ces densités par des histogrammes tandis que Moon et al. [Moo(1995)] proposent de les estimer à l'aide d'estimations par noyaux gaussiens. Il faut noter que cette estimation est positive et est censée être égale à 0 lorsque les deux variables aléatoires X_t et $X_{t-\tau}$ sont indépendantes. Le premier minimum local de l'information mutuelle moyenne d'une série temporelle est donc le temps de retard le plus proche possible de l'instant présent t pour lequel les données aux temps de retard précédant et succédant à t ont, en moyenne, une dépendance plus forte avec la donnée à l'instant t .

4.2. EXTRAIRE UNE COMPOSANTE CHAOTIQUE SOUS-JACENTE À L'ÉVOLUTION DES INDICATEURS DE RISQUE

Estimation de la dimension de plongement Tant que la dimension de plongement est choisie plus grande qu'une dimension de plongement minimale, la trajectoire des vecteurs reconstruits dans l'espace des phases reflète l'évolution de l'ensemble des états du système chaotique (voir théorème 2). De nombreux algorithmes sont disponibles pour déterminer cette dimension de plongement minimale (voir par exemple [Boc(2002)], [Cao(1997)], [Ken(1992)]).

Une des méthodes les plus populaires est la méthode des plus proches voisins (voir [Fra(1986)], [Kan(2004)]). Pour chaque $Y(t, \tau, d)$, on détermine son plus proche voisin $Y(t_{nn(t)}, \tau, d)$ au sens d'une certaine distance (par exemple la distance euclidienne) pour τ et d fixés. Quand $\|Y(t, \tau, d+1) - Y(t_{nn(t)}, \tau, d+1)\|$ est significativement plus élevé que $\|Y(t, \tau, d) - Y(t_{nn(t)}, \tau, d)\|$, c'est peut-être le résultat de points qui ne sont pas proches dans un espace des phases de dimension $d+1$ mais deviennent des points proches dans l'espace de dimension inférieure d . En conséquence, $Y(t_{nn(t)}, \tau, d)$ est appelé « faux plus proche voisin » de $Y(t, \tau, d)$ si

$$\frac{\|Y(t, \tau, d+1) - Y(t_{nn(t)}, \tau, d+1)\|}{\|Y(t, \tau, d) - Y(t_{nn(t)}, \tau, d)\|} - 1 \geq s$$

où s est un seuil déterminé par la situation pratique. Afin de déterminer d , on utilise donc l'algorithme suivant :

1. $d = 1$
2. On détermine le nombre de faux plus proches voisins. Si ce nombre est faible et ne décroît plus quand d augmente, on stoppe.
3. $d = d + 1$ puis étape 1.

Néanmoins, s reste un paramètre subjectif délicat à sélectionner dans cette approche.

La méthode de Cao permet d'éviter le choix de ce paramètre. Elle se base sur le même principe que celui des faux plus proches voisins mais considère les quantités suivantes :

$$E_1(d) = \frac{E(d+1)}{E(d)} \quad \text{et} \quad E_2(d) = \frac{E^*(d+1)}{E^*(d)}$$

où

$$E(d) = \frac{1}{T - (d-1)\tau} \sum_{t=(d-1)\tau+1}^T \frac{\|Y(t, \tau, d+1) - Y(t_{nn(t)}, \tau, d+1)\|}{\|Y(t, \tau, d) - Y(t_{nn(t)}, \tau, d)\|}$$

et

$$E^*(d) = \frac{1}{T - d\tau} \sum_{t=(d-1)\tau+1}^{T-d\tau} \|y(t + d\tau) - y(t_{nn(t)} + d\tau)\|$$

En principe, $E_1(d)$ converge vers 1 lorsque d augmente si la série est chaotique. On peut alors considérer qu'il existe une valeur de saturation d_0 après laquelle $E_1(d)$ sera presque constant. Cette valeur dépend du comportement prévisible de la série y_t . Dans le cas où la série est chaotique et où $E_1(d)$ converge après une valeur de saturation d_0 , $d_0 + 1$ peut être choisi comme étant la dimension d'enfoncement minimale.

Lorsqu'on travaille sur des échantillons finis d'observations, $E_1(d)$ augmente parfois de manière lente et il est difficile de savoir si une valeur de saturation est atteinte ou si $E_1(d)$ continue à augmenter lentement lorsque d augmente. Il peut ainsi arriver que $E_1(d)$ semble converger alors que la série n'est pas chaotique mais aléatoire. C'est pourquoi la quantité $E_2(d)$ est intéressante. D'après Cao, $E_2(d)$ sera toujours proche de 1 lorsque les données sont imprévisibles mais il existera des valeurs d pour lesquelles $E_2(d)$ sera significativement éloigné de 1 si y_t est prévisible. Avant de déterminer la dimension d'enfoncement minimale, il faut donc s'assurer que la série considérée est chaotique.

Améliorer la vitesse de l'algorithme recherchant les k plus proches voisins

La recherche d'un ou de plusieurs plus proches voisins d'un vecteur $Y_{t,d,\tau} = (y_t, \dots, y_{t-\tau(d-1)})$ est une opération qui peut s'avérer coûteuse en temps de calcul. Ce temps de calcul dépend à la fois de la longueur de ce vecteur et du nombre de vecteurs auquel il peut être comparé, c'est-à-dire de la longueur de la série temporelle. Nous introduisons ici ce sujet et présentons un algorithme efficace pour diminuer ce temps de calcul. Il existe toutefois de nombreuses méthodes de recherche de plus proches voisins et le lecteur intéressé pourra se référer à [Ber(2004)] pour une vue plus exhaustive de ces méthodes.

Approche classique L'approche classique pour rechercher les k plus proches voisins d'un vecteur $Y_{t,d,\tau}$ consiste à le comparer à l'ensemble des vecteurs $\{Y_{k,d,\tau} | k \in]\tau(d-1), t-\tau(d-1)[u]t+\tau(d-1), T]\}$ à l'aide d'une distance prédéfinie, par exemple euclidienne. Il faut alors calculer $T-3\tau(d-1)$ distances. S'il faut rechercher les k plus proches voisins pour chacun des vecteurs $Y_{t,d,\tau}$ pour t appartenant à $]\tau(d-1), T]$, cela fait un total de $(T-\tau(d-1)+1) \times (T-3\tau(d-1))$ distances à calculer. On dit également que ce type de calcul est de l'ordre de $O(T^2)$. Pour une série temporelle de longueur 5000, cela fait donc environ 25 millions de distances à calculer, ce qui correspond à un temps de calcul incompatible avec une utilisation en temps réel. On propose, dans la suite, une méthode diminuant l'ordre de ce calcul à $O(T \log(T))$ ce qui, pour notre exemple, réduirait le calcul à environ 42600 distances à calculer et rend cette application plus compatible avec une application en temps réel. Il faut toutefois noter que le temps de calcul d'une seule distance est fonction de la dimension du vecteur $X_{t,d,\tau}$, c'est-à-dire de d . Si celui-ci devient trop grand, le temps de calcul peut également devenir ingérable en temps réel (on parle également du problème de la « malédiction de la dimension »). Il est toutefois peu probable que l'application à des vecteurs de phase se confronte à cette problématique car d est généralement estimé, au maximum, aux alentours de 10 sur les séries financières que nous avons étudiées.

Méthode des arbres k-d. (ou k-d. trees) La méthode des arbres k-d. consiste à partitionner l'espace de dimension d , auquel appartiennent les vecteurs $Y_{t,d,\tau}$, suivant une seule dimension à chaque étape (voir [Ben(1975)] [Fri(1977)]). Chaque branche de l'arbre est créée en sélectionnant un pivot partitionnant la dimension $j \in [1, \dots, d]$ en deux.

Notons y_j^{max} et y_j^{min} le maximum et le minimum atteints par l'ensemble des vecteurs $Y_{t,d,\tau}$ au niveau de la $j^{\text{ème}}$ dimension (on a donc $y_j^{max} = \max_{j \in [1, \dots, d]} Y_{t,d,\tau}(j)$ et $y_j^{min} = \min_{j \in [1, \dots, d]} Y_{t,d,\tau}(j)$). La dimension j est sélectionnée si $\frac{y_j^{max} - y_j^{min}}{2}$ est plus élevé pour cette dimension que pour les autres dimensions. Le vecteur pivot est alors le vecteur $Y_{t,d,\tau}$ dont la $j^{\text{ème}}$ coordonnée est la plus proche de $\frac{y_j^{max} + y_j^{min}}{2}$. Les vecteurs dont la $j^{\text{ème}}$ coordonnée est inférieure à celle du vecteur pivot sont alors regroupés dans un ensemble et celle dont la coordonnée est supérieure dans un second ensemble. On réitère cette dichotomie dans chaque sous-ensemble de vecteurs jusqu'à ce qu'il ne reste plus qu'un seul vecteur dans chaque ensemble.

La recherche du plus proche voisin consiste alors à descendre le long de cet arbre. On calcule à chaque étape la distance entre le vecteur pivot et le vecteur dont on recherche le plus proche voisin (que l'on appellera vecteur objet). Si la boule de centre l'objet et de rayon cette distance n'intersecte pas un ensemble alors le plus proche voisin de l'objet n'appartient pas à cet ensemble. A chaque niveau de l'arbre, il suffit donc de déterminer dans quel ensemble le plus proche voisin appartient. On s'intéresse ensuite aux sous-ensembles de l'ensemble sélectionné. Cette méthode permet alors de

4.2. EXTRAIRE UNE COMPOSANTE CHAOTIQUE SOUS-JACENTE À L'ÉVOLUTION DES INDICATEURS DE RISQUE

limiter les calculs à $O(T \log(T))$ [Fri(1977)].

4.2.2 Exposant maximal de Lyapunov et horizon de prédiction

Considérons le système dynamique suivant :

$$x_{t+1} = f(x_t), \quad \forall t \in [t_0, t_n] \text{ et pour } x_t \in \mathbb{R}$$

La séparation entre deux dynamiques engendrées par deux points initiaux $x_0^{(1)}$ et $x_0^{(2)}$ après N périodes de temps est alors la suivante :

$$|x_N^{(1)} - x_N^{(2)}| = \left| \overbrace{f \circ f \circ f \dots \circ f}^{N \text{ fois}}(x_0^{(1)}) - \overbrace{f \circ f \circ f \dots \circ f}^{N \text{ fois}}(x_0^{(2)}) \right| \quad (4.6)$$

$$= |f^{(N)}(x_0^{(1)}) - f^{(N)}(x_0^{(2)})| \approx |x_0^{(1)} - x_0^{(2)}| \left| \frac{d}{dx} f^{(N)}(x_0^{(1)}) \right| \quad (4.7)$$

Or, $\frac{d}{dx} f^{(N)}(x_0^{(1)}) = \frac{d}{dx} f(x_{N-1}^{(1)}) \times \frac{d}{dx} f(x_{N-2}^{(1)}) \times \dots \times \frac{d}{dx} f(x_0^{(1)})$.

Si on suppose que $\left| \frac{d}{dx} f(x_{N-1}^{(1)}) \right| \approx \left| \frac{d}{dx} f(x_{N-2}^{(1)}) \right| \approx \dots \approx \left| \frac{d}{dx} f(x_0^{(1)}) \right| \leq c$ où $c > 0$, on peut écrire que

$$|x_N^{(1)} - x_N^{(2)}| \leq |x_0^{(1)} - x_0^{(2)}| c^N = |x_0^{(1)} - x_0^{(2)}| e^{\lambda N} \quad (4.8)$$

Le λ minimum vérifiant (4.8) est appelé exposant maximal de Lyapunov. Il caractérise la divergence d'évolution du système dynamique lorsque les états initiaux du système sont différents. Si λ est positif, cela signifie que l'erreur entre deux états initiaux du système ira en augmentant les instants suivants. En revanche, s'il est négatif, cette erreur ira en diminuant.

En pratique, Rosenstein, Collins et DeLuca [Ros(1993)] ont proposé une méthode pour calculer cet exposant. L'idée consiste à effectuer la régression suivante :

$$\forall t \in [t_0, t_n], \quad \log(|y_{t+\tau \times k} - y_{t_{nn(t)}+\tau \times k}|) = \log(|y_t - y_{t_{nn(t)}}|) + \lambda_t \times k$$

où $t_{nn(t)}$ correspond à l'instant du plus proche voisin du vecteur des phases $Y(t, d, \tau)$. Matriciellement, on peut réécrire ceci comme :

$$\mathbf{P} = \mathbf{Q}\beta$$

où

$$\mathbf{P} = \begin{pmatrix} \log(|y_{t_0+\tau \times 1} - y_{t_{nn(t_0)}+\tau \times 1}|) & \log(|y_{t_0+\tau \times 2} - y_{t_{nn(t_0)}+\tau \times 2}|) & \dots & \log(|y_{t_0+\tau \times k_{max}} - y_{t_{nn(t_0)}+\tau \times k_{max}}|) \\ \vdots & \vdots & & \vdots \\ \log(|y_{t_N+\tau \times 1} - y_{t_{nn(t_N)}+\tau \times 1}|) & \log(|y_{t_N+\tau \times 2} - y_{t_{nn(t_N)}+\tau \times 2}|) & \dots & \log(|y_{t_N+\tau \times k_{max}} - y_{t_{nn(t_N)}+\tau \times k_{max}}|) \end{pmatrix}$$

$$\mathbf{Q} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ \vdots & \vdots \\ 1 & k_{max} \end{pmatrix} \quad \text{et} \quad \beta = \begin{pmatrix} \log(|y_{t_0} - y_{t_{nn(t_0)}}|) & \log(|y_{t_1} - y_{t_{nn(t_1)}}|) & \dots & \log(|y_{t_N} - y_{t_{nn(t_N)}}|) \\ \lambda_{t_0} & \lambda_{t_1} & \dots & \lambda_{t_N} \end{pmatrix}$$

avec $k_{max} \geq (d - 1)$.

4.2. EXTRAIRE UNE COMPOSANTE CHAOTIQUE SOUS-JACENTE À L'ÉVOLUTION DES INDICATEURS DE RISQUE

Afin d'estimer un exposant λ moyen à partir de petits ensembles de données, Roseinstein, Collins et Deluca [Ros(1993)] se ramènent à une simple régression linéaire en moyennant les écarts constatés :

$$\bar{\mathbf{P}} = \left(\log(\overline{|y_{t+\tau \times 1} - y_{t_{nn(t)+\tau \times 1}|}) \quad \log(\overline{|y_{t+\tau \times 2} - y_{t_{nn(t)+\tau \times 2}|}) \quad \dots \quad \log(\overline{|y_{t+\tau \times k_{max}} - y_{t_{nn(t)+\tau \times k_{max}}|}) \right)$$

$$\bar{\beta} = \left(\frac{\log(\overline{|y_t - y_{t_{nn(t)}}|})}{\lambda} \right)$$

$$\text{où } \overline{|y_t - y_{t_{nn(t)+\tau \times k}|}} = \frac{1}{t_n - t_0} \sum_{t=t_0}^{t_n} |y_t - y_{t_{nn(t)+\tau \times k}|.$$

Dans ce cadre, on effectue la régression linéaire $\bar{\mathbf{P}} = \mathbf{Q}\bar{\beta} + \epsilon$. Ce qui donne l'estimateur suivant pour $\bar{\beta}$:

$$\hat{\bar{\beta}} = (\mathbf{Q}^{Tr} \mathbf{Q})^{-1} \mathbf{Q}^{Tr} \bar{\mathbf{P}}$$

Horizon de prédiction Si l'exposant maximal de Lyapunov estimé λ_{max} est positif, on peut essayer d'estimer un horizon maximum de prédiction N^* . Comme nous l'avons précisé plus haut, $|x_0^{(1)} - x_0^{(2)}| e^{\lambda_{max} N}$ représente une estimation d'un majorant de l'écart $|x_N^{(1)} - x_N^{(2)}|$. On peut supposer que

$$\max |x_N^{(1)} - x_N^{(2)}| \approx |x_0^{(1)} - x_0^{(2)}| e^{\lambda_{max} N}$$

D'où

$$N \approx \frac{1}{\lambda_{max}} \log\left(\frac{\max |x_N^{(1)} - x_N^{(2)}|}{|x_0^{(1)} - x_0^{(2)}|}\right)$$

Pour estimer l'horizon maximal de prédiction, une méthode consiste à fixer une valeur pour $\max |x_N^{(1)} - x_N^{(2)}|$ proportionnelle à $|x_0^{(1)} - x_0^{(2)}|$. On a alors :

$$N^*(\alpha) = \frac{1}{\lambda_{max}} \log(\alpha)$$

où $\max |x_N^{(1)} - x_N^{(2)}| = \alpha |x_0^{(1)} - x_0^{(2)}|$. Si on considère deux séries temporelles ayant des exposants de Lyapunov respectifs λ_1 et λ_2 tels que $\lambda_1 < \lambda_2$, alors, pour n'importe quel alpha fixé, l'horizon maximal de prédiction de la première série sera plus élevé que celui de la seconde série. Néanmoins, α est une valeur assez subjective et délicate à choisir. DeDomenico et Ghorbani [DeD(2010)] suggèrent d'utiliser $\alpha = 1.96$. Cette valeur repose sur l'idée que $|x_N^{(1)} - x_N^{(2)}|$ est réparti suivant une loi $\mathcal{N}(0, |x_0^{(1)} - x_0^{(2)}|)$.

4.2.3 Tester l'imprévisibilité d'une série temporelle

Dans ce paragraphe, on s'intéresse à tester l'hypothèse selon laquelle une série temporelle est imprévisible. Dans un premier temps, on présente le test de blancheur de Ljung-Box, couramment utilisé pour tester une hypothèse de décorrélation. On propose ensuite un nouveau test de cette imprévisibilité basé sur la théorie du Chaos.

Test de Portmanteau de Ljung-Box

Ce test consiste à tester la corrélation linéaire entre la valeur de la série temporelle à l'instant t et ses valeurs précédentes. Si cette corrélation est nulle, on peut considérer que l'observation à l'instant t ne dépend pas linéairement de celles observées aux instants passés considérés. En conséquence, on pourrait en déduire abusivement que la série temporelle n'est pas prévisible puisque les observations passées ne permettent pas d'en déduire les observations futures. Néanmoins, ce test ne mesure qu'une dépendance linéaire : la série peut donc être prévisible alors que ce test estime que la corrélation entre données passées et futures est significativement nulle. On va cependant utiliser ce test pour comparer nos résultats : nous allons ainsi montrer que la prédictibilité n'est pas forcément synonyme de décorrélation.

L'hypothèse nulle de ce test est la suivante :

$$H_0 : \ll \rho_y(1) = \dots = \rho_y(k) = 0 \gg$$

La statistique de Ljung-Box [Lju(1978)] proposée pour ce test est la suivante :

$$S_{LB} = T(T+2) \sum_{h=1}^k \frac{\hat{\rho}_y^2(h)}{T-h}$$

où

- $\hat{\rho}_y(h) = \frac{T}{T-h} \frac{\sum_{i=1}^{T-h} y_i y_{i+h}^2}{\sum_{i=1}^T y_i^2}$ est une estimation de la corrélation entre l'observation en t et celle en $t+h$.
- k correspond au nombre d'auto-corrélations testées.

S_{LB} converge en loi, lorsque $T \rightarrow +\infty$, vers une loi du χ^2 à k degrés de liberté. L'hypothèse nulle peut donc être rejetée à un niveau α si :

$$S_{LB} > q_{\chi_k^2}(1-\alpha)$$

Ce test possède deux inconvénients : il nécessite de fixer k et il ne mesure qu'une indépendance linéaire au sein des observations. Dans le prochain paragraphe, on propose un nouveau test statistique de l'imprévisibilité d'une série temporelle évitant ces deux inconvénients.

Un test de prévisibilité basé sur les vecteurs de phase

Soit $(y_t)_{t \in [1, \dots, T]}$ la série temporelle observée. Supposons qu'il existe un système dynamique ϕ générant des états x_t ($x_{t+\tau} = \phi(x_t)$) inobservés mais mesurés dans \mathbb{R} grâce à une fonction notée g (on a $y_t = g(x_t)$). Si le système dynamique et la fonction de mesure sont suffisamment réguliers (C^k -difféomorphismes) et si deux vecteurs de l'espace des phases z_{t_1} et z_{t_2} sont égaux (z_t étant égal à $(y_{t-(d-1)\tau} \ y_{t-(d-2)\tau} \ \dots \ y_t)^{Tr}$), on peut conclure, par le théorème de Takens, qu'ils correspondent à un même état du système : $x_{t_1-(d-1)\tau} = x_{t_2-(d-1)\tau}$. En conséquence, on aura

$$\overbrace{\phi \circ \phi \circ \dots \circ \phi}^{n \text{ fois}}(x_{t_1-(d-1)\tau}) = \overbrace{\phi \circ \phi \circ \dots \circ \phi}^{n \text{ fois}}(x_{t_2-(d-1)\tau}) = x_{t_1-(d-1)\tau+n\tau} = x_{t_2-(d-1)\tau+n\tau}$$

Les états à venir seront donc similaires pour les deux vecteurs. Par conséquent, les vecteurs de phases à venir seront également similaires du fait de la bijectivité de g . Si on trouve un vecteur des phases

4.2. EXTRAIRE UNE COMPOSANTE CHAOTIQUE SOUS-JACENTE À L'ÉVOLUTION DES INDICATEURS DE RISQUE

passé $z_{t_{nn(t)}} (t_{nn(t)} < t)$ égale au vecteur des phases z_t , le vecteur des phases à venir $z_{t+\Delta\tau}$ sera le même que $z_{t_{nn(t)}+\Delta\tau}$ pour $\Delta > 0$. La série temporelle sera donc totalement prévisible.

L'idée du test est donc la suivante : pour chaque instant t , on recherche le plus proche voisin de z_t , noté $z_{t_{nn(t)}}$, puis on observe les valeurs $y_{t+\Delta}$ et $y_{t_{nn(t)}+\Delta}$. Si ϕ est un système suffisamment régulier (c.-à-d. déterministe) et si z_t et $z_{t_{nn(t)}}$ sont suffisamment proches, la série devrait être prévisible : $y_{t+\Delta}$ et $y_{t_{nn(t)}+\Delta}$ devraient être suffisamment proches.

Si une série temporelle n'est prévisible que jusqu'à un certain horizon de temps, $y_{t+\Delta}$ et $y_{t_{nn(t)}+\Delta}$ devraient être proches lorsque Δ est faible. Ensuite, il devrait exister un $\Delta_0 > 0$ tel que $\forall \Delta > \Delta_0$, on ne peut savoir si $y_{t+\Delta}$ et $y_{t_{nn(t)}+\Delta}$ sont proches ou pas. Si la série est imprévisible, $y_{t+\Delta}$ et $y_{t_{nn(t)}+\Delta}$ ne devraient pas être spécifiquement proches lorsque Δ est faible ou lorsqu'il augmente, donc pour tout $\Delta > 0$. Plus particulièrement, dans le cas d'une série imprévisible, on devrait avoir

$$\frac{1}{T-\Delta-d\tau} \sum_{t=d\tau}^{T-(\Delta+1)} |y_{t+(\Delta+1)\tau} - y_{t_{nn(t)}+(\Delta+1)\tau}| \approx \frac{1}{T-\Delta-d\tau} \sum_{t=d\tau}^{T-(\Delta+1)} |y_{t+\Delta\tau} - y_{t_{nn(t)}+\Delta\tau}|$$

pour tout $\Delta > 0$, alors que dans le cas où le système est déterministe, il devrait exister au moins un $\Delta > 0$ pour lequel

$$\frac{1}{T-\Delta-d\tau} \sum_{t=d\tau}^{T-(\Delta+1)} |y_{t+(\Delta+1)\tau} - y_{t_{nn(t)}+(\Delta+1)\tau}| \neq \frac{1}{T-\Delta-d\tau} \sum_{t=d\tau}^{T-(\Delta+1)} |y_{t+\Delta\tau} - y_{t_{nn(t)}+\Delta\tau}|$$

On va donc utiliser la quantité suivante pour distinguer une série prévisible d'une série imprévisible :

$$E_2(\Delta) = \frac{\overline{|y_{t+(\Delta+1)\tau} - y_{t_{nn(t)}+(\Delta+1)\tau}|}}{\overline{|y_{t+\Delta\tau} - y_{t_{nn(t)}+\Delta\tau}|}}$$

où $\overline{|y_{t+\Delta\tau} - y_{t_{nn(t)}+\Delta\tau}|} = \frac{1}{T-\Delta-d\tau} \sum_{t=d\tau}^{T-(\Delta+1)} |y_{t+\Delta\tau} - y_{t_{nn(t)}+\Delta\tau}|$. Puisque l'enjeu est de savoir si $E_2(\Delta)$ est proche de 1 pour tout Δ , on va s'intéresser à la statistique suivante :

$$S(d_{max}) = \frac{1}{d_{max}} \sum_{\Delta=1}^{d_{max}} |E_2(\Delta) - 1|$$

Étude de la répartition de la statistique S lorsque la série est imprévisible Afin de construire un test statistique permettant de tester l'hypothèse d'imprévisibilité d'une série temporelle, on va étudier la répartition de S lorsque la série est imprévisible. On peut considérer que les $E_2(\Delta)$ sont i.i.d. si la série est imprévisible. On a alors, par le théorème central-limite :

$$S(d_{max}) \xrightarrow{d_{max} \rightarrow \infty} \mathcal{N}(\mathbb{E}[|E_2(d_{max}) - 1|], \frac{1}{d_{max}} \mathbb{V}[|E_2(d_{max}) - 1|])$$

Il suffit alors d'estimer $\mathbb{E}[|E_2(d_{max}) - 1|]$ et $\mathbb{V}[|E_2(d_{max}) - 1|]$ pour en déduire une estimation de la loi de probabilité de la statistique sous l'hypothèse d'imprévisibilité. On peut notamment utiliser les estimateurs sans biais suivant :

•

$$\mathbb{E}[|E_2(\widehat{d_{max}}) - 1|] = S\left(\frac{T}{\tau} - d - 1\right)$$

4.2. EXTRAIRE UNE COMPOSANTE CHAOTIQUE SOUS-JACENTE À L'ÉVOLUTION DES INDICATEURS DE RISQUE

•

$$\mathbb{V}[|E_2(\widehat{d_{max}}) - 1|] = \frac{1}{(\frac{T}{\tau} - d - 1)(\frac{T}{\tau} - d - 2)} \sum_{\Delta=1}^{\frac{T}{\tau} - d - 1} (|\frac{|y_{t+(\Delta+1)\tau} - y_{t_{nn(t)}+(\Delta+1)\tau}|}{|y_{t+(\Delta)\tau} - y_{t_{nn(t)}+(\Delta)\tau}|} - 1| - S(\frac{T}{\tau} - d - 1))^2$$

Afin de mesurer ces quantités sur des échantillons finis, on va générer des séries indépendamment distribuées à l'aide de différentes lois de probabilité puis on va estimer $S(d_{max})$ à partir de ces séries pour différentes valeurs de d_{max} . On note \widehat{S}_n la $n^{\text{ème}}$ estimation de S . Chaque estimation de S peut être supposée indépendante des autres estimations et on peut donc supposer, par le théorème central-limite, que la loi de \widehat{S}_n converge vers une loi normale lorsque n tend vers l'infini.

On génère 1000 séries indépendamment distribuées de longueur 1000 à l'aide d'une loi uniforme $U([1, 100])$, d'une loi gaussienne $\mathcal{N}(0, 1)$ et d'une loi de Student $St(5)$, puis on étudie la répartition de \widehat{S}_n dans ces trois cas. La figure 4.12 représente la densité, estimée par des noyaux gaussiens, des estimations dans les trois cas et pour des valeurs de d_{max} différentes : 10, 20 et 30. On constate que la densité estimée de \widehat{S}_n est quasiment la même quelque soit d_{max} et la loi ayant servie à générer les données. D'une manière générale, la loi empirique estimée peut être approximée par une loi gaussienne de moyenne 0.02624 et d'écart-type 0.005. Le tableau 4.5 représente la statistique et la p.valeur du test de Kolmogorov-Smirnov entre cette loi gaussienne et la loi empirique des \widehat{S}_n .

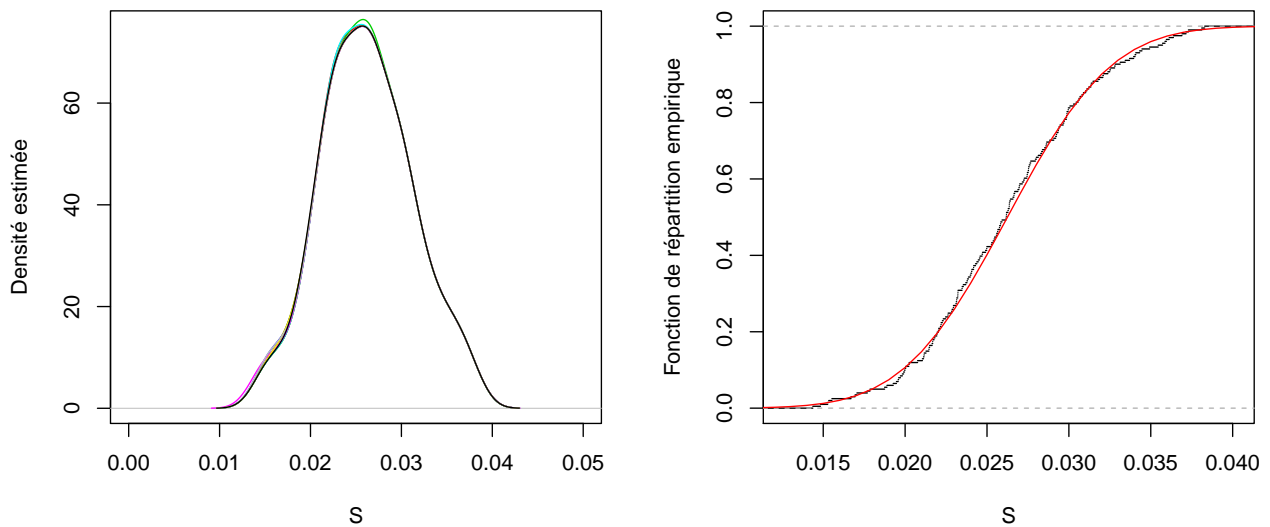


FIGURE 4.12 – Densités estimées par des noyaux gaussiens (à gauche) de la statistique S à partir de séries indépendamment distribuées à partir des lois $U([1, 100])$, $\mathcal{N}(0, 1)$ et $St(5)$ et pour $d_{max} \in [10, 20, 30]$. On constate que la répartition de \widehat{S}_n est quasiment la même pour toutes les lois et pour tous les d_{max} choisis. La fonction de répartition empirique de \widehat{S}_n (à droite) est très proche de celle d'une loi normale de moyenne 0.02624 et d'écart-type 0.005 (en rouge).

4.2. EXTRAIRE UNE COMPOSANTE CHAOTIQUE SOUS-JACENTE À L'ÉVOLUTION DES INDICATEURS DE RISQUE

Loi	d_{max}	D	p.value
$U([1, 100])$	10	0.03486229	0.9674991
$\mathcal{N}(0, 1)$	10	0.03486229	0.9674991
$St(5)$	10	0.03424513	0.9724648
$U([1, 100])$	20	0.03486229	0.9674991
$\mathcal{N}(0, 1)$	20	0.03486229	0.9674991
$St(5)$	20	0.03486229	0.9674991
$U([1, 100])$	30	0.03486229	0.9674991
$\mathcal{N}(0, 1)$	30	0.03486229	0.9674991
$St(5)$	30	0.03486229	0.9674991

TABLEAU 4.5 – D représente la statistique du test de Kolmogorov-Smirnov. La p.value du test de Kolmogorov-Smirnov indique que l'hypothèse selon laquelle la répartition de \widehat{S}_n est semblable à celle d'une loi $\mathcal{N}(0.02624, 0.005)$ est largement acceptée ($p.value \approx 0.97$) dans tous les cas.

Afin d'étudier le comportement de la statistique sous l'hypothèse de prévisibilité entre 1 et d_{max} , on peut supposer l'existence d'une quantité $\gamma > 0$ telle que

$$\forall \Delta \in [1, \dots, d_{max}], \quad \left| \frac{|y_{t+(\Delta+1)\tau} - y_{t_{nn(t)+(\Delta+1)\tau}|}{|y_{t+(\Delta)\tau} - y_{t_{nn(t)+(\Delta)\tau}|} - 1 \right| \geq \gamma > 0$$

Par conséquent, $S(d_{max}) \geq \frac{1}{d_{max}} \gamma d_{max} = \gamma > 0$. La valeur γ dépendra du comportement plus ou moins prévisible de la série temporelle. Si la série est proche de l'hypothèse d'imprévisibilité, γ sera très proche de 0. Si la série est prévisible jusqu'à un certain horizon, cette quantité sera significativement non nulle. Sous l'hypothèse de prévisibilité, on supposera donc, d'une manière générale, que γ est supérieure à un quantile de la loi de probabilité suivie par la statistique sous l'hypothèse d'imprévisibilité à un niveau $1 - \alpha$.

Test de l'hypothèse d'imprévisibilité On va donc construire un test statistique pour tester

$$H_0 : \ll y_t \text{ est une série imprévisible} \gg \text{ contre } H_1 : \ll y_t \text{ est une série prévisible} \gg \quad (4.9)$$

Sous H_0 , $S(10)$ est supposée suivre une loi $\mathcal{N}(0.02624, 0.005)$. H_0 peut donc être rejetée à un niveau α si $S(10) > q_{\mathcal{N}(0.02624, 0.005)}(1 - \alpha)$. La probabilité d'observer $S(10)$ sous H_0 vaudra $1 - \Phi_{\mathcal{N}(0.02624, 0.005)}(S(10))$.

Exemple sur des séries générées Afin d'étudier les résultats de ce test, on génère différentes séries considérées plus ou moins prévisibles. Sur chaque série générée, on va étudier la statistique et la p.value de notre test comparées à celles du test de Ljung-Box.

Les séries générées sont les suivantes :

- Le modèle 1 correspond à une suite chaotique de Feigenbaum :

$$y_t = 3.57y_{t-1}(1 - y_{t-1})$$

avec $y_0 = 0.7$.

4.2. EXTRAIRE UNE COMPOSANTE CHAOTIQUE SOUS-JACENTE À L'ÉVOLUTION DES INDICATEURS DE RISQUE

- Le modèle 2 est un bruit blanc gaussien de moyenne nulle et de variance 1.
- Le modèle 3 un processus ARMA(2,1) de la forme :

$$y_t = 0.45y_{t-1} + 0.2y_{t-2} + u_t + 0.3u_{t-1}$$

avec $y_0 = 1$, $y_1 = 0.7$ et u_t un bruit blanc gaussien de moyenne nulle et de variance unitaire.

- Le modèle 4 est un processus ARMA(2,1) de la forme :

$$y_t = 0.8y_{t-1} + 0.15y_{t-2} + u_t + 0.3u_{t-1}$$

avec $y_0 = 1$, $y_1 = 0.7$ et u_t est un bruit blanc gaussien de moyenne nulle et de variance unitaire.

Le modèle 1 est déterministe et donc la suite d'observations générées par ce modèle est entièrement prévisible. Le modèle 2 génère des observations indépendamment, ce qui les rend imprévisibles. Le modèle 4 possède une partie autorégressive qui rend dépendant y_t de ses réalisations passées. De plus, les coefficients reliant y_t à y_{t-1} et y_{t-2} sont significatifs et entraîne une forte dépendance au sein des données. En conséquence, on peut s'attendre à ce que ce modèle soit « presque prévisible ». Les coefficients du modèle 3 sont moins significatifs que ceux du modèle 4, ce qui réduit la dépendance au sein des données. On peut donc s'attendre à ce que le modèle 3 soit également « presque prévisible » mais peut être un peu moins prévisible que le modèle 4.

On génère 10 séries de longueur 1000 pour les modèles aléatoires 2 à 4 et une série de longueur 1000 pour le modèle 1. Le tableau présente les résultats du test de Ljung-Box pour l'hypothèse $H_0 : \ll \rho(1) = 0 \gg$ et ceux de notre test. Pour la série générée à l'aide du modèle 1, les deux tests rejettent clairement leurs hypothèses nulles. Les résultats concernant les 10 séries générées à l'aide du modèle 2 montrent que notre test semble plus fiable pour détecter le fait que la série est imprévisible. Par exemple, le test de Ljung-Box pour la série n°6 générée par le modèle 2 retourne une p.value de 0.05206 alors que notre test détecte clairement l'imprévisibilité avec une p.value de 0.9687226. Pour le modèle 3, le test de Ljung-Box rejette clairement l'hypothèse qu'il existe une corrélation nulle entre la série à l'instant t et sa valeur en $t + 1$ tandis que notre test ne rejette pas clairement l'hypothèse d'imprévisibilité pour toutes les séries générées. Par exemple, la p.value de la série n°6 générée par ce modèle est de 0.57 pour notre test contre 0 pour celui de Ljung-Box. Ceci peut s'expliquer par le fait qu'une série peut avoir une autocorrélation d'ordre 1 significativement non nulle mais suffisamment faible pour que la série ne soit pas suffisamment prévisible. Enfin, pour les séries générées par le modèle 4, les deux tests rejettent clairement leurs hypothèses nulles, ce qui correspond à nos attentes.

4.2. EXTRAIRE UNE COMPOSANTE CHAOTIQUE SOUS-JACENTE À L'ÉVOLUTION DES INDICATEURS DE RISQUE

	D	p.value (KS)	τ	$S(10)$	p.value (Notre test)
Modèle 1	831.3955	0	2	0.1831444	0
Modèle 2 - 1	0.1488	0.6997	2	0.01510256	0.9870426
Modèle 2 - 2	0.2275	0.6334	1	0.01777755	0.9547237
Modèle 2 - 3	0.0332	0.8554	3	0.02070474	0.8658643
Modèle 2 - 4	0.002	0.9641	1	0.02088008	0.8581366
Modèle 2 - 5	0.9114	0.3398	2	0.02693322	0.4448656
Modèle 2 - 6	3.7739	0.05206	1	0.01692828	0.9687226
Modèle 2 - 7	3.091	0.07873	2	0.02818527	0.3486183
Modèle 2 - 8	0.6868	0.4072	1	0.02108596	0.8486844
Modèle 2 - 9	0.5487	0.4589	1	0.01754708	0.9589461
Modèle 2 - 10	5e-04	0.9827	2	0.02875231	0.3076714
Modèle 3 - 1	562.1152	0	8	0.04510748	8e-05
Modèle 3 - 2	546.1707	0	5	0.0385378	0.006955388
Modèle 3 - 3	540.8324	0	9	0.05069207	5e-07
Modèle 3 - 4	520.6916	0	4	0.0391259	0.004980502
Modèle 3 - 5	518.5832	0	9	0.049505	1.635706e-06
Modèle 3 - 6	530.6779	0	4	0.02525544	0.5780521
Modèle 3 - 7	543.9824	0	6	0.04911516	2.380739e-06
Modèle 3 - 8	448.2261	0	6	0.03707567	0.01511235
Modèle 3 - 9	518.2484	0	5	0.03907735	0.005122186
Modèle 3 - 10	538.3926	0	5	0.02932376	0.2686994
Modèle 4 - 1	937.2192	0	16	0.1902716	0
Modèle 4 - 2	929.7966	0	10	0.1521744	0
Modèle 4 - 3	932.1397	0	19	0.198466	0
Modèle 4 - 4	919.3836	0	14	0.1677501	0
Modèle 4 - 5	924.4112	0	16	0.1838989	0
Modèle 4 - 6	934.5846	0	19	0.2133339	0
Modèle 4 - 7	902.9872	0	14	0.1714908	0
Modèle 4 - 8	920.6976	0	23	0.2253676	0
Modèle 4 - 9	921.4119	0	15	0.175603	0
Modèle 4 - 10	921.5528	0	15	0.180064	0

TABLEAU 4.6 – Résultats du test de Ljung-Box pour l'hypothèse $H_0 : \ll \rho(1) = 0 \gg$ et ceux de notre test pour les différentes séries générées (τ est une estimation du temps de retard).

4.2.4 Estimation non-linéaire de la composante chaotique

Nous avons montré, dans notre deuxième chapitre, que la régression non-paramétrique par décomposition modale empirique présentait des résultats intéressants. Néanmoins, la technique de seuillage que nous avons utilisée pour estimer la tendance était un critère de validation croisée ordinaire. Cette technique est extrêmement coûteuse en temps de calcul et, si elle permet d'obtenir une tendance proche des données, cette dernière semble parfois « trop » proche des données. Afin d'estimer la composante prévisible de la série temporelle, nous allons ici effectuer un test de prédictibilité sur chaque nouvelle fonction modale intrinsèque estimée : d'une part, ce critère est plus rapide en temps de calcul et, d'autre part, la tendance estimée est assurée de ne pas être imprévisible.

Soient y_t la série d'observations et $f(t)$ une estimation de la tendance de cette série. On rappelle que, dans le cadre de la décomposition modale empirique,

$$y_t - \sum_{i=1}^{\lambda} f_i(t) \approx f(t) \tag{4.10}$$

où $f_i(t)$ correspond à la $i^{\text{ème}}$ fonction modale intrinsèque issue de la décomposition de y_t .

L'idée est donc de soustraire à y_t , l'ensemble des fonctions modales intrinsèques considérées imprévisibles. En effet, la première fonction modale intrinsèque estimée a la bande de fréquence la plus élevée des fonctions modales intrinsèques, cette bande de fréquence diminuant au fur et à mesure que de nouvelles fonctions modales intrinsèques sont estimées. La p.value du test de prédictibilité devrait donc être élevée pour les premières fonctions modales intrinsèques puis diminuer au fur et à mesure. Les composantes imprévisibles de la série temporelle devraient donc correspondre aux premières fonctions modales intrinsèques estimées.

Exemple sur les rendements du S&P 500 On teste cette méthode sur les rendements quotidiens su S&P 500 entre le 01/01/1989 et le 27/10/2010.

Série	p.value	$S(10)$	$\hat{\tau}$
y	0.9974	0.0122	1
$y - f_1$	0.9688	0.0169	2
$y - f_1 - f_2$	6.33e-11	0.0584	4
$y - f_1 - f_2 - f_3$	0	0.0821	8
$y - f_1 - f_2 - f_3 - f_4$	0	0.097	14
$y - f_1 - f_2 - f_3 - f_4 - f_5$	0	0.107	22
$y - f_1 - f_2 - f_3 - f_4 - f_5 - f_6$	0	0.121	43

TABLEAU 4.7 – La p.value et la statistique $S(10)$ de notre test. y représente la série d'observation et f_i la fonction modale intrinsèque numéro i . $\hat{\tau}$ correspond à un temps de retard estimé.

On obtient une composante prévisible en soustrayant les deux premières fonctions modales intrinsèques estimées à la série d'observation (p.value = 6.33e-11). La composante prévisible résultante possède un temps de retard estimé de 4 (voir tableau 4.7) et une dimension de plongement minimale estimée de 9 (voir figure 4.13). Elle est présentée dans la figure 4.14.

4.2. EXTRAIRE UNE COMPOSANTE CHAOTIQUE SOUS-JACENTE À L'ÉVOLUTION DES INDICATEURS DE RISQUE

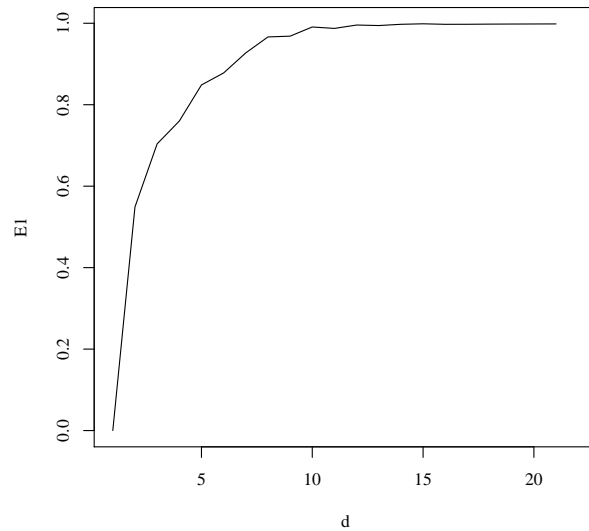


FIGURE 4.13 – Évolution de la quantité $E1$ proposée par Cao [Cao(1997)] en fonction de la dimension de plongement. On considère que pour $d \geq 9$, cette quantité est suffisamment proche de 1. La dimension de plongement minimale estimée est donc de 9.

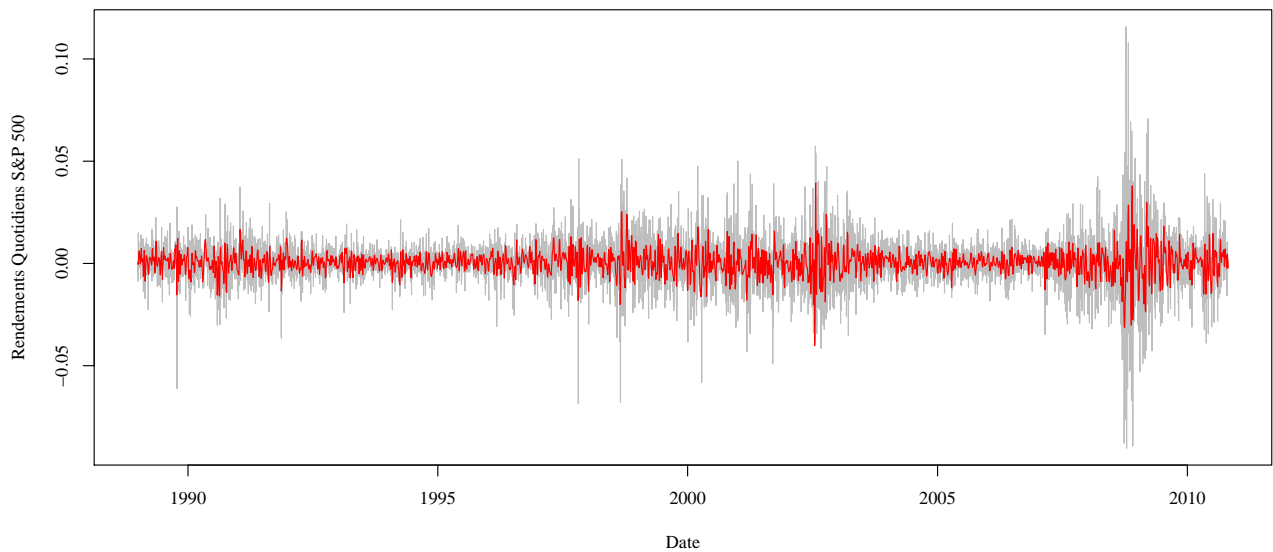


FIGURE 4.14 – Les rendements de l'indice de prix S&P 500 (en gris) et sa composante prévisible estimée (en rouge).

Exemple sur les rendements de l’approximation de prime de risque liée au marché action

On teste cette méthode sur les variations de l’approximation de prime de risque liée au marché action US entre le 01/01/1990 et le 26/10/10. Cette approximation correspond à la différence entre les rendements moyens en dividendes du marché action US et le taux des obligations d’état à maturité 10 ans aux États-Unis.

Série	p.value	$S(10)$	$\hat{\tau}$	\hat{d}_{min}	$\hat{\lambda}_L$
y	0.7769	0.0224	1	10	$7.86e^{-04}$ [$1.22e^{-04}$]
$y - f_1$	0.6829	0.0239	3	9	$5.06e^{-04}$ [$8.31e^{-05}$]
$y - f_1 - f_2$	2.83e-13	0.0623	5	8	$1.74e^{-04}$ [$3.08e^{-05}$]
$y - f_1 - f_2 - f_3$	0	0.0846	10	7	$4.16e^{-05}$ [$8.2e^{-06}$]
$y - f_1 - f_2 - f_3 - f_4$	0	0.1034	19	6	$1.38e^{-05}$ [$1.78e^{-06}$]
$y - f_1 - f_2 - f_3 - f_4 - f_5$	0	0.1114	29	5	$2.74e^{-06}$ [$5.29e^{-07}$]
$y - f_1 - f_2 - f_3 - f_4 - f_5 - f_6$	0	0.1254	53	4	$1.98e^{-06}$ [$5.99e^{-07}$]

TABLEAU 4.8 – y représente la série d’observation et f_i la fonction modale intrinsèque numéro i . \hat{d}_{min} représente la dimension de plongement minimale estimée. $\hat{\lambda}_L$ représente une estimation de l’exposant maximal de Lyapunov (voir paragraphe 4.2.2).

On obtient une composante prévisible en soustrayant les deux premières fonctions modales intrinsèques estimées à la série d’observation (p.value = 2.83e-13). La composante prévisible résultante possède un temps de retard estimé de 5 et une dimension de plongement minimale estimée de 9 (voir figure 4.15). La figure 4.16 représente l’évolution des variations de la prime de risque du marché action aux États-Unis et la composante prévisible estimée.

4.3 Prédiction non-linéaire des composantes chaotiques

Dans ce paragraphe, on s’intéresse à la prédiction non-linéaire des composantes prévisibles des indicateurs de risque. Pour cela, on utilise une méthode basée sur le comportement chaotique de ces composantes. La régression non-paramétrique polynomiale locale offre des solutions intéressantes pour cette prédiction.

4.3.1 Prédiction d’une seule composante chaotique

Soit \hat{m} la composante prévisible estimée. On souhaite prédire $\hat{m}(t + \omega)$ avec $\omega \in \mathbb{N}^*$ en sachant l’information disponible en t . On doit donc estimer \mathbf{F} telle que :

$$\hat{m}(t + \omega) = \mathbf{F} \begin{pmatrix} \hat{m}(t) \\ \vdots \\ \hat{m}(t - (d - 1)\tau) \end{pmatrix} + \eta_t \tag{4.11}$$

où $\mathbf{F} : \mathbb{R}^d \rightarrow \mathbb{R}$ est une fonction linéaire ou non-linéaire associant le dernier vecteur retard de l’espace des phases disponible à une donnée future. η est un bruit tel que $\mathbb{E}[\eta] = \mathbf{0}$. La fonction \mathbf{F} est estimée localement dans un voisinage du vecteur retard au sein de l’espace des phases (farmer et al. 1988). En conséquence, une régression non-paramétrique polynomiale locale semble être une méthode appropriée pour réaliser cette prédiction. Cette méthode débute par les deux étapes suivantes :

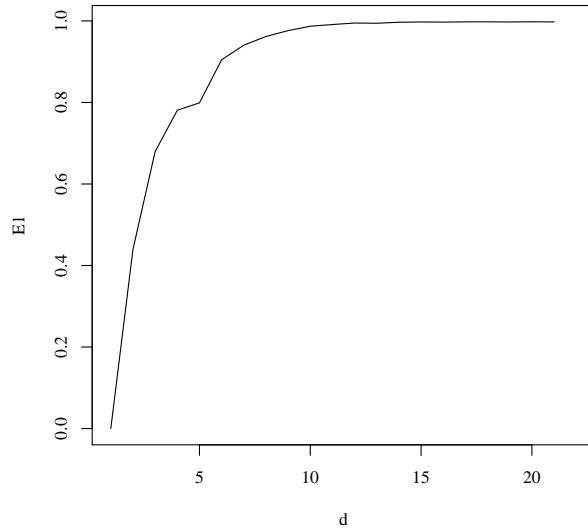


FIGURE 4.15 – La quantité $E1$ proposée par Cao [Cao(1997)] pour estimer la dimension de plongement de la composante prévisible. On considère que pour $d \geq 9$, cette quantité est suffisamment proche de 1. La dimension de plongement minimale estimée est donc de 9.

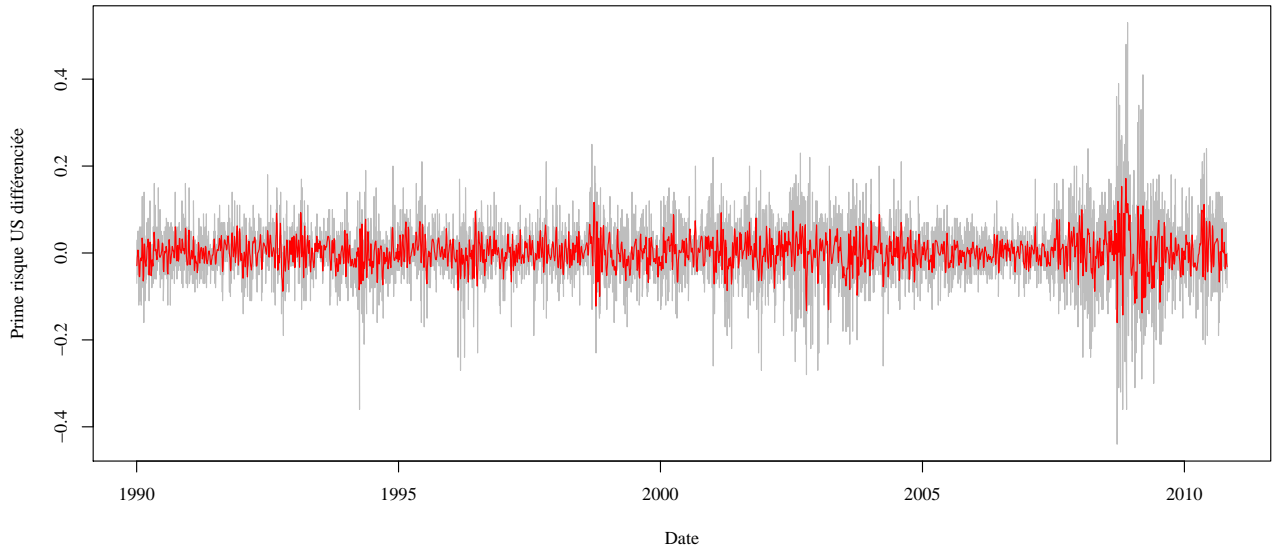


FIGURE 4.16 – Les variations de l’approximation de prime de risque liée au marché action (en gris) et sa composante prévisible estimée (en rouge).

1. Reconstruire l'espace des phases : estimer le temps de retard τ puis la dimension de plongement d . Les vecteurs retard s'écrivent alors comme :

$$G(t, \tau, d) = (\widehat{m}(t), \widehat{m}(t - \tau), \dots, \widehat{m}(t - (d - 1)\tau))^{Tr}$$

2. Rechercher les k plus proches voisins de chaque vecteur de retard en utilisant une distance euclidienne ($k > 1$).

Afin de simplifier les notations de ce paragraphe, le vecteur retard $G(t, \tau, d)$ sera dorénavant noté $G(t)$. Dans un voisinage de $G(t)$, noté $V(G(t))$, \mathbf{F} peut, par exemple, être approximée par un développement de Taylor à l'ordre 2 :

$$\forall G(t_0) \in V(G(t)), \quad \mathbf{F}(G(t_0)) \approx \mathbf{F}(G(t)) + D\mathbf{F}(G(t))(G(t_0) - G(t)) + \frac{1}{2}(G(t_0) - G(t))^{Tr} D^2\mathbf{F}(G(t))(G(t_0) - G(t))$$

où $D\mathbf{F}$ est le gradient de \mathbf{F} et $D^2\mathbf{F}$ correspond à la matrice Hessienne de \mathbf{F} . L'équation peut alors être réécrite sous forme d'un produit de deux vecteurs :

$$\forall G(t_0) \in V(G(t)), \quad \mathbf{F}(G(t_0)) \approx C_{G(t)}^{Tr} \mathbf{B}_{G(t)}$$

où $C_{G(t)}$ et $\mathbf{B}_{G(t)}$ sont deux vecteurs de longueur $(1 + d + d^2)$ définis par :

$$C_{G(t)} = \begin{pmatrix} 1 \\ (G_1(t_0) - G_1(t)) \\ \vdots \\ (G_d(t_0) - G_d(t)) \\ (G_1(t_0) - G_1(t))^2/2 \\ (G_1(t_0) - G_1(t))(G_2(t_0) - G_2(t))/2 \\ \vdots \\ (G_d(t_0) - G_d(t))(G_{d-1}(t_0) - G_{d-1}(t))/2 \\ (G_d(t_0) - G_d(t))^2/2 \end{pmatrix}$$

et

$$\mathbf{B}_{G(t)} = \begin{pmatrix} \mathbf{F}(G(t)) \\ \partial\mathbf{F}/\partial G_1(t_0)(G(t)) \\ \vdots \\ \partial\mathbf{F}/\partial G_d(t_0)(G(t)) \\ \partial^2\mathbf{F}/\partial G_1(t_0)^2(G(t)) \\ \partial^2\mathbf{F}/\partial G_1(t_0)\partial G_2(t_0)(G(t)) \\ \vdots \\ \partial^2\mathbf{F}/\partial G_d(t_0)\partial G_{d-1}(t_0)(G(t)) \\ \partial^2\mathbf{F}/\partial G_d(t_0)^2(G(t)) \end{pmatrix}$$

où $G_i(t)$ correspond à la $i^{\text{ème}}$ coordonnée de $G(t)$, c'est-à-dire $\widehat{m}(t - (i - 1)\tau)$.

C'est pourquoi, la régression non-paramétrique polynomial locale consiste à minimiser la somme pondérée des erreurs de prédiction :

$$\widehat{\mathbf{B}}_{G(t)} = \arg \min_{\mathbf{B}_*} (\Delta(\omega) - \mathbf{D}_{G(t)}\mathbf{B}_*)^{Tr} K_{H,G(t)}(\Delta(\omega) - \mathbf{D}_{G(t)}\mathbf{B}_*) \quad (4.12)$$

où

- \mathbf{B}_* est un vecteur de dimension $(1 + d + d^2)$.
- $\mathbf{D}_{G(t)} = \left(C_{G_*^1}^{Tr} \quad \dots \quad C_{G_*^l}^{Tr} \quad \dots \quad C_{G_*^k}^{Tr} \right)^{Tr}$ avec G_*^l le $l^{\text{ème}}$ plus proche voisin de $G(t)$.
- $\Delta(\omega) = \begin{pmatrix} \widehat{m}(t_{G_*^1} + \omega) \\ \vdots \\ \widehat{m}(t_{G_*^k} + \omega) \end{pmatrix}$

où $t_{G_*^l}$ est défini par $G_*^l = (\widehat{m}(t_{G_*^l}), \widehat{m}(t_{G_*^l} - \tau), \dots, \widehat{m}(t_{G_*^l} - (d-1)\tau))^{Tr}$.

- une matrice des poids
 $K_{H,G(t)} = (K_H(G_*^1 - G(t)) \quad K_H(G_*^2 - G(t)) \quad \dots \quad K_H(G_*^k - G(t))) \mathbf{I}_k$
 avec \mathbf{I}_k une matrice carrée identité de taille $k \times k$ et $K_H(\cdot)$ une fonction noyau.

La dérivée première par rapport à \mathbf{B}_* de la somme pondérée des erreurs de prédiction permet d'en déduire que :

$$\widehat{\mathbf{B}}_{G(t)} = (\mathbf{D}_{G(t)}^{Tr} K_{H,G(t)} \mathbf{D}_{G(t)})^{-1} \mathbf{D}_{G(t)}^{Tr} K_{H,G(t)} \Delta(\omega) \quad (4.13)$$

Preuve.

$$\frac{\partial}{\partial \mathbf{B}_*} (\Delta(\omega) - \mathbf{D}_{G(t)} \mathbf{B}_*)^{Tr} K_{H,G(t)} (\Delta(\omega) - \mathbf{D}_{G(t)} \mathbf{B}_*) \quad (4.14)$$

$$= \frac{\partial}{\partial \mathbf{B}_*} \Delta(\omega)^{Tr} K_{H,G(t)} \Delta(\omega) - 2\mathbf{B}_*^{Tr} \mathbf{D}_{G(t)}^{Tr} K_{H,G(t)} \Delta(\omega) + \mathbf{B}_*^{Tr} \mathbf{D}_{G(t)}^{Tr} K_{H,G(t)} \mathbf{D}_{G(t)} \mathbf{B}_* \quad (4.15)$$

$$= 0 - 2\mathbf{D}_{G(t)}^{Tr} K_{H,G(t)} \Delta(\omega) + 2\mathbf{D}_{G(t)}^{Tr} K_{H,G(t)} \mathbf{D}_{G(t)} \mathbf{B}_* \quad (4.16)$$

Ce qui permet d'aboutir au résultat. \square

On en déduit le vecteur de biais et la matrice de variance-covariance de cet estimateur :

- $\widehat{\mathbf{B}}_{G(t)} - \mathbf{B}_{G(t)} = (\mathbf{D}_{G(t)}^{Tr} K_{H,G(t)} \mathbf{D}_{G(t)})^{-1} \mathbf{D}_{G(t)}^{Tr} K_{H,G(t)} \eta$ où η correspond au vecteur des résidus.
- $\mathbb{V}[\widehat{\mathbf{B}}_{G(t)}] = (\mathbf{D}_{G(t)}^{Tr} K_{H,G(t)} \mathbf{D}_{G(t)})^{-1} \mathbf{D}_{G(t)}^{Tr} K_{H,G(t)} \mathbb{E}[\eta \eta^{Tr}] K_{H,G(t)}^{Tr} \mathbf{D}_{G(t)} (\mathbf{D}_{G(t)}^{Tr} K_{H,G(t)}^{Tr} \mathbf{D}_{G(t)})^{-1}$

Preuve.

$$\widehat{\mathbf{B}}_{G(t)} = (\mathbf{D}_{G(t)}^{Tr} K_{H,G(t)} \mathbf{D}_{G(t)})^{-1} \mathbf{D}_{G(t)}^{Tr} K_{H,G(t)} \Delta(\omega) \quad (4.17)$$

$$= (\mathbf{D}_{G(t)}^{Tr} K_{H,G(t)} \mathbf{D}_{G(t)})^{-1} \mathbf{D}_{G(t)}^{Tr} K_{H,G(t)} (\mathbf{T} + \eta) \quad (4.18)$$

où $\mathbf{T} = \begin{pmatrix} \mathbf{F}(M_{t_{G_*^1}}) \\ \vdots \\ \mathbf{F}(M_{t_{G_*^k}}) \end{pmatrix}$ avec $M_t = (\widehat{m}(t), \dots, \widehat{m}(t - (d-1)\tau))$. Or, $\mathbf{T} = \mathbf{D}_{G(t)} \mathbf{B}_{G(t)}$, on a donc :

$$\widehat{\mathbf{B}}_{G(t)} = \mathbf{B}_{G(t)} + (\mathbf{D}_{G(t)}^{Tr} K_{H,G(t)} \mathbf{D}_{G(t)})^{-1} \mathbf{D}_{G(t)}^{Tr} K_{H,G(t)} \eta \quad (4.19)$$

$$(4.20)$$

Pour la matrice de variance-covariance de l'estimateur,

$$\mathbb{V}[\widehat{\mathbf{B}}_{G(t)}] = \mathbb{E}[(\widehat{\mathbf{B}}_{G(t)} - \mathbb{E}[\widehat{\mathbf{B}}_{G(t)} | G_*^1, \dots, G_*^k]) (\widehat{\mathbf{B}}_{G(t)} - \mathbb{E}[\widehat{\mathbf{B}}_{G(t)} | G_*^1, \dots, G_*^k])^{Tr}] \quad (4.21)$$

$$= \mathbb{E}[(\mathbf{D}_{G(t)}^{Tr} K_{H,G(t)} \mathbf{D}_{G(t)})^{-1} \mathbf{D}_{G(t)}^{Tr} K_{H,G(t)} \eta \eta^{Tr} K_{H,G(t)}^{Tr} \mathbf{D}_{G(t)} (\mathbf{D}_{G(t)}^{Tr} K_{H,G(t)}^{Tr} \mathbf{D}_{G(t)})^{-1}] \quad (4.22)$$

$$= (\mathbf{D}_{G(t)}^{Tr} K_{H,G(t)} \mathbf{D}_{G(t)})^{-1} \mathbf{D}_{G(t)}^{Tr} K_{H,G(t)} \mathbb{E}[\eta \eta^{Tr}] K_{H,G(t)}^{Tr} \mathbf{D}_{G(t)} (\mathbf{D}_{G(t)}^{Tr} K_{H,G(t)}^{Tr} \mathbf{D}_{G(t)})^{-1} \quad (4.23)$$

\square

L'estimation de la matrice de variance-covariance de l'estimateur dépend donc d'une estimation de la matrice de variance-covariance des résidus, notée $\mathbb{E}[\eta\eta^{Tr}]$.

Au final, cette manière de prédire la valeur de la composante prévisible repose sur quatre ingrédients : l'ordre du développement de Taylor, le choix de la fonction noyau $K_H(\cdot)$, le nombre k de plus proches voisins sélectionnés ainsi que la fenêtre H de la fonction noyau. Pour choisir ces ingrédients, on peut se référer au paragraphe 3.3.2.

Remarque sur la fonction noyau Une fonction noyau dans le cas multivarié est une fonction de m variables satisfaisant

$$\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} K(x) dx_1 \dots dx_m = 1$$

où $x = (x_1, \dots, x_m)^{Tr}$. Des choix généralement fait pour la fonction $K(x)$ sont (voir paragraphe 3.3.2) :

- La densité normale standard :

$$K(x) = \frac{1}{(2\pi)^{-m/2}} e^{-\|x\|^2/2}$$

- La fonction de Epanechnikov :

$$K(x) = \frac{m(m+2)\Gamma(m/2)}{4\pi^{m/2}} (1 - \|x\|^2) \mathbf{1}_{\|x\| \leq 1}$$

Si on cherche le plus proche voisin de x , il faut donner un poids plus important aux vecteurs les plus proches de x par rapport à l'ensemble des autres vecteurs auxquels il peut être comparé. Or, cet ensemble de comparaison peut varier : deux vecteurs proches dans un ensemble peuvent ne pas l'être dans un autre ensemble. Il peut donc être utile d'utiliser une fenêtre afin de pondérer les fonctions $K(x)$ suivant les ensembles de comparaison. Soit H une matrice symétrique définie positive que l'on nommera « matrice de fenêtrage ». La fonction noyau pondérée utilisée au sein de la matrice des poids est alors définie par :

$$K_H(x - y) = \frac{1}{\det(H)} K(H^{-1}(x - y))$$

Dans ce cadre, H est une matrice carrée $m \times m$. Afin de diminuer le nombre d'hyperparamètres, il est plus simple d'utiliser, en pratique, une fenêtre $h \in \mathbb{R}$ telle que $H = h\mathbf{I}_m$.

Remarque sur l'inversion de la matrice $\mathbf{D}_{G(t)}^{Tr} K_{H,G(t)} \mathbf{D}_{G(t)}$ En pratique, il peut arriver que la matrice $\mathbf{D}_{G(t)}^{Tr} K_{H,G(t)} \mathbf{D}_{G(t)}$ ne soit pas inversible ($\det[\mathbf{D}_{G(t)}^{Tr} K_{H,G(t)} \mathbf{D}_{G(t)}] = 0$). Puisque $K_{H,G(t)}$ est une matrice diagonale par construction, cela signifie qu'une ou plusieurs colonnes de la matrice $\mathbf{D}_{G(t)}$ sont colinéaires. Plusieurs méthodes ont été proposées pour résoudre ce type de problème. Les méthodes les plus courantes sont la régression ridge (voir [Bjo(2001)]) et la régression sur composantes principales (voir [Jol(2002)]).

Dans le cas de la régression ridge, l'estimateur de $\mathbf{B}_{G(t)}$ est le suivant :

$$\widehat{\mathbf{B}}_{G(t)} = (\mathbf{D}_{G(t)}^{Tr} K_{H,G(t)} \mathbf{D}_{G(t)} + \lambda \mathbf{I}_{1+d+d^2})^{-1} \mathbf{D}_{G(t)}^{Tr} K_{H,G(t)} \Delta(\omega)$$

où $\lambda \in \mathbb{R}^*$ est un hyperparamètre supplémentaire à déterminer (voir paragraphe suivant). Dans ce cadre, $\det[\mathbf{D}_{G(t)}^{Tr} K_{H,G(t)} \mathbf{D}_{G(t)} + \lambda \mathbf{I}_{1+d+d^2}] = (1 + d + d^2)\lambda \neq 0$ et cette matrice est bien inversible. Le choix de λ étant assez délicat, il peut être préférable d'utiliser une régression sur composantes principales afin d'obtenir une estimation assez robuste de $\widehat{\mathbf{B}}_{G(t)}$.

Dans le cas de la régression sur composantes principales, on effectue une analyse en composantes principales sur les colonnes de la matrice $\sqrt{K_{H,G(t)}} \mathbf{D}_{G(t)}$ où $\sqrt{K_{H,G(t)}}$ correspond à la matrice $(\sqrt{K_H(G_*^1 - G(t))} \quad \sqrt{K_H(G_*^2 - G(t))} \quad \dots \quad \sqrt{K_H(G_*^k - G(t))}) \mathbf{I}_k$. Cette matrice peut alors être transformée en une matrice dont les colonnes sont orthonormées $\mathbf{Z}_{G(t)} = (\sqrt{K_{H,G(t)}} \mathbf{D}_{G(t)}) W$ où W est une matrice orthogonale dont les vecteurs colonnes sont les vecteurs propres de la matrice $\Sigma = \mathbf{D}_{G(t)}^{Tr} K_{H,G(t)} \mathbf{D}_{G(t)}$. On a alors $\sqrt{K_{H,G(t)}} \mathbf{D}_{G(t)} W W^{Tr} \mathbf{B}_* = \mathbf{Z}_{G(t)} \gamma$, avec $\gamma = W^{Tr} \mathbf{B}_*$. La fonction à minimiser en (4.12) peut donc se réécrire comme :

$$(\sqrt{K_{H,G(t)}} \Delta(\omega) - \mathbf{Z}_{G(t)} \gamma)^{Tr} (\sqrt{K_{H,G(t)}} \Delta(\omega) - \mathbf{Z}_{G(t)} \gamma) \quad (4.24)$$

L'idée est alors de limiter la régression aux vecteurs propres dont les valeurs propres sont les plus élevées. On procède donc de la façon suivante :

1. Calcul des vecteurs propres de la matrice Σ .
2. Sélection de a vecteurs propres tels que

$$\frac{\sum_{i=1}^a \lambda_i}{\sum_{i=1}^{1+d+d^2} \lambda_i} > s$$

où λ_i est la valeur propre associée au $i^{\text{ème}}$ vecteur propre de Σ et s est un certain pourcentage de la variance totale expliquée par les vecteurs sélectionnés, par exemple 99%.

3. Calcul de $\mathbf{Z}_{G(t),a} = (\sqrt{K_{H,G(t)}} \mathbf{D}_{G(t)}) W_a$ où W_a est une matrice dont les colonnes sont les vecteurs propres sélectionnés.
4. Calcul de $\widehat{\gamma} = (\mathbf{Z}_{G(t),a}^{Tr} \mathbf{Z}_{G(t),a})^{-1} \mathbf{Z}_{G(t),a}^{Tr} \sqrt{K_{H,G(t)}} \Delta(\omega)$.
5. $\widehat{\mathbf{B}}_{G(t)} = W_a \widehat{\gamma}$.
6. Puisque $\mathbb{V}[\widehat{\gamma}] = W_a^{Tr} \mathbb{V}[\widehat{\mathbf{B}}_{G(t)}] W_a$, $\mathbb{V}[\widehat{\mathbf{B}}_{G(t)}] = W_a \mathbb{V}[\widehat{\gamma}] W_a^{Tr}$. Or, le biais de $\widehat{\gamma}$ est égal à :

$$\widehat{\gamma} - \gamma = (\mathbf{Z}_{G(t),a}^{Tr} \mathbf{Z}_{G(t),a})^{-1} \mathbf{Z}_{G(t),a}^{Tr} \sqrt{K_{H,G(t)}} \eta$$

Ce qui permet d'en déduire la variance de $\widehat{\gamma}$:

$$\mathbb{V}[\widehat{\gamma}] = (\mathbf{Z}_{G(t),a}^{Tr} \mathbf{Z}_{G(t),a})^{-1} \mathbf{Z}_{G(t),a}^{Tr} \sqrt{K_{H,G(t)}} \mathbb{E}[\eta \eta^{Tr}] \sqrt{K_{H,G(t)}}^{Tr} \mathbf{Z}_{G(t),a} (\mathbf{Z}_{G(t),a}^{Tr} \mathbf{Z}_{G(t),a})^{-1}$$

Remarque sur le nombre de plus proches voisins Lorsqu'on tente de prédire certaines séries, notamment des séries non-stationnaires, il est parfois possible de ne pas trouver suffisamment de plus proches voisins dans le passé pour effectuer des prévisions. Il faut alors différencier suffisamment la série afin d'obtenir une série différenciée ayant suffisamment de plus proches voisins dans le passé.

Détermination des hyperparamètres Les hyperparamètres peuvent être sélectionnés afin de minimiser la somme des erreurs de prédictions obtenue sur plusieurs échantillons de données passées (méthode de validation croisée *d - delete*). Dans le cas où h et k sont les seuls hyperparamètres, on les sélectionne en minimisant la somme des erreurs de prédictions suivante :

$$\sum_{t=a}^b (\Delta(\omega) - \mathbf{D}_{G(t)} \widehat{\mathbf{B}}_{G(t)})^{Tr} K_{H,G(t)} (\Delta(\omega) - \mathbf{D}_{G(t)} \widehat{\mathbf{B}}_{G(t)}) \quad (4.25)$$

où $a > (d - 1)\tau + 1$ et $b < T - w$.

Pour effectuer cette minimisation, on présélectionne un ensemble de valeurs que peut prendre k , par exemple $[5, 10, 15, 20, 25, 30]$, puis on estime la fenêtre optimale h , pour chaque valeur de cet ensemble, en utilisant un algorithme de minimisation ne nécessitant pas de calcul de dérivées (voir [Bre(1973)]). Le couple (h, k) minimisant la somme des erreurs de prédiction est alors sélectionné.

Exemple sur une série générée On génère une série de 500 points par

$$y_t = \mu_t + \epsilon_t$$

où $t \in [1, \dots, 500]$, $\mu_t = \sin(\frac{t}{10})$ et ϵ_t suit une loi $\mathcal{N}(0, 0.1)$. On utilise un développement de Taylor à l'ordre 2 et une densité normale standard comme fonction noyau. Le nombre de plus proches voisins et la fenêtre sont déterminés à la manière du paragraphe précédent. La figure 4.17 représente la série y_t et les prédictions de $t + 1$ à $t + 50$.

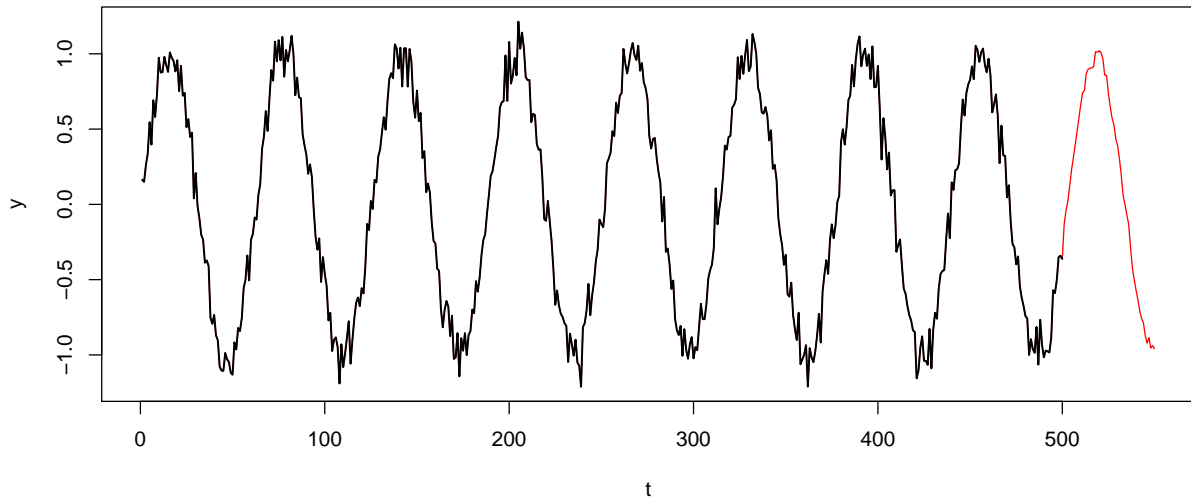


FIGURE 4.17 – La série y_t est représentée en noir et les 50 prédictions de y_t sont représentées en rouge. On estime $\tau = 5$ et $d = 6$. Le couple (k, h) utilisé est celui minimisant les erreurs de prédictions sur les données passées, à savoir $k = 15$ et $h = 0.1453361$.

4.3.2 Prédiction de plusieurs composantes chaotiques

On s'intéresse d'abord au cas où on possède deux composantes prévisibles puis ensuite au cas de p composantes prévisibles.

Cas de deux composantes chaotiques et développement de Taylor à l'ordre 1

Dans le paragraphe précédent, nous avons supposé que $\mathbf{F} : \mathbb{R}^d \rightarrow \mathbb{R}$. Ici, nous allons supposer que $\mathbf{F} : \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \rightarrow \mathbb{R}^2$:

$$(\widehat{m}_1(T + \omega) \quad \widehat{m}_2(T + \omega)) = \mathbf{F} \begin{pmatrix} \widehat{m}_1(T) & \widehat{m}_2(T) \\ \vdots & \vdots \\ \widehat{m}_1(T - (d_1 - 1)\tau_1) & \widehat{m}_2(T - (d_2 - 1)\tau_2) \end{pmatrix} + \eta_T \quad (4.26)$$

où η_T est un bruit centré sur 0.

L'idée est alors de se ramener à l'estimation de deux fonctions \mathbf{F}_1 et \mathbf{F}_2 ne prenant en compte qu'un seul vecteur appartenant à $\mathbb{R}^{d_1+d_2}$. On va supposer que ces deux fonctions $\mathbf{F}_1 : \mathbb{R}^{d_1+d_2} \rightarrow \mathbb{R}$ et $\mathbf{F}_2 : \mathbb{R}^{d_1+d_2} \rightarrow \mathbb{R}$ sont définies par :

$$\widehat{m}_1(T + \omega) = \mathbf{F}_1 \begin{pmatrix} V_1(T) \\ V_2(T) \end{pmatrix} + \eta_{1,T} \quad (4.27)$$

$$\widehat{m}_2(T + \omega) = \mathbf{F}_2 \begin{pmatrix} V_1(T) \\ V_2(T) \end{pmatrix} + \eta_{2,T} \quad (4.28)$$

où $\eta_{1,T}$ et $\eta_{2,T}$ sont des bruits centrés sur 0 et où $V_1(T) = (\widehat{m}_1(T) \quad \dots \quad \widehat{m}_1(T - (d_1 - 1)\tau_1))^{Tr}$ et $V_2(T) = (\widehat{m}_2(T) \quad \dots \quad \widehat{m}_2(T - (d_2 - 1)\tau_2))^{Tr}$. Une fois que l'on a estimé les temps de retard τ_1 et τ_2 ainsi que les dimensions de plongement d_1 et d_2 , il faut rechercher les k plus proches voisins du vecteur $G(T) = \begin{pmatrix} V_1(T) \\ V_2(T) \end{pmatrix}$ dans le passé. Ceci peut être fait à l'aide d'une distance Euclidienne.

Notons $(t_{G_*^l})_{l \in [1, \dots, k]}$ les instants correspondant aux instants des plus proches voisins de $G(T)$. Ces derniers doivent être recherchés dans l'intervalle $[(d_{i^*} - 1)\tau_{i^*}, T - \omega]$ où $i^* = \arg \max_{i \in [1, 2]} (d_i - 1)\tau_i$ de façon à pouvoir extraire des vecteurs de phases pour \widehat{m}_1 et \widehat{m}_2 ainsi que pour avoir des valeurs disponibles pour $\widehat{m}_1(t_{G_*^l} + \omega)$ et $\widehat{m}_2(t_{G_*^l} + \omega)$.

La démarche est ensuite la même que dans le cas univarié. Par exemple, si on utilise un développement de Taylor à l'ordre 1 pour approximer les fonctions \mathbf{F}_1 et \mathbf{F}_2 ,

$$\forall t_0 \in (t_{G_*^l})_{l \in [1, \dots, k]}, \quad \begin{pmatrix} \mathbf{F}_1(G(t_0)) \\ \mathbf{F}_2(G(t_0)) \end{pmatrix} \approx \begin{pmatrix} \mathbf{F}_1(G(T)) + D\mathbf{F}_1(G(T))(G(t_0) - G(T)) \\ \mathbf{F}_2(G(T)) + D\mathbf{F}_2(G(T))(G(t_0) - G(T)) \end{pmatrix} \approx \mathbf{B}_{G(T)}^{Tr} C_{G(t_0)} \quad (4.29)$$

où :

$$C_{G(t_0)} = \begin{pmatrix} 1 \\ \hat{m}_1(t_0) - \hat{m}_1(T) \\ \hat{m}_1(t_0 - \tau_1) - \hat{m}_1(T - \tau_1) \\ \vdots \\ \hat{m}_1(t_0 - (d_1 - 1)\tau_1) - \hat{m}_1(T - (d_1 - 1)\tau_1) \\ \hat{m}_2(t_0) - \hat{m}_2(T) \\ \hat{m}_2(t_0 - \tau_2) - \hat{m}_2(T - \tau_2) \\ \vdots \\ \hat{m}_2(t_0 - (d_2 - 1)\tau_2) - \hat{m}_2(T - (d_2 - 1)\tau_2) \end{pmatrix}$$

et

$$\mathbf{B}_{G(T)} = \begin{pmatrix} \mathbf{F}_1(G(T)) & \mathbf{F}_2(G(T)) \\ \partial\mathbf{F}_1/\partial G_1(t_0)(G(T)) & \partial\mathbf{F}_2/\partial G_1(t_0)(G(T)) \\ \partial\mathbf{F}_1/\partial G_2(t_0)(G(T)) & \partial\mathbf{F}_2/\partial G_2(t_0)(G(T)) \\ \vdots & \vdots \\ \partial\mathbf{F}_1/\partial G_{d_1}(t_0)(G(T)) & \partial\mathbf{F}_2/\partial G_{d_1}(t_0)(G(T)) \\ \partial\mathbf{F}_1/\partial G_{d_1+1}(t_0)(G(T)) & \partial\mathbf{F}_2/\partial G_{d_1+1}(t_0)(G(T)) \\ \vdots & \vdots \\ \partial\mathbf{F}_1/\partial G_{d_1+d_2}(t_0)(G(T)) & \partial\mathbf{F}_2/\partial G_{d_1+d_2}(t_0)(G(T)) \end{pmatrix}$$

Le vecteur $\mathbf{B}_{G(T)}$ peut alors s'estimer par

$$\hat{\mathbf{B}}_{G(T)} = (D_{G(T)}^{Tr} K_{H,G(T)} D_{G(T)})^{-1} D_{G(T)}^{Tr} K_{H,G(T)} \Delta(\omega)$$

où

- $\Delta(\omega) = \begin{pmatrix} \hat{m}_1(t_{G_1^*} + \omega) & \hat{m}_2(t_{G_1^*} + \omega) \\ \vdots & \vdots \\ \hat{m}_1(t_{G_k^*} + \omega) & \hat{m}_2(t_{G_k^*} + \omega) \end{pmatrix}$
- $D_{G(T)} = \begin{pmatrix} C_{G_1^*}^{Tr} \\ \vdots \\ C_{G_k^*}^{Tr} \end{pmatrix}$ est une matrice de taille $k \times (1 + d_1 + d_2)$.
- $K_{H,G(T)}$ est une matrice de taille $k \times k$ composée des fonctions noyaux.

Cas de p composantes chaotiques et développement de Taylor à l'ordre N

Afin de généraliser au cas de p composantes prévisibles, il suffit d'estimer les fonctions $\mathbf{F}_1, \dots, \mathbf{F}_p$ allant de $\mathbb{R}^{d_1+\dots+d_p}$ dans \mathbb{R} et associant un vecteur $G(T) = (V_1(T), \dots, V_p(T))^{Tr}$ à, respectivement, $\hat{m}_1(t+\omega), \dots, \hat{m}_p(t+\omega)$. Dans ce cadre, on utilise la formule de Taylor à l'ordre N pour des fonctions

multivariées :

$$\forall t_0 \in (t_{G_*^l})_{l \in [1, \dots, k]}, \quad \begin{pmatrix} \mathbf{F}_1(G(t_0)) \\ \mathbf{F}_2(G(t_0)) \\ \vdots \\ \mathbf{F}_p(G(t_0)) \end{pmatrix} \approx \begin{pmatrix} \mathbf{F}_1(G(T)) + \sum_{m=1}^N \sum_{\substack{\alpha_1, \dots, \alpha_q \\ t.q. |\alpha|=m}} \frac{D^\alpha \mathbf{F}_1(G(T))}{\alpha!} (G(t_0) - G(T))^\alpha \\ \mathbf{F}_2(G(T)) + \sum_{m=1}^N \sum_{\substack{\alpha_1, \dots, \alpha_q \\ t.q. |\alpha|=m}} \frac{D^\alpha \mathbf{F}_2(G(T))}{\alpha!} (G(t_0) - G(T))^\alpha \\ \vdots \\ \mathbf{F}_p(G(T)) + \sum_{m=1}^N \sum_{\substack{\alpha_1, \dots, \alpha_q \\ t.q. |\alpha|=m}} \frac{D^\alpha \mathbf{F}_p(G(T))}{\alpha!} (G(t_0) - G(T))^\alpha \end{pmatrix} \quad (4.30)$$

où :

- $q = \sum_{j=1}^p d_j$, $|\alpha| = \sum_{i=1}^q \alpha_i$ et $\alpha! = \prod_{i=1}^q \alpha_i!$
- $\forall i \in [1, q]$, $\alpha_i \in [0, \dots, |\alpha|]$
- $(G(t_0) - G(T))^\alpha = \prod_{i=1}^q (G_i(t_0) - G_i(T))^{\alpha_i}$ où $G_i(T)$ correspond à la $i^{\text{ème}}$ coordonnée du vecteur $G(T)$.
- $D^\alpha \mathbf{F}(G(T)) = \frac{\partial^{|\alpha|} \mathbf{F}(G(T))}{\partial G_1(t_0)^{\alpha_1} \partial G_2(t_0)^{\alpha_2} \dots \partial G_q(t_0)^{\alpha_q}}$

On peut ensuite exprimer cette matrice comme un produit $\mathbf{B}_{G(T)}^{Tr} C_{G(t_0)}$ où $\mathbf{B}_{G(T)}$ est une matrice ayant pour première ligne le vecteur $(\mathbf{F}_1(G(T)) \dots \mathbf{F}_p(G(T)))$ et $C_{G(t_0)}$ est un vecteur dont la première coordonnée est 1 et contenant les produits $(G(t_0) - G(T))^\alpha$. La démarche est ensuite la même que dans le cas précédent. La première ligne de l'estimateur $\mathbf{B}_{G(T)}$ correspond aux prédictions :

$$\widehat{\mathbf{B}}_{G(T)} = (D_{G(T)}^{Tr} K_{H,G(T)} D_{G(T)})^{-1} D_{G(T)}^{Tr} K_{H,G(T)} \Delta(\omega)$$

où

- $\Delta(\omega) = \begin{pmatrix} \widehat{m}_1(t_{G_1^*} + \omega) & \widehat{m}_2(t_{G_1^*} + \omega) & \dots & \widehat{m}_p(t_{G_1^*} + \omega) \\ \vdots & \vdots & & \\ \widehat{m}_1(t_{G_k^*} + \omega) & \widehat{m}_2(t_{G_k^*} + \omega) & \dots & \widehat{m}_p(t_{G_k^*} + \omega) \end{pmatrix}$
- $C_{G_*^1}$ est un vecteur de taille $(1 + \sum_{m=1}^N \sum_{\substack{\alpha_1, \dots, \alpha_q \\ t.q. |\alpha|=m}} \frac{|\alpha|}{\alpha!})$ dont la première coordonnée est 1 et contenant les produits $(G(t_0) - G(T))^\alpha$ pour les vecteurs α vérifiant $|\alpha| \in [1, \dots, N]$.
- $D_{G(T)} = \begin{pmatrix} C_{G_*^1}^{Tr} \\ \vdots \\ C_{G_*^k}^{Tr} \end{pmatrix}$ est une matrice de taille $k \times (1 + \sum_{m=1}^N \sum_{\substack{\alpha_1, \dots, \alpha_q \\ t.q. |\alpha|=m}} \frac{|\alpha|}{\alpha!})$.
- $K_{H,G(T)}$ est une matrice de taille $k \times k$ composée des fonctions noyaux.

Exemple sur des séries générées On va considérer deux séries générées d’après l’attracteur de Ikeda ([Ham(1985)]). La fonction de Ikeda est la suivante :

$$x_{n+1} = p + \mu(x_n \cos(t) + y_n \sin(t))$$

$$y_{n+1} = \mu(x_n \sin(t) + y_n \cos(t))$$

où $p = 1$, $\mu = 0.9$, $t = k - \frac{\alpha}{1+x_n^2+y_n^2}$ avec $k = 0.4$ et $\alpha = 6$. Cette fonction est également utilisée dans l’article de Cao [Cao(1997)].

On génère des séries $x = (x_n)_{n \in [1,700]}$ et $y = (y_n)_{n \in [1,700]}$ puis on compare les prédictions réalisées à partir de ces données aux vraies données. Pour $t \in [100, \dots, 700]$, l’erreur de prédiction est évaluée au fur et à mesure par $\|\hat{x}_{t+\omega} - x_{t+\omega}\| + \|\hat{y}_{t+\omega} - y_{t+\omega}\|$ où $\hat{x}_{t+\omega}$ et $\hat{y}_{t+\omega}$ correspondent à des prédictions réalisées en $t + \omega$ à partir des données appartenant à $[1, t]$. L’erreur de prédiction moyenne obtenue est reportée entre crochets dans le tableau 4.9 pour des prédictions en $t + \omega$ avec $\omega = 1, 10, 20, 30$ et 40. Nous avons utilisé un développement de Taylor à l’ordre 2 pour faire ces prédictions.

$k \backslash \omega$	1	10	20	30	40	<i>Err.Moy.</i>
10	0.622 [8.83]	0.578 [53.89]	0.708 [40.27]	0.622 [35.49]	0.611 [62.84]	40.27
15	0.767 [8.02]	0.633 [50.84]	0.631 [38.03]	0.632 [35.93]	0.797 [51.63]	36.89
20	0.660 [7.26]	0.679 [48.16]	0.537 [43.53]	0.337 [40.69]	0.691 [58.01]	39.53
25	0.417 [9.63]	0.619 [48.95]	0.629 [42.97]	0.825 [33.92]	0.843 [50.61]	37.21

TABLEAU 4.9 – Fenêtre h minimisant l’erreur de prédiction pour différents nombre de plus proches voisins (k) et pour différents horizons de prédiction (ω). L’erreur de prédiction moyenne obtenue pour chaque k et chaque ω est exprimée entre crochets. La colonne *Err.Moy.* correspond à la moyenne de ces erreurs de prédictions en fonction du nombre de plus proches voisins.

Les résultats présentés dans le tableau 4.9 semblent indiquer qu’un nombre de plus proches voisins de 15 pour une fenêtre approximativement égale à 0.63 permet d’obtenir de bonnes prédictions entre $t + 10$ et $t + 40$. La figure 4.18 présente les prédictions réalisées pour $k = 15$ et $h = 0.63$.

4.4 Correction des effets de bord

4.4.1 Formulation du problème

En pratique, ce sont souvent les dernières estimations de tendance qui sont les plus importantes pour la prise de décision du gérant de portefeuille. Malheureusement, ces estimations sont soumises à des effets de bord. En d’autres termes, les dernières estimations de tendance sont sujettes à des révisions significatives lorsque de nouvelles données sont disponibles. Concernant la décomposition modale empirique, ceci est due aux fonctions d’interpolation des extremas qui créent des distorsions au niveau des bords. On s’intéresse donc ici à l’aspect adaptatif de l’estimation de la tendance.

On va supposer que l’on se trouve dans le cadre suivant :

$$y_t = \hat{\mu}_t + \hat{\epsilon}_t$$

où $\hat{\mu}_t$ est le paramètre de tendance et $\hat{\epsilon}_t$ celui du bruit.

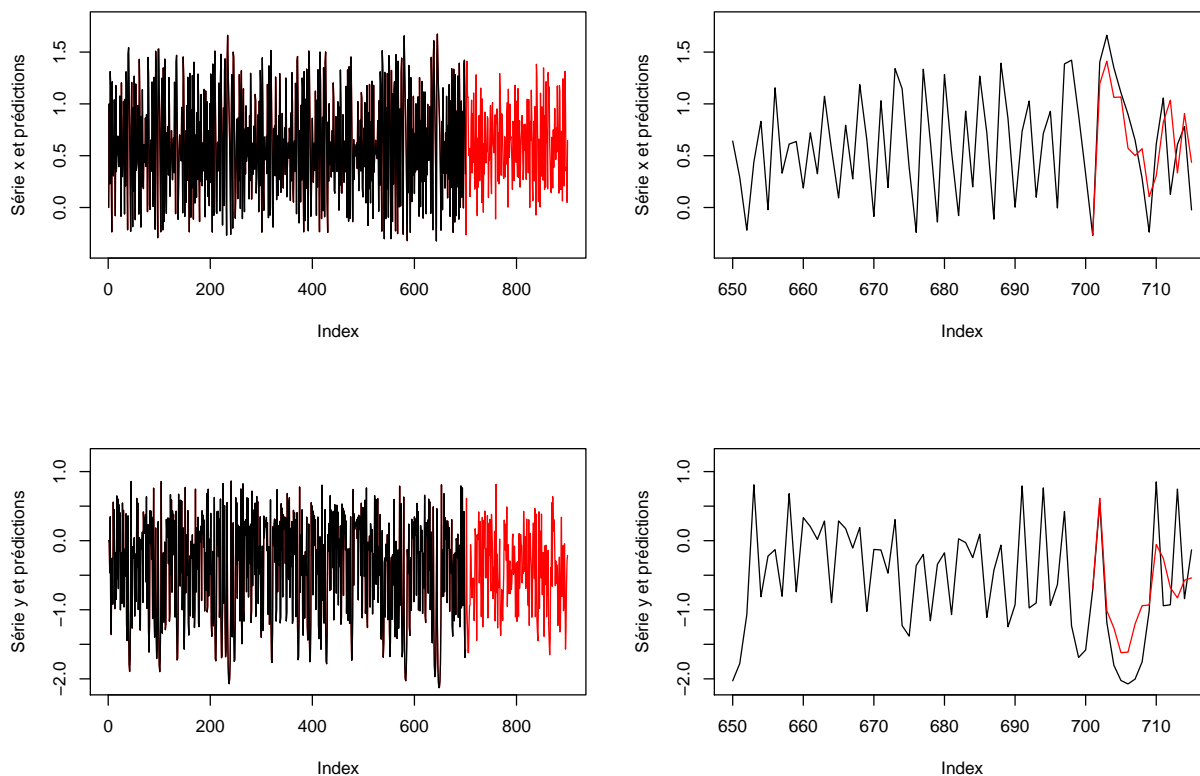


FIGURE 4.18 – Les graphiques de gauche présente les 200 prédictions réalisées (en rouge) à partir de la 700 ème donnée pour $h = 0.63$ et $k = 15$. Les graphiques de droite comparent les quinze prédictions réalisées (en rouge) à partir de la 700 ème donnée aux vraies données.

Notons \mathcal{F}_t l'information disponible à l'instant t . On suppose que l'on a estimé une tendance $\hat{\mu}_t|_{\mathcal{F}_t}$ avec l'information disponible à l'instant t et on s'intéresse à la révision de cette estimation, notée $\hat{\mu}_t|_{\mathcal{F}_{t+\Delta}}$ où $\Delta > 0$. On a alors :

$$y_t|_{\mathcal{F}_{t+\Delta}} - y_t|_{\mathcal{F}_t} = 0 = \hat{\mu}_t|_{\mathcal{F}_{t+\Delta}} - \hat{\mu}_t|_{\mathcal{F}_t} + \hat{\epsilon}_t|_{\mathcal{F}_{t+\Delta}} - \hat{\epsilon}_t|_{\mathcal{F}_t}$$

La révision de l'estimation de tendance est donc égale à $\hat{\epsilon}_t|_{\mathcal{F}_t} - \hat{\epsilon}_t|_{\mathcal{F}_{t+\Delta}}$. Lorsqu'on utilise une décomposition modale empirique, on a donc $\hat{\epsilon}_t|_{\mathcal{F}_t} - \hat{\epsilon}_t|_{\mathcal{F}_{t+\Delta}} = \sum_{i=1}^{\lambda} f_i(t)|_{\mathcal{F}_t} - f_i(t)|_{\mathcal{F}_{t+\Delta}}$ où f_i correspond à la $i^{\text{ème}}$ fonction modale intrinsèque estimée et λ est un seuil permettant l'estimation non-linéaire de la tendance.

Dans la littérature, plusieurs méthodes ont été proposées pour atténuer les effets de bord engendrés par la décomposition modale empirique (voir par exemple [Hua(1998)] [Zen(2004)]). Parmi ces méthodes, on trouve :

- La méthode « symétrique » consistant à répliquer les premières données disponibles de la série de données au niveau des dernières données disponibles. On effectue ensuite la décomposition modale empirique sur l'ensemble de ces données. Cela permet de corriger les fonctions modales intrinsèques au niveau du bord gauche de l'échantillon en tenant compte des propriétés de la

série de départ.

- La méthode « symétrique-oppoé » (ou even-odd) combinant la méthode symétrique à une méthode réflexive consistant à répliquer l'oppoé des données disponibles. Pour chaque méthode, on calcule les enveloppes supérieures et inférieures lors de la décomposition modale empirique. La moyenne des quatre enveloppes correspond alors à la fonction modale intrinsèque.

Néanmoins, ces techniques paraissent très empiriques et peuvent être inadapées dans le cas d'une série non-stationnaire.

Puisque le problème vient de l'interpolation des extremas à l'aide de splines cubiques, qui est une méthode de collocation des extremas, certains ont également pensé à modifier cette méthode d'interpolation en utilisant une méthode d'approximation. On peut ainsi utiliser une moyenne mobile calculée sur les extremas observés par le passé pour construire les enveloppes inférieures et supérieures lors de la décomposition modale empirique [Lin(2009)]. De cette manière, les fonctions modales intrinsèques estimées ne seront pas modifiées lorsque de nouvelles données seront disponibles.

On propose une méthode différente consistant à s'intéresser à la série temporelle des révisions constatées par le passé $(\hat{\epsilon}_t|_{\mathcal{F}_t} - \hat{\epsilon}_t|_{\mathcal{F}_{t+\Delta}})_{t \in [\Delta+1, T-\Delta]}$ pour estimer la révision de l'estimation de tendance.

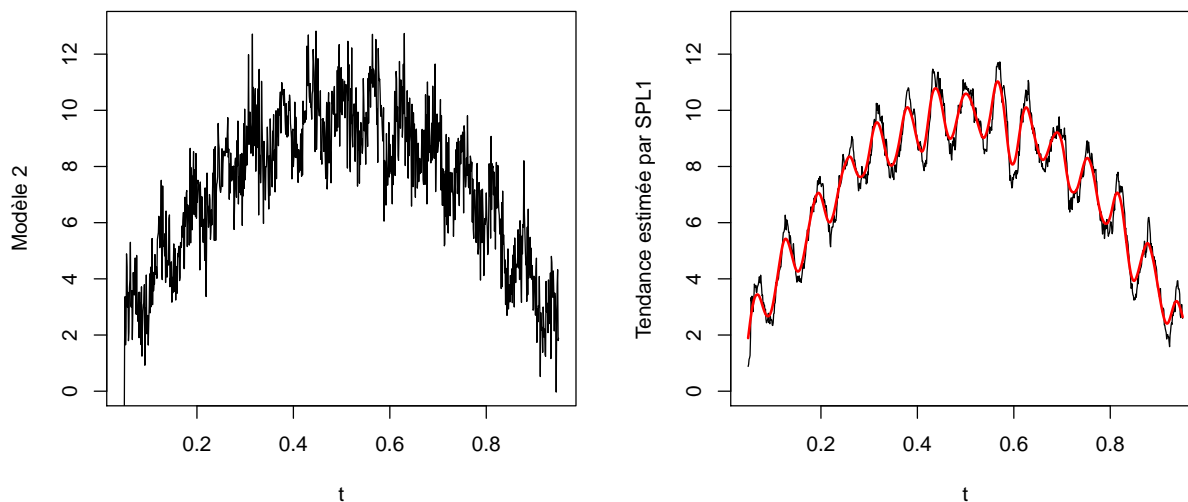


FIGURE 4.19 – Le graphique de droite correspond à une série générée à l'aide du modèle 2 présenté dans le paragraphe. Le graphique de gauche compare $\hat{\mu}_t|_{\mathcal{F}_t}$ (en noir) à $\hat{\mu}_t|_{\mathcal{F}_{t+20}}$ (en rouge)

4.4.2 Prédiction de la révision de l'estimation à l'aide d'un processus ARMA(p, q)

On peut raisonnablement supposer que la série $(\hat{\epsilon}_t|_{\mathcal{F}_t} - \hat{\epsilon}_t|_{\mathcal{F}_{t+\Delta}})_{t \in [\Delta+1, T-\Delta]}$ est stationnaire au second ordre pour un $\Delta > 0$ fixé.

On note $w_T(\Delta) = \hat{\epsilon}_T|_{\mathcal{F}_T} - \hat{\epsilon}_T|_{\mathcal{F}_{T+\Delta}}$. On va supposer que $w_t(\Delta)$ suit un processus ARMA(p, q), ce qui s'exprime comme :

$$w_t(\Delta) - \mu - \phi_1(w_{t-1}(\Delta) - \mu) - \dots - \phi_p(w_{t-p}(\Delta) - \mu) = \delta_t + \gamma_1\delta_{t-1} + \dots + \gamma_q\delta_{t-q}$$

μ correspond à la moyenne du processus $w_t(\Delta)$ et où $(\delta_t)_{t \in [\Delta+1, T-\Delta]}$ est un bruit blanc de variance σ_δ^2 . Cette expression peut se réécrire comme :

$$\phi(L)w_t(\Delta) = \gamma(L)\delta_t$$

où ϕ et γ sont des polynômes de degrés p et q et ont pour coefficients les paramètres ϕ_i et γ_i avec $\phi_0 = \gamma_0 = 1$.

La vraisemblance d'un tel modèle peut être estimée à l'aide d'un filtre de Kalman. Pour cela, il faut estimer une représentation état-espace du processus. On peut par exemple utiliser la représentation état-espace suivante (voir [Ham(1994)]) :

$$w_t(\Delta) = \mu + H^{Tr} \Psi_t \quad \text{avec } \Psi_{t+1} = F\Psi_t + Z_t$$

$$\text{où } H^{Tr} = (1 \quad \gamma_1 \quad \dots \quad \gamma_{r-1}), F = \begin{pmatrix} 1 & \phi_1 & \phi_2 & \dots & \phi_{r-1} & \phi_r \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ \vdots & & \vdots & \ddots & & \vdots \\ 0 & 0 & & 0 \dots & 1 & 0 \end{pmatrix} \text{ et } Z_t = (\delta_{t+1} \quad 0 \quad \dots \quad 0) \text{ et avec}$$

$r = \max(p, q + 1)$. Le lecteur intéressé par le calcul détaillé de cette vraisemblance pourra ensuite se reporter à [Ham(1994)].

Une fois déterminé les paramètres γ_i et ϕ_i maximisant la vraisemblance de ce modèle, les ordres p et q optimaux peuvent se déterminer en minimisant un critère BIC égale à

$$BIC = -2 \ln(L) + k \ln(T)$$

où L représente la vraisemblance du modèle considéré, T est le nombre d'observations et k est le nombre de paramètres libres à estimer. Dans ce cas, on a $k = 2r - 1$.

Une prédiction de $w_T(\Delta)$ peut ensuite être faite en écrivant :

$$w_T(\Delta) = \mu + \phi_1(w_{T-1}(\Delta) - \mu) + \dots + \phi_p(w_{T-p}(\Delta) - \mu) + \delta_T + \gamma_1 \delta_{T-1} + \dots + \gamma_q \delta_{T-q}$$

On a alors

$$\mathbb{E}[w_T(\Delta) | \mathcal{F}_T] = \mu + \phi_1(\mathbb{E}[w_{T-1}(\Delta) | \mathcal{F}_T] - \mu) + \dots + \phi_p(\mathbb{E}[w_{T-p}(\Delta) | \mathcal{F}_T] - \mu)$$

4.5 Application à des indicateurs liés aux croyances des investisseurs

Nous avons précédemment remarqué que les indicateurs PRUS, PRDEF, VIX et ARMS (MM 1m) sont positivement corrélés avec les rendements au carré des marchés actions. Il en est de même pour les opposés des indicateurs AVERS et PMINFL.

On différencie une fois ces indicateurs puis on estime une composante prévisible au sein de l'évolution de ces séries différenciées en soustrayant les 3 premières fonctions modales intrinsèques estimées sur chacune des séries. Pour tous les indicateurs, la composante ainsi estimée a un comportement chaotique qui peut se caractériser par les temps de retard et dimensions de plongement présentés dans le tableau 4.10.

Nous avons utilisé un développement de Taylor à l'ordre 2 pour estimer les prédictions de ces composantes. Afin d'obtenir une fenêtre h et un nombre k de plus proches voisins permettant de

	PRUS	PRDEF	VIX	ARMS (MM 1m)	-AVERS	-PMINFL
$\hat{\tau}$	10	12	9	7	9	8
\hat{d}	8	10	9	10	9	9

TABLEAU 4.10 – Temps de retard et dimensions de plongement estimés pour la composante prévisible en soustrayant les trois premières fonctions modales intrinsèques au sein des séries différenciées pour l’ensemble des indicateurs.

minimiser les erreurs de prédictions, on réutilise la même méthode que précédemment utilisée pour l’attracteur de Ikeda. On estime que pour $h \approx 1.2$ et $k = 15$, les erreurs de prédictions sont les plus faibles. La figure 4.20 compare 100 prédictions aux valeurs obtenues a posteriori pour l’estimation des composantes prévisibles.

Si les composantes prévisibles des variations des différents indicateurs sont toutes négatives, les facteurs de risque sont en phase baissière et on en déduit que le contexte est favorable au marché action. Lorsque la composante prévisible coupe l’axe des abscisses, cela correspond donc à un retournement de direction au sein de la tendance du facteur de risque. On peut ainsi prévoir des retournement au sein des tendances des facteurs de risque.

4.6 Conclusion

Dans ce chapitre, nous avons présenté un certain nombre d’indicateurs dont la signification semble très utile pour un gérant de portefeuille. Nous avons montré qu’un certain nombre de ces indicateurs semblent particulièrement intéressants car ils sont significativement et positivement corrélés avec les rendements au carré de différents indices de prix du marché action. Nous avons présenté une manière originale d’estimer la composante prévisible sous-jacente à l’évolution de ces indicateurs. Cette méthode se base sur une décomposition modale empirique ainsi que sur un test d’imprévisibilité des fonctions modales intrinsèques. On s’assure ainsi qu’il est possible d’utiliser le comportement chaotique de la composante prévisible pour la prédire. Nous avons ensuite présenté une méthode de prédiction non-linéaire de cette composante chaotique se basant sur une régression non-paramétrique polynomiale locale. Plusieurs aspects pratiques ont été discutés : l’estimation des temps de retard et des dimensions de plongement afin de reconstruire l’espace des phases, l’amélioration de la vitesse de recherche de plus proches voisins dans le passé, éviter les problèmes d’inversion de matrices lors de la phase de prédiction. Nous avons également discuté de la correction à apporter aux dernières estimations de tendance disponibles. Nous avons proposé d’estimer cette correction en prédisant, à l’aide d’un modèle ARMA(p, q), les révisions des estimations à partir de celles déjà observées.

4.6. CONCLUSION

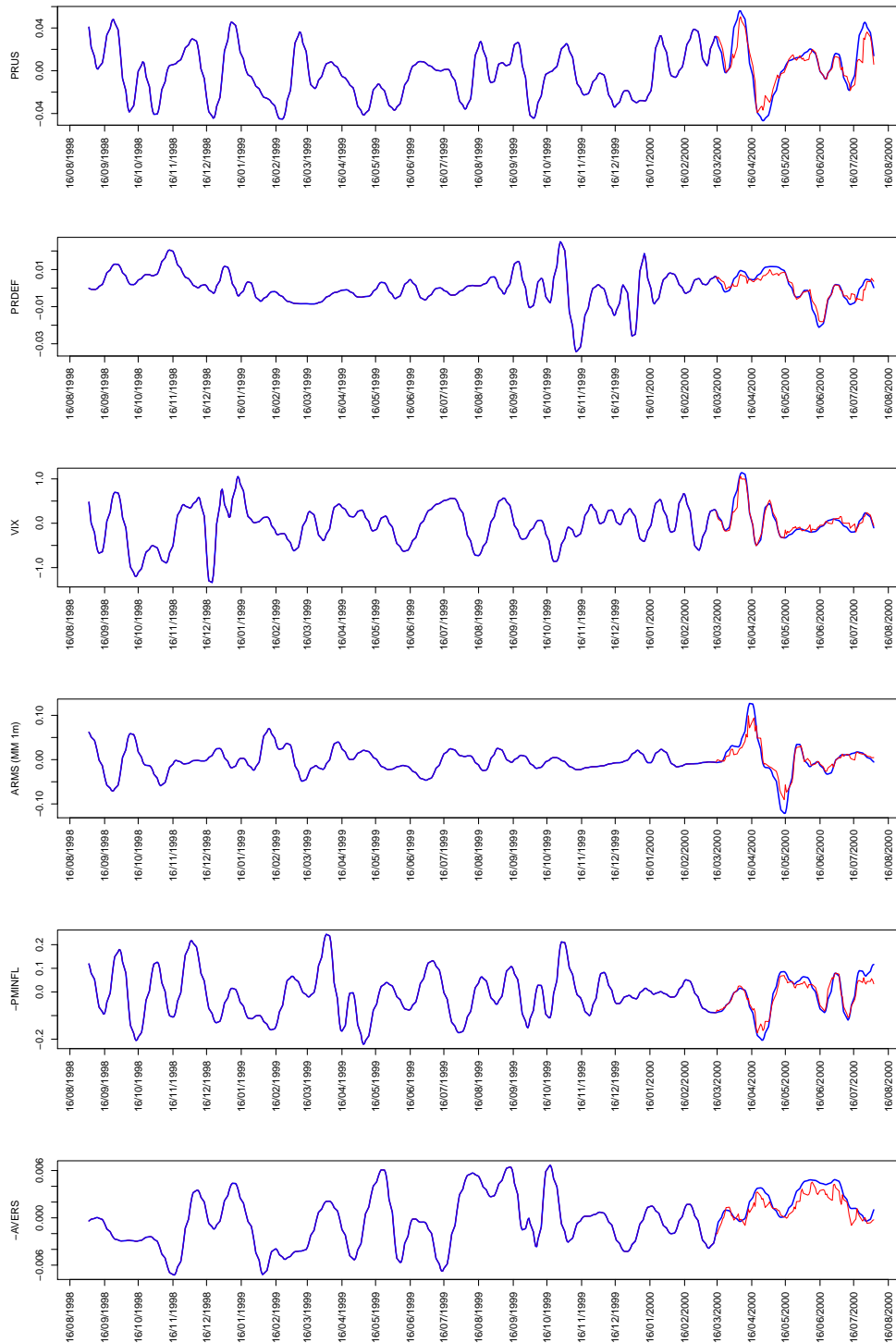


FIGURE 4.20 – La composante prévisible de la série différenciée des différents indicateurs est en bleu. Les prédictions effectuées pour un horizon $\omega = 1$ à partir du 13/03/2000 sont en rouge.

4.6. CONCLUSION

4.6. CONCLUSION

Construire un indicateur synthétique de l'augmentation de facteurs de risque à partir d'une chaîne de Markov cachée

Sommaire

5.1	Coder qualitativement les indicateurs	142
5.2	Fusionner les indicateurs à l'aide d'une chaîne de Markov cachée	143
5.2.1	L'approche de Grégoir et Lenglard	143
5.2.2	Présentation du modèle de chaîne de Markov cachée	144
5.2.3	Présentation de l'algorithme d'Espérance-Maximisation	145
5.2.4	Étape d'espérance et algorithme forward-backward	146
5.2.5	Étape de maximisation	147
5.2.6	Insertion d'une étape Stochastique	148
5.2.7	Construction d'un indicateur synthétique d'augmentation des indicateurs	149
5.3	Applications et résultats	150
5.3.1	Construction d'une horloge d'investissement	150
5.3.2	Construction d'un indicateur de retournement du risque lié aux croyances des investisseurs	154
5.4	Conclusion	154

Introduction

Afin d'anticiper l'évolution des marchés financiers, il faut pouvoir synthétiser l'information macroéconomique et financière de manière synthétique de la manière la plus objective possible. Dans le chapitre précédent, nous avons présenté un certain nombre d'indicateurs du risque lié au marché action, également appelés facteurs de risque. Dans ce chapitre, nous proposons une méthodologie originale permettant de fusionner ces facteurs de risque afin de distinguer le plus clairement possible des phases d'augmentations ou de baisses généralisées de ces facteurs. On espère ainsi résumer l'information disponible à travers des indicateurs synthétiques.

Pour se faire, et compte-tenu des remarques faites dans le chapitre précédent, nous allons considérer trois types d'indicateurs. Le premier est lié à l'information macroéconomique et se base sur des indicateurs avancés de l'économie mondiale ainsi que sur des enquêtes d'opinions telles que l'IFO en Allemagne. Le second est un indicateur lié à l'inflation dans les pays développés. Enfin, le dernier repose sur des indicateurs de risque liés aux croyances des investisseurs. Les indicateurs synthétiques ainsi construits évoluent entre 0 et 1. Lorsqu'ils valent 0 (resp. 1), cela signifie que les facteurs de risque considérés par l'indicateur synthétique sont dans une phase baissière (resp. haussière). Pour qu'ils soient utilisables en temps réel et aient une réelle signification, il faut que ces indicateurs synthétiques ne fluctuent pas trop souvent de manière à ce que, lorsqu'un indicateur indique une phase haussière (ou baissière), il soit fort probable qu'il reste dans cet état les instants suivants. Nous allons donc suivre l'approche faite par Chauvet et Potter [Cha(2000)] consistant à construire des indicateurs avancés du contexte financier à l'aide de facteurs de risque reflétant les croyances des investisseurs.

Un des premiers papiers concernant la construction d'indicateurs avancés de l'économie fut celui de Burns et Mitchell [Bur(1946)]. Le sujet est devenu récurrent en économie appliquée (voir par exemple [Ham(1989)], [Gré(2000)], [Bel(2002)]) car il concerne la distinction des phases d'expansions, de récessions ou de stagnations des économies nationales ou internationales.

Dans ce chapitre, on développe une méthodologie originale pour construire de tels indicateurs. Elle se base sur une chaîne de Markov cachée dont les lois de probabilités associées aux états de la chaîne sont des lois binomiales (on parlera également de lois « d'attaches » aux données). On suppose alors que le niveau d'augmentation ou de baisse des facteurs de risque est généré par ces lois binomiales.

Le chapitre est composé de 4 paragraphes. Le premier paragraphe traite de la transformation des facteurs de risque en variables qualitatives représentant le niveau d'évolution de ces facteurs. Le second présente le modèle de chaîne de Markov cachée utilisé et l'algorithme d'Espérance-Maximisation permettant de l'estimer. Il traite également du calcul de l'indicateur synthétique à partir de ce modèle. Le troisième paragraphe est une application de cette méthodologie afin de construire les trois indicateurs synthétiques désirés.

5.1 Coder qualitativement les indicateurs

La phase de codage consiste à transformer l'évolution des indicateurs en variables qualitatives. On cherche alors à obtenir une série de valeurs entières comprises entre 1 et n_{max} représentant une échelle d'augmentation du facteur de risque considéré. En d'autres termes, cette série vaut 1 lorsque le facteur est en forte baisse et n_{max} lorsqu'il est en forte augmentation. Plusieurs méthodes ont été proposés dans la littérature pour coder les données.

Grégoir et Lengart [Gré(2000)] supposent que ces indicateurs suivent un modèle autorégressif d'ordre p :

$$y_{i,t} = \sum_{k=1}^p \phi_k^i y_{i,t-k} + \epsilon_{i,t}$$

où $y_{i,t}$ est la valeur de l'indicateur $i \in [1, \dots, M]$ à l'instant t et $\epsilon_{i,t}$ un bruit blanc réparti suivant une loi normale centrée. Ces innovations sont ensuite codées par +1 lorsqu'elles sont positives ou -1 lorsqu'elles sont négatives. En effet, lorsque cette innovation est positive, cela signifie que la nouvelle donnée économique disponible est supérieure à son évolution « naturelle » représentée par le modèle autorégressif. On peut donc supposer qu'il s'agit d'une expansion de l'activité économique dans le

cadre des indicateurs précédemment cités. On obtient une nouvelle variable codée $x_{i,t}$ égale à plus ou moins 1 :

$$x_{i,t} = \mathbf{1}_{\epsilon_{i,t} > 0} - \mathbf{1}_{\epsilon_{i,t} \leq 0}$$

Néanmoins, supposer que l'évolution « naturelle » de la série est un modèle autorégressif peut parfois être inadapté. Il existe alors d'autres manières de coder l'évolution des indicateurs. Par exemple, [Bar(2009)] procèdent de la manière suivante :

$$x_{i,t} = \mathbf{1}_{y_{i,t} - y_{i,t-l} > q_{0.5}(l,i)} - \mathbf{1}_{y_{i,t} - y_{i,t-l} \leq q_{0.5}(l,i)}$$

où l correspond à un temps de retard et $q_{0.5}(l,i)$ correspond à la médiane des variations $y_{t,i} - y_{i,t-l}$ constatées par le passé.

5.2 Fusionner les indicateurs à l'aide d'une chaîne de Markov cachée

5.2.1 L'approche de Grégoir et Lenglart

Grégoir et Lenglart [Gré(2000)] supposent qu'il existe une variable aléatoire Z_t suivant une chaîne de Markov à 2 états générant le signe des innovations. En d'autres termes, Z_t est censée décrire l'état de l'économie ($Z_t = 0$ par une contraction et $Z_t = 1$ pour une expansion de l'activité économique). La répartition des $(x_t)_{t \in [1, \dots, T]} = (x_{1,t} \dots x_{M,t})_{t \in [1, \dots, T]}$ est supposée entièrement déterminée par Z_t et les $x_{i,t}$ sont supposés être générés indépendamment les uns des autres :

$$\mathbb{P}(X_t = x_t | Z_t = k) = \prod_{i=1}^M \mathbb{P}(X_{i,t} = x_{i,t} | Z_t = s) \quad (5.1)$$

où $s = 0$ ou 1 . La vraisemblance de ce modèle peut alors s'écrire comme :

$$\mathbb{P}(X_T = x_T, \dots, X_1 = x_1) = \mathbb{P}(X_1 = x_1) \prod_{t=2}^T \mathbb{P}(X_t = x_t | I_{t-1}) \quad (5.2)$$

où $I_{t-1} = \{X_1 = x_1, \dots, X_{t-1} = x_{t-1}\}$. L'enjeu consiste à estimer les probabilités $\mathbb{P}(Z_t = 1 | I_t)$ et $\mathbb{P}(Z_t = 0 | I_t)$, en se basant sur les observations $(x_t)_{t \in [1, \dots, T]}$, afin de savoir dans quel état se trouve l'économie (expansion ou contraction) à un instant t . Hamilton [Ham(1989)] a proposé d'utiliser un algorithme d'Espérance-Maximisation (EM) afin d'estimer une telle chaîne de Markov. Nous discutons plus amplement de cet algorithme dans le paragraphe 5.2.3 de ce chapitre.

Grégoir et Lenglart [Gré(2000)] introduisent également une seconde variable aléatoire, notée W_t , et suivant également une chaîne de Markov à deux états. Cette variable représente l'incertitude sur le codage des données. Elle vaut donc 1 lorsque le codage est significatif et 0 sinon. De cette manière, l'indicateur final à l'instant t , noté Ind_t , vaut :

$$Ind_t = \widehat{\mathbb{P}}(Z_t = 1 | I_t) - \widehat{\mathbb{P}}(Z_t = 0 | I_t)$$

où $\widehat{\mathbb{P}}(Z_t = s | I_t) = \widehat{\mathbb{P}}((Z_t, W_t) = (s, 1) | I_t) + \widehat{\mathbb{P}}((Z_t, W_t) = (s, 0) | I_t)$ pour $s = 0$ ou 1 .

Dans les prochains paragraphes, nous présentons une approche différente de celle de Grégoir et Lenglart [Gré(2000)]. On considère qu'il existe une chaîne de Markov à N états, et non pas seulement 2 états, cachée derrière les données codées et dont chaque état i est associé à une loi

binomiale $\mathcal{B}(n_{max}, p_i)$. On fixe alors les paramètres $(p_i)_{i \in [1, \dots, N]}$ de ces lois binomiales afin de donner une signification à ces états. De cette manière, il est possible d'estimer une probabilité que l'on se trouve dans un des états de la chaîne de Markov générant des données associées à une augmentation des facteurs de risque. Cette probabilité estimée correspond à notre indicateur.

5.2.2 Présentation du modèle de chaîne de Markov cachée

Le principe général d'une chaîne de Markov cachée est de supposer qu'une suite d'observations $(y_t)_{t \in [1, \dots, T]}$ « cache » une suite d'états $(z_t)_{t \in [1, \dots, T]}$ qui correspond à la réalisation d'une variable aléatoire $Z = (Z_t)_{t \in [1, \dots, T]}$ suivant une chaîne de Markov. Un modèle de chaîne de Markov cachée est donc défini par un quintuplet $\lambda = (S, E, A, B, \Pi)$, où

- S est un ensemble de N états.
- $E = [1, \dots, T]$ correspond à l'ensemble d'indexation des observations.
- $A = S \times S$ est la matrice de transition de la chaîne de Markov.
- $B = S \times E$ est la matrice des lois d'attache aux données.
- $\Pi : S \rightarrow [0, 1]$ est le vecteur des distributions initiales.

Toutefois, il est courant de caractériser un tel modèle par le triplet (A, B, Π) . Dans ce cadre, on a :

$$\Pi_i = \mathbb{P}(Z_1 = i) \quad A_{ij} = \mathbb{P}(Z_{t+1} = j | Z_t = i) \quad B_i(y_t) = p_{Y_t | Z_t = i}(y_t)$$

On note $p_{Y_t | Z_t = i}(y_t)$ car si y_t est une variable discrète, $p_{Y_t | Z_t = i}(y_t) = \mathbb{P}(Y_t = y_t | Z_t = i)$. En revanche, si y_t est une variable continue, $p_{Y_t | Z_t = i}(y_t)$ correspond à la densité de Y_t en y_t sachant que l'on est dans l'état i en t , ce qui se note également comme $f_{Y_t | Z_t = i}(y_t)$.

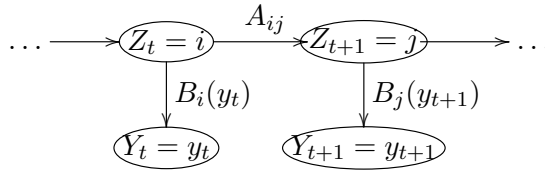


FIGURE 5.1 – Graphique conditionnel d'une chaîne de Markov cachée : chaque flèche représente une probabilité conditionnelle.

L'ensemble des paramètres du modèle λ doit donc répondre aux contraintes stochastiques suivantes :

- $\sum_{j=1}^N A_{ij} = 1$
- $\sum_{j=1}^N \Pi_j = 1$
- $\sum_{y_t \in \mathbb{Y}} B_i(y_t) = 1$ si Y_t est une variable aléatoire discrète et \mathbb{Y} est l'ensemble des valeurs qu'elle peut prendre.
- $\int_{\mathbb{R}} B_i(y) dy = 1$ si Y_t est une variable aléatoire continue.

Dans notre cas, nous avons M indicateurs de risque qui constituent des mesures de l'évolution du risque lié au marché action et nous allons supposer que la variable aléatoire Z_t , suivant la chaîne de Markov, représente un certain niveau d'évolution du risque lié au marché action « dans l'absolu ». Z_t prend N valeurs discrètes et chacune de ces valeurs peut être associée à un niveau global d'augmentation des différents facteurs de risque pris en compte. Nous allons supposer que ces différents indicateurs de risque ont été codés (voir paragraphe 5.1) et constituent donc des observations discrètes.

L'objectif consiste à déterminer le modèle λ permettant de maximiser la vraisemblance du modèle, notée $L_Y(y|\lambda)$ avec $y = (y_t)_{t \in [1, \dots, T]}$ les observations discrètes issues du codage des différents indicateurs. Pour cela, il faut présupposer une loi de probabilité discrète afin d'estimer B . L'algorithme d'Espérance-Maximisation mis au point par [Bau(1966)] va nous permettre d'estimer A , B et Π .

5.2.3 Présentation de l'algorithme d'Espérance-Maximisation

On va faire deux hypothèses essentielles :

- Les variables aléatoires $y = (y_t)_{t \in [1, \dots, T]}$ sont indépendantes entre elles conditionnellement à $Z = (Z_t)_{t \in [1, \dots, T]}$.
- La distribution de y_t conditionnellement à Z est égale à sa distribution conditionnellement à Z_t .

On appelle « vraisemblance complétée par Z » du modèle λ la fonction $L_{Y,Z}(y, z|\lambda)$ définie comme suit :

$$L_{Y,Z}(y, z|\lambda) = \mathbb{P}(Z = z|\lambda) p_{Y|Z=z, \lambda}(y) = \Pi_{z_1} B_{z_1}(y_1) \prod_{t=2}^T A_{z_{t-1}, z_t} B_{z_t}(y_t)$$

où z correspond à une réalisation de Z .

On obtient ce résultat à l'aide des deux hypothèses précédentes puisqu'elles impliquent que :

$$p_{Y|Z=z, \lambda}(y) = \prod_{t=1}^T p_{Y_t|Z_t=z_t, \lambda}(y_t) = \prod_{t=1}^T B_{z_t}(y_t)$$

La vraisemblance complétée par Z du modèle λ est une fonction de y réelle, positive, mesurable et intégrable en y pour λ fixé. On peut donc écrire que la vraisemblance du modèle s'écrit comme

$$L_Y(y|\lambda) = \int_{z \in \mathbb{E}^Z} L_{Y,Z}(y, z|\lambda) d\mu(z)$$

où \mathbb{E}^Z correspond à l'ensemble des vecteurs z possibles ($Card(\mathbb{E}^Z) = N^T$). Z étant une variable aléatoire discrète, on utilise une mesure de comptage, ce qui nous mène à

$$L_Y(y|\lambda) = \sum_{z \in \mathbb{E}^Z} L_{Y,Z}(y, z|\lambda) = \sum_{z \in \mathbb{E}^Z} \Pi_{z_1} B_{z_1}(y_1) A_{z_1, z_2} B_{z_2}(y_2) \dots A_{z_{T-1}, z_T} B_{z_T}(y_T)$$

Malheureusement, le calcul de cette vraisemblance est très fastidieux en pratique (de l'ordre de $2T \times N^T$ opérations). C'est pourquoi, Baum et al. [Bau(1970)] introduisent une fonction auxiliaire Q donnant l'espérance, par rapport à Z , du logarithme de la vraisemblance complétée d'un modèle λ_* sachant les données observées y et le modèle λ . Puisque z ne peut être observée, cette espérance remplacera le logarithme de la vraisemblance complétée des données :

$$Q(\lambda, \lambda_*) = \mathbb{E}_Z[\log(L_{Y,Z}(y, z|\lambda_*)|Y = y, \lambda)] = \sum_{z \in \mathbb{E}^Z} \log(L_{Y,Z}(y, z|\lambda_*) L_{Y,Z}(y, z|\lambda))$$

Baum et al. [Bau(1970)] montrent alors le théorème suivant :

Théorème 3. *Soit y une série d'observations fixée. Si $Q(\lambda, \lambda_*) \geq Q(\lambda, \lambda)$ alors $L_Y(y|\lambda_*) \geq L_Y(y|\lambda)$. L'inégalité étant stricte sauf si $L_{Y,Z}(y, z|\lambda) = L_{Y,Z}(y, z|\lambda_*) d\mu(z)$ -presque partout.*

Nous avons discuté de la preuve de ce théorème dans [Gia(2007)]. Chercher à maximiser $L_Y(y|\lambda)$ en λ revient donc à chercher le λ_* maximisant $Q(\lambda, \lambda_*)$. L'algorithme d'Espérance-Maximisation trouve son fondement dans ce théorème. En effet, notons λ^r le modèle estimé à la $r^{\text{ème}}$ itération de cet algorithme. L'algorithme EM se déroule en deux étapes :

- Étape E : calcul de $Q(\lambda^{r-1}, \lambda)$. Cela revient à estimer $L_{Y,Z}(y, z|\lambda^{r-1})$.
- Étape M : choix de λ^r tel que $\lambda^r = \arg \max_{\lambda} Q(\lambda^{r-1}, \lambda)$.

Baum et al. [Bau(1970)] montrent que cet algorithme permet d'augmenter le log-vraisemblance du modèle à travers la propriété suivante :

Propriété 1.

Soit λ^{r-1} défini un point critique de $L_Y(y|\lambda)$ et $\lambda^r = \lambda^{r-1}$.

Soit $L_Y(y|\lambda^r) \geq L_Y(y|\lambda^{r-1})$.

Ainsi, l'algorithme d'Espérance-Maximisation permet d'augmenter la valeur du log-vraisemblance jusqu'à ce qu'il atteigne un point critique. Toutefois, on ne sait pas si le modèle atteint constitue un point de selle, un maximum local ou global du log-vraisemblance. De ce fait, la convergence vers un maximum global dépend essentiellement de l'initialisation de l'algorithme, c'est-à-dire de λ_0 . Il faut également noter que la fonction de vraisemblance n'est pas bornée, l'algorithme EM peut donc dégénérer vers une solution singulière [Eve(1981)].

Une manière d'éviter ce problème de convergence de l'algorithme EM consiste à insérer une étape dite « Stochastique » dans l'algorithme. Une fois l'étape d'Espérance effectuée, on génère une suite d'états z_g à l'aide de la chaîne de Markov estimée par le modèle λ^{r-1} . On produit alors un échantillon dit complété (y, z_g) . On passe ensuite à l'étape M en calculant l'estimateur du maximum de vraisemblance λ^r sur la base de l'échantillon (y, z_g) (voir paragraphe 5.2.6).

5.2.4 Étape d'espérance et algorithme forward-backward

Le but de l'étape d'espérance est d'estimer les probabilités conjointes ($\mathbb{P}(Z_t = i, Z_{t+1} = j|Y = y)$) et marginales ($\mathbb{P}(Z_t = i|Y = y)$) a posteriori. L'algorithme forward-backward permet d'estimer ces probabilités. Il décompose la série d'observations en deux sous-séries : une première allant de 1 à t en attribuant l'état i à Z_t (probabilités « forward ») et la seconde allant de $t + 1$ à T sachant que l'on se trouve dans l'état i en t (probabilités « backward »). Pour éviter d'obtenir des probabilités trop petites pour la capacité d'un ordinateur (problème d'*underflow*), on utilise les probabilités forward et backward modifiées suivantes (introduites par [Dev(1985)]) :

- Probabilités forward :

$$\alpha_t(i) = \mathbb{P}(Z_t = i | (Y_1, \dots, Y_t) = (y_1, \dots, y_t))$$

- Probabilités backward :

$$\beta_t(i) = \frac{p_{Y_{t+1}, \dots, Y_T | Z_t = i}(y_{t+1}, \dots, y_T)}{p_{Y_{t+1}, \dots, Y_T | (Y_1, \dots, Y_t) = (y_1, \dots, y_t)}(y_{t+1}, \dots, y_T)}$$

On rappelle que l'on note $p_{Y_{t+1}, \dots, Y_T | Z_t = i}(y_{t+1}, \dots, y_T)$ avec un petit p car cette quantité peut soit être une probabilité si y_t est une série de valeurs discrètes, soit être la valeur d'une densité si y_t est une série à valeurs réelles. L'algorithme forward-backward fonctionne alors de la manière suivante :

- Étape 1 :

$$\alpha_1(i) = \frac{\Pi_i B_i(y_1)}{\sum_{j=1}^N \Pi_j B_j(y_1)} \quad \beta_T(i) = 1$$

- Étape t ($\forall t \in [1, \dots, T-1]$) :

$$\alpha_{t+1}(i) = \frac{B_i(y_{t+1}) \sum_{j=1}^N \alpha_t(j) A_{ji}}{\sum_{l=1}^N [B_l(y_{t+1}) \sum_{j=1}^N \alpha_t(j) A_{jl}]} \quad \beta_t(i) = \frac{\sum_{j=1}^N B_j(y_{t+1}) \beta_{t+1}(j) A_{ij}}{\sum_{l=1}^N [B_l(y_{t+1}) \sum_{j=1}^N \alpha_t(j) A_{jl}]}$$

Une démonstration de ces formules a été donnée dans [Gia(2007)]. On peut alors obtenir les probabilités conjointes a posteriori à l'aide des probabilités forward et backward :

$$\forall t \in [1, \dots, T-1], \quad \Psi_t(i, j) = \mathbb{P}(Z_t = i, Z_{t+1} = j | Y = y) \quad (5.3)$$

$$= \frac{\alpha_t(i) A_{ij} B_j(y_{t+1}) \beta_{t+1}(j)}{\sum_{l=1}^N [B_l(y_{t+1}) \sum_{j=1}^N \alpha_t(j) A_{jl}]} \quad (5.4)$$

Ce qui nous permet de déduire les probabilités marginales a posteriori suivantes :

$$\forall t \in [1, \dots, T-1], \quad \gamma_t(i) = \mathbb{P}(Z_t = i | Y = y) = \sum_{j=1}^N \Psi_t(i, j)$$

5.2.5 Étape de maximisation

Réestimation de Π_i et A_{ij} La méthode du multiplicateur de Lagrange, en utilisant les différentes contraintes stochastiques précédemment évoquées pour Π et A , permet d'obtenir les formules suivantes de réestimation (voir [Gia(2007)]) :

$$\Pi_i^r = \frac{1}{T} \sum_{t=1}^T \gamma_t(i) \quad A_{ij}^r = \frac{\sum_{t=1}^{T-1} \Psi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)}$$

Réestimation des lois d'attaches aux données Pour les lois d'attaches aux données, on va supposer que les données discrètes issues des M indicateurs de risque sont générées par des lois binomiales de paramètres $p_1 \dots p_N$ différents pour chaque état de la chaîne de Markov cachée. En supposant une hypothèse d'indépendance entre les réalisations des variables $Y_{1,t}, \dots, Y_{M,t}$, on a :

$$B_i(y_t) = \mathbb{P}(Y_t = y_t | Z_t = i) = \mathbb{P}(Y_{1,t} = y_{1,t}, \dots, Y_{M,t} = y_{M,t} | Z_t = i) \quad (5.5)$$

$$= \mathbb{P}(Y_{1,t} = y_{1,t} | Z_t = i) \mathbb{P}(Y_{2,t} = y_{2,t} | Z_t = i) \dots \mathbb{P}(Y_{M,t} = y_{M,t} | Z_t = i) \quad (5.6)$$

On peut alors supposer que, pour tout $m \in [1, \dots, M]$, $\mathbb{P}(Y_{m,t} = y_{m,t} | Z_t = i)$ correspond à la réalisation d'une loi binomiale de paramètre p_i . Pour réestimer les paramètres p_1, \dots, p_N , on dérive la fonction $\hat{Q}(\lambda^{r-1}, \lambda)$.

On peut remarquer que $\frac{\partial}{\partial p_i} \log(B_i(y_t)) = \sum_{m=1}^M \frac{y_{m,t}}{p_i} - \frac{n_{max} - y_{m,t}}{1 - p_i} = \sum_{m=1}^M \frac{y_{m,t} - p_i n_{max}}{p_i(1 - p_i)}$ où n_{max} correspond à la valeur maximale que peut prendre $y_{m,t}$. On a alors

$$\frac{\partial}{\partial p_{k,i}} [\widehat{Q}(\lambda^{r-1}, \lambda)] = \sum_{z \in \mathbb{E}^Z} L_{Y,Z}(y, z | \lambda^{r-1}) \times \sum_{t=1}^T \sum_{m=1}^M \frac{y_{m,t} - p_i n_{max}}{p_i(1-p_i)} \mathbf{1}_{z_t=i} = 0 \quad (5.7)$$

$$\Leftrightarrow \sum_{t=1}^T \mathbb{P}(Z_t = i | Y = y, \lambda^{r-1}) \times \sum_{m=1}^M y_{m,t} - p_i n_{max} = 0 \quad (5.8)$$

$$\Leftrightarrow \sum_{t=1}^T \gamma_t(i) \times \sum_{m=1}^M y_{m,t} - p_i n_{max} = 0 \quad (5.9)$$

Ce qui nous permet d'arriver au résultat suivant :

$$p_i^r = \frac{\sum_{t=1}^T \gamma_t(i) \sum_{m=1}^M y_{m,t}}{\sum_{t=1}^T \gamma_t(i) M n_{max}}$$

5.2.6 Insertion d'une étape Stochastique

L'insertion d'une étape Stochastique entre l'étape d'Espérance et celle de Maximisation va permettre de faire converger en loi le modèle stochastique estimé vers le modèle dont le log-vraisemblance correspond au maximum global.

Supposons que $\gamma_t(i)$ soit soumis aux contraintes stochastiques suivantes :

- $\forall i \in [1, \dots, N], \quad 0 < \gamma_t(i) < 1$
- $\sum_{i=1}^N \gamma_t(i) = 1$

Posons $e(y_t) = \{e_i(y_t) | i \in [1, \dots, N]\}$ avec $e_i(y_t) = \mathbf{1}_{z_t=i}$. Cette étape consiste à générer un certain nombre de $\widehat{e}(y_t)$ à l'aide d'une loi multinomiale de paramètres les $(\gamma_t(i))_{i \in [1, \dots, N]}$:

$$\mathbb{P}(e(y_t) = (e_1(y_t), \dots, e_N(y_t))) = \frac{(e_1(y_t) + \dots + e_N(y_t))!}{e_1(y_t)! \dots e_N(y_t)!} \gamma_t^{e_1(y_t)}(1) \dots \gamma_t^{e_N(y_t)}(N)$$

Les réalisations $\widehat{e}(y_t)$ générées à l'aide de cette loi définissent une partition Q_1, \dots, Q_N de l'échantillon y_1, \dots, y_T où $Q_i = \{y_t | \widehat{e}_i(y_t) = 1\}$.

Il est alors possible, si $\gamma_t(i)$ est très faible pour un certain état i , qu'une de ces partitions constitue un ensemble vide. Dans ce cas, un des états de la chaîne de Markov est inutile puisqu'aucune des observations ne lui est liée. Cet algorithme permet donc de sélectionner le nombre d'états de la chaîne de Markov maximisant le log-vraisemblance, ce que ne fait pas l'algorithme d'Espérance-Maximisation « basique ». Il suffit de préciser un nombre maximal d'états de la chaîne de Markov que l'on souhaite estimer lors de l'initialisation de l'algorithme pour obtenir un nombre d'état optimal.

En pratique, il est parfois souhaitable d'avoir tout de même un certain nombre d'états. Puisqu'on suppose que $\forall i \in [1, \dots, N], \quad 0 < \gamma_t(i) < 1$, il suffit de sélectionner un $\epsilon > 0$ suffisamment petit de façon à imposer, dans l'algorithme, que $\epsilon < \gamma_t(i) < 1 - \epsilon$ tout en respectant la deuxième contrainte stochastique. Ensuite, on recommence la génération des $\widehat{e}(y_t)$ tant qu'une des partitions Q_i est vide ou ne contient pas suffisamment de données.

Si on insère une étape Stochastique, le vecteur des probabilités stationnaires et les paramètres des lois binomiales peuvent se réestimer à l'aide des vecteurs $\widehat{e}(y_t)$:

$$\Pi_i^r = \frac{1}{T} \sum_{t=1}^T \widehat{e}_i(y_t) \quad p_i^r = \frac{\sum_{t=1}^T \widehat{e}_i(y_t) \sum_{m=1}^M y_{m,t}}{M n_{max} \sum_{t=1}^T \widehat{e}_i(y_t)}$$

La réestimation de la matrice de transition A peut être faite sans prendre en compte les vecteurs $\hat{e}(y_t)$.

5.2.7 Construction d'un indicateur synthétique d'augmentation des indicateurs

On suppose que l'on possède M facteurs de risque qui ont été codés de 1 à n_{max} , avec n_{max} un nombre pair. On suppose que si l'indicateur vaut $n > \frac{n_{max}}{2}$ (resp. $n \leq \frac{n_{max}}{2}$) alors le facteur de risque considéré est en hausse (resp. en baisse). On suppose également qu'il existe une chaîne de Markov composée de N états cachés derrière les valeurs des indicateurs codés. Afin de construire un indicateur ayant du sens, il peut être utile de donner une signification à chacun des états de la chaîne de Markov. Pour cela, on peut fixer les paramètres p_i des lois binomiales de façon à caractériser les états.

La fonction de densité d'une loi binomiale $f : p \mapsto \binom{n_{max}}{l} p^l (1-p)^{n_{max}-l}$ est une fonction croissante si $n_{max}p \leq l$ et décroissante dans le cas contraire. En effet, on a

$$f'(p) = \binom{n_{max}}{l} (l - n_{max}p) p^{l-1} (1-p)^{n_{max}-l-1}$$

qui est positive (resp. négative) si et seulement si $n_{max}p \leq l$ (resp. $n_{max}p \geq l$). Par conséquent, si $p_i < p_{i+1} \leq \frac{l}{n_{max}}$ ou si $p_i > p_{i+1} \geq \frac{l}{n_{max}}$, on a

$$\forall m \in [1, \dots, M], \quad \mathbb{P}(Y_{m,t} = l | Z_t = i) < \mathbb{P}(Y_{m,t} = l | Z_t = i + 1)$$

Dans le cas où $l = n_{max}$ et si $p_i < p_{i+1}$ pour tout $i \in [1, \dots, N-1]$ on a

$$\forall i \in [1, \dots, N-1], \quad \forall m \in [1, \dots, M], \quad \mathbb{P}(Y_{m,t} = n_{max} | Z_t = i) < \mathbb{P}(Y_{m,t} = n_{max} | Z_t = i + 1)$$

La figure 5.2 représente l'évolution des différentes probabilités $\mathbb{P}(Y_{m,t} = n | Z_t = i)$ en fonction de p_i et pour $n \in [1, \dots, n_{max}]$ avec $n_{max} = 2$ et 4.

On peut alors fixer une valeur $q_i = p_i^{n_{max}} = \mathbb{P}(Y_{m,t} = n_{max} | Z_t = i)$. Pour que l'état i soit caractérisé par de fortes probabilités que les indicateurs soient égaux à n_{max} , on peut donc fixer $p_i = 0.999^{\frac{1}{n_{max}}}$ au sein de l'algorithme. En revanche, si on veut que cet état ou un autre soit caractérisé par de faibles probabilités que les indicateurs soient égaux à n_{max} , on peut fixer $p_i = 0.001^{\frac{1}{n_{max}}}$. Une manière de faire pour fixer les p_i pour $i \in [1, \dots, N]$, consiste à fixer

$$p_i = 0.001^{\frac{1}{n_{max}}} + (i-1) \times \frac{0.999^{\frac{1}{n_{max}}} - 0.001^{\frac{1}{n_{max}}}}{N-1}$$

de façon à avoir des p_i équirépartis entre $0.001^{\frac{1}{n_{max}}}$ et $0.999^{\frac{1}{n_{max}}}$.

Pour construire un indicateur synthétique d'augmentation des indicateurs, on va utiliser la probabilité que l'état i de la chaîne de Markov associé à l'observation à l'instant t soit un état caractérisé par une augmentation des indicateurs, c'est-à-dire que

$$\forall m \in [1, \dots, M], \quad \mathbb{P}(Y_{m,t} = \frac{n_{max}}{2} + 1 \cup Y_{m,t} = \frac{n_{max}}{2} + 2 \cup \dots \cup Y_{m,t} = n_{max} | Z_t = i) > 0.5$$

ce qui peut également s'écrire comme :

$$\forall m \in [1, \dots, M], \quad \sum_{n=\frac{n_{max}}{2}+1}^{n_{max}} \mathbb{P}(Y_{m,t} = n | Z_t = i) > 0.5$$

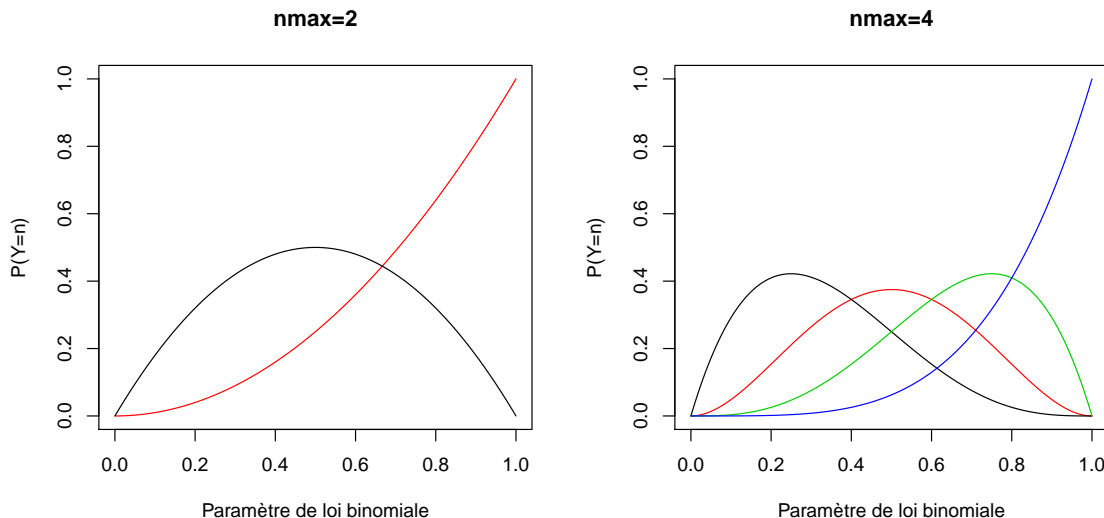


FIGURE 5.2 – Soit Y une variable suivant une loi binomiale $\mathcal{B}(n_{max}, p)$. Les graphiques ci-dessus représentent la probabilité $\mathbb{P}(Y = n)$ pour n compris entre 1 et n_{max} (différentes couleurs) et pour p variant entre 0 et 1 (en abscisse). Le graphique de gauche représente les probabilités $\mathbb{P}(Y = 1)$ (courbe noire) et $\mathbb{P}(Y = 2)$ (courbe rouge) pour $n_{max} = 2$ et celui de droite les probabilités $\mathbb{P}(Y = 1)$ (courbe noire), $\mathbb{P}(Y = 2)$ (courbe rouge), $\mathbb{P}(Y = 3)$ (courbe verte) et $\mathbb{P}(Y = 4)$ (courbe bleue) pour $n_{max} = 4$.

On note $\sum_{n=\frac{n_{max}}{2}+1}^{n_{max}} \mathbb{P}(Y_{m,t} = n | Z_t = i) = S_i^a$. L'indicateur correspond alors à la probabilité suivante :

$$I_t^a = \mathbb{P}(Z_t = i \cup \dots \cup Z_t = j | Y = y, S_i^a > 0.5, \dots, S_j^a > 0.5) = \sum_{i=1}^N \gamma_t(i) \times \mathbf{1}_{S_i^a > 0.5}$$

On peut également construire un indicateur synthétique de baisse des indicateurs en raisonnant d'une manière similaire :

$$I_t^b = \mathbb{P}(Z_t = i \cup \dots \cup Z_t = j | Y = y, S_i^b > 0.5, \dots, S_j^b > 0.5) = \sum_{i=1}^N \gamma_t(i) \times \mathbf{1}_{S_i^b > 0.5}$$

où $S_i^b = \sum_{n=1}^{\frac{n_{max}}{2}} \mathbb{P}(Y_{k,t} = n | Z_t = i)$. Néanmoins, puisque $S_i^b = 1 - S_i^a$, $I_t^b = 1 - I_t^a$.

5.3 Applications et résultats

5.3.1 Construction d'une horloge d'investissement

Afin de construire une horloge d'investissement, on va utiliser deux indicateurs : un indicateur du contexte économique mondial et un indicateur de l'inflation dans les pays développés.

Pour l'indicateur du contexte économique mondial, on utilise les indicateurs macroéconomiques suivants :

- Indicateurs avancés de l'économie (Composite Leading Indicators) dans un certain nombre de pays représentatifs de l'économie mondiale, choisis d'après l'importance de leur PIB : États-Unis, Japon, Chine, Allemagne, France, Royaume-Uni, Italie, Espagne, Canada, Australie.
- Indicateurs avancés d'enquêtes d'opinions : ISM Purchasing Managers (États-Unis), IFO (Allemagne).

Ces données sont de fréquence mensuelle et ont été recueillies sous Datastream. Elles vont d'Avril 1991 à Décembre 2009.

On utilise une méthode inspirée de celle de Bardaji et Tallet [Bar(2009)] consistant à créer la série $x_{i,t}$ telle que :

$$x_{i,t} = \mathbf{1}_{y_{i,t}-y_{i,t-l}>q_{0.75}(l,i)} + \mathbf{1}_{y_{i,t}-y_{i,t-l}>q_{0.5}(l,i)} + \mathbf{1}_{y_{i,t}-y_{i,t-l}>q_{0.25}(l,i)} + 1$$

où l correspond à un temps de retard et $q_{\alpha}(l, i)$ correspond au quantile empirique de niveau α évalué sur les variations $y_{i,t} - y_{i,t-l}$ constatées par le passé. l est un temps de retard de 3 mois. On a donc $n_{max} = 4$. On suppose qu'il existe une chaîne de Markov à $N = 9$ états cachée derrière ces données codées et dont les lois de probabilités associées aux états sont des lois binomiales de paramètres $(p_k)_{k \in [1, \dots, 9]}$ tels que

$$\forall k \in [1, \dots, 9], \quad p_k = 0.001^{\frac{1}{4}} + (k - 1) \times \frac{0.999^{\frac{1}{4}} - 0.001^{\frac{1}{4}}}{8}$$

On applique l'algorithme d'Espérance-Maximisation en fixant ces paramètres p_i pour estimer la chaînes de Markov cachée. On obtient la matrice de transition suivante :

$$A = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.703 & 0.297 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.05 & 0.765 & 0.184 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.087 & 0.721 & 0.137 & 0 & 0.054 & 0 & 0 \\ 0 & 0 & 0 & 0.201 & 0.620 & 0.179 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.266 & 0.583 & 0.151 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.153 & 0.737 & 0.11 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.291 & 0.651 & 0.058 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.308 & 0.692 \end{pmatrix}$$

Et le vecteur des probabilités initiales suivant :

$$\Pi = \begin{pmatrix} 0 \\ 0.0165 \\ 0.0976 \\ 0.2059 \\ 0.1963 \\ 0.1717 \\ 0.2194 \\ 0.0780 \\ 0.0144 \end{pmatrix}$$

On calcul ensuite l'indicateur synthétique d'augmentation des indicateurs qui est présenté dans la figure 5.3.

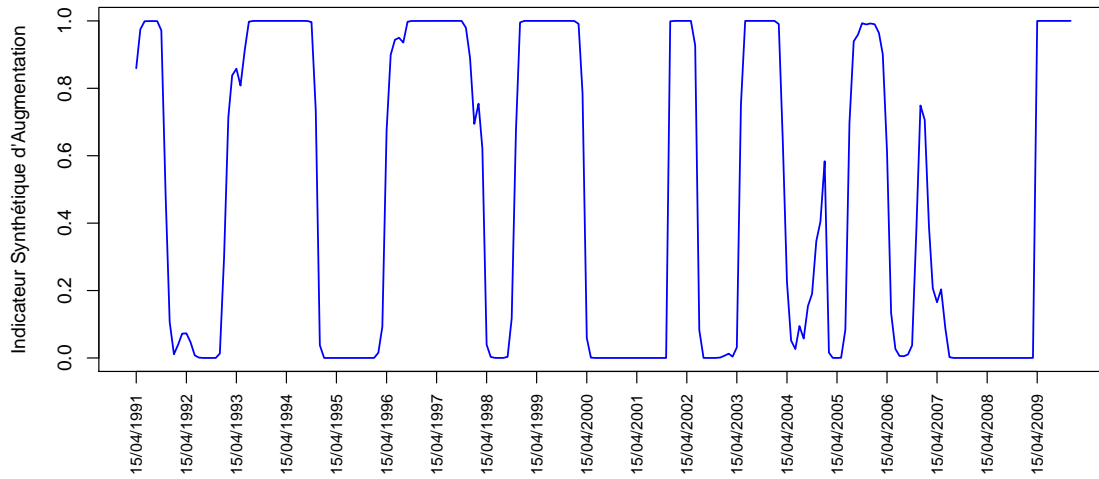


FIGURE 5.3 – Indicateur synthétique d’augmentation des indicateurs macroéconomiques du contexte économique mondial

L’indicateur semble cohérent avec l’indicateur proposé par Eurostat (voir figure 4.1) : il est non nul au niveau des périodes où l’indicateur proposé par Eurostat est supérieur à 0.75, ce qui est synonyme d’une hausse généralisée des indicateurs pris en compte par Eurostat. Notre indicateur possède néanmoins l’avantage d’être beaucoup moins volatile et reste peu entre 0 et 1. Lorsqu’il est proche de 1, cela signifie que les indicateurs macroéconomiques considérés sont en phase de hausse, et donc que l’on est dans une période de croissance économique.

Dans un second temps, afin de construire un indicateur synthétique de l’inflation, on utilise les indicateurs suivant :

- Indicateurs des prix à la consommation aux Etats-Unis, en France et en Allemagne.
- Indicateur du point mort d’inflation aux Etats-Unis.

Les données sont de fréquence mensuelle et recueillies à l’aide de Datastream entre Avril 1998 et Décembre 2009.

On suppose également qu’il existe une chaîne de Markov à 9 états cachée derrière les données codées. On applique l’algorithme d’Espérance-Maximisation en fixant les paramètres p_i des lois binomiales associées aux états de la chaîne de Markov cachée de la même manière que précédemment. L’indicateur d’augmentation obtenu est présenté dans la figure 5.4. On obtient la matrice de transition suivante :

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.758 & 0 & 0.242 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.629 & 0 & 0.371 & 0 & 0 & 0 \\ 0 & 0 & 0.249 & 0.682 & 0.068 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.608 & 0.336 & 0 & 0.056 \\ 0 & 0 & 0 & 0 & 0.512 & 0 & 0.488 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.308 & 0.692 \end{pmatrix}$$

Et le vecteur des probabilités initiales suivant :

$$\Pi = \begin{pmatrix} 0 \\ 0.1309 \\ 0.0255 \\ 0.2669 \\ 0.0987 \\ 0.2453 \\ 0.1557 \\ 0.0513 \\ 0.0256 \end{pmatrix}$$

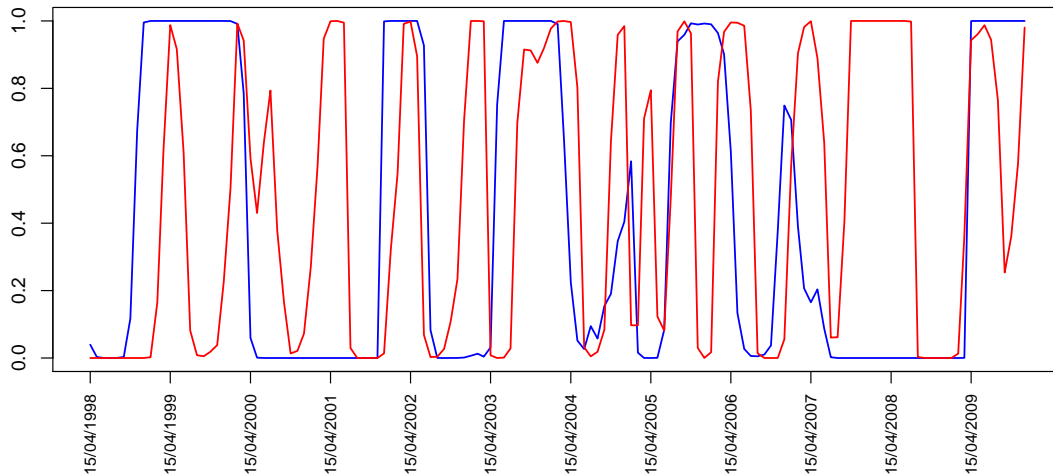


FIGURE 5.4 – Indicateur synthétique d’augmentation de l’inflation (en rouge) comparé à celui du contexte économique mondial.

On peut alors construire une horloge d’investissement en reprenant l’idée présentée dans le paragraphe 4.1.1. Pour cela, il suffit de poser :

$$\theta_t = \begin{cases} \arcsin\left(\frac{I_t^i - 0.5}{\sqrt{(I_t^c - 0.5)^2 + (I_t^i - 0.5)^2}}\right) & \text{si } I_t^i > 0.5 \\ \frac{\pi}{2} + \arcsin\left(\frac{I_t^c - 0.5}{\sqrt{(I_t^c - 0.5)^2 + (I_t^i - 0.5)^2}}\right) & \text{si } I_t^i < 0.5 \text{ et si } I_t^c > 0.5 \\ -\frac{\pi}{2} + \arcsin\left(\frac{I_t^c - 0.5}{\sqrt{(I_t^c - 0.5)^2 + (I_t^i - 0.5)^2}}\right) & \text{si } I_t^i < 0.5 \text{ et si } I_t^c < 0.5 \end{cases}$$

où $I_t^c \neq 0.5$ correspond à la valeur de l’indicateur synthétique d’augmentation du contexte économique mondial et $I_t^i \neq 0.5$ celui de l’indicateur synthétique d’augmentation de l’inflation. On peut alors faire un parallélisme entre l’angle θ_t défini sur le cercle trigonométrique et l’angle de l’aiguille sur l’horloge d’investissement présentée dans la figure 4.2.

Dans le cas où $I_t^c = 0.5$ ou dans le cas où $I_t^i = 0.5$, on est dans l’incertitude concernant l’évolution d’un des indicateurs : on ne peut pas donner une direction précise sur l’horloge d’investissement.

5.3.2 Construction d'un indicateur de retournement du risque lié aux croyances des investisseurs

Afin de construire un indicateur de retournement du risque lié aux croyances des investisseurs, on utilise les facteurs de risque suivants :

- La différence entre les rendements en dividendes moyens sur le marché action US et les taux d'obligations d'état US à maturité 10 ans.
- La différence entre les taux moyens des obligation d'entreprises US notées BAA et celles notées AAA.
- L'indice de volatilité implicite VIX.
- Un indicateur d'aversion au risque des investisseurs estimé d'après une fenêtre glissante de 6 mois.
- L'opposé de l'indicateur ARMS estimé à l'aide d'une moyenne mobile de 1 mois.
- L'opposé du point mort d'inflation aux USA.

Ces indicateurs sont de fréquence quotidienne et ont été recueillis à l'aide de Datastream. Ils vont du 16 Avril 1998 au 31 Octobre 2010.

L'ensemble de ces indicateurs ont des corrélations significativement positives avec les rendements au carré du marché action, que ce soit au niveau quotidien, hebdomadaire ou mensuel. Afin de coder ces indicateurs, on commence par les différencier puis on estime une composante prévisible au sein de ces séries différenciées à l'aide d'une décomposition modale empirique. On note $\hat{m}_{i,t}$ la valeur de cette composante prévisible évaluée pour l'indicateur i à l'instant t . Ces séries sont ensuite codées par :

$$x_{i,t} = \mathbf{1}_{\hat{m}_{i,t} > q_{0.75}(i)} + \mathbf{1}_{\hat{m}_{i,t}^i > q_{0.5}(i)} + \mathbf{1}_{\hat{m}_{i,t} > q_{0.25}(i)} + 1$$

où $q_\alpha(i)$ correspond au quantile de la répartition des $\hat{m}_{i,t}$ au niveau α .

On estime ensuite, de même que pour les précédents indicateurs, une chaîne de Markov cachée à 9 états dont les paramètres p_k des lois binomiales ont été fixés. La figure 5.5 présente différents indicateurs synthétiques d'augmentation obtenus à partir de différentes estimations de tendance au sein des séries d'indicateurs différenciées.

On constate que si la composante prévisible estimée possède une bande de fréquence trop élevée, l'indicateur synthétique d'augmentation obtenu fluctue trop rapidement pour qu'il constitue une information significative pour l'investisseur. En revanche, il est possible d'obtenir un indicateur cohérent si on choisit une composante prévisible fluctuant « plus doucement », c'est-à-dire en soustrayant aux données de départ des fonctions modales intrinsèques supplémentaires.

5.4 Conclusion

Dans ce chapitre, nous avons présenté une méthodologie originale permettant de fusionner l'évolution de différents indicateurs pour n'en retenir qu'un seul correspondant à une probabilité que les indicateurs soient, d'une manière générale, en phase haussière ou baissière. Cette méthodologie se base sur un modèle de chaînes de Markov cachée dont les lois d'« attache » aux données sont des lois binomiales. Les réalisations de ces lois binomiales correspondent aux degrés d'augmentation des indicateurs fusionnés. Nous avons ainsi montré qu'il est possible de construire des indicateurs synthétiques de l'augmentation de l'inflation ou du contexte économique mondial mesuré à l'aide de certaines variables macroéconomiques. Ces indicateurs peuvent notamment servir à construire une

5.4. CONCLUSION

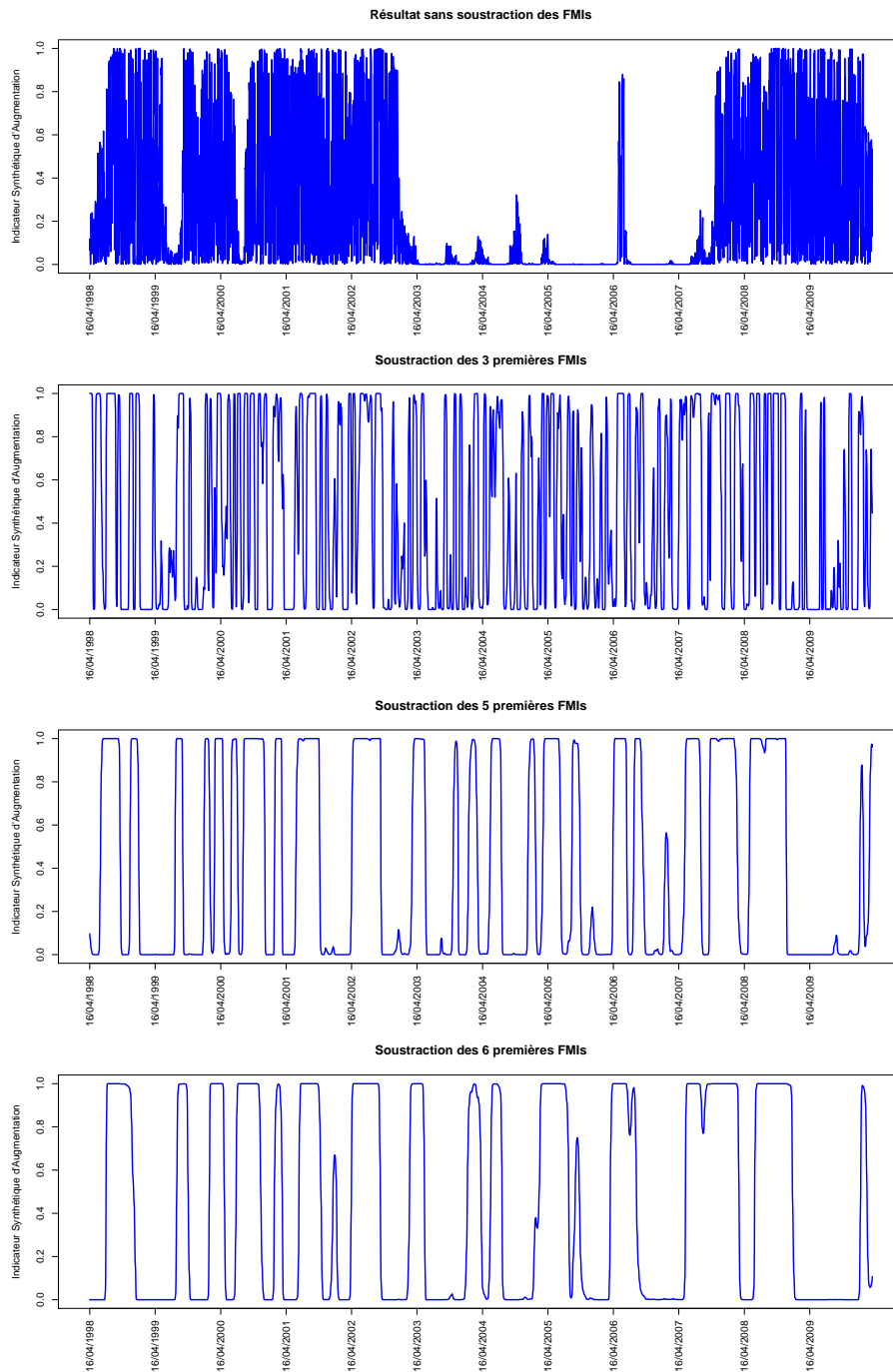


FIGURE 5.5 – Le graphique tout en haut correspond à une estimation de l'indicateur synthétique d'augmentation appliqué directement sur les séries d'indicateurs différenciés, sans aucune estimation d'une composante prévisible. Les autres graphiques correspondent à l'indicateur synthétique d'augmentation obtenu d'après des composantes prévisibles, estimées sur les séries différenciées, en soustrayant à ces séries les 3, 5 et 6 premières fonctions modales intrinsèques estimées.

horloge d'investissement. On fait également le lien avec les chapitres précédents en construisant un indicateur synthétique lié aux croyances des investisseurs. Cet indicateur synthétique repose sur un ensemble d'indicateurs positivement corrélés avec les rendements au carré des marchés actions, que ce soit au niveau quotidien, hebdomadaire ou mensuel (voir Chapitre 4). Il est obtenu en utilisant une composante prévisible estimée par une décomposition modale empirique des séries différenciées des différents indicateurs.

5.4. CONCLUSION

5.4. CONCLUSION

Allocation stratégique de portefeuille à partir de stratégies momentums réglées par une horloge d'investissement

Sommaire

6.1	Sélectionner un univers d'investissement diversifié	160
6.1.1	La diversification	160
6.1.2	Classification ascendante hiérarchique des actifs	162
6.1.3	Détermination d'un portefeuille composé d'un nombre minimal de classes nécessaires à sa diversification	164
6.2	Stratégies « momentums » réglées par une horloge d'investissement . . .	170
6.2.1	Présentation d'une stratégie momentum classique	170
6.2.2	Allocation en fonction d'une horloge d'investissement	174
6.2.3	Allocation par le « momentum »	175
6.2.4	Backtesting de l'allocation stratégique proposée	178
6.3	Conclusion	180

Introduction

Ce chapitre se compose de deux parties : dans la première, nous proposons une méthode permettant de sélectionner un univers d'investissement diversifié dans le sens où les tendances des rendements des différents actifs sélectionnés évoluent de manière indépendante. Dans la seconde partie du chapitre, nous partons de cet univers d'investissement sélectionné pour construire une allocation stratégique de portefeuille basée sur une horloge d'investissement (telle que proposée dans le Chapitre 5) et sur des stratégies momentums entre les différentes classes d'actifs.

6.1 Sélectionner un univers d'investissement diversifié

6.1.1 La diversification

La diversification d'un portefeuille consiste à sélectionner un certain nombre d'actifs dont les rendements sont faiblement corrélés. On espère ainsi réduire le risque associé au portefeuille.

Soient $N \in \mathbb{N}\{0, 1\}$ le nombre d'actifs détenus dans un portefeuille, noté P_N , et $0 < \sigma_{P_N} < +\infty$ l'écart-type des rendements de ce portefeuille. Nous allons montrer, dans un premier temps, qu'il existe une limite à la diversification de ce portefeuille :

Proposition 1. *Si N tend vers l'infini, on a équivalence entre σ_{P_N} et l'écart-type d'un portefeuille contenant les mêmes actifs que P_N mais également un actif supplémentaire, noté a_{N+1} , ayant un écart-type fini ($\sigma_{a_{N+1}} < +\infty$).*

Preuve. Le rendement de P_{N+1} peut s'écrire comme

$$\begin{aligned} R_{t,t+\Delta t}^{P_{N+1}} &= \rho_{P_{N+1}}^{a_1}(t)R_{t,t+\Delta t}^{a_1} + \dots + \rho_{P_{N+1}}^{a_N}(t)R_{t,t+\Delta t}^{a_N} + \rho_{P_{N+1}}^{a_{N+1}}(t)R_{t,t+\Delta t}^{a_{N+1}} \\ &= (1 - \rho_{P_{N+1}}^{a_{N+1}}(t))R_{t,t+\Delta t}^{P_N} + \rho_{P_{N+1}}^{a_{N+1}}(t)R_{t,t+\Delta t}^{a_{N+1}} \end{aligned}$$

où $\rho_P^{a_k}$ correspond au poids accordé à l'actif a_k dans le portefeuille P . On en déduit que

$$\sigma_{P_{N+1}}(t)^2 \leq \sigma_{P_N}(t)^2 + \rho_{P_{N+1}}^{a_{N+1}}(\sigma_{a_{N+1}}(t)^2 - \sigma_{P_N}(t)^2)$$

Lorsque $N \rightarrow +\infty$, on a $\sum_{i=1}^N \rho_{P_N}^{a_i} = 1$ où $\rho_{P_N}^{a_i} > 0$ désigne le poids de l'actif a_i dans le portefeuille P_N . Par conséquent, $\forall \epsilon > 0, \exists n_0 \in [1, \dots, N], \forall n \in [n_0, \dots, N], \rho_{P_n}^{a_n} < \epsilon$. D'où :

$$\begin{aligned} \forall \epsilon > 0, \exists n_0 \in [1, \dots, N], \forall n \in [n_0, \dots, N], \\ |\sigma_{P_{n+1}}(t)^2 - \sigma_{P_n}(t)^2| \leq \rho_{P_{n+1}}^{a_{n+1}}|\sigma_{a_{n+1}}(t)^2 - \sigma_{P_n}(t)^2| < \epsilon|\sigma_{a_{n+1}}(t)^2 - \sigma_{P_n}(t)^2| = \delta \end{aligned}$$

Ainsi, à la condition que les écart-types $\sigma_{P_n}(t)$ et $\sigma_{a_{n+1}}(t)$ soient finis, on a :

$$\forall \delta > 0, \exists n_0 \in [1, \dots, N], \forall n \in [n_0, \dots, N], |\sigma_{P_{n+1}}(t)^2 - \sigma_{P_n}(t)^2| < \delta$$

□

Puisque l'écart-type des rendements converge lorsque le nombre d'actifs augmente, il peut être intéressant de savoir vers quelle limite. Cette dernière va néanmoins dépendre des poids de chacun des actifs détenus dans le portefeuille ainsi que de la variance de leurs rendements. On ne peut donc en obtenir un résultat précis sans faire quelques hypothèses. Une manière de faire consiste à se limiter à l'étude d'un portefeuille équilibré. Nous allons ainsi montrer que la variance d'un portefeuille équilibré peut se décomposer en une partie diversifiable (ou spécifique), c'est-à-dire négligeable lorsque le nombre d'actifs du portefeuille devient grand, et une partie non-diversifiable (ou systématique) équivalente à la covariance moyenne entre les tendances (ou composantes prévisibles) des rendements des différents actifs du portefeuille :

Proposition 2. *Soit $R_{t,t+\Delta t}^{a_i} = \mu_{t,t+\Delta t}^{a_i} + \epsilon_{t,t+\Delta t}^{a_i}$ où $R_{t,t+\Delta t}^{a_i}$ correspond au rendement de l'actif a_i entre t et $t + \Delta t$, $\mu_{t,t+\Delta t}^{a_i}$ correspond à la composante déterministe de $R_{t,t+\Delta t}^{a_i}$ et $\epsilon_{t,t+\Delta t}^{a_i}$ à sa composante aléatoire.*

Sous l'hypothèse que

$$\text{cov}(\mu_{t,t+\Delta t}^{a_i}, \epsilon_{t,t+\Delta t}^{a_j}) = \text{cov}(\mu_{t,t+\Delta t}^{a_i}, \epsilon_{t,t+\Delta t}^{a_j}) = \text{cov}(\epsilon_{t,t+\Delta t}^{a_i}, \mu_{t,t+\Delta t}^{a_j}) \quad (6.1)$$

$$= \text{cov}(\epsilon_{t,t+\Delta t}^{a_i}, \epsilon_{t,t+\Delta t}^{a_j}) = \text{cov}(\mu_{t,t+\Delta t}^{a_i}, \epsilon_{t,t+\Delta t}^{a_i}) = 0 \quad \forall (i, j) \in [1, \dots, N] \quad (6.2)$$

et dans le cas d'un portefeuille équipondéré,

$$\sigma_{P_N}^2 \text{ est équivalent à } \frac{2}{N(N-1)} \sum_{1 \leq i < j \leq N} \text{cov}(\mu_{t,t+\Delta t}^{a_i}, \mu_{t,t+\Delta t}^{a_j}) \text{ lorsque } N \rightarrow +\infty.$$

Ainsi, un portefeuille équipondéré composé d'un très grand nombre d'actifs dont les tendances des rendements sont totalement décorrélées entre elles aura une variance quasi-nulle.

Preuve.

$$\mathbb{V}[R_{t,t+\Delta t}^{P_N}] = \sum_{i=1}^N (\rho^{a_i}(t))^2 (\mathbb{V}[\mu_{t,t+\Delta t}^{a_i}] + \mathbb{V}[\epsilon_{t,t+\Delta t}^{a_i}] + 2 \times \text{cov}(\mu_{t,t+\Delta t}^{a_i}, \epsilon_{t,t+\Delta t}^{a_i})) \quad (6.3)$$

$$+ 2 \times \sum_{1 \leq i < j \leq N} \rho^{a_i}(t) \rho^{a_j}(t) \times [\text{cov}(\mu_{t,t+\Delta t}^{a_i}, \mu_{t,t+\Delta t}^{a_j}) + \text{cov}(\mu_{t,t+\Delta t}^{a_i}, \epsilon_{t,t+\Delta t}^{a_j}) \quad (6.4)$$

$$+ \text{cov}(\epsilon_{t,t+\Delta t}^{a_i}, \mu_{t,t+\Delta t}^{a_j}) + \text{cov}(\epsilon_{t,t+\Delta t}^{a_i}, \epsilon_{t,t+\Delta t}^{a_j})] \quad (6.5)$$

On en déduit, d'après les hypothèses, que

$$\mathbb{V}[R_{t,t+\Delta t}^P] = \sum_{i=1}^N (\rho^{a_i}(t))^2 (\mathbb{V}[R_{t,t+\Delta t}^{a_i}]) \quad (6.6)$$

$$+ 2 \times \sum_{1 \leq i < j \leq N} \rho^{a_i}(t) \rho^{a_j}(t) \times \text{cov}(\mu_{t,t+\Delta t}^{a_i}, \mu_{t,t+\Delta t}^{a_j}) \quad (6.7)$$

Dans le cadre d'un portefeuille équipondéré, ceci peut se réécrire comme

$$\mathbb{V}[R_{t,t+\Delta t}^P] = \frac{1}{N^2} \sum_{i=1}^N \mathbb{V}[R_{t,t+\Delta t}^{a_i}] \quad (6.8)$$

$$+ \frac{N(N-1)}{N^2} \frac{2}{N(N-1)} \times \sum_{1 \leq i < j \leq N} \text{cov}(\mu_{t,t+\Delta t}^{a_i}, \mu_{t,t+\Delta t}^{a_j}) \quad (6.9)$$

$\frac{1}{N^2} \sum_{i=1}^N \mathbb{V}[R_{t,t+\Delta t}^{a_i}]$ représente la partie de la variance du portefeuille dite « diversifiable ». En effet, on peut raisonnablement supposer que

$$\frac{1}{N} \sum_{i=1}^N \mathbb{V}[R_{t,t+\Delta t}^{a_i}] = o_{N \rightarrow +\infty}(N)$$

La variance diversifiable du portefeuille tend vers 0 lorsque le nombre d'actifs augmente.

$\frac{N(N-1)}{N^2} \frac{2}{N(N-1)} \times \sum_{1 \leq i < j \leq N} \text{cov}(\mu_{t,t+\Delta t}^{a_i}, \mu_{t,t+\Delta t}^{a_j})$ représente la partie de la variance non-diversifiable.

Lorsque $N \rightarrow +\infty$, elle est équivalente à $\frac{2}{N(N-1)} \times \sum_{1 \leq i < j \leq N} \text{cov}(\mu_{t,t+\Delta t}^{a_i}, \mu_{t,t+\Delta t}^{a_j})$. Ce qui représente

la covariance moyenne entre les tendances des rendements de l'ensemble des titres du portefeuille. \square

Pour réduire la variance du portefeuille équilibré lorsque le nombre d'actifs est élevé, il faut réduire la partie non-diversifiable de la variance car la partie diversifiable sera négligeable. Puisque la partie non-diversifiable tend vers la covariance moyenne entre les tendances des rendements, il faut tenter de minimiser cette covariance moyenne. Nous proposons, dans le prochain paragraphe, une méthode de sélection d'un ensemble de titres maximisant l'indépendance entre les tendances de leurs rendements.

6.1.2 Classification ascendante hiérarchique des actifs

Il existe deux grands types d'approches pour classifier un ensemble de données. Le premier type consiste à partitionner l'ensemble des données en fixant, au préalable, un nombre de classes. C'est par exemple le cas de la méthode des centres mobiles et de ses variantes [Tho(1953)] [Bal(1967)]. Le second type d'approche fournit une classification en fonction d'un niveau de regroupement, chaque niveau étant déduit de l'agrégation des classes du niveau précédent. C'est ce qu'on appelle la classification hiérarchique (voir par exemple la classification hiérarchique ascendante de Ward [War(1963)]). Ces méthodes sont présentées dans [Leb(2006)] et [Sap(2006)].

Dans ce paragraphe, on propose d'utiliser une méthode de classification hiérarchique ascendante basée sur un indice de dissimilarité original. Celui-ci est basé sur un coefficient de corrélation de Spearman (corrélation des rangs) calculé entre les tendances des rendements des actifs. Cet indice estime la dépendance, linéaire ou non-linéaire, entre les tendances des rendements. L'enjeu est alors de construire un portefeuille avec des actifs ayant des tendances fluctuant le plus indépendamment possible les unes des autres. Une fois l'arbre de classification (ou dendrogramme) calculé, on propose une manière originale d'obtenir un portefeuille suffisamment diversifié en nous basant sur le résultat de la proposition 2. En partant d'un univers d'investissement très large, on obtient un algorithme permettant de sélectionner un univers d'investissement réduit mais suffisamment diversifié.

Indice de dissimilarité La classification hiérarchique ascendante d'un ensemble de données consiste à partitionner cet ensemble en regroupant, à chaque étape, les deux éléments les plus proches selon un indice de dissimilarité choisi. Cet indice, notée δ , est d'autant plus petit que deux données sont semblables. Il devient grand lorsque deux données sont éloignées. Dans notre cas,

$$\delta(a_i, a_j) = 1 - |c_{Spear}(\mu_{t,t+\Delta t}^{a_i}, \mu_{t,t+\Delta t}^{a_j})|$$

où c_{Spear} représente le coefficient de corrélation de Spearman, $\mu_{t,t+\Delta t}^{a_i}$ correspond à la tendance des rendements de l'actif a_i . On rappelle que le coefficient de corrélation de Spearman correspond au coefficient de corrélation de Pearson appliqué sur les variables ordonnées. L'indice de dissimilarité est donc d'autant plus petit que deux actifs ont leurs tendances de rendements dépendantes l'une de l'autre.

Dans notre étude, $\mu_{t,t+\Delta t}^{a_i}$ correspond à la composante prévisible estimée à l'aide de la méthode présentée dans le chapitre 4.

Positionnement multidimensionnel Néanmoins, l'indice de dissimilarité présenté ne correspond pas à une distance euclidienne, ce qui pose problème pour effectuer une classification ascendante hiérarchique par méthode de Ward par exemple (il faut pouvoir appliquer le théorème de décomposition de Huygens). Il existe alors une méthode, appelée positionnement multidimensionnel (ou *multidimensional scaling*), qui permet de trouver un ensemble de N points repérés par leurs

coordonnées euclidiennes et ayant des distances euclidiennes semblables aux indices de dissimilarité (voir [Kru(1964)]).

Pour N actifs, on obtient une matrice de dissimilarité

$$\Delta = \begin{pmatrix} \delta(a_1, a_1) & \delta(a_1, a_2) & \dots & \delta(a_1, a_N) \\ \delta(a_2, a_1) & \delta(a_2, a_2) & \dots & \delta(a_2, a_N) \\ \vdots & \vdots & & \vdots \\ \delta(a_N, a_1) & \delta(a_N, a_2) & \dots & \delta(a_N, a_N) \end{pmatrix} = \begin{pmatrix} 0 & \delta(a_1, a_2) & \dots & \delta(a_1, a_N) \\ \delta(a_2, a_1) & 0 & \dots & \delta(a_2, a_N) \\ \vdots & \vdots & & \vdots \\ \delta(a_N, a_1) & \delta(a_N, a_2) & \dots & 0 \end{pmatrix}$$

Pour laquelle $\delta(a_i, a_j) = \delta(a_j, a_i)$ pour tout $(i, j) \in [1, \dots, N]$. Le positionnement multidimensionnel consiste alors à rechercher un ensemble de points $(x_1, \dots, x_N) \in \mathbb{R}^q$ tels que $\|x_i - x_j\| \approx \delta(a_i, a_j)$ où $\|\cdot\|$ représente la distance euclidienne. Le choix de la dimension q est arbitraire mais plus il est élevé et plus les approximations des distances seront bonnes. En pratique, on peut choisir arbitrairement $q = \frac{N}{2}$ afin d'obtenir des résultats suffisamment cohérents. Le positionnement multidimensionnel peut donc se reformuler comme un problème d'optimisation :

$$(x_1, \dots, x_N) = \arg \min_{(x_1^*, \dots, x_N^*)} \sum_{i < j} (\|x_i^* - x_j^*\| - \delta(a_i, a_j))^2$$

Néanmoins, cette minimisation est complexe et ne possède pas une solution explicite. Pour résoudre ce problème, l'algorithme SMACOF (Scaling by Majorizing a Complex Function) donne de bons résultats ([DeL(1980)], [Bor(2005)]). Il se base sur des majorisations successives de la fonction à minimiser.

Classification ascendante hiérarchique de Ward Supposons que les données aient des coordonnées dans un espace euclidien et que cet ensemble soit partitionné en K classes non vides. Notons M le nombre total de données, C_k la $k^{\text{ème}}$ classe, m_k le nombre de données appartenant à cette classe, et g_k le centre de gravité des données de la classe C_k . On appelle inertie *intraclasse*, la quantité

$$I_{intra} = \sum_{k=1}^K \sum_{i \in C_k} \frac{1}{m_k} d(x_i, g_k)^2$$

où x_i correspond à la $i^{\text{ème}}$ donnée de la classe C_k et où $d(\cdot, \cdot)$ représente la distance Euclidienne.

Si on note g le centre de gravité de l'ensemble des données, on appelle inertie *interclasse* la quantité

$$I_{inter} = \sum_{k=1}^K \frac{m_k}{M} d(g_k, g)^2$$

Lorsque la distance d est euclidienne, on a le résultat suivant dû au théorème de décomposition de Huygens :

$$I_{tot} = I_{inter} + I_{intra}$$

où $I_{tot} = \frac{1}{M} \sum_{k=1}^K \sum_{i \in C_k} d(x_i, g)^2$ correspond à l'inertie totale des données.

La méthode de Ward consiste à agréger, à chaque étape, deux classes minimisant l'inertie *intraclasse* ou, de façon équivalente, maximisant l'inertie *interclasse*.

Considérons le passage de K classes à $K - 1$ classes en agrégeant deux classes C_{k_1} et C_{k_2} obtenues lors de la partition en K classes. La nouvelle classe $C_{k_1+k_2}$ formée par l'agrégation de ces deux classes aura les caractéristiques suivantes :

- $m_{k_1+k_2} = m_{k_1} + m_{k_2}$
- $g_{k_1+k_2} = \frac{m_{k_1}g_{k_1} + m_{k_2}g_{k_2}}{m_{k_1} + m_{k_2}}$

L'inertie *interclasse* lors de la partition en K classes peut se noter comme

$$I_{inter}(K) = \sum_{k=1}^K \frac{m_k}{M} d(g_k, g)^2 = \frac{m_{k_1}}{M} d(g_{k_1}, g)^2 + \frac{m_{k_2}}{M} d(g_{k_2}, g)^2 + \sum_{k=1, k \neq k_1, k \neq k_2}^K \frac{m_k}{M} d(g_k, g)^2$$

Après l'agrégation des classes C_{k_1} et C_{k_2} , elle s'écrit donc comme :

$$I_{inter}(K - 1) = \frac{m_{k_1+k_2}}{M} d(g_{k_1+k_2}, g)^2 + \sum_{k=1, k \neq k_1, k \neq k_2}^K \frac{m_k}{M} d(g_k, g)^2$$

Par conséquent, $I_{inter}(K) - I_{inter}(K - 1) = \frac{m_{k_1}}{M} d(g_{k_1}, g)^2 + \frac{m_{k_2}}{M} d(g_{k_2}, g)^2 - \frac{m_{k_1+k_2}}{M} d(g_{k_1+k_2}, g)^2$. Or, par le théorème de Huygens, on sait que

$$m_{k_1} d(g_{k_1}, g)^2 + m_{k_2} d(g_{k_2}, g)^2 = m_{k_1} d(g_{k_1}, g_{k_1+k_2})^2 + m_{k_2} d(g_{k_2}, g_{k_1+k_2})^2 + m_{k_1+k_2} d(g_{k_1+k_2}, g)^2$$

La diminution de l'inertie *interclasse* peut donc se réécrire comme

$$I_{inter}(K) - I_{inter}(K - 1) = \frac{m_{k_1}}{M} d(g_{k_1}, g_{k_1+k_2})^2 + \frac{m_{k_2}}{M} d(g_{k_2}, g_{k_1+k_2})^2$$

D'après la définition de $g_{k_1+k_2}$, on en déduit que

$$I_{inter}(K) - I_{inter}(K - 1) = \frac{m_{k_1} m_{k_2}}{M(m_{k_1} + m_{k_2})} d(g_{k_1}, g_{k_2})^2$$

Puisque $I_{inter}(K) - I_{inter}(K - 1)$ correspond à une estimation de la dérivée de $I_{inter}(K)$, on souhaite minimiser celle-ci afin d'atteindre un maximum local de l'inertie interclasse (ou, par équivalence, un minimum local de l'inertie intraclasse). Lorsqu'on possède K classes, on va donc rechercher les deux classes minimisant $\frac{m_{k_1} m_{k_2}}{(m_{k_1} + m_{k_2})} d(g_{k_1}, g_{k_2})^2$, également appelé « écart de Ward », afin de les agréger.

L'algorithme de classification ascendante hiérarchique basé sur la méthode de Ward consiste à partir de M classes, chaque classe contenant une donnée. Lors de la première étape, on regroupe les deux données ayant un écart de Ward le plus faible. Ensuite, à chaque niveau de regroupement, on agrège les deux classes minimisant cet écart. On peut alors construire un arbre de classification, également appelé dendrogramme. On aligne sur l'axe horizontal des points représentant les différentes données de départ puis, on les joint deux à deux, successivement, en suivant cet algorithme. La hauteur de chaque noeud de l'arbre correspond alors à l'écart de Ward.

6.1.3 Détermination d'un portefeuille composé d'un nombre minimal de classes nécessaires à sa diversification

Afin de déterminer les classes d'actifs suffisantes pour diversifier un portefeuille d'investissement, on pré-sélectionne 50 indices de prix, chacun représentant l'évolution d'un investissement dans une

certaine classe d'actifs. Les indices sont présentés dans le tableau et sont évalués quotidiennement du 01/01/1999 au 18/02/2011. Le dendrogramme obtenu après le calcul de la matrice des dissimilarités Δ , le positionnement multidimensionnel puis la classification ascendante hiérarchique de Ward est présenté dans la Figure 6.1. Pour sélectionner le nombre de classes nécessaire à une bonne diversification du portefeuille, on va parcourir le dendrogramme issu de la classification de Ward.

6.1. SÉLECTIONNER UN UNIVERS D'INVESTISSEMENT DIVERSIFIÉ

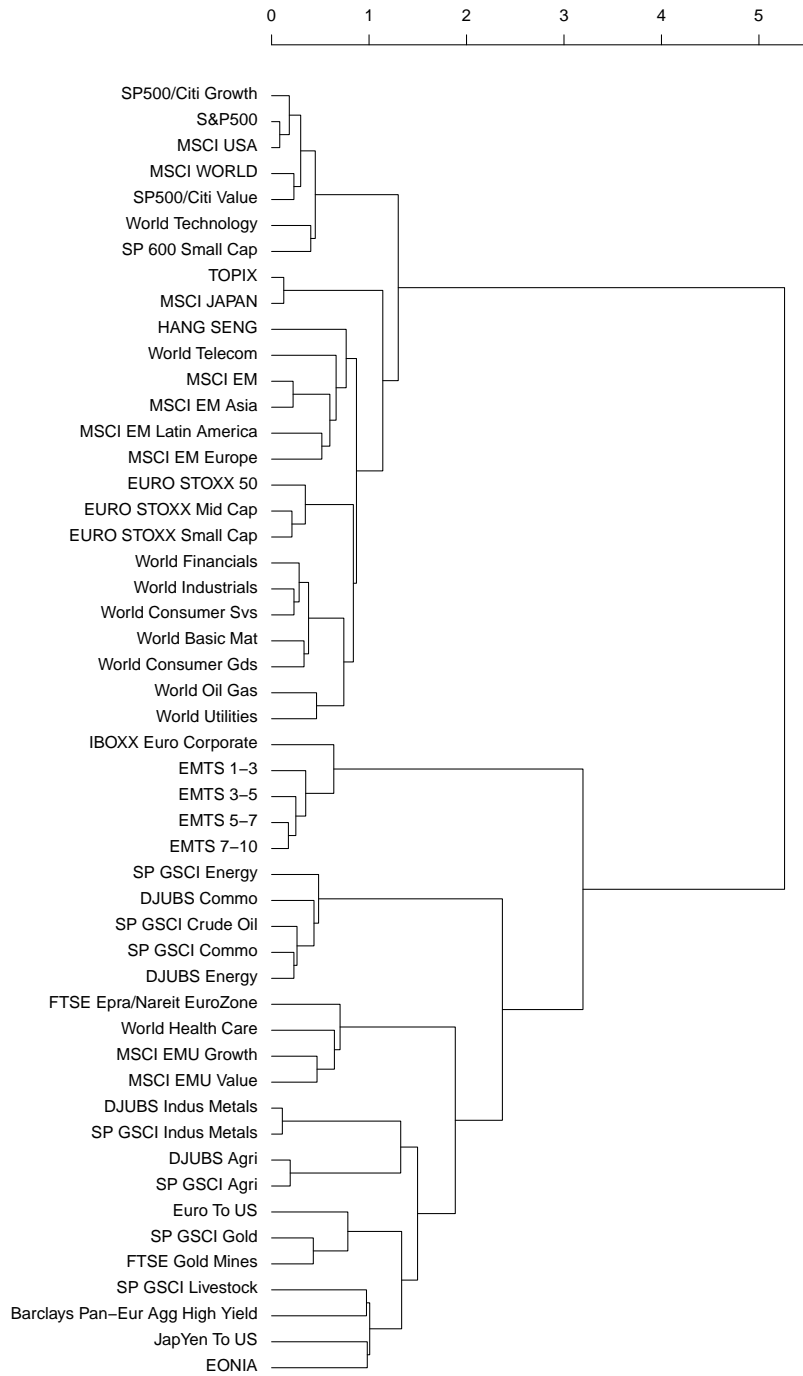


FIGURE 6.1 – Dendrogramme obtenu d'après la méthode de classification de Ward en utilisant notre indice de dissimilarité et un positionnement multidimensionnel appliquée sur un portefeuille composé de 50 indices de prix présélectionnés.

Nombre minimal nécessaire à l'homogénéité des classes A chaque niveau de regroupement, on souhaite obtenir des classes suffisamment homogènes vis-à-vis de la diversification du portefeuille. N'importe quel actif sélectionné dans une classe doit donc avoir le même impact sur la réduction de la partie diversifiable du portefeuille qu'un autre actif sélectionné dans cette même classe. Pour un nombre de classes sélectionné, on va tirer aléatoirement un actif dans chaque classe afin de composer un portefeuille. On réitère ce processus un certain nombre de fois afin de générer aléatoirement plusieurs portefeuilles associés à un même nombre de classes. Si les classes sont homogènes, la partie diversifiable de la variance de ces portefeuilles générés doit être la même pour tous. La figure 6.2 représente l'évolution de la variance de la partie diversifiable, calculée à partir de 1000 portefeuilles générés, en fonction du nombre de classes.

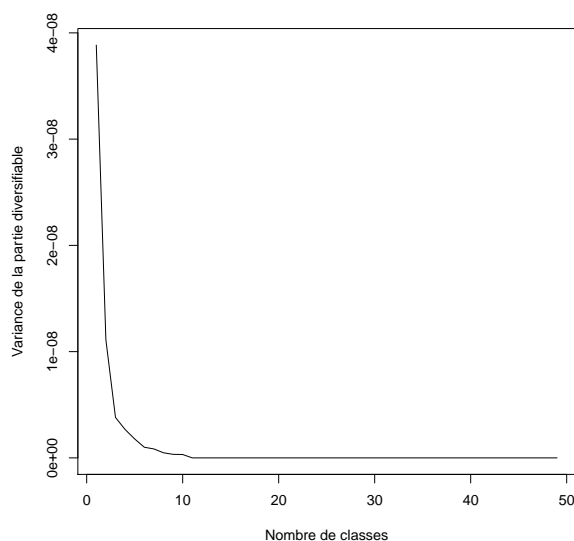


FIGURE 6.2 – Estimation de la variance de la partie diversifiable évaluée sur 1000 portefeuilles sélectionnés aléatoirement.

On constate qu'à partir de 12 classes, cette variance est quasi-nulle. Lorsqu'on observe la partie diversifiable de la variance du portefeuille, on conclue donc que lorsque le portefeuille se compose de 12 classes d'actifs ou plus, la méthode de sélection de l'actif représentatif de la classe ne modifie pas la partie diversifiable de la variance du portefeuille. En conclusion, le nombre de classes doit être supérieur ou égal à 12 afin d'obtenir des classes homogènes vis-à-vis de la diversification du portefeuille.

Nombre minimal permettant d'atteindre la limite de diversification du portefeuille On recherche maintenant un nombre de classes supérieur ou égal à 12 tel que lorsqu'on augmente ce nombre, la partie diversifiable du risque du portefeuille ne diminue plus significativement. On recherche donc un nombre de classes suffisamment élevé pour atteindre la limite de diversification du portefeuille. Dans ce but, on va estimer la dérivée seconde de la partie diversifiable en fonction du nombre d'actifs (voir figure 6.3). Lorsque cette dérivée seconde est proche de 0, on peut considérer que la partie diversifiable du risque ne diminue plus significativement. Dans notre étude, on constate

6.1. SÉLECTIONNER UN UNIVERS D'INVESTISSEMENT DIVERSIFIÉ

que rajouter des classes après en avoir sélectionné environ 23 ne diminue plus significativement la partie diversifiable du risque du portefeuille. Les 50 indices présélectionnés regroupés en 23 classes sont présentés dans le tableau 6.1.

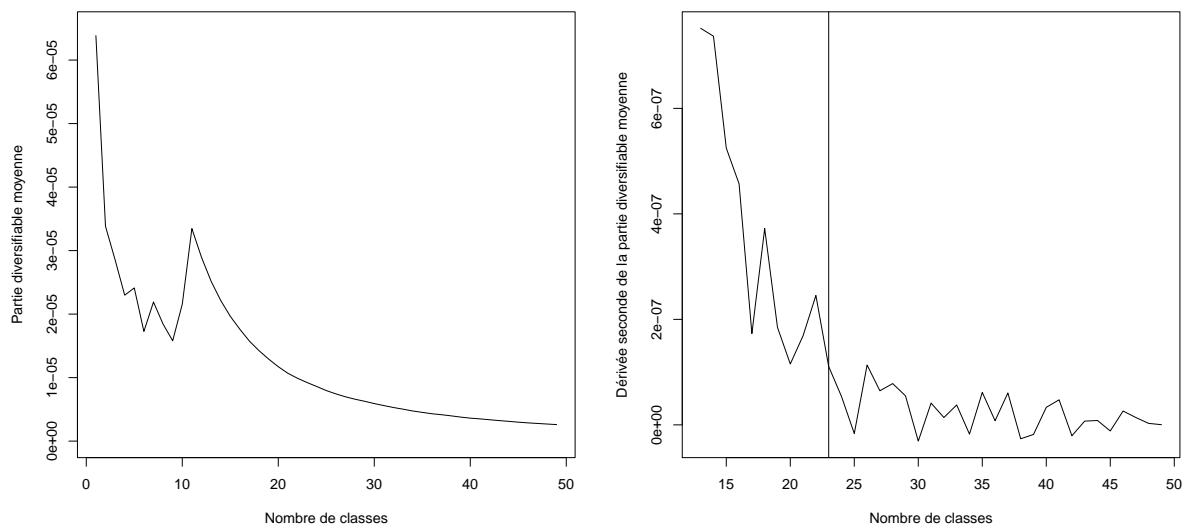


FIGURE 6.3 – Évolution de la dérivée seconde de la partie diversifiable moyenne évaluée sur 1000 portefeuilles sélectionnés aléatoirement. Après 23 classes, la dérivée seconde de la partie diversifiable de la variance du portefeuille semble évoluer dans un voisinage de 0.

6.1. SÉLECTIONNER UN UNIVERS D'INVESTISSEMENT DIVERSIFIÉ

Classe	Indices sélectionnés	Caractérisation
1	SP500, MSCI USA, MSCI World, World Consumer Svs, World Technology, SP 600 Small Cap, SP 500/Citi Growth, SP 500/Citi Value	Actions Marché US et secteurs Services à la consommation, Technologie
2	TOPIX, MSCI Japan	Actions Marché Japonais
3	Euro Stoxx 50, Euro Stoxx Small Cap, Euro Stoxx Mid Cap	Actions Zone Euro
4	Hang Seng	Actions Marché de Hong Kong
5	MSCI EM, MSCI EM Asia	Actions Marchés émergents, Asie
6	DJUBS Commo, SP GSCI Commo, DJ UBS Energy, SP GSCI Energy, SP GSCI Crude Oil	Matières premières, Énergies
7	DJUBS Agri, SP GSCI Agri	Matières premières Agriculture
8	DJ UBS Industrial Metals, SP GSCI Industrial Metals	Matières premières Métaux industriels
9	SP GSCI Gold, FTSE Gold Mines	Matières premières Or
10	EMTS 1-3, EMTS 3-5, EMTS 5-7, EMTS 7-10	Taux obligataires long-terme
11	Euro To US	Taux de change Euro/\$
12	Jap Yen To US	Taux de change JapYen/\$
13	FTSE Epra/NAREIT EuroZone	Actions Zone Euro secteur Immobilier
14	EONIA	Taux courts
15	World Oil Gas, World Utilities	Actions Internationales secteurs Pétroles, gaz et Services
16	World Basic Mat, World Industrial, World Consumer Gds, World Financials	Actions Internationales autres secteurs
17	World Health Care	Actions Internationales secteur Santé
18	World Telecom	Actions Internationales secteur Télécommunications
19	MSCI EMU Value, MSCI EMU Growth	Actions Zone Euro Growth et Value
20	MSCI EM Europe, MSCI EM Latin America	Actions Marchés émergents Europe et Amérique Latine
21	SP GSCI Livestock	Matières premières Bétail
22	IBOXX Euro Corporate	Taux Obligataires Entreprises Européennes Investment Grade
23	Barclays Pan Eur Agg High Yield	Taux Obligataires Entreprises Européennes High Yield

TABLEAU 6.1 – Les 50 indices regroupés en 23 classes d'après notre méthodologie

Recherche d'un portefeuille ayant un nombre de classes optimal et des tendances évoluant le plus indépendamment possible On recherche un portefeuille composé de 23 indices et dont la covariance moyenne entre les tendances de rendement est la plus faible possible. D'après l'étude précédente, on a $8 \times 2 \times 4 \times 2 \times 5 \times 2 \times 2 \times 2 \times 4 \times 2 \times 4 \times 2 \times 2 = 655360$ portefeuilles de 23 actifs possibles. Afin de rechercher le portefeuille ayant la covariance moyenne entre les actifs minimale, on génère 100 000 portefeuilles de manière aléatoire en sélectionnant au hasard un actif dans chaque classe. La densité, estimée par des noyaux gaussiens, de la répartition des covariances moyennes obtenues est présentée dans la figure 6.4.

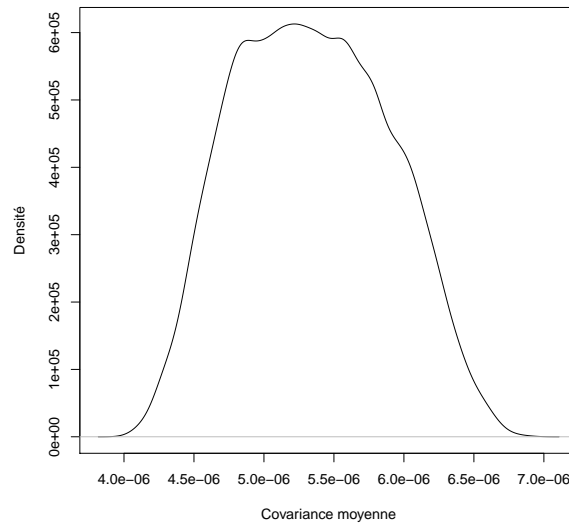


FIGURE 6.4 – Densité estimée par des noyaux gaussiens de la covariance moyenne obtenue à l'aide de 100 000 portefeuilles composés de 23 actifs piochés aléatoirement dans chacune des 23 classes.

Le portefeuille présenté dans le tableau 6.2 permet d'obtenir une covariance moyenne minimale de $4.009695e - 06$.

6.2 Stratégies « momentums » réglées par une horloge d'investissement

6.2.1 Présentation d'une stratégie momentum classique

En finance de marché, on parle généralement de momentum (qui signifie « dynamique » en anglais) des titres pour désigner le fait que les titres ayant eu une bonne (resp. mauvaise) performance dans le passé auront une bonne (resp. mauvaise) performance dans le futur. Ce phénomène a été testé à plusieurs reprises sur les marchés actions et les marchés des matières premières ([Jeg(1993)], [Jeg(1999)], [Jif(2006)]) : les titres ayant eu les meilleures performances sur les 1 à 12 mois passés font partie des meilleures performances constatées les 1 à 12 mois suivants.

Une stratégie momentum consiste donc à investir dans les actifs ayant eu les meilleures performances les x mois passés puis à détenir ces actifs pendant y mois avant de rebalancer le portefeuille.

6.2. STRATÉGIES « MOMENTUMS » RÉGLÉES PAR UNE HORLOGE D'INVESTISSEMENT

Classe	Actif sélectionné	Caractérisation
1	S&P500/Citi Growth	Actions US - Growth
2	TOPIX	Actions Japonaises
3	Euro Stoxx Mid Cap	Actions Zone Euro - Petites Capitalisations
4	Hang Seng	Actions Marché de Hong Kong
5	MSCI EM	Actions Internationales Marchés émergents
6	S&P GSCI Energy	Matières premières - Énergie
7	DJUBS Agri	Matières premières - Agriculture
8	S&P GSCI Industrial Metals	Matières premières - Métaux Industriels
9	S&P GSCI Gold	Matières premières - Or
10	EMTS 7-10	Taux obligataires Zone Euro - maturité 7-10 ans
11	Euro To US	Taux de change <i>EURO/US\$</i>
12	Jap Yen To US	Taux de change <i>JapYen/US\$</i>
13	FTSE EPRA/NAREIT EuroZone	Actions Zone Euro secteur Immobilier
14	EONIA	Taux Interbancaire Zone Euro - court terme
15	World Utilities	Actions Internationales secteur Services
16	World Consumer Goods	Actions Internationales secteur Biens de consommation
17	World Health Care	Actions Internationales secteur Santé
18	World Telecom	Actions Internationales secteur Télécommunications
19	MSCI EMU Growth	Actions Zone Euro - Growth
20	MSCI EM Latin America	Actions Marchés émergents Amérique Latine
21	S&P GSCI Livestock	Matières premières - Bétails
22	IBOXX Euro Corporate	Taux Obligataires Entreprises Européennes - Investment Grade
23	Barclays Pan Eur Agg High Yield	Taux Obligataires Entreprises Européennes - High Yield

TABLEAU 6.2 – Les 23 indices de prix obtenus après diversification de notre portefeuille. Ils possèdent une covariance moyenne de 4.009695e-06.

Par conséquent, elle se caractérise par une période d'observation et une période de détention (définies en nombre de mois) ainsi que par une stratégie d'allocation concernant les actifs ayant le mieux performés. Une méthode courante consiste à retenir les 3 actifs ayant eu la meilleure performance les x mois passés puis à associer un poids de 60% à l'actif ayant le mieux performé, 30% au second et 10% au troisième. En reprenant le portefeuille diversifié obtenu dans le tableau 6.2, il est possible de déterminer des périodes d'observation et de détention optimales (voir tableau 6.3).

Les résultats obtenus sur la période indiquent qu'une stratégie momentum 60-30-10 offre les meilleures performances en utilisant une période d'observation de 5 mois et une période de détention de 1 mois.

Impact de l'allocation des trois actifs les plus performants Afin d'étudier l'impact de l'allocation des trois actifs les plus performants, on génère aléatoirement des allocations entre ces trois actifs puis on étudie l'espérance des rendements annualisée des stratégies momentum correspondantes

6.2. STRATÉGIES « MOMENTUMS » RÉGLÉES PAR UNE HORLOGE D'INVESTISSEMENT

		Période d'observation (en mois)											
		1	2	3	4	5	6	7	8	9	10	11	12
Période de détention (en mois)	1	14.54	16.53	18.67	19.49	20.29	16.32	18.61	12.87	16.72	12.22	12.00	18.11
	2	13.79	17.37	16.13	18.58	17.78	16.42	15.04	13.15	13.77	12.10	11.84	15.33
	3	12.68	14.72	15.33	15.23	15.58	12.29	13.56	10.67	12.85	11.83	12.39	13.80
	4	10.59	14.13	14.17	12.58	11.97	10.64	12.37	10.85	12.17	10.07	11.56	13.37
	5	11.04	12.86	12.28	9.72	11.42	9.76	11.83	10.48	11.32	10.11	11.37	11.66
	6	11.07	10.40	9.38	8.60	10.28	8.90	11.19	9.37	10.15	9.09	9.28	9.36
	7	10.65	9.52	5.91	5.82	9.04	8.29	10.15	8.91	9.67	7.32	7.21	7.36
	8	9.87	8.41	5.44	6.24	9.08	8.04	9.95	8.73	8.43	7.06	6.58	6.32
	9	9.80	8.41	6.23	7.36	9.14	8.27	9.98	8.24	8.09	6.45	5.70	5.82
	10	10.19	8.94	7.35	8.20	9.41	8.30	9.23	7.73	7.32	5.44	5.31	5.56
	11	10.25	9.35	6.88	8.37	8.94	7.44	8.69	6.91	6.47	4.75	6.37	5.31
	12	10.28	8.50	7.54	9.34	8.55	7.84	9.30	7.21	6.89	6.21	6.21	5.48

TABLEAU 6.3 – Espérances des rendements annualisés des différentes stratégies momentum 60-30-10 testées (en %) sur la période 01/06/99-01/12/09.

		Période d'observation (en mois)											
		1	2	3	4	5	6	7	8	9	10	11	12
Période de détention (en mois)	1	7.00	7.02	6.52	6.61	7.10	7.41	7.24	7.31	7.50	7.59	7.32	7.39
	2	4.74	4.75	4.61	4.67	5.13	5.43	5.10	5.57	5.68	5.59	5.52	5.63
	3	4.33	3.91	4.12	4.11	4.27	4.61	4.45	4.68	4.58	4.61	4.63	4.71
	4	3.58	3.66	3.84	3.73	3.64	4.22	4.10	4.25	4.18	4.07	4.20	4.26
	5	3.48	3.35	3.47	3.39	3.52	3.90	3.85	3.92	3.89	3.86	3.93	3.91
	6	3.34	3.06	3.25	3.31	3.37	3.69	3.65	3.66	3.66	3.61	3.55	3.57
	7	3.06	2.85	3.10	3.11	3.25	3.45	3.40	3.33	3.33	3.20	3.24	3.16
	8	2.81	2.74	2.86	3.02	2.98	3.15	3.17	3.12	3.02	3.06	3.12	3.00
	9	2.72	2.65	2.77	2.85	2.79	2.95	3.05	2.90	2.89	2.90	2.92	2.84
	10	2.62	2.63	2.74	2.75	2.69	2.86	2.85	2.67	2.72	2.65	2.68	2.72
	11	2.58	2.48	2.54	2.71	2.66	2.68	2.68	2.52	2.53	2.47	2.57	2.53
	12	2.41	2.34	2.50	2.68	2.50	2.54	2.54	2.39	2.45	2.38	2.42	2.39

TABLEAU 6.4 – Écarts-types des rendements mensualisés des différentes stratégies momentum 60-30-10 testées (en %) sur la période 01/06/99-01/12/09. Il faut multiplier ces rendements par $\sqrt{12}$ pour obtenir des écart-types annualisés.

ayant une période d'observation de 5 mois et une période de détention d'un mois. La répartition des espérances de rendements annualisées est présentée dans la figure 6.5. Le tableau 6.5 représente les 10 allocations générées entraînant les plus fortes espérances de rendement. On constate que les stratégies accordant un poids le plus faible possible au deuxième actif ayant le mieux performé sur la période d'observation offrent les meilleures performances. En revanche, les poids accordés à l'actif ayant le mieux performé et au troisième meilleur performeur ne semblent pas avoir un impact important sur l'espérance de rentabilité de la stratégie. Ce résultat un peu surprenant peut être le fait d'un événement isolé, par exemple une forte baisse du deuxième actif par rapport aux autres, constaté sur

l'intervalle de temps dont nous disposons.

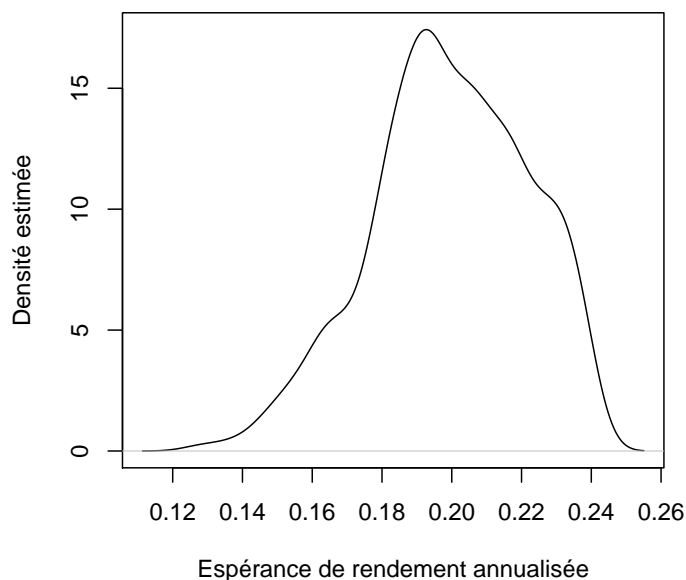


FIGURE 6.5 – Densité estimée à l'aide de noyaux gaussiens des espérances de rendements annualisés dont les poids accordés aux trois actifs les plus performants sont générés aléatoirement pour des stratégies momentum avec une période d'observation de 5 mois et une période de détention d'un mois.

Poids Actif 1	Poids Actif 2	Poids Actif 3	Esp. Rdt Ann. (en %)
50.00	0	50.00	23.98
46.94	0.18	52.87	23.96
56.19	0.22	43.58	23.96
84.57	0.24	15.18	23.95
41.48	0.27	58.23	23.95
89.96	0.27	09.75	23.95
40.60	0.49	58.89	23.92
31.82	0.74	67.43	23.89
54.42	0.77	44.79	23.89
54.58	0.87	44.54	23.88

TABLEAU 6.5 – Les allocations générant les plus fortes espérances de rendement parmi les allocations générées.

En conclusion, une stratégie momentum nécessite de fixer un certain nombre de paramètres : la période d'observation, la période de détention, le nombre d'actifs à sélectionner ainsi que l'allocation des poids à accorder aux actifs sélectionnés. Dans le prochain paragraphe, on expose une méthode

permettant de mettre en place une telle stratégie de gestion tout en prenant en compte les cycles macroéconomiques tels que mesurés par les indicateurs synthétiques d'augmentation présentés dans le Chapitre 5.

6.2.2 Allocation en fonction d'une horloge d'investissement

On souhaite prendre en compte l'information issue de l'horloge d'investissement présentée dans la figure 4.2 et dont l'heure est indiquée par la méthode présentée dans le paragraphe 5.3.1. Pour cela, on va regrouper les indices présentés dans le tableau 6.2 suivant leur nature (Ex. : Actions, Obligations, Matières Premières, Liquidités,...) :

Groupe	Indices sélectionnés	Caractérisation du groupe
1	S&P500/Citi Growth, TOPIX, Euro Stoxx Mid Cap, Hang Seng, MSCI EM, FTSE EPRA/NAREIT Euro-Zone, World Utilities, World Consumer Goods, World Health Care, World Telecom, MSCI EM Latin America, MSCI EMU Growth	Actions
2	EMTS 7-10, IBOXX Euro Corporate, Barclays Pan Eur Agg High Yield	Taux obligataires
3	S&P GSCI Energy, DJUBS Agri, S&P GSCI Industrial Metals, S&P GSCI Gold, S&P GSCI Livestock, Euro To US, Jap Yen To US	Matières premières et devises
4	EONIA	Taux Interbancaire Zone Euro - court terme

TABLEAU 6.6 – Les 4 groupes d'indices sélectionnés d'après le portefeuille diversifié obtenu dans le paragraphe 6.1.3

On va créer un vecteur poids définissant une allocation entre ces 4 groupes et défini à partir de l'horloge d'investissement. Pour cela, on note $\theta_t \in [0, 2\pi[$ l'angle de l'aiguille sur l'horloge d'investissement et on pose :

$$\rho_1(\theta_t) = \begin{pmatrix} (1 - \frac{|\frac{3\pi}{4} - \theta_t|}{\pi/2}) \mathbf{1}_{\theta_t \in [\frac{\pi}{4}, \frac{5\pi}{4}]} \\ (1 - \frac{|\frac{5\pi}{4} - \theta_t|}{\pi/2}) \mathbf{1}_{\theta_t \in [\frac{3\pi}{4}, \frac{7\pi}{4}]} \\ (1 - \frac{|\frac{\pi}{4} - \theta_t|}{\pi/2}) \mathbf{1}_{\theta_t \in [0, \frac{3\pi}{4}]} + (1 - \frac{|\frac{9\pi}{4} - \theta_t|}{\pi/2}) \mathbf{1}_{\theta_t \in [\frac{7\pi}{4}, 2\pi[} \\ (1 - \frac{|-\frac{\pi}{4} - \theta_t|}{\pi/2}) \mathbf{1}_{\theta_t \in [0, \frac{\pi}{4}]} + (1 - \frac{|\frac{7\pi}{4} - \theta_t|}{\pi/2}) \mathbf{1}_{\theta_t \in [\frac{5\pi}{4}, 2\pi[} \end{pmatrix}$$

La $i^{\text{ème}}$ coordonnée du vecteur $\rho_1(\theta_t)$ correspond alors à un poids accordé au groupe i . Cette stratégie d'allocation ne permet d'investir que dans deux groupes à un même instant. La figure 6.6 représente l'évolution des poids ainsi accordés aux différents groupes d'après notre horloge d'investissement.

Si on souhaite qu'il soit possible d'avoir un poids non nuls au niveau des 4 groupes d'actifs, on

peut utiliser le vecteur des poids défini par :

$$\rho_2(\theta_t) = \frac{1}{2} \begin{pmatrix} 1 - \left| \frac{3\pi - \theta_t}{4} \right| \\ 1 - \left| \frac{5\pi - \theta_t}{4} \right| \\ 1 - \left| \frac{\pi - \theta_t}{4} \right| \\ 1 - \left| \frac{7\pi - \theta_t}{4} \right| \end{pmatrix}$$

L'allocation correspondante à ρ_2 est affichée dans la figure 6.7. Enfin, on propose un troisième type d'allocation :

$$\rho_3(\theta_t) = \frac{1}{S} \begin{pmatrix} \left| 1 - \frac{3\pi - \theta_t}{4} \right|^2 \\ \left| 1 - \frac{5\pi - \theta_t}{4} \right|^2 \\ \left| 1 - \frac{\pi - \theta_t}{4} \right|^2 \\ \left| 1 - \frac{7\pi - \theta_t}{4} \right|^2 \end{pmatrix}$$

où $S = \sum_{k \in [1,3,5,7]} \left| 1 - \frac{k\pi - \theta_t}{4} \right|^2$. L'allocation obtenue est affichée dans la figure 6.8.

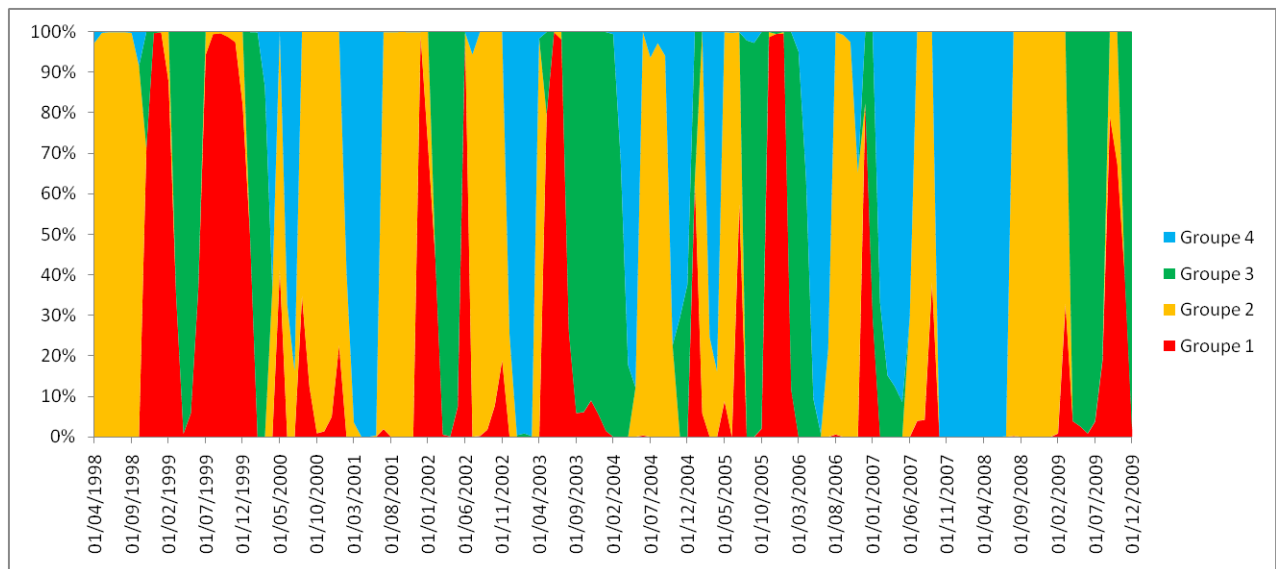


FIGURE 6.6 – Évolution de l'allocation définie par ρ_1 à une fréquence mensuelle et accordée aux différents groupes d'après l'horloge d'investissement.

Les résultats obtenus par les trois méthodes d'allocation semblent bien anticiper le krach boursier d'Octobre 2008 par exemple. En effet, chacune des stratégies d'allocation investie le plus qu'elle peut dans les liquidités (Groupe 4) entre Octobre 2007 et Juillet 2008 puis en obligations (Groupe 2) entre Aout 2008 et Février 2009.

6.2.3 Allocation par le « momentum »

Au sein de chaque groupe dont l'allocation fournie par l'horloge d'investissement n'est pas nulle, on applique une stratégie momentum afin de choisir les actifs les plus performants du groupe sur une

6.2. STRATÉGIES « MOMENTUMS » RÉGLÉES PAR UNE HORLOGE D'INVESTISSEMENT

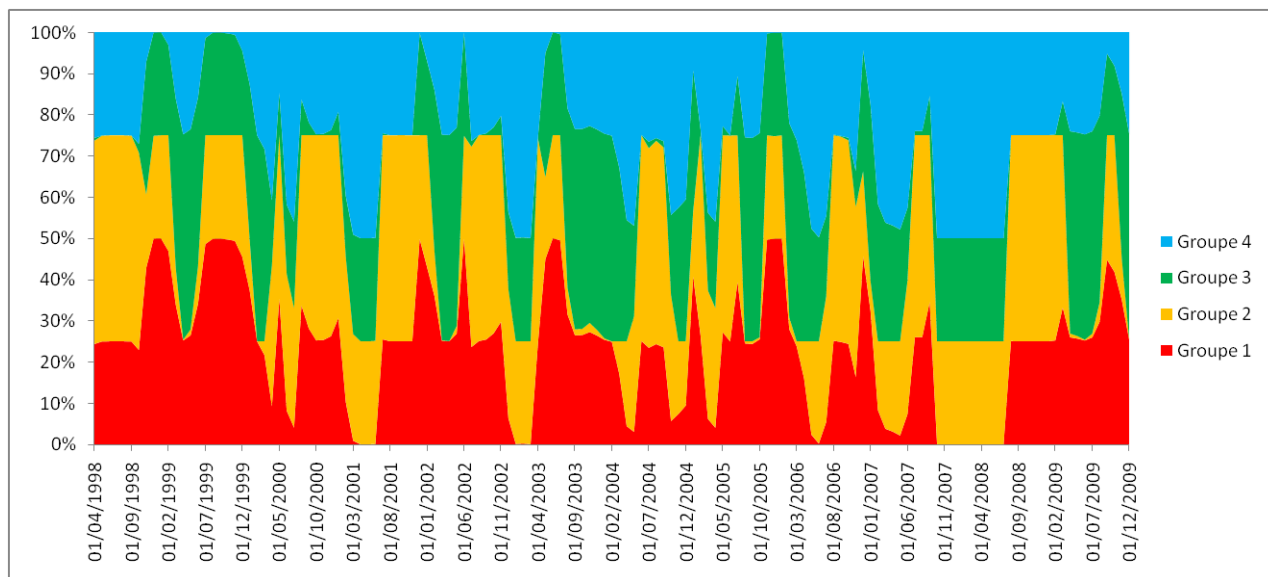


FIGURE 6.7 – Évolution de l'allocation définie par ρ_2 à une fréquence mensuelle et accordée aux différents groupes d'après l'horloge d'investissement.

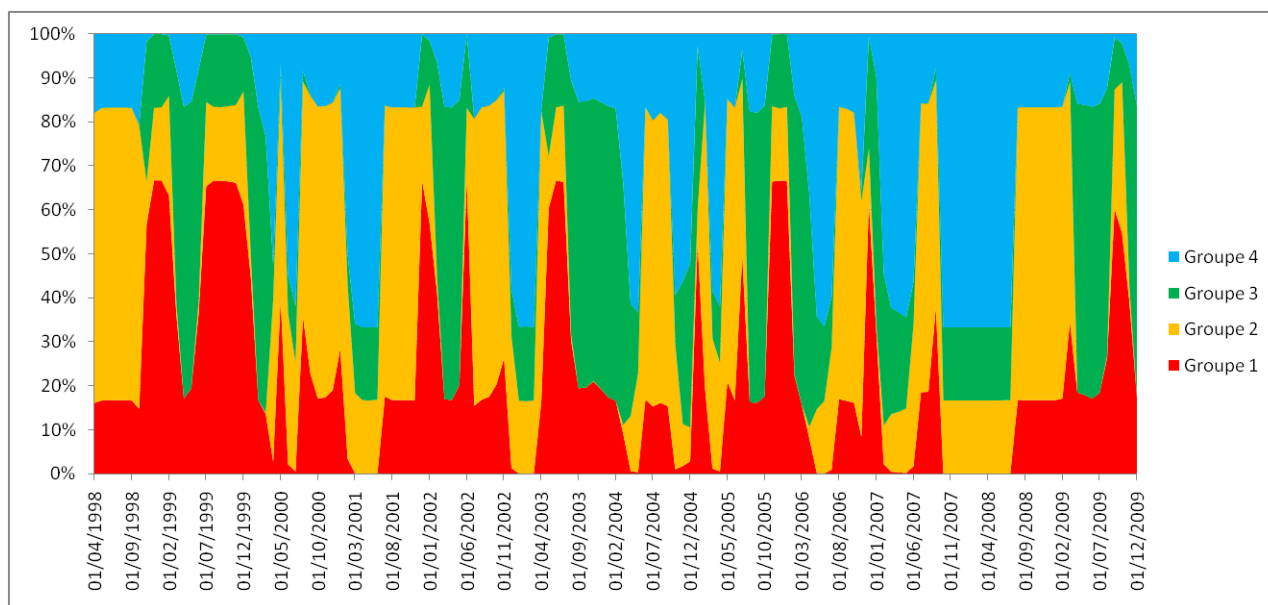


FIGURE 6.8 – Évolution de l'allocation définie par ρ_3 à une fréquence mensuelle et accordée aux différents groupes d'après l'horloge d'investissement.

période d'observation passée. Il faut donc déterminer, pour chaque groupe, les périodes d'observation et de détention optimales ainsi que le nombre d'actifs à retenir et les poids à leur accorder. Dans ce but, on génère des allocations aléatoires de portefeuille pour un nombre d'actifs variant de 1 à $[\frac{N_g}{2}] + 1$, où N_g correspond au nombre d'actifs inclus dans le groupe g et $[\cdot]$ représente la partie entière, et pour des périodes d'observation et de détention allant de 1 à 12 mois. On retient alors la

stratégie donnant l'espérance de rendement la plus élevée.

Groupe 1 Le tableau 6.7 présente les stratégies momentum permettant d'obtenir les espérances de rendement les plus élevées en fonction du nombre d'actifs retenus. On constate qu'un nombre d'actifs retenus entre 1 et 3 permet de créer des stratégies ayant une espérance de rendement annualisée d'environ 24%. La période d'observation permettant d'obtenir les meilleures stratégies est de 2 mois et celle de détention de 1 mois. Les allocations qui ont permis d'offrir les meilleures espérances de rendement donnent un poids important au premier actif et des poids plus faibles aux autres. L'écart-type des rendements des stratégies retenues diminue lorsque le nombre d'actifs retenus augmente. On va donc retenir une stratégie consistant à investir dans les trois actifs ayant offert les meilleures performances sur les deux mois passés et à préserver ces actifs dans le portefeuille durant un mois suivant une allocation 95-4-1.

	Nombre d'actifs retenus						
	1	2	3	4	5	6	7
Esp. Rdt Ann. Max.	24.49	24.55	23.59	16.44	15.66	14.64	13.83
E.-T. Mens.	29.09	28.75	28.26	24.04	19.43	19.71	19.53
Pér. Obs.	2	2	2	1	1	1	2
Pér. Dét.	1	1	1	1	1	1	1
p_1	100	98.02	94.29	75.32	26.76	35.29	31.35
p_2		1.98	4.43	12.21	28.75	21.82	27.79
p_3			1.26	1.33	41.17	26.69	1.06
p_4				11.13	2.25	2.64	30.95
p_5					1.07	1.46	2.60
p_6						12.07	2.68
p_7							3.55

TABLEAU 6.7 – Les stratégies ayant générées les plus fortes espérances de rendement annualisées (en %) pour le groupe 1 en fonction du nombre d'actifs. Pour chaque nombre d'actifs sélectionnés, on détaille l'espérance de rendement annualisée maximale obtenue ainsi que l'écart-type annualisé (en %) correspondant à cette stratégie, sa période d'observation, sa période de détention et l'allocation générée aléatoirement qui a permis d'obtenir cette espérance de rendement maximale (poids p_i (en %)).

Groupe 2 Le tableau 6.8 présente les stratégies momentum permettant d'obtenir les espérances de rendement les plus élevées en fonction du nombre d'actifs retenus. La période d'observation permettant d'obtenir les meilleures stratégies est de 5 mois et celle de détention de 1 mois. Les résultats suggèrent de retenir une stratégie consistant à investir dans l'actif ayant offert les meilleures performances sur les cinq mois passés et à préserver cet actif dans le portefeuille durant un mois.

Groupe 3 Le tableau 6.9 présente les stratégies momentum permettant d'obtenir les espérances de rendement les plus élevées en fonction du nombre d'actifs retenus. On constate que plus le nombre d'actifs sélectionnés est faible et plus la stratégie a une espérance de rendement élevée et un écart-type de rendement élevé. La période d'observation permettant d'obtenir les meilleures stratégies est de 5 mois et celle de détention de 1 mois. Les allocations qui ont permis d'offrir les meilleures espérances

6.2. STRATÉGIES « MOMENTUMS » RÉGLÉES PAR UNE HORLOGE D'INVESTISSEMENT

	Nombre d'actifs retenus	
	1	2
Esp. Rdt Ann. Max.	9.88	9.44
E.-T. Ann.	8.59	8.49
Pér. Obs.	5	5
Pér. Dét.	1	1
p_1	100	97.08
p_2		2.92

TABLEAU 6.8 – Les stratégies ayant générées les plus fortes espérances de rendement annualisées (en %) pour le groupe 2 en fonction du nombre d'actifs. Pour chaque nombre d'actifs sélectionnés, on détaille l'espérance de rendement annualisée maximale obtenue ainsi que l'écart-type des rendements annualisés (en %) correspondant à cette stratégie, sa période d'observation, sa période de détention et l'allocation générée aléatoirement qui a permis d'obtenir cette espérance de rendement maximale (poids p_i (en %)).

de rendement donnent un poids important au premier actif et des poids plus faibles aux autres. On va donc retenir une stratégie consistant à investir dans les deux actifs ayant offert les meilleures performances sur les cinq mois passés et à préserver ces actifs dans le portefeuille durant un mois suivant une allocation 95-5.

	Nombre d'actifs retenus			
	1	2	3	4
Esp. Rdt Ann. Max.	27.82	25.92	22.39	18.55
E.-T. Ann.	25.29	24.04	15.55	16.59
Pér. Obs.	5	5	5	5
Pér. Dét.	1	1	2	1
p_1	100	93.17	86.52	58.21
p_2		6.82	11.73	18.89
p_3			1.74	17.44
p_4				5.45

TABLEAU 6.9 – Les stratégies ayant générées les plus fortes espérances de rendement annualisées (en %) pour le groupe 3 en fonction du nombre d'actifs. Pour chaque nombre d'actifs sélectionnés, on détaille l'espérance de rendement annualisée maximale obtenue ainsi que l'écart-type des rendements annualisés (en %) correspondant à cette stratégie, sa période d'observation, sa période de détention et l'allocation générée aléatoirement qui a permis d'obtenir cette espérance de rendement maximale (poids p_i (en %)).

6.2.4 Backtesting de l'allocation stratégique proposée

Les poids des actifs sont déterminés en multipliant les poids obtenus à la suite de l'allocation momentum, par le poids accordé à l'aide de l'horloge d'investissement au groupe auquel les actifs appartiennent. En d'autre termes, le poids accordé à l'actif i , noté p_i , est égale à :

$$p_i(t) = \rho_1(\theta_t)p_{mom(g)}(t)$$

6.2. STRATÉGIES « MOMENTUMS » RÉGLÉES PAR UNE HORLOGE D'INVESTISSEMENT

si on applique l'allocation définie par $\rho_1(\theta_t)$ et où $p_{mom(g)}(t)$ correspond au poids accordé par la stratégie momentum appliquée dans le groupe g auquel appartient l'actif i . La figure 6.9 présente les différentes évolutions de valeurs liquidatives obtenues et le tableau 6.10 présente les profils rendement-risque des différentes stratégies.

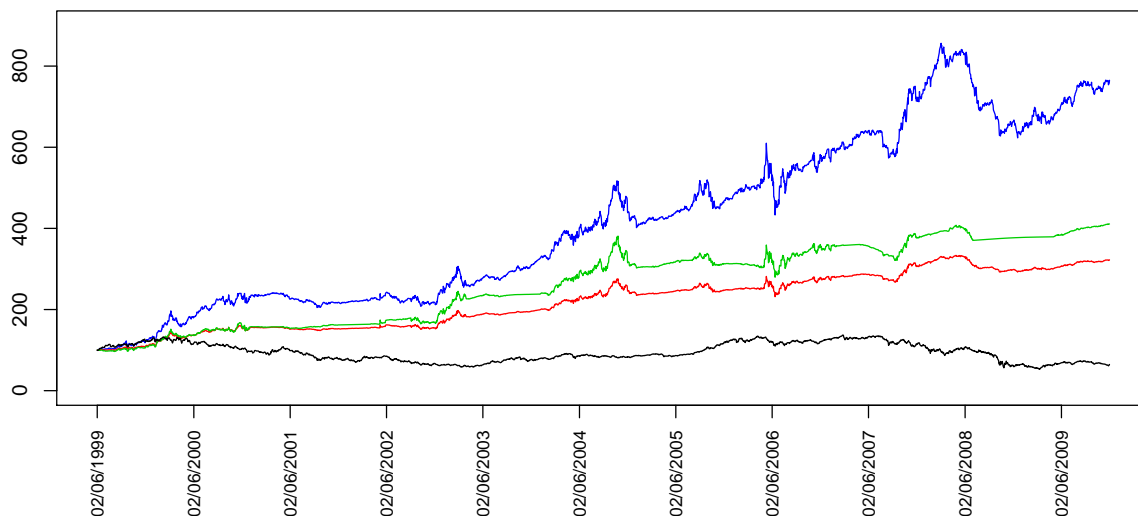


FIGURE 6.9 – Évolution de l'allocation stratégique proposée en utilisant les allocations ρ_1 , ρ_2 et ρ_3 , entre le 01/06/99 et le 01/12/09, en investissant 100 le 01/06/99. La stratégie utilisant ρ_1 est en vert, celle utilisant ρ_2 est en bleu et celle utilisant ρ_3 est en rouge. On a représenté, en noir, l'évolution d'un investissement dans le *S&P* 500.

	Esp. Rdt Ann.	E.-T. Ann.
ρ_1	14.09%	14.73%
ρ_2	20.59%	19.31%
ρ_3	11.39%	11.17%
S&P 500	-0.01%	22.65%

TABLEAU 6.10 – Profil rendement/risque des stratégies utilisant ρ_1 , ρ_2 et ρ_3 comparées au profil du S&P 500 sur la période allant du 01/06/99 au 01/12/09.

Les résultats montrent que l'allocation stratégique combinant horloge d'investissement et stratégies momentum permet d'obtenir des espérances de rendement bien supérieures à celles d'un investissement « buy-and-hold » dans le S&P 500 pour des écart-types de rendements inférieurs à ceux du S&P 500.

Néanmoins, ces performances sont à relativiser puisqu'aucun coût de rebalancement n'a été pris en compte. Ces coûts de rebalancement dépendent de la différence entre la vente des titres détenus durant le mois passé et l'achat des titres que l'on détiendra durant le mois suivant. Puisque les rebalancements du portefeuille sont mensuels, il est possible d'estimer que le rendement moyen mensuel de la stratégie utilisant l'allocation ρ_2 est de $(1.2059)^{\frac{1}{12}} - 1 \approx 1.57\%$. Il est alors possible d'estimer un coût de

rebalancement moyen que peut supporter chaque rebalancement mensuel en utilisant la formule suivante $(1 - c) \times V_t \times (1 + 1.57\%) = V_{t+1}$ où c correspond au coût de rebalancement. Puisque l'on souhaite $V_{t+1} > V_t$, il faut que $1.57\% - c - 1.57\%c > 0$, ce qui nous donne finalement $\frac{1.57\%}{1+1.57\%} > c$. Il faut donc que les coûts de rebalancement moyens appliqués chaque mois n'excèdent pas $\frac{1.57\%}{1+1.57\%} = 1.54\%$ de la valeur totale du portefeuille pour que cette stratégie est une espérance de rendement non nulle.

6.3 Conclusion

Dans ce chapitre, nous avons proposé une méthode pour diversifier un portefeuille d'investissement basée sur une classification ascendante hiérarchique construite en utilisant une mesure de l'indépendance de l'évolution des tendances des différents actifs considérés. La méthode, appliquée sur cinquante indices de prix représentant chacun un investissement dans une classe spécifique d'actifs, nous permet de retenir 23 classes d'actifs différentes constituant un portefeuille suffisamment diversifié. Nous avons ensuite discuté de l'allocation de ce portefeuille. Nous avons proposé d'utiliser des stratégies momentum réglées par une horloge d'investissement construite à partir des indicateurs synthétisant l'information macroéconomique qui ont été présentés dans le Chapitre 5. Il est intéressant de remarquer que l'allocation issue de l'horloge d'investissement anticipe correctement le krach boursier d'Octobre 2008 dû à la crise des subprimes en faisant investir le portefeuille en liquidités et obligations. Elle offre par ailleurs une espérance de rendement annualisée bien supérieure à celle du marché action.

Conclusion générale

Cette thèse aborde un certain nombre de points essentiels en gestion de portefeuille : l'estimation de tendances, la prédiction de tendances, la synthétisation de l'information disponible aux investisseurs, la diversification de portefeuille et son allocation.

Estimation non-paramétrique de tendances Concernant l'estimation de tendances, nous avons fait le lien entre la définition d'une stratégie d'investissement optimale et l'estimation de tendances au sein des rendements d'un investissement (Chapitre 2). Nous apportons également une contribution en comparant plusieurs types d'estimation non-paramétrique de tendances (Chapitre 3). Nous avons conclu que la régression non-paramétrique par décomposition modale empirique fournissait des résultats intéressants pour cette estimation. En effet, la tendance ainsi estimée est en général proche des données et le bruit résultant est, lui, stationnaire au second ordre. Néanmoins, afin d'opérer cette régression, il faut prédéfinir un seuil permettant l'estimation non-linéaire de la tendance. Nous avons proposé de définir ce seuil par rapport au comportement imprévisible des fonctions modales intrinsèques. Pour cela, nous avons développé un test statistique d'imprévisibilité d'une série temporelle se basant sur la reconstruction de l'espace des phases de cette série (Chapitre 4). Les résultats empiriques de ce test ont été comparés à celui du test de Portmanteau de Ljung-Box sur des séries générées. Les résultats obtenus ont montré sa pertinence. La tendance ainsi estimée est assurée d'être prévisible.

Prédiction des tendances Concernant la prédiction de tendances, nous avons utilisé le comportement chaotique de la tendance estimée afin de la prédire (Chapitre 4). Dans ce but, nous avons présenté une méthode de prédiction non-linéaire basée sur une régression non-paramétrique polynomiale locale. Nous avons notamment fait part des difficultés pratiques que peut engendrer une telle méthode de prédiction lors de son implémentation. Nous avons donc présenté un certain nombre de solutions concernant l'amélioration de la vitesse de recherche des plus proches voisins dans le passé, les problèmes d'inversion de matrices ou la détermination des hyperparamètres. Nous avons également étendu cette méthode au cas multivarié afin de prédire des ensembles de tendances qui ne sont pas indépendantes les unes des autres.

La diversification de portefeuille Concernant la diversification de portefeuille, nous avons présenté une méthodologie permettant de composer un portefeuille possédant des actifs ayant des tendances de rendements évoluant le plus indépendamment possible (Chapitre 6). Cette méthodologie se base sur une classification ascendante hiérarchique. Elle se déroule en trois étapes : détermination

d'un nombre de classes minimale nécessaire à l'homogénéisation de ces classes vis-à-vis de la réduction de la partie diversifiable de la variance du portefeuille, minimisation de cette partie diversifiable de la variance afin de déterminer le nombre exact de classes nécessaires, puis minimisation de la partie non-diversifiable de la variance du portefeuille afin de sélectionner les actifs qui feront partis du portefeuille diversifié.

Synthétisation de l'information disponible aux investisseurs Concernant la synthétisation de l'information disponible aux investisseurs, nous avons présenté une méthodologie basée sur des chaînes de Markov cachées multivariées qualitatives afin de construire des indicateurs de points de retournement de tendances que nous avons également appelé « indicateurs synthétique d'augmentation » (Chapitre 5). Ces indicateurs permettent de distinguer clairement des phases de hausses ou de baisses parmi un ensemble de variables sous-jacentes. Nous avons ainsi pu construire des indicateurs liés au contexte économique mondial, à l'inflation et aux croyances des investisseurs.

L'allocation stratégique de portefeuille Concernant l'allocation stratégique de portefeuille, nous avons proposé une manière originale d'allouer un portefeuille composé de plusieurs classes d'actifs en combinant des stratégies « momentums » à une stratégie basée sur une horloge d'investissement (Chapitre 6). Cette horloge d'investissement est fondée sur les indicateurs synthétique d'augmentation présentés dans le Chapitre 5.

Perspectives

Tendance locale et effets de bord La régression non-paramétrique par décomposition modale empirique possède un certain désavantage au niveau du bord de l'échantillon puisque la tendance ainsi estimée peut être révisée de manière significative lorsque de nouvelles données sont disponibles. Nous avons proposé de prédire les révisions d'estimation à l'aide d'un modèle ARMA(p, q) (voir paragraphe 4.4). Il semblerait intéressant de tester d'autres méthodes de prédiction des révisions d'estimation afin d'atténuer l'effet de bord.

Utilisation d'indicateurs de croyances des investisseurs Dans le Chapitre 6, nous avons utilisé des indicateurs liés à la macroéconomie. Cependant, dans le Chapitre 5, nous avons également présenté un indicateur lié aux croyances des investisseurs. Il semblerait donc intéressant de prendre également en compte ce type d'indicateur dans l'allocation stratégique proposée. Il faudrait alors combiner une allocation définie par l'horloge d'investissement avec celle définie par les croyances des investisseurs. Ce type d'indicateur servirait à limiter les investissements dans les actifs risqués.

Allocation stratégique de portefeuille Dans le Chapitre 6, nous avons présenté une allocation stratégique de portefeuille basée sur des indicateurs synthétisant l'information disponible et sur les momentums des différentes classes d'actifs. On pourrait également basée cette allocation sur une assurance de portefeuille à coussin. L'assurance de portefeuille à coussin est un arbitrage entre actifs risqués et non risqués. C'est une stratégie d'allocation d'actifs dynamique dont le but est de maximiser l'exposition dans les actifs risqués tout en garantissant un rendement minimum à l'issue de la période prédéfinie. En fonction de l'évolution des actifs risqués, des règles mécaniques d'achat et de vente sont appliquées grâce à divers paramètres définis au démarrage de façon à conserver une part constante d'exposition au risque. En fonction de la nature de la classe d'actifs risqués à laquelle on recherche

6.3. CONCLUSION

une exposition, on définit un niveau de « coefficient multiplicateur » déterminant le niveau de l'allocation en actifs risqués. Il semblerait intéressant d'utiliser les indicateurs synthétisant l'information disponible présentés dans notre Chapitre 5 afin d'estimer un niveau de coefficient multiplicateur au sein de ce type d'allocation.

Bibliographie

- [Ach(1995)] S.B.Achelis, *Technical analysis from A to Z : covers every trading tool – from the Absolute Breath Index to the Zig Zag*, Probus Pub., 1995
- [Ale(1961)] Alexander, *Price movements in speculative market : Trends or random walks*, Industrial Management Review, Vol 2, 1961, pp.7-26
- [Ale(1964)] Alexander, *Price movements in speculative market : Trends or random walks, Number 2*, Industrial Management Review, Vol 5, pp.25-46
- [Alp(1982)] M.Alpert, H.Raiffa, *A progress report on the training of probability assessors*, paru dans D.Kahneman, P.Slovic, A.Tversky, *Judgment under uncertainty : Heuristics and Biases*, New York : Cambridge University Press, 1982, pp.294-305
- [Ant(2001)] A.Antonadis, J.Fan, *Regularization of wavelet approximations*, Journal of the American Statistical Association, Vol.96, 2001, pp.939-955
- [Ari(1987)] R.A.Ariel, *A monthly effect in stock returns*, Journal of Financial Economics, Vol.17, 1992, pp.161-174
- [Art(1999)] P.Artzner, F.Delbaen, J.-M. Eber, D.Heath, *Coherent measures of risk*, Mathematical Finance, Vol.9, Issue 3, Juillet 1999, pp.203-228
- [Art(2002)] L.M.Artiles Martinez, B.Y. Levit, *Adaptive minimax regression on the interval*, report of the Eindhoven University of Technology, 2002
- [Avr(2006)] D.Avramov, T.Chordia, *Asset pricing models and financial market anomalies*, Review of Financial Studies, Vol.19, N°3, 2006, pp.1001-1040.
- [Baj(1998)] I.Bajeux-Besnaiou, R.Portait, *Dynamic Asset Allocation in a Mean-Variance Framework*, Management Science, Vol.44, N°11, Novembre 1998, pp.S79-S95
- [Bal(1967)] G.H.Ball, D.J.Hall *A clustering technique for summarizing multivariate data*, Behavioral science, Vol.12, N°2, Mars 1967, pp.153-155
- [Ban(2005)] A.Banerjee, G.Urga, *Modelling structural breaks, long memory and stock market volatility : an overview*, Journal of Econometrics, Vol.129, 2005, pp.1-34
- [Ban(1981)] R.W.Banz, *The relationship between return and market value of common stocks*, Journal of Financial Economics, Vol.9, 1981, pp.3-18
- [Bar(1998)] N.Barberis, A.Shleifer, R.Vishny, *A model of investor sentiment*, Journal of Financial Economics, Vol.49, 1998, pp.307-343

- [Bar(2009)] J. Bardaji, F. Tallet, *Detecting economic regimes in France : a qualitative markov-switching indicator using mixed frequency data*, Document de travail, INSEE, disponible à http://www.insee.fr/fr/publications-et-services/docs_doc_travail/G2009-12.pdf
- [Bau(1966)] L.E.Baum, T.Petrie, *Statistical inference for probabilistic functions of finite state Markov chains*, Annals of Mathematical Statistics, Vol. 37, 1966, pp.1554-1563
- [Bau(1970)] L.E.Baum, T.Petrie, G.Soules, N.Weiss, *A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains*, Annals of Mathematical Statistics, Vol. 41, N°1, 1970, pp.164-171
- [Bax(1999)] M.Baxter, R.G.King, *Measuring business cycles : approximate band-pass filters for economic time series*, The Review of Economics and Statistics, MIT Press, Vol.81, N°4, Novembre 1999, pp.575-593
- [Bel(1982)] D.E.Bell, *Regret in decision making under uncertainty*, Operations research, Vol.30, Issue 5, Sept.-Oct. 1982, pp.961-981
- [Bel(2002)] Bellone, Gautier, *Predicting economic downturns through a financial qualitative hidden Markov model*, Working Paper, disponible à <http://bellone.ensae.net/bellonepaper.html>
- [Ben(1995)] S.Benartzi, R.H.Thaler, *Myopic loss aversion and the equity premium puzzle*, The Quarterly Journal of Economics, MIT Press, Vol.110, Issue 1, February 1995, pp.73-92
- [Ben(1975)] J.L.Bentley, *Multidimensional binary search trees used for associative searching*, Commun. ACM 18, Vol.9, pp.509-517
- [Ber(2004)] S.A.Berrani, *Recherche approximative de plus proches voisins avec contrôle probabiliste de la précision ; application à la recherche d'images par le contenu*, Thèse de l'université de Rennes 1, 2004
- [Bik(2000)] S.Bikhchandani, S.Sharma, *Herd behavior in financial markets : a review*, IMF Working Paper N°00/48, Mars 2000
- [Bjo(2001)] A.Björkström, *Ridge regression and inverse problems*, Technical report, Stockholm University, 2001
- [Bla(1976)] F.Black, *Studies in stock price volatility changes*, American Statistical Association, Proceedings of the Business and Economic Statistics Section, 1976, pp.177-181
- [Bla(1973)] F.Black, M.Scholes *The pricing of options and corporate liabilities*, Journal of Political Economy, Vol.81, Mai-Juin 1973, pp.637-654
- [Boc(2002)] S.Boccaletti, D.L.Valladares, L.M.Pecora, H.P.Geffert, T.Carroll, *Reconstructing embedding spaces of coupled dynamical systems from multivariate data*, Physical Review E, Vol.65, 2002, 035204-pp.1-4
- [Bol(1996)] T.Bollerslev, H.O.Mikkelsen, *Modeling and pricing long memory in stock market volatility*, Journal of Econometrics, Vol.73, N°1, Juillet 1996, pp.151-184
- [Bol(2008)] T.Bollerslev, G.Tauchen, H. Zhou, *Expected stock returns and variance risk premia*, Review of Financial Studies, Forthcoming
- [Bor(2005)] I. Borg, P.J.F. Groenen, *Modern multidimensional scaling : theory and applications, 2nd edition*, Springer, New York
- [Bou(2001)] J-P.Bouchaud, A.Matacz, M.Potters, *Leverage effect in financial markets : the retarded volatility model*, Physical Review Letters, Vol.87, Issue 22, 26 novembre 2001, 228701

- [Bre(1989)] W.Breen, L.R.Glosten, R.Jagannathan, *Economic significance of predictable variations in stock index returns*, Journal of Finance, Vol.44, 1989, pp.1177-1189
- [Bre(1973)] R.P.Brent, *Algorithms for minimization without derivatives*, Prentice-Hall, Englewood Cliffs, N.J., 1973
- [Bro(1993)] M.Brockmann, T.Gasser et E.Herrmann, *Locally adaptive bandwidth choice for kernel regression estimators*, Journal of the American Statistical Association, Vol.88, N°424, Décembre 1993, pp.1302-1309
- [Bru(2004)] J.Brusa, P.Liu, *The day-of-the-week and the week-of-the-month effects : an analysis of investors' trading activities*, Review of Quantitative Finance and Accounting, Vol.23, Issue 1, 2004, pp.19-30
- [Bur(1946)] A.F. Burns, W.C.Mitchell, *Measuring Business Cycles*, NBER, 1946, New York
- [Cam(1997)] J.Y.Campbell, A.W.Lo, A.C.MacKinlay *The econometrics of financial markets*, Princeton
- [Cam(1988)] J.Campbell, R.Shiller, *Stock prices, earnings and expected dividends*, Journal of Finance, Vol.46, 1988, pp.661-676
- [Cam(1989)] J.Y.Campbell, R.J.Shiller, *The dividend-price ratio and expectations of future dividends and discount factors*, Review of Financial Studies, Vol.1, N°3, 1989, pp.195-228.
- [Cao(1997)] L.Cao, *Practical method for determining the minimum embedding dimension of a scalar time series*, Physica D : Nonlinear Phenomena, Vol.110, Décembre 1997, pp.43-50
- [Car(1997)] M.M.Carhart, *On persistence in mutual fund performance*, Journal of Finance, Vol.52, 1997, pp.57-82.
- [Cex(2005)] J.C.Cexus, *Analyse des signaux non-stationnaires par Transformation de Huang, Opérateur de Teager-Kaiser, et transformation de Huang-Teager*, Thèse de l'université de Rennes 1, 2005
- [Cha(2000)] M.Chauvet, S.Potter, *Coincident and leading indicators of the stock market*, Journal of Empirical Finance, Vol. 7(1), Mai 2000, pp.87-111
- [Che(1986)] N-F.Chen, R.R.Roll, S.A.Ross *Economic forces and the stock market*, Journal of Business, Vol.59, N°3, 1986, pp.383-404.
- [Che(2005)] T.Chellathurai, D.Draviam, *Dynamic Portfolio selection with nonlinear transaction costs*, Proceedings of the Royal Society Mathematical, Physical and Engineering Sciences, Vol.461, N°2062, 2005, pp.3183-3212
- [Che(1997)] M.-Y.Cheng, J.Fan, J.S.Marron, *On automatic boundary corrections*, Annals of Statistics, Vol.25, N°4, pp.1691-1708
- [Chr(2003)] L.J.Christiano, T.J.Fitzgerald, *The band-pass filter*, International Economic Review, Vol.44, N°2, 2003, pp.435-465
- [Cla(1994)] R.G.Clark, S.Krase, M.Statman *Tracking errors, regret, and tactical asset allocation*, Journal of Portfolio Management, Vol.20, Issue 3, Spring 1994, pp.16-24
- [Cra(1961)] H.Cramér, *On some classes of nonstationary stochastic processes*, Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 2, Juin-Juillet 1960, pp. 57-78

- [Cra(1962)] H.Cramér, *Décompositions orthogonales de certains processus stochastiques*, Annales scientifiques de l'université de Clermont-Ferrand 2, tome 8, série Mathématiques, N°2, 1962, p.15-21
- [Dan(1993)] G.B.Dantzig, G.Infanger, *Multi-stage stochastic linear programs for portfolio optimization*, Annals of Operations Research, Vol.45, pp.59-76
- [Dau(1992)] I.Daubechies, *Ten lectures on wavelets*, SIAM, 1992
- [DeB(1985)] W.F.M.DeBondt, R.Thaler, *Does the stock market overreact?*, Journal of Finance, Vol.40, 1985, pp.793-808
- [DeD(2010)] M.De Domenico, M. Ali Ghorbani, *Chaos and scaling in daily river flow*, arXiv, 2010
- [DeL(1980)] J. de Leeuw, W. Heiser, *Multidimensional scaling with restrictions on the configurations*, in P.R. Krishnaiah, Multivariate Analysis V, North-Holland, 1980, pp.501-522
- [Dev(1985)] P.A. Devijver, *Baum's forward backward algorithm revisited*, Pattern Recognition Letters, Vol.3, 1985, pp.369-373
- [Dhr(1984)] P.J.Dhrymes, I.Friend, N.B.Gultekin *A critical reexamination of the empirical evidence on the Arbitrage Pricing Theory*, Journal of Finance, Vol.39, Juin 1984, pp.323-346.
- [Din(1993)] Z.Ding, C.W.Granger, R.F.Engle, *A long memory property of stock market returns and a new model*, Journal of Empirical Finance, Vol.1, N°1, Juin 1993, pp.83-106
- [Don(1995)] D.L.Donoho, *Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition*, Applied and computational harmonic analysis, 1995, pp.101-126
- [Dyb(2005)] P.H.Dybvig, *Mean-Variance Portfolio Rebalancing with Transaction Costs*, Working Paper, <http://phildybvig.com/papers/tcost1per2.pdf>, Washington University, Janvier 2005
- [Edw(2001)] R.D.Edwards, J.Magee, *Technical analysis of stock trends*, W.H.C. Bassetti, 2001
- [Efr(1983)] B.Efron, G.Gong, *A leisurely look at the bootstrap, the jackknife and cross-validation*, The American Statistician, Vol.37, N°1, Février 1983, pp.36-48
- [Emb(2003)] P.Embrechts, F.Lindskog, A.J.McNeil, *Modelling dependence with copulas and applications to risk management*, in S.T.Rachev, Handbook of heavy tailed distributions in finance, Elsevier/North-Holland, Amsterdam, pp.329-384.
- [Eng(1994)] J.Engle, *A simple wavelet approach to nonparametric regression from recursive partitioning schemes*, Journal of Multivariate Analysis, Vol. 49, 1994, pp.242-254
- [Eng(2002)] R.F.Engle, *Dynamic conditional correlation - a simple class of multivariate GARCH models*, Journal of Business and Economic Statistics, Vol. 20, Issue 3, pp.339-350
- [Eve(1981)] B.S.Everitt, D.J. Hand, *Finite Mixture Distributions*, 1981, Chapman and Hall, London
- [Fam1(1965)] E.F.Fama, *The behaviour of stock market prices*, Journal of Business, Vol.38, N°1, Janvier 1965, pp.34-105
- [Fam2(1965)] E.F.Fama, *Random walks in stock market prices*, Financial Analysis Journal, Vol.51, N°1, Septembre-Octobre 1965, pp.55-59
- [Fam(1970)] E.F.Fama, *Efficient capital markets : a review of theory and empirical work*, Journal of Finance, Vol.25, N°2, Mai 1970, pp.383-417
- [Fam(1981)] E.F.Fama, *Stock returns, real activity, inflation and money*, American Economic Review, Vol.71, 1981, pp.545-564.

- [Fam(1988)] E.F.Fama, K.R.French, *Dividend yields and expected stock returns*, Journal of Financial Economics, Vol.22, N°1, Octobre 1988, pp.3-25.
- [Fam(1992)] E.F.Fama, K.R.French, *The cross-section of stock returns*, Journal of Finance, Vol.47, 1992, pp.427-465
- [Fam(1993)] E.F.Fama, K.R.French, *Common risk factors in the returns on stocks and bonds*, Journal of Financial Economics, Vol.33, N°1, Février 1993, pp.3-56.
- [Fam(2003)] E.F.Fama, K.R.French, *The capital asset pricing model : theory and evidence*, The Journal of Economic Perspectives, Vol.18, N°3, Été 2004, pp.25-46.
- [Fam(1977)] E.F.Fama, G.W.Schwert, *Asset returns and inflation*, Journal of Financial Economics, Vol.5, 1977, pp.115-146.
- [Fan(1996)] J.Fan, I.Gijbels, *Local polynomial modelling and its applications*, Chapman & Hall, London, 1996
- [Fan(2001)] J.Fan, R.Li, *Variable selection via nonconcave penalised likelihood and its oracle properties*, Journal of the American Statistical Association, Vol.96 , 2001, pp.1348-1360
- [Fle(1975)] W.H.Fleming, R.W.Rishel, *Deterministic and stochastic optimal control*, Springer, New-York
- [Fra(1986)] A.Fraser, H.Swinney, *Independent coordinates for strange attractors from mutual information*, Physical Review A, Vol.33, pp.1134-1140
- [Fre(1980)] K.R.French, *Stock returns and the week-end effect*, Journal of Financial Economics, Vol.8, 1980, pp.55-70
- [Fri(1977)] J.H.Friedman, J.L.Bentley, R.A.Finkel, *An algorithm for finding best matches in logarithmic expected time*, ACM Transactions on Mathematical Software, Vol.3, Septembre 1977, pp.209-226
- [Gab(1939)] G.Szegö, *Orthogonal Polynomials*, American Mathematical Society Colloquium Publications, Vol.23, New York, 1939
- [Gal(1973)] A.R.Gallant, W.A.Fuller, *Fitting segmented polynomial regression models whose joint points have to be estimated*, Journal of the American Statistical Association, Vol.68, N°341, Mars 1973, pp.144-147
- [Gei(1975)] S.Geisser, *The predictive sample reuse method*, Journal of the American Statistical Association, Vol.70, N°350, Juin 1975, pp.320-328
- [Gem(2002)] H.Geman, *Pure jump Lévy processes for asset price modelling*, Journal of Banking and Finance, Vol.26, Issue 7, 2002, pp.1297-1316
- [Gew(1983)] J.Geweke, S.Porter-Hudak, *The estimation and application of long-memory time series models*, Journal of Time Series Analysis, Vol.4, 1983, pp.221-238
- [Gia(2007)] Q. Gai Gianetto, *Analyse de la surperformance des fonds*, Mémoire de Master 2, Telecom Bretagne - Université de Rennes 1, 2007
- [Gib(1981)] M.R.Gibbons, P.Hess, *Day of the week effects and asset returns*, Journal of Business, Vol.54, 1981, pp.579-596
- [Gom(2001)] V.Gomez, *The use of Butterworth filters for trend and cycle estimation in economic time series*, Journal of Business and Economic Statistics, Vol.19, N°3, Juillet 2001, pp.365-373
- [Gor(1956)] M.Gordon, E.Shapiro, *Capital equipment analysis : the required rate of profit*, Management Science, Vol.3, 1956, pp.102-110.

- [Gor(1962)] M.Gordon, *The investment, financing, and valuation of the corporation*, Homewood, 1962
- [Gra(1949)] B.Graham, *The intelligent investor*, 4ème édition révisée, Harper & Row, Publishers, Inc., New York
- [Gré(2000)] S.Grégoir, F.Lenglart, *Measuring the probability of a business cycle turning point by using a multivariate qualitative hidden markov model*, Journal of Forecasting, Vol.19, 2000, pp.39-58
- [Gri(1993)] W.E.Griffiths, R.C.Hill, G.G. Judge, *Learning and practicing econometrics*, John Wiley & Sons, New-York, 1993
- [Gri(1995)] M.Grinblatt, S.Titman, R.Wermers, *Momentum investment strategies, portfolio performance, and herding : a study of mutual fund behavior*, The American Economic Review, Vol.85, Issue 5, 1995, pp.1088-1105
- [Ham(1922)] W.P.Hamilton, *The stock market barometer*, Harper & Brothers, New York. Reprint in 1998 by John Wiley & Sons, Inc., New York
- [Ham(1989)] J.D. Hamilton, *A new approach to the economic analysis of nonstationary time series and the business cycle*, Econometrica, Vol.57, 1989, pp.357-384
- [Ham(1994)] J.D.Hamilton, *Time Series Analysis*, Chapitres 1-3, Princeton University Press, Princeton, 1994
- [Ham(1985)] S.M.Hammel, C.K.R.T.Jones, J.V.Moloney, *Global dynamical behavior of the optical-field in a ring cavity*, Journal of Optical Society American B, Vol. 2, pp.552-564
- [Har(2003)] A.C.Harvey, T.M.Trimbur, *General model-based filters for extracting cycles and trends in economic time series*, The Review of Economics and Statistics, Vol.85, N°2, Mai 2003, pp.244-255
- [Hau(1988)] R.A.Haugen, J.Lakonishok, *The incredible January effect*, Dow Jones-Irwin, Homewood, IL.
- [Hen(1996)] C.R.Hensel, W.T.Ziemba, *Investment results from exploiting turn-of-the-month effects*, Journal of Portfolio Management, Vol.22, Issue 3, 1996, pp.17-23
- [Her(2005)] F.Herzog, *Strategic portfolio management for long term investments : an optimal control approach*, PhD Thesis, Diss. ETH N°16137, ETH Zurich
- [Hod(1981)] R.J. Hodrick, E.C. Prescott, *Postwar U.S. business cycles : an empirical investigation*, Discussion Papers 451, Northwestern University, Center for Mathematical Studies in Economics and Management Science
- [Hua(1998)] N.E.Huang, Z.Shen, S.R.Long, M.C.Wu, H.H.Shih, Q.Zheng, N.-C. Yen, C.C.Tung, H.H. Liu *The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis*, Proceedings of the Royal Society London A., Vol.454, 1998, pp.903-995
- [Hub(2001)] G.Huberman, T.Regev, *Contagious speculation and a cure for cancer : a nonevent that made stock prices soar*, The Journal of Finance, MIT Press, Vol.56, Issue 1, 2001, pp.387-396
- [Huk(1993)] J.P. Huke, *Embedding nonlinear dynamical systems : a guide to Takens' theorem*, Technical Report, DRA Malvern, UK, 1993
- [Hut(1990)] M.F.Hutchinson, *A stochastic estimator of the trace influence matrix for laplacian smoothing splines*, Communication in Statistics - Simulation and Computation, Vol.19, N°2, 1990, pp.433-450
- [Jay(1994)] A.W.Jayawardena, F.Lai, *Analysis and prediction of chaos in rainfall and streamflow time series*, Journal of Hydrology, Vol.153, 1994, pp.23-52

- [Jeg(1993)] N.Jegadeesh, S.Titman, *Returns to buying winners and selling losers : implications for stock market efficiency*, Journal of Finance, Vol.48, 1993, pp.65-91
- [Jeg(1999)] N.Jegadeesh, S.Titman, *Profitability of momentum strategies : an evaluation of alternative explanations* NBER Working paper N°7159, 1999
- [Jen(1978)] M.C.Jensen, *Some anomalous evidence regarding market efficiency*, Journal of Financial Economics, Vol.6, 1978, pp.95-101.
- [Jol(2002)] I.T.Jolliffe, *Principal component analysis, second edition*, Springer-Verlag, New-York, 2002
- [Jup(1978)] D.L.B. Jupp, *Approximation to data by splines with free knots*, SIAM Journal on Numerical Analysis, Vol.15, N°2, Avril 1978, pp.328-343
- [Kah(1979)] D.Kahneman, A.Tversky, *Prospect theory : an analysis of decisions under risk*, Econometrica, Vol.47, Issue 2, 1979, pp.263-291
- [Kah(1982)] D.Kahneman, P.Slovic, A.Tversky, *Judgment under uncertainty : Heuristics and Biases*, New York : Cambridge University Press, 1982
- [Kal(1993)] A.J.Kalotay, G.O.Williams, F.J.Fabozzi, *A model for valuing bonds and embedded options*, Financial Analysts Journal, Vol.49, N°3, Mai-Juin 1993, pp.35-46.
- [Kan(2004)] H.Kantz, T.Schreiber, *Nonlinear time series analysis*, Second Edition, Cambridge University Press, Cambridge, 2004
- [Kei(1984)] D.B.Keim, R.F.Stambaugh, *A further investigation of the weekend effect in stock returns*, Journal of Financial Economics, Vol.12, 1984, pp.13-32
- [Kei(1986)] D.B.Keim, R.F.Stambaugh, *Predicting returns in the stock and bond markets*, Journal of Financial Economics, Vol.17, N°2, Décembre 1986, pp.357-390.
- [Ken(1953)] M.G.Kendall, *The Analysis of Economic Time Series-Part I : Prices*, Journal of The Royal Statistical Society, Vol.116, Issue 1, 1953, pp.11-34
- [Ken(1992)] M.B.Kennel, R.Brown, H.D.I.Abarbanel, *Determining embedding dimension for phase-space reconstruction using a geometrical construction*, Physical Review A, Vol.45, N°6, 1992, pp.3403-3411
- [Kim(1994)] C-W.Kim, J.Park, *Holiday effects and stock returns : further evidence*, Journal of Financial and Quantitative Analysis, Vol.29, 1994, pp.145-157
- [Koh(1992)] R.K.Kohli, T.Kohers, *The week-of-the-month effect in stock returns : the evidence from the S&P Composite Index*, Journal of Economics and Finance, Vol.16, Issue 2, 1992, pp.129-137
- [Kri(2006)] M.Krishnan, S.C.Lim, P.Zhou, *Analysts' herding propensity : theory and evidence from earnings forecasts*, AAA 2007 Financial Accounting & Reporting Section Meeting Paper, September 2006
- [Kro(2002)] P.Krokhmal, J.Palmquist, S.Uryasev, *Portfolio Optimization with conditional value-at-risk objective and constraints*, The Journal of Risk, Vol.4, N°2, Hiver 2001/2002, pp.43-68
- [Kru(1964)] J.B.Kruskal, *Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis*, Psychometrica, Vol.29, pp.1-27, 1964
- [Lai(1997)] D.Laibson, *Golden eggs and hyperbolic discounting*, The Quarterly Journal of Economics, Vol.112, Issue 2, Mai 1997, pp.443-477
- [Lak(1984)] J.Lakonishok, S.Smidt, *Volume and turn-of-the-year behavior*, Journal of Financial Economics, Vol.13, 1984, pp.435-456

- [Lam(1987)] C.G.Lamoureux, P.Poon, *The market reaction to stock splits*, Journal of Finance, Vol.42, Issue 5, Décembre 1987, pp.1347-1370
- [LeB(1992)] B.LeBaron, *Some relations between volatility and serial correlations in stock market returns*, Journal of Business, Vol.65, Issue 2, Avril 1992, pp.199-219
- [Leb(2006)] L.Lebart, M.Piron, A.Morineau, *Statistique exploratoire multidimensionnelle*, 4ème édition, Dunod, 2006
- [Lic(1982)] S.Lichtenstein, B.Fischoff, L.D.Phillips, *Calibration of probabilities : the state of the art to 1980*, paru dans D.Kahneman, P.Slovic, A.Tversky, *Judgment under uncertainty : Heuristics and Biases*, New York : Cambridge University Press, 1982, pp.54-67
- [Lin(1965)] J.Lintner, *The valuation of risk assets and the selection of risky investments in stock portfolio and capital budgets*, Review of Economic and Statistics, Vol.13, 1965, pp.13-37
- [Lin(2009)] L.Lin, Y.Wang, H.Zhou, *Iterative filtering as an alternative for empirical mode decomposition*, Advances in Adaptive Data Analysis, Vol.4, 2009, pp.543-560
- [Lju(1978)] G.M.Ljung, G.E.P.Box, *On a measure of a lack of fit in time series models*, Biometrika, Vol.65, N°2, 1978, pp.297-303
- [Lo(2004)] A.Lo, *The adaptative market hypothesis : market efficiency from an evolutionary perspective*, 2004
- [Lo(1999)] A.W.Lo, A.C.MacKinlay, *A Non-Random Walk Down Wall Street*, Princeton, NJ : Princeton University Press, 1999
- [Loa(1999)] C.R. Loader, *Bandwidth selection : classical or plug-in ?*, Annals of Statistics, Vol.27, N°2, Avril 1999, pp.415-438
- [Lob(1998)] I.N.Lobato, N.E.Savin *Real and spurious long-memory properties of stock market data*, Journal of Business & Economic Statistics, Vol.16, N°3, Juillet 1998, pp.261-268
- [Lob(2006)] M.Lobo, M.Fazel, S.Boyd, *Portfolio optimization with linear and fixed transactions costs*, The Annals of Operations Research, Vol.152, N°1, 2006, pp.341-365
- [Mal(1996)] B.G.Malkiel, *A random walk down wall street*, W.W.Norton & Company, 1996
- [Mal(2003)] B.G.Malkiel, *The efficient market hypothesis and its critics*, Journal of Economic Perspectives, Vol.17, N°1, hiver 2003, pp.59-82
- [Mal(1999)] S.Mallat, *A wavelet tour of signal processing*, 2nd. Edition, Academic Press, San Diego, California, 1999
- [Mal(1973)] C.L.Mallows, *Some comments on C_p* , Technometrics, Vol.15, N°4, Nov.1973, pp.661-675
- [Man(1963)] Mandelbrot, *The Variation of Certain Speculative Prices*, Journal of Business, Vol 36, pp.394-419
- [Mar(1952)] H.Markowitz, *Portfolio selection*, Journal of Finance, Vol.7, Issue 1, Mars 1952, pp.77-91
- [Meh(1985)] R.Mehra, E.C.Prescott, *The equity premium : a puzzle*, Journal of Monetary Economics, Vol.15, 1985, pp.145-161
- [Meh(2003)] R.Mehra, *The equity premium : why is it a puzzle ?*, Mimeo, University of Chicago, Mai 2001
- [Mer(1971)] R.C.Merton, *Optimum consumption and portfolio rules in a continuous-time model*, Journal of Economic Theory, Vol.3, Issue 4, Décembre 1971, pp.373-413

- [Mer(1974)] R.C.Merton, P.A.Samuelson, *Fallacy of the log normal approximation to optimal portfolio decision-making over many periods*, Journal of Financial Economics, Vol.1, Issue 1, Mai 1974, pp.67-84
- [Mer(1976)] R.C.Merton, *Option pricing when underlying stocks are discontinuous*, Journal of Financial Economics, Vol.3, 1976, pp.125-144.
- [Meu(1992)] L.K.Meulbroeck, *An empirical analysis of illegal insider trading*, Journal of Finance, Vol.7, N°5, 1992, pp.1661-1699.
- [Mey(1992)] Y.Meyer, *Wavelets and operators*, Vol.37 of Cambridge Studies in Advanced Mathematics, Cambridge University Press, 1992
- [Jif(2006)] J.Miffre, G.Rallis, *Momentum strategies in commodity futures markets*, Cass Business School Research Papers, Aout 2006
- [Mil(1975)] D.T.Miller, M.Ross, *Self-serving biases in the attribution of causality : fact or fiction ?*, Psychological Bulletin, Vol.82, Issue 2, pp.213-225
- [Mit(2004)] M.Mitchell, T.Pulvino, E.Stafford, *Price pressure around mergers*, Journal of Finance, Vol.59, Issue 1, Février 2004, pp.31-63
- [Moi(2008)] S.Moinas, *Le carnet d'ordres : une revue de littérature*, Finance, Vol.29, N°1, Juin 2008
- [Moo(1995)] Y.-I. Moon, B.Rajagopalan, U.Lall *Estimation of mutual information using kernel density estimators*, Physical Review E, Vol.52, N°3, 1995, pp.2318-2321
- [Mos(1966)] J.Mossin, *Equilibrium in a capital market*, Econometrica, Vol.34, Octobre 1966, pp.768-783.
- [Mül(1987)] H.-G.Müller, U.Stadt Müller, T.Schmitt, *Bandwidth choice and confidence intervals for derivatives of noisy data*, Biometrika, Vol.74, N°4, 1987, pp.743-749
- [Mye(1994)] S.Myers, L.Borucki, *Discounted cash flow estimates of the cost of equity capital*, Financial Markets and Institutions, Blackwell Publishing, Oxford.
- [O'D(1999)] T.O'Donoghue, M.Rabin, *Doing it now or later*, The American Economic Review, Vol.89, Issue 1, Mars 1999, pp.103-124
- [Osl(1995)] C.L.Osler, P.H.K. Chang, *Head and shoulders : not just a flaky pattern*, Reserve bank of New-York, Staff Reports, N°4, Aout 1995
- [Ost(2007)] E.Ostrovsky, Y.Zelikov, *Adaptive Optimal Nonparametric Regression and Density Estimation based on Fourier-Legendre expansion*, Arxiv preprint, Juin 2007
- [Pai(2010)] R.L.Paige, A.A. Trindade, *The Hodrick-Prescott filter : a special case of penalized spline smoothing*, Electronic Journal of Statistics, Vol.4, 2010, pp.856-874
- [Pes(1995)] M.H.Pesaran, A.G.Timmermann, *Predictability of stock returns : robustness and economic significance*, Journal of Finance, Vol.50, 1995, pp.1201-1228
- [Pes(2001)] M.H.Pesaran, A.G.Timmermann, *How costly is it to ignore breaks when forecasting the direction of a time series ?*, Manuscript, Cambridge University and UCSD, 2001
- [Pri(1998)] M.Pring, *Martin Pring's introduction to technical analysis*, McGraw-Hill Professional, 1998
- [Pri(1997)] R.Storn, K.Price, *Differential evolution : a simple and efficient heuristic for global optimization over continuous spaces*, Journal of Global Optimization, Vol.11, N°4, 1997, pp.341-359

- [Rah(2001)] R.Rahi, P.Gottardi, *Efficiency properties of rational expectations equilibria with asymmetric information*, Discussion Paper, N°381, Financial Market Group, London School of Economics and Political Science, 2001.
- [Rei(1983)] M.R.Reinganum, *The anomalous stock market behavior of small firms in January : empirical evidence for tax-loss effects*, Journal of Financial Economics, Vol.12, Issue 1, Juin 1983, pp.89-104
- [Rhe(1932)] R.Rhea, *The Dow theory*, Barron's, New York. Reprint in 1993 by Fraser Publishing Company, Burlington
- [Rit(1989)] J.R.Ritter, N.Chopra, *Portfolio rebalancing and the turn of the year effect*, Journal of Finance, Vol.44, 1989, pp.149-166
- [Rob(1959)] H.V.Roberts, *Stock-Market Patterns and Financial Analysis : Methodological Suggestions*, Journal of Finance, Vol.14, Issue 1, Mars 1959, pp.1-10
- [Rob(1995)] P.M.Robinson, *Gaussian semiparametric estimation of long range dependence*, Annals of statistics, Vol.23, N°5, Octobre 1995, pp.1630-1661
- [Rol(1977)] S.A.Ross, *A critique of the asset pricing theory's tests : part I : on past and potential testability of the theory*, Journal of Financial Economics, Vol.4, 1977, pp.129-176.
- [Rol(1980)] R.Roll, S.A.Ross, *An empirical investigation of the arbitrage pricing theory*, Journal of Finance, Vol.35, 1980, pp.1073-1103.
- [Rol(1984)] P.J.Dhrymes, I.Friend, N.B.Gultekin, *A critical reexamination of the empirical evidence on the Arbitrage Pricing Theory : A reply*, Journal of Finance, Vol.39, N°2, Juin 1984, pp.347-350.
- [Ros(1976)] S.A.Ross, *The arbitrage theory of capital asset pricing*, The Journal of Economic Theory, Vol.13, 1976, pp.341-360.
- [Ros(1993)] M.T.Rosenstein, J.J.Collins, C.J.DeLuca, *A practical method for calculating largest Lyapunov exponents from small data sets*, Physica D, Vol.65, 1993, pp.117-134
- [Rou(1998)] K. Geert Rouwenhorst, *International Momentum Strategies*, Journal of Finance, Vol.53, N°1, Février 1998, pp.267-284
- [Roz(1976)] M.S.Rozeff, W.R.Kinney, *Capital market seasonality : the case of stock returns*, Journal of Financial Economics, Vol.3, 1976, pp.379-402
- [Rup(1995)] D.Ruppert, S.J.Sheather, M.P.Wand, *An effective bandwidth selector for local least squares regression*, Journal of the American Statistical Association, Vol.90, N°432, Décembre 1995, pp.1257-1270
- [Sam(1965)] P.Samuelson, *Proof that properly anticipated prices fluctuate randomly*, Industrial Management Review, Vol.6, N°2, printemps 1965, pp.41-49
- [Sam(1969)] P.A.Samuelson, *Lifetime portfolio selection by dynamic stochastic programming*, Review of Economics and Statistics, pp.68-126
- [Sap(2006)] G.Saporta, *Probabilités, analyse des données et statistique*, Éditions Technip, 2006
- [Sau(1991)] T.Sauer, J.A.Yorke, M.Casdagli, *Embedology*, Journal of Statistical Physics, Vol.65, 1991, pp.579-616
- [Sei(1996)] B.Seifert, T.Gasser, *Finite-sample variance of local polynomials : analysis and solutions*, Journal of the American Statistical Association, Vol.91, N°433, Mars 1996, pp.267-275
- [Sha(1964)] W.Sharpe, *Capital asset prices : a theory of market equilibrium under conditions of risk*, Journal of Finance, Vol. 19, Septembre 1964, pp.425-442

- [Shl(2000)] A.Shleifer, *Inefficient markets. An introduction to behavioral finance*, Oxford University Press, 2000
- [Sve(1981)] O.Svenson, *Are we all less risky and more skilful than our fellow drivers ?*, Acta Psychologica, Vol.47, 1981, pp.143-148
- [Sti(1985)] R.A. Stine, *Bootstrap prediction intervals for regression*, Journal of the American Statistical Association, Vol.80, N°392, Décembre 1985, pp.1026-1031
- [Tak(1981)] F.Takens, *Detecting strange attractors in turbulence*, in D.A.Rand, L.-S.Young, *Dynamical Systems and Turbulence*, Lecture Notes in Mathematics, Vol.898, Springer-Verlag, 1981, pp.366-381
- [Tay(1986)] S.J.Taylor, *Modeling financial time series*, Chichester, Wiley, 1986
- [Tay(1988)] S.E.Taylor, J.D.Brown, *Illusion and well-being : a social psychological perspective on mental health*, Psychological Bulletin, Vol.103, 1988, pp.193-210
- [Tha(1981)] R.H.Thaler, H.M.Shefrin, *An economic theory of self-control*, Journal of Political Economy, Vol.89, Issue 1, Avril 1981, pp.392-406
- [Tha(1999)] R.H.Thaler, *Mental accounting matters*, Journal of Behavioral Decision Making, Vol.12, 1999, pp.183-206
- [Tho(1953)] R. Thorndike, *Who belongs in the family ?*, Psychometrika, Vol.18, N°4, Décembre 1953, pp.263-276
- [Trz(1986)] C.A.Trzinka, *On the number of factors in the Arbitrage Pricing Model*, Journal of Finance, Vol.43, pp.881-892
- [Tsy(2004)] A.B.Tsybakov, *Introduction à la régression non-paramétrique*, collection Mathématiques et applications, Vol.41, 2004
- [Tuk(1960)] J.W.Tukey, *Curves as parameters, and touch estimation*, Proceedings of the fourth berkeley symposium on mathematical statistics and probability, University of California Press, Vol.1, 20 Juin - 30 Juillet 1960, pp.681-694
- [Tve(1974)] A.Tversky, D.Kahneman, *Judgment under uncertainty : Heuristics and Biases*, Science, New Series, Vol.185, Issue 4157, 27 Septembre 1974, pp.1124-1131
- [Tve(1991)] A.Tversky, D.Kahneman, *Loss aversion in riskless choice : a reference-dependent model*, The Quarterly Journal of Economics, MIT Press, Vol.106, Issue 4, Novembre 1991, pp.1039-1061
- [Vie(1991)] P.Vieu, *Nonparametric regression : local optimal bandwidth choice*, Journal of Royal Statistical Society, Série B, Vol.53, pp.453-474
- [Vil(2005)] V.Totik, *Orthogonal polynomials*, Surveys in Approximation Theory, Vol. 1, Novembre 2005, pp.70-125
- [War(1963)] J.H.Ward, *Hierarchical grouping to optimize an objective function*, Journal of the American Statistical Association, Vol.58, pp.238-244
- [Wei(1989)] P.Weil, *The equity premium puzzle and the riskfree rate puzzle*, Journal of Monetary Economics, Vol.24, Issue 2, Novembre 1989, pp.401-421
- [Wei(1980)] N.D.Weinstein, *Unrealistic optimism about future life events*, Journal of Personality and Social Psychology, Vol.39, pp.806-8204
- [Whi(2000)] H.White, *A reality check for data snooping*, Econometrica, Vol.68, 2000, pp.1097-1126

- [Wie(1964)] N.Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*, The MIT Press, 1964
- [Wol(1975)] R.J.Wolosin, S.J.Sherman, A.Till, *Effects of cooperation and competition on responsibility attribution after success and failure*, Journal of Experimental Social Psychology, Vol.9, Issue 3, pp.220-235
- [Wor(1934)] H.Working, *A Random Difference Series for Use in The Analysis of Time Series*, Journal of the American Statistical Association, Vol.29, Issue 185, Mars 1934, pp.11-24
- [Wu(1986)] C.F.J Wu, *Jackknife, bootstrap and other resampling methods in regression analysis*, Annals of Statistics, Vol.14, N°4, 1986, pp.1261-1295
- [Zen(2004)] K.Zeng, M.-X.He, *A simple boundary process technique for Empirical Mode Decomposition*, Proceedings of IEEE International Geoscience and Remote Sensing Symposium, Vol.6, 2004, pp.4258-4261
- [Zha(1993)] P.Zhang, *Model selection via multifold cross validation*, Annals of Statistics, Vol.21, N°1, 1993, pp.299-313
- [Zhou(1999)] X.Y.Zhou, D.Li, *Continuous-time mean-variance portfolio selection : a stochastic LQ framework*, Applied Mathematics & Optimization, Vol.42, N°1, Novembre 1999, pp.19-33