



HAL
open science

Stratégies d'arbitrage systématique multi-classes d'actifs et utilisation de données hétérogènes

Yohan Fereres

► **To cite this version:**

Yohan Fereres. Stratégies d'arbitrage systématique multi-classes d'actifs et utilisation de données hétérogènes. Economies et finances. Université Paris-Est, 2013. Français. NNT : 2013PEST0075 . tel-00987635

HAL Id: tel-00987635

<https://theses.hal.science/tel-00987635v1>

Submitted on 6 May 2014

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.



Thèse
pour l'obtention du titre de
DOCTEUR EN SCIENCES DE GESTION

présentée et soutenue publiquement le 10 décembre 2013 par :

Yohan FERERES

**Stratégies d'arbitrage systématique multi-classes
d'actifs et utilisation de données hétérogènes**

JURY

Directeurs de thèse : **Monsieur Pierre CHOLLET,**
Professeur, Université Montpellier 1

Monsieur Jacques PRINTEMS,
Maître de conférences HDR, Université Paris-Est Créteil

Rapporteurs : **Monsieur Jean-Paul LAURENT,**
Professeur, Université Paris 1

Monsieur Jean-Laurent VIVIANI,
Professeur, Université Rennes 1

Suffragants : **Monsieur Gilles PAGES,**
Professeur, Université Paris 6

Monsieur Olivier RAMOND,
Professeur, Université Paris-Est Créteil

L'Université n'entend donner aucune approbation ni improbation aux opinions émises dans les thèses ; ces opinions doivent être considérées comme propres à leurs auteurs.

Remerciements

Je tiens à remercier mes directeurs de thèse qui m'ont accompagné et soutenu dans mon travail de recherche. Cette initiation à la recherche fut dans un premier temps difficile mais finalement nécessaire à ma soif d'apprentissage, de compréhension, d'analyse et de synthèse.

Le soutien de mes directeurs a été un facteur déterminant dans la conduite de cette expérience humaine et intellectuelle. Précisément, je tiens à remercier Monsieur le Professeur Chollet pour la confiance qu'il a eue en ma capacité à convertir ma pensée pragmatique à un univers plus académique dont l'exigence est élevée. Il a régulièrement su orienter ce travail d'investigation pour lui donner toute sa consistance. Je tiens également à exprimer toute ma gratitude et mes remerciements à Monsieur Printems qui m'a aidé à structurer ma pensée, à progresser sur divers aspects quantitatifs, informatiques et m'a toujours soutenu dans mes diverses démarches. Sans vous cette thèse ne serait restée qu'un projet de recherche abstrait.

Je tiens à témoigner toute ma gratitude envers Monsieur Thauvron qui m'a donné une seconde chance académique lors d'une reprise d'étude en Master 2 de gestion de portefeuille. Je suis reconnaissant envers Monsieur le Professeur Ramond qui m'a exprimé avec franchise et considération un certain nombre de conseils en matière de recherche en Finance. Je tiens également à remercier Monsieur Blanchard qui m'a permis de compléter ma formation en statistique et en Econométrie ainsi que Monsieur le Professeur Legendre qui a cru en ma capacité à mener à bien mon travail de thèse. J'exprime toute ma gratitude à Madame la Professeure Pigeyre qui m'a permis de perfectionner la présentation de mes recherches. Je remercie également Monsieur Cellier qui a su orienter ma recherche sur les marchés boursiers.

Je remercie Monsieur le Professeur Laurent d'accepter d'être rapporteur sur mon travail de recherche dont certains aspects concernent à la fois la gestion quantitative de portefeuille et la gestion du risque. Je remercie Monsieur le Professeur Viviani pour sa participation à mon jury de thèse en tant que rapporteur. Ses travaux sur l'optimisation m'ont aidé à avancer dans mes travaux. Je tiens finalement à remercier Monsieur le Professeur Pagès d'accepter de faire partie de mon jury de thèse, ses travaux sur la quantification ayant inspiré mes recherches.

Je tiens à remercier ma famille qui m'a toujours soutenu moralement et encouragé dans la voie de la connaissance et des études. Parmi ma famille, je pense surtout à mes parents, mon frère et mes grands-parents. Leur mérite et leur bienveillance seront toujours là. Je tiens à exprimer toute ma reconnaissance à Albert Nahon qui m'a encouragé lors de mon

travail de thèse de doctorat en Sciences de Gestion. Un remerciement particulier est destiné à Monsieur Abib. Ces derniers ont toujours hissé le savoir comme valeur essentielle.

De nombreux échanges avec divers interlocuteurs du monde de l'entreprise ayant une sensibilité académique ont également participé à l'avancement de mes recherches. Je tiens à remercier Jacques Henry (Pictet Asset Management), Daniel Fermon, Stéphane Mattatia et Nicolas Elbaz (Société Générale CIB), Isaac Chebar (DNCA Finance), Joel Couse et Damien Reboul-Salze (Total Trading), Baruk Pilo (Icône Corporate Finance), Hervé Pilo (ex Partner Exane BNP Paribas), François Chauvet (Aptimum Conseil), Ouri Mimran (Natixis CIB), Djaafar Aballeche et Sylvain Lemaire (Aurel Bgc), Cedric Abid, Charly Abid et Yacine Sikkadour (AJC Formation), Lionel Pupille (Kepler Capital Markets), Steeve Bitton (ex Société Générale CIB), Jérémy Serror (John Locke Investments), Ludovic Cohen (Deutsche Bank), Elie Sellam (Allianz), Eric Edery et Thomas Padovani (Webinfluence Group) et Stéphane Guillot (SGA Finance).

Je n'aurai pas de mots suffisants et assez forts pour remercier ma femme Gaëlle, mon fils Jonas et ma fille Léa pour leur patience, leur soutien et le courage qu'ils m'ont communiqués. Je tiens à leur exprimer mes plus profonds remerciements pour avoir traversé avec moi les aléas qu'engendre un travail de recherche.

SOMMAIRE

<i>Introduction Générale</i>	7
1. Contexte de la recherche.....	7
2. Problématique générale et questions de recherche	9
3. Fondements théoriques et cadre conceptuel de la recherche	10
4. Contributions de la recherche	15
5. Démarche méthodologique de la recherche	16
6. Plan de la recherche	18
<i>Chapitre 1. Littérature et concepts clés</i>	22
1. Hétérogénéité des données en Sciences de Gestion.....	23
2. Données hétérogènes en univers multi-classes d'actifs	30
3. Arbitrage statistique : stratégies historiques et limites	34
4. Diversification de portefeuille	43
5. Outils quantitatifs de la recherche.....	49
<i>Chapitre 2. Méthodes d'arbitrage systématique et prise en compte de données hétérogènes</i>	59
1. Prémices d'une approche systématique	60
2. Arbitrage systématique : une approche paramétrique.....	68
3. Arbitrage systématique : une approche non-paramétrique	76
4. Analyse comparative des approches : paramétrique <i>versus</i> non-paramétrique	80
<i>Chapitre 3. Échantillon de l'étude</i>	83
1. Périodes de l'échantillon.....	83
2. Recueil des données de l'échantillon.....	88
3. Traitements appliqués aux données	104
4. Composition et caractéristiques de l'échantillon	112
<i>Chapitre 4. Portefeuille d'arbitrage « euro neutre » en diversification « naïve »</i>	127
1. Influence des données hétérogènes sur des portefeuilles d'arbitrage « euro neutre » en allocation « naïve ».....	127
2. Éléments de robustesse	157
<i>Chapitre 5. Portefeuille d'arbitrage « euro neutre » en diversification optimale</i>	166
1. Cadre et contraintes de la diversification optimale.....	166
2. Influence des données hétérogènes sur des portefeuilles d'arbitrage « euro neutre » en allocation optimale.....	170
3. Éléments de robustesse	187
4. Analyse comparative des résultats empiriques : diversification « naïve » <i>versus</i> diversification optimale	194
<i>Conclusion Générale</i>	201
1. Principaux résultats de la recherche.....	201

2. Apports de la recherche	209
3. Limites de la recherche	214
4. Perspectives de recherche	215
<i>Bibliographie</i>	218
<i>Liste des tableaux</i>	228
<i>Liste des figures</i>	230
<i>Table des matières</i>	231
<i>Liste des Annexes</i>	234

INTRODUCTION GÉNÉRALE

1. Contexte de la recherche

L'information joue un rôle crucial sur les marchés financiers, notamment dans la formation des prix des actifs. Cette information s'est largement développée et diversifiée au cours des récentes décennies. Elle est de toutes natures: financière, économique, sociale, sociétale... La place de l'information dans un environnement financier qui évolue, constamment et rapidement, est devenue stratégique pour les parties prenantes des marchés d'actifs cotés. Face à une telle évolution, le choix des informations pertinentes demeure une question essentielle tant pour les professionnels des marchés que pour les chercheurs en Finance.

Dans ce contexte, nous souhaitons montrer l'intérêt de l'utilisation en Finance de données de tous types, que nous qualifions de « données hétérogènes ». Notre recherche a pour objet de montrer l'apport de la prise en compte des différents types d'information dans l'élaboration des stratégies d'arbitrage systématique et ceci particulièrement pour une période de temps incluant une phase de crise. Dans notre recherche, la période de crise des *subprimes* est étudiée afin de mieux comprendre comment évolue la volatilité des portefeuilles en présence d'informations multiples. Nous étudions l'influence des différents flux d'information avant, pendant et après la période de crise des *subprimes*. L'enjeu de nos recherches est de mesurer quantitativement l'effet de la prise en compte de données hétérogènes sur la composition et la performance des portefeuilles d'arbitrage en univers multi-classes d'actifs, notamment au travers d'allocations « naïve », puis optimale.

En Finance, le débat autour du clivage quantitatif/qualitatif a été abordé (Gregory-Allen, Stangl, et Shawky, 2009) tant sur le type de données mobilisées que sur le type d'approche en matière de gestion de portefeuille. Cependant, un champ alternatif de réflexion semble être la prise en compte de l'hétérogénéité des variables d'information considérées. Dans le cadre de notre recherche, nous considérons deux types d'information : les données de cotation des marchés financiers et une catégorie de données que nous regroupons sous l'appellation « autres informations de marché ». Une donnée de cotation, est dans notre approche, une mesure qui permet en l'état de prendre une décision sur une base tangible. Les données de cotation résultent principalement de transactions financières¹, et sont considérées

¹ Le prix d'une action d'une entreprise est une donnée de cotation, tout comme le prix d'une obligation du

comme des données homogènes. Les « autres informations de marché » correspondent notamment à l'expression de sentiments d'agents économiques et/ou d'investisseurs. Il s'agit d'indicateurs tels que la confiance (optimisme/pessimisme) dans la conjoncture économique, le sentiment haussier (ou baissier) des investisseurs sur les bourses, la « nervosité » des marchés financiers². Dans le cadre de nos travaux, les données hétérogènes sont définies comme la combinaison de données de cotation et d'« autres informations de marché ».

L'introduction d'« autres informations de marché » comme paramètres formels de nos travaux est motivée par deux éléments. D'une part, les investisseurs ne réagissent pas toujours rationnellement lorsque la peur ou les émotions interviennent dans leur processus de décision (Whaley, 2009)³. D'autre part, les allocations d'actifs ne sont pas toujours effectuées dans des perspectives économiques et financières dans la mesure où des informations « fondamentales » seraient encore sous-exploitées (Wermers et al., 2010). Ainsi, notamment lors d'un moment de peur, les « autres informations de marché » peuvent contribuer à un processus de décision plus pertinent. Le manque d'exploitation des différentes informations disponibles peut être comblé par un travail d'investigation et de tri des données. Certains chercheurs analysent ainsi le lien entre l'économie réelle et l'évolution des marchés financiers (Gregory-Allen, Stangl, et Shawky, 2009 ; Brenner, Pasquariello, et Subrahmanyam, 2009). D'autres recherches soulèvent le manque de travaux entre l'univers-multi-classes d'actifs et les informations « fondamentales » (Blitz et Vliet, 2008 ; Andersen et al., 2007). Nous proposons d'explorer ces perspectives de recherches sur le lien entre données fondamentales et univers multi-classes d'actifs financiers cotés.

Nos recherches portent l'espoir que la gestion dite « quantitative », qui repose que l'utilisation de modèles, soit enrichie par un ensemble d'informations dont les contenus diffèrent et peuvent se compléter. Un des buts de nos investigations est de participer à la recherche contemporaine en Finance en apportant des éléments de réponse tangibles aux discussions relatives aux modèles d'arbitrage statistique en univers multi-classes d'actifs et à leurs paramètres (Khandani et Lo, 2007). Plus techniquement, nous souhaitons mesurer comment la présence de données hétérogènes permet grâce à une plus grande diversification de réduire le risque de portefeuilles d'arbitrage et les pertes probables. Nous souhaitons également évaluer si la performance de portefeuilles d'arbitrage peut être optimisée grâce à l'utilisation de données hétérogènes.

Trésor français.

² Nous détaillerons dans le chapitre 3 l'ensemble des données de notre échantillon d'étude.

³ Les agents économiques ne raisonnent pas toujours comme l'*homo œconomicus* (Pareto, 1906).

2. Problématique générale et questions de recherche

Nos travaux portent sur l'apport des données hétérogènes en Finance. Ce questionnement se forme dans un contexte incluant la crise des *subprimes* qui a accru la sensibilité des intervenants financiers au risque (Khandani et Lo, 2007). En substance, notre problématique générale consiste à s'interroger sur la valeur ajoutée mesurable d'une combinaison de données hétérogènes dans un processus de gestion de portefeuille, plus particulièrement dans un contexte d'arbitrage multi-classes d'actifs financiers. Concrètement, notre recherche vise à évaluer la création de valeur liée à la présence de données hétérogènes, à travers son impact sur l'allocation d'actifs et le *market timing*, et ses effets en termes de rentabilité, de volatilité et de couple rentabilité/risque. Cette valeur potentiellement créée doit être estimable, quantifiable afin de répondre à notre problématique. De celle-ci sont issues trois questions principales qui structurent notre recherche.

Question 1. Quels sont les fondements d'une combinaison de données hétérogènes en Sciences de Gestion et plus particulièrement en Finance ?

Nous posons les bases théoriques et conceptuelles de nos travaux de recherche. Il paraît pertinent de comprendre ce qui motive les recherches antérieures lorsqu'elles considèrent différents types d'information d'un point de vue global en Sciences de Gestion mais également du point de vue spécifique de la Finance. La compréhension de cette combinaison de données, dans divers champs des Sciences de Gestion et notamment en Finance, permet d'anticiper les effets de levier de l'utilisation d'informations plurielles et du croisement de données hétérogènes en univers multi-classes d'actifs.

Pourquoi avoir recours à ces différents types de données ? Nous analysons les motivations de cette combinaison de données hétérogènes. Nous répondons à cette interrogation initiale à la fois par certains éléments suggérés dans la littérature existante et d'autres issus de nos résultats empiriques. Nous sommes amenés à nous questionner sur les limites des paramètres utilisés dans les modèles d'arbitrage statistique et nous envisageons de compléter ces paramètres « classiques » avec d'« autres informations de marché » exprimées sous forme de variables quantitatives⁴.

⁴ Nous définissons ces variables quantitatives dans le chapitre 1.

Question 2. Dans quelle mesure, la combinaison d'informations hétérogènes peut-elle s'intégrer à un modèle d'évaluation systématique en univers multi-classes d'actifs financiers cotés ?

Nous explorons les processus qui permettent de prendre en compte des données hétérogènes au sein de méthodes systématiques d'arbitrage multi-classes. Nous développons nos recherches en nous appuyant sur le modèle multifactoriel de Ross (1976) en y intégrant les « autres informations de marché ». Nous construisons un modèle empirique dérivé de ces travaux dans une configuration « euro neutre »⁵. Nous cherchons théoriquement et empiriquement comment d'« autres informations de marché » peuvent s'intégrer dans notre modélisation. A partir de travaux sur les méthodes de quantification (Pagès et Printems, 2008), des données hétérogènes sont intégrées à notre modèle d'évaluation d'un portefeuille d'arbitrage.

Question 3. Comment mesurer les effets de l'utilisation de données hétérogènes sur la composition et la performance des portefeuilles d'arbitrage « euro neutre » multi-classes d'actifs ? Quelle est la valeur ajoutée de la combinaison de données hétérogènes?

Nous cherchons à mesurer l'impact des données hétérogènes sur les différentes caractéristiques des portefeuilles d'arbitrage multi-classes d'actifs. Il s'agit d'établir les effets en termes de rentabilité, de volatilité, d'allocations d'actifs et de corrélation entre les classes d'actifs. Cette troisième question principale de recherche, nous permet d'envisager des réponses au travers d'une diversification de portefeuille d'abord « naïve », puis « optimale ».

3. Fondements théoriques et cadre conceptuel de la recherche

Notre recherche repose sur des éléments théoriques et conceptuels que nous analysons ci-après. Nous mentionnons les principaux travaux de références à l'aide desquels nous progressons dans nos discussions et nos découvertes empiriques.

Le cadre probabiliste en Finance

Dans le cadre de nos travaux, nous commençons par souligner l'influence des probabilités sur l'anticipation des mouvements de prix d'actifs financiers⁶ et donc des rendements. Depuis le début du XX^{ième} siècle, des travaux de recherche sont fondés sur la

⁵ Notion différente de celle de « marché neutre » usuellement proposé par certaines recherches en arbitrage statistique. Ces éléments sont définis dans le chapitre 2. Voir Avellaneda et Lee (2010) et Khandani et Lo (2007) pour plus de détails sur ces deux définitions.

⁶ Nous verrons que ce point est crucial dans les chapitres 2, 4 et 5.

place déterminante des probabilités dans la prévision des fluctuations de prix d'actifs dans une fenêtre de temps précise (Bachelier, 1900). Il s'agit de comprendre « l'état statique » du marché et de déterminer « la loi de probabilité des variations de cours » à un instant de la cotation⁷.

Certains travaux de quantification vectorielle permettent d'examiner l'influence des probabilités en Finance dans le cadre de méthodes numériques (Pagès et Printems, 2008). Brièvement⁸, les méthodes de quantification aident à déterminer la probabilité que le prix d'un actif prenne une valeur précise au travers de calculs de distances minimums (Kohonen, 1995 ; Kohonen, 1982 ; Lloyd, 1982). Historiquement, les méthodes de quantification (probabilistes) ont émergé dans les années 1950 afin de discrétiser un signal émis (Gersho et Gray 1992 ; Graf et Luschgy 2000). *In fine*, nous pouvons anticiper les évolutions de prix d'actifs entre deux dates, ce qui équivaut à considérer un mouvement de prix d'un jour sur l'autre comme « plus ou moins probable ».

Par ailleurs, notre recherche revêt un caractère markovien (Bremaud, 2010) puisque nous sommes intéressés par la détermination des prix futurs à l'aide d'informations passées relativement « récentes ». Ces informations peuvent *a priori* intégrer ou non des données hétérogènes.

La prise en compte de données hétérogènes en Finance

La pertinence de la prise en compte de données hétérogènes en Finance est au centre de nos questionnements. Ces dernières années, plusieurs recherches échangent sur ce point. L'existence d'une pluralité d'approches ou styles du gestionnaire d'actifs, permet d'envisager que ce dernier n'est pas uniquement « quantitatif » ou « fondamentaliste » (Gregory-Allen, Stangl, et Shawky, 2009). Les auteurs rappellent qu'un gestionnaire « quantitatif » peut combiner des approches « quantitatives, *Top Down* et fondamentales ». Dans les faits, le mélange des techniques et des données existe déjà sous l'appellation d'« approche hybride » (Fabozzi, Focardi, et Jonas, 2008).

L'introduction de données hétérogènes peut également être motivée par un souci d'amélioration dans l'évaluation et la prévision de phénomènes financiers tels que la prédiction de défaut de paiements (Lehman, 2003) ou l'évaluation d'entreprise (Damodaran, 2001).

Des liens existent entre l'économie et la sphère financière. Certains travaux établissent un lien entre la prédiction des rendements d'actifs et le cycle de l'économie (Brenner, Pasquariello, et Subrahmanyam, 2009 ; Dahlquist et Campbell, 2001 ; Jones, Lamont, et

⁷ Par exemple à un instant de clôture (cotation de clôture).

⁸ Le chapitre 2 est en partie consacré aux méthodes de quantification vectorielles.

Lumsdaine, 1998). Les rendements du marché actions peuvent réagir à des « forces externes » (Chen, Roll, et Ross, 1986) comme les variables macroéconomiques. Cette constatation empirique laisse penser que certains facteurs économiques tels que la production industrielle, l'inflation non-anticipée ou le changement du niveau d'inflation anticipée peuvent expliquer statistiquement les rendements attendus des prix des actions. D'autres travaux de recherches incitent à l'élargissement des données utilisées en allocation d'actifs en considérant des variables « prédictives », non spécifiques à une seule classe d'actifs ou encore des « indicateurs macro-économiques » (Blitz et Vliet, 2008). Ainsi, les prix des actifs ne suffisent pas toujours à modéliser le risque de portefeuille. Ce travail d'évaluation peut être complété par l'intégration d'une hétérogénéité dans les données utilisées.

Equilibre de marché et processus de retour à la moyenne

Notre modélisation d'arbitrage systématique multi-classes d'actifs s'inspire initialement de plusieurs travaux en arbitrage statistique. Des recherches ont été menées sur des modèles de retour à la moyenne ou *mean reversion* avec l'application de stratégies « contrariantes » (Lehmann, 1990 ; Lo et MacKinlay, 1990) bénéficiant de la « sous » ou « sur-réaction » d'un marché. Ces stratégies consistent à acheter des « perdants » et à vendre des « gagnants » d'aujourd'hui, en espérant que des « perdants » d'aujourd'hui deviennent des « gagnants » de demain et les « gagnants » d'aujourd'hui des « perdants » de demain.

Le phénomène de retour à la moyenne résulte de l'« autocovariance croisée » (Lo et MacKinlay, 1990) entre deux actifs distincts et de la sur ou sous-évaluation des titres par rapport au marché. Nous comparons le taux de variation du prix d'un actif à celui du marché afin de savoir si un ordre d'achat ou de vente peut être envisagé.

Une modélisation gaussienne du phénomène de retour à la moyenne par le biais d'un processus d'Ornstein-Uhlenbeck (Uhlenbeck et Ornstein, 1930) est envisagé dans un travail de référence sur l'arbitrage statistique (Avellaneda et Lee, 2010). Des recherches issues de l'économétrie des séries non-stationnaires (Granger et Newbold, 1974) ont permis de proposer d'autres formes de processus de retour à la moyenne, notamment dans le cadre de la cointégration (Granger, 1981) plus précisément des modèles à correction d'erreur (Engle et Granger, 1987) et des modèles vectoriels à correction d'erreur (Johansen et Juselius 1990; Johansen 1991 ; Osterwald-Lenum, 1992) afin de modéliser des dynamiques de retour vers un équilibre de long terme.

Nous verrons comment notre modèle empirique s'engage dans une voie théorique relativement plus souple à l'aide d'une modélisation sous-jacente qui ne suit pas de loi de probabilité *a priori* (caractère non-paramétrique) et où l'équilibre du marché est déterminé statistiquement à l'aide d'analyses factorielles.

L'allocation tactique comme stratégie de gestion de portefeuille

Le périmètre de notre recherche est celui de l'allocation tactique ou des « décisions tactiques » puisque nous exploitons des méthodes permettant de modifier le poids des divers actifs d'un portefeuille à une fréquence quotidienne. L'allocation tactique est une allocation de court-terme qui « consiste à ajuster le portefeuille régulièrement en fonction des opportunités et des menaces du marché... Le gérant fait ainsi preuve d'une plus ou moins bonne aptitude en matière de *market timing* »⁹.

Dahlquist et Campbell (2001) définissent l'allocation tactique comme une stratégie fondée sur des informations « conditionnelles » facilitant la prévision des rendements futurs de manière quotidienne ou intra-journalière. Ils soulignent également le caractère dynamique de l'allocation lié aux fréquents ajustements de court-terme. L'allocation tactique vient modifier dynamiquement, dans les limites fixées au gestionnaire, les choix effectués au niveau de l'allocation stratégique qui elle s'étend sur un horizon plus long terme.

En amont de l'allocation d'actifs, les décisions relatives à l'allocation stratégique peuvent être discrétionnaires (Bodson et al., 2010) relayant les traitements quantitatifs pour la phase d'allocation tactique. La distinction entre l'allocation stratégique et tactique peut être quantitativement définie, assurant en général une performance de long terme autour de laquelle il est possible de surperformer le portefeuille de marché, de référence ou d'« équilibre » (Black et Litterman, 1992) sur le court terme à l'aide d'ajustements conjoncturels du portefeuille (Brinson, Singer, et Beebower 1991 ; Brinson, Hood, et Beebower 1986 ; Stockton et Shtekhman 2010 ; Tokat 2005).

Dans le cadre de nos recherches, nous adoptons une approche quantitative de l'allocation tactique en construisant des méthodes de *market timing* systématique. Nous cherchons à estimer quotidiennement la position relative d'un actif par rapport à une certaine tendance ou à l'équilibre de marché.

Les modèles multifactoriels et la gestion du risque spécifique

Le cadre théorique de l'arbitrage entre les classes d'actifs prend forme au sein des modèles multifactoriels et plus précisément de celui de Ross (1976) fondé sur l'« *arbitrage pricing theory* » (APT). Notons que l'auteur décrit une théorie de *pricing* sans arbitrage dans un cadre multifactoriel, l'utilisation d'un modèle multifactoriel n'étant que le point de départ

⁹ Chollet, P., donne cette définition dans le chapitre 19 « Gestion d'un portefeuille titres » de l'ouvrage coordonné par Thauvron (2011) *Gestion de patrimoine - Stratégies juridiques, fiscales et financières*. 3^e éd., Dunod, 2011. Le *market timing* fait référence au moment de l'investissement. Nous revenons sur le concept de *market timing* dans les chapitres 1 et 2.

de sa réflexion. Le cadre générique d'un modèle multifactoriel correspond au calcul du rendement attendu d'un actif en fonction d'un ensemble de facteurs¹⁰. Dans ce cadre, chaque coefficient explicatif (bêta) de la régression traduit la sensibilité de l'actif à un facteur (de référence). C'est au sein de ce cadre que nous intégrons les « autres informations de marché » d'une manière distincte de celle employée par l'auteur¹¹.

Si le prix de l'actif diverge, l'arbitrage doit en théorie le faire ultérieurement converger vers son « vrai » prix. La stratégie principale est de définir la position d'un actif par rapport à « son marché » pour savoir si l'actif est sur ou sous-évalué par rapport à « son marché ».

Soulignons qu'empiriquement, la manière de définir le marché de référence et donc le *benchmark* est un enjeu de recherche sur lequel un consensus ne semble pas encore émerger. La définition du marché n'est pas forcément la même selon les sensibilités du gestionnaire. Par ailleurs, plusieurs travaux évoquent la complexité d'identifier et de définir un *benchmark* (Xu et Malkiel, 2004 ; Black et Litterman, 1992 ; Ackermann, McEnally, et Ravenscraft, 1999).

Notre approche du marché est statistique dans la mesure où nous utilisons une analyse en composantes principales pour séparer les facteurs de risques principaux des facteurs de risques non-principaux. Dans cette optique, les facteurs de risques principaux incarnent le risque de marché, le risque systématique en gestion, et les facteurs non-principaux deviennent le risque spécifique représentant le risque propre à l'actif financier. Partant d'un modèle multifactoriel, un type de gestion « marché neutre » assure théoriquement de n'être exposé qu'au risque spécifique car le risque systématique est par construction « neutralisé ». Lors des recherches, nous nous sommes concentrés sur l'estimation de hausse ou de baisse (au travers de calculs de probabilités) du risque spécifique¹². Une rentabilité positive est générée si l'estimation (probabilité de hausse ou de baisse) de l'évolution du risque spécifique est réalisée (et inversement).

¹⁰ Ces facteurs peuvent être des facteurs macroéconomiques ou des indices de marché théoriques.

¹¹ Nous expliquons en détail notre manière de procéder dans les chapitres 2 et 3.

¹² Bien que nous ne soyons pas dans un cadre « marché neutre » dans notre approche, le rendement spécifique est assimilé à la rentabilité de nos portefeuilles d'arbitrage « euro neutre ».

4. Contributions de la recherche

Bien que reposant sur de nombreuses recherches existantes, notre recherche se différencie des recherches antérieures à plusieurs égards. Il n'existe pas de modèle d'évaluation d'allocations multi-classes d'actifs communément accepté (Blitz et Vliet, 2008). L'influence des données hétérogènes sur nos portefeuilles d'arbitrage permet de créer ce lien entre les classes d'actifs avec la présence d'« autres informations de marché », censée impacter toutes les classes d'actifs. Nous tentons de communiquer des éléments de réponses relatifs à l'évaluation de portefeuilles d'arbitrage multi-classes d'actifs en présence de données hétérogènes.

Des recherches sur la quantification ont été menées dans le cadre de l'évaluation d'options (Corlay, 2011 ; Pagès et Printems, 2008). À ce jour, à notre connaissance, aucune recherche sur la quantification vectorielle n'est dédiée à des portefeuilles d'arbitrage multi-classes d'actifs en présence de données hétérogènes. Proposer une version alternative d'un modèle d'arbitrage (Ross, 1976) à l'aide de la quantification permet de comparer nos résultats à ceux des travaux originels dans un cadre probabiliste.

Comme le rappellent Andersen et al. (2007), peu de recherches se concentrent à la fois sur plusieurs classes d'actifs et sur plusieurs pays dans le monde. Les auteurs intègrent l'univers multi-classes sans prendre en compte les matières premières et prennent en considération de nombreuses enquêtes économiques sans intégrer le sentiment des investisseurs dans leurs données. Certains travaux portent leur attention sur le lien entre l'économie et les rendements d'actifs (Brenner, Pasquariello, et Subrahmanyam, 2009 ; Andersen et al., 2007 ; Boyd, Jagannathan, et Hu, 2005 ; Jones, Lamont, et Lumsdaine, 1998) quand d'autres se concentrent sur le lien entre les indicateurs de sentiments et les rendements d'actifs (Fisher et Statman, 2003 ; Fisher et Statman, 2000 ; Thorp, 2004). Prenant en compte des données hétérogènes dans un univers multi-classes d'actifs, nous nous investissons dans une voie de recherche où de nombreuses questions subsistent.

Plusieurs définitions sont disponibles pour définir les types de variables utilisées dans les modèles financiers mais aucun consensus ne semble se former au sujet d'une sémantique commune. Dans la littérature existante en Finance, nous ne trouvons pas de termes standardisés pour qualifier les données qui ne sont pas propres aux prix d'actifs. Ces différentes définitions sont reprises et discutées dans le chapitre 1. Pour évoquer les données qualifiées (à tort ou à raison) de « quantitatives », les termes employés sont variés : « *hard facts* » (Lehman, 2003), « éléments quantitatifs » ou « information quantitative » (Gregory-Allen et al., 2009). Pour faire référence aux données plus qualitatives, plus subjectives, les

mots utilisés peuvent être « facteurs fondamentaux », « information macroéconomique » (Brenner et al, 2009), « indicateurs macro-économiques », « variables prédictives » (Blitz et Vliet, 2008), « *economic drivers* », « données autres que des prix passés », « indicateurs économiques » (Khandani et Lo, 2007), « variables qualitatives », « *qualitative drivers* », « facteurs qualitatifs » (Magni et al, 2006) ou « variable d'information » (Dahlquist et Campbell, 2001). Cette brève analyse vient mettre en lumière le fait que la réponse donnée à cette question n'est pas consensuelle, d'où l'utilité de proposer une qualification alternative et plus générique des données intégrant un système d'informations plus exhaustif.

Nous intégrons des indicateurs de trois types: de conjoncture, de sentiment et de « nervosité » des marchés financiers. La qualification générique « autres informations de marché » est une proposition sémantique en vue de regrouper l'ensemble des informations¹³ qui peuvent interagir avec les rendements d'actifs.

5. Démarche méthodologique de la recherche

Notre démarche de recherche est celle d'une approche inductive, puisque nous souhaitons comprendre si l'utilisation de données hétérogènes influence la composition (allocation tactique) et la gestion (*market timing*) d'un portefeuille d'arbitrage. Comme le rappellent Evrard et al. (2009), « lorsque l'objectif est de comprendre, une approche inductive est appropriée puisque l'on part des observations pour arriver à une meilleure compréhension globale des phénomènes ». Les mêmes auteurs recommandent de partir d'une base de données et de la réduire (par exemple) à l'aide d'une analyse en composantes principales afin de mieux décrire en aval les relations entre variables. Nous sommes partis du même schéma méthodologique, à l'aide d'une analyse factorielle, pour isoler le risque spécifique des portefeuilles d'arbitrage. Nous sommes dans le cadre général de l'approche inductive puisque nous mettons en place un ensemble de méthodes, nous effectuons un travail de recueil de données, nous cherchons à analyser et à interpréter nos résultats puis nous utilisons ces derniers comme une aide à la décision.

Nos travaux partagent également certaines caractéristiques d'une approche prédictive puisque nous cherchons, par l'intermédiaire des méthodes de quantification¹⁴, à estimer la probabilité que le prix d'un actif augmente ou baisse d'une période à une autre (par exemple d'un jour sur l'autre).

Nous adoptons une posture épistémologique proche du positivisme. Nous souhaitons

¹³ Qui ne sont pas des prix d'actifs.

¹⁴ Méthodes probabilistes décrites dans le chapitre 1.

découvrir une réalité déterminée, celle de l'influence des données hétérogènes dans la composition et la gestion d'un portefeuille d'arbitrage systématique en univers multi-classes d'actifs. Nous répondons aux critères de validité du positivisme : la vérifiabilité car nous testons empiriquement ce que nous supposons être la réalité et la confirmabilité car nous ne considérons pas que notre proposition est « universellement vraie » (Evrard, Pras, et Roux, 2009) mais qu'elle est vérifiée pour nos tests empiriques dans le cadre de notre échantillon¹⁵.

Nous nous rattachons au courant de recherche dominant en Finance de marché selon lequel la plupart des réalisations financières sont mesurables scientifiquement. Certes, il existe des phénomènes non-prévisibles, notamment liés aux facteurs humains tels que les comportements des investisseurs. Cependant, les progrès scientifiques en recherche¹⁶ favorisent une meilleure mesure des phénomènes financiers que sur le passé tant sur le raffinement de nombreuses modélisations que sur la capacité informatique à effectuer un nombre élevé de *backtestings* à une vitesse de calcul optimisée.

Nous commençons par étudier la façon dont la recherche en Finance présente les méthodes d'allocations et de *market timing* en présence de données hétérogènes : c'est-à-dire des modèles quantitatifs très sophistiqués, parfois peu lisibles pour des investisseurs (Black et Litterman, 1992), ainsi que des modèles trop discrétionnaires pour être traçables.

Nous nous demandons si les investisseurs prêtent attention à des informations extra-financières telles que par exemple la conjoncture économique ou les sentiments des investisseurs, pour composer leurs portefeuilles d'actifs. L'idée sous-jacente à notre réflexion est que ces « autres informations de marché » peuvent indiquer des arbitrages possibles entre classes d'actifs (autres que celles déterminées en l'absence d'« autres informations de marché ») car ces dernières ne semblent pas réagir de la même manière à la publication de ces indicateurs (Brenner, Pasquariello, et Subrahmanyam, 2009 ; Andersen et al., 2007).

A partir d'une analyse de la littérature, nous cherchons à établir un état de l'existant concernant la définition des différents types de données utilisées en Finance. Nous n'obtenons pas sur ce sujet de définition unique. Par ailleurs, nous analysons les travaux relatifs à l'arbitrage statistique, ce qui nous permet de comprendre le fonctionnement des modèles sous-jacents au « déclenchement » d'ordres systématiques d'achat et de vente. Nous faisons le constat qu'à l'heure actuelle, en notre connaissance, aucune recherche n'utilise un ensemble de données hétérogènes en vue de constituer des portefeuilles d'arbitrage « euro neutre » systématique multi-classes d'actifs financiers incluant des obligations, des matières premières, des actions et des devises.

¹⁵ La composition de l'échantillon ainsi que les périodes de notre étude sont précisées dans le chapitre 3.

¹⁶ Dans les domaines de la probabilité, de la statistique, de l'Econométrie, de la Sociologie, de la Socioéconomie, de la Microéconomie, de la Gestion et de l'Informatique.

Nous cherchons à comprendre les différentes recherches proposées sur les modèles d'allocation qui peuvent combiner des données de cotation et certaines « autres informations de marché »¹⁷. Pour ce faire, nous définissons les concepts clés de notre travail de recherche comme l'arbitrage statistique, les données de cotation, les « autres informations de marché », l'univers multi-classes d'actifs, l'« allocation tactique » ou le *market timing*.

Nous souhaitons répondre le plus précisément possible à notre problématique de recherche et espérons contribuer à une voie de recherche encore peu explorée. Dans cette optique, nous associons systématiquement des données d'univers différents, les données de cotation et les « autres informations de marché ». Nous mettons en place un processus de traitement des données recueillies sur notre échantillon¹⁸, dans le but de pouvoir traiter (informatiquement) l'ensemble des données de manière systématique (c'est-à-dire non-discretionnaire). Afin d'apporter des réponses quantitatives à notre problématique générale, nous utilisons des outils statistiques (tels que l'analyse factorielle) et des méthodes probabilistes (quantification) pour générer des portefeuilles d'arbitrage en univers multi-classes. Ces techniques quantitatives nous permettent de mettre en place une approche non-paramétrique agrégeant des données hétérogènes.

Le processus de contrôle des résultats empiriques sur le passé est effectué à travers différents tests de robustesse menés à la fois pour les allocations « naïves » et pour les allocations optimales. Il s'agit, d'une part, de tests de robustesse propres aux techniques de simulation. Ces tests nous permettent de multiplier un « grand » nombre de fois des tests sur le passé et/ou de faire appel à des variables aléatoires, afin d'estimer le comportement moyen des stratégies d'arbitrage multi-classes d'actifs en présence ou non de données hétérogènes. D'autre part, un autre type de tests porte sur l'extension de la taille de l'échantillon d'étude.

6. Plan de la recherche

La présentation de notre recherche est organisée en cinq chapitres (figure 1). Le premier chapitre introduit les éléments de littérature et les outils numériques nécessaires à notre champ d'investigation. Nous présentons une revue de littérature sur les types de données utilisées en Sciences de Gestion et plus particulièrement en Finance. Cet état des lieux de l'existant permet d'envisager des associations possibles de données en univers multi-classes d'actifs. La prise en compte de cette combinaison de types de données permet notamment de

¹⁷ Les publications de recherche d'entreprises sur ce sujet restent peu transparentes car elles ont souvent le savoir-faire opérationnel pour générer de la rentabilité et souhaitent conserver une grande confidentialité sur leur processus d'allocation.

¹⁸ Nous détaillons le recueil des données dans le chapitre 3.

répondre à certaines limites décrites dans la littérature. Nous expliquons comment les modèles d'arbitrage statistique peuvent être limités par l'utilisation quasi-exclusive de données de cotation (données homogènes). Nous prolongeons cette synthèse de la littérature par la présentation de l'arbitrage statistique et des limites qui y sont associées, notamment en ce qui concerne l'utilisation quasi-exclusive de données de cotation comme paramètre (d'entrée) des modèles. Nous proposons de découvrir les deux principales « familles » d'allocations d'actifs que sont la diversification « naïve » (DeMiguel, Garlappi, et Uppal, 2009) et la diversification optimale de portefeuille, en conduisant une analyse comparée des ces deux types d'allocations d'actifs. Nous retrouvons ces formes d'allocations lors des tests empiriques. Nous clôturons ce premier chapitre par l'introduction de deux outils quantitatifs utilisés lors des tests empiriques : l'analyse factorielle (ACP) et les algorithmes de quantification optimale (Lloyd, 1982 ; Kohonen, 1982 ; Kohonen, 1995).

Le deuxième chapitre permet d'introduire le modèle empirique d'arbitrage systématique approprié pour un univers multi-classes d'actifs en présence ou non de données hétérogènes. Dans un premier temps, nous analysons l'approche systématique comparativement à l'approche discrétionnaire, afin de justifier notre choix en faveur d'une voie de recherche systématique. En nous inspirant d'un cadre multifactoriel, nous détaillons le processus d'identification du risque spécifique incorporant potentiellement des données hétérogènes. Le processus d'isolation du risque spécifique est globalement inspiré de Ross (1976) et nous y agrégeons un cadre probabiliste (quantification). Nous analysons comparativement deux méthodes d'arbitrage systématique (paramétrique et non-paramétrique) et nous examinons la valeur ajoutée de l'approche non-paramétrique retenue dans le cadre des tests empiriques.

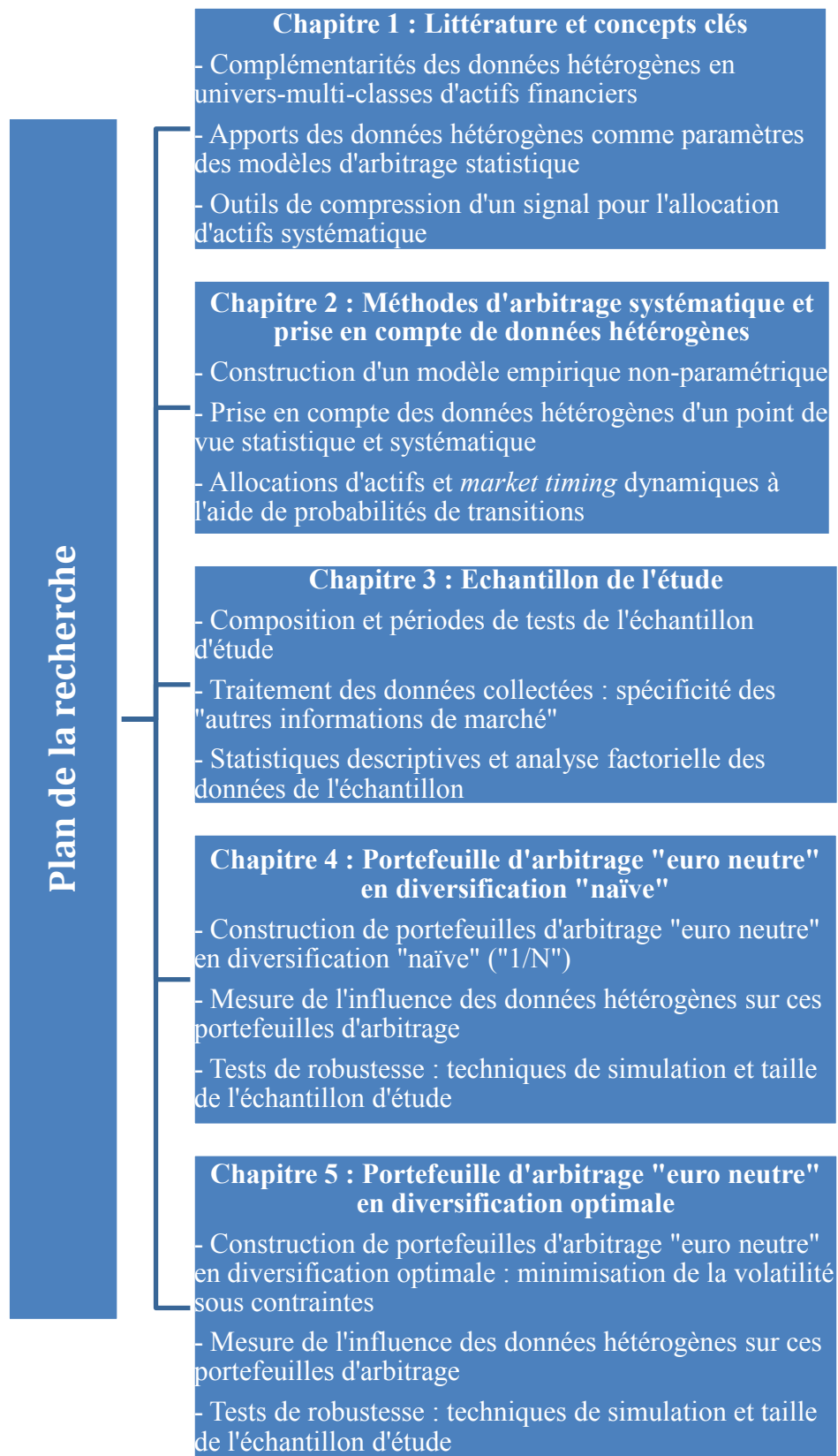
Le troisième chapitre permet de présenter l'échantillon d'étude à partir duquel sont réalisés des tests empiriques. Il a également pour objet d'expliquer la méthodologie de recueil des données. Le choix des périodes et sous-périodes de l'échantillon est justifié à la fois par le souhait de prendre en considération des périodes *pré* et *post* crise des *subprimes* et par la disponibilité de données. Ainsi, les tests sont conduits sur la période allant de 2004 à 2011. Notre étude avant, pendant et après la période des *subprimes*, permet de mesurer l'importance des données hétérogènes dans une période incluant un temps de crise. Nous présentons et analysons les critères de choix des données qui diffèrent selon le type de données (données homogènes et données hétérogènes). Nous détaillons les traitements appliqués aux données collectées. Nous accompagnons la découverte de notre échantillon d'une analyse statistique des données, à la fois descriptive et factorielle, permettant de comprendre le comportement passé.

Les chapitres quatre et cinq reposent sur des tests empiriques menés sur les données passées. L'objet du quatrième chapitre est de réaliser des tests, afin de mesurer l'influence des données hétérogènes sur les stratégies d'arbitrage multi-classes d'actifs pour des portefeuilles « euro neutre » en diversification « naïve » (DeMiguel, Garlappi, et Uppal, 2009). Nous entendons ainsi quantifier la réaction des portefeuilles lorsque nous considérons un univers de données hétérogènes par rapport à un univers de données homogènes (données de cotation). Par réaction des portefeuilles, nous faisons notamment référence à la rentabilité, à la volatilité, à l'allocation d'actifs, la contribution de chaque actif à la rentabilité ou à l'estimation des pertes monétaires probables des portefeuilles en présence de données hétérogènes et homogènes. Nous estimons la distorsion engendrée par l'introduction de données hétérogènes que nous avons sélectionnées (chapitre 3). Nous concluons ce chapitre par trois éléments de robustesse de notre méthode d'arbitrage systématique non-paramétrique de façon à mesurer la stabilité des résultats obtenus par notre approche. Ces tests de robustesse sont effectués afin de vérifier si nous obtenons des éléments de cohérence avec nos résultats empiriques.

Le cinquième chapitre introduit la diversification optimale. Il permet d'étudier l'influence des données hétérogènes en effectuant des tests sur le passé destinés à mesurer cet impact sur les stratégies d'arbitrage multi-classes d'actifs pour des portefeuilles « euro neutre » en diversification optimale (DeMiguel, Garlappi, et Uppal, 2009). La procédure d'optimisation retenue consiste à minimiser la volatilité sous contraintes. Nous procédons de manière identique, avec les mêmes tests sur le passé que ceux du quatrième chapitre, à la différence que l'univers de recherche introduit dorénavant des procédures d'optimisation dans la phase d'allocation des portefeuilles. Les portefeuilles générés comportent des pondérations d'actifs différentes de celles obtenues après diversification « naïve ». Les tests de robustesse présentés, dans ce dernier chapitre, sont également les mêmes que ceux du quatrième chapitre, seul le cadre optimal diffère de la diversification « naïve ». Pour l'ensemble de ces tests (tests sur le passé et tests de robustesse), nous soulevons les points de convergence et de divergence des résultats obtenus entre une diversification « naïve » et une diversification optimale des portefeuilles d'arbitrage en présence de données hétérogènes.

La conclusion générale permet de synthétiser les principaux résultats empiriques quelle que soit la diversification de portefeuille (« naïve » ou optimale) choisie. Nous soulignons les apports empiriques, méthodologiques et managériaux issus de nos recherches. Nous analysons les limites que comportent nos travaux et proposons des perspectives de recherche, comme éléments de réponses envisageables à ces limites.

Figure 1 - Plan de la thèse par chapitre : principaux contenus par chapitre



CHAPITRE 1. LITTÉRATURE ET CONCEPTS CLÉS

Ce chapitre explore de manière synthétique les différents travaux de recherche existant sur la combinaison de données hétérogènes en Sciences de Gestion, notamment lorsqu'il est question d'intégrer des informations subjectives comme le sentiment des agents économiques et celui des investisseurs. Notre analyse de la littérature est centrée sur l'apport d'informations de différentes natures en Finance. Elle porte ensuite sur les modèles d'arbitrage statistique et l'allocation d'actifs. Enfin, diverses techniques quantitatives appliquées en Finance et mobilisées dans le cadre de cette thèse sont présentées.

Dans le cadre de notre recherche, nous distinguons sur le fond des éléments d'information qui sont directement issus de transactions financières et des éléments qui rendent compte de la conjoncture et intègrent (partiellement ou totalement) le sentiment, l'appréciation subjective des personnes sondées, tels que des réponses sur la « normalité » de la conjoncture économique et la direction haussière ou baissière des marchés boursiers. Pour rendre compte de la différence entre ces informations, mais également de leur complémentarité, les notions d'hétérogénéité et d'homogénéité sont introduites afin de répondre à l'enjeu de la qualification de la nature des données. Cette classification est d'autant plus importante que nous cherchons à mesurer l'importance des « autres informations de marché » au-delà d'un univers de données de cotation.

Bien que des recherches prennent déjà en considération l'utilisation de données hétérogènes en Finance (Brenner, Pasquariello, et Subrahmanyam, 2009 ; Fermon, 2006 ; Gregory-Allen, Stangl, et Shawky, 2009 ; Lehman, 2003), notre but est de souligner et d'enrichir cette approche dans le cadre d'un univers multi-classes d'actifs¹⁹ financiers cotés. Nous présentons les principales caractéristiques de l'arbitrage statistique avant d'en soulever les limites. Nous cherchons à analyser comment l'intégration de données hétérogènes peut partiellement répondre aux limites identifiées et participer à l'évolution des modèles d'arbitrage statistique.

Nous introduisons les deux principaux cadres d'allocation d'actifs, autrement appelés la diversification de portefeuille. Nous distinguons la diversification « naïve » de la diversification optimale. Nous proposons également une analyse comparative des ces deux types de diversification de portefeuilles afin présenter leurs caractéristiques respectives.

Enfin, ce chapitre présente plusieurs concepts quantitatifs nécessaires à la compréhension des stratégies d'arbitrage utilisées dans notre recherche. Ces outils quantitatifs

¹⁹ Nous verrons dans le chapitre 3 quelles sont précisément les classes d'actifs concernées.

sont nécessaires à la compréhension de la modélisation empirique proposée dans le cadre de nos recherches. Soulignons que le recours à ces outils statistiques et probabilistes est un moyen de mesurer l'efficacité de notre approche.

1. Hétérogénéité des données en Sciences de Gestion

La diversité des Sciences de Gestion²⁰ permet de faire appel à différents types de données. Dans ce contexte, le choix de l'information et son traitement occupent une place déterminante dans un processus de décision de gestion. Il est question de comprendre que le contenu des variables sollicitées en Sciences de Gestion peut être profondément hétérogène. Notons que de nombreuses variables qualitatives sont numériques, comme pour les variables à plusieurs modalités. Ces variables à « échelle nominale » (Evrard, Pras, et Roux, 2009) peuvent également exprimer un sentiment (par exemple, avoir confiance en l'avenir ou non). Nous définissons et analysons les types de données sollicitées dans le cadre de notre recherche. Nous décomposons les contenus de ces données hétérogènes. Nous cherchons à comprendre la complémentarité des données hétérogènes dans le cadre des Sciences de Gestion.

1.1. Structure des données hétérogènes

Dans l'Introduction Générale, nous distinguons les données de cotation des données relatives aux « autres informations de marché ». Nous différencions les données de cotation des « autres informations de marché », notamment à l'aide de la littérature existante sur les types de données en Sciences de Gestion.

D'un point de vue statistique, tant les données de cotation que les « autres informations de marché » sont des variables dites « quantitatives » bien qu'elle résument dans la plupart des cas des opinions ou des sentiments. Evrard et al. (2009) précisent que « les données quantitatives » utilisent des échelles de proportion et d'intervalle regroupées sous l'appellation de « variable quantitative ». Les échelles de proportion, également appelées échelles de « ratio » sont très utilisées dans les Sciences « dures » comme les Sciences Physique et les Sciences Naturelles. Les variables quantitatives sont des mesures « idéales » car il est possible de leur appliquer des opérateurs mathématiques tels que les additions, les soustractions, les multiplications et les divisions. La liberté de calcul offerte par ce type de

²⁰ Les Sciences de Gestion couvrent des champs tels que le Marketing et la Finance.

variables permet d'effectuer divers traitements de données dans le cadre de nos travaux²¹. Nous calculons par la suite des taux de variation sur les données de notre échantillon en faisant appel à certains de ces opérateurs mathématiques.

En plus des variables « quantitatives », il existe des variables « qualitatives », qui sont mesurées par des échelles « nominales » et « ordinales » (Evrard, Pras, et Roux, 2003). Les échelles nominales indiquent qu'une variable ou une donnée appartient à un groupe ou à une classe. Pour ce type de variable, il est uniquement possible d'afficher des relations d'équivalence. En Sciences de Gestion mais également en Sciences Économiques, ce sont par exemple des variables binaires, booléennes ou *dummy* prenant la valeur numérique 1 ou 0 pour des événements se réalisant ou pas. Les échelles ordinales permettent de classer des objets par direction ou par dimension en affichant un ordre de préférence ($A > B > C$). Les opérations arithmétiques sur les variables mesurées par des échelles ordinales sont la médiane ou le quartile : des critères de position. Les données qualitatives peuvent être exprimées par des mesures ou un code binaire. Dans cette perspective, les données « qualitatives » deviennent numériques mais mesurent des phénomènes pour partie intangibles et subjectifs. Dans notre troisième chapitre, nous verrons que des éléments comme la confiance des investisseurs, des consommateurs ou l'appréciation de la conjoncture économique par différents agents peuvent être formalisés au travers de chiffres. Cette transformation quantitative permet notamment de rendre tangibles et exploitables ces différentes informations.

L'hétérogénéité des données provient du fait que nous considérons des données de cotation, comme les cotations de clôture des marchés financiers, et d'« autres informations de marché ». Nous rappelons que les « autres informations de marché » incluent des indicateurs de conjoncture économique, de sentiment et des indices de volatilité²². Certaines « autres informations de marché » que nous sollicitons²³, transforment dans la majorité des cas, des réponses qualitatives (textuelles ou vocales) en mesures quantitatives.

Le contenu de ces « autres informations de marché » diffère de celui de données de cotation. Cette segmentation envisage l'existence d'une pluralité, d'une diversité et d'une complexité au sein des données en Sciences de Gestion qui ne s'exprime pas assez exhaustivement par le clivage quantitatif/qualitatif. Bien qu'au cours des travaux de recherche nous dépasserons le clivage quantitatif/qualitatif, nous soulignons que cette dichotomie n'est pas uniquement applicable aux types de données. Ce clivage peut également être applicable

²¹ Dans le troisième chapitre.

²² Indices de volatilité prospective. Voir le chapitre 3.

²³ Dans le chapitre 3 nous analyserons en détail la composition de l'échantillon des « autres informations de marché » ainsi que celle des données de cotation.

aux approches méthodologiques.

L'éclectisme des variables que nous intégrons au cadre de nos recherches provient du fait que certaines d'entre elles combinent tout autant une analyse du passé, de la situation présente et des attentes, anticipations pour un futur proche²⁴. De nombreuses séries contiennent, dans une plus ou moins large mesure, une lecture subjective de la conjoncture financière et économique²⁵. Ainsi, certaines enquêtes font état de réponses « bonne » ou « moins bonne » pour la situation conjoncturelle actuelle et « plus favorable » ou « inchangée » pour les six mois à venir. Il convient de souligner que la différence d'interprétation et d'utilisation de ces informations illustre que pour deux chefs d'entreprises et/ou investisseurs sondés, la signification de ces termes ne sera pas obligatoirement la même. Les réponses à ces enquêtes étant celles d'êtres humains, ces indicateurs bien qu'informatifs peuvent rester incomplets pour retranscrire l'aspect parfois intangible de l'évolution de l'environnement économique et financier. Notons qu'un traitement statistique est appliqué à de nombreuses séries économiques et de sentiment par les entités qui les publient. Le but de ces traitements est globalement de communiquer des données interprétables, utilisables pour les utilisateurs finaux de ces informations.

Les différents points qui précèdent ont permis d'introduire le concept d'hétérogénéité des données en Sciences de Gestion. L'étape suivante de notre réflexion est de comprendre la complémentarité qui peut exister au sein de ces données hétérogènes et explorer les raisons qui incitent à une possible combinaison. Nous détaillons comment cette combinaison devient opérante à l'aide de différents travaux en Sciences de Gestion.

1.2. Complémentarité des données hétérogènes en Sciences de Gestion

Les données hétérogènes auxquelles nous faisons référence véhiculent des contenus différents qui peuvent potentiellement se compléter. Sans obligation de liens entre elles, ces données peuvent être nécessaires les unes aux autres. Nous décrivons et analysons ci-après la nature de ces synergies puis les réalisations concrètes de ces multiples combinaisons de données. Nous tentons d'expliquer pourquoi ces deux familles de données (données de cotation et « autres informations de marché ») peuvent être associées puis comment ce regroupement est mis en œuvre au travers des tests empiriques issues de travaux de recherches en Sciences de Gestion.

²⁴ Généralement pour les six mois à venir pour les séries retenues dans notre échantillon. Se reporter au chapitre 3.

²⁵ Se reporter au chapitre 3.

La complémentarité entre données hétérogènes permet, dans une certaine mesure, de mieux évaluer les risques de pertes financières. Certains modèles quantitatifs n'avaient pas décelé l'arrivée de la crise des *subprimes* (Khandani et Lo, 2007) alors qu'ils étaient censés évaluer des probabilités de défaut. Les auteurs trouvent empiriquement que les événements financiers et économiques ne sont pas toujours gouvernés par des distributions gaussiennes et que des événements rares peuvent se produire. Ils font référence à un manque de précision en expliquant : « il n'est pas ingénieux d'affirmer qu'une stratégie est couronnée de succès sauf dans l'évènement du vingt cinquième écart type. Vue l'improbabilité d'un tel évènement, la conclusion est que : soit la distribution réelle des rendements est extraordinairement leptokurtique, soit l'écart-type est variable dans le temps et présente des pics occasionnels ». L'idée émergente de leurs recherches est qu'une stratégie d'investissement devrait évaluer et préciser son « *risque de ruine* ». Leurs conclusions ne démontrent pas pour autant que les modèles quantitatifs portent la responsabilité de la crise de l'été 2007 : il y a peut-être eu un manque d'informations. Dans la liste des éléments manquants est recensé le risque d'« illiquidité » (concernant les instruments financiers négociés) plutôt que le fait de considérer la volatilité comme mesure du risque dans les stratégies d'arbitrage. D'autres éléments comme « l'instinct de survie » ou la sélection naturelle dans l'industrie financière et le vieux concept économique de concurrence imparfaite ont un fort impact sur les prix d'actifs quand il y a des débouchements massifs de positions comme le rappelle Montier (2007) en évoquant le « risque endogène ». Ces facteurs semblent pouvoir accroître la précision dans l'évaluation de phénomènes financiers.

Lehman (2003) étudie la qualité des modèles bancaires de notations des entreprises (clientes). L'auteur identifie l'apport et les synergies créés par la prise en compte de sources d'informations financières (« dures ») et « qualitatives » (« molles ») comme le jugement subjectif des analystes financiers. Ces recherches sont conduites sur l'estimation de probabilité de défaut de crédit d'une entreprise en présence de sources d'informations hétérogènes intégrées dans une modélisation de régressions logistiques (Evrard, Pras, et Roux, 2009 ; King et Zeng, 2001)²⁶. L'élément principal de ce travail est d'apporter une réponse tangible sur les facteurs explicatifs qui augmentent la probabilité de détecter le défaut de paiement d'une entreprise. Le modèle intégrant des variables « dures » et « molles » augmente plus fortement la détection de probabilité de défaut que celui ne prenant en compte que des facteurs « durs ». Ces conclusions sont encadrées statistiquement : la courbe *Receiver Operating Characteristic* (ROC) confirme également que l'intégration de facteurs « mous »

²⁶ Ils expliquent comment est effectuée une régression logistique et quels en sont les enjeux : trouver les variables explicatives qui augmentent la probabilité qu'un évènement se réalise.

participe à une meilleure prédiction de défaut que lorsque seules des données comptables (appartenant au groupe des facteurs « durs ») sont prises en compte. Le modèle intégrant de l'information plus qualitative permet par ailleurs de réduire la variabilité des résultats. Une traduction de ce constat : le modèle qui a permis de mieux détecter un défaut intègre des variables « molles » parmi ses variables explicatives et que le paramètre « mou » améliore significativement un modèle qui n'est nourri que de facteurs « durs ». Notons que l'identification des facteurs à considérer pour étudier des phénomènes financiers reste un questionnement ouvert. Certains travaux ont initialement tenté de répondre à ce questionnement sur les mesures adéquates à l'analyse des rapports financiers (Altman, 1968 ; Altman, Haldeman, et Narayanan, 1977).

L'association de données hétérogènes permet d'approcher différemment la valorisation d'actifs ou d'entreprises. Certains travaux (Chollet et Cellier, 2011 ; Chollet, Cellier, et Gajewski, 2011) soulignent l'importance de la prise en compte d'informations extra-financières en analysant les rendements des actions autour des dates de publication de notations sociales des entreprises. L'apport de l'hétérogénéité des informations est matérialisé par des changements de rendements lors (et autour) de la publication d'informations extra-financières. La prise en compte d'autres types d'information que les données de cotation, permet de constater quantitativement leur influence sur l'évolution des rendements d'actions.

Dans un autre environnement, les modèles d'actualisation de flux de trésorerie ne prennent pas toujours en compte d'« importantes » variables (Magni, Malagoli, et Mastroleo, 2006). D'autres types d'information peuvent être nécessaires pour évaluer un environnement financier complexe et changeant rapidement (Damodaran, 2001) ; il existe toutefois des éléments influençant cette valorisation qui sont difficiles à modéliser quantitativement : « l'analyse de certains segments où opère l'entreprise sont ignorés, or les relations entre les fournisseurs, les clients, les distributeurs, les barrières à l'entrée et à la sortie, l'intensité concurrentielle sont fondamentaux pour l'évaluation de ces entreprises. Les méthodes d'actualisation de flux de trésorerie négligent essentiellement les facteurs stratégiques et d'affaires qui s'avèrent être vitaux pour déterminer la valeur d'une entreprise ».

Dans notre recherche, l'introduction d'indicateurs de conjoncture et/ou de sentiment peut répondre aux insuffisances décrites car ces indicateurs constituent des données numériques exploitables dans une approche quantitative modélisée. En effet, ce type d'information synthétise, de manière certes imparfaite, différents aspects d'activité affectant les relations entre agents économiques (ménages, entreprises) dont il semble primordial de tenir compte.

La combinaison de données de cotation et d'« autres informations de marché » ont fait

l'objet de recherches empiriques lorsque la « valeur » d'un dirigeant d'entreprise entre en considération pour expliquer les rendements des marchés actions. En associant des critères de gouvernance d'entreprise et de « valeur » d'un dirigeant à des données de valorisations boursières, des portefeuilles d'actions rentables peuvent être constitués (Fermon, 2006). Selon l'auteur, plusieurs tests sur le passé ont révélé que la « valeur » d'un dirigeant peut augmenter ou réduire les rendements d'une action cotée sur les marchés boursiers. Trois voies d'exploration sont proposées afin d'évaluer l'impact de la « valeur » d'un dirigeant sur le prix de l'action : la corrélation entre la communication du dirigeant²⁷ et le rendement du prix de l'action, le calcul de la valorisation financière de chaque action d'un secteur²⁸ permettant d'identifier des différences imputables à la « valeur » du dirigeant puis finalement le suivi des rendements après le remplacement du dirigeant.

Il existe une distinction entre les approches des gestionnaires « fondamentalistes » et « quantitatifs » (Gregory-Allen, Stangl, et Shawky, 2009). Ces auteurs tempèrent cette dichotomie et expliquent comment ce genre de typologie omet de comprendre l'ensemble des approches mobilisées. Ils mentionnent qu'il existe une « croyance » que les managers quantitatifs sont moins exposés aux erreurs cognitives (et autres biais) et qu'ils réalisent des performances plus élevées que des performances aléatoires. Ils expliquent que l'association des deux approches crée de la valeur. Précisément, ils soulignent que la contribution marginale de l'approche « fondamentale » est la seule à permettre l'accroissement des rendements lorsque cette dernière vient en complément d'une première approche (quantitative ou autre). La combinaison d'approches et de types de données distinctes peut accroître la performance des portefeuilles : en examinant la performance qui est imputable à une approche secondaire, seule « l'approche fondamentale » est significativement²⁹ créatrice de valeur. Notons que la classification entre « fondamentalistes » et quantitatifs semble biaisée. L'approche quantitative n'implique pas de n'employer que des techniques quantitatives, et inversement pour les approches « fondamentales ».

Après identification des processus de sélection d'actifs de type fondamental ou quantitatif, Zhao (2006) constate que les performances des deux types d'approches sont équivalentes pour des fonds sur le marché actions entre 1994 et 2003 : la sélection quantitative n'a pas, dans ce spectre, une prédominance absolue. Tout en conservant initialement l'opposition entre analyse fondamentale et recherche quantitative, Wermers et al. (2010) expliquent que la capacité de sélection d'actifs est enrichie par l'analyse fondamentale : les gestionnaires pourraient ainsi capter des informations qui ne seraient pas reflétées dans

²⁷ Prenons l'exemple d'une publication de chiffres trimestriels.

²⁸ Il s'agit ici de secteurs d'activité tels que la banque ou les télécommunications.

²⁹ La significativité est statistique (Gregory-Allen, Stangl, et Shawky, 2009).

les prix d'actifs. Par inférence, les auteurs estiment que toute surperformance peut être attribuée à l'utilisation d'information « fondamentale ». Ces auteurs complètent leur analyse en constatant que des « signaux quantitatifs » ne suffisent pas toujours pour estimer les rendements futurs.

Dans une logique de synthèse, une étude empirique menée par Fabozzi et al. (2008) rend compte de l'émergence d'une « approche hybride » utilisant des techniques quantitatives (probabilistes) et « fondamentales ». La prise en compte de données hétérogènes est envisageable dans le cadre de cette approche.

D'autres travaux ne permettent pas d'attribuer de valeur à l'approche fondamentale. Ahmed et Nanda (2005) mettent en avant l'idée que l'approche exclusivement quantitative, en dehors de toute considération humaine et émotionnelle, permet d'adopter une allocation d'actifs plus « objective ».

En élargissant notre champ d'investigation à d'autres Sciences « molles », nous constatons que certaines complémentarités subsistent également entre différents types d'information. En Sciences Sociales, il existe une combinaison de données hétérogènes lorsqu'il s'agit de compléter l'analyse d'enquêtes à partir d'informations brutes. En Sociologie, plusieurs recherches ont souligné la nécessité de combiner différents types de données. Les données de sondage, d'enquêtes peuvent être adossées à des informations complémentaires pour analyser certains contextes ou certaines opinions. Le traitement statistique (quantitatif) est nécessaire mais l'analyse du contexte permet d'affiner la compréhension de faits sociaux (Bourdieu, 2002). Cette complémentarité s'illustre surtout lorsque ces autres types d'information permettent de pondérer plus précisément les réponses communiquées par les personnes interrogées. Le « capital social » et l'origine sociale peuvent influencer sur les réponses données : considérer d'autres types d'information permet de mieux approcher le phénomène étudié. L'analyse du contexte, plus qualitative, vient assouplir le biais de « désirabilité sociale » (Evrard, Pras, et Roux, 2009) qui incite les personnes sondées à répondre en fonction de ce que la société voudrait qu'elles pensent et non en fonction de ce qu'elles pensent réellement. D'autres biais tels que la versatilité³⁰ de l'opinion peuvent également apparaître dans des enquêtes. Labov (2006), le fondateur de la Sociolinguistique quantitative, a identifié à travers ses travaux qu'une volatilité des réponses émerge lorsque les contextes sociaux changent. L'interprétation de résultats quantitatifs, de sondages, peut être facilitée avec la prise en considération d'informations complémentaires.

³⁰ Changement fréquent d'une opinion. Bourdieu (2002) critique à ce titre le concept d'opinion publique.

2. Données hétérogènes en univers multi-classes d'actifs

La prise en compte de données hétérogènes peut être déterminante dans un contexte multi-classes d'actifs financiers. Nous expliquons pourquoi nous considérons un univers multi-classes d'actifs. Nous analysons le lien qui peut exister entre les données hétérogènes et l'univers multi-classes d'actifs. Un enjeu de notre recherche est de poursuivre les travaux de ceux qui tentent de réunir des données hétérogènes au sein d'un univers multi-classes d'actif financiers. Comme évoqué dans l'Introduction Générale, nous sommes dans un contexte d'allocation tactique internationale multi-classes d'actifs où nous cherchons à mesurer l'influence que peut exercer la présence de données hétérogènes.

2.1. Complexité du cadre multi-classes d'actifs

Des travaux mettent en exergue le manque de recherche en Finance sur l'univers multi-classes d'actifs financiers, particulièrement sur le lien entre les annonces économiques et les mouvements de prix d'actifs. Andersen et al. (2007) recensent différents cas de figure dans la littérature et montrent que les recherches sur le lien entre informations « fondamentales » et l'univers multi-classes d'actifs restent limitées. Selon les auteurs, certains étudient le lien entre les publications macroéconomiques et les variations de prix d'actifs sur une classe et un pays (Balduzzi, Elton, et Green, 2001). Certaines recherches analysent les réactions de plusieurs classes d'actifs aux informations économiques mais sur un seul pays (Brenner, Pasquariello, et Subrahmanyam, 2009 ; Boyd, Jagannathan, et Hu, 2005) et d'autres s'intéressent à plusieurs pays mais sur une seule classe d'actifs (Andersen et al., 2003). Les auteurs rappellent que les recherches actuelles commencent désormais à se concentrer sur plusieurs classes d'actifs et plusieurs pays.

Des travaux soulignent la difficulté de prendre en considération un univers multi-classes d'actifs et des éléments explicatifs de ce constat. Comme le rappellent Blitz et Vliet (2008), la plupart des participants sur les marchés financiers se concentrent sur une seule classe d'actifs jugeant que l'univers multi-classes est un objectif trop « ambitieux » à atteindre. La plupart des intervenants se concentrent sur « l'allocation et la sélection de titres pour une certaine classe d'actifs ». Chaque classe d'actifs a, dans une certaine mesure, sa propre logique : « construire une stratégie d'allocation multi-classes basée sur la valeur de l'actif n'est pas quelque chose d'immédiat, car il n'y a pas une seule mesure de valorisation applicable à toutes les classes ». Ils estiment qu'il est parfois délicat de choisir des variables qui ont un lien avec un univers multi-classes d'actifs. Une variable comme les révisions

d'estimations de bénéfices des sociétés faites par les analystes est appropriée pour le marché des actions. Le marché obligataire ne possède pas de mesure équivalente. Ils soulèvent une question névralgique de notre réflexion : ils estiment qu'il n'existe pas de modèle d'évaluation unique pour toutes les classes d'actifs communément accepté.

L'introduction d'une combinaison de données hétérogènes permettrait de créer le lien manquant entre des variables « prédictives »³¹, non-spécifique à une seule classe d'actifs, et un modèle d'évaluation multi-classes d'actifs financiers en univers mondial. Dans la mesure où il n'y a pas un seul modèle généralisé de « *pricing* » pour toutes les classes d'actifs, la présence d'« autres informations de marché » aux cotés de données de cotation peut faciliter l'évaluation de classes d'actifs différentes par essence³².

Black et Litterman (1992) soulignent l'univers restreint des investisseurs qui ne sont concernés que par quelques actifs. Monarcha (2008) mentionne la spécialisation des gérants se concentrant uniquement sur quelques segments : « le nombre de données économiques étant en effet très important, une recherche intensive est nécessaire. C'est pourquoi les gérants tendent à se spécialiser sur quelques marchés ».

La prise en compte de données hétérogènes peut générer des situations d'arbitrage multi-classes mondiales. Nous nous intéressons aux interactions entre données hétérogènes et évolutions des prix d'actifs en univers multi-classes.

2.2. Interactions entre données hétérogènes en univers multi-classes d'actifs

La présence de données hétérogènes peut générer des mouvements distincts et/ou contraires sur les prix d'actifs financiers en univers multi-classes d'actifs. Afin d'envisager concrètement les stratégies de gestion multi-classes d'actifs en présence de données hétérogènes, nous commençons par faire référence au *global macro*³³ qui représente historiquement ces stratégies de manière la plus générique qui soit. Cette stratégie *global macro* est d'autant plus représentative qu'elle cherche à établir un lien entre des événements économiques (et autres) et une allocation multi-classes d'actifs mondiale. Nous rappelons que les champs d'investigations ne sont pas uniquement limités au *global macro* mais que cette stratégie est la plus répandue au sein de la gestion multi-classes d'actifs.

Joseph G. Nicholas³⁴ propose la définition suivante : « la stratégie *global macro hedge fund* possède le mandat le plus large de toutes les stratégies *hedge fund* où les gestionnaires

³¹ Blitz et Vliet (2008).

³² D'actifs entre classes et d'actifs au sein d'une même classe.

³³ Notons toutefois que n'est pas la seule stratégie multi-classes d'actifs.

³⁴ Fondateur et du Président du groupe *Hedge Fund Research* (Drobny, 2006).

ont la capacité de prendre des positions sur tous les marchés et sur tous les instruments financiers ». En adoptant une lecture sémantique du *global macro*, l'auteur explique que la partie *macro* de l'expression vient de la tentative des gestionnaires d'exploiter des principes économiques afin d'identifier des dislocations entre les prix d'actifs et la partie *global* de l'expression indique que ces dislocations peuvent se produire partout dans le monde³⁵. Au travers de cette définition, nous comprenons que les termes « *global* » et « *macro* » peuvent se compléter. Le terme « dislocations » fait également référence à la notion d'arbitrage, ici essentielle dans notre contexte de recherche sur l'univers multi-classes d'actifs.

Chan et al. (2005) poursuivent dans l'analyse du lien entre ces données hétérogènes et les arbitrages possibles : « les gestionnaires *global macro* tiennent des positions d'achats et de ventes sur tous les marchés mondiaux de capitaux ou dérivés. Ces positions reflètent leur vues sur la direction globale du marché, influencée par les principales tendances économiques et principaux événements [...] partout dans le monde dans des pays développés et des marchés émergents ». Cette définition intègre une composante géographique qui permettrait d'envisager différentes sources de rentabilités potentielles à l'échelle mondiale.

L'univers international et multi-classes d'actifs accorde une place conséquente à l'analyse *top down* et à l'Economie. Un complément de définition est donné par Stulz (2007) qui souligne le lien précédemment décrit : « les gestionnaires ont tendance à utiliser une approche *top-down* globale dans le but de prévoir comment les événements macroéconomiques et politiques vont affecter les valorisations des instruments financiers ». Il précise que des situations d'arbitrage multi-classes peuvent se manifester en évoquant : « de mauvaises évaluations de prix que les gestionnaires ont tendance à identifier [...] affectant la valorisation des instruments financiers ». Selon Edhec Risk Institute³⁶, la stratégie *global macro* « permet de participer à tous les marchés importants actions, obligations, devises et matières premières pas forcément au même moment ». D'après le même institut, « les fonds *global macro* ont pour but de tirer avantage des changements des économies mondiales ». Nous nous appuyons, entre autres, sur ces éléments de définition pour établir la pertinence d'une présence d'« autres informations de marché » au sein d'un univers multi-classes d'actifs.

La prise en compte de données hétérogènes peut engendrer des opportunités d'arbitrage entre classes d'actifs (et entre actifs d'une même classe). Il existe des réactions contraires entre classes d'actifs lors de la publication d'un même ensemble d'indicateurs économiques (Brenner, Pasquariello, et Subrahmanyam, 2009 ; Andersen et al., 2007).

³⁵ Cette définition nous aidera partiellement à expliquer et commenter le choix de données pour notre étude empirique du chapitre 3.

³⁶ <http://www.edhec-risk.com/>

Un premier type de recherches mesure l'impact (et sa persistance) de la publication d'enquêtes économiques sur les performances quotidiennes d'actifs financiers ainsi que sur la volatilité « conditionnelle » à ces annonces. Ils tentent d'établir un lien entre différents marchés financiers (actions et obligations) et l'« économie réelle » pour en conclure que l'« arrivée d'informations macroéconomiques a un impact significatif statistiquement et économiquement » sur différentes classes d'actifs aux Etats-Unis. Les auteurs évaluent l'impact de ces annonces économiques sur les rendements en excès d'un *benchmark*³⁷ ainsi que sur la volatilité « conditionnelle » estimée par un modèle “*Dynamic Conditional Correlation*” (DCC) (Engle, 2002) flexible³⁸.

Un deuxième type de recherches participe également à la discussion sur la complémentarité entre données hétérogènes en univers multi-classes d'actifs. Dans un cadre de gestion « haute fréquence », les auteurs observent que les cotations internationales de contrats à terme³⁹ de plusieurs classes d'actifs sont liées aux « fondamentaux » tels que des informations macroéconomiques trimestrielles, mensuelles et hebdomadaires aux Etats-Unis. Ces auteurs soulignent que les réactions à ces annonces économiques diffèrent selon les classes d'actifs. Ils estiment que la corrélation entre différentes classes d'actifs évolue en fonction des cycles économiques. Ces différentes phases d'activités peuvent être estimées par des séries choisies au sein d'« autres informations de marché »⁴⁰. Ces différences de réponses des rendements laissent penser que pour une même information (annoncée), des arbitrages entre classes et entre actifs au sein de ces classes sont possibles.

D'autres recherches mettent en avant cette notion d'« arbitrage macroéconomique » entre classes laissant la possibilité d'exploiter aussi des données microéconomiques (Monarcha, 2008). Cette définition ne mentionne pas s'il convient d'utiliser les données économiques par une approche discrétionnaire ou systématique. Blitz et Vliet (2008) proposent d'exploiter des variables comme le sentiment des consommateurs, qui sont supposées avoir une influence sur les rendements de plusieurs classes d'actifs. Des opportunités d'arbitrage se matérialisent lorsque des réactions asymétriques entre classes d'actifs (par exemple actions et obligations) se manifestent pour la publication de la même information économique ou lorsque des inefficiences entre classes d'actifs sont causées par

³⁷ Bon à trois mois du Trésor américain.

³⁸ Le DCC ici spécifié est une extension du modèle GARCH(1,1) (Bollerslev, 1986). Selon Brenner et al. (2009), les modèles GARCH ne sont pas flexibles et le nombre de paramètres est trop nombreux pour introduire des formes complexes de comouvements conditionnels entre classes d'actifs. D'après ces auteurs, le DCC a la flexibilité d'un GARCH univarié sans la complexité des spécifications du modèle GARCH multivarié. Notons que la significativité statistique est mesurée au travers des estimations du quasi maximum de vraisemblance (QML) et leurs « robustes » *t*-stats tels que Bollerslev et al. (1992) les calculent.

³⁹ Comme nous le verrons dans le chapitre 3, une partie de nos données correspondent à des contrats à terme.

⁴⁰ Voir le chapitre 3.

des effets comportementaux des investisseurs avec un montant de *smart money* insuffisamment disponible pour « arbitrer activement » ces inefficiences.

3. Arbitrage statistique : stratégies historiques et limites

Une partie de notre cadre théorique est lié aux travaux sur l'arbitrage statistique. Comme cela est précisé dans l'Introduction Générale, cette technique d'arbitrage est au cœur de nombreuses discussions et inspire nos principaux développements empiriques⁴¹.

Nous rappelons que l'hypothèse d'absence d'opportunité d'arbitrage est à la base de nombreux travaux en Finance. S'appuyant sur des travaux fondateurs (Sharpe, 1964 ; Markowitz, 1956, 1952), Ross (1976) élabore une théorie de fixation des prix à partir de modèles multi-facteurs reposant sur le postulat d'absence d'opportunité d'arbitrage. Au sein d'un univers de marchés financiers, cette théorie démontre l'existence d'une relation linéaire entre les primes de risque des actifs risqués (et donc leur rendement moyen) et les primes de risque des facteurs de risques principaux. Ceci permet d'évaluer le prix d'équilibre des actifs risqués de l'univers considéré.

L'arbitrage statistique s'inscrit dans une logique différente. Sous sa forme la plus quantitative, il consiste par exemple à exploiter les propriétés de retour à la moyenne des prix ou des rendements d'un actif. En pratique, sa mise en œuvre en terme de stratégies de gestion revêt différentes formes plus ou moins quantitatives qui ont, de plus, évoluées dans le temps. Pour la classe d'actifs des actions, Khandani et Lo (2007) distinguent trois catégories de stratégies d'arbitrage : l'arbitrage statistique, le « *quantitative equity market-neutral* » et le « *long/short equity* ». Cette dernière stratégie fait référence à tout type de portefeuille impliquant des ventes à découvert et pouvant être marché-neutre ou non, quantitative ou non⁴².

Nous nous concentrerons sur les stratégies du premier type, soit l'arbitrage statistique. Nous pouvons toutefois trouver une caractéristique commune aux trois types de stratégies décrites par les auteurs : déterminer un équilibre statistique autour duquel oscillent les prix d'actifs de manière aléatoire. L'objectif final est de capturer numériquement un équilibre puis de comprendre statistiquement l'aléa autour de cet équilibre.

Dans ce chapitre, nous décrivons et analysons les concepts relatifs à l'arbitrage statistique en faisant référence à l'existant dans la littérature consacrée au sujet. Ces éléments viennent principalement de la classe d'actifs des marchés actions, car historiquement, les

⁴¹ Voir les chapitres 4 et 5.

⁴² La stratégie non-quantitative est souvent qualifiée de fondamentale. Dans le monde de la gestion d'actifs, il est fait référence à la gestion de conviction.

travaux de recherches sur l'arbitrage statistique portaient principalement sur ces marchés.

Khandani et Lo (2007) proposent une définition de l'arbitrage statistique qui résume l'ensemble de ce type de stratégies : « des stratégies court-terme hautement techniques de retour à la moyenne impliquant un nombre important de titres, avec des courtes périodes de tenues de positions (de l'ordre de quelques secondes à plusieurs jours) et des infrastructures conséquentes de *trading*, de calculs et technologiques ».

Spécifiquement, l'arbitrage statistique est un moyen de générer des stratégies de gestion systématique en composant des portefeuilles d'actifs financiers. Par stratégie de gestion, il est entendu que les techniques d'arbitrage statistique permettent de produire des indicateurs de *market timing* (moment d'achat et/ou de vente). La composition ou la construction de portefeuilles fait référence aux proportions, quantités que nous souhaitons allouer sur les actifs d'un portefeuille. Plusieurs travaux de référence sur l'arbitrage statistique (Avellaneda et Lee, 2010 ; Lo et MacKinlay, 1990) permettent de comprendre les principales dynamiques de ce type d'arbitrage.

3.1. Stratégies historiques

Historiquement, le concept d'arbitrage repose sur le concept de « paris contrariants » (Lehmann, 1990 ; Lo et MacKinlay, 1990). Les « paris contrariants » consistent à acheter les « perdants » d'hier et à vendre les « gagnants » d'aujourd'hui en espérant qu'un processus de retour à la moyenne se produise. Les auteurs ont étudié l'impact sur la rentabilité des actions du phénomène contrariant. Nous définissons les « gagnants » (respectivement « perdants ») comme des actifs surévalués (sous-évalués) par rapport à un marché de référence⁴³. Le fil

⁴³ Un retour sur différentes notations utilisées par Lo et MacKinlay (Lo et MacKinlay, 1990), nous permet d'expliquer leurs travaux. Soit N le nombre d'actifs du « marché » $R_{i,t}$, $1 \leq i \leq N$ le rendement de l'actif i en date t , calculé entre t et $t-1$ (base quotidienne). L'actif i est considéré comme surévalué si :

$$R_{i,t} > R_{m,t} := \frac{1}{N} \sum_{j=1}^N R_{j,t}$$

et sous-évalué dans le cas contraire. Dans ce cadre $R_{m,t}$ est la moyenne arithmétique de tous les rendements du marché considéré. C'est une première manière de définir le marché de référence. Les auteurs suggèrent également une allocation d'actifs sur ce critère : soit $Q_{i,t}$ le montant (dans la monnaie domestique, par exemple en euros ou en dollars) investi dans l'actif i . Ils proposent l'allocation suivante :

$$Q_{i,t} = -\frac{1}{N} (R_{i,t} - R_{m,t}) = -\frac{1}{N} \left(R_{i,t} - \frac{1}{N} \sum_{j=1}^N R_{j,t} \right)$$

conducteur des travaux des auteurs est que cette stratégie de gestion bénéficie du paradigme du retour à la moyenne associé à la fois au « *market overreaction* » et à l'effet « *cross autocovariance* ». Concernant, l'effet « *market overreaction* », si un actif est jugé sous-évalué alors l'adoption d'une position *long* implique que le prix va augmenter et inversement si le rendement de l'actif est jugé surévalué. L'effet « *cross-autocovariance* » décrit une relation de covariance faisant intervenir deux actifs distincts à deux périodes de temps différentes. En supposant que X est un « gagnant », c'est-à-dire qu'il croît à partir de la date t et qu'il est déjà surévalué, la position cohérente avec la logique des auteurs des stratégies « contrariantes » est de vendre X à la date t . Le profit sur cette position doit être positif si $X_{t+h} < X_t$. Maintenant, soit Y un autre actif tel que $Cov(X_t, Y_{t+h}) > 0$. Si X croît à la date t alors Y doit probablement croître en date $t+h$. En terme de gestion, une position *long* est adoptée sur l'actif Y en date t afin de bénéficier d'un profit dû à une hausse de Y entre t et $t+h$ ⁴⁴.

Citons également le *pair trading*⁴⁵ souvent décrit comme l'ancêtre de l'arbitrage statistique tel qu'il est pratiqué aujourd'hui. Originellement, ce type de gestion a également émergé sur la classe actions. Le *pair trading* vise à arbitrer une action contre une autre⁴⁶. Le *pair trading* est une déclinaison de la stratégie *long/short* qui est envisageable sur toutes les classes d'actifs. Do et al. (2006) analysent un ensemble de méthodes permettant de développer quantitativement des stratégies de *Pair trading*. Pole (2007) analyse également le *pair trading* et fournit une revue globale de l'arbitrage statistique. Il existe la possibilité de mettre en place une stratégie *long/short* en dehors de toute considération statistique, en utilisant une approche fondamentale.

Nous avons précédemment décrit l'arbitrage d'un actif par rapport à un marché puis l'arbitrage entre deux actifs. Un modèle multi-facteurs permet une approche plus synthétique. Un tel modèle permet de décomposer le rendement $R_{i,t}$ d'un actif i au sein d'un univers de

Un tel choix respecte automatiquement une contrainte « dollar neutre ». Nous remarquons qu'avec ce type d'allocation, chaque actif du marché considéré entre dans la composition du portefeuille. Une exposition *long* ou *short* dépend du signe de $Q_{i,t}$.

⁴⁴ Constatée en $t+h$.

⁴⁵ Si X et Y sont les séries temporelles de prix d'actifs ayant les mêmes caractéristiques, le *pair trading* s'écrit comme :

$$\frac{X_t - X_{t-1}}{X_{t-1}} = \beta \frac{Y_t - Y_{t-1}}{Y_{t-1}} + \frac{Z - Z_{t-1}}{Z_{t-1}},$$

où Z est un processus stationnaire ou de retour à la moyenne. Le processus fait référence au résidu de la régression. Un tel portefeuille *long/short* peut osciller autour d'un équilibre statistique dépendant des propriétés statistiques de Z .

Une telle modélisation suggère d'être *long* d' 1 € sur l'actif X et *short* de $\beta \text{ €}$ sur Y si Z est significativement négatif (ou faible). Le cas contraire est d'être *short* d' 1 € sur X and *long* de $\beta \text{ €}$ sur Y si Z est significativement positif (fort).

⁴⁶ Généralement du même secteur.

marchés financiers, entre deux dates $t-1$ et t par exemple, par une combinaison linéaire finie d'un nombre fini de facteurs F_j et d'un terme résiduel indépendant :

$$\frac{X_t - X_{t-1}}{X_{t-1}} = \alpha_i + \sum_{j=1}^m \beta_{i,j} F_{j,t} + (Z_{i,t} - Z_{i,t-1}).$$

Les facteurs F_j sont supposés être centrés et décorréllés, c'est-à-dire $E(F_j F_k) = 0$ si $j \neq k$.

Dans ce type de régression, le terme α_i désigne le rendement moyen de l'actif i et

$\varepsilon_{i,t} = (Z_{i,t} - Z_{i,t-1})$ désigne le risque spécifique de l'actif. Ce risque spécifique est supposé être indépendant des facteurs F_j .

Les m premiers termes de la relation sont assimilés au risque systématique en gestion de portefeuille (c'est-à-dire au risque non-diversifiable ou risque de marché), en quantifiant l'exposition du rendement de l'actif i au marché. Les derniers termes $\varepsilon_{i,t} = (Z_{i,t} - Z_{i,t-1})$ renvoient au risque idiosyncratique ou spécifique de l'actif i c'est-à-dire le risque diversifiable, « diluable ».

En pratique, l'identification et l'évaluation de ces facteurs à partir seulement des rendements des actifs sont délicates. Dans son article séminal sur l'« *arbitrage pricing theory* », Ross (1976) suppose l'existence d'un tel modèle multi-facteurs, avec des facteurs donnés de manière exogène, afin d'en déduire, grâce à des hypothèses d'absence d'opportunité d'arbitrage, une relation entre la prime de risque des actifs risqués et celles des facteurs de risques non-diversifiables (systématique). Ces facteurs de risque sont souvent remplacés par des indices de marché (indices actions, indices sectoriels, fonds indiciels cotés par exemple) afin de pouvoir régresser le rendement des actifs risqués sur ces mêmes indices. Différents auteurs ont montré que ce modèle multi-facteurs pouvait également être obtenu, de façon endogène, après une analyse en composantes principales des rendements sans que cela ne change la conclusion initiale de Ross (1976)⁴⁷.

Par ailleurs, dans les travaux d'Avellaneda et Lee (2010) il est fait appel à un processus paramétrique pour quantifier le phénomène de retour à la moyenne (*mean reversion*) du résidu Z_i qui est dans un processus d'Ornstein-Uhlenbeck (Uhlenbeck et Ornstein, 1930). Un tel processus est solution de l'équation différentielle stochastique (EDS) suivante :

⁴⁷ Voir également les travaux (Chamberlain et Rothschild, 1983).

$$dZ_i(t) = k_i(m_i - Z_i)dt + \sigma_i dB_i(t),$$

où k_i est le taux de retour à la moyenne (alternativement, on pose $\tau_i = \frac{1}{k_i}$ qui représente la longueur de temps caractéristique pendant laquelle se produit le phénomène de retour à la moyenne), m_i la moyenne autour de laquelle Z_i oscille et σ_i un paramètre de dispersion. Ici B_i est un mouvement brownien standard, ainsi Z_i est un processus gaussien⁴⁸. Les auteurs montrent comment s'effectue l'estimation de ces paramètres (k_i, m_i, σ_i) en utilisant des données sur des fenêtres glissantes. Les actifs dont le taux de retour est trop « faible »⁴⁹ sont assimilés à un processus de racine unitaire (marche aléatoire, processus non stationnaire) et sont exclus du portefeuille composé. En utilisant ces paramètres, le *spread* $Z_i - m_i$ peut être estimé précisément, permettant d'obtenir des signaux de *market timing*. Au final, leur allocation est effectuée sur la base de portefeuilles « dollar neutre » (Avellaneda et Lee, 2010 ; Khandani et Lo, 2007) et « 1/N » (diversification « naïve »).

Notons que Do et al. (2006) dévoilent des résultats empiriques de trois couples d'actions impliqués dans une stratégie de *pair trading* pour les bourses australiennes, américaines et anglaises. Ces auteurs modélisent le comportement de retour à la moyenne comme statistiquement significatif dans le cadre de *mispricing* de paires d'actions : le but restant d'être *long* sur une action et *short* sur une autre. En outre, les auteurs utilisent la modélisation de Vasicek (1977)⁵⁰. La stationnarité des résidus $Z_{i,t} = \sum_{s \leq t} \varepsilon_{i,s}$ construite sur le risque spécifique ε_i est centrale dans les études précédentes.

Une autre stratégie historique en relation avec l'arbitrage statistique permet de composer avec la non-stationnarité de nombreuses séries financières et économiques : l'approche en cointégration, qui est une modélisation possible du processus de retour à la moyenne. L'idée principale de la cointégration appliquée aux séries financières est que les marchés financiers doivent évoluer conjointement (avec une tendance commune) s'ils sont intégrés⁵¹. Dans ce cadre théorique, il est supposé que les situations d'arbitrage se présentent lorsque les marchés ne sont pas intégrés et des stratégies *long/short* peuvent permettre d'en tirer profit. Les séries peuvent être cointégrées, mais il est souvent préférable que les relations

⁴⁸ En utilisant le calcul standard d'Itô, nous identifions la distribution du résidu.

⁴⁹ Avellaneda et Lee (2010) prennent comme définition d'un temps de retour trop « faible » lorsqu'il est supérieur à la moitié de la largeur de la fenêtre mobile utilisée pour les estimations. Notons que cette restriction nécessite de travailler sur un ensemble homogène de titres afin que la distribution des temps de retour entre les actifs soit bien identifiée.

⁵⁰ Qui est une version modifiée du processus d'Ornstein-Uhlenbeck (Uhlenbeck et Ornstein, 1930).

⁵¹ Il s'agit de la notion d'intégration en Économétrie (Bourbonnais et Terraza, 2008 ; Bourbonnais, 2008). Ces références correspondent respectivement à une synthèse sur les séries temporelles et à une explication du concept d'intégration.

découvertes aient un sens économique, correspondent à une explication en lien avec des fondamentaux⁵². L'objectif de la cointégration est d'estimer des coefficients de régression supposés stables dans le temps. La cointégration a été introduite par Granger (1981) dans le but d'étudier les propriétés de séries non-stationnaires. La théorie de la cointégration porte sur la transformation de séries non-stationnaires en une combinaison de ces mêmes variables, qui devient elle-même stationnaire. Murray (1994) illustre la relation de cointégration par l'image d'un homme ivre tenant un chien par une laisse. Leur distance respective est ainsi bornée puisque la laisse a une extension maximale et le « va-et-vient » entre les deux êtres matérialise ce phénomène de retour.

Il existe deux approches majeures de la cointégration : celle d'Engle et Granger (1987) et celle de Johansen (1991). La première approche permet, dans un cadre univarié, d'estimer des modèles à correction d'erreur (Davidson et al., 1978)⁵³ alors que la deuxième est développée dans un cadre multivarié. La méthode d'estimation de ces deux approches diffère, mais l'avantage que présente l'approche de Johansen est de distinguer plusieurs relations de cointégration alors que celle de Engle et Granger (1987) ne permet de n'en détecter qu'une seule. Le modèle le plus répandu est celui du « *Vector Error Correction model* » (*VECM*) de Johansen (1991). Les tests permettant la détection de relations de cointégration s'effectuent dans un cadre économétrique⁵⁴. Ces tests sont détaillés par Johansen et Juselius (1990)⁵⁵.

Nous complétons notre revue des principales stratégies historiques par une analyse de l'intérêt de l'analyse technique, laquelle fait l'objet de nombreuses réserves quant à sa validité scientifique. L'analyse technique permet à l'aide de certains indicateurs de tenter une identification d'un retour à la moyenne d'un actif. Elle repose avant tout sur l'hypothèse d'efficience des marchés en supposant que toute l'information disponible est impactée instantanément dans les prix d'actifs, que les prix d'actifs suivent des tendances et que les cycles économiques se reproduisent (Béchu, Bertrand, et Nebenzhal, 2008).

Cependant, il est difficile de quantifier l'effet de l'analyse technique car prédire le comportement humain est délicat et les tendances ne sont pas toujours identifiables uniquement avec l'analyse technique. En analyse technique toute l'information est supposée être incorporée dans les prix d'actifs financiers. L'information fondamentale (résultats

⁵² Les tests proposés par deux approches de (Johansen et Juselius, 1990 ; Johansen, 1991) permettant de distinguer n relations mais ne permettent pas d'identifier quelles sont les relations de cointégration entre les variables. C'est à cette phase de la procédure que peut intervenir la recherche de sens économique, fondamental entre les séries.

⁵³ MCE ou plus communément *ECM* pour « *error correction model* » (Davidson et al., 1978).

⁵⁴ Les logiciels EViews 7 ou SAS permettent par exemple d'effectuer des tests de spécification de modèle pour une approche en cointégration ainsi que des tests de détection de variables cointégrées.

⁵⁵ Les deux tests permettant de détecter éventuellement des relations de cointégration entre des séries sont de la trace de la trace (Johansen et Juselius, 1990 ; Osterwald-Lenum, 1992) et celui de la valeur propre maximale.

d'entreprise, enquêtes économiques) n'a aucune place⁵⁶ en tant que telle. La variation de prix d'un actif n'est pas toujours due à un phénomène psychologique et l'effet supposé auto-réalisateur dû à l'utilisation de ces techniques biaiserait toute démonstration sur l'efficience. L'analyse technique porte en elle-même un caractère auto-réalisateur. Malgré son caractère discutabile sur un plan scientifique, de nombreux gestionnaires y prêtent une attention particulière en intégrant tout ou partie de ces signaux dans leurs décisions de gestion.

Des approches utilisant la statistique existent en analyse technique pour la détection de tendance et l'identification de processus de retour à la moyenne⁵⁷. Ces indicateurs suivent le retournement de tendance, la dynamique de prix d'un actif va s'inverser ou non. Nous pensons notamment aux différents indicateurs de suivi de tendance *trend following* qui cherchent à « épouser » les tendances et leur changement au plus près. Des indicateurs de lissage ont pour vocation d'identifier également des processus de retour à la moyenne en tentant de réduire le « bruit » contenu dans les séries de prix d'actifs : c'est ce que le *moving average convergence divergence* (MACD) se propose d'effectuer. Les stratégies d'arbitrage statistique présentent, comme d'autres approches, certaines limites théoriques et empiriques qu'il convient d'analyser.

3.2. Limites et interrogations

Les processus de gestion relève du *market timing*. En amont de ce déclenchement d'ordre d'achat et/ou de vente, la phase d'allocation est souvent préliminaire. La stationnarité des résidus semble être le facteur clé à évaluer concernant le *market timing*. Nous revenons dans cette section sur quelques concepts statistiques et économétriques relatifs à la stationnarité ou à la non-stationnarité des séries temporelles. Nous analysons certaines difficultés de mise en œuvre lorsque des différences entre hypothèses théoriques et réalités empiriques existent.

Commençons par un exemple de stationnarité des résidus pour souligner les limites que cette dernière comporte. Dans les travaux d'Avellaneda et Lee (2010), nous étudions les propriétés du résidu Z_i , définies précédemment. Nous omettons à présent l'indice i par souci de simplicité. Le processus Z est défini en temps continu mais ses valeurs sont généralement obtenues en temps discret (à partir de prix de clôtures quotidiens par exemple). Dans ce cas, soit $\Delta t = \frac{1}{252}$ l'intervalle de temps (en base quotidienne). En considérant les valeurs de $Z(t)$ à un instant $t = t_k = k\Delta t$, il est possible d'estimer les paramètres (k, m, σ) à l'aide d'une

⁵⁶ Puisque toute l'information est censée être impactée dans les prix d'actifs financiers.

⁵⁷ Voir Béchu et al. (2008).

régression de type moindres carrés ordinaires (MCO)⁵⁸ :

$$Z(t_{k+1}) = aZ(t_k) + b + \varepsilon_{k+1}.$$

A l'aide des précédentes relations, nous pouvons calculer les paramètres a , b et ε_{k+1} . Il en est de même pour les paramètres k, m, σ . Nous avons identifié au travers de l'équation ci-dessus un processus discret $Z_k \sim Z(t_k)$ ⁵⁹.

Les concepts de stationnarité et de stabilité des coefficients de régression ne sont pas toujours empiriquement vérifiés pour les séries financières et économiques. L'étude historique de Nelson et Plosser (1982) rappelle la non-stationnarité de nombreuses séries macroéconomiques américaines qui indiquent une instabilité des coefficients de régression dans les régressions linéaires. Le calcul du coefficient de corrélation entre deux séries est avantageux par rapport à la recherche d'estimations plus complexes car il est relativement simple à obtenir.

Pour autant, nous demeurons attentifs au fait que les matrices de variances-covariances des rendements d'actifs ne sont pas stables dans le temps comme nous le rappellent Boyer, Gibson et Loretan (1997), Forbes et Rigobon (2002). Ces derniers ont trouvé que le coefficient de corrélation est toujours biaisé en périodes de fortes volatilités. Le coefficient de corrélation ne semble pas, dans cette optique, adapté pour prendre en compte des relations non-linéaires ou des distributions « leptokurtiques ».

Selon Roll (1992), le coefficient de corrélation n'est pas suffisant pour estimer l'intégration des séries. Certaines corrélations entre indices sont « faibles » car les marchés, représentés par les indices actions, ont des structures industrielles différentes. Cette analyse peut également renvoyer à la nécessité d'informations fondamentales pour comprendre et anticiper l'évolution des rendements d'actifs financiers.

Bien que les approches en cointégration aient ouvert la voie à la modélisation de séries non-stationnaires, certaines réserves peuvent être formulées quant à leur mise en place. La spécification d'un modèle en cointégration peut être confrontée aux difficultés de calcul de la moyenne vers laquelle se produit la convergence, de détermination de la période qui peut potentiellement dévoiler une relation de cointégration ou une destruction de cette relation. Les différents tests de racines unitaires présentent des limites⁶⁰ comme l'autocorrélation des résidus, la mauvaise spécification de tendance déterministe ou la non-prise en compte de

⁵⁸ La régression MCO réclame de vérifier au préalable des hypothèses stochastiques et structurelles. Se reporter à Bourbonnais (1998).

⁵⁹ C'est un exemple d'un processus autorégressif à l'ordre 1, soit un processus AR(1).

⁶⁰ Voir d'autres travaux (Bourbonnais et Terraza, 2008 ; Bourbonnais, 2008).

ruptures de tendances. Les approches en cointégration nécessitent parfois d'être complétées par des procédures d'optimisation de signal (Alexander et Dimitriu, 2003) ainsi que par des politiques de gestion du risque.

Dans le cadre de l'arbitrage statistique, la cointégration est une des techniques historiques possible mais c'est une approche qui impose que certains critères économétriques soient strictement respectés tels qu'un même ordre d'intégration des séries. Nous verrons dans les chapitres 2 et 3 que notre approche de l'arbitrage statistique ne repose pas sur des hypothèses économétriques aussi strictes.

Enfin, les tests de la trace ainsi que celui de la valeur propre maximale, permettant de déterminer le nombre de relations de cointégration dans l'approche de Johansen (1991), utilisent des fonctions de vraisemblance à partir de distributions gaussiennes. Or le caractère gaussien n'est pas toujours adapté aux séries économiques et financières.

Interrogeons-nous sur la portée et la pertinence d'un modèle de retour à la moyenne dans un univers multi-classes d'actifs. Les travaux que nous avons mentionnés ont pu dans certaines configurations apporter des réponses dans un univers mono-classe. Lorsque Riva (2008) évoque la notion de processus de prix permettant de modéliser la dynamique d'un actif sous-jacent reposant sur l'utilisation d'un mouvement brownien géométrique⁶¹, ce dernier insiste sur le fait que ce type de modélisation est adapté pour les actions et les indices boursiers mais qu'il est impropre pour modéliser la dynamique des taux d'intérêt.

Partant de cette constatation, les modèles d'arbitrage statistique reposant sur les prix ne peuvent pas avoir un taux de retour unique à la moyenne pour toutes les classes d'actifs agrégeant d'autres informations de marché au prix des actifs. Il serait intéressant de pouvoir définir un modèle d'arbitrage statistique permettant d'estimer les processus de prix d'actifs, à l'aide d'indicateurs comme les « autres informations de marché », en univers multi-classes d'actifs. Les enquêtes d'opinion, de conjoncture et les enquêtes de sentiment des marchés financiers ne sont pas spécifiques à une seule classe d'actifs : ceci engendre la possibilité que toutes les classes d'actifs puissent être impactées (pareillement ou non) par ces « autres informations de marché » (Blitz et Vliet, 2008 ; Brenner, Pasquariello, et Subrahmanyam, 2009).

Les politiques de *market timing* peuvent s'inscrire dans des processus d'allocation. Nous étudions les deux principales formes de diversification à l'aide d'une analyse comparative.

⁶¹ Dans le cadre de simulations de Monte Carlo. Voir Riva (2008), p. 226.

4. Diversification de portefeuille

Nous introduisons les deux principales familles de diversification de portefeuille. A partir de la deuxième moitié du vingtième siècle, la diversification « naïve » de type « 1/N » et la diversification optimale ont connu de nombreuses évolutions tant d'un point de vue théorique qu'empirique. Nous étudions également les deux principales formes de diversification au travers d'une analyse comparative.

4.1. Diversifications « naïve » et optimale

Pour définir une stratégie d'allocation optimale⁶², nous prenons comme référentiel de départ le cas de la diversification « naïve » afin de mettre en exergue leurs différences fondamentales. Nous rappelons que les stratégies « naïves » sont analysées par des recherches diverses (Benartzi et Thaler, 2001 ; DeMiguel, Garlappi, et Uppal, 2009 ; Huberman et Jiang, 2006). A la différence de la diversification « naïve », l'allocation optimale attribue des poids différents aux actifs qui composent un portefeuille. Cette composition peut être obtenue par une procédure d'optimisation. Il est techniquement possible de réaliser cette pondération (inégaie⁶³) à travers une décision discrétionnaire⁶⁴.

Un concept essentiel pour les stratégies d'allocations optimales repose sur l'idée que les marchés financiers ne sont pas statiques mais dynamiques. Lorsque Merton (1980) rend compte du changement dans le temps de la variance des rendements du marché, Stockton et Shtekhman (2010) font référence à des ajustements de manière active en fonction de prévisions de marché à court-terme.

La diversification optimale de portefeuille peut être une déclinaison de l'allocation tactique visant à s'adapter aux fluctuations de court-terme des prix d'actifs financiers : il s'agit d'après les auteurs d'une « stratégie dynamique qui ajuste de manière active l'allocation stratégique des actifs d'un portefeuille en fonction de prévisions de marché à court terme. Cette méthode d'allocation vise à exploiter de manière systématique les inefficiences ou déviations temporaires de la valeur d'équilibre de différentes classes ou sous-classes d'actifs ». Ces changements se traduisent concrètement par des modifications de pondérations pour les actifs composant un portefeuille d'arbitrage ou un portefeuille qui n'a aucune vocation à réaliser des arbitrages. L'allocation tactique vient modifier dynamiquement les

⁶² Une multitude de stratégies d'allocations optimales sont disponibles. L'objectif n'étant pas dans nos recherches de toutes les traiter mais d'en cerner les principaux aspects.

⁶³ Dans la majeure partie des cas, les poids du portefeuille ne sont pas égaux.

⁶⁴ Non-systématique.

choix effectués au niveau de l'allocation stratégique qui elle s'étend sur un horizon plus long-terme. Les poids du portefeuille varient dans le temps et ne sont pas forcément égaux : des procédures d'optimisation peuvent prendre en charge ces changements dans le temps. Pour comprendre ce que sont les sur ou sous-pondérations maîtrisées, nous pouvons faire appel au « *processus d'optimisation des pondérations d'actifs* » (Stockton et Shtekhman, 2010) qui aboutit à un changement dynamique des poids d'un portefeuille.

Depuis la deuxième moitié du XX^{ième} siècle, les procédures d'optimisation de portefeuille ont connu de nombreuses évolutions. Les différentes formes d'optimisation des poids d'un portefeuille ainsi que leurs évolutions successives à travers le temps sont détaillées dans plusieurs travaux (Maillard, Roncalli, et Teleitche, 2010 ; DeMiguel, Garlappi, et Uppal, 2009 ; Clarke, Silva, et Thorley, 2006 ; Viviani, 2001).

D'un point de vue plus global, Park et al. (2009) évoquent d'ailleurs « une augmentation rapide de la quantité et de la complexité de la recherche quantitative en Finance due à des avancées de la théorie financière et des techniques statistiques ». Des approches « moyenne-variance » (Markowitz, 1956 ; Markowitz, 1952) ont permis de sélectionner les poids d'un portefeuille qui, pour un niveau de variance donné, maximisent l'espérance de rentabilité. Bien que cette approche comporte certaines limites⁶⁵, elle reste un travail fondateur de l'optimisation de portefeuille.

D'autres formes d'optimisations correspondent à la résolution partielle d'un programme d'optimisation sous contraintes (Elton et Gruber, 1995). Il s'agit de maximiser l'équivalent du ratio de Sharpe sous contrainte que la somme des poids soit égale à 1. Ces formes de résolution partielle d'un programme d'optimisation permettent notamment de converger vers la solution plus rapidement que l'utilisation d'un solveur « classique »⁶⁶. Plus tardivement, l'avènement de l'approche bayésienne⁶⁷ de l'erreur d'estimation des rendements attendus a confirmé le besoin d'évolution des méthodes d'optimisation du cadre « moyenne-variance ». Brièvement, l'approche bayésienne repose sur l'estimation des rendements et de la matrice de variance covariance calculée en utilisant une distribution prédictive des rendements attendus. Différents développements ont vu le jour dont les principaux sont le « *Bayesian diffuse-prior portfolio* » (Brown, 1979 ; Klein et Bawa, 1976) et « *shrinkage estimators* » (Jorion, 1986), ces derniers étant utilisés pour encadrer l'erreur d'estimation des rendements espérés (Stein, 1956).

Pour s'adapter à ces évolutions, différents programmes d'optimisation constituent un moyen tangible pour gérer systématiquement le choix des poids d'un portefeuille dans le

⁶⁵ Nous analysons ces limites dans la section suivante.

⁶⁶ Un solveur de type Excel 2010.

⁶⁷ DeMiguel et al. (2009) propose une revue de littérature sur les approches bayésiennes.

temps. A la différence de portefeuilles « 1/N », où le poids de chaque actif est le même⁶⁸, les avancées scientifiques sur les stratégies optimales sont en partie dues aux progrès sur les formes d'optimisation appliquées à l'allocation de portefeuille. L'approche par optimisation permet de répondre à l'émergence du paradigme du « *risk budgeting* » (Maillard, Roncalli, et Teleitche, 2010) qui consiste à évaluer monétairement⁶⁹ l'impact de tout changement de poids d'un actif sur le risque total d'un portefeuille. Le passage du paradigme de la performance au paradigme du « *risk budgeting* » est devenu « une pratique standard pour les investisseurs institutionnels. Il s'agit d'avoir le même niveau de risque sur chaque actif du portefeuille malgré des différences de poids pour chaque actif.

Pour appréhender la diversification optimale de portefeuille, la compréhension quantitative du problème permet d'évaluer l'importance de tout changement de poids, d'allocation. Pour obtenir cette allocation dynamique, la résolution du programme d'optimisation à différents instants t au cours du temps peut être résumée à un objectif : la minimisation de la variance des rendements d'un portefeuille (sous contrainte(s) ou non), ce qui revient à rechercher le vecteur de poids pour lequel la variance du portefeuille sera minimum⁷⁰. Cette procédure permet de rechercher les poids qui feront que la variance du portefeuille sera minimisée. Il s'agit de trouver le vecteur de poids X_{optim} composé d'autant de poids que d'actifs du portefeuille :

$$X^{optim} = \operatorname{argmin}_{X \in C} X'VX$$

avec X le vecteur de poids des actifs, X' le vecteur transposé de X , V la matrice de variances covariances des rendements des actifs et C un ensemble de contraintes. L'écriture mathématique d'une fonction à minimiser ainsi que l'écriture matricielle d'un programme d'optimisation sous contraintes est évoquée dans le chapitre 5.

4.2. Analyse comparative : diversification « naïve » versus diversification optimale

Nous cherchons à comprendre quels sont les apports et les limites de chacun des types de diversification. Nous débutons par la présentation des principaux apports de chacune des allocations puis nous soulevons les limites de ces dernières. Notre analyse comparative favorise la compréhension des environnements respectifs qui sont notamment ceux des tests

⁶⁸ Sauf dans le cas d'un portefeuille d'arbitrage « euro neutre ».

⁶⁹ En euros par exemple.

⁷⁰ Dans d'autres cas, nous pourrions procéder à d'autres formes d'optimisation comme lorsque nous souhaitons maximiser le rendement d'un portefeuille sous contrainte(s).

empiriques des chapitre 4 et 5.

La diversification « naïve » (« 1/N ») permet principalement d'être accessible à un large public d'investisseurs ne maîtrisant pas obligatoirement les subtilités de l'ingénierie financière. Ceci se traduit par la « facilité » d'exécution et de compréhension par rapport à une diversification optimale. La facilité de développement des stratégies « naïves » basées sur des « règles simples d'allocation »⁷¹ repose sur le fait que nous n'estimons pas les moments des rendements du portefeuille et qu'aucune optimisation n'est calculée (DeMiguel, Garlappi, et Uppal, 2009 ; Huberman et Jiang, 2006 ; Benartzi et Thaler, 2001). Les premiers auteurs rappellent que certains praticiens continuent d'investir sur des portefeuilles « 1/N » pour répartir leurs capitaux parmi différents actifs plutôt que de bénéficier du développement de méthodes avancées d'optimisation. Certains investisseurs ont des lacunes en la matière liées à « la lourdeur des calculs supplémentaires ...car ils ont besoin de trouver des solutions à travers un large éventail de scénarios » (Maillard, Roncalli, et Teleitche, 2010).

Les minimisations ou maximisations de fonction font appel à des notions de mathématiques que tout public ne maîtrise pas. La diversification « naïve », peut malgré sa « simplicité », être à l'origine de performances supérieures que celles obtenues en présence de diversification optimale. Cette surperformance est caractérisée par une supériorité du ratio de Sharpe et du rendement équivalent certain pour une allocation « naïve » (DeMiguel, Garlappi, et Uppal, 2009). Les auteurs expliquent que les stratégies optimales sont en général moins coûteuses (que les stratégies « naïves ») car moins de titres sont sélectionnés pour composer le portefeuille. Pour les portefeuille « 1/N », il faut un nombre important de titres (d'actifs) pour assurer un effet de « dilution » du risque (Chouiefaty et Coignard, 2008).

Les portefeuilles optimaux offrent également la possibilité d'innovations financières en gestion du risque. La gestion du risque peut être améliorée en prenant en compte la volatilité des rendements. La volatilité liée à un investissement peut être quantitativement prise en compte dans l'allocation de portefeuille; ce qui n'est pas le cas pour un portefeuille équipondéré où chaque titre a le même poids, ces derniers n'étant pas choisis en fonction de leur volatilité. Des portefeuilles issus d'une diversification optimale permettent d'ajuster la contribution de chaque actif au risque total du portefeuille. Si initialement les contributions des actifs au risque total du portefeuille ne sont pas équivalentes, l'optimisation tente de les rendre proches les unes par rapport aux autres.

Maillard et al. (2010) redéveloppent une approche qui permet de rendre égales les contributions de chaque actif au risque total du portefeuille. Ce principe est autrement nommé « *equally-weighted risk contributions* » ou « *ERC* ». L'univers optimal permet une gestion du

⁷¹ (DeMiguel, Garlappi, et Uppal, 2009).

risque où le « budget risque » est réparti également sur tous les actifs du portefeuille. Les portefeuilles « *ERC* », autrement appelés « *Risky Parity portfolios* », permettent de limiter la surexposition au risque d'une classe d'actifs et de fournir une large exposition à toutes les classes d'actifs en éliminant toute allocation ne fournissant pas de diversification supplémentaire en terme de risque (Qian, 2010)⁷².

La souplesse de spécification et de paramétrage que propose la diversification optimale est plus grande que celle des portefeuilles « naïfs ». La prise en compte de la volatilité est réalisée à travers une procédure d'optimisation, avec ou sans contraintes, ce qui laisse à chaque gérant l'opportunité de préciser ses propres restrictions (contraintes réglementaires, contraintes de signe des positions)⁷³. La flexibilité tient aussi au fait qu'il est envisageable de calculer des programmes d'optimisation sans considérer les rendements espérés d'un portefeuille tel que dans le cadre « *minimum variance* » (Clarke, Silva, et Thorley, 2006), où seule la minimisation de la volatilité est recherchée sans problématique d'estimation des rendements futurs (Viviani, 2001). D'autres recherches (Maillard, Roncalli, et Teleitche, 2010 ; Choueifaty et Coignard, 2008) appliquent certes des programmes d'optimisation différents, mais partagent cette volonté d'indépendance aux rendements estimés.

Concluons cette analyse comparative par les principales limites de la diversification « naïve » et de la diversification optimale. La première est plus coûteuse car le nombre de titres nécessaire pour diversifier le risque du portefeuille est plus important⁷⁴. Les innovations financières de construction de portefeuille, dans le cadre « 1/N », sont plus limitées que dans le cadre optimal : l'optimisation est un champ de recherche en constante évolution⁷⁵. La diversification du risque est théoriquement moins pertinente en diversification « naïve ».

Une première limite d'une diversification optimale est la complexité de certains programmes d'optimisation. Maillard et al. (2010) évoquent le « fardeau supplémentaire de calcul qui est imposé » pour faire référence à cette barrière à l'entrée. Black et Litterman (1992) mènent un travail sur des modèles quantitatifs d'allocation d'actifs et supposent que « leurs utilisateurs les ont trouvés d'utilisation difficile ». Les mêmes auteurs remettent en

⁷² D'autre part, les contributions au risque d'un portefeuille peuvent expliquer la contribution de chaque position (poids) aux pertes *expost* du portefeuille (Qian, 2010).

⁷³ L'objectif de l'optimisation peut être atteint en intégrant des contraintes qui peuvent différer selon les problématiques de gestion. Nous pouvons par exemple minimiser la volatilité d'un portefeuille en imposant des poids uniquement positifs soit *long only* ou respecter que la somme des poids n'excède pas 100% (Choueifaty et Coignard, 2008 ; Clarke, Silva, et Thorley, 2006). Un autre apport est de limiter le poids d'un actif au sein d'un portefeuille (Clarke, Silva, et Thorley, 2006).

⁷⁴ (DeMiguel, Garlappi, et Uppal, 2009).

⁷⁵ Notons cependant qu'une grande proportion de praticiens favorisent des solutions « heuristiques » plus simple à calculer et supposées « robustes » car elles ne dépendent pas de l'estimation des rendements futurs (Maillard, Roncalli, et Teleitche, 2010).

cause l'utilisation d'un modèle quantitatif optimal pouvant devenir une « *black box* ».

Un enjeu théorique et empirique de la diversification optimale⁷⁶ est la difficulté d'estimer les rendements futurs et la sensibilité des procédures d'optimisation à l'estimation des rendements futurs⁷⁷. La sensibilité des paramètres d'un modèle en optimisation, notamment aux espérances de rendements, entraîne des changements considérables dans le choix des poids d'un portefeuille. Cette sensibilité sur la composition du portefeuille aux erreurs d'estimations est d'autant plus forte que les programmes d'optimisation « moyenne-variance »⁷⁸ maximisent les erreurs d'estimation (Michaud, 1989). Il convient de vérifier que le modèle soit prédictif et que les contraintes « contiennent l'optimiseur » puisque les approches « moyenne-variance » sont fortement sensibles à certains paramètres (Stockton et Shtekhman, 2010). Ces programmes augmentent la pondération des titres à forts rendements estimés et diminuent les poids de ceux à rendements estimés plus faibles sur lesquels portent le plus fort risque d'estimation (Viviani, 2001)⁷⁹. Ce dernier conclut que de nombreuses études mettent en garde sur la gestion « moyenne-variance » : « l'effort de prévision doit essentiellement porter sur les espérances de rendements ; en présence de risque d'estimation les ventes à découvert sont à proscrire ; plus le risque d'estimation est important plus il faut s'orienter vers des portefeuilles équipondérés »⁸⁰.

Des concentrations de poids dans le cadre d'allocation optimale peuvent également se produire entraînant des effets contraires à la diversification du risque souhaitée (Maillard, Roncalli, et Teleitche, 2010 ; Clarke, Silva, et Thorley, 2006).

Dans le cadre d'une diversification optimale, les propriétés des estimateurs ne sont pas toujours adaptées aux séries financières et économiques. Les estimateurs doivent prendre en compte l'hétéroscédasticité (Merton, 1980), la non-stationnarité étant observée pour de nombreuses séries financières et économiques⁸¹. DeMiguel et al. (2009) estiment que dans le cas de modèle dépendant d'estimateurs biaisés, le principe de diversification optimale est remis en cause. Cette même conclusion est effectuée avec des estimateurs bayésiens dans le cadre de leur étude.

⁷⁶ Quel que soit le cadre de diversification optimal dont celui « moyenne variance » (Markowitz, 1956 ; Markowitz, 1952).

⁷⁷ Ce sont les deux analyses de Black et Litterman (1992).

⁷⁸ (Markowitz, 1956 ; Markowitz, 1952).

⁷⁹ Ces résultats sont tempérés par l'ajout d'une contrainte de « non-négativité » au problème d'optimisation (Best et Grauer, 1991).

⁸⁰ Le choix d'un portefeuille équipondéré devient une « solution extrême » pour limiter les risques d'estimation sur le portefeuille (Viviani, 2001).

⁸¹ Voir la section précédente du chapitre 1.

5. Outils quantitatifs de la recherche

Nous introduisons plusieurs techniques quantitatives qui sont reprises dans le cadre des tests empiriques (chapitres 4 et 5). Ces outils nous aideront à estimer l'évolution probable des actifs financiers d'un univers multi-classes d'un jour sur l'autre. Nous présentons deux types de d'outils quantitatifs originellement dédiés à la compression d'un signal tels que l'analyse en composantes principales (ACP) et la quantification vectorielle.

La première technique est une analyse factorielle historique des données (ACP), fréquemment utilisée en arbitrage statistique afin d'isoler les facteurs de risque principaux (Avellaneda et Lee, 2010 ; Plerou et al., 1999 ; Laloux et al., 2000).

Le deuxième outil quantitatif permet d'isoler des modèles non-aléatoires à partir d'un phénomène aléatoire autrement appelés « *codebook* » en terminologie de la théorie du signal (Gersho et Gray, 1992). Cette deuxième famille de techniques permettra de quantifier les facteurs résiduels (risque spécifique) pour évaluer les actifs⁸² et déterminer leur sur ou sous-évaluation par rapport à un niveau moyen, censé représenter le niveau d'équilibre du marché.

5.1. Analyse factorielle des données historiques : analyse en composantes principales

Nous recourons à l'analyse factorielle pour distinguer statistiquement les facteurs principaux des facteurs résiduels. Le principe est de pouvoir isoler un signal résiduel afin de l'utiliser dans le deuxième type d'outils quantitatifs de quantification. Parmi les techniques d'analyse factorielle, nous choisissons l'analyse en composantes principales (ACP) qui a été initiée par deux travaux principaux (Spearman, 1904 ; Pearson, 1901) au début du vingtième siècle puis complétée par d'autres travaux plus tardivement (Jolliffe, 1973 ; Hotelling, 1933).

L'introduction et le développement de l'informatique en statistique ont permis de traiter un nombre important de données et d'accompagner la rapide exécution de ces procédures. L'ACP est un pan de l'analyse de données, elle-même un sous-champ de l'analyse statistique multivariée. Concrètement, l'ACP permet de résumer l'information contenue dans une table de données en compressant le nombre de variables initiales en un nombre réduit de variables (Evrard, Pras, et Roux, 2009). Les champs d'application de l'ACP sont divers. Elle est utilisée en Sciences Economiques, en Sciences de Gestion, en Sciences Sociales, en Biologie et en Médecine. Bien que cette technique d'analyse soit relativement ancienne, de nombreuses recherches empiriques continuent d'être menées à l'aide d'ACP dans divers champs de recherche. Notons que l'ACP est une analyse de données répandue en

⁸² On parlera également de *pricing* pour l'évaluation de prix d'actifs financiers.

Finance⁸³, en Economie⁸⁴ ou en Sociologie⁸⁵. L'objectif principal de l'ACP est de tirer avantage des corrélations entre les composants d'un vecteur aléatoire⁸⁶ afin d'engendrer des composants (rendements d'actifs par exemple) décorrelés dans une nouvelle base (vectorielle) constituées de combinaisons linéaires des directions initiales (de la base canonique).

Chaque composant X_i peut s'écrire sous la forme d'une somme pondérée de facteurs décorrelés deux à deux. Un nombre fini de facteurs, ceux qui expliquent une « majeure » partie de la variance, sont souvent interprétés comme ceux qui ont un sens économique et sont assimilés aux facteurs de risques principaux du marché (risque systématique d'un portefeuille d'actifs). Par exemple, une ACP sur les composants d'un indice actions aura souvent comme premier facteur de risque l'indice lui-même (Avellaneda et Lee, 2010 ; Laloux et al., 2000 ; Plerou et al., 1999). Les facteurs de risque suivants de l'ACP peuvent être décorrelés du premier, ce qui explique pourquoi l'ACP est régulièrement utilisée en arbitrage statistique pour générer des portefeuilles « marché-neutre »⁸⁷ au sein de stratégies contrariantes. En conséquence, les portefeuilles créés évolueront de manière orthogonale à ce qui est perçu comme l'évolution du marché.

Dans notre recherche, l'ACP aura l'utilité d'un outil numérique pour construire des *proxies*⁸⁸ des différents facteurs entrant dans la décomposition du rendement d'un actif de l'équation. Dans un premier temps, nous isolerons les facteurs (de risque) dit principaux F_j , $1 \leq j \leq m$ parmi l'ensemble des informations à disposition puis nous construirons des portefeuilles d'arbitrage⁸⁹.

Soit d un nombre de composants (d'actifs financiers) et T la taille d'un échantillon en termes de temps. Considérons un historique de plusieurs échantillons $\{X_{i,t}\}$ où $1 \leq t \leq T$ et $1 \leq i \leq d$ où chaque échantillon $X_i := \{X_{i,t}\}$, $1 \leq t \leq T$ représente par exemple un composant (actif) financier ou économique. $R_{i,t}$ est le taux de variation quotidien dans cet échantillon :

$$R_{i,t} = \frac{X_{i,t} - X_{i,t-1}}{X_{i,t-1}}, \quad 1 \leq t \leq T, \quad 1 \leq i \leq d.$$

⁸³ Pour le calcul du H-Score ou le calcul de *Value-at-Risk (VaR)* en Finance de marché.

⁸⁴ Etudes de conjoncture.

⁸⁵ Sondage utilisant des échelles de Likert.

⁸⁶ L'historique d'une série de prix d'un actif peut être contenu dans un vecteur aléatoire.

⁸⁷ Nous n'exploiterons pas le cadre « marché neutre » mais nous servirons de ce cadre comme point de départ de notre cadre théorique dans le chapitre 2.

⁸⁸ Le terme *proxies* signifie des mesures équivalentes.

⁸⁹ Cette étape sera détaillée dans le chapitre 2.

\bar{R}_i (respectivement $\bar{\sigma}_i$) est la moyenne de l'échantillon (respectivement l'écart-type) des rendements de l'échantillon $X_i := \{X_{i,t}\}, 1 \leq t \leq T$.

Nous pouvons réécrire $R_{i,t}$:

$$R_{i,t} = \bar{R}_i + \bar{\sigma}_i Y_{i,t}$$

Nous rappelons que :

$$\bar{R}_i = \frac{1}{T} \sum_{t=1}^T R_{i,t} \text{ et}$$

$$\bar{\sigma}_i^2 = \text{Var}(R_i) := \frac{1}{T-1} \sum_{t=1}^T (R_{i,t} - \bar{R}_i)^2 .$$

Dans cette présentation, l'ACP est effectuée sur des rendements normalisés (Plerou et al., 1999 ; Greene, 1997). Nous introduisons le vecteur Y en dimensions d (actifs) dont les composants dans la base canonique $\{e_i\}, 1 \leq i \leq d$ sont les échantillons

$$Y_i := \{Y_{i,t}\}, 1 \leq i \leq d :$$

$$Y = \sum_{i=1}^d Y_i e_i$$

L'ACP cherche à définir une base vectorielle de l'espace R^d tel que les composants de Y soient décorrélés.

Nous cherchons une base $\{p_j\}, 1 \leq j \leq d$ de l'espace d en dimension R^d telle que :

$$Y = \sum_{j=1}^d Z_j p_j$$

$$\sum_{i=1}^d Z_{i,t} Z_{j,t} = 0 \text{ pour } i \neq j.$$

Une telle base constitue une famille de vecteurs propres de la matrice empirique de variance covariance C ,

$$C_{i,j} = \frac{1}{T-1} \sum_{t=1}^T Y_{i,t} Y_{j,t}.$$

Matriciellement, cela revient à écrire :

$$C = \frac{1}{T-1} \sum_{t=1}^T YY'$$

avec Y' la transposée de Y .

Il existe des valeurs propres réelles non-négatives λ_i , $1 \leq i \leq d$ de la matrice C telles que

$$Cp_j = \lambda_j p_j, \quad p_i' p_j = \delta_{i,j}, \quad \lambda_j = \text{Var}(Z_j).$$

De plus, il est supposé que les valeurs propres λ_i sont triées en ordre décroissant.

Soit $p_{i,j} = p_j' e_i$. Alors $PP' = P'P = I$ et $Z = P'Y$ impliquant $Y = PZ$.

Finalement, nous posons

$$F_{j,t} = \sum_{i=1}^d \frac{p_{i,j}}{\sigma_i} R_{i,t}.$$

À l'inverse, nous avons

$$R_{i,t} = \sum_{j=1}^d p_{i,j} \sigma_i F_{j,t} = \sum_{j=1}^d \beta_{i,j} F_{j,t}$$

où nous avons posé $\beta_{i,j} = \sigma_i p_{i,j}$.

Les taux de variation des données sont segmentés en plusieurs facteurs F_j dont la matrice de covariances peut être aisément calculée :

$$\text{Cov}(F_i F_j) = \frac{1}{T-1} \sum_{t=1}^T (F_{i,t} - \bar{F}_i)(F_{j,t} - \bar{F}_j)$$

$\text{Cov}(F_i F_j) = \lambda_i$ si $i = j$, 0 sinon.

Ainsi, les variances de ces facteurs décorrelés sont donnés par les λ_i .

Les facteurs F_j peuvent être lus comme des *proxies* numériques des facteurs de risque entrant

dans la décomposition du rendement d'un actif risqué. Chaque facteur F_j peut s'interpréter comme le rendement d'un portefeuille dont les poids sont proportionnels aux vecteurs propres P_j .

En arbitrage statistique, il est courant de distinguer les facteurs principaux des facteurs non-principaux (résiduels). Nous devons donc à présent estimer le paramètre m , qui nous donne le seuil qui sépare les facteurs principaux (risque systématique) des facteurs résiduels (risque spécifique). Ce seuil, m , peut être déterminé tel que :

$$m = \inf \left\{ k \geq 1, \sum_{i=1}^k \lambda_i \geq c \sum_{i=1}^d \lambda_i \right\},$$

pour un nombre $C \in [0 ; 1]$ correspondant à un seuil donné.

Des procédures optimales permettent de trouver le nombre de valeurs propres à retenir. La procédure de Cattell « *scree test* » (Evrard, Pras, et Roux, 2009 ; Cattell, 1966) permet de représenter, dans un plan à deux dimensions, en abscisses les facteurs et en ordonnées le % de variance expliquée. Cet algorithme cherche le changement de concavité dans la courbe appartenant au plan, ce qui est équivalent à chercher un point d'inflexion. Si en ajoutant une composante (facteur) le taux de variance expliqué n'augmente pas, alors il n'est pas nécessaire d'ajouter un facteur supplémentaire. Les facteurs supplémentaires ne sont pas retenus si l'efficacité marginale devient négative. Selon Evrard et al (2009) : « la variance restituée par chaque facteur va en diminuant. La règle d'arrêt consiste à chercher quel est - dans une analyse marginale - le premier facteur dont l'élimination conduit à une perte d'information minimum. Pour cela il suffit de porter un graphique (« *scree test* » de Cattell ou test du coude), en abscisse le numéro des axes factoriels, en ordonnée le pourcentage d'inertie qu'il(s) restitue(nt) et d'éliminer les facteurs situés après le changement de concavité de la courbe (c'est-à-dire après le point d'inflexion). La recherche du point d'inflexion peut être systématisée en examinant le signe des écarts entre les différences de valeurs propres consécutives; le point d'inflexion correspond à un changement de ce signe. Si on note $\delta_k = \lambda_k - \lambda_{k+1}$ et $\epsilon_k = \delta_k - \delta_{k+1}$ on retiendra $k+1$ facteurs, k étant le premier indice pour lequel ϵ_k est négatif ».

D'autres procédures existent comme celles faisant intervenir des résultats théoriques sur les matrices aléatoires (Laloux et al., 2000 ; Plerou et al., 1999). Il s'agit par exemple de la technique dite de *density of states (DOS)* appliquée à la classe des indices actions. On remarque que la technique *DOS* a tendance à sous-estimer les facteurs (de risques) principaux (Avellaneda et Lee, 2010).

Afin de compléter notre présentation de l'analyse en composantes principale appliquée à la gestion de portefeuille, nous mettons en exergue la définition du risque résiduel (non-principal, spécifique) $\varepsilon_{i,t}$:

$$R_{i,t} = \sum_{j=1}^d \beta_{i,j} F_{j,t} + \varepsilon_{i,t}$$

où

$$\varepsilon_{i,t} = \sum_{j=m+1}^d \beta_{i,j} F_{j,t}, \quad \beta_{i,j} = \sigma_i p_{i,j}.$$

Ces deux dernières équations permettent de réaliser une approximation numérique des facteurs donnés par la théorie. Nous reprendrons ces définitions du risque résiduel dans le chapitre 2.

5.2. Algorithmes de quantification optimale : processus de compression d'un signal

Nous présentons à présent une technique qui, bien qu'elle soit également une technique de compression de signal, est assimilable à une procédure de classification. Elle permet de regrouper un ensemble d'individus en classes homogènes. Il s'agit de quantification vectorielle de variables aléatoires. A l'aide de cette méthode de quantification, nous obtenons des grilles de probabilité résumant le comportement des actifs. Cette méthode nous amène, dans le chapitre 2, à construire un modèle empirique non-paramétrique reposant sur le concept de retour à la moyenne. Une des hypothèses principales est que les prix des actifs cotés suivent une dynamique markovienne : le futur s'écrivant à l'aide du passé très récent⁹⁰.

La quantification est une méthode visant à discrétiser un signal aléatoire. La théorie de la quantification vectorielle optimale trouve ses origines dans les années 1950 et a été découverte pour discrétiser l'émission de signaux (Graf et Luschgy, 2000 ; Gersho et Gray, 1992). Cette méthode consiste à obtenir la meilleure approximation d'un vecteur aléatoire par un autre vecteur prenant un nombre fini de valeurs.

⁹⁰ Pour plus de détails sur les chaînes de Markov voir Bremaud (2010).

Prenons un vecteur aléatoire $X : \Omega \rightarrow R^d$, une norme $\|\cdot\|$ dans R^d (norme euclidienne) et un N -uplets $x = (x^1, x^2, \dots, x^N)$ de points de R^d , nous définissons le N -quantificateur de Voronoï de X par

$$\widehat{X}^x = \sum_{k=1}^N x^k 1_{C_k(x)}(X)$$

où $C_k(x)$ indique la $k^{\text{ième}}$ cellule de Voronoï du N -uplet x , soit

$$C_k(x) = \{y \in R^d, \|y - x^k\| < \|y - x^i\|, i \neq k\},$$

et où 1_A est la fonction indicatrice de l'ensemble A , c'est-à-dire prenant des valeurs dans $\{0,1\}$ de façon que $1_A(z) = 1$ si $z \in A$ et 0 sinon. L'ensemble des cellules de Voronoï du N -uplet est appelé diagramme de Voronoï.

Soit P_x la distribution de X définie comme $P_x(A) = P(X \in A)$ pour tout borélien $A \subset R^d$. La distribution de \widehat{X}^x en tant que vecteur aléatoire est donnée par le N -uplet $\{p_k\}$, $1 \leq k \leq N$. où

$$p_k = P_x(C_k(x)) = P(\widehat{X}^x = x^k) = P(X \in C_k(x)), 1 \leq k \leq N$$

Pour un N -uplet x donné, il est relativement aisé de voir que, par définition de la projection au plus proche voisin, l'erreur obtenue quand $X(\omega)$ est remplacé par $\widehat{X}^x(\omega)$ est

$$\|X(\omega) - \widehat{X}^x(\omega)\| = \min_{1 \leq k \leq N} \|X(\omega) - x^k\|.$$

La distorsion du N -uplet x par rapport à X est obtenue en prenant la moyenne quadratique de l'erreur « trajectorielle » ci-dessus, soit :

$$D_N^X(x) = E\|X - \widehat{X}^x\|^2 = E \min_{1 \leq k \leq N} \|X - x^k\|^2$$

Pour un vecteur aléatoire X donné ou une loi équivalente sur P_x et étant donné un entier N , nous pouvons chercher maintenant à optimiser par rapport aux N -uplet x l'erreur de distorsion précédente, c'est-à-dire chercher les distorsions les plus faibles. Le lecteur intéressé peut se reporter aux travaux de référence sur la quantification (Pagès et Printems, 2003 ;

Pagès et Printems, 2008) et leurs références associées.

Soit une loi P_X sur R^d et un entier N . Notre objectif est de minimiser la fonction à $N * d$ variables suivante :

$$x \mapsto D_N^X(x) = \int_{R^d} \min_{1 \leq k \leq N} \|u - x^k\|^2 P_X(du) = \sum_{k=1}^N \int_{C_k(x)} \|u - x^k\|^2 P_X(du)$$

où x est dans $(R^d)^N$. Nous sommes amenés à résoudre un problème d'optimisation non-quadratique⁹¹.

Premièrement, notons que la condition nécessaire d'Euler pour un optimum implique que l'optimum x^* doit être stationnaire, c'est-à-dire qu'il doit satisfaire pour tout $1 \leq k \leq N$:

$$\int_{C_k(x^*)} u P_X(du) = p_k x^{*,k} \text{ ou } E(X|\hat{X}) = X.$$

L'équation ci-dessous signifie que x^* doit être le barycentre de sa propre cellule de Voronoï quant à sa mesure P_X . Cette déclaration est le point de départ de la première méthode que nous allons introduire : l'algorithme de Lloyd.

L'algorithme de Lloyd

Soit $x = (x_0^1, x_0^2, \dots, x_0^N)$ un N -uplet initial de points dans R^d . L'idée principale consiste à voir x^* comme des points fixes de fonctions contractantes.

Pour chaque $n \geq 0$, nous calculons les barycentres suivant :

$$x_{n+1}^k = \frac{1}{P(X \in C_k(x))} = \int_{C_k(x^*)} u P_X(du)$$

pour tout $1 \leq k \leq N$ (de façon explicite en dimension 1 ou bien par une méthode de Monte Carlo en dimension plus grande).

Il peut être démontré que l'application $x_n \mapsto f(x_n) = x_{n+1}$ est strictement contractante sur $(R^d)^N$ et donc la suite converge vers x^* . Cette procédure est répandue car elle ne requiert le « calage » d'aucun paramètre.

⁹¹ Malgré les apparences, il ne s'agit pas d'un programme quadratique .

L'algorithme de Kohonen

En revenant sur le calcul de la distorsion $D_N^X(x)$, nous utilisons une méthode de descente de gradient : nous partons d'un point initial N -uplet $x_0 \in (R^d)^N$, nous construisons une séquence de N -uplet $\{x_n\}, n \geq 0$ telle que

$$x_{n+1} = x_n - \gamma_{n+1} \nabla D_N^X(x_n).$$

Le gradient de $D_N^X(x)$ peut être calculé autrement (Pagès, 1998) et nous obtenons :

$$\frac{\partial D_N^X(x)}{\partial x^k} = 2 \int_{C_k(x^*)} (u - x^k) P_X(du), \quad 1 \leq k \leq N.$$

Le principal inconvénient de ce type de méthode est qu'il est délicat en général de calculer exactement les intégrales dans l'équation en dimension supérieure à 1.

L'algorithme de Kohonen fait référence à un algorithme de descente de gradient stochastique. Une conséquence de l'équation précédente est que le gradient de $D_N^X(x)$ a une représentation stochastique :

$$\frac{\partial D_N^X(x)}{\partial x^k} = 2E((X - x^k) 1_{C_k(x)}(X)).$$

Ainsi, nous remplaçons l'expression x_{n+1} par l'équation suivante :

$$X_{n+1} = X_n - \gamma_{n+1} \sum_{k=1}^N (X_n^k - \xi_{n+1}) 1_{C_k(X_n)}(\xi_{n+1}).$$

où $\{\xi_n\}$ est une suite de vecteurs aléatoires identiquement et indépendamment distribués (i.i.d) selon la distribution de X . C'est un exemple d'algorithme stochastique qui repose sur la possibilité de générer un grand nombre de copies indépendantes de variables aléatoires. Dans les chapitres suivants, nous utilisons l'historique des données de l'échantillon⁹² pour effectuer ces simulations sur le passé.

⁹² Voir le chapitre 3.

Dans ce premier chapitre, nous synthétisons et analysons les concepts théoriques clés nécessaires à la compréhension de notre recherche. Cette étape nous permet de recenser les travaux de référence dont nous nous inspirons pour développer notre recherche. C'est également à partir de ces travaux que nous identifions les voies d'investigation peu ou faiblement empruntées ; nous souhaitons participer aux débats soulevés, sur lesquels il ne semble pas y avoir de consensus scientifique. Nous définissons statistiquement le type de données que nous considérons pour nos travaux, puis nous notons les éléments de complémentarité entre des données hétérogènes. Nous analysons la pertinence de développer notre réflexion dans un univers multi-classes d'actifs pour identifier des contextes d'arbitrage potentiels. Nous soulignons l'intérêt particulier des données hétérogènes en univers multi-classes, notamment du fait des différences de réactions que peuvent avoir les actifs lors de publications d'informations économiques et extra-financières. Nous introduisons les principaux aspects de l'arbitrage statistique, découvrons l'intérêt de développer ces stratégies en univers multi-classes et notons les limites de ce pan de la gestion quantitative. Dans le contexte d'allocation tactique, qui est le nôtre, nous présentons et comparons deux types de diversifications de portefeuille qui seront utilisés lors des tests empiriques menés dans les chapitres 4 et 5. Enfin, ce chapitre présente les outils quantitatifs de compression de signal que nous utiliserons dans le chapitre 2, en vue de proposer un modèle empirique d'arbitrage systématique.

CHAPITRE 2. MÉTHODES D'ARBITRAGE SYSTÉMATIQUE ET PRISE EN COMPTE DE DONNÉES HÉTÉROGÈNES

Dans ce deuxième chapitre, nous souhaitons introduire le modèle empirique qui sera utilisé pour nos tests sur des données passées. La spécification de ce modèle empirique est le fruit de différentes recherches. Nous étudions deux méthodes d'arbitrage systématique inspirées par différents travaux théoriques et empiriques. Celles-ci sont complétées par le cadre innovant de la quantification dans un univers multi-classes d'actifs en présence de données hétérogènes. Une innovation de nos travaux est de formaliser numériquement la combinaison de données hétérogènes au sein de stratégies d'arbitrage. Avant toute chose, signalons que nos méthodes d'arbitrage ont un caractère générique dans le sens où elles sont applicables aussi bien à l'univers multi-classes d'actifs qu'à un univers mono-classe.

Avant de détailler ces deux modélisations, il nous semble essentiel d'expliquer pourquoi nous avons choisi une approche systématique lors des recherches plutôt qu'une approche discrétionnaire. Dans l'Introduction Générale, nous expliquons comment nous réduisons le biais discrétionnaire de notre approche, en intégrant des paramètres chiffrés pour l'ensemble des données relatives aux « autres informations de marché ».

Nous soulignons certains avantages d'une gestion systématique par rapport à une gestion discrétionnaire pour davantage expliquer notre choix de l'approche systématique. Par la suite, nous exposerons deux méthodes d'arbitrage alliant détermination du risque spécifique et détection de signaux d'investissement. Notre première méthode est une approche paramétrique alors que la seconde aura la caractéristique d'être non-paramétrique. Pour conclure, nous menons une analyse comparative des deux approches afin d'aboutir au choix de celle qui est employée pour l'étude empirique du chapitre 3. La mise en comparaison de ces approches permet d'introduire les apports et les limites de chacune : ce travail d'analyse nous oriente vers le choix d'une voie de recherche non-paramétrique.

Nous précisons en amont de tous développements, que notre approche de l'arbitrage s'inspire initialement de la logique du *long short equity* sur plusieurs classes d'actifs distinctes avec la possibilité de vente à découvert⁹³. Bien que le schéma du *long short equity* soit théoriquement applicable à un univers multi-classes, en pratique les classes d'actifs ne suivent pas toujours toutes les mêmes règles de fonctionnement ni les mêmes processus de prix comme cela a été rappelé dans le chapitre 1.

⁹³ Nous verrons en détail, dans le chapitre 3, les classes d'actifs impliquées lors du processus de collecte des données.

Nous choisissons de construire une méthode d'arbitrage systématique fondée sur une allocation d'actifs dynamique. L'aspect dynamique de notre approche consiste à utiliser les probabilités pour bâtir une allocation de portefeuille qui peut évoluer dans le temps.

1. Prémices d'une approche systématique

Nous mentionnons les principaux éléments de l'orientation systématique de nos recherches. Nous proposons une analyse comparative des deux principales approches en gestion puis nous expliquons comment le calcul du risque spécifique des portefeuilles d'arbitrage est effectué de manière systématique.

1.1. Approche systématique *versus* approche discrétionnaire en univers multi-classes d'actifs

Comparer deux styles historiques de gestion, systématique et discrétionnaire, permet de préciser les raisons qui nous conduisent à choisir une approche systématique permettant d'intégrer des informations « fondamentales ». Cette discussion permet de comprendre synthétiquement les principaux éléments de comparaisons de ces styles de gestion. Comme dans le chapitre 1, nous proposons d'explorer une stratégie multi-classes d'actifs « historique » : la stratégie *global macro*. A titre d'exemple, nous analysons et comparons chacune des deux approches de gestion (systématique et discrétionnaire) au sein de cette même stratégie. Aux antipodes l'une de l'autre, nous disposons de deux catégories de stratégies multi-classes d'actifs, de type stratégie *global macro* : le *global macro* systématique et le *global macro* discrétionnaire. Nous prendrons pour exemple la stratégie *global macro* car elle matérialise les aspects internationaux, multi-classes et économiques du cadre de notre recherche⁹⁴. La comparaison de cette stratégie multi-classes sous l'aspect discrétionnaire et systématique peut toutefois être conceptuellement généralisée à toutes les autres stratégies multi-classes. Nous pouvons décliner les éléments de discussion qui suivront à toutes les stratégies multi-classes d'actifs.

Nous rappelons qu'il existe également des stratégies « hybrides » (Fabozzi, Focardi, et Jonas, 2008) comportant une partie discrétionnaire et une partie quantitative. Chacune des stratégies, discrétionnaire ou quantitative, peut intégrer une part de l'autre (Gregory-Allen, Stangl, et Shawky, 2009). Ainsi, au sein de la communauté des gérants « quantitatifs », certains utilisent également des analyses « fondamentales » de ratios financiers. Dans le cas

⁹⁴ Voir le chapitre 1.

d'une gestion discrétionnaire, une grande liberté d'appréciation est laissée aux gestionnaires. Dans un contexte systématique, les décisions d'investissement sont prises par l'intermédiaire de règles mécaniques avec (quasiment) aucune intervention humaine sauf pour le paramétrage des modèles.

Le *global macro* systématique recouvre plusieurs types de gestion. Cette « hétérogénéité » est évoquée dans certaines recherches. Dans certains cas, la mise en œuvre de la stratégie multi-classes de type *global macro*, peut être exclusivement discrétionnaire : « l'analyse fondamentale du contexte économique global repose sur la perception propre du gestionnaire. Les stratégies mises en place sont donc basées sur une vue subjective de l'évolution des fondamentaux, ce qui rend cette approche d'investissement complètement discrétionnaire » (Monarcha, 2008). C'est en ce sens qu'une gestion systématique vient réduire cet espace discrétionnaire. Le même auteur insiste sur l'importance de sélectionner les données « les plus pertinentes » et sur celle de la capacité à les traiter. Ces différents courants systématiques se rejoignent en quelques points : « bien qu'il existe divers programmes de *global macro* systématiques, ils partagent beaucoup de caractéristiques communes...les décisions d'investissement sont prises par des algorithmes, basés sur des règles mécaniques fondées sur une analyse statistique et historique » (Park, Tanrikulu, et Wang, 2009). Ils soulignent que dans le cadre d'une gestion discrétionnaire « le contrôle des émotions », à des phases critiques de marchés, peut être difficile à conserver. Lors d'une gestion systématique, nous ne nous trouvons pas régulièrement confrontés à ce cas de figure.

Une critique, centrée sur la gestion discrétionnaire, tient au changement probable d'allocation tactique dans un contexte économique bruité et instable : la prise de décision devient dans cette optique un véritable défi. Nous tempérons ce point de vue car l'effet de panique peut également se produire dans le cas où des modèles divergent de leurs objectifs numériques (limites de risque, performances) conduisant parfois à vouloir « déconnecter » le modèle mis en place⁹⁵. La courbe d'expérience peut être mise en avant pour le cas discrétionnaire, évitant parfois des erreurs stratégiques de gestion. Néanmoins, l'outil systématique assure une transparence totale couplée à une règle d'investissement homogène.

L'autre avantage majeur de l'approche systématique est que le développement et la maintenance de programmes d'arbitrage systématique peuvent être effectués dans le temps par des personnes distinctes. Le risque de dépendance est quant à lui beaucoup plus élevé dans le cas discrétionnaire avec le besoin d'un « homme clé » (Park, Tanrikulu, et Wang, 2009) pour générer de la performance. L'utilisation de règles de décisions mécaniques basées

⁹⁵ Cette situation est extrême et requiert selon les structures différentes approbations hiérarchiques avant de procéder ainsi.

sur des « signaux extraits des marchés » (aussi discutables soient-elles) permet d'éviter d'être confronté à des risques « exogènes » aux marchés car dans le cas discrétionnaire, la nature atypique des risques est due au processus d'investissement qui repose sur l'analyse d'informations fondamentales.

Notre approche tente d'éviter le risque d'interprétation discrétionnaire des données d'« autres informations marché » mais aussi de se laisser mécaniquement influencer par des données de cotation. Enfin, la capacité d'expertise d'un gérant se limite souvent à une, voire deux, classes d'actifs, surtout dans le cas discrétionnaire⁹⁶. La plupart des analystes ou gestionnaires se concentrent sur « l'allocation et la sélection de titres pour une certaine classe d'actifs » (Blitz et Vliet, 2008).

1.2. Détermination systématique d'un risque spécifique

Dans le cadre de notre modèle empirique, nous nous intéressons au risque spécifique pour différentes raisons que nous abordons au cours de cette section. C'est à partir de ce risque résiduel que sont estimés les taux de variation futurs des données homogènes (de cotation). Xu et Malkiel (2004) soulignent l'importance du risque « idiosyncratique » (spécifique) dans l'évaluation d'actifs ou *pricing* d'actifs. Les auteurs testent leurs hypothèses sur des univers actions aux Etats-Unis et au Japon. En moyenne, la gestion du risque « idiosyncratique » permet d'augmenter le rendement d'un actif.

1.2.1. Cadre historique de portefeuilles « marché neutre »

Nous présentons le cadre « marché neutre » car ce dernier est historiquement celui des travaux d'arbitrage statistique (Avellaneda et Lee, 2010). Bien que notre modèle empirique n'appartienne pas au cadre « marché neutre », de nombreuses similarités existent entre notre cadre de recherche et celui du « marché neutre » tant pour les régressions multifactorielles que pour la gestion du risque spécifique. Ce cadre sert de référence à notre modélisation et permet au lecteur d'appréhender le contexte et l'objectif de notre modèle empirique. La compréhension du concept de « marché neutre » permet également de mesurer l'importance d'une exposition au risque spécifique d'un portefeuille d'arbitrage. Nous insistons sur ce dernier point car notre approche méthodologique se concentre sur la modélisation et la gestion du risque spécifique d'un portefeuille d'arbitrage.

⁹⁶ (Park, Tanrikulu, et Wang, 2009).

Soit $\{R_{i,t}\}, 1 \leq i \leq d, t \in \Delta$ un ensemble de rendements sur un intervalle de temps Δ satisfaisant un modèle multi-facteurs de la forme

$$R_{i,t} = \alpha_i + \sum_{j=1}^m \beta_{i,j} F_{j,t} + \varepsilon_{i,t}$$

où les m facteurs $F_{j,t}$ et les résidus ε_i sont tels que $Cov(F_i F_j) = 0$ pour $i \neq j$,

$$E(\varepsilon_{i,t}) = Cov(F_j, \varepsilon_i) = 0 \text{ pour tous } i, j.$$

On peut supposer pour les besoins de l'exposé que les facteurs sont centrés, $E(F_i) = 0$. Dans ce cas, α_i représente le rendement moyen de l'actif i et $R_{i,t} - \alpha_i$ le rendement non-attendu. Les facteurs $F_j, 1 \leq j \leq m$ désignent ce que l'on appelle le risque systématique du rendement non attendu et ε_i le résidu ou risque spécifique.

La connaissance de la dynamique des résidus nous permet de détecter si un actif est sur ou sous-évalué par rapport à son « marché ». Ce raisonnement est le point de départ de l'élaboration des stratégies contrariantes et plus globalement des portefeuilles *long/short* « marché neutre ». A cet effet, nous avons un portefeuille sur lequel nous disposons de Q_i euros investis sur l'actif i et $-\sum_{i=1}^d \beta_{i,j} Q_i$ dollars sur le facteur j . Cette stratégie nous mènera en date t à un profit ou une perte de

$$\sum_{i=1}^d Q_i R_{i,t} - \sum_{j=1}^m \left(\sum_{i=1}^d Q_i \beta_{i,j} \right) F_{j,t} = \sum_{i=1}^d Q_i (\alpha_i + \varepsilon_{i,t}).$$

Le portefeuille est désormais « marché-neutre » et uniquement exposé à l'évolution des résidus (risque spécifique). Cette stratégie d'arbitrage repose donc sur les propriétés statistiques des résidus, comme la stationnarité, suggérées dans la théorie APT (Ross, 1976)⁹⁷.

La théorie de fixation des prix par arbitrage (APT) donne une assise théorique aux travaux précédents en gestion de portefeuille (Sharpe, 1964 ; Markowitz, 1956) portant sur l'analyse en moyenne et variance des portefeuilles dans un univers d'actifs risqués. La théorie de prix par arbitrage tire parti de l'hypothèse d'absence d'opportunité d'arbitrage et permet d'initier des travaux à partir d'un modèle multi-facteurs⁹⁸. Cette théorie introduit en particulier une mesure de probabilité sur l'ensemble des rendements dont l'existence implique la stationnarité des résidus.

⁹⁷ Se reporter à l'Introduction Générale pour plus de détails sur l'APT.

⁹⁸ Les coefficients ou *beta* sont proportionnels aux covariances des rendements d'actifs avec les facteurs.

Remarquons que la stationnarité est difficilement vérifiée en pratique pour les séries financières⁹⁹. D'autre part, si une telle théorie peut être validée pour une classe d'actifs, rien ne présuppose que d'un point de vue économique elle puisse être vérifiée dans un cadre multi-classes d'actifs. Rien ne nous assure qu'*a priori*, les facteurs de risque d'un marché actions soit comparables à ceux d'un marché de devises par exemple. Il est désormais envisageable d'explorer la manière d'intégrer des données hétérogènes au risque spécifique d'un portefeuille.

1.2.2. Incorporation de données hétérogènes au risque spécifique d'un portefeuille d'arbitrage

L'enjeu des données hétérogènes en univers multi-classes d'actifs est déjà compris¹⁰⁰ ainsi que l'importance de déterminer le risque spécifique de notre portefeuille pour le contexte d'arbitrage. Concentrons-nous à présent sur la manière d'agréger nos données.

Considérons dès lors un univers dont les d_q premiers composants sont des données homogènes (de cotation) et dont les derniers d_Q composants sont constitués des données relatives aux « autres informations de marché ». On désigne par $S_{i,t}$ (respectivement $I_{i,t}$) les valeurs prises par le premier type de données (respectivement prises par le second type) pour $1 \leq i \leq d_q$ (respectivement d_Q) et $t \in \Delta$. La notation $Z_{i,t}$ signifiera souvent l'ensemble des valeurs avec $1 \leq i \leq d = d_q + d_Q$.

Il existe différentes techniques pour dériver numériquement l'équation de $R_{i,t}$:

$$R_{i,t} = \alpha_i + \sum_{j=1}^m \beta_{i,j} F_{j,t} + \varepsilon_{i,t}.$$

Nous pouvons connaître les facteurs nommément¹⁰¹ (fonds indiciels cotés ou *ETF*, indices ou des portefeuilles *benchmark* représentant les facteurs de risque systématique) et nous calculons les β en utilisant des techniques standardisées de régression (multiple). Le cas échéant, nous ne connaissons pas ces facteurs et nous tentons de les extraire à partir des données à l'aide d'une analyse factorielle comme l'analyse en composantes principales (ACP). Nous choisissons cette voie de recherche en nous concentrant sur une analyse en composantes principales¹⁰².

Les valeurs des taux de variation des prix d'actifs Z sont représentés dans une matrice

⁹⁹ Voir le chapitre 1.

¹⁰⁰ Voir chapitre 1.

¹⁰¹ Dans le sens où ces facteurs correspondent un indice actions tel que par exemple le Dax pour représenter le marché des actions en Allemagne.

¹⁰² On peut consulter Chamberlain et Rothschild (1983) pour une justification de l'analyse en composantes principales dans le contexte d'un modèle multi-facteurs.

$d * T$ de R où :

$$R_{i,t} = \frac{Z_{i,t} - Z_{i,t-1}}{Z_{i,t-1}}, \quad t \in \Delta, \quad 1 \leq t \leq T$$

où Z peut être un des symboles $\{S, I\}$. Nous calculons des taux de variation pour l'ensemble des données $\{S, I\}$. Nous travaillons avec des taux de variation normalisés pour se concentrer uniquement sur les corrélations entre les taux de variation de prix d'actifs financiers et non sur leurs valeurs (prix).

Toute la procédure d'ACP est détaillée dans le chapitre 1 pour obtenir les facteurs non-principaux qui correspondent au risque résiduel (ou spécifique) de notre portefeuille d'actifs financiers. Comme cela est mentionné dans le chapitre 1, une étude numérique des valeurs propres est possible : cette dernière dévoile qu'un spectre contient quelques grandes valeurs propres détachées du reste du spectre (Avellaneda et Lee, 2010), ce qui permet en pratique de séparer les valeurs propres « significatives » associées au risque systématique de celles associées au risque spécifique. Nous choisissons de ne retenir qu'un seuil empirique communiqué par les auteurs.

Dans les faits, nous utiliserons une règle systématique pour déterminer numériquement un seuil permettant de discriminer un signal (facteurs principaux, représentant le risque systématique) du bruit (les facteurs restants ou non-principaux, représentant le risque spécifique) : nous choisirons un nombre de vecteurs propres qui correspond à un pourcentage de la variance totale de l'univers de données. Avec le nombre m , nous renseignons le nombre de valeurs propres. Pour chaque $1 \leq j \leq m$, nous considérons les « portefeuilles propres » dont les rendements sont

$$F_{j,t} = \sum_{i=1}^d \frac{p_{i,j}}{\sigma_i} R_{i,t}.$$

Cela correspond aux montants respectifs investis dans chacun des actifs¹⁰³ défini comme

$$\frac{p_{i,j}}{\sigma_i}, \quad 1 \leq i \leq d.$$

¹⁰³ Pour les actifs négociables de notre univers de données. Voir chapitre 3.

Nous vérifions que les rendements des « portefeuilles propres » sont décorrélés et que leurs variances sont égales à λ_i , $1 \leq j \leq m$.

Nous pouvons exprimer $\varepsilon_{i,t}$ tel que :

$$\varepsilon_{i,t} = \sum_{j=m+1}^d p_{i,j} (F_{j,t} - \bar{F}_j), \quad \bar{F}_j = \frac{1}{N} \sum_{t=1}^T F_{j,t},$$

$$\beta_{i,j} = \sigma_i p_{i,j},$$

$$\alpha_i = \sigma_i \sum_{j=m+1}^d p_{i,j} \bar{F}_j.$$

Numériquement, l'ACP est exécutée avec des taux de variation quotidiens (extrait des prix de clôture des données de cotation).

Notons que la procédure d'ACP que nous retenons lors des tests empiriques, permet d'intégrer ou non des données hétérogènes. L'ACP peut être mise en œuvre avec des données de cotation (données homogènes) ou une combinaison de données de cotation et d'« autres informations de marché » (données hétérogènes).

Le point commun entre ces deux ACP est que seuls les taux de variation des séries sont retenus. L'ACP sur un groupe de données hétérogènes permet d'isoler des facteurs de risque non-principaux (risque spécifique) pouvant incorporer une trace des informations économiques et du sentiment des investisseurs. Nous découvrons progressivement dans ce chapitre, comment les procédures de quantification sont appliquées au risque spécifique calculé à partir de notre ACP.

En terme de gestion, nous souhaitons vérifier que la stratégie introduite est liée *in fine* au risque spécifique (résidu) d'un portefeuille (ou d'un actif).

Illustrons notre propos.

Si

$$\sum_{i=1}^d R_{i,t} - \sum_{j=1}^m \left(\sum_{i=1}^d \beta_{i,j} \right) F_{j,t} < 0$$

alors le $i^{\text{ème}}$ actif est sous-évalué par rapport au marché représenté, le marché étant représenté par l'expression :

$$\sum_{j=1}^m \left(\sum_{i=1}^d \beta_{i,j} \right) F_{j,t}.$$

Pour tirer parti de cette sous-évaluation une position¹⁰⁴ *long* (achat) est adoptée pour l'actif i . À l'inverse, une position *short* (vente) est prise lorsque

$$\sum_{i=1}^d R_{i,t} - \sum_{j=1}^m \left(\sum_{i=1}^d \beta_{i,j} \right) F_{j,t} > 0$$

En théorie, pour demeurer strictement « marché neutre » une position de couverture doit être prise, pour couvrir les positions *long/short*

$$\sum_{i=1}^d Q_i R_{i,t}.$$

Cette opération d'achat et de vente simultanée permet de constituer un portefeuille « marché neutre » puisque la rentabilité de notre stratégie ne dépendra désormais que du risque spécifique $\varepsilon_{i,t}$.

Nous prenons une certaine distance par rapport au cadre « marché neutre » dans la mesure où aucune couverture « parfaite » ne sera effectuée dans notre approche : nous resterons « euro neutre ». C'est une distance prise par rapport à la théorie APT mais la logique générale de notre approche méthodologique reste proche : déterminer le risque spécifique à partir du risque systématique et se concentrer sur la modélisation puis la gestion du risque spécifique.

L'innovation de notre recherche tient surtout à l'estimation d'un jour sur l'autre de l'évolution de ce risque spécifique à l'aide de données homogènes et hétérogènes. Lors des tests empiriques nous analysons si, malgré l'exemption du cadre « marché neutre », nos résultats montrent que les portefeuilles évoluent avec une certaine « indépendance » au marché¹⁰⁵.

Nous estimons implicitement, de manière *ex ante*, que l'effet de couverture (*hedging*) peut provenir des ordres de vente de chaque portefeuille « euro neutre » que nous constituons à l'aide des méthodes d'arbitrage présentés dans ce chapitre¹⁰⁶. Ceci revient à dire que pour notre exemple les positions *short* pourraient remplacer la vente du marché. Dans la théorie APT, la couverture « parfaite » proviendrait d'une vente du marché.

¹⁰⁴ Par position nous signifions un ordre en bourse.

¹⁰⁵ Nous évoquons cet aspect dans les chapitres 3, 4 et 5.

¹⁰⁶ Des éléments statistiques corroborent ce point dans le chapitre 3.

Dans notre cadre de recherche, les facteurs $F_{j,t}$ ne sont pas des actifs négociables. Comme nous l'avons déjà mentionné, les facteurs $F_{j,t}$ peuvent être des indices de marché tels que le CAC 40. Dans un cadre mono-classe, il s'agit d'arbitrer une action (un titre tel que Saint Gobain par exemple) contre un indice actions (tel que le CAC 40). Comme le CAC40 n'est pas facilement négociable en l'état, sauf en achetant tous les titres de l'indice et en gérant toutes les opérations sur titres¹⁰⁷, il est commun d'investir sur un fonds indicier coté ou un portefeuille *benchmark*. Dans un cadre multi-classes d'actifs, l'indice en question est plus complexe à identifier. Nous poursuivons l'analyse de ce point à l'aide d'autres éléments dans les chapitres 3 et 4.

2. Arbitrage systématique : une approche paramétrique

Une approche paramétrique est une approche qui dépend des paramètres d'une loi de probabilité (une distribution). Les outils numériques présentés dans le chapitre 1 sont utilisés dans nos méthodes d'arbitrage qui nous servent à bâtir notre modèle empirique. Notre première approche, comme la deuxième, reposent sur des méthodes de quantification permettant de générer des signaux d'investissement à partir de signaux statistiques déterminés par une analyse factorielle.

2.1. Quantification d'un signal résiduel

Nous allons quantifier le signal résiduel $\{\varepsilon_{i,t}\}, 1 \leq i \leq d, 1 \leq t \leq T_0$ obtenu par ACP. Nous prenons une fenêtre de temps ($T_0 = 5$ ans par exemple, mais cette durée peut être variable selon la période d'analyse retenue). Pour tout i , nous avons $\varepsilon_{i,t}$, comme échantillon d'une variable aléatoire ε_i .

Nous définissons ensuite un vecteur aléatoire $\varepsilon = (\varepsilon_1, \dots, \varepsilon_d)$. Nous considérons un quantificateur $\hat{\varepsilon} = (\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_d)$ où chaque composant est obtenu par l'intermédiaire d'algorithmes indépendants unidimensionnels de Kohonen (voir chapitre 1) avec N points.

¹⁰⁷ Ce qui d'un point de vue de gestion est coûteux et fastidieux.

Par

$$\varepsilon_i^k, 1 \leq k \leq N,$$

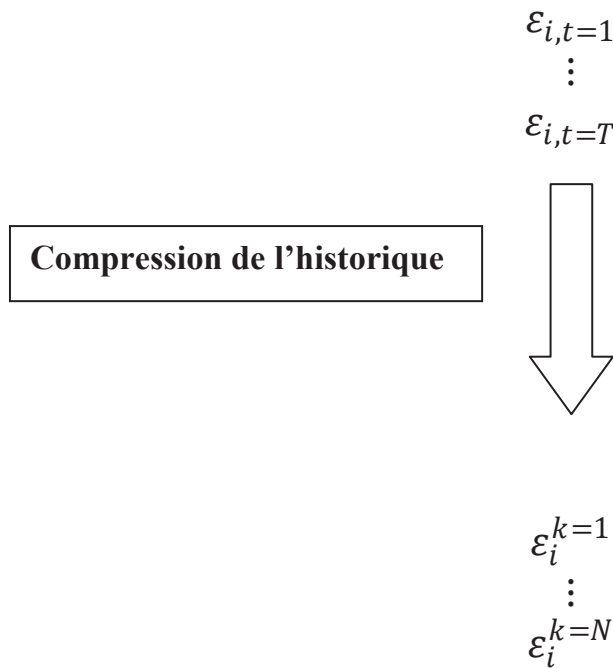
nous signifions les points de quantification de ε_i après que les procédures d'optimisation aient été réalisées et par

$$q_l^k = P(\hat{\varepsilon}_l = \varepsilon_l^k), 1 \leq k \leq N, 1 \leq l \leq d$$

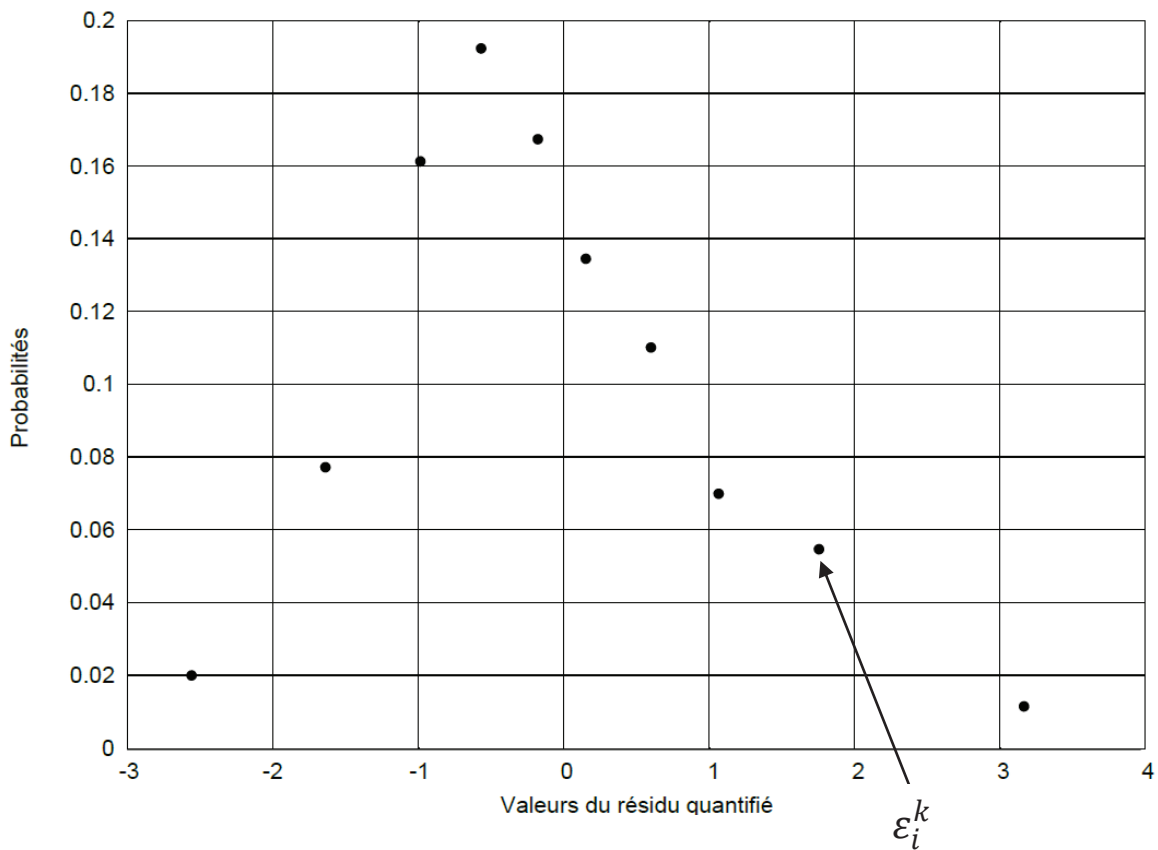
la distribution de $\hat{\varepsilon}_l$.

Figure 2 - Processus de quantification vectorielle (N = 10 points)

$\varepsilon_{i,t}$ est le rendement résiduel ou le risque spécifique d'un actif à l'instant t obtenu après analyse en composantes principales (ACP).



ε_i^k est le k ième point de quantification de $\hat{\varepsilon}_i$, de la distribution ci-dessous.



La figure ci-dessus montre comment la procédure de quantification permet de compresser un historique de données en un nombre finis de points. Ici nous quantifions l'historique avec N = 10 points.

Nous obtenons ainsi une grille de probabilités pour un historique de données aléatoires : comme les données de cotation. Par définition, cela signifie que

$$\hat{\varepsilon}_l = \sum_{k=1}^N \varepsilon_l^k \mathbf{1}_{(\varepsilon^{k-1/2}, \varepsilon^{k+1/2})}(\varepsilon_l)$$

et

$$q_l^k = P\left(\varepsilon^{k-\frac{1}{2}} < \varepsilon_l < \varepsilon^{k+\frac{1}{2}}\right)$$

où $\varepsilon^{k\pm\frac{1}{2}} = (\varepsilon^k + \varepsilon^{k\pm 1})$. Nous nous reportons à la figure 2 pour un exemple de quantification.

2.2. Quantification d'un signal résiduel cumulé

Nous allons, dans une deuxième étape, quantifier le résidu cumulé $X_{i,s}$. Les résidus cumulés $X_{i,t} = \sum_{s<t} \varepsilon_{i,s}$ sont tels que

$$R_{i,t} = \sum_{j=1}^m \beta_{i,j} F_{j,t} + \Delta X_{i,t} + \alpha_i, 1 \leq i \leq d.$$

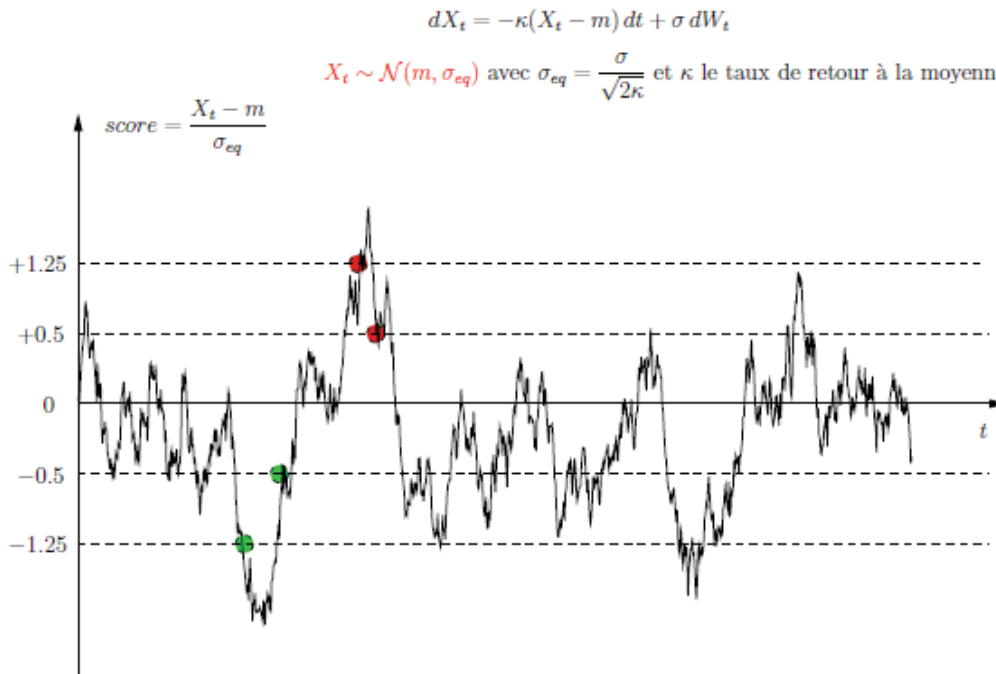
Dans un contexte d'arbitrage statistique, le *market-timing* peut reposer sur l'analyse des résidus cumulés. L'équation ci-dessous suggère une stratégie contrariante dans laquelle nous sommes *long* de 1 dollar sur l'actif i *short* de $\beta_{i,j}$ dollars sur les facteurs F_j si $X_{i,t}$ est « petit » et inversement, nous devenons *short* de l'actif i et *long* des facteurs F_j si $X_{i,t}$ est « grand ».

Usuellement, la modélisation de $X_{i,t}$ s'effectue à travers un processus de retour à la moyenne gaussien ou *mean reverting*, tel que le processus Ornstein-Uhlenbeck (1930)¹⁰⁸ et nous permet de quantifier quelles sont les positions *long* et *short* et quand (*timing*) elles sont déclenchées. La figure 3 représente la méthode usuelle pour extraire des signaux de gestion à partir de la dynamique de $X_{i,t}$.

¹⁰⁸ Voir Avellaneda et Lee (2010) également.

Figure 3 - Signaux d'ouverture et de fermeture de positions pour des résidus de type Ornstein-Uhlenbeck

Signaux de gestion pour des résidus de type processus Ornstein-Uhlenbeck



La figure ci-dessus est un exemple de résidus cumulés modélisés par un processus *mean reverting* gaussien. Les points rouges sont les moments où nous ouvrons/fermons des positions *short* et les points verts ceux où nous ouvrons/fermons des positions *long*.

Nous présentons ci-après la méthode visant à générer des signaux de gestion. Elle repose sur une méthode de quantification dont la présentation a été introduite dans le chapitre 1. Nous utilisons l'information dont nous disposons pour quantifier les processus $X_{i,t}$. Nous considérons de « petites » fenêtres de temps de 60 jours ($T_1 = 60$ jours). Par t , nous indiquons le jour présent. Chaque jour t , en utilisant des données « EOD »¹⁰⁹, nous effectuons une ACP sur une fenêtre de temps $[t - T_1, t]$ et nous obtenons les valeurs de $X_{i,t}$.

Les dynamiques de $X_{i,t}$ peuvent être modélées par un modèle AR(1) (auto-régressif d'ordre 1¹¹⁰) :

$$X_{i,s+1} = a_i X_{i,s} + b_i + \eta_{i,s}$$

où $\eta_{i,s}$ est une variable aléatoire de moyenne nulle. Les coefficients a_i et b_i sont

¹⁰⁹ End of day ou cours de clôture.

¹¹⁰ Voir Bourbonnais et Terraza (2008).

calculés en utilisant une régression linéaire. Nous utilisons l'équation ci-dessous afin de quantifier les variables aléatoires $X_{i,s}$ à chaque jour $s \in [t - T_1, t]$.

Soit N_X un entier et $\widehat{X}_{l,s}$ un quantificateur de Voronoï N_X de $X_{i,t}$. Par définition, nous avons

$$\widehat{X}_{l,s} = \sum_{k=1}^{N_X} x_{i,s}^k 1_{(x_{i,s}^{k-1/2}, x_{i,s}^{k+1/2})}(X_{i,s}),$$

avec le quantificateur $\widehat{X}_{l,s}$, nous pouvons également calculer des paramètres importants : la distribution de $\widehat{X}_{l,s}$ et la distribution de $\widehat{X}_{l,s+1}$ étant donné $\widehat{X}_{l,s}$, soit les nombres :

$$p_{i,s}^k = P\left(x_{i,s}^{k-\frac{1}{2}} < X_{i,s} < x_{i,s}^{k+\frac{1}{2}}\right)$$

et

$$p_{i,s}^{k \rightarrow l} = P\left(x_{i,s+1}^{l-\frac{1}{2}} < X_{i,s+1} < x_{i,s+1}^{l+\frac{1}{2}} \mid x_{i,s}^{k-\frac{1}{2}} < X_{i,s} < x_{i,s}^{k+\frac{1}{2}}\right).$$

Nous rappelons que les valeurs optimales de $p_{i,s}^k$ et $x_{i,s}^k$ sont obtenues en utilisant l'algorithme de Kohonen et les dynamiques de X_i . Nous négligeons l'indice i car chaque quantification est effectuée de manière indépendante.

Nous nous concentrons sur la quantification de X_t sur l'intervalle $[0, T_1]$:

- $x_{T_1}^k = 0, p_{T_1}^k = 0, k = 1, \dots, N_X$.
- Pour ω allant de 1 à MC .
 - $Z_0 = 0$
 - pour s de 0 à $T_1 - 1$ faire
 - $Z_{s+1} = aZ_s(\omega) + b + \widehat{\varepsilon}$.

- Trouver $k^*(s + 1)$ qui soit tel que $x_{s+1}^{k^*(s+1)}$ est le plus proche voisin de Z_{s+1} .
- $x_{s+1}^{k^*(s+1)} \leftarrow Z_{s+1} + (1 - \gamma)(x_{s+1}^{k^*(s+1)} - Z_{s+1})$, où $\gamma > 0$.
- $p_{s+1}^{k^*(s+1)} \leftarrow p_{s+1}^{k^*(s+1)} + 1$
- $p_s^{k^*(s) \rightarrow k^*(s+1)} \leftarrow p_s^{k^*(s) \rightarrow k^*(s+1)} + 1$.

La boucle en ω est une simulation de Monte Carlo (Metropolis et Ulam, 1949). Cette itération permet de simuler un nombre important MC de « copies » indépendantes de processus $s \rightarrow X_s$. Pour ces simulations l'incrément $\hat{\varepsilon}$ simulé selon la distribution $\{q^k\}$ calculée à la première étape. Chaque point sur une couche de temps t représente un point de quantification x_t^k à l'instant t . Les lignes entre les couches représentent la simulation des incréments $\hat{\varepsilon}$.

2.3. Détermination de signaux d'investissement

Nous nous intéressons à présent aux règles de *market timing*¹¹¹, de déclenchement des signaux de gestion. Nous souhaitons précisément connaître les critères de génération de ces signaux.

Pour chaque actif négociable $1 \leq i \leq d_q$, nous utilisons la distribution $\{p_{i,T_1}^k\}$ de \hat{X}_{i,T_1} , afin d'extraire des ordres d'ouverture et de clôture de positions¹¹².

Précisément, pour chaque i , soient $x_{i,OS}$, $x_{i,CS}$, $x_{i,CL}$ et $x_{i,OL}$, les seuils de déclenchement d'ordre respectivement définis tels que :

$$\begin{aligned}
 P(\hat{X}_{i,T_1} < x_{i,OS}) &= P_{OS} = 0,9 \\
 P(\hat{X}_{i,T_1} < x_{i,CS}) &= P_{CS} = 0,75 \\
 P(\hat{X}_{i,T_1} < x_{i,CL}) &= P_{CL} = 0,25 \\
 P(\hat{X}_{i,T_1} < x_{i,OL}) &= P_{OL} = 0,10.
 \end{aligned}$$

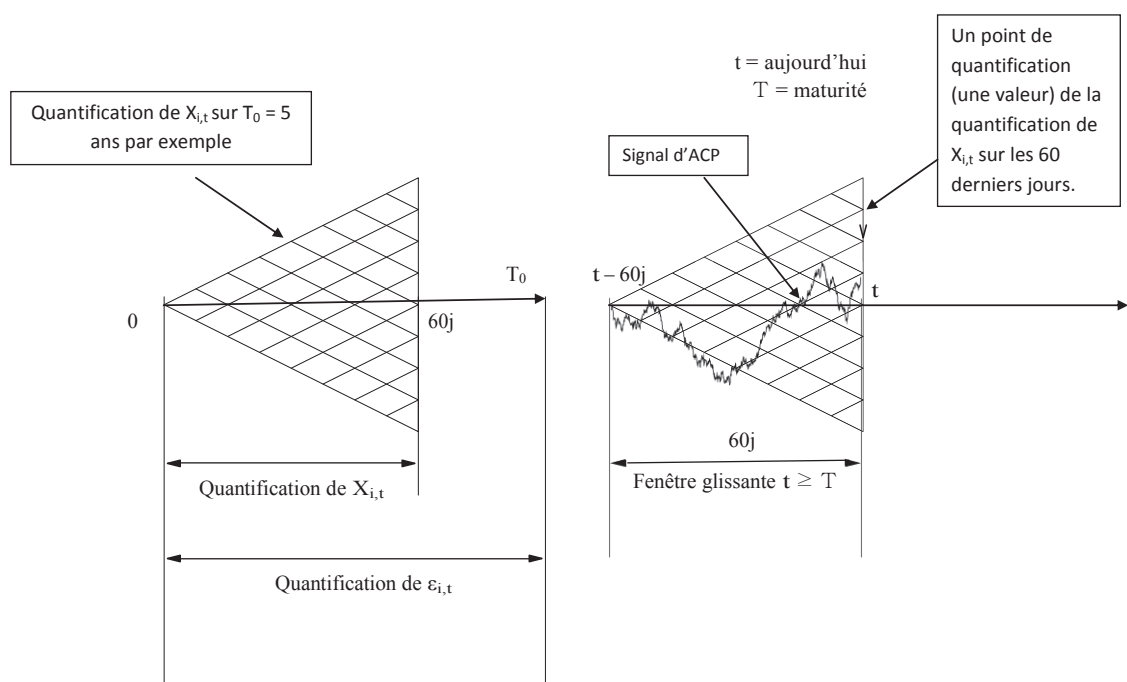
Avec : OS = *Open Short*, CS = *Close Short*, CL = *Close Long*, OL = *Open Long*.

¹¹¹ Signaux d'achat et de vente.

¹¹² Achat et/ou vente.

Ces niveaux servent à prévoir l'évolution future de X_{i,T_1} . Par exemple, si $X_{i,T_1} \sim x_{i,OS}$, alors le processus a une probabilité de 10% de croître. Cela signifie que l'actif i est surévalué et que nous pouvons le vendre. Ces valeurs sont calculées à l'aide de la fonction de répartition de \hat{X}_{i,T_1} ¹¹³. Une fois que ces niveaux, ces seuils sont calculés, nous les comparons avec les valeurs observées de X_{i,T_1} . Nous avons trois choix devant nous : ouvrir une position, la fermer ou ne rien faire. Un schéma synoptique facilite la compréhension de notre approche méthodologique paramétrique (figure 4).

Figure 4 - Approche paramétrique de *market timing* par ACP et par quantification



Etape 1 : quantification de $\varepsilon_{i,t}$ sur l'intervalle $[0, T_0]$ ($T_0 = 5$ ans par exemple).

Etape 2 : quantification de $X_{i,t} = \sum_{s < t} \varepsilon_{i,s}$ en utilisant les incréments $\varepsilon_{i,t}$ sur l'intervalle $[0, T_1]$, $T_1 = 60$ jours par exemple (Avellaneda et Lee, 2010).

Etape 3 : le signal X_t est déterminé par ACP sur une fenêtre glissante $[t, t-T_1]$. Trouver son plus proche voisin \hat{X}_t et générer le signal de gestion *market timing*.

¹¹³ Ces valeurs sont arbitraires et peuvent être changées, par exemple selon la sensibilité du gestionnaire par exemple.

3. Arbitrage systématique : une approche non-paramétrique

Une approche alternative à la modélisation paramétrique consiste à appliquer une méthode qui ne fait référence à aucune loi *a priori*. Cette autre approche, non-paramétrique, permet de déterminer des positions *long* et *short* en intégrant des probabilités de transition (figure 5). Ces probabilités de transition contribuent au caractère dynamique de notre approche dans le sens où elles permettent d'estimer l'évolution haussière ou baissière d'un actif en date t connaissant sa position en date $t-1$. Cette approche exploite la stationnarité du résidu ε_t , en supposant que les probabilités de transition restent invariantes dans le temps. Notre méthode reste cohérente avec le cadre inter-journalier proposé par Avellaneda et Lee (2010). Dans le cadre de la présente recherche, nous analysons une variation entre une date t et une date $t-1$, mais notre approche peut être développée sur un horizon de temps différent. Ce point sera abordé en Conclusion Générale.

3.1. Evaluation directionnelle des données homogènes à l'aide de probabilités de transition

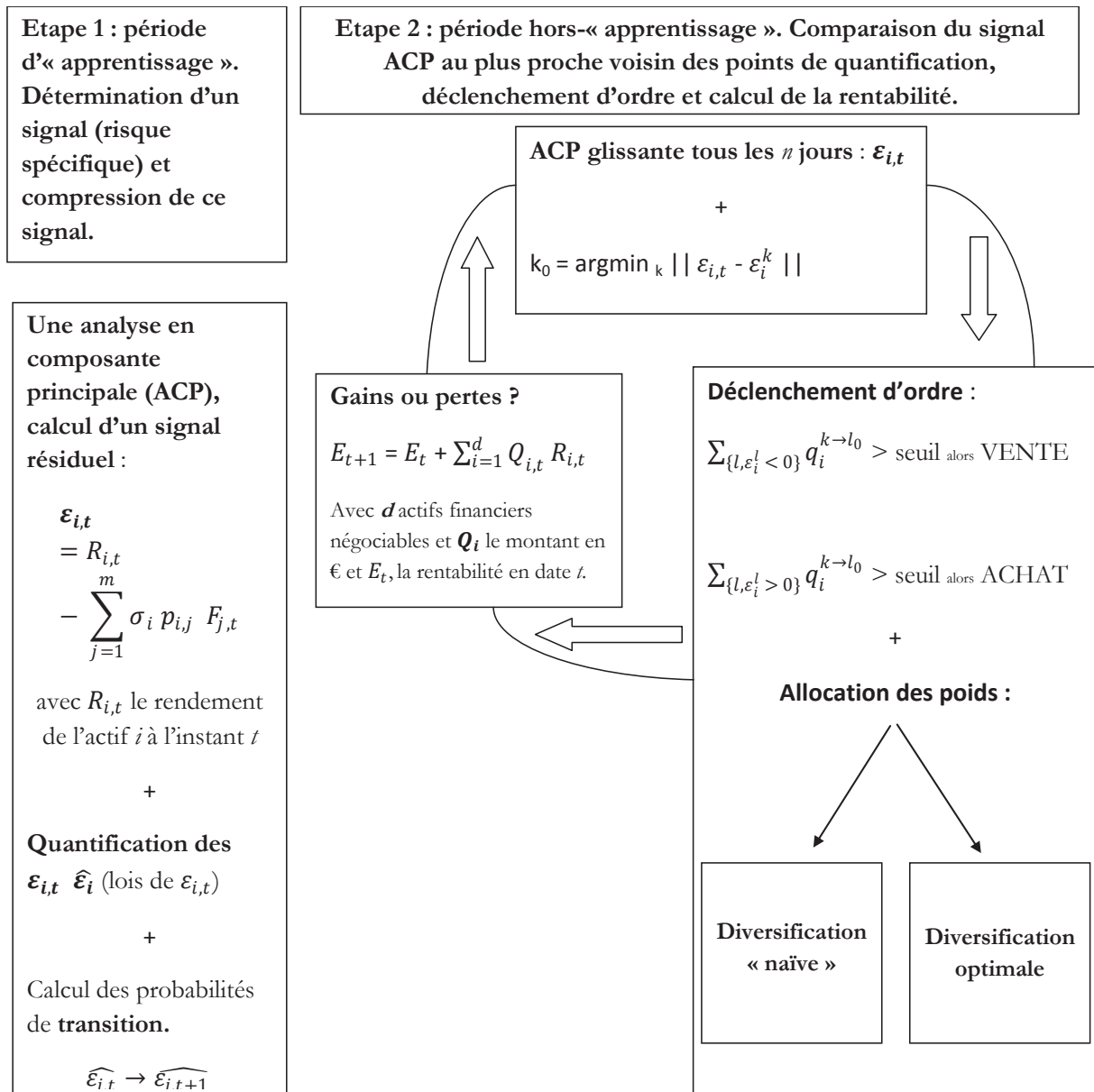
Nous élargissons le cadre probabiliste défini préalablement par notre approche paramétrique à l'aide de probabilités de transition. Ceci nous permet de capturer, autrement qu'en univers paramétrique, l'influence des « autres informations de marché » sur la gestion d'un portefeuille.

Nous cherchons à mesurer la dynamique du taux de variation d'un actif i entre un instant $t-1$ et un instant t , en considérant le rendement passé de ce même actif. En l'espèce, nous mesurons le taux de variation du risque spécifique d'un actif i d'un instant à l'autre : cette variation permet d'analyser le comportement du rendement de l'actif i .

Dans notre contexte d'arbitrage, les positions *long* et *short* peuvent être initiées pour chaque actif i en sommant tous les scénarios haussiers et baissiers provenant du processus de quantification à un instant $t-1$ en espérant qu'en t nous bénéficions¹¹⁴ de l'effet prédictif calculé par notre modèle empirique. Cet effet prédictif provient des distributions obtenues par quantification pour chaque actif i mais aussi des probabilités de transition que nous avons calculées.

¹¹⁴ Par anticipation en $t-1$ pour la date future t .

Figure 5 - Market timing et allocations d'actifs d'un portefeuille d'arbitrage issu d'une approche non-paramétrique par ACP et par quantification



Après avoir déterminé N (prenons par exemple $N = 10$ points de quantification) et le quantificateur $\hat{\varepsilon}_{i,N}$ (figure 2), nous obtenons une matrice de transition dont la taille sera 10×10 . Cette matrice donne la probabilité d'évolution (de transition) d'un point de quantification vers un autre point de cette même grille de quantification : du premier point $\hat{\varepsilon}_{i,1}$ au second point $\hat{\varepsilon}_{i,2}$ par exemple. Ceci indique l'évolution probable que peut connaître le taux de variation d'un actif i en date t sachant qu'il était à un niveau donné en date $t-1$. Cette probabilité de transition est calculée de la façon suivante :

$$P(\hat{\varepsilon}_{i,t} < 0 \mid \hat{\varepsilon}_{i,t-1} = \varepsilon_i^k) = p_{short}^i(\varepsilon_i^k) = \sum_{l, \varepsilon_i^l < 0} q_{i,t}^{k \rightarrow l}$$

où

$$q_{i,t}^{k \rightarrow l} = P\left(\varepsilon_i^{l-\frac{1}{2}} < \varepsilon_{i,t} < \varepsilon_i^{l+\frac{1}{2}} \mid \varepsilon_i^{k-\frac{1}{2}} < \varepsilon_{i,t-1} < \varepsilon_i^{k+\frac{1}{2}}\right).$$

avec ε_i^k , le $k^{\text{ième}}$ point de quantification pour l'actif i , pour $k = 1, \dots, 10$, points quantifiés du $i^{\text{ième}}$ actif et $q_{i,t}^{k \rightarrow l}$ la probabilité pour $\hat{\varepsilon}_i$ d'aller du point k au point l . La somme est effectuée sur l'indice des l pour avoir des scénarios négatifs et nous conduire à de potentiels signaux de vente.

À l'inverse,

$$P(\hat{\varepsilon}_{i,t} > 0 \mid \hat{\varepsilon}_{i,t-1} = \varepsilon_i^k) = p_{long}^i(\varepsilon_i^k) = \sum_{l, \varepsilon_i^l > 0} q_{i,t}^{k \rightarrow l}$$

La somme est effectuée sur l'indice des l pour avoir des scénarios positifs et nous conduire à de potentiels signaux d'achats.

3.2. Interprétation des signaux d'investissement

Nous détaillons les critères de *market timing* issus de l'utilisation de nos probabilités de transition. Nous expliquons comment nous déclenchons des ordres d'achat ou de vente à partir de ces signaux précédemment décrits. Pour illustrer la possibilité d'être *short*, par exemple, nous découvrons la relation ci-dessous.

Si :

$$p_{short}^i(\varepsilon_i^k) > seuil$$

alors la probabilité pour que le $i^{\text{ième}}$ actif baisse est supérieur à *seuil*.

Traduisons cette équation : avec une probabilité supérieure à *seuil* nous avons une certaine chance (x %) que le taux de variation de l'actif i baisse. Si nous vendons l'actif i en date $t-1$ alors en date t et si cette décroissance est avérée *ex post*, nous générons un profit.

Si $p_{short}^i(\varepsilon_i^k) > seuil$ alors une position *short* peut être initiée.

Le paramètre *seuil* est d'importance et peut être modifié. Trop « faible » ou trop

« élevé », il ne permet pas la détection de signaux. Ceci implique qu'un actif peut ne jamais atteindre le niveau de déclenchement calculé si le paramètre *seuil* est à un niveau « extrême ». Ce niveau est également optimisable : nous aborderons ce point en Conclusion Générale.

Pour l'ensemble des classes d'actifs, nous déterminons le niveau moyen de *seuil* ultérieurement dans le chapitre 4. Pour le déclenchement d'une position *long*, si

$$p_{long}^i(\varepsilon_i^k) = 1 - p_{short}^i(\varepsilon_i^k) > \textit{seuil}$$

alors un ordre d'achat est identifié. Nous soulignons que le calcul du *seuil* est dynamique, évolue en fonction de chaque actif *i* dans le temps.

Explicitement, le *seuil* est une fraction de l'écart-type quantifié des probabilités baissières :

$$\textit{seuil} = 0,5 + \theta \left(\sum_{k=1}^N q_i^k (p_{short}^i(\varepsilon_i^k) - 0,5)^2 \right)^{1/2}$$

où θ est un paramètre à choisir

et 0,5 étant la probabilité inconditionnelle sur notre échantillon que ε_i^k soit positif.

Dans notre approche, il n'est pas requis que les écart-types suivent une distribution gaussienne. La flexibilité d'une telle approche est liée au fait que le niveau *seuil* peut être choisi soit sous d'autres hypothèses probabilistes, soit par une approche discrétionnaire en fixant arbitrairement par exemple *seuil* = 60%. La fixation d'un *seuil* discrétionnaire peut se faire en fonction d'un profil de gestion.

Nous ne considérons pas un niveau absolu pour le *seuil*. Poser un *seuil* fixe présente l'avantage d'avoir un niveau unique pour notre univers de données, mais présente l'inconvénient de considérer que tous les actifs ont le même comportement. Nous perdrons ainsi la dynamique propre à l'évolution de chaque actif. Une voie théorique qui permet d'éviter le recours à un *seuil* unique, est d'adapter le *seuil* en fonction de la volatilité de chaque actif en prenant en compte sa variabilité propre. Nous proposons de l'effectuer au travers du calcul du *seuil*.

4. Analyse comparative des approches : paramétrique *versus* non-paramétrique

Nous sommes en présence de deux approches de l'arbitrage statistique, paramétrique et non-paramétrique, auxquelles nous agrégeons des procédures de quantification, usuellement appliquées à la discrétisation des variables aléatoires. L'ensemble de ces méthodes permet d'obtenir un modèle empirique d'arbitrage systématique intégrant ou non des données hétérogènes comme paramètres¹¹⁵ d'entrée. Nous présentons les apports et limites de ces deux modélisations afin de mettre en évidence les avantages comparatifs de la méthode non-paramétrique.

4.1. Convergence des approches

Les deux approches sont des déclinaisons des stratégies historiques d'arbitrage statistique auxquelles nous ajoutons un cadre probabiliste non-paramétrique. Ces deux approches tentent d'estimer la sur ou sous-évaluation des actifs, actif par actif d'un portefeuille, par rapport à un « marché » défini statistiquement dans les deux cas.

Le point de convergence principal des deux approches est qu'elles cherchent à quantifier un risque résiduel, le risque spécifique en gestion. Les algorithmes de quantification, présentés dans le chapitre 1, assurent qu'une compression de l'information est effectuée et que des lois de probabilité sont calculées pour chaque actif.

Quelle que soit l'approche (paramétrique ou non-paramétrique) la composante aléatoire est toujours présente dans le cadre de la quantification, assurant que les points de quantifications obtenus ne sont pas toujours générés de la même façon.

4.2. Singularité des approches

Le caractère paramétrique ou non-paramétrique marque la singularité principale d'une approche par rapport à l'autre. Dépendre d'une loi de probabilité implique un cadre plus « stricte » qu'un cadre où nous ne suivons aucune loi *a priori*.

La stabilité de notre approche paramétrique repose sur la stationnarité du processus de l'actif. Comme nous l'avons rappelé, la stationnarité n'est pas souvent vérifiée pour les séries financières et économiques. Nous pouvons considérer de « courtes » fenêtres de temps à l'intérieur desquelles nous tentons d'estimer les paramètres (moyenne, taux de retour à la

¹¹⁵ *Input.*

moyenne) de quelques modèles *mean reverting* utilisés pour capturer les dynamiques de $X_{i,t}$. Dans une telle approche, dépendante d'un modèle, il peut être difficile d'agréger d'« autres informations de marché » dans le calcul des paramètres.

Notre première approche (paramétrique) cherche à quantifier un signal résiduel différent de celui envisagé par notre deuxième approche (non-paramétrique). Dans le premier cas, il s'agit de compresser un risque résiduel cumulé $X_{i,t}$ au moyen de ses incréments alors que dans la deuxième configuration il s'agit de quantifier le risque résiduel uniquement (non-cumulé).

Le type de quantification diffère d'une approche à l'autre. Pour l'approche paramétrique, nous faisons appel à une quantification par « gerbe ». Notre approche paramétrique peut connaître une période de réallocation des poids des portefeuilles plus longue que dans le cas d'une approche non-paramétrique. Dans notre approche non-paramétrique, la rotation des poids de portefeuilles peut se produire potentiellement d'un jour sur l'autre.

Le retour à la moyenne est plus long pour la première approche (paramétrique) que pour la deuxième (non-paramétrique), notamment car la convergence s'effectue sur des intervalles de temps plus larges. Ceci implique que notre première approche est structurellement moins consommatrice de frais de transactions et que la rentabilité peut être potentiellement moins impactée par ces coûts d'intermédiation.

Une autre différence majeure réside dans le fait que la deuxième approche (non-paramétrique) exploite des probabilités de transition, avec pour apport la considération d'une information *a priori*. Cette information *a priori* est la connaissance du rendement résiduel d'un actif en date $t-1$. Cette approche est plus dynamique que la première (paramétrique), les changements opérés se faisant sur un intervalle de temps plus réduit.

D'un point de vue algorithmique, le temps de calcul de notre approche non-paramétrique est significativement moins long que pour notre première approche (paramétrique). Ainsi, le temps de traitement est réduit avec notre deuxième approche (non-paramétrique) et l'allocation d'actifs est plus rapidement accessible.

Dans ce deuxième chapitre, notre réflexion s'oriente vers l'exploration d'une voie non-paramétrique (dans un cadre d'arbitrage systématique) plutôt que vers une voie paramétrique, largement empruntée dans le cadre de l'arbitrage statistique et de la littérature

existante¹¹⁶. Nous détaillons les tenants et les aboutissants de la détermination de signaux de *market timing* en expliquant les règles quantitatives de génération d'indications de scénarios haussiers et baissiers des taux de variation des prix d'actifs financiers cotés. L'approche non-paramétrique que nous avons développée permet de circonscrire le cadre dans lequel sont effectués nos travaux empiriques. Avant d'explorer l'impact des données hétérogènes sur la gestion d'un portefeuille d'arbitrage selon les types de diversification de portefeuille (chapitres 4 et 5), il convient de présenter dans le chapitre 3 l'échantillon sur lequel sont menées les études empiriques.

¹¹⁶ Voir le chapitre 1.

CHAPITRE 3. ÉCHANTILLON DE L'ÉTUDE

Ce chapitre a pour objet de présenter l'échantillon utilisé dans le cadre de nos études empiriques qui feront l'objet des deux chapitres suivants. Nous commençons par la définition des périodes de l'étude. Nous expliquons le processus de sélection de données, puis détaillons le traitement appliqué aux données retenues. Nous exposons également les critères de choix retenus pour la construction de notre univers de données.

L'échantillon est constitué de manière à nous permettre d'effectuer des tests empiriques sur les données passées. Nous soulevons l'importance d'homogénéiser la fréquence des données, qui du fait de leur périodicité de publication peut être (avant tout traitement) quotidienne, hebdomadaire ou mensuelle. Après des traitements de données, les variables sont toutes quotidiennes.

Enfin, dans la dernière partie du chapitre, nous spécifions précisément la composition et les caractéristiques de notre échantillon au travers de statistiques descriptives et d'analyses factorielles.

1. Périodes de l'échantillon

Notre période d'étude est considérée à la fois dans sa globalité et par sous-périodes. La segmentation de cet échantillon en sous-périodes permet d'analyser l'impact de données hétérogènes sur la gestion d'un portefeuille multi-classes d'actifs pour des périodes économiques et financières distinctes. Notre période d'étude inclut une période de crise: celle des *subprimes*.

1.1. Périodes de l'étude : avant, pendant et après la crise des *subprimes*

Notre échantillon de données couvre la période 2004-2011. Nous présentons ci-après la décomposition de cette période 2004-2011 en trois sous-périodes. Notre échantillon est décomposé en trois sous-périodes, notamment définies en référence à la crise des *subprimes*. La segmentation retenue est la suivante: avant, pendant et après la crise des *subprimes*.

L'idée directrice de notre recherche est de mettre en évidence de manière tangible que la combinaison de données hétérogènes modifie ou non la gestion de portefeuille en univers multi-classes d'actifs. Ce questionnement est soulevé dans des contextes hors-crisis, avant et

après la crise des *subprimes*, ainsi qu'en période de crise des *subprimes*.

Notre échantillon couvre environ sept années d'historique: notre période d'étude débute le 4 août 2004 et s'achève le 3 mars 2011. Précisons que certaines recherches définissent un cycle macroéconomique sur sept années d'étude (Ackermann, McEnally, et Ravenscraft, 1999). En outre, nous soulignons qu'entre 2004 et 2011 la période de notre échantillon comprend des phases de hausses et de baisses boursières. Pour juger de l'efficacité d'une méthode, «il convient de disposer d'un historique de cours qui reflète des configurations de marché aussi variées que possible» (Riva, 2008). Cette alternance permet d'éviter l'obtention de résultats empiriques dont l'analyse pourrait être biaisée par une phase unique de marché.



Notre date de début de période est soumise à la contrainte de disponibilités de données, notamment pour les variables du marché obligataire recueillies (sous-section « 2.1. Recueil des données de cotation »). La date à partir de laquelle toutes les données sont disponibles est le 4 août 2004. La date finale de notre échantillon est le 3 mars 2011, date à partir de laquelle nous effectuons notre dernière requête. A partir de cette date finale, intervient le traitement (section 3 « Traitement des données ») nous permettant d'effectuer des tests et des analyses empiriques (chapitre 4 et 5). Les trois sous-périodes de notre échantillon sont ci-après définies.

Dans notre approche, la sous-période précédant les *subprimes* débute le 4 août 2004 et se termine le 31 janvier 2007, date à partir de laquelle interviendra le premier avertissement sur les résultats de la banque HSBC en raison d'une hausse des provisions pour créances immobilières douteuses aux Etats-Unis (le 8 février 2007)¹¹⁷. Au sein de nos travaux, la sous-période des *subprimes* commence en 1^{er} février 2007 et se termine le 31 mars 2009. Des

¹¹⁷ Les Echos du 22 septembre 2008.

travaux en Finance examinent et définissent ces sous-périodes. Allen et Carletti (2010) situent le début de la crise à partir de l'été 2007 lorsque Khandani et Lo (2007) évoquent les mois tumultueux pour les marchés financiers mondiaux qui ont conduit au déboulement massif de positions financières en août 2007. Dans le cadre de notre recherche, nous prenons en considération les premiers signes d'annonce de la crise des *subprimes* et initions cette période des *subprimes* au 1er février 2007. La date de fin de la sous-période des *subprimes* est le 31 mars 2009 car à partir du 2 avril 2009, le G20 se réunit pour tenter de juguler la crise financière et le ralentissement économique mondial. Allen et Carletti (2010) évoquent un déclin de l'activité économique, aux Etats-Unis et dans d'autres pays, entre la faillite de la banque Lehman Brothers (septembre 2008) et le premier trimestre de l'année 2009. La troisième sous-période est celle *post subprimes* qui débute le 1^{er} avril 2009 et se termine le 3 mars 2011, la date de finale de notre échantillon comme évoqué précédemment. Nous remarquons que nous n'avons pas fait débiter la période des *subprimes* le 8 février 2007 mais quelques jours avant. La même logique est conservée pour la fin de période de la crise des *subprimes*. Des chercheurs (Campbell, Lo, et MacKinlay, 1997 ; Chollet et Cellier, 2011) suggèrent notamment de conserver une zone d'amortissement de quelques jours avant une fenêtre d'estimation.

Nos travaux cherchent à déterminer la valeur ajoutée de l'intégration de données hétérogènes dans un processus d'allocation d'actifs. La conduite de tests empiriques lors des trois sous-périodes susmentionnées (avant, pendant et après la crise des *subprimes*) permet notamment de suivre la portée de notre approche lorsque les contextes changent.

Après avoir situé le contexte économique et financier de la période de notre échantillon, nous définissons les périodes de tests empiriques. Nous segmentons notre échantillon en une période d'« apprentissage » des données et plusieurs périodes de tests, au sein desquelles sont effectuées des études empiriques.

1.2. Période d'« apprentissage » et périodes de tests

Afin de procéder aux tests empiriques, nous divisons notre échantillon en deux périodes : une période d'« apprentissage » des données historiques et des périodes de tests. Les périodes de tests correspondent aux trois sous-périodes suivantes: avant, pendant et après la crise des *subprimes*.

Tableau 1 - Périodes de l'échantillon (2004-2011)

Période de l'échantillon	Nombre de jours de la période	Nombre de jours (% de la période)
Période d'apprentissage (1)	504	29%
Périodes de tests (2)	1213	71%
Total (1) + (2)	1717	100%
Périodes de tests		
Avant la crise des <i>subprimes</i> (3)	147	12%
Pendant la crise des <i>subprimes</i> (4)	564	46%
Après la crise des <i>subprimes</i> (5)	502	41%
Total (3) + (4) + (5)	1213	100%

En partant de la gauche du tableau, dans la première colonne se trouvent les différentes périodes de l'échantillon. La deuxième colonne indique le nombre de jours que comporte chaque période. La troisième colonne fournit le pourcentage du nombre de jours d'une sous-période rapporté au nombre de jours de la période.

Notre période d'« apprentissage » intervient avant le début de la crise des *subprimes*. Cette phase d'« apprentissage », située dans une période hors-crise¹¹⁸, permet de déterminer une loi de probabilités des rendements d'un actif financier en dehors de toutes perturbations. Nous souhaitons que l'estimation de cette loi de probabilités soit le moins sensible possible à tout choc tel qu'un évènement de crise comme celui des *subprimes*: cette imperméabilité témoigne du caractère intemporel de cette loi. Notre période d'« apprentissage » débute en août 2004 et s'achève en juillet 2006. Soulignons que la période d'« apprentissage » choisie est relativement courte en comparaison de celles retenues dans d'autres travaux sur la quantification (Pagès et Printems, 2003 ; Pagès et Printems, 2008). Ces derniers travaux simulaient massivement des variables aléatoires alors que nous travaillons avec des historiques de prix journaliers. Nous analyserons ultérieurement les effets de cette période d'« apprentissage » sur les résultats empiriques dans les chapitres 4 et 5.

Au sein de notre période d'« apprentissage », nous générons des « grilles » de probabilités de rendements de chaque donnée de cotation sélectionnée (voir chapitre 2). Les méthodes de quantification (Pagès et Printems, 2008 ; Pagès et Printems, 2003) mobilisées pour les tests empiriques requièrent une période d'« apprentissage » des données afin de calculer ces « grilles » de probabilités. Cette période d'« apprentissage » équivaut à une période « dans l'échantillon » ou *in sample*. Au sein de notre période d'« apprentissage »,

¹¹⁸ En dehors de toute crise, indépendamment de celle des *subprimes*.

nous réduisons le nombre de données initiales afin d'obtenir ces « grilles » de probabilités à 10 points par exemple. Ces « grilles », stabilisées en période d'« apprentissage », constituent un ensemble d'informations exploitables dans la période de tests « en dehors de l'échantillon ».

Les périodes de tests¹¹⁹ commencent un jour après la fin de notre période d'« apprentissage » et se terminent le 3 mars 2011, date d'arrêt des requêtes de données. Ces périodes de tests incluent une période *pré* crise des *subprimes*, une période de crise des *subprimes* et une période *post subprimes* (tableau 1). Les périodes de tests permettent de comparer un signal statistique¹²⁰ (calculé en périodes de tests) à une loi de probabilité déterminée par quantification en période d'« apprentissage ».

Dans notre approche, nous calculons un signal statistique¹²¹ en dehors de la période d'« apprentissage » des données. Ce signal est ensuite comparé à la « grille » de probabilités des rendements d'un actif¹²². Cette « grille » de probabilités est déterminée depuis la période d'« apprentissage » des données: en ce sens la dichotomie *in sample out, of sample* est conservée. Cette comparaison permet d'anticiper potentiellement d'un jour sur l'autre¹²³ la probabilité que le rendement d'un actif financier soit croissant ou décroissant. Les périodes étant définies, nous nous concentrons maintenant sur le recueil des variables.

¹¹⁹ Constituant l'équivalent d'une période « en dehors de l'échantillon » ou *out of sample*.

¹²⁰ Calculé par ACP. Voir chapitre 2.

¹²¹ Par ACP.

¹²² Initialement obtenue en période d'« apprentissage ».

¹²³ Nous pouvons élargir la longueur de cet intervalle de temps à plus d'une journée. Voir la Conclusion Générale.

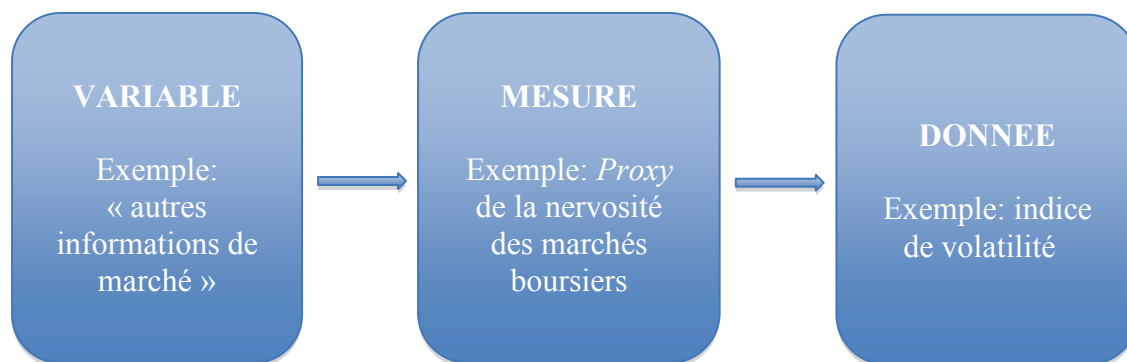
2. Recueil des données de l'échantillon

Dans cette section, nous détaillons la manière dont nous avons sélectionné les variables et recueilli les données de notre échantillon, tant pour les données de cotation que pour les « autres informations de marché ».

Nous rappelons que le recueil de données, bien qu'orienté par des éléments tangibles d'investigation, a été effectué de manière qualitative. Nous basons notre processus de recueil sur une sélection non-formelle telle qu'effectuée par Blitz et Vliet (2008). Les auteurs expliquent que, bien que leur sélection d'actifs ne repose pas sur un ensemble de règles formelles, ils choisissent cependant un univers de données selon certains critères clairement identifiés : ce qui est également notre cas.

Nous soulignons, par ailleurs, que le lien entre les variables et les données sélectionnées n'est pas direct dans le cadre de notre recherche. En l'espèce, une variable est associée à une mesure qui elle-même est matérialisée par une donnée (figure 6).

Figure 6 - Passage d'une variable à une donnée



La collecte des données est intégralement réalisée à partir de la base de données Thomson Reuters Datastream¹²⁴ afin d'obtenir des séries en historique. Les principaux avantages à faire appel à la base de données Thomson Reuters Datastream par rapport à d'autres fournisseurs de données économiques et financières sont: la profondeur (longueur) des historiques de données, la capacité à convertir systématiquement un historique de prix d'actifs en une devise souhaitée (par exemple en euros) et la possibilité d'automatiser des

¹²⁴ En utilisant des tables de requêtes ou *table request* dans l'univers Thomson Reuters Datastream.

requêtes de données. Par automatisation de requête de données, nous faisons référence à des tables de requêtes paramétrables et accessibles sous Excel *via* une interface développée par Thomson Reuters Datastream.

Nous constituons notre univers de données en regroupant deux types de séries: les données de cotation (tableau 2) et les « autres informations de marché » (tableau 3). Comme cela est évoqué dans le chapitre 1, si ces deux types de séries sont des « variables quantitatives » (Evrard, Pras, et Roux, 2009) elles se distinguent cependant l'une de l'autre car les premières sont issues de transactions financières telles que des prix d'actifs financiers cotés¹²⁵ et les deuxièmes regroupent des indicateurs de conjoncture, de sentiment et des indices de volatilité (tableau 3). Nous expliquons pour chacune de ces variables (données de cotation et « autres informations de marché ») quels ont été les critères de choix de données.

2.1. Recueil des données de cotation

Dans notre approche, les données de cotation sont des prix quotidiens de clôture d'actifs financiers cotés, tels qu'Avellaneda et Lee (2010) le font pour définir leur cadre d'arbitrage statistique. Selon les auteurs, le prix de clôture ou « *end of day* » (« *EOD* ») est supposé prendre en compte l'intégralité du signal de prix en agrégeant toutes les informations de la journée. La nature de l'information agrégée reste une zone de débats. Non par hasard, Brenner et al. (2009) considèrent des données de cotation quotidiennes car ces dernières sont adéquates pour évaluer la significativité économique et statistique de l'influence des informations fondamentales sur les classes d'actifs. De plus, cette fréquence quotidienne semble correspondre à un horizon de temps pertinent pour les analystes, les investisseurs et les entités réglementaires. Les auteurs évoquent le fait que les données « haute fréquence » sont exposées à des frictions de microstructure, pouvant biaiser les analyses du lien entre l'économie « réelle » et les marchés financiers.

Dans le cadre de notre recherche, nous n'intégrons pas des prix intra-journalières (*intraday*) car ils ont des structures de corrélations qui ne sont pas les mêmes que pour des prix journaliers (prix de clôture). Avec des prix intra-journaliers, les rendements sont fortement corrélés et les variances le sont également (Mandelbrot, 1963 ; Plerou et al., 1999). Rappelons que le phénomène sous-jacent dont bénéficie notre approche de gestion est notamment la corrélation négative d'un jour sur l'autre de l'actif (voir chapitre 2) qui n'entre pas obligatoirement dans le cadre *intraday*.

¹²⁵ Sur les marchés financiers avec un prix de clôture par exemple.

Comme nous l'avons mentionné en Introduction Générale, le contexte de notre recherche est celui de l'arbitrage en univers mondial et multi-classes d'actifs. C'est au sein de cet environnement mondial et multi-classes d'actifs que nous sélectionnons des données de cotation (tableau 2). La couverture géographique de notre échantillon de données de cotation est mondiale pour tenter de capter, entre autres, les arbitrages inter-continentaux. Dahlquist et Campbell (2001) indiquent qu'il existe des écarts de rendements sur les marchés actions entre les Etats-Unis et les autres pays dans le monde. Ces écarts sont des sources potentielles d'arbitrage lorsqu'ils sont détectés au(x) moment(s) opportun(s).

Tableau 2 - Données de cotation (données homogènes) de l'échantillon

Données homogènes	Classe d'actifs	Type de cotation
France 2004 4% 25/04/14	Obligations	Au comptant
US 2004 4% 15/02/14	Obligations	Au comptant
Japan 2004 1,5% 20/03/14	Obligations	Au comptant
Brazil 2004 10,5% 14/07/14	Obligations	Au comptant
Brent (pétrole)	Matières premières	A terme
West Texas Intermediate (pétrole)	Matières premières	A terme
Or	Matières premières	Au comptant
Blé	Matières premières	A terme
Maïs	Matières premières	A terme
Bovespa (Brésil)	Actions (indice)	Au comptant
Bse sensx (Inde)	Actions (indice)	Au comptant
Cac 40 (France)	Actions (indice)	Au comptant
Dax (Allemagne)	Actions (indice)	Au comptant
Ftse 100 (Royaume Uni)	Actions (indice)	Au comptant
S&P 500 (Etats-Unis)	Actions (indice)	Au comptant
Nasdaq 100 (Etats-Unis)	Actions (indice)	Au comptant
Asx 200 (Australie)	Actions (indice)	Au comptant
Shanghai SE (Chine)	Actions (indice)	Au comptant
Nikkei 225 (Japon)	Actions (indice)	Au comptant
Dollar américain / euro	Devises	Au comptant
Sterling /euro	Devises	Au comptant
Franc suisse / euro	Devises	Au comptant
Yen / euro	Devises	Au comptant
Yuan /euro	Devises	Au comptant
Couronne norvégienne / euro	Devises	Au comptant
Real brésilien / euro	Devises	Au comptant
Roupie indienne / euro	Devises	Au comptant
Won coréen / euro	Devises	Au comptant
Dollar australien / euro	Devises	Au comptant
Rand sud africain / euro	Devises	Au comptant
Dollar néo-zélandais / euro	Devises	Au comptant
Dollar canadien / euro	Devises	Au comptant

Source : Thomson Reuters Datastream

En partant de la gauche du tableau, le nom de la donnée homogène est en première colonne. La classe d'actifs à laquelle appartient l'actif est en deuxième colonne. Le type de cotation de l'actif est en troisième colonne.

Dans le chapitre 1, nous avons fait référence aux stratégies d'investissement multi-classes d'actifs qui s'inscrivent souvent dans un univers mondial, comme cela est le cas pour le *global macro*. Comme le rappellent Blitz et Vliet (2008), les candidats à l'arbitrage dans un univers multi-classes d'actifs sont vraisemblablement les fonds alternatifs de type « global macro ».

Les différentes classes d'actifs que nous retenons sont « traditionnellement » celles des stratégies multi-classes : les obligations, les actions, les matières premières et les devises. Ces classes d'actifs correspondent aux différentes classes sur lesquelles sont effectuées les allocations multi-classes, notamment pour les stratégies *global macro*. Chan et al. (2005) définissent le *global macro* comme une stratégie visant à initier des positions sur toutes classes d'actifs dans le monde. Stulz (2007) propose une définition de la catégorie « *macro hedge fund* » : « une stratégie qui vise à identifier des mauvaises évaluations¹²⁶ sur les marchés actions, de taux, de devises et de matières premières ».

Alors que certains auteurs (Brenner, Pasquariello, et Subrahmanyam, 2009) rappellent la pertinence d'inclure un univers composé d'actions et d'obligations, d'autres (Andersen et al., 2007 ; Black et Litterman, 1992) élargissent cette sélection en intégrant également des devises. Par notre sélection multi-classes élargie, nous souhaitons compléter l'approche de Blitz et Vliet qui soulignent que les classes des devises et des matières premières sont difficilement modélisables au regard de leur valorisation.

Au sein de ces classes d'actifs, nous découvrons les axes de recueil des données de cotation de notre échantillon (tableau 2). Les données de cotation sont choisies selon deux axes au sein de notre univers multi-classes d'actifs (obligations d'Etat, matières premières, indices actions et devises) : les actifs sont majoritairement des instruments financiers « non complexes »¹²⁷ au sens de la Directive MIFID 2004¹²⁸ et au sein de sa classe d'actifs, l'actif recueilli est censé être un actif « majeur ».

Avec notre premier critère, nous cherchons notamment à montrer qu'il est possible de générer de la rentabilité sans obligatoirement recourir à des instruments financiers « complexes »¹²⁹ et à effet de levier. Nous vérifions ce point dans les chapitres 4 et 5.

Précisons que pour ne pas exclure la classe des matières premières selon le critère de

¹²⁶ L'auteur semble faire référence à une sous-évaluation ou une surévaluation.

¹²⁷ Hormis les contrats à terme de la classe des matières premières.

¹²⁸ Transposée au droit français en 2007.

¹²⁹ De type options exotiques et autres produits dérivés complexes.

« complexité » au sens MIFID, nous avons choisi d'utiliser des contrats à terme fermes¹³⁰ pour certains actifs¹³¹, qui sont pourtant assimilés *stricto sensu* à des instruments financiers « complexes »¹³². Exclure la classe d'actifs « matières premières » est en contradiction avec l'univers multi-classes et la corrélation entre des marchés financiers « majeurs » (voir les chapitres 4 et 5). En tant que chercheur en Sciences de Gestion, et spécifiquement en gestion de portefeuille, il nous semble plus adéquat d'effectuer des tests empiriques sur des données de contrats à terme fermes plutôt que des données d'actifs physiques (marchandises) qui n'entrent pas dans le périmètre conceptuel des actifs financiers.

Notre deuxième axe de sélection signifie que l'actif doit être un actif important de sa classe d'actifs dans le sens où il est représentatif de sa classe. Le fait qu'un actif soit « majeur » au sein de sa classe peut, de manière générale, être approché avec le concept de la découverte de prix (Biais, Bisiere, et Decamps, 1999). Certains facteurs comme la découverte du prix peuvent aider à déterminer l'importance d'un actif au sein de sa classe dans le sens où une découverte de prix se produit sur un marché principal. Ce critère est général et nous nous appuyons sur d'autres sources d'information pour expliquer la composition de notre échantillon de données de cotation.

Pour la classe « actions », nous avons sélectionné des places boursières « historiques » (dont les Etats-Unis, l'Angleterre, la France ou le Japon) ainsi que des places boursières plus récemment actives (dont la Chine ou le Brésil). La majorité des indices actions recueillis sont notamment choisis par d'autres recherches (Andersen et al., 2007 ; Dahlquist et Campbell, 2001 ; Ibbotson et Kaplan, 2000 ; Ferson et Harvey, 1993 ; Black et Litterman, 1992) et sont également recensés dans l'univers *MSCI World Index*.

Pour la classe « matières premières », nous avons sélectionné certains composants (tableau 2) du *S&P GSCI Commodity index*. Dans la littérature existante en Finance, des recherches (Bruder et al., 2013 ; Gorton et Rouwenhorst, 2006) incluent les matières premières que nous sélectionnons pour notre univers de tests. D'un point de vue fondamental, les données de cotation de la classe « matières premières » (tableau 2) sont globalement influencées par des facteurs économiques et géopolitiques.

Au sein de la classe « devises », nous retenons les devises qui représentent une part

¹³⁰ Au sens MIFID.

¹³¹ Sauf pour l'or.

¹³² Notons que les contrats à terme sont les produits dérivés les moins sophistiqués au niveau du *pricing*. Les contrats à terme ou *futures* sont a priori plus accessibles que les options.

significative (>5%¹³³) du chiffre d'affaires mondial du marché des changes en avril 2010 (Von Kleist et al., 2010), les devises qui sont impliquées dans des stratégies de « *carry trade* »¹³⁴, les devises des « marchés émergents » et les « nouvelles paires de devises » d'après ledit rapport. Les devises les plus contributrices du chiffre d'affaires mondial du marché des changes sont le dollar américain, l'euro, le yen, le franc suisse et la livre sterling. Nous retrouvons par ailleurs certaines de ces devises¹³⁵ dans les travaux d'Andersen et al. (2007).

Les obligations d'Etat recueillies sont celles d'Etats qui sont parmi les plus emprunteurs au niveau mondial (tableau 2)¹³⁶. Les obligations sont des émissions à 10 ans, qui correspondent à une des maturités négociées de manière soutenue (Brenner, Pasquariello, et Subrahmanyam, 2009). Jones et al. (1998) choisissent, entre autres, des obligations d'Etat américain dont la maturité est de 10 ans, car des enquêtes économiques peuvent avoir un impact sur les prix de ces actifs. Un rapport de la Banque de France (Bourgey, 2012), souligne la montée tendancielle de l'endettement public depuis le début des années soixante-dix. Ce rapport permet d'insister sur le fait que « nombre de pays industrialisés ont abordé la crise de 2007 avec un niveau d'endettement élevé ». Ce rapport évoque le niveau d'endettement des pays industrialisés très élevé notamment au moment où se déclenche la crise financière pour les Etats-Unis, la zone euro et le Japon. Les ratios d'endettement de pays émergents entre 2004 et 2014 (estimé) montrent que le Brésil est parmi les pays les plus endettés.

¹³³ (Von Kleist et al., 2010).

¹³⁴ Stratégie de portage.

¹³⁵ Le dollar américain, la livre sterling, l'euro et le yen.

¹³⁶ La date à partir de laquelle toutes les données obligataires étaient disponibles correspond au 4 août 2004.

2.2. Recueil des données relatives aux « autres informations de marché »

Parmi les variables de notre échantillon, nous incluons d'autres types de variables que des données homogènes. Ces autres variables sont composées de trois types d'indicateurs : des indicateurs de conjoncture économique, de sentiment et des indices de volatilité (tableau 3). Nous précisons que les différents sentiments exprimés au sein de ces indicateurs sont formalisés numériquement. Nous tentons de mesurer la complémentarité entre des données homogènes et d'« autres informations de marché » dans le cadre d'une gestion de portefeuille d'arbitrage multi-classes d'actifs.

Tableau 3 - Données relatives aux « autres informations de marché » par type d'indicateur

Données	Source	Type d'indicateur
Sentiment économique des investisseurs pour l'économie allemande	Zentrum für Europäische Wirtschaftsforschung (Zew)	Indicateur de conjoncture
Sentiment des consommateurs américains	Université du Michigan, Thomson Reuters	Indicateur de sentiment
Climat des affaires allemand	Ifo Institute	Indicateur de conjoncture
Industrie manufacturière belge	Banque nationale de Belgique	Indicateur de conjoncture
Directeurs d'achat	Institute of Supply Management	Indicateur de conjoncture
Sentiment haussier des marchés boursiers américains	American Association of Individual Investors	Indicateur de sentiment
Sentiment baissier des marchés boursiers américains	American Association of Individual Investors	Indicateur de sentiment
Sentiment neutre des marchés boursiers américains	American Association of Individual Investors	Indicateur de sentiment
Indice de volatilité VIX américain	CBOE	Indice de volatilité
Indice de volatilité VDAX allemand	Euronext	Indice de volatilité
Indice de volatilité VCAC français	Euronext	Indice de volatilité

En partant de la gauche, le type de donnée d'« autres informations de marché » est en première colonne. La source de la donnée se trouve dans la deuxième colonne. Le type d'indicateur est en troisième colonne.

Nous détaillons et analysons les axes de recueil des données liées aux variables de type « autres informations de marché ». Nous recensons, par la suite, les similarités entre les données recueillies.

2.2.1. Critères de choix des données relatives aux « autres informations de marché »

Dans la même logique que dans la sous-section « Recueil des données de cotation », un tri non-formel des données (Blitz et Vliet, 2008) est effectué parmi les variables de type « autres informations de marché ». Trois critères équipondérés permettent d'orienter notre choix de données parmi les nombreuses séries de la famille des « autres informations de marché » (figure 7).

Figure 7 - Critère de choix des données de la variable « autres informations de marché »

1^{er} critère : expression totale ou partielle d'un sentiment au sein de l'indicateur

2^{ème} critère : impact de l'indicateur sur les données de cotation de l'échantillon

3^{ème} critère : fréquence de publication de l'indicateur \leq fréquence mensuelle

Le premier critère est que la série doit exprimer, au moins partiellement, un sentiment. Le deuxième critère est que la série doit avoir un impact direct ou indirect sur les données de cotation de notre échantillon. Le troisième critère est que la série doit avoir une fréquence de publication inférieure ou égale au mois¹³⁷ pour rester cohérent avec un univers de gestion relativement « court-terme »¹³⁸ car nos données de cotation sont journalières.

Le premier critère permet de retenir des séries qui intègrent le sentiment des agents économiques et/ou des investisseurs. Pour l'enquête de sentiment des investisseurs individuels américains (tableau 3), historiquement un pic d'optimisme précède souvent une forte correction à la baisse des marchés boursiers (Thorp, 2004) : il s'agit d'un indicateur pouvant avoir un effet dit « contrariant ». Dans cette perspective, les sentiments des

¹³⁷ D'une fréquence journalière à mensuelle.

¹³⁸ Comparé à une fréquence *intraday*, un horizon quotidien peut être vu comme « long-terme ».

investisseurs peuvent communiquer des informations s'agréant à un processus de décision final.

Le deuxième axe (critère) de notre recueil de données est que les séries sélectionnées aient un impact direct ou indirect sur toutes les classes d'actifs. Afin de conserver une cohérence dans notre recherche, il est nécessaire que les données triées aient un lien avec les séries d'actifs de notre échantillon. Comme le préconisent Blitz et Vliet (2008), nous souhaitons éviter le fait qu'une série soit spécifique à une classe d'actifs.

Avant de comprendre comment certains travaux abordent ce lien, nous pouvons souligner les influences potentielles entre les différents types d'indicateurs et les classes d'actifs de notre périmètre d'investigation. Les indicateurs de conjoncture peuvent avoir un impact sur toutes classes d'actifs puisque l'économie a des interactions avec la sphère financière. Les indicateurs de sentiment sur la consommation peuvent être des indicateurs avancés du niveau d'achats des ménages et de leur contribution la croissance d'un pays. La consommation est un moteur de l'économie et son recul peut aussi être un indicateur avancé d'une dégradation des conditions économiques qui auront *in fine* des répercussions sur l'univers des marchés financiers. Les indices de volatilités des marchés boursiers sont *a priori* impactant pour la classe « actions ». Dans la littérature existante ainsi que dans les faits, ces indices de volatilités peuvent être perçus comme des *proxy* de la nervosité, de la « peur » sur les marchés financiers dépassant le cadre *stricto sensu* des marchés « actions ».

Revenons sur quelques éléments de travaux existant en Finance qui explorent le lien entre les indicateurs et les données de cotation que nous avons sélectionnés. Comme mentionné précédemment, les séries sélectionnées peuvent exercer une influence sur différentes classes d'actifs.

Concernant les indicateurs de conjoncture, Andersen et al. (2007) étudient notamment les différents impacts des publications économiques mensuelles (dont consommation, production industrielle, inflation, confiance des consommateurs, achats des entreprises) sur les rendements des classes « actions », « obligations » et « devises » aux Etats-Unis, en Angleterre et en Allemagne. Dahlquist et Campbell (2001) soulignent que l'impact du cycle des affaires, d'activités¹³⁹ des Etats-Unis sur les rendements des actions est plus fort sur les pays étrangers que sur les rendements des actions aux Etats-Unis. Les rendements des marchés actions des pays étrangers seraient plus sensibles au cycle des affaires que les rendements américains. Cette découverte empirique explique la raison pour laquelle nous

¹³⁹ Au sens économique.

introduisons plusieurs indicateurs aux Etats-Unis. Plusieurs de nos indicateurs examinent la conjoncture et donnent des informations relatives à ce cycle (tableau 3).

Allen et Carletti (2010) montrent au travers d'enquêtes comment les dirigeants d'entreprises de pays comme la France, les Etats-Unis ou le Japon affichent une préférence soit pour la sécurité de l'emploi, soit pour la distribution de dividendes. Ce choix semble entraîner une instabilité macroéconomique en période de crise qui peut affecter potentiellement toutes classes d'actifs. Ferson et Harvey (1993) montrent que certaines variables d'« information », comme l'inflation ou la production industrielle, peuvent entraîner des différences de rendements entre marchés financiers internationaux.

Les indicateurs de sentiment économique du Zew et du climat des affaires de l'Ifo (tableau 3) peuvent à leur tour impacter les cotations de l'indice actions du Dax en Allemagne (Entorf, Gross, et Steiner, 2011). Selon Fisher et Statman (2003), la confiance des consommateurs américains croît lorsque les rendements du S&P 500 augmentent : cette relation tend à montrer qu'un lien mesurable existe entre les données de cotation et des données de type « autres informations de marché ». D'autre part, les auteurs concluent qu'un faible niveau de confiance des consommateurs permet de prédire de « hauts » rendements des actions du Nasdaq ainsi que les rendements d'actions à plus faible capitalisation. Dans leur optique, le sentiment des investisseurs peut être décisif dans une allocation tactique d'actifs financiers. Ils ont découvert qu'il existe une relation négative et statistiquement significative entre le sentiment des investisseurs, individuels et institutionnels, et les rendements futurs du S&P 500. Blitz et Vliet (2008) identifient certaines variables prédictives affectant toutes les classes d'actifs financiers non encore prises en compte dans le cadre de modèles d'allocations comme la confiance des consommateurs et des producteurs.

Concernant les indices de volatilité, l'indice VIX peut permettre d'évaluer « la peur » des investisseurs dans un marché baissier (Whaley, 2009). L'auteur évoque un pic du niveau du VIX en période de nervosité des marchés financiers. L'impact statistique de l'indice VIX, théoriquement sur la classe « actions », se manifeste également pour les obligations et les matières premières (Munenzon, 2010). Nous soulignons qu'il peut également exister des relations explicatives au sein des variables « autres informations de marché ». La confiance des consommateurs peut augmenter à mesure que le sentiment haussier des investisseurs croît et cette relation peut être statistiquement significative (Fisher et Statman, 2003).

A l'aide de notre troisième critère de sélection, les indicateurs retenus sont tous publiés à une fréquence régulière inférieure ou égale au mois en lien avec un univers de données de cotations relativement « court-terme ». Les indicateurs « d'autres informations de marché » peuvent être quotidiens, hebdomadaires ou mensuels (tableau 4).

Notons que parmi les données relatives aux « autres informations de marché », les indicateurs des directeurs d'achat et de sentiment des consommateurs (tableau 3) sont des composants de l'indice économique avancé du *Conference Board* aux Etats-Unis. Le fait que ces deux séries soient des composants de cet indice, témoigne du caractère anticipatif de ces indicateurs sur la conjoncture. Au niveau européen, les indicateurs du Zew et de l'Ifo (tableaux 3, 4 et 5) sont des indicateurs avancés du cycle de l'économie allemande (Entorf, Gross, et Steiner, 2011).

Parmi les indicateurs, certains communiquent des informations relatives aux relations économiques internationales des zones géographiques industrialisées (Chine, Europe) : l'indicateur des directeurs d'achat américains (tableau 3) notamment avec les équipements de transports échangés dans le monde.

Par l'intermédiaire de ces trois critères (figure 7), nous collectons un ensemble de données matérialisant la variable « autres informations de marché » (tableau 3). Notons que ces données constituent des indicateurs basés sur des éléments passés, prospectifs, voire une combinaison des deux éléments dans certains cas (tableau 5). La série des directeurs d'achat américains est la seule qui ne prend en compte que des éléments historiques mais elle reste construite comme la plupart des autres enquêtes de notre sélection. Les séries de notre échantillon composées d'éléments historiques et prévisionnels (climat des affaires allemand, industrie manufacturière belge) se distinguent de celles qui sont exclusivement composées d'éléments prospectifs (sentiment économique des investisseurs pour l'économie allemande, sentiment des consommateurs américains, sentiment haussier, baissier et neutre des marchés boursiers, indices de volatilité).

Tableau 4 - Données recueillies pour la variable « autres informations de marché »

Données	Source	Fréquence de publication
Sentiment économique des investisseurs pour l'économie allemande	Zentrum für Europäische Wirtschaftsforschung (Zew)	Mensuelle
Sentiment des consommateurs américains	Université du Michigan, Thomson Reuters	Mensuelle
Climat des affaires allemand	Ifo Institute	Mensuelle
Industrie manufacturière belge	Banque nationale de Belgique	Mensuelle
Directeurs d'achat	Institute of Supply Management	Mensuelle
Sentiment haussier des marchés boursiers américains	American Association of Individual Investors	Hebdomadaire
Sentiment baissier des marchés boursiers américains	American Association of Individual Investors	Hebdomadaire
Sentiment neutre des marchés boursiers américains	American Association of Individual Investors	Hebdomadaire
Indice de volatilité VIX américain	CBOE	Quotidienne
Indice de volatilité VDAX allemand	Euronext	Quotidienne
Indice de volatilité VCAC français	Euronext	Quotidienne

En partant de la gauche, le type de donnée d'« autres informations de marché » est en première colonne. La source de la donnée se trouve dans la deuxième colonne. La fréquence de publication de la donnée est en troisième colonne.

Tableau 5 - Composantes historique et prospective des données relatives aux « autres informations de marché »

Données	Source	Composante historique de la donnée	Composante prospective de la donnée
Sentiment économique des investisseurs pour l'économie allemande	Zentrum für Europäische Wirtschaftsforschung (Zew)	Non	Oui
Sentiment des consommateurs américains	Université du Michigan, Thomson Reuters	Non	Oui
Climat des affaires allemand	Ifo Institute	Oui	Oui
Industrie manufacturière belge	Banque nationale de Belgique	Oui	Oui
Directeurs d'achat	Institute of Supply Management	Oui	Non
Sentiment haussier des marchés boursiers américains	American Association of Individual Investors	Non	Oui
Sentiment baissier des marchés boursiers américains	American Association of Individual Investors	Non	Oui
Sentiment neutre des marchés boursiers américains	American Association of Individual Investors	Non	Oui
Indice de volatilité VIX américain	CBOE	Non	Oui
Indice de volatilité VDAX allemand	Euronext	Non	Oui
Indice de volatilité VCAC français	Euronext	Non	Oui

En partant de la gauche, le type de donnée d'« autres informations de marché » est en première colonne. La source de la donnée se trouve dans la deuxième colonne. L'indication de composante historique (passée) de la donnée est en troisième colonne. L'indication de composante prospective est en quatrième colonne.

Nous notons que l'indice des nouvelles commandes de l'enquête des directeurs d'achat ainsi que le sentiment des consommateurs américains (tableau 4) sont deux composants de l'indice des indicateurs avancés de l'économie du *Conference Board* aux Etats-Unis (sur 10 composants). La prise en compte de ces deux indicateurs par le *Commerce Department* américain confirme leurs capacités à comprendre et prévoir les changements économiques. Chaque série appartenant à l'indice des indicateurs avancés du *Conference Board* est choisie pour sa performance sur critères rigoureux dont : la significativité économique, l'adéquation statistique, la cohérence du *timing* lors de pics de cycles économiques et la rapide disponibilité des chiffres.

Bien que nos critères mettent en avant l'expression d'un sentiment en lien avec les marchés financiers, il est intéressant de constater que certains des indicateurs présentent un caractère anticipatif de la conjoncture économique.

2.2.2. Similarités entre données recueillies

Certaines séries de notre échantillon partagent des caractéristiques communes de contenu (tableau 3) et de méthodologie.

En termes de contenu, le sentiment de confiance dans l'économie est mesuré au travers de différents indicateurs (directeurs d'achat américains, industrie manufacturière belge, climat des affaires allemand, sentiment économique des investisseurs pour l'économie allemande et sentiment des consommateurs américains). A titre d'exemple, cette confiance se mesure pour la production (industrie manufacturière belge, climat des affaires allemand et sentiment des consommateurs américains) et pour les attentes économiques sur la conjoncture (industrie manufacturière belge, climat des affaires allemand, sentiment économique des investisseurs pour l'économie allemande et sentiment des consommateurs américains).

Le sentiment de « nervosité » sur les marchés financiers peut être approché au travers d'indices de volatilité (indices de volatilité VIX, VDAX, VCAC). Les anticipations directionnelles sur les marchés financiers (sentiment haussier, neutre ou baissier des marchés boursiers américains) mesurent la confiance des investisseurs pour les mois à venir sur les marchés boursiers.

La dimension humaine, traduite par l'ensemble des données recueillies, introduit par construction de la subjectivité, des opinions dans l'interprétation des enjeux économiques et l'expression de sentiments divers¹⁴⁰. Notre recherche tente de mesurer la valeur d'une combinaison de prix d'actifs financiers et d'informations dont le contenu est dans une certaine mesure qualitatif. L'apport vient également du fait que nous quantifions cette influence sur des portefeuilles d'arbitrage en univers multi-classes d'actifs. En conservant cette dimension subjective dans le cadre de notre recherche, le but est de mesurer quantitativement cet impact dans un processus de décision.

La subjectivité est constatée au travers de la majorité de nos indicateurs (l'ensemble des données hormis les indices de volatilité¹⁴¹) qui fonctionnent avec le même type de réponses. Ces questions proposent trois types de possibilités de réponses : positives, égales ou négatives.

Quelques illustrations viennent en compléments de nos propos. Pour évaluer le contexte actuel des affaires dans le cadre du climat des affaires allemand (tableau 3), les réponses sont de la forme suivante : « bon », « satisfaisant » ou « pauvre ». Pour ce même

¹⁴⁰ Sentiments sur la conjoncture économique mais aussi sur les marchés boursiers.

¹⁴¹ Pour les indices de volatilités, la dimension subjective n'est appréciée que par la caractéristique de *proxy* de la « peur », de la « nervosité » sur les marchés financiers affectée à ces indices.

indicateur, les réponses proposées pour évaluer les attentes sur le climat des affaires pour les six prochains mois suivent la même logique : « plus favorables », « inchangées » ou « moins favorables ». L'indicateur du sentiment économique des investisseurs pour l'économie allemande (tableau 3) proposent des réponses telles que : « s'améliorer », « aucun changement » ou « s'aggraver ». Pour le sentiment des marchés boursiers américains (tableau 3), les trois choix de réponses sont : « haussier », « neutre » ou « baissier ». Ces réponses mettent en exergue le caractère subjectif de la « normalité », du bon ou du mauvais par exemple. D'autre part, le sentiment d'appréciation, de stagnation ou de recul comporte également une composante subjective.

Dans plusieurs cas, la durée sur laquelle porte l'estimation des données relatives aux « autres informations de marché » est de six mois maximum (sentiment haussier, neutre ou baissier des marchés boursiers américains, industrie manufacturière belge, climat des affaires allemand et sentiment économique des investisseurs pour l'économie allemande).

Deux similarités méthodologiques existent dans le calcul des données publiées. La première similarité repose sur le principe de soldes ou *balances* : ce qui consiste à soustraire les opinions « positives » aux opinions « négatives », sans tenir compte des opinions « neutre » lorsqu'elles existent dans l'indicateur¹⁴². Les indicateurs employant ce mode de calcul sont : l'industrie manufacturière belge, le climat des affaires allemand, le sentiment des consommateurs américains, le sentiment économique des investisseurs pour l'économie allemande et sentiment des consommateurs américains et les directeurs d'achats américains (tableau 3). Par extension, nous incluons également le sentiment haussier, neutre ou baissier des marchés boursiers américains car la somme des trois réponses est égale à 100% (tableau 6) : nous pouvons savoir quel est le pourcentage dominant dans un cas où nous n'avons pas les mêmes pourcentages de réponses sur chacune des trois réponses. Pour donner un exemple de solde, prenons l'indicateur du sentiment économique des investisseurs pour l'économie allemande (tableau 3). Nous obtenons 25,7% d'opinions « s'améliorer » et 11,6% d'opinions « s'aggraver » en mars 2011. Le solde de l'enquête est égal à 14,1%, ce qui situe le résultat de cet indicateur dans une zone positive interprétée comme une zone d'optimisme.

Un autre point commun est que certains indicateurs (directeurs d'achats américains, industrie manufacturière belge) ont recours aux indices de diffusion¹⁴³. Ces indices de

¹⁴² Nous excluons de cette explication les indices de volatilité qui ne sont pas concernés par ce mode de calcul. Les indices de volatilité sont basés sur des calculs de volatilité prospective. Nous rappelons que nous utilisons ces indices de volatilité comme des *proxy* de sentiment sur les marchés financiers.

¹⁴³ Un indice de diffusion se calcule comme suit : le pourcentage des réponses positives plus la moitié du pourcentage de réponses neutres. L'idée implicite est que 50% des réponses sont des avis positifs.

diffusion prennent en considération les réponses de type « neutre ». Les indices de diffusion ont surtout la caractéristique d'être des indicateurs avancés et sont des mesures qui permettent de fournir la direction et l'ampleur d'un changement. A partir des données recueillies, des traitements de données sont appliqués pour que ces dernières deviennent exploitables dans le cadre des tests empiriques.

3. Traitements appliqués aux données

Les différentes données collectées de notre échantillon présentent des échelles et des fréquences distinctes. Ces différences ne permettent pas en l'état l'utilisation de ces données. Dans le cadre des tests, nous procédons à une homogénéisation de l'ensemble des données de l'échantillon afin de les rendre comparables entre elles. D'un point de vue du traitement des données, nous constituons un univers de taux de variation quotidiens. Nous définissons également la manière de traiter spécifiquement des données relatives aux « autres informations de marché », car ces dernières ne sont pas toujours aux normes des données homogènes (fréquence, intervalle numérique des données).

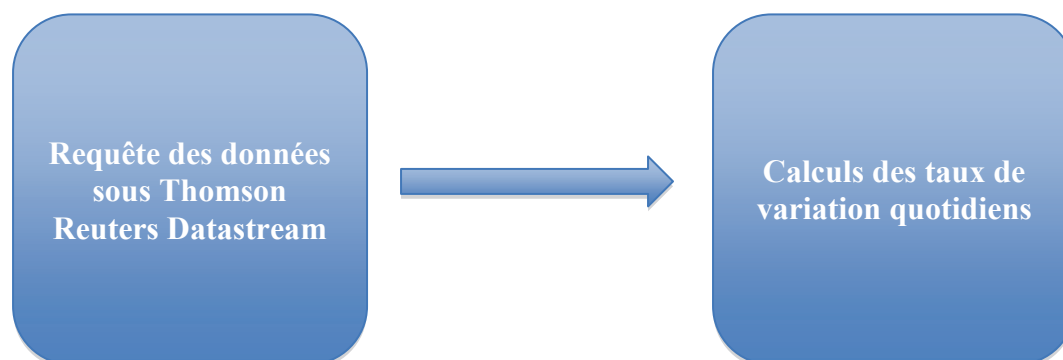
3.1. Comparaison et agrégation de données hétérogènes

Après avoir analysé le contenu de notre échantillon, nous souhaitons agréger l'ensemble de données recueillies afin qu'elles soient utilisables dans le cadre des tests empiriques¹⁴⁴. L'idée conductrice de nos traitements est de rendre ces données hétérogènes compatibles entre elles avec une fréquence quotidienne commune.

L'objectif final de ces traitements est de calculer des taux de variation quotidiens pour chacune des séries (figure 8). Ces taux de variation sont notamment nécessaires pour différentes procédures d'analyse factorielle dans ce chapitre ainsi que dans les chapitres 4 et 5. Ces taux de variation sont des paramètres d'entrée de notre modélisation (chapitre 2).

¹⁴⁴ Les tests empiriques sont menés dans les chapitres 4 et 5.

Figure 8 - Objectif final des traitements appliqués aux données de l'échantillon



Nous calculons des taux de variation quotidiens car nous souhaitons calibrer la fréquence quotidienne des données homogènes sur les données hétérogènes. Nous rappelons que nous étudions l'impact d'une combinaison de données hétérogènes sur la gestion d'un portefeuille en univers multi-classes d'actifs. Les données hors-cotation n'appartiennent pas au périmètre des données négociables (données homogènes) dans notre approche¹⁴⁵. Il est relativement « intuitif » de procéder aux traitements de données homogènes lorsque ces dernières sont obtenues en prix de clôture : il suffit de calculer un taux de variation quotidien d'un jour sur l'autre. Nous reviendrons sur ce calcul ultérieurement dans cette section pour les données homogènes et les trois indices de volatilités publiés quotidiennement (tableau 4).

Pour les « autres informations de marché »¹⁴⁶, nous rencontrons différentes échelles et différentes fréquences.

L'enjeu des traitements opérés est de rendre comparables deux univers de données dont les échelles et les fréquences de publication sont généralement différentes (tableau 6). Nous procédons au traitement (figures 9, 10, 11 et 12) d'une série en identifiant au préalable son intervalle numérique d'appartenance (tableau 6) ainsi que sa fréquence de publication. Les données homogènes (les prix d'actifs cotés) appartiennent à un intervalle unique et sont disponibles quotidiennement en prix de clôture, ce qui n'est pas toujours le cas pour les « autres informations de marché ». Par soucis de présentation et de calculs, nous segmentons les « autres informations de marché » en trois sous-groupes : les données mensuelles, hebdomadaires et quotidiennes comme l'indiquent les séries du tableau 6.

¹⁴⁵ Des produits mis en place par des institutions financières peuvent rendre négociables des phénomènes non négociables en l'état. Deux exemples : un « catastrophe bond » ou un « ETF VIX ».

¹⁴⁶ Hormis les indices de volatilité (tableau 6).

La première difficulté est de résoudre le conflit d'intervalles entre les données homogènes et certaines « autres informations de marché » pour obtenir des taux de variation quotidiens cohérents. Si la valeur d'une série croît d'une date à une autre, le taux de variation doit être positif et inversement. Les données homogènes sont toutes quotidiennes (cotation de clôture) et les cotations évoluent dans un intervalle borné entre $[0 ; +\infty]$. Le traitement consiste à calculer, à partir des données homogènes brutes, un taux de variation arithmétique $R_{i,t}$ tel que :

$$R_{i,t} = \frac{X_{i,t} - X_{i,t-1}}{X_{i,t-1}},$$

avec $X_{i,t}$ la cotation quotidienne de la série i à l'instant t (à la clôture en date t),
 $X_{i,t-1}$ la cotation quotidienne de la série i à l'instant $t-1$ (à la clôture en date $t-1$).

Figure 9 - Traitement des données de cotation

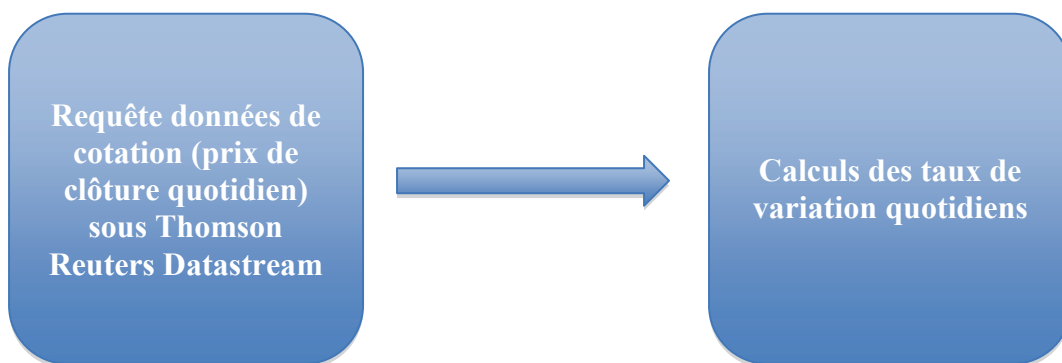


Figure 10 - Traitement des indices de volatilités

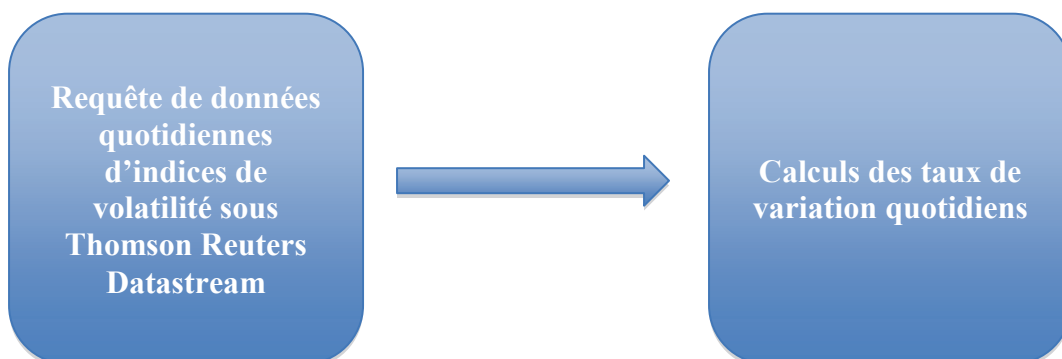


Figure 11 - Traitement des « autres informations de marché » mensuelles

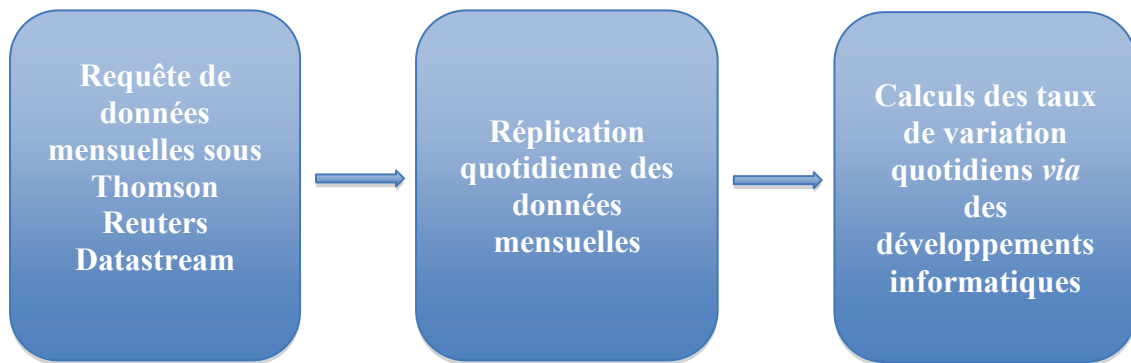


Figure 12 - Traitement des « autres informations de marché » hebdomadaires



Parmi les « autres informations de marché », seuls les indices de volatilités (tableau 6) peuvent être traités d'une manière similaire (figure 10). Certaines données présentent à la fois des fréquences et des échelles d'appartenance distinctes (tableau 6).

Tableau 6 - Disparités des fréquences et des intervalles d'appartenance des données relatives aux « autres informations de marché »

Données	Source	Fréquence de publication	Echelle
Sentiment économique des investisseurs pour l'économie allemande	Zentrum für Europäische Wirtschaftsforschung (Zew)	Mensuelle	[-100 ; +100]
Sentiment des consommateurs américains	Université du Michigan, Thomson Reuters	Mensuelle	[0 ; +100]
Climat des affaires allemand	Ifo Institute	Mensuelle	[0 ; +∞]
Industrie manufacturière belge	Banque nationale de Belgique	Mensuelle	[-100 ; +100]
Directeurs d'achat	Institute of Supply Management	Mensuelle	[0 ; +100]
Sentiment haussier des marchés boursiers américains	American Association of Individual Investors	Hebdomadaire	[0 ; +100]
Sentiment baissier des marchés boursiers américains	American Association of Individual Investors	Hebdomadaire	[0 ; +100]
Sentiment neutre des marchés boursiers américains	American Association of Individual Investors	Hebdomadaire	[0 ; +100]
Indice de volatilité VIX américain	CBOE	Quotidienne	[0 ; +∞]
Indice de volatilité VDAX allemand	Euronext	Quotidienne	[0 ; +∞]
Indice de volatilité VCAC français	Euronext	Quotidienne	[0 ; +∞]

En partant de la gauche, le type de donnée d'« autres informations de marché » est en première colonne. La source de la donnée se trouve dans la deuxième colonne. La fréquence de publication de la donnée est en troisième colonne. L'intervalle d'appartenance de la donnée est en quatrième colonne.

Le deuxième enjeu de traitements de données, est celui de l'homogénéité calendaire (temporelle) entre les données homogènes et les données d'« autres informations de marché ». Nous cherchons à obtenir des taux de variation quotidiens pour l'ensemble des données hétérogènes.

Pour l'ensemble de ces données, nous souhaitons répliquer la logique des prix de clôture « *EOD* » utilisés pour le cadre d'arbitrage statistique d'Avellaneda et Lee (2010). Notons que la plus haute fréquence de certaines données hors-cotation (les indices de volatilités) est quotidienne, tout comme les prix « *EOD* ». Ces dernières ne posent aucun problème de traitement en termes d'échelle ou de fréquence ; il convient de calculer un taux de variation $R_{i,t}$ pour les trois indices de volatilité (tableau 6). Les autres données sont mensuelles et hebdomadaires (tableau 6). Elles posent un problème d'échelle, un problème de fréquence, ou à la fois les deux. Nous analysons la résolution de ces problèmes posés dans la sous-section suivante, en soulignant la spécificité de ces « autres informations de marché ».

3.2. Traitement spécifique des données relatives aux « autres informations de marché »

Pour calculer des taux de variation quotidiens des « autres informations de marché », nous identifions les données mensuelles et les données hebdomadaires (tableau 6) afin de les traiter séparément (figures 11 et 12). L'étape qui précède les calculs de taux de variation est celle de la répétition quotidienne des valeurs de chacune de ces séries. Ce n'est qu'ensuite que nous calculons les taux de variation.

Pour les données mensuelles comme pour les données hebdomadaires, nous avons fait appel à différentes fonctions¹⁴⁷ sous Excel 2010 ainsi que différents développements informatiques sous VBA/Excel 2010 (annexes 1 et 2) pour répliquer quotidiennement les valeurs de ces séries. Les fonctions matricielles ont permis de répliquer quotidiennement des valeurs qui étaient publiées une fois par mois ou une fois par semaine. En procédant de la sorte, nous répliquons quotidiennement le taux de variation de la série et non sa valeur lors de la publication.

Le développement informatique a facilité le calcul de taux de variation quotidiens. Notons que sans l'intervention du développement informatique, le calcul de taux de variation est effectué une seule fois pour les séries mensuelles et hebdomadaires. Ceci implique que pour les autres jours du mois ou de la semaine, la variation est nulle ce qui est en contradiction avec l'esprit de notre recherche. Face à nos données de cotation, nous faisons « vivre » ces autres types de données en répliquant leurs variations respectives (mensuelles ou hebdomadaires).

Pour les séries mensuelles, nous répétons la valeur du mois tant que la nouvelle valeur n'est pas publiée. Pour les séries hebdomadaires, la logique de traitement est identique. Une fois ces données converties à une fréquence quotidienne, nous calculons les variations quotidiennes à l'aide du taux $R_{i,t}$ ¹⁴⁸. Deux exceptions se présentent au calcul du taux de variation $R_{i,t}$.

¹⁴⁷ Notamment les fonction SI(), ESTERREUR() et les fonctions matricielles telles que RECHERCHEV().

¹⁴⁸ Le même taux que celui utilisé pour les données de cotation.

Pour les séries dont l'intervalle d'appartenance est $[-100 ; +100]$ (tableau 6), nous calculons un autre taux de variation $R^*_{i,t}$:

$$R^*_{i,t} = \frac{X^*_{i,t} - X^*_{i,t-1}}{|X^*_{i,t-1}|},$$

avec $X^*_{i,t}$ la valeur de la série i à l'instant t et $|X^*_{i,t-1}|$, la valeur absolue de $X^*_{i,t-1}$.

La présence de la valeur absolue au dénominateur de cette formule est due au fait que si la valeur d'une série passe de -5 à -3, le calcul de $R_{i,t}$ est égal à -40%. Ce taux de variation $R_{i,t}$ est inexact pour nos deux exceptions puisqu'en passant de -5 à -3, nous devons obtenir un taux de variation positif. Le calcul de $R^*_{i,t}$ est égal à +40%, ce qui correspond à l'évolution exacte entre les valeurs -5 et -3. Les taux de variation peuvent être calculés et être rapprochés des taux de variation quotidiens des données homogènes (de cotation).

Nous envisageons le fait qu'une donnée mensuelle ou hebdomadaire soit répliquée de manière quotidienne. Dans un cadre d'« efficacité forte » (Fama, 1970), l'information publiée est immédiatement et totalement intégrée dans les prix des actifs. Dans ce cadre théorique, la répétition des taux de variation des séries d'« autres informations de marché » ne semble pas avoir de sens puisque le taux de variation ne vaut que pour la date de publication. Répliquer ce rendement au-delà de cette date n'apporte aucune information supplémentaire puisque cette dernière est déjà impactée dans les prix d'actifs.

La réplification quotidienne d'un taux de variation¹⁴⁹ après sa date de publication, implique que nous supposons un impact *post* publication de cette information sur les rendements de données homogènes comme le présentent certaines recherches. Brenner et al. (2009) rapportent qu'il existe des fenêtres de temps avant, pendant et après la publication d'informations « fondamentales » à l'intérieur desquelles l'impact sur les actifs est matérialisé : la volatilité et « l'activité de marché » varient à l'intérieur de ces fenêtres de temps. Les auteurs montrent que la persistance de l'information sur les rendements et la volatilité au-delà de son instant de publication est statistiquement mesurable et significatif. Ils mettent en exergue les effets *ex ante* et *ex post* de la publication d'indicateurs économiques sur les actifs financiers. La persistance de l'information sur les prix d'actifs après l'instant de publication permet également d'analyser les différences de comportement par classe d'actifs. Chollet et Cellier (2011) montrent l'impact significativement positif d'informations extra-financières¹⁵⁰

¹⁴⁹ Mensuel ou hebdomadaire issus d'une série mensuelle ou hebdomadaire.

¹⁵⁰ L'information *Corporate Social Responsibility* à travers notamment des notations de la base de données

sur les performances d'actifs financiers deux jours précédant et suivant l'annonce de ces informations extra-financières. En ce qui concerne des données économiques, Jones et al. (1998) considèrent à l'inverse que la volatilité du marché obligataire le jour de publication ne perdure pas après l'instant d'annonce. Les auteurs constatent, dans le cadre de leur étude, une incorporation instantanée de l'information aux prix d'obligation d'Etats aux Etats-Unis.

Dans le cadre des traitements de données, nous répétons quotidiennement chaque variation (mensuelle et hebdomadaire) jusqu'à ce que la donnée suivante (mois suivant et semaine suivante) de la série soit publiée. Jusqu'à ce que la nouvelle information soit disponible, nous conservons l'information publiée.

Nous pourrions alternativement considérer des données de consensus comme les prévisions d'analystes de la base IBES Thomson Reuters Datastream. Les données de consensus peuvent être sujettes à une forte volatilité due notamment à la versatilité des analystes ou à d'autres biais comportementaux. Trueman (1994) évoque le comportement « grégaire » de certains analystes financiers qui peuvent publier des prévisions de résultats proches de celles d'autres analystes alors que leurs informations « privées » leurs indiquent d'autres signaux. Williams (2013) trouve que les révisions de prévisions d'analystes financiers sont liées à la surestimation de leur ressemblance avec celles des autres analystes. Ces changements d'estimation peuvent se matérialiser par des ajustements des prévisions d'analyste. En général, les consensus IBES ne sont pas aisément accessibles, leurs coûts étant élevés. Comme le rappelle Lehman (2003), il est toujours coûteux pour un chercheur d'acquérir de l'information.

4. Composition et caractéristiques de l'échantillon

Nous abordons l'analyse factuelle et historique de notre échantillon tant sur sa structure que sur les statistiques descriptives qui permettent de mieux cerner le comportement passé de nos variables. Nous appréhendons la composition de l'échantillon sous ses différentes formes puis procédons à une analyse historique de données.

4.1. Composition de l'échantillon

Comme cela a été rappelé dans ce chapitre, la composition de notre échantillon résulte de critères non-formels (Blitz et Vliet, 2008) de sélection. Une logique et des axes de recueil sous-tendent le choix de nos données (section « Recueil des données de l'échantillon »).

Tableau 7 - Composition de l'échantillon selon le type de données

Composition de l'échantillon par type de données	
	N
Obligations	4
Matières premières	5
Actions (indices)	10
Devises	13
Indicateurs de conjoncture	4
Indicateurs de sentiment	4
Indices de volatilité	3
Total	43

En partant de la gauche, les types de données sont en première colonne. Le nombre de données est en deuxième colonne.

Nous n'avons pas déterminé *ex ante* de contraintes ni sur le nombre de types de données à recueillir, ni sur les pondérations entre les données homogènes et les données relatives aux « autres informations de marché ». Le nombre de données obtenues et les pondérations de chaque type de données sont le résultat *ex post* de notre recueil de données (tableau 7).

Nous sommes en présence de 32 types de données de cotation et de 11 indicateurs « autres

informations de marché » (tableau 7), ce qui équivaut à respectivement à 74% et 26% de l'ensemble des données.

La variable données homogènes est mesurée par des données de cotation, elles-mêmes représentées par différentes classes d'actifs : les classes « obligations », « matières premières », « actions » et « devises » (tableau 7). Les données homogènes forment un panier équipondéré de données de cotation : chacune des 32 données à la même pondération. Cependant, les pondérations agrégées par classes d'actifs ne sont pas équivalentes. Au sein des données homogènes, la classe « devises » est la plus représentée avec 41% du total des données de cotation (tableau 8). Notons que le marchés des devises est un des marchés les plus liquides (Borgy, Idier, et Le Fol, 2010). La classe « devises » et la classe « actions » équivalent, à elles deux, à plus de 70% des données de cotation recueillies (tableau 8).

Tableau 8 - Composition des données homogènes de l'échantillon par classe d'actifs

Données homogènes par classe d'actifs	% du total des données
Obligations	12%
Matières premières	16%
Acrions (indice)	31%
Devises	41%
Total	100%

En partant de la gauche, les classes d'actifs sont en première colonne. Le pourcentage de représentation de chaque classe d'actifs est en deuxième colonne.

Les données « autres informations de marché » forment également un panier équipondéré d'indicateurs mesurant l'état de la conjoncture économique, les sentiments des investisseurs et des agents économiques ainsi que la « peur », la nervosité sur les marchés financiers¹⁵¹. Chacune des 11 séries a le même poids. Lorsque nous regroupons chacune de ces 11 séries par type d'indicateur (conjoncture, sentiment, volatilité), nous sommes en présence d'une pondération relativement équilibrée (tableau 9). Ces pondérations offrent une couverture relativement exhaustive des différents types d'indicateurs pouvant influencer

¹⁵¹ Au travers d'indices de volatilité pour la peur sur les marchés financiers.

l'évolution des actifs de notre échantillon. Le poids des indicateurs de conjoncture est le même que celui des indicateurs de sentiment (tableau 9). Le poids des indices de volatilité est inférieur aux deux autres types d'indicateurs (tableau 9).

Tableau 9 - Composition des données relatives aux « autres informations de marché » de l'échantillon par type d'indicateur

Données "autres informations de marché" par type d'indicateur	% du total des données
Indicateurs de conjoncture	36%
Indicateurs de sentiment	36%
Indices de volatilité	28%
Total	100%

En partant de la gauche, le type de données relatives aux « autres informations de marché » est en première colonne. Le pourcentage de représentation de chaque type de données est en deuxième colonne.

La répartition des indicateurs par fréquence de publication montre que les séries sont majoritairement mensuelles, avec plus de 45% d'entre elles (tableau 10). Dans un contexte multi-classes d'actifs international, Andersen et al. (2007) collectent une majorité d'enquêtes économiques mensuelles. Brenner et al. (2009) recueillent également une large proportion de séries économiques mensuelles pour un univers multi-classes d'actifs aux Etats-Unis.

Les indicateurs de la zone des Etats-Unis sont en relative supériorité numérique par rapport à ceux de la zone européenne, même si l'écart reste « tempéré ». Les indicateurs américains représentent 55% du total des indicateurs alors que 45% des indicateurs restants sont de la zone Europe. L'importance des indicateurs de la zone des Etats-Unis, dans notre sélection de données, matérialise l'écho international qu'ont ces informations américaines sur le reste du monde. Selon Dahlquist et Campbell (2001) : « il n'y a nulle part où se réfugier lors que les Etats-Unis sont en crise ». D'après les auteurs, lorsque l'économie américaine est en récession les autres pays dans le monde sont également affectés. Ils montrent ainsi que les rendements des marchés actions des pays non-américains sont plus sensibles au cycle de l'économie américaine que les rendements des actions aux Etats-Unis.

Tableau 10 - Composition des données relatives aux « autres informations de marché » de l'échantillon par fréquence de publication

Données "autres informations de marché" par fréquence de publication	% du total des données
Mensuelle	46%
Hebdomadaire	27%
Quotidienne	27%
Total	100%

En partant de la gauche, le type de fréquence des données relatives aux « autres informations de marché » est en première colonne. Le pourcentage de représentation par type de fréquence est en deuxième colonne.

4.2. Caractéristiques de l'échantillon

Nous terminons l'étude de notre échantillon par un ensemble de statistiques descriptives sur l'ensemble de la période de tests ainsi que par sous-périodes. Cette segmentation permet d'analyser le comportement des données de l'échantillon sur le passé. Nous poursuivons cette analyse historique par deux analyses factorielles : une sur les données homogènes et une autre sur les « autres informations de marché ». L'ensemble de ces travaux statistiques permet de cerner les caractéristiques principales de notre échantillon de données.

4.2.1. Statistiques descriptives

L'analyse des statistiques descriptives permet de comprendre le comportement historique des données retenues pour la période de tests de l'échantillon et les sous-périodes. Nous analysons les données homogènes puis nous nous concentrons sur les données relatives aux « autres informations de marché ». Nous précisons que l'ensemble de ces statistiques est calculé à partir des données des taux de variation (section 3 « Traitement des données »).

Pour l'ensemble de la période de tests, les taux de variation quotidiens les plus élevés sont, en moyenne, ceux de la classe « matières premières » (tableau 11). Le classement des 10 premiers taux de variation les plus élevés permet de constater que : les 3 premiers appartiennent à la classe « matières premières », et que 4 appartiennent à la classe « actions ». Nous soulignons que parmi ces 10 premiers taux de variation, plusieurs zones géographiques

du monde sont représentées dont la Chine, l'Inde et l'Allemagne (tableau 11). L'inclusion de ces zones géographiques, en accord avec l'univers mondial multi-classes d'actifs, permet d'envisager que ces pays deviennent des places financières « majeures » à considérer. Notons que pour toute la période de test de l'échantillon, comme avant et après les *subprimes*, l'indice Dax (tableau 11, annexes 3 et 5) figure parmi les 10 premières données. Nous rappelons que nous avons intégré l'indice de volatilité du Dax aux données « autres informations de marché » (tableau 3).

Pendant et après la crise des *subprimes*, les 10 taux de variation les plus élevés en moyenne comptent à la fois certains composants de la classe « matières premières » et de la classe « actions » (annexes 3 et 5). Pendant la crise des *subprimes*, les 10 taux de variation les plus élevés en moyenne incluent toutes les classes d'actifs de notre échantillon. La présence de toutes les classes d'actifs laisse penser que la crise des *subprimes* a impacté l'ensemble des classes et que la hiérarchie des taux de variation des cotations, hors-périodes des *subprimes*, n'est pas conservée. La hiérarchie avant et après la crise des *subprimes* est modifiée : l'or est en deuxième position et la majorité des indices actions ne sont plus présents (annexe 4). Le prix du blé peut être considéré comme un indicateur avancé d'inflation et l'or comme une valeur refuge en temps de crise (annexe 4).

Pour toute la période de tests de l'échantillon, les 10 écarts-types les plus élevés intègrent des composants de la classe « matières premières » et de la classe « actions » (tableau 11). En l'occurrence, Gorton et Rouwenhorst (2006) évoquent le fait que les matières premières ont une saisonnalité et une volatilité « prononcées ». Les 5 premiers écart-types sont exclusivement ceux de la classe « matières premières » (tableau 11). En dixième position, nous trouvons le Cac 40 (tableau 11) : rappelons que nous avons intégré l'indice de volatilité VCAC dans la sélection des données relatives aux « autres informations de marché ». La couverture géographique de ces 10 premières données intègre l'Amérique du Nord, l'Amérique du Sud, l'Asie et l'Europe.

Pour l'ensemble de la période de tests de l'échantillon comme pour les trois sous-périodes, la volatilité des taux de variation des « matières premières » est toujours parmi les 10 premières (tableau 11, annexes 3, 4 et 5). Ces 10 premières positions incluent également l'univers « actions » mais la hiérarchie change selon les périodes (annexes 3, 4 et 5). Les volatilités du blé et du *west texas intermediate* (pétrole) qui étaient les deux premières avant la crise des *subprimes* (annexe 3), ont augmenté pendant la crise des *subprimes*, puis baissé après celle-ci. *A posteriori*, ces indicateurs avancés d'inflation communiquent des signaux de crise.

Tableau 11 - Statistiques descriptives des taux de variation des données homogènes pour la période de tests de l'échantillon (2006-2011)

	Moyenne	Ecart-type	Médiane	Minimum	Maximum
France 2004 4% 25/04/14	0,0045%	0,2431%	0,0000%	-0,9979%	1,4801%
US 2004 4% 15/02/14	0,0084%	0,7596%	-0,0019%	-5,0138%	3,7007%
Japan 2004 1,5% 20/03/14	0,0293%	0,9527%	-0,0480%	-4,1565%	7,4929%
Brazil 2004 10,5% 14/07/14	-0,0003%	0,7472%	-0,0039%	-7,8651%	5,6262%
Brent (pétrole)	0,0517%	2,1473%	0,0424%	-10,7920%	16,9479%
West Texas Intermediate (pétrole)	0,0546%	2,6834%	0,0000%	-13,9394%	22,1112%
Or	0,0681%	1,2850%	0,0818%	-7,4416%	6,4977%
Blé	0,1161%	3,2087%	0,0000%	-20,3704%	11,2299%
Maïs	0,1157%	2,3684%	0,0000%	-10,3806%	11,3514%
Bovespa (Brésil)	0,0732%	2,0460%	0,0430%	-11,3931%	14,6592%
Bse sensex (Inde)	0,0642%	1,8862%	0,0627%	-10,9564%	17,3393%
Cac 40 (France)	-0,0026%	1,6631%	0,0000%	-9,0368%	11,1762%
Dax (Allemagne)	0,0319%	1,5611%	0,0688%	-7,1639%	11,4020%
Ftse 100 (Royaume Uni)	0,0125%	1,4791%	0,0000%	-8,8493%	9,8388%
S&P 500 (Etats-Unis)	0,0168%	1,5906%	0,0732%	-9,0350%	11,5800%
Nasdaq 100 (Etats-Unis)	0,0498%	1,6685%	0,1014%	-10,5195%	12,5799%
Asx 200 (Australie)	0,0038%	1,3617%	0,0000%	-8,3362%	5,7896%
Shanghai SE (Chine)	0,0626%	2,0116%	0,0932%	-8,8449%	9,4537%
Nikkei 225 (Japon)	-0,0149%	1,7574%	0,0000%	-11,4064%	14,1503%
Dollar américain / euro	-0,0045%	0,6828%	-0,0079%	-4,5122%	3,9189%
Sterling / euro	-0,0159%	0,5949%	-0,0019%	-2,8650%	3,2233%
Franc suisse / euro	0,0165%	0,4351%	0,0000%	-2,7670%	2,0513%
Yen / euro	0,0244%	0,9723%	-0,0269%	-7,7777%	6,6571%
Yuan / euro	0,0120%	0,6921%	0,0000%	-3,4515%	2,6159%
Couronne norvégienne / euro	0,0040%	0,5484%	0,0127%	-3,5479%	4,0341%
Real brésilien / euro	0,0221%	1,1130%	0,0458%	-5,7295%	7,2875%
Roupie indienne / euro	-0,0028%	0,6739%	0,0000%	-3,4283%	3,8984%
Won coréen / euro	-0,0143%	1,1249%	-0,0159%	-9,9360%	12,0678%
Dollar australien / euro	0,0217%	0,7803%	0,0404%	-5,2175%	6,1638%
Rand sud africain / euro	0,0019%	1,1219%	0,0306%	-13,8904%	6,2513%
Dollar néo-zélandais / euro	0,0137%	0,8250%	0,0296%	-4,6993%	3,7114%
Dollar canadien / euro	0,0066%	0,7096%	0,0000%	-3,2444%	4,5696%

En partant de la gauche, les données homogènes sont en première colonne. De la deuxième à la sixième colonne, nous avons la moyenne, l'écart-type, la médiane, le minimum et le maximum des taux de variation quotidiens.

La classe « actions » a le plus fortement subi l'évènement *subprimes* (annexe 4). Les prémices des problèmes sur les crédits hypothécaires, qui n'ont *a priori* pas de lien direct avec les actions, ont engendré un choc sur les marchés actions (Khandani et Lo, 2007). Bien que les marchés actions ne soient pas les plus imposants en termes de volumes de transactions, ces derniers peuvent devenir un baromètre du comportement de l'ensemble des marchés financiers.

Tableau 12 - Statistiques descriptives des taux de variation des données relatives aux « autres informations de marché » pour la période de tests de l'échantillon (2006-2011)

Données	Source	Moyenne	Ecart-type	Médiane	Minimum	Maximum
Sentiment économique des investisseurs pour l'économie allemande	Zentrum für Europäische Wirtschaftsforschung (Zew)	10,96%	106,80%	-1,54%	-296,43%	471,43%
Sentiment des consommateurs américains	Université du Michigan, Thomson Reuters	0,21%	7,23%	-0,29%	-19,13%	17,94%
Climat des affaires allemand	Ifo Institute	0,09%	1,86%	0,27%	-5,37%	4,33%
Industrie manufacturière belge	Banque nationale de Belgique	-8,01%	103,04%	5,66%	-437,50%	350,00%
Directeurs d'achat	Institute of Supply Management	0,32%	4,35%	0,20%	-13,17%	10,56%
Sentiment haussier des marchés boursiers américains	American Association of Individual Investors	2,49%	23,17%	-2,23%	-43,26%	102,92%
Sentiment baissier des marchés boursiers américains	American Association of Individual Investors	2,65%	20,70%	1,86%	-43,52%	77,50%
Sentiment neutre des marchés boursiers américains	American Association of Individual Investors	1,46%	24,99%	-2,02%	-49,45%	150,98%
Indice de volatilité VIX américain	CBOE	0,27%	7,21%	-0,49%	-29,57%	64,22%
Indice de volatilité VDAX allemand	Euronext	0,15%	5,51%	-0,11%	-23,69%	34,96%
Indice de volatilité VCAC français	Euronext	0,24%	6,91%	-0,05%	-31,05%	62,75%

En partant de la gauche, le type de donnée d'« autres informations de marché » est en première colonne. La source de la donnée est en deuxième colonne. De la troisième à la septième colonne, nous avons la moyenne, l'écart-type, la médiane, le minimum et le maximum des taux de variation quotidiens.

Pour l'ensemble de la période de tests de l'échantillon et les trois sous-périodes de tests (avant, pendant et après la crise des *subprimes*), le sentiment haussier et baissier des marchés boursiers américains figurent parmi les 4 taux de variation les plus élevés en moyenne (tableau 12, annexes 6, 7 et 8).

Avant la crise des *subprimes*, les 4 taux de variation les plus élevés en moyenne comptent uniquement des indicateurs de sentiment aux Etats-Unis (annexe 4). Ces 4 indicateurs sont des indicateurs de sentiment des marchés boursiers et de sentiment des consommateurs : la sphère financière comme celle de l'économie « réelle » sont représentées. Avant cette crise, la variation la plus élevée est celle du sentiment neutre des marchés boursiers américains. Ce chiffre peut être considéré comme une information d'incertitude

avant un contexte de crise.

Pendant la crise des *subprimes*, les trois types d'indicateurs (conjoncture, sentiment, volatilité) sont représentés parmi les 4 taux de variation les plus élevés en moyenne (annexe 7). De plus, la hiérarchie *pré-crise* des *subprimes* est modifiée (annexes 6 et 7). Nous effectuons une analyse similaire à celle proposée pour les données « homogènes » en période de crise des *subprimes* : tous les indicateurs sont représentés tout comme l'étaient toutes les classes d'actifs (annexes 4 et 7). Pendant la crise des *subprimes*, le sentiment baissier des marchés boursiers américains est en deuxième position¹⁵² alors qu'avant crise il occupait la quatrième position. Nous remarquons la remontée du sentiment baissier en période de crise des *subprimes*.

Après la crise des *subprimes*, deux indicateurs de conjoncture sont les deux premiers taux de variation et le troisième est le sentiment haussier des marchés boursiers américains. Cette nouvelle hiérarchie laisse penser que des signaux favorables de l'économie et des investisseurs témoignent d'une sortie possible de crise des *subprimes*.

Pour l'ensemble de la période de tests de l'échantillon comme pour les trois sous-périodes, quatre niveaux d'écart-types les plus élevés intègrent des indicateurs de conjoncture et de sentiments (tableaux 12, 3, annexes 6, 7 et 8). Andersen et al. (2007) constatent que les enquêtes économiques peuvent avoir une volatilité élevée avant toutes réductions¹⁵³. Les écart-types élevés des indicateurs de conjoncture européens sont en partie dus aux traitements des données¹⁵⁴ qui font que des oscillations entre des valeurs négatives et positives entraînent de fortes variations.

Avant et pendant la crise des *subprimes*, le sentiment neutre des marchés boursiers américains est en troisième position des écart-types les plus élevés (annexe 6 et 7). Après la crise des *subprimes*, c'est le sentiment haussier des marchés boursiers américains qui se trouve en troisième position (annexe 8). Le sentiment baissier des marchés boursiers américains est en quatrième position avant la crise des *subprimes*. Après la crise des *subprimes*, cet indicateur retrouve la cinquième position et son niveau de volatilité est inférieur à celui d'avant crise (annexe 6 et 8).

En analysant la moyenne des valeurs publiées¹⁵⁵ sur l'ensemble de la période de tests de l'échantillon (tableau 13), nous pouvons constater l'évolution moyenne des séries. Cette

¹⁵² Des taux de variation les plus élevés en moyenne sur la période des *subprimes*.

¹⁵³ « Standardisation » des données.

¹⁵⁴ Voir la section 3 « Traitements appliqués aux données ».

¹⁵⁵ Et non pas les taux de variation des valeurs publiées.

moyenne est effectuée pour une période de tests qui inclut notamment une période de crise des *subprimes*. En moyenne pour l'ensemble de la période de tests de l'échantillon, la plupart des indicateurs publiés en l'état décrivent un cadre « neutre » voire légèrement « pessimiste » du contexte économique et financier pour toute la période de tests de l'échantillon (tableau 13). La seule exception est la zone relativement « optimiste » dans laquelle se situe le sentiment des consommateurs américains (tableau 13). Les indicateurs de conjoncture publient des opinions plus ou moins négatives (industrie manufacturière belge et sentiments économiques des investisseurs pour l'économie allemande) et relativement « neutre » (climat des affaires allemand et directeurs d'achat américains). Les indicateurs de sentiment rapportent des informations faiblement pessimistes sauf pour le sentiment des consommateurs américains qui est plutôt dans une zone « optimiste » (tableau 13). Les indices de volatilités sont en dessous des niveaux de « stress »¹⁵⁶ des marchés boursiers, bien que ces derniers ne soient pas non plus à des niveaux de faibles volatilités (tableau 13).

Tableau 13 - Statistiques descriptives des valeurs des données relatives aux « autres informations de marché » pour la période de tests de l'échantillon (2006-2011)

Données	Source	Moyenne
Sentiment économique des investisseurs pour l'économie allemande	Zentrum für Europäische Wirtschaftsforschung (Zew)	-0,27
Sentiment des consommateurs américains	Université du Michigan, Thomson Reuters	67,19
Climat des affaires allemand	Ifo Institute	103,77
Industrie manufacturière belge	Banque nationale de Belgique	-7,65
Directeurs d'achat	Institute of Supply Management	50,67
Sentiment haussier des marchés boursiers américains	American Association of Individual Investors	38,82
Sentiment baissier des marchés boursiers américains	American Association of Individual Investors	39,00
Sentiment neutre des marchés boursiers américains	American Association of Individual Investors	22,19
Indice de volatilité VIX américain	CBOE	24,30
Indice de volatilité VDAX allemand	Euronext	23,30
Indice de volatilité VCAC français	Euronext	25,34

En partant de la gauche, le type de donnée d'« autres informations de marché » est en première colonne. La source de la donnée est en deuxième colonne. La moyenne quotidienne est en troisième colonne.

¹⁵⁶ Au-dessus de 30, le niveau du VIX peut être perçu comme un proxy du « stress » sur les marchés financiers. Munenzon (2010) analyse notamment les niveaux du VIX qui dépassent le seuil de 30.

4.2.2. Analyse des données de l'échantillon

Suite à l'analyse des statistiques descriptives des données de notre échantillon, nous poursuivons notre travail d'introspection par deux analyses factorielles sur l'ensemble de la période de l'échantillon (2004-2011). Nous procédons à cette analyse sur la période totale de l'échantillon (période d'« apprentissage » plus période de tests) afin de prendre en considération l'ensemble du passé disponible.

La première analyse de données consiste à effectuer une analyse en composantes principales sur les taux de variation des données homogènes. La deuxième analyse factorielle permet de réaliser les mêmes travaux sur les taux de variation des « autres informations de marché ». L'ACP sur l'univers de données homogènes permet de calculer le premier « portefeuille propre »¹⁵⁷ (tableau 14), qui est l'équivalent d'un premier vecteur propre normalisé par les volatilités. Dans le tableau 14, les poids associés à chaque actif correspondent au poids du premier « portefeuille propre »¹⁵⁸. Nous rappelons que la variance la plus élevée est portée par le premier « portefeuille propre »¹⁵⁹.

Ce premier « portefeuille propre » permet de considérer que, d'après nos données homogènes sur le passé, des signes contraires (+ et -) de pondérations existent. Ceci laisse présager que des arbitrages entre classes d'actifs et entre actifs sont possibles (tableau 14).

¹⁵⁷ Voir le calcul du « portefeuille propre » dans le chapitre 2.

¹⁵⁸ Dont la formule du rendement est précisée dans le chapitre 2.

¹⁵⁹ Vecteur propre usuellement.

Tableau 14 - Analyse en composantes principales des taux de variation des données homogènes pour la période totale de l'échantillon (2004-2011)

Données	Premier portefeuille propre
France 2004 4% 25/04/14	50,51
US 2004 4% 15/02/14	29,62
Japan 2004 1,5% 20/03/14	31,72
Brazil 2004 10,5% 14/07/14	11,75
Brent (pétrole)	-5,49
West Texas Intermediate (pétrole)	-4,69
Or	2,71
Blé	-1,83
Maïs	-3,43
Bovespa (Brésil)	-13,72
Bse sensex (Inde)	-10,92
Cac 40 (France)	-21,38
Dax (Allemagne)	-21,99
Ftse 100 (Royaume Uni)	-23,92
S&P 500 (Etats-Unis)	-18,73
Nasdaq 100 (Etats-Unis)	-15,80
Asx 200 (Australie)	-14,11
Shanghai SE (Chine)	-4,83
Nikkei 225 (Japon)	-11,04
Dollar américain / euro	24,81
Sterling / euro	0,42
Franc suisse / euro	51,09
Yen / euro	28,84
Yuan / euro	21,00
Couronne norvégienne / euro	-22,85
Real brésilien / euro	-11,73
Roupie indienne / euro	9,73
Won coréen / euro	-3,51
Dollar australien / euro	-15,97
Rand sud africain / euro	-14,83
Dollar néo-zélandais / euro	-17,54
Dollar canadien / euro	-11,61

En partant de la gauche, la donnée homogène est en première colonne. Pour ce premier facteur propre, les poids de chaque actif appartenant au premier vecteur propre sont en deuxième colonne.

Nous obtenons globalement des signes positifs pour les « obligations » alors que nous constatons des signes négatifs pour certains composants des classes « matières premières », « actions » et « devises » (tableau 14). Entre actifs de la même classe, des signes contraires (+ et -) laissent penser que des arbitrages intra-classe peuvent également se produire (tableau 14). Au sein de la classe « matières premières », nous obtenons un signe positif pour l'or alors que des signes négatifs sont constatés pour les autres actifs de la même classe. Ce phénomène de signes contraires (+ et -) au sein de la classe « devises » est également constaté. Seule la classe « actions » ne suit pas cette dynamique car tous les indices sont représentés par des signes négatifs (tableau 14).

Une analyse similaire peut être faite pour le calcul du premier vecteur propre des « autres informations de marché » (tableau 15). Nous n'envisageons pas d'arbitrage comme cela est mentionné pour les données « homogènes » car les séries d'« autres informations de marché » ne sont pas en l'état négociable¹⁶⁰. Les trois indices de volatilité ainsi que l'indicateur de sentiment haussier des marchés boursiers américains (signes -) n'ont pas les mêmes pondérations que les autres indicateurs de sentiment et de conjoncture (signes +). Il semble que l'ensemble de ces indicateurs ne suivent pas la même direction. Nous souhaitons recueillir une certaine hétérogénéité dans les informations complémentaires aux données homogènes.

¹⁶⁰ Sauf pour certains indices de volatilité. D'autre part, l'ingénierie financière permet de rendre négociable des éléments qui ne le sont pas en l'état.

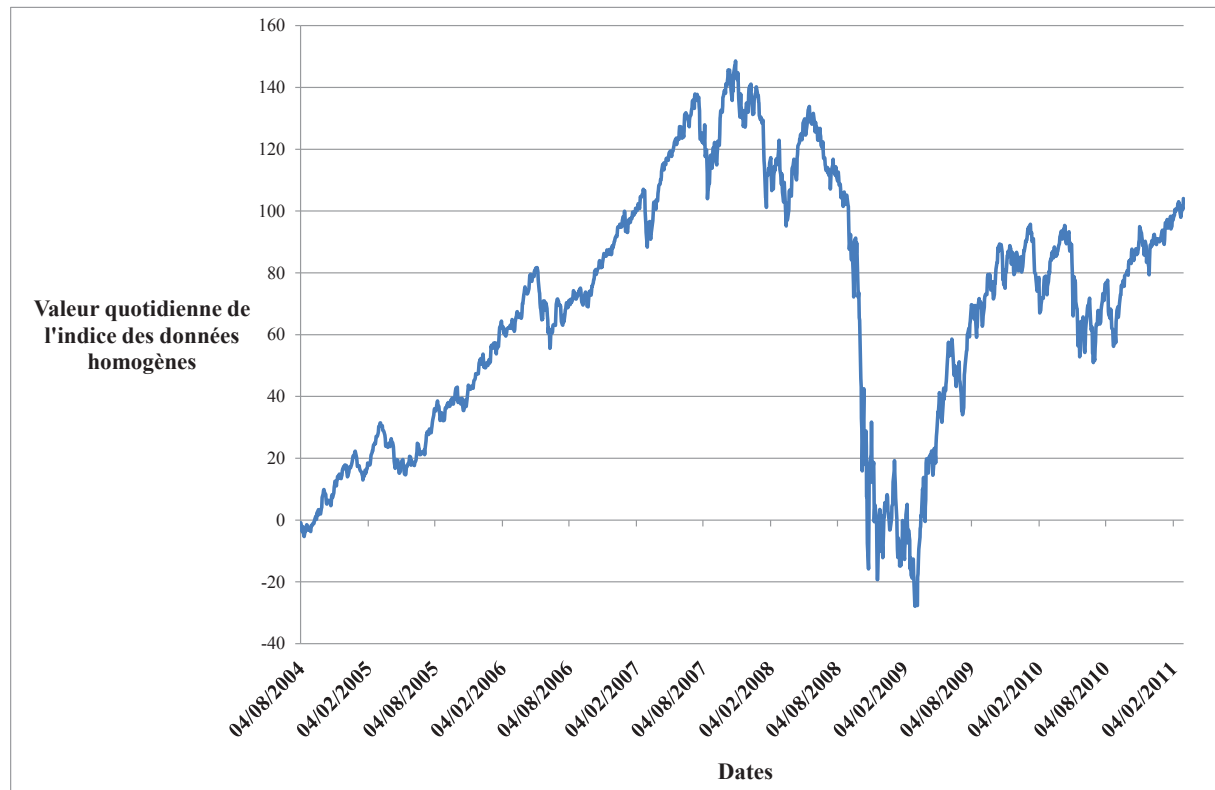
Tableau 15 - Analyse en composantes principales des taux de variation des données relatives aux « autres informations de marché » pour la période totale de l'échantillon (2004-2011)

Données	Premier vecteur propre
Sentiment économique des investisseurs pour l'économie allemande	0,13
Sentiment des consommateurs américains	0,95
Climat des affaires allemand	9,23
Industrie manufacturière belge	0,10
Directeurs d'achat	4,48
Sentiment haussier des marchés boursiers américains	-0,14
Sentiment baissier des marchés boursiers américains	0,25
Sentiment neutre des marchés boursiers américains	0,04
Indice de volatilité VIX américain	-7,06
Indice de volatilité VDAX allemand	-11,36
Indice de volatilité VCAC français	-9,11

En partant de la gauche, le type de donnée d'« autres informations de marché » est en première colonne. La pondération du premier vecteur propre pour chaque donnée est en deuxième colonne.

Nous terminons notre analyse de données homogènes en constituant un portefeuille de « marché » tel que le premier « portefeuille propre ». Nous calculons la valorisation quotidienne de ce portefeuille de « marché » (figure 13), censé répliquer statistiquement le marché constitué des données homogènes recueillies. Qualitativement, nous constatons que l'indice de « marché » suit globalement les phases de hausses et de baisses de la crise des *subprimes*. La première phase de baisse de l'indice se produit le 3 juillet 2007 et le relèvement de l'indice dans une valorisation strictement positive se produit à partir du 19 mars 2009.

Figure 13 - Valeur quotidienne du premier facteur de risque issu d'une analyse en composantes principales (ACP) des données homogènes pour la période totale de l'échantillon (2004-2011)

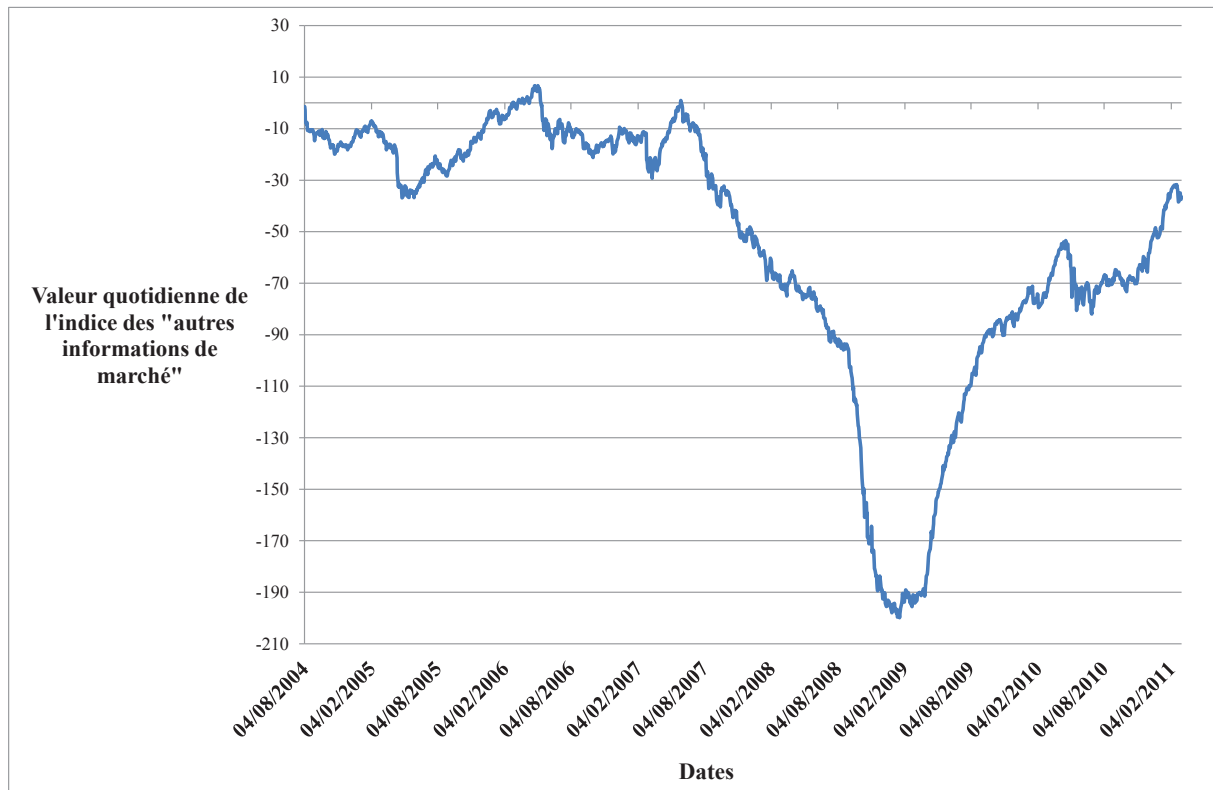


Les dates quotidiennes sont en abscisses et la valeur quotidienne du portefeuille telle que l'ACP nous la suggère est en ordonnée. La valorisation est calculée du 04/08/2004 au 03/03/2011.

Pour la période de tests (2006-2011), la rentabilité annualisée de l'indice de « marché » est de 11,36% pour une volatilité annualisée de l'indice atteignant les 38,9% soit un ratio rentabilité/volatilité égal à 0,29. Nous reviendrons dans les chapitres 4 et 5 sur ces résultats pour analyser l'impact de l'introduction d'« autres informations de marché » sur la volatilité et le ratio rentabilité/volatilité des portefeuilles empiriques.

Sur le même principe, nous avons également calculé un indice synthétique des données relatives aux « autres informations de marché » avec une publication quotidienne (figure 14). Nous constatons une baisse de la valeur de cet indice sur une période de temps relativement similaire que celle de l'indice de « marché ». Précisément, le début de la baisse de l'indice est le 4 juin 2007 et la reprise définitive commence le 31 mars 2009.

Figure 14 - Valeur quotidienne du premier facteur de risque issu d'une analyse en composantes principales (ACP) des données relatives « autres informations de marché » pour la période totale de l'échantillon (2004-2011)



Les dates quotidiennes sont en abscisses et la valeur quotidienne du portefeuille telle que l'ACP nous la suggère est en ordonnée. La valorisation est calculée du 04/08/2004 au 03/03/2011.

Le chapitre 3 nous permet de définir les périodes d'« apprentissage » et de tests de notre échantillon. Nous présentons et analysons les différentes données recueillies en expliquant la sélection effectuée par type de données. Nous détaillons ensuite le processus de traitements des données recueillies en soulignant les spécificités de traitements qu'entraînent les données relatives aux « autres informations de marché ». Nous explorons finalement la composition et les caractéristiques de notre échantillon. Cette dernière analyse historique est réalisée au travers de statistiques descriptives et d'analyses factorielles pour les données homogènes et les données hétérogènes. Dans le chapitre suivant, des tests empiriques sont menés dans un cadre de diversification « naïve ». Ces différents tests mesurent l'impact de la combinaison de données hétérogènes sur les caractéristiques d'un portefeuille d'arbitrage multi-classes d'actifs.

CHAPITRE 4. PORTEFEUILLE D'ARBITRAGE « EURO NEUTRE » EN DIVERSIFICATION « NAÏVE »

Dans notre quatrième chapitre, nous effectuons des tests empiriques sur les données de notre échantillon¹⁶¹ dans un cadre de diversification « naïve » (DeMiguel, Garlappi, et Uppal, 2009) de portefeuilles d'arbitrage « euro neutre ». Par diversification « naïve »¹⁶², nous faisons référence à un portefeuille équipondéré de type « 1/N ». Nous identifions et mesurons les éléments d'influence des données hétérogènes sur les portefeuilles d'arbitrage. Notre étude empirique est complétée par différents tests de robustesse permettant de vérifier la stabilité des résultats de nos recherches. Dans l'ensemble de ce quatrième chapitre, nous conservons une allocation « naïve ». Ultérieurement, dans le cadre du chapitre suivant, la diversification optimale sera abordée.

1. Influence des données hétérogènes sur des portefeuilles d'arbitrage « euro neutre » en allocation « naïve »

Nous choisissons de composer des portefeuilles d'arbitrage « euro neutre » dans la continuité de stratégies historiques d'arbitrage statistique (Avellaneda et Lee, 2010). Cette « euro neutralité » permet notamment de constituer des portefeuilles qui ne sont biaisés ni en *long*¹⁶³ ni en *short*¹⁶⁴. La hausse ou la baisse de la valeur des portefeuilles ne peut pas être imputée à une sur ou sous-exposition aux marchés financiers de notre univers de données. Nous choisissons d'appliquer une diversification « naïve » au sein d'une équipartition de portefeuille. Le concept d'équi-répartition est le suivant : chaque euro de rentabilité généré est réparti à 50% sur des positions *long* et à 50% (restant) sur des positions *short*. La totalité de la somme obtenue est également investie d'un jour sur l'autre : nous ne conservons pas une partie de la rentabilité. Nous sommes dans le cadre d'un portefeuille « faisable » ou « réalisable » (Riva, 2008) où la totalité de la rentabilité est réinvestie d'un jour à l'autre. C'est au sein de cette équi-répartition que nous appliquons la diversification « naïve » des poids du portefeuille. Le choix de l'allocation « 1/N » est une manière « naïve », sans *a*

¹⁶¹ Voir chapitre 3.

¹⁶² Autrement appelée diversification « 1/N ».

¹⁶³ Biais acheteur du portefeuille.

¹⁶⁴ Biais vendeur du portefeuille.

priori, d'attribuer des poids aux portefeuilles d'arbitrage. Par « naïveté », la construction de portefeuille consiste en l'attribution du même poids à l'ensemble des actifs d'un portefeuille. Nous déterminons une absence de leviers dans le cadre des tests empiriques. L'effet de levier peut amplifier la rentabilité et la volatilité des portefeuilles d'arbitrage or nous souhaitons également éviter ce biais dans le cadre de notre recherche.

Nos tests sur le passé ont été réalisés avec des frais de transaction globaux, et calculés à un même taux pour toutes les classes d'actifs. D'une part, le choix d'un taux de frais unique est justifié par un souci de simplicité. Cependant, la détermination d'un taux de transaction pour chaque classe d'actifs est empiriquement et théoriquement un enjeu de recherche car même au sein d'une classe d'actifs les frais peuvent être très variables¹⁶⁵. D'autre part, la réglementation sur les frais de transaction évoluant constamment, nous optons pour une vision générique du taux de frais de transactions pour notre univers multi-classes d'actifs qui inclut l'ensemble des frais liés aux transactions réalisées, dont les frais de courtage et les frais de gestion. Dans le cadre des tests empiriques, la rentabilité calculée est considéré comme nette de frais. Nos portefeuilles d'arbitrage sont toujours : « euro neutre », diversifiés de manière « naïve » au sein d'une équi-répartition des montants investis, sans levier et avec un taux de transactions de 3% annualisé.

L'alternance des allocations est réalisée dans un cadre d'arbitrage systématique proche de l'univers de certains fonds alternatifs ou *hedge funds*¹⁶⁶. Le but des *hedge funds* est de réaliser une performance indépendante de tout indicateur de référence (*benchmark*)¹⁶⁷. Ces *hedge funds* ou fonds alternatifs, cherchent à dégager une performance indépendamment de l'évolution du marché en utilisant des positions *long* et *short* sur des « supports » constitués d'« actifs financiers usuels » (Monarcha, 2008).

Les résultats des tests empiriques ne sont pas comparés à un indicateur de référence ou *benchmark*. Les difficultés d'identifier et de détenir un portefeuille de « marché » sont des enjeux contemporains de la recherche en Finance. Xu et Malkiel (2004) évoquent notamment l'incapacité pour des investisseurs de détenir le portefeuille de marché. Dans un univers de gestion alternative, Ackermann et al. (1999) notent également un problème d'identification d'un « bon benchmark ». Dans un cadre multi-classes d'actifs, Black et Litterman (1992)

¹⁶⁵ Nous ajoutons que les stratégies d'arbitrage statistique sont souvent mise en œuvre par des automates de *trading*. Le taux de courtage de ces plateformes est proportionnel aux volumes négociés.

¹⁶⁶ Chan et al. (2005) font référence au journal spécialisé *the Journal of Alternative Investments*, crée après les nombreuses recherches publiées sur la gestion alternative. Plusieurs travaux se sont concentrés sur de larges études empiriques des performances historiques des *hedge funds* (Ackermann, McEnally, et Ravenscraft, 1999 ; Fung et Hsieh, 1997), évoquant le faible « faibles » bêta des *hedge funds*.

¹⁶⁷ Capocci (2008) propose une revue de littérature sur les *hedge funds*.

soulignent également cette difficulté d'identification.

L'ensemble des tests empiriques est réalisé à l'aide d'une programmation informatique sous Scilab¹⁶⁸. Nous développons plusieurs algorithmes adaptés à la gestion d'un portefeuille d'arbitrage. Afin d'estimer si les « autres informations de marché » exercent une influence sur les portefeuilles d'arbitrage, nous considérons deux univers : un univers de données incluant des données de cotation et d'« autres informations de marché » (données hétérogènes) et un deuxième ne considérant que des données de cotation (données homogènes).

Nous souhaitons comprendre l'évolution de la structure des portefeuilles d'arbitrage en analysant leurs rentabilités, volatilités, allocations d'actifs et contributions des classes d'actifs à leurs rentabilités. Ces analyses sont menées en présence ou non d'« autres informations de marché » afin de mesurer l'impact d'une combinaison de données hétérogènes sur les portefeuilles d'arbitrage¹⁶⁹. Des tests de robustesse sont effectués pour valider les résultats empiriques obtenus dans ce chapitre.

1.1. Modification de l'allocation entre classes d'actifs

Avant d'évaluer la modification d'allocation d'actifs induite par la combinaison de données hétérogènes, nous précisons le calcul des pondérations des portefeuilles d'arbitrage. Dans notre recherche, le processus d'allocation d'actifs pour des poids *long* est défini par la relation suivante :

$$Q^{Long}_{i,t} = \frac{E_t}{n_L},$$

avec $Q^{Long}_{i,t}$, le poids *long* de l'actif i à l'instant t dans le portefeuille¹⁷⁰,

E_t , la valorisation du portefeuille à la date t

et n_L , le nombre de positions *long* du portefeuille d'arbitrage.

¹⁶⁸ Voir www.scilab.org. Il s'agit d'un logiciel de calcul scientifique développé par l'INRIA et l'ENPC.

¹⁶⁹ Dont les caractéristiques ont été définies précédemment.

¹⁷⁰ Ce poids peut être exprimé en euros.

Pour les poids *short*, nous réécrivons la relation ci-dessous quasiment à l'identique :

$$Q^{short}_{i,t} = - \frac{E_t}{n_S},$$

avec $Q^{short}_{i,t}$, le poids *short* de l'actif i à l'instant t dans le portefeuille¹⁷¹,

E_t , la valorisation du portefeuille à la date t

et n_S , le nombre de positions *short* du portefeuille d'arbitrage. Le caractère « euro neutre » signifie que la somme des poids est nulle à l'instant t :

$$\sum_{i=1}^{32} Q_{i,t} = 0.$$

L'indice i évolue de 1 à 32 car nous avons 32 données homogènes négociables dans notre échantillon.

L'influence des « autres informations de marché » sur les portefeuilles d'arbitrage se manifeste par un changement d'allocation sur certaines classes d'actifs (tableau 16). En présence de données « hétérogènes », les pondérations sur les classes « obligations » et « matières premières » évoluent par rapport à un univers de données « homogènes » (tableau 16). Soulignons que l'allocation d'actifs représente globalement l'importance « relative » des actifs parmi l'ensemble des 32 données de cotation recueillies (tableau 8)¹⁷². L'introduction des données hétérogènes dans notre analyse modifie cette allocation, mais d'un point de vue global, la photographie (vue statique) de l'allocation ressemble à celle des données recueillies (voir la section 4.1 « Composition de l'échantillon » dans le chapitre 3).

¹⁷¹ Ce poids peut être exprimé en euros.

¹⁷² Voir le chapitre 3.

Tableau 16 - Comparaison de l'allocation d'actifs d'un portefeuille d'arbitrage selon le type de données pour la période de tests (2006-2011)

Type de données	Classe d'actifs				Total
	Obligations	Matières premières	Actions	Devises	
Données hétérogènes	7%	20%	32%	41%	100%
Données homogènes	13%	16%	32%	40%	100%

En partant de la gauche, le type de données est en première colonne. L'allocation d'actifs par classe d'actifs est renseignée de la deuxième à la cinquième colonne. Le total des pondérations est en sixième colonne.

En introduisant de l'hétérogénéité dans l'analyse statistique de nos données, nous sous-pondérons les « obligations » de 6%¹⁷³ (tableau 16) et surpondérons les « matières premières » de 4% (tableau 16) pour la période de tests (2006-2011). Pour la classe d'actifs « devises », l'allocation ne varie que très « légèrement » au regard des changements obtenus pour les classes « obligations » et « matières premières ». En l'espèce, nous obtenons une surpondération de 1% pour les « devises ». L'allocation sur la classe « actions » ne varie pas en introduisant des données « hétérogènes » dans le cadre des tests empiriques.

Empiriquement, le choix des données de cotation qui entrent dans la composition du portefeuille varie selon que nous décidons d'inclure ou non d'« autres informations de marché » dans notre analyse de données¹⁷⁴. Cette étape d'analyse de données est préalable à la détection de signaux d'investissement¹⁷⁵. Les modifications d'allocations semblent être caractérisées par des changements de décisions sur la répartition des fonds (entre les quatre classes d'actifs) en présence d'une analyse de données « hétérogènes ».

Suite à une étude empirique sur l'allocation tactique multi-classes d'actifs, incluant des actions et des obligations, Blitz et Vliet (2008) expliquent que les marchés peuvent être « macro inefficients » car les montants de « *smart money* » sont insuffisamment alloués sur les arbitrages potentiels, eux-mêmes issus d'effets comportementaux. Ces effets comportementaux, qui peuvent entraîner ces situations d'arbitrage, sont : le fait de ne pas s'éloigner de l'investissement stratégique de long terme¹⁷⁶, le fait d'avoir une allocation fixe et de pas vouloir s'en éloigner, le comportement grégaire (ne pas s'éloigner des *peers*) ou le fait de vouloir rester investi sur une classe d'actifs qui a récemment eu une performance

¹⁷³ A partir du tableau 16, le calcul suivant est effectué : 13% - 7% = 6%.

¹⁷⁴ Voir le chapitre 2, où toutes les indications sur le processus de *market timing* sont renseignées. Notre analyse de données peut inclure ou non des données hétérogènes. Si l'analyse factorielle n'inclut pas de données hétérogènes, seules des données homogènes sont utilisées.

¹⁷⁵ Voir le chapitre 2.

¹⁷⁶ L'allocation tactique multi-classes d'actifs est de relativement court-terme.

historique forte. Ces auteurs pensent qu'il existe deux principales raisons qui font que les situations d'arbitrage ne sont pas exploitées. Une allocation multi-classes d'actifs peut dissuader les investisseurs car ils n'ont pas d'éléments théoriques et empiriques de la « bonne évaluation » de toutes les classes d'actifs. De plus, la plupart des professionnels se spécialisent sur l'allocation et la sélection de titres à l'intérieur d'une classe d'actifs. Ils proposent de prendre en considération des variables « prédictives » telle que des indicateurs macroéconomiques et de confiance qui peuvent exercer des effets sur toutes les classes d'actifs.

Nous poursuivons les travaux Blitz et Vliet (2008) en proposant des conclusions empiriques semblables mais dans le cadre des arbitrages mondiaux effectués sur un échantillon multi-classes d'actifs plus large. Notre univers multi-classes d'actifs a un périmètre plus étendu que celui de l'étude des précédents auteurs car nous intégrons les classes « devises » et « matières premières ». Dans notre processus d'allocation, l'introduction des « autres informations de marché » s'inscrit dans le cadre des « variables prédictives » suggérées par les auteurs. Ainsi, les changements d'allocation des portefeuilles d'arbitrage semblent dus à l'introduction de données hétérogènes, qui incluent des séries propres à la confiance des consommateurs, des investisseurs et des séries mesurant la conjoncture économique.

En partant de la conclusion des auteurs sur le manque de prise en compte de « variables prédictives » en allocations d'actifs, nous suggérons que la présence de données hétérogènes permet d'orienter les choix d'allocations d'actifs multi-classes pour que les portefeuilles d'arbitrage constitués représentent la « *smart money* » disponible pour capturer le « cross asset allocation alpha ». La présence de données hétérogènes dans le cadre d'allocations multi-classes d'actifs est une solution supplémentaire à la présence des *hedge funds*, proposé par les auteurs.

Nous élargissons la portée de leurs conclusions dans une logique d'arbitrage *long/short*. Nous construisons des portefeuilles d'arbitrage en prenant en considération des données hétérogènes, ces dernières nous permettant parfois de modifier les allocations (tableau 16). Ces modifications, qui résultent d'un changement de décision, peuvent entraîner un changement de rentabilité. La conséquence implicite de la prise en compte des « autres informations de marché » est que les allocations effectuées en présence de ces dernières semblent permettre de « capturer » ces « inefficiences » dès qu'elles se manifestent, sans que nous soyons limités par les contraintes comportementales observées dans la pratique, auxquelles les auteurs font référence. La prise en compte d'« autres informations de marché »

paraît être une solution pour éloigner la crainte d'entreprendre une allocation d'actifs, en ouvrant l'esprit des participants financiers à d'autres classes d'actifs que celles qu'ils connaissent particulièrement. Les « autres informations de marché » comme les « variables prédictives » sont censées irradier toutes les classes d'actifs. Cette propagation peut permettre d'améliorer la « *fair valuation* » sur plusieurs classes d'actifs, l'impact de ces informations modifiant l'anticipation des arbitrages éventuels.

Afin de calculer la rentabilité des portefeuilles d'arbitrage, nous définissons au préalable la méthode de valorisation des portefeuilles. Lors des tests empiriques, le calcul de la valeur de notre portefeuille d'arbitrage est réalisé quotidiennement à l'aide de l'équation suivante :

$$E_{t+\Delta t} = (1 + r\Delta t)E_t + \sum_i^{d_q} Q_{i,t}R_{i,t} - r\Delta t \left(\sum_{i=1}^{d_q} Q_{i,t} \right),$$

où $\Delta t = 1/252$,

r , le taux de frais sur les transactions (3%/252),

$Q_{i,t}$, le montant en euros investi sur l'actif i à l'instant t

et $R_{i,t}$, le taux de variation de l'actif i sur l'intervalle de temps $[t ; t + \Delta t]$. Nous rappelons que nous calculons $R_{i,t}$ de la manière suivante :

$$R_{i,t} = \frac{X_{i,t} - X_{i,t-1}}{X_{i,t-1}},$$

avec $X_{i,t}$ la cotation de l'actif i à l'instant t (l'instant de clôture en date t par exemple).

L'allocation d'actifs permet de comprendre comment la pondération d'un portefeuille a été effectuée et rend compte des choix d'investissement. L'analyse de la contribution par classe d'actifs à la rentabilité, nous permet d'identifier les classes qui ont le plus, et le moins, fortement contribué à la rentabilité du portefeuille.

1.2. Modification de la contribution par classe d'actifs à la rentabilité

L'allocation d'actifs permet de déterminer les pondérations, les montants de monnaie¹⁷⁷ investis sur chaque actif d'un portefeuille. L'introduction d'« autres informations de marché » dans nos travaux, modifie les sources de rentabilité des portefeuilles d'arbitrage. Nous nous intéressons à la contribution annualisée de chaque actif à la rentabilité annualisée des portefeuilles d'arbitrage. Nous effectuons la même démarche pour la contribution annualisée de chaque classe d'actifs à la rentabilité du portefeuille.

Le calcul de C_t , la contribution par classe d'actifs (de chaque actif i de la classe) à la rentabilité à l'instant t , est la suivante :

$$C_t = \sum_{i \in \text{Classe}} \sum_{s \leq t} Q_{i,s} R_{i,s},$$

avec i , l'indice du nombre d'actifs de la classe¹⁷⁸,

Classe, la classe d'actifs,

$Q_{i,s}$, le montant en euros investi sur l'actif i à l'instant s

et $R_{i,s}$, le taux de variation de l'actif i à l'instant s . La contribution moyenne annualisée est définie ci-dessous :

$$\frac{1}{T} \sum_{t=1}^T \left(\frac{C_t - C_{t-1}}{C_{t-1}} \right) \times 252,$$

avec C_t la contribution en date t ,

et T , la date finale de notre échantillon.

Les contributions par classe d'actifs sont des rentabilités de sous-portefeuilles qui, agrégés, égalisent la rentabilité annualisée du portefeuille comprenant les 32 données de cotation ($i=1, \dots, 32$) de notre échantillon.

L'influence exercée par une combinaison de données hétérogènes sur les contributions à la rentabilité est étudiée par classes d'actifs (tableau 17) puis par actif (tableau 18). La contribution par classe d'actifs à la rentabilité des portefeuilles d'arbitrage varie selon que

¹⁷⁷ En euros ou une autre monnaie par exemple.

¹⁷⁸ Prenons par exemple la classe d'actifs « obligations ». Nous sommes l'indice i jusqu'à 4 car nous avons 4 actifs recueillis dans la classe d'actifs « obligations ».

nous considérons ou non une combinaison de données hétérogènes¹⁷⁹ pour toute la période de tests de notre échantillon. Nous cherchons à identifier les classes d'actifs qui ont le plus (et le moins) fortement contribué à la rentabilité, tout en considérant l'allocation d'actifs correspondante. Pour une allocation d'actifs relativement proche (tableau 8), les contributions par classes d'actifs changent (tableau 17). Il semble que les signaux de *market timing* en présence de données hétérogènes ne soient pas identiques à ceux obtenus en présence de données homogènes. Ceci peut induire des décisions différentes en matière de gestion de portefeuille.

Tableau 17 - Comparaison de la contribution par classe d'actifs à la rentabilité d'un portefeuille d'arbitrage selon le type de données pour la période de tests (2006-2011)

Type de données	Classe d'actifs				Total
	Obligations	Matières premières	Actions	Devises	
Données hétérogènes (1)	3,5%	20,6%	52,6%	23,3%	100,0%
Données de homogènes (2)	0,4%	26,7%	44,6%	28,3%	100,0%
Différence de contribution : (1) - (2)	3,2%	-6,1%	8,0%	-5,1%	0,0%

En partant de la gauche, le type de données est en première colonne. La contribution par classe d'actifs à la rentabilité est renseignée de la deuxième à la cinquième colonne. Le total des pondérations est en sixième colonne.

Pour chaque classe d'actifs, la contribution à la rentabilité varie en présence d'une combinaison de données hétérogènes (tableau 17). Alors que l'allocation d'actifs de la classe « actions » est identique¹⁸⁰ (tableaux 16), la contribution de cette dernière augmente de 8% en présence de données hétérogènes (tableau 17). Cette différence de contribution est la plus élevée parmi les quatre classes d'actifs (tableau 17). Ce chiffre indique qu'en moyenne sur l'ensemble de notre période de tests, 52,6% de la rentabilité est expliquée par la classe « actions » en présence de données hétérogènes alors que nous obtenons une contribution de 44,6% en présence de données homogènes (tableau 17). Ce pourcentage semble indiquer qu'une distorsion entre les contributions peut exister en présence de données hétérogènes, toutes choses égales par ailleurs¹⁸¹. Indépendamment de la présence (ou non) de données hétérogènes nous soulignons l'importante contribution de la classe « actions » à la

¹⁷⁹ Voir en détail notre phase d'analyse des données (ACP) dans le chapitre 2.

¹⁸⁰ En présence ou non de données hétérogènes.

¹⁸¹ A périmètre constant.

rentabilité¹⁸². Nous pouvons remarquer que les statistiques descriptives, calculées pour la période de tests de l'échantillon (tableau 11), montrent que la classe « actions » a des taux de variation parmi les plus élevés des données de cotation.

La contribution de la classe « obligations » augmente avec la prise en compte de données hétérogènes (tableau 17) malgré une sous-pondération d'allocation sur cette classe (tableau 16). A l'opposé, la contribution de la classe « matières premières » diminue (tableau 17) pour une surpondération de son allocation (tableau 16) en présence de données hétérogènes. La contribution de la classe « devises » diminue (tableau) pour une relativement « faible » surpondération de son allocation en présence de données hétérogènes (tableau 16). Pour une allocation quasiment semblable, en présence ou non de « données hétérogènes », les contributions des classes « actions » et « devises » ont été modifiées en présence de données hétérogènes.

Avec une granularité plus fine, nous présentons la contribution de chaque actif à la rentabilité (tableau 18). Cette décomposition permet d'obtenir la contribution par actif au sein d'une classe (tableau 17). Les différences de contributions par classe d'actifs (tableau 17) sont analytiquement calculées par actif de la classe (tableau 18). Cette présentation de la contribution par actif permet également de réconcilier le montant agrégé de contribution par classe d'actifs. Pour la classe « obligations », par exemple, la somme des différences de contribution pour les quatre actifs de cette classe est égale à + 3,2%¹⁸³ (tableau 18). Ce chiffre correspond au +3,2% calculé pour la classe « obligations » dans le tableau 17.

Nos données relatives aux « autres informations de marché » des Etats-Unis et d'Europe semblent avoir un impact sur d'autres actifs que ceux de leurs zones géographiques. Certains éléments de contribution par actif observés sur notre échantillon en présence de données hétérogènes sont à étudier. La différence de contributions pour la classe « obligations » (tableau 17) provient d'un accroissement de la contribution des obligations d'Etat américaines et brésiliennes (tableau 18) en présence de « données hétérogènes ». La différence de contribution pour la classe « matières premières » (tableau 17) provient principalement d'une baisse plus prononcée pour le maïs (tableau 18) en présence de données hétérogènes. Notons le relatif accroissement de la contribution du blé en présence de « données hétérogènes » (tableau 18), généralement impactée par les informations économiques dont la demande et l'inflation. Au niveau de la classe « actions », ce sont surtout les accroissements relatifs de contributions des indices indiens, allemands et chinois (tableau

¹⁸² Bien que supérieure en présence de données hétérogènes (tableau 17).

¹⁸³ En tenant compte des arrondis.

18) qui expliquent la différence de contributions. La différence de contributions pour la classe « devises » s'explique principalement par des baisses relatives de contributions des devises won coréen / euro et plus marginalement dollar américain / euro. Notons qu'au sein d'un univers uniquement composés d'actions, Ferson et Harvey (1993) estiment que les rendements d'actifs sont prévisibles dans des pays « développés » et « émergents » lorsque des séries d'inflation et de production de plusieurs pays du G7 sont prises en compte. L'impact d'informations économiques de pays du G7 n'est pas uniquement limité aux marchés actions du G7 mais peut s'étendre à ceux des pays dit « émergents ».

Tableau 18 - Contribution par actif à la rentabilité d'un portefeuille d'arbitrage selon le type de données

	Contribution par actif à la rentabilité en présence de données hétérogènes (1)	Contribution par actif à la rentabilité en présence de données homogènes (2)	(1) - (2)
France 2004 4% 25/04/14	0,1%	0,7%	-0,6%
US 2004 4% 15/02/14	2,1%	-0,8%	2,9%
Japan 2004 1,5% 20/03/14	0,1%	0,9%	-0,9%
Brazil 2004 10,5% 14/07/14	1,2%	-0,5%	1,7%
Brent (pétrole)	16,0%	15,1%	0,9%
West Texas Intermediate (pétrole)	-0,8%	0,1%	-0,9%
Or	6,6%	6,7%	-0,1%
Blé	12,4%	8,7%	3,7%
Maïs	-13,6%	-3,9%	-9,7%
Bovespa (Brésil)	-1,5%	6,3%	-7,9%
Bse sensex (Inde)	7,4%	-2,1%	9,5%
Cac 40 (France)	11,7%	12,7%	-1,0%
Dax (Allemagne)	3,9%	-2,6%	6,5%
Ftse 100 (Royaume Uni)	2,9%	2,7%	0,2%
S&P 500 (Etats-Unis)	8,8%	13,4%	-4,6%
Nasdaq 100 (Etats-Unis)	4,2%	8,6%	-4,4%
Asx 200 (Australie)	4,3%	3,1%	1,2%
Shanghai SE (Chine)	7,0%	-2,2%	9,3%
Nikkei 225 (Japon)	3,9%	4,7%	-0,7%
Dollar américain / euro	3,2%	7,6%	-4,4%
Sterling /euro	0,1%	0,0%	0,1%
Franc suisse / euro	2,1%	0,8%	1,3%
Yen / euro	1,0%	-2,1%	3,2%
Yuan /euro	2,4%	3,3%	-0,9%
Couronne norvégienne / euro	-2,5%	-0,8%	-1,7%
Real brésilien / euro	2,4%	2,7%	-0,2%
Roupie indienne / euro	7,4%	5,5%	1,8%
Won coréen / euro	-2,1%	10,5%	-12,6%
Dollar australien / euro	4,7%	-1,4%	6,1%
Rand sud africain / euro	1,9%	-5,7%	7,6%
Dollar néo-zélandais / euro	1,2%	4,1%	-2,8%
Dollar canadien / euro	1,3%	3,8%	-2,5%

En partant de la gauche, les actifs sont en première colonne. La contribution par actif à la rentabilité en présence de données hétérogènes est en deuxième colonne. La contribution par actif à la rentabilité en présence de données homogènes est en troisième colonne. La différence entre la deuxième et la troisième colonne est en quatrième colonne.

Nous complétons les découvertes empiriques de divers travaux (Brenner, Pasquariello, et Subrahmanyam, 2009 ; Andersen et al., 2007 ; Boyd, Jagannathan, et Hu, 2005), qui dans des configurations différentes, mesurent l'impact d'informations économiques sur les prix d'actifs cotés et/ou leurs autres moments statistiques (rendements, volatilités). Ces travaux montrent également en quoi la publication de ces enquêtes économiques n'entraîne pas les mêmes réactions pour toutes les classes d'actifs.

Dans une autre direction, Jones et al. (1998) trouvent que la volatilité du marché obligataire américain¹⁸⁴ le jour de l'annonce d'informations économiques ne perdure pas. Par extension, l'impact *post* publication n'entre pas dans le cadre de l'analyse des auteurs.

Boyd et al. (2005) constatent qu'aux Etats-Unis, la publication d'enquête sur l'emploi n'entraîne pas les mêmes réactions de prix sur les actions (S&P 500) et les obligations (obligations d'Etat). Ils tempèrent ainsi ces deux effets en expliquant qu'ils varient selon l'état de l'économie. Selon Andersen et al. (2007), le marché obligataire est celui qui réagit le plus fortement aux « nouvelles » économiques par rapport aux marchés mondiaux des actions et des devises.

Nos résultats empiriques abondent dans le sens de travaux qui soutiennent que les rendements d'actifs sont partiellement prévisibles en fonction du cycle de la conjoncture (Andersen et al., 2007 ; Dahlquist et Campbell, 2001). Notons que l'information « conditionnelle » à l'estimation des rendements d'actifs d'un jour sur l'autre intègre l'information contenue dans les « autres informations de marché » de notre échantillon.

Dans notre cadre de recherche, les différences de réactions par classe d'actifs aux informations économiques sont mesurées notamment par les différences de contributions à la rentabilité des portefeuilles d'arbitrage (tableaux 17 et 18). Empiriquement, les réactions des actifs ne sont pas les mêmes en présence de données hétérogènes qu'en présence de données homogènes.

D'après nos résultats empiriques, nous divergeons de l'approche de Jones et al. (1998) car nous considérons la persistance des effets des « autres informations de marché » sur nos données de cotation. Cette trace *post* publication se manifeste conceptuellement dans le cadre des traitements de données (chapitre 3) ainsi que dans l'allocation et les contributions par classe d'actifs à la rentabilité des portefeuilles.

Nous nous inscrivons dans la continuité des découvertes en univers « haute fréquence » (Brenner, Pasquariello, et Subrahmanyam, 2009 ; Andersen et al., 2007) ou

¹⁸⁴ Obligations d'Etat américain.

quotidienne (Boyd, Jagannathan, et Hu, 2005) mais en nous plaçant dans un cadre d'arbitrage multi-classes d'actifs utilisant des cotations de clôture. A une fréquence plus « basse », nous confirmons l'existence d'interactions entre les « autres information de marché » et l'évolution d'actifs financiers cotés, et ceci dans un cadre mondial qui intègre les « nouvelles » zones de croissance économique (Brésil, Chine, Inde et Australie).

Nous montrons que la prise en compte des données hétérogènes impacte sensiblement la contribution des actifs à la rentabilité des portefeuilles d'arbitrage. Ce résultat est d'autant remarquable, que nous avons trouvé que la combinaison de données de cotation et d'« autres informations de marché » a relativement peu d'influence sur la composition des portefeuilles d'arbitrage (tableau 16). L'analyse de données que nous avons effectuée en période « apprentissage »¹⁸⁵ semble être discriminante. La prise en compte des « autres informations de marché » permet en effet de générer de nouveaux signaux.

1.3. Réduction en moyenne de la variabilité des rentabilités

Pour la période de tests de notre échantillon, la volatilité annualisée des portefeuilles d'arbitrage est inférieure (4,70%) en présence de données hétérogènes par rapport à celle obtenue en présence de données homogènes (5,55%, tableau 19). A l'exception de la sous-période de tests précédant la crise des *subprimes*¹⁸⁶, la volatilité décroît en présence de données hétérogènes (tableau 19). La volatilité annualisée des portefeuilles d'arbitrage est calculée de la manière suivante :

$$\sqrt[2]{\frac{1}{T-1} \sum_{t=1}^T \left(\frac{E_t - E_{t-1}}{E_{t-1}} - \overline{R_T} \right)^2} * \sqrt[2]{252}$$

avec E_t , la valorisation du portefeuille à l'instant t

et $\overline{R_T}$, la rentabilité moyenne annualisée telle que :

$$\overline{R_T} = \frac{1}{T} \sum_{t=1}^T \left(\frac{E_t - E_{t-1}}{E_{t-1}} \right) * 252.$$

¹⁸⁵ Voir chapitre 2.

¹⁸⁶ Dans notre approche, nous définissons cette période entre juillet 2006 et janvier 2009.

Tableau 19 - Rentabilités et volatilités annualisées d'un portefeuille d'arbitrage selon le type de données et les périodes de tests

Données	Période	Rentabilité (1)	Volatilité (2)	(1) / (2)
Données hétérogènes	2006 - 2011	11,79%	4,70%	2,5
Données homogènes		13,93%	5,55%	2,5
Données hétérogènes	Avant la crise des <i>subprimes</i> (juillet 2006 - janvier 2007)	10,32%	3,78%	2,7
Données homogènes		12,08%	3,38%	3,6
Données hétérogènes	Pendant la crise des <i>subprimes</i> (février 2007 - mars 2009)	11,96%	5,61%	2,1
Données homogènes		19,68%	6,78%	2,9
Données hétérogènes	Après la crise des <i>subprimes</i> (avril 2009 - mars 2011)	12,03%	3,72%	3,2
Données homogènes		8,02%	4,37%	1,8

En partant de la gauche, le type de données est en première colonne. La période de tests de l'échantillon est en deuxième colonne. La rentabilité annualisée est en troisième colonne. La volatilité annualisée est en quatrième colonne. Le ratio rentabilité annualisée/volatilité annualisée est en cinquième colonne.

La volatilité diminue en présence de « données hétérogènes » pendant et après la crise des *subprimes*¹⁸⁷ (tableau 19). L'écart de volatilité le plus élevé, entre nos deux univers de données hétérogènes et homogènes, se manifeste pendant la période des *subprimes* (tableau 19). La période *post subprimes*¹⁸⁸ est l'intervalle de temps où la volatilité est la plus basse (tableau 19) en présence de données hétérogènes. Les baisses relatives de volatilité en présence de données hétérogènes lors de ces sous-périodes de tests (tableau 19) permettent en partie d'expliquer la baisse de la volatilité pour l'ensemble de la période de tests de l'échantillon.

En présence de données hétérogènes, l'abaissement du niveau de la volatilité peut également s'expliquer par le fait que certaines corrélations¹⁸⁹ entre les contributions à la rentabilité par classe d'actifs diminuent de manière significative¹⁹⁰ (tableau 20).

¹⁸⁷ Telles que nous délimitons ces périodes dans notre approche.

¹⁸⁸ Dans notre approche, nous délimitons cette période entre avril 2009 et mars 2011.

¹⁸⁹ Faiblement positives ou légèrement négatives.

¹⁹⁰ Au sens statistique.

Pour estimer si la corrélation est significativement différente de 0, nous calculons la t statistique de Student :

$$t = \frac{|\rho_{x,y}|}{\sqrt{\frac{1-\rho_{x,y}^2}{n-2}}}$$

avec x , une contribution à la rentabilité pour une classe d'actifs,

y , une autre contribution à la rentabilité pour une autre classe d'actifs et

n le nombre d'observations.

La corrélation entre les contributions indique comment évolue la contribution d'une classe en fonction d'une autre classe. Ces éléments de corrélation permettent d'expliquer partiellement la baisse de la volatilité des portefeuilles d'arbitrage mais également de mesurer d'éventuels éléments de diversification du risque des portefeuilles d'arbitrage.

Décrivons l'effet de la baisse relative des corrélations, en présence de données hétérogènes, sur la volatilité. Pour la période de tests de notre échantillon, les corrélations entre contributions en présence de données hétérogènes sont majoritairement plus faibles¹⁹¹ (tableau 20), entraînant une diminution de la variance des portefeuilles d'arbitrage. Cette conclusion est renforcée par le caractère significatif de certaines corrélations (tableau 20). La corrélation entre contributions de la classe « obligations » et de la classe « actions » est « légèrement » positive et statistiquement significative¹⁹² (tableau 20). En d'autres termes, quand la contribution à la rentabilité de la classe « obligations » croît, celle des « actions » augmente et inversement. La corrélation entre les contributions de la classe « obligations » et de la classe « devises » est significativement négative (tableau 20) ainsi que celle entre les contributions de la classe « devises » et de la classe « actions » (tableau 20). Bien que significative, en présence de données hétérogènes comme en présence de données homogènes, la corrélation entre « obligations » et « actions » est plus faible en présence de données hétérogènes (tableau 20). Il en est de même pour la corrélation entre « devises et actions » : en présence de données hétérogènes, la corrélation devient négative (tableau 20).

Brenner et al. (2009) évoquent des différences de volatilités « conditionnelles » entre les marchés actions et obligations aux Etats-Unis lors des publications d'enquêtes économiques américaines. Les auteurs analysent également les « co-mouvements » entre les

¹⁹¹ Qu'en présence de données homogènes.

¹⁹² A un seuil de confiance de 95%.

rendements : au sein des actifs de la classe et entre classes d'actifs. Les différents impacts des publications économiques sur les volatilités peuvent engendrer des mouvements contraires entre rendements.

Plusieurs recherches identifient l'effet « surprise » des annonces¹⁹³ (Brenner, Pasquariello, et Subrahmanyam, 2009 ; Andersen et al., 2007 ; Andersen et al., 2003) sur les rendements et/ou les volatilités d'actifs, alors que l'on a l'habitude de traiter les séries publiées historiquement sans chercher à capter l'effet inattendu d'une information. Nos résultats semblent indiquer que la présence des données hétérogènes peut influencer des rendements et des volatilités d'actifs alors même que les informations ne sont pas issues d'un effet « surprise ».

La baisse de corrélations entre contributions par classe d'actifs à la rentabilité (tableau 20) ou l'accroissement de la négativité des corrélations (tableau 20) en présence de données hétérogènes converge avec les découvertes empiriques des différents auteurs. Nous obtenons des conclusions similaires sur le fait qu'une « autre information de marché » entraîne des différences de corrélations et de rendements entre classes d'actifs. Nous considérons ces résultats dans un contexte de classes d'actifs incluant les « devises » et les « matières premières » et dans une modélisation d'arbitrage statistique différente des modélisations économétriques proposées par les différents travaux.

¹⁹³ Généralement, cet effet « surprise » correspond à la différence entre le chiffre estimé (à partir d'un consensus d'analystes) et le chiffre publié.

Tableau 20 - Significativité statistique des corrélations entre les contributions par classe d'actifs à la rentabilité d'un portefeuille d'arbitrage selon le type de données pour la période de tests de l'échantillon (2006-2011)

Données homogènes	Type de données	Coefficient de corrélation de Pearson	<i>t stat</i>
Obligation et matières premières	Données hétérogènes	0,13%	0,04
	Données homogènes	-2,57%	0,89
Obligations et actions	Données hétérogènes	6,75%	2,35 **
	Données homogènes	23,95%	8,58 ***
Obligations et devises	Données hétérogènes	-6,66%	2,32 **
	Données homogènes	0,77%	0,27
Matières premières et actions	Données hétérogènes	0,32%	0,11
	Données homogènes	0,88%	0,31
Matières premières et devises	Données hétérogènes	-0,78%	0,27
	Données homogènes	-6,08%	2,12 **
Devises et actions	Données hétérogènes	-6,13%	2,13 **
	Données homogènes	14,87%	5,23 ***

** : $p < 0,05$; *** : $p < 0,01$

En partant de la gauche, les données homogènes sur lesquelles porte le calcul de la corrélation sont en première colonne. Le type de données utilisé pour composer nos portefeuilles d'arbitrage est en deuxième colonne. Le coefficient de corrélation de Pearson entre contributions de classes d'actifs à la rentabilité annualisée est en troisième colonne. La *t-stat* de Student est en quatrième colonne.

Une conséquence de cette baisse significative¹⁹⁴, en présence de données hétérogènes, est que les covariances entre classes d'actifs sont négatives ou (faiblement) positives. Ceci implique que la volatilité globale du portefeuille d'arbitrage diminue en présence de données hétérogènes (tableau 19) par rapport à un univers de données homogènes où les corrélations

¹⁹⁴ Au sens statistique.

significatives¹⁹⁵ sont plus fortement positives (tableau 20). Les covariances entre contributions par classes d'actifs sont inférieures pour un univers de données hétérogènes, ce qui implique une baisse de la volatilité des rentabilités des portefeuilles d'arbitrage (tableau 19). L'effet de la baisse des corrélations en présence de données hétérogènes, contribue à la diversification du risque de nos portefeuilles d'arbitrage. Pour la période de tests, les corrélations baissent et peuvent devenir négatives en présence de données hétérogènes (tableau 20).

Nous effectuons également le même calcul de corrélations pour les données de cotation (par classe d'actifs) telles que nous les avons collectées sous Thomson Reuters Datastream, donc sans période d'« apprentissage ». L'objectif de ce nouveau calcul de corrélations (tableau 21) est de comprendre en quoi l'introduction de données hétérogènes vient modifier la structure des corrélations originelles. De cette manière, nous calculons des corrélations en considérant uniquement des prix de clôture (données de cotation) sans faire appel à nos procédures d'analyse de données et de quantification. Pour ce nouveau calcul de corrélations, nous utilisons donc les données de cotation en l'état pour la période de tests de notre échantillon. Les corrélations sont à l'origine toutes significatives et relativement « fortement » positives ou négatives (tableau 21).

Les modifications de corrélations peuvent notamment être expliquées par la prise en compte de données hétérogènes. Après une période d'« apprentissage », nous constatons que les corrélations entre classes d'actifs ont été plus fortement abaissées en présence de données hétérogènes qu'en présence de données homogènes (tableau 20). Précisément, les corrélations convergent autour de 0% dans des zones positives ou négatives (tableau 20). La corrélation entre les classes « obligations » et « actions » passe de -49,34% (tableau 21) à -6,75% (tableau 20), celle entre les classes « obligations » et « devises » passe de 44,24% (tableau 21) à -6,66% (tableau 20) et celle entre les classes « devises » et « actions » passe de 18,34% (tableau 21) à -6,13% (tableau 20).

¹⁹⁵ Au sens statistique.

Tableau 21 - Significativité statistique des corrélations entre données homogènes pour la période de tests de l'échantillon (2006-2011)

Données homogènes	Coefficient de corrélation de Pearson	<i>t stat</i>
Obligation et matières premières	-11,33%	3.97 ***
Obligations et actions	-49,34%	19.75 ***
Obligations et devises	44,24%	17.18 ***
Matières premières et actions	27,79%	10.07 ***
Matières premières et devises	19,12%	6.78 ***
Devises et actions	18,34%	6.49 ***

** : $p < 0,05$; *** : $p < 0,01$

En partant de la gauche, les données homogènes sur lesquelles porte le calcul de la corrélation sont en première colonne. Le coefficient de corrélation de Pearson est en deuxième colonne. La *t-stat* de Student est en troisième colonne.

Ce résultat empirique semble indiquer que l'introduction d'« autres informations de marché » peut favoriser l'indépendance statistique des classes d'actifs. Dans cette configuration, l'évolution d'une classe d'actifs est de moins en moins sensible à l'évolution d'une autre. La sensibilité des variations d'une classe d'actif à une autre diminue en présence de données hétérogènes. Les contributions par classes d'actifs à la rentabilité devenant faiblement corrélées, l'évolution de la rentabilité d'une classe n'impacte pas obligatoirement l'évolution de la rentabilité d'une autre classe. En ce sens, la diversification du risque des portefeuilles d'arbitrage peut être améliorée en présence de données hétérogènes.

Dans une logique similaire, nous notons que la rentabilité de notre portefeuille

d'arbitrage évolue en sens inverse de celle d'un portefeuille des données de cotation obtenu par analyse factorielle (ACP) dans le chapitre 3 (figure 13). Ce portefeuille fondé sur des données de cotation et issu d'une ACP n'est pas soumis à notre méthode d'arbitrage. Ce portefeuille ne constitue pas en soi un *benchmark* mais peut être considéré comme l'évolution statistique historique de l'ensemble des données de cotation sur le passé. A ce titre, ce portefeuille représente l'évolution des actifs sans prise en compte des innovations proposées par notre approche en présence de données hétérogènes et par des procédures de quantification. En présence de données hétérogènes la rentabilité quotidienne de notre portefeuille d'arbitrage « euro neutre » d'allocation « naïve » ne semble pas subir la « chute » de valorisation que rencontre le portefeuille de données de cotation lors de la crise des *subprimes* (figure 13, chapitre 3). L'évolution de la valorisation quotidienne de notre portefeuille en présence de données hétérogènes (annexe 9) ne semble pas affectée par l'avènement de la crise des *subprimes* puisque sa rentabilité n'a pas connu de décroissance « marquée » pour cette période (tableau 19).

Certains travaux mettent en exergue la hausse de la volatilité des rendements d'indices actions mondiaux (*MSCI World*) et de la corrélation entre un indice actions national et des indices mondiaux en périodes de récession. Dahlquist et Campbell (2001) tentent de mesurer le changement de volatilité et de corrélations en périodes d'expansion et de contraction économique. Les auteurs estiment que les volatilités et également les corrélations augmentent en périodes de contraction économique. Ceci sous-entend implicitement qu'une forte hausse des covariances a lieu en période de contraction économique. Les auteurs soulignent que les variables qui aident à prévoir les rendements permettent aussi de prévoir le cycle de l'économie.

Andersen et al. (2007) partagent un autre point de vue sur les corrélations. Ils découvrent une corrélation négative entre les « obligations » et les « actions » en présence d'informations économiques et ceci en période de contraction économique. Gorton et Rouwenhorst (2006) montrent que, dans un cadre « 1/N », les matières premières apportent une solution de diversification du risque d'un portefeuille qui intègre également des actions et des obligations. D'une part, les auteurs insistent sur les corrélations négatives entre les matières premières et les deux autres classes d'actifs. Ils se basent sur ces corrélations pour souligner les évolutions divergentes des classes d'actifs. D'autre part, ils évoquent également la corrélation positive entre les matières premières et l'inflation. D'après eux, les corrélations négatives entre la classe des « matières premières » et les deux autres classes sont majoritairement imputables aux réactions aux cycles économiques.

Nos résultats se situent dans le prolongement de ceux de Gorton et Rouwenhorst (2006) dans la mesure où nous obtenons des corrélations qui diminuent en présence de données hétérogènes (tableau 20), et que l'utilisation de ce type de données nous permet de prendre en compte l'état de la conjoncture économique et indirectement d'intégrer les cycles économiques. Cependant, notre étude innove par rapport aux travaux des précédents auteurs en intégrant la classe des « devises ».

Nos travaux se situent dans la continuité de ceux qui établissent un lien entre volatilité, corrélation et informations économiques ou « fondamentale ». A cet effet, nous utilisons des données hétérogènes qui intègrent tout autant des indicateurs de conjoncture que de sentiment et nous mesurons leur influence sur les volatilités des portefeuilles et les corrélations entre les classes d'actifs. Nous élargissons le spectre de la conclusion selon laquelle la volatilité augmente en temps de crise. Nous obtenons des volatilités qui diminuent en période de *subprimes* ainsi que sur une période de tests qui inclut des périodes hors-crise (tableau 19). Nous expliquons ces résultats par le fait que dans le cadre de l'arbitrage statistique l'effet de la baisse des corrélations est renforcé par la prise en compte de données hétérogènes. Nous poursuivons par ailleurs ces travaux sur les liens existants entre les actifs et les « autres informations de marché » en proposant un panier d'indicateurs mondiaux et de différentes natures (conjoncture, sentiment et volatilité).

Nous analysons à présent l'influence de données hétérogènes sur le ratio rentabilité/volatilité, sachant désormais que celles-ci peuvent permettre une réduction de la volatilité des portefeuilles.

1.4. Accroissement du ratio rentabilité/volatilité après la crise des *subprimes*

La contraction de la volatilité en présence de données hétérogènes a été mise en évidence dans la section précédente. Il s'agit maintenant de comprendre comment cette contraction influence le ratio rentabilité/volatilité.

Sur l'ensemble de la période de tests de notre échantillon, le ratio rentabilité/volatilité est égal à 2,5. Il est identique en présence de données hétérogènes ou homogènes (tableau 19), ce qui signifie que les effets de la prise en compte des « autres informations de marché » sur la rentabilité et le risque se compensent (approximativement). La rentabilité du portefeuille issu d'une analyse de données hétérogènes est inférieure de 2,14% à celle du portefeuille pour un univers homogène de données, alors que sa volatilité est inférieure de 0,85%. En valeurs relatives, la baisse de la rentabilité est de 15,36 % et celle de la volatilité de 15,32%.

Sur la période qui précède la crise des *subprimes* le ratio rentabilité/volatilité avec données hétérogènes est inférieur de 0,9 par rapport à celui avec données homogènes (tableau 19). L'écart se réduit légèrement lors la crise des *subprimes* (0,8). Par contre, après la crise, le l'écart s'inverse et devient plus important. Le ratio obtenu avec données hétérogènes, qui connaît sa valeur la plus élevée (3,2) sur toute la période de tests, est supérieur de 1,4 par rapport à celui avec données homogènes. Lors de cette période, l'écart des ratios a augmenté en faveur des portefeuilles composés en présence de données hétérogènes : la rentabilité étant plus forte et la volatilité plus faible (tableau 19). Une interprétation possible de ce résultat est que les « autres informations de marché » ont pris une dimension décisive. Notre pensée est qu'en temps de crise des indicateurs autres que des données de cotation peuvent impacter les décisions de gestion. L'analyse des ratios rentabilité/volatilité des sous-périodes permet de rendre compte de l'importance d'autres types d'information en gestion après qu'un évènement de crise soit survenu.

Dans leurs travaux, Ferson et Harvey (1993) abordent la relation empirique entre les rendements futurs d'actifs et leurs sensibilités à un marché. Notons que la mesure du « bêta » inclut notamment la variance. Ils participent à la discussion sur cette relation en expliquant que les rendements attendus d'actions internationales évoluent dans une faible proportion en fonction de mouvements de « beta », qui eux changent en fonction d'informations locales. Selon ces auteurs, les rendements attendus évoluent plus fortement en fonction d'une prime de risque international. Au travers de leur étude empirique, les auteurs décèlent que les primes de risque international dépendent de « variables d'information » internationales, comme l'inflation et le taux de croissance de la production industrielle. Ils supposent que si les marchés « actions » diffèrent dans leur exposition à l'évolution des perspectives d'inflation, c'est qu'il existe une prime de risque inflation pour les marchés « actions ». Jones et al. (1998) pensent que les rendements futurs sont prévisibles car le risque « fondamental » l'est également : les prix d'actifs varient alors avec le risque macroéconomique.

Nous complétons les travaux de Ferson et Harvey (1993) en estimant les rendements futurs en fonction d'informations passées qui ne sont pas spécifiques à un pays en particulier et qui peuvent avoir une portée internationale. Bien que nous ne développions pas des tests empiriques en référence à un « benchmark »¹⁹⁶, nous estimons les sensibilités entre classes d'actifs au travers des corrélations de contributions (tableau 20). La sensibilité entre classes d'actifs baissant en présence de données hétérogènes, nous complétons leur analyse en

¹⁹⁶ Nous avons expliqué pourquoi précédemment dans ce chapitre.

univers multi-classes d'actifs. Il semble que l'impact des données relatives aux « autres informations de marché » dépasse le cadre mono-classe d'actifs défini par les auteurs. Les mêmes phénomènes d'exposition semblent se produire sauf que la résonance des « autres informations de marché » s'effectue sur plusieurs de nos classes d'actifs.

L'accroissement de ratio rentabilité/volatilité des portefeuilles d'arbitrage après la crise des *subprimes* (tableau 19) est principalement dû à une rentabilité supérieure en présence de données hétérogènes. Cependant, l'effet bénéfique est complété par la réduction de risque (sous toutes ses formes) liée à la prise en compte des données relatives aux « autres informations de marché ». En ce sens, nous convergions vers les conclusions de Jones et al. (1998) sur l'estimation des rendements futurs à l'aide de l'estimation du risque « fondamental » ou macroéconomique. Après une période de crise, le fait de considérer des données hétérogènes semble expliquer, au moins partiellement, l'obtention d'une rentabilité supérieure en présence de données hétérogènes, ce qui n'est jamais le cas pour les autres périodes où la rentabilité est toujours supérieure en présence de données homogènes (tableau 19). Les « autres informations de marché » semblent permettre d'estimer ce risque « fondamental », différentes enquêtes de conjoncture économique étant incluses dans cette variable¹⁹⁷. Nous contribuons à l'élargissement du champ d'analyse des précédents auteurs dans deux directions. D'une part, nous incluons d'autres types de séries (confiance des consommateurs et des investisseurs, nervosité des marchés financiers) qui permettent d'estimer potentiellement le risque « fondamental ». D'autre part, nous menons l'analyse dans un nouveau contexte : celui de l'arbitrage dans un univers multi-classes d'actifs.

1.5. Réduction des pertes financières probables pendant la crise des *subprimes*

La question de la gestion du risque est implicitement abordée au travers des corrélations entre les contributions des classes d'actifs à la rentabilité (tableau 20). La réduction de celles-ci en présence de données hétérogènes constitue en effet un élément de diversification du risque des portefeuilles d'arbitrage¹⁹⁸. Un autre aspect de la gestion du risque concerne l'estimation des pertes probables. Nous examinons comment peuvent évoluer les pertes financières probables de portefeuilles d'arbitrage « euro neutre » diversifiés de manière « naïve » en présence ou non de données hétérogènes.

¹⁹⁷ Variable d'« autres informations de marché ».

¹⁹⁸ Dans notre recherche nous n'avons pas comme priorité d'analyser l'évolution de la gestion du risque en présence de données hétérogènes. Nous ne soulevons que très partiellement les nombreux aspects de la gestion du risque au sein de notre étude empirique.

Nous utilisons la mesure de la *Value-at-Risk* (Linsmeier et Pearson, 1996), autrement appelée *VaR*, pour estimer ces pertes probables. Nous appliquons un test de Kupiec (1995) à ces estimations de pertes financières probables pour valider ou non la significativité statistique de nos résultats. Nous définissons la *VaR*, les modalités du test de Kupiec (1995) puis nous présentons les résultats statistiques du test.

La *VaR* peut être estimée de manière paramétrique et non-paramétrique. Dans le premier cas, l'estimation de la *VaR* dépend d'une loi de probabilité. Dans le deuxième cas, l'estimation de la *VaR* ne dépend pas d'une loi de probabilité. De manière générale, la *VaR* est définie comme le niveau minimal γ tel que la probabilité de perte du portefeuille $-r(\omega, x)$ ne dépasse pas ε^{199} :

$$Prob\{\gamma \leq -r(\omega, x)\} \leq \varepsilon.$$

Le rendement total d'un portefeuille d'actifs financiers peut être vu comme une variable aléatoire $r(\omega, x) = \sum_{i=1}^n w_i x_i$ avec w_i l'allocation sur un actif i et x_i le rendement de l'actif i , avec x modélisé tel qu'un vecteur aléatoire de dimension n (El Ghaoui, Oks, et Oustry, 2003).

Nous calculons une *VaR* paramétrique quotidienne avec un intervalle de confiance de 99% ($\varepsilon \leq 1\%$) à l'aide d'une loi normale standard²⁰⁰. L'approche de la *VaR* suppose que nous connaissions toute la distribution des rendements : dans notre cas elle sera gaussienne.

Le calcul de notre *VaR* paramétrique quotidienne à 99% est décomposé ci-dessous comme le proposent Racicot et Théorêt (2006) :

$$VaR^{99\%} = A * z * \sigma,$$

Avec A , le montant d'un portefeuille en euros,
 z , le quantile de la loi normale standard²⁰¹ ou l'inverse de la fonction de répartition de la loi normale standard $\varphi^{-1}(1 - \varepsilon) = z$, avec φ la fonction de répartition de la loi normale standard) et

σ , l'écart-type des taux de variation de notre portefeuille d'arbitrage.

L'écart-type σ , a été calculé avant la crise des *subprimes* (juillet 2006 à janvier 2007) puis utilisé comme paramètre fixe des calculs de *VaR* pour les autres sous périodes de tests de

¹⁹⁹ $\varepsilon \in [0 ; 1]$.

²⁰⁰ D'espérance nulle et d'écart-type nul.

²⁰¹ A 99% le quantile est tabulé : $z \sim 2,326$.

l'échantillon (février 2007 à mars 2011).

Nous conservons la valeur de σ , calculée pendant la période de crise des *subprimes*, et l'utilisons dans les calculs de la $VaR^{99\%}$ quotidienne pour les autres sous périodes de tests. Dans notre calcul de la $VaR^{99\%}$, nous n'intégrons pas les rendements quotidiens car ces derniers sont « négligeables ».

En souhaitant rester cohérent avec le cadre de nos travaux, nous choisissons de calculer σ avant la crise des *subprimes* pour mesurer si une combinaison de données hétérogènes peut avoir une influence sur les pertes estimées pendant les autres sous-périodes de tests.

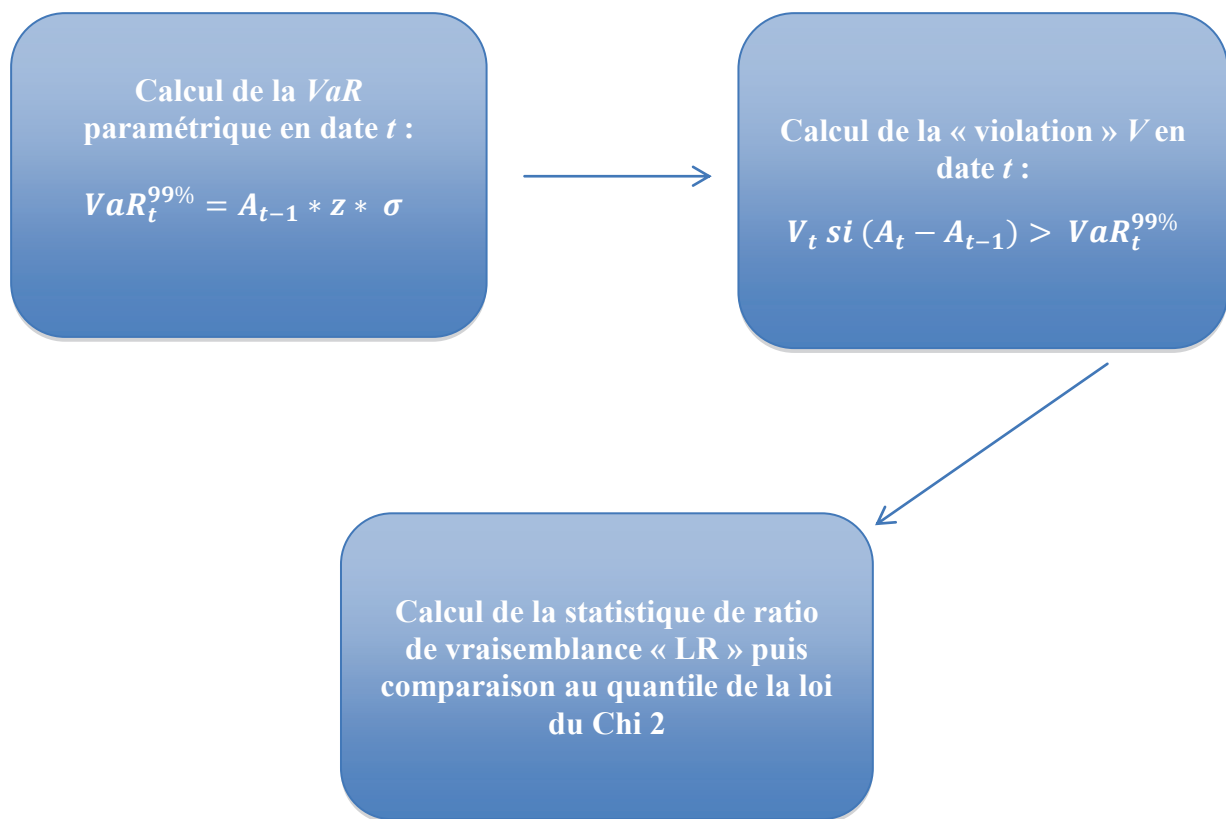
Une VaR paramétrique, reposant sur une distribution normale centrée réduite, est une fonction de σ : une « fonction monotone croissante de la variance des rendements » (Bodson et al., 2010). Telle qu'elle est strictement calculée, la VaR est une mesure « relative » (Racicot et Théorêt, 2006) car l'espérance (probabiliste) mathématique des rendements du portefeuille n'est pas prise en compte selon les auteurs²⁰².

Le test de Kupiec (1995) est un test de couverture non-conditionnelle qui permet d'estimer statistiquement si les pertes réalisées sont supérieures aux pertes estimées (figure 15) : si tel est le cas, nous sommes en présence d'une « violation ». Ce test repose sur une hypothèse forte d'indépendance des « violations » entre elles : d'où le caractère non-conditionnel²⁰³.

²⁰² Pour obtenir une mesure « absolue » Racicot et Théorêt (2006). Les auteurs détaillent comment s'opère le calcul d'une VaR par une méthode historique.

²⁰³ Pour plus d'informations voir Christoffersen (1998).

Figure 15 - Principales étapes du test de Kupiec (1995)



Nous traitons V_t (figure 15) comme un booléen : nous pouvons affecter une valeur 1 à une violation, une valeur 0 sinon (pas de violation). Nous sommes toutes les violations N pour la période des *subprimes* que nous divisons par le nombre d'observations (jours) T de la période.

Le taux « d'échec » ou *failure* est défini par la fréquence empirique des violations V_t :

$$\frac{N}{T}$$

L'objectif du test de ratio de vraisemblance est de déterminer si les pertes estimées sont statistiquement significatives. Le taux d'échec est un estimateur convergent du taux de couverture (ici 99%). La statistique de vraisemblance (« Likelihood ratio » ou « LR ») associé au test de Kupiec (1995) est :

$$LR = 2 \ln \left(\frac{N}{T} \right)^N \left(1 - \left(\frac{N}{T} \right)^{T-N} \right) - 2 \ln \left(\alpha^N - (1 - \alpha)^{T-N} \right),$$

avec $\alpha = 1\%$.

Avec un taux de couverture de la *VaR* à $\alpha\%$, l'hypothèse nulle du test de Kupiec (1995) est :

$$H_0: E(V_t) = \alpha.$$

Nous comparons la statistique « LR » au quantile associée à la loi du Chi 2 à 1 degré de liberté²⁰⁴.

La significativité statistique n'est validée que pour les portefeuilles composés en présence de données hétérogènes et ceci pendant la période des *subprimes* (tableau 22). Les tests empiriques, établissent que l'introduction de données hétérogènes a permis de réduire les probabilités de pertes financières pendant la crise des *subprimes*. Les pertes²⁰⁵ quotidiennes sont réduites lors de cette période, ce qui n'a pas été observé pour les autres périodes de tests de notre échantillon.

²⁰⁴ 3,84 à 5%.

²⁰⁵ En unité monétaire, l'euro dans le cadre de nos tests.

Tableau 22 - Test de Kupiec (1995) d'un portefeuille d'arbitrage selon le type de données et la période de tests

Type de données	Période de tests	Statistique "LR" (5%)	Valeur critique du Chi 2 (1 ddl, 5%)	Significativité
Données hétérogènes	Pendant la crise des <i>subprimes</i> (février 2007 - mars 2009)	3,55 **	3,84 **	Significatif
Données homogènes		5,40		Non-significatif
Données hétérogènes	Après la crise des <i>subprimes</i> (avril 2009 - mars 2011)	28,43	3,84 **	Non-significatif
Données homogènes		6,11		Non-significatif

** : $p < 0,05$.

Le type de données est en première colonne. La période de tests en dehors de l'échantillon (juillet 2006 – janvier 2007) est en deuxième colonne. La statistique de ratio de vraisemblance de Kupiec (1995) est en troisième colonne. Le quantile, valeur critique du Chi2 à 1 degré de liberté et à 5% de risque de première espèce est en quatrième colonne. Le résultat du test (caractère statistiquement significatif ou non) est en cinquième colonne.

En période de crise des *subprimes*, l'estimation des pertes probables est statistiquement significative uniquement en présence de données hétérogènes (tableau 22). Après la crise des *subprimes*, aucune significativité statistiques n'est recensée ni en présence de données hétérogènes ni en présence de données homogènes (tableau 22). Prendre en considération des données hétérogènes semble atténuer le risque de pertes financières uniquement en temps de crise pour des portefeuilles d'arbitrage diversifiés de manière « naïve ».

A titre d'information, un document de recherche de la Banque de France (Bourgey, 2012) permet d'analyser les impacts de la crise de l'endettement sur les marchés et le système bancaire. Cette institution analyse cet impact sur les banques en observant que de nombreuses banques européennes ont connu une forte hausse de leur *Value-at Risk (VaR)* entre mi-2007 et mi-2009.

L'expansion des fonds alternatifs *hedge funds* suivie par la contraction de leur nombre a été un phénomène important dans l'évolution de la gestion du risque dans le sens où le comportement des rendements devenait fortement « skewed »²⁰⁶ et « leptokurtique »²⁰⁷. Des enjeux croissants de mesure du risque s'en sont suivis et des questions ont émergé sur la mesure de la *VaR* (Bodson et al., 2010) telles que la non-normalité des densités de rendements d'actifs et la probabilité de pertes extrêmes au-delà d'un intervalle de confiance²⁰⁸. Bien que

²⁰⁶ Distribution asymétrique.

²⁰⁷ Queue de distribution épaisse.

²⁰⁸ La *Modified VaR* et la *Conditonal VaR* ont par exemple respectivement répondu à ces enjeux. Des

la *VaR* reste une mesure débattue, elle demeure un indicateur fréquemment utilisé par les régulateurs et les professionnels (Hull, 2009).

Parmi les facteurs qui ont contribué aux ventes massives d'actifs sur le marchés actions en août 2007, Khandani et Lo (2007) évoquent le *timing* inconnu des bouleversements sur le marchés du crédit lié aux *subprimes*. La prise en compte de données hétérogènes est effectuée dans l'optique d'anticiper si possible de futurs bouleversements sur les marchés financiers et d'éviter « le climat de peur et de panique » occasionné par ces irrégularités hypothécaires.

L'intégration de données hétérogènes, dans notre périmètre de recherche, semble permettre l'anticipation de pertes monétaires massives dans un univers d'arbitrage multi-classes d'actifs qui dépasse le cadre des marchés actions étudiés par les auteurs. En élargissant le spectre d'analyse de nos données, nous participons à la réflexion sur l'« accroissement de la sensibilité au risque des *managers* et des investisseurs » qu'ont entraîné les *subprimes*. Nos résultats empiriques viennent compléter et tempérer les conclusions des auteurs sur le manque de pertinence des outils « standards » comme la *VaR* et les distributions normales.

L'introduction de données hétérogènes vient compenser le risque d'« endogénéité » lorsque des stratégies de portefeuilles peuvent être pénalisées par leur propre évolution et non pas par des facteurs exogènes. La considération d'indicateurs de conjoncture, de sentiment et de volatilité semble générer un signal anticipé pour notamment contourner ce risque d'« endogénéité » et rendre plus pertinent des mesures de risque dites « standards » (dont la *VaR*). Nous participons à la discussion ouverte sur le « risque de ruine » que toute stratégie d'investissement fructueuse est censée incorporer. Afin de modéliser de manière plus sophistiquée les dynamiques des marchés, les « autres informations de marché » viennent enrichir l'estimation du risque de liquidité et l'estimation des facteurs communs d'évolution des actifs, qui permettent d'anticiper un risque de liquidation d'une position d'investissement. L'ensemble de ces éléments permet une meilleure mesure et gestion du risque.

indicateurs plus récents ont été développés comme le ratio Oméga (Keating et Shadwick, 2002) ou le ratio de Sortino (Sortino et Meer, 1991). Une extension des travaux de Sortino et de Van der Meer permet d'introduire les moments partiels inférieurs (Kaplan et Knowles, 2004).

2. Eléments de robustesse

L'étude de la robustesse est une étape déterminante de nos recherches dans la mesure où cette phase nous permet d'exposer différents éléments de stabilité de nos méthodes d'arbitrage issues d'une allocation « naïve ». Les résultats empiriques ont une portée plus large si ces derniers sont issus de méthodes dites « robustes ». L'analyse de cette robustesse s'effectue à trois niveaux : en vérifiant que les « autres informations de marché » ont un effet « réel » sur les contributions des classes d'actifs à la rentabilité et non un effet de *data mining*, en simulant un nombre important de *backtestings* (intégrant une composante aléatoire) pour calculer l'erreur moyenne de notre approche et en étendant la période d'« apprentissage » des *backtestings* pour observer la persistance possible des caractéristiques des portefeuilles d'arbitrage.

2.1. Simulations en présence de variables aléatoires

Nous souhaitons valider le fait que l'introduction d'« autres informations de marché » dans notre processus d'analyse de données apporte une valeur ajoutée « non-artificielle ». Nous souhaitons éviter d'observer une valeur ajoutée « artificielle » sur les données, produite par un effet de « *data mining* » (Avellaneda et Lee, 2010 ; Blitz et Vliet, 2008). Les tests empiriques faisant intervenir un nombre important de données peuvent produire des effets « *data mining* ». Afin d'éviter tout biais provoqué par le traitement et l'utilisation des données hétérogènes, nous décidons de remplacer les « autres informations de marché » par des variables aléatoires²⁰⁹. En introduisant ce phénomène aléatoire, nous souhaitons vérifier que l'influence des données relatives aux « autres informations de marché » sur le comportement des portefeuilles d'arbitrage issus d'une allocation « naïve » n'est pas fallacieuse.

Le but de ce test de robustesse est de vérifier si les résultats précédemment obtenus sur les contributions par classe d'actifs peuvent être confirmés ou non. La présence de variables aléatoires en remplacement des « autres informations de marché » permet d'observer si les portefeuilles d'arbitrage évoluent de manière similaire ou non. Si l'évolution des portefeuilles d'arbitrage en présence de données aléatoires est proche de celle constatée en présence d'« autres informations de marché », nous estimons que les « autres informations de marché » ne sont pas remplaçables par des données quelconques (de types variables aléatoires).

²⁰⁹ Les variables aléatoires ont été générées à l'aide du simulateur variable aléatoire RAND() ou Rnd sous VBA/Excel.

Nos tests sur le passé sont effectués avec les 32 séries de données de cotation et 11 séries de variables aléatoires qui remplacent les 11 anciennes séries « autres informations de marché » (tableau 23).

Tableau 23 - Composition de l'échantillon du test de robustesse selon le type de données

Composition de l'échantillon par type de données	N
Obligations	4
Matières premières	5
Actions (indices)	10
Devises	13
Variables aléatoires	11
Total	43

Sources : Thomson Reuters Datastream,
Simulateur de variables aléatoires sous
VBA/Excel

En partant de la gauche, les types de données sont en première colonne. Le nombre de données est en deuxième colonne.

Les 11 séries de variables aléatoires remplacent les 11 séries d'« autres informations de marché ». A la suite des tests empiriques en présence de variables aléatoires, nous constatons des différences de significativité statistique au niveau des contributions à la rentabilité par classes d'actifs par rapport à notre échantillon incluant d'« autres informations de marché ». En présence de variables aléatoires, les corrélations significatives ne sont pas les mêmes qu'en présence d'« autres informations de marché » (tableau 24)²¹⁰.

En présence de variables aléatoires, deux corrélations entre contributions par classe d'actifs sont statistiquement significatives : celle entre les classes « matières premières » et « devises », puis celle entre les classes « devises » et les « actions » (tableau 24). Différemment, en présence d'« autres informations de marché » nous obtenons une significativité entre « obligations » et « actions », « obligations » et « devises », puis « devises » et « actions ».

La seule corrélation entre contributions communément significative en présence de

²¹⁰ Les résultats des tests pour les données hétérogènes sont repris du tableau 20.

données aléatoires et d'« autres informations de marché ») est celle qui existe entre les classes « devises » et « actions » (tableau 24). Cette significativité est elle-même de nature différente, en présence d'« autres informations de marché » la corrélation est négative (-6,13%) alors qu'elle est positive en présence de données aléatoires (13,60%).

Tableau 24 – Test de robustesse en présence de variables aléatoires : significativité statistique des corrélations entre les contributions par classe d'actifs à la rentabilité d'un portefeuille d'arbitrage selon le type de données pour la période de tests de l'échantillon (2006-2011)

Données homogènes	Type de données	Coefficient de corrélation de Pearson	t stat
Obligation et matières premières	Données hétérogènes	0,13%	0,04
	Données homogènes et variables aléatoires	4,60%	1,60
Obligations et actions	Données hétérogènes	6,75%	2,35 **
	Données homogènes et variables aléatoires	0,79%	0,28
Obligations et devises	Données hétérogènes	-6,66%	2,32 **
	Données homogènes et variables aléatoires	0,59%	0,21
Matières premières et actions	Données hétérogènes	0,32%	0,11
	Données homogènes et variables aléatoires	-2,61%	0,91
Matières premières et devises	Données hétérogènes	-0,78%	0,27
	Données homogènes et variables aléatoires	-7,18%	2,51 **
Devises et actions	Données hétérogènes	-6,13%	2,13 **
	Données homogènes et variables aléatoires	13,60%	4,77 ***

** : $p < 0,05$; *** : $p < 0,01$

En partant de la gauche, les données homogènes sur lesquelles porte le calcul de la corrélation sont en première colonne. Le type de données utilisé pour composer nos portefeuilles d'arbitrage est en deuxième colonne. Le coefficient de corrélation de Pearson entre contributions de classes d'actifs à la rentabilité annualisée est en troisième colonne. La *t-stat* de Student est en quatrième colonne.

Lorsque les significativités sont obtenues pour les mêmes classes et a fortiori de mêmes signes, nous envisageons un effet *data mining* de la présence des « autres informations de marché ». Or cela n'est pas le cas.

Le test de robustesse révèle que la prise en compte de variables aléatoires ne permet pas de reproduire les caractéristiques de notre portefeuille d'arbitrage en présence d'« autres informations de marché ». L'influence des « autres informations de marché » sur notre portefeuille d'arbitrage ne semble pas être, au niveau de ce test de robustesse, un artefact : l'apport des « autres informations de marché » est ainsi quantifié.

2.2. Simulations de Monte Carlo

Une deuxième voie d'investigation en matière de robustesse pour notre approche appartient aux techniques de simulation. Ces techniques permettent de répéter des tests sur le passé (*backtestings*). Il s'agit de simuler plusieurs fois la construction quotidienne d'un portefeuille d'arbitrage en diversification « naïve » et d'obtenir un historique de rentabilités quotidiennes pour la période de tests de notre échantillon. Chaque fois que nous effectuons un test sur le passé, nous simulons une trajectoire. De manière arbitraire, nous simulons 100 trajectoires et estimons l'erreur moyenne²¹¹ associée à ces différentes simulations.

La composante aléatoire de notre expérience tient au fait que, lorsque nous répétons les tests sur le passé, nous introduisons pour chaque trajectoire un nouveau nombre aléatoire²¹² lors des procédures de quantification²¹³ du risque spécifique.

Nous effectuons ces simulations en présence de données hétérogènes puis en présence de données homogènes. Nous suivons la même logique que la technique de Monte Carlo (Metropolis et Ulam, 1949). Nous choisissons de simuler 100 trajectoires : soit 100 fois le test sur le passé qui consiste à composer quotidiennement un portefeuille d'arbitrage « euro neutre » en diversification « naïve ». Pour une trajectoire, nous avons 1213 valeurs de notre portefeuille : une valeur à chaque date quotidienne de clôture entre juillet 2006 et mars 2011.

²¹¹ Nous définissons cette erreur moyenne plus loin dans cette section.

²¹² A l'aide du simulateur de nombres aléatoires grand(). Nous simulons des lois normales standard.

²¹³ Voir chapitre 2.

Nous calculons chaque jour la variation du portefeuille pour la simulation i de l'ensemble des simulations S :

$$S = \sum_{i=1}^{100} S_i$$

L'objectif final est de calculer M , la moyenne des variations de ces 100 trajectoires, et leurs écart-types σ pour estimer l'erreur moyenne σ_{MC} . Pour calculer M , nous effectuons la moyenne arithmétique annualisée des 100 portefeuilles : nous aurons 100 moyennes puisque nous simulons 100 trajectoires. Pour obtenir M , nous calculons la moyenne arithmétique de ces 100 moyennes. Nous calculons ensuite l'écart-type σ de ces 100 moyennes. L'erreur moyenne σ_{MC} est alors calculée comme suit :

$$\sigma_{MC} = \frac{\sigma}{\sqrt{S}}$$

avec S , le nombre de trajectoire soit $S = 100$. Chaque trajectoire i de S est considérée comme une variable aléatoire.

L'influence des données hétérogènes se manifeste par une erreur moyenne σ_{MC} de 9,83% inférieure à celle obtenue pour un univers de données homogènes de 11,53% (tableau 25). En introduisant d'« autres informations de marché » dans notre spectre d'analyse, nous obtenons en moyenne une erreur plus faible.

Tableau 25 - Erreur moyenne des simulations de Monte Carlo selon le type de données pour la période de tests de l'échantillon (2006 – 2011)

Type de données	Moyenne des taux de variations	Erreur moyenne
Données hétérogènes	11,95%	9,80%
Données homogènes	12,70%	11,53%

Le type de données est en première colonne. La moyenne des taux de variation du portefeuille est en deuxième colonne. L'erreur moyenne est en troisième colonne. L'erreur moyenne se calcule comme indiquée ci-dessous : $\frac{\sigma}{\sqrt{S}}$, avec σ , l'écart-type des taux de variation et S , le nombre de trajectoires. Nous simulons 100 trajectoires ($S = 100$).

Les résultats des simulations de Monte Carlo peuvent être mis en perspective de la baisse de la volatilité moyenne des rendements des portefeuilles d'arbitrage obtenus dans les sections précédentes. Pour un nombre répété de simulations, nous obtenons des éléments de volatilité et de rentabilité cohérents avec les résultats des tests empiriques. En moyenne, la rentabilité et la volatilité sont inférieures en présence de données hétérogènes (tableaux 19 et 25). La répétition des trajectoires n'a pas modifié en moyenne les caractéristiques des portefeuilles d'arbitrage issus d'une allocation « naïve ». La baisse de la volatilité des portefeuilles d'arbitrage semble être liée à une influence durable et mesurable des données hétérogènes pour la période de tests de notre échantillon après la simulation de plusieurs scénarios.

Indépendamment du type de données, la stabilité de notre processus se matérialise par deux erreurs²¹⁴ moyennes « relativement » faibles au regard de la composante aléatoire introduite par les procédures de quantification²¹⁵. Les algorithmes de quantification peuvent générer une « forte » volatilité des taux de variation des données, ce qui n'est pas le cas avec notre approche lors des tests (tableau 25). D'autre part, le nombre relativement « restreint » de données dans l'échantillon²¹⁶ utilisé pour nos procédures de quantification peut entraîner une volatilité élevée : ce qui n'est pas le cas lors des simulations de robustesse. Pour obtenir des grilles de probabilités dites « stables » pour chacun des actifs, il est préférable d'avoir un nombre plus élevé de données en phase d'« apprentissage » que celui que nous avons utilisé. Malgré le nombre de données relativement limité pour nos procédures de quantification, nous obtenons des grilles de probabilités qui semblent engendrer des signaux robustes pour la constitution des portefeuilles d'arbitrage.

2.3. Extension de la période d'« apprentissage » de l'échantillon

Une troisième investigation de robustesse porte sur la longueur de la période d'« apprentissage » de l'échantillon²¹⁷ lors des tests empiriques. Nous rallongeons dans ce qui suit cette période d'« apprentissage ». Ceci a pour conséquence de décaler mécaniquement (retarder) la date de départ des tests empiriques.

Nous souhaitons mesurer si le rallongement de cette période « apprentissage » modifie

²¹⁴ Une en présence de données hétérogènes, une autre en présence de données homogènes.

²¹⁵ Voir le chapitre 1.

²¹⁶ Entre août 2004 et fin juin 2006.

²¹⁷ Voir les chapitres 1 et 2 pour la référence à la quantification.

le comportement des portefeuilles d'arbitrage obtenus lors des tests empiriques²¹⁸. Nous changeons le point de départ des tests sur le passé, en reculant la date finale de la période « apprentissage » : nous retardons ainsi la date de départ des tests. Nous analysons par la suite les différences et similarités des portefeuilles d'arbitrage pour la période initiale de tests (juillet 2006 à mars 2011) et la nouvelle période de tests issue de l'extension de la période d'« apprentissage ».

En remplacement d'une période d'« apprentissage »²¹⁹ couvrant l'intervalle de temps 2004-2006²²⁰, nous considérons une période d'« apprentissage » plus longue : d'août 2004 à février 2009. Nous augmentons la longueur de cette période « apprentissage » d'environ 11 cycles de publications de résultats trimestriels²²¹ conformément à Avellaneda et Lee (2010). Chaque cycle trimestriel de publications de résultats dure environ 60 jours selon les auteurs.

Le décalage dans le temps de la date de départ des tests empiriques permet de conserver les principales caractéristiques de l'allocation d'actifs et de volatilité au sein des portefeuilles d'arbitrage (tableau 26 et 27).

Tableau 26 – Test de robustesse à l'aide d'une période d'« apprentissage » étendue : comparaison de l'allocation d'actifs d'un portefeuille d'arbitrage selon le type de données pour une période de tests décalée (février 2009 - mars 2011)

Type de données	Classe d'actifs				Total
	Obligations	Matières premières	Actions	Devises	
Données hétérogènes	17%	15%	33%	35%	100%
Données homogènes	11%	19%	30%	40%	100%

En partant de la gauche, le type de données est en première colonne. L'allocation d'actifs par classe d'actifs est renseignée de la deuxième à la cinquième colonne. Le total des pondérations est en sixième colonne.

Nous vérifions que la prise en compte de données hétérogènes induit aussi un changement d'allocation lorsque nous élargissons la taille de l'échantillon d'« apprentissage » (tableau 26). Une constante entre les tests initiaux (tableau 16) et les tests avec un rallongement de la période d'« apprentissage » (et donc un décalage du point de départ des

²¹⁸ Voir les sections précédentes.

²¹⁹ Nous rappelons que la période d'« apprentissage » fait référence à une période où le processus de quantification permet de compresser des informations (des données). Cette période est composée d'un ensemble de variables supposées aléatoires qui sera ramené à un nombre fini de points dans le cadre d'une grille de probabilité.

²²⁰ Voir chapitre 3.

²²¹ Soit un rallongement de 683 jours de la période initiale d'« apprentissage ».

tests), est que les classes « actions » et « devises » restent celles sur lesquelles la majorité de l'allocation est portée (tableaux 16 et 26). Dans chacune des configurations (avec et sans rallongement de la période d'« apprentissage »), en présence de données hétérogènes ou non, ces deux classes totalisent presque 70% de l'allocation d'actifs des portefeuilles d'arbitrage (tableaux 16 et 26).

En rallongeant la période d'« apprentissage », nous conservons globalement les mêmes caractéristiques de rentabilité et de volatilité des portefeuilles d'arbitrage. Nous obtenons un ratio rentabilité/volatilité équivalent en présence de données hétérogènes et en présence de données homogènes (tableau 27). Bien que la valeur de ce ratio (2,2) diffère « légèrement » par rapport à celui de la période initiale de tests (2,5) sur notre échantillon (tableaux 19 et 27), nous conservons l'équivalence du ratio quelles que soit les données et les périodes de tests.

Tableau 27 – Test de robustesse à l'aide d'une période d'« apprentissage » étendue : rentabilité et volatilité annualisées d'un portefeuille d'arbitrage selon le type de données pour une période de tests décalée (février 2009 - mars 2011)

Données	Période	Rentabilité (1)	Volatilité (2)	(1) / (2)
Données hétérogènes	Février 2009 -	12,61%	5,86%	2,2
Données homogènes	mars 2011	14,49%	6,46%	2,2

En partant de la gauche, le type de données est en première colonne. La période de tests de l'échantillon est en deuxième colonne. La rentabilité annualisée est en troisième colonne. La volatilité annualisée est en quatrième colonne. Le ratio rentabilité annualisée/volatilité annualisée est en cinquième colonne.

Le phénomène qui explique cette équivalence reste identique pour les deux périodes de tests de notre échantillon : une baisse à la fois de la volatilité et de la rentabilité, en présence de données hétérogènes par rapport à la seule prise en compte de données homogènes (tableau 19 et 27). Avec une extension de la période d'« apprentissage », le ratio rentabilité/volatilité est le fruit des mêmes manifestations que celles des premiers tests empiriques²²² : une volatilité moins élevée (5,86% contre 6,46%, tableau 27) compensant une rentabilité moins forte (12,61% contre 14,49%, tableau 27) en présence de données hétérogènes.

Le contexte économique et financier change entre la période initiale des tests

²²² Voir les sections précédentes.

empiriques (tableau 1) et la période déterminée pour notre test de robustesse (février 2009 – mars 2011). Notre échantillon d'« apprentissage » rallongé (pour les besoins de notre test de robustesse) inclut la quasi-totalité de la période de crise des *subprimes*, ce qui n'est pas le cas pour période d'« apprentissage » des tests empiriques²²³ hors-robustesse. Malgré ce changement de contexte, l'évolution des portefeuilles en termes de rentabilité et de volatilité d'une part, et de « couple » rentabilité/volatilité d'autre part, reste semblable.

Des tests empiriques sur des données passées permettent de mesurer l'influence de la prise en compte de données hétérogènes sur les portefeuilles d'arbitrage « euro neutre » diversifiés de manière « naïve ». A l'aide de différents tests de robustesse, nous mettons en évidence des éléments de stabilité de notre approche de l'arbitrage statistique en univers multi-classes d'actifs en présence de données hétérogènes et homogènes. Nous trouvons des éléments de cohérence par rapport aux résultats obtenus lors des tests empiriques précédents, notamment en termes de rentabilité, de volatilité, d'allocation et de corrélation des données de cotations.

L'ensemble des résultats empiriques obtenu dans ce chapitre se situe dans le cadre d'un portefeuille d'arbitrage diversifié de manière « naïve ». D'autres cadres théoriques, permettant d'utiliser des programmes d'optimisation, rompent avec l'allocation « naïve ». Les allocations ainsi permises, que nous qualifions d'allocations de « diversification optimale », sont étudiées dans notre cinquième chapitre. Les mêmes tests empiriques y sont menés dans un cadre de diversification optimale.

²²³ Voir le chapitre 3.

CHAPITRE 5. PORTEFEUILLE D'ARBITRAGE « EURO NEUTRE » EN DIVERSIFICATION OPTIMALE

Nous poursuivons notre étude empirique en construisant des portefeuilles d'arbitrage « euro neutre » issus d'une allocation optimale en présence de données hétérogènes et homogènes. Nous analysons l'impact d'une combinaison de données hétérogènes sur les portefeuilles d'arbitrage. Nous présentons le cadre de la diversification optimale de nos portefeuilles d'arbitrage : le contexte théorique et les contraintes empiriques de notre processus d'optimisation. Nous mesurons les effets des données hétérogènes sur l'évolution des portefeuilles d'arbitrage avant d'effectuer différents tests de robustesse des résultats obtenus. Nous terminons ce chapitre par une analyse comparative des résultats empiriques obtenus dans le cadre d'une diversification « naïve » avec ceux issus d'une diversification optimale.

1. Cadre et contraintes de la diversification optimale

Les tests empiriques sont menés dans un cadre d'allocation optimale. Cette diversification optimale des portefeuilles d'arbitrage fait intervenir des procédures mathématiques plus complexes que la stratégie « 1/N ». Nous définissons le cadre empirique des tests réalisés et nous formalisons l'optimisation mise en place pour progresser dans nos travaux d'allocation d'actifs et de *market timing*.

1.1. Cadre théorique de la diversification optimale

Plusieurs types d'allocations sont envisageables en gestion. Nous détaillons ces types de diversification de portefeuille dans le chapitre 1. Dans ce premier chapitre, nous analysons les atouts et les limites que comporte la gestion de portefeuille dans un cadre optimal. Nous rappelons que les apports de l'allocation « 1/N » en termes de rentabilité, de ratio de Sharpe et de rendement équivalent certain sont étayés par DeMiguel et al. (2009), surtout dans le cas où le nombre d'actifs choisi est élevé.

Dans le chapitre 4, nous faisons appel au cadre de diversification « naïve » au sein duquel nous effectuons divers tests empiriques. Ici, nous proposons un cadre de

diversification alternatif : celui de l'optimisation. Nous nous concentrons sur une approche optimale afin de mesurer l'apport de données hétérogènes sur les portefeuilles d'arbitrage. A la seule différence du caractère optimal de l'allocation, les critères de constitution des portefeuilles d'arbitrage retenus pour les tests empiriques de ce chapitre sont les mêmes que ceux du chapitre 4. Les portefeuilles d'arbitrage sont toujours : « euro neutre », diversifiés de manière optimale au sein d'une équi-répartition des montants investis, sans levier et avec un taux de coût de transactions de 3% annualisé.

Nous explorons la spécificité d'une gestion de portefeuille optimisée. Nous retenons une approche plus proche de celle en termes de portefeuilles « *minimum variance* » (Clarke, Silva, et Thorley, 2006) que de l'optimisation « moyenne-variance » (Markowitz, 1956 ; Markowitz, 1952).

1.2. Contraintes de la diversification optimale

En présence de données hétérogènes et de données homogènes, une nouvelle voie d'investigation nous permet d'agir sur la variance des portefeuilles en résolvant un programme de minimisation sous contraintes. Suite à cette procédure d'optimisation, nous analysons la structure des portefeuilles (rentabilité, volatilité, allocation, contribution par classe d'actifs et corrélations).

Précisons quelques généralités sur les procédures d'optimisation avant de présenter les contraintes de notre programme de minimisation. L'objectif est de trouver le minimum global d'une fonction²²⁴ qui va s'annuler en un point x , ce dernier pouvant être vu comme un vecteur de poids d'actifs Q du point de vue des Sciences de Gestion. Géométriquement, le point x aura d coordonnées dans un plan et chacune des coordonnées représentera un composant du vecteur Q . Un composant est un poids du portefeuille. La procédure de minimisation, telle que l'approche « *minimum variance* » (Clarke, Silva, et Thorley, 2006), revient à choisir le vecteur de poids (dans la matrice de covariances) qui correspond au portefeuille (d'actifs) dont la variance est minimale. Plusieurs travaux détaillent les différentes formes d'optimisation en gestion de portefeuille (Broquet et al., 2004 ; Viviani, 2001 ; Elton et Gruber, 1995).

²²⁴ Tel que $f'_x = 0$ et $f''_x > 0$.

La traduction mathématique du programme d'optimisation quadratique²²⁵ est présentée ci-dessous :

$$X^{optim} = \underset{X \in C}{\operatorname{argmin}} Q'YQ$$

L'écriture matricielle de la variance d'un portefeuille correspond à $Q'YQ$, avec Q le vecteur de poids des actifs, Q' , le vecteur transposé de Q et Y , la matrice de covariances des taux de variation des actifs. La minimisation de la fonction est réalisée sous contraintes²²⁶. Les contraintes sont divisées en deux groupes : les contraintes d'égalité et les contraintes d'inégalités (figure 16).

Le but principal de notre programme d'optimisation est de réduire la volatilité des portefeuilles d'arbitrage en conservant l'ensemble des critères de notre approche mentionnés précédemment. Afin de respecter une logique d'arbitrage²²⁷, nous n'imposons pas la « non-négativité » (Best et Grauer, 1991) au sein de nos contraintes. Le critère de « non-négativité » consiste à proscrire les ventes d'actifs, autrement dit à interdire des poids négatifs dans un vecteur Q des poids du portefeuille²²⁸. Cette contrainte de « non-négativité » est en contradiction avec la logique de l'arbitrage. De même, les travaux de Best et Grauer (1991) n'intègrent pas toujours la « *budget constraint* » (somme des poids du portefeuille est égale à 1) dans la liste des contraintes. Notons que la non prise en compte de la « *budget constraint* » permet notamment de conserver une part de montants non investis dans le portefeuille²²⁹. Les auteurs rappellent, qu'en pratique, la « *budget constraint* » n'est pas obligatoire. Nous rappelons que nous investissons systématiquement l'ensemble des montants des portefeuilles d'arbitrage de manière équi-répartie et la somme des poids est égale à 0 (figure 16).

Afin de limiter les effets du risque d'estimation sur le portefeuille, Viviani (2001) propose de « relâcher le lien entre les indicateurs statistiques et les pondérations choisies » et de choisir des méthodes qui « malgré leur manque de fondement théorique...donnent d'assez bon résultats dans la pratique surtout dans des périodes de grandes incertitudes ».

²²⁵ Cette technique de minimisation est similaire celle des travaux de Lagrange du 18^{ÈME} siècle, utilisée dans divers champs de recherche y compris en Sciences de Gestion.

²²⁶ Les contraintes sont toutes linéaires.

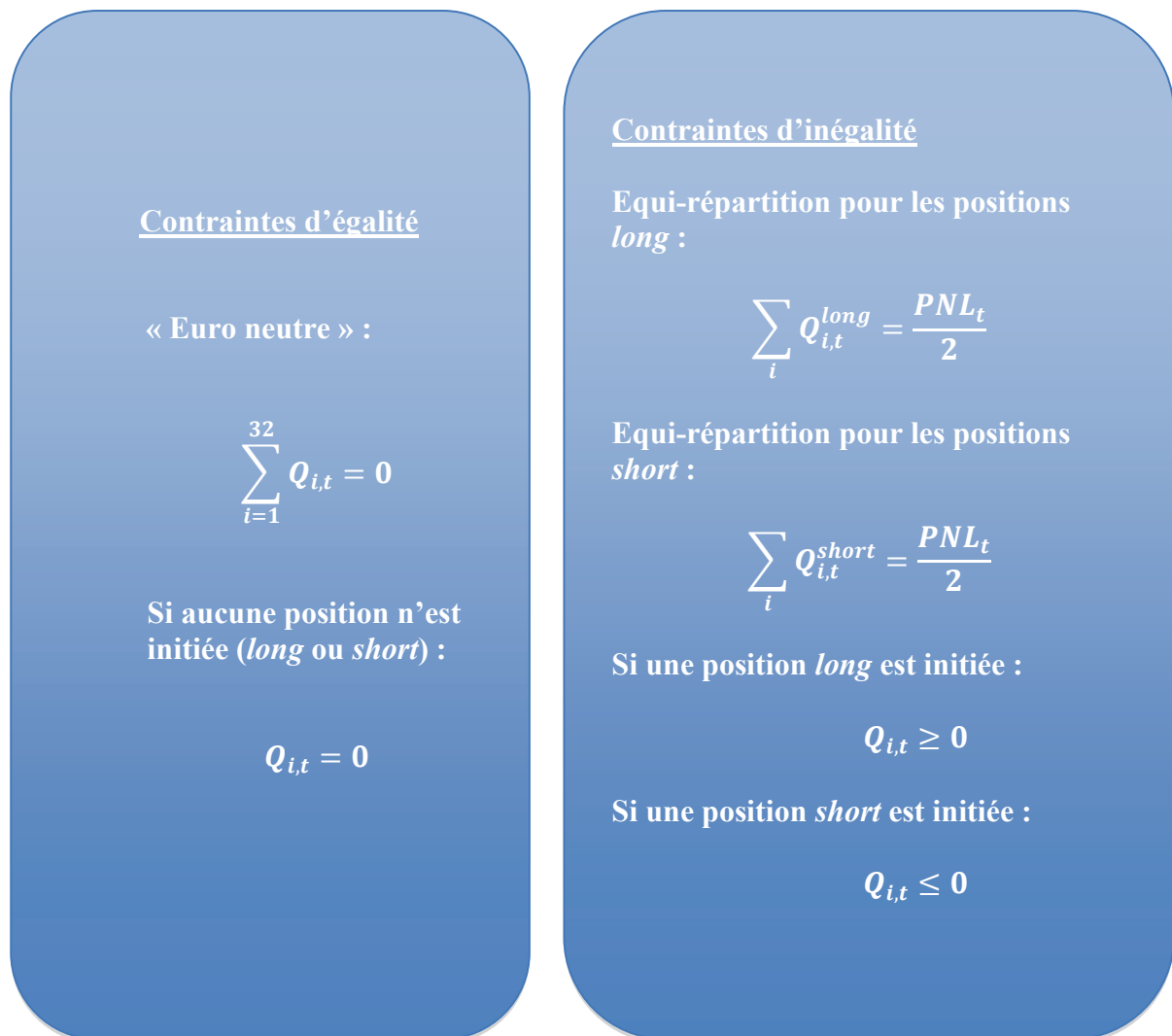
²²⁷ Achat et vente.

²²⁸ Notons que pendant la crise des *subprimes*, certaines ventes d'actifs à découvert étaient proscrites sur les marchés des actions : pour certains titres du secteur financier.

²²⁹ Dans la pratique, la conservation d'une part de liquidité est réglementaire, notamment pour faire face aux rachats de parts de fonds, mais peut aussi résulter d'un choix de gestion. Pour le cas des banques, les ratios de liquidité et de solvabilité sont imposés par les régulateurs nationaux et transnationaux.

L'équi-répartition permet de limiter les montants investis sur des positions *long* et *short* : or les contraintes sur les proportions peuvent limiter les effets des erreurs d'estimations sur la structure du portefeuille.

Figure 16 - Liste des contraintes de la diversification optimale de notre portefeuille d'arbitrage : minimisation de la volatilité sous contraintes



Dans notre approche, nous pondérons la position sur chaque actif par l'inverse de sa volatilité. Schématiquement, ce raisonnement vise à réduire l'importance d'une position (poids) dont la volatilité est élevée et d'augmenter l'importance d'une position dont la volatilité est relativement faible²³⁰. Cette technique n'est pas uniquement utilisée pour

²³⁰ Les procédures d'optimisations empruntent souvent des algorithmes tels que celui du gradient conjugué ou celui de Newton-Raphson comme c'est le cas pour le solveur sous Excel (Riva, 2008). Nous utilisons la fonction d'optimisation *qpsolve* sous Scilab. *Qpsolve* est la contraction de *quadratic program*

l'allocation de portefeuille, mais est aussi mobilisée dans d'autres contextes. Par exemple, la détermination des poids de l'indicateur avancé de l'économie aux Etats-Unis, publié par le *Conference Board*, est également régie selon ce même principe. Les séries composant cet indicateur (indice) ont des poids équivalents à l'inverse de l'écart-type du mois précédent.

Dans une logique similaire, Maillard et al. (2010) expliquent que pour un portefeuille « ERC » le poids x_i d'un actif i est inversement proportionnel à son β_i , $i=1,\dots,n$ tel que : $x_i = \frac{\beta_i^{-1}}{n}$. Plus le β d'un actif est élevé plus son poids dans le portefeuille sera faible et inversement. Les composants ou actifs ayant une « volatilité élevée, ou une forte corrélation avec les autres seront pénalisés ».

Dans la mesure où notre processus d'optimisation s'effectue dans un cadre globalement proche du « *Minimum-Variance Portfolio* » (MVP)²³¹, l'estimation des rendements n'entre pas en considération comme une information *a priori*. Notons que la matrice de variances-covariances ne sera pas la même selon que nous intégrons ou non les taux de variation des « autres informations de marché ».

Les poids des portefeuilles qui vérifient le fait que la variance soit minimale changent, compte tenu de la structure de la matrice de variances-covariances qui peut évoluer en présence de données hétérogènes ou non. Nous nous concentrons sur l'influence des données hétérogènes exercée sur les portefeuilles d'arbitrage issus d'une diversification optimale.

2. Influence des données hétérogènes sur des portefeuilles d'arbitrage « euro neutre » en allocation optimale

Nous ne portons pas encore de regard sur le type de données qui entrent comme paramètres dans ces modèles d'optimisation. Nous analysons ci-après la nature des données qui influencent la répartition des poids et l'évolution des portefeuilles d'arbitrage.

Dans un cadre optimal, l'élargissement du périmètre des informations (autres que les rendements passés de prix d'actifs) peut permettre d'améliorer l'estimation des paramètres nécessaires à l'optimisation. Les informations en question sont tout autant comptables, financières qu'économiques. Elles peuvent provenir d'horizons divers, tels que par exemple les séries économiques et sectorielles. La prise en compte d'autres données que les

solve. La fonction *qpsolve* permet de résoudre un programme quadratique linéaire comportant des contraintes (égalité, d'inégalité). Pour plus de détails sur la fonction *qpsolve* voir : http://help.scilab.org/docs/5.4.0/fr_FR/qpsolve.html.

²³¹ Nous nous différencions de ce cadre car nous autorisons des ventes et souhaitons rester « euro neutre ».

rendements historiques des actifs cotés revient à considérer ce que nous appelons les « autres informations de marché ».

Pour nos tests empiriques, l'optimisation consiste à minimiser la volatilité d'un portefeuille sous certaines contraintes²³² que nous définirons par la suite. Nous dépassons le paradigme « moyenne-variance » (Markowitz, 1956 ; Markowitz, 1952) pour nous rapprocher du portefeuille optimal « *minimum variance* » (Clarke, Silva, et Thorley, 2006) dans la mesure où nous cherchons à minimiser la volatilité à l'aide d'informations distinctes.

Nous décrivons et analysons l'apport de données hétérogènes en termes d'allocation d'actifs, de rentabilité, de volatilité, de « couple » rentabilité/volatilité à nos portefeuilles d'arbitrage « euro neutre » optimisés.

2.1. Changement d'allocations entre classes d'actifs

La minimisation de la volatilité (sous contraintes²³³) en présence de données hétérogènes impacte notre allocation d'actifs. Indépendamment de la présence ou non de données hétérogènes dans le processus de construction du portefeuille d'arbitrage, l'optimisation a permis de modifier la hiérarchie de l'allocation d'actifs définie dans un scénario de diversification « naïve » (tableaux 16 et 28). Nous analysons le lien entre le processus général de diversification optimale et les rotations d'allocations engendrées en présence de données hétérogènes.

La hiérarchie décroissante de l'allocation est identique en présence de données hétérogènes et en présence de données homogènes : « devises », « obligations », « actions » et « matières premières » (tableau 28). Au sein de cette hiérarchie, des changements de pondérations ont lieu en présence de données hétérogènes. L'effet des données hétérogènes sur les portefeuilles d'arbitrage devient tangible lorsque nous constatons des différences d'allocation entre un univers de données hétérogènes et homogènes. Alors que les allocations sont identiques pour les classes « obligations » et « matières premières », les différences d'allocation se manifestent pour celles des « actions » et des « devises » (tableau 28). En présence de données hétérogènes, la diversification optimale permet de surpondérer la classe d'actifs des « devises » (tableau 28) et de sous-pondérer la classe des « actions » par rapport à la présence de données homogènes (tableau 28).

²³² Décrites précédemment.

²³³ Précisées précédemment.

Tableau 28 - Comparaison de l'allocation d'actifs d'un portefeuille d'arbitrage optimisé selon le type de données pour la période de tests (2006-2011)

Type de données	Classe d'actifs				Total
	Obligations	Matières premières	Actions	Devises	
Données hétérogènes	35%	2%	11%	51%	100%
Données homogènes	35%	2%	24%	39%	100%

En partant de la gauche, le type de données est en première colonne. L'allocation d'actifs par classe d'actifs est renseignée de la deuxième à la cinquième colonne. Le total des pondérations est en sixième colonne.

Après avoir examiné la trace quantitative laissée par la présence des données hétérogènes, nous nous concentrons sur l'explication des changements d'allocation. Sans prendre en considération les « autres informations de marché », l'explication vient du fait que le processus d'optimisation attribue un poids faible à une classe qui a une volatilité élevée et inversement, un poids élevé lorsqu'une volatilité faible est historiquement calculée.

En prenant en compte la volatilité pour ajuster les pondérations des portefeuilles, une classe d'actifs ayant une volatilité « importante » a un poids plus « faible » et inversement. Ce processus permet notamment d'expliquer que l'allocation d'actifs sur la classe « matières premières » est « très faible » (tableau 28).

Les statistiques descriptives²³⁴ (tableau 11) sur la période de tests de l'échantillon permettent de recenser les plus forts écart-types des taux de variation des données de cotation de cette classe. Les écart-types de la classe « actions » sont moins élevés que ceux des « matières premières » mais restent en deuxième position décroissante (tableau 11). A l'inverse, les écart-types des taux de variation des classes « devises » et « obligations » sont relativement plus « faibles » d'où l'importance de leur allocation dans un univers optimal (tableau 28). Pour les « devises », la présence de données hétérogènes semble participer à l'augmentation de cette allocation (tableau 28) et à la réduction de celle obtenue sur les « actions » (tableau 28).

Dans le cadre d'une allocation tactique (« actions » et « obligations ») de type « moyenne variance » (Markowitz, 1952), Dahlquist et Campbell (2001) procèdent à une minimisation de la variance d'un portefeuille en présence d'information « conditionnelle » et

²³⁴ Calculées dans le chapitre 3.

inconditionnelle²³⁵. En intégrant des informations « conditionnelles » telles que des sources d'informations « fondamentales »²³⁶, les auteurs signalent que des changements d'allocations d'actifs peuvent se produire. Les auteurs soulignent l'imprévisibilité des marchés actions qui sont fortement volatiles et qui incorporent rapidement l'information, ce qui ne veut pas dire que l'information est intégrée instantanément. Ces derniers préfèrent s'employer à comprendre les variables qui pourraient influencer la prévisibilité des rendements d'actifs. Les « variables d'information » historiques peuvent impacter l'allocation internationale d'actifs surtout quand elles sont liées à des éléments de conjoncture. Du point de vue des auteurs, le fait que les rendements soient prévisibles permet de justifier la pertinence d'une allocation tactique²³⁷ qui surperformera une allocation passive. Notons que cette allocation tactique permet d'ajuster les pondérations des portefeuilles de manière fréquente²³⁸ à l'aide notamment de procédures d'optimisation. Les modifications d'allocations sont effectuées pour tenter de réaliser cette surperformance et bénéficier de rendements supérieurs à ceux obtenus avec une allocation statique.

Black et Litterman (1992) construisent une allocation d'actifs quantitative, multi-classes et optimale évoluant en fonction des « vues » ou opinions des investisseurs. Comme nous l'avons mentionné précédemment²³⁹, les auteurs rappellent la sensibilité des programmes d'optimisation à l'erreur d'estimation des rendements futurs qui favorise la construction de portefeuilles aux comportements « déraisonnables ». Ce type de dérèglement se traduit souvent par d'importantes positions (poids) *long* et *short*. Ils proposent ces « vues » pour tenter de pallier à l'erreur d'estimation des rendements futurs. Ces « vues » sont des sentiments « subjectifs » qui permettent de formuler des opinions quant aux valeurs relatives que les investisseurs attribuent aux classes d'actifs concernées. Les investisseurs conservent toutefois la liberté de ne prononcer aucune « vue », ce qui équivaut à une opinion neutre ou inchangée. Notons que ces opinions neutres sont également intégrées dans la plupart des données « autres informations de marché » de notre échantillon. Les auteurs intègrent ces préférences de marché à une allocation multi-classes d'actifs modélisée pour l'univers des actions, des devises et des obligations.

²³⁵ Prix d'actifs historiques.

²³⁶ Comme le positionnement d'une entreprise dans son cycle d'activité ou cycle économique.

²³⁷ Qui effectue des changements d'allocation à court terme pour s'ajuster à la conjoncture économique et l'évolution des marchés financiers. Voir l'Introduction Générale.

²³⁸ Les auteurs parlent de fréquences intra-journalières et journalières.

²³⁹ Voir chapitre 1.

Nous poursuivons nos recherches dans l'esprit des travaux précédemment cités par le fait que nous considérons des informations *a priori* (« autres informations de marché ») et que nous cherchons à évaluer leur impact sur les allocations d'actifs. Nous nous inscrivons dans la continuité de ces travaux car la présence de données hétérogènes semble modifier les allocations des portefeuilles optimisés. Nous progressons vers l'idée d'une pertinence d'informations complémentaires aux données de cotation et nous développons nos explorations dans un cadre d'arbitrage systématique multi-classes d'actifs.

Lorsque les portefeuilles sont issus d'une diversification optimale, le phénomène recherché est un rapprochement relatif entre les allocations²⁴⁰. Cette tentative de convergence d'allocations entre classes d'actifs permet maintenant de distinguer les éventuels écarts mutuels de contributions à la rentabilité, en présence de données hétérogènes. Le principe est de comparer les classes d'actifs entre elles après avoir effectué une normalisation par les écarts-types des variations. Les allocations (tableau 28) sur les classes d'actifs sont mises en perspectives des contributions à la rentabilité des portefeuilles d'arbitrage en présence de données hétérogènes et homogènes.

2.2. Modification de contribution par classes d'actifs à la rentabilité

La présence de données hétérogènes semble continuer d'exercer une influence mesurable sur les portefeuilles d'arbitrage. L'analyse des contributions par classe d'actifs à la rentabilité des portefeuilles optimisés permet de souligner cette empreinte. Pour toutes les classes d'actifs, les contributions à la rentabilité ont été modifiées en présence de données hétérogènes (tableau 29). Nous rappelons que la contribution d'une classe d'actifs à la rentabilité peut être définie comme une sous-poche de la rentabilité globale des portefeuilles d'arbitrage constitués pour la période de tests de l'échantillon.

²⁴⁰ Sans pour autant chercher à les égaliser.

Tableau 29 - Comparaison de la contribution par classe d'actifs à la rentabilité d'un portefeuille d'arbitrage optimisé selon le type de données pour la période de tests (2006-2011)

Type de données	Classe d'actifs				Total
	Obligations	Matières premières	Actions	Devises	
Données hétérogènes (1)	32,7%	7,9%	29,7%	29,7%	100%
Données homogènes (2)	25,6%	-1,8%	36,3%	39,9%	100%
Différence de contribution : (1) - (2)	7,2%	9,6%	-6,6%	-10,2%	0%

En partant de la gauche, le type de données est en première colonne. La contribution par classe d'actifs à la rentabilité est renseignée de la deuxième à la cinquième colonne. Le total des pondérations est en sixième colonne.

La présence de données hétérogènes semble modifier les contributions, notamment pour des allocations équivalentes. Pour une allocation identique et « quasi nulle » en présence de données hétérogènes et homogènes, la contribution de la classe « matières premières » à la rentabilité devient positive en présence de données hétérogènes, alors qu'avec uniquement des données de cotation elle est négative. La différence de contribution est de 9,6% (tableau 29). Un phénomène similaire se produit pour la classe « obligations ». Pour une allocation équivalente (tableau 28), la contribution de la classe « obligations » est surpondérée en présence de données hétérogènes (tableau 29). En présence de données hétérogènes, les contributions des classes « actions » et « devises » sont inférieures à celle obtenues en présence de données homogènes (tableau 29). La sous-pondération de l'allocation sur la classe « actions » (tableau 28) en présence de données hétérogènes semble engendrer une contribution également inférieure à celle obtenue en présence de données homogènes (tableau 29). La surpondération de l'allocation sur la classe « devises » (tableau 28), n'accroît pas sa contribution (tableau 29) à la rentabilité du portefeuille d'arbitrage, mais au contraire la réduit fortement.

Analytiquement, le calcul des contributions par actif à la rentabilité (tableau 30) permettent de comprendre l'évolution des contributions par classe d'actifs (tableau 29). En présence de données hétérogènes, la supériorité de la contribution de la classe « obligations » provient de taux variations des obligations brésiliennes et américaines. Pour la classe des « matières premières », ce sont principalement les accroissements des cotations du pétrole (brent) et du blé qui expliquent la supériorité de la contribution de cette classe en présence de données hétérogènes. Les contributions inférieures en présence de données hétérogènes pour les classes « actions » et « devises » proviennent respectivement de baisses des taux de

variation d'un indice américain (Nasdaq 100) et du taux de change dollar américain/euro.

A l'inverse, notons que le taux de variation de la devise dollar américain/euro contribue le plus fortement à la contribution de la classe « devises » en présence de données homogènes. Pour la classe « devises », on peut remarquer que le taux de variation de la roupie indienne/euro est relativement élevé en présence de données hétérogènes. Les résultats montrent que l'apport informationnel des données hétérogènes à travers son impact sur les contributions à la rentabilité de portefeuilles d'arbitrage multi-classes d'actifs dans un contexte international. L'influence de ces variables supplémentaires d'information semble également quantifiable pour de « nouvelles » zones de croissance économique (Brésil, Inde).

Tableau 30 - Contribution par actif à la rentabilité d'un portefeuille d'arbitrage optimisé selon le type de données

	Contribution par actif à la rentabilité en présence de données hétérogènes (1)	Contribution par actif à la rentabilité en présence de données homogènes (2)	(1) - (2)
France 2004 4% 25/04/14	4,4%	7,7%	-3,3%
US 2004 4% 15/02/14	9,6%	1,2%	8,4%
Japan 2004 1,5% 20/03/14	-1,0%	7,1%	-8,1%
Brazil 2004 10,5% 14/07/14	19,7%	9,5%	10,1%
Brent (pétrole)	5,9%	7,1%	-1,2%
West Texas Intermediate (pétrole)	1,7%	-2,0%	3,8%
Or	1,5%	-7,2%	8,7%
Blé	3,0%	1,3%	1,7%
Maïs	-4,3%	-0,8%	-3,4%
Bovespa (Brésil)	10,1%	3,5%	6,5%
Bse sensex (Inde)	10,6%	-1,4%	12,0%
Cac 40 (France)	13,7%	27,1%	-13,4%
Dax (Allemagne)	0,5%	3,3%	-2,8%
Ftse 100 (Royaume Uni)	-12,1%	-11,7%	-0,3%
S&P 500 (Etats-Unis)	15,7%	14,2%	1,5%
Nasdaq 100 (Etats-Unis)	-12,8%	-0,1%	-12,7%
Asx 200 (Australie)	3,7%	-1,3%	4,9%
Shanghai SE (Chine)	3,3%	-0,3%	3,6%
Nikkei 225 (Japon)	-3,0%	3,0%	-6,0%
Dollar américain / euro	-28,5%	39,2%	-67,7%
Sterling /euro	-0,4%	0,6%	-1,0%
Franc suisse / euro	9,4%	2,2%	7,2%
Yen / euro	-2,2%	-9,9%	7,6%
Yuan /euro	2,6%	6,9%	-4,3%
Couronne norvégienne / euro	-6,7%	-1,3%	-5,4%
Real brésilien / euro	-2,7%	4,4%	-7,1%
Roupie indienne / euro	56,3%	-0,8%	57,1%
Won coréen / euro	-3,4%	1,2%	-4,5%
Dollar australien / euro	13,5%	-3,7%	17,3%
Rand sud africain / euro	2,0%	-2,3%	4,3%
Dollar néo-zélandais / euro	-10,2%	2,4%	-12,7%
Dollar canadien / euro	0,0%	1,0%	-1,0%

En partant de la gauche, les actifs sont en première colonne. La contribution par actif à la rentabilité en présence de données hétérogènes est en deuxième colonne. La contribution par actif à la rentabilité en présence de données homogènes est en troisième colonne. La différence entre la deuxième et la troisième colonne est en quatrième colonne.

L'analyse de corrélations entre contributions permet de rendre compte des différentes réactions entre classes d'actifs en présence de données hétérogènes. L'apport discriminant des données hétérogènes réside dans le fait que les corrélations significatives ne sont pas toujours les mêmes selon le type de données utilisées (tableau 31). D'autre part, dans le cas où la significativité est vérifiée en présence de données hétérogènes et homogènes (« obligations » et « actions », « devises » et « actions »), le coefficient de corrélation est plus fortement négatif en présence de données hétérogènes. Ce dernier élément est déterminant dans la diversification du risque des portefeuilles d'arbitrage optimisés. La présence de données hétérogènes semble accroître la diversification du risque du fait de corrélations plus fortement négatives entre contributions.

Tableau 31 - Significativité statistique des corrélations entre les contributions par classe d'actifs à la rentabilité d'un portefeuille d'arbitrage optimisé selon le type de données pour la période de tests de l'échantillon (2006-2011)

Données homogènes	Type de données	Coefficient de corrélation de Pearson	<i>t stat</i>
Obligations et matières premières	Données hétérogènes	-2,18%	0,76
	Données homogènes	-7,29%	2,54 ***
Obligations et actions	Données hétérogènes	-23,03%	8,23 ***
	Données homogènes	-8,34%	2,91 ***
Obligations et devises	Données hétérogènes	2,21%	0,77
	Données homogènes	-9,32%	3,26 ***
Matières premières et actions	Données hétérogènes	-8,55%	2,98 ***
	Données homogènes	1,48%	0,52
Matières premières et devises	Données hétérogènes	3,84%	1,34
	Données homogènes	-11,85%	4,15 ***
Devises et actions	Données hétérogènes	-22,23%	7,93 ***
	Données homogènes	-9,43%	3,29 ***

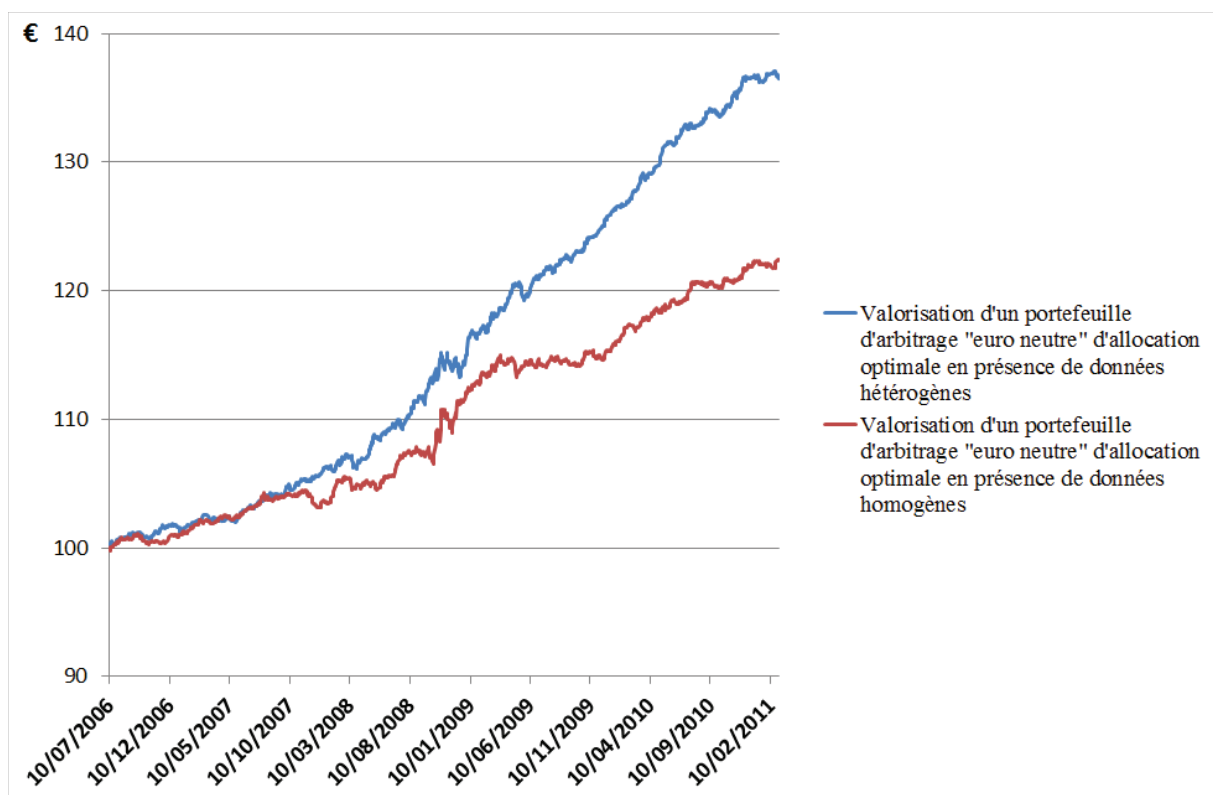
** : $p < 0,05$; *** : $p < 0,01$

En partant de la gauche, les données homogènes sur lesquelles porte le calcul de la corrélation sont en première colonne. Le type de données utilisées pour composer nos portefeuilles d'arbitrage est en deuxième colonne. Le coefficient de corrélation de Pearson entre contributions de classes d'actifs à la rentabilité annualisée est en troisième colonne. La *t-stat* de Student est en quatrième colonne.

Dans une logique connexe, nous remarquons que la rentabilité de notre portefeuille d'arbitrage évolue en sens inverse de celle d'un portefeuille des données de cotation obtenu par analyse factorielle (ACP) dans le chapitre 3 (figure 13). Ce portefeuille de données de cotation issu d'une ACP n'est pas soumis à notre méthode d'arbitrage. Ce portefeuille ne constitue pas en soi un *benchmark* mais peut être considéré comme l'évolution statistique

historique de l'ensemble des données de cotation sur le passé. A ce titre, ce portefeuille représente l'évolution des actifs sans prise en compte des innovations proposées par notre approche en présence de données hétérogènes et par des procédures de quantification. En présence de données hétérogènes la rentabilité quotidienne de notre portefeuille d'arbitrage « euro neutre » d'allocation optimale (figure 17) ne semble pas subir la « chute » de valorisation que rencontre le portefeuille de données de cotation lors de la crise des *subprimes* (figure 13).

Figure 17 - Valorisation quotidienne d'un portefeuille d'arbitrage « euro neutre » issue d'une diversification optimale selon le type de données (2006 – 2011)



Les dates sont en abscisse et la valorisation quotidienne en euros est en ordonnée pour 100 euros investis en date de départ. La courbe bleue correspond à la valorisation d'un portefeuille « euro neutre » en allocation optimale (diversification optimale) composé à l'aide d'une analyse de données hétérogènes. La courbe rouge correspond à la valorisation d'un portefeuille « euro neutre » en allocation optimale (diversification optimale) composé à l'aide d'une analyse de données homogènes.

L'évolution de la valorisation quotidienne de notre portefeuille en présence de données hétérogènes ne semble pas avoir été fortement impacté par la crise des *subprimes* puisque sa rentabilité n'a pas connu de décroissance « prononcée » pour cette période (tableau 32, figure 17).

Différents travaux mesurent l'impact d'informations autres que les données boursières sur les rendements de portefeuilles. Dans le cadre d'une allocation tactique (« actions » et « obligations ») intégrant des informations « conditionnelles », autres que les rendements d'actifs historiques, Dahlquist et Campbell (2001) cherchent à mesurer l'existence d'un impact du cycle de l'économie américaine sur le reste du monde. Ils constatent que la plupart des rendements d'actions en dehors des Etats-Unis sont plus sensibles au cycle de la conjoncture des Etats-Unis que les rendements des actions aux Etats-Unis.

Kandel et Stambaugh (1996) étudient l'allocation d'actifs lorsqu'elle dépend de variables « prédictives » dans le cadre d'un modèle bayésien (*diffuse prior*)²⁴¹. Dans leurs travaux, les auteurs soulignent que les rendements du *New York Stock Exchange* (NYSE) peuvent être prévisibles en faisant dépendre l'allocation d'actifs de la valeur historique la plus récente du taux de rendement en dividende. Cette dernière valeur est considérée comme une variable « prédictive » puisqu'elle entraîne des changements d'allocations et de rendements. Cependant, les résultats obtenus par ces auteurs ne sont pas toujours validés statistiquement.

Dans le cadre d'une diversification optimale des portefeuilles d'arbitrage, nous prolongeons les conclusions selon lesquelles des informations passées peuvent avoir un pouvoir « prédictif » sur les rendements d'actifs. Nous convergions vers l'idée que les informations passées, autres que des données de cotation, peuvent participer à l'estimation des rendements futurs. Nous complétons la découverte selon laquelle l'impact de ces informations historiques sur les actifs peut se former sur plusieurs zones dans le monde. Les influences exportées sont ainsi mesurées sur d'autres zones que celles qui publient ces informations économiques. Nous mettons en évidence un impact multiple dans un contexte multi-classes d'actifs avec prise en compte d'informations variées, qui ne se résument pas à des données comptables ou économiques. Les interactions mises en évidence sont mesurées dans un contexte d'arbitrage que la plupart des travaux ignorent.

Nous participons à la littérature existante sur l'optimisation de portefeuille en mesurant les changements d'allocations et de contributions à la rentabilité provoqués par la présence de données hétérogènes. Nous mesurons ce lien dans un contexte (peu répandu dans la littérature) d'allocation tactique à l'aide de méthodes d'arbitrage systématique intégrant des données hétérogènes comme paramètres quantitatifs.

Dans un cadre optimal, nos travaux soulignent deux types d'éléments favorables en matière de diversification du risque en présence de données hétérogènes. Le premier élément

²⁴¹ Voir chapitre 1.

est que la prise en compte de données hétérogènes peut générer une contribution supérieure d'une classe d'actifs pour une allocation équivalente. Sans augmenter son allocation, il est possible d'augmenter sa contribution à la rentabilité globale des portefeuilles d'arbitrage. Le deuxième élément bénéfique est que l'introduction de données hétérogènes peut accroître la négativité de la corrélation entre les contributions de deux classes d'actifs. L'évolution contraire de ces deux classes d'actifs permet qu'une baisse de rentabilité sur une classe n'expose pas l'autre classe à la même évolution.

Dans la section suivante, nous nous concentrons sur l'évolution du ratio rentabilité/volatilité des portefeuilles d'arbitrage en présence de données hétérogènes et homogènes.

2.3. Accroissement du ratio rentabilité/volatilité

L'impact des données hétérogènes semble également se manifester sur la rentabilité d'un portefeuille d'arbitrage optimisé et, dans une moindre mesure, sur sa volatilité. La rentabilité augmente en présence de données hétérogènes (tableau 32). La volatilité (écart-type) est « légèrement » inférieure en présence de données hétérogènes, sauf pour la période précédant la crise des *subprimes* (tableau 32).

Tableau 32 - Rentabilités et volatilités annualisées d'un portefeuille d'arbitrage optimisé selon le type de données et les périodes de tests

Données	Période	Rentabilité (1)	Volatilité (2)	(1) / (2)
Données hétérogènes	Juillet 2006 - mars 2011	6,49%	2,00%	3,2
Données homogènes		4,22%	2,10%	2,0
Données hétérogènes	Avant la crise des <i>subprimes</i> (juillet 2006 - janvier 2007)	2,97%	1,43%	2,1
Données homogènes		2,63%	1,42%	1,8
Données hétérogènes	Pendant la crise des <i>subprimes</i> (février 2007 - mars 2009)	6,88%	2,41%	2,9
Données homogènes		5,41%	2,55%	2,1
Données hétérogènes	Après la crise des <i>subprimes</i> (avril 2009 - mars 2011)	7,08%	1,60%	4,4
Données homogènes		3,34%	1,65%	2,0

En partant de la gauche, le type de données est en première colonne. La période de tests de l'échantillon est en deuxième colonne. La rentabilité annualisée est en troisième colonne. La volatilité annualisée est en quatrième colonne. Le ratio rentabilité annualisée/volatilité annualisée est en cinquième colonne.

Dans le cadre de la diversification optimale, le ratio rentabilité/volatilité d'un portefeuille varie en présence de données hétérogènes et ce, quelles que soient les périodes de tests de l'échantillon²⁴². Pour la période de tests de l'échantillon, le ratio rentabilité/volatilité est supérieur en présence de données hétérogènes : soit 3,24 contre 2,01 en présence de données homogènes (tableau 32). Avant, pendant et après la crise des *subprimes*, ce ratio est supérieur en présence de données hétérogènes : respectivement 2,1 contre 1,8, 2,9 contre 2,1 et 4,4 contre 2 (tableau 32). Quelles que soient les périodes de tests de l'échantillon, l'écart (en valeur) entre les ratios²⁴³ s'explique par une rentabilité supérieure pour une volatilité plus ou moins²⁴⁴ équivalente en présence de données hétérogènes (tableau 32).

La supériorité de la rentabilité en présence de données hétérogènes est marquée pour l'ensemble de la période de tests et la sous-période d'après crise des *subprimes* (tableau 33) : respectivement une surperformance de 53,8%²⁴⁵ et de 111,7% en présence de données hétérogènes par rapport à un univers de données homogènes²⁴⁶. Pour ces deux mêmes périodes, l'écart de ratio entre le portefeuille alloué en présence de données hétérogènes et celui diversifié à l'aide de données homogènes est respectivement de 61% et de 118%. Ce sont les deux écarts les plus élevés des différentes périodes de tests.

La procédure de minimisation permet globalement de contenir la volatilité des rentabilités quel que soit le type de données considérées. Il semble que les changements d'allocations (tableau 28) aient participé à cet accroissement (tableaux 32).

La comparaison de ces résultats avec ceux obtenus en diversification « naïve » (tableaux 19 et 32) montre que la prise en compte de données hétérogènes dans le cadre d'une diversification optimale permet d'augmenter sensiblement le ratio rentabilité/volatilité, sauf sur la période qui précède la crise des *subprimes*. Par contre, sur la base de données homogènes la diversification optimale aboutit à un ratio plus faible, sauf pour la période d'après la crise.

Ces résultats sont à rapprocher de ceux de DeMiguel et al. (2009) qui constatent la supériorité de performance de l'allocation de portefeuille « 1/N » sur différentes formes d'optimisations. Les auteurs constatent un plus fort ratio de Sharpe pour la stratégie « 1/N » en période *out of sample*. Ils affirment que « plus d'énergie » doit être consacrée pour

²⁴² De juillet 2006 à mars 2011.

²⁴³ Rentabilité/volatilité en présence de données hétérogènes et ce même ratio en présence de données homogènes.

²⁴⁴ Plus de précisions sont apportées sur la volatilité dans cette sous-section.

²⁴⁵ Une performance annualisée de 6,49% en présence de données hétérogènes et de 4,22% en présence de données homogènes (tableau 33).

²⁴⁶ Une performance annualisée de 7,08% en présence de données hétérogènes et de 3,34% en présence de données homogènes (tableau 33).

améliorer l'estimation des premiers et deuxièmes moments en n'utilisant pas uniquement les statistiques mais aussi « d'autres informations disponibles ». En dehors de toute considération de type de données utilisé, les auteurs aboutissent au constat que la volatilité est inférieure pour une stratégie de portefeuille « naïve » par rapport à une composition optimale²⁴⁷.

Nous obtenons des résultats empiriques inversés en présence de données hétérogènes comme en présence de données homogènes. Nos travaux, à travers ce constat et le précédent concernant les ratios rentabilité/volatilité, tendent à montrer que la conclusion quant à la supériorité des méthodes d'optimisation par rapport à une allocation naïve dépend du type de données utilisé.

Dans l'esprit d'une utilisation accrue « d'autres informations disponibles », Black et Litterman (1992) définissent un cadre théorique optimal pour introduire des « vues » d'investisseurs. Cette approche, ne se limitent pas à exprimer une estimation d'un rendement futur en excès d'un taux court domestique²⁴⁸. L'aspect « innovant » de cette approche, « non-traditionnellement reconnue », est que les auteurs proposent une estimation des rendements futurs reposant sur deux sources d'informations : les « vues » ou opinions des investisseurs et l'équilibre du marché. Dans la cadre de nos recherches, les correspondances des « vues » et de l'équilibre de marché sont respectivement les « autres informations de marché » et la détermination de facteurs systématiques (chapitre 2).

Dans une version modifiée du « CAPM », Xu et Malkiel (2004) mènent une étude empirique sur les marchés « actions » aux Etats-Unis et au Japon. Ils suggèrent que les rendements de leurs portefeuilles « co-varient » avec les rendements de portefeuilles « couvrant » le risque idiosyncratique. Cette constatation empirique précise l'importance de considérer le rendement d'un actif par le biais de son risque spécifique, ce que nous effectuons lors de nos travaux, en présence ou non de données hétérogènes. Nous calculons l'évolution du risque résiduel à partir de notre analyse de données pour estimer le rendement d'un actif d'un jour sur l'autre. En ce sens, nous nous rapprochons les travaux de ces derniers²⁴⁹ et renforçons leurs conclusions dans le cadre de notre allocation systématique multi-classes d'actifs dans un cadre géographique élargi, qui dépasse les Etats-Unis et le Japon. Le principe de centrer l'attention de nos travaux sur le risque spécifique, comme élément d'anticipation des rentabilités futures, est dans la continuité de leurs avancées

²⁴⁷ Nous revenons sur ce point dans le prochain chapitre ainsi qu'en Conclusion Générale.

²⁴⁸ Voir la note de bas de page n°5 de Black et Litterman (1992) pour le calcul du rendement en excès d'une classe d'actifs.

²⁴⁹ Nous proposons une modélisation du rendement d'un titre à un instant donné, proche de celle de Xu et Malkiel (2004). Voir la section 4 de leur étude.

empiriques. La présence des données hétérogènes permet de poursuivre les découvertes de leurs travaux pour des portefeuilles d'arbitrage « euro neutre » issus d'allocations optimales. Dans une allocation tactique multi-classes « actions » et « obligations », Blitz et Vliet (2008) proposent de solliciter des programmes d'optimisation pour obtenir des corrélations optimales entre classes d'actifs et surtout accroître la rentabilité ajustée au risque du portefeuille. Le caractère prédictif de l'allocation optimale est une question en suspens pour les auteurs qui ne font pas référence aux données requises pour ce type d'approche optimale. Nous proposons un premier élément de développement avec la constatation d'un accroissement du ratio rentabilité/volatilité des portefeuilles d'arbitrage lorsque la variance est minimisée en présence de données hétérogènes (tableau 32).

L'importance que nous attribuons au premier moment (rentabilités) est déterminante dans l'accroissement du rapport rentabilité/volatilité. Nous nous concentrons sur la détermination des rendements futurs (d'un jour à l'autre) qui en présence de données hétérogènes, d'après nos tests empiriques, semble être *ex ante* plus précise.

Par analogie avec les travaux cités, les « autres informations de marché » intègrent des « vues » d'investisseurs et l'équilibre statistique du marché correspond au « *market equilibrium* » décrit par les auteurs. La prise en compte de ces facteurs informationnels peut favoriser une meilleure anticipation des rendements futurs en présence de données hétérogènes. La présence de données hétérogènes permet, entre autres, d'accroître le ratio rentabilité/volatilité²⁵⁰ pour une volatilité « contenue » (tableau 32).

Nos travaux portent également sur la détermination d'un risque spécifique en présence de données hétérogènes et homogènes. Cette estimation du risque spécifique est censée devenir un élément prédictif des rendements futurs des portefeuilles d'arbitrage. Il semble que l'évaluation du risque spécifique en présence de données hétérogènes permet une meilleure approximation des rendements futurs qu'en présence de données homogènes dans le cas d'une diversification optimale²⁵¹. La généralisation de ce phénomène au contexte d'arbitrage multi-classes d'actifs²⁵² contribue à souligner l'importance de la mesure du risque spécifique comme *proxy* de la rentabilité future y compris pour un contexte plus large que celui défini par les différents auteurs. L'optimisation en présence de données hétérogènes semble correspondre à deux objectifs distincts fixés par Blitz et Vliet (2008) : intégrer des variables

²⁵⁰ La minimisation de la volatilité sous contraintes ayant été spécifiée au préalable.

²⁵¹ Ce n'est pas le cas pour une allocation « naïve » sauf pour la période *post subprimes* où le ratio rentabilité / volatilité est plus élevé en présence de données hétérogènes.

²⁵² Notons que Black et Litterman (1992) n'incluent pas la classe des matières premières dans le cadre de leurs travaux.

informationnelles prédictives et réaliser des travaux d'optimisation afin d'augmenter la rentabilité ajusté du risque d'un portefeuille issue d'une allocation tactique multi-classes d'actifs. Alors que les auteurs suggèrent ces deux voies de recherche séparément, nous les intégrons de manière conjointe afin d'obtenir un ratio rentabilité/volatilité plus élevé en présence de données hétérogènes surtout après la crise des *subprimes*.

2.4. Réduction des pertes financières probables

Le calcul de corrélations entre contributions (tableau 31) est une première étape pour mesurer la diversification du risque des portefeuilles d'arbitrage. Nous examinons à présent comment peuvent évoluer les pertes financières probables de portefeuille d'arbitrage « euro neutre » diversifiés de manière optimale en présence ou non de données hétérogènes. Nous utilisons la mesure de la *Value-at-Risk* (Linsmeier et Pearson, 1996) et nous appliquons un test de Kupiec (1995) à ces estimations de pertes financières probables pour vérifier la significativité statistique de nos résultats. Nous suivons la même méthodologie que celle de la section 1.5 « Réduction des pertes financières probables durant la crise des *subprimes* » du chapitre 4.

Tableau 33 - Tests de Kupiec (1995) d'un portefeuille d'arbitrage optimisé selon le type de données et la période de tests

Type de données	Période de tests	Statistique "LR" (5%)	Valeur critique du Chi 2 (1 ddl, 5%)	Significativité
Données hétérogènes	Pendant la crise des <i>subprimes</i> (février 2007 - mars 2009)	0,05 **	3,84 **	Significatif
Données homogènes		0,05 **		Significatif
Données hétérogènes	Après la crise des <i>subprimes</i> (avril 2009 - mars 2011)	32,46	3,84 **	Non-significatif
Données homogènes		12,27		Non-significatif

** : $p < 0,05$.

Le type de données est en première colonne. La période de tests en dehors de l'échantillon (juillet 2006 – janvier 2007) est en deuxième colonne. La statistique de ratio de vraisemblance de Kupiec (1995) est en troisième colonne. Le quantile, valeur critique du Chi2 à 1 degré de liberté et à 5% de risque de première espèce est en quatrième colonne. Le résultat du test (caractère statistiquement significatif ou non) est en cinquième colonne.

L'estimation des pertes probables est statistiquement significative en présence de données hétérogènes et homogènes pendant la période de crise des *subprimes* (tableau 33).

Ceci peut être cohérent avec le fait que la *VaR* paramétrique²⁵³ que nous estimons est largement influencée par la volatilité historique. Suite à notre processus d'optimisation, les volatilités ont été minimisées et deviennent plus proches les unes des autres en présence de données hétérogènes comme en présence de données homogènes. D'autre part, nous obtenons un taux « d'échec » ou *failure* (fréquence empirique des violations V_t)²⁵⁴ égal en présence de données hétérogènes et homogènes en période de la crise des *subprimes*.

Après la crise des *subprimes*, aucune significativité statistique n'est recensée ni en présence de données hétérogènes ni en présence de données homogènes (tableau 33). Le risque de pertes financières en période *post* crise n'a pas été atténué dans le cadre d'une diversification optimale des portefeuilles d'arbitrage.

Depuis la crise des *subprimes* de nombreuses réserves sont émises sur l'approche paramétrique de la *VaR*, notamment à l'aide de la loi gaussienne. En pratique les rendements ne sont pas gaussiens (El Ghaoui, Oks, et Oustry, 2003). Les auteurs rappellent que la formalisation « classique » de l'optimisation de portefeuille peut rendre les portefeuilles extrêmement sensibles aux erreurs d'estimations des moyennes et des matrices de covariances.

La significativité statistique des pertes financières probables (en *backtestings*) permet de penser que les méthodes de quantification²⁵⁵ utilisées pour estimer les rendements futurs (d'un jour sur l'autre) constituent une voie d'investigation pour limiter la sensibilité de ces procédures d'optimisation aux estimations des rendements. En d'autres termes, la quantification permet d'estimer « finement » les rendements à venir : cette précision permet de répondre à certaines limites des portefeuilles optimaux, notamment l'extrême sensibilité des allocations d'actifs aux erreurs d'estimation des rendements futurs.

3. Eléments de robustesse

La robustesse de notre approche est une étape déterminante de nos recherches dans la mesure où cette phase de recherche nous permet d'exposer différents éléments de stabilité de nos méthodes d'arbitrage issues d'une allocation optimale. Les résultats empiriques ont une portée plus large si ces derniers sont issus de méthodes dites « robustes ». L'analyse de cette robustesse s'effectue à trois niveaux : en vérifiant que les « autres informations de marché »

²⁵³ Voir le chapitre 4.

²⁵⁴ Voir le chapitre 4.

²⁵⁵ Voir les chapitres 1 et 2.

ont un effet « réel » sur les contributions des classes d'actifs à la rentabilité et non un effet de *data mining*, en simulant un nombre important de *backtestings* (intégrant une composante aléatoire) pour calculer l'erreur moyenne de notre approche et en étendant la période d'« apprentissage » des *backtestings* pour observer le maintien possible des caractéristiques des portefeuilles d'arbitrage initialement construit hors-étape de robustesse.

3.1. Simulations en présence de variables aléatoires

Nous souhaitons mesurer si l'introduction d'« autres informations de marché » dans notre processus d'analyse de données apporte une réelle valeur ajoutée. Nous souhaitons éviter d'observer une valeur ajoutée « artificielle » sur nos données, produite par un possible effet de « *data mining* » (Avellaneda et Lee, 2010 ; Blitz et Vliet, 2008). Les tests empiriques faisant intervenir un nombre important de données peuvent produire des effets « *data mining* ».

Afin d'éviter tout biais provoqué par le traitement et l'utilisation des données hétérogènes, nous décidons de remplacer les « autres informations de marché » par des variables aléatoires²⁵⁶. En introduisant ce phénomène aléatoire, nous cherchons à vérifier que l'influence des données relatives aux « autres informations de marché » sur le comportement des portefeuilles d'arbitrage issus d'une allocation optimale n'est pas fallacieuse. Le but de ce test de robustesse est de vérifier si les résultats précédemment obtenus sur les contributions par classe d'actifs peuvent être confirmés ou non.

La méthodologie suivie pour mener ces simulations est la même que celle suivie dans le chapitre 4. Ces tests de robustesse permettent d'effectuer les mêmes travaux que dans le chapitre 4 sauf que des portefeuilles d'arbitrage sont issus d'un processus d'optimisation.

²⁵⁶ Les variables aléatoires ont été générées à l'aide du simulateur variable aléatoire RAND() ou Rnd sous VBA/Excel.

Tableau 34 – Test de robustesse en présence de variables aléatoires : significativité statistique des corrélations entre les contributions par classe d’actifs à la rentabilité d’un portefeuille d’arbitrage optimisé selon le type de données pour la période de tests de l’échantillon (2006-2011)

Données homogènes	Type de données	Coefficient de corrélation de Pearson	<i>t stat</i>
Obligation et matières premières	Données hétérogènes	-2,18%	0,76
	Données homogènes et variables aléatoires	-0,54%	0,19
Obligations et actions	Données hétérogènes	-23,03%	8,23 ***
	Données homogènes et variables aléatoires	-3,97%	1,38
Obligations et devises	Données hétérogènes	2,21%	0,77
	Données homogènes et variables aléatoires	-68,45%	32,66 ***
Matières premières et actions	Données hétérogènes	-8,55%	2,98 ***
	Données homogènes et variables aléatoires	-4,44%	1,55
Matières premières et devises	Données hétérogènes	3,84%	1,34
	Données homogènes et variables aléatoires	0,50%	0,17
Devises et actions	Données hétérogènes	-22,23%	7,93 ***
	Données homogènes et variables aléatoires	-18,58%	6,58 ***

** : $p < 0,05$; *** : $p < 0,01$

En partant de la gauche, les données homogènes sur lesquelles porte le calcul de la corrélation sont en première colonne. Le type de données utilisées pour composer nos portefeuilles d’arbitrage est en deuxième colonne. Le coefficient de corrélation de Pearson entre contributions de classes d’actifs à la rentabilité annualisée est en troisième colonne. La *t-stat* de Student est en quatrième colonne.

A la suite des tests empiriques en présence de variables aléatoires, nous constatons des différences de significativité statistique au niveau des contributions à la rentabilité par classes d'actifs par rapport à notre échantillon incluant d'« autres informations de marché », à l'exception de la corrélation entre « devises » et « actions » (tableau 34).

Les corrélations significatives ne sont pas les mêmes en présence de variables aléatoires et en présence d'« autres informations de marché » (tableau 34). En présence de variables aléatoires, deux corrélations entre contributions par classe d'actifs sont statistiquement significatives : celle entre les classes « obligations » et « devises », puis celle entre les classes « devises » et les « actions » (tableau 34) alors qu'en présence d'« autres informations de marché » (données hétérogènes) nous obtenons trois corrélations significatives : entre « obligations » et « actions », « matières premières » et « actions », puis « devises » et « actions » (tableau 34).

En présence d'« autres informations de marché », les corrélations entre contributions sont « fortement » négatives et significatives (tableau 34). Le remplacement de ces « autres informations de marché » par des variables aléatoires permet de constater des significativités statistiques différentes lors des calculs de corrélations entre classes d'actifs (tableau 34). Ces différences de significativités sont obtenues pour la corrélation entre « obligations » et « actions » et celle entre « obligations » et « devises » (tableau 34).

Si les significativités sont obtenues pour les mêmes classes, nous envisageons un effet *data mining* produit par la présence des « autres informations de marché ». A l'issue de ce test de robustesse, l'effet *data mining* est donc écarté. La prise en compte de variables aléatoires ne permet pas de reproduire les caractéristiques de notre portefeuille d'arbitrage en présence d'« autres informations de marché ». L'influence des « autres informations de marché » sur notre portefeuille d'arbitrage ne semble pas être, au niveau de ce test de robustesse, un artefact : l'apport des « autres informations de marché » est ainsi confirmé.

3.2. Simulations de Monte Carlo

Ces simulations permettent de répéter plusieurs fois des tests sur le passé (*backtestings*). L'objectif visé est le même que dans le chapitre 4 mais dans le cadre d'un portefeuille d'arbitrage en diversification optimale. La méthodologie suivie pour mener les simulations est la même que celle du chapitre 4. Les tests de robustesse réalisés permettent de vérifier si l'empreinte des données hétérogènes est validée lorsque de nombreuses trajectoires sont simulées en vue de constituer des portefeuilles d'arbitrage à partir d'un processus

d'optimisation.

Tableau 35 - Erreur moyenne des simulations de Monte Carlo d'un portefeuille optimisé selon le type de données pour la période de tests de l'échantillon (2006 – 2011)

Type de données	Moyenne des taux de variations	Erreur moyenne
Données hétérogènes	7,92%	9,49%
Données homogènes	5,36%	10,94%

Le type de données est en première colonne. La moyenne des taux de variation du portefeuille est en deuxième colonne. L'erreur moyenne est en troisième colonne. L'erreur moyenne se calcule comme indiquée ci-dessous : $\frac{\sigma}{\sqrt{S}}$, avec σ , l'écart-type des taux de variation et S , le nombre de trajectoires. Nous simulons 100 trajectoires ($S = 100$).

Les résultats des simulations de Monte Carlo peuvent être mis en perspective de l'accroissement de la rentabilité des portefeuilles d'arbitrage dans les sections précédentes. Pour un nombre répété de simulations, nous obtenons des éléments de volatilité et de rentabilité cohérents avec les résultats des tests précédents. En moyenne, la rentabilité est supérieure en présence de données hétérogènes (tableaux 32 et 35). Notons qu'à l'issue des simulations, en moyenne la volatilité baisse plus fortement que lors des tests précédents (tableaux 32 et 35).

La répétition des trajectoires n'a pas modifié en moyenne les caractéristiques des portefeuilles d'arbitrage issus d'une allocation optimale. L'augmentation de la rentabilité semble être également accompagnée d'une baisse de la volatilité pour nos portefeuilles d'arbitrage. Ces caractéristiques statistiques semblent refléter une influence durable et mesurable des données hétérogènes pour la période de tests de notre échantillon après la simulation de plusieurs scénarios.

Indépendamment du type de données, la stabilité de notre processus se matérialise par deux erreurs²⁵⁷ moyennes « relativement » faibles au regard de la composante aléatoire introduite par les procédures de quantification²⁵⁸. Les algorithmes de quantification peuvent engendrer une « forte » volatilité, ce qui n'est pas le cas ici (tableau 35). D'autre part, le

²⁵⁷ Une en présence de données hétérogènes, une autre en présence de données homogènes.

²⁵⁸ Voir le chapitre 1.

nombre relativement « restreint » de données dans l'échantillon²⁵⁹ utilisé pour nos procédures de quantification peuvent entraîner une volatilité élevée : ce qui n'est pas le cas lors des simulations de robustesse. Pour obtenir des grilles de probabilités dites « stables » pour chacun des actifs, il est préférable d'avoir un nombre plus élevé de données en phase d'« apprentissage » que celui que nous avons utilisé. Malgré le nombre de données relativement limité pour nos procédures de quantification, nous obtenons des grilles de probabilités qui semblent engendrer des signaux robustes pour les portefeuilles d'arbitrage.

3.3. Extension de la période d'« apprentissage » de l'échantillon

Une troisième voie d'investigation de robustesse porte sur la longueur de la période d'« apprentissage » de l'échantillon²⁶⁰ lors des tests empiriques. Nous rallongeons cette période d'« apprentissage » de notre échantillon. La date de départ des tests empiriques se trouve mécaniquement décalée, retardée. La procédure suivie pour le rallongement de la période d'« apprentissage » de notre échantillon est la même que celle du chapitre 4. Le seul changement est que nous sommes dans le cadre d'une allocation optimale.

Le décalage dans le temps de la date finale de la période « apprentissage » (et de la date de départ) des tests empiriques permet de conserver les principales caractéristiques d'allocations d'actifs et de rentabilité des portefeuilles d'arbitrage optimisés (tableaux 36 et 37).

Tableau 36 – Test de robustesse à l'aide d'une période d'« apprentissage » étendue : comparaison de l'allocation d'actifs d'un portefeuille d'arbitrage optimisé selon le type de données pour une période de tests décalée (février 2009 - mars 2011)

Type de données	Classe d'actifs				Total
	Obligations	Matières premières	Actions	Devises	
Données hétérogènes	20%	3%	35%	42%	100%
Données homogènes	28%	2%	19%	52%	100%

En partant de la gauche, le type de données est en première colonne. L'allocation d'actifs par classe d'actifs est renseignée de la deuxième à la cinquième colonne. Le total des pondérations est en sixième colonne.

²⁵⁹ Entre août 2004 et fin juin 2006.

²⁶⁰ Voir les chapitres 1 et 2, faisant référence à la quantification.

Le changement d'allocation obtenu en présence de données hétérogènes se produit également lorsque nous élargissons la fenêtre de l'échantillon « apprentissage » (tableau 36). Une constante entre les tests initiaux (tableau 28) et les tests avec un rallongement de la période d'« apprentissage » (et donc du point de départ des tests), est que les classes « devises » et « matières premières » gardent des pondérations diamétralement opposées (tableau 36). L'allocation sur la classe « devises » est la plus élevée alors que celle sur la classe « matières premières » est la plus faible (tableau 36). Soulignons que l'allocation sur la classe « matières premières » demeure quasiment inchangée (tableau 28 et 36).

Dans chacune des configurations (avec et sans rallongement de la période d'« apprentissage »), en présence de données hétérogènes ou non, les classes « devises » et « actions » totalisent plus de 60% de l'allocation d'actifs des portefeuilles d'arbitrage (tableaux 28 et 36).

Avec un rallongement de la période d'« apprentissage », nous conservons globalement les mêmes caractéristiques de rentabilité de nos portefeuilles d'arbitrage optimisés. Nous obtenons une rentabilité plus élevée en présence de données hétérogènes (tableaux 32 et 37). Nous obtenons également un ratio rentabilité/volatilité supérieur en présence de données hétérogènes (tableau 32 et 37).

Tableau 37 – Test de robustesse à l'aide d'une période d'« apprentissage » étendue : rentabilité et volatilité annualisées d'un portefeuille d'arbitrage optimisé selon le type de données pour une période de tests décalée (février 2009 - mars 2011)

Données	Période	Rentabilité (1)	Volatilité (2)	(1) / (2)
Données hétérogènes	Février 2009 -	5,77%	2,04%	2,83
Données homogènes	mars 2011	3,76%	1,48%	2,54

En partant de la gauche, le type de données est en première colonne. La période de tests de l'échantillon est en deuxième colonne. La rentabilité annualisée est en troisième colonne. La volatilité annualisée est en quatrième colonne. Le ratio rentabilité annualisée/volatilité annualisée est en cinquième colonne.

Pour les tests empiriques hors-robustesse sur les portefeuilles d'arbitrage optimisés comme pour les tests de robustesse, nous établissons plusieurs points communs entre les portefeuilles. Ces points convergents permettent de mesurer des segments de stabilité des résultats empiriques dans notre cadre optimal. Nous proposons de comparer les résultats empiriques obtenus dans le cadre d'une diversification « naïve » avec ceux obtenus lors de ce

chapitre.

4. Analyse comparative des résultats empiriques : diversification « naïve » versus diversification optimale

Les différents résultats empiriques des chapitres 4 et 5 semblent montrer qu'il existe une « trace quantitative » laissée par les données hétérogènes sur les portefeuilles d'arbitrage, que ces derniers soient issus d'une allocation « naïve » ou optimale. Nous proposons d'agrèger ces différentes découvertes de manière dynamique. Nous synthétisons les résultats qui tendent à aller dans la même direction en présence de données hétérogènes pour une diversification « naïve » comme une diversification optimale des portefeuilles d'arbitrage. Nous effectuons cette même synthèse pour nos résultats qui divergent.

4.1. Convergences en présence de données hétérogènes

Nous synthétisons ci-après l'ensemble des résultats empiriques des chapitres 4 et 5 qui évoluent dans une même direction en présence de données hétérogènes. La convergence des résultats empiriques s'effectue lorsque, pour des stratégies d'allocations distinctes (diversification « naïve », diversification optimale), nous obtenons des résultats similaires en présence de données hétérogènes. Ce travail permet de mettre en exergue les points communs de nos découvertes malgré les changements de cadres d'allocations (« naïve » et optimale).

4.1.1. Changement d'allocation d'actifs

La présence de données hétérogènes modifie l'allocation d'actifs des portefeuilles d'arbitrage issus d'une allocation « naïve » et optimale (tableaux 16 et 28). L'introduction des données hétérogènes dans la phase d'analyse factorielle des données entraîne une modification de l'allocation d'actifs par rapport à un univers de données homogènes. Soulignons que l'allocation sur la classe des « devises » reste dans les deux cadres d'allocations (« naïve » et optimale) la plus élevée (tableaux 16 et 28) avec plus de 40% des pondérations.

Dans notre cadre « naïf », la classe des « matières premières » est surpondérée alors que celle des « obligations » est sous-pondérée (tableau 16). Dans notre cadre optimal, à l'inverse de leurs volatilités historiques, les classes « obligations » et « devises » sont surpondérées alors que la classe « matières premières » est sous-pondérée. Malgré la stratégie

d'optimisation, nous constatons que l'allocation d'actifs varie également en présence de données hétérogènes (tableau 28).

4.1.2. Variations des contributions par classe d'actifs à la rentabilité

Malgré la stratégie d'optimisation, nous constatons que les contributions par classe d'actifs à la rentabilité globale des portefeuilles d'arbitrage varient en présence de données hétérogènes tout comme nous l'identifions pour une diversification « naïve ».

Dans le cadre de la diversification « naïve » des portefeuilles d'arbitrage, certaines contributions changent en présence de données hétérogènes (tableau 17) pour une allocation pourtant proche (tableau 16). Pour une allocation similaire sur la classe « actions » (en présence de données hétérogènes et homogènes), la contribution augmente en présence de données hétérogènes et représente plus de 50% de la rentabilité des portefeuilles d'arbitrage. Nous observons le phénomène inverse pour une allocation équivalente sur la classe « devises » : une augmentation de la contribution, mais en présence de données homogènes. La surpondération (« matières premières ») comme la sous-pondération (« obligations ») ont entraîné respectivement une baisse et une hausse des contributions (tableau 17).

Dans le cadre de la diversification optimale des portefeuilles d'arbitrage certaines contributions changent en présence de données hétérogènes (tableau 29) pour une allocation pourtant identique (tableau 28). Pour une allocation équivalente sur la classe « obligations » et « matières premières » (en présence de données hétérogènes et homogènes), les contributions augmentent respectivement en présence de données hétérogènes à un niveau de 32,7% et 7,9% (tableau 29). La surpondération des « devises » comme la sous-pondération des « actions » entraînent une baisse des contributions (tableau 29).

4.1.3. Accroissement du ratio rentabilité/volatilité après la crise des *subprimes*

L'accroissement du ratio rentabilité/volatilité des taux de variation des données de cotation en présence de données hétérogènes est un résultat constant de nos tests empiriques. Dans le cadre d'une diversification « naïve », ce ratio est égal à 3 contre 1,8 en présence de données homogènes (tableau 19). En univers optimal, ce ratio atteint la valeur de 4,4 en présence de données hétérogènes contre 2 en présence de données homogènes.

Il s'agit dans les deux cas (« naïf » et optimal) de l'écart le plus élevé entre les ratios pour l'ensemble de la période de tests (tableaux 19 et 32). Après la crise des *subprimes*, la différence de ratio est la plus élevée. En allocation « naïve », l'écart de ratio est de +76% en

faveur d'un portefeuille composé en présence de données hétérogènes. En allocation optimale, l'écart de ratio est de +118% en faveur d'un portefeuille composé en présence de données hétérogènes.

Cette supériorité du ratio en présence de données hétérogènes est vérifiée pour toute la période de tests ainsi que les autres sous-périodes (avant et pendant la crise des *subprimes*) lorsqu'il est question de diversification optimale. La sensibilité supposée aux données hétérogènes semble se matérialiser le plus distinctement après la crise des *subprimes* au travers du prisme rentabilité/volatilité.

4.1.4. Baisse du niveau des corrélations entre contributions par classes d'actifs à la rentabilité

Les corrélations entre les contributions à la rentabilité ont tendance à diminuer lorsque nous sommes en présence de données hétérogènes pour des portefeuilles d'arbitrage issus d'une allocation « naïve » et optimale (tableaux 20 et 31). Ce phénomène a pour effet de faire baisser la volatilité des portefeuilles d'arbitrage issus d'une diversification « naïve ». Dans les deux cadres de diversification (« naïve » et optimale), l'introduction de données hétérogènes, dans notre analyse de données, contribue à l'abaissement du niveau de corrélation des contributions entre les « obligations » et les « actions », et les « devises » et les « actions » (tableaux 20 et 31).

Cette baisse du niveau de corrélation en présence de données hétérogènes contribue à la diversification du risque et donc à une meilleure gestion du risque. Lorsque la corrélation baisse pour se rapprocher de zéro ou devenir de plus en plus négative, l'évolution des rentabilités d'une classe d'actifs n'impacte pas obligatoirement l'évolution des rentabilités d'une autre classe. Dans un cas où les corrélations sont négatives, la hausse des variations d'une classe entraînera probablement la baisse des variations d'une autre au regard de l'analyse statistique historique. Dans cette optique, une politique de couverture des risques peut être « naturellement » engagée à un moindre coût, en faisant reposer l'analyse uniquement sur les corrélations. L'accroissement du ratio rentabilité/volatilité participe également à une première approche de la gestion du risque des portefeuilles d'arbitrage.

4.1.5. Contributions des nouveaux pays industrialisés à la rentabilité

L'analyse de la contribution par actif à la rentabilité des portefeuilles d'arbitrage permet de détailler la contribution agrégée par classe d'actifs. Dans des contextes de diversification « naïve » et optimale, la contribution des actifs des nouveaux pays industrialisés ou « développés » est notable en présence de données hétérogènes, et ce quelle que soit l'allocation choisie. L'évolution de leurs contributions permet d'expliquer de nombreux mouvements de contribution par classe d'actifs en présence de données hétérogènes. Nous nous intéressons à la contribution par actif en présence de données hétérogènes.

Dans les deux cas de diversification (« naïve » et optimale) de portefeuilles, l'évolution de la contribution de la classe « obligations » est dictée par les variations des obligations brésiliennes et américaines (tableaux 18 et 30). Pour la classe des « matières premières », ce sont communément les accroissements sur le blé qui explique une partie de la contribution au niveau de cette classe. L'évolution du blé est généralement impactée par les informations économiques comme la demande et l'inflation mondiale, souvent reliées aux nouveaux pays « développés ». Notons que les contributions de la classe « devises » sont, dans les deux cas de diversification influencée par des devises comme la roupie indienne (tableaux 18 et 30) et le won coréen (tableaux 18 et 30). Lors d'une allocation « naïve », l'évolution de la contribution de la classe « actions » est en partie dictée par les variations d'indices indiens et chinois. Ceci n'est pas le cas pour une allocation optimale au sein des « actions ».

4.1.6. Tests de robustesses convergents

Nos différents tests de robustesse valident dans l'ensemble les résultats obtenus en présence de données hétérogènes pour les deux types de diversifications de portefeuilles (« naïve » et optimale). Le remplacement des « autres informations de marché » par des variables aléatoires permet de montrer que les corrélations changent et que la significativité statistique n'est pas la même lorsque les données hétérogènes sont remplacées (tableaux 24 et 34). Le risque d'un effet « artificiel » des données « autres informations de marché » sur les portefeuilles d'arbitrage semble écarté.

Les simulations de Monte Carlo permettent de mesurer une erreur moyenne inférieure en présence de données hétérogènes, octroyant une « prime » à la stabilité des portefeuilles

construits en présence de données hétérogènes. Dans le cadre d'une diversification « naïve », ces simulations permettent de rendre cohérent la baisse de la volatilité (tableau 19) et la hausse de la rentabilité (tableau 32) dans le cadre d'une diversification optimale en présence de données hétérogènes.

L'extension de la période d'« apprentissage » et de la date de départ des tests ne modifie pas les principales caractéristiques de l'allocation, de la rentabilité et de la volatilité en présence de données hétérogènes. Bien que des différences existent, la majorité de l'allocation est portée par les classes « devises » et « actions » (tableau 26 et 36) dans des contextes de diversification « naïve » et optimale. La baisse de la volatilité (diversification « naïve ») et la hausse de la rentabilité (diversification optimale) sont également maintenues malgré le décalage temporel du calendrier de *backtesting*.

4.2. Divergences en présence de données hétérogènes

Nous agrégeons ci-après l'ensemble des résultats empiriques des chapitres 4 et 5 qui n'évoluent pas dans le même sens en présence de données hétérogènes. Ce travail permet de mettre en exergue les divergences de nos découvertes lorsque nous comparons les allocations « naïve » et optimale en présence de données hétérogènes.

4.2.1. Différences d'analyse du niveau de ratio rentabilité/volatilité

Dans le cadre d'une diversification « naïve » des portefeuilles d'arbitrage, la supériorité du ratio rentabilité/volatilité en présence de données hétérogènes pour l'après crise des *subprimes* provient de la baisse de volatilité des portefeuilles d'arbitrage (tableau 19). Pour cette même période, dans le cadre d'une diversification optimale la supériorité de ce ratio provient d'une hausse de la rentabilité (tableau 32). Notons qu'en présence de données hétérogènes et pour une allocation « naïve », ce ratio est toujours inférieur (avant et pendant la crise des *subprimes*) ou égal (toute la période de tests) au ratio en présence de données homogènes (tableau 19). Dans le cadre d'une diversification optimale, ces ratios sont supérieurs en présence de données hétérogènes pour toutes les périodes de tests (tableau 32).

Nous analysons les différentes évolutions au cours du temps du ratio des portefeuilles d'arbitrage selon les contextes et le type de données utilisé.

En présence de données hétérogènes et pour une allocation « naïve », les rentabilités sont principalement inférieures à celles obtenues en présence de données homogènes sauf pour la

période *post subprimes* (tableau 19). Cette hiérarchie est respectée pour la période précédant les *subprimes* ainsi que pour la période des *subprimes* mais la période *post subprimes* permet d'inverser cette hiérarchie : la rentabilité annualisée en présence de données hétérogènes est supérieure à celle obtenue en présence de données homogènes (respectivement 12,03% contre 8,02%) en période *post subprimes*. Pour ce même cadre de gestion, les volatilités sont principalement inférieures à celles de l'univers homogènes sauf pour la période précédant les *subprimes* où la volatilité en présence de données hétérogènes est supérieure (tableau 19).

En présence de données hétérogènes et pour une allocation optimale, le schéma est inversé. Les rentabilités sont toujours plus élevées en présence de données hétérogènes (tableau 32). Par rapport au cadre « naïf », les volatilités sont généralement plus faibles et plus proches les unes des autres en présence de données hétérogènes et homogènes (tableaux 32 et 19). Le processus d'optimisation a permis de « contenir » les volatilités des portefeuilles d'arbitrage, offrant moins de pouvoir d'influence aux données hétérogènes sur ces dernières.

4.2.2. Modification de la hiérarchie d'allocations par classe d'actifs

Nous rappelons que dans le cadre de portefeuilles d'arbitrage issus d'une diversification « naïve », l'allocation des pondérations sur chaque classe d'actifs reflète globalement le poids respectif de chacune des classes dans la sélection des données de cotation effectuée dans le chapitre 3 : l'allocation reflète la sélection.

La présence de données hétérogènes permet de modifier la hiérarchie des poids alloués sur chaque classe d'actifs pour des portefeuilles d'arbitrage « 1/N ». En présence de données hétérogènes et en ordre décroissant (de la pondération la plus élevée à celle la moins élevée), nous obtenons : les classes « devises », « actions », « matières premières » (surpondérée) et « obligations » (sous-pondérée) (tableau 16). Cette hiérarchie est changée dans le cadre de portefeuilles d'arbitrage issus d'une diversification optimale et en présence de données hétérogènes : « devises » (surpondérées), « obligations », « actions » (sous-pondérées) et « matières premières » (tableau 28). L'allocation est (généralement) inversement définie par les volatilités historiques de chaque actif dans le cadre d'une diversification optimale d'un portefeuille d'arbitrage.

L'allocation sur la classe d'actifs « obligations » est surpondérée par rapport à une diversification « naïve » et occupe désormais la deuxième place en termes de répartition des poids alors qu'elle occupait la dernière place pour des portefeuilles issus d'une diversification naïve » (tableaux 16 et 28). Bien que la classe « devises » conserve la plus grande allocation,

la classe « actions » occupe désormais la troisième position pour des portefeuilles optimisés et notons que la classe « matières premières » est pourvue de la plus faible allocation (tableau 28).

4.2.3. Différences de significativité des pertes financières probables par période

Les pertes probables financières sont estimées au travers de l'estimation *VaR*. La significativité statistique de cette perte probable est calculée *via* la procédure de *backtesting* du test de Kupiec (1995). L'influence des données hétérogènes sur l'estimation des pertes probables n'est vérifiée que pour des portefeuilles d'arbitrage « naïfs » pendant la période des *subprimes* (tableau 22). Pour cette même période, les pertes ne sont pas significatives en présence de données homogènes (tableau 22).

Dans le cadre d'une diversification optimale, les pertes probables estimées sont significatives en période de crise des *subprimes* tant en présence de données hétérogènes qu'homogènes. L'effet discriminant des données hétérogènes sur l'estimation des pertes probables n'est pas observé dans le cadre d'une allocation optimale puisque la significativité est également constatée en présence de données homogènes. Remarquons que, quelle que soit la diversification choisie pour les portefeuilles d'arbitrage, les pertes probables ne sont pas significatives après la crise des *subprimes* (tableaux 22 et 33) et ce quel que soit le type de données (hétérogènes, homogènes).

Les tests empiriques de notre recherche sont menés dans le cadre d'une diversification « naïve » (chapitre 4) et d'une diversification optimale pour les portefeuilles d'arbitrage « euro neutre ». Dans le cadre d'une allocation optimale, nous mesurons les éléments d'influence des données hétérogènes sur lesdits portefeuilles d'arbitrage. La présence de données hétérogènes modifie l'allocation d'actifs et les contributions à la rentabilité par classes d'actifs. L'impact des données hétérogènes est également tangible dans l'accroissement de la rentabilité et du ratio rentabilité/volatilité. Les corrélations statistiquement significatives entre classes d'actifs changent lorsque nous introduisons d'« autres informations de marché ». Les différents tests de robustesse permettent globalement de vérifier la stabilité des résultats, renforçant la portée des résultats empiriques en univers optimal.

CONCLUSION GÉNÉRALE

1. Principaux résultats de la recherche

Nos travaux de recherche ont pour objectif d'évaluer l'intérêt de la prise en compte de données hétérogènes en Finance. La nature de l'information et son degré d'influence sur les prix des actifs financiers constituent des zones que la recherche en Finance explore depuis plusieurs décennies. La manière d'identifier ces informations reste une question sur laquelle il ne semble pas exister de consensus entre les différents travaux. Nous tentons de comprendre ce lien entre évolution des marchés financiers et publication d'informations. Notre problématique générale porte sur la valeur ajoutée d'une combinaison de données hétérogènes dans un processus de gestion de portefeuille par arbitrage multi-classes d'actifs financiers.

Nos recherches permettent d'élaborer deux méthodes différentes d'arbitrage systématique qui se complètent. La première méthode est paramétrique alors que la deuxième est non-paramétrique. La comparaison de ces approches met en évidence la souplesse et le caractère innovant de notre méthode non-paramétrique. La souplesse réside dans le fait que notre approche est modulable tant sur la longueur des périodes que sur les spécifications de l'allocation d'actifs et du *market timing*. L'innovation repose sur deux éléments : l'application des méthodes de quantification au cadre de l'arbitrage statistique et l'intégration de données hétérogènes comme paramètres d'une allocation systématique multi-classes d'actifs.

Nous retenons un modèle empirique non-paramétrique, incluant des probabilités de transition et permettant de prendre en considération une combinaison de données hétérogènes dans un cadre multi-classes d'actifs. Cette modélisation non-paramétrique est notamment un réceptacle pour l'intégration de données hétérogènes. A l'aide de cette approche, nous menons des tests empiriques sur des portefeuilles d'arbitrage « euro neutre » en présence ou non de données hétérogènes et pour deux types de diversification (« naïve » et optimale). Nous menons ces recherches dans deux contextes différents : l'allocation « naïve » et l'allocation optimale. Les résultats empiriques sont ainsi mis en perspective dans des contextes fondamentalement éloignés. En considérant ces deux contextes, nous élargissons la portée des résultats issus des tests, qui ne sont pas forcément attribués à une stratégie de

gestion.

Notre choix se porte sur une voie non-paramétrique du fait de certains avantages comparatifs par rapport au cadre paramétrique : des « epsilons » (risque spécifique, facteurs non-principaux) non-paramétriques, des rentabilités supérieures, un temps de calcul algorithmique plus court et la possibilité d'exploiter des probabilités de transition servant à un insuffler une dynamique d'alternance d'achat/vente propre à une logique d'arbitrage.

Les portefeuilles étudiés sont construits dans une logique d'arbitrage : les actifs sont achetés (vendus) en fonction de leur position par rapport à l'équilibre du « marché ». Cette position relative au « marché » est définie d'un point de vue probabiliste et le niveau du « marché » est estimé statistiquement (chapitres 1 et 2). L'équilibre de « marché » est défini à l'aide de travaux théoriques (Avellaneda et Lee, 2010 ; Engle et Granger, 1987 ; Johansen, 1991 ; Lo et MacKinlay, 1990 ; Ross, 1976), de calculs statistiques (ACP) et d'outils probabilistes de quantification (Pagès et Printems, 2008).

Nos différentes réflexions s'orientent autour de nos trois questions de recherche²⁶¹. Une première question de recherche porte sur le fondement d'une combinaison de données hétérogènes en Sciences de Gestion et plus particulièrement en Finance. En associant différents travaux empiriques de la littérature existante, l'intérêt de cette combinaison d'informations de tout type est analysé à l'aide de recherches menées principalement en Sciences de Gestion (Finance, Marketing et Sociologie) et en Sciences Economiques. Le croisement de ces recherches, provenant d'horizons divers, permet de comprendre que la prise en compte de données hétérogènes en univers multi-classes d'actifs peut favoriser l'anticipation d'arbitrages entre actifs et/ou entre classes d'actifs. Au regard de cette synthèse de recherches, la prise en compte de données hétérogènes semble (entre autres) permettre une évaluation plus fine de phénomènes financiers tels que les opportunités d'arbitrage ou le risque de défaut. Une deuxième question de recherche porte sur l'adaptabilité d'informations hétérogènes à un modèle d'évaluation multi-classes d'actifs. Autrement dit, l'enjeu est de savoir comment ces différentes informations peuvent s'agréger à un processus d'allocation tactique multi-classes d'actifs. A partir de travaux sur l'analyse de données, les modèles multifactoriels et la quantification, nous définissons des méthodes d'arbitrage systématique pouvant prendre en compte des informations de tout type (données de cotation et « autres informations de marché »). Nous élaborons une méthode d'arbitrage systématique à l'aide de génération de signaux de *market timing* servant *in fine* d'outil d'allocation tactique multi-

²⁶¹ Se reporter à l'Introduction Générale, section 2 « Problématique générale et questions de recherche ».

classes d'actifs. Nous constituons un modèle empirique (non-paramétrique) ayant la capacité d'intégrer à la fois des prix d'actifs, provenant d'un univers multi-classes, et des données informationnelles. Une troisième interrogation porte sur la manière de mesurer l'apport de l'utilisation de données hétérogènes pour l'allocation et la performance de portefeuilles d'arbitrage. Notre modélisation (chapitre 2) permet de combiner quantitativement des informations de toutes natures et de comprendre, au moins partiellement, les changements d'allocations d'actifs en présence de données hétérogènes. Nous mesurons globalement cette valeur ajoutée par des changements d'allocations d'actifs, de rentabilités, de volatilités, de corrélations (entre classes d'actifs) et de pertes probables pour des portefeuilles d'arbitrage « euro neutre » multi-classes d'actifs en présence de données hétérogènes. En l'espèce, nous estimons notamment l'influence des données hétérogènes sur des portefeuilles d'arbitrage par l'accroissement de la rentabilité et/ou la baisse de la volatilité (chapitres 4 et 5). Ces manifestations de la création de valeur engendrée par la présence de données hétérogènes, sont accompagnées d'éléments de robustesse de notre approche.

Préalablement à la présentation des résultats, nous rappelons que tous les portefeuilles d'arbitrage sont uniquement composés d'actifs négociables (données de cotation) mais qu'en amont de notre modèle empirique, un processus d'analyse factorielle (ACP) peut, soit faire appel à des données hétérogènes, soit ne retenir que des données homogènes (données de cotation). Il convient également d'insister sur le caractère « dollar neutre » (Avellaneda et Lee, 2010 ; Khandani et Lo, 2007)²⁶² ou « euro neutre » des portefeuilles testés sur le passé à l'aide des données de notre échantillon (chapitre 3). Le caractère « dollar neutre » ou « euro neutre » permet d'assurer que la somme des montants investis en euros, sur les positions acheteuses et vendeuses, est nulle.

Les tests empiriques menés dans les chapitres 4 et 5, montrent que plusieurs résultats convergent indépendamment de l'allocation choisie (« naïve » ou optimale). Ces tests aboutissent majoritairement à des réactions communes en présence de données hétérogènes : une sur ou une souspondération de l'allocation pour certaines classes d'actifs, un changement des contributions à la rentabilité, un changement des corrélations entre classes d'actifs, une augmentation de la valeur du ratio rentabilité/volatilité après la crise des *subprimes* et un abaissement relatif de la volatilité.

²⁶² En fait « euro neutre » dans notre cadre de travail. Le principe est le même, seule la devise change.

1.1. Changement d'allocation d'actifs en présence de données hétérogènes

Quel que soit le type de diversification (« naïve » ou optimale), l'allocation d'actifs est modifiée en présence de données hétérogènes. Le changement d'allocation d'actifs des portefeuilles d'arbitrage « euro neutre » diversifiés de façon « naïve » est notable en présence de données hétérogènes surtout pour les classes « obligations » et « matières premières ». Respectivement, la première classe est sous-pondérée et la deuxième est surpondérée en présence de données hétérogènes (par rapport à la présence de données homogènes). Pour des portefeuilles d'arbitrage « euro neutre » optimisés, l'allocation d'actifs est également modifiée en présence de données hétérogènes par un effet de sous-pondération de la classe « actions » et de surpondération de la classe « devises » (par rapport à un univers de données homogènes). L'ensemble des classes d'actifs est donc exposé à un changement d'allocations dans le cadre des tests empiriques, que la diversification soit « naïve » ou optimale. L'influence des données hétérogènes sur l'allocation ne semble pas confinée à un type de diversification de portefeuille.

Nos résultats empiriques accompagnent les travaux qui suggèrent qu'une information à *priori* permet de modifier une allocation d'actifs. Nous élargissons les cadres de réflexion analysés dans la littérature existante en proposant un contexte d'arbitrage statistique qui inclut un univers de données hétérogènes plus exhaustif que les informations économiques et l'analyse financière fondamentale²⁶³. De plus, nous considérons un champ multi-classes à quatre classes d'actifs (« obligations », « matières premières », « actions », « devises ») ce qui n'est pas le cas de toutes les recherches en Finance. En général, nombreuses sont les découvertes empiriques qui incluent une à deux classes d'actifs, mais rares sont celles qui vont au-delà de ce cadre²⁶⁴.

²⁶³ Par exemple la prise en compte d'un taux de dividendes estimés ou du ratio *Price Earning Ratio* (*PER*).

²⁶⁴ Relativement peu de travaux considèrent trois classes d'actifs (Andersen et al., 2007 ; Black et Litterman, 1992).

1.2. Modification de la contribution par classes d'actifs à la rentabilité en présence de données hétérogènes

Les différences de contributions (par classe d'actifs) à la rentabilité des portefeuilles d'arbitrage « euro neutre » permet de comprendre comment l'introduction de données hétérogènes (dans un processus d'analyse de données²⁶⁵) peut influencer la structure des portefeuilles. Ces changements de contributions en présence de données hétérogènes sont matérialisés pour les deux familles d'allocation (diversification « naïve » et diversification optimale²⁶⁶). Pour la période de test de l'échantillon²⁶⁷, la présence de données hétérogènes permet de faire varier la contribution de chaque classe d'actifs à la rentabilité du portefeuille. Pour une allocation équivalente (en présence de données hétérogènes et homogènes), la contribution de la classe « actions » pour un portefeuille « naïf » augmente en présence de données hétérogènes. En allocation optimale, les contributions des classes « obligations » et « matières premières » sont aussi supérieures en présence de données hétérogènes alors que les allocations sont identiques en présence de données hétérogènes et homogènes.

En présence de données hétérogènes, les décisions d'investissements et les rentabilités par classes d'actifs peuvent être modifiées au niveau des classes d'actifs pour une allocation équivalente²⁶⁸. Il semble que l'effet des « autres informations de marché » soit de nature à influencer la gestion d'un portefeuille d'arbitrage « euro neutre ».

Nous complétons les travaux qui formalisent les différences de réponses par classes d'actifs aux cycles économiques, et ceci dans un univers d'arbitrage systématique. Nous proposons des éléments supplémentaires d'analyse des différences de réaction par classe d'actifs. Nous élargissons la sélection des séries de conjoncture économique en intégrant des indicateurs traduisant les sentiments des investisseurs (sentiments des marchés boursiers, de volatilité).

²⁶⁵ Nous faisons référence à l'ACP effectuée dans la phase d'« apprentissage » de nos données. Se reporter au chapitre 2.

²⁶⁶ Minimisation de la volatilité d'un portefeuille d'arbitrage sous contraintes. Les contraintes sont énumérées et explicitées dans le chapitre 3.

²⁶⁷ De juillet 2006 à mars 2011.

²⁶⁸ Voir le chapitre 3.

1.3. Baisse des corrélations entre les rentabilités par classe d'actifs en présence de données hétérogènes

Les corrélations entre les contributions par classe d'actifs à la rentabilité ont tendance à baisser et/ou à devenir négatives en présence de données hétérogènes, que les portefeuilles d'arbitrage soient issus d'une allocation « naïve » ou optimale. Un élément constant des tests empiriques est également que la corrélation entre les classes « obligations » et « actions » baisse significativement en présence de données hétérogènes, quelle que soit la diversification de portefeuille envisagée (« naïve » ou optimale).

Pour une allocation « naïve », la corrélation devient significativement²⁶⁹ plus fortement négative en présence de données hétérogènes pour les classes suivantes : « obligations » et « actions » (également en allocation optimale pour cette paire de classes), « obligations » et « devises », « devises » et « actions ». En allocation optimale, nous obtenons ce même phénomène également pour la corrélation entre les classes « matières premières » et « actions ». Les différences de significativité des corrélations en présence ou non de données hétérogènes permettent de comprendre les évolutions directionnelles des classes d'actifs.

Différents travaux de recherche en Finance portent sur les corrélations négatives entre classes d'actifs. Nos résultats empiriques prolongent les découvertes de ces recherches dans un cadre de diversification « naïve » et optimale. Le fait qu'une baisse des corrélations soit obtenue pour ces deux types d'allocation en présence de données hétérogènes est un élément supplémentaire qui permet de souligner l'existence potentielle de réactions contraires des classes d'actifs lorsque des informations de tous types sont publiées. D'autre part, la baisse des corrélations est envisagée dans un cadre d'arbitrage multi-classes d'actifs, ce qui n'est pas le cas pour les recherches passées et actuelles.

1.4. Baisse relative de la volatilité des rendements d'un portefeuille d'arbitrage en présence de données hétérogènes

La présence de données hétérogènes semble participer à la baisse de la variabilité des rendements des portefeuilles d'arbitrage « euro neutre », que ces derniers soient issus d'une diversification « naïve » ou optimale. Cette baisse est relative en comparaison du niveau de

²⁶⁹ Au sens statistique.

volatilité obtenu en présence de données homogènes. Ce constat de baisse relative de la volatilité est notable pour la période de tests, la sous-période de crise des *subprimes* et la sous-période *post* crise des *subprimes*. Les tests empiriques font apparaître qu'en présence de données hétérogènes, l'écart-type de la moyenne des rentabilités est inférieur à celui calculé sur la base de données homogènes.

Cette baisse est également validée par une technique de simulation (Monte Carlo) permettant de confirmer ce résultat après une répétition des simulations de portefeuilles d'arbitrage sur le passé. A l'issue de ces simulations, nous constatons l'atténuation relative de la volatilité des rentabilités en présence de données hétérogènes. Pour une allocation « naïve » comme pour une allocation optimale, nous conservons une infériorité et une stabilité de l'erreur moyenne en présence de données hétérogènes²⁷⁰.

Nous participons à la discussion sur le niveau de la volatilité en temps de crise à deux égards. Le premier élément proposé consiste à soutenir l'idée que la présence de données hétérogènes influence le niveau de la volatilité en général. Le deuxième élément apporté par nos résultats est la mise en évidence d'une baisse de la volatilité des portefeuilles d'arbitrage en présence de données hétérogènes et en allocation « naïve ». Cette baisse est notamment remarquée pour la sous-période de crise des *subprimes*. Ce deuxième point vient compléter la vision de plusieurs recherches qui évaluent une hausse de la volatilité en temps de crise. Ce résultat de baisse de la volatilité met en exergue le fait que des arbitrages peuvent être générés sur la base de données hétérogènes, y compris en temps de crise, et que la prise en compte d'autres flux informationnels (que les prix des actifs) semble inciter à plus de prudence. Cette découverte empirique est effectuée dans un cadre multi-classes d'actifs relativement plus large que celui employé dans différents travaux en Finance sur la volatilité.

1.5. Accroissement du ratio rentabilité/volatilité après la crise des *subprimes*

L'influence des données hétérogènes sur le ratio rentabilité/volatilité se mesure principalement lors de la période *post* crise des *subprimes*²⁷¹ pour un portefeuille d'arbitrage « euro neutre » de type « naïf » ou optimal. Pendant la période *post* crise des *subprimes*, l'accroissement de ce ratio est le plus élevé (par rapport aux autres périodes et sous-périodes de tests) en présence de données hétérogènes. L'introduction de données hétérogènes dans notre processus d'analyse des données (ACP) semble faire augmenter encore plus fortement

²⁷⁰ Entre 9,50% et 9,80%.

²⁷¹ Avril 2009 à mars 2011.

la valeur du « couple » rentabilité/volatilité des portefeuilles optimisés après la crise des *subprimes*. L'importance prise par les « autres informations de marché » semble s'être renforcée à partir de la période de crise des *subprimes* et ces indicateurs paraissent plus discriminants après cette période de crise²⁷².

Nous complétons les différentes recherches qui trouvent une surperformance du ratio de Sharpe pour les stratégies « naïves » par rapport aux stratégies optimales. Dans une version alternative du ratio de Sharpe, nous obtenons un ratio rentabilité/volatilité plus élevé pour une allocation optimale que pour celle « naïve » en présence de données hétérogènes et en période *post* crise des *subprimes*. Notre procédure de minimisation (d'optimisation) de la volatilité semble avoir été accompagnée d'un accroissement des rentabilités en présence de données hétérogènes. La portée de ces résultats se situe dans un contexte de portefeuilles d'arbitrage *long/short*.

Nous participons également aux différents travaux qui suggèrent l'amélioration des méthodes d'estimation des moments des rendements d'actifs dans le cadre de portefeuilles optimisés, le but étant de recueillir des informations pertinentes pour assurer une estimation significative des premiers moments. Dans le cadre des tests empiriques, la présence de données hétérogènes semble capter des phénomènes qui permettent de mieux estimer les rendements futurs, occasionnant la définition d'une allocation optimale en conséquence.

1.6. Diminution des pertes probables en présence de données hétérogènes pendant la crise des *subprimes*

L'estimation des pertes probables est un enjeu réglementaire et financier de la gestion d'un portefeuille. Nous utilisons la méthode de la *VaR* (paramétrique) pour calculer des pertes potentielles journalières en présence de données hétérogènes et homogènes pendant et après la période de la crise des *subprimes*. Nous effectuons des *backtestings* avec le test de Kupiec (1995), qui est un test de couverture non-conditionnelle permettant d'estimer la significativité des pertes probables. Le calcul de pertes probables n'est statistiquement significatif qu'en présence de données hétérogènes pour des portefeuilles d'arbitrage « euro neutre » en allocation « naïve » pendant la crise des *subprimes*. Le risque de pertes financières²⁷³ journalières potentiellement subies par ce portefeuille d'arbitrage semble être atténué par la présence de données hétérogènes. Pendant la crise des *subprimes*, en *backtesting* les pertes

²⁷² Celle des *subprimes*.

²⁷³ En euros.

probables estimées pour un portefeuille composé en présence de données hétérogènes sont inférieures à celles estimées en présence de données homogènes. Pendant la crise des *subprimes* et pour un portefeuille en allocation optimale, l'apport des données hétérogènes n'est pas constaté puisque la significativité est acceptée également en présence de données homogènes.

Une constante des résultats est que l'importance des données hétérogènes ne semble mesurée que pendant la crise des *subprimes*, car les pertes probables ne sont pas significatives après cette crise, quelle que soit l'allocation (« naïve » ou optimale).

Plusieurs travaux soulignent l'importance du choix des lois de probabilités dans le calcul de la *VaR*, en expliquant que la loi normale n'est pas une photographie « fidèle » des rendements d'un portefeuille. Malgré l'utilisation d'une loi normale dans le cadre de nos recherches, nous estimons des pertes probables significatives pendant une période de crise « majeure ». Ces résultats laissent penser que, hormis les progrès effectués sur la découverte des distributions appropriées, la prise en compte d'informations de divers horizons peut également contribuer à enrichir l'utilisation de modèle de gestion du risque.

2. Apports de la recherche

Nos travaux permettent de contribuer au développement de la recherche contemporaine en Finance sur trois aspects. Nous présentons les apports empiriques, méthodologiques et managériaux de notre recherche. L'ensemble de ces apports souligne l'importance des données hétérogènes en Finance.

2.1. Apports empiriques

Nous présentons les principaux apports empiriques issus des résultats de notre recherche. Le premier apport porte sur la terminologie associée aux données sollicitées dans le cadre de nos travaux. L'information est au cœur des processus de décision. Il semble primordial de savoir identifier et qualifier cette information. Notre travail sur la terminologie permet de rassembler l'ensemble des informations, hors données de cotation, pouvant affecter des prix des actifs sous une même appellation : les « autres informations de marché ». La combinaison de différents types de données n'est pas limitée au clivage quantitatif/qualitatif ou quantitatif/fondamental. Afin de répondre au manque de consensus sur la nature des données en Finance et de dépasser les clivages précédemment mentionnés, les concepts

d'hétérogénéité et d'homogénéité sont suggérés dans notre discussion.

Par l'utilisation de termes « autres informations de marché » et données hétérogènes, nous tentons de contribuer à la définition d'un type générique de données désignant des enquêtes économiques comme des enquêtes de sentiments des investisseurs. Avec ce travail de recherche sémantique (chapitre 1, Introduction Générale), nous espérons participer au travail de catégorisation des flux informationnels en Finance. Ces propositions de définitions peuvent servir à de futures recherches pour tenter de qualifier les données qui impactent les variables financières.

Le deuxième apport empirique concerne le traitement des données recueillies. Les processus de traitement de données de notre recherche peuvent aider tout chercheur dans une démarche de construction de son échantillon lorsqu'il est en présence d'informations aux caractéristiques différentes. La compréhension des données ainsi que leur traitement est une étape préliminaire et essentielle à toute recherche empirique.

Le troisième apport porte sur la progression proposée dans la construction d'un modèle d'évaluation tactique multi-classes d'actifs en allocation « naïve » et optimale. Nous parlons de « progressions » car il n'existe pas de modèle « communément accepté » de *pricing* qui s'applique à plusieurs classes d'actifs (Blitz et Vliet, 2008). La prise en compte de données hétérogènes semble permettre de créer un lien entre les classes d'actifs qui peuvent avoir des réactions propres à leurs classes. Explicitement, les données hétérogènes collectées sont supposées interagir avec les différentes classes d'actifs de notre échantillon. Ces liens semblent agir comme des facteurs communs permettant d'estimer les variations futures des classes d'actifs.

Le quatrième apport de notre recherche concerne la diversification du risque des portefeuilles d'arbitrage qui semblent se manifester en présence de données hétérogènes. La baisse des corrélations et l'accroissement de la négativité des corrélations entre rentabilités par classes d'actifs permettent que l'évolution d'une classe d'actifs ne dépende pas totalement de celle d'une autre. Dans le cas de corrélations plus fortement négatives, l'évolution positive d'une classe d'actifs engendre l'évolution négative d'une autre classe. Ces évolutions sur les corrélations en présence de données hétérogènes semblent être un élément explicatif de la baisse de la volatilité des portefeuilles d'arbitrage « euro neutre » en allocation « naïve ». La baisse des corrélations peut faire baisser les termes de covariances qui eux-mêmes peuvent faire baisser la volatilité totale d'un portefeuille. Le changement de structure de la matrice des corrélations semble être dû non pas uniquement à l'instabilité des rendements mais aussi à la présence de données hétérogènes.

Le cinquième apport permet de comprendre qu'elle est l'évolution d'un univers de données de cotation, sans prendre en considération notre modèle empirique et les portefeuilles d'arbitrage « euro neutre ». Les premiers constituent exclusivement le résultat d'une analyse factorielle et les seconds ceux de procédures de quantification. En présence de données hétérogènes et homogènes et quel que soit le type d'allocation, les portefeuilles d'arbitrage « euro neutre » ne suivent pas la même direction que le portefeuille propre des données de cotation de notre échantillon (figures 13 et 17, annexe 9). Cette distinction est principalement observée pour la période de la crise des *subprimes*. Pendant cette période, alors que l'« indice » de données de cotation décroît, les portefeuilles d'arbitrage « euro neutre » connaissent une tendance haussière et ceci malgré la crise des marchés financiers (figures 13 et 17, annexe 9).

Le sixième apport empirique porte sur la quantification vectorielle. Les procédures de quantification sont déclinées sur plusieurs classes d'actifs et dans une logique d'arbitrage de produits financiers. Un historique de données relativement « limité » a permis aux algorithmes de quantification de constituer des portefeuilles d'arbitrage performants : avec une hausse de la rentabilité et/ou une baisse de la volatilité. La présence de données hétérogènes semble accentuer le caractère innovant de la quantification malgré le « faible » nombre de points disponibles en période d'« apprentissage ». L'innovation principale de la quantification est de résumer l'historique du risque spécifique des portefeuilles tout en conservant une trace quantitative des « autres informations de marché ».

Le septième apport concerne la valeur ajoutée de la combinaison des méthodes de quantification et de données hétérogènes dans la détermination du risque spécifique, comme *proxy* de la rentabilité future. Dans le cas d'une allocation optimale, l'association de la quantification à des sources d'informations variées permet de définir un risque spécifique qui permet d'anticiper l'évolution des rendements d'actifs. Ce résultat n'est constaté que dans un cadre optimal, la présence de données homogènes n'étant discriminante que sur la rentabilité que pour une diversification « naïve ». Divers travaux, évoqués précédemment sur la classe « actions », se concentrent sur le fait que le risque spécifique permet d'estimer les rentabilités futures. Nous prolongeons les avancées de ces travaux en univers multi-classes pour des portefeuilles d'arbitrage « euro neutre » et optimisés.

2.2. Apports méthodologiques

Les principaux apports méthodologiques de nos travaux sont centrés sur l'élaboration d'un outil de *market timing* et sur la prise en compte de données hétérogènes dans le cadre d'une allocation tactique systématique.

Dans le premier cas, nous expliquons comment générer des signaux d'achats et/ou de ventes de données de cotation. Nous élaborons une méthode d'allocation tactique à l'aide d'un indicateur de *market timing* permettant de générer des portefeuilles d'arbitrage multi-classes d'actifs. La construction de ce modèle empirique de *market timing* permet des ajustements fréquents de pondérations d'actifs d'un portefeuille, comme le rappelle la littérature existante. La mise en place de ce processus de *market timing* est rendu possible par de nombreux échanges théoriques et empiriques. Notre approche permet d'utiliser une combinaison d'outils statistiques et probabilistes qui sont déjà utilisés empiriquement en Finance et dans d'autres Sciences : l'analyse en composantes principales (ACP) et les méthodes de quantification (Kohonen, Lloyd).

Cependant nous exploitons cette combinaison d'outils quantitatifs dans le cadre défini par notre recherche dont le périmètre est celui d'un univers multi-classes d'actifs en présence ou non de données hétérogènes. Un élément important est que nous définissons un équilibre du marché pour identifier une sous ou une surévaluation du rendement d'un titre. Cet équilibre de marché ou *equilibrium* peut varier selon l'introduction ou non de données hétérogènes dans notre processus d'analyse de données (ACP). En l'espèce, nous débutons notre processus de *market timing* par une phase d'analyse de données (ACP), nous définissons un risque spécifique par actif et nous terminons ce processus par la génération de signaux d'arbitrage à l'aide de méthodes de quantification.

La portée de notre approche réside surtout dans le fait qu'elle est paramétrable et libre de toute loi *a priori*. Nous pouvons adapter, entre autres, la fréquence de re-balancement des pondérations (quotidienne, hebdomadaire. . .), le nombre de jours d'ACP, le nombre de jours de quantification (glissante ou non), le niveau de levier, la répartition de la rentabilité sur les positions *long* et *short*²⁷⁴ ou imposer des restrictions sur les poids alloués par classe d'actifs. Nous avons par ailleurs mis en place des procédures permettant de calculer des portefeuilles optimisés : il s'agit de résoudre des programmes de minimisation de la volatilité d'un

²⁷⁴ Effet « euro neutre » ou non.

portefeuille d'arbitrage « euro neutre » sous contraintes²⁷⁵.

Notre approche de l'allocation tactique combine des analyses quantitatives reposant sur des fondements théoriques, ayant connu des applications empiriques en recherche sauf dans le contexte multi-classes d'actifs. Ceci rend le cadre de notre recherche à la fois « classique » (historique) et différent : il s'agit de proposer une extension des méthodes de quantification appliquées à une combinaison d'actifs principalement « primaires » (Riva, 2008) et d'« autres informations de marché » en univers multi-classes d'actifs.

Le deuxième apport méthodologique consiste à expliquer comment nous appréhendons le traitement d'informations dont les contenus, les échelles et les fréquences sont variés (chapitre 3). Au sein des données hétérogènes recueillies, nous sommes en présence de données de cotation (multi-classes d'actifs) et d'« autres informations de marché ». Nous détaillons également une voie méthodologique permettant d'intégrer ces informations hétérogènes au sein d'un univers de données de cotation. Nous notons que les paramètres d'entrée (*input*) de modèles d'allocations tactiques quantitatifs sont généralement des données de cotation, surtout en arbitrage statistique. En ce sens, la manière de rassembler ces données est déterminante pour agréger ces deux types de données entre elles. Nous proposons une voie de regroupement de ces informations, exploitables par la suite pour divers tests empiriques dans des stratégies d'arbitrage systématique.

2.3. Apports managériaux

Les découvertes empiriques de notre recherche peuvent présenter un intérêt pour des décisions managériales. De nos travaux, découlent trois apports managériaux en Finance.

Nous proposons un moyen de prendre en compte systématiquement des données de tous types de contenus dans le cadre institutionnel de la gestion d'actifs. L'aspect systématique permet d'éviter de nombreux biais discrétionnaires de traitements de l'information (données de cotation et « autres informations de marché »).

Nos travaux sur l'allocation d'actifs et le *market timing* permettent de proposer un outil en direction des *managers* des services d'investissements d'institutions financières (banques, assurances, sociétés de gestion d'actifs indépendantes). Les méthodes de gestion de portefeuille et les résultats empiriques mis en évidence, peuvent être des outils d'aide à la décision des *managers* en Finance. Nous définissons concrètement les actifs sur lesquels portent les investissements (achat, vente). Nos recherches permettent de connaître trois

²⁷⁵ Voir le chapitre 5 pour le détail des contraintes.

éléments essentiels d'un investissement : le moment de l'investissement (quand), le choix de l'actif (lequel) et le montant alloué à cet actif (combien).

Nous cherchons à comprendre comment la gestion du risque peut être perfectionnée en prenant en considération des données hétérogènes. Ce questionnement permet de rendre compte du fait que la significativité statistique du calcul de pertes probables peut être influencée par la présence de données hétérogènes. Les décideurs financiers cherchent à évaluer leur risque de pertes, au moins pour des raisons réglementaires (Bâle II). Les résultats empiriques obtenus montrent qu'il est possible de générer une diversification du risque des portefeuilles d'arbitrage en présence de données hétérogènes, à la fois lors de la baisse ou de l'accroissement de la négativité des corrélations entre classes d'actifs. Ces résultats peuvent, pour ces raisons, susciter de l'intérêt auprès des directions de gestion du risque des institutions financières.

3. Limites de la recherche

Toute recherche comporte des limites dans son approche et sa « réalité » est contestable. Nous ne considérons pas que nos résultats et conclusions sont « universellement » vrais, mais que la « réalité » présentée est vérifiée pour l'expérience menée dans le cadre de nos recherches. Les limites que nous identifions portent sur la fréquence des prix d'actifs et le caractère non-optimal des calculs de risque des portefeuilles d'arbitrage.

3.1. Exploitation de fréquences inter-journalières des prix des actifs

Nous considérons des historiques de prix d'actifs quotidiens de clôture « *EOD* »²⁷⁶. Comme mentionné précédemment, une partie de la littérature existante en Finance permet de montrer empiriquement la « pertinence » de fréquences quotidiennes notamment sur les travaux effectuant le lien entre les informations économiques et les rendements d'actifs ou de classe(s) d'actif(s). Nos travaux permettent notamment d'analyser la rentabilité et les corrélations entre de différentes classes d'actifs au sein de portefeuilles d'arbitrage en présence ou non de données hétérogènes²⁷⁷ et ce, à l'aide de prix d'actifs inter-journaliers.

Notre cadre de recherche n'intègre pas la « haute fréquence » comme lorsqu'il est, par

²⁷⁶ (Avellaneda et Lee, 2010).

²⁷⁷ Voir le chapitre 4.

exemple, question de considérer des données de cotations toutes les secondes. La « haute fréquence » implique notamment le fait que les structures de corrélations intra-journalières sont différentes de celles observées en rythme inter-journalier.

3.2. Gestion du risque non-optimisée

La gestion du risque des portefeuilles d'arbitrage n'est envisagée qu'au travers du calcul de *VaR* paramétriques et d'allocations optimales. Nous n'avons considéré qu'une modélisation paramétrique des calculs de pertes probables. Nos travaux d'optimisation portent sur l'allocation d'actifs permettant de générer une rentabilité future. L'optimisation de la gestion du risque n'est pas encore envisagée dans nos travaux.

Certes des phénomènes de diversification du risque, notamment au travers d'une baisse des corrélations entre classes d'actifs, peuvent se manifester *ex post* à nos travaux d'allocation mais il ne s'agit pas d'un but *ex ante* d'encadrement du risque. Compte tenu de l'évolution des réglementations internationales, il paraît nécessaire d'intégrer à nos recherches des procédures optimisées de gestion du risque pour des portefeuilles d'arbitrage notamment en ce qui concerne les calculs de pertes probables. Soulignons que la gestion du risque n'oriente pas actuellement le processus d'allocation dans le cadre de nos travaux.

4. Perspectives de recherche

Les limites propres à nos travaux peuvent susciter des nouvelles voies de recherches. Ces nouveaux champs connexes de recherche peuvent être complémentaires aux découvertes empiriques et/ou ouvrir des voies de réflexion non-envisagées par des questionnements initiaux.

4.1. Utilisation de fréquences intra-journalières des prix des actifs

La transposition de notre cadre de recherche à un univers « haute fréquence » permet d'identifier si le phénomène de baisse des corrélations (ou d'accroissement de la négativité des corrélations) apparaît en présence de données hétérogènes. Les corrélations peuvent être « élevées » lorsque les cotations sont rapprochées. Certains travaux soulignent qu'avec des prix intra-journaliers, les rendements sont fortement corrélés et les variances le sont également (Mandelbrot, 1963 ; Plerou et al., 1999). Certains recherches évoquent également

des biais comme les sauts du *spread bid-ask* ou le regroupement d'ordres (Brenner, Pasquariello, et Subrahmanyam, 2009).

La prise en compte de données hétérogènes dans le cadre de la « haute fréquence » permet de vérifier si les mêmes résultats empiriques sont obtenus dans un autre régime de cotation et de comprendre les éventuelles différences. Il est notamment intéressant de savoir si les fortes corrélations entre rendements d'actifs financiers subsistent malgré la prise en compte d'« autres informations de marché ». L'objectif est d'utiliser des bases de données volumineuses²⁷⁸, stockant les prix des actifs d'une séance de cotation toutes les secondes (par exemple). La prise en compte d'« autres informations de marché » est dans ce contexte proche de l'étude d'évènement puisqu'il semble plus adéquat d'analyser les rendements, les corrélations et les volatilités autour²⁷⁹ du moment de la publication de l'indicateur²⁸⁰.

4.2. Gestion optimale du risque et processus d'allocation des portefeuilles d'arbitrage

L'optimisation de la gestion du risque est enjeu complémentaire à celui de la présente recherche. L'optimisation peut venir du fait que nous souhaitons obtenir une même volatilité sur les quatre classes d'actifs de notre univers de données de cotation. Pour ce faire, il convient d'ajuster les allocations d'actifs des portefeuilles (les pondérations sur chaque classe d'actifs) de manière à obtenir le même niveau de volatilité sur chaque classe d'actifs. Le chemin de l'équipondération en volatilité est une manière de poursuivre les travaux de Maillard et al. (2010).

A partir des travaux de ces auteurs, une perspective de recherche est de construire des portefeuilles « *ERC* » en présence ou non de données hétérogènes. La comparaison avec un portefeuille construit à l'aide de données homogènes permet de constater si malgré l'équipondération en volatilité, les poids du portefeuille changent par rapport à un univers de données hétérogènes. Un changement d'allocation d'actifs des portefeuilles d'arbitrage permettrait de recueillir une trace quantitative de l'influence des données hétérogènes.

Dans une logique similaire des portefeuilles « *ERC* », nous suggérons d'estimer une équipondération en *VaR* pour chaque classe d'actifs des portefeuilles d'arbitrage. L'idée conductrice est d'estimer *ex ante* des pertes probables similaires sur chaque classe d'actifs, de façon à ajuster en conséquence l'allocation d'actifs du portefeuille d'arbitrage. L'objectif est ici aussi d'analyser si les poids du portefeuille obtenus par optimisation évoluent selon la

²⁷⁸ Et coûteuse.

²⁷⁹ Par exemple : quelques heures avant, quelques heures après la publication de l'évènement.

²⁸⁰ De conjoncture, de sentiment et de volatilités.

nature des données employées (données homogènes *versus* données hétérogènes). Dans les cas où les allocations changent en présence de données hétérogènes, nous pourrions y voir un élément supplémentaire en faveur de la prise en compte des données hétérogènes en Finance.

BIBLIOGRAPHIE

- Ackermann C., McEnally R., Ravenscraft D. « The Performance of Hedge Funds: Risk, Return, and Incentives ». *Journal of Finance*. 1999. Vol. 54, n°3, p. 833 - 874.
- Ahmed P., Nanda S. « Performance of Enhanced Index and Quantitative Equity Funds ». *Financial Review*. 2005. Vol. 40, n°4, p. 459 - 479.
- Alexander C., Dimitriu A. *Equity Indexing, Cointegration and Stock Dispersion: A Regime Switching Approach to Market Efficiency*. [s.l.] : ISMA Center Discussion Papers in Finance, 2003.
- Allen F., Carletti E. « An Overview of the Crisis: Causes, Consequences, and Solutions ». *International Review of Finance*. 2010. Vol. 10, n°1, p. 1 - 26.
- Altman E. I., Haldeman R., Narayanan P. « Zeta Analysis : A New Model to Identify Bankruptcy Risk of Corporations ». *Journal of Banking and Finance*. 1977.
- Altman E. I. « Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy ». *Journal of Finance*. 1968. Vol. 23, n°4, p. 589 - 609.
- Andersen T. G., Bollerslev T., Diebold F. X., Vega C. « Real-time price discovery in global stock, bond and foreign exchange markets ». *Journal of International Economics*. 2007. Vol. 73, n°2, p. 251 - 277.
- Andersen T. G., Bollerslev T., Diebold F. X., Vega C. « Micro effects of macro announcements: real-time price discovery in foreign exchange. » *American Economic Review*. 2003. Vol. 93, n°1, p. 38 - 62.
- Avellaneda M., Lee J. H. « Statistical Arbitrage in the U.S. Equities Market ». *Quantitative Finance*. 2010. Vol. 10, n°7, p. 761 - 782.
- Bachelier L. « Théorie de la spéculation ». *Annales scientifiques de l'ENS*. 1900. Vol. 17, n°3, p. 21 - 86.
- Balduzzi P., Elton E. J., Green T. C. « Economic news and bond prices: evidence from the U.S. treasury market ». *Journal of Financial and Quantitative Analysis*. 2001. Vol. 36, n°4, p. 523 - 543.
- Béchu T., Bertrand E., Nebenzhal J. *L'Analyse technique : Pratiques et Méthodes*. 5e éd.[s.l.] : Economica, 2008.
- Benartzi S., Thaler R. « Naive Diversification Strategies in Defined Contribution Saving Plans ». *American Economic Review*. 2001. Vol. 91, n°1, p. 79 - 98.
- Best M. J., Grauer R. R. « On the Sensitivity of Mean-Variance-Efficient Portfolios to Changes in Asset Means : Some Analytical and Computational Results ». *Review of Financial Studies*. 1991. Vol. 4, n°2, p. 315 - 342.

- Biais B., Bisiere C., Decamps J. P. « Short sales constraints, liquidity and price discovery: An empirical analysis on the Paris Bourse ». *European Financial Management*. 1999. Vol. 5, n°3, p. 395 - 409.
- Black F., Litterman R. « Global Portfolio Optimization ». *Financial Analysts Journal*. 1992. Vol. 48, p. 28 - 43.
- Blitz D., Vliet P. V. « Global Tactical Cross-Asset Allocation: Applying Value and Momentum Across Asset Classes ». *The Journal of Portfolio Management*. 2008. Vol. 35, n°1, p. 23 - 38.
- Bodson L., Grandin P., Hübner, G, Lambert M. *Performance de portefeuille*. 2e éd.[s.l.] : Pearson Education, 2010.
- Bollerslev T., Wooldridge J. « Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time-Varying Covariances ». *Econometric Reviews*. 1992. Vol. 11, n°2, p. 143 - 172.
- Bollerslev T. « Generalized Autoregressive Conditional Heteroskedasticity ». *Journal of Econometrics*. 1986. Vol. 31, n°3, p. 307 - 327.
- Borgy V., Idier J., Le Fol G. *Liquidity problems in the FX liquid market : ask for the « bil »*. [s.l.] : Banque de France, Eurosystem, 2010.
- Bourbonnais R., Terraza M. *Analyse des séries temporelles*. 2e éd.[s.l.] : Dunod, 2008.
- Bourbonnais R. *Exercices pédagogiques d'Économétrie*. [s.l.] : Economica, 2008.
- Bourbonnais R. *Économétrie*. [s.l.] : Dunod, 1998.
- Bourdieu P. *Questions de sociologie*. [s.l.] : Les Editions de Minuit, 2002. (Collection Reprise).
- Bourgey C. *La crise de la dette souveraine*. [s.l.] : Banque de France, Eurosystem, 2012. (Documents et débats).
- Boyd J., Jagannathan R., Hu J. « The Stock Market's Reaction to Unemployment News: Why Bad News is Usually Good News for Stocks ». *Journal of Finance*. 2005. Vol. 60, n°2, p. 649 - 672.
- Boyer B., Gibson M., Loretan M. *Pitfalls in tests for changes in correlation*. [s.l.] : Board of Governors of the Federal Reserve System, 1997.
- Bremaud P. *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*. Springer.[s.l.] : [s.n.], 2010.
- Brenner M., Pasquariello P., Subrahmanyam M. « On the Volatility and Comovement of U.S. Financial Markets around Macroeconomic News Announcements ». *Journal of Financial and Quantitative Analysis*. 2009. Vol. 44, n°6, p. 1265 - 1289.
- Brinson G. P., Hood L. R., Beebower G. L. « Determinants of Portfolio Performance ». *Financial Analysts Journal*. 1986. Vol. 42, n°4, p. 39 - 44.

- Brinson G. P., Singer B. D., Beebower G. L. « Determinants of Portfolio Performance II : An Update ». *Financial Analysts Journal*. 1991. Vol. 47, n°3, p. 40 - 48.
- Broquet C., Cobbaut R., Gillet R., Berq A. V. D. *Gestion de portefeuille*. 4e éd.[s.l.] : de boeck, 2004.
- Brown S. « The Effect of Estimation Risk on Capital Market Equilibrium ». *Journal of Financial and Quantitative Analysis*. 1979. Vol. 14, n°2, p. 215 - 220.
- Bruder B., Gaussel N., Richard J.-C., Roncalli T. *Regularization of portfolio allocation*. [s.l.] : Lyxxor Research, 2013.
- Campbell J., Lo A., MacKinlay C. *The econometrics of financial markets*. [s.l.] : Princeton University Press, 1997.
- Capocci D. *An Analysis of Hedge Fund Strategies*. [s.l.] : HEC-ULG Management School, University of Liège (Belgium), 2008.
- Cattell R. B. « The Scree Test for the Number of Factors ». *Multivariate Behavioural Research*. 1966. Vol. 1, n°2, p. 245 - 276.
- Chamberlain G., Rothschild M. « Arbitrage, factor structure, and mean variance analysis on large asset markets ». *Econometrica*. 1983. Vol. 51, n°5, p. 1281 - 1304.
- Chan N. T., Getmansky M., Lo A. W., Haas S. M. « Systemic Risk and Hedge Funds ». In : Carey M, Stulz R (éd.). *The risks of financial institutions and the financial sector*. Chicago : University of Chicago Press, 2005. p. 4535 - 4505.
- Chen N., Roll R., Ross S. « Economic forces and the stock market ». *Journal of Business*. 1986. Vol. 59, n°3, p. 383 - 403.
- Chollet P., Cellier A., Gajewski J. F. « Les annonces de notations extra- financières véhiculent-elles de l'information au marché ». *Finance Contrôle Stratégie*. 2011. Vol. 14, n°3, p. 5 - 38.
- Chollet P., Cellier A. « The Impact of Corporate Social Responsibility Rating Announcements on European Stock Prices ». In : *EFMA Annual Meeting*. Braga Portugal : [s.n.], 2011.
- Choueifaty Y., Coignard Y. « Towards Maximum Diversification ». *Journal of Portfolio Management*. 2008. Vol. 35, n°1, p. 40 - 51.
- Christoffersen P. F. « Evaluating Interval Forecasts ». *International Economic Review*. 1998. Vol. 39, n°4, p. 841 - 862.
- Clarke R., Silva H. D., Thorley S. « Minimum-Variance Portfolios in the U.S. Equity Market ». *Journal of Portfolio Management*. 2006. Vol. 33, n°1, p. 10 - 24.
- Corlay S. *Quelques aspects de la quantification optimale et applications à la finance*. Paris : Université Pierre et Marie Curie, 2011.

- Dahlquist M., Campbell R. H. « Global Tactical Asset Allocation ». *Emerging Markets Quarterly*. 2001. p. 6 - 14.
- Damodaran A. *The Dark Side of Valuation*. Upper Saddle River, NJ : Prentice Hall, 2001.
- Davidson J. E. H., Hendry L. R., Srba F., Yeo J. S. « Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom ». *Economic Journal*. 1978. Vol. 88, p. 661 - 692.
- DeMiguel V., Garlappi L., Uppal R. « Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy ? » *Review of Financial Studies*. 2009. Vol. 22, n°5, p. 1915-1953.
- Do B., Faff R., Hamza K. « A New Approach to Modeling and Estimation for Pairs Trading ». In : *Proceedings of 2006 Financial Management Association European Conference*. Stockholm : [s.n.], 2006.
- Drobny S. *Inside the House of Money: Top Hedge Fund Traders on Profiting in the Global Markets*. [s.l.] : John Wiley & Sons, 2006.
- El Ghaoui L., Oks M., Oustry F. « Worst-Case Value-at-Risk and Robust Portfolio Optimization: A Conic Programming Approach ». *Operations Research*. 2003. Vol. 51, n°4, p. 543 - 556.
- Elton E., Gruber M. *Modern portfolio theory and investment analysis*. 5e éd.[s.l.] : John Wiley, 1995.
- Engle R. F., Granger C. W. J. « Cointegration and error correction: Representation, estimation and testing ». *Econometrica*. 1987. Vol. 55, n°2, p. 251 - 276.
- Engle R. « Dynamic Conditional Correlation - A Simple Class of Multivariate GARCH Models ». *Journal of Business and Economic Statistics*. 2002. Vol. 20, p. 339 - 350.
- Entorf H., Gross A., Steiner C. « Business Cycle Forecasts and their Implications for High Frequency Stock Market Returns ». *Journal of Forecasting*. 2011. Vol. 31, n°1, p. 1 - 14.
- Evrard Y., Pras Y., Roux E. *Market : Fondements et méthodes des recherches en marketing*. 4e éd.[s.l.] : Dunod, 2009.
- Evrard Y., Pras Y., Roux E. *Market : Etudes et recherches en marketing*. [s.l.] : Dunod, 2003.
- Fabozzi J. F., Focardi S. M., Jonas C. *Challenges in Quantitative Equity Management*. [s.l.] : CFA Institute Publications, 2008.
- Fama E. « Efficient Capital Markets : A Review of Theory and Empirical Work ». *Journal of Finance*. 1970. Vol. 25, n°2,.
- Fermon D. *CEO value and corporate governance*. [s.l.] : Societe Generale Investment Banking Research, 2006.
- Ferson W. E., Harvey C. R. « The Risk and Predictability of International Equity Returns ». *The Review of Financial Studies*. 1993. Vol. 6, n°3, p. 527 - 566.

- Fisher K., Statman M. « Consumer confidence and stock returns ». *The Journal of Portfolio Management*. 2003. Vol. 30, n°1, p. 115 - 127.
- Fisher K., Statman M. « Investor Sentiment and Stock Returns ». *The Financial Analysts Journal*. 2000. Vol. 56, n°2, p. 16 - 23.
- Forbes K. J., Rigobon R. « No contagion, only interdependence : Measuring stock market comovements ». *Journal of Finance*. 2002. Vol. 57, n°5, p. 2223 - 2262.
- Fung W., Hsieh D. A. « Empirical Characteristics of Dynamic Trading Strategies: The Case of Hedge Funds ». *Review of Financial Studies*. 1997. Vol. 10, n°2, p. 275 - 302.
- Gersho A., Gray R. M. *Vector Quantization and Signal Compression*. [s.l.] : Kluwer Academic Press, 1992.
- Gorton G., Rouwenhorst G. « Facts And Fantasies About Commodity Futures ». *Financial Analysts Journal*. 2006. Vol. 62, n°2, p. 47 - 68.
- Graf S., Luschgy H. *Foundations of Quantization for Probability Distributions*. [s.l.] : Springer, 2000. (Lect. Notes in Mathematics).
- Granger C., Newbold P. « Spurious regressions in econometrics ». *Journal of Econometrics*. 1974. n°2, p. 111 - 120.
- Granger C. « Some Properties of Time Series Data and Their Use in Econometric Model Specification ». *Journal of Econometrics*. 1981. Vol. 16, n°1, p. 121 - 130.
- Greene W. H. *Econometric Analysis*. 3e éd.[s.l.] : Prentice Hall, 1997.
- Gregory-Allen R. B., Stangl J., Shawky H. A. « Quantitative vs. Fundamental Analysis in Institutional Money Management : Where is the Beef ? ». *Journal of Investing*. 2009. Vol. 18, n°4, p. 42 - 52.
- Hotelling H. « Analysis of a Complex of Statistical Variables Into Principal Components ». *Journal of Educational Psychology*. 1933. Vol. 24, p. 417 - 441 and 498 - 520.
- Huberman G., Jiang W. « Offering vs. Choice in 401(k) Plans : Equity Exposure and Number of Funds ». *Journal of Finance*. 2006. Vol. 61, n°2, p. 763 - 801.
- Hull J. *Options, Futures et Other Derivatives*. 6e éd.[s.l.] : Pearson Education, 2009.
- Ibbotson R. G., Kaplan P. D., « Does Asset Allocation Explain 40, 90 or 100 Percent of Performance ? », *Financial Analysts Journal*. J. 2000, Vol. 56, n°1, p. 26 - 33.
- Johansen S., Juselius K. « Maximum Likelihood Estimation and Inference on Cointegration With Applications to the Demand for Money ». *Oxford Bulletin of Economics and Statistics*. 1990. Vol. 52, n°2, p. 169 - 210.

- Johansen S. « Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models ». *Econometrica*. 1991. Vol. 59, n°6, p. 1551 - 1580.
- Jolliffe I. T. « Discarding Variables in a Principal Component Analysis. II : Real Data ». *Applied Statistics*. 1973. Vol. 22, n°1, p. 21 - 31.
- Jones C. M., Lamont O., Lumsdaine R. L. « Macroeconomic news and bond market volatility ». *Journal of Financial Economics*. 1998. Vol. 47, p. 315 - 337.
- Jorion P. « Bayes-Stein Estimation for Portfolio Analysis ». *Journal of Financial and Quantitative Analysis*. 1986. Vol. 21, n°3, p. 279 - 292.
- Kandel S., Stambaugh R. F. « On the Predictability of Stock Returns: An Asset-Allocation Perspective ». *Journal of Finance*. 1996. Vol. 51, n°2, p. 385 - 424.
- Kaplan P. D., Knowles J. A. « Kappa: A Generalized Downside Risk- Adjusted Performance Measure ». *Journal of Performance Measurement*. 2004. Vol. 8, n°3, p. 42 - 54.
- Keating C., Shadwick W. F. « A Universal Performance Measure ». *Journal of Performance Measurement*. 2002. Vol. 6, n°3, p. 59 - 84.
- Khandani A. E., Lo A. W. « What Happened To The Quants In August 2007? » *Journal of Investment Management*. 2007. Vol. 5, n°4, p. 29 - 78.
- King G., Zeng L. « Logisitic Regression in Rare Events Data ». *Political Analysis*. 2001. Vol. 9, n°2, p. 137 - 163.
- Klein R. W., Bawa V. S. « The Effect of Estimation Risk on Optimal Portfolio Choice ». *Journal of Financial Economics*. 1976. Vol. 3, n°3, p. 215 - 231.
- Kohonen T. *Self-Organizing Maps*. Berlin : Springer, 1995.
- Kohonen T. « Self-organized formation of topologically correct feature maps ». *Biological Cybernetics*. 1982. Vol. 43, n°1, p. 59 - 69.
- Kupiec P. H. « Techniques for Verifying the Accuracy of Risk Measurement Models ». *The Journal of Derivatives*. 1995. Vol. 3, n°2, p. 73 - 84.
- Labov W. *The Social Stratification of English in New York City*. 2e éd. Cambridge : Cambridge U. Press, 2006.
- Laloux L., Cizeau P., Potters M., Bouchaud J. P. « Random matrix theory and financial correlations ». *International Journal of Theoretical and Applied Finance*. 2000. Vol. 3, n°3, p. 391 - 397.
- Lehman B. *Is It Worth the While ? The Relevance of Qualitative Information in Credit Rating*. [s.l.] : University of Konstanz, Centre of Finance and Econometry, 2003.
- Lehmann B. « Fads, Martingales, and Market Efficiency ». *Quarterly Journal of Economics*. 1990. Vol. 105, n°1, p. 1 - 28.

- Linsmeier T., Pearson N. *Risk Measurement: An introduction to value-at-risk*. University of Illinois : [s.n.], 1996.
- Lloyd S. P. « Least squares quantization in PCM. Unpublished memorandum ». *IEEE Transactions on Information Theory*. 1982. Vol. 28, p. 129 - 137.
- Lo A. W., MacKinlay A. C. « When Are Contrarian Profits Due to Stock Market Overreaction ? » *The Review of Financial Studies*. 1990. Vol. 3, n°2, p. 175 - 205.
- Magni C. A., Malagoli S., Mastroleo G. « An alternative approach to firms evaluation: Expert systems and fuzzy logic ». *International Journal of Information Technology*. 2006. Vol. 5, n°1, p. 195 - 225.
- Maillard S., Roncalli T., Teletche J. « The Properties of Equally Weighted Risk Contribution Portfolios ». *Journal of Portfolio Management*. 2010. Vol. 36, n°4, p. 60 - 70.
- Mandelbrot B. « The Variation of Certain Speculative Prices ». *Journal of Business*. 1963. Vol. 36, n°4, p. 394 - 419.
- Markowitz H. « The Optimization of a Quadratic Function Subject to Linear Constraints ». *Naval Research Logistics Quarterly*. 1956. Vol. 3, n°1 - 2, p. 111 - 133.
- Markowitz H. « Portfolio Selection ». *The Journal of Finance*. 1952. Vol. 7, n°1, p. 77 - 91.
- Merton R. C. « On estimating the expected return on the market: An exploratory investigation ». *Journal of Financial Economics*. 1980. Vol. 8, n°4, p. 323 - 361.
- Metropolis N. N., Ulam S. « The Monte Carlo Method ». *Journal of the American Statistical Association*. 1949. Vol. 44, n°247, p. 335 - 341.
- Michaud R. O. « The Markowitz Optimization Enigma: Is “Optimized” Optimal ? » *Financial Analysts Journal*. 1989. Vol. 45, n°1, p. 31 - 42.
- Monarcha G. *L'analyse dynamique des structures de risque des hedge funds*. [s.l.] : Université de la Méditerranée, Faculté des Sciences Économiques et de Gestion, 2008.
- Montier J. *The Myth of Exogenous Risk and the Recent Quant Problems* [En ligne]. [s.l.] : [s.n.], 2007. Disponible sur : < <http://behaviouralinvesting.blogspot.com/2007/09/myth-of-exogenous-risk-and-recent-quant.html> >
- Munenzon M. « 20 Years of VIX: Fear, Greed and Implications for Traditional Asset Classes ». février 2010. Disponible sur : < <http://dx.doi.org/10.2139/ssrn.1583504> >
- Murray M. P. « A Drunk and Her Dog: An Illustration of Cointegration and Error Correction ». *The American Statistician*. 1994. Vol. 48, n°1, p. 37 - 39.
- Nelson C. R., Plosser C. L. « Trend and random walk in macroeconomic times series: Some Evidence and Implications ». *Journal of Monetary Economics*. 1982. Vol. 10, n°2, p. 139 - 162.

- Osterwald-Lenum M. « A Note with Quantiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics ». *Oxford Bulletin of Economics and Statistics*. 1992. Vol. 54, n°3, p. 461 - 472.
- Pagès G., Printems J. « Optimal quantization for Finance: From random vectors to stochastic processes ». In : Ciarlet PG, Bensoussan A, Zhang Q (éd.). *Mathematical Modeling and Numerical Methods in Finance*. [s.l.] : North Holland, 2008. p. 595 - 648.
- Pagès G., Printems J. « Optimal quadratic quantization for numerics: The Gaussian case ». *Monte Carlo Methods and Applications*. 2003. Vol. 9, n°2, p. 135 - 165.
- Pagès G. « A space vector quantization method for numerical integration ». *Journal of Computational and Applied Mathematics*. 1998. Vol. 89, p. 1 - 38.
- Pareto V. *Manuale di economia politica con una introduzione alla scienza sociale*. Milano : [s.n.], 1906.
- Park P., Tanrikulu O., Wang G. *Systematic Global Macro: Performance, Risk, and Correlation Characteristics*. [s.l.] : Graham Capital Management, L.P., 2009.
- Pearson K. « On Lines and Planes of Closest Fit to Systems of Points in Space ». *Philosophical Magazine*. 1901. Vol. 2, n°6, p. 559 - 572.
- Plerou V., Gopikrishnan P., Nunes Amaral L. A., Meyer M., Stanley H. E. « Scaling of the distribution of price fluctuations of individual companies ». *Phys. Rev. E* [En ligne]. décembre 1999. Vol. 60, n°6, p. 6519 - 6529. Disponible sur : < <http://dx.doi.org/10.1103/PhysRevE.60.6519> >
- Pole A. *Statistical arbitrage: Algorithmic trading insights and techniques*. [s.l.] : Wiley Finance, 2007.
- Qian E. E. *Risk Parity: The solution to the unbalanced portfolio*. 470 Atlantic Avenue, 8th floor, Boston, MA 02210 : PanAgora Asset Management, 2010.
- Racicot F. E., Théorêt R. *La Value-at-Risk Modèles de la VaR, simulations en Visual Basic (Excel) et autres mesures récentes du risque de marché*. [s.l.] : École des Sciences de la Gestion, Université du Québec à Montréal, 2006.
- Riva F. *Applications Financières sous Excel en Visual Basic*. 3e éd.[s.l.] : Economica, 2008.
- Roll R. « Industrial Structure and The Comparative Behavior of International Stock Market Indices ». *Journal of Finance*. 1992. Vol. 47, n°1, p. 3 - 41.
- Ross S. « The arbitrage theory of capital asset pricing ». *Journal of Economic Theory*. 1976. Vol. 13, n°3, p. 341 - 360.
- Sharpe W. F. « Capital Asset Prices - A Theory of Market Equilibrium Under Conditions of Risk ». *Journal of Finance*. 1964. Vol. XIX, n°3, p. 425 - 442.
- Sortino F. A., Meer R. V. Der. « Downside risk ». *The Journal of Portfolio Management*. 1991. Vol. 17, n°4, p. 27 - 31.

- Spearman C. « The proof and measurement of association between two things ». *The American Journal of Psychology*. 1904. Vol. 15, n°1, p. 72 - 101.
- Stein C. « Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution ». In : *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley, CA : University of California Press, 1956. p. 197 - 206.
- Stockton K. A., Shtekhman A. *Allocation d'actifs: Avantages, limites et évaluation de l'allocation tactique*. [s.l.] : Vanguard Investment Counseling & Research, 2010.
- Stulz R. M. « Hedge Funds: Past, Present and Future ». *Journal of Economic Perspectives*. 2007. Vol. 21, n°2, p. 175 - 194.
- Thauvron A. *Gestion de patrimoine - Stratégies juridiques, fiscales et financières*. 3e éd.[s.l.] : Dunod, 2011.
- Thorp A. W. « Using Investor Sentiment as a Contrarian Indicator ». *Computerized Investing*. 2004.
- Tokat Y. *Sources of Portfolio Performance-The Enduring Importance of Asset Allocation*. [s.l.] : Vanguard Investment Counseling & Research, 2005.
- Trueman B. « Analyst Forecasts and Herding Behaviour ». *The Review of Financial Studies*. 1994. Vol. 7, n°1, p. 97 - 124.
- Uhlenbeck G. E., Ornstein L. S. « On the theory of Brownian Motion ». *Physical Review*. 1930. Vol. 36, n°5, p. 823 - 841.
- Vasicek O. « An equilibrium characterization of the term structure ». *The Journal of Financial Economics*. 1977. Vol. 5, n°2, p. 177 - 188.
- Viviani J. L. *Gestion de portefeuille*. 2e éd.[s.l.] : Dunod, 2001.
- Von Kleist K., Mallo C., Mesny P., Grouchko S. *Triennial Central Bank Survey - Report on global foreign exchange market activity in 2010*. [s.l.] : Bank for International Settlements, 2010.
- Wermers R., Yao T., Zhao J. *The Investment Value of Mutual Fund Portfolio disclosure*. [s.l.] : University of Maryland, 2010.
- Whaley R. E. « Understanding the VIX ». *The Journal of Portfolio Management*. 2009. Vol. 35, n°3, p. 98 - 105.
- Williams J. « Financial Analysts and the False Consensus Effect ». *Journal of Accounting Research*. 2013. Vol. 51, n°4, p. 855 - 907.
- Xu Y., Malkiel B. « Idiosyncratic Risk and Security Returns ». In : *AFA 2001 New Orleans Meetings* [En ligne]. [s.l.] : [s.n.], 2004. Disponible sur : <
<http://ssrn.com/abstract=255303> >

Zhao J. *Quant Jocks and Tire Kickers : Does the Stock Selection Process Matter.* [s.l.] :
University of Arizona, 2006.

LISTE DES TABLEAUX

Tableau 1 - Périodes de l'échantillon (2004-2011)	86
Tableau 2 - Données de cotation (données homogènes) de l'échantillon.....	91
Tableau 3 - Données relatives aux « autres informations de marché » par type d'indicateur..	95
Tableau 4 - Données recueillies pour la variable « autres informations de marché ».....	100
Tableau 5 - Composantes historique et prospective des données relatives aux « autres informations de marché »	101
Tableau 6 - Disparités des fréquences et des intervalles d'appartenance des données relatives aux « autres informations de marché ».....	108
Tableau 7 - Composition de l'échantillon selon le type de données	112
Tableau 8 - Composition des données homogènes de l'échantillon par classe d'actifs.....	113
Tableau 9 - Composition des données relatives aux « autres informations de marché » de l'échantillon par type d'indicateur	114
Tableau 10 - Composition des données relatives aux « autres informations de marché » de l'échantillon par fréquence de publication	115
Tableau 11 - Statistiques descriptives des taux de variation des données homogènes pour la période de tests de l'échantillon (2006-2011)	117
Tableau 12 - Statistiques descriptives des taux de variation des données relatives aux « autres informations de marché » pour la période de tests de l'échantillon (2006-2011).....	118
Tableau 13 - Statistiques descriptives des valeurs des données relatives aux « autres informations de marché » pour la période de tests de l'échantillon (2006-2011).....	120
Tableau 14 - Analyse en composantes principales des taux de variation des données homogènes pour la période totale de l'échantillon (2004-2011).....	122
Tableau 15 - Analyse en composantes principales des taux de variation des données relatives aux « autres informations de marché » pour la période totale de l'échantillon (2004-2011).	124
Tableau 16 - Comparaison de l'allocation d'actifs d'un portefeuille d'arbitrage selon le type de données pour la période de tests (2006-2011)	131
Tableau 17 - Comparaison de la contribution par classe d'actifs à la rentabilité d'un portefeuille d'arbitrage selon le type de données pour la période de tests (2006-2011).....	135
Tableau 18 - Contribution par actif à la rentabilité d'un portefeuille d'arbitrage selon le type de données	138
Tableau 19 - Rentabilités et volatilités annualisées d'un portefeuille d'arbitrage selon le type de données et les périodes de tests	141
Tableau 20 - Significativité statistique des corrélations entre les contributions par classe d'actifs à la rentabilité d'un portefeuille d'arbitrage selon le type de données pour la période de tests de l'échantillon (2006-2011)	144
Tableau 21 - Significativité statistique des corrélations entre données homogènes pour la période de tests de l'échantillon (2006-2011)	146
Tableau 22 - Test de Kupiec (1995) d'un portefeuille d'arbitrage selon le type de données et la période de tests	155
Tableau 23 - Composition de l'échantillon du test de robustesse selon le type de données..	158
Tableau 24 – Test de robustesse en présence de variables aléatoires : significativité statistique des corrélations entre les contributions par classe d'actifs à la rentabilité d'un portefeuille d'arbitrage selon le type de données pour la période de tests de l'échantillon (2006-2011) .	159
Tableau 25 - Erreur moyenne des simulations de Monte Carlo selon le type de données pour la période de tests de l'échantillon (2006 – 2011)	161
Tableau 26 – Test de robustesse à l'aide d'une période d'« apprentissage » étendue :	

comparaison de l'allocation d'actifs d'un portefeuille d'arbitrage selon le type de données pour une période de tests décalée (février 2009 - mars 2011)	163
Tableau 27 – Test de robustesse à l'aide d'une période d'« apprentissage » étendue : rentabilité et volatilité annualisées d'un portefeuille d'arbitrage selon le type de données pour une période de tests décalée (février 2009 - mars 2011).....	164
Tableau 28 - Comparaison de l'allocation d'actifs d'un portefeuille d'arbitrage optimisé selon le type de données pour la période de tests (2006-2011)	172
Tableau 29 - Comparaison de la contribution par classe d'actifs à la rentabilité d'un portefeuille d'arbitrage optimisé selon le type de données pour la période de tests (2006-2011)	175
Tableau 30 - Contribution par actif à la rentabilité d'un portefeuille d'arbitrage optimisé selon le type de données	177
Tableau 31 - Significativité statistique des corrélations entre les contributions par classe d'actifs à la rentabilité d'un portefeuille d'arbitrage optimisé selon le type de données pour la période de tests de l'échantillon (2006-2011)	179
Tableau 32 - Rentabilités et volatilités annualisées d'un portefeuille d'arbitrage optimisé selon le type de données et les périodes de tests	182
Tableau 33 - Tests de Kupiec (1995) d'un portefeuille d'arbitrage optimisé selon le type de données et la période de tests	186
Tableau 34 – Test de robustesse en présence de variables aléatoires : significativité statistique des corrélations entre les contributions par classe d'actifs à la rentabilité d'un portefeuille d'arbitrage optimisé selon le type de données pour la période de tests de l'échantillon (2006-2011).....	189
Tableau 35 - Erreur moyenne des simulations de Monte Carlo d'un portefeuille optimisé selon le type de données pour la période de tests de l'échantillon (2006 – 2011)	191
Tableau 36 – Test de robustesse à l'aide d'une période d'« apprentissage » étendue : comparaison de l'allocation d'actifs d'un portefeuille d'arbitrage optimisé selon le type de données pour une période de tests décalée (février 2009 - mars 2011)	192
Tableau 37 – Test de robustesse à l'aide d'une période d'« apprentissage » étendue : rentabilité et volatilité annualisées d'un portefeuille d'arbitrage optimisé selon le type de données pour une période de tests décalée (février 2009 - mars 2011)	193

LISTE DES FIGURES

Figure 1 - Plan de la thèse par chapitre : principaux contenus par chapitre.....	21
Figure 2 - Processus de quantification vectorielle (N = 10 points).....	70
Figure 3 - Signaux d'ouverture et de fermeture de positions pour des résidus de type Ornstein-Uhlenbeck.....	72
Figure 4 - Approche paramétrique de <i>market timing</i> par ACP et par quantification.....	75
Figure 5 - <i>Market timing</i> et allocations d'actifs d'un portefeuille d'arbitrage issu d'une approche non-paramétrique par ACP et par quantification.....	77
Figure 6 - Passage d'une variable à une donnée.....	88
Figure 7 - Critère de choix des données de la variable « autres informations de marché ».....	96
Figure 8 - Objectif final des traitements appliqués aux données de l'échantillon.....	105
Figure 9 - Traitement des données de cotation.....	106
Figure 10 - Traitement des indices de volatilités.....	106
Figure 11 - Traitement des « autres informations de marché » mensuelles.....	107
Figure 12 - Traitement des « autres informations de marché » hebdomadaires.....	107
Figure 13 - Valeur quotidienne du premier facteur de risque issu d'une analyse en composantes principales (ACP) des données homogènes pour la période totale de l'échantillon (2004-2011).....	125
Figure 14 - Valeur quotidienne du premier facteur de risque issu d'une analyse en composantes principales (ACP) des données relatives « autres informations de marché » pour la période totale de l'échantillon (2004-2011).....	126
Figure 15 - Principales étapes du test de Kupiec (1995).....	153
Figure 16 - Liste des contraintes de la diversification optimale de notre portefeuille d'arbitrage : minimisation de la volatilité sous contraintes.....	169
Figure 17 - Valorisation quotidienne d'un portefeuille d'arbitrage « euro neutre » issue d'une diversification optimale selon le type de données (2006 – 2011).....	180

TABLE DES MATIÈRES

<i>Introduction Générale</i>	7
1. Contexte de la recherche	7
2. Problématique générale et questions de recherche.....	9
3. Fondements théoriques et cadre conceptuel de la recherche.....	10
4. Contributions de la recherche.....	15
5. Démarche méthodologique de la recherche	16
6. Plan de la recherche.....	18
<i>Chapitre 1. Littérature et concepts clés</i>	22
1. Hétérogénéité des données en Sciences de Gestion	23
1.1. Structure des données hétérogènes.....	23
1.2. Complémentarité des données hétérogènes en Sciences de Gestion.....	25
2. Données hétérogènes en univers multi-classes d'actifs	30
2.1. Complexité du cadre multi-classes d'actifs.....	30
2.2. Interactions entre données hétérogènes en univers multi-classes d'actifs	31
3. Arbitrage statistique : stratégies historiques et limites.....	34
3.1. Stratégies historiques.....	35
3.2. Limites et interrogations.....	40
4. Diversification de portefeuille.....	43
4.1. Diversifications « naïve » et optimale.....	43
4.2. Analyse comparative : diversification « naïve » <i>versus</i> diversification optimale.....	45
5. Outils quantitatifs de la recherche	49
5.1. Analyse factorielle des données historiques : analyse en composantes principales..	49
5.2. Algorithmes de quantification optimale : processus de compression d'un signal	54
<i>Chapitre 2. Méthodes d'arbitrage systématique et prise en compte de données hétérogènes</i>	59
1. Prémices d'une approche systématique.....	60
1.1. Approche systématique <i>versus</i> approche discrétionnaire en univers multi-classes d'actifs.....	60
1.2. Détermination systématique d'un risque spécifique	62
1.2.1. Cadre historique de portefeuilles « marché neutre ».....	62
1.2.2. Incorporation de données hétérogènes au risque spécifique d'un portefeuille d'arbitrage	64
2. Arbitrage systématique : une approche paramétrique	68
2.1. Quantification d'un signal résiduel	68
2.2. Quantification d'un signal résiduel cumulé	71
2.3. Détermination de signaux d'investissement.....	74
3. Arbitrage systématique : une approche non-paramétrique.....	76
3.1. Evaluation directionnelle des données homogènes à l'aide de probabilités de transition.....	76
3.2. Interprétation des signaux d'investissement	78

4. Analyse comparative des approches : paramétrique <i>versus</i> non-paramétrique.....	80
4.1. Convergence des approches	80
4.2. Singularité des approches.....	80
Chapitre 3. Échantillon de l'étude.....	83
1. Périodes de l'échantillon	83
1.1. Périodes de l'étude : avant, pendant et après la crise des <i>subprimes</i>	83
1.2. Période d'« apprentissage » et périodes de tests	85
2. Recueil des données de l'échantillon	88
2.1. Recueil des données de cotation.....	89
2.2. Recueil des données relatives aux « autres informations de marché »	95
2.2.1. Critères de choix des données relatives aux « autres informations de marché ».....	96
2.2.2. Similarités entre données recueillies	102
3. Traitements appliqués aux données.....	104
3.1. Comparaison et agrégation de données hétérogènes.....	104
3.2. Traitement spécifique des données relatives aux « autres informations de marché »	109
4. Composition et caractéristiques de l'échantillon	112
4.1. Composition de l'échantillon	112
4.2. Caractéristiques de l'échantillon	115
4.2.1. Statistiques descriptives	115
4.2.2. Analyse des données de l'échantillon	121
Chapitre 4. Portefeuille d'arbitrage « euro neutre » en diversification « naïve ».....	127
1. Influence des données hétérogènes sur des portefeuilles d'arbitrage « euro neutre » en allocation « naïve »	127
1.1. Modification de l'allocation entre classes d'actifs.....	129
1.2. Modification de la contribution par classe d'actifs à la rentabilité	134
1.3. Réduction en moyenne de la variabilité des rentabilités.....	140
1.4. Accroissement du ratio rentabilité/volatilité après la crise des <i>subprimes</i>	148
1.5. Réduction des pertes financières probables pendant la crise des <i>subprimes</i>	150
2. Eléments de robustesse.....	157
2.1. Simulations en présence de variables aléatoires.....	157
2.2. Simulations de Monte Carlo.....	160
2.3. Extension de la période d'« apprentissage » de l'échantillon	162
Chapitre 5. Portefeuille d'arbitrage « euro neutre » en diversification optimale.....	166
1. Cadre et contraintes de la diversification optimale	166
1.1. Cadre théorique de la diversification optimale	166
1.2. Contraintes de la diversification optimale.....	167
2. Influence des données hétérogènes sur des portefeuilles d'arbitrage « euro neutre » en allocation optimale	170
2.1. Changement d'allocations entre classes d'actifs.....	171
2.2. Modification de contribution par classes d'actifs à la rentabilité.....	174
2.3. Accroissement du ratio rentabilité/volatilité	182
2.4. Réduction des pertes financières probables.....	186
3. Eléments de robustesse.....	187
3.1. Simulations en présence de variables aléatoires.....	188

3.2. Simulations de Monte Carlo.....	190
3.3. Extension de la période d'« apprentissage » de l'échantillon	192
4. Analyse comparative des résultats empiriques : diversification « naïve » <i>versus</i> diversification optimale.....	194
4.1. Convergences en présence de données hétérogènes	194
4.1.1. Changement d'allocation d'actifs.....	194
4.1.2. Variations des contributions par classe d'actifs à la rentabilité	195
4.1.3. Accroissement du ratio rentabilité/volatilité après la crise des <i>subprimes</i>	195
4.1.4. Baisse du niveau des corrélations entre contributions par classes d'actifs à la rentabilité.....	196
4.1.5. Contributions des nouveaux pays industrialisés à la rentabilité.....	197
4.1.6. Tests de robustesses convergents	197
4.2. Divergences en présence de données hétérogènes	198
4.2.1. Différences d'analyse du niveau de ratio rentabilité/volatilité.....	198
4.2.2. Modification de la hiérarchie d'allocations par classe d'actifs	199
4.2.3. Différences de significativité des pertes financières probables par période	200
Conclusion Générale.....	201
1. Principaux résultats de la recherche	201
1.1. Changement d'allocation d'actifs en présence de données hétérogènes.....	204
1.2. Modification de la contribution par classes d'actifs à la rentabilité en présence de données hétérogènes.....	205
1.3. Baisse des corrélations entre les rentabilités par classe d'actifs en présence de données hétérogènes.....	206
1.4. Baisse relative de la volatilité des rendements d'un portefeuille d'arbitrage en présence de données hétérogènes	206
1.5. Accroissement du ratio rentabilité/volatilité après la crise des <i>subprimes</i>	207
1.6. Diminution des pertes probables en présence de données hétérogènes pendant la crise des <i>subprimes</i>	208
2. Apports de la recherche.....	209
2.1. Apports empiriques	209
2.2. Apports méthodologiques	212
2.3. Apports managériaux	213
3. Limites de la recherche.....	214
3.1. Exploitation de fréquences inter-journalières des prix des actifs.....	214
3.2. Gestion du risque non-optimisée.....	215
4. Perspectives de recherche.....	215
4.1. Utilisation de fréquences intra-journalières des prix des actifs.....	215
4.2. Gestion optimale du risque et processus d'allocation des portefeuilles d'arbitrage	216
Bibliographie	218
Liste des tableaux	228
Liste des figures.....	230
Table des matières	231
Liste des Annexes	234

LISTE DES ANNEXES

- Annexe 1 : Code VBA/Excel pour la réplication quotidienne de taux de variation mensuels des données relatives aux « autres informations de marché » (chapitre 3)
- Annexe 2 : Code VBA/Excel pour la réplication quotidienne des taux de variation hebdomadaires des données relatives aux « autres informations de marché » (chapitre 3)
- Annexe 3 : Statistiques descriptives des taux de variation des données homogènes pour la période de tests de l'échantillon précédant la crise des *subprimes* (chapitre 3)
- Annexe 4 : Statistiques descriptives des taux de variation des données homogènes pour la période de tests de l'échantillon de la crise des *subprimes* (chapitre 3)
- Annexe 5 : Statistiques descriptives des taux de variation des données homogènes pour la période de tests de l'échantillon après la crise des *subprimes* (chapitre 3)
- Annexe 6 : Statistiques descriptives des taux de variation des données relatives aux « autres informations de marché » pour la période de tests de l'échantillon précédant la crise des *subprimes* (chapitre 3)
- Annexe 7 : Statistiques descriptives des taux de variation des données relatives aux « autres informations de marché » pour la période de tests de l'échantillon de la crise des *subprimes* (chapitre 3)
- Annexe 8 : Statistiques descriptives des taux de variation des données relatives aux « autres informations de marché » pour la période de tests de l'échantillon après la crise des *subprimes* (chapitre 3)

Annexe 9 : Valorisation quotidienne d'un portefeuille d'arbitrage « euro neutre » en allocation « naïve » selon le type de données (chapitre 4)

ANNEXE 1 - Code VBA/Excel pour la réplification quotidienne de taux de variation mensuels des données relatives aux « autres informations de marché »

```

Sub Replication_Rendement_Quotidien_Donnees_Mensuelles()
Worksheets("Rendements_quotidiens").Select
'Chercher le dernier élément de la plage
n = Range("E7").End(xlDown).Row
'initialisation de la ligne première ligne
first_line = 7
'first_line = Cells(2, 4).Row
Range("F7").Select
'indice de ligne
For i = 7 To n
    'test des dates pour savoir si 1 ou 0 dans la cellule : si 1 on calcule le rendement
    ' sinon nous avons 0 donc nous ne faisons rien
    If Cells(i, 4).Value <> Cells(i + 1, 4).Value And Cells(i + 1, 4).Value <> "" Then
        'selectionner la cellule de calcul du rendement
        Cells(i + 1, 6).Select
        'calcul du rendement
        'rendement « classique » :  $[X_{(i,t)} - X_{(i,t-1)}] / X_{(i,t-1)}$ 
        'ActiveCell.FormulaR1C1 = "=(RC[-1]-R[-1]C[-1])/R[-1]C[-1]"
        'rendement « spécifique » :  $[X_{(i,t)} - X_{(i,t-1)}] / |X_{(i,t-1)}|$  pour les séries BNB et ZEW
        ActiveCell.FormulaR1C1 = "=(R[1]C[-1]-RC[-1])/ABS(RC[-1])"
        'réplication du rendement
        Range(Cells(first_line, 7), Cells(i + 1, 7)).Value = Cells(i + 1, 6).Value
        'changer de la valeur de la variable first_line
        first_line = i + 2
        'trouver la borne inférieure de l'intervalle dans lequel est calculée le rendement
        i = i + 2
    End If
Next i
Application.Goto Reference:=Worksheets("Rendements_quotidiens").Range("A1"), _
scroll:=True
End Sub

```

ANNEXE 2 - Code VBA/Excel pour la réplification quotidienne de taux de variation hebdomadaires des données relatives aux « autres informations de marché »

```
Sub Replication_Rendement_Quotidien_Donnees_Hebdomadaires()
Worksheets("Rendements_quotidiens_bis").Select
'Chercher le dernier élément de la plage
n = Range("E7").End(xlDown).Row
'initialisation de la ligne première ligne
first_line = 7
Range("F7").Select
'indice de ligne
For i = 7 To n
'test des dates pour savoir si 1 ou 0 dans la cellule: si 1 on calcule le rendement
'sinon nous avons 0 donc nous ne faisons rien
If Cells(i, 4).Value <> Cells(i + 1, 4).Value And Cells(i + 1, 4).Value <> "" Then
'selectionner la cellule de calcul du rendement
Cells(i + 1, 6).Select
'calcul du rendement
ActiveCell.FormulaR1C1 = "=(RC[-1]-R[-2]C[-1])/R[-2]C[-1]"
'réplication du rendement
Range(Cells(first_line, 7), Cells(i + 1, 7)).Value = Cells(i + 1, 6).Value
'changer de la valeur de la variable first_line
first_line = i + 2
'trouver la borne inférieure de l'intervalle dans lequel
'est calculée le rendement
i = i + 2
End If
Next i
Application.Goto Reference:=Worksheets("Rendements_quotidiens").Range("A1"), _
scroll:=True
End Sub
```

ANNEXE 3 - Statistiques descriptives des taux de variation des données homogènes pour la période de tests de l'échantillon précédant la crise des *subprimes* (juillet 2006 – janvier 2007)

Données	Moyenne	Ecart-type	Médiane	Minimum	Maximum
France 2004 4% 25/04/14	-0,0033%	0,1854%	-0,0098%	-0,3932%	0,5405%
US 2004 4% 15/02/14	0,0075%	0,4067%	0,0234%	-0,9766%	1,1272%
Japan 2004 1,5% 20/03/14	-0,0289%	0,3474%	0,0000%	-0,8858%	1,2184%
Brazil 2004 10,5% 14/07/14	0,0202%	0,4237%	0,0021%	-1,0651%	1,0310%
Brent (pétrole)	-0,1770%	1,8412%	-0,2049%	-5,2897%	5,3895%
West Texas Intermediate (pétrole)	-0,1526%	2,0036%	-0,1150%	-6,3165%	7,0096%
Or	0,0202%	1,2009%	0,0303%	-4,2063%	3,7178%
Blé	0,1912%	2,4213%	0,0000%	-6,9364%	7,1038%
Maïs	0,3818%	2,1849%	0,0000%	-6,0606%	7,6531%
Bovespa (Brésil)	0,1517%	1,2791%	0,0000%	-4,0301%	4,7111%
Bse sensx (Inde)	0,2047%	1,1280%	0,2056%	-3,6055%	3,4535%
Cac 40 (France)	0,0870%	0,7901%	0,0864%	-1,8052%	2,3656%
Dax (Allemagne)	0,1236%	0,8034%	0,1384%	-1,9605%	2,6393%
Ftse 100 (Royaume Uni)	0,0371%	0,6362%	0,0068%	-1,6302%	1,9954%
S&P 500 (Etats-Unis)	0,0879%	0,5219%	0,0997%	-1,3598%	1,8555%
Nasdaq 100 (Etats-Unis)	0,1094%	0,9118%	0,1495%	-2,2230%	2,6835%
Asx 200 (Australie)	0,0822%	0,7844%	0,0282%	-2,3306%	2,0763%
Shanghai SE (Chine)	0,3340%	1,5497%	0,2169%	-4,9174%	4,7263%
Nikkei 225 (Japon)	0,0904%	0,9483%	0,0000%	-2,7484%	3,0798%
Dollar américain / euro	-0,0082%	0,4090%	0,0039%	-1,0846%	0,9266%
Sterling /euro	0,0289%	0,2624%	0,0428%	-0,6958%	1,0245%
Franc suisse / euro	-0,0231%	0,1594%	-0,0318%	-0,3526%	0,7004%
Yen / euro	-0,0492%	0,3437%	-0,0550%	-0,8497%	0,9853%
Yuan /euro	0,0081%	0,4175%	0,0000%	-1,5275%	0,8989%
Couronne norvégienne / euro	-0,0156%	0,3905%	-0,0093%	-1,2886%	1,0531%
Real brésilien / euro	0,0081%	0,5932%	0,0418%	-1,9642%	2,0771%
Roupie indienne / euro	0,0205%	0,4297%	0,0551%	-1,1944%	1,0024%
Won coréen / euro	-0,0060%	0,4876%	-0,0102%	-1,4916%	1,2032%
Dollar australien / euro	0,0135%	0,3745%	0,0089%	-1,3291%	1,1003%
Rand sud africain / euro	-0,0149%	0,8779%	0,0462%	-3,6235%	1,8157%
Dollar néo-zélandais / euro	0,0717%	0,5985%	0,0550%	-1,5809%	2,3434%
Dollar canadien / euro	-0,0475%	0,4481%	-0,0898%	-1,1759%	1,1187%

En partant de la gauche, les données homogènes sont en première colonne. De la deuxième à la sixième colonne, nous avons respectivement la moyenne, l'écart-type, la médiane, le minimum et le maximum des taux de variation quotidiens.

ANNEXE 4 - Statistiques descriptives des taux de variation des données homogènes pour la période de tests de l'échantillon de la crise des *subprimes* (février 2007 – mars 2009)

Données	Moyenne	Ecart-type	Médiane	Minimum	Maximum
France 2004 4% 25/04/14	0,0122%	0,3050%	0,0098%	-0,9979%	1,4801%
US 2004 4% 15/02/14	0,0278%	0,8660%	0,0000%	-5,0138%	3,7007%
Japan 2004 1,5% 20/03/14	0,0437%	1,1282%	-0,0568%	-4,1565%	7,4929%
Brazil 2004 10,5% 14/07/14	-0,0044%	0,8691%	-0,0097%	-7,8651%	5,6262%
Brent (pétrole)	-0,0081%	2,5220%	0,0000%	-10,7920%	16,9479%
West Texas Intermediate (pétrole)	0,0234%	3,3434%	-0,0397%	-13,9394%	22,1112%
Or	0,0669%	1,4580%	0,0952%	-7,4416%	6,4977%
Blé	0,0722%	3,5260%	0,0000%	-20,3704%	11,0132%
Maïs	0,0224%	2,5381%	0,0000%	-10,3806%	9,0517%
Bovespa (Brésil)	0,0185%	2,6095%	0,0359%	-11,3931%	14,6592%
Bse sensx (Inde)	-0,0399%	2,2823%	0,0000%	-10,9564%	8,2210%
Cac 40 (France)	-0,1027%	2,0001%	-0,0296%	-9,0368%	11,1762%
Dax (Allemagne)	-0,0721%	1,8989%	0,0000%	-7,1639%	11,4020%
Ftse 100 (Royaume Uni)	-0,0636%	1,8707%	-0,0027%	-8,8493%	9,8388%
S&P 500 (Etats-Unis)	-0,0835%	2,0444%	0,0024%	-9,0350%	11,5800%
Nasdaq 100 (Etats-Unis)	-0,0433%	2,1190%	0,0200%	-10,5195%	12,5799%
Asx 200 (Australie)	-0,0704%	1,6846%	-0,0008%	-8,3362%	5,7896%
Shanghai SE (Chine)	0,0014%	2,4487%	0,0174%	-8,8449%	9,4537%
Nikkei 225 (Japon)	-0,1108%	2,2034%	0,0000%	-11,4064%	14,1503%
Dollar américain / euro	-0,0010%	0,7544%	-0,0298%	-4,5122%	3,9189%
Sterling / euro	-0,0567%	0,6574%	-0,0187%	-2,8650%	3,2233%
Franc suisse / euro	0,0137%	0,4601%	0,0000%	-2,7670%	2,0513%
Yen / euro	0,0388%	1,1205%	-0,0124%	-7,7777%	6,6571%
Yuan / euro	0,0227%	0,7379%	0,0122%	-3,4515%	2,6159%
Couronne norvégienne / euro	-0,0136%	0,6107%	0,0000%	-3,5479%	4,0341%
Real brésilien / euro	-0,0080%	1,4344%	0,0226%	-5,7295%	7,2875%
Roupie indienne / euro	-0,0256%	0,7451%	-0,0378%	-3,4283%	3,1933%
Won coréen / euro	-0,0604%	1,3829%	-0,0584%	-9,9360%	12,0678%
Dollar australien / euro	-0,0201%	0,9361%	0,0385%	-5,2175%	6,1638%
Rand sud africain / euro	-0,0431%	1,3879%	0,0202%	-13,8904%	6,2513%
Dollar néo-zélandais / euro	-0,0319%	0,9386%	0,0000%	-4,6993%	3,7114%
Dollar canadien / euro	-0,0122%	0,7721%	0,0000%	-3,2444%	4,5696%

En partant de la gauche, les données homogènes sont en première colonne. De la deuxième à la sixième colonne, nous avons respectivement la moyenne, l'écart-type, la médiane, le minimum et le maximum des taux de variation quotidiens.

ANNEXE 5 - Statistiques descriptives des taux de variation des données homogènes pour la période de tests de l'échantillon après la crise des *subprimes* (avril 2009 – mars 2011)

Données	Moyenne	Ecart-type	Médiane	Minimum	Maximum
France 2004 4% 25/04/14	-0,0018%	0,1683%	0,0000%	-0,6530%	0,5887%
US 2004 4% 15/02/14	-0,0130%	0,7105%	-0,0278%	-2,1982%	2,0245%
Japan 2004 1,5% 20/03/14	0,0302%	0,8548%	-0,0564%	-3,9097%	3,0296%
Brazil 2004 10,5% 14/07/14	-0,0018%	0,6710%	0,0000%	-2,0878%	2,0951%
Brent (pétrole)	0,1863%	1,7275%	0,2527%	-6,0547%	8,0903%
West Texas Intermediate (pétrole)	0,1508%	1,9160%	0,1071%	-7,9575%	8,3571%
Or	0,0836%	1,0884%	0,0715%	-4,4894%	3,1889%
Blé	0,1431%	3,0384%	0,0000%	-10,2459%	11,2299%
Maïs	0,1421%	2,2157%	0,0000%	-8,6466%	11,3514%
Bovespa (Brésil)	0,1116%	1,4129%	0,0730%	-4,7482%	6,5872%
Bse sensex (Inde)	0,1399%	1,5374%	0,0568%	-5,8315%	17,3393%
Cac 40 (France)	0,0835%	1,4133%	0,0417%	-4,5983%	9,6593%
Dax (Allemagne)	0,1218%	1,2791%	0,0940%	-4,0741%	6,0723%
Ftse 100 (Royaume Uni)	0,0908%	1,1102%	0,0584%	-3,1815%	5,1610%
S&P 500 (Etats-Unis)	0,1086%	1,1522%	0,0982%	-4,2790%	4,3965%
Nasdaq 100 (Etats-Unis)	0,1369%	1,1967%	0,1518%	-3,9183%	5,0377%
Asx 200 (Australie)	0,0641%	1,0527%	0,0274%	-3,4413%	3,4676%
Shanghai SE (Chine)	0,0514%	1,5260%	0,0846%	-6,7472%	4,7932%
Nikkei 225 (Japon)	0,0618%	1,3185%	0,0082%	-3,8419%	4,5484%
Dollar américain / euro	-0,0075%	0,6635%	0,0085%	-2,0803%	1,6135%
Sterling / euro	0,0166%	0,5894%	0,0042%	-1,7509%	1,7337%
Franc suisse / euro	0,0313%	0,4609%	0,0284%	-1,7219%	1,8698%
Yen / euro	0,0299%	0,9173%	-0,0211%	-3,1927%	4,9821%
Yuan / euro	0,0013%	0,7047%	0,0000%	-2,0515%	2,5224%
Couronne norvégienne / euro	0,0296%	0,5130%	0,0573%	-3,0205%	1,8740%
Real brésilien / euro	0,0602%	0,7623%	0,0797%	-3,1987%	3,5375%
Roupie indienne / euro	0,0161%	0,6482%	0,0223%	-2,2745%	3,8984%
Won coréen / euro	0,0352%	0,9166%	0,0187%	-3,5234%	4,8899%
Dollar australien / euro	0,0709%	0,6664%	0,0473%	-4,1134%	2,2886%
Rand sud africain / euro	0,0573%	0,8070%	0,0505%	-3,1682%	2,5509%
Dollar néo-zélandais / euro	0,0477%	0,7405%	0,0447%	-2,4413%	2,6790%
Dollar canadien / euro	0,0437%	0,6982%	0,0587%	-2,3220%	2,2543%

En partant de la gauche, les données homogènes sont en première colonne. De la deuxième à la sixième colonne, nous avons respectivement la moyenne, l'écart-type, la médiane, le minimum et le maximum des taux de variation quotidiens.

**ANNEXE 6 - Statistiques descriptives des taux de variation des données relatives aux
« autres informations de marché » pour la période de tests de l'échantillon précédant la
crise des *subprimes* (juillet 2006 – janvier 2007)**

Données	Source	Moyenne	Ecart-type	Médiane	Minimum	Maximum
Sentiment économique des investisseurs pour l'économie allemande	Zentrum für Europäische Wirtschaftsforschung (Zew)	-56,61%	118,03%	-23,42%	-296,43%	81,05%
Sentiment des consommateurs américains	Université du Michigan, Thomson Reuters	3,10%	7,09%	0,69%	-6,21%	15,00%
Climat des affaires allemand	Ifo Institute	0,30%	1,06%	0,18%	-0,97%	1,79%
Industrie manufacturière belge	Banque nationale de Belgique	-4,49%	46,84%	-21,82%	-57,14%	66,67%
Directeurs d'achat	Institute of Supply Management	-0,79%	2,24%	-1,53%	-3,70%	2,19%
Sentiment haussier des marchés boursiers américains	American Association of Individual Investors	2,46%	15,40%	3,42%	-34,66%	29,65%
Sentiment baissier des marchés boursiers américains	American Association of Individual Investors	1,45%	21,19%	-3,91%	-30,95%	46,63%
Sentiment neutre des marchés boursiers américains	American Association of Individual Investors	3,56%	29,45%	-1,16%	-49,45%	79,27%
Indice de volatilité VIX américain	CBOE	-0,08%	4,84%	-0,44%	-13,91%	22,77%
Indice de volatilité VDAX allemand	Euronext	-0,02%	4,55%	-0,03%	-12,87%	13,73%
Indice de volatilité VCAC français	Euronext	0,03%	5,04%	0,00%	-18,09%	15,08%

En partant de la gauche, le type de donnée d'« autres informations de marché » est en première colonne. La source de la donnée est en deuxième colonne. De la troisième à la septième colonne, nous avons la moyenne, l'écart-type, la médiane, le minimum et le maximum des taux de variation quotidiens.

**ANNEXE 7 - Statistiques descriptives des taux de variation des données relatives aux
« autres informations de marché » pour la période de tests de l'échantillon de la crise
des subprimes (février 2007 – mars 2009)**

Données	Source	Moyenne	Ecart-type	Médiane	Minimum	Maximum
Sentiment économique des investisseurs pour l'économie allemande	Zentrum für Europäische Wirtschaftsforschung (Zew)	3,92%	76,86%	0,00%	-166,35%	184,48%
Sentiment des consommateurs américains	Université du Michigan, Thomson Reuters	-1,54%	7,45%	-3,56%	-15,18%	16,06%
Climat des affaires allemand	Ifo Institute	-1,08%	1,69%	-0,83%	-5,37%	1,16%
Industrie manufacturière belge	Banque nationale de Belgique	-33,43%	138,37%	-13,33%	-437,50%	350,00%
Directeurs d'achat	Institute of Supply Management	-1,17%	4,37%	0,00%	-13,17%	5,44%
Sentiment haussier des marchés boursiers américains	American Association of Individual Investors	0,67%	24,47%	-3,97%	-43,26%	102,92%
Sentiment baissier des marchés boursiers américains	American Association of Individual Investors	2,58%	20,89%	-0,25%	-43,52%	77,50%
Sentiment neutre des marchés boursiers américains	American Association of Individual Investors	-0,28%	25,66%	-3,60%	-47,42%	150,98%
Indice de volatilité VIX américain	CBOE	0,57%	8,21%	-0,06%	-24,68%	64,22%
Indice de volatilité VDAX allemand	Euronext	0,35%	6,04%	0,00%	-17,24%	34,96%
Indice de volatilité VCAC français	Euronext	0,45%	7,79%	0,00%	-31,05%	62,75%

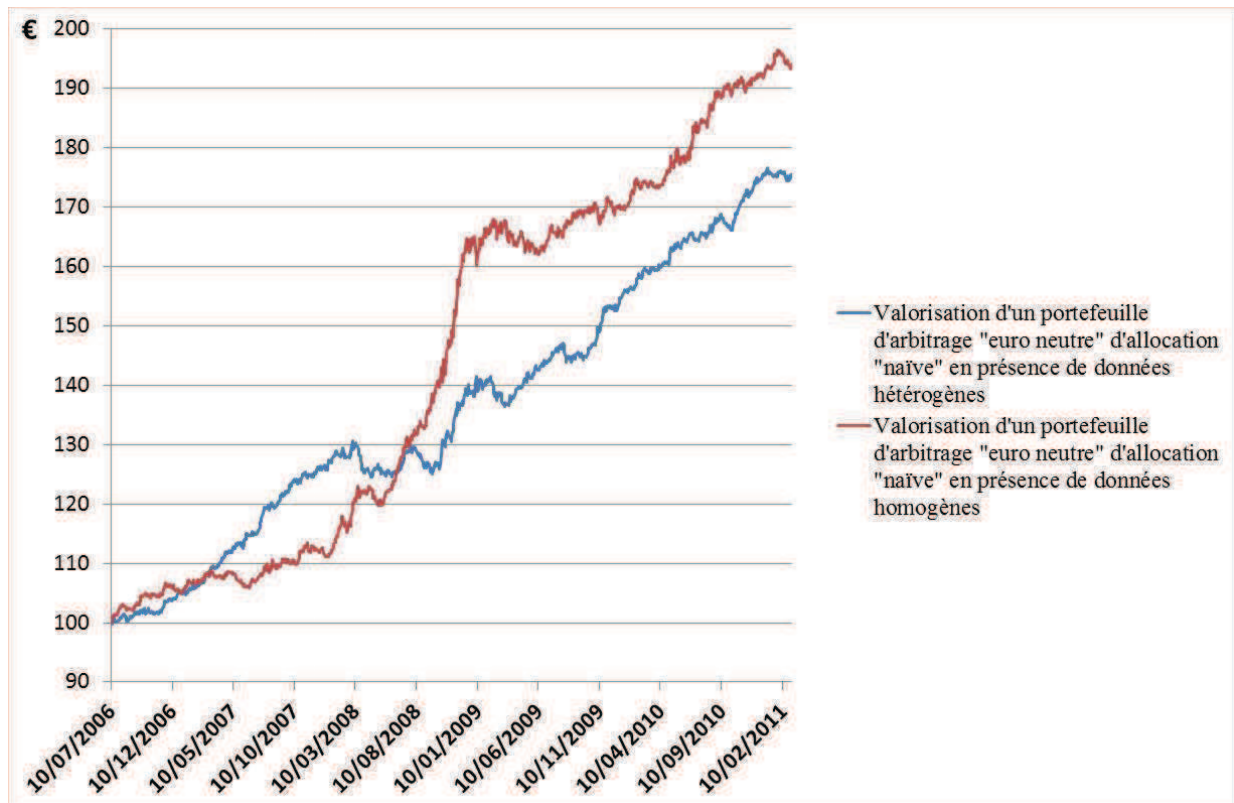
En partant de la gauche, le type de donnée d'« autres informations de marché » est en première colonne. La source de la donnée est en deuxième colonne. De la troisième à la septième colonne, nous avons la moyenne, l'écart-type, la médiane, le minimum et le maximum des taux de variation quotidiens.

**ANNEXE 8 – Statistiques descriptives des taux de variation des données relatives aux
« autres informations de marché » pour la période de tests de l'échantillon après la crise
des *subprimes* (avril 2009 – mars 2011)**

Données	Source	Moyenne	Ecart-type	Médiane	Minimum	Maximum
Sentiment économique des investisseurs pour l'économie allemande	Zentrum für Europäische Wirtschaftsforschung (Zew)	38,80%	121,28%	-1,37%	-130,71%	471,43%
Sentiment des consommateurs américains	Université du Michigan, Thomson Reuters	1,32%	6,51%	1,64%	-19,13%	17,94%
Climat des affaires allemand	Ifo Institute	1,35%	1,28%	0,96%	-0,52%	4,33%
Industrie manufacturière belge	Banque nationale de Belgique	19,50%	45,35%	12,33%	-73,33%	160,87%
Directeurs d'achat	Institute of Supply Management	2,33%	4,00%	1,61%	-4,76%	10,56%
Sentiment haussier des marchés boursiers américains	American Association of Individual Investors	4,55%	23,44%	0,22%	-34,03%	88,01%
Sentiment baissier des marchés boursiers américains	American Association of Individual Investors	3,07%	20,37%	5,09%	-39,56%	50,98%
Sentiment neutre des marchés boursiers américains	American Association of Individual Investors	2,79%	22,59%	0,00%	-47,40%	97,13%
Indice de volatilité VIX américain	CBOE	0,03%	6,56%	-0,83%	-29,57%	31,67%
Indice de volatilité VDAX allemand	Euronext	-0,02%	5,13%	-0,34%	-23,69%	32,76%
Indice de volatilité VCAC français	Euronext	0,07%	6,30%	-0,30%	-21,80%	39,34%

En partant de la gauche, le type de donnée d'« autres informations de marché » est en première colonne. La source de la donnée est en deuxième colonne. De la troisième à la septième colonne, nous avons la moyenne, l'écart-type, la médiane, le minimum et le maximum des taux de variation quotidiens.

ANNEXE 9 – Valorisation quotidienne d'un portefeuille d'arbitrage « euro neutre » issue d'une diversification « naïve » selon le type de données (2006 – 2011)



Les dates sont en abscisse et la valorisation quotidienne en euros est en ordonnée pour 100 euros investis en date de départ. La courbe bleue correspond à la valorisation d'un portefeuille « euro neutre » en allocation « naïve » (diversification « naïve ») composé à l'aide d'une analyse de données hétérogènes. La courbe rouge correspond à la valorisation d'un portefeuille « euro neutre » en allocation « naïve » (diversification « naïve ») composé à l'aide d'une analyse de données homogènes.

Vu le Président
M.....

Vu : les suffragants
MM

Vu et permis d'imprimer :

LE VICE-PRESIDENT DU CONSEIL SCIENTIFIQUE CHARGE DE LA RECHERCHE
DE L'UNIVERSITE PARIS-EST CRETEIL.

STRATEGIES D'ARBITRAGE SYSTEMATIQUE MULTI-CLASSES D'ACTIFS ET UTILISATION DE DONNEES HETEROGENES

Yohan FERERES

Résumé : les marchés financiers évoluent plus ou moins rapidement et fortement au gré des différents types d'information diffusés au cours des périodes d'étude. Dans ce contexte, nous cherchons à mesurer l'influence de tous types d'information sur des portefeuilles d'arbitrage systématique « euro neutres » multi-classes d'actifs, issus soit d'une diversification « naïve » (« 1/N ») soit d'une diversification optimale. Dans le cadre de nos recherches sur l'allocation tactique systématique, ces divers flux informationnels sont regroupés sous le terme de données hétérogènes (données de cotation et « autres informations de marché »). Les données de cotation sont des prix de clôture quotidiens d'actifs tandis que les « autres informations de marché » correspondent à trois types d'indicateurs : de conjoncture, de sentiments et de volatilité. Nous mesurons l'impact d'une combinaison de données hétérogènes sur nos portefeuilles d'arbitrage pour une période de tests incluant la crise des *subprimes*, à l'aide d'analyses de données (ACP) et de techniques probabilistes de quantification vectorielle. L'influence des données hétérogènes sur les portefeuilles d'arbitrage est mesurée notamment au travers d'une hausse de la rentabilité, d'un accroissement du ratio rentabilité/volatilité *post* crise des *subprimes*, d'une baisse de la volatilité ou d'une baisse des corrélations entre classes d'actifs. Ces découvertes empiriques permettent d'envisager la prise en compte des « autres informations de marché » comme élément de diversification du risque d'un portefeuille. Nous formalisons des éléments de réponse au défi posé par l'allocation tactique multi-classes d'actifs (Blitz et Vliet, 2008), en intégrant des variables « prédictives » à un processus systématique de *market timing* qui incorpore de manière quantitative des données hétérogènes.

Mots-clés : combinaison des données hétérogènes ; « autres informations de marché » ; univers multi-classes d'actifs ; gestion de portefeuille ; arbitrage statistique ; allocation tactique internationale ; stratégies d'arbitrage systématique ; quantification

Laboratoire : IRG – Université Paris-Est

MULTI-ASSETS SYSTEMATIC ARBITRAGE STRATEGIES AND USE OF HETEROGENEOUS DATA

Yohan FERERES

Abstract : financial markets evolve more or less rapidly and strongly to all kind of information depending on time period of study. In this context, we intend to measure a broad set of information influence on systematic multi-assets classes “euro neutral” arbitrage portfolios either for “naive” diversification and optimal diversification. Our research focuses on systematic tactical asset allocation and we group these information under the name of heterogeneous data (market data and “other market information”). Market data are “end of day” asset closing prices and “other market information” gather economic cycle, sentiment and volatility indicators. We assess the influence of a heterogeneous data combination on our arbitrage portfolios for a time period including the subprimes crisis period and thanks to data analysis and quantization algorithms. The impact of a heterogeneous data combination on our arbitrage portfolio is materialized by increasing return, increasing return/volatility ratio for the post subprimes crisis period, decreasing volatility and asset class correlations. These empirical findings suggest that “other market information” presence could be an element of arbitrage portfolio risk diversification. Furthermore, we investigate and bring empirical results to Blitz and Vliet (2008) issue on global tactical asset allocation (GTAA) by considering “predictive” variables with a systematic market timing process integrating heterogeneous data thanks to a quantitative data processing.

Key words : heterogeneous data combination; “other market information”; multi-asset classes context; asset management; statistical arbitrage ; global tactical asset allocation ; systematic arbitrage strategies ; quantization

Research Unit : IRG – Université Paris-Est