Facultés Universitaires Catholiques de Mons
et
Université de Perpignan

Rendements intrajournaliers et flux informationnel : Une étude sur Euronext Paris

Thèse pour l'obtention du doctorat en sciences de gestion

présentée et soutenue publiquement par

Alexis CELLIER

17 décembre 2003

JURY

Directeurs de thèse :
***R. De Winne**, Professeur aux FUCaM*
/
***A. François-Heude**, Professeur à l'Université de Perpignan*

Suffragants :
Christophe Bisière,
Professeur à l'Université de Toulouse I,
Laurence Broze,
Professeur à l'Université de Lille III
Michel Levasseur,
Professeur à l'Université de Lille II
Isabelle Platten,
Professeur aux FUCaM,

L'Université et les Facultés n'entendent donner aucune approbation ni improbation aux opinions émises dans cette thèse. Ces opinions doivent être considérées comme propres à l'auteur et n'engageant que sa responsabilité.

Remerciements

Je tiens tout d'abord à remercier Laurence Broze, Christophe Bisière, Michel Levasseur et Isabelle Platten qui ont accepté de participer à ce jury soutenance. Je leur en suis très reconnaissant.

Je ne pourrai jamais remercier comme il le mérite le CIM (Dirk Simoens, Françoise Degembe…) pour m'avoir octroyé une bourse qui m'a permis d'entreprendre et de conclure ce travail.

Un tout grand merci à tous ceux qui ont pu ou dû supporter au jour le jour mes errements concernant SAS, LaTeX, et Gnuplot tout en gardant le sourire (ou presque). La considérable résistance dont ont fait preuve Catherine, Christophe, Marie-France et tous ceux qui ont déjà pris leur envol des Fucam appelle plus que mon respect. Le département finance a toujours été pour moi un lieu dont l'atmosphère calme, chaleureuse et détendue était propice au travail, que ceux qui y ont contribué (Marie-Claire Capiau, Pierre Brehain, Huguette) sachent le souvenir attentionné que j'en conserve.

Mes remerciements vont également vers les membres du département de sciences économiques de l'université de Perpignan (Philipe Mahenc) qui m'ont toujours accueilli avec chaleur lors de mes séjours en dépit de l'exiguïté des bureaux.

Éric Renault m'a fait la faveur de m'accueillir au CIREQ (centre interuniversitaire de recherche en économie quantitative) pendant un an. Ce séjour à l'université de Montréal m'a énormément apporté. Il a contribué tout autant à l'aboutissement de ce travail qu'à mon ouverture d'esprit. Qu'il soit assuré de ma très sincère reconnaissance. Un grand merci également à tous ceux (Élise, Hélène, Pierre, Christian…) qui, par leur chaleur, ont rendu l'hiver canadien plus que supportable.

L'ESA a su fournir un terreau propice à une vocation. La qualité et l'intuition des séminaires dont ceux de Frédéric Lobez n'ont pu que faire éclore cette jeune pousse. Que tous les membres du GERME soient également assurés de ma reconnaissance pour l'ambiance studieuse (On va prendre un autre café ? Non ? Bon, alors un chocolat.) qu'ils ont instillé dans ce lieu.

Le PasFi 51 a été un lieu de discussions stimulantes et dépaysantes. Que ses animateurs (Éric de Bodt, Gérard Desmuliers…) sachent le plaisir que j'ai eu en leur trop rare compagnie.

Un merci tout spécial à David Bourghelle dont la passion s'est révélée contagieuse et m'a

conduit à suivre la voie qui trouve aujourd'hui un achèvement.

Comment oublier ceux qui sont toujours là, les parents et la famille qui ont réussi à me supporter (la performance est appréciable) au long de ce travail tout en me conservant un soutien sans faille (rémunéré par des pralines belges, il est vrai).

Enfin, mes remerciements ne sauraient épargner mes directeurs de thèse Rudy de Winne et Alain François-Heude. Leur disponibilité et la tolérance dont ils ont fait preuve face à mes doutes métaphysiques ont été des atouts sans prix.

Sommaire

Remerciements **i**

Résumé **xii**

Introduction générale **2**

I Déformations temporelles des rendements intrajournaliers : À la recherche du temps perdu **9**

Introduction **10**

1 Conception des séries de rendements et des déformations temporelles **15**

1.1 Aspects Méthodologiques 15
1.2 Analyse de l'activité 28
1.3 Conclusion 37

2 Mise en évidence des configurations intrajournalières **39**

2.1 Revue de la littérature 40
2.2 Méthodologie d'analyse 50
2.3 Groupe HL 53
2.4 Influence du groupe de liquidité 64
2.5 Conclusion 68

3 Propriétés des séries intrajournalières **70**

3.1 Propriétés et méthodologies d'analyse 71
3.2 Groupe HL 77

3.3 Influence du groupe de liquidité 85
3.4 Conclusion 87

Conclusion de la partie I **91**

II Analyse "trade by trade" : Le temps retrouvé **94**

Introduction **95**

4 Construction d'approximations de l'information **97**
4.1 Modélisations de microstructure 98
4.2 Pré-traitements 112
4.3 Description des données 131
4.4 Conclusion 136

5 Relations entre les composantes du processus de transactions **139**
5.1 Méthodologies d'analyse des rendements ultra haute fréquence 141
5.2 Modèles univariés 145
5.3 Modèles bivariés 154
5.4 Modèle trivarié 162
5.5 Conclusion 171

6 Incorporation du carnet dans l'analyse du processus de transactions **174**
6.1 Agrégation de l'information du carnet 177
6.2 Modélisation univariée 183
6.3 Modélisation multivariée 196
6.4 Conclusion 205

Conclusion de la partie II **208**

Conclusion générale **212**

Annexes **217**

A La Base De Marché **217**

B Base de données **221**
B.1 Titres sélectionnables 221
B.2 Influence des retraitements 221

C Prix et calculs **225**
C.1 Comparaison des prix 225
C.2 Comparaison des modalités de calcul 225

D Configurations intrajournalières **229**
D.1 Groupe HL 229
D.2 Groupe ML 232
D.3 Groupe LL 233

E Propriétés des séries **234**
E.1 Dynamique du rendement 234
E.2 Distribution des rendements LL 235

F Dynamique du processus de transactions **237**
F.1 Facteur saisonniers 237
F.2 Modélisations univariées 238
F.3 Modélisations multivariées 240
F.4 Fonction de réponse 243

G Relations carnet - transactions **247**
G.1 Statistiques descriptives du carnet 247
G.2 Relations entre les dimensions 248
G.3 Modélisations GARCH des variables informationnelles 252

Liste des tableaux

1.1 Titres sélectionnés 18
1.2 Longueurs prises en compte 29
1.3 Activité mesurée par les transactions 30
1.4 Activité mesurée par les ordres 31

2.1 Configurations intrajournalières 42
2.2 Configurations de rendement groupe HL
1 heure, *temps horloge* 55
2.3 Configurations de volatilité groupe HL
1 heure, *temps horloge* 56
2.4 Configurations de rendement par horaire de cotation
groupe HL, 1 heure 57
2.5 Configurations de volatilité par horaire de cotation
groupe HL, 1 heure 58
2.6 Configurations de rendement par microstructure
groupe HL, 1 heure 59
2.7 Configurations de volatilité par microstructure
groupe HL, 1 heure 61
2.8 Configurations de volatilité groupe ML
30 minutes, *temps nb. transactions* 67
2.9 Configurations de volatilité groupe LL
30 minutes, *temps nb. transactions* 67

3.1 Persistance de la volatilité : Sanofi, 5 minutes 78
3.2 Décomposition du volume 79
3.3 Absence de significativité du volume moyen : groupe HL 80
3.4 Distribution des rendements : groupe HL 5 minutes 81
3.5 Dynamique des rendements : groupe HL 5 minutes 85
3.6 Persistance de la volatilité : Club Med. 15 minutes 86

3.7 Absence de significativité du volume moyen 86
3.8 Distribution des rendements : groupe ML 15 minutes 87
3.9 Dynamique des rendements : groupe ML 15 minutes 89

4.1 Nombre de transactions par seconde 113
4.2 Relations volatilité - duration - volume : Sanofi 115
4.3 État du carnet de Sanofi
4 janvier 1999 10 :10 :36 118
4.4 Mesures agrégées du carnet de Sanofi 119
4.5 Dimensions de l'information de Sanofi 125
4.6 Rendements observés et désaisonnalisés 134
4.7 Rendements observés et désaisonnalisés
normalisation par la duration 135
4.8 Rendements observés et désaisonnalisés
normalisation par le volume 136
4.9 Durations observées et désaisonnalisées 137
4.10 Volumes observés et désaisonnalisés 138

5.1 Implications des différents modèles 139
5.2 Modèles GARCH 147
5.3 Modèles ACD 150
5.4 Modèles ACV 153
5.5 Influence de la duration sur la volatilité 157
5.6 Influence du volume sur la volatilité 161
5.7 Test de variables omises 164
5.8 Effets des valeurs passées sur les valeurs attendues 166
5.9 Effets des valeurs passées sur les valeurs prédites 169

6.1 Composantes principales des dimensions : Sanofi 178
6.2 Axes factoriels de l'hétérogénéité 179
6.3 Axes factoriels de la direction 180
6.4 Axes factoriels de la profondeur 182
6.5 Comparaison des états du carnet avant transaction 184
6.6 Relations duration - carnet 186
6.7 Relations volume - carnet 186
6.8 Relations rendement - carnet 188
6.9 Relations volatilité - carnet 188
6.10 GARCH et variables informationnelles 192

6.11 ACD et variables informationnelles 193
6.12 ACV et variables informationnelles 195

A.1 La base BDM 217
A.2 BDM1D2 218
A.3 BDCCD2 218
A.4 BDESD2 219
A.5 BDM8D2 220

B.1 Titres sélectionnables 222
B.2 Influence des retraitements 223

D.1 Configurations de rendement groupe HL
15 minutes, *temps nb. transactions* 231
D.2 Configurations de volatilité groupe HL
15 minutes, *temps volume* 231

E.1 Dynamique des rendements EGARCH(1,1)
groupe HL, 5 minutes 234

F.1 UHF-GARCH(1,1) : duration 239
F.2 UHF-GARCH(1,1) : volume 239
F.3 Matrice de transition des types de transactions : Sanofi 241
F.4 Modèle ACD(1,1) trivarié 242
F.5 Modèle ACV(1,1) trivarié 244
F.6 Modèle GARCH(1,1) trivarié 245

G.1 Statistiques descriptives de la dimension hétérogénéité 249
G.2 Statistiques descriptives de la dimension direction 249
G.3 Statistiques descriptives de la dimension profondeur 251
G.4 Hétérogénéité prix 254
G.5 Hétérogénéité activité 254
G.6 Direction prix 255
G.7 Direction activité 255
G.8 Profondeur prix 256
G.9 Profondeur activité 256

Table des figures

1.1 Évolution des indices pendant la période 32
1.2 Effets des déformations temporelles : Sanofi 5 minutes 34
1.3 Effets des déformations temporelles : Club Méditerranée, 15 minutes 36
1.4 Effets des déformations temporelles : Assystem, 30 minutes 36

2.1 Configurations en *temps calendaire* groupe HL, 15 minutes 54
2.2 Configurations intrajournalières groupe HL, 15 minutes 63
2.3 Groupes de liquidité *temps horloge*, 30 minutes 64
2.4 Groupes de liquidité *temps nombre de transactions*, 30 minutes 66
2.5 Groupes de liquidité *temps volume d'ordres*, 30 minutes 66

3.1 Distribution des rendements *temps calendaire*, groupe HL, 5 minutes 82
3.2 Distribution des rendements *temps nb. transactions*, groupe HL, 5 minutes 82
3.3 Autocorrélation de la volatilité groupe HL, 5 minutes 83
3.4 Distribution des rendements *temps horloge*, groupe ML, 15 minutes 88
3.5 Distribution des rendements *temps nb. transactions*, groupe ML, 15 minutes 88

4.1 Distribution des volumes, des nombres et des prix : Sanofi 117

4.2 Distribution des volumes, des nombres et des fourchettes agrégés
Sanofi 119
4.3 Dimensions de l'information : Sanofi 124
4.4 Facteur saisonniers empiriques des transactions : Sanofi 128
4.5 Facteurs saisonniers théoriques des transactions : Sanofi 129
4.6 Facteurs saisonniers empiriques du carnet : Sanofi 130
4.7 Facteurs saisonniers théoriques du carnet : Sanofi 132
4.8 Évolution des titres 133

5.1 Duration surprise : Sanofi 151
5.2 Volume surprise : Sanofi 154
5.3 Relations volatilité conditionnelle - duration : Sanofi 158
5.4 Fonction de réponse : Sanofi 170

6.1 Effets sur l'hétérogénéité : Sanofi 198
6.2 Effets sur la direction : Sanofi 198
6.3 Effets sur la profondeur : Sanofi 199
6.4 Effets sur la duration : Sanofi 201
6.5 Effets sur le volume : Sanofi 201
6.6 Effets sur la volatilité : Sanofi 202

C.1 Comparaison des prix : indice global, 30 minutes
prix fin d'intervalle 226
C.2 Modalités de calcul : indice global, 30 minutes
prix de transaction 227

D.1 Configurations en *temps calendaire*
groupe HL, 5 minutes 229
D.2 Configurations intrajournalières
groupe HL, 5 minutes 230
D.3 Configurations intrajournalières
groupe ML, 15 minutes 232
D.4 Configurations intrajournalières
groupe LL, 1 heure 233

E.1 Distribution des rendements
temps horloge, groupe LL, 30 minutes 236
E.2 Distribution des rendements
temps transaction, groupe LL, 30 minutes 236

F.1 Facteurs saisonniers sans ouverture : Sanofi 237
F.2 Fonction de réponse : St-Gobain 246

G.1 Relations entre les deux premiers axes factoriels : Sanofi 248
G.2 Relations entre les dimensions, 1^{er} axe factoriel : Sanofi 250
G.3 Relations entre les dimensions, 2^{nd} axe factoriel : Sanofi 252

Résumé

L'explosion de la volatilité des marchés financiers pose de manière évidente la question d'un lien à l'information. Cette étude apporte des éclairages empiriques sur la relation entre l'information et le processus d'échange. Nous abordons cette question sous deux angles complémentaires. Nous comparons d'abord plusieurs approximations du flux informationnel. Ensuite, nous analysons les relations entre les différentes composantes du processus d'échange (transactions et carnet d'ordres). La comparaison des différentes approximations de l'information suggère que les meilleures sont par ordre décroissant : - le nombre de transactions ; - le volume de transactions ; - le nombre d'ordres ; - le volume d'ordres. La progressivité de ces études du marché français autorise la comparaison des résultats des différentes méthodologies. Nous avons observé que l'agrégation des données par intervalle et la non prise en compte d'une partie de l'ensemble informationnel conduit à une vision partielle et biaisée des relations. L'organisation du marché parisien, dirigé par les ordres, confère une place centrale au carnet d'ordres. Il ressort que les agents peuvent se limiter à l'observation de celui-ci puisque les caractéristiques des transactions (volume, duration) n'ont plus d'influence significative lorsque les effets des variables du carnet d'ordres sont inclus. Les relations mises en évidence entre les variables soulignent que la principale force est la fourniture de liquidité. Les périodes de forte (faible) activité semblent associées à l'absence (la présence) d'informés.

Mots clefs : Euronext Paris, microstructure des marchés, flux informationnel, haute fréquence, ultra haute fréquence, approximation de l'information, configurations intrajournalières, asymétrie d'information, hétérogénéité, volume, duration.

Introduction générale

> " - Mais voyons mon petit, s'écriait-il, tu sais pourtant aussi bien que moi que la Bourse est absolument un jeu et que le Rio bouge parce qu'il y a une grande opération politique en cours ou parce que quelqu'un a raconté en déjeunant chez Gallopin ou à l'Omnium une histoire inepte sur la hausse des métaux, et que les rentes fichent le camp parce que le premier imbécile venu est plein de secrets sur la chute imminente du ministère ! Tout le marché repose sur des racontars de concierge, comment veux-tu qu'il y ait une science de la Conciergerie ?"
>
> Paul Nizan, *La conspiration*, 1938, Gallimard Collection Folio p. 139

Pour aussi polémique qu'elle semble, cette citation n'en souligne pas moins les principales forces à l'œuvre dans les marchés financiers. Ainsi, on voit que les attentes (les espérances) des agents, quelles soient ou non fondées, influencent le marché. L'information a une place centrale et il peut être difficile de faire le tri entre des racontars (des rumeurs non fondées) et une vraie information. Toutefois, de là à considérer que la bourse est *"absolument un jeu"*, il y a un pas qu'il nous semble abusif de franchir.

Des racontars de concierge aux forums financiers sur internet il ne semble y avoir qu'un saut dans le temps. Pourtant, l'étude d'Antweiler & Frank (2001) prouve que les messages sur ces forums peuvent avoir un contenu informationnel supérieur à celui du Wall Street Journal. On peut alors se demander ce qu'est vraiment l'information et comment l'approcher ou l'estimer. Berry & Howe (1994) étudient les informations diffusées par Reuters (i.e. informations publiques). Ils mettent en évidence une configuration en U des arrivées dans la journée mais ne parviennent pas dégager de lien significatif entre celles-ci et les configurations de volatilité. Il est donc difficile de mettre en évidence une relation significative du flux d'information vers les rendements. Confronté à cette difficulté, on peut prendre le problème à l'envers et tenter de lier un rendement exceptionnel à une information. C'est l'optique choisie par Fair (2002) qui tente de lier les rendements exceptionnels du contrat à terme sur indice Standard & Poor's 500 à des informations. Parmi les 69 événements exceptionnels détectés, seuls 53 sont reliés à des informations identifiables. Il est donc aussi difficile d'associer une information identifiable (e.g. une annonce économique) à un rendement que de trouver quelle information a motivé un rendement donné. Cette difficulté nous semble liée au fait que le flux informationnel arrivant sur le marché provient du traitement par les agents des événements. Or, nous ne savons que peu de choses de ce traitement. Par exemple, l'interprétation d'une nouvelle identique peut varier en fonction d'autres paramètres, un agent peut choisir le moment où il révèle son information (ou son interprétation)... Confrontés à l'impossibilité d'observer le flux informationnel arrivant sur le marché, nous avons choisi d'essayer de comprendre comment, à partir des données de marché observables, on peut inférer la présence d'information.

D'un point de vue plus économique, cette thèse renvoie à la fonction des marchés financiers d'où elle tire ses justifications. L'une des principales missions assignées aux marchés financiers est de permettre une allocation efficiente des capitaux. Pour atteindre ce but le marché doit parvenir à une efficience informationnelle c'est à dire que les prix observés soient en adéquation avec la valeur du titre conditionnellement à l'information diffusée. L'analyse du fonctionnement informationnel du marché pose cependant de nombreux problèmes puisque, les paradigmes qui définissent l'efficience informationnelle d'un marché (dont le concept de rationalité) sont remis en question et difficiles à valider empiriquement.
L'économie classique commence par supposer une structure objective de connaissance commune (tout le monde la connaît et chacun sait que l'autre la connaît). Cette connaissance commune permet alors de déduire l'ensemble des résultats walrassiens qui se sont révélés si fertiles dans le développement de la finance moderne. Ce cadre conceptuel permet une évaluation simple des mécanismes d'incorporation de l'information en raison de la référence à une valeur vraie connaissance commune. Toutefois, il a été depuis longtemps prouvé que la disparition de cette dernière (cf. Akerlof (1970)) remet en cause beaucoup de résultats walrassiens dont tout particulièrement l'existence d'équilibre. Les contradictions empiriques au cadre de modélisation classique ont stimulé l'émergence d'une vaste littérature les étudiant et proposant de nouvelles modélisations ou de nouveaux paradigmes.

Deux grands champs de recherche ont investi les frontières de la rationalité. La finance comportementale réintroduit le facteur humain en finance. Cette voie de recherche propose d'expliquer certaines des "anomalies boursières" par des facteurs d'origine psychologique. On peut regretter que cette voie manque souvent d'un cadre unificateur. De plus, Brav & Heaton (2002) démontrent que, même si l'hypothèse comportementaliste et celle d'existence d'une incertitude structurelle rationnelle relâchent des hypothèses opposées du paradigme de l'espérance rationnelle, elles conduisent toutefois à des implications opérationnelles proches, ce qui les rend ardues à distinguer d'un point de vue empirique.
La finance cognitive (se reporter, par exemple , à Orléan (2002)) tente de repenser le paradigme de base de la finance en prenant en compte le fait que la liquidité des titres rend indispensable la prise en compte de l'opinion d'autrui puisqu'elle affecte la valeur potentielle de revente. La nécessité de cette prise en compte conduit à une redéfinition de la rationalité qui devient auto-référentielle : une action non rationnelle (si on considère les fondamentaux) peut être rationnelle en raison des conditions de marché. Cette nature auto-référentielle peut induire des déviations entre la valeur vraie et la valeur de marché (e.g. si tous pensent que le prix est surévalué mais que les autres investisseurs vont continuer à acheter). En effet, il a été démontré que la cohabitation d'investisseurs rationnels et irrationnels[1] pouvaient conduire à des situations où l'arbitrage

[1]Une telle modélisation pourrait sembler artificielle puisqu'on pourrait croire que le marché implémente une sélection naturelle des investisseurs qui conduit à la disparition des irrationnels. Toutefois, Biais & Shadur (2000)

rationnel ne permet pas de réajuster la valeur fondamentale et le prix[2]. Ce type de comportement est mis en avant par de nombreuses études expérimentales. Le rôle des marchés financiers apparaît alors être la découverte d'une valorisation commune aux différents investisseurs (et fruit de leurs interactions). La différence entre le prix observé et la valeur fondamentale est alors le coût de ce service.

La problématique de cette thèse consiste à comprendre comment l'observation du marché permet d'inférer la présence d'information. Une telle connaissance peut aider un agent à choisir quand et comment traiter en vue de réaliser des profits ou d'éviter des pertes. Nous reproduirons donc le comportement d'un investisseur qui, observant le marché, tente de comprendre si les évolutions qu'il observe sont de nature informationnelle. Il va donc de soi que nous n'emploierons dans cette recherche que des données de marché. Comme le souligne le contenu informationnel des forums financiers, les développements technologiques ont eu des répercussions sur les marchés. Ces effets sont visibles à plusieurs niveaux que ce soit l'organisation des marchés (automatisation), la diffusion d'information plus rapide et complète qu'auparavant ou encore une capacité d'intervention en temps réel des agents même s'ils ne sont pas sur les marchés. Suite à cette accélération des échanges immatériels, il nous semble désormais indéniable que l'échelle de grandeur des phénomènes informationnels que nous tenterons de mettre en lumière est intrajournalière. Nous réalisons donc une étude microstructurelle afin d'essayer de comprendre les relations entre le flux informationnel intrajournalier et les rendements. Pour répondre à cette question, on peut adopter deux optiques proches mais néanmoins différentes.
Ainsi, on peut se demander quelle est la meilleure approximation du flux informationnel. Une telle analyse que nous qualifions de statique, étudie la capacité de plusieurs approximations à se conformer aux relations théoriques que l'on peut attendre entre les rendements et l'information. Une réponse à cette question éclaire l'objectif que nous avons défini. En effet, si un investisseur connaissait une bonne approximation de l'information, il serait à même de dire si oui ou non une variation de prix est d'origine informationnelle.
On peut également étudier les relations entre les différentes composantes du processus d'échange. Cette approche, que nous qualifions de dynamique, analyse comment les différentes variables du processus d'échange (caractéristiques des transactions, états du carnet d'ordres) interagissent. La compréhension des relations entre ces variables permettrait à un investisseur d'évaluer quand les évolutions qu'il observe traduisent la présence d'information ou, au contraire, son absence.

Nous pensons que cette thèse a des intérêts aussi bien théoriques que méthodologiques et empiriques. D'un point de vue théorique, cette thèse propose une synthèse d'une littérature récente, pour partie non encore publiée et non encore structurée en courants. De plus, les résultats

et Mailath & Sandroni (2003), entre autres, démontrent que des investisseurs irrationnels peuvent survivre.

[2]Le lecteur pourra trouver dans le chapitre 6 de Brunnermeir (2001) une revue théorique de ce type de phénomènes.

auxquels nous aboutissons permettent de valider ou d'invalider certaines des hypothèses couramment utilisées dans des modélisations de microstructure. Par exemple, l'hypothèse selon laquelle l'arrivée des agents est exogène aux conditions de marché est clairement démentie par nos résultats. Elle constitue donc un guide pour qui voudrait modéliser l'influence de l'information sur les différentes caractéristiques des échanges.
Ce travail présente également des intérêts méthodologiques. Le traitement de 4 années de données boursières intrajournalières sur 21 titres représente une base de données importante (plusieurs millions d'observations) puisqu'elle a nécessité un traitement exhaustif de toutes les transactions survenues pendant cette période. Afin de garantir la pertinence de ces analyses et leur adéquation avec l'activité des titres, différentes longueurs ont été utilisées, jusqu'à 5 minutes pour les titres du CAC 40. Notre choix d'utiliser des modèles économétriques afin d'analyser les différentes composantes du processus d'échange nous a également permis d'appliquer des méthodes de référence largement utilisées dans la littérature (GMM, GARCH, ACP, VAR...). L'ensemble des traitements a été réalisé avec le logiciel SAS^{TM} dont nous avons acquis une connaissance intime. Confronté à la masse de résultats obtenus pour les différents titres, nous avons choisi de représenter une grande partie des résultats sous forme graphique afin d'améliorer la lisibilité autant que la fluidité des analyses.
Les principaux intérêts de cette recherche nous semblent d'ordre empirique. Cette analyse porte sur le marché français avec des données récentes et pendant une période mouvementée (1997-2000). L'analyse d'un marché gouverné par les ordres permet d'enrichir la connaissance, encore trop rare, sur cette structure de marché. Il autorise en outre la comparaison avec les résultats déjà obtenus sur des marchés dirigés par les prix (ou des structures mixtes). Cette comparaison est d'autant plus fructueuse que les mêmes méthodologies y ont déjà été appliquées. De plus, la constitution d'une telle base de données ouvre de nombreuses possibilités de recherches ultérieures.

Cette thèse débouche sur plusieurs conclusions importantes. La principale conclusion est que le choix de l'ensemble informationnel étudié est critique. Ainsi, le contenu informationnel des caractéristiques des transactions (duration, volume) disparaît lorsque l'on considère l'information apportée par le carnet d'ordres. Les caractéristiques des transactions sont donc des approximations de l'information contenue dans le carnet mais n'ont pas de contenu informationnel propre. Une analyse portant exclusivement sur les transactions conduit à des conclusions partielles et même partiales puisque la décision de négliger des variables décisives (le carnet d'ordres) conduit à attribuer aux caractéristiques des transactions le contenu informationnel pouvant provenir d'autres variables (du carnet d'ordres). L'utilisation d'approximations peut être un choix pertinent. Toutefois, ce travail montre qu'il réduit la portée des résultats et ne permet pas de saisir l'ensemble des mécanismes informationnels à l'œuvre dans le marché.
Le second apport de cette thèse porte sur les méthodologies d'étude des données intrajour-

nalières. Deux grandes catégories d'études cohabitent : - L'analyse de données agrégées par intervalle (données haute fréquence) ; - L'analyse de données désagrégées et donc irrégulièrement espacées (données ultra haute fréquence). Les résultats présentés dans cette thèse plaident pour l'utilisation de cette seconde méthode. Il est vrai que cette méthode d'analyse requiert des méthodologies économétriques spécifiques et conduit souvent à des résultats plus complexes à analyser mais elle est mieux adaptée à la nature des données, elle permet d'étudier plus simplement l'influence du temps entre les observations et est plus robuste. À titre d'exemple, le choix d'une longueur d'intervalle pour l'agrégation n'est pas neutre puisqu'il affecte la significativité des variables. Une analyse haute fréquence doit donc contrôler la robustesse des résultats au choix de cette longueur.
Un troisième apport souligne la différence entre les mécanismes informationnels observés sur des marchés dirigés par les prix (principalement le NYSE) et ceux que nous mettons en évidence sur Euronext Paris. Les dynamiques univariées des différents processus (duration, volume, volatilité) sont globalement similaires toutefois, les relations ente ces divers processus sont différentes. Sur le NYSE la majorité des études s'accorde sur l'association entre les périodes de forte activité (en nombre et en volume de transactions) et les fortes volatilités. Nos résultats soulignent que ce sont les périodes de faible activité (en nombre) qui sont associées à de fortes volatilités. Sur le NYSE les faibles durations sont attribuées à la présence d'information alors que, sur Euronext Paris, ce sont les fortes durations qui traduisent la présence d'information .
Un dernier apport concerne les différentes approximations du flux informationnel. Parmi les quatre approximations étudiées (nombre de transactions, volume de transactions, nombre d'ordres introduits, volumes d'ordres introduits), les différents critères employés suggèrent le classement suivant : - Nombre de transactions ; - Volume de transactions ; - Nombre d'ordres introduits ; - Volume d'ordres introduits.

Comme tout travail, cette thèse n'est pas sans avoir des limites dont nous sommes conscients. Notre choix de faire des analyses exploratoires limite, bien évidemment, la portée théorique de cette recherche. Ainsi, ce travail ne proposera pas de modèle d'équilibre général ou partiel. Plutôt que de définir des catégories d'agents (informés, non-informés...), ce travail s'inscrit dans l'optique d'un continuum d'agents et d'information où une telle discrimination est artificielle. Notre méthodologie repose sur l'application de plusieurs modèles économétriques déjà employés et reconnus dans la littérature. Nous ne proposons ni nouveaux modèles, ni amélioration économétrique, notre intérêt se situant plus sur la mise en regard des différents résultats obtenus et leur analyse du point de vue de la microstructure. Ce choix d'employer des modèles économétriques afin de faire ressortir les modalités d'incorporation de l'information aurait pu être confronté, avec profit, à l'avis des praticiens. L'utilisation d'un questionnaire ou d'entretiens aurait par exemple pu nous permettre de confronter nos résultats et l'avis des agents qui font le marché. Toutefois, cette approche nous semble encore prématurée. Certaines des questions aux-

quelles cette recherche conduit concernant les stratégies de placement des ordres doivent être résolues avant que les bénéfices d'une telle confrontation soient substantiels. Notre décision d'analyser des données de marché intrajournalières implique également des limites. Ainsi, nous avons été obligés d'éliminer les fixages étant donné que ce mécanisme d'échange diffère intrinsèquement des transactions en continu et ne peut être étudié avec la même méthodologie. Cette thèse débouche sur des questions concernant les stratégies de soumission des ordres. Toutefois, nous n'avons pas poursuivi cette voie de recherche étant donné que l'indisponibilité de données précises sur les ordres empêchait l'analyse de ceux-ci. En outre, le volume des données et les contraintes informatiques nous ont obligés à limiter le nombre de titres étudiés et/ou la longueur de la période. Enfin, il aurait été très intéressant de dégager un indicateur informationnel de marché mais notre approche désagrégée (titre par titre) ne le permet pas.

L'organisation de cette recherche reflète une question méthodologique lors de l'analyse des données intrajournalières. Si on peut dire que l'ordre de grandeur intrajournalier des phénomènes informationnels rencontre un certain consensus dans la littérature, la nature exacte de la méthodologie à employer demeure sujette à discussions. Deux écoles s'affrontent autour de la question "Faut-il agréger les données dans des intervalles ?". En effet, les transactions (ou les ordres) sont irrégulièrement répartis. Or, les outils statistiques usuels ne permettent d'étudier que des données espacées régulièrement. Donc, pour analyser des phénomènes intrajournaliers on a le choix entre agréger les données dans des intervalles (données Haute Fréquence) et utiliser des outils standards, ou analyser l'ensemble exhaustif des réalisations (données Ultra Haute Fréquence) avec des outils spécialement conçus pour les données espacées irrégulièrement. Chacune de ces méthodes a des avantages et des inconvénients qui rendent le choix cornélien. Le choix d'une méthode ne relève pas uniquement d'un problème statistique puisque les hypothèses sur les processus étudiés se traduisent également en terme financiers. Afin d'évaluer la pertinence de ces deux méthodologies, nous les emploierons successivement. L'organisation de notre travail repose donc sur une unité autant des données que des méthodes employées. Les deux parties emploient également une organisation similaire. À chaque fois le premier chapitre présente la base de données étudiée, la littérature et les retraitements à effectuer en accord avec le type de données (Haute Fréquence versus Ultra Haute Fréquence). Les deux autres chapitres qui composent chacune des parties exploitent ensuite les données ainsi préparées en réalisant différentes études empiriques.

La première partie emploie des données Hautes Fréquence et essaie de comparer différentes approximations du flux informationnel. Cette comparaison des approximations repose sur l'idée supportant la modélisation de la MDH qui suppose que le vrai temps de fonctionnement des marchés n'est pas le temps horloge mais un temps informationnel. Dans un temps informationnel les données sont en accord avec les prédictions théoriques usuelles (e.g. distribution normale) alors qu'observées en temps horloge elle présentent des caractéristiques dues aux

variations du flux informationnel (e.g. excès de kurtosis). Cette remise en cause du temps est faite en employant la méthodologie des déformations temporelles où le temps n'est plus mesuré en nombre de secondes mais selon une autre variable (e.g. le volume). Le chapitre 1 présente ce courant de la littérature financière et la méthodologie de construction des différentes déformations temporelles à Haute Fréquence ainsi que leurs conséquences temporelles sur les séries. Nous étudions ensuite comment cette distorsion du temps affecte les propriétés des séries. Deux points sont successivement abordés. Dans le chapitre 2, nous vérifions l'impact des déformations temporelles sur les configurations intrajournalières. Si nos hypothèses sont vérifiées et que l'approximation est adéquate on devrait observer une disparition des configurations. Le chapitre 3 se focalise sur la distribution des séries de rendements et étudie comment celle-ci change en fonction de la déformation temporelle. De manière similaire au chapitre 2, on attend, si nos hypothèses et nos approximations sont pertinentes, une distribution normale des rendements.

La seconde partie s'inscrit dans une progression de la précision des données employées. Plutôt que d'essayer de trouver une approximation de la quantité d'information arrivant sur le marché cette partie essaie de comprendre comment les variables du processus d'échange interagissent. Le temps est alors une des caractéristiques des transactions au même titre que le volume, le prix… Nous employons ici des données Ultra Haute Fréquence, soit l'ensemble exhaustif des données intrajournalières, afin d'étudier comment les différentes caractéristiques du processus d'échange peuvent permettre de mieux comprendre ce dernier. Le chapitre 4 revoit la littérature microstructurelle pouvant nous guider, les retraitements spécifiques à ce type de données et leurs effets. Nous analysons ensuite dans le chapitre 5 les relations entre les 3 composantes du processus de transactions (la duration, le volume et le rendement). Cette analyse a une vocation pédagogique. En effet, de manière évidente une quantité importante d'information est négligée (i.e. l'information contenue dans le carnet d'ordres) mais elle permet d'introduire la modélisation qui sera employée par la suite et nous donne une première image qui, bien que partielle, nous sera utile par la suite à titre de comparaison. Le chapitre 6 étend la modélisation employée précédemment en incluant des variables du carnet d'ordres. Cette dernière modélisation propose donc une image nettement plus complète puisque l'ensemble informationnel considéré est beaucoup plus proche de celui dont dispose l'investisseur. Ce dernier chapitre permet de comprendre l'influence que peut avoir le carnet sur les caractéristiques des transactions et dans quelle mesure l'ensemble informationnel considéré affecte notre perception des phénomènes.

Première partie

Déformations temporelles des rendements intrajournaliers : À la recherche du temps perdu

Introduction

Bachelier (1900) a initié la modélisation des cours boursiers en développant l'hypothèse selon laquelle la distribution des rendements était normale. On considérait que les rendements suivaient une marche aléatoire dont les incréments sont indépendants et identiquement distribués selon une loi normale. Cette hypothèse s'est révélée très utile puisqu'elle a permis de construire tout un corpus de valorisation comme le CAPM, le modèle de Black & Scholes et bien d'autres encore. Elle s'est, cependant, révélée inadaptée puisque les propriétés empiriques des séries la contredisent (se référer à Mandelbrot & Taylor (1967) pour une des premières constatations). Par exemple, la volatilité des rendements n'évolue pas selon la racine du temps[3].

De tels phénomènes ont conduit à envisager d'autres distributions pour les rendements. Par exemple, Mandelbrot & Taylor (1967) proposent d'utiliser des distributions de Pareto. Toutefois, même si ces distributions sont plus proches des données empiriques, elles manquent d'un support théorique. Clark (1973) propose un modèle de Mixture of Distributions Hypothesis (MDH) qui fournit un cadre théorique expliquant l'origine de certains phénomènes et permet de rapprocher les prévisions théoriques et les phénomènes observés.

Selon lui, l'échelle d'évolution des rendements n'est pas le temps calendaire tel qu'on l'observe mais un temps informationnel inobservable. Pour être plus précis le changement de prix mesuré entre deux observations ($P_t - P_{t-1}$) n'est pas une fonction du temps les séparant (δt) mais fonction de l'information diffusée dans cet intervalle ($\delta I(t)$) où, $I(t)$ désigne la quantité d'information arrivée sur le marché jusqu'à un instant t (processus stochastique croissant au sens large et à valeur dans $\mathbb{R}^+$)[4]. Le processus $P(I(t))$ est subordonné au processus $P(t)$ et $I(t)$ (le flux informationnel) est alors le processus directeur. La distribution de $\Delta P(I(t))$ est donc subordonnée à celle de $\Delta P(t)$[5]. Cette construction théorique lui permet de mettre en évidence la forte influence du flux informationnel sur la variance du processus subordonné. Cette dernière est la moyenne (des incréments) du flux informationnel $\times$ la variance de $\Delta P(t)$.

[3] Si l'hypothèse Markow-Gaussienne était vérifiée alors la volatilité annuelle devrait être la volatilité journalière $\times\sqrt{250}$

[4] Tout au long de cette introduction I_t conservera cette signification que nous nous abstiendrons donc de rappeler. Par abus de notation, nous écrirons par la suite $|I_t$ pour signifier conditionnellement à ce que la quantité d'information arrivée sur le marché jusqu'au temps t soit I_t.

[5] De manière évidente $P(I(t))$ représente le processus de prix observable sur le marché. $\Delta P(t)$ est l'effet (non observable) d'un choc unitaire sur le prix et $I(t)$ mesure la vitesse d'évolution du prix.

Cette modélisation incite donc à remettre en cause le rôle du temps. En effet, elle démontre que les propriétés des rendements observés en *temps calendaire* sont compatibles avec l'hypothèse selon laquelle les rendements sont dirigés par l'information et que l'intensité de cette dernière varie dans le temps. Si les séries de rendements pouvaient être observées en *temps informationnel*, elles seraient distribuées normalement. En *temps calendaire*, ces séries présentent, entre autres, un excès de kurtosis puisqu'elles résultent alors de la subordination des rendements.

Comme noté par Ross (1989), cette représentation est liée au concept de déformation temporelle puisque la variance évolue dans un temps événementiel (l'arrivée d'informations) plutôt dans un *temps calendaire*[6]. On doit remarquer que, dans ce modèle, l'information n'intervient que comme un nombre de signaux dont les caractéristiques sont sans importance. Qui plus est, la distribution de I_t reste grandement dans l'ombre les caractéristiques principales mises en avant étant l'existence d'une moyenne et l'indépendance avec le processus de rendements. Pour dériver la distribution des séries observées, on somme les variations associées à chaque arrivée d'information. Définir la distribution de cette somme n'est pas trivial puisque les variations du nombre de sommations (i.e. le changement du nombre d'arrivées d'information) rendent impossible l'application du théorème central limite standard. La dérivation de la distribution limite est possible sous des hypothèses faibles concernant le processus à agréger mais des conditions de régularité concernant le nombre d'arrivées d'informations doivent également être satisfaites. Aussi, à notre connaissance, cette modélisation n'est appliquée qu'au niveau journalier.

Il existe deux méthodes pour tester cette hypothèse. On peut, à l'instar de Clark (1973), choisir une approximation du flux informationnel de manière plus ou moins arbitraire comme le volume[7] et vérifier si les séries dans cette échelle de temps déformé se conforment aux prévisions théoriques. On peut également considérer que cette modélisation s'applique à un autre processus dépendant lui aussi du flux informationnel comme le volume. On obtient alors un système d'équations simultanées (pour le rendement et le volume) qu'il est possible d'estimer.

Epps & Epps (1976) emploient le mélange de distribution dans un cadre d'anticipations rationnelles bruitées de type Grossman & Stiglitz (1980). Ils supposent une dépendance fonctionnelle entre le volume et le changement de valorisation. Or, l'existence d'une relation exacte entre la valeur absolue des changements de prix et le volume est peu vraisemblable. Tauchen & Pitts (1983) développent également un modèle de MDH dans le même cadre. Contrairement au précédent modèle, le rendement et le volume sont stochastiquement indépendants mais leur re-

[6]Goodhart & O'Hara (1997) donnent une présentation intuitive de la MDH et des déformations temporelles, une introduction plus pointue est fournie par Stock (1988). Ghysels *et al.* (1995) offrent un exemple d'implémentation et Le Fol & Mercier (1998) comparent les performances de plusieurs déformations temporelles. Gouriéroux & Jasiak (2001, p. 377-385) fournissent également une présentation synthétique des déformations temporelles.

[7]Dans cette étude, Clark (1973) utilise le volume comme une approximation du flux informationnel : **"Trading volume is taken as an instrument for true operational time, or an "imperfect clock" measuring the speed of evolution of the price change process."**.

lation commune au flux informationnel induit une dépendance fonctionnelle de leurs moments. Sous hypothèse de constance du nombre d'investisseurs, ils obtiennent le modèle (1).

$$\begin{aligned} \Delta P_t | I_t &\sim N(0, \sigma_1^2 I_t) \\ V_t | I_t &\sim N(\mu_2 I_t, \sigma_2^2 I_t) \\ Cov(\Delta P_t^2, V_t) &\sim \sigma_1^2 \mu_2 Var(I_t) \end{aligned} \tag{1}$$

L'expression de la covariance entre la volatilité (le carré des rendements) et le volume souligne que la relation observée est directement fonction de la variance du nombre d'arrivées d'information. Si le flux informationnel est constant alors la corrélation entre le volume et la volatilité serait nulle. De plus, le nombre d'investisseurs influence cette relation. En effet, le volume dans ce modèle est fonction du désaccord entre les investisseurs. Toutes choses égales par ailleurs, il devrait donc croître avec le nombre d'investisseurs. A contrario, le changement de prix est fonction de la partie commune aux investisseurs[8] et devrait donc tendre vers une asymptote lorsque le nombre d'investisseurs augmente. La croissance du nombre d'investisseurs doit donc entraîner une modification de la relation. Harris (1986*a*, 1987) définit un modèle légèrement différent également utilisé par Richardson & Smtih (1994). Il relâche l'hypothèse selon laquelle le rendement moyen d'une information est nul et obtient alors le modèle (2).

$$\begin{aligned} \Delta P_t | I_t &\sim N(\mu_1 I_t, \sigma_1^2 I_t) \\ V_t | I_t &\sim N(\mu_2 I_t, \sigma_2^2 I_t) \\ Cov(\Delta P_t^2, V_t) &\sim \sigma_1^2 \mu_2 Var(I_t) \end{aligned} \tag{2}$$

La moyenne des rendements est donc proportionnelle au nombre d'arrivées d'information. Il est également possible de construire une modélisation de MDH en employant un modèle microstructurel. Andersen (1996) exploite le modèle d'asymétrie d'information de Glosten & Milgrom (1985) pour définir le modèle (3).

$$\begin{aligned} \Delta P_t | I_t &\sim N(0, \sigma_1^2 I_t) \\ V_t | I_t &\sim P(\mu_{NI} + \mu_{II} II_t) \\ Cov(\Delta P_t^2, V_t) &\sim \sigma_1^2 \mu_{II} II Var(I_t) \end{aligned} \tag{3}$$

Dans ce modèle, II est le nombre d'investisseurs informés, μ_{II} est le volume moyen de leur transactions alors que μ_{NI} est le volume moyen des non-informés et P la loi de Poisson. Il est supposé que les paramètres non indexés par t sont constants dans le temps (notamment le

[8] Plus précisément, le changement de prix est égal à la révision de prix commune à tous les investisseurs plus la moyenne des révisions idiosynchrastiques de chaque investisseur.

nombre d'informés, leur nombre de transactions ou encore le volume des transactions non informées). Les auteurs construisent également un modèle dynamique en spécifiant la dynamique du nombre d'arrivée d'informations. Lisenfeld (2001) étend le modèle de Tauchen & Pitts (1983)[9] en autorisant une variation temporelle de la sensibilité des investisseurs aux informations. Cette modélisation est donc, elle aussi, dynamique.

On doit distinguer deux grands types de tests. Le premier va s'attacher principalement à vérifier que les distributions empiriques sont qualitativement compatibles avec les propriétés théoriques. On peut par exemple tester que la distribution inconditionnelle des rendements est letptokurtique ou encore qu'il existe une corrélation positive entre la volatilité et le volume. Ces tests ne permettent pas de valider une modélisation de MDH. Ils vérifient uniquement que ce type de modélisation est cohérent avec les données.

Le second type de test exploite un modèle afin de former un système d'équations portant sur les moments et les moments croisés du rendement et du volume. Les moments croisés permettant d'obtenir un système sur-identifié (le nombre de conditions est supérieur au nombre de paramètres inconnus) il est alors avantageux d'utiliser la Méthode des Moments Généralisés (GMM) (Hansen, 1982). En effet, elle utilise les conditions de sur-identification afin de définir une pondération optimale des différentes contraintes, permet de tester la validité de ces dernières et, enfin, requiert moins d'hypothèses distributionnelles que les autres méthodes. Ce type de test permet de valider une modélisation puisqu'il vérifie que les valeurs estimées des paramètres sont celles prévues.

De manière générale, les tests qualitatifs (Harris, 1986*a*, 1987) sont très concluants[10] alors que les conclusions des tests "exacts" (Richardson & Smtih, 1994; Andersen, 1996; Lisenfeld, 2001) sont plus mitigées. Les tests de Richardson & Smtih (1994) ne valident leur modèle de MDH que pour 10 firmes (parmi les 30 du Dow Jones qu'ils étudient). Selon leurs résultats, la distribution du nombre d'informations journalier est vraisemblablement lognormale[11], asymétrique à droite, leptokurtique et a de faibles variations autour de sa moyenne. Andersen (1996) rapporte que son modèle microstructurel de MDH est nettement plus performant que le modèle de Harris (1986*a*, 1987) ou encore Richardson & Smtih (1994). Pour les 5 entreprises du New York Stock Exchange (NYSE) qu'il étudie, le modèle dans sa version statique est accepté alors que le modèle "standard" de MDH est systématiquement rejeté. Le modèle dynamique est encore plus performant que le modèle statique et apporte une importante réduction de la

[9] Il opère également une modification mineure en incluant un terme constant dans la moyenne du volume. Cet ajout a été fait en raison de la significativité du terme μ_{NI} dans le modèle de Andersen (1996).

[10] Un autre test proposé par Lisenfeld (2002) est moins concluant. Suivant l'hypothèse de la MDH les mémoires longues de volatilité et du volume proviendraient de la mémoire longue du flux informationnel. Il vérifie bien que ces deux séries ont des mémoires longues comparables, cependant, son test rejette l'hypothèse selon laquelle ces deux mémoires longues proviendraient d'une dépendance commune à un même processus (le flux informationnel).

[11] Les distributions rejetées sont : gamma inversée, Poisson et lognormale ainsi que les distributions symétriques.

persistance de la volatilité. Le modèle de Lisenfeld (2001) incluant deux composantes inobservables (le processus informationnel et la sensibilité aux informations) est évalué par maximum de vraisemblance simulé sur IBM et Kodak. Les estimations conduisent à accepter cette modélisation[12], elles mettent en évidence que le volume et la volatilité ne répondent pas de la même manière aux paramètres et que cette réponse varie pour le court et le long terme. La volatilité à court terme est principalement gouvernée par l'arrivée d'information alors qu'à long terme elle est dominée par la sensibilité aux nouvelles. À l'opposé, la sensibilité n'est pas pertinente pour le volume qui n'est influencé que par l'arrivée d'information.

On peut donc conclure que la MDH permet de décrire le comportement des séries empiriques[13]. Le flux informationnel est donc une donnée pertinente dont il faut tenir compte. Cependant, étant donné que nous travaillerons sur des données intrajournalières, il nous semble impossible d'utiliser cette modélisation car, nous pensons que les conditions portant sur le nombre d'arrivées d'informations ne sont pas remplies avec des intervalles intrajournaliers.

Aussi, nous utiliserons l'outil des déformations temporelles. L'exposé qui précède avait pour but d'expliquer l'idée supportant la MDH puisque les déformations temporelles reposent également sur celle-ci. Ainsi, une déformation temporelle remet, elle aussi, en cause le temps : on considère que le temps calendaire n'est pas pertinent pour mesurer le temps d'évolution du marché et on lui préfère une autre variable. Les données ne sont donc pas échantillonnées toutes les x secondes mais, par exemple, à chaque fois qu'un volume x a été échangé[14]. Cette optique est proche de celle proposée par Clark (1973). Cet outil de modélisation est plus souple que la MDH et a déjà été à plusieurs reprises utilisé sur des données intrajournalières par exemple par Müller *et al.* (1995).

Cette partie sera organisée comme suit. Le chapitre 1 décrira la construction de l'échantillon et des déformations temporelles que nous étudierons dans cette partie. Les deux chapitres suivants étudieront les conséquences des déformations temporelles sous deux angles complémentaires. Le chapitre 2 analysera leurs effets sur les configurations intrajournalières. Enfin, nous examinerons dans le chapitre 3 leurs impacts sur les distributions des séries.

[12] L'hypothèse d'une variation temporelle de la sensibilité des investisseurs à l'arrivée d'information est donc empiriquement validée. Cela conduit donc au rejet du modèle de base (n'incluant pas une telle variation) initialement proposé par Tauchen & Pitts (1983).

[13] On doit cependant remarquer qu'elle ne parvient pas à prendre compte toutes les caractéristiques des séries. Par exemple, Gallant *et al.* (1992), Gallant *et al.* (1993), Tauchen *et al.* (1996), Qi (2001) montrent que la relation entre le volume et la volatilité est asymétrique ce qui est incompatible avec les modèles présentés.

[14] On peut remarquer qu'une étude "trade by trade" est implicitement une déformation temporelle puisqu'on utilise toutes les transactions, la série est échantillonnée à chaque transaction et non selon une grille temporelle fixe.

Chapitre 1

Conception des séries de rendements et des déformations temporelles

La Société des Bourses Françaises (SBF) diffuse chaque mois sur CD-ROM une base de données dénommée Base de Marché (BDM) (décrite dans l'annexe A) qui contient la presque totalité des événements boursiers du mois et que nous nous exploitons dans cette partie. Ce chapitre détaille d'abord la sélection de notre échantillon à partir de cette base de données puis la conception des différentes déformations temporelles. Nous commencerons par énoncer et motiver nos choix méthodologiques dans la section 1.1. Ensuite, la section 1.2 analysera d'une part l'activité des titres afin de garantir la pertinence des analyses ultérieures et d'autre part les effets des déformations temporelles sur ceux-ci pour être mieux à même de comprendre les résultats subséquents. Enfin, la section 1.3 conclura.

1.1 Aspects Méthodologiques

La prochaine sous-section présentera et analysera les modalités de sélection et de retraitement. Ensuite, nous exposerons les modalités de construction des séries temporelles dans la sous-section 1.1.2.

1.1.1 Sélection - Retraitements

Le prochain point détaille et motive les critères de sélection que nous avons utilisés et l'échantillon auquel nous aboutissons. Les retraitements étant en grande partie guidés par la microstructure du marché étudié, elle est l'objet du point b.. Enfin, le point c. expose les retraitements effectués et analyse leurs effets.

a. Sélection

Nous avons choisi de n'étudier que des actions ordinaires. Notre but est d'apprécier la pertinence d'approximations du flux informationnel fondées sur l'activité du titre. Il est donc indispensable de choisir les titres de manière à minimiser les variations non informationnelles de l'activité qui pourraient obscurcir cette relation. Il a été démontré, entre autres par Tkac (1999), Lo & Wang (2000) sur les marchés américains, ou encore Deininger *et al.* (2000) sur le marché allemand que l'entrée ou la sortie d'un indice induit une modification non informationnelle du volume de transactions[1]. En effet, l'existence d'indices crée une segmentation du marché ce qui implique alors que les modifications d'appartenance indicielle (l'entrée et la sortie) entraîne des transactions non informationnelles dont la motivation est de rebalancer le portefeuille. Nous ne considérons donc que des titres n'ayant pas changé d'indice au cours de la période d'étude.

Il existe une relation entre le flux informationnel et l'activité des titres. Par exemple, les résultats de Easley *et al.* (1996) sur les marchés américains ou encore ceux de Brockman & Chung (2000) sur la bourse de Hong-Kong (marché dirigé par les ordres) mettent en évidence que le rôle des informations privées est d'autant plus important que le titre est moins actif[2]. Il semble donc que la quantité ou la proportion d'asymétrie d'information soit une fonction décroissante de l'activité. Les résultats de Spierdijk (2002) mettent en avant une différence plus profonde entre les titres peu traités et les titres actifs. Ils comparent à l'aide d'un modèle Vector AutoRegressive (VAR) similaire à celui d'Hasbrouck (1991) la relation entre l'activité des titres et l'impact prix des transactions d'IBM et d'autres titres moins liquides. Leurs résultats empiriques soulignent non seulement que la différence d'impact prix entre les périodes calmes et actives est plus importante pour les titres peu traités mais aussi qu'on observe souvent, pour ces titres, un effet de sur-réaction. Les relations entre l'activité et le flux informationnel sont donc profondes et affectent également le mécanisme d'intégration de l'information[3]. La SBF déterminant l'appartenance indicielle en fonction de l'activité des titres[4], il nous semble intéressant de choisir des titres appartenant aux 3 grands indices du marché parisien (le Cotation Assistée en Continu 40 (CAC 40), le Société des Bourses Françaises 120 (SBF 120) et le Société des Bourses Françaises 250 (SBF 250)). Ces indices étant emboîtés (les titres du CAC 40

[1] L'entrée ou la sortie d'un indice induit également de nombreux autres phénomènes notamment sur les rendements, sur le nombre d'analystes suivant la valeur (Davydoff, 1998, p. 83-88)...

[2] Des résultats convergents sont obtenus par nombre d'autres études. Par exemple, Hasbrouck (1991) observe que la persistance des innovations de prix est plus forte pour les titres moins actifs et attribue ce résultat à une plus forte asymétrie d'information pour ces titres. Pour sa part, l'analyse de Kavajecz (1999) suggère que les titres les moins traités ont soit des coûts de transactions supérieurs soit une plus forte asymétrie d'information.

[3] Manganelli (2002) met également en évidence une spécificité des titres actifs sur le NYSE. Les durations (l'occurrence des transactions) ne sont pas exogènes : elles dépendent du volume et de la volatilité passée.

[4] Pour de plus amples détails sur les indices de la bourse de Paris, le lecteur est invité à se reporter au document intitulé "Les Indices de la Bourse de Paris" diffusé par la SBF en 1998 et disponible sur son site internet : `http://www.bourse-de-paris.fr/fr/tools0/pdf/market/b/indices.pdf`. Pour une présentation plus générale des indices bousiers, le lecteur pourra se référer à Davydoff (1998).

appartiennent au SBF 120 lui même inclus dans le SBF 250) nous désignerons désormais, pour plus de clarté, par HL les titres qui appartiennent exclusivement au CAC 40, par ML les valeurs membres du SBF 120 mais pas du CAC 40 et par LL les titres n'appartenant qu'au SBF 250.

Spierdijk *et al.* (2002), en étudiant l'activité de 5 titres de la grande distribution traités sur le NYSE, soulignent que l'activité réagit certes à des nouvelles spécifiques à l'entreprise mais aussi à des nouvelles sectorielles (composante commune des variations). On doit donc considérer que le flux informationnel peut changer en fonction du secteur économique des entreprises. Les causes potentielles de telles différences sont multiples. On pourrait avancer une moins grande lisibilité d'un secteur économique, une différence des intervenants selon le secteur ou encore, la structure concurrentielle du secteur... Nous considérons donc qu'un titre n'est sélectionnable que s'il appartient à un secteur[5] dont on retrouve des membres dans les autres indices.

Nous étudions la période maximale à notre disposition (au commencement de cette étude) c'est à dire : du 1er janvier 1997 au 31 décembre 2000. Ce choix offre la possibilité d'évaluer l'influence des nombreux changements de microstructure opérés pendant cette période (se reporter à la page 19 pour une description de ceux-ci).

Pour résumer, nous considérons comme éligibles les titres faisant partie du même groupe (HL, ML ou LL pendant toute la période) et appartenant à un secteur économique qui présente des membres dans les trois groupes. Ces critères fournissent une première liste de titres sélectionnables repris en annexe dans la table B.1.

Afin de préserver une certaine uniformité de l'échantillon, nous sélectionnons le même nombre de titres pour chaque combinaison (secteur économique, groupe de liquidité). Étant donné le nombre de titres sélectionnables, on ne peut avoir qu'un titre par sous-groupe. À chaque fois qu'un choix se présentait, nous avons privilégié la stabilité de l'activité journalière des titres et un volume journalier important (sur base des historiques journaliers téléchargeables sur le site de la Bourse de Paris[6]). La table 1.1 donne les titres qui résultent de ces critères de sélection.

b. Organisation du marché parisien

Une journée à la Bourse de Paris est organisée en quatre phases successives :

i.) *La pré-ouverture.* Le début des échanges est précédé d'une pré-ouverture. Pendant cette période, les ordres peuvent être introduits et annulés gratuitement mais aucune transaction n'est générée, seul le prix théorique est diffusé.

[5] L'examen des secteurs économiques est fait via le code du secteur économique tel que repris dans la table BDESD2 décrite en annexe B.

[6] `http://www.bourse-de-paris.fr/fr/index_fs.htm?nc=2&ni=12&nom=marche`

TABLE 1.1
Titres sélectionnés

Titre	Indice	Secteur
Thomson-CSF	CAC 40	Aéronautique Espace Armement
Zodiac	SBF 120	Aéronautique Espace Armement
Dassault Aviation	SBF 250	Aéronautique Espace Armement
Schneider	CAC 40	Électricité Électronique Télécom
Sagem	SBF 120	Électricité Électronique Télécom
Radiall	SBF 250	Électricité Électronique Télécom
Lyonnaise des Eaux	CAC 40	Environnement Services aux collectivités
Technip	SBF 120	Environnement Services aux collectivités
Assystem	SBF 250	Environnement Services aux collectivités
Accor	CAC 40	Hôtels Restauration Tourisme
Club Méditerranée	SBF 120	Hôtels Restauration Tourisme
Groupe Partouche	SBF 250	Hôtels Restauration Tourisme
Saint-Gobain	CAC 40	Matériaux
Lapeyre	SBF 120	Matériaux
Vicat	SBF 250	Matériaux
Sanofi	CAC 40	Pharmacie Cosmétiques
Clarins	SBF 120	Pharmacie Cosmétiques
Boiron	SBF 250	Pharmacie Cosmétiques
Lagardère	CAC 40	Sociétés Holdings
Eurafrance	SBF 120	Sociétés Holdings
Dynaction	SBF 250	Sociétés Holdings

Cette table donne les titres que nous analysons dans cette partie ainsi que leurs indices et leurs secteurs économiques.

ii.) *L'ouverture.* À l'ouverture, les ordres en carnet sont confrontés dans un fixage pour déterminer le prix d'ouverture. Les transactions possibles à ce prix sont effectuées et les ordres non-satisfaits restent en carnet. Le fixage agrège l'information accumulée pendant toute la période de pré-ouverture, celle-ci est révélée lors de la diffusion du prix d'équilibre et des transactions effectuées.

iii.) *La journée.* Entre l'ouverture et la clôture (ou la pré-clôture) les transactions sont organisées en continu. Les ordres entrent en carnet suivant leur arrivée. Dès que deux ordres peuvent générer une transaction, celle-ci est automatiquement effectuée, enregistrée et diffusée.

iv.) *La clôture.* Pendant la période deux procédures de clôture ont cohabité, les modifications d'organisation ont fait disparaître la première.
- Les transactions s'arrêtent à la fin de la journée. Il n'y a aucun mécanisme de clôture spécifique. Le prix de clôture est simplement le prix de la dernière transaction en continu.
- Le prix de clôture est déterminé par fixage. Ce dernier est précédé d'une pré-clôture de cinq minutes dont le fonctionnement est similaire à la pré-ouverture.

Entre le 1^{er} janvier 1997 et le 31 décembre 2000, l'organisation du marché a changé à plusieurs reprises :

i.) *Le 2 juin 1998*
L'utilisation d'un fixage de clôture[7] a été généralisée à tous les titres (avis et décision n° 98-2079). La pré-clôture a désormais lieu de 17h à 17:05.

ii.) *Le 20 septembre 1999*
Les horaires de cotation ont été allongés : la pré-ouverture a dorénavant lieu de 7:45 à 9h (avis et décision n° 99-3781).

iii.) *Le 20 mars 2000*
La pré-ouverture a été avancée à 7:30 (avis et décision n° 2000-1146).

iv.) *Le 3 avril 2000*
La pré-clôture est retardée à 17:30-17:35 (avis et décision n° 2000-917).

v.) *Le 25 septembre 2000*
Le système de cotation à règlement mensuel disparaît au profit d'un système de règlement livraison $j+3$ (avis et décision n° 2000-3410).

[7] L'introduction du fixage de clôture date de mai 1996 (avis et décision n° 96-1399) pour les titres à liquidité moyenne (continu B, les actions étrangères et les lignes secondaires). Ces titres avaient alors une pré-clôture de 16:55 à 17h.

c. Retraitements

Nous supprimons toutes les données enregistrées en dehors des horaires théoriques d'ouverture. Nous retirons également les applications. Ces transactions pré-arrangées ne sont qu'enregistrées par le marché. Elles risquent donc de bruiter les données[8].

Le marché parisien est ouvert et clôturé (au moins pendant une partie de la période d'étude) par un fixage. Un fixage avec une période de pré-fixage est une procédure de tâtonnement Walrassien qui permet une allocation optimale des titres entre les investisseurs. Il consiste à accumuler pendant une certaine période la demande afin de définir un prix d'équilibre qui maximise le volume échangé et la continuité des prix. Toutes les transactions possibles sont alors effectuées à ce prix. Toute l'information accumulée pendant la période sans transactions est donc révélée en une fois lors de la diffusion du prix d'équilibre et des transactions réalisées. Ce mécanisme diffère intrinsèquement des transactions en continu où l'information n'est pas agrégée mais incorporée au fur et à mesure. Il est impossible d'étudier simultanément deux mécanismes de transactions fondamentalement différents. Aussi, nous supprimons de la base de données tous les enregistrements liés aux fixages.

Dans une étude intrajournalière la pertinence des journées où on observe des transactions que pendant une infime partie de l'ouverture est plus que douteuse[9]. Nous excluons donc les journées dont la durée effective (temps entre la première et la dernière transaction en continu) est inférieure à deux heures.

L'annexe B.2 analyse les effets de ces retraitements sur la base de données en comparant les caractéristiques journalières des séries retraitées et les historiques journaliers téléchargeables sur le site de la Bourse de Paris[10]. L'effet des retraitements sur le rendement ou la volatilité est faible. À l'inverse, la réduction du volume est importante. De plus l'activité est très variable d'un groupe de liquidité à l'autre, suggérant qu'une même longueur d'intervalle n'est pas adaptée à tous les groupes.

Nous opérons également des retraitements sur le fichier des ordres que nous utiliserons par la suite pour définir des déformations temporelles. Ce fichier contient tous les événements concernant les ordres. Un nouvel enregistrement est généré à chaque fois qu'une des caractéristiques d'un ordre est modifiée. Étant donné que nous désirons déterminer si le volume ou le nombre d'ordres introduits est une bonne approximation du flux informationnel, nous ne conservons que les enregistrements traduisant l'introduction d'un d'ordre.

[8] Le retrait des applications est une procédure courante, par exemple Gouriéroux *et al.* (1999*b*) dans une étude du marché parisien font de même.

[9] Il est vrai que l'absence de transaction est une information. Nous devons toutefois remarquer que l'utilisation de telle journée pourrait biaiser les configurations intrajournalières que nous étudierons dans le chapitre 2.

[10] `http://www.bourse-de-paris.fr/fr/index_fs.htm?nc=2&ni=12&nom=marche`

1.1.2 Construction des déformations temporelles

Les données intrajournalières sont par nature irrégulièrement espacées. L'occurrence d'une transaction est un événement dont la date, le prix et le volume sont tous les trois aléatoires. Face à cette incertitude, la pratique usuelle consiste à ne retenir que les transactions suivant une grille temporelle prédéfinie (i.e. à échantillonner les transactions toutes les x minutes). Cependant, si on considère que le flux informationnel est variable dans le temps, cela introduit une distortion. En effet, la quantité d'information diffusée entre deux transactions retenues peut être différente. Or, si on émet l'hypothèse que les caractéristiques des transactions sont fonction de l'information, les séries construites ne reflètent plus alors le processus d'intérêt (e.g. le processus de rendement) mais un mélange de distribution (e.g. du rendement et du flux informationnel). Clark (1973) propose ces hypothèses afin d'expliquer le caractère non Markov-Gaussien des distributions empiriques des rendements. Il introduit en outre la notion de temps opérationnel (que nous nommons *temps informationnel*). On n'échantillonne plus les séries en fonction d'une grille temporelle prédéfinie mais en accord avec une approximation de la quantité d'information diffusée. Le but de cette procédure est de ne retenir que des événements séparés par une quantité x d'information diffusée. Les séries sont alors espacées régulièrement en *temps informationnel*. Par exemple, si on prend comme approximation du flux informationnel le volume de transactions, on échantillonnera alors la série à chaque fois qu'un volume x aura été échangé. On peut appeler cette pratique une déformation temporelle, le temps étant mesuré non en nombre de secondes mais, dans l'exemple précédent, en volume échangé.

Le point a. présente et discute les différentes pistes pour choisir des approximations de l'information. Ensuite, nous expliquons la construction des déformations temporelles dans le point b.. Enfin, le point c. traite du calcul des prix et des rendements.

a. Approximations du flux informationnel

La construction d'une déformation temporelle repose sur une approximation du flux informationnel. Aussi, avant de concevoir des déformations temporelles, nous présentons ici différentes études qui donnent des indications quant aux approximations pertinentes. Nous ne présenterons ici que les études exploratoires qui suggèrent des approximations du flux informationnel. Nous laissons volontairement de côté les études qui lient des approximations du flux informationnel et les configurations intrajournalières (abordées dans la section 2.1), des approximations du flux informationnel et les distributions des rendements (présentées dans la section 3.1) ou encore les modélisations de microstructure (détaillées dans la section 4.1).

De nombreuses études ont mis en évidence que des paramètres de l'activité (e.g. le volume) permettent d'expliquer une part significative de la volatilité. Ce type de test consiste le plus souvent à régresser les rendements ou la volatilité sur des approximations en considérant que la

meilleure approximation a le plus fort pouvoir explicatif. On trouve également des études de la distribution conjointe de deux variables.

La variable dont le contenu informationnel a été le premier reconnu est le volume. Depuis longtemps, l'existence d'une relation positive entre le volume et la volatilité est connue au niveau journalier[11] (cf. Karpoff (1987) pour une revue de la littérature). Le volume a alors été considéré comme une approximation naturelle du flux informationnel. On doit toutefois remarquer qu'à partir du moment où il y a agrégation des données sur un intervalle (par opposition à des données trade by trade) alors le volume a deux composantes : le nombre de transactions $\times$ le volume moyen. Une conclusion globale sur le volume n'indique pas si la relation est due à l'une de ces composantes ou au deux. La décomposition du volume a été longtemps impossible en raison de l'indisponibilité de l'information concernant le nombre de transactions. Ce fait pourrait alors expliquer partiellement la place dominante du volume comme approximation de l'information puisqu'il a longtemps été la seule mesure d'activité disponible.

En analysant 853 actions cotées sur le National Association of Security Dealers (NASDAQ) Jones *et al.* (1994*b*) établissent que le nombre de transactions explique mieux la volatilité des rendements journaliers que ne le fait le volume[12]. Ce résultat est original puisqu'il remet en cause le fait que les volumes dirigent les prix. Chan & Fong (2000) étudient le même thème avec deux échantillons pairés d'actions cotées sur le NYSE et le NASDAQ. Ils démontrent que le nombre de transactions à une influence prépondérante sur la volatilité mais que le volume joue également un rôle dans cette relation[13]. Ils attribuent le résultat de Jones *et al.* (1994*b*) à l'influence non linéaire du volume qui ne peut être prise en compte avec une régression classique. Xu & Wu (1999) confirment également les résultats de Chan & Fong (2000) avec une méthodologie de régression GMM. En effet, suivant leurs résultats, le volume a un contenu informationnel au delà du nombre de transactions. Un résultat similaire est obtenu par Huang & Masulis (2003) sur le London Stock Exchange (LSE). On doit donc s'interroger sur la généralité du phénomène mis en évidence par Jones *et al.* (1994*b*) qui pourrait être dû à un comportement de fractionnement des ordres. De plus, cette relation semble également dépendre de l'ensemble informationnel considéré. En effet, Li & Kryzanowski (1997) rapportent qu'à la bourse de Toronto la relation volume volatilité disparaît lorsque l'on conditionne sur le nombre de transactions et la profondeur.

Kempf & Korn (1999) étudient à l'aide réseaux de neurones la relation entre le contenu in-

[11] Maillet & Michel (1997) et Giot (2000) démontrent que cette relation est également observée sur des données haute fréquence.

[12] Le volume est mesuré par le taille moyenne des transactions du jour mais les résultats sont globalement robustes à l'utilisation d'autres mesures.

[13] Plus spécifiquement, ils obtiennent, tout particulièrement pour le NASDAQ, des résultat compatibles avec l'hypothèse de Barclay & Warner (1993, cf. sous-section 4.1.1) selon laquelle les ordres de taille moyenne ont une influence plus marquée sur la volatilité. Ils mettent aussi en évidence une influence significative des ordres approchés par le déséquilibre entre les transactions acheteuses et vendeuses.

formationnel des transactions (l'impact prix permanent d'une transaction) et plusieurs approximations du flux informationnel. Leur analyse sur les contrats à terme sur l'indice DAX entre septembre 1993 et 1994 met en évidence que la relation est non linéaire, que le flux d'ordres net (approché par le différence en les volumes acheteurs et vendeurs) contient une information et que les gros ordres contiennent en terme absolu plus d'information mais moins en terme relatif[14]. Le flux d'ordres net et le nombre de transactions sont des approximations de qualités équivalentes.

Lin *et al.* (1999) étudient, sur Intel et IBM en, janvier 1994, la performance de trois approximations du flux informationnel (le volume, le nombre de transactions et nombre de changements de prix). Ils concluent que les nombres de transactions et de changements de prix sont, dans cet ordre, les meilleures approximations alors que le volume est qualifié d'inférieur.

D'un point de vue pratique, il semble que les trois éléments que sont le nombre de transactions, le volume de transactions et le flux d'ordres interagissent avec la volatilité et ont donc un contenu informationnel. Tout ceci montre cependant qu'il faut faire preuve de prudence. Nous sommes confronté à un choix. L'étude séparée des différents paramètres peut permettre de mettre en évidence des résultats clairs mais ne fournit qu'une vision partielle et donc partiale. L'étude simultanée, si elle est a priori plus pertinente, pose d'évidents problèmes d'interprétation en raison de la complexité et de la variabilité des relations.

Il semble donc indéniable que le rendement est une fonction de l'information[15] et que les différentes approximations ont un lien statistique certain avec les caractéristiques des transactions. Toutefois, un examen attentif conduit à minorer la portée théorique des précédents résultats.

En effet, une approximation du flux informationnel est une fonction de l'information. On peut alors avoir des relations temporelles variables entre l'information et l'approximation. Par exemple, He & Wang (1995) obtiennent un modèle où le volume peut être autocorrélé même si le flux informationnel ne l'est pas. Si on considère que les relations temporelles entre le flux informationnel et ses diverses manifestations peuvent être différentes alors, sans hypothèses supplémentaires, la mise en évidence d'une relation temporelle entre celles-ci ne suffit pas pour conclure qu'une de ces variables est une approximation du flux informationnel[16].

La relation entre rendement et information est par nature inobservable. Il est donc difficile d'échapper à un test joint : on teste en même temps l'existence d'une relation entre le rendement et l'information et la qualité de l'approximation utilisée. Aussi, les résultats de ce type d'étude ne sont que des indications et non des preuves. Considérons le modèle de rendement suivant :

[14]Cela signifie que les gros ordres ont un contenu informationnel plus important que les petits mais que, rapportés à leurs volumes les petits ordres apportent plus d'information.

[15]Easley *et al.* (2002) mettent en évidence sur un large échantillon de titres cotés sur le NYSE que le risque informationnel est une des composantes qui influence le rendement demandé par les investisseurs.

[16]Par exemple, la mise en évidence d'une relation entre le rendement et le volume ne suffit pas pour conclure que celui-ci est une approximation du flux informationnel.

$R_t = f(I_t) + \varepsilon_{1t}$ où, R_t est le rendement, I_t l'information, ε_{1t} est un bruit blanc et $f(.)$ une fonction quelconque pouvant être non-linéaire, avoir des effets de retard...

De manière générale, une approximation de l'information se définit par sa capacité à améliorer la prévision du processus de rendement (du rendement en lui même ou de la volatilité). À l'extremum toute variable ayant cette caractéristique pourrait être vue comme une approximation du flux informationnel. Une approximation du flux informationnel est donc une variable A_t décrite par : $A_t = g(I_t) + \varepsilon_{2t}$ où, $g(.)$ est une fonction quelconque et ε_{2t} est un bruit blanc tel que : $R_t|A_t = h(I_t|A_t) + \varepsilon_{3t}$.

C'est à dire que le rendement conditionnellement à une approximation de l'information est une fonction de l'information non encore prise en compte et que ε_{3t} soit stochastiquement dominé par $\varepsilon_{1,t}$. Cela signifie que l'utilisation d'une approximation de l'information réduit l'erreur moyenne de prédiction.

La relation entre R_t et A_t est a priori inconnue, elle dépend des fonctions $f(.)$ et $g(.)$ et du processus I_t. Sans hypothèses additionnelles sur ces composantes, il est a priori impossible de spécifier plus avant la relation entre R_t et A_t que ce soit en terme temporel (relation de "lead-lag") ou en terme de distribution.

Les conclusions de ce type de test dépendent donc des méthodologies utilisées. Par exemple un test par régression linéaire pourra échouer à mettre en évidence une relation non-linéaire pourtant avérée. De manière plus générale, un test utilisant une formulation paramétrique dépendra de la validité de celle-ci. Le type de test le plus général est non paramétrique mais on risque alors de manquer les relations les plus ténues (peut-être les plus intéressantes...).

b. Calcul des déformations

Nous considérons 5 hypothèses concernant le flux informationnel :

i.) *Constant dans le temps :* **Temps calendaire**
Cette hypothèse conduit au temps usuel mesuré en minutes et secondes (le temps n'est pas déformé il est considéré tel que donné dans les fichiers). C'est l'hypothèse commune, souvent non explicitée, de toutes les études qui utilisent un intervalle fixe.

ii.) *Proportionnel au nombre de transactions :* **Temps nombre de transactions**
Nous considérons ici une déformation temporelle utilisant le nombre de transactions. La série sera échantillonnée toutes les x transactions. Nous supposons ici que chaque transaction apporte au marché la même quantité d'information quelles que soient ses autres caractéristiques.

iii.) *Proportionnel au volume de transactions :* **Temps volume de transactions**
Nous définissons ici une déformation utilisant le volume de transactions en nombre de titres. Les séries seront échantillonnées à chaque fois qu'un volume x aura été échangé. On

considère que le contenu informationnel d'une transaction est proportionnel à son volume. Ainsi une transaction de 100 titres a un contenu informationnel 10 fois supérieur à une transaction de 10 titres.

iv.) *Proportionnel au nombre d'ordres introduits :* **Temps nombre d'ordres**

v.) *Proportionnel au volume d'ordres introduits :* **Temps volume d'ordres**

Ces deux dernières déformations temporelles sont similaires à celles définies sur les transactions mais utilisent les ordres.

Par la suite, le terme "*temps activité*" désignera l'ensemble des temps construits en utilisant une mesure d'activité (tous les temps à l'exception du *temps calendaire*). De même, les termes "*temps nombre* (*temps volume*)" désigneront les temps utilisant le nombre (volume) de transactions ou le nombre (volume) d'ordres.

Le marché parisien ferme pendant la nuit. Pendant cette période les échanges sont impossibles, les investisseurs ont le temps de synthétiser l'information diffusée pendant la journée. Cette fermeture induit une discontinuité dans le processus informationnel. Aussi, nous pensons qu'il faut respecter cette nature journalière de l'activité. Notre but sera donc de construire des déformations temporelles respectant la structure journalière des données :

Les différents temps que nous définissons n'interviennent que dans le découpage de la journée. Le nombre d'intervalles composés pour un même jour, l'heure de la première et de la dernière transaction sont tous trois identiques quelle que soit la déformation temporelle. Cette dernière n'intervient donc que dans l'échantillonnage des prix à l'intérieur de la journée.

Afin de mesurer le temps en fonction de l'activité, nous divisons l'activité observée jusqu'à la t^{e} transaction $(\sum_{i=1}^{t} A_{id}^{a})$ par le total journalier $(\sum_{i} A_{i,d}^{a})$ puis multiplions ce ratio par la longueur de la journée $(\max_{i}(H_{i,d}) - \min_{i}(H_{i,d}))$[17] enfin, nous alignons tous les horodatages sur l'heure de la première transaction du jour $(\min_{i}(H_{i,d}))$. L'équation (1.1) synthétise la formule de calcul de temps que nous appliquons.

$$\tau_{t,d}^{a} = \min_{i}(H_{i,d}) + \frac{\sum_{i=1}^{t} A_{i,d}^{a}}{\sum_{i} A_{i,d}^{a}} \times (\max_{i}(H_{i,d}) - \min_{i}(H_{i,d})) \tag{1.1}$$

Où, le premier indice $(_t)$ est la position ordinale de la transaction dans la journée alors que le second $(_d)$ est le jour. Enfin l'exposant $(^a)$, lorsqu'il est présent, donne le nom du temps employé. De plus, H est l'heure d'une transaction alors que A^a est sa valeur informationnelle (dans le temps a). Afin de garantir que les horodatages de la première transaction du jour soient identiques pour tous les temps nous avons adapté la méthode de calcul de cette valeur qui est

[17] De manière évidente, $\max_{i}(H_{i,d})$ et $\min_{i}(H_{i,d})$ se réduisent à l'heure de la dernière et de la première transaction le jour d.

donnée par l'équation (1.2).

$$A^a_{t,d} = \begin{cases} 0 & \text{Si } t = 1 \\ \text{Sinon} & \begin{cases} 1 & \text{En temps nombre de transactions} \\ VT_t & \text{En temps volume de transactions} \\ NO_{t' \in [H_{t-1}, H_t]} 1 & \text{En temps nombre d'ordres} \\ \sum_{t' \in [H_{t-1}, H_t]} VO_{t'} & \text{En temps volume d'ordres} \end{cases} \end{cases} \tag{1.2}$$

La seule adaptation est donc de considérer que la première transaction de chaque jour n'a pas de contenu informationnel ce qui contraint leurs horodatages à coïncider avec leurs *temps calendaire*[18]. À l'exception de cette première transaction la valeur informationnelle d'une transaction dans un temps donné est la valeur de la mesure d'activité utilisée pour désigner ce temps.

Nous avons donc défini une méthodologie de construction de déformation temporelle intra-journalière. Les horodatages de la première et de la dernière transaction du jour sont identiques à ceux observés en *temps calendaire* alors qu'entre ces deux bornes le temps est mesuré en fraction de l'activité journalière. Nous déformons donc le temps à l'intérieur de la journée pour qu'une même fraction de l'activité journalière sépare chaque donnée échantillonnée.

c. Approximations des prix

Une fois définies les modalités de calcul des temps et la longueur de l'intervalle (nous traiterons de ce dernier point lors de la sous-section 1.2), il faut déterminer le prix représentatif de l'intervalle. Cette définition s'articule autour de deux points :

i.) Choix du prix. Pour un même actif on peut considérer différents prix : le prix de transaction, le prix milieu de fourchette, le prix acheteur ou vendeur.

ii.) Le traitement optionnel dont ce prix peut faire l'objet :
- De manière classique on utilise le dernier prix de l'intervalle. Celui-ci est alors artificiellement considéré comme simultané à la fin de l'intervalle.
- On peut aussi tenter de prendre en compte l'asynchronisme en approchant le vrai prix par ajustement linéaire comme Harris (1986*b*)[19]. Cet ajustement conduit à répartir le rendement sur les intervalles qui y ont contribué. À la différence de la méthode fin d'intervalle, on peut donc avoir un rendement non nul même si aucune transaction n'est observée dans l'intervalle.

[18] L'alignement des horodatages de la dernière transaction dans les différents temps n'est pas un problème puisque on a toujours $\sum_{i=0}^{t_{Max}} A^a_{i,d} = \sum_i A^a_{i,d}$ d'où un alignement mécanique sur le *temps horloge* des dernières transactions.

[19] Cette pratique est courante dans les études des marchés des changes comme, entre autres, Bollerslev & Domowitz (1993), Andersen & Bollerslev (1997*b*) ou encore Müller *et al.* (1995).

– On peut se demander si le prix obtenu par la méthode précédente est le plus représentatif. L'instabilité du processus ou l'existence de bruits (comme la danse de la fourchette) peuvent rendre celui-ci plus représentatif de choc de courte durée que de réelles modifications de la valeur du titre. Par exemple, Chauveau *et al.* (1999) en étudiant l'existence de dynamiques non-linéaires sur Alcatel et France Telecom démontrent que celles-ci, bien que toujours présentes, sont plus faibles avec un prix moyen qu'avec un prix fin d'intervalle.

Pour déterminer le prix (P_i) représentatif de l'intervalle (i), nous utilisons deux prix : le prix fin d'intervalle et le prix moyen. Des versions synchronisées par ajustement linéaire de ces prix sont également construites. La synchronisation consiste à ajouter la fraction du rendement jusqu'à la transaction suivante attribuable à l'intervalle courant ($((i-t')/(t-t')\times(P_{t>i}-P_{t'\leqslant i})$). Ces divers prix suivent donc l'équation (1.3).

$$P_i = P_{t'\leqslant i} + D_s \times \frac{i-t'}{t-t'} \times (P_{t>i} - P_{t'\leqslant i}) \tag{1.3}$$

Où, t' et t sont les horodatages des transactions entourant la fin de l'intervalle i et D_s une variable muette valant 1 si le prix est synchronisé et 0 sinon.

Enfin, il existe deux méthodes pour calculer les rendements. La formule la plus largement utilisée est l'approximation par les logarithmes : $R_t = \ln(P_t) - \ln(P_{t-1})$. Cette formule a l'avantage de conférer aux rendements une propriété d'additivité. Une alternative peu utilisée[20] est la formule exacte : $R_t = (P_t - P_{t-1})/P_{t-1}$. Cette formule ne connaît pas le biais baissier des logarithmes mais on perd alors la propriété d'additivité des rendements[21]. Afin d'éviter le biais baissier des logarithmes qui pourrait être important sur des données intrajournalières nous utilisons la formule exacte.

Étant donné le faible nombre de titres (21 répartis en 3 groupes de liquidité et 7 secteurs), la constitution de portefeuilles sectoriels ne nous semble pas pertinente (chacun ne contiendrait que 3 titres). Nous ne construisons que des portefeuilles par groupe de liquidité, chaque groupe de liquidité reprend 7 titres. La construction des séries temporelles est faite individuellement pour chaque titre. Ensuite, le rendement d'un groupe est calculé comme la moyenne des rendements de ses composantes pour chaque intervalle (chaque date et intervalle du jour). Enfin, la volatilité est la valeur absolue du rendement ainsi défini[22].

[20] Harris (1986*b*) utilise cette formule.

[21] Pour une discussion des avantages respectifs de ces deux formules le lecteur peut se référer à Campbell *et al.* (1997, p. 9-11) ou Gouriéroux & Jasiak (2001, p. 12)

[22] Il est également possible de calculer la volatilité en prenant le carré des rendements. Toutefois nous préférons ne pas utiliser cette méthode. En effet, Andersen *et al.* (2001) soulignent que l'utilisation du carré rend la volatilité beaucoup plus sensible aux valeurs extrêmes et rend d'autant plus critiques les tests asymptotiques.

1.2 Analyse de l'activité

Comme l'a démontré l'analyse des effets des retraitements, présentée en annexe B.2, l'activité journalière est très variable entre les groupes et même entre les titres d'un même groupe. La pertinence d'une étude intrajournalière est directement liée à cette activité. Une série de rendements à 5 minutes pour un titre n'ayant qu'un prix à l'ouverture et un à la clôture n'aurait aucun sens ! La sous-section 1.2.1 analysera l'activité journalière afin de définir des longueurs pertinentes pour les déformations temporelles intrajournalières. La sous-section 1.2.2 examinera ensuite les effets des déformations temporelles sur les différents titres.

1.2.1 Activité journalière

La table 1.2 donne le temps entre la première et la dernière transaction en continu exprimé en pourcentage de la journée de bourse[23].

La longueur effective de la journée augmente avec la liquidité du groupe considéré. Elle est la plus longue pour le groupe HL (environ 99%) vient ensuite le groupe ML (environ 96%) et, elle est la plus courte pour le groupe LL (environ 83 %). On peut noter une certaine uniformité à l'intérieur des groupes. Cependant, les disparités entre les titres sont d'autant plus importantes que la liquidité du groupe est faible. Deux titres se distinguent nettement en raison de la petitesse de leurs journées : Vicat et Radiall. La perte est inégalement répartie entre l'ouverture et la clôture cette dernière étant toujours beaucoup plus faible[24] (sauf pour Vicat et Radiall). Le fixage semble permettre de satisfaire la demande et il faut alors un certain temps pour qu'elle se reconstitue[25].

L'étude de la longueur effective de la journée souligne un biais de sélection possible lors de l'étude des configurations intrajournalières en *temps calendaire* pour les titres peu liquides (groupes ML et LL). Par exemple, pour le groupe LL, il est courant que la première transaction après ouverture intervienne après un délai d'une vingtaine de minutes (environ 5%). Aussi, l'analyse des effets d'ouverture sur ce groupe avec un intervalle plus court serait très influencée par les quelques jours exceptionnels où la première transaction intervient "rapidement" après le fixage d'ouverture.

[23]Nous avons choisis d'exprimer cette longueur en % de la journée boursière afin de rendre comparables les chiffres obtenus sur une période où la durée de la journée de bourse a changé. À titre de comparaison, 5% représentent 21 minutes avec les horaires de début de période (10h-17h) et 25 minutes 30 secondes avec les horaires de fin de période (9h-17:30).

[24]Le rapport va du simple au décuple. Ce résultat est en contradiction avec celui de Stoll & Whaley (1990) qui rapportent que, sur le NYSE entre 1982 et 1986, le temps entre la dernière transaction et la clôture est supérieur à celui entre l'ouverture et la première transaction.

[25]Cela pourrait également être interprété comme le temps d'incorporation de l'information véhiculée par le fixage d'ouverture.

TABLE 1.2
Longueurs prises en compte

	Longueur de la journée				Ouverture	Clôture
Titre	Q1	μ	σ	Q3	μ	μ
Thomson	98.84	98.78	2.22	99.34	1.10	0.12
Zodiac	96.98	97.42	3.26	99.02	1.77	0.81
Dassault	86.66	89.41	11.64	96.96	5.63	4.96
Schneider	99.02	98.96	2.04	99.40	0.96	0.08
Sagem	95.55	96.28	5.36	99.20	2.84	0.87
Radiall	64.41	76.03	18.55	91.19	12.91	11.06
Lyonnaise	99.08	99.03	1.88	99.43	0.89	0.07
Technip	97.78	97.90	2.85	99.10	1.70	0.39
Assystem	81.53	86.54	13.91	96.31	7.94	5.51
Accor	99.02	99.00	1.87	99.41	0.92	0.08
Club Med.	97.09	97.58	3.11	99.15	1.86	0.56
Partouche	83.86	87.81	12.61	96.68	6.25	5.93
St-Gobain	99.05	99.00	1.87	99.40	0.92	0.08
Lapeyre	93.58	94.99	5.50	98.46	3.30	1.71
Vicat	59.70	71.35	18.70	86.83	13.70	14.95
Sanofi	98.98	98.94	1.91	99.40	0.97	0.08
Clarins	96.94	97.42	3.12	99.01	1.79	0.78
Boiron	83.17	87.94	11.97	96.64	7.47	4.59
Lagardère	98.99	98.91	2.33	99.40	1.00	0.09
Eurafrance	94.19	95.13	5.71	98.43	3.35	1.52
Dynaction	87.81	90.84	9.94	97.50	4.81	4.35

Cette table donne pour les différents titres les paramètres (premier quartile Q1, moyenne μ, écart-type σ et troisième quartile Q3) de la distribution de la longueur effective de la journée (temps entre la première et la dernière transaction en continu) exprimée en % de la journée de bourse. Elle fournit également le temps moyen (dans la même unité) entre l'ouverture (la clôture) et la première (la dernière) transaction.

Pour mieux comprendre quelle longueur pourrait être utilisable dans une déformation temporelle il est également nécessaire d'étudier la répartition des mesures d'activité sur lesquelles les déformations temporelles sont fondées. La table 1.3 donne les principaux paramètres de la distribution de l'activité journalière (nombre et volume de transactions). La table 1.4 donne les mêmes informations concernant les ordres.

TABLE 1.3
Activité mesurée par les transactions

	Nombre de transactions				Volume de transactions			
Titre	Q_1	μ	σ	Q_3	Q_1	μ	σ	Q_3
Thomson	314	465	229	569	95 501	176 258	138 600	215 013
Zodiac	90	157	105	192	2 673	6 311	5 835	7 894
Dassault	17	41	41	50	316	1 389	2 400	1 652
Schneider	599	920	469	1 108	182 957	313 616	188 822	390 277
Sagem	41	205	324	263	942	6 161	15 456	5 675
Radiall	5	12	12	15	176	794	1 521	886
Lyonnaise	861	1 377	774	1 654	209 945	332 490	194 074	399 173
Technip	121	196	115	243	13 428	29 733	23 752	37 564
Assystem	11	22	18	28	637	2 740	5 143	3 111
Accor	624	1 090	671	1 370	84 169	214 930	214 591	267 080
Club Med.	89	173	161	205	12 609	30 749	49 511	35 519
Partouche	14	32	32	39	580	2 074	2 638	2 739
St-Gobain	679	934	400	1 091	102 078	165 246	97 742	196 995
Lapeyre	36	68	60	80	3 306	9 176	11 219	10 698
Vicat	5	12	15	14	162	546	890	625
Sanofi	574	822	338	1 003	150 045	398 025	336 213	573 170
Clarins	89	160	131	182	6 111	13 069	13 683	15 542
Boiron	12	23	17	29	660	2 064	2 475	2 504
Lagardère	507	850	507	1 016	174 099	303 308	212 055	367 098
Eurafrance	39	65	45	79	703	2 184	2 862	2 680
Dynaction	21	48	47	57	1 990	6 932	10 194	7 710

Cette table donne pour chaque titre les paramètres de l'activité journalière des transactions (premier quartile Q_1, moyenne μ, écart-type σ, et troisième quartile Q_3). Le volume de transactions est le nombre d'actions échangées.

L'analyse de la répartition des transactions par titre fait ressortir le même classement des groupes de liquidité que la comparaison de la longueur effective de la journée. La différence entre le groupe HL et les autres est plus tranchée pour le volume que pour le nombre de transactions. Ainsi, pour les groupes de moindre liquidité (ML et LL), on observe moins de transactions et elles sont d'un volume plus faible. La différence de volume moyen par transaction entre ces deux groupes est cependant faible. Ce qui les distingue est principalement le nombre de transactions. On doit à nouveau remarquer qu'il existe des disparités importantes à l'intérieur des groupes. Ainsi, le volume de transactions du titre Dynaction le rapproche nettement du groupe

ML (son volume de transactions est comparable à Zodiac ou Sagem) alors que son nombre de transactions l'identifie clairement comme un titre du groupe LL. Il diffère donc de son groupe de liquidité par des transactions d'un volume moyen important. De manière semblable, Eurafrance se rapproche du groupe LL alors qu'il appartient au groupe ML.

TABLE 1.4
Activité mesurée par les ordres

	Nombre d'ordres				Volume d'ordres			
	Q_1	μ	σ	Q_3	Q_1	μ	σ	Q_3
Thomson	780	1 128	520	1 347	738 914	1 276 231	887 749	1 508 297
Zodiac	199	335	223	405	18 256	49 941	60 623	59 337
Dassault	43	86	71	106	2 627	9 624	13 529	11 514
Schneider	1 367	1 826	656	2 140	1 172 045	1 856 543	1 068 449	2 186 152
Sagem	96	500	781	629	7 123	59 033	143 474	49 290
Radiall	10	26	26	33	1 271	6 105	10 623	7 150
Lyonnaise	1 392	2 261	1 316	2 759	991 865	1 698 867	2 097 023	1 959 131
Technip	271	418	218	515	101 182	213 823	168 491	267 398
Assystem	23	50	40	62	5 240	20 592	28 856	24 582
Accor	1 133	1 861	971	2 335	432 208	1 320 435	3 457 618	1 402 354
Club Med.	199	385	340	465	91 307	233 084	342 114	260 734
Partouche	32	68	63	84	3 381	14 795	18 719	18 212
St-Gobain	1 306	1 788	687	2 104	580 587	955 146	612 665	1 124 546
Lapeyre	84	163	133	200	27 098	73 670	81 598	88 045
Vicat	7	19	23	22	494	2 707	5 678	2 353
Sanofi	1 030	1 684	785	2 168	723 256	2 431 555	2 335 161	3 737 389
Clarins	193	303	178	367	40 171	87 618	85 458	106 233
Boiron	27	50	35	62	4 894	16 353	22 085	18 906
Lagardère	1 052	1 763	1 169	2 001	1 082 662	2 014 175	1 921 504	2 313 023
Eurafrance	90	154	105	181	6 504	19 956	25 836	23 911
Dynaction	49	95	77	117	11 552	43 994	70 086	45 735

Cette table donne pour chaque titre les paramètres de l'activité journalière des ordres (premier quartile Q_1, moyenne μ, écart-type σ, et troisième quartile Q_3). Le volume d'ordres est mesuré en nombre d'actions.

On observe une très grande similarité entre les ordres et les transactions. La disparité entre les groupes de liquidité est plus marquée (nettement plus si on considère le volume). À l'inverse les groupes de liquidité sont plus homogènes exception faite des volumes d'ordres de Dynaction et Eurafrance.

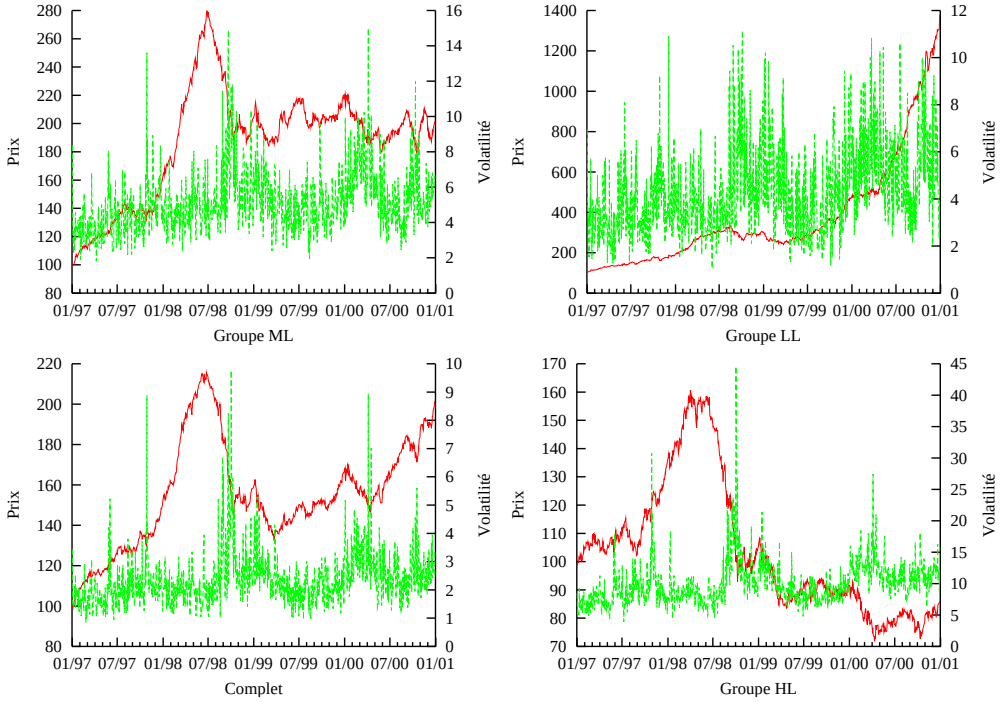
Lorsque l'on examine la distribution de l'activité des titres, il devient évident qu'une même longueur ne peut être adaptée aux différents groupes de liquidité. Le troisième quartile du nombre journalier de transactions du groupe LL se répartit entre 14 pour Vicat et 57 pour Dynaction. À l'évidence, un intervalle de 5 minutes (ou même de 15) serait inadapté pour ce groupe. En effet, elle nécessiterait 12 intervalles par heure d'activité dont une majorité serait alors vide. Cela est d'autant plus vrai que nous n'avons pas pris en compte le clustering des transactions qui augmenterait encore le nombre d'intervalles vides. En raison de ces considérations, nous

décidons de réaliser les analyses avec les longueurs suivantes :

- *Groupe HL :* 5 minutes
- *Groupe ML :* 15 minutes
- *Groupe LL :* 30 minutes

La figure 1.1 représente l'évolution du 1er janvier 1997 au 31 décembre 2000 des groupes en terme de volatilité et de prix (base 100 en début période). Ce prix est la valeur d'un portefeuille qui serait investi chaque matin (à la première transaction en continu) dans les indices de liquidité et serait vendu chaque soir (à la dernière transaction en continu). Il tient donc exclusivement compte des variations observées entre la première et la dernière transaction en continu[26].

FIGURE 1.1
Évolution des indices pendant la période



Cette figure donne pour chaque groupe (en dessous du graphique), l'évolution des prix (courbe rouge axe de gauche) et des volatilités (courbe verte hachurée axe de droite). Le prix, calculé sur une base 100 en début de période, est obtenu en composant les rendements intrajournaliers (par multiplication puisque nous avons utilisé la formule exacte). La volatilité est la somme journalière des valeurs absolues des rendements intrajournaliers. Nous avons à chaque fois employé les longueurs adaptées aux groupes. Le groupe complet agrège tous les titres de l'échantillon, il est calculé avec une heure.
Seules les variations intrajournalières entre la première et la dernière transaction en continu entrent dans le calcul de ces évolutions.

[26]Les variations observées pendant la nuit, entre le fixage d'ouverture et la première transaction en continu ou entre la dernière transaction en continu et le fixage de clôture ne sont donc pas prises en compte. Le prix indiqué est donc différent de celui observable sur le marché.

Les évolutions des groupes sont très différentes. Ainsi, le groupe LL n'a cessé d'augmenter pendant la période, cette augmentation s'accélérant nettement en fin de période. À l'inverse, les groupes ML et HL ont connu une importante augmentation en début de période (jusqu'à mi 1998) avant de connaître une chute prononcée. Si cette chute s'est stabilisée rapidement pour le groupe ML autour d'un niveau de 200, le groupe HL n'a cessé de diminuer par paliers successifs jusqu'à un niveau inférieur de 20% à sa valeur de départ. La volatilité connaît elle aussi des évolutions différentes selon les groupes. On observe pour les groupes HL et ML des pics de volatilité importants autour de certaines périodes de chutes. On voit donc que la période d'étude est loin d'être uniforme dans le temps ou entre les groupes.

Les résultats pour le groupes LL illustrent la spécificité des rendements intrajournaliers pour les titres peu liquides. Le marché ouvre souvent trop bas et augmente trop pendant la journée. Il clôture alors sur une valeur trop forte qui est sur-compensée par une baisse dans la nuit. De manière récurrente les rendements pendant la journée sont donc très positifs alors que les rendements pendant la fermeture sont très négatifs. Pour ce groupe, l'évolution représentée traduit donc plus des coûts de transactions récurrents que la hausse des titres.

1.2.2 Effets des déformations temporelles

Par la suite, nous allons comparer plusieurs caractéristiques des séries construites avec différentes déformations temporelles. Afin de pouvoir comprendre les propriétés de ces séries, il est nécessaire de savoir quels sont les effets temporels des déformations (quelle partie de la journée est compressée, étendue ...). Étant donné qu'un groupe est construit par agrégation des séries définies individuellement sur ses composantes, il n'est pas possible d'estimer l'effet d'une déformation directement sur un groupe. Aussi, les figures 1.2 à 1.4 illustrent sur un titre de chaque groupe l'effet des déformations temporelles.

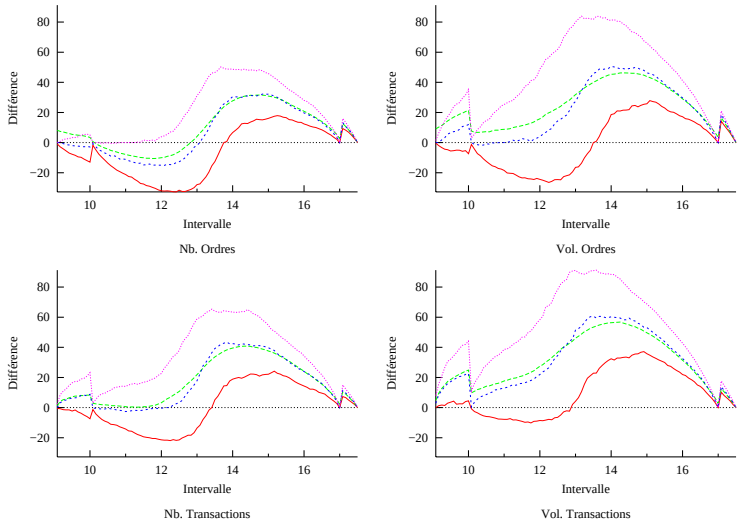
Les courbes représentées exprimant $\delta =$*temps activité* - *temps horloge*, une valeur positive (négative) signifie que le temps déformé est en avance (retard) sur le *temps horloge*. Cela revient à dire que la période précédente a été compressée[27] (étendue) en raison de sa moindre (forte) contribution à l'activité journalière.

Le cas du dernier intervalle du jour est spécifique puisque l'horodatage de la dernière transaction est identique pour tous les temps. Or, jusqu'au 3 avril 2000 17h a été le dernier intervalle (environ 80% de l'échantillon) donc, la majorité des différences de cet intervalle mécaniquement nulle. Cela explique alors la plus faible dispersion de la différence entre un *temps activité* et le *temps horloge* pour cet intervalle. Une constatation similaire peut être faite pour le premier intervalle de la journée (10h dans environ 70% de l'échantillon) dont la dispersion est, elle

[27] Cela signifie que nous avons été obligés d'utiliser une transaction postérieure car l'activité avait jusqu'ici été trop faible.

aussi, mécaniquement limitée.

FIGURE 1.2
Effets des déformations temporelles : Sanofi 5 minutes



Cette figure donne pour Sanofi les principaux paramètres de la distribution par intervalle du jour de la différence d'horodatage entre *le temps déformé* (en dessous des graphiques) et le temps horloge. Cette différence (δ = *temps activité - temps horloge*) est exprimée en minute. La courbe rouge continue est le premier quartile de la distribution, la verte hachurée la moyenne, la bleue hachurée la médiane et la violette pointillée le troisième quartile de la distribution.

Pour Sanofi le fait principal est que, quelle que soit la mesure d'activité, la fin de la journée contribue d'une manière très importante à l'activité journalière. Cette période est toujours étendue en *temps activité*. Le milieu de la journée est toujours compressé. Le fait que cette compression soit plus importante pour les *temps volume* que les *temps nombre* souligne que pendant cette période il a y d'une part, moins de transactions, et qu'elles sont d'autre part d'un volume plus faible.

La différence est plus importante et plus dispersée dans les *temps volume*. Cela montre que le volume d'activité a une plus forte saisonnalité intrajournalière mais aussi que sa répartition dans la journée est plus instable.

L'effet sur l'ouverture est différent selon le temps utilisé :

- Un *temps nombre d'ordres* étend cette période. Cela signifie que, pendant cette période, un nombre important d'ordres est introduit.
- Le *temps nombre de transactions* affecte peu cette période. Le nombre de transactions à l'ouverture est dans la normale.
- Le *temps volume de transactions* et, dans une moindre mesure le *temps volume d'ordres*, compresse cette période. Cela signifie que la fraction du volume journalier observée pen-

dant cette période est inférieure à la normale.

On comprend donc que la journée se divise en 3 périodes aux caractéristiques bien distinctes :

i.) Le matin d'une longueur d'environ 2 heures où on observe un nombre de transactions normal et un nombre d'ordres légèrement supérieur à la normale. Cependant, les volumes de transactions et des ordres sont inférieurs à la normale.

ii.) Le milieu de la journée (12:30-14h) où l'activité est faible particulièrement en volume.

iii.) L'après midi où l'activité augmente jusqu'à la clôture. Ce phénomène est plus important pour les volumes.

Ces comportements sont très robustes puisque le premier quartile, la moyenne, la médiane et le troisième quartile suivent tous la même évolution.

On observe pour Club Med. (membre du groupe ML) des comportements similaires à Sanofi mais moins tranchés. L'accroissement de l'écart inter-quartile reflète la plus grande dispersion de l'activité intrajournalière pour les titres de moindre liquidité ainsi que l'existence de chocs d'activité en proportion plus importants (distance plus importante entre les paramètres centraux et les quartiles de la distribution).

Les résultats sont beaucoup plus préoccupants pour Assystem (membre du groupe LL. Les différents paramètres de la distribution n'ont pas tous la même évolution. Il arrive que la moyenne soit supérieure au troisième quartile, la médiane est le plus souvent nulle et l'écart interquartile l'est aussi parfois. De plus aucune évolution représentative ne se dégage. Les résultats concernant Assystem ne sont pas généralisables à l'ensemble du groupe.

Les effets des déformations temporelles sont si instables pour ce groupe qu'il faut différencier les titres et les temps pour produire une tentative de classification approximative :

- En *temps volume de transactions* tous les titres présentent approximativement la même forme d'évolution : une augmentation plus ou moins erratique, jusqu'à un maximum en début d'après-midi, puis une diminution jusqu'à la clôture.
- En *temps nombre d'ordres* on peut distinguer d'un côté Dynaction, Boiron, Dassault, Partouche et Assystem qui ont des courbes assez semblables à celles du groupe ML. D'un autre côté Vicat et Radial pour lesquels l'ouverture est compressée alors que la clôture est étendue (courbe en cloche). Cela signifie que la majorité des ordres sont introduits à la clôture et qu'ils sont peu nombreux à l'ouverture.
- En *temps nombre de transactions*, on observe approximativement pour tous les titres le même comportement : une absence d'effet d'ouverture, une augmentation en milieu de journée puis une diminution progressive en s'approchant de la clôture.
- Le *temps volume d'ordres* est le plus erratique. Une diminution à la clôture est courante mais de faible amplitude, une augmentation en milieu de journée l'est également. Les

FIGURE 1.3
Effets des déformations temporelles : Club Méditerranée, 15 minutes

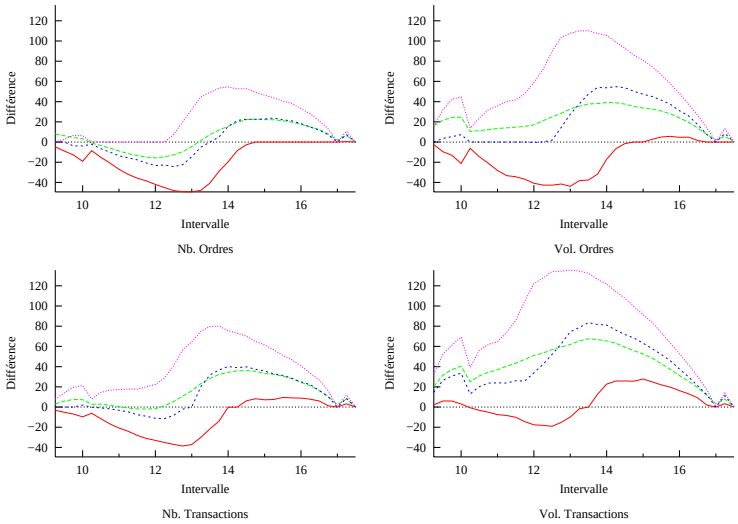
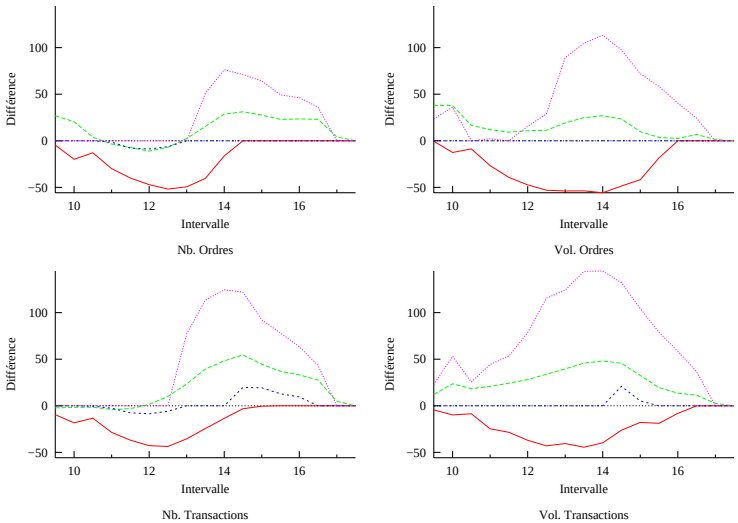


FIGURE 1.4
Effets des déformations temporelles : Assystem, 30 minutes



Ces figures donnent pour Club Med. et Assystem les principaux paramètres de la distribution de la différence d'horodatage entre le temps activité (en dessous des graphiques) et le temps horloge par intervalle du jour. La différence ($\delta =$ *temps activité* - *temps horloge*) est exprimée en minute. La courbe rouge continue est le premier quartile de la distribution, la verte hachurée la moyenne, la bleue hachurée la médiane et la violette pointillée le troisième quartile.

évolutions au cours de la journée ne suivent pas forcément une courbe en cloche.

L'activité des titres de ce groupe est trop erratique et dissemblable pour qu'une conclusion générale sur l'influence temporelle des déformations soit possible.

1.3 Conclusion

Dans ce chapitre, nous avons sélectionné un échantillon de manière à former trois groupes de liquidité composés de titres des mêmes secteurs économiques et membres des trois plus gros indices du marché parisien. Cette méthodologie de sélection devrait nous permettre d'étudier au mieux les effets des déformations temporelles sur ces groupes puisqu'elle minimise a priori les changements d'origine externe (e.g. changement de l'appartenance indicielle, effet propre à un secteur économique...). Nous avons ensuite défini quatre déformations temporelles utilisant des mesures d'activité en nombre et en volume sur les ordres et sur les transactions.

L'analyse de l'activité journalière souligne la forte corrélation entre celle-ci et le groupe de liquidité. Cet effet était attendu puisque la liquidité est un des critères de sélection indicielle. Cependant, les niveaux de liquidité observés pour les différents groupes nous oblige à conclure que la faible liquidité de certains groupes contraints la longueur de l'intervalle minimum pertinent (i.e. la longueur qui n'induise pas trop de biais d'asynchronisme ou d'infrequent trading). Aussi, nous avons choisi de réaliser les analyses avec, au mieux, des longueurs de 5 minutes pour le groupe HL, de 15 minutes pour le groupe ML et de 30 minutes pour le groupe LL.

Nous avons enfin analysé les effets temporels des déformations c'est à dire la différence entre les quatre temps construits pour une même transaction. Cette analyse souligne que les effets varient en fonction de la déformation (i.e. de l'approximation du flux informationnel) et du groupe considéré.

Les effets temporels majeurs des déformations sont d'étendre la fin de la journée et de compresser le milieu de journée. L'effet sur le début de la journée est une légère expansion (contraction) dans les *temps nombre* (*temps volume*).

Plus la liquidité du groupe est faible plus ces effets sont importants. Toutefois, en dessous d'une certaine activité, les effets temporels des déformations deviennent instables. Au vu des résultats, il semble que le groupe LL soit en dessous de cette limite. Il conviendra donc d'analyser avec précaution les résultats de ce groupe.

Enfin, nous avons noté une dispersion importante des effets des déformations temporelles. Ainsi, même si leurs influences suivent globalement les mêmes tendances au cours des différentes journées, il est également manifeste que ces impacts varient selon les jours. On peut alors se poser deux questions :

"Les configurations intrajournalières des déformations temporelles font-elles échos à

celles du flux informationnel. ?[28]

Les déformations temporelles permettent-elles de se rapprocher du processus d'intérêt : le rendement ?"

On observe d'étonnantes similitudes entre les configurations intrajournalières en activité (en volume ou en nombre) et en prix ou volatilité (que nous présenterons dans le prochain chapitre). Les déformations temporelles exploitant des mesures d'activité, on retrouve ces mêmes similitudes entre les effets des déformations temporelles et les configurations intrajournalières en prix ou en volatilité. Il est alors intéressant d'analyser si les configurations intrajournalières sont affectées par les déformations temporelles. Selon l'hypothèse de la MDH, la volatilité évolue en temps informationnel. Donc, si les configurations intrajournalières sont causées par l'information, que l'hypothèse de la MDH est vérifiée et que l'approximation du flux informationnel est pertinente alors, on devrait observer une disparition des configurations intrajournalières de volatilité en *temps déformé*. Le chapitre 2 analysera l'effet des différentes déformations temporelles sur les configurations intrajournalières.

La MDH fait également des prévisions sur la distribution des séries. Les mesures en *temps informationnel* doivent permettre de supprimer, par exemple, l'excès de kurtosis dans les séries étant donné qu'il provient de la subordination du processus de rendement au flux informationnel. Si le temps est déformé selon une approximation pertinente du flux informationnel et que l'hypothèse de la MDH est pertinente alors nous devrions observer des séries en *temps informationnel* sans excès de kurtosis. Le chapitre 3 analysera, entre autres, l'effet des déformations temporelles sur les distributions des rendements.

[28]Cette question proche de celle posée par Lüders & Peisl (2000) les conduit à lier les anomalies boursières et le flux informationnel.

Chapitre 2

Mise en évidence des configurations intrajournalières

Dans ce chapitre, nous évaluons l'influence des déformations temporelles définies dans le précédent chapitre sur les configurations intrajournalières. Pour rappel, nous considérons 5 déformations temporelles :

i.) Le *temps calendaire*[1]

ii.) Le *temps nombre de transactions*

iii.) Le *temps nombre d'ordres* introduits

iv.) Le *temps volume de transactions*

v.) Le *temps volume d'ordres* introduits

Nous but est double. D'une part, nous examinons les causes des configurations intrajournalières. L'existence de changement de microstructure pendant notre étude nous permet d'évaluer certaines des hypothèses avancées pour expliquer les configurations intrajournalières. D'autre part, nous évaluons la pertinence des approximations du flux informationnel. Plus une approximation est pertinente, plus elle doit entraîner une réduction importante des configurations intrajournalières dues à l'information. Nous analysons donc les effets des différentes déformations temporelles sur les configurations intrajournalières.

Ce chapitre sera organisé comme suit. Dans la section 2.1 nous résumerons les configurations mises en évidence dans la littérature et les causes avancées. La section 2.2 exposera la méthodologie d'analyse ainsi que nos hypothèses. Nous analyserons les résultats dans les sections 2.3 pour le groupe HL et 2.4 pour les autres groupes. Enfin, la section 2.5 conclura.

[1]Le temps calendaire n'est pas à proprement parler une déformation mais, par simplicité d'exposition, nous l'assimilons à une déformation identité.

2.1 Revue de la littérature

Toute analyse empirique est indissociable des données étudiées qu'on peut classer en quatre catégories : *i)* les actions ; *ii)* les dérivés ; *iii)* les obligations ; *iv)* les changes.

Les informations pertinentes pour ces différents produits sont potentiellement différentes et les réactions ne sont pas forcément identiques entre les marchés ou les produits. Par exemple, Andersen *et al.* (2000) mettent en évidence une faible influence des annonces macroéconomiques sur les cours du NIKKEI 225 (NIKKEI). Cependant, tous les résultats ne vont pas dans ce sens. Ainsi, Chen *et al.* (1999) démontrent que les variations de taux d'intérêt non attendues ont une influence significative sur les actions américaines. Il y a cependant un large consensus pour reconnaître que les marchés de taux ou de devises répondent fortement et rapidement aux annonces macroéconomiques. Par exemple, Ederington & Lee (1993) soulignent la très forte réaction des marchés de taux (bond ou Foreign eXchange future market) aux annonces macroéconomiques ou encore Goodhart *et al.* (1993) qui soulignent que la prise en compte de ces annonces est cruciale pour obtenir des estimations pertinentes sur les taux de change[2]. Pour une comparaison des configurations de volatilité entre les différents marchés, le lecteur pourra se référer à Andersen *et al.* (2000).

Quand bien même les ensembles pertinents seraient identiques, rien n'indique que leurs effets sur des produits différents soient comparables. De telles variations pourraient être dues aux spécificités des produits, à des différences de comportements des investisseurs[3] ou à des différences de microstructure. L'organisation des marchés fixe les règles d'échange, elle conditionne donc la relation entre l'information et les transactions. Aussi, des différences de microstructure entre les marchés[4] modifient probablement les relations entre information et rendement. À titre d'exemple, Bollerslev & Domowitz (1993) mettent en évidence que, pour les intervenants sur le marché interbancaire dont la présence est continue on observe une fourchette globalement constante dans la journée alors que, pour les intervenants ayant une présence limitée dans le temps la fourchette suit une courbe en U. On peut donc penser qu'il existe un lien entre l'absence de marché (la fermeture ou la présence des intervenants) et l'existence des configurations.

Le flux informationnel doit donc être considéré comme propre à chaque produit. Aussi, les études sur des produits autres que les actions, sujet de cette étude, ne peuvent fournir que des indications générales ou méthodologiques. Par conséquent, nous ne présenterons de manière

[2]Dans leur étude du taux de change Sterling Dollar, l'inclusion des annonces macroéconomiques dans la modélisation permet de passer d'un Generalized Auto Regressive Conditionnal Heteroskedasticity (GARCH) quasi intégré à une version stable.

[3]Les résultats présentés par Berglund & Westerholm (2002) soulignent qu'il est plausible que différents investisseurs (selon la taille de leur portefeuille) aient des réactions différentes aux informations.

[4]On peut songer, par exemple, au fait que les marchés des changes sont interbancaires et ouverts $24h/24h$ alors que les marchés d'actions sont généralement centralisés et à fermeture périodique.

détaillée que les études des actions[5].

La sous-section 2.1.1 fera le point sur les configurations mises en évidence dans la littérature. Nous essayerons de démêler les causes avancées dans la sous-section 2.1.2.

2.1.1 Des configurations convergentes mais différentes

Nous ne traiterons que des configurations observées sur les prix. Des configurations intrajournalières ont également été mises en évidence pour d'autres dimensions du processus d'échange[6].

De manière évidente, l'étude des configurations intrajournalières requiert des données intrajournalières. Le développement de ce courant de recherche a été subordonné à la disponibilité de ces données (à partir du milieu des années 80). La table 2.1 synthétise les données et les résultats des principales études des configurations intrajournalières à notre connaissance.

Les études sur des séries longues et anciennes (Smirlock & Starks, 1986; Jain & Joh, 1988; Gerety & Mulherin, 1994; Lockwood & Linn, 1990) ont utilisé des indices dont les niveaux horaires étaient diffusés dans la presse. Un indice est un panier de valeurs qu'on calcule en employant le dernier prix des composantes. Cette méthode pose deux problèmes. D'une part, en fonction des instants des transactions, on utilise des données plus ou moins récentes et d'autre part, les intervalles pendant lesquels des transactions se produisent ont une probabilité supérieure de générer un rendement non nul puisque le rendement calculé entre deux transactions échantillonnées est attribué uniquement à l'intervalle de la dernière. Cela ne reflète que la concentration des transactions, mais il est problématique de ne pouvoir quantifier cet effet puisqu'on ne dispose que de niveaux agrégés. L'utilisation d'intervalles plus courts, une minute pour Niemeyer & Sandås (1993), est certes un progrès mais pas une solution.

Pour mesurer l'effet de l'agrégation temporelle, il faut disposer de l'heure et du prix de chaque transaction : d'un historique des transactions comme la base BDM diffusée par la SBF ou la base de données Trade and Quotes (TAQ) diffusée par le NYSE. Il faut ensuite définir une méthodologie qui répartisse les rendements sur l'ensemble des intervalles entre les deux transac-

[5]Concernant les études des configurations observées sur les marchés de dérivés sur actions ou indices (sur le marché des changes) le lecteur pourra se référer, à Ho & Lee (1998), Chan *et al.* (1995*a*) (Andersen & Bollerslev (1997*b*) ou Ballochi *et al.* (1997)).

[6]La fourchette et les mesures d'activité ont des configurations majoritairement en U : fourchette [Brock & Kleidon (1992), Chan *et al.* (1995*b*), Chung *et al.* (1999), Hasbrouck (1991), Mc Inish & Wood (1992), Foster & Viswanathan (1993), Lee *et al.* (1993), Ahn & Cheung (1999), Bildik (2000), Declerck (2000*b*), Daníelsson & Payne (2001)] ; le volume [Ahn & Cheung (1999), Lee *et al.* (1993), Bildik (2000)] ; le nombre de transactions [Daníelsson & Payne (2001)] ; l'introduction des ordres [Ahn *et al.* (2001), Chung *et al.* (1999), Biais *et al.* (1995), Bisière & Kamionka (2000)]. À l'inverse, la forme dominante des configurations de profondeur [Ahn *et al.* (2001), Ahn & Cheung (1999), Brockman & Chung (1999), Fung & Patterson (1999), Lee *et al.* (1993), Daníelsson & Payne (2001)] partagé également par la duration [Engle & Russell (1997, 1998), Hafner (1999)] est le U renversé.

Ces différentes configurations sont souvent liées. Par exemple, Lee *et al.* (1993) observent une corrélation positive (négative) entre le spread et le volume de transactions (la profondeur cotée).

TABLE 2.1
Configurations intrajournalières

			Rendement			Volatilité		
Étude	Intervalle	Marché	Forme	Ouverture	Clôture	Forme	Ouverture	Clôture
Wood, Mc Inish & Ord (1985) 946 et 1138 actions, 09\71-02\72 et 1982	1mn	NYSE	J	30mn	1mn-5mn	J	30mn	1mn-5mn
Harris (1986*b*) 1616 actions, 12\81-01\83	15mn	NYSE	J	45mn	15mn	U	30mn	45mn
Smirlock & Starks (1986) Indice DJIA, 01\62-12\83	1h	NYSE	non étudié			non étudié		
Jain & Joh (1988) Actions 78-84	1h	NYSE	?	?	1h	non étudié		
Harris (1989) 1617 actions, 12\81-01\83 et 03\85-04\85	"trade to trade"	NYSE	n.e.	3 first	last	non étudié		
Lockwood & Linn (1990) Indice DJIA, 01\64-02\89	1h	NYSE	non étudié			J	2h	2h
Hamon & Jacquillat (1991*b*) 205 actions, 04\88-12\89	5mn	SBF	J	15 mn	10mn	non étudié		
Foster & Viswanathan (1993) Toutes actions, année 1988	60mn	NYSE	non étudié			∪	1h	1h
Gerety & Mulherin (1994) Indice DJCI et actions, 1952-1992 et 07\89-03\92	1h	NYSE	non étudié			déclin dans la journée		
Niemeyer & Sandås (1993) Indice OMX, 12\91-01\92	1mn	SSE	ø	ø	ø	ø	10mn	ø

L'acronyme n.e. signifie que le point n'est pas étudié, ø indique une absence d'effet et ? traduit notre incapacité à conclure en raison de l'inclusion des rendements entre la clôture et l'ouverture.

tions et non sur le dernier comme avec un indice. Wood *et al.* (1985) sont, à notre connaissance, les premiers à avoir utilisé ce type de données. Leur utilisation nous semble cependant limitative[7] puisqu'ils ne répartissent pas les rendements et ne mettent donc pas à profit la richesse de telles données. Harris (1986*b*) étudie lui aussi ce type de données mais avec une méthodologie innovante puisqu'il répartit les rendements prorata temporis par ajustement linéaire entre les différents intervalles de quinze minutes concernés. Hamon & Jacquillat (1991*b*) appliquent la même méthodologie à la Bourse de Paris avec un intervalle de cinq minutes.

Deux niveaux d'analyse cohabitent : - une optique strictement intrajournalière recherche les variations du processus de rendement en fonction de *l'heure du jour* ; - une autre à mi-chemin entre intrajournalier et interjournalier analyse les différences en fonction de *l'heure du jour de la semaine*. Cette dernière est une des voies à explorer lors de l'interprétation des variations intrajournalières puisqu'elles pourraient n'être dues qu'à un jour dans la semaine.

Certaines études, dont Jain & Joh (1988), intègrent aux rendements intrajournaliers le rendement de la nuit. Cette pratique nous semble abusive puisque le rendement pendant la nuit totalise les variations sur une période de plusieurs heures pendant laquelle les transactions sont impossibles. Il nous semble donc sans commune mesure avec des rendements intrajournaliers. Nous ne tiendrons donc pas compte de ce type de résultats. Les configurations sont souvent représentées par des lettres qui symbolisent la forme des courbes (∪, ∩, J, ɻ). Bien qu'opérationnelle, cette pratique est souvent peu précise[8]. Aussi, nous convenons qu'une courbe est en U si les niveaux matin et soir sont comparables à la différence d'une courbe en J.

Toutes les études à l'exception de Niemeyer & Sandås (1993) mettent en évidence des variations intrajournalières significatives sur les points qu'elles étudient. Ces configurations concernent au moins l'une des deux dimensions du processus (moyenne ou volatilité) et ont été résumées dans la table 2.1.

L'analyse de Gerety & Mulherin (1994) est particulière puisqu'ils calculent la volatilité des rendements journaliers à différentes heures du jour. Leurs résultats ne sont donc pas comparables à ceux strictement intrajournaliers (volatilité d'un intervalle dans la journée). Leur principale conclusion est le déclin de la volatilité au cours de la journée. À l'exception du lundi, la volatilité du premier intervalle est la plus forte de la journée, elle décline ensuite jusqu'à la clôture. Ces résultats sont similaires pour l'indice et les titres individuels bien que, dans ce second cas, les déviations soient plus nombreuses. Pour expliquer cette forme, ils arguent de la résolution progressive de l'asymétrie d'information au cours de la journée.

La majorité des études (voire également les résultats présentés par Xu & Wu (1999) ou encore Bildik (2000)...) penche pour des configurations en J pour les rendements et la volatilité.

[7] Cette remarque est aussi valable pour Gerety & Mulherin (1994).

[8] De nombreuses études qualifient des configurations en U alors que les valeurs à la clôture sont supérieures à celles de l'ouverture. Nous rebaptisons des telles configurations de J.

Elles se caractérisent donc par des niveaux supérieurs, positifs pour les rendements, à la clôture et à l'ouverture. Seules ces deux périodes fournissent des différences significatives. Les phénomènes d'ouverture sont souvent plus longs mais moins intenses que les effets de clôture. Les études ne s'accordent ni sur la longueur ni sur l'amplitude exacte de ces phénomènes.

Concernant l'influence du jour de la semaine sur les configurations intrajournalières, seuls les rendements négatifs le lundi à l'ouverture sont significatifs. Encore une fois, il y a désaccord quant à la longueur du phénomène. Harris (1986*b*) avance une durée de 45 minutes contre 6 heures 15 minutes pour Hamon & Jacquillat (1991*b*). On peut remettre en question la stabilité ou la généralité de ces configurations puisque Smirlock & Starks (1986) font ressortir le rapprochement vers l'ouverture des rendements négatifs du lundi. Ceux-ci étaient observés toute la journée dans les premières sous-périodes de leur échantillon et uniquement lors des premiers intervalles d'ouverture dans les dernières sous-périodes.

Il ne semble pas y avoir d'autres variables d'intervention. Les divers tests de stabilité dont entre autres la fréquence de transaction pour Hamon & Jacquillat (1991*b*), la capitalisation pour Harris (1986*b*), le volume de transactions monétarisé pour Foster & Viswanathan (1993) n'ont pas révélé d'effet notable sur les configurations.

Le plus souvent un test de Fisher est utilisé pour déceler les variations de niveaux entre les intervalles. Il est aussi possible, à l'instar de Werner & Kleidon (1996), d'utiliser une régression pour mettre en évidence les variations. Ces derniers étudient les conséquences de la double cotation[9] sur les configurations. Ils utilisent des régressions avec correction d'hétéroskédasticité et des variables muettes temporelles. La méthode d'estimation peut encore être améliorée en utilisant les GMM comme Xu & Wu (1999) ou, avant eux, Foster & Viswanathan (1993). L'analyse de Werner & Kleidon (1996) a ceci d'intéressant qu'elle met en évidence des configurations qui, même si elles changent d'amplitude, sont de formes similaires que l'action fasse ou non l'objet d'une double cotation. La volatilité des cotations suit une courbe en ʟ, celle des transactions une courbe en J. Cette différence est due selon eux à des transactions pré-arrangées en fin de journée. Ce résultat remet en cause l'intégration des deux marchés. Si les deux marchés étaient intégrés les configurations devraient être réparties non sur l'ouverture d'un seul marché mais sur la réunion des périodes d'ouverture des deux marchés. Chan *et al.* (1996) mettent également en évidence des résultats qui confirment cette hypothèse. Les auteurs mettent en évidence que les configurations d'actions doublement cotées (Amérique - Europe, Amérique - Japon) sont semblables.

En guise de conclusion nous pouvons remarquer que l'existence des configurations n'est plus sujette à controverse mais que leurs amplitudes et leurs causes demeurent objet de discussion. De plus, la majorité des études utilise des tests de Fischer peu adaptés aux caractéristiques

[9]Ils utilisent 13 actions cotées sur le Stock Exchange Automated Quotation (SEAQ) et sur le NYSE via des American Despositary Receipt (ADR).

des données haute fréquence (cf. sous-section 3.1.1). À cet égard, une méthodologie de régression avec des variables muettes (Werner & Kleidon, 1996), (Foster & Viswanathan, 1993) permet facilement de prendre en compte les spécificités de ces données (autocorrélation hétéroskédasticité). Les méthodologies employées par Xu & Wu (1999) et Foster & Viswanathan (1993) nous semblent les plus abouties.

2.1.2 Causes des configurations

De nombreuses causes ont été avancées pour expliquer l'existence des configurations intrajournalières. Les points suivants détaillent celles qui ont rencontré le plus de support. L'énoncé suivant n'est pas exhaustif puisque d'autres causes ont également été avancées mais ne seront que peu ou pas développées. À titre d'exemple, Kramer (2000) relie les configurations intrajournalières et intrahebdomadaires à l'existence de cycles dans l'activité humaine. Le lecteur désireux d'avoir une optique différente de la notre concernant les relations entre la volatilité à l'ouverture et à la clôture peut se référer à Minguet (2003, section 4 p. 278-284).

a. Structure de marché et mécanismes d'échange

Amihud & Mendelson (1987) comparent la volatilité des prix d'ouverture issus de fixages à celle des prix de clôture issus d'un marché continu. À partir des titres du Dow Jones Industrial List (DJIL) de février 1982 à février 1983, ils concluent que le mécanisme de fixage génère plus de volatilité que les transactions continues. L'attribution exclusive de la différence au mécanisme de transaction nous semble excessive. En effet, le moment de la journée change lui aussi, nous pourrions aussi bien attribuer cette différence à l'heure du jour[10]. En comparant une ouverture par fixage et une ouverture en continu sur le marché milanais, Amihud *et al.* (1990) rejettent leur ancienne hypothèse de volatilité excessive induite par le fixage. Ils mettent alors en avant l'existence d'une période de fermeture précédente pour expliquer la plus forte volatilité. Enfin, dans une étude du Tokyo Stock Exchange (Tokyo S.E.)[11], Amihud & Mendelson (1991) démontrent que la volatilité à l'ouverture est liée à la longueur de la fermeture. Hamao & Hasbrouck (1995) corroborent ces résultats[12].

L'hypothèse d'une volatilité induite par le fixage est donc à rejeter. Aujourd'hui le fixage est même considéré comme un mécanisme stabilisant. Ainsi, afin de réduire la volatilité observée

[10] Selon nous, pour qu'une telle conclusion puisse être valide, il faudrait comparer la volatilité d'un prix issu de fixage et de marché continu au même instant de la journée et bien sûr pour le même type de titres. Cela pourrait être fait à lors d'un changement de fonctionnement du marché mais il y aurait alors un biais historique. Une comparaison est donc loin d'être aisée.

[11] Cette bourse a la particularité d'avoir deux sessions (9h-11h et 13h-15h) ouvertes toutes les deux par fixage.

[12] Ils démontrent que, sur le même marché, la volatilité du fixage matinal est supérieure à celle du second fixage, ce dernier ne se distinguant pas des transactions en continu. Une observation similaire est réalisée par Ho & Lee (1998) concernant la bourse de Hong-Kong et Bildik (2000) sur la bourse d'Istanbul.

à la clôture, la Bourse de Paris a introduit un fixage de clôture. Thomas & Demarchi (1999) mettent en évidence que la généralisation de la clôture par fixage a entraîné, quelle que soit la liquidité des titres, une réduction des effets de clôture en rendement et en volatilité[13].

Chung *et al.* (1999) mettent en évidence sur le NYSE que les configurations de spread sont influencées par les variations intrajournalières de la concurrence entre le teneur de marché et le carnet d'ordre. La microstructure du marché expliquerait donc en partie l'existence de configurations. Stoll & Whaley (1990) avancent une explication similaire pour expliquer l'effet d'ouverture : le plus grand pouvoir de discrimination du spécialiste à ce moment. Ces deux études tout en fournissant des éclaircissements fructueux sur la nature des phénomènes et leur modalités d'action repoussent en fait le problème. En effet, elles débouchent sur de nouvelles questions : "D'où proviennent les variations intrajournalières de la concurrence entre le carnet d'ordre et le spécialiste ? ou encore, Pourquoi le pouvoir de discrimination du spécialiste est plus important à l'ouverture ?". Elles constituent donc une avancée importante sans pour autant apporter une réponse définitive à la question de recherche.

L'existence de périodes de fermeture du marché pourrait à elle seule expliquer les configurations. Brock & Kleidon (1992) suggèrent que l'existence de période de fermeture induit une moindre élasticité de la demande avant (impossibilité à rebalancer son portefeuille) et après (accumulation d'information) ces périodes. Cet accroissement de la demande d'immédiateté à l'ouverture et à la clôture induit alors les configurations de la volatilité et de l'activité. Slezak (1994) modélise également l'influence des périodes de fermeture[14]. En présence d'agents averses aux risques, une période de fermeture induit une augmentation du risque pour tous les agents puisqu'il y a moins d'opportunités de traiter (risque de non-exécution). Cette augmentation est cependant plus importante pour les informés qui ne pourront plus observer les prix, ce qui induit en réaction des transactions d'assurance des informés. Ces phénomènes augmentent le risque à la clôture. De même l'avantage des informés est plus important à l'ouverture (les non-informés ne pouvaient pas observer les prix qui sont une source d'information). L'aversion aux risques des agents induit alors des configurations en rendement et en volatilité des transactions. Toutes choses égales par ailleurs, ces modélisations démontrent que l'existence d'une période de fermeture et l'aversion aux risques des agents sont suffisantes pour générer des configurations de rendement en moyenne et en volatilité.

L'introduction de procédures spécifiques d'ouverture et de clôture vient cependant réduire l'influence potentielle de la fermeture du marché. Par exemple, le fixage de clôture diminue sen-

[13]Ce résultat est à mettre en relation avec l'augmentation de la volatilité et de la fourchette avant l'introduction du fixage de clôture mise en évidence par Hillion & Suominen (1998). Ces résultats (ainsi que des suspicions de manipulation de cours) avaient incité la Bourse de Paris à généraliser le fixage de clôture. Cette plus forte pression à la clôture peut être expliquée par une plus forte demande d'immédiateté mise en évidence par Cushing & Madhavan (2000) sur les firmes du Russel 100. Pendant les 5 dernières minutes de la journée la demande d'immédiateté est plus importante et les prix sont plus sensibles au flux d'ordres.

[14]Voir également Hong & Wang (2000) à ce sujet.

siblement le risque de non-exécution mentionné par Slezak (1994). De même, la pré-ouverture est destinée à permettre une découverte du prix plus efficiente. Biais *et al.* (1999) démontrent le rôle d'apprentissage de cette procédure[15]. Il y a convergence du prix théorique vers le prix d'ouverture. On pourrait donc attendre une moindre volatilité à l'ouverture si cette procédure remplit un rôle de découverte du prix comme on peut s'y attendre[16]. Cependant il semble y avoir peu a gagner à l'allongement de la pré-ouverture. Le modèle proposé par Broccard & Calcagno (1999) démontre théoriquement qu'en présence d'asymétrie d'information et d'agents stratégiques, la découverte de prix n'a lieu que pendant une phase active. Ce modèle apporte un support théorique aux résultats empiriques de Biais *et al.* (1999) selon lesquels la découverte a majoritairement lieu à la fin de la pré-ouverture.

b. Flux informationnel et interventions des informés

Les configurations observées pourraient également être dues à des périodicités du flux informationnel ou des biais dans les prix. Lüders & Peisl (2000) proposent une modélisation des rendements qui leur permet d'expliquer les anomalies boursières par un flux informationnel dont la volatilité est stochastique. La microstructure propose également plusieurs courants de recherche abordant cette hypothèse.

Les configurations pourraient, par exemple, être dues à une modification des proportions d'achat et de vente entraîne qui entraînerait une tendance dans le prix. Porter (1992) sur le Toronto Stock Exchange et le NYSE décèle que la probabilité de transaction à l'achat et les rendements évoluent de concert : ils sont sont faibles le lundi à l'ouverture et élevés en fin de journée et à l'ouverture (excepté le lundi). Il infère que les configurations de rendement sont surestimées et prétend que ce motif pourrait expliquer l'ensemble du phénomène mis en évidence par Wood *et al.* (1985). Aitken *et al.* (1995) analysent le même thème sur l'Australian Stock Exchange (ASX) de septembre 1990 à août 1993 (267 actions cotées via le système Stock Exchange Automated Trading System (SEATS)). Ils démontrent l'existence de variations intrajournalières : la probabilité de transaction à l'achat augmente à la fin de la journée mais ce phénomène n'explique, selon les auteurs, qu'au maximum 50% des rendements de clôture. La danse de la fourchette conjuguée à une modification de la fréquence relative d'achat semble donc partiellement responsable des configurations intrajournalières. Cependant, rien ne prouve que cette modification de probabilité ne résulte pas d'une variation du flux informationnel. Ce type d'explication est compatible avec le modèle proposé par Admati & Pfleiderer (1989). Dans ce modèle, les teneurs de marché découragent pendant certaines périodes les transactions

[15]Comme démontré par Cao *et al.* (2000), la procédure d'ouverture du NASDAQ permet également une découverte de prix. La procédure de pré-ouverture utilisée à la bourse de Paris n'est donc pas la seule possible.

[16]L'expérience menée par Biais & Pouget (1999) montre qu'une pré-ouverture permet aux agents de mieux apprendre les stratégies d'équilibre et donc améliorent la qualité du marché.

à l'achat ou à la vente afin de résoudre l'asymétrie d'information[17]. Cela permet alors au teneurs de marché de mieux inférer le contenu informationnel des transactions. Ce modèle a l'avantage de justifier l'existence de configuration de rendement en moyenne. On voit donc que les explications proposées ne sont pas exclusives, ce qui complique d'autant l'analyse et l'inférence.

Admati & Pfleiderer (1988) développent des modèles d'information privée où des configurations de volatilité ou de rendement surviennent en raison d'interactions entre les agents informés et non informés. Pour justifier l'existence des configurations à l'ouverture et à la clôture, les auteurs suggèrent alors que l'impossibilité de traiter pendant la fermeture peut induire une plus forte activité des non-informés non-discrétionnaires (ces derniers peuvent choisir leur instant d'intervention). Cette plus forte liquidité induirait en retour une plus forte activité des non-informés discrétionnaires et des informés. Dans ce modèle si on considère que l'acquisition d'information est endogène, alors les prix sont plus informatifs pendant les périodes où la liquidité est plus forte. Cela signifierait que les configurations coïncident avec un flux informationnel plus important. Cependant, la validité de ce modèle est mise en cause par Foster & Viswanathan (1993). Ces derniers mettent en évidence que les coûts de sélection adverse sont plus importants lorsque l'activité est plus importante alors que ce modèle prévoyait le contraire.

Back & Pedersen (1998) étendent le précédent modèle en considérant que l'information à une durée de vie longue. On est alors obligé de différencier l'arrivée auprès de l'informé et l'arrivée dans le marché (lorsque l'informé exploite son information). À l'instar du modèle de Admati & Pfleiderer (1988), dans cette structure, la volatilité suit la même évolution que la liquidité. Il y cependant deux différences, ni les prix ni l'impact des ordres n'ont de variations systématiques et l'arrivée auprès des informés n'a pas d'importance.

Les modèles de décomposition des coûts de transactions peuvent également éclairer les causes des configurations. Grossièrement, les coûts se répartissent entre trois composantes : l'asymétrie d'information, l'effet d'inventaire (ou d'incitation) et un coût de traitement fixe. L'évolution au cours de la journée du premier coût peut donc permettre d'évaluer dans quelle mesure les configurations observées sont dues à une périodicité du flux informationnel. Cependant, avant de présenter les résultats de plusieurs études, on doit remarquer que ceux-ci sont très dépendants du modèle utilisé[18].

Hasbrouck (1991) décompose la variance des changements de prix d'équilibre à l'aide d'une méthodologie VAR afin de séparer les variations entre des composantes non corrélées (l'apport informationnel) et la partie corrélée (bruit). Sur un échantillon de titre du NYSE et de l'American Stock Exchange (AMEX), il met en évidence que l'apport informationnel absolu des tran-

[17]Cette discrimination est faite par les teneurs de marché en orientant la fourchette. Cela génère alors des périodes orientées à l'achat ou à la vente. Pendant ces périodes, les informés sont prépondérants puisque les non-informés sont découragés par l'orientation défavorable de la fourchette.

[18]Le lecteur intéressé trouvera dans De Winne & Majois (2003) et dans Biais *et al.* (1997, partie 3, p. 151-269) une mise en perspective des différents modèles de décomposition et de leurs résultats.

sactions décroît au cours de la journée alors que l'apport relatif augmente. Cela suggère donc que la volatilité matinale est influencée par l'arrivée d'information publique qui est incorporée progressivement dans le prix. À l'inverse, l'importance de l'asymétrie d'information augmente au cours de la journée. Lin *et al.* (1995) analysent la fourchette de 150 actions cotées sur le NYSE en 1988. Leur modèle met en évidence un déclin au cours de la journée de l'asymétrie d'information. Madhavan *et al.* (1997) décomposent les fourchettes selon les 3 composantes déjà mentionnées. Ils estiment ensuite ce modèle sur 171 actions de la base de données Institute for the Study of Security Markets (ISSM)[19]. Comme Hasbrouck (1991) ils mettent en évidence un déclin au cours de la journée de la part de la volatilité due à l'information publique. Cependant ils mettent également en évidence un déclin de contribution de l'information privée au cours de la journée. La courbe en U des coûts de transactions provient selon eux de l'augmentation au cours de la journée des coûts d'incitation ; ce qui est en contradiction avec les résultats de Hasbrouck (1991). Declerck (2000*a*) en estimant le modèle de Huang & Stoll (1994) sur les titres du CAC 40 entre janvier et juin 1998 met en évidence (cf. p. 189-191) une courbe en U de l'asymétrie d'information alors que les coûts d'inventaire et de traitement des ordres suivent une courbe en $\cap$ (voire en J), un résultat similaire est obtenu par Ahn *et al.* (2002) en appliquant le modèle de Madhavan *et al.* (1997) sur la Bourse de Tokyo (marché dirigé par les ordres). Ces résultats sont donc proches de ceux de Hasbrouck (1991) où l'information (publique plus privée) était plus importante le matin et le soir. Chan (2000) analyse les déterminants de l'impact prix des transactions dans le marché de Hong Kong (marché dirigé par les ordres), il montre que le coût d'inventaire et de sélection adverse sont tous les deux significatifs mais que ce dernier est prépondérant. L'impact prix des transactions est inversement lié à la capitalisation, au taux de rotation et aux transactions de liquidité. D'autre part l'impact prix intrajournalier suit une courbe en U alors que sur le NYSE une tendance décroissante au cours de la journée a été mise en évidence. Ahn *et al.* (2002), en utilisant le modèle de Madhavan *et al.* (1997), soulignent également que la décomposition et la dynamique du spread sur la bourse de Tokyo est différente de celle observée sur les marchés américains. Spécifiquement, ils mettent en évidence sur les composants du NIKKEI 225 de janvier à mars 2000 que les coûts de sélection adverse et de traitement suivent tous les deux des courbes en U mais que ces dernières sont indépendantes. La taille de transaction influence positivement (négativement) le coût de sélection adverse (de traitement des ordres).

Si toutes les études s'accordent sur le fait que les transactions véhiculent de l'information et donc augmentent la volatilité[20], en revanche il n'y a pas unanimité sur les configurations du flux informationnel. Les résultats suggèrent que le poids de l'information publique décline au

[19]Cette estimation est réalisée par GMM à partir des prix et ensuite avec les variances et les covariances. L'utilisation de plusieurs méthodes d'évaluation permet de vérifier la robustesse de cette modélisation.

[20]Elles sont donc toutes en accord avec French & Roll (1986).

cours de la journée. Les conclusions sur l'information privée sont plus problématiques. Certaines études mettent en évidence une augmentation dans la journée alors que d'autres penchent pour un déclin. La relation entre les configurations intrajournalières et l'information est donc incertaine. Chung *et al.* (1999) en analysant les interactions entre le spécialiste et les ordres à cours limite sur le NYSE concluent que les modèles de fourchette fondées sur l'information sont plus en accord avec leurs résultats. Ils suggèrent en outre que les similitudes entre les séries (cotations par le spécialiste, introductions d'ordres à cours limite...) peuvent résulter d'une relation commune à une variable non observable comme l'intervention d'informés. Il est donc tout à fait plausible que les configurations intrajournalières de rendement et/ou de volatilité soient induites par le flux informationnel.

2.2 Méthodologie d'analyse

Notre méthode d'analyse s'inspire au départ de celle utilisée par Foster & Viswanathan (1993), Gerety & Mulherin (1994) ou Xu & Wu (1999) qui emploient une régression avec des variables muettes identifiant les intervalles. À titre d'exemple, le modèle de régression que nous utilisons pour une longueur 30 minutes est :

$$\begin{aligned} R_i = & \, Intercept \\ & + H_{09:30} \times D_{09:30} + \cdots + H_{17:30} \times D_{17:30} \\ & + J_{Lundi} \times D_{Lundi} + \cdots + J_{Vendredi} \times D_{Vendredi} \\ & + O_{-4} \times D_{-4} + \cdots + O_4 \times D_4 + \varepsilon \end{aligned} \tag{2.1}$$

où, R_i est le rendement observé pour l'intervalle i, *Intercept* une constante, $D_{09:30}$ (D_{Lundi})[21] sont des variables muettes qui identifient les intervalles du jour (les jours de la semaine).

Les variables $J_{\cdot}$ ne permettent d'identifier que des différences de niveaux entre les jours de la semaine. Ces variables ne permettent donc pas de mettre en évidence des changements dans la répartition intrajournalière en fonction du jour de la semaine. Pour identifier ce genre d'effet, il faudrait avoir des variables muettes qui traduisent en même temps le jour de la semaine et l'heure du jour ce qui produirait un trop grand nombre de variables. Nous expliquerons ultérieurement comment nous testons ce type d'effet. Au cours de la période d'étude les horaires de cotation ont changé. Il n'est alors plus possible de comparer les résultats obtenus pour un même intervalle qui, au début, était l'intervalle d'ouverture (de clôture) et qui, suite aux changements d'horaire, se situe désormais à l'intérieur de la journée. Les variables muettes $D_{\cdot}$ identifient la

[21] Pour éviter que les variables soient colinéaires, 13h et mercredi ne donnent pas lieu à la création de variable muette.

position relative d'un intervalle dans la journée sur une longueur de 2 heures[22] à l'ouverture et à la clôture. Par exemple, la variable muette D_{-1} (D_1) identifie le dernier (le premier) intervalle de la journée. Nous cherchons donc à évaluer les paramètres $H_\cdot, J_\cdot$, et $O_\cdot$ dont l'indice dénote quels effets ils capturent.

Lors de l'analyse de telles données il y des problèmes économétriques spécifiques. Ronen (1997) et Andersen *et al.* (2001) soulignent particulièrement les problèmes de recouvrement des données et de la corrélation. Puisque nos données sont contiguës, il n'y pas de problème de recouvrement. Toutefois, afin de prendre en compte la corrélation, nous estimons cette régression par la méthode des GMM (Hansen, 1982) en utilisant comme conditions l'orthogonalité entre les variables muettes et les résidus[23]. Nous sommes donc dans le cas d'un système juste identifié[24]. Nous utilisons une matrice de variance-covariance robuste (Newey & West, 1987) afin de prendre en compte l'hétéroskédasticité et l'autocorrélation des séries, caractéristiques largement mises en évidence sur des données intrajournalières. Nous analysons ensuite les paramètres obtenus avec les outils standards[25].

Pour évaluer des effets liés aux changements de microstructure (changement de la procédure d'ouverture ou de clôture, modification des horaires de cotation) ou les différences de configurations intrajournalières entre les jours de la semaine, nous divisons la base de données selon le critère pertinent (e.g. selon le jour de la semaine pour le dernier cas évoqué) et estimons le modèle sur les différents sous-échantillons ainsi formés[26].

2.2.1 Hypothèse sur les configurations

D'après les résultats présentés dans la revue de la littérature, on peut attendre les configurations suivantes :

- Le rendement du dernier intervalle est le seul significativement différent de zéro, il est même significativement positif. La configuration la plus probable est un J. Cette spécificité est concentrée dans les dernières minutes de la journée mais assez importante pour influencer un intervalle plus long. On peut également attendre un rendement négatif le lundi à l'ouverture. La durée de ce phénomène est très incertaine.
- Les configurations de volatilité sont symétriques. La volatilité est plus forte à l'ouverture et à la clôture, ces deux phénomènes sont de même amplitude (configuration en U). Le

[22] Cette longueur de 2 heures (pour identifier les premiers et les derniers intervalles de la journée) a été choisie en fonction de la longueur des configurations mises en évidence dans la littérature.

[23] Cette méthode est identique à celle employée par Huang & Stoll (1994) dans un cas similaire.

[24] C'est à dire qu'il y a autant de conditions de moments que de paramètres à estimer.

[25] Les estimateurs GMM ont l'avantage d'être asymptotiquement normaux, ont peut donc utiliser des tests standards d'autant plus que la volatilité est calculée par la valeur absolue et qu'il n'y a pas de recouvrement des données (se référer à Andersen *et al.* (2001) pour une discussion plus générale de ce point).

[26] Nous retirons au passage toute variable qui ne serait plus pertinente. Dans le précédent exemple, les variables se référant au jour de la semaine ne seraient plus pertinentes.

milieu de la journée se distingue donc par une moindre volatilité.

Les effets des changements de microstructure peuvent permettre d'éclairer les causes des configurations.

Hypothèse 1.1 Si les configurations sont dues au flux informationnel d'origine économique *(e.g. la diffusion d'information à heure régulière), toutes choses égales par ailleurs, les modifications d'horaires ne devraient entraîner aucun changement des configurations.*

On devrait donc observer que les variables muettes $H_{\cdot}$ sont significatives et sont les mêmes quelles que soient les caractéristiques des périodes d'évaluations (e.g. horaire ou procédure d'ouverture du marché).

Hypothèse 1.2 Si les configurations sont dues à l'existence de périodes de fermeture, *l'extension des horaires de cotation devrait impliquer un déplacement des configurations sans autres modifications. Si l'amplitude des configurations est liée à la longueur de la période de non cotation, l'extension des horaires devrait se traduire par une réduction des configurations. Si l'efficience de la pré-ouverture n'est pas liée à sa longueur, l'allongement de la pré-ouverture ne devrait entraîner aucune réduction des configurations d'ouverture.*

D'après cette hypothèse seules les variables $O_{\cdot}$ devraient être significatives. De plus, s'il existe une relation entre la longueur des fermetures et l'importance des configurations alors on devrait obtenir des paramètres plus faibles dans les dernières périodes ou les horaires ont été substantiellement allongés. Cette hypothèse devrait également se traduire par des configurations plus importantes avant (ou après) le week-end.

Hypothèse 1.3 Si l'effet de clôture est dû à un déséquilibre, *l'introduction du fixage de clôture sur les titres à forte liquidité devrait entraîner une baisse des configurations à la clôture tout spécialement pour la volatilité.*

On devrait alors observer une baisse des paramètres observés à la clôture après la date d'introduction du fixage de clôture.

Hypothèse 1.4 Si les configurations sont dues à l'asymétrie d'information, *les configurations devraient être plus importantes pour les titres les moins capitalisés. En effet, l'asymétrie d'information est plus importante pour les titres peu traités.*

Cette relation n'est pas univoque puisque les coûts de transactions sont également plus importants pour les titres peu traités. Une relation entre l'amplitude des configurations et l'activité des titres peut donc avoir d'autres interprétations que l'asymétrie d'information.

La principale prédiction de la MDH, sur laquelle s'accordent toutes les modélisations, est que la volatilité évolue en temps informationnel[27]. On peut donc émettre l'hypothèse suivante :

Hypothèse 1.5 *L'utilisation d'une approximation du flux informationnel pour mesurer le temps induit une réduction des configurations de volatilité. De plus, si l'hypothèse 1.4 est vérifiée, alors la réduction des configurations en temps activité devrait être plus importante pour les titres peu traités.*

Selon cette hypothèse, les coefficients caractérisant l'amplitude des configurations de volatilité (i.e. la somme des variables qui s'appliquent à un intervalle) devraient être plus faible en temps activité.

Nous nous devons de souligner que ces hypothèses ne sont pas les seules explications possibles des phénomènes envisagés. Leur validation empirique constituerait au mieux une indication mais pas l'invalidation d'hypothèses concurrentes.

Suite aux résultats de la pré-analyse, présentée en annexe C, nous n'utiliserons par la suite que deux prix (le prix de transaction et le prix milieu de fourchette) et qu'une modalité de calcul (fin d'intervalle synchronisé par ajustement linéaire). Nous commencerons par présenter les résultats concernant le groupe HL dans la section 2.3. Ensuite, nous résumerons les résultats obtenus pour les autres groupes de liquidité dans la section 2.4.

2.3 Configurations groupe HL

Dans la sous-section 2.3.1, nous commencerons par analyser les résultats obtenus en *temps calendaire*. Ensuite, nous étudierons l'influence des déformations temporelles sur les configurations dans la sous-section 2.3.2.

2.3.1 Configurations en temps calendaire

La figure 2.1 illustre les configurations observées pour le groupe HL avec une longueur de 15 minutes[28].

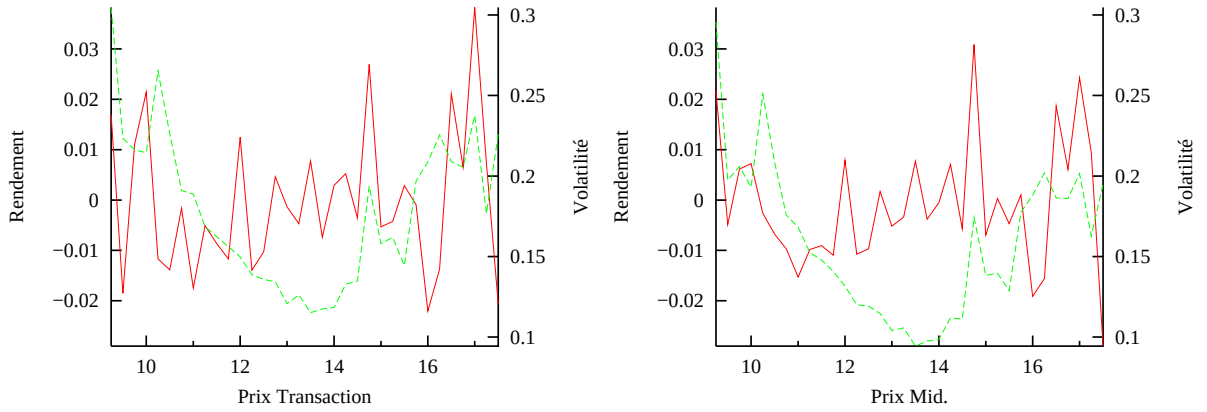
L'observation de cette figure fait ressortir plusieurs faits :

– Il semble à première vue difficile de parler de configuration de rendement. Les courbes sont très heurtées et la clôture se distingue à peine du reste de la journée.

[27] Une relation entre le rendement et le flux informationnel est mise en avant par Harris (1986*a*, 1987) et Richardson & Smtih (1994). Cependant cette relation est, d'après les auteurs, faible en raison de la plus forte dispersion des rendements. Aussi, nous n'attendons d'effet que sur la volatilité.

[28] La faible longueur de l'intervalle nuisant à la lisibilité des figures, le lecteur trouvera en annexe D la même figure pour une longueur de 5 minutes.

FIGURE 2.1
Configurations en *temps calendaire*
groupe HL, 15 minutes



Cette figure donne, pour les différents prix, les configurations observées en temps calendaire. Le prix utilisé est indiqué en bas de chaque graphique. La courbe rouge continue est le rendement (axe de gauche), la courbe verte hachurée est la volatilité (axe de droite). Les deux graphiques de cette figure ont la même échelle.

- À l'inverse une configuration de volatilité est nettement visible (surtout si on considère les changements d'horaires intervenus). On observe une configuration en ι (l'effet d'ouverture est plus important que l'effet de clôture).
- Les changements de microstructure ont visiblement eu des conséquences sur les configurations comme l'indiquent les fortes discontinuités aux marges de la journée.

Le prix milieu de fourchette réduit notablement l'amplitude des oscillations au cours de la journée. Aussi, par la suite, nous nous focaliserons sur ce prix.

La table 2.2 (2.3) donne les paramètres du modèle (2.1) estimés sur les rendements (la volatilité) pour le prix milieu de fourchette en *temps horloge*. Par souci de concision et de clarté, nous ne présentons que les paramètres estimés avec un intervalle d'une heure et indiquerons les modifications observées avec un intervalle de 5 minutes.

Seuls trois intervalles du jour sont significativement positifs, 14h-15h dont la spécificité est à rapprocher de l'ouverture du NYSE[29], et les deux intervalles qui terminent la journée (17h et 18h). Toutefois, on doit également noter que le dernier et l'avant dernier intervalle de la journée (O_{-1} et O_{-2}) sont eux aussi significatifs (au seuil de 5%) mais négatifs. Cette conjonction ne permet pas de comprendre quelles sont exactement les configurations. L'analyse par horaire de cotation permettra d'éclaircir ce point. L'unique effet notable que l'on peut dès à présent retenir est l'influence significative et positive de l'ouverture du NYSE sur le rendement.

[29]Gouriéroux *et al.* (1999*b*) ont également mis en évidence l'effet de l'ouverture du NYSE sur des titres du CAC 40.

TABLE 2.2
Configurations de rendement groupe HL
1 heure, *temps horloge*

Paramètre	Estimation	Écart-type	T-stat
Int.	$-0{,}038^{**}$	0,016	-2,34
10:00	0,045	0,061	0,74
11:00	$-0{,}001$	0,042	-0,03
12:00	0,035	0,026	1,35
14:00	0,024	0,015	1,59
15:00	$0{,}049^{***}$	0,018	2,74
16:00	$0{,}061^{*}$	0,033	1,86
17:00	$0{,}176^{***}$	0,051	3,48
18:00	$0{,}133^{**}$	0,061	2,17
O_{-2}	$-0{,}075^{**}$	0,034	-2,19
O_{-1}	$-0{,}129^{**}$	0,053	-2,44
O_1	0,009	0,046	0,20
O_2	$-0{,}048$	0,030	-1,62
Lundi	0,018	0,018	1,01
Mardi	0,024	0,016	1,48
Jeudi	0,011	0,017	0,66
Vendredi	0,020	0,017	1,22

Cette table fournit les paramètres du modèle (2.1) estimés sur les rendements. Int. désigne la constante. Une variable significative au seuil de 1% est marquée par ***, au seuil de 5% par ** et au seuil de 10% par *.

Le nombre d'intervalles significativement différents de O est beaucoup plus important pour la volatilité que pour le rendement. On peut noter l'existence d'un effet premier intervalle du jour, O_1, 10h et 11h sont tous trois positifs et significatifs respectivement au seuil de 1%, 5% et 5%. Un effet dernier intervalle du jour est présent. Celui-ci semble avoir disparu lors du recul de l'heure de clôture puisque 18h est significatif mais de signe opposé à O_{-1}. Enfin, le milieu de journée se caractérise par une volatilité particulièrement faible.

L'étude des configurations en fonction des horaires de cotation permet d'éclairer les interactions troubles entre les positions relatives ($O_.$) et les positions absolues ($H_.$) ainsi que les causes des configurations. En fonction des changements de microstructure on peut définir trois périodes pendant lesquels les horaires de cotation en continu[30] n'ont pas changé :

i.) 02/01/97 - 17/09/99 (682 jours de bourse) : cotation en continu de 10h à 17h

ii.) 20/09/99 - 31/03/00 (139 jours de bourse) : cotation en continu de 9h à 17h

iii.) 03/04/00 - 29/12/00 (188 jours de bourse) : cotation en continu de 9h à 17:30

[30]Cette classification ne tient pas compte des modifications intervenues sur les modalités d'ouverture, de clôture ou encore, sur l'avancement de la pré-ouverture.

TABLE 2.3
Configurations de volatilité groupe HL
1 heure, *temps horloge*

Paramètre	Estimation	Écart-type	T-stat
Int.	0,260***	0,012	20,89
10:00	0,099**	0,042	2,35
11:00	0,072**	0,028	2,55
12:00	0,050***	0,017	3,00
14:00	−0,048***	0,010	-4,80
15:00	0,036***	0,012	3,05
16:00	0,032	0,023	1,40
17:00	0,054	0,037	1,45
18:00	−0,100**	0,043	-2,31
O_{-2}	0,016	0,024	0,66
O_{-1}	0,092**	0,039	2,33
O_1	0,133***	0,032	4,13
O_2	0,014	0,020	0,69
Lundi	0,013	0,014	0,93
Mardi	−0,003	0,012	-0,27
Jeudi	0,018	0,013	1,36
Vendredi	0,018	0,014	1,28

Cette table fournit les paramètres du modèle (2.1) estimés sur la volatilité. Int. désigne la constante. Une variable significative au seuil de 1% est marquée par ***, au seuil de 5% par ** et au seuil de 10% par *.

Les coefficients du modèle (2.1) significatifs au seuil de 10% pour les rendements (la volatilité) en divisant la base selon ces sous-périodes sont repris dans la table 2.4 (2.5).

Les configurations de rendement par horaire de cotation sont particulièrement instables puisqu'aucun des paramètres n'est significatif dans les trois sous-périodes considérées. Aussi, nous n'analyserons pas plus avant ces résultats.

Les configurations de volatilité sont plus régulières. L'effet ouverture est observable dans chacune des sous-périodes. Il a suivi les modifications d'heure d'ouverture. Le milieu de la journée est toujours caractérisé par une plus faible volatilité. L'effet dernier intervalle du jour est lui aussi toujours significatif. Son absence de la table 2.5 est due à l'utilisation d'un intervalle d'une heure alors que le dernier intervalle ne fait que 30 minutes[31].

Lorsqu'on on définit les sous-périodes d'après les modifications de microstructure (tout changement de réglementation affectant les modalités de cotation) des titres on obtient alors les périodes suivantes :

i.) 02/01/97 - 29/05/98 (352 jours de bourse) :

[31]Si on utilise un intervalle de 30 minutes (tous les intervalles ont la même longueur) alors l'effet dernier intervalle du jour est toujours présent. Par exemple, on obtient dans la dernière sous-période 0,086 (T-value 4,41) pour l'intervalle 17:30. On observe donc une diminution de cet effet mais il demeure très significatif.

TABLE 2.4
Configurations de rendement par horaire de cotation
groupe HL, 1 heure

Paramètre	Estimation	Écart-type	T-stat
i.)2 janvier 1997 - 17 septembre 1999			
15:00	0,043*	0,022	1,95
ii.) 20 septembre 1999 - 31 mars 2000			
Int.	−0,104***	0,040	-2,62
10:00	0,146**	0,060	2,43
12:00	0,134***	0,047	2,88
14:00	0,091**	0,039	2,36
15:00	0,129***	0,045	2,84
17:00	0,132**	0,054	2,45
iii.) 3 avril 2000 - 29 décembre 2000			
11:00	−0,082**	0,041	-2,01
17:00	0,074*	0,043	1,70
Lundi	0,063*	0,038	1,68
Mardi	0,083**	0,033	2,53
Vendredi	0,085**	0,034	2,48

Cette table fournit les paramètres du modèle (2.1) estimés sur le rendement lorsque la base de données est divisée par horaire de cotation. Int. désigne la constante. Une variable significative au seuil de 1% est marquée par ***, au seuil de 5% par ** et au seuil de 10% par *. Seuls les coefficients significatifs sont repris dans cette table.

ii.) 02/06/98 - 17/09/99 (331 jours de bourse) : Introduction d'un fixage de clôture avec pré-clôture de 17 à 17:05.

iii.) 20/09/99 - 17/03/00 (129 jours de bourse) : Extension des horaires de cotations à 9h 17h.

iv.) 20/03/00 - 31/03/00 (9 jours de bourse) : Recul de la pré-clôture à 17:30.

v.) 03/04/00 - 22/09/00 (120 jours de bourse) : Avancement du début de la pré-ouverture à 7:30.

vi.) 25/09/00 - 29/12/00 (68 jours de bourse) : Disparition du Règlement Mensuel

Les coefficients du modèle (2.1) significatifs au seuil de 10% pour les rendements (la volatilité) obtenus en divisant la base selon la réglementation sont repris dans la table 2.6 (2.7). Les longueurs de ces périodes étant très dissemblables, les résultats doivent être analysés avec précaution.

Cette subdivision souligne encore l'instabilité des configurations de rendement. On peut remarquer que, pour la période du 02 juin 1998 au 17 septembre 1999 qui correspond à l'effondrement de groupe HL (cf. figure 1.1 page 32), aucune configuration n'est observable. De manière générale, les configurations sont instables, on ne retrouve pas les mêmes paramètres dans les différentes sous-périodes et quand on obtient les mêmes paramètres, les signes peuvent

TABLE 2.5
Configurations de volatilité par horaire de cotation
groupe HL, 1 heure

Paramètre	Estimation	Écart-type	T-stat
i.) 2 janvier 1997 17 septembre 1999			
Int.	0,239***	0,016	14,85
11:00	0,211***	0,018	11,66
12:00	0,070***	0,016	4,42
14:00	−0,046***	0,013	-3,60
15:00	0,045***	0,015	2,98
16:00	0,050***	0,015	3,39
17:00	0,155***	0,018	8,54
Lundi	0,035*	0,018	1,92
Jeudi	0,039**	0,017	2,24
Vendredi	0,035*	0,019	1,81
i.) 20 septembre 199 - 31 mars 2000			
Int.	0,309***	0,026	11,77
10:00	0,195***	0,037	5,24
11:00	0,093***	0,033	2,82
12:00	0,051*	0,030	1,71
14:00	−0,049*	0,025	-1,92
17:00	0,108***	0,037	2,90
iii.) 3 avril 2000 - 29 décembre 200			
Int.	0,292***	0,027	10,79
10:00	0,238***	0,035	6,72
11:00	0,062**	0,025	2,46
14:00	−0,056***	0,021	-2,71
17:00	0,063**	0,029	2,20

Cette table fournit les paramètres du modèle (2.1) estimés sur la volatilité lorsque la base de données est divisée par horaire de cotation. Int. désigne la constante. Une variable significative au seuil de 1% est marquée par ***, au seuil de 5% par ** et au seuil de 10% par *. Seuls les coefficients significatifs sont repris dans cette table.

TABLE 2.6
Configurations de rendement par microstructure
groupe HL, 1 heure

Paramètre	Estimation	Écart-type	T-stat
i.) 2 janvier 1997 - 29 mai 1998			
11:00	0,068**	0,033	2,05
20 septembre 1999 - 17 mars 2000			
Int.	−0,100**	0,041	-2,43
10:00	0,140**	0,063	2,24
12:00	0,129***	0,048	2,71
14:00	0,089**	0,040	2,22
15:00	0,135***	0,048	2,82
17:00	0,128**	0,054	2,35
Jeudi	0,077*	0,046	1,68
iii.) 20 mars 2000 - 31 mars 2000			
Lundi	0,241*	0,143	1,69
iv.) 03 avril 2000 - 22 septembre 2000			
Int.	−0,101**	0,047	-2,14
15:00	0,090*	0,054	1,67
17:00	0,115**	0,057	2,01
Lundi	0,087*	0,048	1,80
Mardi	0,084**	0,042	2,00
Vendredi	0,079*	0,044	1,81
v.) 25 septembre 2000 - 29 décembre 2000			
11:00	−0,181***	0,060	-3,04
14:00	−0,110**	0,045	-2,47
15:00	−0,124**	0,061	-2,05
18:00	−0,133**	0,058	-2,28
Vendredi	0,093*	0,054	1,72

Cette table fournit les paramètres du modèle (2.1) estimés sur le rendement lorsque la base de données est divisée par changement de microstructure. Int. désigne la constante. Une variable significative au seuil de 1% est marquée par ***, au seuil de 5% par ** et au seuil de 10% par *. Seuls les coefficients significatifs sont repris dans cette table.

changer. Il semble donc excessif de parler de configurations de rendement. Celles-ci dépendent de la période d'estimation et les résultats sont très instables. Elles constituent donc au mieux une propriété des périodes et des marchés étudiés mais pas une propriété générale (ou un fait stylisé pour anticiper sur le chapitre 3)[32].

Comparées à l'instabilité des configurations de rendement, celles de volatilité apparaissent particulièrement stables. L'effet premier intervalle du jour se distingue dans toute les sous-périodes par une très forte volatilité (la plus forte de la journée). La faiblesse de volatilité en milieu de journée est significative dans presque toutes les sous-périodes. L'effet de clôture semble légèrement plus instable. Sa disparition pour certaines sous-périodes est due à l'utilisation d'une longueur d'une heure[33], avec un intervalle de 30 minutes, le dernier intervalle du jour a toujours une volatilité importante et significative.

On peut déjà apporter certaines réponses sur les configurations et leurs origines. Les configurations de rendement nous semblent trop instables pour qu'une conclusion claire soit possible. Les configurations de volatilité sont stables. Sur le marché parisien elles sont une forme en **l**. La plus forte volatilité est observée à l'ouverture et on constate qu'elle est particulièrement faible en milieu de journée pour remonter avant la clôture.

Concernant les causes des configurations, les résultats conduisent à attribuer l'origine des configurations à l'existence de périodes de fermeture (hypothèse 1.2)[34]. En effet, lors des changements d'horaires, nous avons constaté un déplacement des configurations sans autres modifications. Nous rejetons donc l'hypothèse 1.1. Certes l'allongement de la pré-ouverture n'a pas eu d'effet sur les configurations mais l'extension des horaires de cotation n'en n'a pas eu non plus. On doit donc conclure que la longueur de la fermeture nocturne ou de la pré-ouverture n'ont pas d'influence sur l'amplitude des configurations. Cette conclusion est également en accord avec l'absence d'effet jour de la semaine dans les configurations intrajournalières.

L'introduction du fixage a réduit l'effet de clôture sur la volatilité. Aussi, on peut donc croire que cet effet est en partie dû à l'existence d'un déséquilibre pendant cette période (hypothèse 1.3). Cependant on doit souligner que cette interprétation n'est pas la seule possible. Elle pourrait aussi être liée à des effets informationnels.

[32]Un effet de clôture des rendements n'est observable et stable que si on utilise le prix de transaction. Aussi, celui-ci doit être considéré comme résultant de frictions dans le processus de transaction. On observe alors que ce phénomène est concentré dans les dernières minutes de la journée mais suffisamment important pour influencer un intervalle plus long. Donc, sans être en désaccord avec la littérature sur les configurations de rendement, nous sommes obligés d'en souligner la portée limitée.

[33]La clôture ayant lieu à 17:30, avec un intervalle d'une heure le dernier intervalle du jour est deux fois moins long que les autres. Avec une longueur de 30 minutes tous les intervalles ont la même longueur.

[34]Cette interprétation est renforcée par le fait que les configurations de rendement ne sont significatives et stables qu'avec les prix de transaction.

TABLE 2.7
Configurations de volatilité par microstructure
groupe HL, 1 heure

Paramètre	Estimation	Écart-type	T-stat
i.)2 janvier 1997 - 29 mai 1998			
Int.	0,211***	0,018	11,79
11:00	0,196***	0,020	9,63
12:00	0,042***	0,015	2,71
14:00	−0,038***	0,014	-2,66
15:00	0,080***	0,019	4,28
16:00	0,077***	0,018	4,34
17:00	0,178***	0,024	7,49
ii.) 02 juin 98 - 17 septembre 1999			
Int.	0,269***	0,027	9,97
11:00	0,226***	0,030	7,50
12:00	0,100***	0,028	3,57
14:00	−0,054**	0,021	-2,54
17:00	0,130***	0,027	4,75
Lundi	0,058**	0,029	1,99
Mardi	0,039*	0,023	1,66
Jeudi	0,067**	0,028	2,37
Vendredi	0,079**	0,033	2,39
iii.) 20 septembre 1999 - 17 mars 2000			
Int.	0,312***	0,027	11,55
10:00	0,187***	0,039	4,81
11:00	0,083**	0,034	2,47
14:00	−0,056**	0,027	-2,12
17:00	0,095***	0,037	2,59
Mardi	−0,053*	0,030	-1,78
iv.) 20 mars 2000 - 31 mars 2000			
Int.	0,261**	0,114	2,29
10:00	0,300**	0,138	2,17
v.) 03 avril 2000 - 22 septembre 2000			
Int.	0,298***	0,036	8,23
10:00	0,198***	0,044	4,55
11:00	0,080**	0,034	2,35
14:00	−0,059**	0,028	-2,09
vi.) 25 septembre 2000 - 29 décembre 2000			
Int.	0,282***	0,039	7,14
10:00	0,310***	0,061	5,09
14:00	−0,052*	0,029	-1,78
17:00	0,083*	0,043	1,91

Cette table fournit les paramètres du modèle (2.1) estimés sur la volatilité lorsque la base de données est divisée par changement de microstructure. Int. désigne la constante. Une variable significative au seuil de 1% est marquée par ***, au seuil de 5% par ** et au seuil de 10% par *. Seuls les coefficients significatifs sont repris dans cette table.

2.3.2 Configurations en temps activité

Étant donné ce que nous avons observé en *temps calendaire*, nous focaliserons notre analyse sur la volatilité en ne traitant le rendement que de manière annexe. La figure 2.2 présente pour le groupe HL les configurations observées sur l'ensemble de la période d'étude en fonction des prix et des temps utilisés.

L'observation de cette figure fait ressortir plusieurs faits :

- Les *temps activité* réduisent l'amplitude des oscillations des rendements à l'intérieur de la journée ce qui à pour effet de faire ressortir un peu les configurations.
- L'effet des *temps activité* est beaucoup plus important sur la volatilité :
  - L'effet d'ouverture semble préservé
  - La chute en milieu de journée n'est plus observée
  - L'effet de clôture disparaît.

  Au total en *temps activité*, la volatilité décline de manière presque monotone au cours de la journée.

À titre d'exemple, les tables D.1 et D.2, reprises en annexe, donnent les paramètres significatifs au seuil de 10% pour le rendement (en *temps nombre de transactions*) et pour la volatilité (en *temps volume de transactions*) avec un intervalle de 15 minutes. Les résultats confirment les conclusions de l'observation visuelle de la figure D.2. Les configurations de rendement sont toujours erratiques. En revanche les configurations de volatilité se distinguent par une décroissance sur les deux premières heures puis par l'absence de différence significative. Il ne reste donc qu'un effet ouverture.

L'hypothèse 1.5 est donc vérifiée : **Un temps informationnel réduit les configurations de volatilité**. Force est de remarquer que cette réduction affecte différemment l'ouverture et la clôture. Les études intrajournalières des composantes de la fourchette s'accordent pour reconnaître que l'information publique est forte le matin et décroît au cours de la journée. Il semble donc que les déformations temporelles échouent à réduire les configurations dues à l'information publique. Concernant les configurations de clôture il n'y a pas unanimité sur leurs origines (attribuable selon certains modèles aux frictions et selon d'autres à l'asymétrie d'information). Quelle que soit la cause de l'effet de clôture, il est indéniable que les déformations temporelles entraînent sa presque totale disparition.

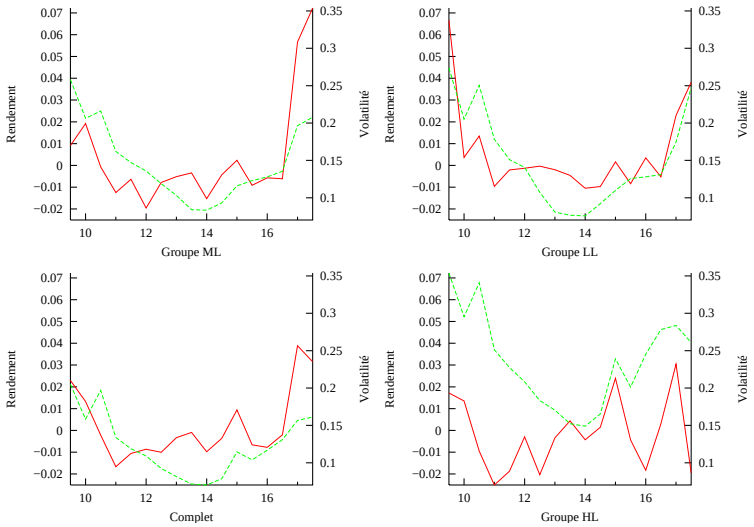
FIGURE 2.2
Configurations intrajournalières
groupe HL, 15 minutes

Cette figure donne pour le prix de transaction et le prix milieu de fourchette les configurations observées dans les différents temps. Le prix utilisé est indiqué à gauche de chaque ligne, le temps en bas de chaque colonne. La courbe rouge continue est le rendement (axe de gauche), la courbe verte hachurée est la volatilité (axe de droite). Tous les graphiques dans cette figure ont la même échelle. La première colonne de cette figure reprend deux les graphiques déjà présentés dans la figure 2.1.

2.4 Influence du groupe de liquidité

Comme nous l'avons vu, le prix milieu de fourchette réduit significativement les configurations de rendement et limite les problèmes de danse de la fourchette. Cependant, les configurations mises en évidence restent globalement similaires. Cela demeure vrai dans les cas que nous allons exposer. Nous continuerons à ne présenter que des résultats sur le prix milieu de fourchette. De même, l'influence des différents *temps nombre* et des différents *temps volume* est très similaire. Nous ne présenterons donc que trois temps : le *temps horloge*, le *temps nombre de transactions* et le *temps volume d'ordres*. La figure 2.3 donne les configurations observées en *temps horloge* pour les différents groupes de liquidité avec un intervalle de 30 minutes.

FIGURE 2.3
Groupes de liquidité
temps horloge, 30 minutes



Cette figure illustre les configurations des différents groupes de liquidité. Le rendement est représenté par la courbe rouge continue (axe de gauche) et la volatilité par la courbe verte hachurée (axe de droite). Tous les graphiques présentés dans cette figure ont la même échelle.

On voit que les configurations de rendement sont différentes selon les groupes de liquidité : un J pour le groupe ML et un l pour le groupe LL. À l'inverse, les configurations de volatilité, même si elles changent d'amplitude, ont la même forme pour tous les groupes de liquidité. La relation entre l'amplitude des configurations de volatilité et la liquidité du groupe n'est pas monotone. Les configurations sont plus fortes pour les groupe HL, vient ensuite le groupe LL et elles sont les plus faibles pour le groupe ML[35]. On doit donc rejeter l'hypothèse 1.4. On peut

[35]Les configurations de volatilité sont influencées par nos modalités de conception des portefeuilles. La volatilité est la valeur absolue de la moyenne des rendements des différentes composantes. Donc, la faible volatilité du

également remarquer que l'effet de l'ouverture du NYSE n'est visible que pour le groupe HL cela est logique puisque seuls les titres de ce groupe sont cotés sur NYSE via des ADR. La figure 2.4 (2.5) donne les configurations, en *temps nombre de transactions* (*volume d'ordres*), pour les différents groupes.

Quel que soit le groupe, un *temps nombre de transactions* a peu d'influence sur la forme ou le niveau des rendements. En revanche ses effets sur les configurations de volatilité changent beaucoup selon le groupe. Pour le groupe LL on observe presque une ligne droite faiblement décroissante. Pour le groupe ML la forme de la courbe s'apparente à un $\smile$ dont l'effet ouverture est légèrement plus prononcé.

À la différence du temps *nombre de transactions*, les effets du *temps volume d'ordres* sont similaires pour les différents groupes. L'effet d'ouverture est préservé, les autres effets s'ils ne disparaissent pas sont d'une significativité très faible. L'effet de cette déformation étant similaire à celui observé sur le groupe HL que nous avons déjà décrit, nous ne l'analyserons pas plus avant.

La table 2.8 (2.9) donne les paramètres des configurations de volatilité significatifs au seuil de 10% pour le groupe ML (LL).

On doit noter que même si les configurations semblent visuellement faibles dans la figure 2.4, la table 2.8 montre qu'elles n'en sont pas moins très significatives[36]. On retrouve les différents effets déjà mis en évidence pour le groupe HL. La valeur négative des variables $O_{\cdot}$ est due à une légère baisse d'amplitude des configurations suite aux modifications d'horaires.

Pour le groupe LL le nombre de paramètres significatifs est plus faible, les configurations sont donc moins importantes. Toutefois, on retrouve une configuration d'ouverture. Au vu de la figure 2.4 un tel résultat peut sembler étrange. La figure D.4 en annexe illustre le phénomène à l'œuvre. On voit en effet que si on considère exclusivement le groupe LL, les effets des déformations temporelles sont plus faibles qu'ils n'apparaissaient dans la figure 2.4. La présence du groupe HL dont les configurations de volatilité sont plus prononcées et moins affectées par les déformations temporelles accroissaient l'importance visuelle des réductions pour les autres groupes. Ces dernières sont effectivement sensibles mais l'effet d'ouverture demeure toujours significatif.

Nous avons observé que l'effet d'une déformation temporelle s'accroissait avec l'illiquidité du groupe. Toutefois, une volatilité plus forte est toujours observée à l'ouverture même si cette augmentation est de moindre importance avec une déformation temporelle (particulièrement en *temps nombre*) qu'en *temps calendaire*.

groupe complet s'explique par un effet de diversification plus important que sur les autres groupes (21 titres contre 7 pour les autres). Pour aller plus loin, cette volatilité est également sensible à la corrélation entre les membres d'un groupe.

[36] La figure D.3, en annexe, montre sans ambiguïtés que les configurations sont encore significatives même en *temps activité*.

FIGURE 2.4
Groupes de liquidité
temps nombre de transactions, 30 minutes

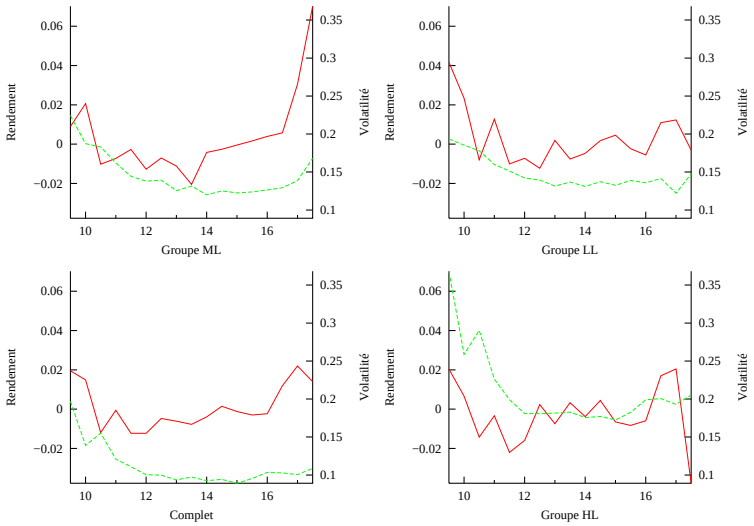
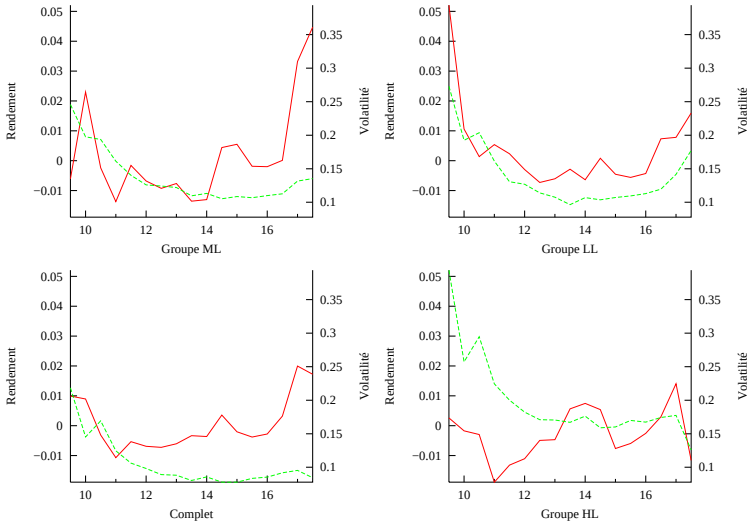


FIGURE 2.5
Groupes de liquidité
temps volume d'ordres, 30 minutes



Ces figures illustrent les configurations des différents groupes de liquidité en *temps nombre de transactions* et en *temps volume d'ordres*. Le rendement est représenté par la courbe rouge continue (axe de gauche) et la volatilité par la courbe verte hachurée (axe de droite). Dans chacune de ces 2 figures les graphiques ont la même échelle.

TABLE 2.8
Configurations de volatilité groupe ML
30 minutes, *temps nb. transactions*

Paramètre	Estimation	Écart-type	T-stat
Int.	0,124***	0,004	28,53
09:30	0,117***	0,019	6,11
10:00	0,098***	0,018	5,49
10:30	0,073***	0,013	5,61
11:00	0,069***	0,014	5,10
11:30	0,027***	0,008	3,39
12:00	0,029***	0,008	3,45
12:30	0,014***	0,005	2,79
16:00	0,031**	0,014	2,16
16:30	0,063***	0,019	3,38
17:00	0,099***	0,022	4,51
17:30	0,133***	0,027	4,95
O_{-3}	−0,035**	0,014	-2,39
O_{-2}	−0,065***	0,019	-3,45
O_{-1}	−0,091***	0,023	-4,02
O_2	−0,037**	0,015	-2,53
O_4	−0,025***	0,009	-2,70

Cette table fournit les paramètres du modèle (2.1) estimés sur les volatilités. Int. désigne la constante. Une variable significative au seuil de 1% est marquée par ***, au seuil de 5% par ** et au seuil de 10% par *. Seuls les paramètres significatifs sont repris dans cette table.

TABLE 2.9
Configurations de volatilité groupe LL
30 minutes, *temps nb. transactions*

Paramètre	Estimation	Écart-type	T-stat
Int.	0,128***	0,005	25,82
09:30	0,043*	0,023	1,89
10:00	0,058***	0,018	3,18
10:30	0,033**	0,014	2,27
11:00	0,033**	0,013	2,50
11:30	0,017*	0,009	1,94
12:00	0,015*	0,008	1,76
Vendredi	0,009*	0,005	1,69

Cette table fournit les paramètres estimés à l'aide du modèle (2.1) appliqué aux volatilités. Int. désigne la constante. Une variable significative au seuil de 1% est marquée par ***, au seuil de 5% par ** et au seuil de 10% par *.

2.5 Conclusion

L'étude des configurations intrajournalières a mis en évidence plusieurs faits intéressants.

Les configurations de rendement sont très instables et changent selon la liquidité du groupe considéré. Il est donc difficile de considérer l'existence des ces configurations comme une caractéristique stable (un fait stylisé) des données haute fréquence.

À l'inverse, les configurations de volatilité sont relativement stables en forme de U, de J ou de ɩ selon le cas. On observe donc trois effets : une forte volatilité à l'ouverture, une volatilité particulièrement faible en milieu de journée et une volatilité élevée à la clôture.

En profitant des changements de procédures d'ouverture et de clôture intervenus pendant la période d'étude, nous avons pu explorer certaines causes possibles des configurations. La cause dominante semble être l'existence d'une fermeture du marché puisque les configurations se sont déplacées dans la journée suite aux modifications des horaires. La longueur de cette fermeture (de la pré-ouverture) ne semble pas avoir d'effet notable puisque l'extension des horaires (de la longueur de la pré-ouverture) n'a pas induit de réduction significative des configurations. L'effet de clôture est dû en partie à un déséquilibre à l'achat : l'introduction d'un fixage de clôture l'a réduit, il est plus marqué sur le prix d'achat.

On peut se demander si les configurations sont dues à l'asymétrie d'information. En effet, plusieurs études ont mis en évidence une plus forte asymétrie d'information pour les titres peu liquides. On aurait donc pu attendre des configurations plus prononcées pour les titres peu actifs si l'origine des configurations avait été l'asymétrie d'information. Or, si les configurations de rendement sont effectivement plus prononcées pour les titres peu liquides, les configurations de volatilité ne le sont pas. Ainsi, la question demeure ouverte.

Nous avons ensuite analysé l'influence de déformations temporelles utilisant des mesures d'activité sur les configurations intrajournalières. Ces mesures d'activité peuvent être vues comme des approximations du flux informationnel. L'hypothèse étant que, si les configurations ont pour origine l'information et si les approximations sont pertinentes, alors on devrait observer une diminution ou une disparition des configurations.

Nous avons mis en évidence que toutes les déformations temporelles réduisaient les configurations intrajournalières. L'effet sur les rendements est trouble principalement en raison de l'instabilité des configurations de rendement. Les déformations induisent systématiquement une réduction de l'amplitude des variations à l'intérieur de la journée. L'effet des déformations sur la volatilité est plus importante : elles permettent de réduire significativement les configurations de volatilité. De manière générale, l'effet d'ouverture n'est que peu ou pas affecté alors que les effets de milieu de journée et de clôture sont réduits ou disparaissent. L'effet des déformations temporelles est plus important sur les titres moins liquides. C'est sur ces titres que l'on observe une réduction de l'effet d'ouverture. Si on classe les approximations en fonction de leur capacité

à réduire les configurations alors les approximations fondées sur le nombre d'événements ont la préférence sur celles employant le volume. Ensuite, l'utilisation des transactions semble préférable à celle des ordres. Le classement provisoire de ces quatre approximations serait donc : nombre de transactions, nombre d'ordres introduits, volume de transactions, volume d'ordres.

La question de savoir si ces déformations temporelles reflètent effectivement le flux informationnel intrajournalier est encore ouverte. S'il y a un accord sur l'influence prépondérante de l'information publique en début de journée, la dynamique de l'asymétrie d'information est sujette à controverse. Pour certains, elle décline au cours de la journée (les configurations sont alors dues à l'augmentation des frictions) alors que pour d'autres, elle augmente dans la journée (c'est alors la cause des configurations). Or, le principal effet des déformations temporelles est une réduction de l'effet de clôture, nous n'observons de réduction des configurations d'ouverture de la volatilité que pour les titres les moins actifs.

Afin de compléter le classement que nous avons établi entre les différentes approximations, le chapitre 3 comparera les caractéristiques des distributions des rendements dans les différentes déformations temporelles et les propriétés théoriques si les rendements étaient échantillonnés en temps informationnel. Cela nous donnera un second critère d'évaluation des approximations.

Chapitre 3

Propriétés des séries intrajournalières

Dans le chapitre 2 nous avons comparé la performance de plusieurs approximations du flux informationnel en utilisant l'outil des déformations temporelles pour étudier leur influence sur les configurations intrajournalières. Tout en restant dans le même cadre d'analyse (données intrajournalières haute fréquence), nous complétons ici cette comparaison en étudiant deux points : d'une part, la distribution des rendements dans les différents *temps déformés* et d'autre part l'influence de ces approximations du flux informationnel sur la persistance de la volatilité en *temps horloge*.

Dans l'hypothèse de la MDH, les rendements sont subordonnés au flux informationnel. Cela induit que les séries observées en *temps horloge* dévient de la distribution normale et présentent, entre autres, un excès de kurtosis. À l'inverse, si les séries sont échantillonnées suivant le processus directeur (i.e. si une même quantité d'information diffusée sépare chacune des transactions retenues) alors les séries devraient être distribuées selon une loi normale.

La persistance de la volatilité est associée par certains auteurs à l'omission du flux informationnel (d'approximation(s) de celui-ci) dans le modèle. On peut alors considérer que plus une approximation est pertinente, plus elle réduit la persistance.

L'analyse de ces deux points nous permettra d'avoir une vue plus complète sur les qualités et défauts des approximations du flux informationnel. Ce chapitre sera organisé comme suit. Nous commencerons par rappeler les faits stylisés relatifs aux rendements à haute fréquence et les méthodologies d'analyse dans la section 3.1. Ensuite, nous présenterons les résultats concernant le groupe HL dans la section 3.2 puis pour les autres groupes de liquidité dans la section 3.3. Enfin, la section 3.4 conclura.

3.1 Propriétés et méthodologies d'analyse

Les propriétés des rendements haute fréquence (intrajournaliers), détaillées dans la sous-section 3.1.1, ont conduit au développement d'outils spécifiques pour les étudier que nous présenterons dans la sous-section 3.1.2. Dans cette dernière sous-section, nous présenterons en particulier notre méthodologie d'analyse.

3.1.1 Propriétés des rendements HF.

Il est reconnu depuis longtemps que l'hypothèse Markov-Gaussienne, c'est à dire l'approximation des cours boursiers par une marche aléatoire aux incréments indépendants et identiquement distribués selon une loi normale est inadéquate. Les propriétés statistiques des séries ont donc été étudiées par de nombreux chercheurs (se reporter à Pagan (1996), Campbell *et al.* (1997), Cont (2001) ou Gouriéroux & Jasiak (2001) pour des revues de la littérature) et on dispose aujourd'hui d'une bonne connaissance de leurs principales caractéristiques.

Suivant la définition employée par Cont (2001), un fait stylisé est une propriété commune à différents marchés et produits. Pour atteindre une certaine généralité, ces propriétés sont surtout formulées en terme qualitatif. Suivant cette définition et notre intérêt pour les séries univariées de rendements, les faits stylisés qui retiennent notre attention sont[1] :

- Les queues de distributions sont épaisses : la distribution des rendements est leptokurtique.
- Lorsqu'on diminue la fréquence d'échantillonage des séries (on accroît le temps entre les observations retenues), on se rapproche de la normalité. On parle parfois de normalité agrégée. La distribution des rendements varie en fonction du pas temporel d'échantillonnage : plus il est petit, plus on s'éloigne de la normalité.
- L'autocorrélation de la valeur absolue des rendements exprimée en fonction du temps séparant les observations décroît lentement (moins rapidement que l'hypothèse Markov-Gaussienne ne le prévoit). Cela est souvent considéré comme une évidence de mémoire longue des rendements.
- Il y a un phénomène de clustering de la volatilité, c'est à dire qu'une forte (faible) volatilité tend à être suivie par une forte (faible) volatilité.
- Même après avoir pris en compte le clustering de volatilité, la distribution conditionnelle des rendements a encore des queues épaisses, bien que moins épaisses que la distribution inconditionnelle.

[1] D'autre faits stylisés sont connus notamment les relations entre le volume et la volatilité (cf. entre de nombreux autres Gallant *et al.* (1992), Ghysels *et al.* (2000)...). Étant donné que nous nous focalisons ici sur les séries univariées de rendements, nous ne les traiterons pas. Ces points sont en partie abordés dans l'introduction de la première partie et dans revue de la littérature du chapitre 4.

Chacun de ces points a donné naissance à une vaste littérature tentant soit de les expliquer, soit de les modéliser. Nous n'avons pas ici pour but de faire une revue exhaustive de ces différents phénomènes mais nous tenterons plutôt de présenter quelques études permettant de comprendre ce cadre de recherche et d'autres mettant en évidence la pertinence de l'approche que nous employons.

Bachelier (1900) a introduit l'hypothèse Markov-gaussienne en finance. En effet, si on considère que les changements de prix sont IID, de variance finie et répartis uniformément dans le temps, alors le théorème central limite prédit que les changements de prix échantillonnés à une fréquence donnée sont distribués suivant une loi normale. Les violations empiriques de l'hypothèse de normalité ont conduit Fama (1963, 1965) à préférer les distributions de Lévy α stables[2] (ou lois de Pareto). Ces distributions découlent de l'agrégation de variables dont la variance est infinie. L'hypothèse relâchée porte donc sur la variance des changements de prix[3]. On peut regretter que de telles lois même si elles décrivent mieux les données n'aient pas forcément de motivation théorique c'est à dire, pas de fondements économiques qui pourraient permettre de comprendre l'origine des phénomènes observés. Nous n'avons pas pour objectif de trouver la distribution qui se rapproche le mieux des données, mais d'essayer de comprendre pourquoi on observe ces phénomènes. Nous tenterons, tout particulièrement, d'évaluer si une hypothèse de variation du flux informationnel peut les expliquer en exploitant l'idée de la MDH.

Sans vouloir suggérer que cette hypothèse soit celle qui décrive le mieux les distributions empiriques, on peut remarquer que plusieurs études (cf. Westerfield (1977) , Hall *et al.* (1989)) confirment la pertinence de ce type de modélisation. De plus, Mandelbrot (1973) souligne la proximité opérationnelle entre l'hypothèse α stable et la subordination proposée par Clark (1973). Il relève principalement deux désaccords. Le premier porte sur l'approximation du flux informationnel à utiliser : Mandelbrot & Taylor (1967) emploient le nombre de transactions alors que Clark (1973) préfère le volume. Le second désaccord porte sur la nature de la variance qui est finie dans l'hypothèse de la MDH alors qu'elle peut être infinie dans l'hypothèse α stable.

Le concept de subordination des rendements au flux informationnel de la MDH peut aussi permettre de comprendre la nature gaussienne des rendements agrégés. Plus on augmente le temps entre deux données échantillonnées, plus la quantité d'information diffusée entre ces deux données est stable (par application du théorème centrale limite). On peut donc attendre que les séries mesurées à très faible fréquence (e.g. mensuelle) soient normalement distribuées alors qu'à haute fréquence elles devraient s'éloigner de manière importante de la normalité. Le concept du subordination suggère donc une explication vraisemblable et intuitive aux relations

[2]La distribution normale est un cas particulier de cette famille lorsque l'indice α est égal à 2.

[3]Le lecteur peut se reporter à Le Courtois (2001) ou à Gouriéroux & Jasiak (2001, p. 428-433) pour d'autres exemples de distributions. On trouve également dans Farmer (1999) une revue des apports de "l'econophysics" à l'étude des données financières et tout particulièrement des distributions issues du champ de la physique.

entre la distribution des rendements et la fréquence d'échantillonage des séries.

Le phénomène de clustering de la volatilité a d'abord été modélisé avec des Auto Regressive Conditionnal Heteroskedasticity (ARCH) (Engle, 1982). Cette classe de modèles a ensuite été étendue par les GARCH (se référer à Bollerslev (1987) pour l'introduction des GARCH et à Bollerslev *et al.* (1991) pour une revue de la littérature en français)[4]. La spécification d'un GARCH(1,1) est :

$$\begin{aligned} r_t &= \mu_t + \varepsilon_t \\ h_t &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1} \end{aligned} \tag{3.1}$$

Ce type de modélisation postule donc que les agents prédisent la variance de la période courante via une moyenne pondérée de la volatilité moyenne de long terme (la constante α_0), de la variance conditionnelle de la période précédente (h_{t-1} le terme GARCH) et de la volatilité observée dans la période précédente (ε_{t-1}^2 le terme ARCH). Si le rendement observé dans la période précédente a été important (à la hausse ou à la baisse), les investisseurs vont réviser à la hausse leurs prédictions de la volatilité pour la période suivante.

Cette classe de modèles a connu un développement très important. Le nombre d'auteurs l'ayant enrichie est si important qu'il est aujourd'hui difficile d'en faire une liste exhaustive. Par exemple, on trouve les EGARCH, les IGARCH, les FIGARCH, les APARCH, les HYGARCH[5]... Ces modélisations divergent par leurs capacités à prendre en compte certains phénomènes. Par exemple, un GARCH classique ne permet pas de prendre en compte le fait que, pour les actions, la réaction de la volatilité est différente selon le signe du choc de rendement, alors que cela est possible avec un EGARCH. Ces modèles diffèrent également par leur généralité : certaines formulations plus complexes englobant de nombreux cas particuliers simples.

Les modèles GARCH posent principalement deux problèmes[6]. D'une part, l'origine de cette dynamique (l'autocorrélation au sens large) demeure incertaine. Par exemple, Brock & Lebaron (1996) lient la persistance de la volatilité au fait que les transactions sont plus fréquentes que les révisions de croyances. Comme résumé par Brunnermeir (2001, p. 113-117)[7], l'autocorrélation pourrait également provenir de l'apprentissage réalisé par les investisseurs en observant les marchés. D'autre part, la persistance de la volatilité mise en évidence par ces modèles ($\alpha + \beta$ dans un GARCH(1,1)) est souvent élevée (de l'ordre de 0,90) et est proche de la contrainte

[4]Le lecteur intéressé trouvera également dans Hamilton (1994, p. 657-676) une présentation économétrique détaillée.

[5]Le lecteur intéressé par l'univers GARCH pourra trouver dans Hansen & Lunde (2002) un catalogue impressionnant, mais non exhaustif, des différentes extensions développées.

[6]Le lecteur est invité à consulter Goodhart & O'Hara (1997) pour une plus ample discussion financière des enjeux et problèmes posés par les modèles GARCH appliqués à des données haute fréquence.

[7]Cette possibilité est également mentionnée par Goodhart & O'Hara (1997).

assurant l'existence du second moment de la distribution ($\alpha + \beta < 1$). Une telle persistance de la volatilité est souvent considérée comme peu vraisemblable. Lamoureux & Lastrapes (1990) attribuent celle-ci à l'omission d'une variable d'information dans l'équation de la volatilité conditionnelle. L'inclusion d'une variable informationnelle, le volume dans leur étude, entraîne une forte chute de la persistance. Des conclusions similaires sont mises en évidence par Locke & Sayers (1993) avec le volume, le nombre de changements de prix ou le déséquilibre du volume, par Donaldson & Kamstran (2000) avec le volume et la volatilité implicite extraite d'options, par Marsh & Wagner (2000) sur les principaux indices boursiers ou enfin par Berg (2000) sur le petit indice suédois. Ces résultats obtenus à partir de données journalières semblent également valides au niveau intrajournalier. Par exemple, Maillet & Michel (1997) mettent en évidence, sur Alcatel, que l'inclusion du volume induit une réduction de la persistance y compris au niveau intrajournalier. Une conclusion similaire est atteinte par Giot (2000) avec le volume ou le nombre de transactions d'IBM. La conclusion globale de ces études est que la persistance de la volatilité mise en évidence par les modèles GARCH est surestimée en raison de l'omission d'une ou plusieurs variables informationnelles dans l'équation de la volatilité conditionnelle.

3.1.2 Méthodologie d'analyse

Comme nous venons de le signaler, la persistance de la volatilité est souvent attribuée à l'omission d'une variable d'information dans l'équation de la volatilité conditionnelle. L'inclusion d'une telle variable fait baisser la persistance de la volatilité. On peut émettre l'hypothèse que plus l'approximation de l'information est pertinente plus cette réduction devrait être importante.

On peut facilement comparer la réduction de la persistance entraînée par les différentes approximations en comparant les résultats des estimations de modèles de type GARCH contenant à chaque fois une approximation différente dans l'équation de la volatilité conditionnelle. Toutefois, le choix d'une modélisation parmi la classe GARCH est difficile devant la multitude des possibilités. Les articles cités par Hamilton (1994, p. 674) mettent en avant des modélisations simples de type GARCH(1,1) ou EGARCH(1,1). Dans une comparaison large (330 modèles sont comparés), Hansen & Lunde (2002) mettent en évidence que, sur le marché des changes, la spécification GARCH(1,1) n'est pas inférieure aux autres. À l'inverse sur les actions (IBM), une modélisation prenant en compte l'effet de levier (e.g. un EGARCH) est préférable. En dépit

ces suggestions nous employons un GARCH(1,1)[8] donné par l'équation (3.2).

$$\begin{aligned} r_t &= \mu_t + \varepsilon_t \\ \varepsilon_t | (A_t^a, \varepsilon_{t-1}, \dots) &\sim N(0, h_t) \\ h_t &= \alpha_0 + \beta_1 h_{t-1} + \alpha_1 \varepsilon_{t-1}^2 + \delta_1 A_t^a \end{aligned} \tag{3.2}$$

où, r_t est le rendement en *temps horloge* observé dans l'intervalle t et A_t^a est l'une des approximations du flux d'informations diffusé dans l'intervalle t et le terme d'erreur est normalement distribué (de moyenne 0 et de variance h_t). Nous utilisons les quatre mesures d'activité qui nous ont servies à définir des déformations temporelles : le nombre de transactions, le volume de transactions, le nombre d'ordres introduits et le volume d'ordres introduits. La meilleure approximation du flux informationnel étant, par hypothèse, celle qui entraîne la plus forte réduction de la persistance.
Les fondements de ce test proviennent de l'hypothèse de la MDH. Une hypothèse additionnelle est que la persistance de la volatilité mise en évidence par un GARCH capture en fait la dépendance temporelle du flux informationnel. Dans cette modélisation, A_t^a est utilisée comme approximation du processus informationnel (le processus directeur des rendements). Elle est dès lors faiblement exogène (se reporter à Lamoureux & Lastrapes (1990, p. 223)). En l'état cette spécification pourrait souffrir de biais de spécification et/ou de biais de simultanéité au sujet de ce dernier point voir la note de bas page 4 p. 228 de ce même article.

D'une manière intuitive l'hypothèse de la MDH postule que le temps d'évolution des marchés financiers n'est pas le *temps calendaire* mais un autre temps (le *temps informationnel*). Sous cette hypothèse, c'est dans ce temps que l'on peut le mieux comprendre ou étudier la dynamique du processus d'intérêt puisque celle-ci n'est plus alors influencée par les modifications du flux informationnel. Plus formellement, la MDH modélise la distribution des rendements par l'équation (3.3).

$$R_t | I_t = N(0, \sigma^2 I_t) \tag{3.3}$$

En relâchant l'hypothèse selon laquelle le rendement associé à une information est 0, Harris (1986*a*, 1987) et Richardson & Smtih (1994) dérivent un modèle légèrement différent repris dans l'équation (3.4).

$$R_t | I_t = N(\mu I_t, \sigma^2 I_t) \tag{3.4}$$

Sous l'hypothèse de variation du flux informationnel le modèle (3.3) prévoit, entre autres, que la distribution inconditionnelle des rendements en *temps calendaire* présente un excès de kurtosis

[8] Nous sommes contraints d'utiliser ici un GARCH car le logiciel que nous employons (SAS^{TM}) ne permet pas d'inclure des variables dans l'équation de h_t avec un modèle EGARCH. Pour un rappel plus précis des hypothèses de la modélisation GARCH, le lecteur peut se reporter à la section 5.2.1.

par rapport à la loi normale, propriété partagée par de nombreux actifs financiers. Le modèle (3.4) fait intervenir le flux informationnel dans la moyenne et la variance de la distribution et prévoit donc qu'elle doit être asymétrique (de même signe que μ) .

Les rendements en *temps déformés* étant mesurés selon une approximation de I_t alors, si le modèle est exact, les rendements devraient être normalement distribués. Une hypothèse plus faible est que les séries en *temps déformés* ont une kurtosis plus faible qu'en *temps calendaire*. En effet cette hypothèse ne requiert pas que le modèle (3.3) soit exact, elle nécessite juste que le concept à l'œuvre dans la MDH (la subordination des rendements au flux informationnel) soit pertinent[9]. Si on considère le modèle (3.4), on peut ajouter à ces hypothèses une réduction de l'asymétrie[10]. Afin de vérifier ces hypothèses, nous produirons les statistiques descriptives des séries. La normalité des séries étant un point central, nous présenterons trois tests de normalité des séries[11] :

i.) Kolmogorov-Smirnov

ii.) Cramer-von-Mises

iii.) Anderson-Darling

Dans les modèles de MDH que nous avons déjà présentés, seul le volume était considéré. Maillet & Michel (1997) confirment le rôle important du volume, ils mettent en évidence que la convergence de la distributions des rendements vers la normalité est plus rapide lorsque les séries sont construites en *temps volume*. Cependant cette prééminence est remise en question par Ané & Geman (2000) qui calculent les moments d'une déformation temporelle stochastique telle que la distribution conditionnelle des rendements soit normale. Or, les moments calculés sont proches de la distribution empirique du nombre de transactions et non de celle du volume. Si on en croit ces auteurs, le *temps informationnel* est le *temps nombre de transactions*. La diversité des résultats motive la comparaison de différentes approximations du flux informationnel.

Même si la normalité des séries n'est pas atteinte, les déformations temporelles peuvent avoir un intérêt opérationnel si elles apportent une simplification de la dynamique des processus étudiés. Par exemple, Chauveau *et al.* (1999) observent sur deux actions françaises (Alcatel et France Télécom) une réduction de la non-linéarité des séries lorsqu'on emploie un *temps volume*. De plus, Vesserau (2000) démontre que les dynamiques non-linéaires des actions françaises sont bien prises en compte par un modèle GARCH(1,1). On peut essayer de marier ces deux approches pour voir si les déformations temporelles conduisent à une simplification des dynamiques mises en évidence par un modèle de type GARCH. Suivant les résultats de Hansen

[9]Même si la forme de la distribution théorique des rendements postulée dans le modèle de MDH est erronée on doit observer une diminution de la kurtosis si le processus de rendement est subordonné au flux informationnel.

[10]Selon les auteurs ayant utilisé ce modèle, cet effet est probablement très faible puisqu'il est lié à la valeur de μ (rendement d'une information) qui ne peut être que minime.

[11]Nous avons omis le test du χ^2 car celui-ci donne toujours les mêmes conclusions que les trois tests présentés.

& Lunde (2002), nous considérons un modèle EGARCH(p,q) (3.5) :

$$\begin{aligned} r_t &= \mu_t + \varepsilon_t \\ \log(h_t) &= \alpha_0 + \sum_{i=1}^{q} \beta_i \log(h_{t-1}) + \sum_{i=1}^{p} \alpha_i g(z_{t-1}) \\ g(z_t) &= \theta z_t + |z_t| - E(z_t) \\ z_t &= \varepsilon_t / \sqrt{h_t} \end{aligned} \tag{3.5}$$

Afin de déterminer si une déformation temporelle apporte une simplification, il faut une méthode pour choisir un modèle. En nous focalisant sur des modèles ayant les mêmes ordres p et q, plusieurs critères peuvent être utilisés[12]. Nous préférons le test du ratio de vraisemblance. Si on considère deux modèles EGARCH d'ordres respectifs i et $i+1$ et de vraisemblances respectives L_i et L_{i+1}, le test du ratio de vraisemblance $LR = -2 \times \ln(L_i/L_{i+1})$ suit une loi du χ^2 dont le nombre de degrés de liberté est le nombre de contraintes, soit 2[13]. La valeur critique au seuil de 5% est alors égale à 5,99. En partant du modèle le plus simple, nous choisissons le premier modèle dont l'ordre supérieur n'apporte pas d'amélioration significative au seuil de 5%.

Nous commençons par étudier en détail le groupe HL dans la section 3.2. Puis, dans la section 3.3, nous analyserons l'influence du groupe de liquidité sur ces propriétés. À l'image du chapitre précédent, tous les résultats présentés par la suite utilisent un prix milieu de fourchette synchronisé par ajustement linéaire.

3.2 Propriétés du groupe HL

Nous étudierons successivement chacun des trois points que nous avons discutés : la pertinence des approximations du flux informationnel (sous-section 3.2.1), la distribution des rendements (3.2.2) et la dynamique des rendements dans les différentes déformations temporelles (3.2.3).

3.2.1 Approximations du flux informationnel

Comme nous l'avons déjà signalé, les mesures d'activité ne sont pas disponibles pour les groupes de liquidité. Aussi, nous présentons ici des résultats pour les titres individuels. La table 3.1 fournit les estimations du modèle (3.2) pour Sanofi avec un intervalle de 5 minutes et les

[12] On peut songer par exemple à divers critères d'information comme le Akaike, le Schwartz, ou le Bayésien.

[13] En effet, le modèle i est une version contrainte du modèle $i+1$ où les paramètres α_{i+1} et β_{i+1} sont contraints à 0.

données brutes. La désaisonnalisation des séries[14] ne modifie pas qualitativement les résultats : la persistance est légèrement diminuée par cette procédure, elle passe à 0,96 en *temps horloge* mais les évolutions de α_1 et β_1 demeurent semblables lorsque des variables sont incluses dans l'équation de h_t.

TABLE 3.1
Persistance de la volatilité
Sanofi, 5 minutes

	Calendaire	*Nb. Transactions*	*Nb. Ordres*	*Volume*	*Vol. Ordres*
μ	0,0005	-0,0009***	-0,0006	-0,0009	-0,0006
	(0,60)	(-3,91)	(-0,86)	(-1,53)	(-1,00)
α_0	0,0031***	0,0000	0,0020***	0,0029***	0,0052***
	(9,97)	(0,40)	(5,83)	(6,76)	(12,37)
α_1	0,199***	0,239***	0,278***	0,325***	0,345***
	(17,10)	(28,66)	(17,04)	(27,86)	(20,11)
β_1	0,778***	0,136***	0,106***	0,205***	0,129***
	(62,19)	(15,06)	(4,88)	(8,69)	(6,23)
δ_1		0,0036***	0,0016***	1,44E-6***	1,66E-5***
		(32,76)	(25,26)	(18,46)	(18,28)
Log Likelihood	9 795	22 158	17 784	16 530	14 798
Jarque-Béra	515 734***	536 097***	246 275***	303 091***	295 595***

Cette table donne les résultats de l'estimation du modèle (3.2) page 75 sur les séries de rendements en *temps horloge*. Une variable ou une statistique significative au seuil de 10% est indiquée par *, au seuil de 5% par ** et au seuil de 1% par ***. Les valeurs entre parenthèses en dessous des paramètres estimés sont les T-values. L'entête des colonnes donne l'approximation du flux informationnel incluse dans l'équation de la volatilité conditionnelle (δ_1).

Il y a une baisse significative de la persistance de la volatilité qui passe de 0,977 sans approximation du flux informationnel à environ 0,380 lorsqu'un nombre d'événements est inclus dans la volatilité conditionnelle et environ 0,500 lorsque c'est un volume qui est inclus. Les approximations fondées sur un nombre (de transactions ou d'ordres) apportent une plus grande réduction de la persistance : elles sont plus performantes. Ce résultat est en accord avec Lin *et al.* (1999) qui démontrent que l'utilisation du nombre de transactions induit une réduction beaucoup plus importante de la persistance de la volatilité haute fréquence que celle engendrée par le volume.

L'inclusion d'approximation du flux informationnel dans l'équation de la variance conditionnelle affecte différemment les paramètres α_1 et β_1. On a en effet une nette réduction de β_1 et une moindre augmentation de α_1. C'est à dire que l'approximation de l'information explique l'autocorrélation de la volatilité conditionnelle mais elle modifie aussi sa dynamique. Cette dernière devient moins GARCH et plus ARCH. Ces résultats sont similaires à ceux obtenus par,

[14]Cette désaisonnalisation suit la méthode adoptée par Andersen & Bollerslev (1997*a*) ou encore Maillet & Michel (1997) : les valeurs observées sont normées par leur moyenne sur l'intervalle. Les rendements sont divisés par la moyenne de la volatilité.

entre autres, Maillet & Michel (1997) ou Lamoureux & Lastrapes (1990). Les approximations fondées sur un volume sont inférieures sur chacun de ces points, la réduction de β_1 est moindre et l'augmentation de α_1 est plus importante.

Le fait que le volume et le nombre d'événements aient un impact significatif pose une question intéressante. En effet, le volume est le produit du nombre d'événements (de transactions ou d'ordres) par leur taille moyenne. On peut alors se demander si l'effet observé globalement pour le volume provient de l'une de ces composantes ou des deux simultanément. Afin d'évaluer ce point, nous décomposons le volume entre le nombre de transactions et le volume moyen des transactions dans l'intervalle, puis nous incluons simultanément ces deux variables dans l'équation de h_t. La table 3.2 donne les résultats de cette décomposition du volume des transactions pour plusieurs titres du groupe HL.

TABLE 3.2
Décomposition du volume

	Sanofi	Accord	Lyonnaise	St Gobain
μ	-0,0009***	0,0003**	0,0000	-0,0001***
	(-98,03)	(2,43)	(1,35)	(-3,63)
α_0	1,32E-8	1,47E-6	1,13E-8*	3,97E-7***
	(1,10)	(0,84)	(1,80)	(4,01)
α_1	0,240***	0,249***	0,240***	0,188***
	(29,39)	(33,15)	(29,82)	(12,49)
β_1	0,136***	0,140***	0,153***	0,069***
	(13,51)	(113,56)	(10,94)	(2,74)
δ_1	0,0036***	0,0019***	0,0011***	0,0024***
	(53,24)	(539,08)	(38,06)	(6,81)
δ_2	1,95E-7	-3,86E-7***	2,81E-6***	3,74E-5
	(0,12)	(-3576,25)	(4,38)	(0,77)

Cette table donne pour différents titres les paramètres du modèle (3.2) page 75 en décomposant le volume entre le nombre de transactions (δ_1) et le volume moyen des transactions (δ_2). Une variable significative au seuil de 10% est indiquée par *, au seuil de 5% par ** et au seuil de 1% par ***. Les valeurs entre parenthèses en dessous des paramètres estimés sont les T-values.

Le nombre de transactions est toujours très significatif au seuil de 1% ; à l'inverse le volume moyen est parfois non significatif. Cela suggère que, parfois, le contenu informationnel du volume est entièrement capturé par le nombre de transactions. Ces résultats viendraient donc corroborer ceux de Jones *et al.* (1994*b*). Toutefois, l'absence de significativité du volume moyen est instable. Comme on peut le voir dans la table 3.3, les titres pour lesquels le volume moyen n'est pas significatif changent en fonction de la longueur employée. Il est donc difficile de tirer une conclusion ferme sur le contenu informationnel du volume moyen.

La capacité à réduire la persistance de la volatilité est souvent associée à l'autocorrélation de la variable employée. La non-significativité du volume moyen pourrait être due à l'ab-

TABLE 3.3
Absence de significativité du volume moyen
groupe HL

Longueur	Titre
60	Schneider Lyonnaise
30	Lyonnaise
15	Thomson Accor Lyonnaise
5	St Gobain Sanofi Thomson

Cette table fournit, en fonction de la longueur, la liste des titres du groupe HL pour lesquels le volume moyen n'est pas significatif au seuil de 5%. La longueur est exprimée en minutes.

sence d'autocorrélation de cette variable. Nous avons donc estimé la régression suivante $A_t^a = \mu + \alpha A_{t-1}^a + \varepsilon_t$ c'est à dire un AR(1) sur les approximations du flux informationnel envisagées (nombre de transactions, volume de transactions, volume moyen ...). Toutes les variables ont une autocorrélation très significative (T-values supérieures à 50). La non-significativité du volume moyen ne peut donc pas être due à ce fait.

Nous avons également testé d'autres décompositions (e.g. le nombre de transactions par le volume, la décomposition du nombre de transactions par le nombre d'ordres, ou encore des décompositions des ordres). Celles-ci produisent à chaque fois deux coefficients significatifs. On peut donc conclure que les ordres et les transactions ont des contenus informationnels spécifiques. Le nombre de transactions a incontestablement un contenu informationnel mais celui du volume est sujet à caution. De manière générale, une approximation utilisant le nombre d'événements est donc préférable à celle employant le volume.

3.2.2 Distribution des rendements

La table 3.4 donne la distribution des rendements du groupe HL dans les différentes déformations temporelles, .

La différence de moyenne entre les temps est due à l'utilisation de la formule exacte pour calculer les rendements. En effet, avec cette formule de calcul nous n'avons pas la propriété d'additivité des logarithmes. Si les logarithmes avaient été utilisés alors toutes les moyennes auraient été identiques.

Les séries en *temps activité* ont toutes une volatilité, une kurtosis et une dispersion (étendue) plus faibles qu'en *temps calendaire*. L'effet d'une déformation temporelle sur la skewness est plus faible. Cette dernière est systématiquement inférieure en *temps activité* mais elle change parfois de signe. L'hypothèse de non-normalité est à chaque fois acceptée au seuil de 1%. Aucune des séries étudiées n'est normalement distribuée[15]. Nos données rejettent donc la modé-

[15]Ce résultat peut sembler en contradiction avec la normalité obtenue par Maillet & Michel (1997) sur Elf-

TABLE 3.4
Distribution des rendements
groupe HL, 5 minutes

	Calendaire	*Nb. Trans.*	*Nb. Ordres*	*Vol. Trans.*	*Vol. Ordres*
Moyenne	-0,00046	-0,00047	-0,00048	-0,00047	-0,00050
Écart-type	0,12	0,09	0,09	0,09	0,09
Skewness	-0,10	0,07	0,00	-0,08	-0,10
Kurtosis	11,18	5,64	6,53	7,96	7,23
Min	-1,64	-1,33	-1,29	-1,48	-1,33
Max	2,02	0,91	1,41	1,00	0,99
Kolmogorov-Smirnov	0,08***	0,05***	0,05***	0,06***	0,06***
Cramer-von Mises	300,48***	90,92***	100,80***	141,88***	126,14***
Anderson-Darling	1675,14***	557,47***	600,51***	844,82***	751,94***

Cette table donne les principales caractéristiques de la distribution des rendements en fonction de la déformation temporelle utilisée (en haut de chaque colonne). Une statistique significative au seuil de 1% est marquée par ***.

lisation proposée par la MDH (3.3). Toutefois, les statistiques obtenues en *temps activité* sont toujours nettement inférieures à celles obtenues en *temps calendaire*. Cela souligne un rapprochement de la normalité avec les séries en *temps activité*. L'hypothèse d'une subordination du rendement au flux informationnel est donc vraisemblable.

On peut remarquer que la série ayant la plus faible kurtosis et la plus faible étendue est observée en *temps nombre de transactions*. Ce fait corrobore la prééminence des approximations du flux informationnel fondées sur le nombre d'événements.

Pour mieux comprendre les effets des déformations temporelles, les figures 3.1 et 3.2 présentent les distributions[16] en *temps calendaire* et en *temps nombre de transactions*.

La saisonnalité intrajournalière de la volatilité ($|r_t|$) est nettement visible de même que l'autocorrélation positive des rendements à l'ordre 1. La densité empirique présente un pic particulièrement important autour de 0 alors que le QQ plot signale en plus l'existence de queues de distributions très épaisses.

Aquitaine. On doit cependant remarquer plusieurs différences. Tout d'abord, la longueur de l'échantillon qu'ils construisent : 6 mois de mars à août 1996 contre 4 ans dans notre étude. Ensuite, la déformation temporelle qu'ils emploient ne respecte pas la nature journalière des données (elle est calculée sur toute la base de données en une fois). Enfin, les auteurs excluent de l'analyse les rendements de la première demi-heure (période dont on a vu dans le chapitre 2 qu'elle était critique dans une déformation temporelle). Prises ensemble, ces différences peuvent aisément expliquer la distance avec nos résultats.

[16]La densité empirique est estimée par noyau. Nous employons un noyau gaussien et une fenêtre de lissage calculée par $3\sigma(1/(70\sqrt{\pi}))^{1/5}$. Cette fenêtre de lissage correspond à l'option "oversmoothed" de la procédure KDE de SAS^{TM}. La densité théorique est calculée en simulant le même nombre de données tirées d'une loi normale de même moyenne et écart-type. Ensuite nous estimons la densité théorique avec le même noyau et une fenêtre de lissage standard ($\sigma(4/(3n))^{1/5}$). Ainsi, l'étendue et les discontinuités de la densité théorique sont celles que l'on pourrait observer avec un échantillon de cette taille si les données étaient distribuées selon une une loi normale de même moyenne et écart-type.

FIGURE 3.1
Distribution des rendements
temps calendaire, groupe HL, 5 minutes

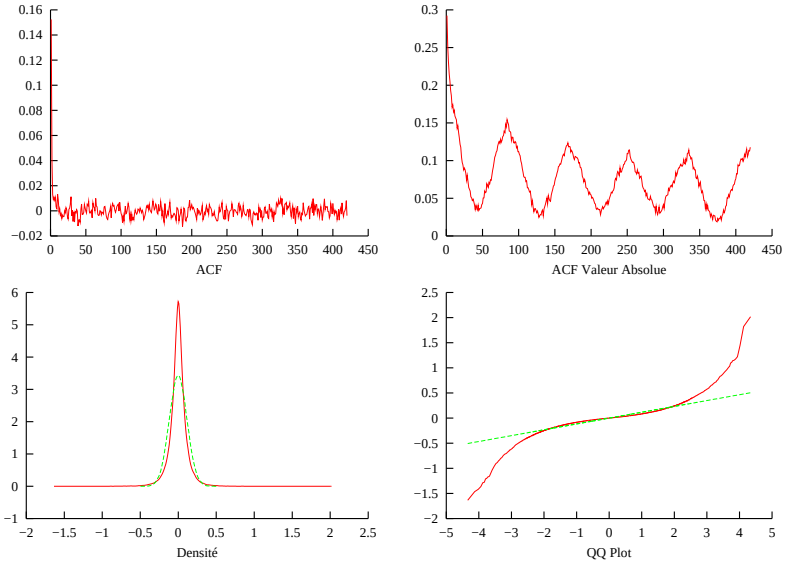
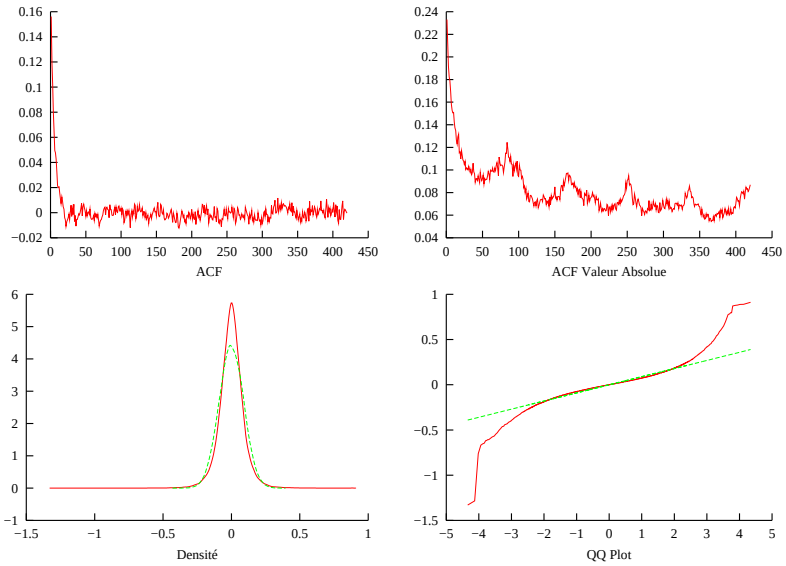


FIGURE 3.2
Distribution des rendements
temps nb. transactions, groupe HL, 5 minutes



Ces figures présentent les principales caractéristiques de la distribution des rendements du groupe HL en *temps calendaire* et en *temps nb. transactions*. Le nom de la statistique calculée est donné en dessous de chaque graphique. La courbe rouge continue est la valeur empirique de la statistique, la courbe verte hachurée est, lorsqu'elle est présente, la valeur théorique si la distribution était normale de même moyenne et écart-type.

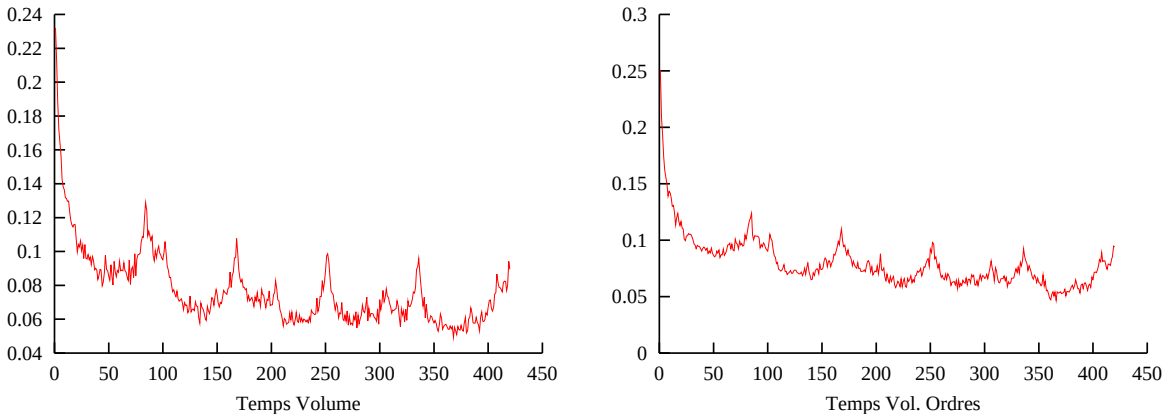
Un *temps nombre de transactions* a plusieurs effets :

– il diminue l'amplitude des queues de distributions
– il diminue l'autocorrélation de la volatilité
– il prolonge à plusieurs décalages l'autocorrélation des rendements.

La comparaison des deux densités fait ressortir un rapprochement de celles-ci autour de 0 même si le niveau ne change pas. La comparaison des QQ plot permet de comprendre la cause de cet effet : la déformation temporelle réduit l'amplitude des queues de distribution (cette réduction est plus prononcée pour les mouvements haussiers que pour les mouvements baissiers). Cela implique en retour une plus grande densité théorique autour de 0 (en raison de la moindre dispersion) qui se rapproche alors de la densité empirique.

Ces effets sont globalement observés pour toutes les déformations. Cependant, l'influence sur l'autocorrélation de la volatilité varie. En *temps nombre*, elle est similaire à celle que nous venons de décrire. La figure 3.3 donne l'autocorrélation de la volatilité pour les *temps volume*.

FIGURE 3.3
Autocorrélation de la volatilité
groupe HL, 5 minutes



Cette figure donne pour les *temps volume* la fonction d'autocorrélation de la volatilité ($|r_t|$). Le temps employé est indiqué en dessous de chaque graphique.

On voit que, pour le *temps volume de transactions*, la périodicité intrajournalière est plus marquée qu'en *temps nombre de transactions*. À l'inverse, en *temps volume d'ordres* cette périodicité est moins marquée mais la baisse de l'autocorrélation en fonction du décalage est plus lente.

L'utilisation d'une déformation temporelle utilisant une approximation de l'information fondée sur l'activité permet un rapprochement de la normalité. Ce rapprochement est principalement le fait d'une réduction des queues de distribution (tout particulièrement pour les rendements positifs) ce qui réduit alors la dispersion. Les effets que nous venons de décrire sont

globalement robustes à l'agrégation temporelle (on les observe quelle que soit la longueur employée). On doit toutefois remarquer que, plus la longueur est importante, moins la réduction d'amplitude des mouvements extrêmes en *temps activité* est prononcée.

L'examen des autres titres du groupe HL fait ressortir le même type d'effet : une réduction de l'amplitude des queues de distribution surtout pour les mouvements positifs. La longueur a un effet plus important sur les titres : - pour les petites longueurs la réduction des queues est importante mais la déformation n'a pas d'effet sur la densité empirique autour de 0 ; - lorsque la longueur s'accroît, la réduction des queues de distribution est moins prononcée, mais on observe alors une diminution de la densité empirique autour de 0. Une déformation temporelle, en éliminant les transactions n'ayant pas de contenu informationnel, induit une réduction du nombre de rendements nuls. Il y a alors un phénomène de substitution entre la réduction des queues de distribution et la réduction de la densité empirique autour de 0. Cela souligne que le flux informationnel intrajournalier est discret et qu'une déformation temporelle ne peut, avec un intervalle trop court, parvenir à des séries régulièrement espacées en *temps informationnel*.

3.2.3 Dynamique des rendements

La table 3.5 donne, pour les différentes déformations temporelles, les paramètres (3.5). À chaque fois, le nombre de paramètres du modèle est choisi par le test du ratio de vraisemblance celui-ci peut donc varier en fonction des déformations temporelles.

Le test du ratio de vraisemblance ne soutient pas l'hypothèse d'une simplification des dynamiques en *temps déformés*. En effet, dans toutes les déformations, à l'exception du *temps volume*, les modèles suggérés sont plus complexes (ont plus de paramètres) que celui obtenu en *temps calendaire*.

Toutes les déformations s'accordent sur une valeur négative de θ de l'ordre de $-0,04$. Il y a donc des effets de levier et d'asymétrie significatifs (les innovations négatives de rendement augmentent plus la volatilité que les innovations positives).

Enfin, si le test de normalité des résidus rejette toujours fortement celle-ci, on constate cependant de grandes disparités. Les valeurs obtenues en *temps ordre* sont plus importantes que celle observées en *temps calendaire* alors qu'elles sont inférieures pour les *temps transactions* (nombre ou volume de transactions). Cela suggère que l'utilisation d'une déformation fondée sur l'activité des transactions ne supprime pas la dynamique EGARCH de la volatilité mais rend cette modélisation plus en adéquation avec les données.

La diversité des ordres obtenus pour les différentes déformations temporelles et les différences de valeurs des paramètres (que ce soit en signe ou en amplitude) rendent difficile sinon impossible la mise en évidence d'une relation claire entre les déformations temporelles et la

TABLE 3.5
Dynamique des rendements
groupe HL, 5 minutes

	Calendaire	*Nb. Transactions*	*Nb. Ordres*	*Volume*	*Vol. Ordres*
μ	-0,0 000	0,0003*	0,0027***	0,0003	0,0021***
	(-0,00)	(1,82)	(11,40)	(1,28)	(8,85)
α_0	-0,199***	-0,003***	-1,177***	-0,228***	-0,808***
	(-10,85)	(-13,18)	(-34,90)	(-29,85)	(-37,57)
α_1	0,411***	0,299***	0,215***	0,264***	0,248***
	(56,08)	(78,63)	(44,79)	(54,24)	(49,24)
α_2	-0,168***	-0,098***	0,353***		0,284***
	(-7,77)	(-45,04)	(47,92)		(55,81)
α_3		-0,191***	0,250***		
		(-45,85)	(50,61)		
β_1	1,146***	1,173***	-0,627***	0,952***	-0,084***
	(18,65)	(206,33)	(-328,68)	(619,59)	(-39,44)
β_2	-0,194***	0,427***	0,451***		0,920***
	-3,39	49,44	149,59		424,81
β_3		(-0,601***)	(0,932***)		
		(-75,58)	(465,59)		
θ	-0,047***	-0,059***	-0,048***	-0,040***	-0,042***
	(-6,90)	(-8,50)	(-5,93)	(-5,56)	(-5,26)
Log Likelihood	78 747	98 146	95 729	98 805	101 379
Jarque-Béra	410 465***	11 534***	600 546***	72 382***	550 154***

Cette table donne les résultats de l'estimation d'un EGARCH sur les séries de rendements. Une variable ou une statistique significative au seuil de 10% est indiquée par *, au seuil de 5% par ** et au seuil de 1% par ***. Les valeurs entre parenthèses en dessous des paramètres estimés sont les T-values. L'entête des colonnes donne la déformation temporelle utilisée.

dynamique des rendements[17].

3.3 Influence du groupe de liquidité

De manière générale, les effets des déformations temporelles sont d'autant plus prononcés que la liquidité du groupe est faible. La table 3.6 illustre l'effet de l'inclusion d'approximation du flux informationnel dans l'équation de la volatilité conditionnelle pour un titre du groupe ML (Club Méditerranée).

Les effets déjà notés sur le groupe HL sont également observables pour ce titre du groupe ML mais de manière plus prononcée[18]. Ainsi, la réduction de β_1 et l'augmentation de α_1 sont, toutes les deux, beaucoup plus importantes. Au final, une réduction de la persistance est encore observée mais de moindre ampleur que sur le groupe HL.

[17] La comparaison des estimations d'un EGARCH(1,1), présentée en annexe E.1, suggère une légère baisse de $\alpha_1 + \beta_1$.

[18] Cette observation est encore vraie pour le groupe LL.

TABLE 3.6
Persistance de la volatilité
Club Med., 15 minutes

	Calendaire	*Nb. Transactions*	*Nb. Ordres*	*Volume*	*Vol. Ordres*
μ	-0,0007	0,0000	-0,0019	-0,0000*	-0,0032**
	(-0,34)	(1,03)	(-0,86)	(-1,83)	(-2,17)
α_0	0,0134***	0,0000*	0,0041***	0,0000	0,0065***
	(4,47)	(1,74)	(3,30)	(0,37)	(3,75)
α_1	0,315***	0,463***	0,411***	0,423***	0,429***
	(5,79)	(14,32)	(8,89)	(7,82)	(12,81)
β_1	0,651***	0,063***	0,079***	0,320***	0,176***
	(11,35)	(6,53)	(3,32)	(6,13)	(3,23)
δ_1		0,0115***	0,0045***	0,0001***	6,61E-6***
		(27,58)	(17,39)	(7,45)	(8,41)
Log Likelihood	-6 502	1 372	-2 007	-2 352	-2 162

Cette table donne les résultats de l'estimation du modèle (3.2) sur les séries de rendements en *temps horloge*. Une variable significative au seuil de 10% est indiquée par *, au seuil de 5% par ** et au seuil de 1% par ***. Les valeurs entre parenthèses en dessous des paramètres estimés sont les T-values. L'entête des colonnes donne l'approximation du flux informationnel incluse dans l'équation de la volatilité conditionnelle (δ_1). Le test de Jarque-Béra (non montré) rejette systématiquement au seuil de 1% la normalité des résidus.

Les observations que nous avions faites sur le volume moyen demeurent vraies pour les groupes ML et LL. Comme on peut le voir dans la table 3.7, il arrive souvent que le volume moyen ne soit pas significatif. Les autres décompositions que nous avions mentionnées produisent cependant des résultats trop instables pour être analysables.

TABLE 3.7
Absence de significativité du volume moyen

Longueur	Titre
	Groupe ML
60	Sagem Lapeyre Clarins
30	Club Med. Lapeyre Clarins
15	Sagem Technip Eurafrance
	Groupe LL
60	Dassault Radial Assystem Partouche Vicat Boiron
30	Vicat Boiron

Cette table donne, en fonction de la longueur, la liste des titres des groupes ML et LL lesquels le volume moyen n'est pas significatif au seuil de 5%. La longueur est exprimée en minutes.

La table 3.8 donne, pour les différentes déformations temporelles, la distribution des rendements pour le groupe ML avec un intervalle de 15 minutes.

On note les mêmes effets que ceux observés pour le groupe HL. La figure 3.4 (3.5) donne, avec un intervalle de 15 minutes, la distribution des rendements en *temps horloge* (*temps nombre*

TABLE 3.8
Distribution des rendements
groupe ML, 15 minutes

	Calendaire	*Nb. Transaction*	*Nb. Ordre*	*Volume*	Vol. Ordre
Moyenne	-0,00042	-0,00049	-0,00054	-0,00059	-0,00056
Écart-type	0,14	0,13	0,13	0,13	0,13
Skewness	0,34	0,05	-0,15	-0,01	-0,21
Kurtosis	9,38	5,80	18,62	9,17	14,17
Min	-1,56	-1,30	-2,88	-1,44	-1,95
Max	1,69	1,43	2,14	1,94	2,14
Kolmogorov-Smirnov	0,08***	0,05***	0,06***	0,07***	0,07***
Cramer-von Mises	93,97***	37,49***	52,79***	70,81***	66,89***
Anderson-Darling	522,28***	221,97***	308,64***	404,49***	385,12***

Cette table donne les principales caractéristiques de la distributions des rendements en fonction de la déformation temporelle utilisée (donnée en haut de chaque colonne). Une statistique significative au seuil de 1% est marquée par ***.

de transactions) pour le groupe ML[19].

Les effets observés sont similaires à ceux notés sur le groupe HL mais on remarque cependant quelques différences. La comparaison des deux densités fait ressortir, à la différence du groupe HL, une diminution de la densité empirique autour de 0. L'effet de la déformation temporelle sur l'autocorrélation de la volatilité est frappant. En *temps nombre de transactions*, la périodicité journalière est presque invisible alors qu'elle est manifeste en *temps horloge*. Il y a même une réduction des niveaux observés tant pour l'autocorrélation des rendements que de la volatilité. Les autres effets notés à partir du QQ plot ou de la densité restent semblables à ceux déjà décrits pour le groupe HL.

Pour ce qui est de la dynamique du rendement, la table 3.9 donne les estimations du modèle (3.2) sur le groupe ML.

On peut remarquer qu'avec le groupe ML, les déformations temporelles utilisant les transactions (les ordres) mettent en avant des modèles plus simples (complexes) qu'en *temps calendaire*. Cependant, on doit également remarquer que les différentes estimations de θ ne sont comparables ni en amplitude ni en signe. Ce paramètre peut même être non significatif (en *temps volume de transactions*). Les estimations sont donc encore plus instables que pour le groupe HL.

3.4 Conclusion

Au cours de ce chapitre nous avons évalué la pertinence d'approximations du flux informationnel en étudiant d'une part, leur effet sur la persistance de la volatilité (dans le cadre d'une

[19]Le lecteur pourra trouver en annexe E.2 les même figures pour le groupe LL.

FIGURE 3.4
Distribution des rendements
temps horloge, groupe ML, 15 minutes

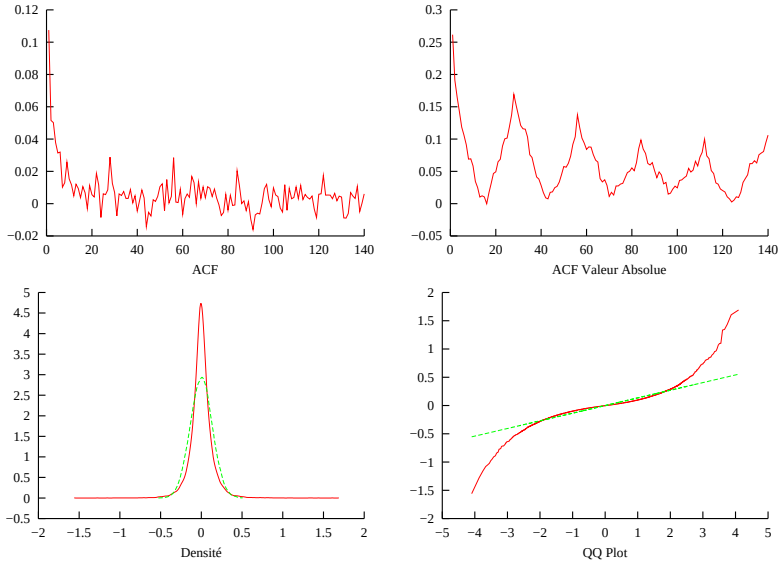
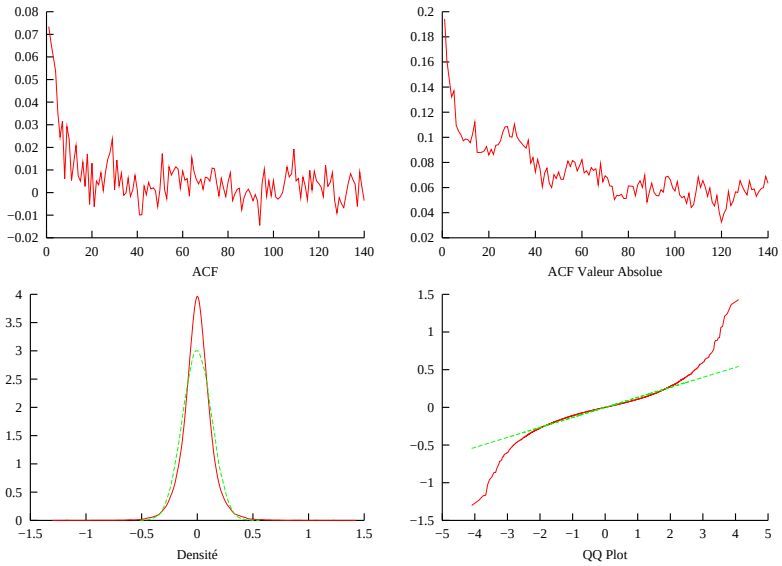


FIGURE 3.5
Distribution des rendements
temps nb. transactions, groupe ML, 15 minutes



Ces figures présentent les principales caractéristiques de la distribution des rendements du groupe ML en *temps horloge* et en *temps nb. transactions*. Le nom de la statistique calculée est donné en dessous de chaque graphique. La courbe rouge continue est la valeur empirique de la statistique, la courbe verte hachurée est, lorsqu'elle est présente, la valeur théorique si la distribution était normale de même moyenne et écart-type.

TABLE 3.9
Dynamique des rendements
groupe ML, 15 minutes

	Calendaire	*Nb. Transactions*	*Nb. Ordres*	*Volume*	*Vol. Ordres*
μ	-0,0021***	0,0005	-0,0032	-0,0012*	-0,0031***
	(-2,97)	(0,72)	(-0,03)	(-1,82)	(-6,20)
α_0	-1,384***	-0,043***	-0,000	-0,395***	-1,861***
	(-28,42)	(-7,29)	(- 0,00)	(-16,02)	(-20,96)
α_1	0,456***	0,103***	0,259***	0,300***	0,343***
	(40,94)	(17,17)	(3,48)	(28,37)	(29,54)
α_2	0,461***		-0,252***		0,282***
	(41,58)		(-3,36)		(25,89)
α_3					0,377***
					(31,24)
β_1	-0,176***	0,988***	1,605***	0,899***	-0,051***
	(-29,48)	(683,08)	(22,44)	(149,07)	(-6,95)
β_2	0,823***		-0,605***		-0,246***
	(137,67)		(-8,46)		(-37,34)
β_3					0,841***
					(114,36)
θ	0,026**	-0,105***	-0,170*	-0,004	-0,043***
	(2,11)	(-5,65)	(-1,70)	(-0,25)	(-3,10)
Log Likelihood	20 065	20 089	20 389	20 161	21 562
Jarque-Béra	96 847***	34 031***	410 103***	56 103***	213 726***

Cette table donne les résultats de l'estimation d'un EGARCH sur les séries de rendements. Une variable ou une statistique significative au seuil de 10% est indiquée par *, au seuil de 5% par ** et au seuil de 1% par ***. Les valeurs entre parenthèses en dessous des paramètres estimés sont les T-values. L'entête des colonnes donne la déformation temporelle utilisée.

modélisation GARCH) et d'autre part, la distribution et la dynamique des rendements dans les différentes déformations temporelles.

Quelle que soit la longueur de l'intervalle ou la liquidité du titre, l'inclusion d'une des quatre approximations de l'information dans l'équation de la volatilité conditionnelle d'un GARCH réduit toujours la persistance de la volatilité. En première analyse nous avons noté que les approximations utilisant le nombre d'événements sont préférables à celles employant le volume. Nous avons décomposé le volume de transaction afin de vérifier si celui-ci a un contenu informationnel distinct du nombre de transactions. Cette décomposition fait ressortir que le nombre de transactions est toujours significatif alors qu'il arrive régulièrement que le volume moyen ne le soit pas. Nous rejoignons donc, avec quelques réserves, les conclusions de Jones *et al.* (1994*a*), Lin *et al.* (1999) ou encore de Ané & Geman (2000) qui mettent en évidence le contenu informationnel du nombre de transactions aux dépens du volume.

L'analyse de la distribution des rendements dans les différentes déformations temporelles souligne que toutes les déformations modifient la distribution et la dynamique des processus. L'utilisation d'une déformation temporelle, dans notre étude et avec notre méthodologie, ne

permet jamais d'obtenir des séries normalement distribuées. Toutefois, tous les *temps activité* provoquent un rapprochement vers la normalité. On rejette donc la modélisation de la MDH sans toutefois rejeter l'hypothèse d'une subordination des rendements au flux informationnel. Les effets notables des déformations temporelles sont une réduction de la dispersion et surtout de la kurtosis des séries (ce qui était la principale prédiction de la MDH). Si on considère le critère de réduction de la kurtosis, la meilleure déformation temporelle est le *temps nombre de transactions*. Cela renforce donc l'observation faite à partir de la réduction de la persistance de la volatilité.

Au regard des résultats présentés dans ce chapitre, le classement des approximations diffère selon la liquidité du groupe :

- **groupe HL :** nombre de transactions, nombre d'ordres, volume de transactions et volume d'ordres.
- **groupe ML et LL :** nombre de transactions, volume de transactions, nombre d'ordres et volume d'ordres.

Nous avons aussi souligné l'influence de la liquidité sur la performance des déformations temporelles. Plus la liquidité du groupe est faible, plus l'effet des déformations est important. L'influence des déformations subit un effet de substitution. La réduction de la dispersion est plus prononcée pour les petites longueurs alors que la baisse du nombre de rendements nuls est plus importante pour les longueurs élevées. Ces faits suggèrent que le processus informationnel est discontinu et que si la longueur utilisée est trop courte alors la déformation temporelle ne permet pas de construire des séries espacées régulièrement suivant la quantité d'information diffusée.

Enfin, la dynamique du processus de rendement (dans une modélisation EGARCH) est souvent plus complexe en *temps déformés* qu'en *temps calendaire*. Ce fait suggère que notre hypothèse d'une relation linéaire entre le flux informationnel et les approximations est sans doute trop simplificatrice.

Conclusion de la partie I

Au cours de cette partie nous avons évalué la pertinence de quatre approximations du flux informationnel (le nombre de transactions, le volume de transactions, le nombre d'ordres et le volumes d'ordres). Nous avons à chaque fois postulé une relation linéaire constante dans le temps entre ces approximations et le flux informationnel.

L'analyse a été réalisée sur un échantillon sélectionné de manière à minimiser les variations exogènes de l'activité (stabilité de l'appartenance indicielle et secteurs économiques identiques). Nous avons utilisé l'outil des déformations temporelle afin de déformer le temps selon les approximations du flux informationnel étudiées. Nous avons ensuite analysé plusieurs propriétés des séries construites dans ces différentes déformations temporelles. Nous avons, dans le chapitre 2, analysé les configurations intrajournalières dans les différentes déformations temporelles. Ensuite, dans le chapitre 3, nous avons comparé les propriétés statistiques et la dynamique des séries de rendements dans ces déformations temporelles.

Nous avons également apprécié l'effet des changements de microstructure opérés pendant la période (généralisation du fixage de clôture, extension des horaires de cotation, allongement de la pré-ouverture) sur les configurations. Sur une période longue (de début 1997 à fin 2000), nous avons observé que les configurations de rendement étaient particulièrement instables et peu significatives. Elles constituent au mieux une spécificité des périodes étudiées ou traduisent l'existence de frictions systématiques dans le processus de transactions (tout particulièrement si on utilise un prix de transaction). À l'inverse, les configurations de volatilité sont stables et se caractérisent par une plus forte volatilité à l'ouverture et à la clôture (l'amplitude respective de ces deux hausses varie en fonction de la liquidité des titres). La généralisation du fixage de clôture a entraîné une réduction de l'effet de clôture alors que l'extension de la pré-ouverture n'a pas eu d'effet notable. Enfin, le résultat de l'extension des horaires est un simple déplacement des configurations sans autre modification. Aucune modification des configurations n'est notée en fonction des jours de la semaine. Ces résultats éclairent les causes des configurations.

Les configurations semblent intrinsèquement dues à l'existence de périodes de fermeture. La longueur de ces dernières n'ayant qu'une faible importance[20] puisque l'on observe les même

[20]Cela n'entre pas en contradiction avec les études mettant en évidence la moindre volatilité du fixage d'après-midi à la Bourse de Tokyo. Dans ce dernier cas, la fermeture de mi-journée n'est que de quelques heures et diffère

configurations pour les différents jours de la semaine ou après l'extension des horaires de cotations. Toutefois, cette relation n'exclut pas que le phénomène soit causé des variations du flux informationnel ou encore par des variations des frictions du processus d'échange puisque l'existence de fermeture peut, théoriquement, affecter ces deux paramètres.

L'analyse des configurations intrajournalières a permis de mettre en évidence une influence significative des différentes déformations temporelles construites. Leur impact est assez proche et consiste principalement à réduire l'effet de clôture alors que l'effet d'ouverture est, globalement, préservé. Même si les différences entre les *temps activité* sont faibles, le nombre de transactions est préférable. En effet, c'est cette déformation qui entraîne la plus forte réduction des configurations.

L'étude des distributions montre que la normalité est systématiquement rejetée. Toutefois, toutes les déformations temporelles induisent un rapprochement des distributions empiriques vers la loi normale. Ce rapprochement, observé pour tous les titres et toutes les longueurs, est causé par la diminution des queues de distributions (surtout lorsque la longueur de l'intervalle est faible) et par la moindre occurrence de rendements nuls (principalement lorsque la longueur est plus importante). La substitution entre ces deux effets souligne le caractère probablement discret du processus informationnel. La réduction de la kurtosis des séries est très nette. La réduction la plus importante est observée en *temps nombre de transactions*.

Dans le chapitre 3, nous avons également évalué la capacité des différentes approximations à réduire la persistance de la volatilité. Les quatre approximations entraînent toutes une forte réduction de la persistance. Celle-ci est la plus importante avec le nombre de transactions. De plus, si on décompose le volume entre le nombre de transactions et le volume moyen, il arrive que ce dernier ne soit pas significatif. Cette analyse, tout en confirmant l'intérêt des variables que nous avons étudiées met donc, elle aussi, en avant le nombre de transactions.

Il y a donc une convergence des résultats des différentes études. Elles reconnaissent toutes la significativité des variables étudiées et concluent à chaque fois que le nombre de transactions est l'approximation du flux informationnel à privilégier.

Nous avons également noté certaines limites. Tout d'abord, aucune des approximations étudiées ne s'approche des critère d'acceptation (disparition des configurations, distribution normale des rendements). Cela suggère que notre hypothèse d'une relation linéaire avec l'information est sans doute trop contraignante. De plus, l'analyse des distributions des rendements dans les différents temps met en évidence l'influence importante de la procédure d'agrégation. Loin d'être neutre, cette dernière peut modifier la significativité d'une variable ou altérer un effet. Une étude haute fréquence ne semble donc pas être la panacée lorsque l'on souhaite étudier les relations informationnelles. Enfin, on a noté que les différentes approximations entretenait

d'une fermeture pendant la nuit ou le week-end. Nos résultats montrent que la longueur de la "nuit" du marché n'influence pas les configurations.

des relations proches mais néanmoins distinctes au flux informationnel. Il semble dès lors plus profitable de tirer partie de cette diversité en essayant de comprendre comment ces variables interagissent.

La prochaine partie tentera de lever les limites que nous venons de noter. Cette étude sera ultra haute fréquence, elle s'affranchira donc des biais d'agrégation. Qui plus est, elle permettra d'étudier de manière plus simple l'effet des périodes sans transactions sur le processus de rendements. Les hypothèses de travail seront moins contraignantes. Tout particulièrement, nous intégrerons la possibilité de relations non linéaires. Enfin, cette étude utilisera simultanément les différentes variables. Nous essaierons donc de comprendre comment elles interagissent et révèlent à leurs manières des aspects différents de l'information. L'analyse sera donc beaucoup plus riche que le cadre univarié - linéaire - haute fréquence que nous avons employé jusqu'ici.

Deuxième partie

Analyse "trade by trade" :
Le temps retrouvé

Introduction

Dans la partie I, nous avons évalué des approximations du flux informationnel en utilisant l'outil des déformations temporelles. Cet outil nous a permis non seulement de démontrer la pertinence des approximations proposées, de les classer (de la meilleure à la moins bonne : le nombre de transactions, le volume de transaction, le nombre d'ordres et le volume des ordres) mais aussi d'apprécier les limites de l'approche utilisée.

Nous avons observé que les effets des déformations temporelles changeaient selon les titres ou la mesure d'activité. Cela prouve que chacune de ces mesures a une relation spécifique avec le processus de rendements et que celle-ci varie en fonction des titres. Nous avons également noté qu'aucune des approximations ne s'approchait des critères d'acceptation (disparition des configurations intrajournalières, non rejet de l'hypothèse de normalité). L'analyse critique de la méthodologie fait alors apparaître trois limites principales.

i.) La relation entre le flux informationnel et les approximations peut varier dans le temps. Par exemple, l'effet différentiel de la déformation temps nombre de transactions sur les configurations d'ouverture (maintien) et de clôture (disparition) prouve que la relation entre volatilité et nombre n'est pas constante dans le temps et/ou est non linéaire.

ii.) Les approximations proposées étaient unidimensionnelles. Or, il semble naturel de considérer que le flux informationnel se révèle via plusieurs variables. Il serait préférable que la modélisation utilise simultanément ces sources d'informations.

iii.) L'agrégation des données sur un intervalle, fut-il de longueur variable, peut altérer la perception des phénomènes réels (e.g. les configurations d'ouverture et de clôture semblent identiques en temps horloge alors qu'elles ne le sont pas).

L'analyse souffre également du fait que les déformations temporelles ne sont, avec notre méthodologie, calculables qu'ex-post (après la fin de la journée). Nous utilisons donc une information qui n'est pas à la disposition de l'investisseur lors de la prise de décision. Nous adopterons, dans cette partie, une méthodologie tenant compte de ces enseignements.

Si on considère l'information diffusée en temps réel par le marché parisien, l'investisseur dispose de deux types des données :

i.) L'historique des transactions passées

ii.) L'état des meilleures limites [21]

L'information exacte dont dispose l'investisseur varie en fonction de sa qualité. S'il est membre du marché, il connaît de l'état complet du carnet. S'il ne l'est pas, il ne dispose que des 5 meilleures limites à l'achat et à la vente[22].

Le processus d'apprentissage peut se réaliser à partir de l'ensemble de ces informations. C'est à dire que la réaction à un événement donné peut varier en fonction de l'ensemble des historiques disponibles. Pour comprendre l'interprétation que les investisseurs font des différentes informations dont ils disposent il est alors nécessaire de se demander comment ces variables, toutes observables, interagissent. Si en plus on considère qu'une agrégation a des effets difficilement estimables alors on est amené à n'étudier que des données ultra haute fréquence (i.e. l'historique intégral des transactions et des meilleurs limites)[23]. L'information pouvant être véhiculée par plusieurs variables il importe en plus que le modèle soit simultané. Enfin, la modélisation adoptée devra tenir compte des différentes propriétés statistiques déjà mises en évidence.

La mise à disposition par la SBF de l'historique des 5 meilleures limites nous offre l'opportunité d'étudier les interactions entre les transactions et le carnet d'ordres. Notre objectif est de comprendre comment les variables de microstructure interagissent dans le processus d'apprentissage et, de là, comment elles peuvent être utilisées pour approcher le flux informationnel. Nous serons donc amenés à utiliser des modèles économétriques afin d'analyser et de comprendre les interactions entre ces variables de microstructure.

Dans un souci autant méthodologique que pédagogique nous adopterons une démarche progressive aussi bien pour les variables considérées que pour la complexité des modèles. Le chapitre 4 présentera les fondements théoriques[24] du contenu informationnel des variables employées et nous exposerons ensuite les aspects méthodologiques du pré-traitement d'une base de données Ultra Haute Fréquence (UHF). Le chapitre 5 analysera la dynamique du processus de transactions en passant progressivement de modélisations univariées à multivariées. Ensuite, nous étendrons cette modélisation en y incorporant les meilleures limites dans le chapitre 6.

[21] L'internaute pourra par exemple trouver ce type d'information sur des sites comme `http://www.boursorama.com/`.

[22] Les données peuvent également être diffusées avec un certain retard (généralement de 15 minutes). Toutefois cela n'a rien de systématique et dépend des accords entre le fournisseur de données, l'entreprise et le client. Nous considérerons donc toujours une information en temps réel.

[23] L'utilisation de ces type de données permet, en outre, de mieux étudier comment la baisse de l'intensité des transactions affecte les rendements (bonne ou mauvaise nouvelle ?).

[24] Pour rappel, les fondements empiriques de cette relation ont été présentés dans la sous-section 1.1.2.

Chapitre 4

Construction d'approximations de l'information

L'hypothèse principale supportant les études empiriques réalisées dans cette partie est que le marché réalise un apprentissage. En observant les événements se déroulant sur le marché les agents sont capables d'inférer le prix vrai de l'actif. Cette notion d'apprentissage ne se réduit pas au simple cas d'asymétrie d'information puisqu'il peut aussi résoudre une hétérogénéité des valorisations.

On s'intéresse donc à l'un des objectifs principaux de la microstructure qui est de comprendre les mécanismes par lesquels l'information est intégrée dans les prix, c'est à dire comment les interactions entre des agents et une structure de marché permet la découverte de la valeur vraie. Nous essaierons de contribuer à une réponse en apportant un éclairage empirique sur cette question dans le marché parisien.

La diffusion grandissante des données boursières a permis d'analyser cette problématique avec une précision accrue. On arrive aujourd'hui à une limite puisqu'il est possible de disposer de tous les événements, ce que Engle (2000) dénomme "ultra-high frequency data". Ce type de données peut être vu comme un "mark point process" : un processus dont les instants de réalisation (les points) sont aléatoires, les marques étant alors les caractéristiques de ces réalisations.

Pour analyser ces données espacées irrégulièrement, on peut (comme dans la partie I) les agréger sur un intervalle. Toutefois, la perte d'information inhérente à cette procédure est difficilement estimable[1]. Il semble donc plus pertinent de prendre en compte la nature irrégulière de ces données lors de l'analyse. On est alors confronté à plusieurs nouveaux problèmes dont la méthode économétrique, les pré-traitements à effectuer ou encore la quantité de données. En effet, l'analyse de données espacées irrégulièrement requiert des outils spécifiques. De plus,

[1] Par exemple, nous avons vu dans le chapitre 3 que la capacité du volume moyen à apporter une information en plus du nombre de transactions dépendait de la longueur de l'intervalle. Cela souligne que la perte d'information liée à l'agrégation des données par intervalle influence les résultats.

la nature désagrégée de ces données les rend d'autant plus sensibles à des phénomènes d'origine microstructurelle et à des bruits qu'il convient de prendre en compte. Enfin, le nombre d'informations disponibles contraint à faire sélectionner les données analysées.

Nous présenterons, dans la section 4.1 des modélisations microstructurelles qui lient les caractéristiques des transactions et du carnet à l'information. Ensuite, la section 4.2 aura deux objectifs. Elle discutera les différents traitements indispensables à une analyse Ultra Haute Fréquence puis elle proposera la construction de variables informationnelles à partir du carnet d'ordres. Enfin, la section 4.3, analysera les caractéristiques des transactions en comparant les données brutes et les données après les retraitements.

4.1 Modélisations de microstructure

L'hypothèse principale qui soutient l'analyse réalisée dans cette partie est que le marché permet de réaliser un apprentissage. C'est à dire que l'existence et le fonctionnement d'un marché boursier permet aux investisseurs d'apprendre un ou plusieurs paramètres affectant la valeur du titre. La littérature nous fournit deux grands paradigmes liés à ce phénomène :

– *L'asymétrie d'information* postule qu'un ou plusieurs agents informés traitent avec des agents non informés et étudie alors comment l'information est transférée ou quelles en sont les conséquences sur le comportement des différents agents. Bagehot (1971) le premier a considéré un scénario avec des investisseurs disposant d'informations privées. Selon lui, le spécialiste ne possède que l'information publique et est confronté à deux types d'investisseurs qu'il ne peut distinguer : les informés et les non-informés. Les non-informés sont également appelés *liquidity traders* puisque leurs transactions résultent de besoins de consommation, de stratégies de portefeuilles, de règles de transaction spécifiques ou, reflètent simplement leur sensibilité personnelle au prix (Easley & O'Hara, 1992). Le teneur de marché perd systématiquement lorsqu'il traite avec des informés. Il fixe alors des prix à l'achat et à la vente qui lui permettent d'égaliser les profits qu'il réalise auprès des non-informés et les pertes qu'il enregistre avec les informés. Ces éléments constitutifs des modèles d'asymétrie d'information ont par la suite été analysés par Copeland & Galai (1983) et plus tard formalisés par Kyle (1985), Glosten & Milgrom (1985), Easley & O'Hara (1987), Admati & Pfleiderer (1988) et Foster & Viswanathan (1990, 1994).

  Un élément clef dans les modèles d'asymétrie d'information est que les transactions véhiculent une information. Le spécialiste en observant le flux de transactions infère graduellement l'information détenue par les informés de telle manière qu'à chaque point dans le temps le prix est l'espérance de la valeur terminale du titre conditionnellement à l'information publique (qui inclut les actions des informés). Les prix ont alors une forme

d'efficience semi-forte[2]. Cela implique que les prix sont des martingales et qu'il y a, au sens stricte du terme, apprentissage d'un nouveau prix.

Les modèles initiaux n'accordaient que peu d'importance au temps ou au volume dans ce processus d'apprentissage. Ainsi, dans le modèle de Kyle (1985), les ordres sont accumulés puis confortés dans un processus de type fixage. Par conséquent, l'arrivée des ordres ne peut avoir d'influence et seul le déséquilibre entre les volumes à l'achat et à la vente affecte le prix. Dans le modèle de Glosten & Milgrom (1985), les ordres arrivent selon une loi de probabilité exogène connue du spécialiste. Dans ce dernier modèle ou encore dans celui de Copeland & Galai (1983), le volume est contraint à 1 et n'a donc aucun contenu informationnel. Comme démontré par Grossman & Stiglitz (1980), il est impossible que les marché parviennent à une efficience forte car elle impliquerait une contradiction logique. De même, pour qu'un autre paramètre que les prix ait un contenu informationnel sans que l'on soit confronté à ce type de problème, il faut également ajouter une autre source d'incertitude. Ce nouveau paramètre peut alors fournir des indications sur cette nouvelle source d'incertitude.

– *L'hétérogénéité* considère que la valeur est dès le départ correcte et qu'aucun agent n'a d'information supérieure sur le titre. C'est la différence des valorisations (l'hétérogénéité) qui, jusqu'à la découverte de la valorisation commune, génère les transactions. Cette hétérogénéité peut provenir de plusieurs sources. On peut considérer que deux groupes d'investisseurs ont des valorisations a priori différentes, qu'ils interprètent différemment des signaux identiques[3] Ou encore, qu'elle est générée par des opportunités individuelles d'investissement dans la sphère non financière (Wang, 1994). Il n'y a donc pas à proprement parler d'apprentissage. On peut toutefois remarquer que le marché converge vers un prix : il apprend un paramètre affectant la dispersion des prix.

On peut alors rapprocher ces deux concepts de caractéristiques du flux informationnel. L'asymétrie d'information représente l'occurrence d'un signal non distribué identiquement entre les agents alors que l'hétérogénéité représente le cas d'un signal bruité mais observable par tous les agents. Ces deux concepts s'apparentent à certaines combinaisons de paramètres statistiques[4]. Par exemple, un signal de moyenne nulle et de forte dispersion pourrait être interprété comme la preuve d'une faible asymétrie et d'une forte hétérogénéité.

[2]Cette propriété est en accord avec Grossman & Stiglitz (1980) qui démontrent l'impossibilité d'efficience forte des marchés.

[3]À l'évidence, il peut être observationnellement équivalent de considérer un signal sans bruits mal interprété ou un signal bruité bien interprété : c'est uniquement la résultante de l'interprétation que nous sommes capables d'étudier à partir de données boursières. Les signaux concernés peuvent être privés (Wang, 1998) ou publics (Harris & Raviv, 1993).

[4]À titre d'exemple, on peut songer à sa valeur absolue (la force du signal), son signe (le sens du signal) ou sa dispersion, (le bruit du signal).

La relation exacte entre ces paradigmes et les caractéristiques du carnet ou des transactions est cependant trouble. En effet, selon les modèles considérés, les conclusions peuvent être différentes et les évidences empiriques sont sinon contradictoires du moins complexes. De plus, de nombreuses études ne se focalisent que sur un aspect de la relation (i.e. l'effet de l'asymétrie sur les prix, le contenu informationnel du nombre de transactions, de la profondeur ou du volume…). Pour plus de clarté, nous présenterons dans les sous-sections les différents modèles groupés selon que l'aspect étudié porte sur les caractéristiques des transactions (sous-section 4.1.1) ou sur le carnet d'ordres (sous-section 4.1.2). Nous discuterons ensuite dans la sous-section 4.1.3 les modélisations et les études qui traitent explicitement de marchés dirigés par les ordres. La problématique du placement d'ordres nous semblant particulière, elle ne sera traitée que dans cette sous-section.

4.1.1 Caractéristiques des transactions

Les modèles séquentiels proposés par Easley & O'Hara (1987), Easley & O'Hara (1992), Easley *et al.* (1997*a*) sont parmi les premiers à reconnaître un rôle informationnel aux caractéristiques des transactions. Pour l'essentiel ces modélisations supposent toutes une incertitude sur l'occurrence d'une information : au début de chaque journée il peut y avoir une bonne, une mauvaise nouvelle ou pas de nouvelle du tout. La variable étudiée (le volume ou le nombre de transactions) permet d'inférer de l'information sur l'évènement survenu. Même si ces modèles mettent en avant des relations positives entre l'occurrence d'une information et les deux variables d'activité (le volume et l'intensité des transactions) celles-ci reposent cependant sur des mécanismes différents. En effet, la corrélation positive entre l'intensité des transactions et l'occurrence d'une information repose sur le fait que les informés ne traitent que lorsqu'il y a une information. Une faible fréquence de transaction diminue la vraisemblance d'un événement informationnel en début de journée. La relation positive avec le volume viendrait de ce que les informés préfèrent les grosses quantités.

Les différents résultats empiriques (Easley & O'Hara (1992), Easley *et al.* (1996), Easley *et al.* (1997*a*) et Easley *et al.* (1997*b*)) soulignent d'une part l'importance de l'asymétrie d'information et le caractère non markovien du processus de prix. En outre, le contenu informationnel du temps entre les transactions est toujours mis en valeur alors que la contribution du volume est incertaine. On doit noter que la plus grande partie de ces résultats souffrent d'un problème indisponibilité des données. En effet, ces modèles requièrent que les transactions soient signées en fonction de leur origine mais les auteurs ne disposent que des prix de transaction. Ils emploient donc un algorithme de signature. Or, il a été démontré ultérieurement que ces procédures induisaient des biais dans la classification (le lecteur peut se reporter à la discussion de ce point dans Minguet (2003) pages 126-127). On doit en plus noter que ces études reposent sur l'agré-

gation des données sur un intervalle dont on a vu dans les précédentes études qu'elle affectait de manière importante la significativité du volume dans ce type de relation.

En outre, de nombreux résultats empiriques soulignent une relation de type quadratique entre le volume et de nombreuses caractéristiques des transactions (impact prix des transactions, composante asymétrie d'information du spread, ou participation aux changements de prix...). Pour n'illustrer que ce dernier point Barclay & Warner (1993), Chakravarty (2001) ou encore Huang & Masulis (2003) démontrent tous les trois que les transactions de taille moyenne représentent une part disproportionnée des changements de prix. Cela les pousse à émettre l'hypothèse que les informés camouflent leurs transactions en les fractionnant. La non-linéarité de cette relation pourrait être à l'origine du moindre support empirique apporté au volume dans les modèles ci-dessus.

L'une des restrictions déterminante dans les précédents modèles est qu'il n'y a pas de contrainte sur les ventes à découvert. En effet, lorsque qu'une contrainte interdisant cette pratique est ajoutée, cela modifie la relation entre l'information et l'activité des transactions. Une telle contrainte empêche les investisseurs informés de vendre le titre s'il ne le possédaient pas auparavant. Une baisse de l'intensité des transactions peut alors être attribuée à une mauvaise nouvelle ou à l'absence de nouvelle. Diamond & Verrecchia (1987) ne considérant que deux évènements au début de chaque journée (bon ou mauvais) concluent à une association entre la diminution de l'intensité des transactions et les mauvaises nouvelles soit alors des rendements négatifs. On pourrait aussi songer que les fortes intensités de transactions signalent les bonnes nouvelles et qu'elles sont associées à des rendements positifs.

L'un des points communs des deux modèles précédents est qu'ils ignorent l'influence potentielle d'agents ayant un comportement stratégique. Les investisseurs de liquidité peuvent éventuellement choisir leur moment d'intervention. L'existence de cette liquidité flottante peut affecter significativement les relations évoquées plus haut si le comportement de ce flottant est lié à des facteurs informationnels. Or, Admati & Pfleiderer (1988) proposent ce type de modélisation avec trois types d'agents : des informés, des non-informés discrétionnaires et non-discrétionnaires et démontrent que dans ce cadre il y a un phénomène de concentration des investisseurs :

"However, despite the complexity of the strategic interaction among traders, our analysis shows that whenever an equilibrium exists, it is characterized by the concentration of liquidity and informed trading and by the resulting patterns in volume and price behavior. The actual timing and shape of trading patterns in financial markets are determined by a number of factors and parameters that are exogenous to the model, such as the rate of arrival of public information, the amount of nondiscretionary liquidity trading, and the length of the interval within which each discretionary liquidity trader trades."

L'interprétation de ce modèle est délicate et repose grandement sur les hypothèses concer-

nant l'accès à l'information privée (exogène vs endogène). Si le nombre d'informés est fixe (l'acquisition d'information est exogène) ni la volatilité ni la quantité d'information révélée par les prix ne change lors des périodes de concentration de l'activité. Si l'acquisition d'information privée est endogène, plus d'agents choisissent de devenir informés[5] quand l'intensité de transaction est importante et le prix devient plus informatif à ces périodes. Les différentes implications testables sont :

1. Le coefficient de profondeur du marché (le λ de Kyle) est plus important quand le volume est faible.
2. Les prix sont plus informatifs lorsque le volume est élevé.
3. En considérant deux états (H haut, B bas) la relation entre les variances mesurées devrait être la suivante : $V_{B\dashrightarrow H} \geqslant V_{B\dashrightarrow B} \geqslant V_{H\dashrightarrow B}$

On doit noter que dans les versions les plus élaborées de ce modèle l'occurrence d'une activité de liquidité concentrée modifie le flux d'informations privées. Les auteurs soulignent par ailleurs que plus la concurrence entre les investisseurs informés est importante, plus ce phénomène de concentration devrait être grand et moins la sélection adverse devrait être sévère pour les non-informés.

Dans le précédent modèle, l'information acquise par les informés est révélée à la fin de la période. Les informés n'ont donc que peu de choix quand à la stratégie à suivre. Cette limite est levée par Back & Pedersen (1998) qui proposent un modèle où la durée de vie de l'information est longue. On doit alors remarquer qu'il y a une différence entre l'instant d'arrivée de l'information auprès des informés et le moment où le marché la reçoit (où l'informé l'utilise). L'instant d'arrivée auprès des informés n'affecte pas les dynamiques observées. À l'instar du précédent modèle, les auteurs mettent en évidence les évolutions conjointes de la volatilité et de la liquidité. La différence majeure est qu'alors, l'impact prix des ordres ne varie pas. Ce modèle considère le cas d'un informé monopoliste. À l'opposé, si on considère une concurrence entre des informés[6] ou si les informés sont averses au risque, l'équilibre pourrait être compatible avec des variations systématiques de la profondeur.

Les différents modèles convergent donc vers une association globale entre la volatilité et la liquidité. Toutefois ils ne s'accordent ni sur la cause exacte des phénomènes ni sur leur nature précise (i.e. les variations de l'impact prix des transactions).

[5] De manière plus réaliste, on pourrait aussi songer que des agents déjà informés décident d'intervenir à ce moment.

[6] La concurrence entre les informés est étudiée en autres par Holden & Subrahmanyam (1992), Foster & Viswanathan (1996) en temps discret ou Back *et al.* (2000) en temps continu. Ces études admettent que la compétition entre plusieurs agents informés peut améliorer ou détériorer l'efficience du marché tout en soulignant que la profondeur doit avoir des variations systématiques et que la liquidité du marché doit être très réduite avant diffusion publique d'une information.

L'hétérogénéité propose également une interprétation de la relation entre le volume et l'information. D'une manière générale, le courant "hétérogénéité" conclut à une relation positive entre l'hétérogénéité et l'activité (plus particulièrement le volume).

Harris & Raviv (1993) proposent un modèle de différence d'opinion où les transactions sont générées par l'hétérogénéité entre les investisseurs[7]. Les croyances préalables et l'information reçue sont identiques pour tous mais les investisseurs diffèrent dans leurs interprétations des annonces. Tous les investisseurs sont rationnels mais étant donné que cette rationalité n'appartient pas à la connaissance commune alors le seul comportement rationnel devient une maximisation. Ce modèle permet de reproduire beaucoup des régularités empiriquement constatées : corrélation positive entre la volatilité et le volume, autocorrélation négative (positive) des prix (des volumes). Bien que ces résultats soient dérivés en utilisant une forme paramétrique spécifique, les propriétés centrales du modèle, à savoir une plus grande sensibilité au flux informationnel futur lorsque les croyances sont diffuses et une plus forte probabilité de transaction non informée lorsque l'hétérogénéité est élevée, semblent particulièrement robustes. Dans cette modélisation, le volume traduisant l'absence de consensus initial entre les investisseurs, un volume plus important signale une liquidité accrue tant des transactions (volume de transactions) que du carnet (profondeur cotée)[8].

Les modèles He & Wang (1995), Wang (1994), Wang (1998) s'inscrivent tous dans une optique d'anticipation rationnelle. Ce dernier modèle envisage à la fois une asymétrie d'information et une hétérogénéité (des croyance préalables[9]). Il démontre que sous ces hypothèses les transactions de liquidité (exploitant l'hétérogénéité) disparaissent à la limite, la vitesse de convergence étant indépendante de la quantité de liquidité enfin, l'informé traite peu. L'équilibre atteint est séparant puisque les transactions informationnelles dominent les premières périodes alors que les transactions dues à l'hétérogénéité dominent les dernières. Ce modèle prévoit une corrélation entre le volume et volatilité ainsi qu'une autocorrélation positive des volumes même si le flux informationnel ne l'est pas. L'étude empirique menée par Brockman & Chung (2000) sur la bourse de Hong-Kong sépare les échanges selon qu'ils sont d'origine informationnelle ou non informationnelle à l'aide d'une extension du modèle de Easley *et al.* (1996). Ils étudient ensuite les relations entre l'asymétrie d'information, l'hétérogénéité et les caractéristiques des transactions. Ils mettent en évidence la pertinence du concept d'hétérogénéité et soulignent globalement l'accord entre leurs résultats et le modèle de Wang (1998). Tout particulièrement, l'effet de clôture viendrait principalement de la saisonnalité de l'hétérogénéité alors que l'effet

[7]Kandel & M.Pearson (1995) proposent un modèle légèrement différent où les croyances préalables et les fonctions de vraisemblance sont toutes les deux différentes. Le principal changement par rapport au modèle de Harris & Raviv (1993) est que le volume peut être positif même si le prix ne change pas.

[8]Une relation similaire est mise en évidence par Grundy & Mc Nichols (1989), Shalen (1993), He & Wang (1995), Biais & Bossaerts (1998) qui suggèrent tous une relation positive entre d'une part l'hétérogénéité et d'autre part le volume et la volatilité.

[9]En comparaison, le modèle de He & Wang (1995) se focalise sur une hétérogénéité de l'information privée.

de l'asymétrique d'information serait globalement constant dans la journée.

Les différents modèles cités auparavant[10] s'accordent pour reconnaître une relation positive entre l'hétérogénéité et l'activité des transactions. On doit tout d'abord remarquer que cette interprétation induit une relation positive entre le volume et la liquidité du marché[11] alors que les modèles de type Easley & O'Hara (1992) s'orientent vers une relation négative entre la profondeur et le volume[12] étant donné que ce dernier traduit la présence d'asymétrie d'information et donc une augmentation du risque informationnel. Il y a donc une contradiction, au moins partielle, entre deux optiques d'analyse ayant toutes les deux réuni des supports empiriques importants.

Pour réconcilier ces résultats il faut faire appel à plusieurs faits. L'hétérogénéité, même si elle constitue le motif de transaction de certains agents, peut également en conduire d'autres (averses au risque) à réduire leurs opérations comme le suggèrent Domowitz *et al.* (2000) en étudiant la relation entre liquidité et volatilité[13].

La production de résultats empiriques sur cette relation souffre, à l'heure actuelle, de l'absence de consensus sur les approximations de l'hétérogénéité à utiliser. Si un certain accord se dégage autour de variables externes aux marchés financiers comme les prévisions des analystes pour Bamber *et al.* (1999) ou encore Ghysels & Juergens (2001), l'unanimité est loin d'être atteinte à propos des variables de marché à employer.

Deux grandes mesures ont émergé : le volume de transactions et le déséquilibre des ordres (i.e. de la position ouverte)[14].

L'étude de Zhang *et al.* (2000) peut permettre de réconcilier des modèles théoriques et des résultats empiriques conflictuels sur la relation volume - hétérogénéité - information. Ils analysent empiriquement les déterminants informationnels des cotations. Le modèle de départ est celui de Hasbrouck (1999). La dynamique des cotations est décomposée entre la dynamique du prix efficient (qui suit un E-GARCH) et la dynamique de la tenue de marché (qui suit un processus autorégressif). Les cotations proviennent donc des interactions entre ces deux processus et d'un mécanisme d'arrondi asymétrique. À chaque fois des variables de marché (fondée sur les volumes) approchant l'asymétrie d'information entrent dans les équations. Ces variables permettent de réduire la persistance du coût de tenue de marché et il semble donc que la liquidité

[10] D'autres modèles tout aussi intéressants n'ont pas été mentionnés dans un souci de brièveté. Par exemple, Blume *et al.* (1994) proposent un modèle où le volume est corrélé à la précision du signal reçu…

[11] Une relation similaire est mise en évidence par Grundy & Mc Nichols (1989), Shalen (1993), He & Wang (1995), Biais & Bossaerts (1998) qui suggèrent tous une relation positive entre d'une part l'hétérogénéité et d'autre part le volume et la volatilité.

[12] Le lecteur a déjà pu s'apercevoir que ce modèle avait bénéficié d'un important support empirique. D'autres études confortant ce modèle seront encore présentées par la suite.

[13] On peut effet émettre l'hypothèse que la volatilité est une approximation grossière de l'hétérogénéité puisque ce sont toutes les deux des mesures de dispersion.

[14] La dispersion des transactions utilisée par Goetzmann & Massa (2001) est une mesure intéressante mais nécessite des données sur les positions individuelles des investisseurs dont nous ne disposons pas.

soit directement liée à l'asymétrie d'information. De plus, une décomposition du volume selon le sens (achat ou vente) souligne que la relation est asymétrique et non-linéaire. Ainsi, le lien entre fourchette et volume est loin d'être aussi évident que la corrélation entre ces deux variables peut le laisser croire. L'effet des variables de marché sur la dynamique du prix efficient est encore plus important : elles capturent la plus grande partie de la persistance. Il semble donc que l'asymétrie d'information soit une des sources majeures du clustering de volatilité.

Les variables de marché utilisées sont de deux catégories. Elles indiquent soit une direction (volume ou profondeur signés) soit un sentiment général (la volatilité). Toutes sont le résultat d'une agrégation par intervalle de 15 minutes et sont transformées en logarithme. Il ressort que l'information contenue dans le volume est multidimensionnelle, ils proposent donc une "décomposition" du volume entre d'une part une mesure de pression : le déséquilibre entre les volume des 2 côtés de la fourchette et, d'autre part, une mesure d'intérêt : le volume d'un des côté de l'offre soit, une décomposition du volume entre une composante asymétrique (comme l'information) et une composante symétrique (comme l'hétérogénéité).

Il se dégage de cette revue de la littérature une conclusion en demi-teinte. Les implications liées au courant asymétrie d'information reposent sur les restrictions imposées aux capacités d'action des agents[15] (présence d'agents de liquidité discrétionnaire, contraintes sur les ventes à découvert...) et peuvent selon les hypothèses choisies aller jusqu'à se contredire. D'un autre côté, les évidences empiriques semblent soutenir le modèle de Easley & O'Hara (1992) et le rôle du temps (i.e. de l'intensité des transactions) plutôt que du volume. On doit toutefois remarquer que le courant de l'hétérogénéité suggère lui aussi une relation entre hétérogénéité et volume et a également bénéficié d'un support empirique. Le relatif insuccès empirique du volume peut provenir des interactions entre asymétrie d'information et hétérogénéité qui troublent l'analyse (partielle) d'une seule composante de la relation. Une décomposition du volume semble préférable.

4.1.2 Caractéristiques du carnet

Nous avons vu dans la précédente sous-section que nombreuses étaient les études qui établissaient ou étudiaient un lien entre l'information et les caractéristiques des transactions. Le sujet d'un lien entre l'état du carnet (l'offre résiduelle exprimée sur le marché) et l'information est plus récent.

Le modèle de Kavajecz (1998) lie explicitement l'information et le choix d'une position en carnet. Kavajecz (1999) démontre empiriquement qu'il est possible d'inférer de l'information à partir de la profondeur du carnet. Il compare, pour 144 titres traités sur le NYSE de novembre

[15] On peut évoquer de la même manière des contraintes sur l'information en terme de durée de vie, d'amplitude ou de précision.

1990 à janvier 1991, les contributions du spécialiste et du carnet d'ordres à la fourniture de liquidité. Il met en évidence que la profondeur est utilisée de manière stratégique par le spécialiste. Ce dernier en affichant une profondeur inférieure ou égale à celle offerte en carnet se débarrasse des transactions non désirées. De plus, tous les fournisseurs de liquidité réduisent leur exposition (diminution de la profondeur cotée) avant un événement informationnel et reviennent ensuite sur le marché. Cela souligne une relation entre les mesures de profondeur et l'occurrence d'une information, les facteurs avancés étant soit l'asymétrie d'information (autour des annonces de résultats) soit le coût trop important de la volatilité (de l'incertitude autour des annonces FOMC[16]). Enfin, certains résultats (surtout sur les titres actifs) soulignent que l'orientation de la fourchette est corrélée avec la direction de l'annonce renforçant l'hypothèse que le spécialiste parvient à inférer des informations sur l'annonce future.

L'analyse empirique réalisée par Kavajecz & Odders-White (2001) apporte un meilleur éclairage sur les sources d'informations employées par les spécialistes du NYSE pour actualiser leur cotations puisqu'elle étudie la relation dynamique entre les transactions (prix, volume et activité), l'activité du carnet (prix des cotations, profondeur, nombre d'ordres, annulations...), et la fourchette ou la profondeur disponible en carnet. Elle met en évidence que les deux dimensions du carnet (la fourchette et la profondeur) réagissent à différents événements, que le carnet d'ordres agrège bien une information, et que le spécialiste contribue à la découverte de prix dans la mesure où ses activités dépassent le cadre du simple mimétisme du carnet et ont une puissance prédictive. De manière plus précise, ils émettent deux hypothèses qui sont vérifiées par les résultats empiriques :

i.) Les changements de cotations ne sont pas dûs exclusivement aux transactions.

Les cotations répondent à d'autres déterminants que les transactions. L'influence des transactions et de l'activité cumulée est inférieure à celle du carnet d'ordres (beaucoup plus importante pour les titres moins traités).

ii.) Le spécialiste réagit aux changements de son environnement en actualisant ses prix et sa profondeur de manière éventuellement différente.

Le spécialiste actualise ses cotations prix et profondeur, le processus étant différent pour ces deux dimensions. Les changements de la meilleure offre en carnet (prix et profondeur) affectent les cotations du spécialiste. La profondeur cotée est révisée suite à une transaction (quelle que soit sa taille) alors que le prix coté ne l'est que si la transaction dépasse la profondeur cotée.

Des résultats allant dans le même sens, globalement compatibles avec le modèle de Easley & O'Hara (1992), avaient déjà été obtenus par Lee *et al.* (1993). En analysant des actions cotées sur le NYSE, les auteurs mettent en évidence que la profondeur est réduite et la fourchette

[16]Ce phénomène pourrait être interprété comme le départ des agents trop averses à l'hétérogénéité.

importante (i.e. une réduction de la liquidité) après les périodes caractérisées par des volumes importants ou avant une annonce de résultat. Si on ajoute le fait que sur le NYSE les périodes de forte activité (volume important) sont attribuées à la présence d'informés, il ressort alors une relation négative entre la liquidité et le niveau d'incertitude (asymétrie d'information + hétérogénéité associée à l'interprétation d'une nouvelle publique). La liquidité serait donc un indicateur du sentiment d'absence d'incertitude de court terme[17].

Concernant un sujet connexe Hasbrouck & Seppi (2001) présentent une analyse descriptive des variations communes entre la découverte du prix et la fourniture de liquidité pour les 30 firmes du Dow. Ils démontrent d'une part que les rendements et le flux d'ordres ont des variations communes, ce dernier explique environ $1/3$ des rendements. Ils mettent également en évidence une relation similaire mais plus faible[18] entre des approximations de la liquidité (fourchette, profondeur) et l'impact prix des transactions. Cette étude souligne donc qu'il existe un lien statistique entre la liquidité (le carnet d'ordres) et une mesure de la découverte du prix. Le carnet d'ordres semble donc modifier la manière dont l'incorporation est réalisée.

Muranaga (1999) établit clairement l'existence d'un lien entre la liquidité et l'intensité des transactions des titres cotés sur la première section du Tokyo S.E. entre octobre 1995 et septembre 1996. Il étudie trois indicateurs de liquidité inspirés de Kyle (1985) : - l'étroitesse du marché (approchée par le spread) ; - l'impact de marché (pour la profondeur approchée par les changements de cotation divisé par le volume) ; - la résilience (mesurée par la vitesse de retour à l'équilibre de la fourchette après une transaction). Pour chacun d'eux, il met en évidence une corrélation positive avec l'intensité des transactions.

Étant donné que le prix et la profondeur ont un impact différent, qu'ils réagissent à des événements distincts[19], que le carnet véhicule une information différente des transactions, cette revue nous incite à conclure que l'étude d'un seul aspect de cette relation conduit à une image, sinon fausse, du moins partielle. Les interactions entre ces variables[20] doivent être prises en compte pour que les résultats soient complets.

La revue précédente nous a permis d'appréhender les différentes possibilités d'interprétation mais aussi de comprendre l'importance d'une modélisation simultanée des différents aspects de

[17] D'autres études considèrent également la liquidité comme l'indicateur d'un sentiment toutefois l'interprétation de ce dernier est très variable. Par exemple, Baker & Stein (2000) proposent un modèle ou la liquidité agrégée provient de la sous réaction aux informations d'une classe d'investisseurs irrationnels. En cas de contrainte sur les ventes à découvert, une forte liquidité est alors le signal que le marché est dominé par ces investisseurs et que le titre est surévalué.

[18] Ainsi la significativité économique des covariations des mesures de liquidité des titres est faible.

[19] Dupont (2000) propose également un modèle où les réactions de la fourchette et de la profondeur sont différentes. Dans cette modélisation mono périodique de formation des cotations (prix et profondeur), sous hypothèse d'asymétrie d'information, le volume est plus sensible à une augmentation de l'asymétrie d'information que la fourchette. La profondeur peut aller de 0 à ∞ selon, l'asymétrie d'information, la volatilité et la force de la demande de liquidité alors que la fourchette est toujours positive et finie.

[20] Harris (1994) à l'aide d'un modèle stylisé de la relation échelon de cotation - spread - profondeur met également en évidence que ces deux leviers d'action (cotation profondeur) sont utilisés simultanément.

la relation. Toutefois, les études citées jusqu'ici reflètent majoritairement le fonctionnement d'un marché dirigé par les prix : un teneur de marché est confronté à des demandes émanant d'investisseurs dont il ne connaît pas le type (informé ou non-informé), il propose alors des cotations tenant compte de la probabilité d'être confronté à un informé ; les résultats étant le plus souvent, dérivés sous hypothèse de neutralité au risque et de concurrence. Les spécificités d'un marché dirigé par les ordres induisent des prévisions et des problématiques subtilement différentes de celles rencontrées dans les marchés dirigés par les prix. La prochaine section s'attachera à synthétiser les connaissances des marchés dirigés par les ordres utiles à la poursuite de notre étude.

4.1.3 Marchés dirigés par les ordres

Par la suite et par abus de langage mais dans un souci de plus grande clarté, nous désignerons sous les termes : - **ordre de marché** : un ordre à exécution immédiate qui consomme la liquidité ; - **ordre limite** : ordre qui fournit de la liquidité en carnet.

Le fonctionnement d'un marché dirigé par les prix est significativement différent de celui d'un marché dirigé par les ordres. Comme souligné par Handa *et al.* (1998), ce dernier système d'échange s'apparente à un milieu écologique où la viabilité du marché est assurée par un équilibre entre les offreurs de liquidité (les investisseurs qui introduisent des ordres limite) et les consommateurs de liquidité (les investisseurs opérant avec des ordres de marché). Les apporteurs de liquidité sont confrontés à un risque[21] rémunéré par de meilleures conditions d'exécution moyennes. À l'inverse, les consommateurs de liquidité ne supportent aucun risque mais sont obligés d'accepter les conditions actuelles. Ils rémunèrent en cela les investisseurs patients qui ont fournit le service de liquidité.

Une hypothèse commune aux modélisations de microstructure est que les informés utilisent exclusivement des ordres de marché. Or, cette hypothèse est difficile à justifier dans un marché dirigé par les ordres où l'investisseur a toujours le choix entre fournir et consommer de la liquidité (i.e. introduire un ordre limite ou de marché). Il est d'ailleurs difficile d'apporter un support théorique ou empirique à cette hypothèse. Par exemple, Parlour (1998, p. 807) note :

"One could imagine an informed trader submitting a limit order to get the best possible execution price while rationnaly anticipating the future order flow."

Kaniel & Liu (2002) démontrent que les informés peuvent préférer utiliser des ordres à cours limite. Leurs résultats sur 144 titre du NYSE entre novembre 1990 et janvier 1991 démontrent que les ordres limite sont perçus par le marché et le teneur de marché comme étant plus informationnels. L'intuition du modèle est que tant que l'horizon de vie de l'information

[21] L'investisseur qui entre un ordre à cours limite offre au marché une option gratuite. Pour une plus ample discussion de ce point, le lecteur peut se référer à Copeland & Galai (1983) qui a introduit ce concept ainsi qu'à la discussion de cet article dans O'Hara (1995, chap. 2 p. 13-52).

est suffisamment long, le bénéfice conditionnel de meilleurs coûts d'exécution peut dépasser le coût potentiel de non exécution. À mesure que l'horizon de vie de l'information privée augmente, la probabilité d'exécution augmente ce qui renchérit l'attractivité des ordres limite. Cette capacité des informés à fournir de la liquidité complique singulièrement l'analyse[22].

Ce type de structure induit une relation entre la volatilité transitoire et la fourniture de liquidité également observables dans les marchés mixtes comme le NYSE. Comme démontré empiriquement sur la Bourse de Paris par Biais *et al.* (1995), sur le NYSE par Handa & Schwartz (1996) ou Bae *et al.* (2003), ou encore sur la bourse de Hong Kong par Ahn *et al.* (2001), l'occurrence d'une volatilité transitoire stimule l'introduction d'ordres limite alors que la faible rémunération de la liquidité stimule sa consommation.

En raison de leur place centrale dans cette structure de marché de nombreux modèles se consacrent aux stratégies de placement des ordres. Parlour (1998) et Foucault (1999) étudient des modèles dynamiques. Parlour (1998) démontre comment le choix de placement dépend de la profondeur disponible dans le carnet. Foucault (1999) applique une modélisation séquentielle à l'analyse des conséquences de l'hétérogénéité[23] sur la formation des prix et le choix entre ordre limite et ordre de marché. L'évaluation de ces deux opportunités est faite en terme de meilleur prix d'exécution pour les ordres limite au prix d'un risque de non exécution et de malédiction du vainqueur.

La principale relation mise en évidence est l'influence de la volatilité sur le rapport entre ordres de marché et ordres limite. L'augmentation de la volatilité a des effets à plusieurs niveaux : elle augmente le risque supporté par un ordre limite (donc la rémunération demandée), ce renchérissement du coût de l'immédiateté rend moins intéressante l'utilisation d'ordres de marché ce qui diminue alors la probabilité d'exécution des ordres en carnet. Ces relations impliquent alors une corrélation positive entre la proportion d'ordres entrant en carnet et la volatilité, une relation négative entre la volatilité et le ratio des ordres limites entrant en carnet sur le total des ordres limite soumis. Ces effets ayant pour conséquence un élargissement du spread, on pourrait également tester la relation positive entre la proportion d'ordres à cours limite et la taille du spread.

Il met en outre en évidence que le coût total de transaction[24] (somme des coûts à l'achat et à la vente) est une fonction concave du ratio des ordres d'achat sur les ordres de vente qui atteint

[22]On doit remarquer que l'absence d'un agent de référence non informé peut avoir des conséquences importantes sur la diffusion d'information. En effet, les actions de cet agent pouvaient être interprétées en connaissance de son type non informé et donc servir de point de référence. À la différence, dans un marché dirigé par les ordres, l'incertitude est plus grande. Tous les agents quels que soient leurs rôles peuvent être informés. L'interprétation d'une action spécifique n'en est que plus difficile.

[23]Dans ce modèle il n'y a pas d'asymétrie d'information et les transactions proviennent de l'existence de deux groupes ayant des valorisations a priori différentes.

[24]Le coût de transaction d'un côté du marché est d'autant plus faible que ce côté représente une faible fraction des ordres introduits.

son maximum pour la valeur 1. Les quantités traitées étant contraintes à l'unité, ce modèle ne nous donne aucun renseignement sur les relations avec le volume. Enfin, la durée de vie des ordres limite n'étant que d'une période (le carnet est plein ou vide), ce modèle ne permet pas d'analyser les interactions entre les changements de l'état du carnet et le flux d'ordres. Il néglige également les effets des coûts d'attente avant exécution.

Cette limite est levée par le modèle de Foucault *et al.* (2003). Ce modèle étudie cet unique point puisque les investisseurs sont tous non informés et ne diffèrent que par leur impatience (assignée aléatoirement). Ils démontrent qu'un choc de liquidité (la succession de plusieurs ordres de marché) élargit la fourchette. Ils mettent également en évidence que la résilience de la fourchette est maximale (même niveau que dans un cadre compétitif) si les investisseurs ont des coûts d'attente similaires. À défaut, une proportion significative des transactions a lieu à des conditions moins favorables que le niveau concurrentiel. En fait, toute la distribution de la fourchette est fonction de la population des investisseurs. Si les impatients dominent alors la distribution est asymétrique à droite et la fourchette moyenne est plus élevée. Ce modèle conduit à plusieurs implications testables. En particulier, si on contrôle le taux d'arrivée des ordres alors on devrait observer une corrélation positive entre la fourchette et le nombre de transactions. Cela pourrait également suggérer que les configurations intrajournalières de la liquidité peuvent être dues à des variations de l'impatiente des investisseurs.

Handa *et al.* (2003) proposent un modèle de formation des prix dans un marché dirigé par les ordres où on retrouve à la fois de l'asymétrie d'information (un groupe d'investisseurs reçoit une information non connue des autres) et de l'hétérogénéité (il y a au départ deux groupes avec des valorisations hautes et basses) le signal informationnel reçu peut être soit haut soit bas. La structure informationnelle étant simpliste puisque l'information expire après la transaction on considère que tous les informés utilisent des ordres de marché alors que les non informés utilisent les deux types d'ordres.

À l'aide de cette structure ils démontrent que le prix de réservation optimal est une moyenne pondérée des prix de réservation corrigée du risque d'une information et de la probabilité d'être confronté à un informé. La fourchette s'accroît à mesure que la différence de valorisation initiale entre les deux groupes augmente et elle est maximale lorsque ces deux groupes sont de même taille. On voit donc là une extension par rapport au résultat de Foucault (1999) puisqu'ils démontrent que la fourchette est une fonction de l'asymétrie d'information et de l'hétérogénéité. Par contraste, la fourchette optimale d'un teneur de marché risque neutre est indépendante de la proportion des investisseurs dans les groupes de valorisation.

Ce modèle est évalué sur les actions du CAC 40 pendant l'année 1995 en agrégeant les données par intervalle de 5 minutes. La répartition des investisseurs est mesurée par le rapport entre le nombre de transactions à l'achat + le nombre d'ordres d'achat à cours limite sur le nombre total de transactions et le nombre total d'ordres à cours limite introduit. L'expression

exacte de la fourchette en fonction de 4 paramètres inconnus (le déséquilibre entre les deux groupes, la probabilité d'être confronté à un informé, la différence de valorisation entre les groupes, et le choc informationnel) leur permet de conduire un test GMM de leur modélisation qui n'est rejeté que pour 22% des actions.

Ranaldo (2004) présente, à l'aide d'un modèle probit, une analyse empirique de l'influence du carnet sur les décisions des investisseurs. L'analyse de 15 actions cotées sur le marché suisse de mars à avril 1997[25] souligne que les investisseurs deviennent plus agressifs lorsque leur côté est fourni et moins agressifs lorsque le côté opposé est dégarni. Les périodes de forte introduction d'ordres sont déclenchées principalement par des ordres passifs. Une volatilité transitoire importante et des fourchettes larges stimulent l'introduction d'ordres limite. La volatilité et l'agressivité évoluent dans le même sens. Les ordres de marché et à cours limite ont des réactions inverses aux changements de conditions. Ces résultats démontrent que les deux côtés du carnet sont importants pour comprendre le choix des agents et que ces derniers réagissent aux changements de probabilité d'exécution. Il met également en évidence certaines dissymétries entre les réactions à l'achat et à la vente en attribuant celles-ci à une moindre information véhiculée par les ventes ainsi qu'une plus forte activité des institutionnels.

Coppejans & Domowitz (2002) analysent les relations entre le carnet d'ordres (en terme de flux d'ordres et de demande insatisfaite) et les transactions sur les contrats à terme suédois OMX. Le choix d'un contrat à terme sur indice n'est pas dû au hasard mais a pour but, d'après les auteurs, de minimiser l'asymétrie d'information. Le cadre d'analyse est un Auto regressive Conditionnal Duration (ACD)-GARCH (ce type de modélisation sera présenté dans le prochain chapitre). Cette analyse met en avant l'existence de liens entre les différents évènements ou variables. Ainsi, il est manifeste que le carnet apporte une information différente des transactions et qu'elle affecte le comportement des investisseurs. Par exemple, les transactions passées et le niveau de la fourchette touchent directement les ordres et les annulations mais les transactions ne semblent réagir à aucun des autres éléments (niveau de la fourchette, occurrence d'un ordre ou annulation). Seules deux variables pertinentes émergent pour expliquer la duration : la profondeur et les durations passées.

Coppejans *et al.* (2003) analysent la dynamique de la liquidité et sa relation à la volatilité et au rendement. En utilisant des données sur contrat à terme, ils mettent en évidence que la liquidité de marché a des variations systématiques qu'ils attribuent à des investisseurs stratégiques renforçant une concentration naturelle. Ce fait corrobore le modèle de Admati & Pfleiderer (1988). Les chocs de volatilité réduisent la liquidité et l'efficience. Cette réduction est cependant moindre en raison de la résilience du marché (rapidité du retour à la normale après un choc).

Sans vouloir prétendre à une quelconque exhaustivité, la revue de la littérature présentée ci

[25]Les données utilisées sont similaires à celles de la base TAQ.

dessus fait ressortir quelques point saillants. L'organisation d'un marché dirigé par les ordres induit une instabilité de l'offre de liquidité puisque qu'aucun intervenant n'est désigné pour la fournir. Lorsqu'un choc se produit il faut attendre que la hausse de la rémunération proposée attire de nouveaux fournisseurs de liquidité. Les mécanismes de fourniture de la liquidité induisent une dynamique temporelle et il n'est pas certain qu'elle ait le moindre rapport avec celle de l'information.

Les différents modèles ou études présentés établissent tous un lien entre des caractéristiques des transactions ou du carnet et l'information. Cependant, la diversité des approches employées et des résultats obtenus obscurcit l'analyse.

Pour être synthétique, les principales conclusions sont que, dans un marché dirigé par les ordres, un informé peut rationnellement se porter apporteur de liquidité. De manière évidente, la durée de vie de l'information devient un élément crucial puisqu'elle conditionne la capacité des investisseurs à exploiter leurs informations. Les concepts d'asymétrie d'information et d'hétérogénéité ont tous les deux été appliqués dans des modèles de marché dirigés par les ordres. Les coûts de transaction sont une fonction concave (avec un maximum à 1) du ratio de la demande[26] (Foucault, 1999) mais aussi une fonction croissante de l'asymétrie d'information (Handa *et al.*, 2003). Les quelques études empiriques mentionnées ici confirment que le carnet d'ordres est une des variables participant à l'incorporation de l'information puisqu'il permet d'expliquer une partie des caractéristiques des transactions. Toutefois, son rôle exact demeure trouble.

Il apparaît alors que la prise en compte des deux composantes que sont l'asymétrie d'information et l'hétérogénéité est indispensable afin d'obtenir une image plausible du fonctionnement du marché[27]. Enfin, comme souligné par Ranaldo (2004), les deux côtés du carnet sont importants et ont leur place dans l'analyse.

4.2 Pré-traitements

Nous travaillons à partir de deux fichiers : l'historique des transactions et celui des 5 meilleures limites[28]). Nous ne disposons des 5 meilleures limites que sur les titres du CAC 40 pendant les 4 premiers mois de 1999. Aussi, nous analyserons ici les 7 titres du CAC 40 déjà étudiés dans la première partie et restreindrons la période d'analyse.

[26]Intuitivement, les coûts d'exécution d'un ordre de marché à l'achat sont d'autant plus importants que la proportion de vendeurs est faible : les ordres doivent consentir un effort plus important pour attirer un vendeur. Ainsi, le coût d'un ordre de marché à l'achat est croissant avec la proportion d'ordres achat. Le même argument pour les ordres de vente donne in fine une relation concave entre le coût total et le ratio de la demande.

[27]Des conclusions similaires sont mises en avant par le modèle de Hollifield *et al.* (2003*a*) et dans l'étude de Hollifield *et al.* (2003*b*) qui s'en inspire.

[28]Ces données ont gracieusement été mises à notre disposition par la SBF vu que le fichier diffusé dans la base BDM ne reprend que la meilleure limite.

Les retraitements effectués sur les transactions sont identiques à ceux de la première partie (suppression des transactions en dehors des horaires, exclusion des transactions aux fixages et des applications). Les effets sur cet échantillon sont similaires à ceux décrits dans la première partie.

L'analyse de données tick par tick requiert que le prix à un instant donné soit unique. Or, on observe sur le marché plusieurs transactions au même horodatage. Nous devons donc définir une procédure de traitement des événements simultanés. La sous-section 4.2.1 analysera les différentes procédures de traitement des événements synchrones. L'information apportée par le carnet est importante et répartie dans un grand nombre de variables. La sous-section 4.2.2 proposera la construction de plusieurs variables tirées du carnet et interprétables en terme de paramètres du flux informationnel. Notre but dans cette partie est d'étudier la dynamique du processus d'échange. L'existence de configurations intrajournalières obscurcit cette dynamique. Comme souligné par Engle & Russell (1998), on peut incorporer cette dynamique dans la modélisation ou désaisonnaliser les données avant analyse. À l'instar de Engle & Russell (1997, 1998) ou encore Manganelli (2002) nous préférerons, pour plus de simplicité, désaisonnaliser les données avant traitement. Nous présenterons les méthodes de désaisonnalisation dans la sous-section 4.2.3.

4.2.1 Traitement des événements simultanés

Avant de s'intéresser à la manière dont on peut traiter les événements simultanés, il convient d'apprécier l'ampleur du phénomène. La table 4.1 donne pour les différentes titres la distribution du nombre de transactions dans une même seconde lorsqu'il y en a.

TABLE 4.1
Nombre de transactions par seconde

	St-Gobain	Schneider	Sanofi	Thomson	Accor	Lyonnaise	Lagardère
Moyenne	1,74	1,81	1,60	1,70	1,78	1,67	1,77
Écart-type	3,11	3,57	1,58	2,33	2,96	3,01	3,46
Médiane	1	1	1	1	1	1	1
Q3	2	2	2	2	2	2	2
90%	3	3	3	3	3	3	3
95%	4	4	4	4	4	4	4
99%	10	11	7	9	12	8	11
Max	173	258	79	147	171	216	303

Cette table donne pour les différents titres les principaux paramètres de la distribution du nombre de transactions dans une même seconde lorsqu'au moins une transaction est observée.

Nous notons que l'occurrence de plus d'une transaction dans la seconde est fréquente (plus de 25% des cas). Il est rare d'observer un très grand nombre de transactions dans la même

seconde (le 95^{e} centile n'est que de 4) mais les maxima sont importants (de 79 à 303).

Ce phénomène est en partie dû à l'organisation du marché parisien en marché électronique dirigé par les ordres : dans cette structure de marché, le nombre de transactions généré par un ordre est fonction du nombre de contreparties nécessaires à sa satisfaction[29]. Une séquence de transactions[30] peut résulter de l'arrivée simultanée sur le marché de deux ordres ou de l'arrivée sur le marché d'un ordre exécuté contre plusieurs contreparties éventuellement à des prix différents.

Le premier problème posé par les séquences provient d'une déficience des données. En effet, on dispose dans le fichier des transactions et des meilleures limites d'un horodatage dont la précision est la seconde et d'un numéro de séquence identifiant l'ordre des événements à l'intérieur des secondes. Cependant, les numéros de séquence dans ces deux historiques ne sont pas communs et sont donc difficiles à rapprocher. De plus, les meilleures limites sont rarement actualisées à l'intérieur des séquences de transactions. Si le prix milieu de fourchette est utilisé de manière naïve (le dernier prix milieu de fourchette en vigueur est appliqué à toutes les transactions qui suivent) alors celui-ci ne change pas à l'intérieur d'une séquence : les rendements des transactions à l'intérieur d'une séquence sont nuls. Pour résoudre ces problèmes, nous utilisons la procédure suivante :

i.) Nous ne retenons que les meilleures limites à la fin de chaque seconde puis les introduisons dans l'historique des transactions avant ces dernières.

ii.) Si la transaction n'appartient pas à une séquence, nous utilisons le dernier prix milieu de fourchette en vigueur. Si la transaction appartient à une séquence, alors nous défalquons des quantités disponibles celles déjà consommées par les précédentes transactions de la séquence afin d'actualiser le prix milieu de fourchette à l'intérieur des séquences comme la moyenne des meilleurs prix à l'achat et à la vente pour lesquels la quantité disponible est strictement positive. Cela revient à dire que nous annulons les meilleurs limites déjà consommées. Le prix milieu de fourchette reflète donc les conditions à l'instant précis de la transaction.

On trouve dans la littérature deux types de traitements pour résoudre le problème des événements simultanés (Gouriéroux & Jasiak, 2001, cf. p. 358-359). Les séquences sont agrégées en une seule transaction[31]. Les horodatages sont artificiellement modifiés afin que les événements ne soient plus exactement synchrones (Coppejans & Domowitz, 1999; Darolles *et al.*, 2000). On doit remarquer qu'il n'est pas possible de modifier les horodatages d'un nombre fixe.

[29]Dans un marché dirigé par les prix, un ordre ne génère qu'une transaction à un prix unique.

[30]Nous désignerons par la suite sous le terme de séquence un ensemble de transactions enregistrées à la même seconde.

[31]Le volume est alors la somme des volumes et le prix peut être soit la moyenne des prix pondérés par les volumes ou le dernier prix de la séquence (Aubier & Miloudi, 2001).

En effet, étant donné le nombre important de transactions réalisées dans une même seconde, cette pratique résoudrait certes une partie des problèmes de simultanéité mais en créerait également de nouveaux. Une solution peut être de répartir les transactions simultanées dans la même seconde[32]. Afin d'explorer les effets de ces deux méthodes de traitement des transactions simultanées nous évaluons le modèle (4.1) sur les séries construites avec ces différentes méthodes.

$$\sigma_t = \mu + \alpha_1 d_t + \alpha_2 d_{t-1} + \alpha_3 v_t + \alpha_4 v_{t-1} + \varepsilon_t \tag{4.1}$$

où σ_t est la volatilité de la t^e transaction mesurée par la valeur absolue du rendement, d_t est la duration (temps séparant une transaction de la précédente) et v_t le volume de la transaction (en nombre de titres). Nous régressons donc la volatilité de la transaction sur sa duration et son volume. La table 4.2 donne les résultats de l'estimation du modèle (4.1) sur Sanofi.

TABLE 4.2
Relations volatilité - duration - volume
Sanofi

	Non Séquence	Répartition	Agrégation Moyen	Agrégation Fin
R^2 Aju.	0,22	0,19	0,25	0,24
μ	0,05***	0,03***	0,05***	0,05***
	(50,72)	(66,06)	(55,86)	(53,18)
d_t	0,00021***	0,00030***	0,00020***	0,00020***
	(16,40)	(29,15)	(18,38)	(17,94)
d_{t-1}	-0,00004***	-0,00005***	-0,00004***	-0,00004***
	(4,37)	(8,01)	(4,95)	(4,51)
v_t	3,25E-6***	1,41E-6*	4,92E-6***	9,94E-6***
	(2,38)	(1,37)	(5,59)	(9,44)
v_{t-1}	4,70E-6***	6,18E-6***	1,55E-5***	1,42E-5***
	(3,35)	(7,42)	(11,04)	(9,60)

Cette table fournit les paramètres du modèle (4.1) estimés par moindres carrés ordinaires. Les T-values données entre parenthèses en dessous des paramètres sont calculées en utilisant les écarts types robustes de White (1980). Un paramètre significatif au seuil de 1% est indiqué par ***, au seuil de 5% par ** et au seuil de 10% par *. L'entête de la colonne indique la méthode utilisée pour construire les séries (moyen ou fin se réfèrent à la méthode de sélection du prix cf. note de bas de page 31). La colonne "Non Séquence" est le résultat d'une estimation faite exclusivement sur les transactions n'appartenant pas à une séquence.

Si on répartit les transactions dans la seconde, la relation entre la volatilité d'une transaction et son volume est peu significative. La faiblesse de cette significativité semble être la conséquence de cette méthode de traitement puisque cette relation est très significative avec toutes les autres méthodes, y compris pour les transactions n'appartenant pas à une séquence et qui ne

[32]Par exemple, si on a 2 transactions dans la séquence, la seconde est incrémentée de 1/2 seconde, si on en a 3, la deuxième est incrémentée de 1/3 de seconde et la troisième de 2/3....

sont donc pas affectées par cette méthode.

L'influence du prix choisi pour représenter les séquences est plus trouble. D'un point de vue théorique, un prix moyen représente mieux le prix payé par l'investisseur initiateur de la séquence. Il conduit cependant une mauvaise estimation du rendement qui suit la séquence puisque le prix précédent n'est pas celui observé sur le marché mais le prix moyen. À l'inverse, le choix du dernier prix surestime le prix payé par l'initiateur (cela se traduit par la forte augmentation du paramètre v_t) mais évalue bien le rendement suivant les séquences.

Étant donné les effets indésirables d'une répartition des transactions dans la seconde et la prépondérance dans la littérature de prix pondérés par les volumes, nous utiliserons la méthode d'agrégation des transactions dans la même seconde avec un prix moyen pondéré par les volumes.

Pour mesurer le volume, on trouve dans la littérature deux choix. On peut utiliser le volume tel quel (nombre de titres de la transaction) ou le transformer par une fonction (e.g. le logarithme pour Zhang *et al.* (2000), la racine carrée pour Hasbrouck & Seppi (2001)). Cette transformation peut rendre les résultats difficilement interprétables[33] mais peut permettre de linéariser une relation. Nous utiliserons des mesures non transformées (i.e. le nombre de titres pour le volume).

La méthode de calcul des rendements la plus populaire (cf. Engle (2000)) est une mesure ajustée par la duration (pour être plus précis par la racine carrée de la duration). Cette normalisation permet alors d'avoir pour chaque transaction une mesure du rendement (ou de la volatilité) par seconde qui doit, théoriquement, être comparable quelle que soit la durée séparant les transactions. On peut tout d'abord remarquer que cette normalisation est proposée dans le cadre d'un modèle qui étudie exclusivement l'influence de la duration sur la volatilité[34]. Donc, si on veut étudier l'influence du volume sur la volatilité, il est alors logique de normaliser le rendement par la racine du volume, ce qui fournit alors un rendement par unité de transaction.

Ce type de mesure peut cependant introduire des biais. Par exemple, si on considère, comme Diamond & Verrecchia (1987), qu'un temps sans transactions (une longue duration) est révélateur d'une mauvaise nouvelle et a donc un important impact et négatif sur les prix, cette normalisation réduit artificiellement cette relation. En effet, les durations importantes sont mécaniquement associées à des rendements ajustés faibles. La normalisation pourrait donc fausser les résultats. Si cette méthode est pertinente et même nécessaire lorsque le but est exclusivement d'étudier l'effet de la duration sur la volatilité (comme le fait Engle (2000)), elle ne semble pas l'être pour analyser l'effet conjoint des durations et des volumes sur les rendements.

Nous utilisons donc, selon l'objectif, trois mesures différentes de rendement. La mesure de base est la différence des logarithmes entre deux prix successifs $r_t = \ln(P_t) - \ln(P_{t-1})$. Les deux autres mesures sont des rendements ajustés par les racines carrées soit de la duration

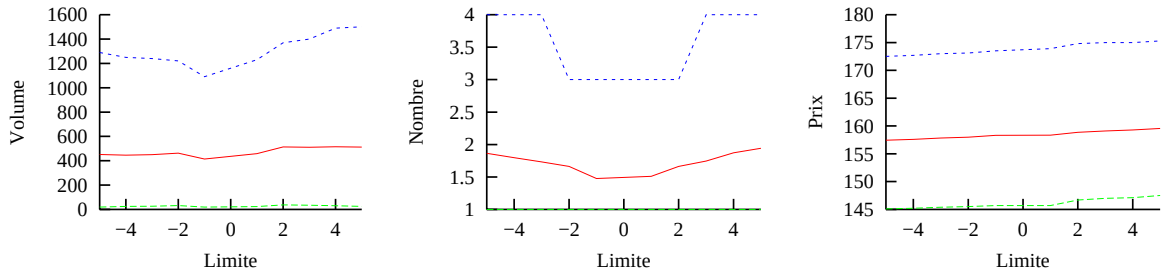
[33] Par exemple, après cette transformation, le volume d'une transaction d'un titre est nul.
[34] Ce modèle sera présenté en détail dans la sous-section 5.3.1.

$r_{d,t} = r_t/\sqrt{d_t}$, soit du volume $r_{v,t} = r_t/\sqrt{v_t}$. Les premières données du jour sont calculées par référence au fixage d'ouverture[35].

4.2.2 L'information du carnet

Nous disposons de 30 variables qui décrivent toutes les modifications d'une des 5 premières limites à l'achat et à la vente en terme de nombre (nombre d'ordres), de prix et de volume (nombre de titres). La figure 4.1 décrit la distribution des 3 variables pour les différentes limites et la table 4.3 donne en exemple l'état du carnet de Sanofi après le fixage d'ouverture le 1^{er} janvier. Cet exemple nous servira par la suite à illustrer notre méthodologie.

FIGURE 4.1
Distribution des volumes, des nombres et des prix
Sanofi



Dans chaque graphique, l'axe Y indique la variable traitée. La courbe rouge continue est la moyenne alors que les courbes hachurées délimitent un échantillon de confiance empirique à 90%. Une valeur positive (négative) de limite indique une limite à la vente (à l'achat) dans le carnet d'ordres. Le volume est le nombre d'actions offertes, le nombre est le nombre d'ordres présents, et le prix est exprimé en €.

On peut remarquer une dissymétrie importante du volume des deux côtés du carnet. Ainsi, le 95^e centile de la distribution du volume par limite est plus important à la vente qu'à l'achat. On doit remarquer la très forte dispersion des variables. Le prix sous cette forme n'apporte que peu d'information. Il est plus intéressant d'étudier l'évolution de la fourchette (différence entre le prix acheteur et vendeur). Nous remplacerons désormais le prix par cette construction. L'un des intérêts majeurs des données sur les limites est qu'elles permettent de reconstituer les courbes d'offre et demande. Pour calculer ces courbes, on retraite les données afin de mettre en évidence les variations marginales. Cette image est plus fine puisqu'elle met en relief la sensibilité des variables.

[35]Le premier rendement du jour est la différence entre le logarithme du prix milieu de fourchette lors de la première transaction en continu et le logarithme du prix de fixage. De la même manière, la première duration du jour est le temps séparant la première transaction en continu et le fixage d'ouverture.

TABLE 4.3
État du carnet de Sanofi
4 janvier 1999 10:10:36

Achat				Vente		
Prix	Volume	Nombre	n°	Prix	Volume	Nombre
143,6	100	1	1	144,9	182	4
143,5	200	1	2	145	602	2
143,3	40	1	3	145,5	6	1
143,1	150	1	4	145,6	20	2
141,7	8	1	5	146	650	1

Cette table donne l'état des 5 premières limites du carnet d'ordres de Sanofi le 4 janvier 1999 à 10:10:36. Le volume désigne le nombre de titres offerts et le nombre donne le nombre d'ordres présents à une limite. n° indique la profondeur de la limite en carnet.

Il est usuel de calculer une courbe d'offre exprimant les prix en fonction des volumes (cf. Gouriéroux & Jasiak (2001, p. 386-387)). Afin de traiter de la même manière les trois variables dont nous disposons, nous exprimerons celles-ci en fonction des différentes limites[36].

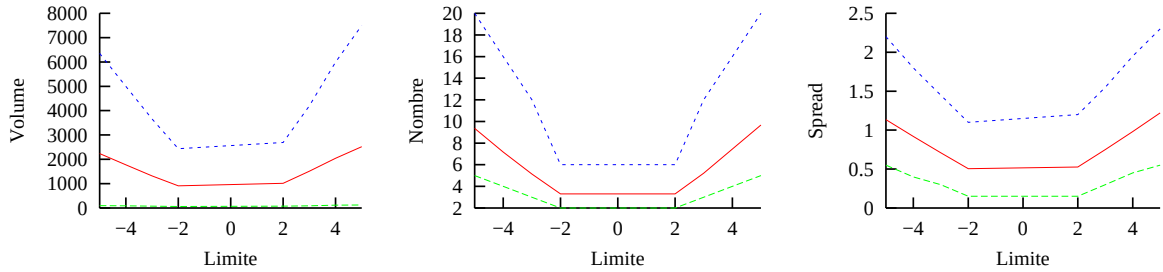
Le calcul des courbes d'offre est simple puisqu'il suffit de sommer les valeurs observées aux limites inférieures (i.e. le volume agrégé à la troisième limite est la somme des volumes proposés aux limites 1, 2 et 3). Cette procédure de sommation n'est pertinente que pour les volumes et les nombres. Le calcul de la courbe d'offre pour les prix est légèrement différent et se déroule en deux étapes. Les meilleures limites à l'achat et à la vente nous permettent de calculer le prix milieu de fourchette. On calcule ensuite pour les autres limites la différence absolue entre le prix proposé et le prix milieu de fourchette. Une telle mesure est bien équivalente aux sommes construites sur les volumes et les nombres puisque le chiffre obtenu reflète bien le coût total pour atteindre la x^{e} limite en carnet[37]. La figure 4.2 donne les courbes d'offre agrégées et la table 4.4 les mesures agrégées découlant de l'état du carnet présenté dans la table 4.3.

Les courbes d'offre et de demande soulignent une légère asymétrie de toutes les variables mettant en évidence des valeurs plus élevées du côté vente. Ce phénomène assez ténu est principalement observable sur le 95$^{\text{e}}$ centile de la distribution. On voit donc que les volumes, les nombres et que les fourchettes ont une légère tendance à augmenter plus vite à la vente. Ces considérations n'ont pas la même signification. Il est usuel de considérer que la liquidité a au moins deux dimensions : la profondeur et le prix. La liquidité volume est d'autant plus importante que le volume l'est alors que la liquidité prix est une fonction de l'inverse de la fourchette (elle est d'autant plus élevée que la fourchette est faible). Il apparaît alors que le côté vente

[36] Si nous avions voulu agréger l'information prix et volume nous aurions pu par exemple utiliser la mesure proposée Hasbrouck & Seppi (2001) qui emploient la pente de la droite entre les deux cotations (dans un repère log volume, prix).

[37] Par définition cette mesure n'est donc pas disponible pour la meilleure limite.

FIGURE 4.2
Distribution des volumes, des nombres et des fourchettes agrégés
Sanofi



Dans chaque graphique, l'axe Y indique la variable traitée. La courbe rouge continue est la moyenne alors que les courbes hachurées délimitent un échantillon de confiance empirique à 90%. Les variables représentées ici sont agrégées. Pour le volume et le nombre la valeur associée à une limite est la somme des valeurs brutes observées aux limites inférieures ou égale. La fourchette est la différence absolue entre le prix milieu de fourchette et le prix proposé à une limite. Une valeur positive (négative) de limite indique le prix vendeur (acheteur) dans le carnet d'ordres. Les unités sont identiques à celles du précédent graphique.

est, dans notre échantillon, plus liquide en volume mais moins en prix. Cela suggère l'existence d'une relation de substitution entre ces deux variables.

TABLE 4.4
Mesures agrégées du carnet de Sanofi
4 janvier 1999 10:10:36

Achat				Vente		
Fourchette	Volume	Nombre	n°	Fourchette	Volume	Nombre
0,65	100	1	1	0,65	182	4
0,75	300	2	2	0,75	784	6
0,95	340	3	3	1,25	790	7
1,15	490	4	4	1,35	810	9
2,55	498	5	5	1,75	1460	10

Cette table donne les mesures agrégées construites à partir du carnet d'ordres présenté dans la table 4.3. Les variables ont les mêmes significations que dans cette table. Fourchette désigne la différence entre le prix proposé et le prix milieu de fourchette (144,25 €).

Les mesures agrégées sont, de manière évidente, plus pertinentes dans notre étude puisqu'elles reflètent mieux la mécanique du carnet d'ordres où une limite ne peut être touchée que lorsque les limites inférieures en carnet ont été déjà consommées[38]. On peut en outre noter que

[38]Nous renvoyons le lecteur au document intitulé "Les ordres de bourses" diffusé par la SBF en 2001 qui détaille les différents types d'ordres et leurs effets respectifs en fonction du carnet d'ordres, disponible à l'adresse : `http://www.bourse-de-paris.fr/centredoc/pdf/ordresbourse_nov2001.pdf`.

ces mesures permettent mieux d'appréhender l'état du carnet comme l'analyse de la figure 4.2 l'a montré.

On peut toutefois regretter que ces mesures ne soient pas interprétables en terme de paramètres du flux informationnel. La revue de la littérature précédente souligne que le carnet d'ordres fournit des indications pertinentes sur le flux informationnel. Toutefois, on aussi vu que cette relation est complexe et que l'utilisation de mesures brutes du carnet peut conduire à des résultats trompeurs en raison de la nature de la relation qui peut être non linéaire, asymétrique, multiforme...

Nous formalisons la construction de mesures informationnelles à partir des 5 meilleures limites agrégées en les reliant à des caractéristiques interprétables du flux informationnel telles qu'utilisées dans la littérature. Il apparaît alors nécessaire d'employer des mesures comparant ou agrégeant les données des deux côtés du carnet afin de faire ressortir les différentes dimensions de l'information contenue en carnet. Nous définissons 3 dimensions à partir du carnet d'ordres :

i.) La symétrie des courbes d'offre et de demande

ii.) La direction d'un éventuel déséquilibre entre l'offre et de la demande

iii.) L'intérêt global du marché.

Ces 3 mesures peuvent être interprétées en terme de paramètres du flux informationnel que nous avons évoqués précédemment. La symétrie mesure la ressemblance entre les deux côtés du carnet. Plus les deux côtés se ressemblent, plus les valorisations sont hétérogènes. Cette mesure n'est pas signée et indique le niveau du désaccord entre les investisseurs. Elle est donc une mesure apparentée au concept d'hétérogénéité. La direction est une mesure signée, elle indique le sens du déséquilibre observé. Elle est donc apparentée au sens du signal informationnel[39]. Enfin, l'intérêt global du marché traduit une plus forte présence des investisseurs et donc leur acceptation d'une plus forte exposition. On peut donc lier cette mesure à l'incertitude. Moins l'incertitude est élevée, plus l'intérêt global devrait être important. Par souci de clarté ces trois dimensions seront ensuite désignées sous les termes : **hétérogénéité, direction et profondeur**.

Pour définir ces 3 dimensions nous considérons d'abord, pour plus de généralité, une variable N dont on connaît pour une limite, le niveau à l'achat (N_a) et à la vente (N_v). Nous reviendrons ensuite aux formules explicites des différentes mesures à partir des trois données disponibles sur le carnet (le nombre d'ordres, le volume d'ordres et le prix).

L'hétérogénéité mesure le désaccord entre les investisseurs. Ce dernier est d'autant plus important que les deux groupes formés par les investisseurs sont de même type (volume proposé, nombre d'ordres ou distance au prix milieu de fourchette). L'hétérogénéité est donc maximale lorsque les investisseurs se répartissent en 2 groupes de mêmes tailles bien distincts. Il est alors

[39] De manière évidente l'hétérogénéité résulte d'une transformation de type valeur absolue de la direction (au moins sur les données brutes). L'utilisation simultanée de ces deux mesures permet de rendre compte de l'asymétrie de la relation.

logique de retenir comme mesure de l'hétérogénéité une mesure de la symétrie de l'offre des deux côtés du carnet[40]. Nous proposons comme mesure[41] de l'hétérogénéité (4.2).

$$H_N = \min\left(\frac{N_a}{N_a + N_v}, \frac{N_v}{N_a + N_v}\right) \tag{4.2}$$

Cette mesure évolue dans l'intervalle $]0, 0,5[$, plus les deux côtés de l'offre sont similaires plus la valeur est proche de 0,5. Cette mesure s'inspire du ratio utilisé par Foucault (1999). Étant donné qu'il prévoit une relation quadratique entre les coûts de transaction et ce ratio approximation de l'hétérogénéité, nous attendons une relation positive avec notre mesure d'hétérogénéité.

La direction approche la valeur du signal. Contrairement à l'hétérogénéité qui mesure la ressemblance entre les côtés offre et demande, la direction mesure le déséquilibre. La mesure proposée doit donc être une fonction croissante du déséquilibre entre les deux côtés du carnet. Elle doit aussi indiquer de quel côté la transaction suivante devrait avoir lieu soit émettre une hypothèse sur la relation entre le niveau relatif de l'offre d'un côté du carnet et le sens de la prochaine transaction.

Les résultats empiriques ou théoriques ne sont pas convergents. Kavajecz (1999), Kavajecz & Odders-White (2001) suggèrent que la liquidité n'est pas fournie par des informés puisqu'on observe sa diminution avant les événements informationnels. On peut donc penser qu'en dépit de la possibilité théorique que les informés apportent de la liquidité, cette dernière est majoritairement apportée par des non-informés et que ces derniers ont tendance à réduire leur exposition lorsqu'ils perçoivent un risque informationnel. On serait alors tenté de conclure que la prochaine transaction devrait être observée du côté du carnet où l'offre est la moins importante (fuite des non-informés).

Cependant les études de la dynamique des ordres (cf. section 4.1.3) suggèrent une relation différente. Par exemple, Ranaldo (2004) met en évidence que les investisseurs sont plus agressifs lorsque leur côté de l'offre est le plus fourni et qu'ils le sont moins si le côté opposé est dégarni. On est alors amené à émettre l'hypothèse que la prochaine transaction aura lieu du côté ou l'offre est la plus importante.

Étant donné que la première hypothèse propose une interprétation de type informationnel[42] nous préférons celle-ci et considérons par la suite que la direction du signal (le côté de la prochaine transaction) est donnée par le coté où l'offre est la plus faible (c'est le côté où les non-

[40] dans ce cadre, toute dissymétrie de l'offre peut être interprétée comme une réduction de l'hétérogénéité.

[41] Handa *et al.* (1998) emploient une mesure d'hétérogénéité similaire ($V_a/(V_a + V_v)$). Toutefois, cette mesure a le désavantage d'être signée, elle associe donc direction et hétérogénéité. Notre mesure isole l'hétérogénéité qui est maximale lorsque la valeur tend vers 0,5. Les auteurs démontrent alors que la fourchette est plus importante lorsque l'hétérogénéité est forte.

[42] En comparaison, la seconde hypothèse nous semble proposer une interprétation reposant plus sur les mécanismes de fourniture de liquidité. Cette distinction est toutefois un peu artificielle étant donné que la fourniture de liquidité est à l'évidence influencée par l'information.

informés ont le plus intérêt à annuler leur position ou a ne pas en prendre). Afin d'approcher la valeur du signal et sa direction nous proposons la mesure suivante :

$$D_N = (-\mathbb{1}_{N_v \geqslant N_a} + \mathbb{1}_{N_a > N_v}) \times (\max(\frac{N_a}{N_a + N_v}, \frac{N_v}{N_a + N_v}) - 0,5) \tag{4.3}$$

Les deux fonctions indicatrices[43]. ($\mathbb{1}_.$) permettent de signer la variable en fonction du côté où l'offre est la plus faible. La fonction max permet d'obtenir une variable croissante en fonction du déséquilibre[44]. Sans la correction par $0,5$ la mesure évolue dans l'ensemble : $]-1,-0,5[\quad \cup \quad]0,5,1[$, la correction permet de ramener l'ensemble image à $]-1,1[$. On peut remarquer que si la direction n'était pas une variable signée, on aurait direction $+$ hétérogénéité $= 1$. La relation entre ces deux variables doit s'organiser autour d'une courbe de type valeur absolue.

La profondeur mesure la quantité totale de l'offre en carnet. Elle somme donc les offres disponible à l'achat et à la vente. Notre mesure de profondeur est (4.4).

$$P_N = N_a + N_v \tag{4.4}$$

La profondeur n'est pas bornée mais est strictement positive (à valeur dans $\mathbb{R}^{+*}$). Elle est d'autant plus importante que les offres à l'achat et à la vente sont grandes. Le lien entre signal et profondeur est trouble et dépend de la nature de l'apport de liquidité. Si l'apport de liquidité est réalisé majoritairement par des non-informés comme on peut l'attendre, on devrait observer une moindre profondeur pendant l'incorporation d'une information.

Si on considère les différentes variables, on peut penser, par exemple, que l'arrivée d'une information se traduit par une profondeur faible (les non-informés diminuent leur exposition) et une indication directionnelle forte (donc une hétérogénéité faible).

La mesure des différentes dimensions sur le volume ou le nombre d'ordres est directe mais le traitement des prix requiert une étape supplémentaire. Pour calculer les dimensions sur les prix nous décomposons les fourchettes entre la partie à l'achat et à la vente. Étant donné que le point de référence de cette décomposition est le prix milieu de fourchette (moyenne des prix d'achat et de vente à la première limite) les dimensions hétérogénéité prix et direction prix ne sont donc pas disponibles pour la cette limite. Par souci d'uniformité, nous n'utiliserons pas non plus la profondeur prix à la première limite ; toutes les mesures sur les prix débuterons donc à la deuxième limite. L'hétérogénéité prix indique si la fourchette à la x^{e} limite est symétrique par rapport au prix milieu de fourchette. La profondeur est l'inverse de la fourchette, à une limite x elle a pour valeur $1/(P_{Vx} - P_{Ax})$. Nous utilisons l'inverse de la fourchette afin que cette mesure

[43] Une fonction indicatrice a pour valeur 1 si la condition est vérifiée et 0 sinon.

[44] Par opposition, la fonction min dans hétérogénéité permettait d'obtenir une fonction croissante en fonction de la similarité des offres de chaque côté.

soit croissante lorsque la fourchette baisse. Étant donné que nous avons considéré que le sens du signal était le côté où l'offre est la plus faible (la fourchette la plus large) la mesure de direction proposée dans l'équation (4.6) fournit une mesure consistante[45]. Les équations (4.5) à (4.7) fournissent respectivement les formules exactes de calcul de l'hétérogénéité, de la direction et de la profondeur pour le volume, le nombre d'ordres et le prix.

$$\begin{aligned} H_{Volume,l} &= \min\left(\frac{\sum_{i=1}^{l} V_{a,i}}{\sum_{i=1}^{l} V_{a,i}+V_{v,i}}, \frac{\sum_{i=1}^{l} V_{v,i}}{\sum_{i=1}^{l} V_{a,i}+V_{v,i}}\right) \\ H_{Nombre,l} &= \min\left(\frac{\sum_{i=1}^{l} N_{a,i}}{\sum_{i=1}^{l} N_{a,i}+N_{v,i}}, \frac{\sum_{i=1}^{l} N_{v,i}}{\sum_{i=1}^{l} N_{a,i}+N_{v,i}}\right) \\ H_{Prix,l} &= \min\left(\frac{Mid-P_{a,l}}{P_{a,l}-P_{v,l}}, \frac{P_{v,l}-Mid}{P_{v,l}-P_{a,l}}\right) | \forall l > 1 \end{aligned} \tag{4.5}$$

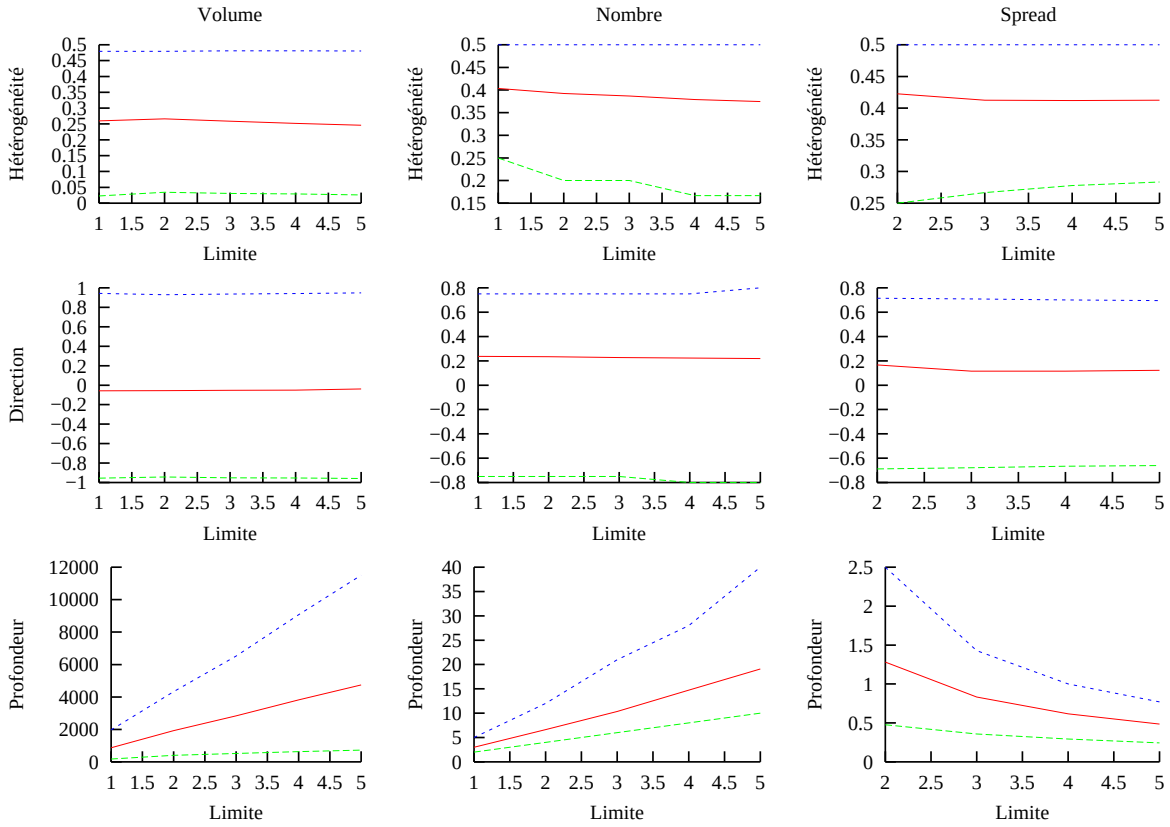
$$\begin{aligned} D_{Volume,l} &= (-\mathbb{1}_{\sum_{i=1}^{l} V_{v,i} \leqslant \sum_{i=1}^{l} V_{a,i}} + \mathbb{1}_{\sum_{i=1}^{l} V_{a,i} < \sum_{i=1}^{l} V_{v,i}}) \times \\ &\quad \left(\max\left(\frac{\sum_{i=1}^{l} V_{a,i}}{\sum_{i=1}^{l}(V_{a,i}+V_{v,i})}, \frac{\sum_{i=1}^{l} V_{v,i}}{\sum_{i=1}^{l}(V_{a,i}+V_{v,i})}\right) - 0,5\right) \\ D_{Nombre,l} &= (-\mathbb{1}_{\sum_{i=1}^{l} N_{v,i} \leqslant \sum_{i=1}^{l} N_{a,i}} + \mathbb{1}_{\sum_{i=1}^{l} N_{a,i} < \sum_{i=1}^{l} N_{v,i}}) \times \\ &\quad \left(\max\left(\frac{\sum_{i=1}^{l} N_{a,i}}{\sum_{i=1}^{l}(N_{a,i}+N_{v,i})}, \frac{\sum_{i=1}^{l} N_{v,i}}{\sum_{i=1}^{l}(N_{a,i}+N_{v,i})}\right) - 0,5\right) \\ D_{prix,l} &= (-\mathbb{1}_{Mid-P_{a,l} \leqslant P_{v,l}-Mid} + \mathbb{1}_{P_{v,l}-Mid < Mid-P_{a,l}}) \times \\ &\quad \left(\max\left(\frac{Mid-P_{a,l}}{P_{V,l}-P_{a,l}}, \frac{P_{v,l}-Mid}{P_{V,l}-P_{a,l}}\right) - 0,5\right) | \forall l > 1 \end{aligned} \tag{4.6}$$

$$\begin{aligned} P_{Volume,l} &= \sum_{i=1}^{l} (V_{a,i}+V_{v,i}) \\ P_{Nombre,l} &= \sum_{i=1}^{l} (N_{a,i}+N_{v,i}) \\ P_{Prix,l} &= \frac{1}{P_{v,l}-P_{a,l}} | \forall l > 1 \end{aligned} \tag{4.7}$$

La figure 4.3 donne pour Sanofi, les dimensions du flux informationnel mesurées aux différentes limites sur les 3 variables du carnet (le volume, le nombre et la fourchette). La table fournit les mesures d'activité construites à partir des mesures agrégées présentées dans la table 4.4.

[45] Si on considère que le côté signalé est celui où l'offre est la plus faible alors, cette formule fournit une indication cohérente puisqu'elle indique le côté où la fourchette est la plus grande.

FIGURE 4.3
Dimensions de l'information
Sanofi



Cette figure donne pour Sanofi les différentes dimensions de l'information (à gauche de chaque ligne) mesurées sur les différentes variables du carnet (en haut de chaque colonne). La valeur de limite indique la profondeur des limites utilisées pour la mesure. Dans chaque graphique, la courbe rouge continue est la moyenne alors que les courbes hachurées délimitent un échantillon de confiance empirique à 90%.

Les mesures d'hétérogénéité et de direction sont à peu près constantes pour les différentes limites. On doit remarquer que le volume fournit une hétérogénéité moyenne plus faible que le nombre et la fourchette. La direction change aussi selon ces deux variables. Le volume fournit une indication directionnelle moyenne proche de 0 alors qu'elle est positive pour le nombre et la fourchette. Étant donnée la méthode de construction des variables, cela signifie que l'offre du côté achat est en moyenne plus faible. Comme on pouvait s'y attendre, la profondeur volume et nombre évoluent de manière similaire (augmentation de la valeur moyenne et de la dispersion quand on s'enfonce dans le carnet) alors que la profondeur prix évolue en sens contraire. De plus, à l'exception du volume, les évolutions en fonction du numéro de la limite ne sont pas linéaires et ont des valeurs extrêmes.

TABLE 4.5
Dimensions de l'information de Sanofi
4 janvier 1999 10:10:36

	Hétérogénéité			Direction			Profondeur		
n°	Fou.	Vol.	Nom.	Fou.	Vol.	Nom.	Fou.	Vol.	Nom.
1	0,50	0,35	0,20	0,00	0,15	0,30	0,77	282	5
2	0,50	0,28	0,25	0,00	0,22	0,25	0,67	1084	8
3	0,43	0,30	0,30	-0,07	0,20	0,20	0,45	1130	10
4	0,46	0,38	0,31	-0,04	0,12	0,19	0,40	1300	13
5	0,41	0,25	0,33	0,09	0,25	0,17	0,23	1958	15

Cette table donne les dimensions de l'information calculées à partir des mesures agrégées présentées dans la table 4.4 en utilisant les formules 4.5 à 4.7. Les indications directionnelles utilisant la fourchette à la première limite ne sont pas utilisée dans le calcul. Ce choix découle de la valeur constante des indications d'hétérogénéité (0,5) et de direction (0) et d'un souci d'uniformité pour la profondeur.

Ces données permettent de connaître à chaque point dans le temps les caractéristiques des 5 meilleures limites à l'achat et à la vente. Ces données sont bien distinctes du flux d'ordres dont elles résultent et déterminent notre cadre d'analyse. Ces variables permettent de caractériser l'état du marché à un instant mais laissent une incertitude quand à leur interprétation (comparé à des données sur les ordres). Cette incertitude affecte autant l'interprétation de l'effet des transactions sur les meilleures limites que l'interprétation de l'état du carnet. Pour donner quelques exemples d'interprétation possibles :

– Effet d'une transaction sur le carnet

  Si la transaction consomme au moins une limite, alors un phénomène de glissement des limites en carnet se produit. De manière évidente la profondeur prix devrait diminuer du côté de la transaction mais le glissement rend impossible la prévision de l'effet d'une telle transaction sur le volume ou le nombre d'ordres). Cependant, si on considère que le volume est croissant en fonction de la profondeur en carnet (un volume plus important

est disponible à la $x+1^{\text{e}}$ limite qu'à la x^{e}) alors qu'on devrait observer une augmentation de la profondeur (activité) du côté où la transaction a eu lieu.

Si la transaction ne consomme que partiellement la première limite alors la profondeur prix est constante et on a une diminution de la profondeur en volume et peut-être en nombre.

– Interprétation des états du carnet

Une hausse de la profondeur prix ne peut être due qu'à l'introduction en carnet d'un ordre améliorant la meilleure limite. À l'inverse une baisse de la profondeur prix peut être attribuée à l'annulation d'une des meilleures limites en carnet, la réalisation d'une transaction consommant au moins une limite.

Une hausse de la profondeur activité peut être due à une transaction ayant entraîné un phénomène de glissement (cela s'accompagne d'une baisse de la profondeur prix), à l'ajout de nouveaux ordres en carnet. Une baisse de la profondeur activité pourrait être attribué à une transaction (qu'elle consomme ou non plusieurs limites), à l'introduction de nouveaux ordres améliorant la profondeur prix mais diminuant la profondeur activité.

Les quelques exemples ci-dessus soulignent que l'interprétation d'une variation de la profondeur du carnet n'est pas univoque. Cela est encore plus vrai des mesures d'hétérogénéité et de direction puisque l'effet moyen dépend de la position moyenne initiale. Les variables que nous employons conditionnent nos capacités d'analyses. Nous aurons pour objectif de comprendre comment les conditions de marché influencent ces réactions ou, plus précisément, comment l'état du carnet influence les caractéristiques des transactions.

Avant de pouvoir regrouper ces variables pour extraire au mieux la valeur des différentes dimensions de l'information via une analyse en composante principale, il convient d'examiner l'existence de saisonnalité intrajournalière. Cela permettra en outre de normaliser les variables qui auront dès lors des moyennes similaires.

4.2.3 Désaisonnalisation

On trouve dans la littérature deux grandes procédures de désaisonnalisation des données intrajournalières. Andersen & Bollerslev (1997*a*) ou Maillet & Michel (1997) norment les valeurs observées par leurs moyennes sur l'intervalle. Andersen *et al.* (2000) calculent la composante intrajournalière en utilisant des formes flexibles de Fourrier (Gallant, 1981, 1982). Ces deux procédures reposent sur l'utilisation d'un intervalle et ne sont donc pas applicables à des données Ultra Haute Fréquence. Pour ces données, la procédure de désaisonnalisation doit produire une valeur théorique à chaque point dans le temps et non pour un intervalle. Cela est généralement fait en utilisant des splines. Les splines utilisées sont des courbes polynomiales (formées de polynômes par morceaux) de degré d dont les $d-1$ premières dérivées sont continues. Les

points de discontinuité de la d^{e} dérivée sont appelés les nœuds[46]. Une spline de degré d ayant n nœuds peut être représentée sous la forme (4.8).

$$\phi(t_i) = \sum_{j=0}^{d} \alpha_i (t_i)^j + \sum_{j=1}^{n} \mathbb{1}_{t_i \geqslant n_j} \beta_j (t_i - n_j)^d \tag{4.8}$$

où t_i est l'instant de la i^{e} transaction, n_j la valeur du j^{e} nœud et $\mathbb{1}_{\cdot}$ une fonction indicatrice donc l'indice dénote l'événement ou l'ensemble concerné. Le premier terme de (4.8) représente la partie commune à toute la courbe alors que le second terme est la modification de l'ordre d à chaque nœuds[47].

Il faut alors choisir les nœuds (point de discontinuité) et le degré des polynômes. Le choix le plus répandu pour les nœuds est 10:30 11h 12h 13h 14h 15h 16h 16:30. L'utilisation de nœuds à 10:30 et 16:30 permet de tenir compte des spécificités des périodes d'ouverture et de fermeture. Concernant le degré, on trouve dans la littérature deux choix : des polynômes de degré 1 (spline linéaire) et des polynômes de degré 3 (splines cubiques). Ce choix n'est pas neutre. En effet, une spline linéaire suppose que le facteur saisonnier est monotone entre deux nœuds alors qu'une spline cubique permet de prendre en compte des dynamiques non monotones.

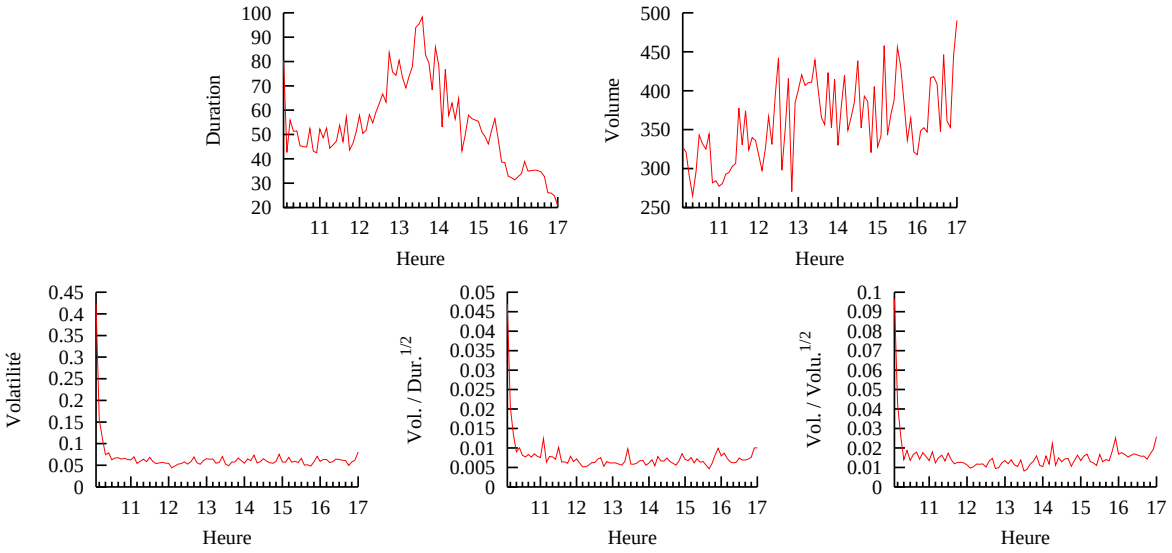
Pour procéder au choix de l'ordre de spline, nous comparons les facteurs saisonniers empiriques et ceux mis en évidence par une spline.
Les facteurs saisonniers empiriques sont estimés en calculant par intervalle de 5 minutes la moyenne des différentes variables. La figure 4.4 illustre les facteurs saisonniers empiriques obtenus pour Sanofi.
Pour calculer le facteur saisonnier théorique, nous estimons une spline sur l'ensemble des données. Un principe de parcimonie guidant nos pas et les résultats obtenus par une spline linéaire nous semblant adéquats, notre choix se portera sur celle-ci. D'un point de vue pratique l'estimation de cette spline est est faite avec la procédure TRANSREG du logiciel SAS$^{\text{TM}}$, elle consiste à calibrer sur l'ensemble des données des segments se connectant aux différents nœuds. La figure 4.5 donne les facteurs saisonniers théoriques des transactions de Sanofi estimés avec une spline linéaire.

On observe que la duration a une très forte saisonnalité intrajournalière. Les durations sont croissantes jusqu'au milieu de la journée puis décroissantes jusqu'à la clôture. Les plus fortes durations sont donc observées pendant le déjeuner. On a également une décroissance de la duration après l'ouverture. Le volume est globalement croissant au cours de la journée avec cependant deux creux aux alentours de 11h et 16h. La volatilité par transaction semble avoir

[46] Le terme de spline est parfois utilisé pour désigner des courbes polynomiales sans nœuds ou pour des courbes dont plus d'une dérivée est discontinue.

[47] Sous les contraintes de continuité que nous avons évoquées plus haut, cette spline se compose donc d'un seul polynôme jusqu'à l'ordre $d-1$ et de différents polynômes de degré d à partir de chaque nœud.

FIGURE 4.4
Facteur saisonniers empiriques des transactions
Sanofi



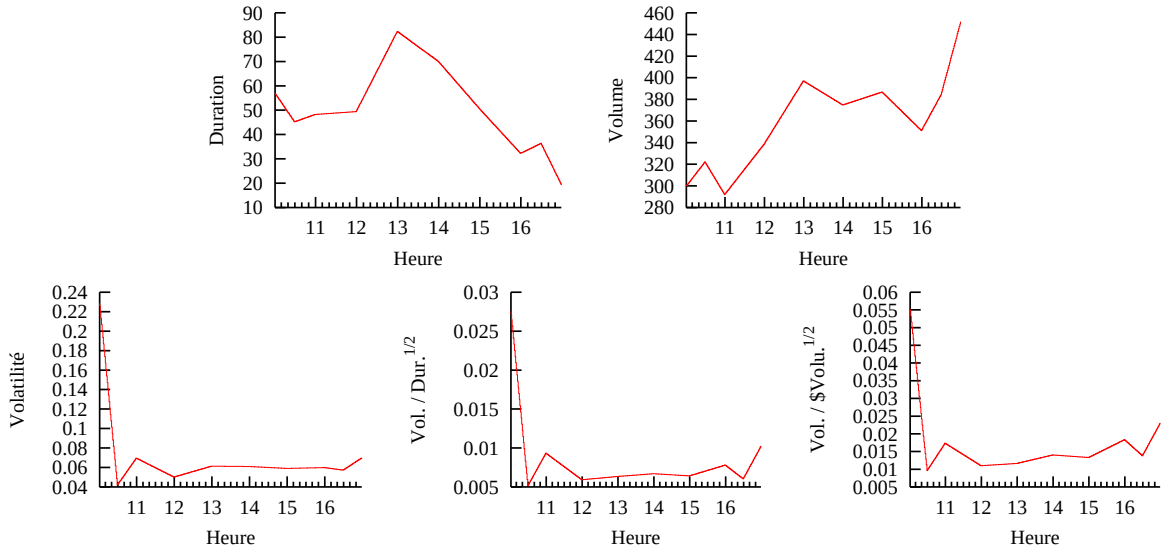
Cette figure donne les facteurs saisonniers empiriques des différentes variables. Ces derniers représentent la moyenne des variables groupées par intervalle de 5 minutes. La duration est exprimée en seconde, le volume en nombre de titres.

une moindre saisonnalité. Les différentes mesures, à l'exception d'une différence de niveau, semblent très proches, le seul effet important parait être une plus forte volatilité après l'ouverture. Cette forte volatilité en début de journée n'est pas due à la première transaction du jour dont le rendement est calculé par référence au fixage fixage. En effet, la suppression de cette donnée ne modifie pas la forme de la courbe[48]. Cette constatation explique la différence entre les configurations d'ouverture et de clôture que nous avons mise en évidence dans la première partie. L'effet d'ouverture est dû à un petit nombre de transactions ayant une volatilité particulièrement élevée alors que l'effet de clôture est dû à un grand nombre de transactions ayant une volatilité commune. C'est alors l'utilisation d'un intervalle fixe qui rendait ces deux configurations similaires. De nombreuses études (par exemple Aubier & Miloudi (2001); Maillet & Michel (1997)) excluent une certaine longueur après ouverture. Nous avons vérifié que l'exclusion des 15 premières minutes après ouverture ne modifie pas sensiblement les résultats. Cependant, cette procédure altères la représentation des facteurs saisonniers. Le lecteur trouvera en annexe F.1 les facteurs saisonniers obtenus suite au retrait du premier $1/4$ d'heure de la journée.

La comparaison des figures 4.4 et 4.5 montre qu'une courbe polynomiale linéaire est suffi-

[48]Même en supprimant les 15 premières minutes de la journée (cf. figure F.1 en annexe), on observe encore une plus forte volatilité à l'ouverture.

FIGURE 4.5
Facteurs saisonniers théoriques des transactions
Sanofi



Cette figure donne les facteurs saisonniers théoriques (calculés à l'aide d'une spline linéaire). La duration est exprimée en seconde et le volume en nombre de titres.

sante pour bien rendre compte des différents facteurs saisonniers. Suivant un principe de parcimonie, nous retenons donc une spline linéaire. Nous calculons la valeur désaisonnalisée en divisant la valeur observée par le facteur saisonnier calculé via la spline linéaire[49].

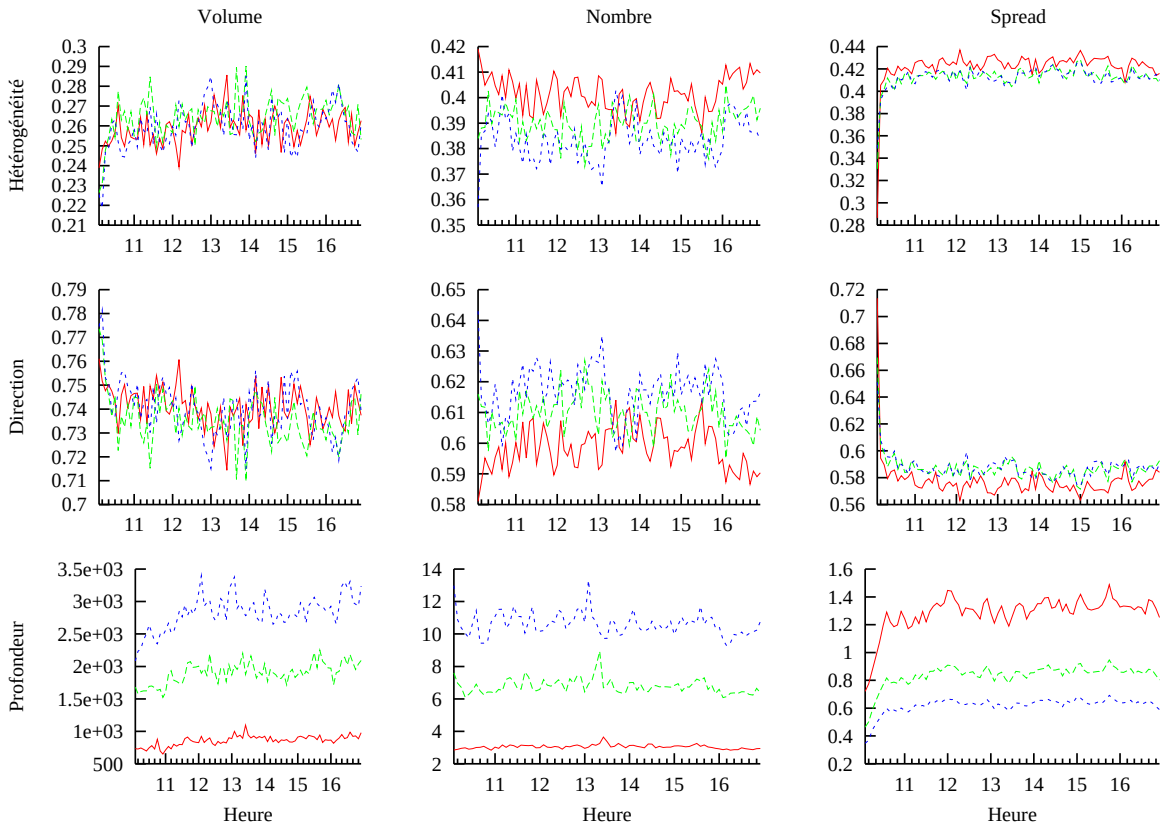
Les variables informationnelles que nous avons définies reposent sur des volumes, des nombres d'ordres ou des fourchettes dont on sait qu'ils ont de fortes saisonnalités intrajournalières. Les mesures issues du carnet d'ordres peuvent également avoir des saisonnalités intrajournalières. Nous appliquons donc la méthodologie de désaisonnalisation que nous venons d'utiliser sur les caractéristiques des transactions aux variables issues du carnet d'ordres. La figure 4.6 donne les facteurs saisonniers empiriques des mesures informationnelles.

Si on compare les facteurs saisonniers empiriques mis en évidence sur le carnet (figure 4.6) et ceux observés sur les transactions (figure 4.4) il apparaît que la saisonnalité des mesures sur le carnet est plus faible que celle des transactions. On peut noter que la forme des courbes obtenues varie aussi bien entre les dimensions qu'entre les variables.

La profondeur est globalement constante dans la journée. On remarque cependant une profondeur prix faible pendant la première demi-heure de la journée ainsi qu'une légère augmentation de la profondeur volume. La première demi-heure se caractérise donc par une profondeur prix et une profondeur volume faibles[50].

[49] Pour le rendement, nous divisons par le facteur saisonnier de la volatilité (valeur absolue du rendement).
[50] Ce résultat suggère que le fixage d'ouverture épuise une partie importante de la demande.

FIGURE 4.6
Facteurs saisonniers empiriques du carnet Sanofi



Cette figure donne pour Sanofi les facteurs saisonniers empiriques (moyenne de la variable par intervalle de 5 minutes) des différentes dimensions de l'information (à gauche de chaque ligne) mesurées sur les différentes variables du carnet (en haut de chaque colonne). La courbe rouge continue est la moyenne observée de la variable à la première limite où elle est calculable, la courbe verte (bleue) hachurée la valeur moyenne à la deuxième (troisième) limite.

La direction et l'hétérogénéité ont des variations complémentaires. L'hétérogénéité prix et volume est faible la première demi-heure alors que la direction est forte. On note également que l'hétérogénéité nombre suit globalement un $\smile$ au cours de la journée. Il semble donc que l'hétérogénéité nombre soit au plus bas en milieu de journée.

Étant donné que les tests (paramétriques et non paramétriques) confirment tous la significativité des variations intrajournalières, nous désaisonnalisons les données. Parmi les différentes dimensions, seule la direction (comme le rendement pour les transactions) peut être négative, c'est d'ailleurs une mesure sensée être liée au rendement. Il est alors logique de procéder comme pour les rendements en calibrant la spline sur la valeur absolue puis en divisant la valeur observée par la valeur absolue théorique. La figure 4.7 donne les facteurs saisonniers théoriques calculés à l'aide d'une spline linéaire.

Les facteurs saisonniers théoriques semblent bien refléter les variations intrajournalières puisque l'on retrouve les grandes évolutions que nous avions notées. Par conséquent, à l'instar des transactions, une spline linéaire nous paraît suffisante pour désaisonnaliser les données issues des 5 meilleures limites.

La figure souligne en outre que la saisonnalité d'une même variable à différentes limites est similaire, qu'on observe bien la relation attendue entre le numéro de la limite et la profondeur : corrélation positive (négative) avec la profondeur volume et nombre (prix) et que les autres variables sont globalement de mêmes niveaux aux différentes limites.

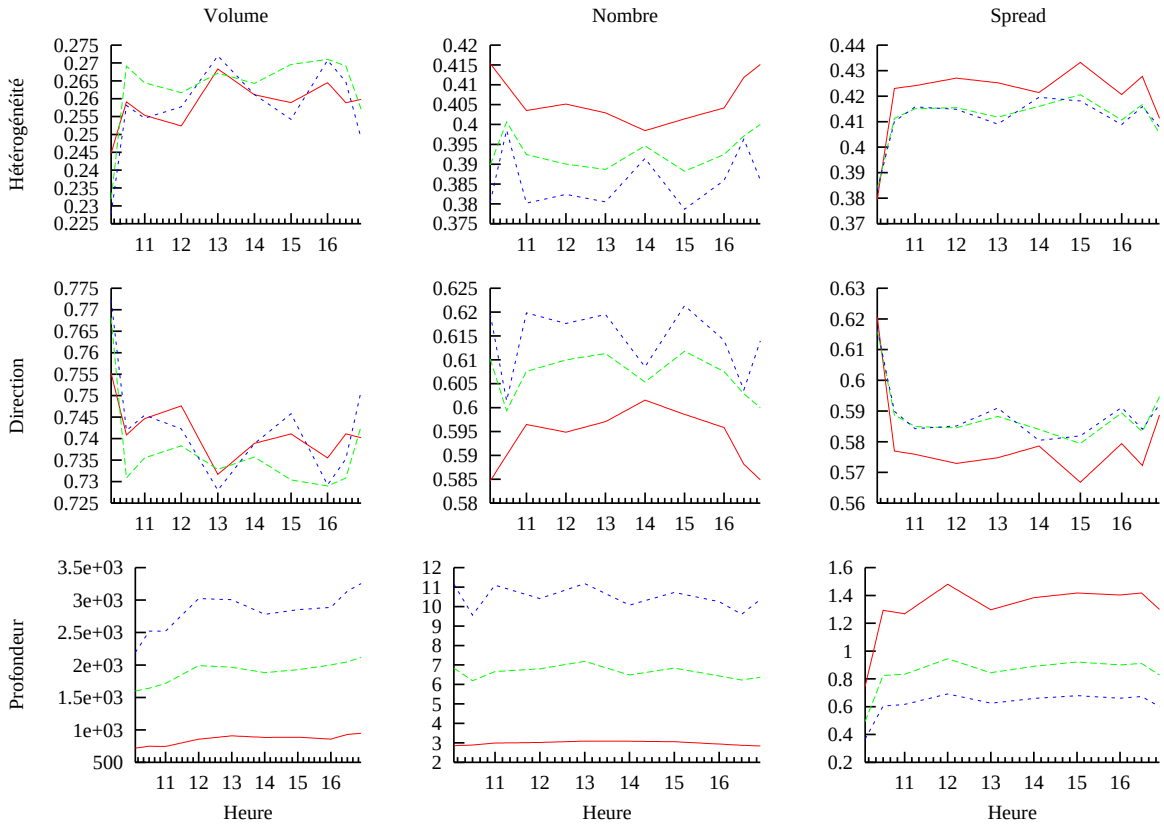
4.3 Description des données

La figure 4.8 donne l'évolution des prix de chacun des titres au cours de la période d'étude.

Cette période d'étude se caractérise par de forts mouvements à la hausse et à la baisse. Les évolutions peuvent être classées en deux grands groupes. St-Gobain, Schneider et Accor ont une tendance haussière (environ +20% pendant l'étude) alors que Lyonnaise, Sanofi et Thomson ont une tendance baissière (de l'ordre de -10% sur la période). Les évolutions haussières sont assez uniformes. Elles sont toutes entrecoupées de périodes de chute et éventuellement de plateau (ce phénomène est patent pour Schneider). Les évolutions baissières sont plus diverses. Pour Sanofi et Lyonnaise, la tendance baissière est entrecoupée de rebonds alors que Thomson est stable sur le premier tiers de la période puis chute vivement et se stabilise. Lagardère est globalement stable sur la période. Ce titre a connu une forte chute suivie d'un rebond en fin de période.

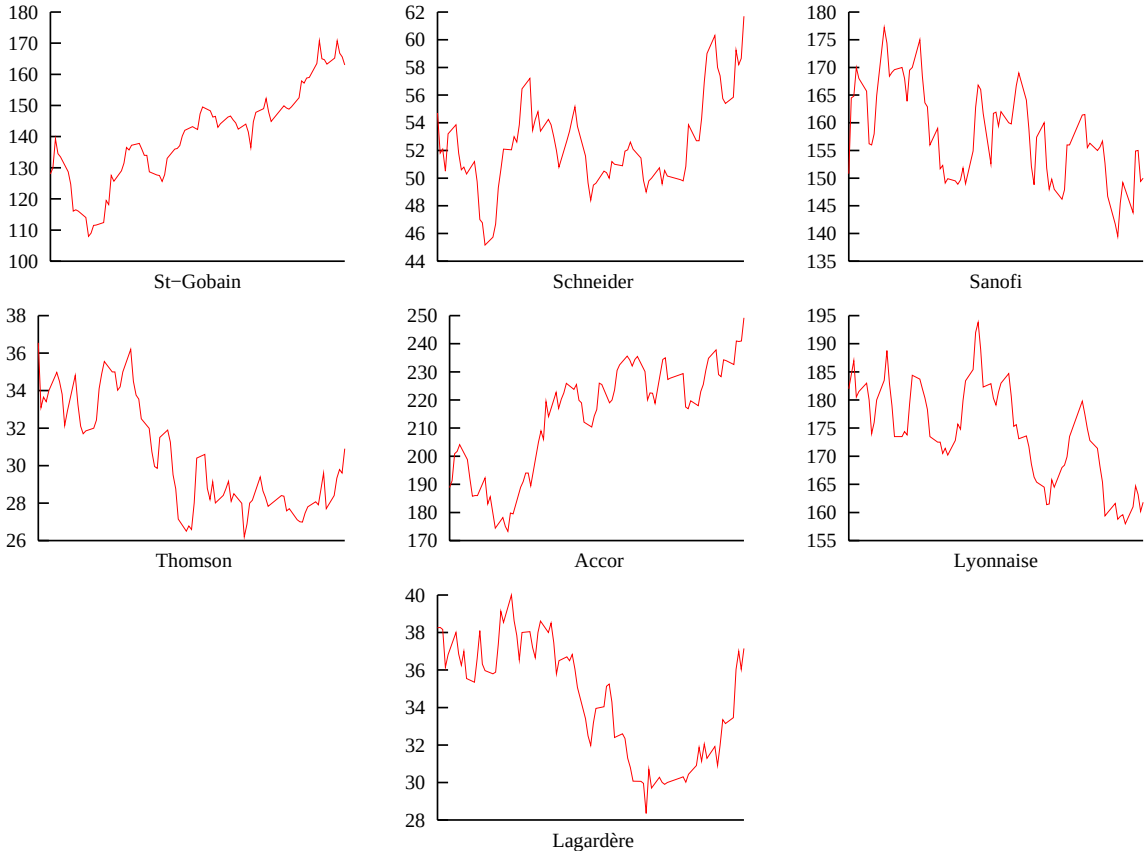
Les sous-sections 4.3.1 et 4.3.2 décrivent l'influence des retraitements respectivement sur les rendements et l'activité des titres. La description des variables du carnet sera présentée dans le chapitre 6.

FIGURE 4.7
Facteurs saisonniers théoriques du carnet Sanofi



Cette figure donne pour Sanofi les facteurs saisonniers théoriques (obtenus avec une spline linéaire) des différentes dimensions de l'information (à gauche de chaque ligne) mesurées sur les différentes variables du carnet (en haut de chaque colonne). La courbe rouge continue est la moyenne observée de la variable à la première limite où elle est calculable, la courbe verte (bleue) hachurée la valeur moyenne à la deuxième (troisième) limite.

FIGURE 4.8
Évolution des titres



Cette figure donne pour les différents titres (le nom est repris en dessous de chaque graphique), l'évolution du cours (en €) pendant la période d'étude (du 1^{er} janvier 1999 au 30 avril 1999).

4.3.1 Rendements

Les tables 4.6 à 4.8 donnent pour chacune des mesures de rendements les caractéristiques des données brutes et celles des données après désaisonnalisation.

TABLE 4.6
Rendements observés et désaisonnalisés

	μ	σ	Min	Max	ρ_1	$Q(20)$
Sanofi	-0,0006	0,0006	-3,702	3,312	-0,120	739
	-0,0070	0,0094	-26,865	31,101	-0,081	406
Accor	-0,0001	0,0004	-3,138	3,247	-0,171	1 802
	-0,0019	0,0085	-25,943	34,447	-0,102	701
Lyonnaise	-0,0003	0,0003	-2,725	3,296	-0,150	1 627
	-0,0072	0,0078	-32,777	56,639	-0,082	536
Thomson	-0,0023	0,0013	-5,129	7,066	-0,216	1 433
	-0,0278	0,0135	-34,757	46,709	-0,164	889
Schneider	-0,0002	0,0005	-6,331	3,650	-0,191	2 562
	-0,0041	0,0085	-43,401	36,620	-0,122	1 108
St-Gobain	0,0004	0,0004	-3,429	3,773	-0,152	1 660
	0,0092	0,0078	-27,201	32,748	-0,090	645
Lagardère	-0,0012	0,0006	-5,222	4,770	-0,293	4 283
	-0,0283	0,0104	-56,651	33,646	-0,216	2 359

La première ligne de chaque titre donne les statistiques sur les données brutes, la seconde ligne sur les données désaisonnalisées. μ (σ) est la moyenne des rendements (l'écart-type). ρ_1 désigne la corrélation d'ordre 1 et $Q(20)$ la statistique de Ljung-Box sur les 20 premiers décalages de la série. Cette statistique suit une loi du χ^2 à 20 degrés de liberté et a pour valeur critique au seuil de 1% 37,57. Elle est donc toujours extrêmement significative et montre donc l'importance de la corrélation des séries.

La moyenne des rendements bruts pourrait sembler étrange au regard de la figure 4.8. En effet, même des valeurs globalement haussières comme Schneider ont un rendement moyen négatif. Il faut pour les comprendre se rappeler que les rendements ne sont calculés qu'à l'intérieur de la journée : ni les variations entre le prix de clôture et le prix d'ouverture, ni celles entre le début (la fin) du continu et le fixages d'ouverture (de clôture) ne sont prises en compte. De plus, comme le rendement est calculé par transaction, on peut comprendre la faiblesse des valeurs obtenues.

La désaisonnalisation du rendement a un effet important sur l'autocorrélation qui diminue fortement (effet similaire à celui observé sur les durations). Celle-ci a également pour effet d'augmenter la moyenne et la dispersion des séries.

La désaisonnalisation des rendements normalisés par la duration a des effets similaires à ceux observés sur les rendements. Toutefois, il est intéressant de noter que ces rendements ont une corrélation beaucoup plus faible que les précédents tant pour les données brutes que pour les données désaisonnalisées. Ce fait suggère donc qu'une partie de la corrélation des rendements

TABLE 4.7
Rendements observés et désaisonnalisés
normalisation par la duration

	μ	σ	Min	Max	ρ_1	$Q(20)$
Sanofi	-0,0002	0,0002	-1,647	1,118	-0,076	426
	-0,0164	0,0130	-115,285	70,864	-0,071	374
Accor	-0,0002	0,0001	-1,589	1,744	-0,049	275
	-0,0235	0,0113	-75,504	82,792	-0,054	326
Lyonnaise	-0,0003	0,0001	-1,202	0,984	-0,069	389
	-0,0377	0,0097	-54,740	122,250	-0,050	235
Thomson	-0,0007	0,0004	-1,991	2,051	-0,109	453
	-0,0401	0,0190	-80,596	105,418	-0,082	303
Schneider	-0,0001	0,0002	-1,379	2,581	-0,098	789
	-0,0165	0,0107	-69,035	63,495	-0,098	819
St-Gobain	-0,0001	0,0001	-1,291	1,134	-0,071	445
	-0,0031	0,0101	-80,014	55,443	-0,058	314
Lagardère	-0,0004	0,0002	-1,359	3,373	-0,177	1 606
	-0,0515	0,0138	-86,802	118,504	-0,168	1 488

La première ligne de chaque titre donne les statistiques sur les données brutes, la seconde ligne sur les données désaisonnalisées. μ (σ) est la moyenne des rendements (l'écart-type). ρ_1 désigne la corrélation d'ordre 1 et $Q(20)$ la statistique de Ljung-Box sur les 20 premiers décalages de la série. Cette statistique suit une loi du χ^2 à 20 degrés de liberté et a pour valeur critique au seuil de 1% 37,57. Elle est donc toujours très significative.

provient de la corrélation de la duration.

La normalisation des rendements par les volumes souligne l'influence de ces derniers : seuls Accor, Schneider et Lagardère ont des rendements moyens négatifs, ce qui contraste fortement avec les résultats présentés dans la table 4.6. Les changements de signes questionnent la relation entre le volume et le rendement. En effet, le fait qu'un rendement moyen soit négatif alors qu'un rendement moyen normalisé par le volume soit positif peut être causé par une association soit entre des rendements négatifs et des volumes importants, soit entre des rendements positifs et les petits volumes.

4.3.2 Activité

Les tables 4.9 et 4.10 fournissent respectivement les caractéristiques (avant et après désaisonnalisation) de la duration et du volume.

La duration moyenne est assez variable selon les titres. Les titres les plus actifs (Lyonnaise, St-Gobain, Schneider) ont une duration moyenne voisine de 30 secondes alors que pour les titres les moins actifs (Lagardère, Sanofi, Thomson), la duration moyenne va de 42 à 74 secondes. Il y a donc de fortes disparités entre les titres. Conformément à la théorie, les durations

TABLE 4.8
Rendements observés et désaisonnalisés
normalisation par le volume

	μ	σ	Min	Max	ρ_1	$Q(20)$
Sanofi	0,0000	0,0001	-0,525	0,740	-0,058	218
	0,0050	0,0138	-54,256	81,607	-0,057	193
Accor	-0,0001	0,0001	-0,563	0,544	-0,093	580
	-0,0047	0,0116	-77,426	56,796	-0,075	396
Lyonnaise	0,0000	0,0001	-0,970	0,602	-0,103	815
	0,0023	0,0112	-77,463	83,902	-0,071	421
Thomson	-0,0004	0,0003	-2,811	3,969	-0,208	2 497
	-0,0550	0,0292	-400,411	294,024	-0,149	1 047
Schneider	0,0000	0,0001	-0,587	0,445	-0,098	797
	0,0041	0,0123	-84,911	100,846	-0,073	463
St-Gobain	0,0001	0,0001	-1,173	0,966	-0,095	735
	0,0107	0,0113	-79,249	85,064	-0,073	479
Lagardère	-0,0002	0,0001	-0,691	0,646	-0,122	839
	-0,0461	0,0156	-116,430	96,023	-0,102	638

La première ligne de chaque titre donne les statistiques sur les données brutes, la seconde ligne sur les données désaisonnalisées. μ (σ) est la moyenne des rendements (l'écart-type). ρ_1 désigne la corrélation d'ordre 1 et $Q(20)$ la statistique de Ljung-Box sur les 20 premiers décalages de la série. Cette statistique suit une loi du χ^2 à 20 degrés de liberté et a pour valeur critique au seuil de 1% 37,57. Elle est donc toujours très significative.

désaisonnalisées ont une moyenne proche de 1. La désaisonnalisation réduit la dispersion des données et entraîne une baisse de la corrélation. Cependant, même après désaisonnalisation, celle-ci demeure très significative. On observe une très forte dispersion de la statistique $Q(20)$ pour les différents titres que la désaisonnalisation ne parvient pas à réduire.

La taille moyenne des transactions est très variable selon les titres : de 219 pour Accor à 638 pour Lagardère. Il n'y a pas de lien évident entre la duration moyenne (l'intensité des transactions) ou l'évolution du titre pendant la période et la taille moyenne des transactions. On peut remarquer que, à la différence des durations, toutes les statistiques d'autocorrélation ont le même ordre de grandeur. Cependant, la désaisonnalisation du volume a beaucoup moins d'effet sur la corrélation de la série puisque nous n'observons qu'une très légère diminution de la statistique $Q(20)$ et la corrélation d'ordre 1 (ρ_1) est identique ou légèrement inférieure.

4.4 Conclusion

Ce chapitre nous a permis dans un premier temps d'apprécier la pertinence et l'importance des concepts d'asymétrie d'information et d'hétérogénéité. Nous avons également pu revoir les modèles théoriques et les études empiriques qui guiderons nos attentes dans les prochains

TABLE 4.9
Durations observées et désaisonnalisées

	μ	σ	Min	Max	ρ_1	$Q(20)$
Sanofi	48,495	0,344	1,000	1 140,000	0,171	9 016
	1,000	0,007	0,012	19,464	0,128	3 888
Accor	35,882	0,209	1,000	1 262,000	0,177	17 747
	1,000	0,005	0,015	23,581	0,141	7 838
Lyonnaise	29,953	0,151	1,000	778,000	0,157	15 362
	1,000	0,005	0,018	18,368	0,108	4 893
Thomson	74,124	0,689	1,000	2 448,000	0,215	8 840
	1,000	0,009	0,008	26,726	0,173	4 616
Schneider	32,683	0,183	1,000	1 013,000	0,211	26 658
	1,000	0,005	0,017	31,796	0,173	15 351
St-Gobain	30,825	0,164	1,000	768,000	0,182	22 275
	1,000	0,005	0,018	17,520	0,144	11 654
Lagardère	42,930	0,281	1,000	1 224,000	0,191	17 442
	1,000	0,006	0,012	24,447	0,151	8 687

La première ligne de chaque titre donne les statistiques sur les données brutes, la seconde ligne sur les données désaisonnalisées. La duration est exprimée en nombre de secondes. μ (σ) est la moyenne (l'écart-type). ρ_1 désigne la corrélation d'ordre 1 et $Q(20)$ la statistique de Ljung-Box sur les 20 premiers décalages de la série. Cette statistique suit une loi du χ^2 à 20 degrés de liberté et a pour valeur critique au seuil de 1% 37,57. Elle est donc toujours très significative.

chapitres.

Nous avons ensuite construit, à partir des données sur les 5 meilleures limites, des variables dérivées des modélisations théoriques et interprétables en terme de caractéristiques du flux informationnel (direction, hétérogénéité et profondeur).

Ce chapitre nous également permis de préparer la base de données pour une analyse tick par tick en opérant les retraitements et désaisonnalisations nécessaires. Nous avons apprécié l'importance de ces procédures dans le traitement des données désagrégées. Celles-ci font ressortir qu'à ce niveau de désagrégation, les caractéristiques des transactions et du carnet ont des saisonnalités fortes qui éclairent certains des résultats obtenus dans la précédente partie.

Au total, la pré-analyse des caractéristiques des transactions souligne le lien entre les fortes périodicités intrajournalières et, notamment, la forte autocorrélation des séries. Parmi les séries traitées, on peut remarquer que c'est la duration qui a la plus forte périodicité intrajournalière et pour laquelle on note la plus grande dispersion de la statistique de Ljung-Box (de 8 800 à 26 600 sur données brutes et de 3 900 à 15 400 sur données retraitées). C'est également sur cette série que la désaisonnalisation a le plus d'effet. Ces faits suggèrent que la dynamique de la duration peut être différente selon les titres. On doit également noter qu'il n'y a pas de lien évident entre les caractéristiques des séries (brutes ou désaisonnalisées) et l'évolution globale du titre. Enfin, les variations observées en fonction de la normalisation appliquée aux séries de

TABLE 4.10
Volumes observés et désaisonnalisés

	μ	σ	Min	Max	ρ_1	$Q(20)$
Sanofi	362,151	3,071	1,000	25 000,000	0,072	1 325
	1,000	0,008	0,002	67,460	0,069	1 151
Accor	219,223	2,031	1,000	20 000,000	0,077	2 928
	1,000	0,009	0,003	92,020	0,074	2 545
Lyonnaise	433,523	3,650	1,000	34 817,000	0,066	1 783
	1,000	0,008	0,002	92,244	0,059	1 275
Thomson	591,210	7,457	1,000	60 000,000	0,071	1 055
	1,000	0,013	0,001	101,463	0,069	1 025
Schneider	630,042	6,571	1,000	62 500,000	0,055	2 047
	1,000	0,010	0,001	103,810	0,055	1 888
St-Gobain	361,517	3,237	1,000	30 000,000	0,069	1 795
	1,000	0,009	0,002	82,074	0,068	1 585
Lagardère	637,990	6,987	1,000	54 500,000	0,081	2 674
	1,000	0,011	0,001	86,677	0,081	2 380

La première ligne de chaque titre donne les statistiques sur les données brutes, la seconde ligne sur les données désaisonnalisées. Le volume est exprimé en nombre de titres. μ (σ) est la moyenne (l'écart-type). ρ_1 désigne la corrélation d'ordre 1 et $Q(20)$ la statistique de Ljung-Box sur les 20 premiers décalages de la série. Cette statistique suit une loi du χ^2 à 20 degrés de liberté et a pour valeur critique au seuil de 1% 37,57. Elle est donc toujours très significative.

rendements suggèrent l'existence de relations complexes entre duration et rendement ou volume et rendement qui peuvent varier selon les titres.

Chapitre 5

Relations entre les composantes du processus de transactions

Avant d'analyser les différentes modélisations économétriques que l'on peut employer pour étudier la relation dynamique entre les caractéristiques ultra haute fréquence des transactions, la table 5.1 synthétise, pour les principaux modèles présentés dans la section 4.1, les signes des corrélations attendues entre les variables étudiées.

TABLE 5.1
Implications des différents modèles

	d_t, σ_t	v_t, σ_t	d_t, v_t
Easley O'Hara et al.	-	+	-
Diamond & Verrecchia (1987)	+	?	?
Admati & Pfleiderer (1988)	-	+	-
Back & Pedersen (1998)	-	+	-
Hétérogénéité	-	+	-

Cette table rassemble les prévisions des principaux modèles déjà présentés concernant le signe des relations. ? désigne l'absence de conclusions du modèle sur ce point. Seules les conclusions concernant les points étudiés à savoir les relations entre la volatilité d'une part et le volume ou la duration d'autre part sont présentées. Concernant les autres prévisions, nous invitons le lecteur à se reporter aux articles concernés. Par abus de langage, nous désignons sous le label Easley, O'Hara et al. les différents modèles produits par ses deux coauteurs avec éventuellement d'autres chercheurs.

L'ensemble des modèles proposés par Easley, O'Hara et d'autres auteurs converge vers une image où l'augmentation de la fréquence de transaction (la diminution des durations) révèle la présence d'informations et donc d'informés ; ceux-ci préférant traiter des quantités plus importantes[1]. Le modèle de Diamond & Verrecchia (1987) propose une interprétation légèrement

[1]D'une manière générale, nous avons pu voir que les résultats empiriques supportaient le rôle attribué à la duration mais que celui du volume était plus sujet à caution.

différente dans la mesure où l'existence de contraintes sur les ventes à découvert modifie l'interprétation d'une diminution de la fréquence de transaction : une mauvaise nouvelle[2]. L'interprétation d'une très faible duration devrait alors coïncider avec une bonne nouvelle. Même si l'explication proposée par Admati & Pfleiderer (1988) est toute différente puisque la concentration provient des non-informés et attire les informés, les prévisions sont assez proches et donc difficiles à distinguer des modèles de Easley, O'Hara. Back & Pedersen (1998) proposent sur les transactions des conclusions similaires au précédent modèle. Enfin, d'une manière générale, l'hétérogénéité suggère une association positive entre volatilité, volume et intensité des transactions.

On peut d'ores et déjà remarquer les limites d'une étude portant exclusivement sur les transactions. En effet, il apparaît que plusieurs modèles proposent sur le triplet duration - volume - volatilité des prévisions similaires. Il ne semble donc pas qu'il faille attendre d'une étude des caractéristiques de transactions une discrimination très fine entre ces différents modèles. Nous verrons cependant que cette étude apporte, à nos yeux, des connaissances nouvelles sur le fonctionnement du marché parisien et les caractéristiques de transactions qui en découlent.

Le choix entre les différentes causes avancées est d'autant plus compliqué que les relations temporelles entre ces variables sont difficiles à déterminer étant donné que leur dépendance au flux informationnel peut induire des dynamiques différentes[3]. Nous adopterons par la suite une méthodologie principalement exploratoire afin d'essayer de comprendre ces relations. Comme nous l'avons déjà signalé, les modèles de microstructure ou même les études empiriques ne donnent pas une image univoque de ces relations. Aussi la méthodologie d'analyse est-elle d'autant plus cruciale.

Nous avons débuté le précédent chapitre en présentant le processus de transaction comme un "mark point process" où les points sont les instants de réalisation (de transaction) alors que les marques sont les caractéristiques de ces réalisations (prix, volume). Lorsqu'on considère les processus sous cette forme, il est possible d'étudier non seulement les relations entre les marques (i.e. entre les rendements et les volumes) mais aussi avec les points (i.e. la duration).

Ce chapitre sera organisé comme suit. Nous commencerons par présenter les méthodologies d'analyse des données ultra haute fréquence dans la section 5.1. Nous analyserons ensuite progressivement les différentes composantes du processus de transaction (duration, volume, volatilité). La section 5.2 présentera les résultats de modélisations univariées. Nous présenterons ensuite deux modélisations multivariées. Là encore nous adopterons une démarche progressive. La section 5.3 examinera l'influence d'un paramètre de l'activité sur le rendement à l'aide

[2]Ce signal étant directionnel, il a également des effets négatifs sur les rendements attendus.

[3]Par exemple, dans le précédent chapitre, nous avons mentionné que dans le modèle de He & Wang (1995) le volume pouvait être autocorrélé même si le flux informationnel ne l'était pas ; nous avons aussi noté que certaines variables (du carnet il est vrai) réagissaient de manière différente (Dupont, 2000) à l'asymétrie d'information ou encore que les transactions et le carnet ne réagissaient pas aux mêmes événements (Coppejans & Domowitz, 2002).

du modèle d'Engle (2000)[4]. Nous emploierons enfin la modélisation trivariée proposée par Manganelli (2002) afin d'examiner l'influence de l'activité (la duration **ET** le volume) sur les rendements.

5.1 Méthodologies d'analyse des rendements ultra haute fréquence

On trouve dans la littérature deux types d'approches pour modéliser le processus de transaction.

La première optique d'analyse repose sur une modélisation du processus de rendement. Les rendements sont alors considérés comme des variables discrètes. L'influence de variables d'activité (la duration ou le volume) peut être une composante à part entière de la modélisation (Russell & Engle, 2002) ou être intégrée comme variables explicatives (Rydberg & Shepard, 2003; Hausman *et al.*, 1992).

Rydberg & Shepard (2003) décomposent les variations de prix en trois composantes : l'occurrence d'un changement de prix, son signe et son amplitude. Ils modélisent ensuite dans un cadre multivarié chacune des composantes avec des modèles Generalized Linear Auto Regressive Moving Average (GLARMA). Ils analysent ensuite l'influence de différentes variables (les logarithmes du volume et de la duration, des variables muettes de l'heure du jour, du jour de la semaine, du mois de l'année…) sur les composantes du processus de prix en les intégrant en tant que variables exogènes. Ils estiment cette modélisation sur les prix de transaction d'IBM pendant l'année 1995. Ils mettent en évidence que les variables d'activité ont un impact faible ou non significatif sur le signe du rendement de la transaction alors qu'elles ont un effet important sur l'occurrence d'un changement de prix ou son amplitude. L'influence des valeurs courantes (du volume et de la duration) est toujours positive. Cela signifie qu'observer une duration ou un volume important augmente la probabilité d'un changement de prix et son l'amplitude. L'influence des variables d'activité persiste pendant plusieurs décalages. Les auteurs notent également que : - la duration passée a un impact négatif ; - l'inclusion de variables explicatives faire chuter la persistance temporelle des modèles et apporte une amélioration significative de la vraisemblance. On peut remarquer que les principales relations démontrent l'influence de l'activité sur la volatilité (l'occurrence et l'amplitude des changements de prix).

Russell & Engle (2002) modélisent les changements de prix avec un modèle Auto regressive Conditionnal Multinomial (ACM). Succinctement, la densité jointe des durations et des prix est décomposée entre la densité des durations et la densité conditionnelle des changements de prix. La distribution marginale des durations est modélisée par un ACD alors que les changements de

[4]Cette modélisation est bivariée : les rendements et un paramètre d'activité.

prix sont considérés comme des variables multinomiales où les probabilités de transition d'un prix à l'autre sont autorégressives. Ce modèle est évalué sur les transactions d'un titre (Airga) traité sur le NYSE pendant l'année 1999. Ils observent que la volatilité s'accroît avec la duration mais que cet accroissement n'est observé que pour les petites valeurs de la duration : dès que la duration est supérieure ou égale à sa moyenne, la volatilité est constante. Ce dernier point serait cohérent avec les prédictions de Easley & O'Hara (1992) où une absence de transactions indique l'absence d'informations et diminue donc la volatilité[5]. Ils obtiennent un résultat compatible avec le modèle de Diamond & Verrecchia (1987) où les durations longues sont associées à des prix décroissants.

La précédente modélisation peut être vue comme une extension du modèle de Hausman *et al.* (1992) en incluant une dynamique temporelle. Dans cet article, Hausman *et al.* (1992) proposent une modélisation des changements de prix par un modèle probit qu'ils testent sur 100 titres choisis aléatoirement dans la base de données ISSM en 1988. Ils mettent en évidence que l'impact prix d'une transaction augmente de manière non-linéaire avec son volume mais soulignent également l'influence importante de l'ordre des événements[6].

Darolles *et al.* (2000) modélisent les changements de prix avec des chaînes de Markov prenant pour valeur -1, 0, 1[7]. Ils développent plusieurs outils (échelle, vitesse, décomposition spectrale de la matrice de transition) permettant l'analyse de la distribution jointe des durations et des changements de prix. Ces techniques sont appliquées à Elf-Aquitaine et soulignent, en autres, que l'absence de transactions est considérée comme une mauvaise nouvelle corroborant ainsi le modèle de Diamond & Verrecchia (1987).

Les études présentées précédemment[8] se focalisent principalement sur le processus de prix et reposent sur un modèle structurel (stylisé ou économétrique) des changements de prix discrets. À l'inverse, une seconde optique considère que le processus de prix est continu et analyse alors les relations entre les variables d'activité et le processus de rendement.

Le modèle UFH-GARCH proposé par (Engle, 2000) ou (Ghysels & Jasiak, 1998) modélisent les relations entre l'intensité des transactions et la volatilité. La duration est tout d'abord modélisée via un ACD. Ensuite, le rendement est traité comme une variable continue dont la volatilité dépend non-linéairement de la duration. La principale différence entre ces deux ap-

[5]Ce fait peut sembler contre-intuitif. Mais, si on considère que la duration augmente alors la volatilité est constante, il y a bien une diminution de la volatilité par unité de temps, d'où l'accord que nous avons mis en avant.

[6]L'impact prix d'une transaction varie en fonction des événements qui la précèdent. Ce résultat est en accord avec la nature non markovienne du processus de prix soulignée par de nombreux modèles de microstructure (cf. page 98).

[7]À cette fin, après avoir signé les transactions et pris en compte la danse de la fourchette, ils décomposent les transactions entraînant une variation de n ($n > 1$) ticks en n variations de prix de 1 tick réparties en divisant la duration par n.

[8]Meddahi *et al.* (1998) proposent également une modélisation structurelle des rendements à ultra haute fréquence. Cette modélisation n'ayant pas, à notre connaissance été appliquée, nous ne la présenterons pas.

proches est que Ghysels & Jasiak (1998) utilisent un GARCH robuste à l'agrégation temporelle[9] alors qu'Engle (2000) introduit les paramètres de la duration dans l'équation de la volatilité conditionnelle. Ghysels & Jasiak (1998) estiment leur modélisation ACD(2,2)-GARCH sur les transactions d'IBM pendant le mois de novembre 1993. Ils observent une disparition de la persistance de la volatilité ce qui suggère que cette dernière serait induite par la persistance de la duration. Cependant, un test de Granger (1969) souligne l'interdépendance de ces deux variables. La généralité de ces résultats est remise en question par Aubier & Miloudi (2001). Les auteurs estiment le même modèle[10] sur Accor du 1^{er} juillet au 30 septembre 1998 et observent également une très forte réduction de la persistance de la volatilité qui toutefois demeure significative. De plus, le test de Granger (1969) qu'ils utilisent rejette l'hypothèse d'absence de liens de causalité de la duration vers la volatilité. Le lien de causalité semble donc unidirectionnel de la duration vers la volatilité. Engle (2000) estime son modèle sur IBM de novembre 1990 à janvier 1991. Il met en évidence une réduction importante de la persistance de la volatilité lorsque l'influence de la duration est prise en compte. Avec un modèle considérant que la duration n'a pas de contenu informationnel, la persistance est de 0,867 alors qu'elle chute à 0,461 lorsque plusieurs variables représentant l'influence de la duration sont incluses.

Dufour & Engle (2000) étendent le modèle de Hasbrouck (1991) en incluant le temps entre les transactions parmi les déterminants de l'impact prix des transactions. Ils étudient 18 des titres les plus actifs sur le NYSE de novembre 1990 à janvier 1991 et mettent en évidence que, lorsque la duration décroît, l'impact prix, la rapidité d'ajustement des prix et l'autocorrélation positive des prix s'accroissent tous. En rapprochant ces résultats de ceux de Hasbrouck (1991), cela suggère que les périodes de forte activité sont associées à des fourchettes larges, des volumes importants et un fort impact prix des transactions. Sur ce même marché, Spierdijk (2002), et Zebedee (2001) obtiennent, à différentes périodes, des résultats convergents. Les périodes de forte activité semblent donc associées à la présence d'informés. Ces constations sont à rapprocher de celles faites Zhang *et al.* (2001) qui, avec une modélisation ACD à seuil (TACD), mettent en évidence que la dynamique est différente selon que l'activité de la période est faible ou forte. Ces changements structurels sont de plus rapprochables d'événements identifiables.

Manganelli (2002) propose une modélisation VAR des trois variables (rendement, volume duration). Sous certaines hypothèses[11], cette modélisation se réduit à des modèles de type GARCH incluant des variables exogènes dans l'équation de la volatilité conditionnelle. En étudiant 10 titres traités sur le NYSE (5 titres appartiennent au second décile du nombre de

[9] Ils suivent donc l'intuition de Drost & Nijman (1993) et estiment un GARCH faible sur le plus petit intervalle possible (la seconde).

[10] Ils opèrent une légère modification puisqu'ils emploient un Log-ACD(2,2) pour la duration car ce modèle est plus performant dans leur échantillon.

[11] Ces hypothèses qui concernent principalement l'exogénéïté faible des différentes variables sont également faites par la majorité des études citées jusqu'ici.

transactions et 5 autres au 8^{e} décile) du 1er janvier 1998 au 30 juin 1999. Il met en évidence que le volume est persistant pour les titres fortement traités alors qu'il ne l'est pas pour les titres peu actifs. La relation déjà observée à basse fréquence entre le volume et la volatilité est également observée à ultra haute fréquence. Il montre en plus que la duration entre les transactions ne peut être considérée exogène que pour les titres peu actifs. Enfin, le temps de retour à l'équilibre après un choc est d'autant plus court que l'activité est importante.

Les précédentes études sont principalement exploratoires[12] et s'accordent sur certains points. Elles reconnaissent toutes que la variable étudiée (la duration et/ou le volume) a un contenu informationnel. Elles suggèrent également l'existence de périodes conjuguant de faibles durations (une forte intensité des transactions), d'importants volumes et une volatilité élevée. Cela signifierait donc qu'il y a des périodes très actives d'arrivée d'information (et d'informés) alors que les périodes calmes seraient dominées par les non-informés. La majorité des études concernent le NYSE (un marché à structure mixte) et supportent globalement le modèle de Easley & O'Hara (1992).

Les quelques études sur des marchés dirigés par les prix conduisent à des conclusions sensiblement différentes. L'analyse de ces relations sur un marché dirigé par les ordres, tel que la Bourse de Paris, permettra donc d'enrichir les connaissances empiriques sur les relations entre la duration le volume et la volatilité. Cela offrira la possibilité de vérifier si les relations déjà observées sont propres à la structure du NYSE ou, si elles sont similaires sur un marché dirigé par les ordres.

En raison de leur flexibilité, du caractère éprouvé des modèles GARCH et de leur simplicité d'utilisation, nous emploierons les modélisations proposées par Engle (2000) et Manganelli (2002). Ces deux modélisations emploient des modèles de type GARCH avec ou sans variables exogènes dans l'équation de la volatilité conditionnelle. Les estimations que nous réaliserons par la suite ont plusieurs points communs :

– Nous utilisons la procédure Autoreg du logiciel SAS^{TM}. Les écarts-types présentés sont robustes puisqu'ils tiennent compte du fait que l'estimation est réalisée sous hypothèse du pseudo maximum de vraisemblance[13]. On doit cependant remarquer que les écart-types ainsi obtenus sont moins conservatifs que ceux produits par la méthode de Bollerslev & Wooldridge (1992). Toutefois, nous n'avons pas à disposition de logiciel offrant à la fois une souplesse de modélisation (inclusion de variables exogènes) et calculant les écarts-types de Bollerslev & Wooldridge (1992) [14].

[12]Hafner (1999) met en oeuvre une analyse purement exploratoire. Il utilise une modélisation ACD pour les durations puis des modèles non-paramétriques à seuil pour les rendements. Il met en évidence l'importance du volume comme indicateur de tendance et l'influence modératrice de la duration sur cette relation.

[13]Le lecteur peut se référer à l'option "COVEST = QML" pour plus de détails. Des comparaisons avec le package G@RCH 3.0 pour Ox^{TM}donnent sensiblement les mêmes écarts-types.

[14]Cela est possible, par exemple, avec le logiciel Eviews^{TM}.

– Une dynamique de type Auto Regressive Moving Average (ARMA)(1,1) est souvent utilisée pour modéliser les rendements. Ce type de modélisation est en accord avec le fonctionnement d'un marché dirigé par les ordres comme souligné par Handa *et al.* (1998). Le logiciel SAS^{TM} n'offrant pas la possibilité d'estimer directement un ARMA-GARCH, nous procéderons en deux étapes. Dans une première étape, les rendements seront filtrés par un ARMA, ensuite les résidus de cette régression seront modélisés par un GARCH. Cette procédure en deux étapes est similaire à celle utilisée par la toolbox Matlab^{TM} UCSD-GARCH[15].
– Les contraintes imposées changent selon la modélisation. Si le modèle est un GARCH standard, nous employons les contraintes définies par Nelson & Cao (1992). À l'inverse, si l'équation de la volatilité conditionnelle contient des variables exogènes, le modèle est estimé sous contrainte de positivité de la volatilité conditionnelle[16].

5.2 Modèles univariés

Dans cette section, nous analyserons successivement et séparément chacune des composantes du processus de transaction à savoir le rendement, dans la sous-section 5.2.1, puis l'activité (la duration et le volume) dans la sous-section 5.2.2.

5.2.1 Rendement

La modélisation GARCH a été introduite par Bollerslev (1987)[17]. La première hypothèse de cette modélisation est que les rendements sont le produit d'une composante déterministe et d'une composante aléatoire donc une relation du type :

$$r_t = \sigma_t^2 \times \zeta_t \tag{5.1}$$

où r_t est le rendement de la t^{e} transaction, σ_t^2 est la composante déterministe et ζ_t la composante aléatoire. La seconde hypothèse porte sur la nature de σ_t^2. On suppose que celle-ci est l'espérance conditionnelle de la volatilité. Par définition, l'espérance conditionnelle capture toute la dépendance temporelle. Les rendements surprises ($\zeta_t = \frac{r_t}{\sigma_t^2}$) sont donc indépendants et identi-

[15]Cet ensemble de programmes permet de calculer les écarts-types avec la méthode de Bollerslev & Wooldridge (1992) mais ne permet pas d'inclure des variables exogènes.

[16]Cette contrainte signifie que toutes les valeurs estimées de la volatilité conditionnelle des différents processus doivent être positives. Une telle contrainte pourrait s'avérer forte si les variables exogènes ont des valeurs extrêmes.

[17]Le lecteur peut se référer à la section 3.1.2 pour de plus amples détails ou références sur les GARCH.

quement distribués. La dernière hypothèse porte sur la spécification de σ_t^2 :

$$\sigma_t^2 = \omega + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 + \sum_{j=1}^{q} \alpha_j \times r_{t-j} \tag{5.2}$$

Suivant la distribution supposée de ζ, on obtient différentes modélisations. Nous retenons la version la plus simple où ζ est normalement distribué. Le modèle GARCH estimé est donc le suivant.

$$\begin{aligned} r_t &= \sigma_t^2 \times \zeta_t \\ \sigma_t^2 &= \omega + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 + \sum_{j=1}^{q} \alpha_j \times r_{t-j} \end{aligned} \tag{5.3}$$

Avant de procéder à l'estimation du modèle 5.3, les rendements sont filtrés par un ARMA(1,1). La distribution supposée de ζ est normale. Afin de déterminer un modèle "optimal", nous utilisons le test du ratio de vraisemblance. Pour rappel, si on considère deux modèles GARCH d'ordres respectifs i et $i+1$ et de vraisemblances respectives L_i et L_{i+1}, le test du ratio de vraisemblance ($-2 \times \ln(L_i/L_{i+1})$) suit une loi du χ^2 à 2 degrés de liberté (la valeur critique au seuil de 5% est 5,99). Le modèle "optimal" est déterminé en partant du modèle le plus simple (GARCH(1,1)) puis en s'arrêtant dès que l'estimation du modèle suivant (GARCH($i+1$,$i+1$)) ne produit pas un test significatif au seuil de 5%.

La principale image qui se dégage des résultats de ces estimations est une grande disparité. L'ordre des modèles obtenus est très variable puisqu'il va de 1 à 3. Même pour les titres aboutissant à un modèle identique, les différences de valeurs entre les paramètres sont importantes même si les signes ou les degrés de significativité sont similaires. Parmi cette diversité, seuls St-Gobain et Lagardère (les deux titres pour lesquels on obtient un modèle d'ordre 1) semblent relativement proches. Dès lors, on peut se poser la question de la stabilité des résultats.

Toutes les séries s'accordent sur une très forte persistance de la volatilité conditionnelle (les paramètres $\beta_.$) alors que l'effet des chocs sur la volatilité (les paramètres $\alpha_.$) est plus faible. On note ensuite que, lorsque l'ordre du modèle est supérieur à 1, les valeurs des paramètres suggèrent une certaine oscillation des variables (succession de paramètres de signes contraires)[18]. Il est enfin manifeste que plusieurs estimations conduisent à des valeurs pour lesquelles les moments d'ordres supérieurs ou égal à 2 n'existent pas ($\alpha_1 + \beta_1 > 1$ dans le cas d'un GARCH(1,1)).

Les statistiques de Ljung-Box prouvent que cette modélisation parvient mieux à rendre la corrélation du processus de volatilité (résidus des rendements au carré) que celle des rendements. Les modèles GARCH parviennent donc à rendre compte de la dépendance temporelle du carré des rendements (la volatilité) mais reflètent moins bien celle des rendements.

[18] Cette oscillation pourrait traduire un phénomène de convergence vers un point d'équilibre.

TABLE 5.2
Modèles GARCH

	St-Gobain	Schneider	Sanofi	Thomson	Accor	Lyonnaise	Lagardère
α_0	0,1531***	0,0023*	0,0006***	0,0065***	0,0035***	0,0049	0,0832***
	(7,16)	(1,90)	(2,68)	(2,72)	(2,73)	(1,10)	(3,99)
α_1	0,1010***	0,2049***	0,1831***	0,2672***	0,2025***	0,1345***	0,1670***
	(12,81)	(15,04)	(14,35)	(10,59)	(15,50)	(9,13)	(7,08)
α_2		-0,2095***	-0,2863***	-0,2487***	-0,2843***	0,0003	
		(-4,84)	(-10,73)	(-11,58)	(-8,56)	(0,11)	
α_3		0,0085	0,1044***		0,0861***	-0,1281***	
		(0,22)	(6,43)		(3,45)	(-9,53)	
β_1	0,8684***	1,4027***	2,1321***	1,5989***	1,8015***	0,7492***	0,8506***
	(75,40)	(7,56)	(25,73)	(26,82)	(12,86)	(14,91)	(39,60)
β_2		-0,1215	-1,3904***	-0,6156***	-0,8214***	0,9851***	
		(- 0,40)	(-9,44)	(-11,24)	(-3,52)	(141,16)	
β_3		-0,2852**	0,2570***		0,0151	-0,7418***	
		(-2,29)	(3,96)		(0,16)	(-15,59)	
Log Likelihood	-135 251	-127 402	-82 709	-54 175	-115 574	-140 496	-94 848
Ljung Box (20)	395***	186***	328***	187***	383***	214***	207***
Ljung Box^2 (20)	132***	17	50***	14	51***	46***	45***

Cette table donne les paramètres du modèle GARCH (5.3) estimés par la méthode du pseudo maximum de vraisemblance. L'ordre du modèle a été choisi en utilisant le test du ratio de vraisemblance au seuil de 5%. Les valeurs entre parenthèses en dessous des paramètres sont les T-values robustes. La statistique de Ljung-Box (Ljung-Box^2) est calculée sur les rendements surprises (le carré des rendements surprises) avec 20 décalages. Les paramètres ou les statistiques significatifs au seuil de 1% sont signalées par ***, au seuil de 5% par ** et au seuil de 10% par *.

5.2.2 Activité

Le point a. analyse les modélisations de la duration qui sont ensuite appliquées, lors du point b., aux volumes.

a. Duration

Comme la volatilité, les durations présentent une forte autocorrélation. Cette similarité a poussé Engle & Russell (1998) à proposer, sous l'acronyme ACD, une modélisation de type GARCH pour les durations[19].

Cette modélisation est très semblable à la modélisation GARCH et utilise les mêmes hypothèses. Un modèle ACD(p,q) est résumé ci-dessous : La première hypothèse de cette modélisation est que les durations sont le produit d'une composante déterministe et d'une composante aléatoire, soit une relation du type :

$$\begin{aligned} d_t &= \psi_t \times \varepsilon_t \\ \psi_t &= \omega + \sum_{j=1}^{q} \alpha_j \times d_{t-j} + \sum_{j=1}^{p} \beta_j \psi_{t-j} \end{aligned} \tag{5.4}$$

où d_t est la duration de la t^{e} transaction, ψ_t est la composante déterministe et ε_t la composante aléatoire. Par définition, l'espérance conditionnelle capture toute la dépendance temporelle donc, les durations surprises ($\varepsilon_t = \frac{d_t}{\psi_t}$) sont donc indépendantes et identiquement distribuées.

Suivant la distribution supposée de ε, on peut alors obtenir différentes modélisations. La version la plus simple est le EACD où la distribution où ε est exponentielle. On trouve aussi couramment les WACD où la distribution de ε est Weibull[20](ε^{α} est exponentielle). Notre but n'étant pas ici exclusivement la duration mais plutôt la relation qu'elle entretient avec les autres caractéristiques des transactions, nous utiliserons la version la plus simple : le EACD. La proximité avec les GARCH est importante. Comme souligné par (Engle, 2000, Cf. p. 11), on peut, sous hypothèse du pseudo maximum de vraisemblance, estimer ce modèle avec une procédure

[19] Les modèles ACD ont connu le même succès que les modèles GARCH. Par exemple, Jasiak (1999) introduit le FI-ACD. De même, il a également stimulé l'apparition de modèles stochastiques (Bauwens & Veredas (1999) introduisent le stochastic conditionnal duration) à l'instar des "stochastic volatility models". Le lecteur trouvera dans Gouriéroux *et al.* (1999*a*) une présentation et une revue des modèles de duration. Le lecteur intéressé par une analyse critique des modèles ACD pourra également se référer à Rydberg & Shepard (1998, cf. section 5.2).

Nous appliquons ici les modèles ACD aux durations entre les transactions. Ce type de modèle peut être appliqué à des durées séparant d'autres événements. Par exemple, Hautsch (2001) les applique au temps nécessaire à traiter un certain volume ou encore Giot (2000) qui l'applique d'une part au volume et d'autre part au temps nécessaire pour observer un changement de prix supérieur à un certain seuil.

[20] On trouve également les log-ACD développés par Bauwens & Giot (2000, 2001) où c'est le logarithme des durations qui est alors modélisé. Cette spécification permet de relâcher les contraintes portant sur les paramètres du modèle.

pour les GARCH si on prend comme variable endogène la racine carrée de la duration. Le choix du modèle est fait comme précédemment. La table 5.3 donne les estimations des modèles obtenus avec cette méthode.

Comparés aux rendements, il se dégage de ces résultats une image claire. Même si les ordres choisis sont identiques (de 1 à 3 avec domination du modèle le plus simple), chacun des groupes ainsi constitués[21] présente une relative homogénéité.

De manière encore plus prononcée que le rendement, les modèles s'accordent sur une très forte persistance ($\beta_.$) et sur une faible influence des chocs ($\alpha_.$). On retrouve également les phénomènes de changements de signes que nous avions noté. Par exemple, lorsque l'ordre du modèle préféré est supérieur à 1, on obtient systématiquement des valeurs négatives pour α_2 et β_2.

La statistique de Ljung-Box sur les 20 premiers décalages montre que les modèles d'ordres supérieurs à 1 parviennent mieux que les modèles ACD(1,1) à prendre en compte la dépendance temporelle de la duration. Cela suggère que, même si le test du ratio de vraisemblance conduit à un modèle ACD(1,1) la dynamique de la duration est tout de même plus complexe. Globalement, les valeurs observées, même si elles sont significatives, sont faibles pour un échantillon de cette taille et doivent donc être interprétées comme une preuve du succès du modèle.

Le phénomène de grappe de transactions est empiriquement validé. Ce phénomène peut également être rapproché de l'effet diagonal des ordres mis en avant par Biais *et al.* (1995). Ces auteurs, en classant les ordres selon leur agressivité (à exécution immédiatement avec consommation de plusieurs limites, à exécution immédiate sans consommation de plusieurs limites...), mettent en évidence que la probabilité de transition d'un type d'ordre au suivant est toujours plus importante entre deux ordres de même type. Ce phénomène devrait induire une autocorrélation de la duration puisque Bisière & Kamionka (2000) mettent en évidence l'autocorrélation des durations entre l'introduction des ordres.

Deux outils permettent de mieux comprendre la duration. La densité donne la probabilité inconditionnelle d'observer une duration dans le voisinage d'un point. Elle offre donc une image précise de la distribution de la duration. La fonction de hasard donne à chaque point la probabilité d'observer une duration dans ce voisinage conditionnellement à la durée déjà observée (taux instantané de départ par unité de temps). Le lecteur peut se référer à Lancaster (1990, p. 7) pour une discussion de la différence avec la probabilité inconditionnelle. Une interprétation pratique de la fonction de hasard est qu'elle donne à chaque point dans le temps la probabilité d'observer une transaction dans un avenir immédiat sachant qu'il n'y en pas eu jusqu'ici (depuis la précédente transaction). Étant donné que les durations ne sont ni censurées ni contraintes de nombreuses méthodes peuvent être utilisées. Pour estimer la densité, nous employons une

[21] Les groupes sont du plus au moins nombreux : - Sanofi, Thomson, Lyonnaise, Lagardère ; - St-Gobain, Schneider ; - Accor. La constitution de ces groupes ne semble pas liée à des caractéristiques des transactions que nous avons étudiées jusqu'ici.

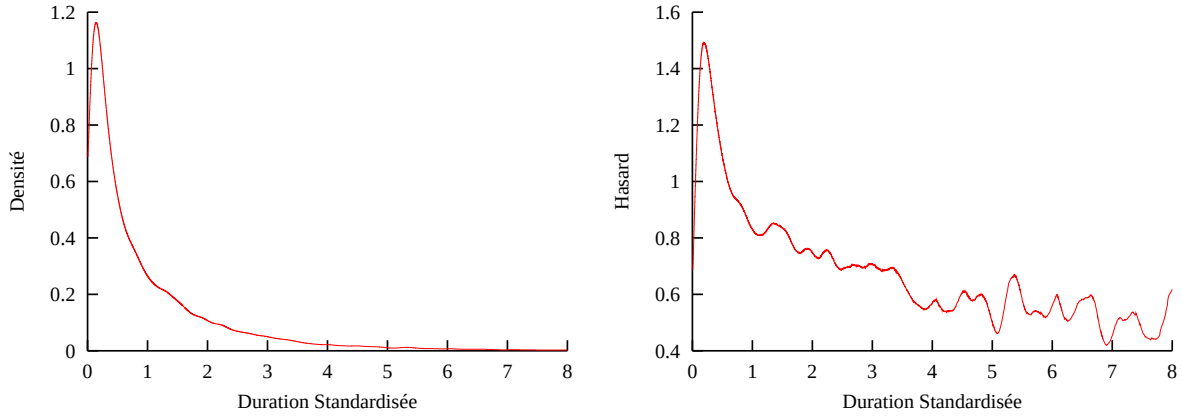
TABLE 5.3
Modèles ACD

	St-Gobain	Schneider	Sanofi	Thomson	Accor	Lyonnaise	Lagardère
α_0	0,0007***	0,0006***	0,0291***	0,0078***	0,0001***	0,0124***	0,0080***
	(4,56)	(3,59)	(3,59)	(4,41)	(7,06)	(5,59)	(5,89)
α_1	0,1037***	0,1152***	0,0729***	0,0670***	0,1150***	0,0404***	0,0527***
	(20,95)	(20,41)	(7,23)	(11,17)	(22,96)	(11,16)	(13,21)
α_2	-0,0943***	-0,1068***			-0,2061***		
	(-21,52)	(-22,34)			(-20,51)		
α_3					0,0916***		
					(17,80)		
β_1	1,5947***	1,6398***	0,8990***	0,9269***	2,4153***	0,9474***	0,9399***
	(44,45)	(44,06)	(50,66)	(129,07)	(525,19)	(168,06)	(186,39)
β_2	-0,6047***	-0,6487***			-1,8761***		
	(-17,50)	(-18,21)			(-260,85)		
β_3					0,4603***		
					(161,22)		
Log Likelihood	-93 209	-87 577	-60 087	-38 647	-80 754	-97 636	-67 011
Ljung Box (20)	25	32**	152***	180***	24	235***	215***
Ljung Box^2 (20)	59***	46***	32**	26	47***	66***	41***

Cette table donne les paramètres du modèle EACD (5.4) estimés par la méthode du pseudo maximum de vraisemblance. L'ordre du modèle a été choisi en utilisant le test du ratio de vraisemblance au seuil de 5%. Les valeurs entre parenthèses en dessous des paramètres sont les T-values robustes. La statistique de Ljung-Box (Ljung-Box^2) est calculée sur les durations surprises (le carré des durations surprises) avec 20 décalages. Les paramètres ou les statistiques significatifs au seuil de 1% sont signalés par ***, au seuil de 5% par ** et au seuil de 10% par *.

estimation par noyau (le noyau utilisé est gaussien avec la fenêtre de lissage proposée par Silverman (1986)). La fonction de survie est calculée en utilisant l'algorithme de Kaplan Meier (aussi connu sous le nom de produit limite). Nous calculons la fonction de hasard en divisant la densité par la fonction de survie. La figure 5.1 donne la densité et la fonction de hasard estimées pour Sanofi.

FIGURE 5.1
Duration surprise
Sanofi



Cette figure donne pour Sanofi la densité et la fonction de hasard estimées sur les durations surprises calculées d'après le modèle présenté dans la table 5.3. Les durations représentées dans cette figure sont, pour plus de lisibilité, tronquées à 8.

Il est manifeste que la fonction de hasard n'est pas monotone. Sur l'étendue représentée, le hasard est dans un premier temps croissant (pour les très petites valeurs) puis décroissant[22]. Ce fait montre que la succession très lente de transactions est un phénomène d'une certaine rareté dans le marché. On peut en outre noter que la pente est plus prononcée pour les petites valeurs que pour les grandes. Enfin, les oscillations de la courbe pour les valeurs $\geqslant 4$ soulignent des difficultés d'estimation en raison de la faible occurrence de ces évènements.

D'un point de vue plus économétrique, la forme de la fonction de hasard exclut que la densité soit de type Weibull (ou exponentiel) puisqu'on devrait alors observer une courbe monotone décroissante[23].

[22]Pour les valeurs extrêmement importantes, non représentées sur le graphique, la fonction de hasard est croissante et ses valeurs sont importantes.

[23]Cette constatation n'affecte pas la consistance des estimations puisque ces dernières le demeurent quelles que soient les distributions. Cela souligne toutefois que, si on voulait procéder à une estimation par maximum de vraisemblance, une hypothèse de type Weibull (un modèle WACD) ne serait pas en adéquation avec les données.

b. Volume

La littérature fournit peu de guides pour choisir une modélisation adaptée aux volumes. On peut cependant remarquer que la dynamique des volumes est similaire à celle observée pour la volatilité et plus encore à celle de la duration. Cette proximité suggère d'utiliser pour le volume une modélisation du même type que celle de la duration. Ce choix facilitera l'intégration ultérieure des 3 caractéristiques des transactions dans une même modélisation trivariée (cf. Section 5.4). Par conséquent, nous retenons à l'instar de Manganelli (2002) une modélisation de type ACD, pour le volume. En raison de sa proximité avec le modèle ACD ce modèle est désigné sous l'acronyme ACV.

Cette modélisation étant identique à celle de la duration, nous nous dispenserons de la commenter. Nous estimons le modèle (5.5) en considérant que la distribution de η est exponentielle et choisissons le modèle optimal selon la procédure déjà décrite pour la duration. La table 5.4 donne les estimations des modèles optimaux.

$$\begin{aligned} v_t &= \phi_t \times \eta_t \\ \phi_t &= \omega + \sum_{j=1}^{q} \alpha_j \times v_{t-j} + \sum_{j=1}^{p} \beta_j \phi_{t-j} \end{aligned} \tag{5.5}$$

De manière évidente, c'est pour le volume que les résultats sont les plus limpides. L'ordre des modèles est 1 ou 2 (avec un domination du modèle le plus simple ACV(1,1)). Les 2 groupes ainsi formés sont très uniformes.

Les effets mis en évidence sont identiques à ceux déjà notés sur la duration nous nous passerons donc d'un commentaire plus approfondi des valeurs obtenues. Notons toutefois que, à la différence des durations, l'autocorrélation du carré des résidus est, la plupart du temps, non significative (aussi bien pour des ACV(1,1) que des ACV(2,2)). Ces résultats suggèrent que la dynamique des volumes est plus simple que celle des durations et nous pousse en outre à croire que la spécification ACV est adaptée à la dynamique du volume.

La dynamique des volumes est très proche de celle des durations. La densité des volumes surprises représentée dans la figure 5.2 montre que le volume et la duration ont également des densités de mêmes formes.

Cette dynamique est peu étudiée dans littérature, l'un des apports de cette étude liminaire est de faire ressortir la pertinence des modélisations de type ACD appliquées au volume. À l'instar de Manganelli (2002) qui avait introduit cette modélisation et vérifié sa pertinence sur le NYSE, nos résultats vont dans le même sens et supportent cette modélisation. Elle permet de souligner que le phénomène de grappe de transactions est également observable concernant les volumes.

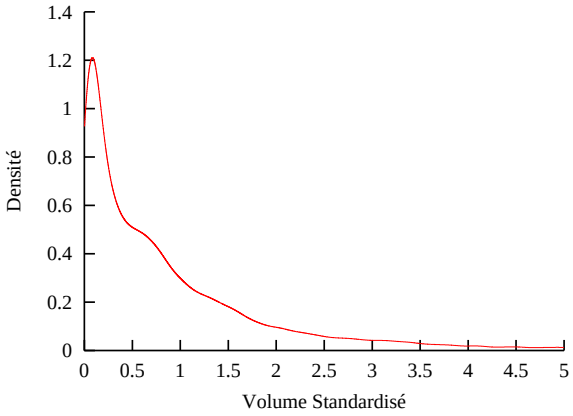
Les durations et les volumes partagent de grandes similarités tant pour leurs dynamiques que pour leurs distributions. Le cadre d'analyse univarié que nous avons utilisé ne permet pas

TABLE 5.4
Modèles ACV

	St-Gobain	Schneider	Sanofi	Thomson	Accor	Lyonnaise	Lagardère
α_0	0,0003	0,0001	0,0001	0,0238***	0,0111***	0,0222***	0,0151***
	(1,59)	(0,34)	(1,27)	(3,43)	(4,27)	(5,23)	(3,55)
α_1	0,0662***	0,0526**	0,0595***	0,0466***	0,0325***	0,0307***	0,0353***
	(11,84)	(2,43)	(10,31)	(6,70)	(8,19)	(9,69)	(6,06)
α_2	-0,0646***	-0,0522**	-0,0587***				
	(-11,96)	(-2,35)	(-10,02)				
β_1	1,7695***	1,8815***	1,8524***	0,9303***	0,9567***	0,9474***	0,9498***
	(53,79)	(40,38)	(135,65)	(71,66)	(153,57)	(134,14)	(97,82)
β_2	-0,7713***	-0,8819***	-0,8532***				
	(-23,89)	(-18,65)	(-62,01)				
Log Likelihood	-93 913	-88 211	-60 403	-39 160	-80 617	-97 602	-66 983
Ljung Box (20)	29*	38***	34**	46***	84***	60***	50***
Ljung Box^2 (20)	1	36**	11	10	125***	5	3

Cette table donne les paramètres du modèle EACV (5.5) estimés par la méthode du pseudo maximum de vraisemblance. L'ordre du modèle a été choisi en utilisant le test du ratio de vraisemblance au seuil de 5%. Les valeurs entre parenthèses en dessous des paramètres sont les T-values robustes. La statistique de Ljung-Box (Ljung-Box^2) est calculée sur les volumes surprises (le carré des volumes surprises) avec 20 décalages. Les paramètres ou les statistiques significatifs au seuil de 1% sont signalées par ***, au seuil de 5% par ** et au seuil de 10% par *.

FIGURE 5.2
Volume surprise
Sanofi



Cette figure donne pour Sanofi la densité des volumes surprises calculée d'après les estimations présentées dans la table 5.4. Les volumes présentés dans cette figure sont, pour plus de lisibilité, tronqués à 5. La procédure d'estimation de cette densité est identique à celle détaillée au précédent point.

d'établir une relation entre ces différents points de convergence et une composante informationnelle. Toutefois, l'existence même de ces ressemblances suggère une composante commune aux différents processus. Il n'y a alors qu'un pas de plus à faire pour suggérer que cette composante commune est l'information. Les modélisations bivariées présentées dans la prochaine section éclairent l'influence d'une variable d'activité sur les rendements.

5.3 Modèles bivariés

Dans la précédente section, nous avons vu que des modèles de type GARCH appliqués aux durations, volumes et rendements parviennent à capturer la dépendance temporelle de ces séries. Nous avons pu noter la proximité des dynamiques mises en évidence pour ces différents processus (très forte persistance, existence d'une forme de clustering des processus, caractère négatif de α et β lorsque l'ordre du modèle est supérieur à 1). La convergence de ces résultats suggère qu'il existe une relation entre ces trois processus. L'analyse de la matrice de transition (présentée dans l'annexe F.3) entre les différents types de transactions[24] prouve qu'on observe, la plupart du temps, la relation suivante : $\forall x \quad P(T_t = x | T_{t-1} = x) \geqslant P(T_t = x | T_{t-1} \neq x)$.

La probabilité de transition d'un type de transaction au suivant est la plupart du temps la plus élevé entre deux types identiques. En raison de l'analogie avec l'effet mis en évidence par Biais *et al.* (1995), nous dénommons cet effet diagonal. Il montre que la proximité entre les dy-

[24]Le type d'une transaction est déterminé selon ses caractéristiques (volatilité, volume et duration).

namiques de la volatilité, du volume et de la duration dépasse le cadre du simple hasard puisque ces variables ont tendance à évoluer de concert. Il est donc logique d'essayer de comprendre les relations entre ces processus qui semblent, à l'instar de l'hypothèse de la MDH, traduire l'existence d'un processus directeur commun.

La sous-section 5.3.1 étudie l'influence de la duration sur les rendements avec le modèle UHF-GARCH d'Engle (2000). Ensuite, dans la sous-section 5.3.2, nous proposons une adaptation de cette modélisation au volume.

5.3.1 Duration - Rendement

Engle (2000) propose une modélisation conjointe de la duration et de la volatilité. Soit le processus de transaction composé de deux marques : la duration (d_t) et le rendement (r_t). Si on suppose que le processus de génération des données peut s'écrire sous la forme $(d_t, r_t) \sim f(d_t, r_t|\Omega_t;\theta)$, cette densité jointe peut être réécrite comme le produit des densités conditionnelles en choisissant la décomposition en fonction du processus d'intérêt. Par exemple, si on s'intéresse au rendement on utilise la décomposition :

$$(d_t, r_t) = g(d_t|\Omega_t;\theta_d)k(r_t|d_t\Omega_t;\theta_r) \tag{5.6}$$

où $\Omega_t = d_{t-1}, r_{t-1}, d_{t-2}, r_{t-2}\ldots$. Donc, la duration actuelle appartient à l'ensemble informationnel utilisé pour estimer la densité des rendements. Dans l'équation (5.6), la densité conditionnelle des rendements (k) est donc conditionnelle à la duration courante et aux valeurs passées des processus.

Suivant la décomposition (5.6), la log-vraisemblance de la densité conjointe est la somme des log-vraisemblances de chacune des densités. Si on considère que les paramètres de (θ_d, θ_r) sont invariants dans le temps[25], une estimation séparée (on estime dans un premier temps θ_d puis θ_r en considérant que θ_d est fixe) des deux vraisemblances produit des résultats efficients. À l'inverse, si ce n'est pas le cas, une estimation séparée des deux termes produit des résultats inefficients par rapport à une estimation conjointe. Or, dans le modèle ci-après, les paramètres de la duration et du rendement sont liés. Une estimation conjointe serait donc préférable. Toutefois, à l'instar Engle (2000), nous procéderons à l'estimation séparée des processus étant donné que les estimations sont inefficientes mais consistantes et que la perte d'efficience doit être faible étant donnée la taille de l'échantillon[26].

Le modèle retenu pour la duration est un ACD. Étant donné qu'aucune variable n'est incluse

[25] Cela revient à considérer que la duration est faiblement exogène.

[26] Cette hypothèse est vraisemblable. Par exemple, Engle & Lunde (2003) proposent une modélisation, du même type, des interactions entre les transactions et les cotations. Pour ce modèle l'estimation séparée n'est pas efficiente au regard de l'estimation conjointe cependant, les différences de résultats entre ces estimations sont faibles.

dans le modèle de duration, on fait l'hypothèse que la volatilité ne cause pas (au sens de Granger (1969)) la duration. Pour le rendement, un modèle de type GARCH est utilisé. Le rendement est ajusté par la duration (se référer à la page 116 pour une discussion de cette procédure d'ajustement). La moyenne des rendements est modélisée par un ARMA(1,1) avec comme variable exogène la duration. La spécification la plus riche[27] (elle permet à la duration et à son espérance d'influencer la volatilité) est donnée par (5.7).

$$\sigma_t^2 = \omega + \beta_1 \sigma_{t-1}^2 + \alpha_1 \frac{r_{t-1}^2}{d_{t-1}} + \gamma_1 \frac{1}{d_t} + \gamma_2 \frac{d_t}{\psi_t} + \gamma_3 \frac{1}{\psi_t} + \gamma_4 \xi_{t-1} \tag{5.7}$$

où ξ est la volatilité de long terme obtenue via un lissage exponentiel des rendements au carré avec un paramètre de 0,995 c'est à dire : $\xi_t = 0,005 r_{t-1}^2/d_t + 0,995\xi_{t-1}$. L'influence de la duration sur la volatilité est traduite par trois termes : l'inverse de la duration $1/d_t$, la duration surprise d_t/ψ_t et l'inverse de l'espérance conditionnelle de la duration $1/\psi_t$. Ces trois termes ont des effets différents sur la prévision de la volatilité. L'espérance de la volatilité est donnée par (5.8).

$$E_{t-1}(\sigma_t^2) = \omega + \beta_1 \sigma_{t-1}^2 + \alpha_1 \frac{r_{t-1}^2}{d_{t-1}} + \gamma_1 E_{t-1}(\frac{1}{d_t}) + \gamma_2 + \gamma_3 \frac{1}{\psi_t} + \gamma_4 \xi_{t-1} \tag{5.8}$$

On remarque donc que l'influence du résidu de la duration (γ_2) n'a pas d'impact persistant. À contrario, les inverses de la duration (γ_1) et de la duration prévue (γ_3) ont un impact persistant puisque la duration est elle même persistante. Avant toute analyse des résultats présentés dans la table 5.5, il convient de se rappeler que les volatilités ont été normalisées par les durations. C'est donc la sensibilité d'une mesure par unité de temps qui est analysée ici.

La comparaison des résultats d'un GARCH(1,1) sans variable exogène (présenté en annexe dans la table F.1) avec les résultats du modèle (5.7) donnés dans la table 5.5 montre que la vraisemblance de ce dernier modèle est toujours inférieure. Un test du ratio de vraisemblance rejetterait donc systématiquement la modélisation (5.7). Cependant, la relation entre ces deux modèles est moins simple. Ainsi, la statistique de Ljung-box sur les résidus est systématiquement (à l'exception de Lagardère) inférieure dans la modélisation (5.7) alors que cette même statistique sur les résidus au carré est plus importante. Il semble donc que le modèle avec variables exogènes parvient mieux à expliquer la dépendance temporelle des rendements et moins bien celle des volatilités. Enfin, bien que toujours significative, la statistique de Jarque-Béra est toujours très inférieure dans le modèle avec variables exogènes.

L'analyse des paramètres obtenus fait ressortir la convergence globale des résultats. Lorsque

[27] Une spécification plus simple supposant que la duration n'est pas informative est également proposée. Le modèle estimé est alors un GARCH sur les rendements normalisés par la duration. Le lecteur trouvera en annexe F.2 les résultats et commentaires de cette estimation.

TABLE 5.5
Influence de la duration sur la volatilité

	St-Gobain	Schneider	Sanofi	Thomson	Accor	Lyonnaise	Lagardère
α_0	0,0820**	0,0000	0,0748	0,0000	0,0000	0,0000	0,0000
	(2,22)	(0,00)	(1,47)	(0,00)	(0,00)	(0,00)	(0,00)
α_1	0,1281***	0,2224***	0,1251***	0,3632***	0,1199***	0,1176***	0,4694***
	(8,75)	(8,48)	(5,24)	(6,29)	(6,61)	(4,95)	(7,81)
β_1	0,0350***	0,0311***	0,0001	0,0741***	0,0159***	0,0135**	0,0210*
	(3,87)	(3,13)	(0,03)	(3,09)	(3,41)	(2,54)	(1,66)
$1/d_t$	1,1642***	1,0225***	1,1114***	0,7721***	1,1617***	1,2847***	0,8362***
	(29,44)	(23,09)	(12,03)	(10,12)	(25,63)	(26,54)	(14,39)
d_t/ψ_t	-0,0123***	0,0668***	0,1508***	0,1853	0,0585***	0,0649***	0,0086
	(-4,14)	(3,19)	(27,19)	(1,26)	(5,31)	(4,24)	(1,23)
$1/\psi_t$	0,0066	-0,4123***	-0,6557***	-0,4235**	-0,4511***	-0,5895***	-0,1805***
	(0,11)	(-9,00)	(-36,47)	(-2,09)	(-13,83)	(-10,21)	(-9,89)
ξ_{t-1}	-0,0018	0,0873***	0,0395***	0,0042	0,0651***	0,1121***	0,0453***
	(-0,86)	(8,21)	(11,83)	(1,18)	(9,91)	(13,12)	(5,50)
Log Likelihood	-160 375	-153 044	-103 623	-72 084	-140 630	-164 202	-121 646
Jarque-Béra	617 147	1 143 821	453 249	677 397	655 050	1 374 019	2 206 530
Ljung Box (20)	65***	87***	177***	30*	72***	34**	48***
Ljung Box^2 (20)	60***	32**	8	1	678***	166***	22

Cette table présente les estimations des paramètres du modèle (5.7). ψ_t est calculé en utilisant les modèles présentés dans la table 5.3. L'ordre des modèles ACD utilisés est donc adapté aux données. Un paramètre ou une statistique significatif au seuil de 1% est indiqué par ***, au seuil de 5% par ** et, au seuil de 10% par *.

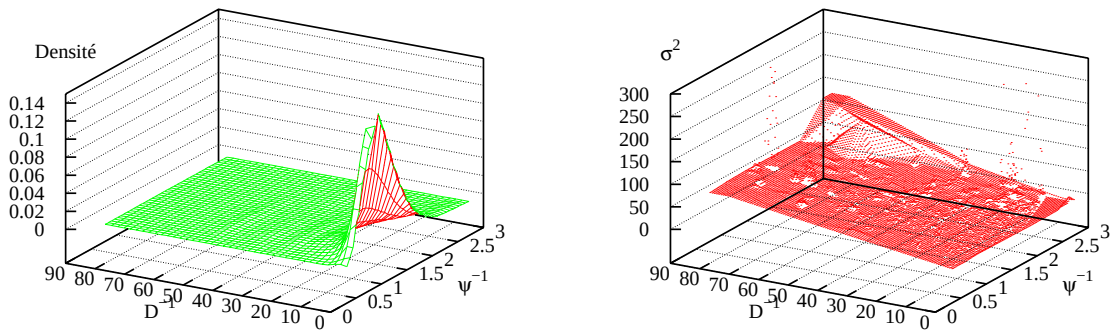
des paramètres incluant l'influence de la duration sont inclus alors la volatilité conditionnelle devient moins persistante (faiblesse de β_1) et plus sensible au chocs (α_1). Les variables explicatives ont une influence globalement similaire entre les titres même s'il leur arrive de ne pas être significatives (au maximum pour 1 ou 2 titres).

De manière synthétique on observe donc que : - la volatilité de long terme, - l'inverse de la duration et la duration surprise ont une influence positive ; - l'effet de la duration espérée est négatif. La valeur des paramètre montre sans ambiguïté que les effets sont du plus au moins importants $1/d_t$, $1/\psi_t$, d_t/ψ_t et enfin ξ_t.

Il apparaît donc qu'une faible duration induit une augmentation de la volatilité par unité de temps et supporte donc le modèle de Easley & O'Hara (1992). L'inverse de la duration espérée a une influence négative sur la volatilité par seconde. Ce paramètre variable a d'autant plus de poids que la duration espérée est faible. Il met donc en évidence que l'influence haussière de l'inverse de la duration est moindre si cette dernière était attendue.

Afin d'essayer d'apporter un peu de clarté, la figure 5.3 illustre la relation entre la volatilité conditionnelle et les deux composantes les plus importantes de la duration : $1/d_t$ et $1/\psi_t$.

FIGURE 5.3
Relations volatilité conditionnelle - duration
Sanofi



Cette figure donne, pour Sanofi, la densité jointe de $1/d_t$ et $1/\psi_t$ obtenue avec une estimation par noyau, ainsi que les valeurs estimées de la volatilité conditionnelle (σ_t^2) en fonction de ces deux dernières variables.

La distribution est concentrée autour des petites valeurs de $1/d_t$ et des valeurs de $1/\psi_t$ voisines de 1. Sur le reste de l'étendue, la densité est faible. Le graphique de la volatilité conditionnelle estimée (σ^2) en fonction de $1/d_t$ et $1/\psi_t$ nous apporte une meilleure compréhension de l'influence de cette dernière variable. La relation positive entre $1/d_t$ et σ_t^2 est évidente. Il est cependant bien difficile d'observer une relation négative entre $1/\psi_t$ et σ^2. De manière globale

l'augmentation de $1/\psi_t$ induit également une augmentation de σ_t^2, au moins pour les valeurs importantes. La relation que nous avons noté provient donc vraisemblablement d'un comportement moyen peu visible à l'œil nu (le rapport de paramètres va de 4 à 20 et est de 7 dans le cas de Sanofi) ou encore de l'influence de quelques valeurs extrêmes bien que cela soit moins probable[28].

L'étude de l'influence de la duration sur la volatilité souligne le très fort impact de cette dernière. On observe que sa prise en compte induit quasiment la disparition des dynamiques GARCH de la volatilité au profit de dynamique ARCH. Cela diminue donc la persistance des modèles au profit d'une influence plus importante laissée aux chocs.

Cette analyse met donc à jour une dynamique complexe où la duration et la volatilité interagissent et où les prévisions formées exclusivement sur la duration modifient la manière dont cette dernière agit sur la volatilité. D'un point de vue général les résultats confortent l'hypothèse avancée par Easley & O'Hara (1992) où l'augmentation de l'intensité des transactions traduit la présence d'agents informés et augmente donc la volatilité. Cette conclusion est renforcée par l'examen de la figure 5.3 qui souligne que, même si la relation globale entre l'inverse de la duration attendue et la volatilité est négative, celle-ci a, au moins pour les grandes valeurs (les petites valeurs de la duration espérée) une influence positive et forte.

On se doit toutefois de remarquer que les résultats dans cette sous-section permettent exclusivement de conclure qu'il y a une relation positive entre la volatilité par seconde et l'inverse de la duration. Cette relation pourrait avoir été induite par la méthode de normalisation utilisée[29]. La seule certitude que l'on peut donc retirer est que la relation empirique entre la volatilité et la duration n'est pas linéaire.

5.3.2 Volume - Rendement

Dans son étude de la relation entre duration et rendement, Engle (2000) remarque que "Since the time is the reciproqual of the transaction rate which is highly correlated with the volume, this study can draw on both the vast litterature on theoretical and empirical litterature on the relation between volatility and volume". Cet argument peut être renversé afin d'appliquer ce modèle à la relation entre le volume et la volatilité plutôt qu'à la relation entre la duration et la volatilité.

Nous reprenons donc le modèle présenté dans la sous-section précédente en l'appliquant au

[28] On note en plus l'existence de quelques points exceptionnels ayant soit une forte duration et une faible espérance conditionnelle, soit une faible duration et une forte espérance conditionnelle et qui ont très certainement un rapport avec la significativité de la duration surprise.

[29] Supposons par exemple que la volatilité n'est pas strictement proportionnelle au temps mais, constante en dessous d'un certain seuil. Dans un tel cas la procédure de normalisation et l'analyse subséquente donnerait probablement ce genre de résultats puisque la volatilité serait importante du fait de la normalisation et que l'inverse de la duration le serait également.

volume et au rendement. C'est à dire que nous estimons le modèle

$$\sigma_t^2 = \omega + \beta_1 \sigma_{t-1}^2 + \alpha_1 \frac{r_{t-1}^2}{v_{t-1}} + \gamma_1 \frac{1}{v_t} + \gamma_2 \frac{v_t}{\phi_t} + \gamma_3 \frac{1}{\phi_t} + \gamma_4 \xi_{t-1} \tag{5.9}$$

Étant donné que nous analysons la relation entre le rendement et le volume, la mesure de rendement (ainsi que la mesure de ξ) utilisée est ajustée par le volume. Il s'agit donc d'une mesure de rendement par unité de transaction. De la même manière, la moyenne des rendements est filtrée par un ARMA(1,1) avec comme variable exogène le volume. Enfin, le volume est modélisé par un ACV et ϕ_t est calculé en utilisant les modèles présentés dans la table 5.4. La table 5.6 donne les résultats des estimations du modèle (5.9).

La comparaison avec le modèle sans variables exogènes (présenté dans la section F.2.2en annexe) souligne que, comme pour le modèle sur la duration, le test du ratio de vraisemblance rejette systématiquement l'ajout de variables exogènes. De plus, si dans le modèle simple β_1 était important, les résultats dans la table 5.6 montrent que l'inclusion de variables concernant le volume conduit à la non-significativité de β_1 (à l'exception de Sanofi et Accor) et une baisse de α_1. On se rapproche donc fortement d'une modélisation de type ARCH.

Nous retrouvons ici le même type de similarités que nous avions déjà observées entre les résultats des ACD et des ACV et cela aussi bien concernant les paramètres GARCH qu'au sujet des variables traduisant l'influence du volume. D'un point de vue global les résultats sont légèrement plus divers. Les cas de non significativité sont plus nombreux mais les mêmes relations sont observées. Seule l'influence de la volatilité de long terme montre une différence importante. Étant donné la faible significativité de cet effet et l'instabilité de son signe, nous n'analyserons pas ce paramètre.

On peut donc conclure que le volume semble avoir sur la volatilité par unité une influence en grande partie similaire à celle de la duration sur la volatilité par unité de temps. Il est cependant intéressant de noter que même si le classement ordinal des effets selon leur importance est identique à celui obtenu pour la duration ($1/v_t$, $1/\phi_t$, v_t/ϕ_t et ξ_t), l'effet modérateur de l'inverse du volume attendu est beaucoup plus important[30] (le rapport entre les deux paramètres est de 3 à 7). Il semble donc que l'augmentation de la volatilité par unité est plus importante lorsque le volume est faible et qu'il n'était pas attendu.

Sans contredire stricto sensu l'hypothèse proposée par Easley & O'Hara (1987) selon laquelle les informés préféraient traiter des quantités importantes, nos résultats montrent cependant que l'augmentation de la volatilité par unité de volume est plus importante pour les petites valeurs. En effet, si on suivait cette hypothèse, on aurait attendu une relation négative entre l'inverse du volume et la sensibilité de la volatilité (une sensibilité plus forte pour les petites va-

[30] On peut se demander si cette plus grande correction apportée à l'influence du volume ne traduit pas la perception par le marché de volumes non-informés n'augmentant pas la volatilité.

TABLE 5.6
Influence du volume sur la volatilité

	St-Gobain	Schneider	Sanofi	Thomson	Accor	Lyonnaise	Lagardère
α_0	0,0832	0,1558***	0,0382**	0,0095	0,0683***	0,1648***	0,0000
	(0,31)	(3,01)	(2,16)	(0,23)	(3,67)	(5,64)	(0,00)
α_1	0,0132*	0,0501**	0,0469***	0,1779***	0,0347***	0,0112***	0,1096***
	(1,74)	(2,25)	(4,89)	(5,29)	(5,65)	(8,03)	(4,96)
β_1	-0,0000	0,0000	0,0015***	0,0000	0,0002***	0,0005***	-0,0000
	(-0,00)	(0,03)	(2,62)	(0,00)	(4,19)	(4,95)	(-0,00)
$1/\nu_t$	0,5225***	0,6177***	0,7013***	0,6769***	0,5835***	0,4829***	0,6099***
	(6,49)	(20,04)	(30,63)	(19,95)	(43,44)	(44,26)	(27,82)
ν_t/ϕ_t	0,0064	0,0210	0,0107***	0,0157***	0,0078***	0,0036***	0,0210***
	(0,15)	(0,58)	(2,75)	(4,55)	(12,87)	(2,70)	(7,54)
$1/\phi_t$	-0,1184	-0,2383***	-0,1111***	-0,1639***	-0,1094***	-0,1642***	-0,1871***
	(- 0,33)	(-3,00)	(-4,67)	(-9,11)	(-15,82)	(-11,50)	(-10,94)
ξ_{t-1}	0,0036	0,0042***	-0,0019***	0,0056*	-0,0011	0,0016*	0,0202***
	(0,30)	(2,87)	(-4,17)	(1,94)	(-1,29)	(1,68)	(7,78)
Log Likelihood	-168 076	-162 010	-106 081	-83 886	-141 713	-173 639	-127 732
Jarque-Béra	543 797	1 609 775	516 725	702 035	607 177	911 201	1 829 147
Ljung Box (20)	103***	297***	334***	106***	410***	218***	13
Ljung Box^2 (20)	0	0	52***	15	9	0	0

Cette table présente les estimations des paramètres du modèle (5.9). Un paramètre ou une statistique significatif au seuil de 1% est indiqué par ***, au seuil de 5% par ** et, au seuil de 10 % par *.

leurs de l'inverse du volume). Des stratégies comme le camouflage (division d'une transaction en plusieurs) par les informés de leurs transactions afin de mieux rentabiliser leur information pourrait expliquer ce phénomène[31]. Tout comme pour la duration, les résultats soulignent donc la non linéarité de la relation entre volatilité par unité et volume. Nous rejoignons donc dans une certaine mesure les différentes études qui mettent en évidence une relation quadratique entre le volume et la découverte de prix (cf. section 4.1).

Les conclusions de cette modélisation bivariée appliquée aux relations duration - volatilité et volume - volatilité[32] sont étrangement proches. À chaque fois, les modèles employés suggèrent que ces variables permettent de capturer une part de la persistance des séries de volatilités en augmentant le poids des chocs. La dynamique entre la volatilité et l'activité est complexe puisqu'elle a des effets directs mais aussi des effets de retour. Cependant, les signes des différentes variables traduisant les effets de l'activité sont globalement similaires pour les deux applications. De là à penser que ces deux variables traduisent, chacune à sa manière, un même type d'information il n'y a qu'un pas que la présente méthodologie ne permet pas encore de franchir. Les cadres d'analyse employés jusqu'ici ne permettent pas ce genre de conclusion étant donné qu'ils négligent les interactions entre les variables d'activité. La prochaine section s'attachera à lever cette limite en utilisant un modèle trivarié dynamique.

5.4 Modèle trivarié

Manganelli (2002) propose une extension du type de modélisation introduit par Engle (2000) aux trois marques du processus de transaction soit, la duration (d), le volume (v) et le rendement (r). En supposant que le processus de génération de ces données peut s'écrire sous la forme $(d_t, v_t, r_t) \sim f(d_t, v_t, r_t|\Omega_t;\theta)$ et sous hypothèse que la causalité va de la duration vers le volume et, de la duration et du volume vers le rendement, il propose alors la décomposition suivante.

$$(d_t, v_t, r_t) = g(d_t|\Omega_t;\theta_d)h(v_{,t}|d_t, \Omega_t;\theta_v)k(r_t|v_t, d_t, \Omega_t;\theta_r) \tag{5.10}$$

En reprenant les notations précédentes (ψ_t l'espérance conditionnelle de la duration en t, ϕ_t celle du volume et σ_t^2 celle du rendement) il estime finalement le modèle (5.11). Dans ce modèle, la moyenne du rendement est modélisée par un ARMA(1,1) avec comme variables exogènes d_t, d_{t-1}, v_t et v_{t-1}. Notre intérêt majeur résidant dans la volatilité, nous ne rapportons pas ces estimations. On peut remarquer que, dans cette modélisation, ψ est représenté par un ACD(1,1),

[31] On pourrait tout aussi bien émettre l'hypothèse que les investisseurs non informés choisissent de traiter des quantités importantes uniquement quand la profondeur en carnet est suffisante pour leur assurer de bonnes conditions d'exécution. Suivant cette hypothèse, la volatilité associée à un volume important sera proportionnellement plus faible et la normalisation donnerait le jour à ce type de relation.

[32] La volatilité est à chaque fois normalisée par la mesure adéquate.

ϕ par un ACV(1,1) et σ^2 par un GARCH(1,1).

$$\begin{aligned}
\psi_t &= \omega_d + \beta_d \psi_{t-1} + \alpha_d d_{t-1} + \gamma_{d,1} v_{t-1} + \gamma_{d,2} \mu_{t-1}^2 \\
\phi_t &= \omega_v + \beta_v \phi_{t-1} + \alpha_v v_{t-1} + \gamma_{v,1} d_t + \gamma_{v,2} d_{t-1} + \gamma_{v,3} \mu_{t-1}^2 \\
\sigma_t^2 &= \omega_r + \beta_r \sigma_{t-1}^2 + \alpha_r \mu_{t-1}^2 + \gamma_{r,1} d_t + \gamma_{r,2} d_{t-1} + \gamma_{r,3} v_t + \gamma_{r,4} v_{t-1}
\end{aligned} \tag{5.11}$$

Cette modélisation[33] est en fait un cas particulier d'un modèle plus général que nous résumerons un peu plus loin. Pour parvenir à ce cas particulier on suppose d'une part l'absence de variations des paramètres et d'autre par l'exogénéïté faible des trois fonctions[34]. Cette dernière restriction pourrait être trop contraignante. Comme suggéré par Dufour & Engle (2000), il est possible de la tester avec la régression suivante :

$$\begin{aligned}
\hat{\varepsilon}_t &= \mu_d + \alpha_{d,1} \hat{\phi}_{t-1} + \alpha_{d,2} \hat{\sigma}_{t-1}^2 \\
\hat{\eta}_t &= \mu_v + \alpha_{v,1} \hat{\psi}_{t-1} + \alpha_{v,2} \hat{\sigma}_{t-1}^2 \\
\hat{\mu}_t / \hat{\sigma}_t^2 &= \mu_r + \alpha_{r,1} \hat{\psi}_{t-1} + \alpha_{r,2} \hat{\phi}_{t-1}
\end{aligned} \tag{5.12}$$

On régresse les résidus standardisés des trois processus sur les variables espérées des deux autres processus. Si le modèle est correctement spécifié, les résidus devraient être IID et les coefficients ne devraient pas être significativement différents de 0. À l'inverse, si les valeurs espérées devaient appartenir au modèle, on devrait observer une corrélation significative entre les résidus et les variables omises. La table 5.7 donne les résultats de l'estimation de la régression (5.12).

Les résultats suggèrent quelques erreurs de spécification de la duration espérée et de la volatilité espérée. Pour la duration on observe 4 cas où $\hat{\sigma}_{t-1}^2$ a une influence significative et positive. Il faudrait donc prendre en compte la volatilité espérée de la transaction précédente pour modéliser la duration. Le problème est plus grave pour la volatilité puisque, dans 3 cas pour la duration et 4 pour le volume, les paramètres sont significatifs et, à une exception près, positifs. Il est remarquable que les problèmes de spécification soient concentrés exclusivement sur quatre titres (St-Gobain, Schneider, Accor et Lyonnaise), les autres titres (Sanofi[35], Thomson, et Lagardère) ne mettent en évidence aucune erreur de spécification. Il semble donc que la relation duration - volume - volatilité change selon le titre. Une comparaison entre l'évolution des titres au cours de la période (figure 4.8) et les erreurs de spécification ne fait pas ressortir de lien

[33] L'intérêt de ce modèle résidant principalement dans la représentation VAR du système (5.11), les estimations des différents modèles ne sont présentées qu'en annexe dans la section F.3.

[34] Ces hypothèses permettent de procéder à l'estimation séparée des trois régressions. Le rejet de celles-ci conduit d'une part à une modélisation différente puisque les variables ψ_t, ϕ_t et σ_t devraient être incluses et d'autre part à l'obligation de procéder à une estimation simultanée des trois processus, ce qui poserait des problèmes d'optimisation numérique.

[35] Ce titre ne met en évidence qu'une erreur de spécification au seuil de 5%.

TABLE 5.7
Test de variables omises

	Duration surprise : $\hat{\varepsilon}_t$			Volume surprise : $\hat{\eta}_t$			Volatitlité surprise : $\hat{\mu}_t/\hat{\sigma}_t^2$		
	μ_d	$\hat{\phi}_{t-1}$	$\hat{\sigma}^2_{t-1}$	μ_v	$\hat{\psi}_{t-1}$	$\hat{\sigma}^2_{t-1}$	μ_r	$\hat{\psi}_{t-1}$	$\hat{\phi}_{t-1}$
St-Gobain	0,9917***	-0,0058	0,0035***	0,9996***	0,0009	-0,0002	-0,0372***	0,0229***	0,0136**
	(71,31)	(-0,43)	(3,45)	(40,17)	(0,04)	(-0,08)	(-2,93)	(3,14)	(2,11)
Schneider	1,0044***	-0,0145	0,0022***	1,0072***	-0,0074	0,0001	-0,0225**	0,0071	0,0133***
	(80,67)	(-1,20)	(3,38)	(37,42)	(-0,30)	(0,04)	(-2,20)	(1,15)	(2,71)
Sanofi	1,0231***	-0,0277	0,0013	1,0377***	-0,0318	-0,0015	0,0104	-0,0197**	0,0042
	(45,02)	(-1,27)	(1,09)	(43,81)	(-1,47)	(-0,95)	(0,63)	(-1,98)	(0,41)
Thomson	1,0384***	-0,0370*	-0,0002	1,0465***	-0,0377	-0,0014*	0,0281	-0,0192	-0,0068
	(48,13)	(-1,91)	(-0,28)	(35,61)	(-1,53)	(-1,80)	(1,40)	(-1,64)	(-0,66)
Accor	0,9881***	0,0042	0,0018**	1,0298***	-0,0238	-0,0014	-0,0469***	0,0210***	0,0175***
	(71,00)	(0,33)	(2,13)	(37,64)	(-0,99)	(-0,60)	(-3,49)	(2,62)	(2,73)
Lyonnaise	1,0042***	-0,0121	0,0019**	1,0185***	-0,0140	-0,0010	-0,0319**	-0,0008	0,0220***
	(63,61)	(-0,77)	(2,08)	(31,51)	(-0,47)	(-0,61)	(-2,07)	(-0,08)	(2,84)
Lagardère	1,0075***	-0,0098	0,0004	1,0150***	-0,0186	0,0007	-0,0038	0,0025	0,0037
	(71,85)	(-0,74)	(0,96)	(33,55)	(-0,71)	(0,86)	(-0,25)	(0,27)	(0,53)

Cette table présente les résultats de l'estimation du modèle (5.12). Les T-values, données en dessous des paramètres estimés, sont d'après les écarts-types robustes de White (1980). Un paramètre significatif au seuil de 1% est indiqué par ***, au seuil de 5% par ** et au seuil de 10% par *.

évident entre ces deux points.

Le modèle (5.11) résulte d'une modélisation VAR. En effet, en posant $\mu_t' = (\psi_t, \phi_t, \sigma_t)$ et $\tau_t' = (d_t, v_t, r_t^2)$, on a le modèle suivant :

$$\mu_t = \gamma + A\mu_{t-1} + B\tau_t + C\tau_{t-1} \tag{5.13}$$

où A, B et C sont des matrices 3×3 de paramètres. La paramétrisation de la densité (5.10), c'est à dire la causalité de la duration vers le volume et de la duration et du volume vers les rendements, impose que B est diagonale inférieure avec des 0 sur la diagonale. Si on impose que A et C soient diagonales et que B soit nulle alors, on retrouve les estimations univariées présentées dans les sous-sections 5.2.2 à 5.2.1. Le modèle (5.11) résulte de l'hypothèse que la matrice A est diagonale, ce qui signifie que les paramètres du système sont constants dans le temps[36].

Pour être explicite et faire le lien entre le système (5.11) et sa représentation matricielle (5.13), les valeurs des matrices A, B et C sont données dans (5.14).

$$A = \begin{pmatrix} \beta_d & \dots & 0 \\ \vdots & \beta_v & \vdots \\ 0 & \dots & \beta_r \end{pmatrix} B = \begin{pmatrix} 0 & \dots & \\ \gamma_{v,1} & 0 & \dots \\ \gamma_{r,1} & \gamma_{r,3} & 0 \end{pmatrix} C = \begin{pmatrix} \alpha_d & \gamma_{d,1} & \gamma_{d,2} \\ \gamma_{v,2} & \alpha_v & \gamma_{v,3} \\ \gamma_{r,2} & \gamma_{r,4} & \alpha_r \end{pmatrix} \tag{5.14}$$

L'effet d'une variation d'une des composantes de τ_{t-1} sur l'espérance des variables est donnée par $AB + C$ et a donc pour valeur (5.15).

$$\begin{pmatrix} \alpha_d & \gamma_{d,1} & \gamma_{d,2} \\ \beta_v\gamma_{v,1} + \gamma_{v,2} & \alpha_v & \gamma_{v,3} \\ \beta_r\gamma_{r,1} + \gamma_{r,2} & \beta_r\gamma_{r,3} + \gamma_{r,4} & \alpha_r \end{pmatrix} \tag{5.15}$$

La table 5.8 donne l'effet des valeurs passées sur les espérances des variables.

Duration

L'effet du volume sur la duration est faible, il est rarement significatif (3 cas sur 7) et change de signe. À l'inverse, la volatilité a un fort impact sur la duration puisqu'elle est toujours significative et positive. Ces résultats contrastent fortement avec ceux de Manganelli (2002) qui trouvait que le volume et la volatilité avaient tout les deux un impact significatif et négatif sur la duration et concluait donc, à la suite de Dufour & Engle (2000), que les durations courtes (les fortes intensités des transactions) suivaient les rendements ou les volumes importants.

Sur le marché parisien, le volume n'a pas d'influence manifeste et les fortes volatilités ont

[36] Le lecteur intéressé trouvera dans Manganelli (2002) le calcul explicite des conditions de stationnarité et la fonction de réponse du système et de son écart-type.

TABLE 5.8
Effets des valeurs passées sur les valeurs attendues

St-Gobain

	d_{t-1}	v_{t-1}	μ^2_{t-1}
ψ_t	0,0427***	-0,0014***	0,0002***
	(15,65)	(-4,87)	(3,86)
ϕ_t	-0,0033***	0,0317***	-0,0002***
	(-2,73)	(7,50)	(-2,63)
σ_t	-0,0188	0,2973***	0,1337***
	(-1,64)	(6,16)	(11,08)

Schneider

	d_{t-1}	v_{t-1}	μ^2_{t-1}
ψ_t	0,0578***	-0,0004	0,0001***
	(13,87)	(-1,22)	(2,82)
ϕ_t	0,0011	0,0384***	-0,0002***
	(1,03)	(7,07)	(-3,02)
σ_t	-0,0157*	0,1702***	0,1342***
	(-1,80)	(5,11)	(9,42)

Sanofi

	d_{t-1}	v_{t-1}	μ^2_{t-1}
ψ_t	0,0737***	0,0032**	0,0005***
	(7,58)	(2,39)	(3,05)
ϕ_t	-0,0006	0,0441***	-0,0001
	(-0,43)	(6,69)	(-0,81)
σ_t	-0,0422***	0,1324***	0,1328***
	(-4,88)	(4,93)	(12,34)

Thomson

	d_{t-1}	v_{t-1}	μ^2_{t-1}
ψ_t	0,0676***	0,0016*	0,0003***
	(11,24)	(1,72)	(3,37)
ϕ_t	0,0022	0,0463***	-0,0000
	(1,32)	(6,48)	(-0,42)
σ_t	-0,0219	0,0734**	0,1613***
	(-1,47)	(2,26)	(4,09)

Accor

	d_{t-1}	v_{t-1}	μ^2_{t-1}
ψ_t	0,0539***	-0,0017***	0,0004***
	(12,21)	(-3,60)	(4,64)
ϕ_t	-0,0007	0,0327***	-0,0002***
	(-0,69)	(8,25)	(-5,12)
σ_t	-0,0127	0,1375***	0,1290***
	(-0,99)	(5,04)	(10,27)

Lyonnaise

	d_{t-1}	v_{t-1}	μ^2_{t-1}
ψ_t	0,0408***	-0,0007	0,0003***
	(10,58)	(-1,47)	(4,21)
ϕ_t	-0,0015	0,0308***	-0,0001*
	(-1,17)	(9,53)	(-1,91)
σ_t	-0,0141	0,2417***	0,1186***
	(-0,26)	(4,91)	(10,05)

Lagardère

	d_{t-1}	v_{t-1}	μ^2_{t-1}
ψ_t	0,0529***	-0,0004	0,0001***
	(13,18)	(-1,01)	(2,75)
ϕ_t	-0,0007	0,0359***	-0,0001***
	(-0,59)	(6,27)	(-3,34)
σ_t	0,0012	0,1123***	0,2078***
	(0,12)	(3,31)	(9,11)

Cette table donne les coefficients de la matrice $AB+C$ soit l'effet des valeurs passées de la duration, du volume et de la volatilité sur l'espérance de ces variables. Un paramètre significatif au seuil de 1% est indiqué par *** au seuil de 5% par ** et au seuil de 10% par *.

pour effet de diminuer l'intensité des transactions. Cette constatation suggère que la dynamique de la duration dans un marché dirigé par les ordres diffère de celle dans un marché dirigé par les prix. Dans un marché dirigé par les ordres, les investisseurs semblent attendre que l'offre en carnet se reconstitue après une forte volatilité. Après une forte volatilité, la rémunération de l'offre de liquidité et le coût d'immédiateté sont plus importants. Les agents ont alors intérêt à introduire des ordres en carnet, ce qui peut expliquer l'influence positive de la volatilité sur la duration.

De manière évidente, nous avons constaté que la duration n'était pas exogène. En effet, la volatilité de la transaction précédente influence positivement la duration. L'hypothèse utilisée dans la sous-section 5.3.1, à savoir que la duration n'était pas causée par la volatilité, est donc fausse.

Volume

Le volume est le processus qui dépend le moins des deux autres. La duration n'a d'impact significatif et négatif que pour St-Gobain. La volatilité a une influence significative et négative dans 4 cas (au seuil de 1%). Il semble donc que les volumes suivent les périodes de moindre volatilité (d'autant plus que l'on a vu que la volatilité augmentait la duration. Cela suggère que les agents choisissent les moments calmes pour faire des transactions plus importantes.

Dans la sous-section 5.3.2, nous avions émis l'hypothèse que le volume n'était pas causé par la volatilité. Globalement, cette hypothèse semble erronée puisqu'on note dans 5 cas (au seuil de 5%) que la volatilité influence négativement le volume.

Volatilité

L'effet de la duration passée sur la volatilité est très rarement significatif. Seul Sanofi met en évidence une influence significative et négative de la duration. Il ne semble donc pas y avoir de relation de l'intensité des transactions vers la volatilité. Ce résultat contraste vivement avec les constatations de Dufour & Engle (2000), Engle (2000) et Manganelli (2002) qui mettaient en évidence une influence négative et significative de la duration sur la volatilité et concluaient alors que les périodes de moindre activité (i.e. de plus longue duration) étaient associées à une moindre activité des informés. Les résultats que nous obtenons suggèrent que, dans un marché dirigé par les ordres, l'intensité de l'activité n'est pas liée à la présence des informés.
Les résultats que nous obtenons ici ne sont pas en contradiction avec ceux que nous avions observés dans la section 5.3.1. Nous avions alors mis en évidence une influence négative et significative de la duration sur la volatilité par seconde. Ces résultats montraient donc que, pour les petites durations on a une relation du type volatilité / duration $= 1 + \alpha$ ($\alpha > 0$) et pour les grandes durations une relations volatilité / duration $= 1 - \beta$ ($\beta > 0$) : la sensibilité de la volatilité à la duration diminue avec la duration.

Le signe du volume est toujours significatif et positif. Ce résultat est à première vue cohérent avec les prédictions faites par Easley & O'Hara (1987) selon lesquelles les transactions de tailles

plus importantes ont plus de chances d'être faites par des informés et ont donc une influence plus importante sur la volatilité. Toutefois, nous avons également vu que la volatilité normalisée par le volume était une fonction croissante de $1/v_t$. Donc, l'influence du volume sur la volatilité par unité de transaction est plus importante pour les petites transactions que pour les transactions importantes. Rapportées à leur taille, les grosses transactions augmentent moins la volatilité que les petites.
On observe donc une relation entre le volume et la volatilité que celle-ci soit ou non normalisée par le volume. On peut donc conclure que le volume influence positivement la volatilité mais que cette relation n'est pas linéaire. L'impact marginal du volume est plus fort lorsque celui-ci est faible que lorsqu'il est important. Si on considère la forme des courbes d'offre et de demande, il est difficile de différencier les relations mises en évidence de la mécanique de fourniture de liquidité.

La fonction de réponse du système (l'effet d'une variation d'une des composantes sur la prévision des variables) permet de calculer pour les transactions qui suivent[37] l'effet d'un des paramètres du système. Elle est donnée par $[(I-B)^{-1}(A+C)]^{i-1}[(I-B)^{-1}(AB+C)]$. Cette fonction calculée pour $i=1$ donne l'effet d'une variation des composantes de τ_{-1} sur la prévision des valeurs de τ soit les valeurs de μ. La table 5.9 donne, pour l'ensemble des titres, les valeurs obtenues pour cet ordre[38]. La fonction de réponse peut également être représentée graphiquement. La figure 5.4 présente la fonction de réponse obtenue pour Sanofi pour i variant de 1 à 100. On représente donc la succession des réponses obtenues pour les différentes transactions.

Les effets des valeurs passées sur l'espérance des variables ou la prévision de l'espérance des variables sont globalement similaires. De même, les résultats représentés par la figure 5.4 confirment ceux observés dans la table 5.9. L'effet d'un choc perdure souvent sur une vingtaine de transactions (soit environ 12 minutes si on multiplie par la moyenne).

Cette figure permet cependant de mieux saisir l'influence réelle de la duration. En effet, si on a bien un effet non significatif sur les premières transactions, celui-ci augmente vivement par la suite et demeure significatif jusqu'à environ la 25^{e} transaction. La duration exerce donc sur la volatilité une influence significative, positive et durable. Ce fait entre alors en contradiction directe avec l'hypothèse de Easley & O'Hara (1992) selon laquelle nous aurions du obtenir une influence négative.

De manière générale, toutes les fonctions de réponse soulignent que l'influence de la dura-

[37] Les prévisions sont en temps transaction. En effet, la duration n'est pas exogène et il faut procéder par simulation pour calculer une fonction de réponse en temps horloge. On est très vite confronté à un problème de dimensionnalité puisqu'il faut générer trois variables aléatoires pour chaque transaction (les erreurs de la duration, du volume et de la volatilité). Cette effet limitant l'intérêt de la simulation, nous ne présentons que la fonction de réponse théorique en temps transaction. Celles-ci confirment toutefois l'importance des relations mises en évidence.

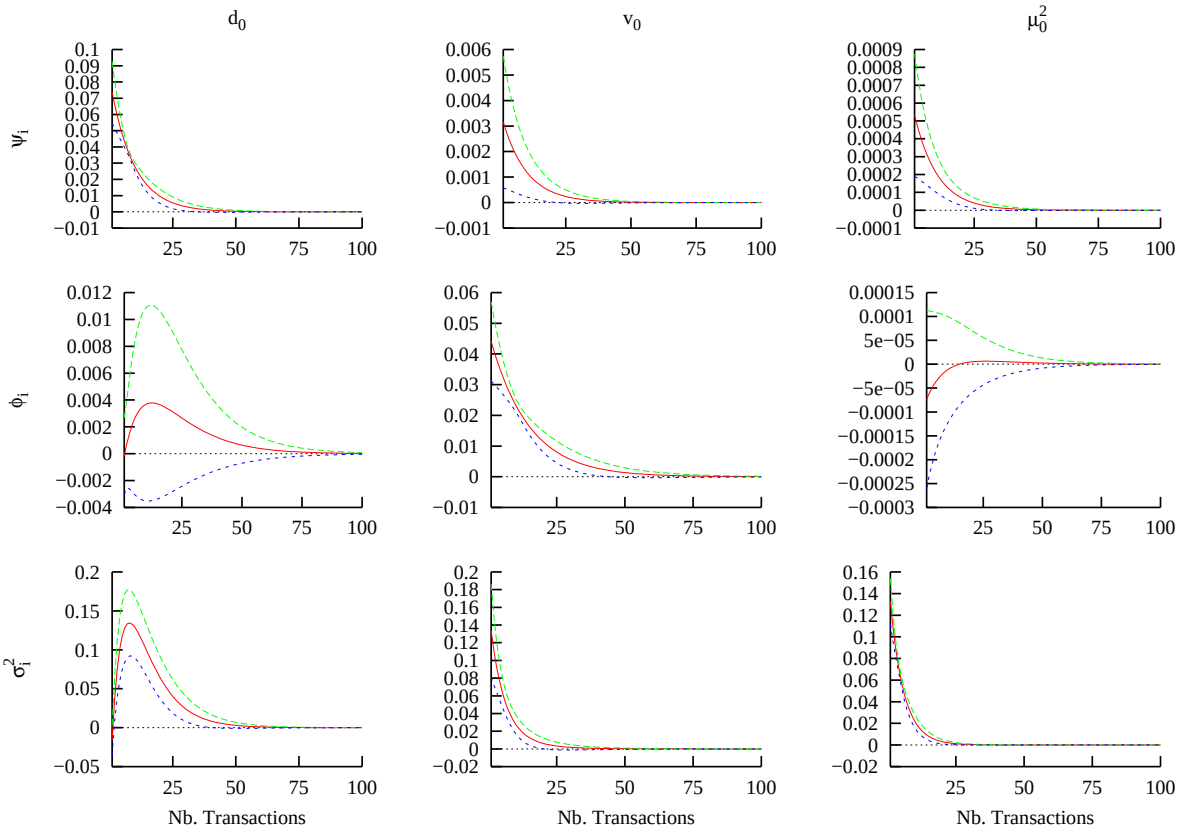
[38] Par construction la première ligne des tables 5.8 et 5.9 est identique.

TABLE 5.9
Effets des valeurs passées sur les valeurs prédites

	St-Gobain			Schneider		
	d_{t-1}	v_{t-1}	μ^2_{t-1}	d_{t-1}	v_{t-1}	μ^2_{t-1}
ψ_t	0,0427***	-0,0014***	0,0002***	0,0578***	-0,0004	0,0001***
	(15,66)	(-4,87)	(3,86)	(13,88)	(-1,22)	(2,82)
ϕ_t	-0,0040***	0,0317***	-0,0002***	0,0003	0,0384***	-0,0002***
	(-3,36)	(7,51)	(-2,68)	(0,27)	(7,07)	(-3,05)
σ_t	-0,0036	0,2989***	0,1337***	0,0166	0,1741***	0,1342***
	(-0,31)	(6,37)	(11,10)	(1,62)	(5,25)	(9,43)
	Sanofi			Thomson		
	d_{t-1}	v_{t-1}	μ^2_{t-1}	d_{t-1}	v_{t-1}	μ^2_{t-1}
ψ_t	0,0737***	0,0032**	0,0005***	0,0676***	0,0016*	0,0003***
	(7,59)	(2,39)	(3,05)	(11,25)	(1,72)	(3,37)
ϕ_t	-0,0002	0,0441***	-0,0001	0,0035**	0,0464***	-0,0000
	(-0,13)	(6,70)	(-0,77)	(2,09)	(6,49)	(-0,35)
σ_t	-0,0160	0,1338***	0,1330***	0,0148	0,0820**	0,1614***
	(-1,58)	(4,97)	(12,37)	(0,99)	(2,55)	(4,10)
	Accor			Lyonnaise		
	d_{t-1}	v_{t-1}	μ^2_{t-1}	d_{t-1}	v_{t-1}	μ^2_{t-1}
ψ_t	0,0539***	-0,0017***	0,0004***	0,0408***	-0,0007	0,0003***
	(12,22)	(-3,60)	(4,64)	(10,59)	(-1,47)	(4,21)
ϕ_t	-0,0012	0,0328***	-0,0002***	-0,0025**	0,0308***	-0,0001**
	(-1,28)	(8,26)	(-5,19)	(-2,04)	(9,55)	(-2,01)
σ_t	0,0073	0,1379***	0,1292***	0,0087	0,2437***	0,1188***
	(0,57)	(5,04)	(10,29)	(0,16)	(4,91)	(10,07)
	Lagardère					
	d_{t-1}	v_{t-1}	μ^2_{t-1}			
ψ_t	0,0529***	-0,0004	0,0001***			
	(13,19)	(-1,01)	(2,75)			
ϕ_t	-0,0021*	0,0359***	-0,0001***			
	(-1,69)	(6,28)	(-3,43)			
σ_t	0,0158	0,1185***	0,2078***			
	(1,53)	(3,53)	(9,11)			

Cette table donne les coefficients de la matrice $(I-B)^{-1}(AB+C)$, soit l'effet des valeurs passées de la duration, du volume et de la volatilité sur la prévision de ces variables. Un paramètre significatif au seuil de 1% est indiqué par ***, au seuil de 5% par ** et au seuil de 10% par *.

FIGURE 5.4
Fonction de réponse
Sanofi



Cette figure donne la fonction de réponse obtenue pour Sanofi. Chaque graphique donne l'effet d'une variable reprise en haut de chaque colonne sur la variable donnée en début de chaque ligne. Cette figure est donc l'analogue de la table 5.9 où chaque graphique donne l'évolution d'une cellule de la table. La ligne rouge est la réponse moyenne alors que les lignes hachurées délimitent un échantillon de confiance à 95%.

tion sur la volatilité n'est pas immédiate (n'affecte pas les premières transactions) mais est tout de même significative, positive et persistante. Il existe donc un effet moyen des durations sur les volatilités mais celui-ci est décalé dans le temps.

5.5 Conclusion

Dans ce chapitre, nous avons étudié les relations entre les trois composantes du processus de transaction (la duration, le volume et la volatilité).

Dans un premier temps nous avons analysé séparément les caractéristiques de ces trois processus. Ces trois processus partagent de grandes similarités : - une modélisation de type GARCH semble capturer de manière adéquate la dépendance temporelle des séries ; les séries ont une très forte persistance... Ces caractéristiques communes suggèrent que la relation entre ces trois processus n'est pas le fruit du hasard. Nous avons alors analysé leurs relations.

Dans une première approche, nous avons employé le modèle d'Engle (2000) afin d'étudier la relation entre la duration et la volatilité. Cette modélisation permet d'évaluer l'effet marginal de la duration sur la volatilité. En effet, la mesure de rendement étant normalisée par la duration, l'analyse de la relation entre le rendement et la duration revient à analyser l'existence de variations marginales à cette relation supposée linéaire. Cette analyse souligne l'influence importante de la duration sur la volatilité. La prise en compte de la duration dans la volatilité entraîne une très forte réduction de la persistance du processus et tout particulièrement du paramètre GARCH. La principale conclusion est que pendant les périodes de forte activité l'augmentation marginale de la volatilité est plus importante que pendant les périodes calmes. Ces résultats sont globalement compatibles avec l'interprétation proposée par Easley & O'Hara (1992) où une forte duration est interprétée comme une absence d'informations.

Nous avons également appliqué cette modélisation à la relation volume rendement et avons observé des résultats similaires : la volatilité par unité de transaction est positivement influencée par l'inverse du volume de la transaction. Les grosses transactions ont donc un effet marginal inférieur à celui des petites transactions. Cette observation va à l'encontre de l'hypothèse de Easley & O'Hara (1987) selon laquelle les informés préfèrent traiter de grosses quantités.

Dans un second temps, nous avons employé le modèle de Manganelli (2002) afin de modéliser les relations entre les trois processus. Cette modélisation étudie simultanément les relations entre duration, volume et volatilité dans un cadre VAR qui admet une représentation de type GARCH. Le rendement employé dans ce modèle n'est pas normalisé. Aussi, il permet d'étudier l'effet moyen des variables. Cette modélisation est en outre dynamique puisqu'elle tient compte des interactions entre les différentes caractéristiques.

Nous avons mis en évidence les relations suivantes. La volatilité influence de manière importante les autres caractéristiques des échanges : - un impact toujours négatif sur la duration ;

- une influence souvent négative sur le volume. Il semble donc que sur le marché parisien une augmentation de la volatilité conduise à des transactions moins fréquentes et plus petites. Ces résultats soulignent que ni la duration ni le volume ne sont exogènes. Il faut donc tenir compte de l'endogénéïté de ces variables dans une modélisation. Concernant la volatilité, la duration n'a pas d'influence immédiate mais un effet décalé positif et durable, le volume a toujours une influence positive. Le volume comme la duration ont pour effet moyen d'augmenter la volatilité.

L'image globale qui émerge de ces analyses est riche. Le volume a un impact moyen positif sur la volatilité toutefois, son impact marginal est décroissant. Aussi, il ne semble pas que l'effet doive être décomposé en deux parties : un effet de liquidité et un effet informationnel. Par effet de liquidité, nous entendons que, toutes choses égales par ailleurs, plus le volume est important, plus la limite touchée en carnet est profonde et donc plus la volatilité est grande. L'effet informationnel étant l'effet marginal du volume. Il ressort alors que l'effet informationnel est plus important pour les petites transactions. Il ne semble donc pas que les volumes importants soient interprétés comme des signaux informationnels puisqu'on pourrait attendre un effet marginal plus grand. Il semble donc que, sur un marché dirigé par les ordres, les informés ne préfèrent pas les transactions de taille importante.
L'effet de la duration sur la volatilité semble similaire. Nous observons un impact moyen positif (même s'il est décalé dans le temps) et un effet marginal décroissant. La volatilité est donc une fonction croissante de la duration mais sa sensibilité est d'autant plus forte que la duration est faible.

Il semble donc que la dynamique du processus de transaction diffère selon que le marché est dirigé par les prix ou par les ordres. Sur les marchés dirigés par les prix on observe généralement que les périodes de forte intensité des transactions suivent les rendements ou les volumes importants et que les périodes de faible duration sont associées à des volatilités plus importantes. Cela suggère que, sur les marché dirigés par les prix, une forte intensité des transactions coïncide avec la présence d'informés (modèle de Easley & O'Hara (1992)). Les résultats que nous avons obtenus fournissent une image différente. En effet, nous observons que, sur un marché dirigé par les ordres, les périodes de forte intensité des transactions suivent les volumes importants et les volatilités faibles et que la duration et le volume ont tous les deux une influence positive sur la volatilité. On a également vu que la sensibilité marginale de la volatilité à ces deux variables était plus forte pour les petites valeurs. La sensibilité de la volatilité est donc plus importante pour les faibles durations et les faibles volumes.
La concentration de l'activité semble donc être majoritairement d'origine non informationnelle (à l'instar des modèles de Admati & Pfleiderer (1988) ou de Back & Pedersen (1998)). D'autre part, l'effet positif de la duration sur la volatilité suggère que ce sont les périodes avec une faible intensité des transactions qui révèlent la présence d'information.

Les relations mises en évidence entre les différents processus sont cohérentes avec l'éco-

logie d'un marché dirigé par les ordres (cf. Handa *et al.* (1998)). En effet, ce type d'organisation nécessite un équilibre dynamique entre la consommation et la fourniture de liquidité qui peuvent induire des relations moyennes et marginales compatibles avec celles observées. Sans autre source que les transactions il est difficile sinon impossible d'attribuer une origine (asymétrie, hétérogénéité, liquidité) aux phénomènes.

Enfin, et ce sera l'objet de notre prochain chapitre, l'examen du problème que nous avons mené jusqu'ici n'est que partiel puisqu'il néglige l'information apportée par le carnet d'ordres. Les réponses auxquelles nous aboutissons ne peuvent donc être que partiales et ne nous permettent pas de comprendre l'origine des relations. Nous espérons que le prochain chapitre, en incluant le carnet d'ordres dans cette analyse, apportera plus de réponses.

Chapitre 6

Incorporation du carnet dans l'analyse du processus de transactions

Nous avons vu dans le chapitre précédent que les relations entre duration , volume et volatilité sur le marché parisien étaient différentes de celles observées sur le NYSE (d'où provient la majorité des résultats). La microstructure semble modifier les relations entre les caractéristiques des transactions. Une première analyse des différences entre ces deux marchés en souligne deux principales :

i.) Le mécanisme de fourniture de la liquidité

Sur le NYSE, la liquidité est fournie conjointement par un spécialiste qui a des obligations contraignantes[1] en terme de quantités minimales et de continuité des prix et par un carnet d'ordres. Ces deux fournisseurs de liquidité sont en concurrence, le carnet ayant la priorité sur le spécialiste (structure mixte)[2]. Le marché de parisien est un marché dirigé par les ordres. La liquidité est fournie exclusivement par les investisseurs finaux qui introduisent des ordres en carnet.

La différence entre ces deux mécanismes semble, à première vue, plus importante qu'elle ne l'est en réalité. En effet, la liberté laissée aux membres du marché parisien leur permet d'agir en compte propre. Ils peuvent donc introduire des ordres en carnet de manière à assumer le rôle d'un teneur de marché. La revue de la littérature concernant l'existence de tenue de marché implicite dans les marchés dirigés par les ordres présentée par Minguet (2003, sections 3.1 et 3.2 p. 295-297) souligne que ce phénomène est observé dans tous les marchés et peut provenir soit de l'incapacité de l'organisme de surveillance à em-

[1] Les obligations du spécialiste sont autant positives (il doit toujours proposer des cotations des deux côtés du marché) que négatives (il ne doit pas empêcher l'exécution d'un ordre en carnet aussi ou plus favorable que les siens). Ces obligations font l'objet d'une évaluation par les bourses. Le lecteur intéressé par de plus amples détails peut se référer à Harris (2003) chapitre 24 pages 494-512.

[2] Chung *et al.* (1999) analysent les apports respectifs des ces deux fournisseurs de liquidité et démontrent que leur compétition a une influence positive et significative sur la fourchette.

pêcher ce comportement (e.g. bourse de Tokyo) ou d'une opportunité laissée aux agents. Les résultats présentés par Declerck (2000*a*) vont plus loin. Ils démontrent que, sur les actions du CAC 40, certains membres du marché ont pour activité principale la tenue de marché. De plus, bien que ces derniers n'aient aucune obligation légale, ils introduisent des ordres qui contribuent significativement à la liquidité des deux côtés de la fourchette dans les périodes de hausse comme de baisse. On peut donc les qualifier de teneurs de marché implicites puisqu'ils remplissent effectivement un rôle similaire à celui auquel sont contraints les spécialistes du NYSE[3]. Il semble donc que ces deux marchés soient des structure mixtes et que seule la dominante du marché change : dirigé par les ordres sur le marché parisien et par les prix sur le marché new-yorkais. Même si les différences pratiques sont faibles, il n'en demeure pas moins que les différences théoriques entre ces deux structures peuvent être importantes. Par exemple, la modélisation proposée par Seppi (1997) arrive à la conclusion qu'une structure mixte fournit une meilleure liquidité aux petits ordres alors qu'un marché dirigé par les ordres offre une liquidité plus importante aux ordres de taille moyenne.

ii.) La diffusion de l'information

On a vu que les deux structures sont en pratique assez proches. Toutefois, les teneurs de marché implicites dans les marché dirigés n'ayant pas les obligations contractuelles du spécialiste, ils n'ont pas non plus les avantages concédés à leur homologues new-yorkais.

Sur le NYSE, les obligations du teneur de marché ne sont que pour des quantités minimales et des fourchettes de prix. Pour traiter on doit s'adresser à lui et négocier un prix (une quantité) éventuellement plus favorable (importante). L'information concernant la fonction de demande n'est donc pas immédiatement accessible à tous les investisseurs.

En comparaison, le carnet d'ordres est transparent. Les teneurs de marché implicites sont soumis aux mêmes règles de diffusion d'information que tous les autres membres. Tous les investisseurs peuvent donc observer les fonctions d'offre et de demande agrégées. La seule incertitude qui pèse concerne les ordres cachés. En effet, il est possible d'introduire un ordre dont seulement une fraction est dévoilée en carnet[4]. Les résultats présentés par D'Hondt *et al.* (2003) démontrent l'importance de la liquidité cachée et suggèrent que,

[3]On peut en outre noter qu'il existe sur le marché parisien des contrats d'animation et de fourniture de liquidité pour les titres de moyenne importance. Pour de plus amples détails, le lecteur peut se reporter au site internet d'Euronext Paris (`http://www.bourse-de-paris.fr`). Declerck & Hazart (2000) démontrent l'influence bénéfique de ces contrats sur la liquidité.

[4]L'ordre apparaît par tranche dans le carnet. Un ordre au prix spécifié et de la quantité dévoilée entre dans le calcul du carnet. Dès que cet ordre est entièrement exécuté, un nouvel ordre identique est automatiquement introduit à cette nouvelle heure : un ordre caché perd la priorité temps à chaque nouvelle tranche. On doit remarquer que ces ordres, s'ils permettent aux investisseurs de masquer l'importance de la quantité demandée, les contraint aussi à renoncer à la priorité temps. Ces ordres ne donnent donc que peu de "pouvoir" comparé au spécialiste du NYSE qui peut de manière discrétionnaire choisir d'offrir, s'il le souhaite, un prix meilleur ou une quantité plus importante que celle présente dans le carnet d'ordres.

sur la bourse de Paris, ces ordres sont principalement utilisés par des investisseurs dans le cadre d'une activité de tenue de marché implicite.

On peut donc dire que le marché parisien offre une plus grande transparence avant transaction que le NYSE.

On voit donc apparaître de légères différences dans le mécanisme d'échange et de diffusion de l'information. Si le mécanisme d'échange et les participants aux marchés semblent proches sur le NYSE[5] et le marché parisien, la quantité et/ou la qualité d'information diffusée sont plus diverses. Après avoir mis en évidence dans le précédent chapitre que les relations entre les caractéristiques des transactions sur le marché parisien étaient différentes de celles observées sur le NYSE, nous essayons d'éclairer l'impact de l'information apportée par le carnet d'ordres sur les évolutions des caractéristiques des transactions. Cette étude sera donc purement exploratoire : au vu des modèles et des études présentés dans le chapitre 4 de cette partie, nous essaierons d'analyser les relations mises en évidence par les modélisations.

De manière intuitive, le modèle trivarié utilisé dans le précédent chapitre sera étendu afin d'incorporer les variables issues du carnet. L'analyse effectuée ici est donc proche de l'objectif poursuivi par Coppejans *et al.* (2003) qui étudient la relation entre la liquidité et la résilience des prix. Notre méthode sera donc similaire puisque nous employons également un système VAR. Cependant une différence majeure est que nous utilisons, comme dans le précédent chapitre modèle de type GARCH (inspiré de Manganelli (2002)) alors que Coppejans *et al.* (2003) préfèrent un modèle structurel.

La quantité d'information apportée par les 5 meilleures limites est importante. En effet, nous disposions au départ de 30 variables décrivant les 5 premières limites à l'achat et à la vente en prix, volume et en nombre d'ordres. La section 4.2.2 nous a déjà permis de construire des variables interprétables en termes de direction, d'hétérogénéité et d'intérêt global à partir des 5 meilleures limites. Le nombre de variables demeurant élevé, il nous semble encore pertinent d'extraire l'information la plus saillante de cet ensemble. La section 6.1 proposera et analysera une méthode d'extraction de l'information des 5 meilleures limites. Nous étudierons ensuite l'influence de ces variables d'abord avec des modèles univariés dans la section 6.2 puis ensuite à l'aide d'une modélisation multivariée dans la section 6.3. Enfin, la section 6.4 conclura.

[5]Cette proposition doit être entendue avec précaution. Par exemple, les résultats obtenus par Corwin (1999) soulignent que les spécialistes ont une influence significative sur les coûts de transaction, la liquidité et le niveau de bruit des prix. De plus, ces effets ne sont éliminés ni par la concurrence ni par les mécanismes de surveillance du NYSE.

6.1 Agrégation de l'information du carnet

Nous avons, dans le chapitre 4, défini 14 mesures pour chaque dimension du flux informationnel. Ce travail nous a déjà permis de construire à partir du carnet des variables interprétables en terme de paramètres du flux informationnel. Toutefois, une simplification (réduction du nombre de variables) permettant d'extraire l'information pertinente de ces variables serait profitable.

Dans un sujet connexe, l'analyse des variations systémiques de la liquidité, il est usuel d'employer une analyse en composante principale (se référer par exemple à Hasbrouck & Seppi (2001)). Notre optique est proche de cette dernière tout étant différente. Nous employons l'analyse en composantes principales sur les différentes mesures d'une même dimension afin de faire ressortir les relations dominantes entre les différents aspects de la liquidité mesurée en terme d'état du carnet. Le but est donc de définir pour chaque dimension un nombre restreint de composantes principales orthogonales entre elles et synthétisant un maximum de l'information initiale.

À chaque fois, l'analyse en composantes principales des différentes dimensions fait ressortir quelques axes interprétables et synthétisant une partie significative de l'information. La sous-section 6.1.1 présentera une analyse détaillée des résultats obtenus pour chaque dimension de Sanofi. Enfin, nous synthétiserons les composantes principales obtenues pour les différents titres dans la sous-section 6.1.2.

6.1.1 Sanofi

La table 6.1 donne pour les 3 dimensions les premières composantes principales (jusqu'à la première ayant une contribution inférieure à 10 %).

Hétérogénéité

La première composante principale de l'hétérogénéité représente 19% de la dispersion initiale des variables. C'est la seule composante dont la contribution est supérieure à 10% puisque la seconde composante ne représente que 8,8%.

Les significations de ces deux axes sont très différentes. En effet, le premier axe mesure l'hétérogénéité de la fourchette alors que le second associe l'hétérogénéité nombre et volume. On peut remarquer que les variables mesurées sur la première, la deuxième et la 5^e meilleure limite sont souvent moins bien représentées que celles sur les 3^e et 4^e. Cela suggère que ces limites fournissent un signal moins bruité que les autres.

Direction

Les premières composantes principales de la direction représentent respectivement 19, 11 et 9,8% de la dispersion initiale.

TABLE 6.1
Composantes principales des dimensions
Sanofi

	Hétérogénéité		Direction			Profondeur		
Variable	H_1	H_2	D_1	D_2	D_3	P_1	P_2	P_3
Valeur propre	2,722	1,235	2,713	1,599	1,366	3,646	1,602	1,365
Contribution	0,194	0,088	0,194	0,114	0,098	0,260	0,114	0,098
N_1	-0,018	0,274	-0,025	0,240	0,661	0,002	0,366	0,182
N_2	-0,018	0,362	0,045	0,321	-0,137	0,042	0,314	-0,076
N_3	-0,013	0,423	0,043	0,355	-0,132	0,065	0,333	-0,114
N_4	-0,013	0,399	0,051	0,388	-0,153	0,066	0,327	-0,368
N_5	-0,017	0,369	0,082	0,370	-0,058	0,068	0,212	0,556
S_2	0,418	0,020	0,414	-0,073	0,045	0,470	-0,091	-0,018
S_3	0,530	0,015	0,519	-0,070	0,028	0,503	-0,076	-0,016
S_4	0,544	0,003	0,536	-0,042	0,017	0,505	-0,059	-0,007
S_5	0,495	-0,002	0,504	-0,024	0,018	0,490	-0,044	0,023
V_1	0,036	0,132	-0,003	0,228	0,664	-0,014	0,385	0,180
V_2	0,015	0,219	0,026	0,284	-0,110	0,035	0,326	-0,101
V_3	-0,004	0,290	-0,001	0,298	-0,140	0,056	0,320	-0,141
V_4	0,012	0,286	0,024	0,324	-0,151	0,072	0,302	-0,347
V_5	0,004	0,296	0,043	0,294	-0,035	0,079	0,193	0,566

Cette table fournit les premières composantes principales (jusqu'à la première ayant une contribution relative inférieure à 10%) des 3 mesures (hétérogénéité, direction, profondeur). *N* désigne les mesures sur le nombre d'ordre, *V* sur le volume et *S* sur la fourchette.

Comme pour l'hétérogénéité, le premier axe synthétise la direction donnée par les mesures de la fourchette alors que le second met en évidence une association entre les mesures sur le nombre et le volume. Le troisième axe oppose les mesures d'activité sur la meilleure limite (positive) à celles sur les autres (la cinquième limite n'est que très peu représentée). Une forte valeur sur cet axe caractérise donc les moments où le déséquilibre de la profondeur de l'activité est important à la meilleure limite et plus faible à l'intérieur du carnet. Il est possible que cette situation soit le résultat d'une transaction ayant consommé les réserves proposées à la meilleure limite ou qu'il soit la conséquence d'une offre en carnet qui se focalise sur cette limite et induise un déséquilibre plus important de celle-ci.

Profondeur

L'analyse des variables de profondeur met également en évidence trois axes dont les contributions respectives sont 26, 11 et 9,8%. Une fois de plus, le premier axe représente la dimension prix alors que le second représente l'activité (en nombre ou en volume).

Une forte valeur sur le 3^{e} axe caractérise une profondeur activité élevée aux limites 1 et 5 et faible aux limites 2, 3 et 4. L'interprétation est donc similaire à celle du 3^{e} axe de la direction ce qui suggère que ces deux limites évoluent de concert.

On peut remarquer que, pour les mesures d'activité, plus la limite est profonde en carnet moins elle est bien représentée. Cela est, au moins en partie, une conséquence de l'utilisation de mesures agrégées. En effet, les variables à une limite l sont les sommes des variables observées aux limites inférieures ou égales. Il est alors logique que les variables soit d'autant mieux représentées qu'elles sont proches de la meilleure limite puisqu'elles sont incluses dans les mesures suivantes.

6.1.2 Composantes principales des différents titres

L'analyse de Sanofi a fait ressortir que les différentes dimensions pouvait être synthétisées en un nombre réduit de composantes principales (environ 3 par dimension) tout en conservant une grande part de l'information initiale (d'environ 30% pour l'hétérogénéité à près de 50% pour la profondeur). De plus, nous avons noté qu'à chaque fois le premier axe représente la dimension fourchette du concept, le second la dimension activité alors que le troisième oppose les mesures d'activité sur la première limite (éventuellement associée à une autre) et celles à l'intérieur du carnet. Ce fait souligne qu'il existe une différence entre le signal apporté par le stock d'ordres non satisfaits (les limites intérieures au carnet) et la meilleure limite.

Les tables 6.2 à 6.4 présentent les caractéristiques des axes factoriels obtenus pour les différents titres et les différentes dimensions. L'interprétation donnée dans les tables qui suivent obéit à la logique que nous avons illustrée lors de l'analyse de Sanofi.

TABLE 6.2
Axes factoriels de l'hétérogénéité

	n°	Contribution	Interprétation
St-Gobain	1	0,21	Spread
	2	0,10	Volume et nombre
Schneider	1	0,22	Spread
	2	0,11	Volume et nombre
	3	0,08	Opposition entre le volume et le nombre aux limites 2 et 4 (+) et aux limites 1 et 5 et le nombre 4 (-)
Sanofi	1	0,19	Spread
	2	0,09	Volume et nombre
Thomson	1	0,21	Spread
	2	0,09	Volume et nombre
Accor	1	0,21	Spread
	2	0,08	Volume et nombre
Lyonnaise	1	0,21	Spread
	2	0,10	Volume et nombre
Lagardère	1	0,21	Spread
	2	0,09	Volume et nombre

Les résultats obtenus sur l'hétérogénéité sont particulièrement stables puisque les deux premiers axes factoriels ont toujours la même interprétation avec des niveaux de significativité voisins et que le nombre d'axes dont la significativité dépasse 10% est toujours de 1 ou 2. Les constatations que nous avions fait concernant Sanofi sont directement généralisables à l'ensemble des titres.

TABLE 6.3
Axes factoriels de la direction

	n°	Contribution	Interprétation
St-Gobain	1	0,21	Association volume et prix.
	2	0,12	Opposition entre le volume et le nombre (+) et spread (-)
	3	0,10	Opposition entre le V. et le N. aux limites 1 et 2 et aux limites 4 et 5 (-)
Schneider	1	0,22	Association (+) entre toutes les variables
	2	0,14	Opposition entre le volume et le nombre (+) et le spread
	3	0,10	Opposition entre le V. et le N. aux limites 1 et 2 (+) et aux limites 4 et 5 (-)
Sanofi	1	0,19	Spread
	2	0,11	Association entre le volume et nombre
	3	0,10	Opposition entre le V. et le N. à la 1^{ere} limite (+) et aux autres
Thomson	1	0,20	Spread
	2	0,11	Association entre le volume et le nombre
	3	0,10	Opposition entre le V. et le N. aux limites 2 à 4 (+) et aux limites 3 et 5 (-)
Accor	1	0,20	Spread
	2	0,10	Association entre le volume et le nombre
	3	0,10	Opposition entre le volume et le nombre aux limites 1 et 5 (+) et au autres (-)
Lyonnaise	1	0,16	Association (+) entre toutes les variables3
	2	0,12	Opposition entre le V. et le N. (+) et le spread (-)
	3	0,10	Opposition entre le V. et le N. aux limites 1 et 2 (+) et le V. et le N. aux deux dernières limites (-)
	4	0,10	Opposition entre le V. et le N. aux limites 1, 4 et 5 (+) et aux limites 2 et 3 (-)
Lagardère	1	0,20	Spread
	2	0,11	Volume et nombre
	3	0,10	Opposition entre le V. et le N. à la 4^{e} limite (+) et à la 5^{e} limite (-)
	4	0,10	Opposition entre le V. et le N. à première limite (+) et aux limites 3 et 4 (-)

Les axes factoriels de la direction sont légèrement plus variables selon les titres puisque le nombre et l'interprétation des axes ayant une contribution supérieure à 10% varient. Selon l'interprétation des 2 premiers axes, on distingue trois grands groupes :

– Schneider et Lyonnaise

  Le premier axe exprime une association de toutes les variables. Ainsi, l'activité et la fourchette donnent principalement des indications directionnelles convergentes. Alors que le second axe représente les cas où l'indication directionnelle de l'activité est forte alors que celle de la fourchette est faible.

– Sanofi et Lagardère

  Le premier axe représente la direction donnée par la fourchette et second celle de l'acti-

vité. Ces titres ne mettent donc pas en évidence d'accord entre les indications directionnelle de la fourchette et de l'activité qui sont orthogonales. Accor est également proche de ce groupe toutefois, le second axe traduit une opposition entre les mesures d'activité et la fourchette.

– St-Gobain et Thomson
  Le premier axe met majoritairement en avant une association entre l'indication directionnelle de la fourchette et du nombre. Le second axe se focalise alors sur les cas où les toutes les mesures d'activité donnent une indication convergente opposée à celle issue de la fourchette.

Les autres axes représentent les relations entre les mesures d'activité aux différentes limites. Ils se fondent toujours sur l'opposition entre deux groupes de limites le volume et le nombre étant le plus souvent représentés de manière équivalente ce qui montre la forte association de ces deux variables. Les associations entre les différentes limites étant variables d'un titre à l'autre, il est difficile de tirer une conclusion.

C'est pour la profondeur qu'on obtient les résultats les moins homogènes. Comme pour l'hétérogénéité ou la direction, les deux premiers axes (excepté Lyonnaise) représentent respectivement la profondeur de la fourchette et celle de l'activité. Les autres axes rendent les relations entre les mesures d'activité aux différentes limites. On doit remarquer que le nombre de composantes ayant une contribution importante est plus élevé et plus variable que les deux autres dimensions, ce qui rend difficile l'interprétation des relations mises en évidence en fonction des titres.

Il ressort donc de cette analyse que les 2 premiers axes factoriels reflètent respectivement le caractère prix de la dimension puis le caractère activité qui sont orthogonaux. La complexité des relations entre les mesures d'activité aux différentes limites est alors rendue par les axes suivants. Il semble donc indispensable d'étudier aux moins deux axes afin de tenir compte des deux composantes des dimensions.

L'analyse des deux premiers axes factoriels de chaque dimension présentée en annexe G souligne la richesse et la diversité de l'information apportée par ces variables. En particulier, cette analyse suggère que la profondeur de la fourchette est plus sensible à l'hétérogénéité alors que celle de l'activité est plus réactive à la direction du signal (asymétrie d'information) soit, un accord avec le modèle de Dupont (2000). Elle confirme en outre que les relations observées entre les variables sont cohérentes avec les concepts qu'elles sont sensées représenter. Aussi, nous considérerons les 2 premiers axes factoriels de chaque dimension.

TABLE 6.4
Axes factoriels de la profondeur

	n°	Contribution	Interprétation
St-Gobain	1	0,26	Spread
	2	0,13	Volume et nombre
	3	0,11	Opposition entre le V. et le N. aux limites 2 et 3 (+) et à la 5^{e} limite plus pour le $V._4$ (-)
	4	0,10	Opposition entre le V. et le N. aux limites 1, 2 et 5 (+) et à la 4^{e} limite (-)
	5	0,10	Opposition entre le V. et le N. aux limites 1, 2 et 4 (+) et aux autres limites
Schneider	1	0,27	Spread
	2	0,17	Volume et nombre
	3	0,13	Opposition entre le V. et le N. aux limites 3 et 5 (+) et aux limites 2 et 4 (-)
	4	0,12	Opposition entre le V. et le N. aux 3 premières limites (+) et aux 2 dernières
	5	0,11	Opposition entre le V. et le N. aux limites 1 et 5 (+) et à la 3^{e} limite
	6	0,10	Opposition entre le V. et le N. aux limites 1, 3 et 4 (+) et aux limites 2 et 5
Sanofi	1	0,26	Spread
	2	0,11	Volume et nombre
	3	0,10	Opposition entre le V. et le N. aux limites 1 et 5 (+) et aux limites 3 et 4 plus $N._2$ (-)
Thomson	1	0,25	Spread
	2	0,13	Volume et nombre
	3	0,11	Opposition entre le V. et le N. aux limites 2 et 4 avc $V._1$ (+) et à la dernière limite (-)
	4	0,10	Opposition entre le V. et le N. aux limites 1, 4 et 5 (+) et aux limites 2 et 3
	5	0,10	Opposition entre le V. et le N. aux limites 1 et 2 (+) et aux limites 3 et 4 (-)
	6	0,10	Opposition entre le V. et le N. aux limites 1 et 3 (+) et aux limites 2, 4 et 5
Accor	1	0,26	Spread
	2	0,12	Volume et nombre
	3	0,11	Opposition entre le V. et le N. aux limites 2 et 3 (+) et aux limites 4 et 5 (-)
	4	0,11	Opposition entre le V. et le N. aux limites 1 et 2 et le V. à la limite 5 (+) et aux limites 3 et 4 (-)
	5	0,11	Opposition entre le V. et le N. aux limites 2 et 4 (+) et aux limites 1, 3 et 5
	6	0,10	Opposition entre le V. et le N. à la 1_{ere} limite (+) et aux limites 2 et 5 plus $N._3$ (-)
	7	0,04	Opposition entre le nombre (+) et le volume
Lyonnaise	1	0,26	Spread et volume 4 et 5
	2	0,13	Opposition entre le V. et le N. (+) et S. aux deux premières limites (-)
	3	0,10	Opposition entre le V. et le N. aux 2 dernières limities (+) et aux 3 premières
	4	0,10	Opposition entre le V. et le N. aux limites 1, 2 et 5 (+) et aux limites 3 et 4
Lagardère	1	0,26	Spread
	2	0,11	Volume et nombre
	3	0,11	Opposition entre le V. et le N. aux limites 2 et 3 (+) et aux limites 1, 4 et 5
	4	0,10	Opposition entre le V. et le N. aux limites 2, 4 et 5 (+) et à la 1_{ere} limite
	5	0,10	Opposition entre le V. et le N. aux limites 3 et 4 (+) et aux limites 2 et 5 (-)
	6	0,10	Opposition entre le V. et le N. aux limites 3 et 5 (+) et à la 4^{e} (-)

6.2 Modélisation univariée

Par la suite nous analysons comment l'état du carnet précédant une transaction permet d'améliorer la compréhension du processus d'incorporation de l'information. Pour ce faire nous analysons la relation entre l'état du carnet et les caractéristiques des transactions. Nous utilisons donc le sous-ensemble des états du carnet suivis d'une transaction.

La première question que l'on doit alors se poser concerne alors l'existence d'une différence entre les états du carnet suivis d'une transaction et ceux suivis d'un autre état du carnet. La table 6.5 montre que les états du carnet suivis d'une transaction sont différents de ceux suivis d'un autre état du carnet.

Quelques faits ressortent de cette table. Les états du carnet avant une transaction se caractérisent par une profondeur de la fourchette importante, et une hétérogénéité (de la fourchette et de l'activité faible) donc un déséquilibre de l'offre. Ce déséquilibre semble légèrement associé à une moindre liquidité du côté vente. Une analyse de la significativité des différences entre les ces deux catégories d'états du carnet révèle que celles-ci sont, dans la très grande majorité des cas, significatives et cela aussi bien avec des tests paramétriques (anova) que non paramétriques (test de la médiane, de Kolmogorov).

Les études présentées dans la section 4.1 suggèrent un ensemble de relations complexes. Les études de l'hétérogénéité penchent pour une association entre l'activité et la profondeur. On devrait donc observer une relation positive entre les mesures d'activité (l'inverse de la duration, le volume) et l'hétérogénéité ou la profondeur. Les prévisions des modèles d'asymétrie d'information sont encore troubles d'autant plus qu'une majorité de ces modèles ne traitent pas directement de la profondeur du carnet. Il faut alors faire une ou des hypothèses supplémentaires sur la fourniture de liquidité pour en dériver des implications testables vis à vis du carnet. Par exemple, si on considère le modèle de Easley & O'Hara (1992) et qu'on suppose que la liquidité est fournie par des non informés alors on devrait observer en même temps une indication directionnelle importante en valeur absolue, de faible duration, une faible profondeur soit une association négative entre l'inverse de la duration et la profondeur. À l'inverse le modèle de Diamond & Verrecchia (1987) prévoit, sous la même hypothèse, une association entre longue duration et faible profondeur (prévision similaire au courant de l'hétérogénéité). Les études de la dynamique des ordres proposent des relations testables mais ne fournissent que peu de guides dans l'interprétation. On relève globalement la forte autocorrélation (effet diagonal) des ordres, la dynamique de la fourniture de liquidité. Les résultats les plus parlants viennent des modèles pour les marchés dirigés par les ordres (Foucault (1999), Handa *et al.* (2003)) qui mettent en avant une relation quadratique entre les coûts de transaction et l'hétérogénéité et une relation positive à l'asymétrie d'information mais ne disent rien du carnet.

TABLE 6.5
Comparaison des états du carnet avant transaction

	Hétérogénéité		Direction		Profondeur	
	1	2	1	2	1	2
Saint-Gobain	-0,082	-0,016	-0,014	-0,002	0,101	-0,002
	(1,76)	(1,16)	(1,73)	(1,32)	(1,9)	(1,37)
	0,070	0,014	0,012	0,002	-0,085	0,002
	(1,64)	(1,15)	(1,67)	(1,31)	(1,88)	(1,34)
Schneider	-0,058	-0,054	-0,026	-0,033	0,079	0,046
	(1,94)	(1,28)	(1,83)	(1,41)	(1,97)	(1,57)
	0,050	0,046	0,023	0,028	-0,068	-0,039
	(1,58)	(1,23)	(1,73)	(1,41)	(1,89)	(1,48)
Sanofi	-0,115	-0,033	-0,033	-0,009	0,202	-0,029
	(1,78)	(1,12)	(1,69)	(1,28)	(1,96)	(1,3)
	0,065	0,019	0,019	0,005	-0,115	0,016
	(1,57)	(1,1)	(1,62)	(1,25)	(1,87)	(1,25)
Thomson	-0,196	-0,034	-0,009	-0,018	0,214	0,001
	(2,01)	(1,12)	(1,75)	(1,25)	(1,95)	(1,43)
	0,105	0,018	0,005	0,010	-0,114	0,000
	(1,55)	(1,1)	(1,66)	(1,23)	(1,83)	(1,27)
Lyonnaise	-0,112	-0,030	-0,005	-0,002	0,117	0,002
	(1,88)	(1,09)	(1,73)	(1,22)	(1,94)	(1,34)
	0,085	0,023	0,003	0,001	-0,089	-0,002
	(1,58)	(1,08)	(1,66)	(1,21)	(1,87)	(1,28)
Accor	-0,073	-0,033	0,020	0,014	0,114	-0,004
	(1,84)	(1,18)	(1,74)	(1,34)	(1,94)	(1,42)
	0,066	0,030	-0,018	-0,013	-0,103	0,004
	(1,57)	(1,15)	(1,69)	(1,32)	(1,91)	(1,32)
Lagardère	-0,109	-0,036	0,000	-0,012	0,184	0,013
	(1,83)	(1,12)	(1,72)	(1,23)	(1,98)	(1,32)
	0,081	0,027	0,000	0,009	-0,137	-0,010
	(1,6)	(1,11)	(1,65)	(1,23)	(1,86)	(1,21)

Dans cette table la première ligne de chaque cellule donne la moyenne des états du carnet avant une transaction. Le deuxième ligne donne les états du carnet non suivis d'une transaction. Les valeurs entre parenthèses sont les écarts-types des variables.

Afin de d'explorer ces relations nous estimons le modèle suivant :

$$Tra_t = \alpha_0 + \alpha_1 Het_1 + \alpha_2 Het_2 + \alpha_3 Dir_1 + \alpha_4 Dir_2 + \alpha_5 Pro_1 + \alpha_6 Pro_2 \quad (6.1)$$

où Tra_t est l'une des caractéristiques des transactions (la duration, le volume, le rendement ou la volatilité). Nous régressons donc les caractéristiques des transactions sur les variables décrivant les états du carnet. Les tables 6.6 à 6.9 détaillent les résultats des estimations par les moindres carrés ordinaires pour chacun des titres et des caractéristiques des transactions.

Duration

La relation entre la duration et les mesures du carnet est particulièrement forte (le R^2 oscille entre 33% et 40%), les paramètres sont souvent très significatifs et leur signe est globalement stable entre les différents titres.

L'hétérogénéité prix a un effet positif très significatif sur la duration alors que l'influence de l'hétérogénéité activité est moins forte mais elle aussi positive et souvent significative. Ainsi, plus l'offre des deux côtés du carnet est symétrique, plus la duration est importante (moins l'intensité des transactions est élevée.

L'interprétation des variables de direction est plus complexe. La direction prix a une influence globalement positive mais n'est significative que dans 3 cas (St-Gobain, Schneider et Lagardère). La direction activité a une influence plus significative (5 cas de significativité) mais elle est 2 fois positive et 3 fois négative. Au total, seul Thomson et Lyonnaise se distinguent par une absence complète de significativité de la direction. Il semble donc que la relation entre l'indication fournie par le carnet et la duration soit asymétrique pour certains titres et que les investisseurs soient plus pressés en lorsqu'il y a une moindre liquidité du côté vente. Une faible (importante) liquidité à l'achat (à la vente) induit une augmentation (une réduction) de la duration pour St-Gobain, Schneider et Lagardère.

Les deux dimensions de la profondeur ont un impact négatif et significatif. Plus la profondeur est élevée, moins la duration est importante. Ainsi, une augmentation de l'intérêt global induit une augmentation de l'intensité de transaction.

Ces résultats suggèrent que l'intensité de transaction est maximale lorsque l'intérêt global est élevé est que celui-ci se caractérise par une hétérogénéité et une direction moyennes (déséquilibre léger de la demande). Les cas de très forte hétérogénéité ou de très forte indication directionnelle réduisent l'intensité des transactions. Cela suggère que les agents sont très averses aux risques que ceux-ci proviennent d'une asymétrie ou d'une hétérogénéité. Toute augmentation de ce risque entraîne alors une diminution de l'intensité des transactions en raison du retrait de certains agents.

TABLE 6.6
Relations duration - carnet

	St-Gobain	Schneider	Sanofi	Thomson	Accor	Lyonnaise	Lagardère
Int	1,0026***	1,0069***	1,0089***	1,0144***	1,0055***	1,0046***	1,0102***
	(199,89)	(190,95)	(149,08)	(114,13)	(184,31)	(214,57)	(164,13)
Het_1	0,0209***	0,0100*	0,0212***	0,0361***	0,0121**	0,0168***	0,0203***
	(4,53)	(1,96)	(3,81)	(5,44)	(2,04)	(3,45)	(5,21)
Het_2	0,0156***	0,0257***	0,0115*	0,0089	0,0162***	0,0212***	0,0016
	(3,16)	(5,10)	(1,76)	(0,99)	(2,92)	(4,84)	(0,28)
Dir_1	0,0120***	0,0119***	0,0055	0,0055	0,0038	-0,0035	0,0086**
	(4,11)	(4,03)	(1,40)	(1,13)	(1,18)	(-1,25)	(2,48)
Dir_2	0,0108***	0,0173***	-0,0214***	-0,0087	-0,0150***	-0,0065*	-0,0095**
	(2,77)	(4,28)	(-4,17)	(-1,26)	(-3,34)	(-1,85)	(-1,96)
Pro_1	-0,0047*	-0,0256***	-0,0285***	-0,0323***	-0,0277***	-0,0188***	-0,0364***
	(-1,66)	(-10,57)	(-8,43)	(-7,80)	(-9,23)	(-6,60)	(-12,31)
Pro_2	-0,0560***	-0,0487***	-0,0093*	-0,0380***	-0,0539***	-0,0160***	-0,0511***
	(-14,24)	(-14,79)	(-1,73)	(-7,17)	(-14,24)	(-4,77)	(-10,83)
$Adjr-sq$	0,3780	0,3715	0,3488	0,3309	0,3773	0,4030	0,3681

TABLE 6.7
Relations volume - carnet

Int	1,0143***	1,0042***	1,0192***	1,0194***	1,0108***	1,0135***	1,0148***
	(109,35)	(95,65)	(113,90)	(76,83)	(106,87)	(117,67)	(91,64)
Het_1	0,0171***	-0,0013	0,0156***	0,0433***	0,0144***	0,0239***	0,0279***
	(3,05)	(-0,29)	(3,07)	(5,89)	(2,95)	(4,55)	(4,25)
Het_2	-0,0066	-0,0217*	-0,0023	0,0243**	0,0330***	-0,0310***	-0,0268**
	(-0,59)	(-1,83)	(-0,28)	(1,96)	(2,62)	(-3,65)	(-2,29)
Dir_1	0,0026	0,0140**	0,0075	-0,0062	-0,0430***	0,0094*	-0,0235***
	(0,46)	(2,12)	(1,53)	(-0,89)	(-7,73)	(1,87)	(-3,67)
Dir_2	0,0104	0,0311***	0,0171***	-0,0068	-0,0161**	0,0009	0,0007
	(1,48)	(4,20)	(2,60)	(-0,68)	(-2,18)	(0,14)	(0,08)
Pro_1	-0,1118***	-0,0883***	-0,0701***	-0,0450***	-0,0672***	-0,1012***	-0,0755***
	(-20,94)	(-16,32)	(-15,47)	(-6,74)	(-13,21)	(-19,43)	(-14,96)
Pro_2	0,1258***	0,0773***	0,0641***	0,0774***	0,1519***	0,0642***	0,1357***
	(8,78)	(4,84)	(8,62)	(7,24)	(9,29)	(7,90)	(9,17)
$Adjr-sq$	0,1674	0,1325	0,2557	0,1847	0,1799	0,1766	0,1597

Ces tables donnent les résultats des estimations du modèle (6.1) entre la duration (le volume) et les variables informationnelles du carnet. Les t-values indiquées entre parenthèses en dessous des paramètres sont calculées en utilisant les écarts-types robustes de White (1980). Un paramètre significatif au seuil de 1% est indiqué par ***, au seuil de 5% par ** et à 10% par *.

Volume

Les résultats concernant le volume sont légèrement moins bons puisque le R^2 évolue entre 13% et 25%.

L'hétérogénéité prix a une influence positive (significative dans 6 cas sur 7) alors que l'hétérogénéité de l'activité a des effets instables (positif et négatif) et souvent peu ou pas significatifs. Ainsi, il apparaît que plus l'hétérogénéité prix est élevée (lorsque la fourchette est symétrique par rapport au milieu de fourchette), plus la taille de transaction augmente.

Il arrive que les variables de direction soient significatives. Ce phénomène est toutefois difficile à interpréter en raison des changements de signes et de l'absence de significativité pour certains titres. On peut cependant remarquer qu'il semble que le volume de transactions ait, pour certains titres, un comportement plus complexe que d'autres avec une composante asymétrique.

La profondeur a toujours un impact significatif. Le prix a une influence négative alors que celle de la profondeur activité est positive. Cela suggère donc une relation de substitution entre ces deux dimensions de la profondeur. Lorsque la profondeur prix est élevée, les transactions sont de taille plus faible que lorsque la profondeur volume est élevée. Cela indique les agents ne déterminent vraisemblablement pas le volume de leurs transactions de manière exogène mais que celui-ci est fonction des conditions de marché.

Le volume est avant tout sensible à la profondeur. Cette influence change cependant en fonction de l'origine de l'intérêt global (hétérogénéité vs direction). En effet, le volume est plus sensible à l'hétérogénéité prix donc à une composante non informée (selon notre hypothèse). Toutefois, les différences entre les paramètres des indications directionnelles pour les différents titres suggère que selon les titres, les informés peuvent avoir des comportements différents : il est possible que le volume de transactions augmente avec l'asymétrie d'information (e.g. Schneider) mais aussi qu'il diminue (e.g. Accor).

Rendement

La relation entre les rendements et le carnet est beaucoup plus faible puisque le R^2 est souvent inférieur à 1% et ne dépasse jamais 2%.

Les variables d'hétérogénéité ont toutes une influence positive mais peu significative (3 cas pour l'hétérogénéité prix et 5 pour l'hétérogénéité activité). Cela souligne donc que rendement d'une transaction est une fonction croissante de la symétrie de la fourchette. Étant donnée la relation entre direction et hétérogénéité, il apparaît que la significativité de l'hétérogénéité capture une composante symétrique (à l'achat et à la vente) de la relation entre rendement et signal.

Les variables de direction ont une influence très significative et négative sur le rendement. Pour rappel cette mesure repose sur l'hypothèse que la transaction doit se trouver du côté où l'offre est la plus faible. La relation négative que nous observons souligne que la transaction a lieu du côté ou l'offre est la plus importante. Cela conforte la conjecture selon laquelle c'est principalement la liquidité qui influence le côté de la prochaine transaction.

TABLE 6.8
Relations rendement - carnet

	St-Gobain	Schneider	Sanofi	Thomson	Accor	Lyonnaise	Lagardère
Int	0,0083	-0,0038	-0,0061	-0,0257*	0,0043	-0,0006	-0,0255**
	(1,03)	(-0,44)	(-0,59)	(-1,70)	(0,48)	(-0,08)	(-2,27)
Het_1	0,0186*	0,0472*	0,0039	0,0168	0,0422***	0,0219**	0,0262***
	(1,72)	(1,77)	(0,34)	(1,13)	(3,49)	(2,28)	(2,58)
Het_2	0,0116	0,0434***	0,0093	0,0340**	0,0291***	0,0374***	0,0421***
	(1,41)	(4,71)	(0,99)	(2,41)	(3,25)	(4,99)	(3,94)
Dir_1	-0,1328***	-0,1211***	-0,0619***	-0,0933***	-0,1320***	-0,1185***	-0,1015***
	(-25,26)	(-18,63)	(-10,01)	(-10,81)	(-23,60)	(-22,18)	(-15,22)
Dir_2	-0,0477***	-0,0281***	-0,0258***	-0,0410***	-0,1069***	-0,0374***	-0,0709***
	(-7,54)	(-3,83)	(-3,62)	(-4,17)	(-15,52)	(-6,30)	(-8,91)
Pro_1	0,0030	0,0083**	-0,0087	0,0028	-0,0102*	-0,0048	0,0025
	(0,54)	(1,99)	(-1,41)	(0,37)	(-1,90)	(-0,83)	(0,44)
Pro_2	-0,0064	0,0053	0,0166**	0,0078	0,0070	0,0112**	0,0037
	(-0,96)	(0,74)	(2,06)	(0,83)	(0,99)	(2,01)	(0,30)
$Adjr-sq$	0,0134	0,0118	0,0033	0,0060	0,0174	0,0114	0,0079

TABLE 6.9
Relations volatilité - carnet

Int	4,2249***	4,6087***	4,1199***	5,6025***	4,3007***	4,3822***	5,5779***
	(58,61)	(47,12)	(44,99)	(26,33)	(54,59)	(51,77)	(35,12)
Het_1	-0,5751***	-2,2465***	-0,2242**	-0,2278	-0,6759***	-0,4739***	-0,4164**
	(-5,02)	(-3,12)	(-1,98)	(-1,33)	(-4,58)	(-3,84)	(-2,34)
Het_2	-0,0940	-0,2275*	-0,0119	-0,0898	0,1036	-0,0184	-0,2083
	(-1,09)	(-1,93)	(-0,16)	(-0,50)	(1,25)	(-0,27)	(-1,40)
Dir_1	0,1673***	0,0647	0,0122	0,0714	-0,0055	-0,2492***	0,2832**
	(3,38)	(0,57)	(0,21)	(0,61)	(-0,10)	(-3,14)	(2,53)
Dir_2	0,0716	0,4419***	-0,0247	0,2506**	0,1511**	0,2836***	0,3664***
	(1,34)	(4,19)	(-0,41)	(2,50)	(2,48)	(5,91)	(3,64)
Pro_1	-1,5580***	-1,4901***	-1,5873***	-2,3287***	-1,5289***	-1,5709***	-2,2687***
	(-24,55)	(-22,22)	(-23,81)	(-17,84)	(-26,03)	(-22,24)	(-23,07)
Pro_2	-0,0819	-0,3996***	0,1136	-0,0027	0,0118	0,1962**	0,3592**
	(-1,32)	(-3,84)	(1,53)	(-0,03)	(0,19)	(2,57)	(2,38)
$Adjr-sq$	0,0836	0,0802	0,0837	0,0462	0,0843	0,0630	0,0453

Ces tables donnent les résultats des estimations du modèle (6.1) entre le rendement (la volatilité) et les variables informationnelles du carnet. Les t-values indiquées entre parenthèses en dessous des paramètres sont calculées en utilisant les écarts-types robustes de White (1980). Un paramètre significatif au seuil de 1% est indiqué par ***, au seuil de 5% par ** et à 10% par *.

Le fait que l'hétérogénéité et la direction soient significatifs en même temps souligne une asymétrie de la réaction du rendement à une signal. À signal de même amplitude, la réaction du rendement est plus vive si le signal est du côté vente que du côté achat.

Volatilité

Comparée aux rendements, la relation entre la volatilité et les mesures du carnet est assez importante puisque le R^2 va de 4% à 8%.

L'impact de l'hétérogénéité des prix est très significatif et négatif. La volatilité diminue donc avec l'hétérogénéité prix.

L'influence des mesures de direction est plus trouble. La direction prix n'est significative que 3 fois et est deux fois positive et une fois négative. La direction activité a un impact plus significatif (5 cas) et toujours positif.

Comme on pouvait s'y attendre[6] la profondeur prix a une influence très importante et négative alors que l'influence de la profondeur activité est rarement significative (3 fois) et est 2 fois positive et 1 fois négative.

Au total, la volatilité est déjà déterminée par la profondeur prix, elle est ensuite une fonction décroissante de l'hétérogénéité et une fonction croissante de l'asymétrie d'information. Les relations mises en évidence sont compatibles avec les prévisions théoriques sur les concepts mesurés. Le fait que les mesures de direction soient significatives (en plus des mesures d'hétérogénéité) souligne que la réaction de la volatilité est asymétrique en ce sens qu'elle augmente plus vite à la vente qu'à l'achat. La volatilité augmentera donc plus si on se trouve dans le cas d'une faible profondeur et d'une faible hétérogénéité, ce qu'on peut associer à la présence d'une information.

Cette analyse de la relation linéaire entre les caractéristiques des transactions et l'état du carnet fait ressortir la très forte influence de la profondeur qui affecte toutes les caractéristiques à l'exception du rendement : duration (-), volume (+), volatilité (-). Les effets négatifs sur la volatilité et la duration sont consistants avec les modèles d'hétérogénéité mais vont à l'encontre des extrapolations que l'on peut faire à partir du modèle de Easley & O'Hara (1992).

L'hétérogénéité (le plus souvent la dimension prix) a pour effet d'augmenter la duration, le volume et le rendement et elle réduit la volatilité (hétérogénéité prix). Il semble donc qu'une incertitude élevée (un équilibre de l'offre) réduise l'activité des transactions qui sont alors plus grosses et de moindre volatilité. Les prévisions du courant hétérogénéité semblent donc en grande partie vérifiées à l'exception de l'effet sur la duration où une fuite des investisseurs averses aux risques semble prépondérante.

La direction a globalement des effets plus instables. Elle est positivement corrélée à la du-

[6]La relation entre la profondeur prix et la volatilité est attendue puisque qu'une profondeur prix importante signifie que la fourchette augmente lentement lorsque l'on descend en carnet. Il est alors logique d'observer une relation négative entre la profondeur et la volatilité.

ration (direction prix), à la volatilité (direction activité) et négativement aux rendements. Étant donné que la variable direction est signée, son effet diffère selon le signe du signal qui est négatif lorsque l'offre la moins importante est du côté vente. La direction permet donc de refléter l'asymétrie de la relation (pour les variables positives). Il apparaît donc que la duration et la volatilité augmentent lorsque l'offre la moins importante (en prix pour la duration et en activité pour la volatilité) est du côté vente et diminue lorsqu'elle est observée du côté achat. Toutefois la relation entre rendement et direction prouve qu'une moindre activité anticipe une transaction de l'autre côté du carnet.

À l'exception du signe de la relation entre rendement et direction, les autres valeurs sont cohérentes avec celles que l'on peut attendre entre des approximations du flux informationnel (asymétrie, hétérogénéité, intérêt global) et les caractéristiques des transactions.

Le cadre d'analyse simplificateur que nous avons employé jusqu'ici a mis en évidence des relations significatives et intéressantes entre les variables informationnelles et les caractéristiques des transactions. Ces résultats confirment à la fois la pertinence des variables ainsi que les relations observées dans le chapitre précédent.

Étant donné que nous avons vu dans le précédent chapitre qu'une dynamique de type GARCH modélise de manière adéquate les processus de duration, de volume et de volatilité, on peut légitimement se demander quel est l'effet de l'inclusion des variables informationnelles dans l'équation de la volatilité conditionnelle. Formellement, nous reprenons dans cet ordre les modèles (5.6), (5.4) et (5.5) auxquels nous intégrons les variables informationnelles. Les modèles évalués sont donc donnés par :

$$\begin{aligned}
\sigma_t^2 &= \omega + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 + \sum_{j=1}^{q} \alpha_j \times r_{t-j} \\
&\quad + \gamma_1 H_1 + \gamma_2 H_2 + \gamma_3 D_1 + \gamma_4 D_2 + \gamma_5 P_1 + \gamma_6 P_2 \\
\psi_t &= \omega + \sum_{j=1}^{q} \alpha_j \times d_{t-j} + \sum_{j=1}^{p} \beta_j \psi_{t-j} \\
&\quad + \gamma_1 H_1 + \gamma_2 H_2 + \gamma_3 D_1 + \gamma_4 D_2 + \gamma_5 P_1 + \gamma_6 P_2 \\
\phi_t &= \omega + \sum_{j=1}^{q} \alpha_j \times v_{t-j} + \sum_{j=1}^{p} \beta_j \phi_{t-j} \\
&\quad + \gamma_1 H_1 + \gamma_2 H_2 + \gamma_3 D_1 + \gamma_4 D_2 + \gamma_5 P_1 + \gamma_6 P_2
\end{aligned} \tag{6.2}$$

où, les variables ont la même signification que dans le chapitre précédent, $H_\cdot$, $D_\cdot$ et $P_\cdot$ désignent les variables informationnelles correspondantes où le numéro est celui de la composante principale (se reporter aux tables 6.2 à 6.4 pour l'interprétation des axes). Le choix de l'ordre de la modélisation et l'estimation sont faites suivant les procédures déjà décrites dans le précédent

chapitre. Les tables 6.10 à 6.12 fournissent les estimations pour la volatilité, la duration et le volume.

Volatilité

La relation observée lorsqu'on inclut les variables informationnelles du carnet dans l'équation de la volatilité conditionnelle des rendements est plus complexe que celle mise en évidence dans les précédentes régressions. On voit que dans la majorité des cas, les variables du carnet ont une influence très significative. L'inclusion des variables du carnet fait chuter de manière importante la persistance de la volatilité. Cette réduction affecte différemment le paramètre β_1 qui diminue substantiellement alors que α_1 augmente quelque peu. La dynamique de la volatilité se rapproche donc d'un ARCH.

Parmi les variables informationnelles, seules trois ont le même signe et une significativité élevée pour tous les titres : la direction prix (-), l'hétérogénéité prix (+ excepté Schneider) et la profondeur prix (-). Ainsi, toutes les variables fondées sur les fourchettes ont une influence significative et stable sur la volatilité. Les mesures d'hétérogénéité et de direction volume confirment globalement les relations déjà observées sur les mesures utilisant la fourchette. La mesure de profondeur activité a une influence de sens opposé à celle de la profondeur prix peut-être en raison de l'effet de substitution qui existe entre ces deux formes de la profondeur.

On pouvait attendre une relation négative entre la profondeur prix et la volatilité. Les relations entre hétérogénéité, direction et volatilité soulignent que la volatilité est une fonction croissante de l'hétérogénéité[7] alors que l'influence additionnelle de la direction prouve que la réaction de la volatilité est plus forte lorsque l'indication directionnelle est du côté vente (i.e. lorsque le côté vente est le moins pourvu).

Duration

La baisse de la persistance est beaucoup plus faible pour la duration. En effet, β_1 a toujours une valeur élevée entre $0,82$ et $0,95$ alors que α_1 demeurent inférieur à $0,10$. On peut remarquer que, pour tous les titres on a $\alpha_1 + \beta_1 < 1$.

Les seules variables informationnelles significatives et de mêmes signes pour tous les titres sont la profondeur prix (-) et la direction prix (+ exception faite de Schneider). Le mécanisme de fourniture et de consommation de liquidité propre aux marchés dirigés par les ordres influence donc tous les titres. Lorsque la profondeur prix est élevée, l'offre de liquidité est peu rémunérée ce qui décourage les nouvelles introductions en carnet et encourage les transactions, d'où l'effet négatif sur la duration (augmentation de la fréquence de transaction). Vu le signe de la direction prix, à hétérogénéité donnée, l'augmentation de la duration est plus importante si le signe de l'indication directionnelle est négatif. Il semble donc que les investisseurs soient plus pressés

[7] On note une contradiction avec le précédent modèle linéaire où on avait observé une relation négative. Il apparaît donc que la manière dont on modélise le processus modifie le signe des relations. Il nous est difficile de déterminer lequel de ces deux résultats reflète le mieux les perceptions des agents.

TABLE 6.10
GARCH et variables informationnelles

	St-Gobain	Schneider	Sanofi	Thomson	Accor	Lyonnaise	Lagardère
α_0	1,0111***	0,3025***	1,8357***	0,1964***	0,7953***	1,2181***	0,2394***
	(120,01)	(151,24)	(212,03)	(75,41)	(104,18)	(111,29)	(140,69)
α_1	0,1598***	0,1444***	0,3266***	0,2009***	0,1861***	0,1401***	0,2253***
	74,80	114,20	71,85	91,28	78,62	76,05	148,62
β_1	0,5575***	0,7980***	0,2008***	0,7842***	0,5996***	0,5358***	0,7596***
	(182,37)	(828,51)	(92,76)	(497,77)	(195,90)	(146,74)	(783,81)
Het_1	0,0447***	-0,0388***	0,0058***	0,0083***	0,0149***	0,0118***	0,0113***
	(40,42)	(-53,94)	(3,82)	(12,72)	(9,92)	(10,84)	(17,93)
Het_2	-0,0295***	0,0166***	0,0180***	-0,0105***	0,0072***	0,0336***	-0,0104***
	(-47,17)	(30,84)	(33,93)	(-10,46)	(4,89)	(34,92)	(-38,04)
Dir_1	-0,0970***	-0,0889***	-0,0292***	-0,0145***	-0,1178***	-0,0810***	-0,0394***
	(-35,02)	(-87,27)	(-12,24)	(-27,46)	(-51,09)	(-30,54)	(-48,96)
Dir_2	0,0013*	-0,0222***	0,0160***	-0,0153***	0,0289***	-0,0337***	-0,0091***
	(1,77)	(-26,56)	(16,87)	(-7,86)	(15,41)	(-30,67)	(-28,51)
Pro_1	-0,1598***	-0,0124***	-0,2646***	-0,0184***	-0,0760***	-0,2185***	-0,0228***
	(-84,87)	(-88,57)	(-172,81)	(-66,23)	(-61,15)	(-84,47)	(-83,42)
Pro_2	0,0072***	-0,0207***	0,0406***	0,0163***	0,0121***	0,0184***	-0,0138***
	(7,59)	(-61,93)	(55,42)	(12,01)	(8,22)	(23,47)	(-39,98)
Jarque-Béra	164173	492111	163125	281989	194591	367698	1303545
Ljung Box (20)	270***	171***	17	217***	345***	154***	172***
Ljung Box^2 (20)	2	416***	0	44***	68***	16	3

Cette table fournit les estimations de la première équation du système (6.2). Soit, l'effet que les variables du carnet ont sur la dynamique de la volatilité. Un paramètre ou une statistique significatif au seuil de 1% est marqué par ***, au seuil de 5% par ** et au seuil de 10% par *.

TABLE 6.11
ACD et variables informationnelles

	St-Gobain	Schneider	Sanofi	Thomson	Accor	Lyonnaise	Lagardère
α_0	0,0066***	0,0105***	0,0739***	0,0150***	0,0189***	0,0238***	0,0118***
	(7,01)	(6,55)	(4,88)	(3,97)	(5,59)	(4,75)	(5,96)
α_1	0,0438***	0,0596***	0,1016***	0,0722***	0,0599***	0,0487***	0,0559***
	14,62	13,59	10,58	8,78	11,54	9,29	12,38
β_1	0,9502***	0,9308***	0,8286***	0,9156***	0,9224***	0,9284***	0,9333***
	(256,01)	(164,04)	(35,24)	(81,60)	(112,61)	(93,40)	(151,74)
Het_1	0,0000	0,0004*	0,0020**	0,0006	0,0010***	0,0010***	0,0008***
	(0,03)	(1,72)	(2,12)	(1,60)	(3,43)	(3,99)	(2,97)
Het_2	-0,0001	0,0006**	-0,0029***	0,0010*	-0,0006	0,0001	0,0000
	(-0,65)	(2,03)	(-2,66)	(1,75)	(-1,30)	(0,16)	(0,10)
Dir_1	0,0012***	0,0004	0,0058***	0,0024***	0,0021***	0,0024***	0,0015***
	(4,12)	(1,49)	(4,71)	(3,56)	(4,08)	(5,86)	(4,32)
Dir_2	0,0003	0,0006	0,0023	-0,0001	0,0013**	0,0016***	0,0001
	(0,99)	(1,38)	(1,50)	(-0,13)	(2,04)	(3,18)	(0,19)
Pro_1	-0,0015***	-0,0006***	-0,0084***	-0,0025***	-0,0026***	-0,0027***	-0,0014***
	(-6,90)	(-3,88)	(-4,83)	(-4,62)	(-5,69)	(-5,26)	(-5,54)
Pro_2	-0,0010***	-0,0009***	0,0012	-0,0004	-0,0005	0,0004	-0,0006
	(-4,60)	(-4,63)	(0,99)	(-1,28)	(-1,50)	(1,27)	(-1,34)
Jarque-Béra	23968	22579	16616	11023	20961	24640	17513
Ljung Box (20)	10685***	7337***	135***	3258***	1647***	609***	3570***
Ljung Box^2 (20)	7275***	4284***	89***	2233***	668***	218***	2050***

Cette table fournit les estimations de la deuxième équation du système (6.2) ; l'effet des variables du carnet sur la dynamique de la duration. Un paramètre ou une statistique significatif au seuil de 1% est marqué par ***, au seuil de 5% par ** et au seuil de 10% par *.

de traiter quand l'offre est moins importante du côté vente.

Concernant les autres variables, on peut noter un accord entre les résultats des différents titres. En effet, s'il arrive que les variables ne soient pas significatives, en revanche, toutes celles qui sont significatives ont le même signe.

Les variables fondées sur l'activité ont globalement peu d'influence : au seuil de 5% l'hétérogénéité n'est significative qu'une fois (-), la direction deux fois (+) et la profondeur deux fois (-). Il semble donc que la duration incorpore presque exclusivement l'aspect fourchette des différentes dimensions alors que l'aspect activité est peu significatif.

Volume

On observe pour le volume une dynamique (paramètres α_1 et β_1) semblable à celle déjà observée pour la duration : la persistance demeure élevée surtout du fait du paramètre β_1.

Le rôle des variables informationnelles est bien distinct de ceux que nous avons déjà observé. L'influence la plus notable est celle de la profondeur activité qui dans 6 cas sur 7 est significative et a une influence positive. La taille de transaction et la profondeur disponible évoluent de concert. Les autres variables sont globalement très peu significatives.

Les mesures d'hétérogénéité n'ont pas d'influence sur le volume. Il semble donc qu'au niveau trade by trade et après prise en compte de la profondeur dans un modèle de type GARCH, le volume de la transaction ne soit plus une fonction croissante de l'hétérogénéité alors qu'il l'était dans le modèle linéaire précédent.

Les mesures de direction on une influence plus forte, positive lorsqu'elle est significative (3/7). Il semble donc que le volume soit, pour certains titres, sensible à l'existence d'un déséquilibre de l'offre en prix ou en activité. Une moindre offre du coté vente se traduit alors par une diminution du volume.

L'inclusion des variables informationnelles dans l'équation de la volatilité conditionnelle (modèles de type GARCH) des différentes caractéristiques des transactions donne des résultats plus complexes que le modèle simple et sont même parfois en contradiction avec ce dernier. En effet, toutes les variables sont significatives pour la volatilité des rendements. Il apparaît que la profondeur (au moins une des dimensions) a un impact sur toutes les caractéristiques.

D'une manière générale la profondeur augmente l'activité du marché (augmentation du volume et diminution de la duration). L'effet de substitution entre profondeur fourchette et activité induit une influence négative (positive) de la profondeur prix (profondeur activité) sur la volatilité. L'hétérogénéité augmente la duration et la volatilité mais n'a pas d'influence sur le volume. Enfin, la direction est souvent significative et a une influence négative (direction prix) sur la duration et la volatilité ce qui souligne l'asymétrie de la réponse de ces deux variables en fonction du sens du choc.

Il ressort de la comparaison des deux modélisations appliquées dans cette section que les

TABLE 6.12
ACV et variables informationnelles

	St-Gobain	Schneider	Sanofi	Thomson	Accor	Lyonnaise	Lagardère
α_0	0,0321**	0,0317***	0,0429***	0,0440***	0,0261***	0,9319***	0,0678**
	(2,21)	(3,57)	(3,30)	(2,88)	(3,57)	(3,05)	(2,33)
α_1	0,0397***	0,0442***	0,0509***	0,0559***	0,0399***	0,0769**	0,0600***
	4,98	7,21	7,58	6,20	7,54	2,11	5,33
β_1	0,9289***	0,9248***	0,9070***	0,9018***	0,9344***	-0,0000	0,8734***
	(42,10)	(63,52)	(48,35)	(39,07)	(76,65)	(-0,00)	(22,24)
Het_1	0,0006	0,0003	0,0000	0,0010	-0,0008	0,0067	0,0001
	(1,31)	(0,51)	(0,04)	(1,25)	(-1,40)	(1,24)	(0,10)
Het_2	0,0010	0,0020***	0,0012	-0,0004	0,0004	-0,0014	0,0001
	(1,29)	(2,89)	(1,45)	(-0,31)	(0,57)	(-0,20)	(0,11)
Dir_1	0,0008	0,0006	-0,0004	0,0036***	0,0008	0,0126***	0,0026**
	(1,34)	(0,96)	(-0,52)	(3,17)	(1,33)	(6,06)	(2,55)
Dir_2	0,0013	0,0007	0,0046***	0,0056**	0,0021*	-0,0146	0,0090***
	(1,02)	(0,66)	(2,90)	(2,20)	(1,77)	(-1,06)	(2,69)
Pro_1	0,0005	-0,0005	0,0007	0,0011	0,0016***	-0,0744**	0,0006
	(1,13)	(-1,09)	(1,40)	(1,53)	(3,22)	(-2,09)	(1,04)
Pro_2	0,0081**	0,0059***	0,0075***	0,0104**	0,0071***	0,0581***	0,0317**
	(2,25)	(2,90)	(3,00)	(2,34)	(3,25)	(5,44)	(2,30)
Jarque-Béra	64339	81865	20143	18331	50355	64953	44028
Ljung Box (20)	251***	199***	52***	134***	397***	2001***	309***
Ljung Box^2 (20)	70***	85***	104***	58***	25	775***	69***

Cette table fournit les estimations de la troisième équation du système (6.2) : l'effet des variables du carnet sur la dynamique de la duration. Un paramètre ou une statistique significatif au seuil de 1% est marqué par ***, au seuil de 5% par ** et au seuil de 10% par *.

variables du carnet ont un effet aussi bien sur la moyenne des processus (modèle linéaire) que sur la volatilité conditionnelle des processus (modélisations GARCH). La première modélisation semble plus en accord avec les hypothèses de départ que les résultats des modélisations GARCH. On observe que les variables informationnelles ont un effet aussi bien sur la moyenne que sur la volatilité conditionnelle des caractéristiques des transactions. De plus, le type de modélisation employée altère les relations mises en évidence. Celles-ci demeurent globalement cohérentes mais certaines variables peuvent ne plus être significatives dans le modèle GARCH.

L'état du carnet semble donc appartenir à l'ensemble informationnel utilisé par les investisseurs. Les principales relations mises en évidence n'apportent de soutien inconditionnel à aucun modèle. Toutes les variables traduisant une incertitude(l'hétérogénéité et la direction) augmentent la duration alors que le volume semble insensible à ces deux facteurs. La principale conclusion est donc l'impact primordial de la profondeur sur les caractéristiques des transactions (duration (-), volume (+) profondeur activité et volatilité (-) profondeur de la fourchette (+) profondeur activité.

Les précédentes modélisations ne tiennent pas compte des interactions possibles entre les variables. La prochaine section essaiera de lever cette limite en adaptant le modèle de Manganelli (2002) pour y intégrer les variables du carnet. Toutefois, il convient auparavant de vérifier la pertinence d'une modélisation de type GARCH pour les variables informationnelles. Ce point est traité dans l'annexe G.3 qui présente les résultats des estimations de modèles GARCH sur l'ensemble de ces variables et suggère que ces modèles sont acceptables pour les variables du carnet.

6.3 Modélisation multivariée

Dans le précédent chapitre, le modèle de Manganelli (2002) était employé afin d'étudier la relation dynamique entre les caractéristiques des transactions. Nous étendons ici cette modélisation afin d'inclure les variables informationnelles. Par abus de notation, les variables informationnelles sont désignées i_t même si elles représentent en fait plusieurs variables différentes (deux pour chaque dimension). Soit le processus des données $(i_t, d_t, v_t, r_t) \sim f(i_t, d_t, v_t, r_t|\Omega, \theta)$

La structure de causalité supposée est très proche de celle déjà utilisée. On considère que les variables informationnelles sont déterminées simultanément avant les variables de transactions puis participent aux différentes caractéristiques des échanges. La densité est donc décomposée de la manière suivante :

$(i_t, d_t, v_t, r_t) = f(i_t|\Omega_t, \theta_i)g(d_t|i_t, \Omega_t, \theta_d)h(v_t|d_t, i_t, \Omega_t, \theta_v)k(r_t|v_t, d_t, i_t, \Omega_t, \theta_r)$

Cette extension du modèle ne modifie aucune des propriétés déjà employées dans le précédent chapitre. Étant donnée la proximité entre la méthodologie employée ici et celle détaillée dans le chapitre précédent, nous analysons directement les fonctions de réponse entre les va-

riables. La sous-section 6.3.1 analysera l'effet des caractéristiques des transactions sur les variables informationnelles puis la sous-section 6.3.2 l'effet des toutes les variables sur les caractéristiques des transactions. Enfin, la sous-section 6.3.3 présentera une synthèse des résultats pour les différents titres.

6.3.1 Effets des transactions sur l'état du carnet

Les figures 6.1 à 6.3 donnent l'effet des caractéristiques des transactions sur chacune des dimensions du flux informationnel. D'une manière générale, les effets des transactions sur les variables du carnet sont de courte durée (au maximum 5 transactions).

Hétérogénéité

L'influence des caractéristiques des transactions sur l'hétérogénéité est trouble. La duration et le volume augmentent l'hétérogénéité prix[8] alors que la volatilité la diminue. Les influences sur l'hétérogénéité activité sont moins significatives : la duration n'est plus significative, le volume a un effet immédiat mais très court et l'influence de la volatilité est très occasionnelle. Cela confirme que la direction du signal est en fait le côté où l'offre est la plus importante. En effet, le volume augmente l'hétérogénéité, c'est à dire qu'il renforce la symétrie de l'offre. Cela révèle que, comme nous l'avions déjà noté, la transaction se produit majoritairement du côté où l'offre est la plus importante et diminue donc la différence entre l'offre des deux côtés.

Direction

L'influence des caractéristiques des transactions sur l'indication directionnelle est peu significative. Seule la direction prix est significativement et positivement affectée par la duration et la volatilité. On peut de plus noter que l'effet de la volatilité sur la direction prix persiste pendant au moins les 12 transactions suivantes. Donc, il semble qu'une augmentation de la duration ou qu'une volatilité élevée entraîne un déséquilibre durable de la fourchette.

Profondeur

Les relations mises en évidence avec la profondeur semblent fortes et tranchées. L'influence de la volatilité traduit la nature compensatoire de la liquidité prix et activité. Cette variable a une influence négative sur la profondeur de la fourchette et positive sur la profondeur activité. Le volume a un effet négatif sur la profondeur de la fourchette. La faiblesse de l'effet du volume sur la profondeur activité peut s'expliquer par un phénomène de glissement des limites[9] La duration a, quant à elle, une influence positive sur les deux dimensions de la profondeur. Il semble donc qu'une augmentation de la duration induit une augmentation de la profondeur fourchette et de la profondeur prix.

[8] On voit sur le graphique une influence positive et significative de la duration et du volume sur le Het_1.

[9] Une limite entièrement consommée lors d'une transaction est remplacée, dans l'état du carnet postérieur, par la limite suivante.

FIGURE 6.1
Effets sur l'hétérogénéité
Sanofi

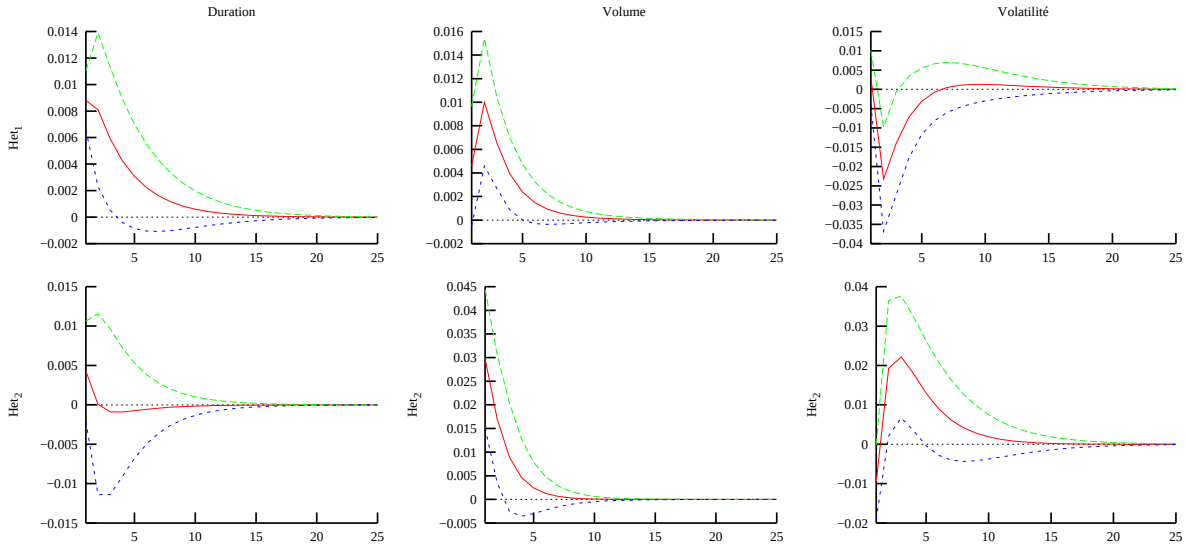
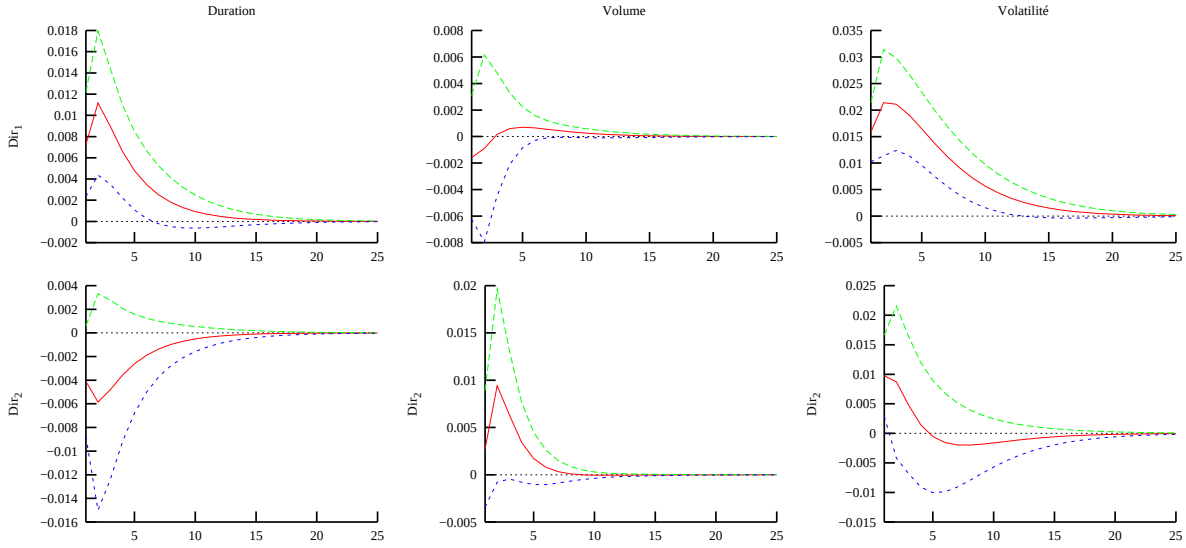
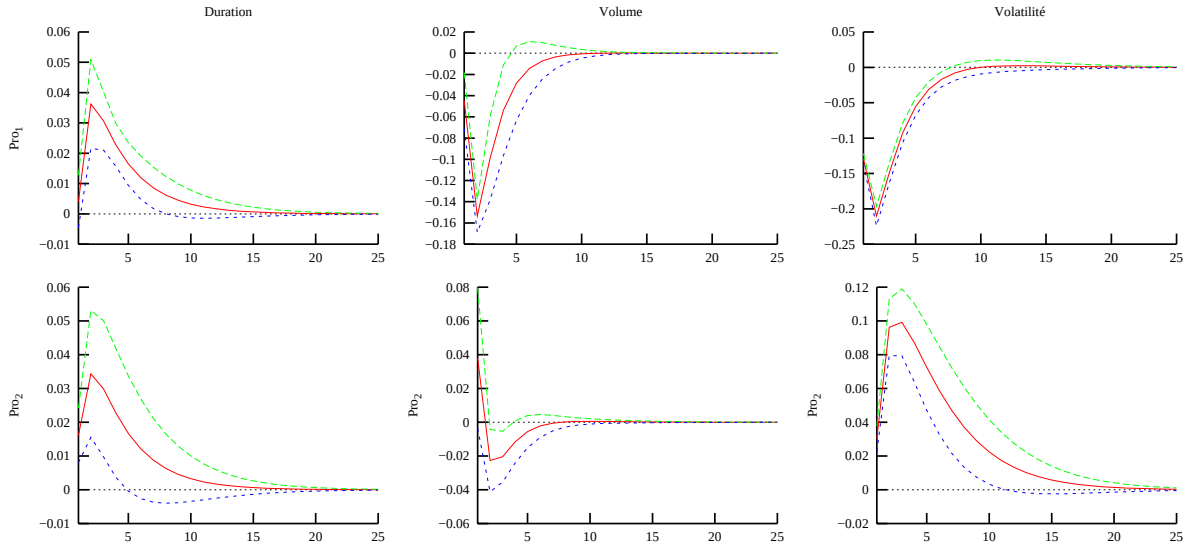


FIGURE 6.2
Effets sur la direction
Sanofi



Ces figures reflètent les effets des caractéristiques des transactions (en haut des colonnes) ont sur l'hétérogénéité et la direction (à gauche de chaque ligne). La courbe rouge continue donne la réponse moyenne alors que les courbes hachurées délimitent un intervalle de confiance à 95%.

FIGURE 6.3
Effets sur la profondeur
Sanofi



Cette figure reflète l'effet que les caractéristiques des transactions (en haut des colonnes) ont sur la profondeur (à gauche de chaque ligne). La courbe rouge continue donne la réponse moyenne alors que les courbes hachurées délimitent un intervalle de confiance à 95%.

Les effets du volume et de la volatilité étaient en grande partie attendus. Il semble logique qu'une transaction ayant un volume important et/ou une volatilité élevée induise un élargissement de la fourchette (consommation de certaines limites en carnet). L'effet de la duration montre qu'un temps sans transaction incite les agents à introduire de nouveaux ordres en carnet qui améliorent le prix proposé et les quantités offertes.

De manière intéressante, l'effet de la volatilité sur la profondeur activité est plus long que celui sur la profondeur fourchette. Il semble donc que la concurrence entre les agents les incite à introduire des ordres améliorant le prix. Un tel comportement pourrait avoir pour but de capter la rémunération associée à la fourniture de liquidité.

En guise de première conclusion, on voit donc que la duration a pour effet global d'augmenter l'aspect prix de l'hétérogénéité, de la direction et les deux aspects de la profondeur. Il est donc difficile de conclure à l'existence d'une association entre duration et information.

Le volume a une influence positive sur l'hétérogénéité, pas d'influence sur la direction et une influence négative sur la profondeur de la fourchette. Il semble donc que les volumes importants ne soient pas le fait d'informés mais plutôt dûs à une différence d'opinion.

Enfin, la volatilité a une influence négative sur l'hétérogénéité prix et la profondeur prix alors qu'elle augmente l'hétérogénéité activité, la direction prix et la profondeur activité. Ces résultats suggèrent que la première force gouvernant la dynamique de fourniture de liquidité

n'est pas l'asymétrie d'information mais plutôt une certaine hétérogénéité des agents.

6.3.2 Effets sur les caractéristiques des transactions

Les figures 6.4 à 6.6 illustrent dans le cas de Sanofi les fonctions de réponse des caractéristiques des transactions.

D'une manière générale, l'inclusion des variables informationnelles dans la modélisation de Manganelli (2002) entraîne une très nette réduction de la persistance des chocs. Dans le modèle ne prenant en compte que les caractéristiques des transactions, on mettait couramment en évidence des effets encore significatifs après plus de 25 transactions, en comparaison lorsque nous tenons compte de l'information apportée par le carnet, la persistance maximale ne dépasse jamais 15 transactions.

Duration

L'inclusion des variables du carnet entraîne une forte chute de la persistance des chocs, les autres caractéristiques des transactions (la volatilité et le volume) conservent toute de même une influence positive et significative pendant les 6 à 12 transactions qui suivent.

Seules la direction activité (D_2) et la profondeur prix (P_1) ont une influence significative et négative. Cette dernière est de courte durée puisqu'elle atteint au maximum la 5^e transaction.

Volume

On retrouve pour le volume le même type d'effets que ceux déjà notés pour la duration : une réduction du temps de retour à l'équilibre. L'influence positive de la volatilité sur le volume n'avait pas été mise en évidence dans le précédent chapitre. On voit donc à nouveau que la modélisation influence les relations mises en évidence. L'inclusion des variables du carnet modifie donc les relations perçues entre les caractéristiques des transactions.

Comme pour la duration seules (D_2) et (P_1) ont une influence significative et négative. Donc, une forte indication directionnelle activité ou une forte profondeur prix entraînent une réduction du volume de transaction.

Volatilité

L'effet le plus manifeste concerne la volatilité. En effet, les autres caractéristiques des transactions (la duration et le volume) n'ont plus, à la différence du précédent chapitre, d'effet significatif. Il semble donc que l'influence que nous avions noté traduisaient le fait que ces caractéristiques des transactions reflétaient certaines dimensions de l'état du carnet.

Une fois de plus, seules deux variables du carnet ont une influence significative : la direction activité (D_2) qui a un impact positif et la profondeur prix (P_1) qui a un impact négatif. La direction prix (D_1) a une influence positive mais celle-ci n'est pas significative.

L'analyse de la fonction de réponse de Sanofi fait ressortir plusieurs faits saillants :

FIGURE 6.4
Effets sur la duration
Sanofi

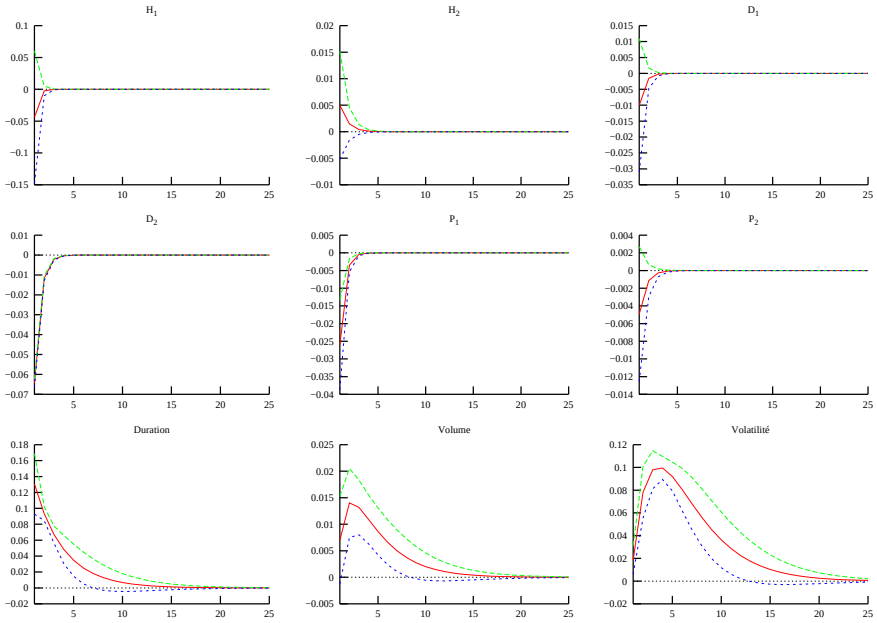
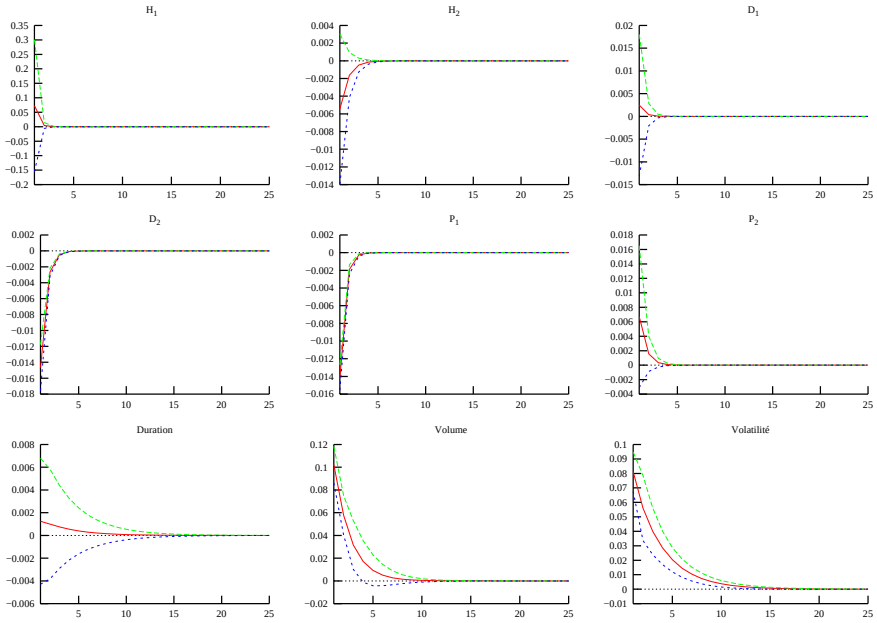
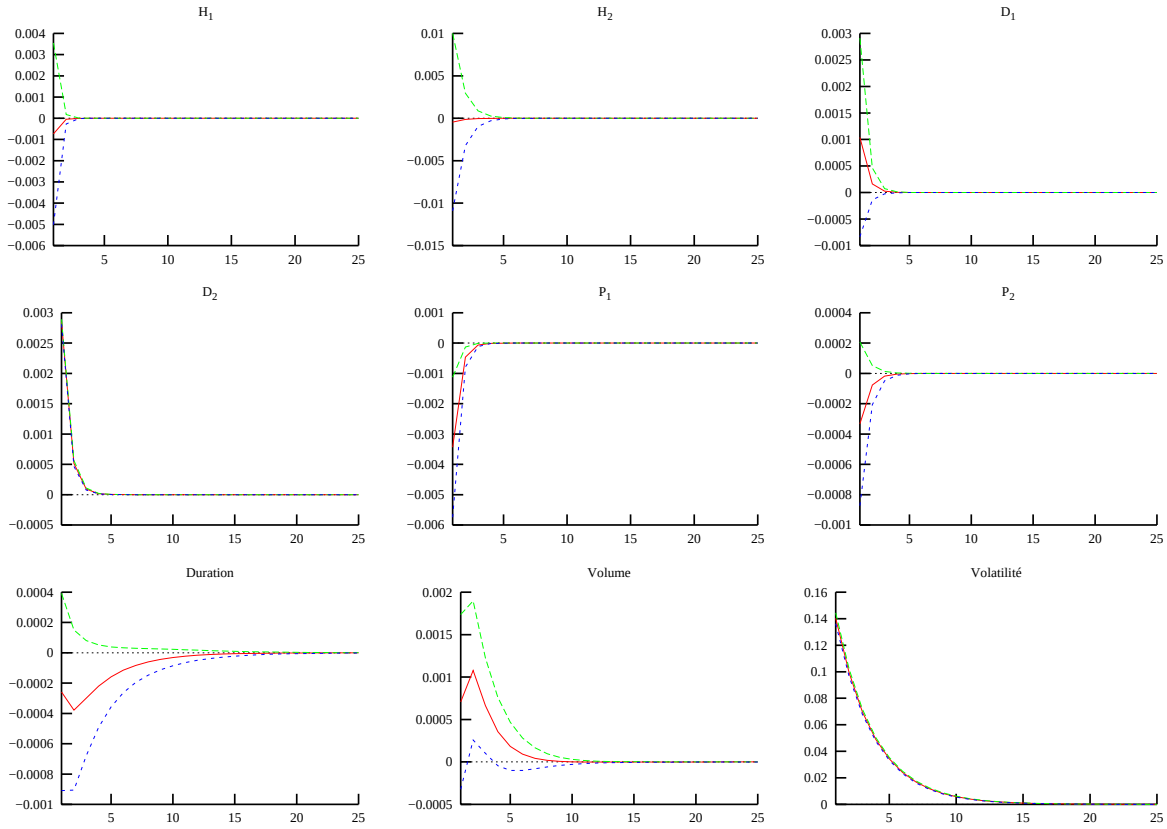


FIGURE 6.5
Effets sur le volume
Sanofi



Ces figures illustrent l'influence des différentes variables (en titre des graphiques) sur la duration et le volume. La courbe rouge continue donne la réponse moyenne alors que les courbes hachurées délimitent un intervalle de confiance à 95%.

FIGURE 6.6
Effets sur la volatilité
Sanofi



Cette figure illustre l'influence des différentes variables (en titre des graphiques) sur la volatilité. La courbe rouge continue donne la réponse moyenne alors que les courbes hachurées délimitent un intervalle de confiance à 95%.

i.) L'inclusion des variables informationnelles entraîne une chute importante de la persistance des chocs.

ii.) Seules les variables de direction activité et de profondeur prix se dégagent par leur significativité, leurs effets changent selon les variables. La profondeur prix a une influence négative sur toutes les variables (volatilité, duration, et volume). La direction activité à une influence positive sur la volatilité et négative sur la duration et le volume.

iii.) L'inclusion des variables informationnelles modifie les relations entre les caractéristiques des transactions. Les principales modifications sont : - L'influence de la duration et du volume sur la volatilité n'est plus significative ; - La volatilité a désormais une influence significative et positive sur le volume.

iv.) Les principales variables significatives du carnet sont la direction activité et la profondeur prix.

6.3.3 Synthèse des différents titres

Après avoir détaillé l'analyse du titre Sanofi, nous synthétisons ici les résultats obtenus pour les différents titres. D'une manière générale, tous les titres confirment que l'inclusion des variables du carnet produit de fortes modifications par rapport aux résultats du précédent chapitre. On doit toutefois remarquer que l'analyse des effets des variables du carnet est loin d'être aisée. En effet, elles ne sont que rarement significatives, la durée de l'effet est souvent faible (au maximum 5 transactions) et le signe de l'influence varie.

Variables du carnet

Les résultats obtenus concernant l'effet des transactions sur l'état du carnet sont globalement convergents et fournissent une image assez claire.

Le premier axe de la profondeur est influencé positivement par la duration et négativement par le volume ou la volatilité. Une transaction d'un volume important ou d'une volatilité élevée induit une réduction de la profondeur (fourchette majoritairement). Le second axe de la profondeur a des liens plus flous toutefois une influence instable mais globalement positive de la volatilité se dégage. Il semble donc qu'une forte volatilité induise une augmentation de la profondeur activité. Cette relation est liée avec fait que la volatilité transitoire stimule la fourniture de liquidité.

Les transactions influencent également la manière dont l'offre se répartit. Ainsi il apparaît que la volatilité a un effet négatif sur le premier axe de l'hétérogénéité, alors que le volume ou la duration ont des effets moins stables mais globalement positifs. L'occurrence de grandes durations ou de transactions importantes a donc pour effet de rendre l'offre en carnet plus symétrique alors qu'une forte volatilité déséquilibre l'offre en carnet. Il est beaucoup plus problématique d'interpréter les liens entre les transactions et le second axe de l'hétérogénéité.

L'influence des transactions sur la direction est, de toutes les variables du carnet, la plus difficile à analyser. Aucune des caractéristiques des transactions n'a d'effet notable, tout juste peut on signaler que la volatilité augmente l'indication directionnelle.

Duration

Pour la duration , nous obtenons des résultats très contrastés entre les différents titres, à tel point qu'un classement en groupe semble pertinent :

- La duration, le volume et la volatilité ont une influence positive et significative sur la duration (Sanofi, Thomson).
- La duration et la volatilité ont une influence positive et significative sur la duration (Accor, Lyonnaise, Schneider, Lagardère).

D'autres parallélismes existent entre les titres. Par exemple, pour Accor et St-Gobain l'influence de la volatilité est dans un premier temps négative puis devient positive et dure au moins jusqu'à la 25^{e} transaction. On voit donc que l'inclusion des variables du carnet modifie singulièrement la complexité des relations entre les caractéristiques des transactions en mettant à jour des variations qui semblent autant quantitatives que qualitatives entre les dynamiques des titres.

La profondeur a majoritairement une influence négative et significative sur la duration (seuls Thomson et Lagardère font exception). La direction est elle aussi souvent significative et a également une influence négative sur la duration. Enfin, c'est pour la duration qu'on observe le plus de cas de significativité de l'hétérogénéité. Cette dernière est significative pour 4 titres et 3 fois négative (principalement le second axe).

Volume

Pour le volume on observe toujours une très forte influence positive des volumes et des volatilités passées. L'influence de la duration est incertaine puisque seuls Accor et St-Gobain mettent en évidence une effet significatif (-) de la duration sur le volume.

L'effet de la profondeur est important puisque pour 4 titres le premier axe de la profondeur a une influence négative et de faible durée (au maximum 5 transactions). Il semble donc que le volume de transactions soit plus sensible à la profondeur de la fourchette (aux coûts de transaction) qu'à la profondeur cotée. L'influence de l'hétérogénéité est faible, elle n'est significative que pour Lyonnaise. L'influence de la direction est incertaine mais présente une certaine régularité. Pour aucun titre les 2 axes ne sont significatifs en même temps. Les deux axes sont significatifs chacun pour 2 titres et le 1^{er} axe a une influence positive alors que le second a une influence négative.

Volatilité

On retrouve le fait déjà marquant pour Sanofi : la duration et le volume n'ont plus d'effets significatifs sur la volatilité. Seul Thomson met encore en évidence une influence positive et significative du volume sur la volatilité. Il semble donc qu'une partie du contenu informationnel des caractéristiques des transactions provient du fait qu'elles reflètent les conditions en carnet.

La relation la plus saillante est l'influence négative de la profondeur (principalement le 1^{er} axe). Il est difficile de conclure quant aux autres variables. L'hétérogénéité n'est significative que dans un cas. La direction est plus souvent significative (pour 3 titres Sanofi, Accor et St-Gobain) mais l'axe significatif ou le signe des influences changent parfois. Cette variabilité des résultats nous amène à conclure que l'asymétrie de la réaction de la volatilité au choc est variable selon les titres.

L'image qui émerge de cette analyse est que le carnet a une place centrale dans la diffusion et l'incorporation de l'information.

Il semble que bien plus que la duration ou le volume de la transaction, c'est le choix d'un agent de traiter alors que la profondeur est faible qui induit une hausse de la volatilité. Cette hausse a pour conséquence d'augmenter l'espérance la duration et du volume. Cela favorise alors reconstitution de l'offre après un choc de liquidité. On peut à partir de là comprendre que c'est le fait qu'un agent persiste à vouloir traiter avant que la profondeur n'ai eu le temps de se reconstituer qui induit les périodes de forts déséquilibres.

La duration n'a donc pas de contenu propre mais ses faibles valeurs capturent les cas où les transactions se succèdent si vite que la liquidité du carnet n'a pas le temps de se reconstituer.

6.4 Conclusion

Dans ce chapitre nous avons analysé l'influence de l'état du carnet d'ordres sur les transactions. Après avoir construit par analyse en composantes principales des variables reflétant les caractéristiques du carnet, nous avons étudié leurs interactions avec les caractéristiques du carnet.

Dans un premier temps, nous avons étudié la capacité des variables du carnet à expliquer les caractéristiques des transactions en régressant ces dernières sur les variables du carnet. Ce test d'une relation linéaire entre les approximations du carnet et les caractéristiques des transactions s'est révélé concluant. Il a mis en évidence la forte puissance explicative des variables du carnet mais surtout des valeurs compatibles avec les relations théoriques.

Dans un second temps, nous avons inclus les variables du carnet dans l'équation de la volatilité conditionnelle (modèles de type GARCH) des caractéristiques des transactions. Cette modélisation a conduit à des résultats différents du précédent modèle. Ainsi, il apparaît dans ce cadre de modélisation que la duration est surtout sensible à la dimension prix du carnet et qu'elle est une fonction croissante de l'hétérogénéité et de la direction et décroissante de la profondeur. Il semble donc que toute incertitude induise une augmentation de la duration. Le volume n'est plus sensible qu'à la profondeur activité. Enfin, les relations entre la volatilité et le carnet sont très complexes. La volatilité est globalement une fonction croissante de l'hé-

térogénéité, décroissante de la direction et négativement (positivement) liée à la profondeur fourchette (activité). On voit donc émerger l'image d'une volatilité asymétrique dont la hausse est plus faible lorsqu'une moindre liquidité est observée du côté vente, ce qui se traduit alors par des transactions à l'achat.

Dans une troisième et dernière étape, nous avons étendu la modélisation trivariée employée dans le chapitre précédent afin d'inclure les variables du carnet. Cette modélisation est dynamique et prend donc en compte les interactions entre les variables. Nous avons été obligés de constater une grande disparité entre les titres. Si, dans tous les cas, l'inclusion des variables du carnet modifie les relations, nous obtenons cependant des résultats difficiles à synthétiser.

La première et la plus importante des conclusions à tirer est que l'examen de l'ensemble des transactions et l'ensemble des états du carnet permettent de donner une image plus fine mais plus complexe. De manière évidente, l'analyse exclusive des transactions ne fournit qu'une image partielle. En effet, l'un des résultats les plus manifestes est que l'influence de la duration et du volume sur la volatilité n'est plus significative lorsqu'on prend en compte le carnet d'ordres. On doit donc conclure que ces deux variables rendaient une information disponible dans le carnet.

À la différence des caractéristiques des transactions dont la significativité présente une certaine régularité, les variables du carnet ont, la plupart du temps, des influences instables en terme de significativité ou de signe. Au total, la deuxième conclusion importante est donc de souligner une grande diversité de dynamique de fourniture de liquidité et/ou d'apprentissage selon les titres. Les conclusions plus précises doivent être prises avec précaution au regard de la diversité des résultats obtenus.

L'interprétation exacte de l'information véhiculée par la duration et le volume est assez complexe. Les interactions entre ces deux variables sont rarement significatives et penchent globalement pour une influence positive du volume sur duration. La volatilité a un effet positif sur le volume et la duration.

L'influence des variables du carnet est globalement peu significative et dans tous les cas de courte durée. Parmi les variables du carnet étudiées, la profondeur est celle qui semble avoir la plus forte influence : - négative sur la volatilité ; - négative sur le volume (profondeur fourchette) ; - négative sur la duration.

On peut donc avancer que la profondeur en temps que mesure d'intérêt par le marché caractérise les situations d'absence d'incertitude (d'hétérogénéité et d'asymétrie) qui entraînent des périodes d'activité concentrée (faible duration), avec des volumes et des volatilités faibles. De la même manière, la décision d'un agent de traiter lorsque la profondeur est faible entraîne une forte volatilité ce qui a pour effet d'augmenter l'espérance de la duration, du volume et de diminuer la profondeur prix.

La direction et l'hétérogénéité ont une influence encore plus instable. À l'exception de la

duration, l'effet de l'hétérogénéité est rarement significatif. Pour cette variable, elle a majoritairement une influence négative. Plus l'offre est symétrique des deux côtés du carnet, plus l'intensité des transactions est importante. Cela renforce l'hypothèse d'un lien entre activité forte et implication des non informés.
La direction est plus souvent significative mais est également dure à analyser en raison de changements de signe des paramètres. Les relations sont brièvement : - effet négatif sur la duration, sur le volume et positif sur la volatilité. Cela suggère que les réponses des variables sont différentes selon le sens du signal et que les agents sont plus pressés, ou traitent de plus forts volumes lorsque le signal est négatif (dans notre cas une plus faible offre du côté vente).

Au total, cette analyse fait ressortir que les états du carnet permettent de comprendre les caractéristiques des transactions. On observe que la dynamique mise en évidence correspond aux mécanismes de fourniture de la liquidité observés dans les études de la dynamique des ordres. Nous sommes également amenés à repousser les hypothèses liant la duration ou le volume et la présence d'informés. En effet, ces deux variables ne servent visiblement que d'approximations de l'état du carnet.

À l'exception des variables de profondeur, l'influence des autres variables est instables cela suggère que d'autres variables ou d'autres modélisations pourraient être plus pertinentes.

Conclusion de la partie II

Dans cette partie, nous avons analysé des données Ultra Haute Fréquence : l'ensemble des transactions et des meilleures limites sur 7 titres du CAC 40 pendant les 4 premiers mois de l'année 1999. Plus précisément, nous avons étudié les relations entre les caractéristiques des transactions puis avons étendu cette modélisation afin de prendre en compte l'état des 5 meilleures limites du carnet. À chaque fois, nous avons commencé par employer des modèles simples avant de déboucher sur des modélisations multivariées plus à même de tenir compte des relations entre les variables. Les résultats de ces dernières modélisations sont donc plus complets mais aussi plus complexes à analyser.

Un apport important des études réalisées dans cette partie est que l'application de différentes modélisations à un même problème souligne l'influence cruciale du modèle employé et de l'ensemble informationnel considéré. L'utilisation d'un ensemble informationnel trop restreint (les transactions dans le chapitre 5) conduit à mettre en évidence des relations fallacieuses dans la mesure où elles peuvent refléter d'autres facteurs non pris en compte. Il faut donc toujours conserver à l'esprit que suggérée par une modélisation dépendent de cette dernière.

Une limite des études que nous avons présentées dans cette partie découle de nos choix méthodologiques. Tout au long de cette partie, nous avons majoritairement employé des modélisations de type GARCH et supposé une structure de causalité (du carnet vers la duration puis vers le volume et enfin vers la volatilité). Nos résultats sont donc inféodés à la validité de ces deux choix. Nous avons pu voir que d'autres choix de modélisation étaient possibles. Les modélisations que nous avons employées semblent avoir une certaine pertinence et étaient largement représentées dans la littérature mais rien ne garantit que ce soient les meilleures.

Dans un chapitre introductif, nous avons pu apprécier la richesse des études de microstructure. Cette richesse même pose question puisque de nombreuses hypothèses concurrentes proposent d'expliquer un même phénomène. Nous avons tout de même mis en évidence deux concepts centraux qui sont l'asymétrie d'information et l'hétérogénéité. Nous avons ensuite tenu compte de ces enseignements et des spécificités des données Ultra Haute Fréquence afin de la rendre propre à une analyse ultérieure (traitement des transactions simultanées, construction des variables du carnet et désaisonnalisation).

L'analyse du processus de transactions, dans le chapitre 5, a mis en évidence un résultat

intéressant. Si sur le NYSE les périodes de forte activité sont généralement associées à des volatilités importantes et à de gros volumes et sont attribuées à l'implication d'informés (modèle de Easley & O'Hara (1992)), cela n'est pas le cas sur le marché parisien. Nos résultats concernant le marché parisien mettent en évidence une association positive entre la duration et la volatilité. Globalement on observe une influence positive du volume et de la duration sur la volatilité et une influence positive de la volatilité sur le volume et la duration. Il semble donc que, sur le marché parisien les périodes de forte activité ne soient pas attribuables à la présence d'informés. L'influence de la volatilité sur les autres processus penchant au contraire pour expliquer ces phénomènes par les mécanismes de fourniture de la liquidité dans un marché dirigé par les ordres (cf. Handa *et al.* (1998)).

Confronté à cette différence, dans le chapitre 6 nous avons étendu l'ensemble informationnel aux variables du carnet afin d'évaluer si elles permettaient d'expliquer les relations entre les caractéristiques des transactions. Nous avons pu observer que tel était le cas. Tout particulièrement, les influences de la duration et du volume sur la volatilité disparaissent lorsque les variables du carnet sont prises en compte. Il apparaît donc que ces deux variables permettent de synthétiser une information contenue dans le carnet mais ne semblent pas avoir de contenu informationnel propre.

L'ensemble des relations mises en évidence souligne que l'interaction principale entre le carnet et les caractéristiques des transactions s'apparente au mécanisme de fourniture de la liquidité dans un marché dirigé par les ordres (cf. Biais *et al.* (1995) ou encore Ranaldo (2004)). La seule variable dont la significativité est importante et l'influence stable est la profondeur qui a une influence négative sur la volatilité et la duration et positive sur le volume. Force est donc de constater que ni la duration ni le volume ne sont exogènes puisqu'ils dépendent directement de la profondeur.

Ce chapitre nous a également permis d'analyser les relations entre les variables du carnet et de souligner que les dimensions liées à l'activité (nombre d'ordres ou volume des ordres à une limite) et celles liées à la fourchette réagissaient de manière différente. Les premières semblent plus sensibles à l'asymétrie alors que les secondes le sont à l'hétérogénéité. Ces considérations vont globalement dans le sens du modèle de Dupont (2000).

Un apport inattendu de l'analyse en composantes principales des différentes dimensions de la liquidité que nous avons définies est de souligner l'existence de très fortes variations selon les titres considérés. Ces changements ont affecté la modélisation subséquente en rendant les résultats instables. Nous ne sommes pas parvenu à relier ces comportements à d'autres caractéristiques. La question demeure donc ouverte.

Nous avons mis en évidence que la dynamique principale dans le marché parisien se résumait à la fourniture de liquidité (qui peut être compatible avec la résolution de l'hétérogénéité).

La question de savoir exactement comment l'information asymétrique est incorporée est encore ouverte. Toujours est-il que l'influence d'une telle incorporation n'est pas visible dans notre échantillon et avec notre méthodologie.

Devant l'absence de modélisation adaptée à des marchés dirigés par les ordres englobant à la fois l'état du carnet et les caractéristiques des transactions tout en considérant en même temps de l'asymétrie d'information et de l'hétérogénéité, il nous est difficile de parvenir à une conclusion plus précise. Les études présentées dans cette partie conduisent donc plutôt à poser de nouvelles questions qu'à apporter des réponses. Il serait par exemple tout à fait pertinent de se demander comment l'asymétrie d'information est résolue dans ce marché.

Notre approche exploratoire nous a permis de confirmer la pertinence de variables reflétant l'état du carnet. On peut toutefois se demander s'il ne serait pas profitable de marier une telle approche avec une analyse de la dynamique des ordres. En effet, les états du carnet, s'ils permettent d'apprécier la qualité du marché et participent à l'ensemble informationnel employé par les agents, souffrent également de difficultés d'interprétations. Aussi, nous pensons qu'un tel rapprochement permettrait d'analyser cette relation avec plus de finesse.

Conclusion générale

Pour conclure le travail de recherche décrit au long de ces pages, nous rappelons succinctement la problématique et les principaux résultats.

L'objectif de ce travail était d'essayer de comprendre comment un investisseur observant le marché pouvait inférer la présence d'information. L'ordre de grandeur des phénomènes informationnels étant selon nous intrajournalier, toutes les études réalisées au cours de cette recherche partagent cette dimension. Notre travail a tenté une exploration empirique progressive de ces phénomènes : - La première partie a analysé des données Haute Fréquence puis la seconde des données Ultra Haute Fréquence ; - La complexité des modélisations employées dans la seconde partie est allée croissante.

La première partie a évalué, dans un cadre Haute Fréquence (i.e. les données sont agrégées dans des intervalles intrajournaliers), la pertinence de plusieurs approximations du flux informationnel déjà reconnues dans la littérature. Elle repose globalement sur l'idée de la MDH selon laquelle le temps réel d'évolution des marchés n'est pas le temps horloge mais un temps informationnel. Nous avons donc déformé le temps selon les différentes approximations afin d'évaluer laquelle permettait un plus grand rapprochement des constatations empiriques et des prévisions théoriques. Le chapitre 1 a préparé les deux analyses suivantes. Dans un premier temps, nous avons présenté l'organisation du marché parisien puis défini les retraitements qui s'imposaient. Nous avons ensuite passé en revue les différentes approximations du flux informationnel proposées dans littérature et expliqué notre méthodologie de construction des déformations. Enfin, nous avons analysé les effets des retraitements et des déformations. Dans les deux chapitres suivants, nous avons étudié les propriétés des séries dans ces différentes déformations temporelles afin d'évaluer la pertinence des différentes approximations. Deux critères ont été employés : dans le chapitre 2 les approximations ont été jugées selon leur capacité à réduire les configurations intrajournalières alors que le chapitre 3 a étudié dans quelle mesure une déformation temporelle permettait de rapprocher la distribution empirique de la normalité.

Cette partie a eu 2 apports principaux. Nous avons noté que le choix d'une longueur d'intervalle modifiait les résultats (capacité du volume à expliquer l'autocorrélation de la volatilité, effets des déformations temporelles sur les distributions). Ces observations soulignent qu'une procédure d'agrégation intrajournalière n'est pas neutre. Elle a des conséquences difficilement estimables. Cette constatation débouche sur deux conclusions importantes pour des études ultérieures. D'une part, une étude intrajournalière qui emploierait une méthode d'agrégation se doit de contrôler l'influence de la longueur de l'intervalle sur les résultats. D'autre part, une étude désagrégée (Ultra haute Fréquence) semble plus adaptée aux caractéristiques des données.
Le second apport est le classement des approximations. Parmi les quatre approximations comparées (nombre de transactions, volume de transactions, nombre d'ordres et volume d'ordres) toutes ont une certaine pertinence mais aucune ne parvient à satisfaire les critères de validation. Elles ne permettent donc pas de construire un indicateur informationnel synthétique satisfaisant.

Nous avons toutefois noté que les classements obtenus dans les études des chapitres 2 et 3 sont globalement convergents. Le classement qui émerge est : -i.) Le nombre de transactions ; -ii.) Le volume de transactions ; -iii.) Le nombre d'ordres, -iv.) Le volume d'ordres. Une étude désirant employer une approximation de l'information devrait donc se diriger de préférence vers le nombre de transactions.

Étant donné les problèmes posés par une agrégation des données dans des intervalles, nous avons adopté dans la seconde partie une approche Ultra Haute Fréquence. Nous avons donc étudié l'ensemble exhaustif des données en utilisant des modélisations adaptées aux données espacées irrégulièrement. L'objectif de cette partie était de comprendre comment les différentes composantes du processus d'échange (caractéristiques des transactions : duration, volume, rendement et états du carnet d'ordres) interagissaient. À l'image du chapitre 1, nous avons dans le chapitre 4 revu la littérature qui pouvait nous guider dans l'interprétation des résultats ou dans les choix méthodologiques. Nous avons également présenté les retraitements indispensables à ce type d'étude. Un point important traité dans ce chapitre est la construction des variables du carnet. À partir des 5 meilleures limites nous avons défini trois ensembles de variables traduisant les caractéristiques de l'offre exprimée en carnet : - l'hétérogénéité (offre que l'on retrouve des deux côtés du carnet) ; - la direction (côté du carnet où l'offre est la plus faible) ; - la profondeur (somme de l'offre des deux côtés du carnet). Nous avons ensuite entrepris une analyse progressive des différentes composantes du processus d'échange. Nous avons dans le chapitre 5 étudié les interactions entre les caractéristiques des transactions. Enfin, dans le chapitre 6, nous avons étendu cette modélisation en y incorporant l'information contenue dans le carnet d'ordres (les variables construites à partir des 5 meilleures limites).

Cette partie a eu trois apports importants. L'analyse du processus de transaction a mis en évidence une différence importante entre le marché parisien et le NYSE. Les études du NYSE concluent à une association entre les périodes de forte activité et la présence d'information (influence négative de la duration sur la volatilité). Au contraire nos résultats suggèrent une association entre les périodes de faible activité et la présence d'informés. Les périodes de forte activité seraient alors le fait d'une concentration des non-informés.

L'intégration des variables du carnet a fait ressortir que les caractéristiques des transactions n'ont pas de contenu informationnel propre mais servent d'approximations de l'information contenue dans le carnet. Il apparaît alors que les dynamiques observées sont principalement en adéquation avec les mécanismes de fourniture de liquidité. Ni le volume, ni la duration ne révèlent à eux seuls la présence d'informations. Ces deux variables doivent être analysées au regard de l'état du carnet.

Le dernier apport est une recommandation pour les modélisations des marchés dirigés par les ordres. Il est manifeste que ni la duration (l'arrivée d'un agent désirant traiter immédiatement)

ni le volume de transactions ne sont exogènes. Aussi, une modélisation se doit d'endogénéiser ces deux choix des agents et les conclusions des modèles qui ne partagent pas ces propriétés sont suspectes.

Ce travail n'est pas sans avoir des limites. Nous avons étudié les mécanismes informationnels pendant les transactions en continu. Nous avons donc exclu toutes les données se référant aux fixages d'ouverture et de clôture. Ceux-ci véhiculent une information que nous avons négligé étant donné qu'il est impossible de les traiter avec les méthodologies utilisées en raison de la différence de mécanisme d'échange. Nous avons également choisi de nous focaliser exclusivement sur des données de marché négligeant l'enrichissement qu'aurait constitué une confrontation avec l'avis des praticiens. Toutefois, nous pensons que certaines questions concernant les stratégies de placement des ordres par les agents doivent être résolues avant que les bénéfices d'une telle confrontation soient substantiels. Les limites principales de cette thèse tiennent à l'optique résolument exploratoire qui a été choisie ne proposant aucune modélisation. Cependant, comme nous avons pu le voir tout au long de ce travail, cette problématique est encore peu étudiée sur les marchés dirigés par les ordres. Aussi, nous pensons que ce travail permet d'ouvrir des pistes de réflexion pour des recherches ultérieures. Par exemple, nous avons regretté l'absence d'une modélisation, adaptée aux marchés dirigés par les ordres, qui tiennent compte en même temps de l'état du carnet, de l'asymétrie d'information et de l'hétérogénéité. Cette recherche débouche sur des interrogations concernant les stratégies de placements d'ordres par les agents. Nous n'avons malheureusement pas pu les étudier en raison de l'indisponibilité de données précises sur les ordres (identités des introducteurs, motifs des transactions).

Les limites que nous avons noté sont autant d'ouvertures potentielles. La disponibilité (malheureusement trop tardive) de données précises sur les ordres nous permet d'envisager une étude des stratégies de placement d'ordres utilisées par les agents. Par exemple, nous avons noté que le volume des transactions ne semblait pas à lié à l'intervention des informés, il serait intéressant de vérifier si ce fait n'est pas dû à un comportement de fractionnement des transactions.
Pour intégrer l'information contenue dans le carnet, nous avons construit des variables reflétant l'état de l'offre. Cette approche nous a permis d'en démontrer l'importance primordiale pour comprendre les caractéristiques des transactions. Il serait également très intéressant d'étudier comment l'offre en carnet interagit avec les introductions d'ordres.
Notre analyse a conclu que, dans leur majorité, les relations entre les caractéristiques des échanges sont attribuables à des phénomènes de liquidité. L'analyse des variations de l'impact prix des transactions en fonction des différentes variables que nous venons d'étudier permettrait certainement de compléter et d'apprécier la portée exacte de nos résultats.
Nos résultats ont également suggéré que plus que les caractéristiques des transactions elles-mêmes, ce sont les relations entre celles-ci et les états du carnet (e.g. décider de traiter un volume important alors que le carnet est peu profond) qui influencent la réaction du marché.

Aussi, une analyse des effets des transactions qui vont à "l'encontre" des signaux observables dans le carnet apporterait sûrement des éclaircissements des plus profitables.
Enfin, à titre plus lointain, il est certain qu'une modélisation prenant en compte en même temps l'asymétrie d'information, l'hétérogénéité des opinions tout en endogénéisant l'arrivée des agents sur le marché et le volume de transactions en fonction de l'état du carnet serait une contribution très significative à la connaissance des marchés.

Annexes

Annexe A

La Base De Marché

La description de la base de données que nous utilisons s'inspire du manuel fourni par la SBF. BDM est la base de données mise en place par la Bourse de Paris[1] depuis 1990. Elle contient l'information de base (caractéristiques de cotations, nombres de titres, dividendes ...) et le détail des négociations (ordres et transactions) effectuées sur les titres cotés sur la place de Paris. Cette base est diffusée sur CD-Rom et contient 15 tables dont le contenu est résumé dans la table A.1.

TABLE A.1
La base BDM

Nom du fichier	Contenu
BDC1D2	Calendrier Annuel
BDM1D2	Transactions
BDM2D2	Meilleures limites
BDM3D2	Évènements de cotations
BDFMD2	F.M.P.
BDM8D2	Ordres de marché
BDIDD2	Codes externes
BDCCD2	Caractéristiques des cotations
BDESD2	Secteurs économiques
BDTAD2	Nombre de titres
BDVAD2	Type de valeur
BDCPD2	Dividendes
BDKAD2	Coefficients d'ajustements
BDM5D2	Informations instantanées indices
BDMNHD2	Opérations hors marché

Dans cette base, nous utilisons principalement 4 tables dont les différents champs sont détaillés dans les tables A.2 à A.5.

La table A.2 contient les caractéristiques des transactions. C'est la principale table que nous

[1]Cette base est le prolongement de la base Association Française de FInance (AFFI)-SBF dont on trouvera une description complète dans Hamon & Jacquillat (1991*a*).

TABLE A.2
BDM1D2

Code valeur interne
Date
Heure
Numéro de séquence
Cours
Nombre de titres
Indicateur d'application 1 dans ce cas 0 sinon

utilisons. Elle nous sert à construire les séries de prix de transaction ainsi que les déformations temporelles fondées sur le nombre ou le volume de transactions. Cette table contient l'information sur les transactions (date, prix, taille . . .) effectuées par le système Cotation Assistée en Continu (CAC). Chaque transaction entraîne la création d'un enregistrement.

TABLE A.3
BDCCD2

Code valeur interne
Date de début de validité
Date de fin de validité
Type de gestion : 1 Comptant ; 2marché à Règlement Mensuel (RM)
Section : 1 Marché officiel ; 2 Second marché
Appartenance indicielle : E CAC 40 R Valeur de remplacement du CAC 40 1 SBF 120 mais pas CAC 40 2 SBF 250 mais pas SBF 120 0 Aucun indice
Groupe de valeur pour la cotation : 01 CAC 40 05 Nouvellement privatisées 11 Continu A RM 12 Françaises Continu B Comptant 13 Françaises Fixing A 14 Fixing B hors cote 16 Françaises Continu A Comptant 14 Continu B RM 21 Étrangères Continu A RM 22 Étrangères Continu A Comptant 23 Étrangères Fixing A 31 Secondaires Continu A 32 Secondaires Continu A RM 33 Secondaires Fixing A 34 Secondaires Fixing B (hors cote)

La table A.3 nous informe sur l'ensemble des caractéristiques des cotations d'un valeur. Elle nous sert à identifier les titres ne changeant pas d'indice pendant la période d'étude. Toute modification d'une caractéristique des cotations entraîne la création d'un nouvel enregistrement.

TABLE A.4
BDESD2

Code valeur interne
Date de début de validité
Date de fin de validité
Code secteur économique

La table A.4 fournit le secteur économique auquel l'entreprise appartient. C'est en croisant les secteurs économiques et l'appartenance indicielle (table A.3) que nous déterminons les titres sélectionnables.

La table A.5 nous sert à définir les déformations temporelles à partir du nombre ou du volume des ordres introduits.

Nous employons également de manière annexe la table fournissant la meilleure limite (BDM2D2) afin d'obtenir pour chaque transaction les prix acheteur, vendeur et milieu de fourchette.

TABLE A.5
BDM8D2

Code valeur interne
Code de la société de bourse qui introduit l'ordre
Date de saisie de l'ordre
Heure de saisie de l'ordre
Type de validité de l'ordre J Jour D Date explicite R A révocation L Liquidation E Exécuté ou éliminé I Illimité
Date de validité de l'ordre
Indicateur du sens de l'ordre : A achat ou V vente
Quantité demandée ou offerte dans l'ordre
Nombre de transactions réalisées sur l'ordre
Quantité de titres restant à exécuter
Code de référence de l'ordre
Code de référence d'un ordre pour chaînage
Date de début de validité de l'enregistrement
Date de fin de validité de l'enregistrement
Indicateur de l'état de l'ordre : Ordre entré en carnet (y compris P) P Partiellement exécuté T (X) Totalement exécuté A Annulé V Validité échue M Modifié D Déclenché de type Stop O Éliminé S Annulé par la surveillance G Éliminé (gel de la valeur) E Éliminé (ordre spécifique)
Indicateur du type de l'ordre : L Prix limite M Prix du marché X Au mieux O Ouverture T Tout ou rien S Stop U Prix limite issu d'un Stop V Au mieux issu d'un Stop 9 Indéterminé ou non significatif
Prix limite de l'ordre (si significatif)
Quantité de titres dévoilés dans l'ordre

Annexe B

Base de données

La section B.1 donne la liste des titres sélectionnables. Dans la section B.2, nous analysons l'effet des retraitements préalables à l'analyse (e.g. le retrait des transactions aux fixages).

B.1 Titres sélectionnables

La table B.1 donne l'ensemble des titres sélectionnables d'après les critères exposés à la page 15. Dans cette liste, nous avons omis les titres appartenant au secteur immobilier.

Le nombre de titre sélectionnables (57) est faible. Si l'on veut préserver une uniformité des titres sélectionnés (nombre identique de titres dans chaque groupe de liquidité et chaque secteur économique) tout en conservant un maximum de secteurs économiques alors, on ne peut sélectionner qu'un titre pour chaque sous-groupe. Pour certains sous-groupes des choix sont offerts alors que pour d'autres aucun choix n'est possible. Par exemple, nous pouvons choisir le titre du sous-groupe HL matériaux entre Lafarge et St-Gobain alors que pour le secteur Aéronautique il n'y qu'un titre pour chaque sous-groupe.

B.2 Influence des retraitements

La table B.2 décrit l'influence des retraitements définis à la page 20. Nous comparons les rendements et les volumes journaliers des données retraitées et des données brutes (téléchargées sur le site de la SBF).

Le critère d'exclusion des jours (au moins deux heures entre la première et la dernière transaction en continu) n'a d'effet notable que sur le groupe LL. Aucun titre du groupe HL n'est exclu. Pour le groupe ML, seuls Eurafrance et Lapeyre sont exclus quelques jours (respectivement 1 et 3 jours supprimés). Dans le groupe LL, le nombre de jours exclus est plus important. Il se situe généralement autour de la dizaine sauf pour les titres Vicat et Radiall où il atteint 300.

TABLE B.1
Titres sélectionnables

	Indice	Secteur
Thomson-CSF	CAC 40	
Zodiac	SBF 120	Aéronautique Espace Armement
Dassault Aviation	SBF 250	
Alcatel Alsthom	CAC 40	
Schneider	CAC 40	
Sagem	SBF 120	Électricité Électronique Télécommunication
Cs Cie Des Signaux	SBF 250	
Radiall	SBF 250	
Gle Des Eaux	CAC 40	
Lyonnaise Des Eaux	CAC 40	Environnement Services aux collectivités
Technip	SBF 120	
Assystem	SBF 250	
Accor	CAC 40	
Club Méditerranée	SBF 120	
Euro Disney Sca	SBF 120	Hôtels Restauration Tourisme
Groupe Partouche	SBF 250	
Louvre	SBF 250	
Lafarge	CAC 40	
Saint-Gobain	CAC 40	
Imetal	SBF 120	Matériaux
Lapeyre	SBF 120	
Vicat	SBF 250	
Oreal	CAC 40	
Sanofi	CAC 40	
Clarins	SBF 120	
Boiron	SBF 250	
Guerbet	SBF 250	Pharmacie Cosmétiques
Labo.Atro.	SBF 250	
Robertet	SBF 250	
Virbac	SBF 250	
Lézard.	CAC 40	
Cgip	SBF 120	
Eurafrance	SBF 120	
Gaz Eaux Fin.Nid.	SBF 120	
Burelle	SBF 250	Sociétés Holdings
Christian Dior	SBF 250	
Dynaction	SBF 250	
Fonc.Fin.Participa.	SBF 250	
Marine Wendel	SBF 250	

Cette table donne les titres qui, sur l'ensemble de la période, n'ont pas changé d'indice et appartiennent à des secteurs économiques qu'on retrouve dans les trois indices (CAC 40, SBF 120 et SBF 250).

TABLE B.2
Influence des retraitements

	Jour	Rendement	Volatilité	Volume
Thomson	1 004	-0,066	2,822	314 853
	1 004	-0,055	2,683	176 257
Zodiac	1 005	-0,144	2,028	10 232
	1 005	-0,146	1,798	6 311
Dassault	1 003	0,209	2,211	3 324
	994	0,022	2,049	1 388
Schneider	1 005	-0,122	2,371	524 875
	1 005	-0,119	2,271	313 616
Sagem	1 004	0,196	2,638	36 086
	1 004	0,079	2,477	6 161
Radiall	1 003	0,223	2,467	964
	671	0,240	2,339	794
Lyonnaise	1 003	0,099	1,688	534 235
	1 003	0,086	1,662	332 490
Technip	1 005	-0,148	2,251	46 088
	1 005	-0,162	2,074	29 733
Assystem	1 004	0,259	2,794	5 334
	955	0,154	2,613	2 739
Accor	1 005	-0,168	2,233	808 395
	1 005	-0,211	2,183	214 929
Club Med.	1 004	0,067	2,307	49 147
	1 004	0,065	2,124	30 748
Partouche	1 005	-0,089	2,028	3 047
	989	-0,128	1,902	2 073
St-Gobain	1 005	-0,056	2,048	297 420
	1 005	-0,013	1,964	165 246
Lapeyre	1 005	-0,011	2,340	17 489
	1 002	-0,043	2,170	9 176
Vicat	988	-0,032	2,108	1 131
	633	0,064	2,017	545
Sanofi	1 005	0,066	2,330	1 174 744
	1 005	0,042	2,280	398 025
Clarins	1 005	0,095	2,000	27 927
	1 005	0,036	1,885	13 068
Boiron	1 005	0,493	1,775	3 723
	985	0,353	1,749	2 064
Lagardère	1 005	-0,075	2,738	489 750
	1 005	-0,099	2,627	303 308
Eurafrance	1 005	-1,173	1,393	45 317
	1 004	0,294	1,664	2 183
Dynaction	1 005	0,052	1,977	10 989
	1 002	0,067	1,783	6 931

Pour chaque titre, la ligne supérieure donne les caractéristiques des données brutes (données téléchargées), la ligne inférieure fournit les paramètres des données retraitées. Un jour n'est admis que si le temps entre la première et la dernière transaction en continu est supérieur à 2 heures. Le rendement est exprimé en %, il est la moyenne des rendements journaliers, la volatilité est l'écart-type des rendements. Le volume est le nombre moyen de titres échangés par jour.

L'activité brute ou retraitée la plus importante est observée pour le groupe HL, ensuite pour le groupe ML enfin le groupe LL[1]. Cette différence de volume d'activité suggère qu'une même la fréquence d'échantillonnage des séries n'est pas adaptée aux différents groupes de liquidité.

L'effet de l'exclusion des fixages sur le volume de transactions est important puisque le volume journalier moyen chute fortement. Par exemple, le volume journalier d'Accor est presque divisé par 4, celui d'Eurafrance par 22,5. La réduction est en général d'autant plus importante que le volume brut est élevé par rapport au groupe d'appartenance[2]. C'est dans le groupe ML que l'on observe les plus fortes diminutions (e.g Eurafrance ou encore le volume journalier moyen de Sagem qui est divisé par 6).

L'influence du retrait des fixages sur le rendement est plus trouble. Les rendements bruts et retraités sont, la plupart du temps, proches. Les différences sont dans 14 cas de l'ordre de 0,01%. Dans les cas où elles sont supérieures à 0,1%, on a toujours un rendement retraité inférieur. Ainsi, il semble que, quand la différence est importante, les fixages ont pour effet d'augmenter une tendance haussière et de stabiliser une tendance baissière.

À l'exception d'Eurafrance, la volatilité des rendements retraités est toujours inférieure à la volatilité des rendements bruts. Cela souligne que la volatilité intrajournalière est inférieure à la volatilité de la clôture. Les valeurs demeurent cependant très proches.

[1]Cette constatation demeure vraie si on considère le taux de rotation (le nombre d'actions échangées divisé par le nombre de titres).

[2]Cela souligne qu'un volume élevé par rapport au groupe d'appartenance est surtout le fait d'une augmentation du volume traité aux fixages et, dans une moindre mesure, d'un développement de l'activité intrajournalière.

Annexe C

Prix et calculs

Afin de retenir pour analyse détaillée les prix et les modalités de calcul les plus intéressants, nous procédons ici à une analyse succincte des différents prix et modalités de calculs exposés dans le point c..

Dans la section C.1 nous analyserons les différences entre les prix puis, dans la section C.2, nous examinerons les différentes modalités de calcul.

C.1 Comparaison des prix

La figure C.1 présente pour le groupe complet (regroupement tous les titres sélectionnés) l'influence du prix utilisé.

On peut remarquer que les configurations de volatilité sont très similaires pour les différents prix. Les configurations de rendement varient selon le prix. Les prix d'achat et de transaction ont des configurations de clôture alors que ces dernières semblent faibles ou inexistantes avec le prix milieu de fourchette ou le prix de vente. Cela suggère que l'effet de clôture des transactions est dû à un déséquilibre du flux d'ordre à l'achat lors de la clôture. L'effet ouverture est plus variable en fonction des prix et des temps. Il apparaît globalement peu significatif.

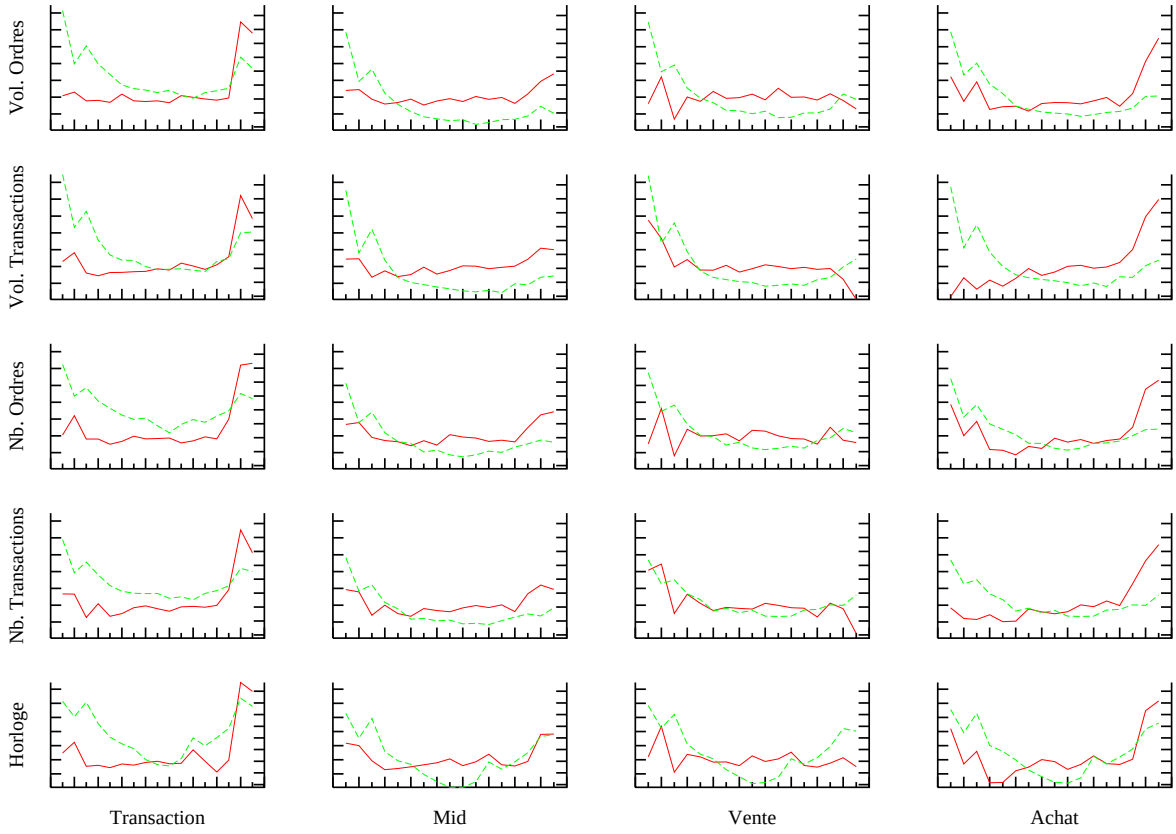
On peut conclure que deux prix résument les différentes configurations : le prix de transaction et le prix milieu de fourchette. Par conséquent, nous nous focaliserons sur ces deux prix.

C.2 Comparaison des modalités de calcul

La figure C.2 compare l'influence de la modalité de calcul sur les configurations intrajournalières de l'indice global avec un prix de transaction.

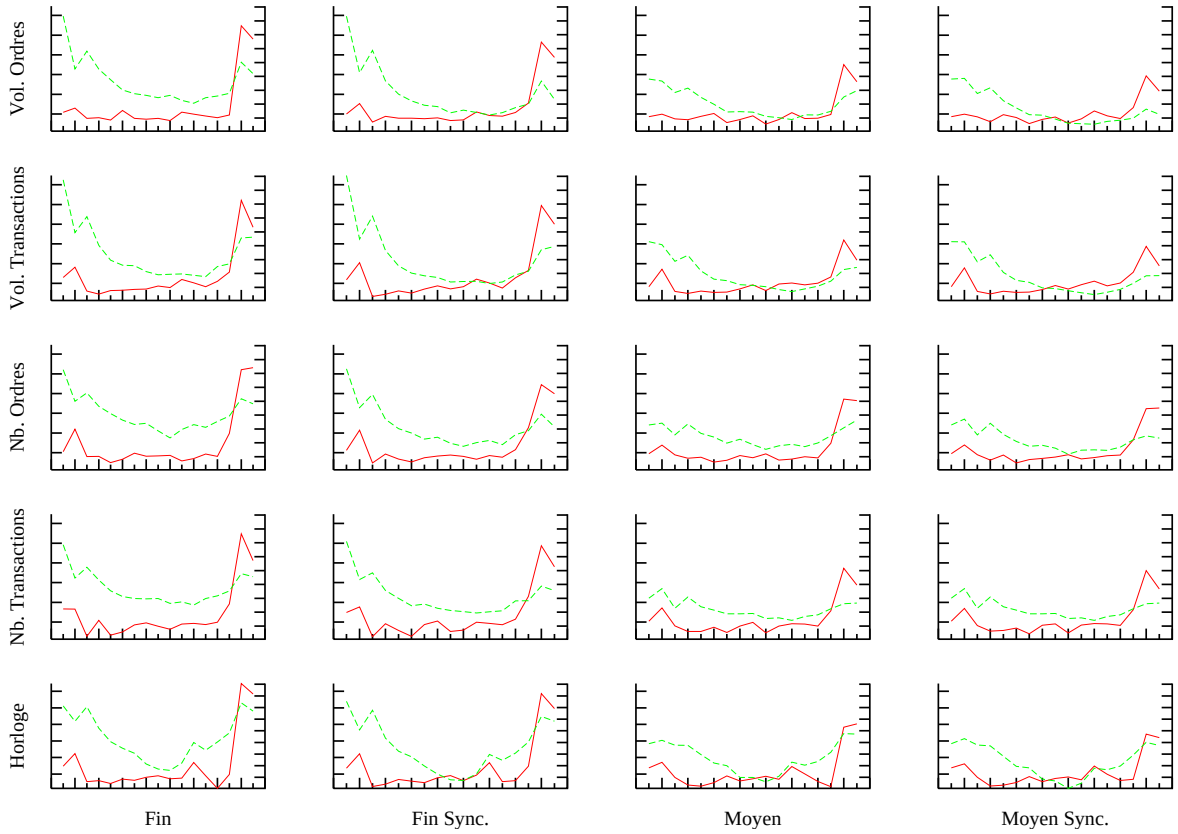
Globalement, un prix moyen réduit l'amplitude des mouvements des rendements et de la

FIGURE C.1
Comparaison des prix : indice global, 30 minutes
prix fin d'intervalle



Ce graphique compare les configurations intrajournalières de l'indice global (tous les titres) avec une longueur de 30 minutes et un calcul fin d'intervalle. Le prix utilisé est donné en bas de chaque colonne, le temps à gauche de chaque ligne. Tous les graphiques dans cette figure ont la même échelle, la courbe rouge continue est le rendement, la courbe verte hachurée la volatilité.

FIGURE C.2
Modalités de calcul : indice global, 30 minutes
prix de transaction



Ce graphique compare les configurations intrajournalières de l'indice global pour une longueur de 30 minutes et un prix de transaction. La modalité de calcul est donnée en bas de chaque colonne, le temps à gauche de chaque ligne. Tous les graphiques dans cette figure ont la même échelle, la courbe rouge continue est le rendement, la courbe verte hachurée la volatilité.

volatilité. L'action conjointe de ces deux effets ne modifie pas au final les configurations mises en évidence. Il n'y a aucune différence notable entre le prix moyen et moyen synchronisé.

La synchronisation du prix fin d'intervalle induit une forte réduction de l'effet de clôture. L'effet de la synchronisation du prix fin d'intervalle et les configurations observées avec le prix moyen soulignent que l'effet de clôture des prix de transaction est concentré à la fin de l'intervalle[1].

La synchronisation d'un prix réduit l'amplitude des mouvements extrêmes (effet de clôture pour la volatilité et le rendement). Elle atténue également l'oscillation des rendements dans la journée. Ces propriétés étant souhaitables car elles permettent de mieux mettre en évidence les résultats significatifs, nous utiliserons un prix fin d'intervalle synchronisé.

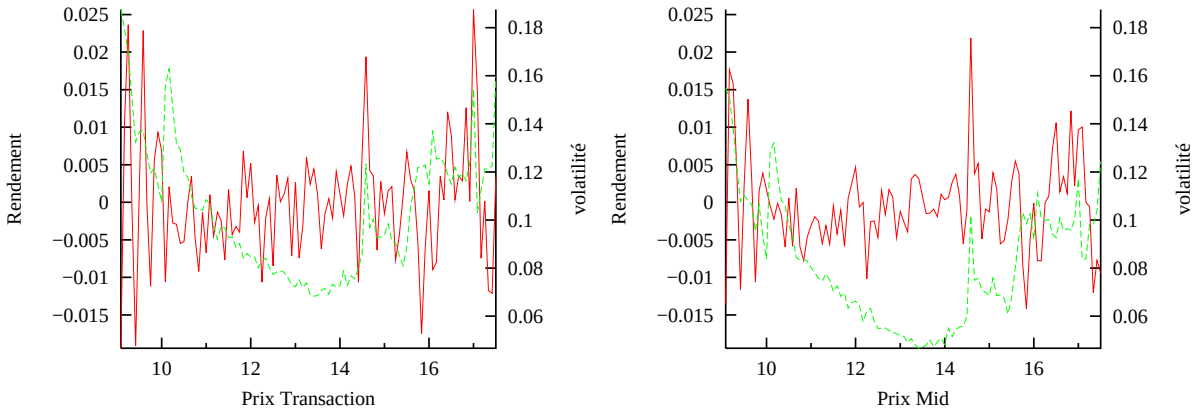
[1]La synchronisation induit une réduction car elle répartit une partie du rendement du dernier intervalle sur l'intervalle précédent. La différence entre le prix moyen et le prix fin d'intervalle prouve que la tendance haussière est observée à l'intérieur de l'intervalle. Ensemble ces constatations montrent que l'effet est plus fort à la fin l'intervalle.

Annexe D

Configurations intrajournalières

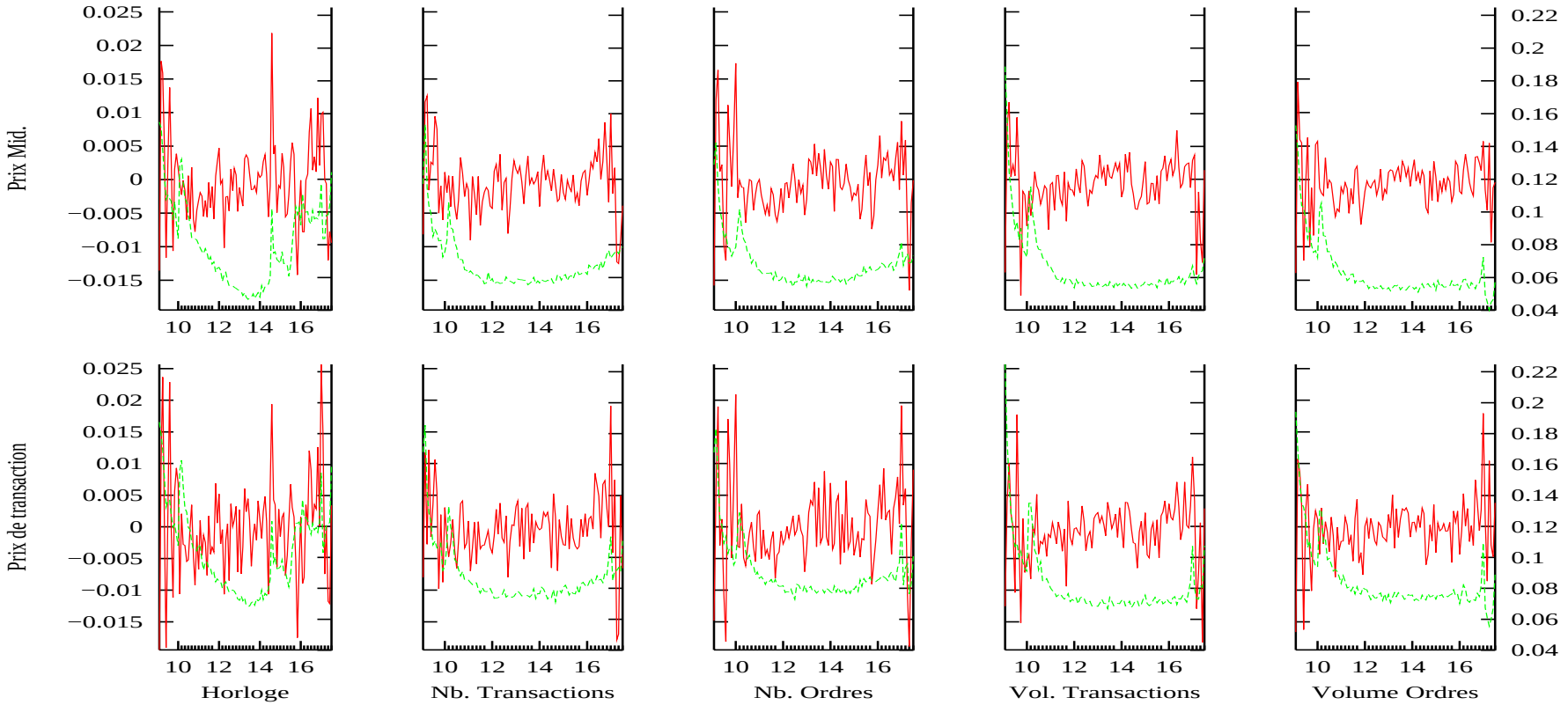
D.1 Groupe HL

FIGURE D.1
Configurations en *temps calendaire*
groupe HL, 5 minutes



Cette figure donne, pour les différents prix, les configurations observées en temps calendaire. Le prix utilisé est indiqué en bas de chaque graphique. La courbe rouge continue est le rendement (axe de gauche), la courbe verte hachurée est la volatilité (axe de droite). Les deux graphiques de cette figure ont la même échelle.

FIGURE D.2
Configurations intrajournalières
groupe HL, 5 minutes



Cette figure donne pour le prix de transaction et le prix milieu de fourchette les configurations observées dans les différents temps. Le prix utilisé est indiqué à gauche de chaque ligne. Le temps est indiqué en bas de chaque colonne. La courbe rouge continue est le rendement (axe de gauche), la courbe verte hachurée est la volatilité (axe de droite). Tous les graphiques dans cette figure ont la même échelle. La première colonne de cette figure reprend donc les graphiques de la figure D.1.

TABLE D.1
Configurations de rendement groupe HL
15 minutes, *temps nb. transactions*

Paramètre	Estimation	Écart-type	T-stat
11:15	$-0{,}017^{*}$	0,010	-1,67
12:00	$-0{,}023^{**}$	0,011	-2,21
12:45	$-0{,}017^{***}$	0,006	-2,82
15:15	$0{,}022^{*}$	0,013	1,70
15:45	$0{,}052^{***}$	0,018	2,89
16:15	$0{,}075^{***}$	0,024	3,08
16:30	$0{,}043^{*}$	0,023	1,84
16:45	$0{,}082^{***}$	0,028	2,93
17:15	$0{,}057^{*}$	0,032	1,77
O_{-8}	$-0{,}037^{***}$	0,013	-2,89
O_{-6}	$-0{,}069^{***}$	0,019	-3,70
O_{-5}	$-0{,}032^{*}$	0,019	-1,70
O_{-4}	$-0{,}077^{***}$	0,025	-3,10
O_{-2}	$-0{,}074^{***}$	0,028	-2,64
O_{7}	$-0{,}018^{*}$	0,010	-1,87

Cette table fournit les paramètres estimés du modèle (2.1) appliqué aux rendements. Int. désigne la constante. Une variable significative au seuil de 1% est marquée par ***, au seuil de 5% par ** et au seuil de 10% par *.

TABLE D.2
Configurations de volatilité groupe HL
15 minutes, *temps volume*

Paramètre	Estimation	Écart-type	T-stat
Int.	$0{,}107^{***}$	0,004	28,72
09:15	$0{,}085^{***}$	0,021	4,10
09:30	$0{,}050^{***}$	0,016	3,19
09:45	$0{,}047^{***}$	0,015	3,07
10:00	$0{,}044^{***}$	0,014	3,26
10:15	$0{,}026^{**}$	0,011	2,44
10:30	$0{,}032^{***}$	0,011	2,97
10:45	$0{,}028^{***}$	0,011	2,68
11:00	$0{,}048^{***}$	0,010	4,72
11:15	$0{,}019^{***}$	0,007	2,59
11:45	$0{,}014^{**}$	0,007	2,00
12:00	$0{,}012^{*}$	0,007	1,80
12:15	$0{,}008^{*}$	0,004	1,79
O_{1}	$0{,}130^{***}$	0,013	9,71
O_{2}	$0{,}047^{***}$	0,012	3,86
O_{3}	$0{,}024^{**}$	0,012	2,03
O_{8}	$-0{,}018^{***}$	0,007	-2,63
Vendredi	$0{,}006^{**}$	0,003	2,06

Cette table fournit les paramètres estimés du modèle (2.1) appliqué aux volatilités. Int. désigne la constante. Une variable significative au seuil de 1% est marquée par ***, au seuil de 5% par ** et au seuil de 10% par *.

D.2 Groupe ML

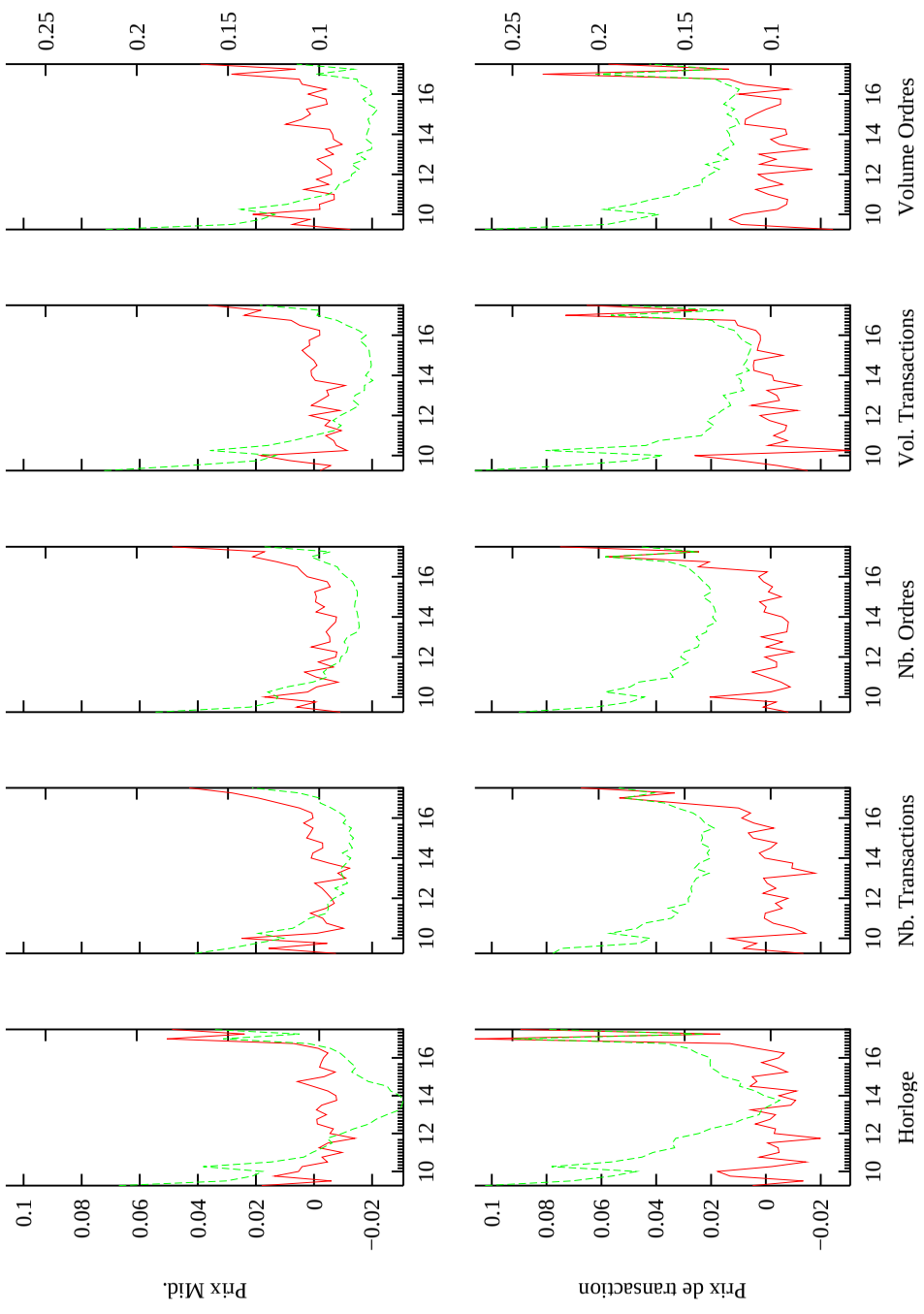


FIGURE D.3: Configurations intrajournalières groupe ML, 15 minutes

Cette figure donne pour le prix de transaction et le prix milieu de fourchette les configurations observées dans les différents temps. Le prix utilisé est indiqué à gauche de chaque ligne. Le temps est indiqué en bas de chaque colonne. La courbe rouge continue est le rendement (axe de gauche), la courbe verte hachurée est la volatilité (axe de droite). Tous les graphiques dans cette figure ont la même échelle.

D.3 Groupe LL

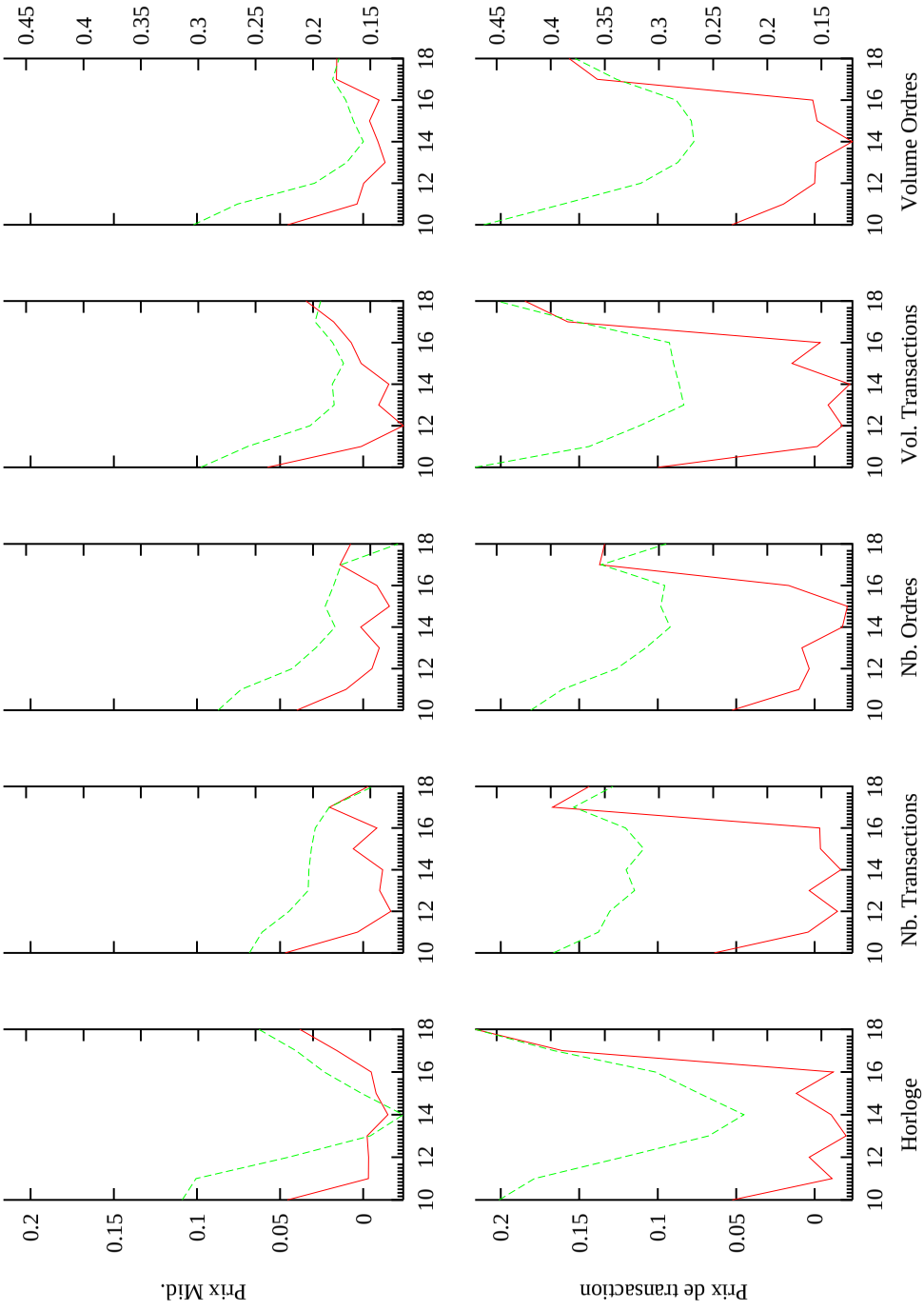


FIGURE D.4: Configurations intrajournalières groupe LL, 1 heure

Cette figure donne pour le prix de transaction et le prix milieu de fourchette les configurations observées dans les différents temps. Le prix utilisé est indiqué à gauche de chaque ligne. Le temps est indiqué en bas de chaque colonne. La courbe rouge continue est le rendement (axe de gauche), la courbe verte hachurée est la volatilité (axe de droite). Tous les graphiques dans cette figure ont la même échelle.

Annexe E

Propriétés des séries

E.1 Dynamique du rendement

La table E.1 donne les estimations d'un EGARCH(1,1) sur les différentes déformation temporelles.

TABLE E.1
Dynamique des rendements EGARCH(1,1)
groupe HL, 5 minutes

	Calendaire	*Nb. Transactions*	*Nb. Ordres*	*Volume*	*Vol. Ordres*
μ	-0,0 000	0,0004***	0,0040***	0,0003	0,0035***
	(-0,00)	(8,88)	(15,41)	(1,28)	(13,84)
α_0	-0,293***	-0,174***	-0,408***	-0,228***	-0,499***
	(-43,06)	(-26,46)	(-34,40)	(-29,85)	(-36,79)
α_1	0,338***	0,236***	0,231***	0,264***	0,261***
	(65,99)	(52,86)	(45,18)	(54,24)	(47,72)
β_1	0,930***	0,964***	0,916***	0,952***	0,899***
	(593,11)	(725,41)	(377,70)	(619,59)	(330,37)
θ	-0,047***	-0,047***	-0,048***	-0,040***	-0,045***
	(-6,75)	(-6,36)	(-5,42)	(-5,56)	(-5,54)
Log Likelihood	78 650	97 732	95 448	98 805	100 945
Jarque-Béra	422 728***	12 278***	806 596***	72 382***	957 157***

Les différentes estimations de θ sont globalement similaires. Toutes les déformations temporelles induisent une diminution α_1. L'effet est différent sur β_1 : les déformations temporelles utilisant les ordres induisent une réduction alors que celles employant les transactions entraîne une augmentation. La résultante est une légère diminution de la persistance de la volatilité en temps activité.

E.2 Distribution des rendements LL

La figure E.1 (E.2) donne, pour le groupe LL, la distribution des rendements en *temps horloge* (*temps nombre de transactions*) avec un intervalle de 30 minutes.

Les phénomènes observés sont similaires à ceux notés sur le groupes ML. Nous laissons le soin au lecteur de se reporter à la sous-section 3.3.

FIGURE E.1
Distribution des rendements
temps horloge, groupe LL, 30 minutes

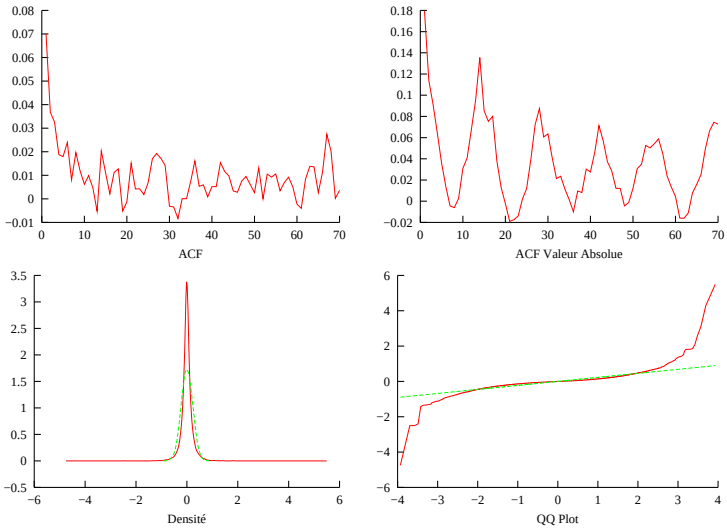
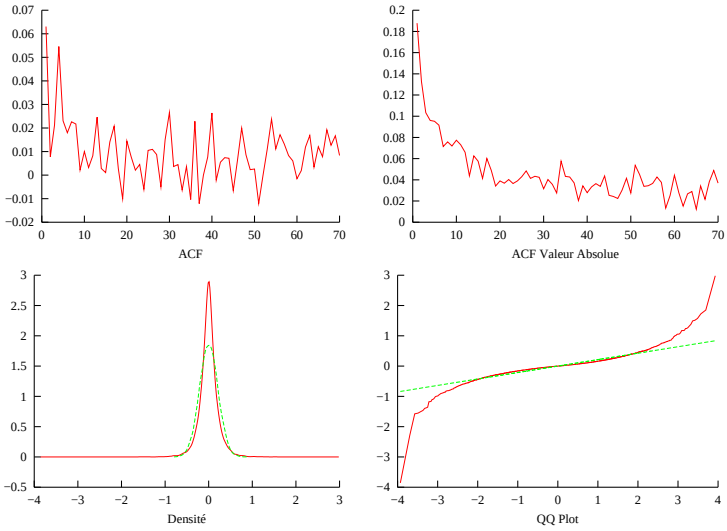


FIGURE E.2
Distribution des rendements
temps transaction, groupe LL, 30 minutes



Cette figure présente les principales caractéristiques de la distribution des rendements du groupe LL en *temps horloge* et en *temps transaction*. Le nom de la statistique calculée est donné en dessous de chaque graphique. La courbe rouge continue est la valeur empirique de la statistique, la courbe verte hachurée est, lorsqu'elle est présente, la valeur théorique si la distribution était normale de même moyenne et écart type.

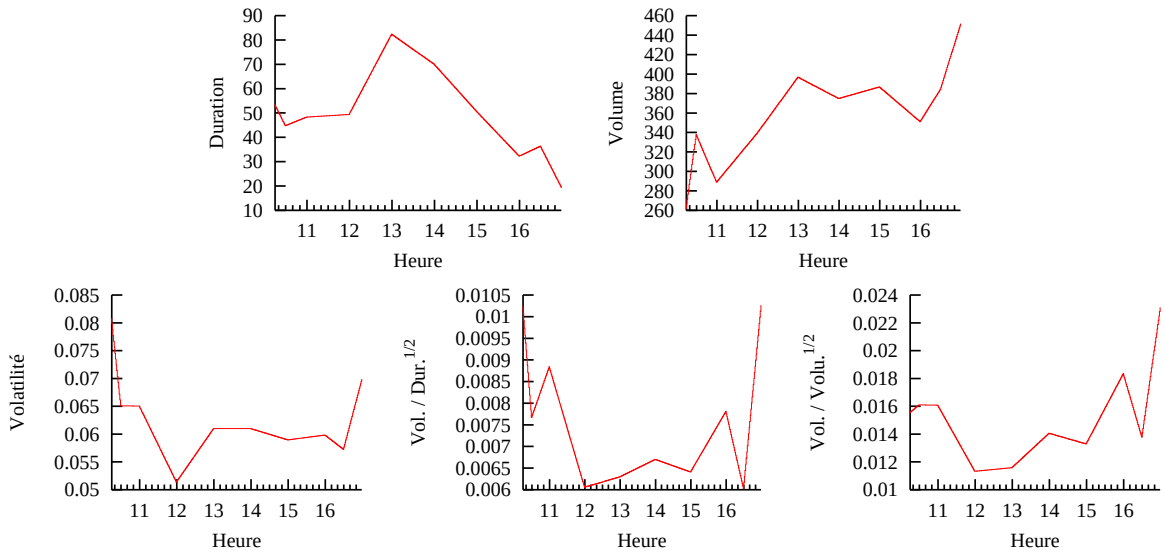
Annexe F

Dynamique du processus de transactions

F.1 Facteur saisonniers

La figure F.1 donne le facteur saisonnier calculé en supprimant toutes les transactions antérieures à 10:15.

FIGURE F.1
Facteurs saisonniers sans ouverture
Sanofi



Lorsqu'on supprime le premier $1/4$ d'heure de la journée les facteurs saisonniers théoriques de la volatilité changent de manière importante (cf. figure 4.5). Les configurations qui émergent alors sont une forte volatilité en début de journée (à l'exception de la standardisation par le volume) et en fin de journée. On peut noter un contraste important avec les configurations mises en avant par Engle (2000, cf. fig. 4 p. 16)[1]. En effet, même en supprimant la période d'ouverture,

[1]De plus, la volatilité que nous obtenons est supérieure d'un ordre de grandeur à celle mise en évidence par

la volatilité est plus forte à l'ouverture puis elle décroît au cours de la journée. Cela suggère que les différences de microstructure (e.g. l'ouverture par un fixage) affectent les caractéristiques intrajournalières des séries. Les différences entre les trois mesures de rendements soulignent qu'une standardisation par la duration ou le volume modifie ou même crée des configurations intrajournalières de volatilité. Une partie de la saisonnalité intrajournalière de la volatilité standardisée est due à la saisonnalité de la variable utilisée la standardisation (la duration ou le volume).

F.2 Modélisations univariées

Pour que le lecteur soit en mesure d'apprécier les résultats présentés dans les sous-sections 5.3.1 et 5.3.2, nous exposons ici les résultat d'un GARCH sur les rendements standardisés. Ces modèles sont les UHF-GARCH les plus simples. La différence avec les modèles (5.7) et (5.9) est qu'aucune variable n'est incluse dans la volatilité conditionnelle. On fait donc à chaque fois l'hypothèse que la variable n'a pas de contenu informationnel pour la prévision de la volatilité.

F.2.1 Duration

La table F.1 présente les estimations d'un UHF-GARCH sur les rendements standardisés par les durations.

Nous n'avons présenté ici que les résultats d'un UHF-GARCH(1,1). Le choix d'un modèle optimal par le test du ratio de vraisemblance conduit le plus souvent à cette modélisation.[2]. L'ordre des modèles mis en évidence sur les rendements standardisés (1 à 3) est donc plus faible que sur les rendements non standardisés (présentés dans la table 5.2). On constate également que les statistiques de Ljung-Box montrent que la corrélation des résidus et résidus au carré est plus faible pour les rendements standardisés par la duration.

Il est usuel d'observer une chute de la persistance de la volatilité sur les données intrajournalières. Or, toutes les estimations mettent en évidence une très forte persistance puisque la somme $\alpha_1 + \beta_1$ n'est inférieure à 1 que pour St-Gobain Accor et Lyonnaise. Ces résultats contrastent vivement avec ceux obtenus par Engle (2000) sur IBM qui observait une persistance de l'ordre 0,80.

Engle (2000).

[2]Les titres qui font exception sont St-Gobain et Schneider (UHF-GARCH(2,2)) et Lyonnaise (UHF-GARCH(3,3)).

TABLE F.1
UHF-GARCH(1,1)
duration

	St-Gobain	Schneider	Sanofi	Thomson	Accor	Lyonnaise	Lagardère
α_0	0,3491***	0,5979***	0,3209***	0,6173***	0,4961*	0,6203***	0,6956***
	(4,28)	(2,93)	(4,65)	(3,45)	(1,68)	(3,01)	(3,08)
α_1	0,1352***	0,2274***	0,1833***	0,7265***	0,1586**	0,1712***	0,6213***
	(4,06)	(3,67)	(4,51)	(2,81)	(2,05)	(2,82)	(3,30)
β_1	0,8508***	0,7640***	0,8297***	0,6609***	0,8220***	0,7856***	0,6509***
	(29,55)	(13,20)	(30,20)	(9,80)	(9,84)	(13,15)	(8,55)
Log Likelihood	-152 603	-143 957	-95 499	-64 430	-133 407	-158 375	-110 019
Ljung Box (20)	146***	114***	242***	26	129***	100***	59***
Ljung Box^2 (20)	59***	14	35**	11	7	57***	12

TABLE F.2
UHF-GARCH(1,1)
volume

	St-Gobain	Schneider	Sanofi	Thomson	Accor	Lyonnaise	Lagardère
α_0	0,4005*	0,2158*	0,0781*	0,1868*	0,2870***	0,2740**	0,1832
	(1,92)	(1,71)	(1,91)	(1,83)	(3,70)	(2,42)	(0,77)
α_1	0,0387**	0,0548**	0,0250***	0,1312**	0,0897***	0,0345***	0,1041
	(2,11)	(2,13)	(2,66)	(2,22)	(4,23)	(3,35)	(1,00)
β_1	0,9181***	0,9368***	0,9686***	0,9205***	0,8946***	0,9378***	0,9216***
	(23,77)	(32,75)	(80,77)	(37,94)	(41,10)	(45,50)	(12,79)
Log Likelihood	-163 500	-155 705	-101 792	-69 982	-135 567	-169 552	-117 888
Jarque-Béra	44 575 085***	60 994 266***	17 676 373***	128 253 339***	10 111 953***	31 880 716***	142 179 815***
Ljung Box (20)	18	32**	27	27	19	12	30*
Ljung Box^2 (20)	9	8	6	1	37**	11	2

Ces tables présentent les estimations d'un GARCH(1,1) sur les rendements normalisés par la variables indiquée. Un paramètre ou une statistique significatif au seuil de 1% est indiqué par ***, au seuil de 5% par ** et, au seuil de 10 % par *.

F.2.2 Volume

La table F.2 présente les estimations d'un UHF-GARCH sur les rendements standardisés par le volume.

Nous n'avons présenté ici que les résultats concernant les UHF-GARCH(1,1). Le choix d'un ordre optimal pour le modèle conduit le plus souvent à cette modélisation. Les seuls titres qui font exception sont St-Gobain et Schneider (UHF-GARCH(2,2))[3] et Largardère (UHF-GARCH(3,3). On observe de grandes similarités entre les résultats obtenus avec des rendements standardisés par les volumes et standardisés par les durations.

À la différence de la duration, on obtient toujours (à l'exception de Thomson) une somme $\alpha_1 + \beta_1$ inférieure à 1. On peut observer que le paramètre α_1 est souvent très faible même s'il est significatif. À l'inverse le paramètre β_1 est très important. La standardisation des rendements par les volume conduit donc à une modification de la dynamique des rendements.

Si on regarde les statistiques de Ljung-Box, les résultats semblent trop bons. En effet, ces statistiques ne sont jamais significatives au seuil de 1%. Ces modèles semblent donc capturer parfaitement la relation temporelle de dépendance des rendements...

F.3 Modélisations multivariées

La table F.3 donne la probabilité d'un type de transaction conditionnellement au type de la transaction précédente. Le type de la transaction est déterminé en classant chacune des variables (duration, volume, volatilité) selon qu'elle est supérieure ou inférieure à sa médiane. Par exemple une transaction de type 1 (8) a une volatilité, un volume et une duration inférieurs (supérieurs) à leur médiane. Le nombre total de types est alors 8 ($2 \times 2 \times 2$).

La table F.3 met en valeur un net effet diagonal. À l'exception des transactions de type 5[4], la probabilité de transition est toujours supérieure entre deux transactions de même type. Pour des raisons de concision la matrice de transition représentée est fondée sur une classification binaire des caractéristiques des transactions. L'effet diagonal que nous avons signalé est globalement robuste à la méthode de classification. Ainsi, quand on détermine les types selon les terciles des variables (on a alors 27 types de transactions) l'élément diagonal est dans 12 cas la plus forte valeur de la colonne et dans 8 cas la deuxième[5]. Cela prouve qu'au delà d'une similarité des processus de duration, de volume et de volatilité, ces trois variables évoluent de concert.

[3] St-Gobain et Schneider se distinguent donc. En effets seuls ces deux titres requièrent pour les deux formes de standardisation des modèles d'ordres supérieurs à 1.

[4] Pour ce type l'élément diagonal est la seconde valeur de la colonne.

[5] De même si les transactions sont classifiées par quartiles et qu'on regroupe les deux quartiles centraux (on a alors également 27 types), alors l'élément diagonal est dans 13 cas la plus grande valeur de chaque colonne, et dans 5 la deuxième plus grande valeur. Des résultats semblables sont obtenus pour les autres titres analysés.

TABLE F.3
Matrice de transition des types de transactions
Sanofi

$Type_{t-1}$				$Type_t$							
$\lvert r_{t-1}\rvert$	V_{t-1}	d_{t-1}	T_{t-1}	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8
0	0	0	1	**21,5**	11,8	16,2	8,4	9,4	12,3	9,3	11,1
0	0	1	2	*17,3*	**17,5**	10,3	*13,7*	5,4	15,9	5,6	14,2
0	1	0	3	15,5	7,7	**23,1**	8,5	9,3	10,8	*13,1*	12,0
0	1	1	4	12,1	12,0	15,0	**12,5**	7,9	14,9	9,4	16,2
1	0	0	5	16,9	8,3	13,3	5,6	*11,4*	16,7	12,9	15,0
1	0	1	6	13,6	*14,1*	10,4	9,9	7,7	**19,4**	8,0	*16,9*
1	1	0	7	11,7	6,2	*16,6*	6,2	**12,1**	15,0	**17,4**	14,8
1	1	1	8	10,1	8,9	13,6	9,6	10,2	*16,9*	12,4	**18,3**

Cette table rapporte, pour Sanofi, la probabilité de chaque type de transaction conditionnellement au type de la transaction précédente. Les trois premières colonnes donnent les caractéristiques de l'état (0 désigne une de la variable inférieure à sa médiane et 1 une valeur supérieure). Chaque colonne $T_.$ correspond au type de la transaction en t. Chaque ligne peut être vue comme un vecteur de probabilité sommant à 100%. Pour plus de lisibilité, la première et la deuxième plus grande valeur de chaque colonne sont respectivement en gras et en italique.

Cependant, le fait que l'effet diagonal diminue avec le nombre de types considéré montre que cette relation est affectée de bruits importants.

Les tables F.4 à F.6 donnent les résultats de l'estimation du modèle (5.13) (Manganelli, 2002) pour les différents titres. Pour la duration et le volume la seule différence entre les modélisations univariées (présentées dans la sous-section 5.2.2) et les modélisations multivariées est que, dans ces dernières, des variables sont incluses dans l'équation de la volatilité conditionnelle. Ces deux modélisations sont donc directement comparables.

Le modèle univarié ACD(1,1) peut être vu comme une version contrainte du modèle multivarié où les paramètres de $v_t - 1$ et μ_{t-1}^2 sont contraints à 0. Un test du ratio de vraisemblance est alors distribué selon une loi du χ^2 à 2 degrés de liberté. Ce test rejette systématiquement le modèle multivarié au profit du modèle univarié les critères d'informations AIC (Akaike) et SIC (Schwartz) conduisent également au rejet du modèle multivarié. On doit cependant remarquer que, pour un même ordre, la statistique de Ljung-Box est systématiquement plus faible pour le modèle multivarié (Manganelli (2002) met en avant un résultat similaire). Cette même statistique sur les résidus au carré produit pour les deux modélisations des résultats très proches. On doit donc conclure que le modèle multivarié parvient moins bien que le modèle univarié à expliquer la duration mais est cependant plus performant pour expliquer la corrélation de la série.

On remarque que l'inclusion de variables exogènes ne modifie pas de manière sensible la dynamique du processus de duration (les valeurs de α_1 et β_1 sont très proches dans les deux

TABLE F.4
Modèle ACD(1,1) trivarié

	St-Gobain	Schneider	Sanofi	Thomson	Accor	Lyonnaise	Lagardère
α_0	0,0059***	0,0074***	0,0244***	0,0059***	0,0132***	0,0129***	0,0082***
	(6,79)	(6,15)	(3,31)	(3,56)	(5,89)	(4,97)	(5,47)
α_1	0,0427***	0,0578***	0,0737***	0,0676***	0,0539***	0,0408***	0,0529***
	(15,65)	(13,87)	(7,58)	(11,24)	(12,21)	(10,58)	(13,18)
β_1	0,9524***	0,9354***	0,8978***	0,9250***	0,9336***	0,9460***	0,9393***
	(293,01)	(184,07)	(52,23)	(126,64)	(149,14)	(152,37)	(183,28)
ν_{t-1}	-0,0014***	-0,0004	0,0032**	0,0016*	-0,0017***	-0,0007	-0,0004
	(-4,87)	(-1,22)	(2,39)	(1,72)	(-3,60)	(-1,47)	(-1,01)
μ^2_{t-1}	0,0002***	0,0001***	0,0005***	0,0003***	0,0004***	0,0003***	0,0001***
	(3,86)	(2,82)	(3,05)	(3,37)	(4,64)	(4,21)	(2,75)
Log Likelihood	-95 638	-90 194	-61 058	-39 770	-82 425	-98 674	-68 597
Ljung Box (20)	297***	343***	142***	175***	227***	228***	212***
Ljung Box^2 (20)	57***	84***	33**	27	49***	67***	40***

Cette table donne les paramètres de l'équation de la duration du modèle (5.11) estimés par la méthode du pseudo maximum de vraisemblance. Les valeurs entre parenthèses en dessous des paramètres sont les T-values robustes. La statistique de Ljung-Box (Ljung-Box^2) est calculée sur les durations standardisées (le carré des durations standardisées) avec 20 décalages. Les paramètres ou les statistiques significatifs au seuil de 1% sont signalées par ***, au seuil de 5% par ** et au seuil de 10% par *.

modèles).

Pour le modèle ACV(1,1), on peut faire les mêmes comparaisons entre le modèle multivarié et le modèle univarié que pour le modèle ACD(1,1). Elles conduisent aux mêmes conclusions : une préférence pour la modélisation univariée même si le modèle multivarié explique mieux la corrélation des séries.

Le modèle multivarié présenté dans la table F.6 n'est pas directement comparable avec le modèle univarié présenté dans la table 5.2. En effet dans le dernier cas les rendements sont filtrés par un ARMA(1,1) alors que dans le second cas on filtre également avec les variables d_t, d_{t-1}, v_t et v_{t-1}.

Si on tente tout de même une comparaison des modélisations univariées et multivariées on observe que les statistiques de Ljung-Box sur les résidus et les résidus au carré sont à chaque fois inférieures dans les modèles multivariés. On remarque également que la persistance $(\alpha_1 + \beta_1)$ est toujours inférieure à 1 dans le modèle multivarié.

F.4 Fonction de réponse

La figure F.2 donne pour St-Gobain la fonction de réponse théorique de la duration, du volume et de la volatilité calculée en utilisant le modèle de Manganelli (2002).

À l'inverse des résultats obtenus pour Sanofi, on doit noter la diversité de longueur des effets. Ainsi, la volatilité a sur elle même un effet limité aux 10 premières transactions alors que l'effet de la duration sur la volatilité persiste environ jusqu'à la 50^{e} transaction.

Le fait le plus important est que St-Gobain confirme le résultat déjà observé sur Sanofi. La duration a un effet décalé, significatif et positif sur la volatilité.

TABLE F.5
Modèle ACV(1,1) trivarié

	St-Gobain	Schneider	Sanofi	Thomson	Accor	Lyonnaise	Lagardère
α_0	0,0211***	0,0191***	0,0259***	0,0199***	0,0139***	0,0266***	0,0188***
	(3,73)	(3,52)	(2,90)	(2,86)	(4,03)	(4,76)	(3,35)
α_1	0,0317***	0,0384***	0,0441***	0,0463***	0,0327***	0,0308***	0,0359***
	(7,50)	(7,07)	(6,69)	(6,48)	(8,25)	(9,53)	(6,27)
β_1	0,9525***	0,9438***	0,9310***	0,9312***	0,9558***	0,9463***	0,9482***
	(113,41)	(94,01)	(66,24)	(70,39)	(150,69)	(125,54)	(96,09)
d_t	-0,0165***	-0,0143***	0,0061	0,0187**	-0,0106**	-0,0259***	-0,0254***
	(-3,66)	(-2,61)	(1,09)	(2,27)	(-2,13)	(-6,72)	(-5,91)
d_{t-1}	0,0124***	0,0146***	-0,0063	-0,0153*	0,0094*	0,0230***	0,0233***
	(2,70)	(2,59)	(-1,13)	(-1,86)	(1,89)	(5,53)	(5,29)
μ^2_{t-1}	-0,0002***	-0,0002***	-0,0001	-0,0000	-0,0002***	-0,0001*	-0,0001***
	(-2,63)	(-3,02)	(-0,81)	(-0,42)	(-5,12)	(-1,91)	(-3,34)
Log Likelihood	-95 638	-90 194	-61 053	-39 772	-82 423	-98 671	-68 586
Ljung Box (20)	93***	53***	46***	47***	80***	55***	49***
Ljung Box^2 (20)	3	35**	8	10	105***	5	5

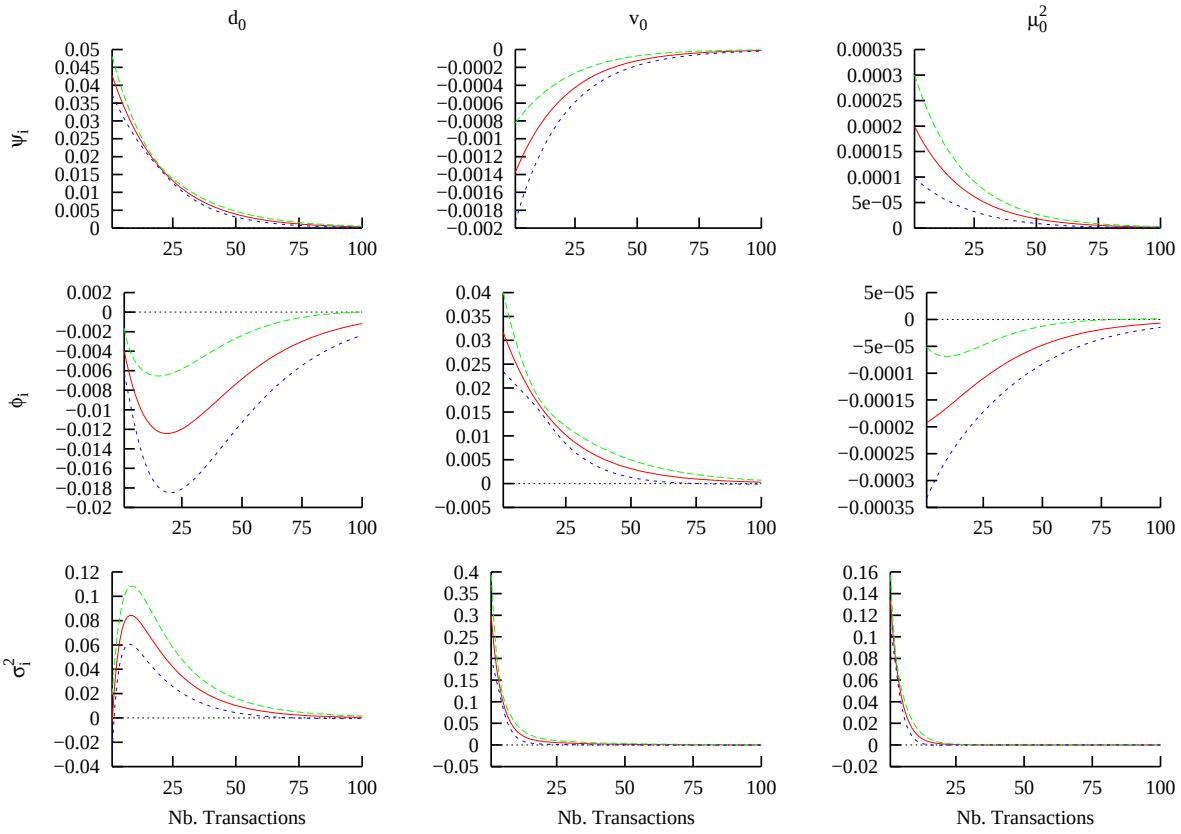
Cette table donne les paramètres de l'équation du volume du modèle (5.11) estimés par la méthode du pseudo maximum de vraisemblance. Les valeurs entre parenthèses en dessous des paramètres sont les T-values robustes. La statistique de Ljung-Box (Ljung-Box^2) est calculée sur les volume (le carré des volume standardisées) avec 20 décalages. Les paramètres ou les statistiques significatifs au seuil de 1% sont signalées par ***, au seuil de 5% par ** et au seuil de 10% par *.

TABLE F.6
Modèle GARCH(1,1) trivarié

	St-Gobain	Schneider	Sanofi	Thomson	Accor	Lyonnaise	Lagardère
α_0	0,0632*	0,0049	0,0617***	0,0044	0,0493**	0,0000	0,0027
	(1,94)	(0,33)	(3,33)	(0,10)	(2,53)	(0,00)	(0,25)
α_1	0,1337***	0,1342***	0,1328***	0,1613***	0,1290***	0,1186***	0,2078***
	(11,08)	(9,42)	(12,34)	(4,09)	(10,27)	(10,05)	(9,11)
β_1	0,7611***	0,8119***	0,8136***	0,8242***	0,8185***	0,8033***	0,7820***
	(24,17)	(33,10)	(42,02)	(19,12)	(36,70)	(28,17)	(29,87)
d_t	0,3616***	0,5584***	0,3551***	0,5350***	0,3735***	0,5628***	0,2837***
	(9,96)	(7,53)	(7,87)	(5,49)	(9,60)	(8,50)	(6,20)
d_{t-1}	-0,2940***	-0,4690***	-0,3311***	-0,4629***	-0,3184***	-0,4662***	-0,2206***
	(-9,79)	(-7,49)	(-8,76)	(-5,51)	(-8,81)	(-10,87)	(-6,07)
v_t	0,0658***	0,1084***	0,0050	0,1671**	0,0328*	0,0757***	0,1765***
	(3,47)	(5,06)	(0,32)	(2,57)	(1,70)	(2,64)	(5,70)
v_{t-1}	0,2472***	0,0822**	0,1284***	-0,0643	0,1107***	0,1809***	-0,0257
	(4,84)	(2,02)	(4,32)	(-0,90)	(3,40)	(4,18)	(-0,57)
Log Likelihood	-142 885	-138 110	-89 645	-62 240	-123 652	-148 491	-107 181
Jarque-Béra	271 482***	580 766***	170 829***	527 316***	300 185***	355 988***	675 757***
Ljung Box (20)	204***	154***	329***	232***	285***	138***	174***
Ljung Box^2 (20)	59***	35**	40***	6	26	29*	27

Cette table donne les paramètres de l'équation de la volatilité du modèle (5.11). Les valeurs entre parenthèses en dessous des paramètres sont les T-values robustes. La statistique de Ljung-Box (Ljung-Box^2) est calculée sur les rendements standardisées (le carré des rendements standardisées) avec 20 décalages. Les paramètres ou les statistiques significatifs au seuil de 1% sont signalées par ***, au seuil de 5% par ** et au seuil de 10% par *.

FIGURE F.2
Fonction de réponse
St-Gobain



Cette figure donne la fonction de réponse obtenue pour Sanofi. Chaque graphique donne l'effet d'une variable reprise en haut de chaque colonne sur la variable donnée en début de chaque ligne. Cette figure est donc l'analogue de la table 5.9 où chaque graphique donne l'évolution d'une cellule de la table. La ligne rouge est la réponse moyenne alors que les lignes hachurées délimitent un intervalle de confiance à 95%.

Annexe G

Relations carnet - transactions

Dans la section 6.1 nous avons définit une méthode d'agrégation des différentes mesures du flux informationnel par analyse en composante principale. Nous analysons ici les propriétés des 2 premiers axes factoriels de chaque dimensions.

La section G.1 donne les statistiques descriptives des 2 premiers axes de chaque dimensions. Par définition, des axes factoriels sont orthogonaux (i.e. leur corrélation dans l'espace des variables est nulle) toutefois différents types de relations non linéaires peuvent exister entre des variables non corrélées. La section G.2 s'attachera à éclairer les relations entre les différentes mesures d'une même dimension, d'une part, et, d'autre part, entre les différentes dimensions.

G.1 Statistiques descriptives du carnet

Nous analysons ici les statistiques descriptives des 2 premiers axes factoriels de chaque dimension données dans les tables G.1 à G.3.

Dans ces différentes tables la moyenne a été omise étant donné qu'une composante principale a, par définition, une moyenne nulle et une variance proportionnelle à la contribution de l'axe. Nous avons également omis les statistiques autocorrélation de Ljung-Box. Ces dernières, de même que les coefficients d'autocorrélation (ρ et ρ^2 présentés dans les tables) sont toujours très significatifs (plus de 1%). Ces résultats prouvent donc l'existence d'une autocorrélation significative des différentes variables informationnelles. On doit noter que toutes les statistiques de corrélation sur les carrés sont inférieures à celle sur les valeurs brutes.

D'une manière générale, les distributions des 2 premiers axes factoriels sont globalement de la même forme pour les différents titres. On note cependant que d'importantes variations de valeurs des paramètres comme le min, le max et de plus rares pour la skewness ou encore le kurtosis.

Le premier axe a une skewness particulièrement faible et une kurtosis élevée, de plus le

mode et la médiane sont tous les deux supérieurs à la moyenne. On est donc en présence d'une distribution asymétrique à droite cependant l'existence de très importantes valeurs négatives induit une skewness négative. Le fait que les valeurs négatives soient beaucoup plus importantes que les valeurs positives souligne qu'il arrive que l'hétérogénéité soit très faible.

Le second axe factoriel de l'hétérogénéité a des propriétés très différentes. En effet, le skewness et le kurtosis sont proches de 0, la médiane est proche de la moyenne et les valeurs extrêmes positives et négatives sont similaires. Cette distribution est donc proche de la symétrie.

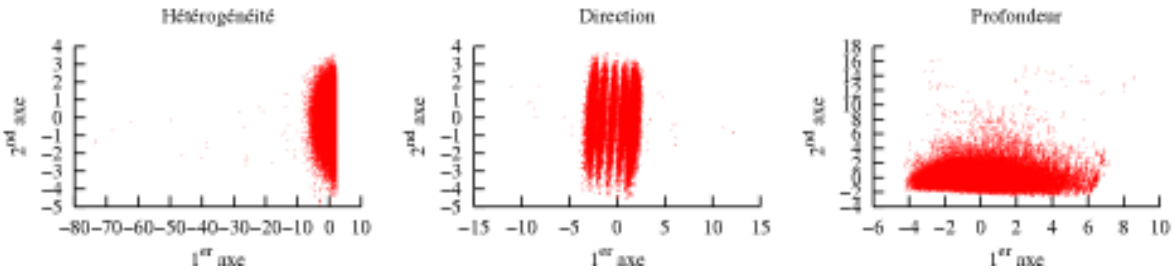
Il existe une relation entre la direction et l'hétérogénéité puisqu'on a $Dir = \pm(1 - Het)$. On retrouve donc pour la direction des résultats "symétriques" à ceux obtenus pour l'hétérogénéité. Ainsi, on observe que la distribution est asymétrique à droite (la médiane est supérieure à la moyenne) mais que les valeurs extrêmes négatives sont souvent (en valeur absolue) supérieures aux valeurs extrêmes positives.

La profondeur est de manière évidente la variable qui change le plus selon le titre puisque les valeurs de la skewness et de la kurtosis sont très variables d'un titre à l'autre. Les valeurs centrales de la distribution demeurent proches pour les différents titres mais la différence sensible des valeurs extrêmes (e.g. le minimum et le maximum) peut affecter de manière sensible des statistiques comme la skewness ou la kurtosis.

G.2 Relations entre les dimensions

La figure G.1 donne pour Sanofi la relation entre les deux premiers axes factoriels de chaque dimension.

FIGURE G.1
Relations entre les deux premiers axes factoriels
Sanofi



Chaque point de cette figure représente l'état du carnet selon le premier axe factoriel de la dimension donnée en dessous du graphique (axe *x*) et le second (axe *y*).

TABLE G.1
Statistiques descriptives de la dimension hétérogénéité

	St-Gobain		Schneider		Sanofi		Thomson		Accor		Lyonnaise		Lagardère	
Numéro	1	2	1	2	1	2	1	2	1	2	1	2	1	2
Écart-type	1,70	1,15	1,76	1,25	1,65	1,11	1,73	1,11	1,72	1,08	1,71	1,17	1,71	1,11
Skewness	-5,29	- 0,37	-22,18	- 0,54	-5,40	- 0,27	-8,12	- 0,29	-4,82	- 0,21	-5,58	- 0,35	-5,70	- 0,20
Kurtosis	137,16	0,15	1354,03	0,65	171,04	0,04	314,23	0,03	106,43	- 0,01	113,60	0,06	125,69	- 0,05
Min	-77,89	-5,04	-140,45	-13,64	-73,66	-4,92	-94,34	-5,12	-65,99	-5,33	-51,21	-5,16	-57,64	-4,99
Q1	-0,78	-0,73	-0,64	-0,76	-0,79	-0,72	-0,85	-0,71	-0,79	-0,71	-0,75	-0,74	-0,77	-0,73
Médiane	0,37	0,07	0,34	0,12	0,35	0,05	0,38	0,06	0,39	0,04	0,33	0,07	0,39	0,04
Mode	0,53	0,25	-1,66	-0,59	0,11	-0,11	0,15	-0,89	1,19	-0,96	-0,74	-1,60	1,29	1,99
Q3	1,14	0,81	0,98	0,89	1,13	0,78	1,20	0,78	1,17	0,76	1,11	0,82	1,16	0,78
Max	2,19	3,62	2,14	3,84	2,41	3,65	2,41	3,83	2,36	4,69	2,14	3,74	2,31	3,75
ρ_1	0,79	0,79	0,78	0,84	0,77	0,72	0,76	0,71	0,76	0,73	0,82	0,81	0,77	0,72
ρ_2	0,66	0,69	0,69	0,76	0,64	0,62	0,60	0,62	0,63	0,61	0,71	0,72	0,63	0,61
ρ_1^2	0,58	0,74	0,66	0,77	0,50	0,63	0,51	0,61	0,38	0,62	0,73	0,75	0,57	0,60
ρ_2^2	0,39	0,62	0,57	0,67	0,30	0,51	0,07	0,50	0,17	0,47	0,61	0,63	0,32	0,46

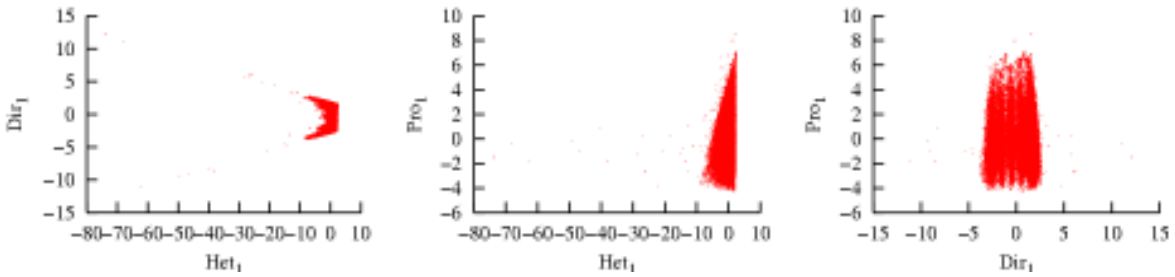
TABLE G.2
Statistiques descriptives de la dimension direction

Numéro	1	2	1	2	1	2	1	2	1	2	1	2	1	2
Écart-type	1,70	1,32	1,78	1,41	1,65	1,26	1,69	1,24	1,69	1,21	1,71	1,33	1,68	1,23
Skewness	-0,25	-0,15	-0,41	0,07	-0,40	-0,23	-0,29	-0,30	-0,33	-0,21	-0,72	-0,11	-0,25	-0,23
Kurtosis	-1,16	-0,28	0,02	0,77	-1,13	-0,21	-0,85	-0,15	-1,22	-0,31	-0,56	-0,27	-1,22	-0,24
Min	-10,90	-4,52	-22,34	-8,88	-11,17	-4,59	-18,30	-5,09	-11,91	-4,31	-10,50	-4,66	-11,43	-4,49
Q1	-1,55	-0,90	-1,44	-0,97	-1,52	-0,85	-1,55	-0,83	-1,56	-0,83	-1,28	-0,91	-1,55	-0,82
Médiane	0,24	0,04	0,26	0,02	0,45	0,05	0,42	0,07	0,41	0,05	0,58	0,03	0,40	0,05
Mode	-1,42	-0,57	-1,29	-1,60	-2,63	-0,77	1,90	1,05	1,36	0,63	0,37	-1,75	-0,20	0,07
Q3	1,53	0,94	1,49	0,97	1,51	0,91	1,60	0,89	1,58	0,88	1,37	0,94	1,64	0,88
Max	8,71	6,66	13,10	17,93	12,14	3,66	11,08	3,25	9,80	3,46	8,35	4,58	9,46	3,51
ρ_1	0,87	0,89	0,90	0,89	0,84	0,88	0,87	0,84	0,86	0,85	0,87	0,89	0,86	0,87
ρ_2	0,78	0,82	0,84	0,82	0,74	0,81	0,78	0,77	0,77	0,76	0,79	0,83	0,77	0,80
ρ_1^2	0,76	0,79	0,76	0,79	0,69	0,78	0,63	0,75	0,70	0,74	0,81	0,81	0,70	0,77
ρ_2^2	0,62	0,68	0,67	0,67	0,52	0,68	0,35	0,63	0,54	0,61	0,69	0,69	0,52	0,65

Le comportement des deux premiers axes de l'hétérogénéité est visiblement différent. Le premier axe est très compact une grande partie de la distribution étant concentrée entre 2 et -10 alors qu'on observe des valeurs extrêmes négatives. Le second axe est lui aussi compact et légèrement asymétrique puisque les valeurs négatives sont plus dispersées mais aussi plus importantes que les valeurs positives. La direction fait apparaître une figure de type rectangle qui suggèrent que l'on trouve toutes les combinaisons entre la direction prix et la direction activité, les indications données par ces deux mesures peuvent ne pas être convergentes. La profondeur ne fait pas ressortir d'association manifeste entre la profondeur prix et la profondeur volume.

La figure G.2 (G.3) donne la relation entre les premiers (seconds) axes factoriels des différentes dimensions.

FIGURE G.2
Relations entre les dimensions : 1^{er} axe factoriel
Sanofi



Chaque point de cette figure représente l'état du carnet selon le premier axe factoriel de la dimension donnée en dessous de l'axe *x* et de celle à gauche de l'axe *y*.

Le graphique du premier axe factoriel de la direction en fonction de l'hétérogénéité fait clairement ressortir la relation de type valeur absolue entre ces deux variables. On peut cependant noter la dispersion des variables autour de cette relation structurelle.

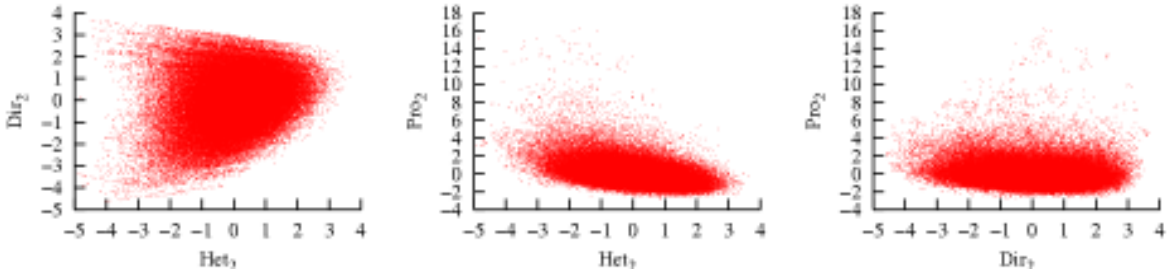
Les graphiques de la profondeur en fonction de l'hétérogénéité ou de la direction mettent en évidence une relation positive entre la profondeur et l'hétérogénéité et une relation négative entre profondeur et direction. Ainsi, un intérêt global plus important semble résulter plus d'une absence d'information (hétérogénéité élevée) que d'une asymétrie d'information (indication directionnelle faible).

La présentation des 2^{nd} axes factoriels (globalement les mesures d'activité) montre que les relations entre ceux-ci sont différentes des celles observées entre les fourchettes. Ainsi, la relation entre direction et hétérogénéité est beaucoup plus bruitée. La relation entre profondeur et hétérogénéité suggère une association négative entre l'hétérogénéité et la profondeur. Donc, il y a une relation positive entre profondeur prix et hétérogénéité prix (1^{er} axes factoriels) et une relation négative entre profondeur activité et hétérogénéité activité (2^{nd} axe factoriel). Cela

TABLE G.3
Statistiques descriptives de la dimension profondeur

	St-Gobain		Schneider		Sanofi		Thomson		Accor		Lyonnaise		Lagardère	
Numéro	1	2	1	2	1	2	1	2	1	2	1	2	1	2
Écart-type	1,89	1,36	1,93	1,53	1,91	1,27	1,88	1,32	1,90	1,31	1,92	1,37	1,92	1,26
Skewness	0,30	2,64	2,54	3,63	0,66	2,23	1,37	4,53	0,72	3,26	0,26	2,99	1,16	7,06
Kurtosis	-0,33	13,35	9,25	19,84	0,26	12,06	3,09	40,02	0,40	19,15	-0,42	16,75	1,42	93,07
Min	-5,42	-2,21	-4,55	-2,03	-4,28	-2,81	-3,03	-2,70	-4,15	-2,27	-5,38	-2,74	-3,93	-1,50
Q1	-1,41	-0,85	-1,10	-0,85	-1,42	-0,81	-1,38	-0,73	-1,40	-0,77	-1,42	-0,81	-1,39	-0,63
Médiane	-0,13	-0,31	-0,44	-0,36	-0,24	-0,22	-0,41	-0,26	-0,27	-0,27	-0,15	-0,27	-0,45	-0,27
Mode	0,66	-1,17	-1,46	-1,66	-2,86	-1,52	-2,19	-1,11	2,56	0,01	-2,11	-0,91	-1,46	-1,10
Q3	1,25	0,43	0,37	0,33	1,16	0,54	0,98	0,37	1,15	0,40	1,31	0,44	0,97	0,26
Max	5,36	18,46	17,89	17,26	8,57	16,18	13,26	18,42	7,30	20,43	6,81	18,08	9,39	30,08
ρ_1	0,94	0,95	0,97	0,97	0,95	0,92	0,96	0,93	0,95	0,94	0,95	0,95	0,97	0,90
ρ_2	0,89	0,91	0,95	0,95	0,91	0,86	0,93	0,89	0,91	0,90	0,91	0,91	0,95	0,82
ρ_1^2	0,86	0,92	0,93	0,95	0,88	0,90	0,91	0,94	0,89	0,92	0,89	0,95	0,92	0,78
ρ_2^2	0,76	0,86	0,88	0,92	0,80	0,84	0,83	0,91	0,81	0,88	0,82	0,90	0,87	0,66

FIGURE G.3
Relations entre les dimensions : 2^{nd} axe factoriel
Sanofi



Chaque point de cette figure représente l'état du carnet selon le second axe factoriel de la dimension donnée en dessous de l'axe *x* et de celle à gauche de l'axe *y*.

suggère donc que les mesures sur la fourchette et celle sur l'activité n'ont pas le même contenu informationnel. La relation entre la profondeur activité et l'indication directionnelle d'activité fait également apparaître une association entre les indications directionnelles importantes et une forte profondeur. Cette relation semble asymétrique[1] une indication directionnelle positive ayant un effet plus important qu'une indication négative.

L'analyse de l'activité suggère que, à la différence des fourchettes, une forte profondeur activité est plus associé à une asymétrie d'information qu'à l'hétérogénéité. Ces constatations supporte le modèle proposé par Dupont (2000) où la profondeur cotée est plus sensible à l'asymétrie d'information que la fourchette. L'analyse des relations entre les différentes mesures informationnelles souligne d'une part qu'elles présentent entre elles des relations compatibles avec les concepts qu'elles sont sensées mesurer et d'autre part que le contenu informationnel de l'activité et des fourchettes est probablement différent.

G.3 Modélisations GARCH des variables informationnelles

Cette section détaille dans les tables G.4 à G.9 les résultats de l'estimation d'un GARCH sur les différentes variables informationnelles.

De manière générale, les modèles de type GARCH se révèlent adaptés aux différentes variables. En effet, pour chaque variable, les paramètres ARCH et éventuellement les paramètres GARCH sont significatifs.

Les dynamiques mises en évidence sont toutefois différentes de celles déjà observées pour les caractéristiques des transactions. Pour les transactions on observe souvent un paramètre

[1] On observe en effet un plus grand nombre de cas traduisant une forte indication directionnelle positive et une forte profondeur.

GARCH (β_1) élevé alors que le paramètre ARCH (α_1) est souvent faible. Nous obtenons souvent l'inverse pour les variables informationnelles : un paramètre α_1 important et très significatif et un paramètre β_1 faible et/ou non significatif. La dynamique des variables du carnet est donc plus proche d'un ARCH que d'un GARCH.

Les valeurs des paramètres pour une même variable sont très proches pour les différents titres à l'exception de l'hétérogénéité prix de Schneider. Ce titre a pour cette variable des paramètres complètement différents des autres titres et en plus, la somme $\alpha_1 + \beta_1$ est supérieure à 1.

TABLE G.4
Hétérogénéité prix

	St-Gobain	Schneider	Sanofi	Thomson	Accor	Lyonnaise	Lagardère
α_0	1,0168***	0,0025	1,3807***	0,4653**	0,5993***	0,6310***	0,9311**
	(4,47)	(0,18)	(3,49)	(2,24)	(2,61)	(5,56)	(2,45)
α_1	0,6525***	0,0810***	0,5226***	0,4734***	0,5526***	0,6911***	0,4548***
	(6,89)	(3,75)	(8,63)	(4,74)	(4,97)	(8,72)	(5,67)
β_1	0,1022	0,9473***	0,0763	0,5145***	0,3794**	0,1776**	0,2895
	(1,10)	(66,82)	(1,02)	(2,89)	(2,49)	(2,37)	(1,55)
Log Likelihood	-133797	-132176	-85937	-59433	-119613	-141119	-97918
Jarque-Béra	169303376***	3051913839***	294834933***	539049956***	241457030***	86757043***	154596826***
Ljung Box (20)	25797***	56097***	12877***	7490***	19715***	30838***	19146***
Ljung Box^2 (20)	1	0	0	0	0	1	0

TABLE G.5
Hétérogénéité activité

	St-Gobain	Schneider	Sanofi	Thomson	Accor	Lyonnaise	Lagardère
α_0	0,3216***	0,2931***	0,3501***	0,3048***	0,3581***	0,2957***	0,3535***
	(28,54)	(15,07)	(24,27)	(15,82)	(28,23)	(29,31)	(24,27)
α_1	0,5006***	0,5150***	0,4100***	0,3970***	0,4479***	0,5118***	0,4309***
	(52,90)	(35,86)	(42,61)	(30,96)	(49,69)	(55,03)	(44,50)
β_1	0,2481***	0,2888***	0,3032***	0,3552***	0,2465***	0,2636***	0,2813***
	(14,93)	(11,62)	(15,70)	(13,35)	(13,92)	(16,97)	(14,43)
Log Likelihood	-105771	-105684	-66024	-42842	-87558	-110394	-73863
Jarque-Béra	54***	14601***	74***	66***	31***	144***	58***
Ljung Box (20)	53300***	71085***	27560***	19476***	35753***	62737***	33429***
Ljung Box^2 (20)	861***	128***	804***	451***	922***	808***	778***

Ces tables donnent les estimations d'un modèle GARCH sur l'hétérogénéité. Un paramètre ou une statistique significatif au seuil de 1% est indiqué par ***, au seuil de 5% par ** et au seuil de 10% par *.

TABLE G.6
Direction prix

	St-Gobain	Schneider	Sanofi	Thomson	Accor	Lyonnaise	Lagardère
α_0	0,7217***	0,4786***	1,0124***	0,8925***	0,8272***	0,5804***	0,8557***
	(28,42)	(19,47)	(20,95)	(18,58)	(28,97)	(20,10)	(21,44)
α_1	0,5494***	0,5486***	0,4638***	0,4769***	0,5206***	0,5317***	0,5002***
	(61,51)	(36,01)	(58,92)	(36,55)	(61,89)	(40,77)	(52,99)
β_1	0,2052***	0,3035***	0,1747***	0,2291***	0,1981***	0,2616***	0,2047***
	(12,81)	(13,32)	(8,82)	(9,41)	(12,19)	(11,86)	(10,13)
Log Likelihood	-132608	-128549	-83618	-55484	-114282	-136955	-94678
Jarque-Béra	3249***	3679***	1270***	667***	2928***	1679***	1712***
Ljung Box (20)	113306***	126755***	52884***	46356***	95095***	95262***	79398***
Ljung Box^2 (20)	707***	108***	298***	308***	724***	324***	339***

TABLE G.7
Direction activité

	St-Gobain	Schneider	Sanofi	Thomson	Accor	Lyonnaise	Lagardère
α_0	0,3453***	0,3644***	0,3195***	0,3522***	0,3589***	0,3093***	0,3083***
	(35,02)	(19,23)	(25,12)	(20,75)	(30,77)	(30,82)	(28,32)
α_1	0,6088***	0,5932***	0,5264***	0,4900***	0,5231***	0,6189***	0,5758***
	(71,25)	(51,88)	(47,32)	(39,50)	(59,99)	(64,02)	(55,26)
β_1	0,1916***	0,2265***	0,2725***	0,2776***	0,2280***	0,2089***	0,2197***
	(15,53)	(12,51)	(15,65)	(13,13)	(15,18)	(15,13)	(14,14)
Log Likelihood	-114397	-112022	-71737	-46023	-93782	-118939	-78708
Jarque-Béra	232***	60815***	356***	237***	413***	114***	224***
Ljung Box (20)	73009***	77595***	43184***	25982***	55956***	73260***	50254***
Ljung Box^2 (20)	590***	53***	545***	419***	814***	510***	498***

Ces tables donnent les estimations d'un modèle GARCH sur direction. Un paramètre ou une statistique significatif au seuil de 1% est indiqué par ***, au seuil de 5% par ** et au seuil de 10% par *.

TABLE G.8
Profondeur prix

	St-Gobain	Schneider	Sanofi	Thomson	Accor	Lyonnaise	Lagardère
α_0	0,6357***	0,0953***	0,5823***	0,2649***	0,4777***	0,5279***	0,2507***
	(37,74)	(14,00)	(27,46)	(12,27)	(30,38)	(37,03)	(23,45)
α_1	0,7440***	0,7125***	0,7328***	0,7550***	0,7790***	0,7480***	0,8487***
	(81,82)	(23,96)	(60,14)	(24,42)	(69,99)	(85,39)	(56,55)
β_1	0,0822***	0,2478***	0,1129***	0,1960***	0,0931***	0,1065***	0,0904***
	(7,28)	(7,65)	(7,58)	(5,67)	(7,25)	(9,93)	(5,71)
Log Likelihood	-139219	-133423	-90412	-58776	-121033	-144858	-101838
Jarque-Béra	1433***	1527***	1566***	1857***	1743***	1146***	867***
Ljung Box (20)	81923***	70114***	56376***	40043***	81816***	104393***	81325***
Ljung Box^2 (20)	318***	357***	164***	69***	132***	257***	188***

TABLE G.9
Profondeur activité

	St-Gobain	Schneider	Sanofi	Thomson	Accor	Lyonnaise	Lagardère
α_0	0,0755***	0,0255***	0,2141***	0,1133***	0,0507**	0,1220***	0,0959***
	(6,52)	(4,67)	(14,43)	(5,01)	(2,57)	(8,33)	(4,30)
α_1	0,5440***	0,5006***	0,6069***	0,4654***	0,4136***	0,6293***	0,5096***
	(14,20)	(8,86)	(26,72)	(10,39)	(4,64)	(16,36)	(7,12)
β_1	0,4201***	0,5024***	0,2382***	0,4488***	0,5561***	0,2741***	0,4276***
	(9,19)	(8,64)	(7,80)	(7,69)	(5,20)	(5,62)	(4,53)
Log Likelihood	-116809	-118685	-72274	-49694	-99176	-123050	-81938
Jarque-Béra	530030***	293436***	39835***	752399***	585776***	212033***	12955121***
Ljung Box (20)	88416***	104263***	37415***	25843***	68098***	76993***	43667***
Ljung Box^2 (20)	12	5	51***	6	78***	16	202***

Ces tables donnent les estimations d'un modèle GARCH sur la profondeur. Un paramètre ou une statistique significatif au seuil de 1% est indiqué par ***, au seuil de 5% par ** et au seuil de 10% par *.

Table des matières

Remerciements **i**

Résumé **xii**

Introduction générale **2**

I Déformations temporelles des rendements intrajournaliers : À la recherche du temps perdu **9**

Introduction **10**

1 Conception des séries de rendements et des déformations temporelles **15**

1.1 Aspects Méthodologiques 15

1.1.1 Sélection - Retraitements 15

a. Sélection 16

b. Organisation du marché parisien 17

c. Retraitements 20

1.1.2 Construction des déformations temporelles 21

a. Approximations du flux informationnel 21

b. Calcul des déformations 24

c. Approximations des prix 26

1.2 Analyse de l'activité 28

1.2.1 Activité journalière 28

1.2.2 Effets des déformations temporelles 33

1.3 Conclusion 37

2 Mise en évidence des configurations intrajournalières **39**
2.1 Revue de la littérature . . . 40
2.1.1 Des configurations convergentes mais différentes . . . 41
2.1.2 Causes des configurations . . . 45
a. Structure de marché et mécanismes d'échange . . . 45
b. Flux informationnel et interventions des informés . . . 47
2.2 Méthodologie d'analyse . . . 50
2.2.1 Hypothèse sur les configurations . . . 51
2.3 Groupe HL . . . 53
2.3.1 Configurations en temps calendaire . . . 53
2.3.2 Configurations en temps activité . . . 62
2.4 Influence du groupe de liquidité . . . 64
2.5 Conclusion . . . 68

3 Propriétés des séries intrajournalières **70**
3.1 Propriétés et méthodologies d'analyse . . . 71
3.1.1 Propriétés des rendements HF. . . . 71
3.1.2 Méthodologie d'analyse . . . 74
3.2 Groupe HL . . . 77
3.2.1 Approximations du flux informationnel . . . 77
3.2.2 Distribution des rendements . . . 80
3.2.3 Dynamique des rendements . . . 84
3.3 Influence du groupe de liquidité . . . 85
3.4 Conclusion . . . 87

Conclusion de la partie I **91**

II Analyse "trade by trade" : Le temps retrouvé **94**

Introduction **95**

4 Construction d'approximations de l'information **97**
4.1 Modélisations de microstructure . . . 98
4.1.1 Caractéristiques des transactions . . . 100

4.1.2 Caractéristiques du carnet . . . 105
4.1.3 Marchés dirigés par les ordres . . . 108
4.2 Pré-traitements . . . 112
4.2.1 Traitement des événements simultanés . . . 113
4.2.2 L'information du carnet . . . 117
4.2.3 Désaisonnalisation . . . 126
4.3 Description des données . . . 131
4.3.1 Rendements . . . 134
4.3.2 Activité . . . 135
4.4 Conclusion . . . 136

5 Relations entre les composantes du processus de transactions 139
5.1 Méthodologies d'analyse des rendements ultra haute fréquence . . . 141
5.2 Modèles univariés . . . 145
5.2.1 Rendement . . . 145
5.2.2 Activité . . . 148
a. Duration . . . 148
b. Volume . . . 152
5.3 Modèles bivariés . . . 154
5.3.1 Duration - Rendement . . . 155
5.3.2 Volume - Rendement . . . 159
5.4 Modèle trivarié . . . 162
5.5 Conclusion . . . 171

6 Incorporation du carnet dans l'analyse du processus de transactions 174
6.1 Agrégation de l'information du carnet . . . 177
6.1.1 Sanofi . . . 177
6.1.2 Composantes principales des différents titres . . . 179
6.2 Modélisation univariée . . . 183
6.3 Modélisation multivariée . . . 196
6.3.1 Effets des transactions sur l'état du carnet . . . 197
6.3.2 Effets sur les caractéristiques des transactions . . . 200
6.3.3 Synthèse des différents titres . . . 203
6.4 Conclusion . . . 205

Conclusion de la partie II 208

Conclusion générale **212**

Annexes **217**

A La Base De Marché **217**

B Base de données **221**
B.1 Titres sélectionnables 221
B.2 Influence des retraitements 221

C Prix et calculs **225**
C.1 Comparaison des prix 225
C.2 Comparaison des modalités de calcul 225

D Configurations intrajournalières **229**
D.1 Groupe HL 229
D.2 Groupe ML 232
D.3 Groupe LL 233

E Propriétés des séries **234**
E.1 Dynamique du rendement 234
E.2 Distribution des rendements LL 235

F Dynamique du processus de transactions **237**
F.1 Facteur saisonniers 237
F.2 Modélisations univariées 238
F.2.1 Duration 238
F.2.2 Volume 240
F.3 Modélisations multivariées 240
F.4 Fonction de réponse 243

G Relations carnet - transactions **247**
G.1 Statistiques descriptives du carnet 247
G.2 Relations entre les dimensions 248
G.3 Modélisations GARCH des variables informationnelles 252

Bibliographie

Admati, A. R. & **P. Pfleiderer** (1988). ‘A Theory of Intra-Day Patterns Volume and Price Variability’, *The Review of Financial Studies* **1(1)**, 3–40. 48, 98, 101, 111, 139, 140, 172

— (1989). ‘Divide and Conquer : A Theory of Intra-Day and Day-of-the-Week Mean Effects’, *The Review of Financial Studies* **2(2)**, 189–223. 47

Ahn, H.-J., **K.-H. Bae** & **K. Chan** (2001). ‘Limit Orders, Depth, and Volatility’, *The Journal of Finance* **56(2)**, 767–788. 41, 109

Ahn, H.-J., **H. Cai**, **Y. Hamao** & **R. T. Ho** (2002). ‘The Components of the Bid-Ask Spread in a Limit-Order Market : Evidence From the Tokyo Stock Exchange’, *Journal of Empirical Finance* **9(4)**, 399–430. 49

Ahn, H.-J. & **Y.-L. Cheung** (1999). ‘The Intra-Day Patterns of the Spread and Depth in a Market Without Market Maker : The Stock Exchange of Hong Kong’, *Pacific-Basin Finance Journal* **7(5)**, 539–556. 41

Aitken, M., **A. Kua**, **P. Brown**, **T. Walter** & **H. Y. Izan** (1995). ‘An Intra-Day Analysis of the Probability of Trading on the ASX at the Asking Price’, *Australian Journal of Management* **20(2)**, 115–154. 47

Akerlof, G. A. (1970). ‘The Market for ’Lemons’ : Quality, Uncertainty, and the Market Mechanism’, *Quaterly Journal of Economics* **84(3)**, 488–500. 3

Amihud, Y. & **H. Mendelson** (1987). ‘Trading Mechanisms and Stock Returns : An Empirical Investigation’, *The Journal of Finance* **42(3)**, 533–555. 45

— (1991). ‘Volatility, Efficiency, and Trading : Evidence from the Japanese Stock Market’, *The Journal of Finance* **46(3)**, 1765–1789. 45

Amihud, Y., **H. Mendelson** & **M. Murgia** (1990). ‘Stock Market Microstructure and Return Volatility’, *Journal of Banking & Finance* **14(2)**, 423–440. 45

Andersen, T. G. (1996). 'Return Volatility and Trading Volume : An Information Flow Interpretation of Stochastic Volatility', *The Journal of Finance* **51(1)**, 169–204. 12, 13

Andersen, T. G. & **T. Bollerslev** (1997*a*). 'Heterogeneous Information Arrivals Returns and Volatility Dynamics : Uncovering the Long Run in High Frequency Returns', *The Journal of Finance* **52(3)**, 975–1005. 78, 126

— (1997*b*). 'Intraday Periodicity and Volatility Persistence in Financial Markets', *Journal of Empirical Finance* **4(2)**, 115–158. 26, 41

Andersen, T. G., **T. Bollerslev** & **J. Cai** (2000). 'Intraday and Interday Volatility in the Japanese Stock Market', *Journal of International Financial Markets, Institutions & Money* **10(2)**, 107–130. 40, 126

Andersen, T. G., **T. Bollerslev** & **A. Das** (2001). 'Variance-Ratio Statistic and High Frequency Data : Testing for Changes in Intraday Volatility Patterns', *The Journal of Finance* **56(1)**, 305–327. 27, 51

Ané, T. & **H. Geman** (2000). 'Order Flow, Transaction Clock and the Normality of Asset Returns', *The Journal of Finance* **55(5)**, 2259–2284. 76, 89

Antweiler, W. & **M. Z. Frank** (2001). 'Is All That Talk Just Noise ? The Information Content of Internet Stock Message Boards', Journal of Finance forthcoming. 2

Aubier, A. & **A. Miloudi** (2001). 'Interactions entre Volatilité des Rendements et Durée entre deux Transactions : Une Modélisation ACD-GARCH', Working Paper Université de Rennes I. 114, 128, 143

Bachelier, M. L. (1900). *Théorie de la Spéculation*, Ph.D. thesis. 10, 72

Back, K., **H. Cao** & **G. A. Willard** (2000). 'Imperfect Competition Among Informed Traders', *The Journal of Finance* **55(5)**, 2117–2156. 102

Back, K. & **H. Pedersen** (1998). 'Long-Lived Information and Intraday Patterns', *Journal of Financial Markets* **1(3–4)**, 385–402. 48, 102, 139, 140, 172

Bae, K.-H., **H. Jang** & **K. S. Park** (2003). 'Traders' Choice between Limit and Market Orders : Evidence from Nyse Stock', *Journal of Financial Markets* **Fortcomming(? ? ?)**, ? ? ? 109

Bagehot, W. (1971). 'The Only Game in Town', *Financial Management Journal* **8**. 98

Baker, M. & **J. C. Stein** (2000). 'Market Liquidity as a Sentiment Indicator', Working Paper Harvard University. 107

Ballochi, G., **M. M. Dacorogna**, **C. M. Ulrich A. Hopman**, **Müller** & **R. B. Olsen** (1997). 'The Intraday Multivariate Structure of the Eurofutures Markets', *Journal of Empirical Finance* **6(5)**, 479–513. 41

Bamber, L. S., **O. E. Baron** & **T. L. Stober** (1999). 'Differential Interpretations and Trading Volume', *Journal of Financial & Quantitative Analysis* **34(3)**, 369–386. 104

Barclay, M. J. & **J. B. Warner** (1993). 'Stealth trading and volatility Which trades move prices ?', *Journal of Financial Economics* **34(3)**, 281–305. 22, 101

Bauwens, L. & **P. Giot** (2000). 'The Logarithmic ACD Model : An Application to Market to the Bid-Ask Quote Process of Three NYSE Stocks', *Annales d'Economie et de Statistique* **60**, 117–149. 148

—, *Econometric Modeling of Stock Market Intraday Activity*, vol. 38 of *Advanced studies in theoritical and applied econometrics* (Kluwer Academic Publishers, 2001). 148

Bauwens, L. & **D. Veredas** (1999). 'The Stochastic Conditional Duration Model : A Latent Factor Model for the Analysis of Financial Duration', . 148

Berg, L. (2000). 'Deterministic Seasonal Volatility in a Small and Integrated Stock Market : The Case of Sweden', . 74

Berglund, T. & **J. Westerholm** (2002). 'Systematic Differences in Trading Patterns between Different Investor Groups', Working Paper Swedish School of Economics and Business Administration Helsinki. 40

Berry, T. D. & **K. M. Howe** (1994). 'Public Information Arrival', *The Journal of Finance* **49**, 1331–1346. 2

Biais, B. & **P. Bossaerts** (1998). 'Asset Prices and Trading Volume in a Beauty Contest', *The Review of Economic Studies* **65(2)**, 307–340. 103, 104

Biais, B., **T. Foucault** & **P. Hillion**, *Microstructure des Marché Financiers*, finance (presses universitaires de france, 1997), 1ère ed. 48

Biais, B., **P. Hillion** & **C. Spatt** (1995). 'An Empirical Analysis of the Limit Order Book and the Order Flow in the Paris Bourse', *The Journal of Finance* **50(5)**, 1655–1689. 41, 109, 149, 154, 209

— (1999). 'Price Discovery and Learning During the Preopening Period in the Paris Bourse', *The Journal of Political Economy* **107(6)**, 1218–1248. 47

Biais, B. & **S. Pouget** (1999). 'Learning to Play Equilibrium Strategies in Experimental Markets : Does Microstructure Matter ?', . 47

Biais, B. & **R. Shadur** (2000). 'Darwinian Selection Does Not Eliminate Irrational Traders', *European Economic Review* **44(3)**, 469–490. 3

Bildik, R. (2000). 'Intra-Day Seasonalities on Stock Returns : Evidence from the Turkish Stock Market', . 41, 43, 45

Bisière, C. & **T. Kamionka** (2000). 'Timing of Orders, Orders Aggressiveness and the Order Book at the Paris Bourse', *Annales d'Economie et de Statistique* **60**, 43–72. 41, 149

Blume, L., **D. Easley** & **M. O'Hara** (1994). 'Market Statistics and Technical Analysis : The Role of Volume', *The Journal of Finance* **49(1)**, 153–181. 104

Bollerslev, T. (1987). 'A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rate of Return', *The Review of Economics and Statistics* **69(3)**, 542–547. 73, 145

Bollerslev, T., **R. Y. Chou**, **N. Jayaranan** & **K. F. Kroner** (1991). 'Les Modèles ARCH en Finance : Un Point sur la Théorie et les Résultats Empiriques', *Annales d'Economie et de Statistique* **24**, 1–59. 73

Bollerslev, T. & **I. Domowitz** (1993). 'Trading Patterns and Prices in the Interbank Foreign Exchange Market', *The Journal of Finance* **48(4)**, 1421–1443. 26, 40

Bollerslev, T. & **J. M. Wooldridge** (1992). 'Quasi-Maximum Likelihood and Inference in Dynamic Models with Time Varying Covariances', *Econometric Reviews* **11(2)**, 143–172. 144, 145

Brav, A. & **J. B. Heaton** (2002). 'Competing Theories of Financial Anomalies', *The Review of Financial Studies* **15(2)**, 575–606. 3

Broccard, N. & **R. Calcagno** (1999). 'Asymmetries of Information in Centralised Order-Driven Markets', Working Paper CORE Discussion Paper n^o 9935. 47

Brock, W. A. & **A. W. Kleidon** (1992). 'Periodic Market Closure and Trading Volume a Model of Intraday Bids and Asks', *Journal of Economic Dynamics & Control* **16(3–4)**, 451–489. 41, 46

Brock, W. A. & **B. D. Lebaron** (1996). 'A Dynamic Structural Model for Stock Return Volatility and Trading Volume', *The Review of Economics and Statistics* **78(1)**, 94–110. 73

Brockman, P. & **D. Y. Chung** (1999). 'An Analysis of Depth Behavior in an Electronic Environment', *Journal of Banking & Finance* **23(12)**, 1861–1886. 41

— (2000). 'An Empirical Investigation of Trading on Asymmetric Information and Heterogeneous Prior Beliefs', *Journal of Empirical Finance* **7(5)**, 417–454. 16, 103

Brunnermeir, M. K., *Asset Pricing under Asymetric Information Bubbles, Crashes, Technical Analysis, and Herding* (Oxford University Press, 2001). 4, 73

Campbell, J. Y., **A. W. Lo** & **A. C. Mackinlay**, *The Econometrics of Financial Markets* (Princeton University Press, 1997). 27, 71

Cao, C., **E. Ghysels** & **F. Hatheway** (2000). 'Price Discovery Without Trading : Evidence from the NASDAQ Pre-Opening', *The Journal of Finance* **55(3)**, 1399–1365. 47

Chakravarty, S. (2001). 'Stealth-trading : Which traders' trades move stock prices ?', *Journal of Financial Economics* **61(2)**, 289–307. 101

Chan, K., **Y. P. Chung** & **H. Jonhson** (1995*a*). 'The Intraday Behavior of Bid-Ask Spreads for NYSE Stocks and CBOE Options', *Journal of Financial & Quantitative Analysis* **30(3)**, 329–346. 41

Chan, K. & **W.-M. Fong** (2000). 'Trade Size, Order Imbalance, and the Volatility-Volume Relation', *Journal of Financial Economics* **57(2)**, 247–273. 22

Chan, K. C., **W. G. Christie** & **P. H. Schultz** (1995*b*). 'Market Structure and the Intraday Pattern of Bid–Ask Spreads for NASDAQ Securities', *Journal of Business* **68(1)**, 35–60. 41

Chan, K. C., **W.-M. Fong**, **B.-C. Kho** & **R. M. Schultz** (1996). 'Information, Trading and Stock Returns : Lessons from Dually-Listed Securities', *Journal of Banking & Finance* **20(7)**, 1161–1187. 44

Chan, Y.-C. (2000). 'The Price Impact of Trading on the Stock Exchange of Hong Kong', *Journal of Financial Markets* **3(1)**, 1–16. 49

Chauveau, T., **J. Damon** & **D. Guegan** (1999). 'Testing for Non-Linearity in Intra-Day Financial Series : The Cases of Two French Stocks', . 27, 76

Chen, C. R., **N. J. Mohan** & **T. L. Steiner** (1999). 'Discount Rate Changes, Stock Market Returns, Volatility, and Trading Volume : Evidence from Intraday Data and Implications for Market Efficiency', *Journal of Banking & Finance* **23(6)**, 897–924. 40

Chung, K. H., **B. F. Van Ness** & **R. A. Van Ness** (1999). 'Limit Orders and the Bid-Ask Spread', *Journal of Financial Economics* **53(2)**, 255–287. 41, 46, 50, 174

Clark, P. K. (1973). 'A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices', *Econometrica* **41(1)**, 135–155. 10, 11, 14, 21, 72

Cont, R. (2001). 'Empirical Properties of Asset Returns : Stylized Facts and Statistical Issues', *Quantitative Finance* **1(2)**, 223–236. 71

Copeland, T. E. & **D. Galai** (1983). 'Information Effects on the Bid Ask Spread', *The Journal of Finance* **38(5)**, 1457–1469. 98, 99, 108

Coppejans, M. & **I. Domowitz** (1999). 'Screen Information, Trader Activity, and Bid-Ask Spreads in a Limit Order Market', Working Paper Penn State University `http://www.smeal.psu.edu/faculty/ihd1/omx4.pdf`. 114

— (2002). 'An Empirical Analysis of Trades, Orders and Cancellations in a Limit Order Market', Working Paper Duke University `http://www.econ.duke.edu/~mtc/papers/omx7.pdf`. 111, 140

Coppejans, M., **I. Domowitz** & **A. Madhavan** (2003). 'Resiliency in an Automated Auction', Working Paper Duke University `http://www.econ.duke.edu/~mtc/papers/OMXCDM28.pdf`. 111, 176

Corwin, S. A. (1999). 'Differences in Trading Behavior Across NYSE Specialist Firms', *The Journal of Finance* **54(2)**. 176

Cushing, D. & **A. Madhavan** (2000). 'Stock Returns and Trading at the Close', *Journal of Financial Markets* **3(1)**, 45–67. 46

Daníelsson, J. & **R. Payne** (2001). 'Measuring and Explaining the Liquidity on an Electronic Limit Order Book : Evidence from Reuters D2000-2', Working Paper London School of Economics and Political Science. 41

Darolles, S., **C. Gouriéroux** & **G. Le Fol** (2000). 'Intraday Transaction Price Dynamics', *Annales d'Economie et de Statistique* **60**, 207–238. 114, 142

Davydoff, D., *Les indices Boursiers*, Gestion (Economica, 1998). 16

De Winne, R. & **C. Majois** (2003). 'A Comparison of Alternative Spread Decomposition Models on Euronext Brussels', Working Paper Facultés Universitaires Catholiques de Mons `http://staff.fucam.ac.be/~finance/finance_working.htm`. 48

Declerck, F. (2000*a*). *Analyse des Meilleurs Limites du Carnet D'ordres : Application À la Bourse de Paris*, Ph.D. thesis, Université de Lille II. 49, 175

— (2000*b*). 'Trading Costs on a Limit Order Book Market : Evidence from the Paris Bourse', . 41

Declerck, F. & **P. Hazart** (2000). 'Impact de la Tenue de Marché sur la Liquidité des Actions du Second Marché', . 175

Deininger, C., **C. Kaeserer** & **S. Ross** (2000). 'Stock Price Effects Associated with Index Replacements in Germany', Working Paper University Fribourg `http://www.unibz.it/archiv/objects/pdf/standard/2/wp_2000-03_nr7_kaserer.pdf`. 16

D'Hondt, C., **R. D. Winne** & **A. François-Heude** (2003). 'Hidden Orders on Euronext : Nothing is Quite as It Seems', Working Paper Fucam `http://staff.fucam.ac.be/~finance/NothingIsQuiteAsItSeems.pdf`. 175

Diamond, D. & **R. E. Verrecchia** (1987). 'Constraints on Short-Selling and Asset Price Adjustment to Private Information', *Journal of Financial Economics* **18(2)**, 277–311. 101, 116, 139, 142, 183

Domowitz, I., **J. Glen** & **A. Madhavan** (2000). 'Liquidity, Volatility, and Equity Trading Costs Across Countries and Over Time', . 104

Donaldson, R. G. & **M. J. Kamstran** (2000). 'Volatility Forecasts Combining ARCH, Trading Volume and Options', . 74

Drost, F. C. & **T. E. Nijman** (1993). 'Temporal Aggregation of GARCH Processes', *Econometrica* **61(4)**, 909–927. 143

Dufour, A. & **R. F. Engle** (2000). 'Time and the Price Impact of a Trade', *The Journal of Finance* **55(6)**. 143, 163, 165, 167

Dupont, D. (2000). 'Market Making, Prices, and Quantity Limits', *The Review of Financial Studies* **13(4)**, 112–1151. 107, 140, 181, 209, 252

Easley, D., **S. Hvidkjaer** & **M. O'Hara** (2002). 'Is Information Risk a Determinant of Asset Returns ?', *The Journal of Finance* **57(5)**, 2185–2221. 23

Easley, D., **N. M. Kiefer** & **M. O'Hara** (1997*a*). 'The Information Content of the Trading Process', *Journal of Empirical Finance* **4(2–3)**, 159–186. 100

— (1997*b*). ‘One day in the life of a very common stock’, *The Review of Financial Studies* **10(3)**, 805–835. 100

Easley, D., **N. M. Kiefer**, **M. O’Hara** & **J. B. Paperman** (1996). ‘Liquidity, Information, and Infrequently Traded Stocks’, *The Journal of Finance* **51(4)**, 1405–1436. 16, 100, 103

Easley, D. & **M. O’Hara** (1987). ‘Price, Trade Size, and Information in Securities Markets’, *Journal of Financial Economics* **19(1)**, 69–90. 98, 100, 160, 167, 171

— (1992). ‘Time and the Process of Security Price Adjustment’, *The Journal of Finance* **47(2)**, 577–605. 98, 100, 104, 105, 106, 142, 144, 158, 159, 168, 171, 172, 183, 189, 209

Ederington, L. H. & **H. Lee** (1993). ‘How Markets Process Information : News Releases and Volatility’, *The Journal of Finance* **48(4)**, 1161–1191. 40

Engle, R. F. (1982). ‘Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation’, *Econometrica* **50(4)**, 987–1008. 73

— (2000). ‘The Econometrics of Ultra High Frequency Data’, *Econometrica* **68(1)**, 1–22. 97, 116, 141, 142, 143, 144, 148, 155, 159, 162, 167, 171, 237, 238

Engle, R. F. & **A. Lunde** (2003). ‘Trades and Quotes : A Bivariate Point Process’, *Journal of Financial Econometrics* **1(2)**, 159–188. 155

Engle, R. F. & **J. R. Russell** (1997). ‘Forecasting the Frequency of Changes in Quoted Foreign Exchange Prices with Autoregressive Conditional Duration Model’, *Journal of Empirical Finance* **4(2–3)**, 187–212. 41, 113

— (1998). ‘Autoregressive Conditional Duration : A New Model for Irregularly Spaced Transaction Data’, *Econometrica* **66(5)**, 1127–1162. 41, 113, 148

Epps, T. W. & **M. L. Epps** (1976). ‘The Stochastic Dependence of Security Price Changes and Transactions Volumes : Implications for the Mixture-of-Distribution Hypothesis’, *Econometrica* **44(2)**, 305–321. 11

Fair, R. C. (2002). ‘Events that Shook the Market’, *Journal of Business* **75(4)**, 713–731. 2

Fama, E. F. (1963). ‘Mandelbrot and the Stable Paretian Hypothesis’, *Journal of Business* **36(4)**, 420–429. 72

— (1965). ‘The Behavior of Stock-Market Prices’, *Journal of Business* **38(1)**, 34–105. 72

Farmer, J. D. (1999). ‘Physicists Attempt to Scale the Ivory Towers of Finance’, *Computing in Science and Engineering* **1(6)**, 26–39. 72

Foster, F. D. & **S. Viswanathan** (1990). 'A Theory of Interday Variations in Volume, Variance, and Trading Costs in Security Markets', *The Review of Financial Studies* **3(4)**, 593–624. 98

— (1993). 'Variations in Trading Volume, Return Volatility, and Trading Costs : Evidence on Recent Price Formation Models', *The Journal of Finance* **48(1)**, 187–211. 41, 42, 44, 45, 48, 50

— (1994). 'Strategic Trading with Asymmetrically Informed Traders and Long-Lived Information', *Journal of Financial & Quantitative Analysis* **29(4)**, 499–518. 98

— (1996). 'Strategic Trading When Agent Forecast the Forecast of Others', *The Journal of Finance* **51(4)**, 1437–1500. 102

Foucault, T. (1999). 'Order Flow Composition and Trading Costs in a Dynamic Limit Order Market', *Journal of Financial Markets* **2(2)**, 99–134. 109, 110, 112, 121, 183

Foucault, T., **O. Kandan** & **E. Kandel** (2003). 'Limit Order Book as a Market for Liquidity', Cahier de Recherche du Groupe HEC, n^o 728. 110

French, K. R. & **R. Roll** (1986). 'Stock Return Variances the Arrival of Information and the Reaction of Traders', *Journal of Financial Economics* **17(1)**, 5–26. 49

Fung, H.-G. & **G. A. Patterson** (1999). 'The Dynamic Relationship of Volatility, Volume, and Market Depth in Currency Futures Markets', *Journal of International Financial Markets, Institutions & Money* **9**, 33–59. 41

Gallant, A. R. (1981). 'On the Bias in Functional Forms and Essential Unbiased Form : The Fourier Flexible Form', *Journal of Econometrics* **15(2)**, 211–245. 126

— (1982). 'Unbiased Determination of Production Technologies', *Journal of Econometrics* **20(2)**, 285–323. 126

Gallant, A. R., **P. E. Rossi** & **G. Tauchen** (1992). 'Stock Prices and Volume', *The Review of Financial Studies* **57(2)**, 199–242. 14, 71

— (1993). 'Nonlinear Dynamic Structures', *Econometrica* **61(4)**, 871–907. 14

Gerety, M. S. & **J. H. Mulherin** (1994). 'Price Formation on Stock Exchanges : The Evolution of Trading Within the Day', *The Review of Financial Studies* **7(3)**, 609–629. 41, 42, 43, 50

Ghysels, E., **C. Gouriéroux** & **J. Jasiak** (1995). 'Trading Patterns, Time Deformation and Stochastic Volatility in Foreign Exchange Markets', . 11

— (2000). ‘Causality Between Returns and Traded Volumes’, *Annales d’Economie et de Statistique* **60**, 189–206. 71

Ghysels, E. & **J. Jasiak** (1998). ‘GARCH for Irregularly Spaced Data : The ACD-GARCH’, *Studies in Nonlinear Dynamics and Econometrics* **2(4)**, 133–149. 142, 143

Ghysels, E. & **J. L. Juergens** (2001). ‘Stock Market Fundamentals and Heterogeneity of Beliefs : Tests Based on a Decomposition of Returns and Volatility’, Working Paper Unuiversity of North Carolina `http://papers.ssrn.com/sol3/delivery.cfm/SSRN_ID301836_code020303530.pdf?abstractid=301` 104

Giot, P. (2000). ‘Time Transformations, Intraday Data and Volatility Models’, *Journal of Computational Finance* **4(2)**. 22, 74, 148

Glosten, L. R. & **P. R. Milgrom** (1985). ‘Bid Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders’, *Journal of Financial Economics* **14(1)**, 71–100. 12, 98, 99

Goetzmann, W. F. & **M. Massa** (2001). ‘Dispersion of Opinion and Stock Returns : Evidence from Index Fund Investors’, Working Paper Yale School of Management. 104

Goodhart, C. A. E., **S. G. Hall**, **S. G. B. Henry** & **B. Pesaran** (1993). ‘News Effects in a High Frequency Model of the Sterling–Dollar Exchange Rate’, *Journal of Applied Econometrics* **8(1)**, 1–13. 40

Goodhart, C. A. E. & **M. O’Hara** (1997). ‘High Frequency Data in Financial Markets : Issues and Applications’, *Journal of Empirical Finance* **4(2–3)**, 73–114. 11, 73

Gouriéroux, C. & **J. Jasiak**, *Financial Econometrics Problems, Models and Methods* (Princeton University Press, 2001). 11, 27, 71, 72, 114, 118

Gouriéroux, C., **J. Jasiak** & **G. Le Fol** (1999*a*). ‘Durations’, . 148

— (1999*b*). ‘Intra-Day Market Activity’, *Journal of Financial Markets* **2(3)**, 193–226. 20, 54

Granger, C. W. J. (1969). ‘Investigating Causal Relations by Econometric Models and Cross Spectral Methods’, *Econometrica* **37(3)**, 424–438. 143, 156

Grossman, S. J. & **J. E. Stiglitz** (1980). ‘On the Impossibility of Informationally Efficient Markets’, *American Economic Review* **70(3)**, 393–408. 11, 99

Grundy, B. D. & **M. Mc Nichols** (1989). 'Trade and the Revelation of Information Through Prices and Direct Disclosure', *The Review of Financial Studies* **2(4)**, 495–526. 103, 104

Hafner, C. M. (1999). 'Durations, Volume and the Prediction of Financial Returns in Transaction Time', Working Paper Erasmus University Rotterdam `http://fmwww.bc.edu/RePEc/es2000/0599.pdf`. 41, 144

Hall, J. A., **B. Wade** & **S. H. Irwin** (1989). 'The Distribution of Futures Prices : A Test of the Stable Paretian and Mixture of Normals Hypotheses', *Journal of Financial & Quantitative Analysis* **24(1)**, 105–116. 72

Hamao, Y. & **J. Hasbrouck** (1995). 'Securities Trading in the Absence of Dealers : Trades, and Quotes on the Tokyo Stock Exchange', *The Review of Financial Studies* **8(3)**, 849–878. 45

Hamilton, J. D., *Time Series Analysis* (Princeton University Press, 1994). 73, 74

Hamon, J. & **B. Jacquillat** (1991*a*). 'Les Données Actions de La Banque de Données Bousières Affi-Sbf', *Finance* **12(1)**, 65–104. 217

— (1991*b*). 'Saisonnalité des Rentabilités dans la Semaine et la Séance à la Bourse de Paris', *Finance* **12(1)**, 105–126. 42, 43, 44

Handa, P. & **R. A. Schwartz** (1996). 'Limit Order Trading', *The Journal of Finance* **51(5)**, 1835–1861. 109

Handa, P., **R. A. Schwartz** & **A. Tiwari** (1998). 'The Ecology of an Order-Driven Market', *Journal of Portfolio Management* **24(2)**, 47–55. 108, 121, 145, 173, 209

Handa, P., **R. A. Schwatz** & **A. Tiwari** (2003). 'Quote Setting and Price Formation in an Order Driven Market', *Journal of Financial Markets* **Forthcoming(? ? ?)**, ? ? ? 110, 112, 183

Hansen, L. P. (1982). 'Large Sample Properties of Generalized Method of Moments Estimators', *Econometrica* **50(4)**, 1029–1054. 13, 51

Hansen, P. R. & **A. Lunde** (2002). 'A Forecast Comparison of Volatility Models : Does anything beat a GARCH(1,1) ?', Working Paper Brown University `http://www.econ.brown.edu/fac/Peter_Hansen/Papers/vola.pdf`. 73, 74, 76

Harris, L., *Trading and Exchanges : Market Microstructure for Practioners* (Oxford University Press, 2003). 174

Harris, L. E. (1986*a*). 'Cross-Security Tests of the Mixture of Distributions Hypothesis', *Journal of Financial & Quantitative Analysis* **21(1)**, 39–46. 12, 13, 53, 75

— (1986*b*). 'A Transaction Data Study of Weekly and Intradaily Patterns in Stock Returns', *Journal of Financial Economics* **16(1)**, 99–117. 26, 27, 42, 43, 44

— (1987). 'Transaction Data Tests of the Mixture of Distributions Hypothesis', *Journal of Financial & Quantitative Analysis* **22(2)**, 127–141. 12, 13, 53, 75

— (1989). 'A Day-End Transaction Price Anomaly', *Journal of Financial & Quantitative Analysis* **24(1)**, 29–45. 42

— (1994). 'Minimum Price Variations, Discrete Bid-Ask Spreads, and Quotation Sizes', *The Review of Financial Studies* **7(1)**, 149–178. 107

Harris, M. & **A. Raviv** (1993). 'Differences of Opinion Make a Horse Race', *The Review of Financial Studies* **6(3)**, 473–506. 99, 103

Hasbrouck, J. (1991). 'The Summary Informativeness of Stock Trades : An Econometric Analysis', *The Review of Financial Studies* **4(3)**, 571–595. 16, 41, 48, 49, 143

— (1999). 'The Dynamics of Discrete Bid and Ask Quotes', *The Journal of Finance* **54(6)**, 2109–2142. 104

Hasbrouck, J. & **D. J. Seppi** (2001). 'Common Factors in Prices, Order Flows and Liquidity', *Journal of Financial Economics* **53(3)**. 107, 116, 118, 177

Hausman, J., **A. W. Lo** & **A. C. Mackinlay** (1992). 'An Ordered Probit Analysis of Transaction Stock Prices', *Journal of Financial Economics* **31(3)**, 319–379. 141, 142

Hautsch, N. (2001). 'Modeling Intraday Trading Activity Using Box-Cox ACD Models', . 148

He, H. & **A. Wang** (1995). 'Differential Information and Dynamic Behavior of Stock Trading Volume', *The Review of Financial Studies* **8(4)**, 919–972. 23, 103, 104, 140

Hillion, P. & **M. Suominen** (1998). 'Deadline Effect on an Order Driven Market : An Analysis of the Last Minute on the Paris Bourse', Working Paper Insead. 46

Ho, R. Y.-K. & **R. S.-K. Lee** (1998). 'Market Closure Effects on Return, Volatility, and Turnover Patterns in the Hong Kong Index Futures Market', *Journal of International Financial Markets, Institutions & Money* **8(3–4)**, 433–451. 41, 45

Holden, C. W. & **A. Subrahmanyam** (1992). 'Long-Lived Information and Imperfect Competition', *The Journal of Finance* **47(1)**, 247–270. 102

Hollifield, B., **R. A. Miller** & **P. Sandås** (2003*a*). 'Empirical Analysis of Limit Order Markets', Working Paper Wharthon Scholl http://finance.wharton.upenn.edu/~sandas/hms_february03.pdf. 112

Hollifield, B., **R. A. Miller**, **P. Sandås** & **J. Slive** (2003*b*). 'Liquidity Supply and Demand in Limit Order Markets', Working Paper Wharthon Scholl http://finance.wharton.upenn.edu/~sandas/vse_main.pdf. 112

Hong, H. & **J. Wang** (2000). 'Trading and Returns Under Periodic Market Closures', *The Journal of Finance* **55(1)**, 297–354. 46

Huang, R. D. & **R. W. Masulis** (2003). 'Trading Activity and Stock Price Volatility : Evidence from the London Stock Exchange', *Journal of Empirical Finance* **10(3)**, 249–269. 22, 101

Huang, R. D. & **H. R. Stoll** (1994). 'Market Microstructure and Stock Return Predictions', *The Review of Financial Studies* **7(1)**, 179–213. 49, 51

Jain, P. C. & **G.-H. Joh** (1988). 'The Dependence Between Hourly Prices and Trading Volume', *Journal of Financial & Quantitative Analysis* **23(3)**, 269–283. 41, 42, 43

Jasiak, J. (1999). 'Persistence in Intertrade Durations', . 148

Jones, C. M., **G. Kaul** & **M. L. Lipson** (1994*a*). 'Information,Trading, and Volatility', *Journal of Financial Economics* **36(1)**, 127–154. 89

— (1994*b*). 'Transactions, Volume and Volatility', *The Review of Financial Studies* **7(4)**, 631–651. 22, 79

Kandel, E. & **N. M.Pearson** (1995). 'Differential Interpretation of Information and Trade in Speculative Markets', *The Journal of Political Economy* **103(4)**, 831–872. 103

Kaniel, R. & **H. Liu** (2002). 'So What Orders Do Informed Traders Use ?', Working Paper University of Texas at Austin http://www.mccombs.utexas.edu/Faculty/Ron.Kaniel/papers/spec.pdf. 108

Karpoff, J. M. (1987). 'The Relation Between Price Changes and Trading Volume : A Survey', *Journal of Financial & Quantitative Analysis* **22**, 109–126. 22

Kavajecz, K. A. (1998). 'A Specialist's Quoted Depth as a Strategic Choice Variable', Working Paper Wharthon School http://finance.wharton.upenn.edu/~kavajecz/theo3.pdf. 105

— (1999). 'A Specialist's Quoted Depth and the Limit Order Book', *The Journal of Finance* **54(2)**, 747–771. 16, 105, 121

Kavajecz, K. A. & **E. R. Odders-White** (2001). 'An Examination of Changes in Specialists' Posted Price Schedules', *The Review of Financial Studies* **14(3)**, 681–704. 106, 121

Kempf, A. & **O. Korn** (1999). 'Market Depth and Order Size', *Journal of Financial Markets* **2(1)**, 29–48. 22

Kramer, L. A. (2000). 'Intraday Stock Returns, Time-Varying Risk Premia and Diurnal Mood Variation', . 45

Kyle, A. S. (1985). 'Continuous Auctions and Insider Trading', *Econometrica* **53(6)**, 1315–1335. 98, 99, 107

Lamoureux, C. G. & **W. D. Lastrapes** (1990). 'Heteroskedasticity in Stock Return Data : Volume Versus Garch Effects', *The Journal of Finance* **45(1)**, 221–229. 74, 75, 79

Lancaster, T., *The Econometric Analysis of Transition Data*, vol. 17 of *Econometric Society Monograph* (Cambridge Univesity Press, 1990). 149

Le Courtois, O. (2001). 'One Century of Lévy processes in Finance', ISFA Université de Lyon. 72

Le Fol, G. & **L. Mercier** (1998). 'Time Deformation : Definition and Comparisons', *Journal of Computational Intelligence in Finance* **6(5)**, 19–33. 11

Lee, C. M. C., **B. Mucklow** & **M. J. Ready** (1993). 'Spreads, Depths, and the Impact of Earnings Information : An Intraday Analysis', *The Review of Financial Studies* **6(2)**, 345–374. 41, 106

Li, J. & **L. Kryzanowski** (1997). 'Trading Activity, Quoted Liquidity, and Stock Volatility', *Multinational Finance Journal* **1(3)**, 199–227. 22

Lin, J.-C., **G. C. Sanger** & **G. G. Booth** (1995). 'Trade Size and the Components of the Bid-Ask Spread', *The Review of Financial Studies* **8(4)**, 1153–1183. 49

Lin, S.-J., **J. Knight** & **S. Satchell**, *Financial Markets Tick By Tick*, chap. 3 Modeling Intra-Day Equity Prices and Volatility Using Information Arrivals : A Comparative Study of Different Choices of Informational Proxies (John Wiley & Sons, 1999), 27–64. 23, 78, 89

Lisenfeld, R. (2001). 'A Generalized Bivariate Mixture Model for Stock Price Volatility and Trading Volume', *Journal of Econometrics* **104(1)**, 141–178. 13, 14

— (2002). ‘Identifying Common Long Range-Dependence in Volume and Volatility Using High Frequency Data’, . 13

Lo, A. W. & **J. Wang** (2000). ‘Trading Volume : Definitions, Data Analysis, and Implications of Portfolio Theory’, *The Review of Financial Studies* **13(2)**, 257–300. 16

Locke, P. R. & **C. L. Sayers** (1993). ‘Intra-Day Futures Prices Volatility : Information Effects and Variance Persistence’, *Journal of Applied Econometrics* **8(1)**, 15–30. 74

Lockwood, L. J. & **S. Linn** (1990). ‘An Examination of Stock Market Return Volatility During Overnight and Intraday Periods 1964-1989’, *The Journal of Finance* **45(2)**, 591–60. 41, 42

Lüders, E. & **B. Peisl** (2000). ‘On the Relationship of Information Processes and Asset Price Processes’, Working Paper University of Konstanz n^o 00/09. 38, 47

Madhavan, A., **M. Richardson** & **M. Roomans** (1997). ‘Why Do Prices Change ? A Transaction-Level Analysis of NYSE Stocks’, *The Review of Financial Studies* **10(4)**, 1035–1064. 49

Mailath, G. J. & **A. Sandroni** (2003). ‘Market Selection and Asymetric Information’, *The Review of Economic Studies* **70(2)**, 343–368. 4

Maillet, B. & **T. Michel** (1997). ‘Mesure de Temps, Information et Distribution des Rendements Intra-Journaliers’, *Journal de la Société Statistique de Paris* **138(4)**, 89–119. 22, 74, 76, 78, 79, 80, 126, 128

Mandelbrot, B. (1973). ‘Comments on : ’a Subordinated Stochastic Process Model with Finite Variance for Speculative Prices’ By Peter Clark’, *Econometrica* **41(1)**, 157–159. 72

Mandelbrot, B. & **H. M. Taylor** (1967). ‘On the Distribution of Stock Price Differences’, *Operation Research* **15**, 1057–1062. 10, 72

Manganelli, S. (2002). ‘Time, Volume and Price Impact of Trades’, Working Paper European Central Bank `http://www.ecb.int/pub/wp/ecbwp125.pdf`. 16, 113, 141, 143, 144, 152, 162, 165, 167, 171, 176, 196, 200, 241, 243

Marsh, T. A. & **N. Wagner** (2000). ‘Return - Volume Dependence and Extremes in International Equity Markets’, Working Paper UC Berkeley `http://haas.berkeley.edu/finance/WP/rpf293.pdf`. 74

Mc Inish, T. & **R. Wood** (1992). ‘An Analysis of Intraday Pattern in Bid/Ask Spread’, *The Journal of Finance* **47(2)**, 753–764. 41

Meddahi, N., **E. Renault** & **J. M. Werker Bas** (1998). 'Modelling High-Frequency Data in Continuous Time', Working Paper Université de Montréal `http://www.mapageweb.umontreal.ca/meddahin/mrw.pdf`. 142

Minguet, A., *La Microstructure des Marchés d'Actions : Une Approche Empirique* (ECONOMICA, 2003). 45, 100, 174

Müller, U. A., **M. M. Dacorogna**, **R. D. Dave**, **R. B. Pictet**, **O. V. Olsen** & **J. R. Ward** (1995). 'Fractals and Intrinsic Time - a Challenge to Econometricians', *Working Paper Olsen & Associates* . 14, 26

Muranaga, J. (1999). 'Dynamics of Market Liquidity of Japanese Stocks : An Analysis of Tick-by-tick Data of the Tokyo Stock Exchange', Working Paper IMES `http://www.imes.boj.or.jp/edps99/99-E-16.pdf`. 107

Nelson, D. B. & **C. Q. Cao** (1992). 'Inequality Constraints in the Univariate Garch Model', *Journal of Business & Economic Statistics* **10**, 229–235. 145

Newey, W. K. & **K. D. West** (1987). 'A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix', *Econometrica* **55(3)**, 703–708. 51

Niemeyer, J. & **P. Sandås** (1993). 'An Empirical Analysis of the Trading Structure at the Stockholm Stock Exchange', *Journal of Multinational Financial Management* **3(3–4)**, 63–101. 41, 42, 43

Nizan, P., *La conspiration*, Folio (Gallimmard, 1938). 2

O'Hara, M., *Market Microstructure Theory* (Blackwell, 1995). 108

Orléan, A. (2002). 'Le Tournant Cognitif en Économie', *Revue d'économie politique* **5**, 717–738. 3

Pagan, A. (1996). 'The Econometrics of Financial Markets', *Journal of Empirical Finance* **3(1)**, 15–102. 71

Parlour, C. A. (1998). 'Price Dynamics in a Limit Order Markets', *The Review of Financial Studies* **11(4)**, 789–816. 108, 109

Porter, D. C. (1992). 'The Probability of a Trade at the Ask : An Examination of Interday and Intraday Behavior', *Journal of Financial & Quantitative Analysis* **27(2)**, 209–227. 47

Qi, R. (2001). 'Return-Volume Relation in the Tail : Evidence from Six Emerging Markets', . 14

Ranaldo, A. (2004). 'Order Aggresssiveness in Limit Order Book Markets', *Journal of Financial Markets* **7(1)**, 53–74. 111, 112, 121, 209

Richardson, M. & **T. Smtih** (1994). 'A Direct Test of Mixture of Distributions Hypothesis : Measuring the Daily Flow of Information', *Journal of Financial & Quantitative Analysis* **29(1)**, 101–116. 12, 13, 53, 75

Ronen, T. (1997). 'Tests of Properties of Variance Ratios in Microstructure Studies', *Journal of Financial & Quantitative Analysis* **32(2)**, 183–204. 51

Ross, S. A. (1989). 'Information and Volatility : The No-Arbitrage Martingale Approach to Timing and Resolution Irrelevancy', *The Journal of Finance* **44(1)**, 1–17. 11

Russell, J. R. & **R. F. Engle** (2002). 'Econometric Analysis of Discrete-Valued Irregularly-Spaced Financial Transactions Data', Working Paper University of Chicago `http://gsb-www.uchicago.edu/fac/jeffrey.russell/research/acm.pdf`. 141

Rydberg, T. H. & **N. Shepard** (1998). 'A Modeling Framework for the Price and Time of Trades Made on the New York Stock Exchange', . 148

— (2003). 'Dynamics of Trade-By-Trade Price Movements : Decomposition and Models', *Journal of Financial Econometrics* **1(1)**, 2–25. 141

SBF (1998). *Les Indices de la Bourse de Paris*, SBF Bourse de Paris, `http://www.bourse-de-paris.fr/fr/tools0/pdf/market/b/indices.pdf`. 16

— (2001). *Les Ordres de Bourse*, SBF Bourse de Paris, `http://www.bourse-de-paris.fr/centredoc/pdf/ordresbourse_nov2001.pdf`. 119

Seppi, D. J. (1997). 'Liquidity Provision with Limit Orders and a Strategic Specialist', *The Review of Financial Studies* **10(1)**, 103–150. 175

Shalen, C. T. (1993). 'Volume, Volatility and the Dispersion of Beliefs', *The Review of Financial Studies* **6(2)**, 405–434. 103, 104

Silverman, B. W., *Density Estimation for Statistics and Data Analysis*, Monographs on Statistics and Applied Probability (London : Chapman and Hall, 1986). 151

Slezak, S. L. (1994). 'A Theory of Dynamics of Security Markets Returns Around Market Closure', *The Journal of Finance* **45(4)**, 1163–12211. 46, 47

Smirlock, M. & **L. Starks** (1986). 'Day-of-the-Week and Intraday Effect in Stock Returns', *Journal of Financial Economics* **17(1)**, 197–210. 41, 42, 44

Spierdijk, L. (2002). 'An Empirical Analysis of the Role of the Trading Intensity in the information Dissemination on the NYSE', Working Paper Tilburg University `http://greywww.kub.nl:2080/greyfiles/center/2002/doc/30.pdf`. 16, 143

Spierdijk, L., **T. E. Nijman** & **A. H. O. van Soest** (2002). 'Modeling Comovements in Trading Intensities to Distinguish Sector and Stock Specific News', Working Paper Tilburg University `http://greywww.kub.nl:2080/greyfiles/center/2002/doc/69.pdf`. 17

Stock, J. H. (1988). 'Estimating Continuous-Time Processes Subject to Time Deformation', *Journal of the American Statistical Association* **83(401)**, 77–85. 11

Stoll, H. R. & **R. E. Whaley** (1990). 'Stock Market Structure and Volatility', *The Review of Financial Studies* **3(1)**, 37–71. 28, 46

Tauchen, G., **H. Zhang** & **M. Liu** (1996). 'Volume, Volatility, and Leverage : A Dynamic Analysis', *Journal of Econometrics* **74(1)**, 177–208. 14

Tauchen, G. E. & **M. Pitts** (1983). 'The Price Variability-Volume Relationship on Speculative Markets', *Econometrica* **51(2)**, 485–505. 11, 13, 14

Thomas, S. & **M. Demarchi**, *The Electronic Call Auction : Market Mechanism and Trading Building a Better Stock Market*, chap. 8 Call Market Mechanism on the Paris Stock Exchange (Kluwer Academic Publishers, 1999), 125–132. 46

Tkac, P. A. (1999). 'A Trading Volume Benchmark : Theory and Evidence', *Journal of Financial & Quantitative Analysis* **34(1)**, 89–114. 16

Vesserau, T. (2000). 'Aspect Non Linéaire du Marché des Actions Françaises', Working Paper Cirano `http://www.cirano.qc.ca/pdf/publication/2000s-45.pdf`. 76

Wang, F. A. (1998). 'Strategic Trading, Asymmetric Information and Heterogeneous Prior Beliefs', *Journal of Financial Markets* **1(3–4)**, 321–352. 99, 103

Wang, J. (1994). 'A Model of Competitive Stock Trading Volume', *The Journal of Political Economy* **102(1)**, 127–168. 99, 103

Werner, I. M. & **A. W. Kleidon** (1996). 'U.K. and U.S. Trading of British Cross–Listed Stocks : An Intraday Analysis of Market Integration', *The Review of Financial Studies* **9(2)**, 619–664. 44, 45

Westerfield, R. (1977). 'The Distribution of Common Stock Price Changes : An Application of Transaction Time and Subordinated Stochastic Models', *Journal of Financial & Quantitative Analysis* **12(5)**, 743–765. 72

White, H. (1980). 'A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity', *Econometrica* **48(4)**, 817–838. 115, 164, 186, 188

Wood, R. A., **T. H. Mc Inish** & **J. K. Ord** (1985). 'An Investigation of Transaction Data for NYSE Stocks', *The Journal of Finance* **40(3)**, 723–741. 42, 43, 47

Xu, X. E. & **C. Wu** (1999). 'The Intraday Relation Between Return Volatility, Transactions, and Volume', *International Review of Economics & Finance* **8(4)**, 375–397. 22, 43, 44, 45, 50

Zebedee, A. A. (2001). 'The impact of a trade on national best bid and offer quotes : a new approach to modeling irregularly spaced data', *Journal of Multinational Financial Management* **11(4–5)**, 363–383. 143

Zhang, M. Y., **J. R. Russel** & **R. S. Tsay** (2000). 'Information Determinants of Bid Ask Quotes', Working Paper University of Chicago `http://gsb-www.uchicago.edu/fac/jeffrey.russell/research/bidask.pdf`. 104, 116

— (2001). 'A Nonlinear Autoregressive Conditional Duration Model with Applications to Financial Transaction Duration', *Journal of Econometrics* **104(1)**, 179–207. 143