High-frequency trading : statistical analysis, modelling and regulation

Pamela Saliba

université PARIS-SACLAY

NNT: 2019SACLX044

Thèse de doctorat

High-frequency trading : Statistical analysis, modelling and regulation

Thèse de doctorat de l'Université Paris-Saclay
préparée à l'Ecole Polytechnique

École doctorale n°574 Ecole doctorale de mathématiques Hadamard (EDMH)
Spécialité de doctorat : Mathématiques appliquées

Thèse présentée et soutenue à Paris, le 18 juillet 2019 par

Pamela Saliba

Composition du Jury :

Nicole El Karoui Professeur émérite, Sorbonne Universités (LPSM)	Présidente
Jean-Philippe Bouchaud Responsable de la recherche, Capital Fund Management	Rapporteur
Alain Chaboud Economiste principal, Federal Reserve Bank	Rapporteur
Frédéric Abergel Professeur, CentraleSupélec (MICS)	Examinateur
Alexandra Givry Directrice de la Surveillance des Marchés, Autorité des Marchés Financiers	Examinatrice
Olivier Guéant Professeur, Université Paris I Panthéon Sorbonne (CES)	Examinateur
Charles-Albert Lehalle Chercheur senior, Capital Fund Management	Examinateur
Mathieu Rosenbaum Professeur, Ecole Polytechnique (CMAP)	Directeur de thèse

THÈSE

présentée pour obtenir

LE GRADE DE DOCTEUR EN MATHÉMATIQUES APPLIQUÉES
DE L'ÉCOLE POLYTECHNIQUE

par

Pamela SALIBA

High-frequency trading: Statistical analysis, modelling and regulation

Soutenue le 18 juillet 2019 devant un jury composé de :

Nicole EL KAROUI	*Sorbonne Universités*	Présidente
Jean-Philippe BOUCHAUD	*Capital Fund Management*	Rapporteur
Alain CHABOUD	*Federal Reserve Bank*	Rapporteur
Fabrizio LILLO	*Scuola Normale Superiore di Pisa*	Rapporteur
Frédéric ABERGEL	*CentraleSupélec*	Examinateur
Alexandra GIVRY	*Autorité des Marchés Financiers*	Examinatrice
Olivier GUÉANT	*Université Paris I Panthéon Sorbonne*	Examinateur
Charles-Albert LEHALLE	*Capital Fund Management*	Examinateur
Mathieu ROSENBAUM	*École Polytechnique*	Directeur de thèse

A Boulos et Camélia ...

Remerciements

Été 2014, j'hésitais énormément entre deux masters différents, l'un appliqué à la finance (master El Karoui) et l'autre à la biologie. A l'époque, j'ignorais encore complètement le monde des mathématiques financières, et face à une telle hésitation, j'ai contacté Mathieu Rosenbaum, le co-responsable du master El Karoui, pour obtenir des renseignements. J'ai été agréablement surprise par sa disponibilité, son objectivité et son implication pour m'orienter vers la bonne voie. Finalement, il a réussi, avec délicatesse, à m'introduire au monde des mathématiques financières, devenant par la suite mon directeur de thèse. Mathieu (ou Prof Rosenbaum), c'est avec toi que j'ouvre le bal des remerciements non par coutume mais par conviction. Je te remercie pour ta bienveillance, ta disponibilité malgré tes nombreuses occupations, et surtout pour ta positivité et tes encouragements tout au long de ma thèse. Au-delà de ce que tu m'as apporté sur le plan scientifique, tu m'as transmis ta curiosité qui touche plusieurs centres d'intérêts à la fois : recherche, histoire, art, séries sans oublier ta passion pour les langues ... domaines qu'à ma grande surprise, tu maîtrises tous aussi bien l'un que l'autre. Mathieu, sans toi cette thèse n'aurait jamais vu le jour, merci infiniment.

Je tiens à remercier Nicolas Megarbane qui a encadré une partie de ma thèse à l'AMF avant qu'il se lance dans la fondation de sa propre startup. Nicolas, merci beaucoup pour le temps que tu m'as consacré et pour ta volonté continue de me tirer vers le haut malgré nos divergences d'avis et nos petits désaccords professionnels. J'ai énormément apprécié nos longs échanges sur le voyage, photographie, philosophie et plein d'autres sujets. Enfin, je suis sûre que ton esprit novateur, ta ténacité et ta rigueur se reflèteront parfaitement dans ta startup !

Je suis très reconnaissante envers Charles-Albert Lehalle avec qui j'ai eu tout au long de ma thèse de nombreuses et enrichissantes discussions scientifiques. Charles, merci pour notre collaboration scientifique, pour tes précieux et brillants conseils livrés toujours avec humilité, et pour ta disponibilité même en plein milieu de tes vacances. Merci également pour avoir accepté d'être membre du jury de ma soutenance.

Je tiens à remercier Agostino Capponi et Othmane Mounjid avec qui j'ai eu la chance d'avoir des collaborations scientifiques et des échanges d'idées très fructueux.

Je remercie chaleureusement Jean-Philippe Bouchaud, Alain Chaboud et Fabrizio Lillo pour avoir accepté d'être les rapporteurs de ma thèse. Merci pour le temps énorme qu'ils m'ont consacré, leur lecture attentive, leur intérêt pour ma recherche et surtout pour leurs commentaires très constructifs et encourageants. Je remercie particulièrement Jean-Philippe Bouchaud qui, lors de nos rares mais si enrichissantes discussions, m'a impressionnée par son humilité et sa grande culture à la fois scientifique et artistique (et surtout sa connaissance des artistes et cinéastes libanais). Tous mes remerciements à Alain Chaboud pour son accueil très chaleureux à la FED, pour m'avoir donné l'opportunité de présenter mes travaux de recherche et d'échanger avec lui ainsi qu'avec d'autres économistes de cette institution, et pour l'agréable

visite de Washington en sa compagnie.

Je remercie sincèrement Frédéric Abergel, Nicole El Karoui et Olivier Guéant, qui me font l'honneur de participer au jury de ma soutenance. Je suis particulièrement touchée par la présence de Nicole El Karoui qui, lors des quelques échanges que j'ai eus avec elle au cours du master, est rapidement devenue «la figure féminine scientifique» inspirante. J'en profite pour la remercier pour son implication très active dans l'encouragement des femmes à s'orienter vers les filières scientifiques, une cause qui m'est personnellement très chère.

Je mesure la chance que j'ai eue d'avoir pu bénéficier des connaissances de mes professeurs du master, en particulier Emmanuel Gobet, Gilles Pages et Nizar Touzi.

Je suis très reconnaissante envers tous mes collègues de l'AMF à qui je présente mes excuses de ne pas pouvoir adresser des remerciements personnels à chacun d'entre eux. Je remercie premièrement Philippe Guillot capable de transformer toute problématique financière en une étude de cas vélo. Au-delà de son humilité, c'est particulièrement cette aisance dans la vulgarisation et la communication de ses idées, si enrichissantes et éclairantes, que j'apprécie énormément chez lui. Je remercie également Alexandra Givry qui derrière son air sérieux et intelligent (aucun doute là-dessus) se cache une personne dotée de valeurs humaines que j'admire énormément. Alexandra, merci de m'avoir soutenue pendant les moments difficiles et d'avoir accepté d'être membre de mon jury. Je remercie également Julien Leprun pour sa légendaire courtoisie, sa lumineuse culture, et surtout sa grande flexibilité qui m'a permis d'effectuer ma thèse dans les meilleures conditions possibles. J'adresse mes remerciements à toute l'équipe d'ingénierie financière de l'AMF : Alexis pour ses blagues de «très haut niveau» et pour notre complicité amicale, Antoine pour ses conseils champagne et technologie, Franck R. pour ses petit cadeaux très expressifs (oui oui ... je parle bien des stylos en forme d'armes). Iris, je te remercie «en tant que personne et en tant qu'amie» pour ta bonne humeur continue et ta très grande bienveillance, Joshua pour nos gentilles taquineries mutuelles, Martin pour tes services de conseil dans la langue de Molière. Enfin, ce serait pêcher de ne pas remercier Franck L. pour toutes ses photos «spontanées», Olivier (ex-AMF) et Pascale (à qui je souhaite un très bon rétablissement) pour nos petites soirées AMF les vendredis et pour tous leurs sages conseils.

Mes grands remerciements vont aux membres de la Team Rosenbaum. Je mentionne en particulier Aditi, dont l'arrivée au sein de la Team, très masculine à l'époque, fut vécue par moi comme un événement victorieux. Aditi, merci pour ta bonne humeur et ton air exotique très new-yorkais. Je tiens à remercier également les inséparables Bastien et Paul pour leur dynamisme, leur bienveillance et surtout leur joie de vivre. Je ne peux surtout pas oublier de remercier Omar pour l'ambiance très agréable qu'il a su instaurer à travers ses inlassables petites taquineries. Omar, Saad, je saurai me souvenir de votre dévouement sans faille dans le transport de mes valises lors de notre conférence à Venise ;) Une pensée particulière à Emel, Marcos et Mehdi.

Je remercie égalemement tout le secrétariat du CMAP, notamment Maud Cadiz-Pena, Nasséra Naar et Alexandra Noiret pour leur disponibilité et leur aide.

Il me tient à coeur de remercier Aline Lefebvre-Lepot et Aldjia Mazari, responsables des TPs dont j'ai été chargée dans le cadre de mes missions d'enseignement à l'Ecole Polytechnique. Merci pour votre dévouement pour l'enseignement et votre flexibilité pendant les périodes les plus chargées. Martin et Mathilde, ce fut un plaisir de collaborer avec vous, merci pour votre entraide et votre esprit d'équipe.

Je souhaiterais remercier également Dr (Karam) Fayad que j'ai contacté la première fois via *Messenger* (merci *Facebook* !) quand j'étais encore au Liban. C'est un ami en commun qui m'a transmis son contact (eh oui, le bouche à oreille à la libanaise fonctionne bien finalement) afin de me renseigner sur la vie et les études en France. Karam, merci pour tous les renseignements que tu m'as donnés à l'époque et surtout pour m'avoir présentée auprès de mes docteurs favoris Dr (Marwa) Banna, Dr (Safaa) El Sayed et Dr (Pierre) Youssef. Je vous remercie tous pour votre amitié, vos conseils, votre soutien et pour avoir supporté mes interminables hésitations.

Je tiens à remercier également mes amis à Paris, Charbel, Collince, Eliane, Hassina, Jinane, Sandrine, Steven, Xavier et Zorba pour leur soutien continu. Je remercie mes amies au Liban, Josiane, Stéphanie et Vola, mon amie aux US Marie-José, et mon amie sans situation géographique fixe Lise pour leur disponibilité malgré les milliers de kilomètres qui nous séparent (merci *Whatsapp* au passage).

Je tiens à exprimer ma gratitude envers mes deux pays: le Liban, mon pays d'origine, et la France, mon pays d'adoption (drôle de coincidence, j'ai été naturalisée le 18 juillet 2018, soit exactement un an avant la date de la soutenance de ma thèse). C'est au Liban que j'ai suivi l'intégralité de ma formation scolaire et une partie de ma formation universitaire. J'en profite pour exprimer ma reconnaissance envers le Collège de la Sainte Famille Française (Jounieh et Fanar) pour son haut niveau accadémique, et l'Université Libanaise ainsi que ses professeurs pour leur implication dans le rayonnement des sciences malgré les moyens financiers très limités qui sont mis à leur disposition. C'est en France que j'ai poursuivi mes études supérieures, à Paris 6 pour mon master et à l'Ecole Polytechnique pour mon doctorat. Comment ne pas être reconnaissante envers ce pays qui a offert avec générosité à la simple étrangère que j'étais encore, l'opportunité d'accéder aux meilleures formations universitaires du monde sans aucune attente en retour ?

Permettez-moi de m'attarder un peu pour remercier ma famille (je vous préviens, nous sommes nombreux !). Je commence par remercier mes frères et sœur par ordre d'âge décroissant pour éviter tout sentiment d'envie. David Fadi, son ambition l'a poussé à quitter le Liban alors que je n'avais encore que quelques mois. Mais depuis mon arrivée à Paris en Septembre 2013, il a réussi à surcompenser tous les moments que nous n'avons pas partagés plus jeunes. Ceci en m'appelant presque quotidiennement pour prendre des nouvelles sur le déroulement de mon master, puis de mon doctorat. David Fadi, merci pour ta GRANDE bienveillance que je n'ai jamais réussi à tarir malgré mes tentatives persistantes et continues. Je ne peux remercier mon frère aîné sans remercier son épouse Darine pour avoir été comme une deuxième soeur pour moi depuis mon arrivée en France. Darine, merci pour ton amabilité, tes plats libanais savoureux, et surtout pour nos soirées de discussions interminables! Chadi, merci pour ton écoute maîtrisée avec art pendant les moments les plus difficiles et les plus joyeux, ta mémoire infaillible, et surtout tes réponses qui se réduisaient parfois à un silence total ... mais un silence si parlant. Chadi, par pitié, pas de scoops ! Plus jeune, j'étais inséparable de Hadi qui m'a appris à nager, à conduire et bien évidemment quelques petites prises de Karaté toujours très importantes à maîtriser. Je lui suis reconnaissante malgré le manque de douceur de ses modes d'apprentissage. Il est resté fidèle à ce rôle d'enseignant qui ressurgit face à tout échec ou succès à travers ses fameuses expressions : «Bravo Pam», «Courage Pam» et «Go forward». Hadi, merci pour tes mots doux et encourageants, à l'opposé de tes moyens d'apprentissage. Je remercie ma sœur Paula pour nos nombreuses discussions téléphoniques, parfois incompatibles avec mon emploi du temps, mais (malheureusement) rien ne l'arrête quand elle veut aborder un sujet. Merci pour tous ses conseils lors des conférences auxquelles j'ai participé ... je parle

évidemment de ses conseils mode (que j'ai pris, parfois, en considération). Paula, merci pour ta bienveillance, toujours masquée par ton air moqueur, et pour tes blagues imbattables qui ont l'art d'alléger la gravité de n'importe quelle situation.

Enfin, je dédie cette thèse (et toutes mes futures réalisations) à mes chers parents Boulos et Camélia. Comment ne pas le faire alors qu'ils sont à l'origine de tous mes accomplissements ? Camélia (oui j'ai cette curieuse habitude d'appeler mes parents par leurs prénoms), merci d'avoir réussi à te délester complètement de cet attachement maternel parfois accaparant pour m'encourager à poursuivre mes études en France. Merci pour ta sagesse, ton ouverture d'esprit et surtout pour ta surveillance méticuleuse et «très» discrète de mon *last seen* sur *Whatsapp* qui déclenche un état d'urgence dès qu'il dépasse les huit heures ! Boulos, par ton ambition et ta pugnacité, tu es parvenu à m'apprendre qu'on peut toujours accomplir nos objectifs avec une petite dose de persévérance, d'audace et d'ambition, sans pour autant que celle-ci n'empiète sur le contentement, me répétant sans cesse ce proverbe qui m'accompagnera toujours «Contentement passe richesse».

Résumé

Cette thèse est constituée de deux parties liées l'une à l'autre. Dans la première, nous étudions empiriquement l'impact de l'activité de trading haute fréquence, définie comme étant une sous-catégorie du trading algorithmique caractérisée par une courte période de détention de titres, sur les marchés financiers européens. Nous utilisons les résultats obtenus afin de construire dans la seconde partie de nouveaux modèles multi-agents. L'objectif principal de ces modèles est de fournir aux régulateurs et plateformes de négociation des outils innovants leur permettant de mettre en place des règles pertinentes pour la microstructure et de quantifier l'impact des divers participants sur la qualité du marché.

Dans la première partie, nous effectuons deux études empiriques sur des données uniques fournies par le régulateur français. Nous avons accès à l'ensemble des ordres et transactions des actifs du CAC 40, à l'échelle de la microseconde, avec par ailleurs les identités des acteurs impliqués. Nous commençons par comparer le comportement des traders haute fréquence à celui des autres intervenants, notamment pendant les périodes de stress, en termes de provision de liquidité et d'activité de négociation. Nous arrivons à montrer dans cette étude une caractéristique cruciale relative aux traders à haute fréquence: leur activité n'est pas limitée à la tenue de marché pure et plus de 50% de leurs sont agressifs (ordres consommant la liquidité). Cette constatation nous pousse à approfondir notre analyse en nous focalisant sur les ordres agressifs. Nous étudions leur impact sur le processus de formation des prix et leur contenu informationnel selon les différentes catégories de flux : traders haute fréquence, participants agissant pour compte client et participants agissant pour compte propre. De plus, nous parvenons dans cette analyse à mettre en évidence empiriquement le lien existant entre les transactions et l'avantage informationnel des traders à haute fréquence.

Dans la seconde partie, nous proposons trois modèles multi-agents. Dans le premier modèle, nous cherchons à incoroprer la relation entre transactions et information. Cette relation est dans le coeur du modèle Glosten-Milgrom, mais ce dernier se limite à la modélisation des meilleurs prix à la vente et à l'achat. À l'aide d'une approche à la Glosten-Milgrom, nous parvenons à construire l'ensemble du carnet d'ordres (spread et volume disponible à chaque prix) à partir des interactions entre trois types d'agents : un agent informé, un agent non informé et des teneurs de marché. Ce modèle nous permet par ailleurs de développer une méthodologie de prédiction du spread en cas de modification du pas de cotation et de quantifier la valeur de la priorité dans la file d'attente. L'inconvénient principal de ce modèle est celui d'agréger les différents participants de marché par groupes. Ainsi, il ne permet pas de mettre en évidence les disparités existantes entre les participants de marché appartenant au même groupe. Afin de se concentrer sur une échelle individuelle, nous proposons une deuxième approche où les dynamiques spécifiques des agents sont modélisées par des processus de type Hawkes non linéaires et dépendants de l'état du carnet d'ordres. Dans ce cadre, nous sommes en mesure de calculer en fonction des flux individuels plusieurs indicateurs pertinents relatifs à la microstructure. Il est notamment possible de classer les

teneurs de marché selon leur contribution propre à la volatilité. Enfin, nous introduisons un modèle où les fournisseurs de liquidité optimisent leurs meilleurs prix à l'achat et à la vente en fonction du profit qu'ils peuvent générer et du risque d'inventaire auquel ils sont confrontés. Nous mettons alors en évidence théoriquement une corrélation négative entre inventaire des teneurs de marché et pression exercée sur les prix. Nous confirmons ce résultat empiriquement en étudiant les inventaires individuels des teneurs de marché dans un contexte multi-plateformes.

Abstract

This thesis is made of two related parts. In the first one, we study the empirical behaviour of high-frequency traders on European financial markets. We use the obtained results to build in the second part new agent-based models for market dynamics. The main purpose of these models is to provide innovative tools for regulators and exchanges allowing them to design suitable rules at the microstructure level and to assess the impact of the various participants on market quality.

In the first part, we conduct two empirical studies on unique data sets provided by the French regulator. It covers the trades and orders of the CAC 40 securities, with microseconds accuracy and labelled by the market participants identities. We begin by investigating the behaviour of high-frequency traders compared to the rest of the market, notably during periods of stress, in terms of liquidity provision and trading activity. We work both at the day-to-day scale and at the intraday level. We then deepen our analysis by focusing on liquidity consuming orders. We give some evidence concerning their impact on the price formation process and their information content according to the different order flow categories: high-frequency traders, agency participants and proprietary ones.

In the second part, we propose three different agent-based models. Using a Glosten-Milgrom type approach, the first model enables us to deduce the whole limit order book (bid-ask spread and volume available at each price) from the interactions between three kinds of agents: an informed trader, a noise trader and several market makers. It also allows us to build a spread forecasting methodology in case of a tick size change and to quantify the queue priority value. To work at the individual agent level, we propose a second approach where market participants specific dynamics are modelled by non-linear and state dependent Hawkes type processes. In this setting, we are able to compute several relevant microstructural indicators in terms of the individual flows. It is notably possible to rank market makers according to their own contribution to volatility. Finally, we introduce a model where market makers optimise their best bid and ask according to the profit they can generate and the inventory risk they face. We then establish theoretically a negative relationship between market makers inventories and price pressure. We confirm this result studying empirically individual market makers inventories in a multi-platform framework.

List of papers being part of this thesis

- Charles-Albert Lehalle, Nicolas Megarbane, Mathieu Rosenbaum and Pamela Saliba *The behaviour of high-frequency traders under different market stress scenarios*, Market Microstructure and Liquidity, Vol. 03, No. 03n04, 1850005, 2017.
- Pamela Saliba, *The information content of high-frequency traders aggressive orders: recent evidences*, submitted, 2019.
- Weibing Huang, Mathieu Rosenbaum and Pamela Saliba *From Glosten-Milgrom to the whole limit order book and applications to financial regulation*, submitted, 2019.
- Othmane Mounjid, Mathieu Rosenbaum and Pamela Saliba, *From asymptotic properties of general point processes to the ranking of financial agents* , submitted, 2019.
- Agostino Capponi, Mathieu Rosenbaum and Pamela Saliba, *Market makers inventories and price pressure: theory and multi-platform empirical evidences*, working paper, 2019.

Contents

Introduction

The guiding principle of this thesis is to understand and model how high-frequency traders (HFTs) impact the market and suggest ways to mitigate their effects. After studying empirically the HFTs behaviour, we aim at proposing models reproducing our empirical observations, and allowing us to design new regulatory tools. Let us begin with presenting and motivating the different questions on which we want to shed some light in this thesis.

Motivations

HFTs are commonly defined as a subset of algorithmic traders using co-location and proximity services to minimise latency, submitting a large number of orders that are cancelled shortly after submission and holding assets over very short periods. The significant growth in high-frequency trading in the recent years has created a debate among regulators, academics and practitioners. They aim at understanding how HFTs affect the markets. Does their activity improve or damage market quality? Some argue that HFTs are beneficial for the market: they increase liquidity provision, do not withdraw from markets in bad times, contribute more to price discovery than non-HFTs and reduce volatility. On the contrary, others assert that they have toxic consequences on the market. This is because when speed disparities exist between traders, adverse selection may increase and liquidity becomes more expensive. We contribute to this debate by considering the following question:

Question 1. *What is the intraday behaviour of HFTs, and how do they react during periods of stress?*

Addressing Question 1, we highlight the behaviours disparities across HFTs. We also give interpretations for the variations of HFTs shares in market depth and traded amounts according to the volatility and to macroeconomic announcements. Furthermore, we show a crucial feature relative to HFTs: their activity is not restrained to pure market making, and more than 50% of their orders are aggressive. Recall that a transaction is derived from the matching of an aggressive order with a passive one: the passive order is the one standing in the order book and providing liquidity, while the aggressive order is the one consuming liquidity. It is well known that HFTs may use aggressive orders for inventory management purpose (we will support this argument by giving empirical proofs later on, when answering Question 5), but the ratio between aggressive and passive orders we obtain, which is higher than 50%, is quite

surprising. This pushes us to wonder whether these aggressive orders can potentially generate profits for HFTs. In addition to this, aggressive orders are rarely studied in the high-frequency trading literature. Consequently, we naturally consider the following question:

Question 2. *Do HFTs trade opportunely when they trade aggressively?*

When answering the question above, we establish that the aggressive trades of HFTs are driven by an informational advantage over the rest of the market. We also show that it is possible to build a classification of market participants based on the potential profit of their aggressive orders. Furthermore, we emphasise the connection between Limit Order Book (LOB) states and price moves. From a theoretical viewpoint, relationships between trades and information are at the heart of the celebrated Glosten-Milgrom model. However, this model is limited to the modelling of the best bid and ask quotes. We wish to model the whole LOB by showing how the latter can emerge from the interaction between market participants with disparities in terms of information. Doing so, our goal is in particular to quantify the queue priority value. This represents the advantage of an order placed on top of a queue in the book compared to an order placed at the bottom. We are interested in quantifying this concept because it has a crucial role in HFTs strategies. This is one of the key elements allowing fast HFTs to generate profits. Therefore, we consider the following question:

Question 3. *How could we extend the Glosten-Milgrom model to the whole LOB?*

As an answer to Question 3, we propose an agent-based model enabling us to forecast the spread according to the tick size, and to quantify the queue priority value. Thanks to the simplicity of the model, computations are explicit. However, it has some limitations. First, it restrains the activity of market makers to passive orders and that of informed traders to aggressive orders, which is not really realistic according to our answers to Questions 1 and 2. Second, it does not take into account the influence of the LOB state on trading behaviours, illustrated in the results of Question 2. Third, in this model, agents are aggregated into three groups: informed traders, noise traders and market makers. Consequently, it is impossible to disentangle disparities among behaviours of market participants belonging a priori to the same group. Therefore, based on this approach we cannot propose any kind of individual regulatory measure or compare market participants across each others. However, these disparities, highlighted in our answer to Question 1, play a key role in the impact of market participants on market quality. Furthermore, we know from our answer to Question 2 that a comparison across market participants is possible based on the potential profit of their aggressive trades. So, we want to propose a general ranking methodology of market participants according to their impact on market quality, based on their individual strategies. This leads us to the following question:

Question 4. *How to build a model for the interactions between strategies of individual market participants and use it to assess their specific contribution to market quality?*

Investigating Question 4, we design a model enabling us to derive a mathematical link between individual agents trading flows (insertion, cancellation and aggressive orders) and market quality, measured for example through the volatility. This relationship is actually in line

with our answer to Question 1 where, at the aggregated level, we empirically find a strong correlation between on the one hand market depth, aggressiveness and market share, and on the other hand volatility. The focus of Question 4 is on high-frequency market makers since they are the most active market participants. The main risk faced by these participants is related to their inventory management, which is not investigated in details when answering Question 4. Hence, the final question of this thesis is the following:

Question 5. *Do high-frequency market makers inventories have an endogenous impact on prices?*

Outline

Each chapter of this thesis addresses one of the questions above. This work can be separated into two main parts. In Part 1, we give empirical answers to Questions 1 and 2 by analysing unique datasets labelled by market participants identities, with microsecond granularity covering the trades and orders of the CAC 40 securities. In Part 2, we deal with Questions 3, 4 and 5 by proposing some suitable agent-based models taking into account empirical observations of Part 1.

In Chapter I, we answer Question 1 by studying the trades and orders of the CAC 40 French securities, on Euronext Paris from November 2015 to July 2016. First, we analyse HFTs contribution to liquidity that we measure in terms of spread and market depth at the three best limits. Second, we evaluate the trading practices of HFTs by assessing their share in traded amounts, and whether they are mainly liquidity providers or liquidity consumers. We show that the behaviour of HFTs during the day can be split into four distinct phases. Furthermore, we identify significant changes of regime in HFTs behaviour in the presence of scheduled news, going beyond the expected reaction to volatility variations. Finally, we shed light on their response towards non-scheduled announcements or pre-market announcements.

In Chapter II, we tackle Question 2 by investigating the same type of data as that of Chapter I, from September 2017 to November 2017. We start with looking at how the price impact of a single aggressive order varies according to the amount of liquidity consumed. This is why we split these orders into three groups: those consuming less than the liquidity available at the first limit (partial aggressive orders), those consuming exactly the liquidity available at the first limit (exact aggressive orders), and those consuming more than the liquidity available at the first limit (n-limit aggressive orders). Then, in contrast to several earlier studies, we show that HFTs aggressive orders have an informational advantage compared to the rest of the market, suggesting an evolution in HFTs activity in the recent years. This result is obtained by computing the potential profits of the different order flow categories (HFTs, agency participants and proprietary participants) over various time horizons, and comparing those of HFTs to the rest of the market. Finally, we display how aggressive orders can be used in order to classify market participants as HFTs or non-HFTs, or to deduce the market participants strategies.

Answer to Question 3 lies in Chapter III. Using a Glosten-Milgrom type approach, we are able to deduce the whole LOB (bid-ask spread and volume available at each price) from the interactions between three different types of agents: an informed trader, a noise trader and several market makers. This model enables us to build several relevant tools for regulators, exchanges and market participants. For instance, we provide a device to forecast consequences of a tick size change on the spread. To validate the efficiency of this method, we predict the spread changes due to the new tick size regime under the recent European regulation MiFID II, and compare our results to the effective spread values. It turns out that our predictions are very accurate: the average relative error we obtain is around 5%. This model also enables us to value quantitatively the queue position of a limit order in the book. We estimate this quantity for five small tick assets from the CAC 40 and show that the values of queue position are of the same order of magnitude as the bid-ask spreads.

We address Question 4 in Chapter IV. For this, we restrict ourselves to the modelling of the best bid and ask limits. We propose an approach based on the individual behaviours of market participants modelled by non-linear and state dependent Hawkes like processes. Our model encompasses the well-known Poisson, Queue-reactive and Hawkes Queue-reactive dynamics for order books. Under mild assumptions, we prove the ergodicity and diffusivity of our model. We also derive semi-closed formulas for the spread, imbalance and market volatility in terms of the intensities of the flows of the different market participants. Thanks to these results, we are for example able to rank market makers according to their contribution to volatility. This ranking is illustrated on several CAC 40 assets. Interestingly the obtained rankings are quite homogeneous from one asset to another.

Question 5 is answered in Chapter V. We extend the literature establishing a negative relationship between price pressure and the inventory of homogeneous market makers to the case where the agents can be heterogeneous.. The price pressure can be defined as the local price impact due to submitted market and limit orders: a highly positive (resp. negative) inventory pushes the price down (resp. up). Thanks to a unique dataset with participants level information, we illustrate this relationship empirically by conducting a multi-platform study of high-frequency market making. Such an empirical study is to our knowledge the first of its kind in the literature.

1 Part 1: Empirical analysis of high-frequency traders behaviour

In Part 1, we conduct empirical analyses in order to answer Questions 1 and 2. A market participant is identified as an HFT if either he belongs to the Supplemental Liquidity Provider (SLP) programme of Euronext, which is a market making programme, or if he meets some criteria related to the lifetime of cancelled orders, detailed in the core of the thesis. When focusing on market making activity, high-frequency market makers are defined as SLP members.

1.1 Chapter I - The behaviour of high-frequency traders under different market stress scenarios

In Chapter I, we answer Question 1: What is the intraday behaviour of HFTs, and how do they react during stress periods? We focus on the behaviour of HFTs on a daily scale (by studying their behaviour day by day) according to the level of stress measured mainly by the implied volatility, on an intraday scale (by studying their seasonal behaviour during a trading day), and around macroeconomic news. To do so, we have access to a unique labelled dataset covering the trades and orders of the CAC 40 French securities, on Euronext Paris from November 2015 to July 2016. On both daily and intraday scales, we analyse liquidity and trading activity. Liquidity is quantified by measuring the bid-ask spread and market depth at the three best limits. The HFTs trading activity is assessed by their aggressive/passive ratio (the ratio of amounts traded aggressively by HFTs over their total traded amounts) and their market share in traded amounts. We deepen our analysis by considering least square regressions for our metrics with respect to the volatility and other explanatory variables where we distinguish between days with and without macroeconomic announcements. We then focus on HFTs behaviour regarding unscheduled and pre-market announcements.

We first analyse how HFTs contribute to liquidity. We give in the following our main findings.

1.1.1 Liquidity analysis

On a daily scale Market depth decreases sharply and spreads widen when implied volatility increases.

Result 1. *Implied volatility affects all participants evenly: the HFTs market share in terms of market depth remains almost constant, independent of the implied volatility level, close to 80% at the three best price limits.*

On an intraday scale Non-HFTs provide a relatively constant amount of market depth in the order book throughout the day. For HFTs, both their market depth share and the amount they offer increase during the first hour of the day. HFTs contribute to spread tightening at the beginning of the day: their share in market depth increases while spread tightens. The levels remain stable during the day, albeit with temporary declines before the usual announcement times. At the end of the day, HFTs market depth share and HFTs amounts offered decrease rapidly. Indeed, HFTs do not generally hold overnight positions.

During macroeconomic announcements Before macroeconomic announcements, HFTs sharply reduce their liquidity provision a few minutes (generally 3 minutes) before the announcements, unlike other market participants who decrease slightly their amount of limit orders. This withdrawal before announcement times contributes to bid-ask spread widening. Then HFTs liquidity returns quickly to the order book after the announcements (generally one or two minutes later).

The main intraday HFTs and non-HFTs liquidity contribution behaviours are summarised in Figure .1.

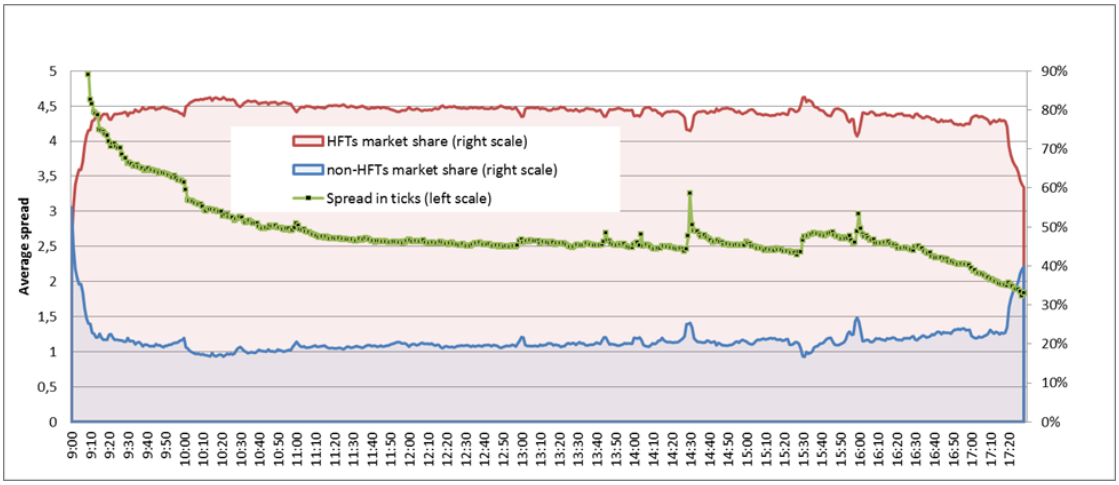


Figure .1 – Intraday evolution of HFTs and non-HFTs market share in terms of market depth compared to the bid-ask spread evolution in ticks.

Second, we analyse the trading behaviour of HFTs by measuring their market share in traded amounts and their aggressive/passive ratio with respect to the volatility and other explanatory variables. Our main findings are summarised below.

1.1.2 Trading activity analysis: market share and aggressiveness

On a daily scale HFTs share in amounts traded is slightly dependent on the implied volatility (positive correlation) while their aggressive/passive ratio is not.

Result 2. *On average, HFTs share in amounts traded oscillates around 60% and they consume more liquidity than they provide, with an aggressive/passive ratio around 53%.*

On an intraday scale The relationship between historical volatility and amounts traded remains true: the higher the historical volatility, the larger the amounts traded over all participants. However, at the very end of the day, this relationship does no longer hold: amounts traded increase sharply while historical volatility tends to decrease. Regarding the behaviour of HFTs, we can distinguish between four distinct phases:

- At the beginning of the day, HFTs share in amounts traded increases gradually from 50% to 58%. At the same time, their aggressive/passive ratio falls from 65% to 55%.
- Before the U.S. market opening, amounts traded by the market as a whole are relatively stable, as is HFTs share in amounts traded (60%) and their aggressive/passive ratio (52%).

- After the U.S. market opening, HFTs share in amounts traded rises from 58% to 65%. Their aggressive/passive ratio also increases from 52% to 55%. This surge in aggressiveness is most likely due to the appearance of arbitrage opportunities. HFTs therefore consume more liquidity than they provide.
- At the end of the day, HFTs gradually withdraw and their share in amounts traded decreases from 60% to 55%. Their aggressiveness increases due to the decrease in their market depth and probably their desire to unwind positions before the close, which tends to trigger aggressive behaviours.

During macroeconomic announcements Before an announcement, HFTs share in amounts traded decreases and their aggressive/passive ratio increases. Both features are mainly due to a decrease in passive orders because HFTs withdraw from the order book. After an announcement, prices are often severely affected. This leads to a sudden increase in aggressive flows from HFTs (as is the case after the U.S. market opening). HFTs share in amounts traded therefore rises, as their aggressive/passive ratio, now because of an increase in their aggressive orders. When the price fluctuates widely, the aggressive/passive ratio for these participants increases significantly. This probably reflects some transformation of mean reversion strategies into short-term directional (momentum) ones, which can hardly be considered as market making.

The main intraday HFTs trading behaviours are summarised in Figure .2.

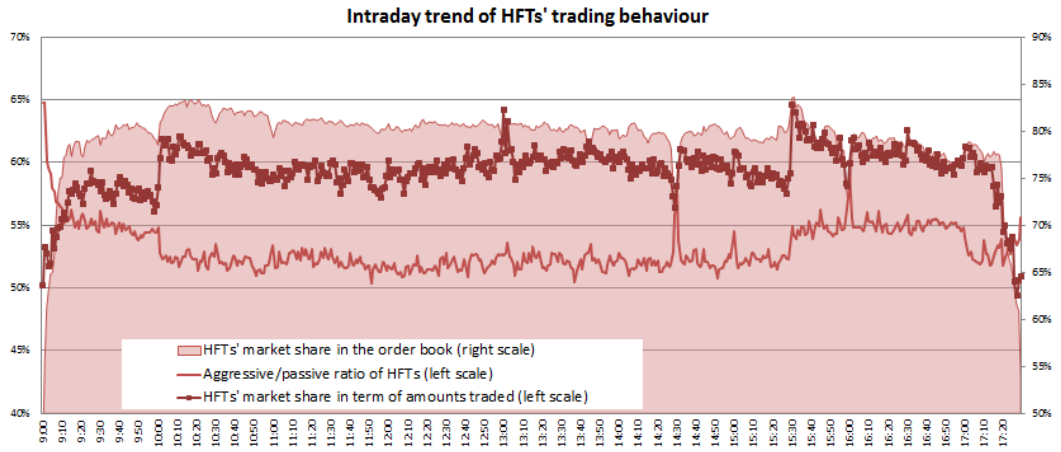


Figure .2 – Intraday evolution of HFTs aggressiveness and HFTs market share in terms of market depth and amounts traded.

Note that all these observations can significantly differ from one HFT to another: each HFT has a very different level of aggressiveness, and the distribution of the share in market depth and traded amounts is unequal across them.

1.1.3 Regression based analysis of these empirical results

We sharpen our observations related to the behaviour of HFTs during macroeconomic announcements by using least square regressions of our metrics with respect to the volatility and other explanatory variables. We distinguish between days with and without scheduled announcements at 4 p.m. All these regression results are in line with the empirical observations, and point out in a quantitative way that:

Result 3. *The behaviour of HFTs around announcements cannot be read as a simple reaction to associated variations of volatility. In fact, HFTs do contribute less to market depth before, during and after announcements, their market share in traded amounts decreases and their aggressive/passive ratio increases before and during announcements even once a volatility effect is taken into account.*

After studying in details the HFTs behaviour with respect to scheduled macroeconomic announcements, we investigate their reaction when the announcement is unscheduled or when it takes place overnight.

1.1.4 Focus on an unscheduled announcement and a pre-market announcement

We focus on the response of HFTs with respect to two specific events. The unexpected and misleading macroeconomic announcement released by the Financial Times on Twitter on December 3^{rd} 2015 concerning the monetary policy of the European Central Bank, and the Brexit referendum announcement occurring before the opening of the markets on June 24^{th} 2016.

Following the unscheduled announcement of December 3^{rd} 2015, liquidity obviously decreased. However, in opposite to what typically happens before the scheduled macroeconomic news, and despite the withdrawal of HFTs and non-HFTs from the order book, HFTs market share increased from 80% to 90%. It is likely that non-HFTs could not track the price drop and hence were unable to update their orders, whereas HFTs were able to quickly move their orders and maintain part of their passive liquidity provision in the order book. The behaviour of market participants is significantly different in the case of the Brexit pre-market announcement: HFTs diminished significantly their presence in the market, while non-HFTs partially offset HFTs reduced liquidity by providing a market depth twice larger than their usual level. We can interpret this finding as follows: when non-HFTs have time to adapt to announcements (they had the overnight in the case of the Brexit, in contrast to the situation on the 3^{rd} of December 2015), they may act as market makers.

One of our remarkable findings in this chapter is that the activity of HFTs is not limited to market making: more than 50% of their orders are aggressive. We aim at analysing these aggressive orders rarely studied in literature: are they used only for inventory management or also for directional trading? We address this issue in Chapter II.

1.2 Chapter II - The Information content of high-frequency traders aggressive orders: recent evidences

In this chapter, we answer Question 2: Do HFTs trade opportunely when they use aggressive orders? We present some evidence concerning the impact of aggressive orders on the price formation process and the information content of these orders according to the different order flow categories (high-frequency traders, agency participants, proprietary participants and retail members). For this, we conduct a study on CAC 40 stocks data from September 2017 to November 2017. Over the analysed period, we use both trade and LOB data to describe the dynamics of the LOB accurately before and after each aggressive order. The whole dataset contains approximately 8 millions aggressive orders and 423 millions events (an event can be an order insertion, an order cancellation, an order modification or a transaction). We summarise in the following our main findings.

1.2.1 Three different groups of aggressive orders

We begin with looking at how the price impact of a single aggressive order varies according to the proportion of liquidity it consumes compared to that present at the best limit. This is why we split aggressive orders into three groups:

- Partial aggressive orders: they consume less than the quantity present at the best limit.
- Exact aggressive orders: they consume exactly the quantity present at the best limit.
- N-limit aggressive orders: they consume more than the quantity present at the best limit.

Partial (resp. exact) aggressive orders constitute approximately 50% (resp. 47%) of aggressive orders in number. Partial and exact aggressive orders are unequally distributed across the different order flow categories. Indeed HFTs send more exact aggressive orders than partial ones: 63% of the exact aggressive orders are sent by HFTs, while only 39% of partial ones are sent by them. Furthermore, we find that exact aggressive orders take place when the LOB is significantly unbalanced. The imbalance at time t, just before the aggressive buy (resp. sell) order takes place, is computed as follows:

$$Imb_t = \frac{Q_t^1 - Q_t^2}{Q_t^1 + Q_t^2},$$

where Q_t^1 denotes the volume available at the best bid (resp. ask) at time t, and Q_t^2 that at the best ask (resp. bid) at time t when it is a buy (resp. sell) aggressive order. We find that the value of the imbalance one microsecond before the exact aggressive trades (on average equal to 27%) is significantly higher than that (on average equal to 3%) before the partial aggressive trades.

1.2.2 Price impact according to the groups of aggressive orders

In order to measure the price impact related to each group of aggressive orders, we first define our price impact measure.

Quantification of the price impact The price impact of an individual buy (resp. sell) aggressive order, taking place at time t and evaluated at time $t+h$ is denoted by PI_{t+h} and defined as follows:

$$PI_{t+h} = \frac{BP_{t+h} - BP_{t^-}}{S} * sign_t$$

where BP_{t^-} denotes the best ask (resp. bid) one microsecond before the buy (resp. sell) aggressive order and S the average spread of the asset. For a given flow of aggressive orders, and for a given time $t+h$, we compute the average of this measure weighted by the executed quantity across all aggressive orders. This is done from 17 minutes before till 17 minutes after the aggressive order. The results, relative to each group of aggressive orders are the following.

Result 4. *As expected, one microsecond after the aggressive order, because of the mechanical impact, the price impact due to n-limit aggressive orders is higher than that of exact ones, which is higher than that of partial ones. One relevant question is whether this mechanical impact is permanent or not. We find that the price impact of exact aggressive orders is permanent: it is above that of partial ones, over all time horizons, higher than two-thirds of the bid-ask spread. On the contrary, n-limit aggressive orders have a temporary component in their price impact: market participants tend to refill the LOB by submitting new orders in place of the consumed ones. Indeed, starting one second after the aggressive order, the price impact begins to attenuate. On a 17 minutes time horizon, the remaining mechanical impact of n-limit aggressive orders is quite equal to that of exact aggressive orders.*

We then shed light on the interest of studying the price impact according to the traded share (relative to the volume available at the best limit) instead of the traded volume.

Importance of analysing the price impact according to the traded share The average traded amount of partial aggressive order (13 k €) is close to the one of exact aggressive orders (11 k €), but their price impact is significantly different. This is a first indicator that when analysing the price impact, one should not only look at the traded volume but also at the traded proportion relative to the quantity present at the best limits. We deepen our analysis by studying the price impact of partial orders according to the consumed proportion (we separate partial orders into 10 groups according to the consumed share). We find that the magnitude of the price impact over all time horizons after the aggressive order is increasing with respect to the consumed share. We then investigate whether the consumed share rather depends on the quantity present at the best limit or on the traded amount. We obtain that the consumed part varies with the traded amount, but also depends significantly on the quantity present at the best limit. This means that the price impact does not depend only on the traded volume, but also implicitly on the volume present at the best limit. Hence the relevance of investigating the price impact according to the traded share.

Now that we know that partial aggressive orders should be studied separately from exact aggressive orders, we focus on analysing the informational advantage according to each order flow category, distinguishing between partial and exact aggressive orders. We first quantify our notion of informational advantage and then give our main results, also illustrated in Figure .3.

1.2.3 Informational advantage

Quantification of the informational advantage To estimate the informational advantage of an agent, we compute the potential profit of a buy (resp. sell) aggressive order that a market participant can realise if he unwinds his position passively at time $t+h$, denoted by PP_{t+h}:

$$PP_{t+h} = \frac{BP_{t+h} - P_t}{S} * sign_t$$

where BP_{t+h} is the best ask (resp. bid) at time $t+h$, P_t the price per share obtained by the aggressive order, S the average spread of the asset and $sign_t$ takes the value 1 (resp. -1) if it is a buy (resp. sell) aggressive order. For a given flow of aggressive orders, and for a given time $t+h$, we compute the average of this measure weighted by the executed quantity across all aggressive orders. This is done from 17 minutes before till 17 minutes after the aggressive order.

Result 5. *The HFT flow stands out with the (significantly) highest potential profit in the case of partial aggressive orders, over all time horizons. One second after partial aggressive orders, HFTs have a potential profit 0.36 spreads higher than that of agency participants, and 0.29 spreads higher than that of proprietary participants, see Figure .3. In addition to this, we show that the aggressive orders of HFTs are less autocorrelated than those of other categories. This allows us to deduce that the high potential profit of HFTs is due to an informational advantage and not to an endogenous price impact. Although HFTs still obtain a better potential profit than other market participants with exact aggressive orders, the difference between the categories is not much significant in this case.*

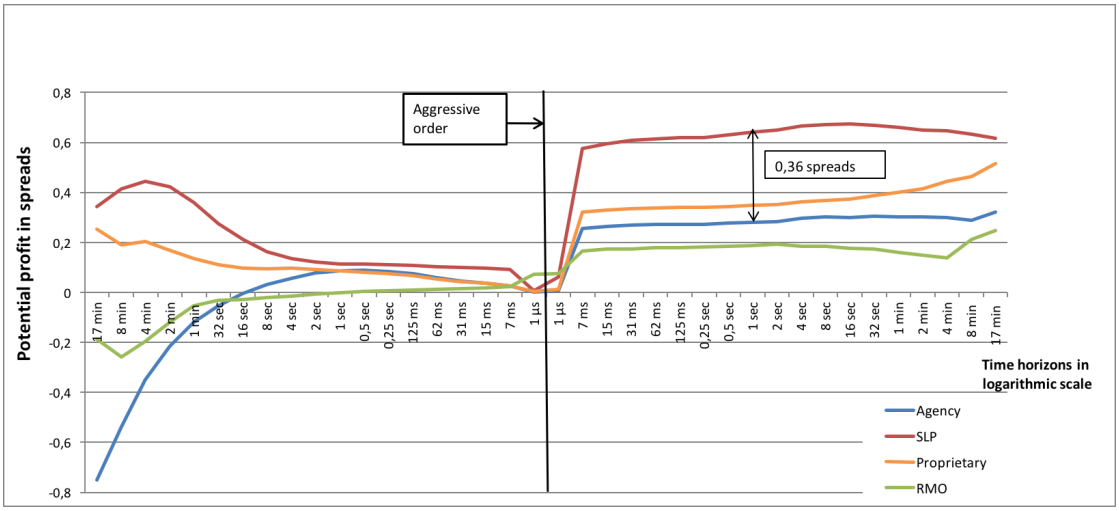


Figure .3 – The potential profit evolution following partial aggressive orders according to the different order flow categories over different time horizons.

Does the potential profit vary among members within the same order flow category? We investigate the potential profit disparities between different members belonging to the same order flow category for partial aggressive orders. We find that over a short time horizon (until 2 minutes after the aggressive order), all HFTs belong to the 25% market participants realising the highest short-term potential profits. Over a longer time horizon, from two minutes after the aggressive order, the proportion of HFTs with potential profit higher than the third quartile starts to decrease to the benefit of proprietary traders. This could be due to the fact that HFTs do not target long-term strategies, high-frequency trading being an activity where participants typically hold positions for very short times.

It turns out that the analysis of aggressive orders is useful to understand other features than price impact and potential profit. For instance, we propose a new classification of market participants based on our investigation of aggressive orders. We also show that we can access to a more granular classification by segmenting member code flows according to the different connectivity channels they use, called SLEs (French acronym for *Serveur Local d'Emission*). Finally, by observing the evolution of the price before the aggressive orders, we deduce the different strategies of member codes, such as mean reverting or trend following.

1.2.4 From single aggressive orders to strategies

A classification tool It is usual to consider cancelled orders to classify members as HFTs or non-HFTs. However, it seems also possible to classify participants by relying on aggressive order potential profits. We propose a new classification: those realising the highest short-term potential profits (one second after the aggressive order) can be considered as HFTs, and those realising the lowest as non-HFTs. It turns out that merging both approaches allows us to

obtain a more complete classification of market participants, see Figure .4. Three different classes can be distinguished:

- Pure HFTs: they are characterised by a high short-term potential profit and a low lifetime of cancelled orders.
- Pure non-HFTs: they are characterised by a high lifetime of cancelled orders and a small short-term potential profit.
- Intermediary agents: they are characterised by a small short-term potential profit and a low lifetime of cancelled orders.

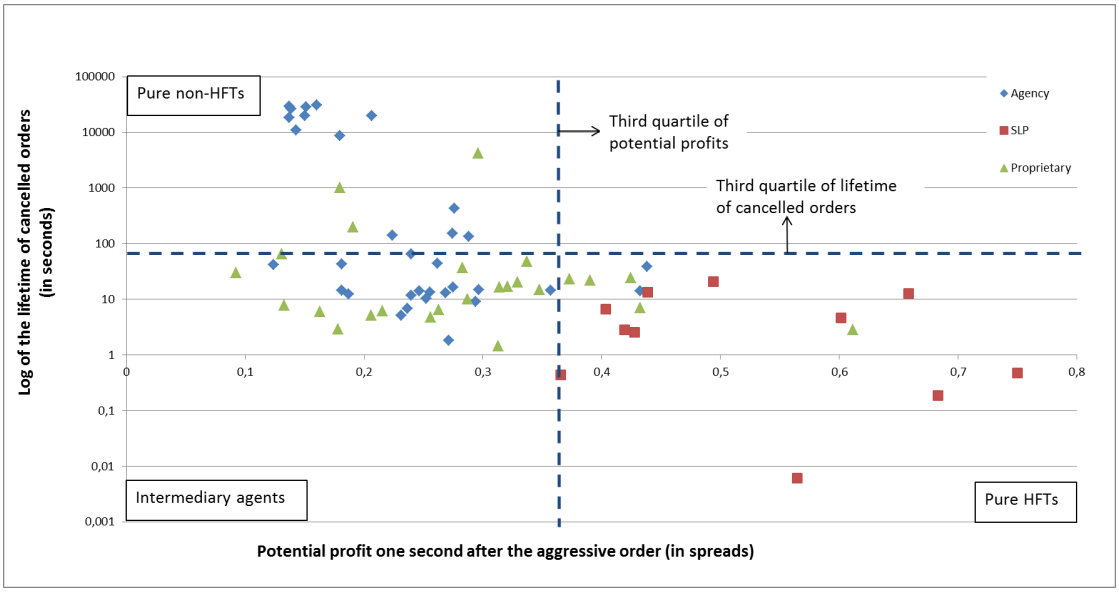


Figure .4 – Merging both classification methods: one relying on cancellations and one relying on potential profit (the flows are grouped by member code and order flow category).

We point out that, as expected, no member code has high short-term potential profits and high lifetime of cancelled orders. Moreover, note that all SLPs belong to the pure HFT category.

A more granular classification using the different connectivity channels Market members connect to Euronext via SLEs to convey their orders. Splitting the flows issued by a same member code, and belonging to the same order flow category according to these connectivity channels can in some cases bring up new information concerning the different activities followed by this member code. For example, we dissociate the flow of a given member code who is an agency broker serving as intermediary for an HFT (among other clients) according to the different SLEs. We compute the potential profit of each of these flows. We find disparities in potential profits according to the different SLEs: one SLE is probably dedicated to the HFT client, while another is dedicated to other type of clients.

Different strategies By observing what happens before the aggressive orders, we find that on average, HFTs and proprietary agents are mean reverting (they buy when the price decreases and sell when the price increases). In contrast, agency members seem on average trend following: they buy when the price increases and sell when the price decreases. Analysing the potential profit by member code, we get that purely mean reverting strategies are not followed by all SLPs. Some of them carry out distinct strategies simultaneously: mean reverting, trend following and another strategy consisting in consuming new orders within the spread.

2 Part 2: From empirical observations to agent-based modelling

In Part 2, we focus on agent-based models with the objective of providing new relevant tools for regulators and exchanges. Also, we wish to take into account the empirical observations obtained in Part 1. We answer in this part to Questions 3, 4 and 5.

2.1 Chapter III - From Glosten-Milgrom to the whole limit order book and applications to financial regulation

In Chapter III, we answer Question 3: How could we extend the Glosten-Milgrom model to the whole LOB? This means that instead of computing only best bid and ask quotes as in the seminal paper by Glosten and Milgrom, we want to be able to build from the interactions between agents the whole LOB: best bid, best ask and volume available at each limit of the LOB.

2.1.1 Efficient price and behaviour of the different market participants

In our model, we assume the existence of an efficient price $P(t)$. It is exogenous, independent of the order book dynamics and satisfies $P(t) = P(0) + Y(t)$, where $Y(t)$ is a compound Poisson process of the form: $Y(t) = \sum_{i=1}^{N(t)} B_i$, with jump rate λ^i and where the B_i are centred price jumps. We consider three different types of market participants as in the Glosten-Milgrom model:

- An informed trader: he receives the value of the price jump size B right before it happens. He then sends his trades based on this information. The informed trader trade size is denoted by Q^i. This market participant can be assimilated to an HFT using aggressive orders when he is more informed than the rest of the market. This feature stems from what we saw in Chapter II when answering Question 2.
- A noise trader: he sends random market orders that follow a compound Poisson process with arrival rate λ^u. The noise trader order size is denoted by Q^u and its cumulative distribution function is denoted by F_{κ^u}.
- Several market makers: they receive the value of the price jump size B right after it happens. We assume that they are risk neutral. They know the proportion of price jumps compared to the total number of events happening in the market which is denoted by $r = \frac{\lambda^i}{\lambda^i + \lambda^u}$.

2.1.2 LOB modelling with a zero tick size

We start with the case where the tick size is assumed to be equal to zero. The obtained results help us to understand those when the tick size is positive. We write L for the cumulative LOB shape function on which we make no a priori assumption (for example it can have a singular part and discontinuities).

The informed trader computes his gain according to the future efficient price. If he knows that the price will increase (resp. decrease), which corresponds to a positive (resp. negative) jump B, he consumes all the sell (resp. buy) orders leading to positive ex-post profit. This is formalised in the assumption below.

Assumption 1. *The informed trader sends his trade in a greedy way such that he wipes out all the available liquidity in the LOB till the level $P(t)+B$. Thus his trade size satisfies: $Q^i = L(B^-)$.*

Regarding the market makers, they compute the conditional average profit of a new infinitesimal order if submitted at price level x knowing that $Q > L(x)$ and without any information about the trade initiator. This quantity is denoted by $G(x)$ and its value is given in the next proposition.

Proposition 1. *The average profit of a new infinitesimal order if submitted at price level x satisfies:*

$$G(x) = \begin{cases} x - \frac{r\mathbb{E}[B1_{B>x}]}{r\mathbb{P}[B>x]+(1-r)\mathbb{P}[Q^u>L(x)]} & \text{if } x \geq 0 \\ -x + r\frac{\mathbb{E}[B\mathbf{1}_{B<x}]}{r\mathbb{P}[B<x]+(1-r)\mathbb{P}[Q^u<L(x)]} & \text{if } x \leq 0. \end{cases}$$

For a given x, market makers compute the theoretical $\hat{L}(x)$ for which $G(x) = 0$. Then they act according to the following zero-profit assumption.

Assumption 2. *For every $x > 0$ (resp. $x < 0$), if $\hat{L}(x) \leq 0$ (resp. $\hat{L}(x) \geq 0$), market makers add no liquidity to the LOB: $L(x) = 0$. If $\hat{L}(x) > 0$ (resp. $\hat{L}(x) < 0$), because of competition, the cumulative order book adjusts so that $G(x) = 0$. We then obtain $L(x) = \hat{L}(x)$.*

Note that under our zero-profit assumption, market makers who place first their orders in the queue, can still make profits. In this setting, we can show the emergence of the bid-ask spread and the LOB shape. We have the following result.

Result 6. *The cumulative LOB satisfies $L(x) = -L(-x)$. Moreover, $L(x) = 0$, for $x \in (-\mu, \mu)$, and $L(x)$ is continuous and strictly increasing for $x \geq \mu$, where μ is the unique solution of the following equation:*

$$\frac{1+r}{2r} = \mathbb{E}[\max(\frac{B}{\mu}, 1)]. \tag{1}$$

For $x > \mu$, $L(x) > 0$:

$$L(x) = F^{-1}_{\kappa^u}(\frac{1}{1-r} - \frac{r}{1-r}\mathbb{E}[\max(\frac{B}{x}, 1)]).$$

For $x < -\mu$, $L(x) > 0$:

$$L(x) = -F^{-1}_{\kappa^u}(\frac{1}{1-r} - \frac{r}{1-r}\mathbb{E}[\max(\frac{B}{-x}, 1)]).$$

In particular, the bid-ask spread is equal to 2μ.

Furthermore, the variance per trade σ^2_{tr} *satisfies:*

$$\sigma^2_{tr} = \mathbb{E}[(P_{\tau_{i+1}} - P_{\tau_i})^2] = \frac{\mathbb{E}[B^2]\mu}{\mathbb{E}[|B|\mathbf{1}_{|B|>\mu}]}.$$

Equation (1) shows that the spread emerges naturally from adverse selection risk. Indeed, it is an increasing function of r. This means that market makers are aware of the adverse selection they face when the number of price jumps increases. As a consequence, they enlarge the spread in order to avoid this effect due to the trades issued by the informed trader just before the price jumps take place. This in line with what we show in Chapter I: market makers withdraw from the market before the announcements because they are aware of the adverse selection risk.

2.1.3 LOB modelling with a non-zero tick size

Now we consider the case where the tick size α is non-zero. We denote by d the distance between the smallest possible price level that is greater than or equal to the current efficient price $P(t)$; $d \in [0, \alpha)$. We write $l^d(i)$ for the quantity placed at the i^{th} limit and the cumulative volume at level i is denoted by $L^d(i)$:

$$L^d(i) = \begin{cases} L(d + (i-1)\alpha) & \text{for } i > 0 \\ L(d + i\alpha) & \text{for } i < 0. \end{cases}$$

Considering the same assumptions as in the case where the tick size is vanishing, and computing market makers profits, the bid-ask spread and LOB shape emerge in the following way.

Result 7. *The LOB shape function satisfies* $l^d(i) = 0$ *for all* $-k^d_l < i < k^d_r$, *where* k^d_l *and* k^d_r *are two positive integers determined by the following equations:*

$$k^d_r = 1 + \lceil \frac{\mu - d}{\alpha} \rceil, \quad k^d_l = \lceil \frac{\mu + d}{\alpha} \rceil,$$

with μ *defined by Equation* (1), *and where* $\lceil x \rceil$ *denotes the smallest integer that is larger than* x *(which can be equal to 0). Furthermore, for* $i \geq k^d_r$:

$$L^d(i) = F^{-1}_{\kappa^u}\Big(\frac{1}{1-r} - \frac{r}{1-r}\mathbb{E}[\max(\frac{B}{d + (i-1)\alpha}, 1)]\Big)$$

and for $i \leq -k^d_l$:

$$L^d(i) = F^{-1}_{\kappa^u}(\frac{-r}{1-r} + \frac{r}{1-r}\mathbb{E}[\max(\frac{B}{d + i\alpha}, 1)]).$$

For a given d, *the bid-ask spread* ϕ^d_α *satisfies:*

$$\phi^d_\alpha = \alpha\Big(\lceil \frac{\mu - d}{\alpha} \rceil + \lceil \frac{\mu + d}{\alpha} \rceil\Big).$$

Considering the approximation that d is uniformly distributed over $[0, \alpha]$, we get that the average spread ϕ_α satisfies:

$$\phi_\alpha = 2\mu + \alpha. \tag{2}$$

Furthermore, the variance per trade is:

$$\sigma_{tr}^2 = \mathbb{E}[(P_{\tau_{i+1}} - P_{\tau_i})^2] = \frac{\mathbb{E}[B^2](\mu + \alpha/2)}{\mathbb{E}[|B|\mathbf{1}_{|B|>\mu+\alpha/2}]}. \tag{3}$$

2.1.4 Relevant tools for regulators and exchanges

Now we shed light on new useful methods that can be derived from our model: a spread forecasting device in case of a change of tick size and the quantitative computation of the priority value, which is the advantage of an order placed on top of a queue in the book compared to an order placed at the bottom.

Spread forecast We obtain that the bid-ask spread is the sum of the tick value and of the intrinsic bid-ask spread arising in the zero tick size case μ, see Equation (2). This shows that the spread is not only a response to adverse selection risk, but that the tick size does play a crucial role. Based on Equation (2), it is possible to predict the spread value after a tick size change since μ is an invariant of the considered asset.

We study the CAC 40 stocks over a six months time period around the implementation of MiFID II regulation: from October 2017 to December 2017 (before the tick size changes) and from January 2018 to March 2018 (after the tick size changes). We consider assets whose tick size has changed after the implementation of MiFID II. This is the case of 14 CAC 40 stocks. We note that for all these assets, the tick size increased. We now forecast the new spreads of our 14 assets due to the new tick size regime, based on pre-MiFID II data. The forecasts are presented in Table .1.

Stock	Tick size before MiF. II	Tick size under MiF. II	Average spread before MiF. II (euros)	Average spread before MiF. II (ticks)	Average spread under MiF. II (euros)	Average spread under MiF. II (ticks)	Expected spread based on our model	Relative error
Accor	0.005	0.01	0.011	2.266	0.016	1.586	0.016	3%
Bouygues	0.005	0.01	0.011	2.277	0.017	1.734	0.016	5%
Kering	0.05	0.1	0.090	1.797	0.141	1.407	0.140	1%
Legrand	0.01	0.02	0.016	1.643	0.029	1.471	0.026	10%
Publicis	0.01	0.02	0.019	1.904	0.030	1.520	0.029	4%
Safran	0.01	0.02	0.019	1.892	0.031	1.556	0.029	7%
Schneider Electric	0.01	0.02	0.016	1.579	0.025	1.235	0.026	4%
TechnipFMC	0.005	0.01	0.010	2.056	0.017	1.677	0.015	10%
Valeo	0.01	0.02	0.018	1.845	0.031	1.568	0.028	10%
Veolia Environnement	0.005	0.01	0.007	1.440	0.012	1.189	0.012	3%
Vinci	0.01	0.02	0.017	1.668	0.026	1.280	0.027	4%
Vivendi	0.005	0.01	0.007	1.408	0.012	1.162	0.012	4%

Table .1 – Forecasting CAC 40 assets spreads under MiFID II.

The forecasts based on our model are very accurate: the average relative error is equal to 5%. Thus it seems that quantitative approaches like ours could be a way to avoid pilot programs on tick sizes as used for example in the United States and Japan.

Priority value quantification The introduction of the tick size enables us to quantify the priority value. The value of queue position at the i^{th} level, denoted by $\tilde{G}^d(i)$, can be formulated in this model as the difference between the expected profit of the order placed on top and that of a new one that would be placed at the bottom of the i^{th} queue. We have the following result.

Result 8. *The value of the queue position $\tilde{G}^d(i)$ for $i \geq k_r^d$ satisfies:*

$$\tilde{G}^d(i) = d + (i-1)\alpha - \frac{r\mathbb{E}[B\mathbf{1}_{B>d+(i-1)\alpha}]}{1 - rF_\psi(d+(i-1)\alpha) - (1-r)F_{\kappa^u}\big(L^d(i-1)\big)}. \tag{4}$$

The formula for $i \leq -k_l^d$ is obviously deduced.

To apply Equation (4) on real data, we need to know the distribution of B. To do so, we resort to Equation (3). Recall that for small tick assets, we should obtain a linear relationship between the volatility per trade and the spread, with a slope coefficient between 1 and 2. So we conclude that a suitable choice, enabling us to satisfy the above relationship is to consider a Pareto distribution for the absolute value of the efficient price jumps with parameters $k > 2$ (the shape) and x_0 (the scale), with $k > 2$ in order to have a finite variance. To ensure that the variance per trade is proportional to the square of the spread we naturally take $x_0 = \mu + \frac{\alpha}{2}$. In this case, the variance per trade becomes:

$$\sigma_{tr}^2 = \frac{k-1}{k-2}(\mu + \alpha/2)^2. \tag{5}$$

For any given small tick asset, we can estimate k by minimising the quadratic error between the actual variance per trade and the one deduced from Equation (5).

We apply these results on all CAC 40 assets with average spread higher than 2 ticks, from January 2018 to March 2018. This corresponds to 5 assets. We find that the values of queue position at the best limits are of the same order of magnitude as the bid-ask spreads, see Table .2.

Stock	Spread (euros)	Spread (ticks)	k	r	First limit priority value (euros)	First limit priority value (spreads)	Second limit priority value (euros)	Second limit priority value (spreads)	Third limit priority value (euros)	Third limit priority value (spreads)	Fourth limit priority value (euros)	Fourth limit priority value (spreads)
Airbus	0.020	2.040	3.478	71%	-0.005	-25%	0.012	61%	0.015	72%	0.014	67%
Lafarge Holcim	0.026	2.438	3.316	70%	-0.008	-31%	0.011	42 %	0.014	55%	0.013	52%
Renault	0.025	2.476	2.866	65%	-0.007	-30%	0.011	45%	0.013	54%	0.013	51%
Saint-Gobain	0.010	2.035	4.776	79%	-0.003	-25%	0.007	65%	0.009	84%	0.008	78%
Société Generale	0.010	2.012	9.910	90%	-0.003	-25%	0.007	73%	0.012	117%	0.013	127%

Table .2 – Queue position values at the four best limits for $d = 0.5\alpha$.

In this model we are able to provide quantitative elements related to the microstructure (LOB shape, spread forecasting device and computation of priority values) at the level of the asset, but not at that of individual agents. Now we want to compare individual market participants, notably in term of their contribution to market quality. This is why we aim at proposing a model based on the individual behaviour of agents, allowing us to rank them with respect to the impact of their trading on market stability.

2.2 Chapter IV - From asymptotic properties of general point processes to the ranking of financial agents

Chapter IV answers Question 4: How to build a model for the interactions between strategies of individual market participants and use it to assess their contribution to market quality? Instead of aggregating agents into groups as in the approach in Chapter III, we want to model the specific flows of each agent taking part in the market. To do so, we restrict ourselves to the modelling of the best bid and ask dynamics. We propose a very general framework where individual behaviours are modelled using point processes that can be seen as state-dependent and non-linear Hawkes-type processes. The well-known Poisson, Queue-reactive and Hawkes Queue-reactive dynamics for order books are particular cases of our setting. Using this approach, we establish theoretical results allowing us to assess the specific contributions of agents to market quality.

2.2.1 Modelling of the best bid and ask dynamics

We use an event by event approach. Each event is characterised by $(T_n, X_n) \in (\mathbb{R}^+, E)$ where:

- T_n is the time of the n^{th} event.
- X_n is a variable encoding the characteristics of the event:

- size s_n: an integer representing the order size.
- price p_n: equals to $k \in \mathbb{N}$ when the order is inserted at the price best bid $+k\tau_0$, where τ_0 is the tick size.
- direction d_n: + if it provides liquidity and − when liquidity is removed or consumed.
- type t_n^o: 1 (resp. 2) when the bid (resp. ask) is modified.
- agent a_n: the market consists in N agents.

The order book state is modelled by the process $U_t = \left(Q_t^1, Q_t^2, S_t\right)$ where Q_t^1 (resp. Q_t^2) is the best bid (resp. ask) quantity and S_t is the spread.

Generalised intensity and market reconstitution The intensity $\lambda_t(e)$ associated to an event $e \in E$ can be informally defined by

$$\lambda_t(e) = \lim_{\delta t \to 0} \frac{\mathbb{P}\left[\#\{T_n \in (t, t+\delta t], X_n = e\} \geq 1 | \mathscr{F}_t\right]}{\delta t},$$

where $\mathscr{F}_t$ is the sigma-algebra representing information from the history of the market.

We assume it depends on the past event and current state of the market in the following way:

$$\lambda_t(e) = \psi\Big(e, U_{t^-}, t, \sum_{T_i < t} \phi(e, U_{t^-}, t - T_i, X_i)\Big),$$

where ψ is a possibly non-linear function, U_{t^-} is the order book state relative to the last event before t and ϕ is the Hawkes-like kernel representing the influence of the past events. The functions ϕ and ψ are both $\mathbb{R}_+$-valued. In absence of the kernel ϕ, this corresponds to the Queue-reactive model (and to the zero-intelligence Poisson model when ψ does not depend on U_{t^-}). When ϕ is non-zero, ψ represents the interaction between the past events and the current order book state. Moreover, we allow ψ to have polynomial growth while in the literature it has typically at most linear growth. In this case, this model generalises the Hawkes Queue-reactive approach. Note that this intensity encompasses the Poisson intensity, the Queue-reactive intensity, the Hawkes Queue-reactive intensity and the Quadratic Hawkes intensity. The market intensity $\lambda_t^M(e')$ of an anonymous event e' (e' does not contain the agent identity) is given by

$$\lambda_t^M(e') = \sum_{a \leq N} \lambda_t\big((e', a)\big).$$

Ergodicity Our first theoretical result is relative to the ergodicity of our limit order book model.

Result 9. *Under suitable assumptions, $\bar{U}_t = (Q_t^1, Q_t^2, S_t, \lambda_t)$ is ergodic: there exists a probability measure $\bar{\pi}$ such that (exponential speed of convergence)*

$$\lim_{t \to \infty} P_t(u, A) = \bar{\pi}(A), \quad \forall u, A,$$

where u is an initial condition which is here a càdlàg function from $(-\infty, 0]$ into $(\mathbb{R}^+)^4$.

From the ergodic propriety, we can derive asymptotic results for long-term behaviour of our system.

Scaling limits The reference price after n jumps P_n writes $P_n = P_0 + \sum_{i=1}^{n} \Delta P_i$ where $\Delta P_i = P_i - P_{i-1} = \eta_i$ and $\mathbb{E}[\eta_i] = 0$. We now assume that η_i is centred and $\eta_i = f(U_i)$ for some measurable function f and consider the process

$$X_n(t) = \frac{P_{\lfloor nt \rfloor}}{\sqrt{n}}, \qquad \forall\, t \geq 0.$$

The next result describes the behaviour of the price at the macroscopic scale (in event time, the result in calendar time is very similar and also provided in the core of Chapter IV).

Result 10. *Under the stationary distribution, the quantity $X_n(t)$ satisfies the following convergence result:*

$$X_n(t) \xrightarrow{\mathscr{L}} \sigma W_t,$$

with $\sigma^2 = \mathbb{E}_\mu[\eta_0^2] + 2\sum_{k\geq 1} \mathbb{E}_\mu[\eta_0 \eta_k]$ and μ the stationary distribution.

This results relates the individual order flow intensities and the macroscopic volatility of the asset. It enables us to rank market participants according to their contribution to volatility. Note that here μ is the stationary distribution on the function space since the marginal law of η_0 is not enough to compute the second term in the volatility in the non-Markov case. However, the first term can indeed be readily obtained from π. We now explain how empirical computations of relevant quantities can be made.

2.2.2 Explicit computations of market quantities

In this model, we can derive semi-explicit formulas for the stationary distribution, expected spread, price volatility and intensities of the fluctuations of the cumulated imbalance. In the following, we provide those relative to the stationary distribution, expected spread and the price volatility.

Stationary probability computation We write π for the stationary distribution of the process U_t. We have the following result which gives us a numerical methodology to compute π.

Result 11. *The distribution π satisfies*

$$\begin{aligned} \pi Q &= 0 \\ \pi \mathbf{1} &= 1. \end{aligned} \tag{6}$$

where the infinite dimensional matrix Q verifies

$$Q(z, z') = \sum_{e \in E(z,z')} \mathbb{E}_\mu[\lambda(e)|],$$

with $E(z, z')$ the set of events directly leading to z' from z.

In order to compute $Q(z,z')$, let us take z and z' two states such that $z \neq z'$, $N_t^{z,z'} = \sum_{T_i<t} \delta^i_{u,u'}$ with $\delta^i_{z,z'} = \mathbf{1}_{U_{T_{i-1}}=z, U_{T_i}=z'}$ and $t^u = \sum_{T_i<t} \Delta T_i \mathbf{1}_{\zeta_{T_{i-1}}=z}$ with $\Delta T_i = T_i - T_{i-1}$. The matrix Q can be estimated the following way.

Result 12. *We have*

$$\hat{Q}(u,u') = \frac{N_t^{u,u'}}{t^u} \underset{t\to\infty}{\to} Q(u,u'), \quad a.s. \tag{7}$$

Note that the form of the estimator of $Q_{u,u'}$: $N_t^{u,u'}/t^u$, and hence π, does not depend on the model.

Spread computation We recall that the spread S is a state variable since $U = (Q^1, Q^2, S)$. Thus, the expected value of the spread $\bar{S}$ under the stationary distribution satisfies

$$\bar{S} = \mathbb{E}_\pi[S] = \sum_{q^1,q^2,s} \pi(q^1,q^2,s)s.$$

We end with the computation of the volatility in practice when restricting ourselves to the Markov case (for simplicity here).

Price volatility computation in the Markov case In the Markov case, the second term in the volatility formula in Result 10 can be easily computed from the marginal stationary distribution π. Let us define P the transition matrix of Markov chain associated to U (transitions after one jump): $P_{u,u'} = -Q_{u,u'}/Q_{u,u}$. In this setting, we have (recall that $\eta_i = f(U_i)$):

$$\mathbb{E}_\mu[\eta_0\eta_k] = \mathbb{E}_\pi[\eta_0\eta_k] = \sum_u \pi(u) f(u) \mathbb{E}_u[\eta_k], \quad \mathbb{E}_u[\eta_k] = \sum_{u'} P^k_{u,u'} f(u'). \tag{8}$$

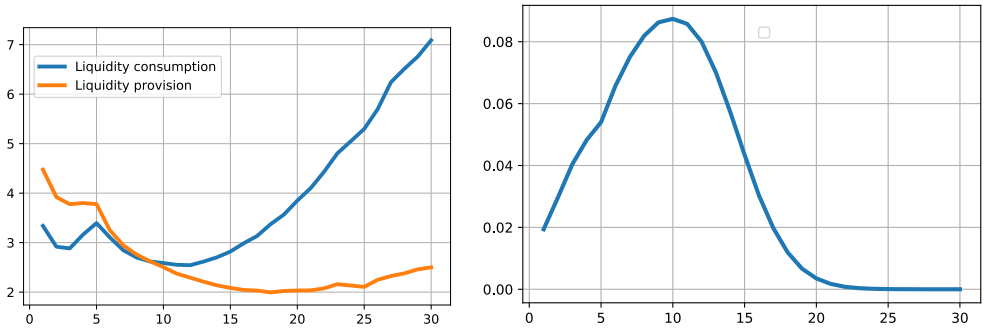
2.2.3 Numerical experiments

Using Result 10 and placing ourselves in the Markov case, we propose a ranking of the nine main market makers based on their impact on volatility on four large tick assets (for which the model is very suitable): Air Liquide, EssilorLuxottica, Michelin and Orange, on Euronext, over a one year period: from January 2017 till December 2017.

For each asset, we compute first the liquidity provision and consumption intensities relative to the whole market using Equation (7). Then, we estimate the stationary measure of the order book using Equation (6). Finally we obtain the macroscopic volatility using Equation (8).

We give in Figure .5 the results relative to Air Liquide.

(a) Intensity of the market (b) Stationary measure Q^1



Long-term price volatility $\sigma^2 = 0.227$.

Figure .5 – (a) Liquidity insertion and consumption intensities (in orders per second) with respect to the queue size (in average event size) and (b) the corresponding stationary distribution of (Q^1) with respect to the queue size (in average event size), proper to Air Liquide.

Thereafter, for each market maker, we estimate the new market volatility based on our model in a situation where we suppose that he withdraws from the market. The market maker who is so that without him the volatility increases (resp. decreases) the most is ranked first (resp. last). This is because he is seen as the most stabilising (destabilising) participant. In the following table, we establish the ranking of market makers which is actually quite stable across assets.

Market maker	Ranking Air Liquide	Market share Air Liquide	Ranking Exilor-Luxottica	Market share Exilor-Luxottica	Ranking Michelin	Market share Michelin	Ranking Orange	Market share Orange
MM1***	4	4%	3	3%	3	4%	3	3%
MM2	9	1%	9	1%	9	1%	7	1%
MM3	6	5%	6	5%	7	4%	5	4%
MM4	5	1%	4	1%	4	0%	4	1%
MM5	7	5%	8	5%	8	5%	9	5%
MM6****	1	3%	2	3%	1	3%	1	4%
MM7****	2	7%	1	12%	2	9%	2	7%
MM8*	3	9%	5	5%	5	5%	6	4%
MM9	8	2%	7	2%	6	2%	8	2%

Table .3 – Market share and ranking of markets makers. We put the symbol * next to the name of the market maker each time he is decreasing the volatility of an asset.

In this chapter, we establish a link between market makers flows and market quality, notably the volatility. One of the main risks that market makers face is related to their inventory management, which is not investigated in details in this chapter. This is why we are interested in Chapter V in studying the relationship between market makers inventories and price pressure.

2.3 Chapter V - Market makers inventories and price pressure: theory and multi-platform empirical evidences

Chapter V answers Question 5: Do high-frequency market makers inventories have an endogenous impact on prices? To address this issue, we propose a model where market makers optimise their quotes in order to maximise their wealth and reduce their inventory risk. In this setting, we establish a negative relationship between market makers inventories and price pressure. Using a unique multi-platform dataset, we provide original empirical evidences that market makers perform cross-market netting of inventories. Furthermore, we show that our negative relationship holds empirically. To do so, we compute the actual inventories of market makers, instead of estimating them from the cumulative net volume as typically done in the literature.

2.3.1 Agents behaviour

We consider that the efficient price S_t follows a Brownian motion with volatility σ. We denote by $S_t + \tilde{a}_t$ and $S_t + \tilde{b}_t$, respectively the ask and bid quotes offered by the competitive market makers. Thus $\tilde{a}_t$ and $\tilde{b}_t$ measure the deviation of ask and bid quotes from the efficient price of the asset.

We distinguish two different types of agents:

- Agents consuming liquidity: buyers and sellers.
- Market makers providing liquidity.

Buy and sell investors Buyers and sellers arrive randomly to the market. Buyers (resp. sellers) have a reservation price $S_t + \tilde{p}$ (resp. $S(t) - \tilde{p}$). When the ask (resp. bid) price is lower (resp. higher) than this reservation price, buyers (resp. sellers) consume all the liquidity present between the reservation price and the ask (resp. bid) price. The demand and supply functions of the buy and sell investors are respectively given by

$$Q^B(\tilde{a}_t) = c(\tilde{p} - \tilde{a}_t), \quad Q^S(\tilde{b}_t) = c(\tilde{b}_t + \tilde{p}), \tag{9}$$

where $c > 0$.

Market makers There are N market makers in the market. For any $l = 1, 2, \ldots, N$, Market Maker l chooses a predictable control strategy $x^l \equiv (x_t^{l,a}, x_t^{l,b})_{t\in[0,\infty)}$ specifying the size of buy and sell orders that will be executed against the randomly arriving buy and sell investors. Market makers being the only counterparty and using Equation (9), we obtain the following ask and bid price pressures at time t:

$$\tilde{a}_t = \tilde{p} - \frac{1}{c}\sum_{l=1}^{N} x_t^{l,a}, \qquad \tilde{b}_t = -\tilde{p} + \frac{1}{c}\sum_{l=1}^{N} x_t^{l,b}.$$

Market maker l chooses his strategy x^l by maximising his expected utility from wealth computed relatively to the efficient price, minus a quadratic penalty denoted by Θ for holding inventory. His value function is given by

$$\bar{v}_l(i) = \sup_{x^l} \mathbb{E}\left[\int_0^\infty e^{-\beta t}(dW_s^{l,(x^l,x^{-l})} - \Theta(I_s^{l,x^l})^2 ds) | I_t^{l,x^l} = i\right], \tag{10}$$

where $\beta > 0$ is a positive constant, $W_t^{l,x}$ is the wealth of Market Maker l at time t and I_t^{l,x^l} the inventory of Market Maker l when applying his strategy x^l.

2.3.2 Theoretical relationship between price pressure and inventory

Now, using the dynamic programming principle, we compute the optimal strategy of each market maker from which we deduce the bid and ask pressures. The solutions we obtain depend on whether market makers are identical or heterogeneous.

Identical market makers We consider first the case where market makers are identical: their costs for holding inventory and their initial inventories are equal. We show that the value function $\bar{v}_l(i)$ is quadratic and concave in i. More precisely we obtain the following result.

Result 13. *In the case of N identical market makers, the ask and bid price pressure policy functions are time-homogeneous, and given respectively by*

$$\begin{cases} \tilde{a}(i) = \dfrac{\tilde{p}(1+2cA^\star) - 2NA^\star i}{N+1+2cA^\star}, \\ \tilde{b}(i) = \dfrac{-\tilde{p}(1+2cA^\star) - 2NA^\star i}{N+1+2cA^\star}, \end{cases}$$

where $A^\star > 0$.

This result extends to the multi-market makers framework the negative relationship between price pressure and inventory of a monopolistic market maker already considered in the literature.

Heterogeneous market makers To go further, we also study the case of two heterogeneous market makers: their costs for holding inventory are the same but their initial inventories are not. In this situation, using approximations, we obtain the following result.

Result 14. *For any $\Theta > 0$ sufficiently small, there exists a solution of the maximisation problem of market makers such that a negative relationship between market makers aggregated net inventories and price pressure holds.*

2.4 Multi-platform empirical analysis

We study cross-platform transactions of two assets: Société Générale and Renault during June 2017 (with 21 trading days) with participant-level information and millisecond time granularity.

2.4.1 Evidence about market makers cross-market inventory netting

Using cross-platform trades allows us to track down the supposed flat position of market makers at the end of the day. For instance, we can see in Figure .6 the difference between the intraday inventory evolution on the 1^{st} of June of one specific market maker on the asset Société Générale when considering Euronext only and when taking into account all the platforms.

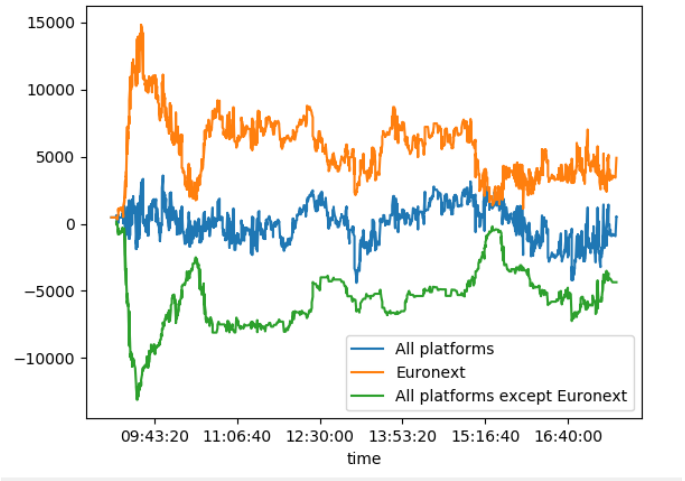


Figure .6 – The intraday inventory (in shares) on the 1^{st} of June of a specific market maker on Euronext vs other platforms (Société Générale).

Based on end-of-day inventory criteria, we identify four market participants playing the role of market makers. Note that one of them having a large part of his activity OTC, he is only partially analysed.

2.4.2 Market makers aggressiveness

In our model, market makers are purely passive, which is not the case in practice. Hence we study whether market makers use aggressive orders for arbitrage opportunities only or also for inventory management.

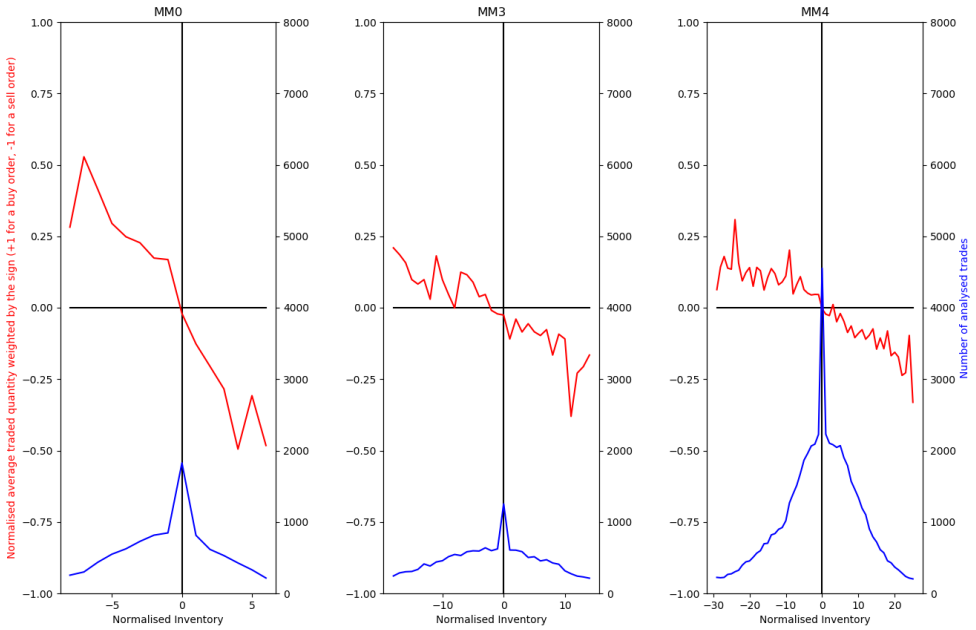


Figure .7 – Aggressive orders and inventory management for three market makers for Société Générale. The red plot corresponds to the average sign of aggressive orders weighted by the traded volume (left y-scale) and the blue plot corresponds to the analysed number of orders (right y-scale).

Figure .7 shows that the average sign of aggressive orders weighted by the traded volume seems to be a decreasing function with respect to the inventory: when the inventory is negative (resp. positive), market makers send more aggressive buy (resp. sell) orders than aggressive sell (resp. buy) orders. This and other conclusions in Chapter V lead us to the following result.

Result 15. *Aggressive orders are not used for arbitrage opportunities only, but for inventory management reason too. But in contrast to the belief that market makers are more aggressive when their inventory is large, we show that their aggressive/passive ratio (which is equal to the volume traded aggressively over the volume traded aggressively and passively) remains approximately constant with respect to their inventory.*

2.4.3 Empirical evidences about the negative relationship between price pressure and market makers inventories

Since we have shown that aggressive orders are also used for inventory management, we study the price pressure of both aggressive and passive orders according to the inventory. The ask

(resp. bid) pressure is considered when the market maker sends an aggressive buy (resp. sell) order or a sell (resp. buy) passive order.

Quantification of the price pressure On a given platform, the ask pressure due to the n^{th} buy aggressive order, denoted by AP_n^{ag}, is measured as follows:

$$AP_n^{ag} = A_{n+1}^{ag} - A_n^{ag},$$

where A_n^{ag} is the best ask value on the considered platform just before the n^{th} buy aggressive order.

On a given platform, the ask pressure due to the n^{th} (executed) sell passive order, denoted by AP_n^{pass}, is measured as follows:

$$AP_n^{pass} = A_n^{pass} - A_{n-1}^{pass},$$

where A_n^{pass} is the best ask value on the considered platform just before the n^{th} executed passive order is executed.

Our measures for the bid pressure are obviously deduced.

Individual regressions For each market maker, we consider all the price pressures corresponding to his aggressive and passive orders. We compute linear regressions to explain price pressures by individual market makers inventories. We give for instance in Table .4 the result relative to the market maker MM0 for Société Générale.

Asset	Variable	Coefficient	Std err	t-stat	p-value	[0.025 0.975]	Number of observations
Société Générale	MM0 inventory	-0.0004	$2.25e^{-5}$	-19.503	0.000	[-0.000 -0.000]	41771

Table .4 – Regression on MM0 price pressure according to his inventory for Société Générale.

The results in Table .4 perfectly agree with the theoretical relationship between price pressure and inventories.

Regression in the case of multiple market makers To show empirically that the negative relationship between market makers inventories and price pressure holds in the case of multiple heterogeneous market makers, we compute a regression to explain price pressure due to all considered market makers together by their cumulated inventory.

Asset	Variable	Coefficient	Std err	t-stat	p-value	[0.025 0.975]	Number of observations
Société Générale	Constant	-0.0002	$8.18e^{-5}$	-2.181	0.029	[-0.000 $-1.81e^{-5}$]	81960
Société Générale	Sum of inventories	$-8.314e^{-5}$	$5.55e^{-6}$	-14.989	0.000	[$-9.4e^{-5}$ $-7.23e^{-5}$]	81960

Table .5 – Regression on price pressures due to all market makers together according to their cumulated inventory for Société Générale .

In agreement with the theory, Table .5 shows that the negative relationship between price pressure and market makers cumulated inventory does hold at the aggregated level.

Part 1

Empirical analysis of high-frequency traders behaviour

CHAPTER I

The behaviour of high-frequency traders under different volatility market stress scenarios

Abstract

This empirical study on European stocks gives evidence about the practices of high-frequency traders (HFTs) under market stress. In the absence of significant news, whatever the market conditions, they are the main contributors to liquidity with a participation of 80% in the market depth. They constitute 60% of the traded amounts, with an aggressive/passive ratio around 53%. We identify a change of regime in the presence of scheduled news that goes beyond the expected reaction to volatility variations. Moreover, in extreme situations, when non-HFTs have time to adjust their tactics, they act as liquidity providers in place of HFTs.

1 Introduction

In the past few decades, technology has transformed financial markets. Before the advent of computers, all trading was conducted between humans. In 1983, Bloomberg introduced the world's first computerised system to provide real-time price feed and analytics to Wall Street firms. In response, market participants began to develop trading algorithms. Over the past 15 years or so, speed has become more and more important in the markets. That is how high-frequency traders (HFTs) have emerged. In the literature, several definitions of HFTs can be found. According to [27], they are a "subset of algorithmic traders that most rapidly turn over their stock positions". They are also characterised by using co-location and proximity services to minimise latency, submitting a large number of orders that are cancelled shortly after submission and by very short holding periods, see [25]. Moreover, they are described as market participants who trade a large number of contracts, consistently maintain a low inventory level, and end the day at or near a zero inventory position, see [11]. This significant growth in high-frequency trading in the recent years has created a debate among regulators, academics and practitioners. They aim at understanding how HFTs affect the markets. For

instance, does HFTs activity improve or damage liquidity? Do they consume liquidity more than they provide? Do they supply more or less liquidity during periods of stress or macroeconomic announcements?

The question of the impact of HFTs on liquidity is an important topic of research in the literature. Before explaining the different positions on this issue, we note that liquidity is a complex and ambiguous notion. However, it is often loosely defined as the ability to trade a large amount of a financial instrument in a short period of time at a price close to the current price. There are several methods to quantify liquidity. Typically, it is evaluated by measuring the bid-ask spread and the market depth which can for example be defined as the cumulative amount placed up to the three best limits on the bid and ask sides, see for example [35]. In our study, liquidity is assessed according to the same two criteria.

Some argue that HFTs are beneficial for the market. For example, [29] studies trading of 26 NASDAQ-labelled HFTs firms on the NASDAQ market in 2008-2010 and show that HFTs tend to improve market quality. It is found that HFTs do not withdraw from markets in bad times, they do not engage in abnormal front-running of large non-HFT trades, contribute more to price discovery than non-HFTs, and reduce volatility. Moreover, the standard view of [60] on market making suggests that competition between HFTs should decrease their profits and lead to lower trading costs for other market participants.

On the contrary, others assert that high-frequency trading has toxic consequences on market quality. In its concept release on equity market structure, the Securities and Exchange Commission, see [104], underlines that differences in speed between market participants, due to the race between HFTs, may damage liquidity. [56] points out that if liquidity providers are even marginally slower than the fastest traders, they are in the position of being adversely selected. Recognising this risk, classical liquidity providers will quote wider spreads. Several theoretical models are in line with this thinking, suggesting that when speed differentials between traders exist, adverse selection may increase and liquidity become more expensive, see [50, 51]. Another closely related negative aspect is discussed in [11]. They argue that regardless of how fast the market as a whole becomes, there is always at least one firm with a relative speed advantage that can adversely select other traders, and they find that high-frequency trading industry is concentrated among a few firms. This idea is supported for example by [19, 30] who show in theory how competition based on latency could lead to market concentration, in particular to a winner-takes-all environment that could cause an inefficient over-investment in speed. In addition, [70] investigate market and limit orders in the U.S. Treasury market around major macroeconomic news announcements, and show that HFTs have a negative effect on liquidity around economic announcements: they widen spreads during the pre-announcement period and lower depth on the order book during the post-announcement period. Moreover, [103] analyse data on the highly liquid S&P 500 ETF traded on NASDAQ from January 6, 2009 to December 12, 2011 and find that in the minute following a macroeconomic news arrival, algorithmic activity increases trading volume and depth at the best quotes, but also increases volatility and leads to a drop in overall depth.

In our paper, based on French data, we contribute to this debate by exploring the impact of HFTs on liquidity and trading activity. It is one of the few studies of HFTs in Europe: it analyses 36 stocks of the CAC 40 index which are traded on Euronext Paris from November 2015 to July 2016 (Note that Euronext is ranked 5^{th} in the world in domestic market capitalization). In addition to this, while most of the works on HFTs are based on trades, each counterpart being flagged depending on whether the trader is HFT or not, our approach also focuses on quotes. We are able to rebuild the limit order book and flag whether orders standing at each level of the book have been submitted by a HFT or a non-HFT. This enables us to compute for example the percentage of HFTs orders at the best quotes and the fraction of the depth submitted by HFTs. We are notably interested in periods of market stress (this notion will be defined with accuracy later), and in their intraday seasonal behaviour: at the beginning of the day, during the day, and at the end of the day. We in particular show how the presence of HFTs has impacted the intraday patterns of liquidity. Moreover, we provide a specific focus on their dynamics before and after macroeconomic announcements, and when historical volatility suddenly increases, for example after the opening of the U.S. market.

A second motivation for this study is the close relationship between HFTs and market making established in the literature and often claimed by HFTs. This is to the extent that HFTs are described as the new market makers in [88]. This latter shows how the success of a new market, Chi-X, critically relied on the participation of a large HFT who acts as a modern market maker. Moreover, [71] notes that high-frequency market makers have largely replaced traditional market makers because the use of speed (co-location) and technology results in a lower adverse selection: they are less susceptible to be seized by market participants having an informational advantage. In addition to this, [11] find that the fastest firms (with the lowest relative latency) are better in risk management and earn a higher realised bid-ask spread when trading passively.

Our paper serves to feed this debate on whether or not HFTs do globally play the role of market makers. This is done by comparing their behaviour to what is typically expected from a market maker.

Some academic studies are also interested in the behaviour of HFTs during specific events. For instance, [74] study the behaviour of high-frequency traders on the E-mini on the 6^{th} of May 2010, the day of the most well-known flash crash. [37] analyses that of the 17^{th} of March 2011 and [113] investigates the U.S. treasury flash crash on the 15^{th} of October 2014. In this paper, we also focus on analysing specific events. Instead of analysing the reaction of HFTs due to flash crashes, we investigate their reaction to surprising news characterised by sudden large historical volatility rise, in particular the 3^{rd} of December 2015 (European Central Bank announcements about the new monetary policy) and the 24^{th} of June 2016 (Brexit day), both days marked by very significant price moves.

Our study is conducted on data provided by the French regulator “Autorité des Marchés

Financiers"(AMF). It covers the trades and orders of the most liquid French securities, specifically the 36 stocks of the CAC 40 index which are traded on Euronext Paris from November 2015 to July 2016.

As we know, a universal definition of HFTs does not exist, and there are several methods to identify HFT members. In our study, an actor is identified as an HFT if either he is a member of the Supplemental Liquidity Provider programme, which is a market making programme specific to Euronext, either he is classified as a pure HFT by a method based on the lifetime of cancelled orders. Details on our classification will be given in Section 2.2.

One of the main findings of our paper is that when the market is under stress, which corresponds to periods of high uncertainty about the future (this will be measured through an implied volatility based metric), market depth offered by all participants tends to decrease, and the bid-ask spreads tend to increase. Such finding is in line with the results in [28] (conducted on 2008-2009 NASDAQ data). However, the rise in global stress affects quite equally HFTs and non-HFTs. We will notably see that the market share of HFTs in terms of market depth and the aggressiveness of HFTs are essentially independent of the level of market stress. However, the HFTs market share in amounts traded depends slightly on it.

Regarding the seasonal behaviour of HFTs during a trading day, our conclusions are the following: at the beginning of the day, HFTs progressively increase their contribution in the order book, gradually tighten spreads and raise their market share in terms of market depth (typically spread is gradually halved while amounts offered by HFTs increase fivefold). Simultaneously, the HFTs aggressive/passive ratio declines as HFTs consume less liquidity and provide more (the aggressive/passive ratio decreases from 65% to 55%). At the end of the day, HFTs essentially withdraw their orders from the order book. The aggressive/passive ratio increases as HFTs consume more liquidity and provide less.

Besides, pre-announcement periods are marked by a strong downturn in market quality: sharp reduction in market depth offered (-35% on average at 2:30 p.m. Paris time, which corresponds to the announcements time of the European Central Bank and U.S figures) and high increase in bid-ask spreads (+32% on average at 2:30 p.m.). During these periods, HFTs reduce their liquidity provision in the order book more significantly than non-HFTs and thus their passive share in amounts traded decreases.

Post-announcement periods are characterised by a sharp increase in overall activity (result in line with [33]), which is more pronounced for HFTs, particularly in terms of share in amounts traded (share in amounts traded rises from 58% to 65%). The aggressive/passive ratio also increases after the announcements. However, contrary to what happens before announcements, this is because HFTs use more liquidity consuming orders, and not because they reduce their presence in the limit order book.

We use linear regressions to analyse scheduled announcements in a quantitative way. We

show that changes in HFTs behaviour around such events are more than a typical reaction to an increase of volatility. A "change of regime" seems to take place: knowing that an announcement is scheduled, on top of usual reactions to volatility fluctuations, HFTs provide 15% less liquidity, are more aggressive, and trade less than usual.

Another important finding of our work is that during periods without much market stress, HFTs consume slightly more liquidity than they provide (aggressive/passive ratio equal to 52%). This stands out more during periods of high historical volatility: at the beginning of the day (between 55% and 65%), around announcement phases (57% during the 2:30 p.m. announcements), and after the U.S. market opening (55%).

One of the main interests of this paper is that it uses recent data to conduct an analysis of the effects of high-frequency trading on market quality. Indeed, there are not many such empirical studies due to the difficulty of obtaining data on HFTs orders, see [18]. All the findings of this study help to understand the impact of HFTs activity on the market, and provide potential insights for new regulation. As a matter of fact, the financial industry has probably not found the adequate calibrations of all the tools offered by high-frequency trading, see [3]. However, note that one limitation of this paper is that we only consider data from the regulated market Euronext. Hence our work should be complemented by a study of the behaviour of HFTs over the same time period on multilateral trading facilities.

The paper is organised as follows. In Section 2, we discuss the data we consider, the methodology used to identify HFTs and the volatility metrics. We also present the main macroeconomic announcement times and highlight the different periods of market stress. We present in Section 3 (resp. Section 4) the liquidity metrics (resp. trading metrics) and the analysis of liquidity provision (resp. trading activity) on a day-to-day and an intraday scale. In Section 5, we provide a more quantitative analysis of HFTs behaviour around the news of 4 p.m. In the last section, we focus on days with particularly high historical volatility: the 3^{rd} of December 2015 in Section 6.1 and the 24^{th} of June 2016 in Section 6.2. We conclude in Section 7.

2 Data description, HFTs identification and volatility metrics

2.1 Description

We recall that the data under study are provided by the French regulator "Autorité des Marchés Financiers" (AMF). They cover the trades and orders on the most liquid French securities, specifically 36[1] of the CAC 40 stocks which represent more than 85% of total amounts traded in all French stocks traded on Euronext Paris, from November 2015 to July

[1]Unibail, Arcelor Mittal, STMicroelectronics and Solvay are listed on Euronext Amsterdam or Euronext Brussels. The AMF does not have order book data for these exchanges. This is why we only analyse 36 and not 40 stocks. Note that Alcatel-Lucent is included for only part of the analysis; it is replaced by Nokia from December 22, 2015.

2016.

Note that according to [48], Euronext accounts for 63% of the traded amounts in the continuous auction phase during this period, while Chi-X accounts for 18%, Turquoise for 12% and BATS for 4% (the remaining market share going to smaller multilateral trading facilities). In our database, all orders (and transactions) are labelled by the name of the owner, which allows us to identify HFTs.

Moreover, we use in some sections a database of scheduled announcements coming from Bloomberg.

Since the amounts traded by HFTs during the opening and closing auctions represent only about 6% of their amount traded during the day, and about 11.5% of the total volume traded at auction[2], the analysis is confined to continuous trading phases (during which HFTs account for more than 60% of the traded amounts).

2.2 HFT identification

A market participant is identified as a high-frequency trader (HFT) if either he is classified as a pure HFT by a method based on the lifetime of cancelled orders, or he is a member of the Supplemental Liquidity Provider (SLP) programme of Euronext. After identifying HFTs, we rebuild the limit order book in order to flag the origin of the depths (HFT or non-HFT) at the best quotes. We describe below first the identification method based on orders, then the SLP programme.

2.2.1 The classification of HFTs based on the lifetime of cancelled orders

The method we now present is one of the criteria used by the AMF to identify HFTs. This classification differentiates three types of market participants in the order book: pure HFTs, mixed HFTs (investment banks with high-frequency trading activity) and non-HFTs. It is based on the lifetime of cancelled orders and determined using two sets of conditions:

- **Condition 1 is based on a comparison with other participants**: the participant must have cancelled at least 100,000 orders during the year, and the average lifetime of his cancelled orders should be less than the average lifetime of all cancelled orders in the book.
- **Condition 2 is based on a set threshold**: the participant must have cancelled at least 500,000 orders with a lifetime of less than 0.1 second (i.e. the participant quickly updates the orders in the limit order book) and the top percentile of the lifetime of its cancelled orders must be less than 500 microseconds (i.e. the participant regularly uses fast access to the market).

[2] See [15, 24] for details on fixing auctions.

A participant is a pure high-frequency trader if he is not an investment bank and he meets one of these conditions. An investment bank meeting one of these conditions is described as mixed. Note that some members satisfy Condition 2 without satisfying Condition 1.

2.2.2 The Supplemental Liquidity Provider programme

Market makers are intermediaries that minimise inventory risk by holding positions for very short periods, see [84]. The market making programme of Euronext Paris, named Supplemental Liquidity Provider (SLP) programme, imposes a market making activity on programme members, including order book presence time at competitive prices. In return, they get favourable pricing and rebates in the form of a maker-taker fees model directly comparable to those of the major competing platforms, see [44].

Each member of the SLP programme is identified as a HFT in this study. Indeed, HFTs are now the only market participants that are able to play the role of market makers on liquid stocks, see [71]. In fact, they are supposedly able to maintain a strong presence at best price limits and operate efficient inventory management in an increasingly fast-moving and fragmented market. Indeed, HFTs can use speed to enhance risk management by avoiding adverse selection, see [72], improving inventory management, see [5] and trading on short-lived information, see [50].

Moreover, according to the classification based on the lifetime of cancelled orders, all SLP members are either pure HFTs or mixed HFTs. So, in this work, we consider HFTs orders as those coming from market participants that are either members of the SLP programme or classified as pure HFTs.

We note that there are 20 members identified as HFTs in this study. SLP activity represents a large part of all estimated high-frequency activity: 65% in number, and 90% in traded amounts. More descriptive statistics are provided in Section 4.2.

2.3 Volatility metrics

2.3.1 Volatility metrics description

We measure market volatility in two different ways: overall market volatility and temporary volatility shocks (due to large price moves that often follow macroeconomic announcements). To do so, we use two volatility metrics, computed every 15 seconds.

- **Our measure for overall market volatility: implied volatility.** Since the study is conducted on the CAC 40 stocks, the implied volatlity we choose is the VCAC index (the CAC 40 volatility index[3]). Its computation is based on option prices and inspired by the VIX methodology, see [45]. It is generally interpreted as a market consensus estimate of the market risk. In this work, we consider that the higher the VCAC, the

[3]Note that the CAC 40 and VCAC indexes are published every 15 seconds.

greater the level of market stress. Note that we mainly use the VCAC volatility measure for the day-to-day scale analysis.

- **Our measure for temporary volatility shocks: historical volatility.** The historical volatility at time t on day d, denoted by $\sigma_{hist}(t,d)$, is taken as the average of the squared increments of the CAC 40 index, sampled every 15 seconds between t and $t-15$ minutes (i.e. 60 measurements)[4]:

$$\sigma_{hist}(t,d) = \frac{1}{60}\sum_{i=0}^{59}\left(P_d(t-i\cdot 15) - P_d(t-(i+1)\cdot 15)\right)^2$$

  where $P_d(t)$ is the CAC 40 level on day d at time t (in seconds).

  This measure covers a short period of time and is somehow representative of instantaneous volatility at time t. It has little sensitivity to the selected sampling frequency. This is because from 15 seconds onward, it is typically not dramatically affected by microstructure effects.

From now on, each day is split into one minute time intervals (bins). One value for each volatility indicator is computed in each bin, this value being its average over the bin. For the day-to-day analysis, we restrict ourselves to the daily average of each metric. For the intraday analysis, for each bin, we average each metric over the days.

2.3.2 Contrasts between the two volatility metrics

Implied volatility is forward-looking while historical volatility is backward-looking. In other words, historical volatility is an ex-post measure of uncertainty while implied volatility is an anticipation of future risk. Another difference between these two notions is that implied volatility is a risk-neutral measure, which is the reason why it incorporates the volatility risk premium (it refers to the phenomenon that options implied volatility tends to exceed realised volatility of the underlying asset). This is related to investors views on market risk.

These two metrics meet different needs: historical volatility captures temporary price shocks, while implied volatility assesses the general level of market stress. On the one hand, these two metrics are clearly related. Indeed, on a day-to-day scale, they both exhibit relatively similar patterns: periods of increasing implied volatility are generally characterised by a significant level of historical volatility, see Figure I.1.

[4] Up to a factor equal to 10^8.

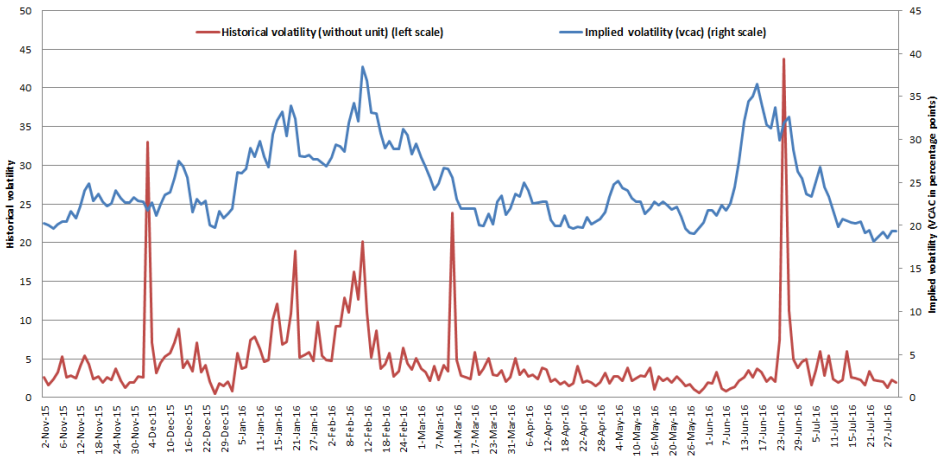


Figure I.1 – Evolution of implied and historical volatilities over the period: the two metrics varied widely. Implied volatility illustrates the overall level of stress while historical volatility highlights the days when there were sharp price changes (notably the 3^{rd} of December 2015 and the 24^{th} of June 2016).

On the other hand, on an intraday scale, scheduled announcements for example impact historical and implied volatilities differently: as implied volatility captures the expected risk of an announcement, it generally tends to decrease when new information reaches the market. In contrast, a price move due to an announcement causes historical volatility to rise.

We give an example of these contrasting variations by displaying in Figure I.2 the evolution of implied and historical volatilities on the 3^{rd} of December 2015.

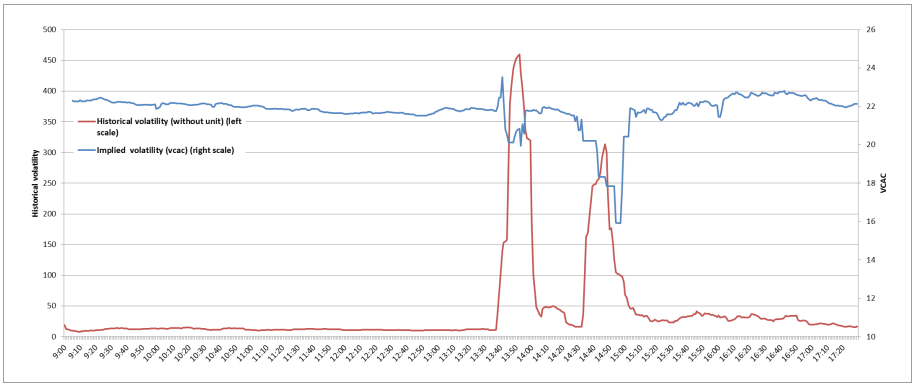


Figure I.2 – Intraday differences between the variations of implied and historical volatilities on the 3^{rd} of December 2015, a day with various announcements.

We can see that after the announcements concerning the new monetary policy at 1:45 p.m. and 2:30 p.m., which have a strong impact on prices, historical volatility increases. On the contrary, implied volatility decreases. This is notably because some put and call prices are unavailable during sharp price variations, which leads to a mechanical decrease of the VCAC followed by an increase, and this is due to the computation method of this index.

2.3.3 A first view on the analysed period

The period under study spans the nine months from November 2015 to July 2016. Figure I.1 shows that French equity market implied volatility swung widely during this period: the year had already started off on a volatile note amid concerns about China's economy and the risks associated with the sharp fall in oil prices, see [111]. Its rise during February 2016 was followed by a more stable period after March 2016. Then implied volatility increased sharply in June 2016 due to the UK referendum.

In addition to the alternations of the level of stress during the period, it is important to bear key figures publication times (Paris time) in mind, in particular macroeconomic announcements, which are often accompanied by local stress of the market:

- **9 a.m.**: Paris market opening and start of continuous trading following the opening auction.
- **11 a.m.**: European figures. For example the industrial confidence indicator or the consumer confidence indicator.
- **11:45 a.m. to 12:00 a.m.**: CAC 40 and Eurostoxx 50 futures price auctions on the third Friday of the month (expiration).
- **1 p.m.**: German consumer confidence figures and UK monetary policy decisions.
- **1:45 p.m.**: European Central Bank (ECB) announcements.
- **2:30 p.m.**: ECB press conferences and U.S. Figures.
- **3:30 p.m.**: Opening of the U.S. market.
- **4 p.m.**: U.S. figures such as leading economic indicators.
- **5:30 p.m.**: End of continuous trading before the closing auction.

It is also worth noticing that very significant sudden price changes occurred throughout the period, notably on the 3^{rd} of December 2015 (ECB announcement) and on the 24^{th} of June 2016 before the market opening (Brexit announcement). Both days are examined in details in Sections 6.1 and 6.2.

Now that volatility metrics are well defined, our goal is to study whether the behaviour of HFTs is affected by the variations of these volatility measures. This is done through day-to-day and intraday analyses.

3 Liquidity provision by HFTs

In this section, we first explain how liquidity is measured, and give some preliminary statistics about liquidity in our dataset. Then we study the contribution of HFTs to liquidity by analysing their behaviour on a day-to-day and an intraday scale. The more the market is stressed, the riskier it is to hold stocks for market makers. So one can expect shallower market depth when volatility rises. However, since certain HFTs (members of the SLP programme) benefit from market makers agreements which can include rebates, it may be feasible for these market members alone to maintain liquidity provision during periods of stress. Thus HFTs behaviour may be less sensitive to the fluctuations of the risk level in the market.

3.1 Liquidity metrics

A lot of definitions of liquidity are possible. However, they all agree on the fact that it is related to the degree to which a financial instrument can be quickly bought or sold in the market without affecting its price. There are different indicators to measure liquidity. In our study, we choose to measure it according to the same criteria as those used in [35]: market depth and bid-ask spread. It is important to not only consider the market depth but also the bid-ask spread since it reflects the cost of an aggressive trade when taking the mid price[5] as reference price, see for example [86, 115]. An upturn (resp. a downturn) in liquidity can be characterised by either an increase (resp. decrease) of market depth at best price limits while spreads stay constant (or tighten, resp. widen) or is marked by spreads tightening (resp. widening) while market depth stays constant (or increases, resp. decreases).

- The **market depth** at time t at the n^{th} best price limit in the order book, denoted by $MD(n,t)$, is the cumulative quantity placed up to the n^{th} best price limit, averaged on the ask and bid sides. This metric used to quantify the actual passive presence of market participants in the order book is expressed in euros and defined as follows:

$$MD(n,t) = \frac{\sum_{i=1}^{n}(V_i^{Ask}(t) + V_i^{Bid}(t))}{2},$$

  where $V_i^{Ask}(t)$ (resp. $V_i^{Bid}(t)$) is the "value" at the i^{th} non-empty limit on the ask (resp. bid) side: the volume multiplied by the price at the i^{th} non-empty limit on the ask (resp. bid) side at time t.

  We denote respectively by $MD^{HFT}(n,t)$ and $MD^{nHFT}(n,t)$ the market depth taking only HFT orders and only non-HFT orders into account.

- The **bid-ask spread in ticks** at time t is equal to the difference between the best ask price (the lowest displayed price at which an investor would sell shares) and the best bid price (the highest displayed price at which an investor would buy shares), divided by the tick size at time t.

[5] The mid price is defined as the average between the best bid and best ask prices.

In the sequel, we compute the spread and the market depth for each security after any event occurring in the limit order book (transaction, insertion of a new order, cancellation or modification of an existing order). For the intraday analysis, bid-ask spreads in ticks and market depths are averaged across all securities, over all the days, and over the same minute bins. For the day-to-day analysis, spreads and market depths are averaged across all securities, and over all the order book events occurring during the day. In this case, the market depth at the n^{th} best price limit on day d is denoted by $MD(n,d)$. The adjusted market depth on day d at the n^{th} best price limit in the order book, denoted by $AMD(n,d)$, is equal to the market depth $MD(n,d)$ multiplied by the variation of the CAC 40 up to day d. More precisely, it is defined by

$$AMD(n,d) = MD(n,d) \cdot \frac{P(1)}{P(d)},$$

where $P(1)$ is the CAC 40 value averaged over the first day of the analysed period: the 2^{nd} of November 2015, and $P(d)$ is the CAC 40 value averaged over day d. Thus it is not affected by the fall in prices in the first part of the period under review.

We denote respectively by $AMD^{HFT}(n,d)$ and $AMD^{nHFT}(n,d)$ the HFTs adjusted market depth and the non-HFTs adjusted market depth at the n^{th} best price limit on day d.

The adjusted market depth will be used for the day-to-day analysis.

3.2 Preliminary statistics

For the whole period and for all the securities taken into account in the analysis, HFTs are present at the best ask or best bid 91% of the time (relatively to all the events occurring during the analysed period). The average HFTs market shares in terms of depth, according to the number of best limits over the period, are summarised in Table I.1.

Presence in the limit order book	Market share in terms of market depth	Amount offered by HFTs
At the best bid and offer prices	70.8 %	EUR 40,826
At the two best price limits	77.3 %	EUR 122,084
At the three best price limits	79.3 %	EUR 224,774

Table I.1 – Average HFTs market shares and average amounts (over all stocks) offered at the best price limit or accumulated for the two or three best bid/ask limits. HFTs show a slightly smaller presence in terms of market share at best limits. The amounts offered (bid/ask average) range from EUR 41,000 to EUR 225,000.

We note that in Table I.1, and for the rest of the study, we exclude amounts provided by Retail Liquidity Providers (RLP) because the liquidity they provide is not visible on the central limit

order book, but is only accessible to retail clients that are members of the Retail Member Organization (RMO), see [43].

Depending on the level of market depth studied, the amounts offered by HFTs range from EUR 41,000 (aggregated for the best bid/ask limits) to EUR 225,000 (aggregated for the three best bid/ask limits). Thus it by far exceeds the average trade size, which ranges from EUR 8,000 to EUR 12,000 for these stocks. Orders therefore rarely breach the best limit: only 2.5% of aggressive orders hit at least two consecutive price limits, and only 0.4% hit three consecutive ones.

We only use the market depth measure built on the three best price limits $MD(3, t)$ throughout the analysis, as it is more stable than the best price limit metric and both sketch similar trends.

We finally note that 50% of liquidity provided by HFTs is essentially contributed by 4 HFT members.

3.3 Day-to-day analysis

Recall that we quantify market stress by the level of implied volatility, see Section 2.3. In Figure I.3, we plot implied volatility and HFTs adjusted market depth AMD^{HFT} over the period, on a day-to-day scale.

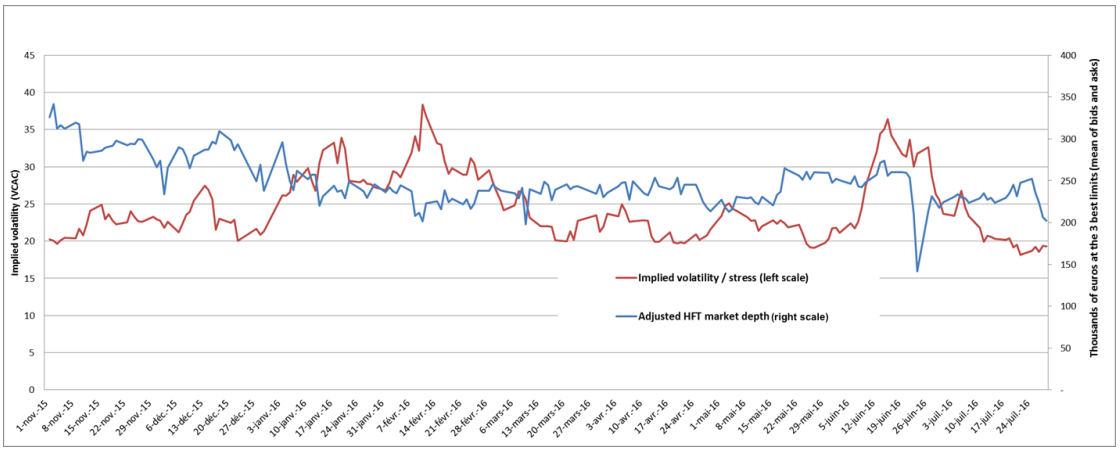


Figure I.3 – Evolution of HFTs adjusted market depth AMD^{HFT} and implied volatility VCAC over the period. Except for the two-week period prior to the Brexit announcement, implied volatility and depth diverged: the higher the volatility, the smaller the quantity offered.

In Figure I.4, we display the evolution of the bid-ask spread in ticks and the adjusted market depth of HFTs and non-HFTs over the period, on a day-to-day scale.

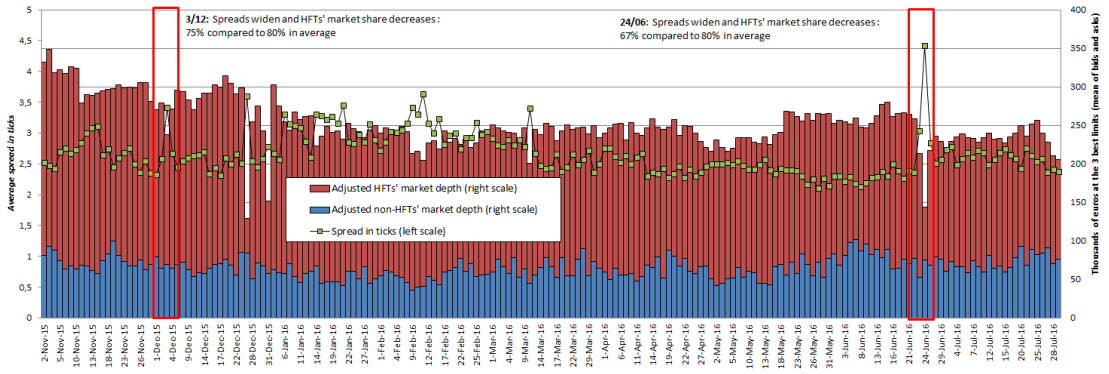


Figure I.4 – Evolution of the bid-ask spread in ticks and adjusted market depth of HFTs AMD^{HFT} and non-HFTs AMD^{nHFT} over the period. In the first part of the period (until February 2016), along with the increase of implied volatility, HFTs and non-HFTs adjusted market depth decreased and bid-ask spreads widened, reflecting a downturn in liquidity.

Figure I.4 shows a negative correlation between HFTs market depth and spread, clearly marked for example on December 3 and June 24. In addition, from November 2015 to February 2016, with the increase of implied volatility, see Figure I.3, market depth decreased sharply (it diminished by more than 30% from its value at the beginning of the period) and spreads widened (average spread of 2.66 ticks in November 2015 versus 3.05 ticks in February 2016). For the rest of the period under review, with the exception of the heightened uncertainty surrounding the Brexit announcement[6], implied volatility and spreads returned to levels close to those observed at the beginning of the period. The VCAC stood at 20, see Figure I.3, spreads were tighter (2.5 ticks on average) and order book depth increased slightly, but did not return to the levels seen in November 2015.

For the entire period under review, excluding the two most volatile days (the 3^{rd} of December 2015 and the 24^{th} of June 2016), Figure I.4 shows that HFTs market share in terms of market depth remains almost constant[7]: close to 80% at the three best price limits.

Figure I.3 and Figure I.4 essentially establish that market depth and implied volatility follow opposite trends over the period: the higher the implied volatility (and thus the higher the risk), the lower the liquidity provision (shallower market depth and wider spreads).

Finally, we observe that implied volatility affects all market participants evenly. So, contrary to what could be expected, HFTs are not less sensitive than other participants to the stress

[6] A specific analysis of the behaviour of HFTs around the Brexit referendum result announcement is provided in Section 6.2.

[7] Although HFTs market share in terms of market depth was relatively constant for the entire period, the days 3^{rd} of December 2015 and 24^{th} of June 2016 showed a sharp decline in market depth provided by HFTs, see Sections 6.1 and 6.2.

in the market, although most of them benefit from rebates, that could be read as incentive to provide liquidity during periods of stress. These findings raise questions about the relevance of fee rebates if they are meant to compensate for adverse selection under stress conditions.

3.4 Intraday analysis

We now consider intraday analysis. We focus in particular on the behaviour of HFTs around scheduled announcements. On the one hand, and by definition, non-directional market making strategies try to avoid sharp price moves. Thus, a withdrawal of HFTs orders before a scheduled announcement seems a rational practice. On the other hand, some HFTs obtain rebates to maintain liquidity. So it is possible that they do not need to reduce their presence in the limit order book to preserve their profit.

Most of the scheduled announcements were enumerated in Section 2.3.3, and can be identified on the next graphs by a fall of the VCAC.

In Figure I.5, we display the bid-ask spread in ticks and the market share of HFTs and non-HFTs in terms of market depth averaged over the whole period.

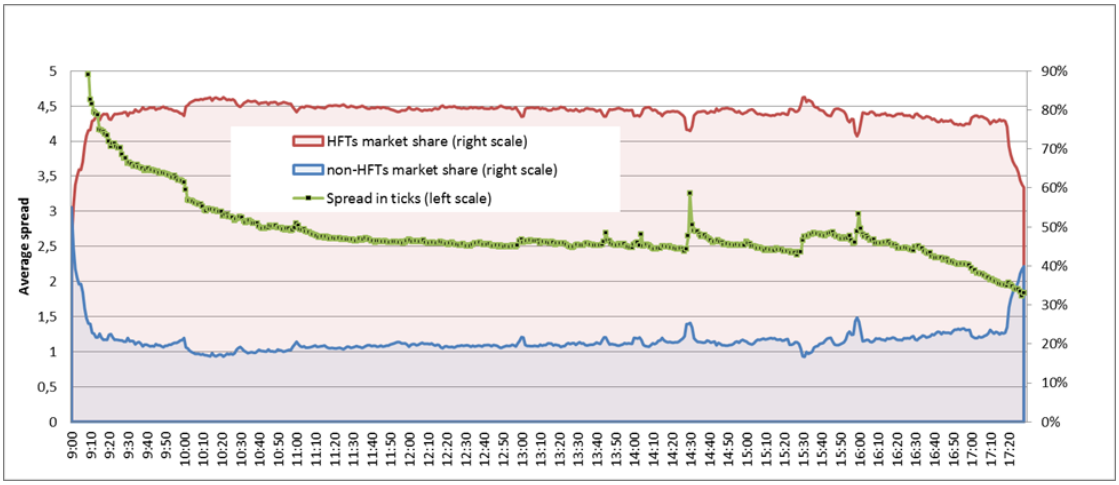


Figure I.5 – Intraday seasonal evolution of market share in terms of depth compared to the bid-ask spread in ticks. HFTs contribute to spread tightening at the beginning of the day (their share increases while spread tightens). They withdraw their orders at the end of the day.

In Figure I.6, we plot the VCAC and the market depth of HFTs MD^{HFT} and of non-HFTs MD^{nHFT}.

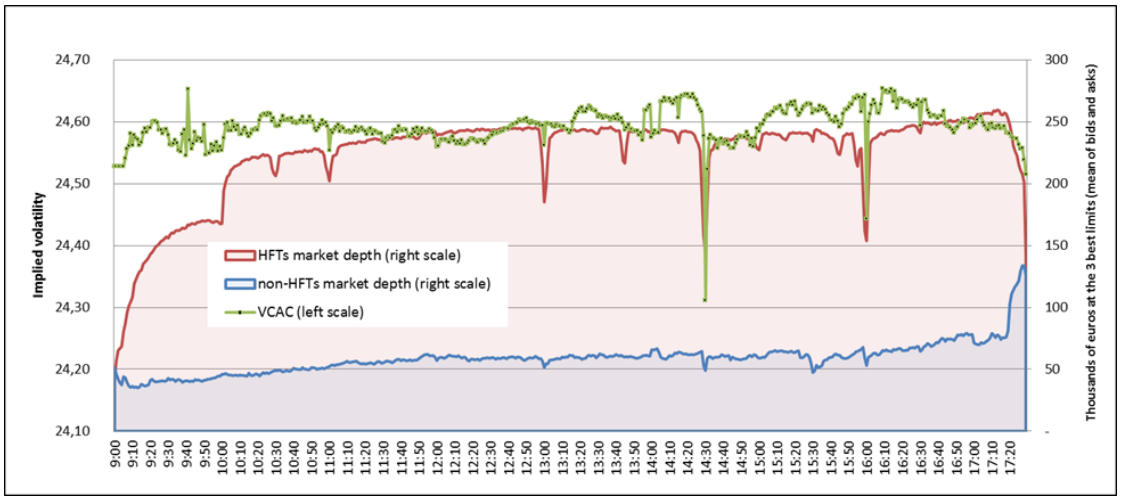


Figure I.6 – Intraday seasonal evolution of HFTs market depth MD^{HFT} and non-HFTs market depth MD^{nHFT}. Market depth offered by HFTs increases at the beginning of the day and decreases at the very end of the day. HFTs withdraw their orders promptly before the announcements, which is rarely the case for non-HFTs. For example at 2:30 p.m., HFTs presence falls by 40% compared to its value before 2:30 p.m.

3.4.1 Intraday behaviour

Figure I.5 and Figure I.6 show that the average intraday evolution of market depth exhibits wide disparities between HFTs and non-HFTs:

- Non-HFTs provide a relatively constant amount (slightly increasing) in the order book throughout the day — except for the last 15 minutes of the day when the amount offered by non-HFTs increases quickly. This behaviour at the end of the day is driven by the strategy adopted by one particular market participant[8], see Figure I.6.
- HFTs behave quite differently: both their share in market depth, see Figure I.5, and the amount they offer, see Figure I.6, increase during the first hour of the day with a sharp rise at 10 a.m. (linked to the arrival of one particular participant). The levels remain stable during the day, albeit with temporary declines before the usual announcement times (identified by a fall in implied volatility in Figure I.6) and sometimes accompanied by spikes in historical volatility, see Figure I.10. At the end of the day, HFTs share in terms of market depth and amounts offered decreases rapidly. Indeed, HFTs do not generally hold overnight positions.

[8]For confidentiality reasons, it is not possible to specify who this one market participant is.

3.4.2 Focus on the behaviour before and after scheduled announcements

Figure I.5 and Figure I.6 clearly demonstrate that the market depth offered by HFTs thins out ahead of scheduled announcements that are likely to cause price shocks. Most of them sharply reduce their liquidity provision a few minutes (generally 3 minutes) before the announcements, unlike other market participants who decrease slightly their amount of limit orders. This is particularly the case for announcements occurring at 1:45 p.m., 2:30 p.m. and 4 p.m. This withdrawal before announcement times contributes to bid-ask spread widening. For example, at 2:30 p.m. spread gains 32% while depth decreases by 35%. Then HFTs liquidity returns quickly to the order book after the announcements (generally one or two minutes later).

Thus, despite the rebates that some SLP members receive to provide liquidity, HFTs sharply reduce their presence before scheduled announcements. Attempts to require market makers to be present during announcements periods (which represent about 1% of the total time markets are open everyday) to smooth price moves have so far been unsuccessful. The presence requirements in the new European directive MiFID II for HFT market makers leave significant scope for withdrawal from the order book: at least 50% of the trading session excluding the opening and closing auctions, see [36]. This means HFTs can remove their orders from the book 50% of the time, a percentage far larger than 1%.

3.4.3 Focus on the 10 most impactful announcements at 2:30 p.m.

We now analyse the changes in liquidity around the 10 announcements at 2:30 p.m. having the greatest impact on prices[9]

In Figure I.7, we display the intraday evolution of the bid-ask spread in ticks together with the HFTs and non-HFTs market shares and depths around 2:30 p.m. This is done over the whole period on the one hand, and on the other hand just considering the 10 most impactful announcements at 2:30 p.m.

[9]The following days are those with highest historical volatility at 2:30 p.m.: December 3, December 4, January 8, January 21, February 5, February 12, March 4, March 10, May 6 and July 8.

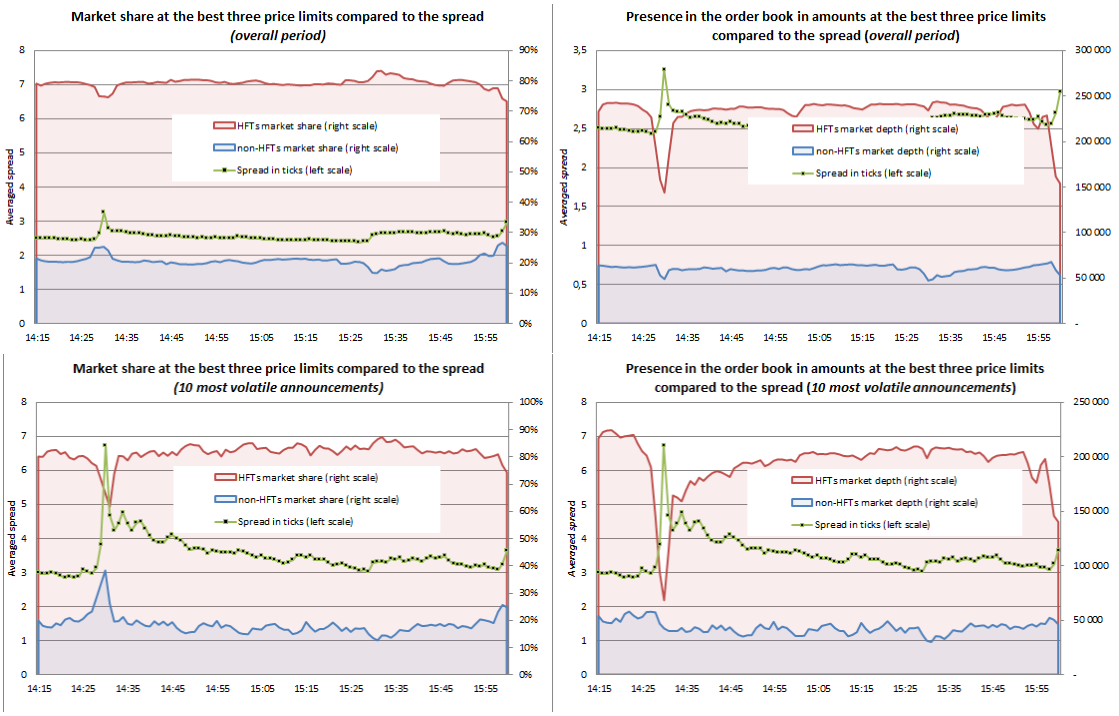


Figure I.7 – Bid-ask spread in ticks together with HFTs and non-HFTs market shares and depths: Comparison between the whole period and the 10 most impactful announcements occurring at 2:30 p.m.

Figure I.7 shows a wider range of spread and market depth fluctuations for the 10 most impactful announcements than on the overall period. Still, the patterns are very similar to those described above. For instance, for the 10 most impactful announcements, HFTs withdraw 70% of their orders from 2:28 p.m. and return to a nearly 80% market share in terms of market depth from 2:35 p.m. Hence, when large volatility spikes are expected, HFTs withdrawal is quite significant compared to other market participants. This can be understood as a fear from adverse selection from market makers.

4 Trading activity of HFTs: Amounts traded and aggressiveness

The previous section focused on limit orders inserted by market participants and compared the behaviour of HFTs to that of non-HFTs. In this section, we are interested in the executed limit orders and their counterparts. In other words, we analyse transactions to evaluate the trading practices of HFTs.

An order is referred to as aggressive when it initiates a trade. A passive order is the counterpart to any aggressive order. We recall that a transaction is derived from the matching of an aggressive order with a passive order. An executed passive order provides liquidity, while

an aggressive order consumes liquidity. In this section, we investigate the HFTs market share in amounts traded and the role they play in transactions: are they mainly liquidity providers or liquidity consumers, in other terms, are their orders mainly aggressive or passive? First, we describe the metrics used to assess the trading behaviour of HFTs. Then we discuss the expected behaviour of a traditional market maker. Thereafter, we analyse the dynamics of HFTs from a day-to-day and an intraday scale. We compare it to that of a traditional market maker, with the objective of understanding whether or not HFTs do really play the role of market markers.

4.1 Metrics used

In order to evaluate the trading behaviour of HFTs, we measure their share in amounts traded and their aggressiveness. We define below these two metrics.

- **The HFTs share in amounts traded** during a time interval $[t, t+h)$, denoted by $MS_{HFT}(t, t+h)$, is the ratio of amounts traded by HFTs (aggressively and passively) across all the stocks during $[t, t+h)$ over all the amounts traded in the market during $[t, t+h)$. It is defined by

$$MS_{HFT}(t, t+h) = \frac{M_{HFT}(t, t+h)}{M_{HFT}(t, t+h) + M_{nHFT}(t, t+h)},$$

  where $M_{HFT}(t, t+h)$ are the amounts traded (volume multiplied by price) by HFTs during $(t, t+h)$ and $M_{nHFT}(t, t+h)$ the amounts traded by non-HFTs during $(t, t+h)$.

- **The HFTs aggressive/passive ratio** during a time interval $[t, t+h)$, denoted by $R_{A/P}(t, t+h)$, is the ratio of amounts traded aggressively by HFTs across all the stocks during $[t, t+h)$ over the total amounts traded by HFTs (aggressively and passively) during the same time period. It is defined by

$$R_{A/P}(t, t+h) = \frac{M_{ag}(t, t+h)}{M_{ag}(t, t+h) + M_{pa}(t, t+h)},$$

  where $M_{ag}(t, t+h)$ are the amounts executed aggressively by HFTs during $(t, t+h)$ and $M_{pa}(t, t+h)$ the amounts executed passively by HFTs during $(t, t+h)$.

  The HFTs aggressive/passive ratio represents the ratio of liquidity consumed by HFTs over the total liquidity consumed and provided by HFTs. A ratio of x% means that x% of the volume of the transactions they are involved in is liquidity consumption.

In the sequel, market share in amounts traded and aggressive/passive ratio are computed every minute of each trading day, so we compute $MS_{HFT}(t, t+60)$ and $R_{A/P}(t, t+60)$. For the intraday scale, these metrics are averaged across all days. For the day-to-day scale, these metrics are averaged across all the minutes for each day.

4.2 Preliminary statistics

Aggressive orders	Passive orders	Percentages (in volume)
HFTs	HFTs	33.6%
non-HFTs	HFTs	22.4%
HFTs	non-HFTs	31.2%
non-HFTs	non-HFTs	12.8%

Table I.2 – The different interactions between HFTs and non-HFTs in the market.

Table I.2 describes the different interactions between HFTs and non-HFTs in the market. Four scenarios are possible. It shows that 60% of liquidity provided by HFTs is consumed by other HFTs, and 40% by non-HFTs. Liquidity provided by non-HFTs is mainly consumed by HFTs (71%). This could indicate information asymmetry among participants since aggressive trades are generally more informed than passive trades, see [20, 52, 99, 105]. In the classical theory of market making (like [76]), market makers are not supposed to act as informed traders. Table I.2 suggests that the behaviour of HFTs is perhaps not what could be expected, in theory, from that of market makers[10], see Section 4.3.

In addition to this, Table I.3 shows that in terms of market share and aggressiveness, all HFTs do not behave similarly. In particular, 50% of the activity among HFTs is held by 4 members, as they have very different levels of aggressiveness.

	Aggressive/Passive ratio over 50%	Aggressive/Passive ratio below 50%
Part in number	40%	60%
Part in traded amounts	55%	45%
Mean of ratio (std)	67% (10%)	25% (18%)

Table I.3 – Disparity in aggressiveness between HFTs.

4.3 Typical behaviour of market makers

HFTs often position themselves as market makers that theoretically act only as intermediaries between buy-side and sell-side flows. This is acknowledged by the academic literature. For

[10] We keep in mind that we provide statistics on all HFTs. We could have some HFTs having a behaviour very similar to what is expected from a market maker, among numerous HFTs not acting like market makers. We are not able to make the difference here between them. Thus we conclude that, as a whole, HTFs do not act as market makers.

example, according to [27]: “Given these results, HFTs appear to be a new form of market makers". Therefore, an intuitive scheme would be an HFT buying from a non-HFT and selling to another non-HFT. In this case, one might expect that the share of HFTs in amounts traded MS_{HFT} is below 50%. It could at most be equal to 50% in case all the members providing liquidity are HFTs. However, Table I.2 shows that this is not what happens in practice. In Section 4.4 and 4.5, we will be interested in how HFTs share in amounts traded varies according to time and scheduled announcements.

As already mentioned above, market makers are intermediaries that minimise inventory risk by holding positions for very short periods (only intraday). An approximation for market makers' average gain per trade is $S/2 - \tilde{\sigma}$, where S is the spread and $\tilde{\sigma}$ is the volatility per trade, see [41, 86, 115]. It is proportional to the spread, but decreases with the volatility per trade. In light of the above, one can think that typical market makers reduce their liquidity provision strategies in anticipation to large price moves triggering volatility bursts (for example at the time of scheduled announcements).

Still, it could be expected that market makers aggressive/passive ratio $R_{A/P}$ does not exceed 50%. Indeed, this value is the one that should be obtained in a scenario of a market consisting only of buyers (or sellers), where the market maker does not succeed in unwinding its position by posting limit orders and use aggressive orders.

4.4 Day-to-day analysis

In Figure I.8, we display the evolution of historical volatility in parallel with that of adjusted amounts traded[11] by HFTs and non-HFTs.

[11]The adjusted amounts traded are the amounts traded multiplied by the variation of CAC 40 index between the first day of the analysis and the day into question. Thus they are not affected by the fall in prices in the first part of the period under review. We apply in fact the same principle as for the definition of AMD in Section 3.1.

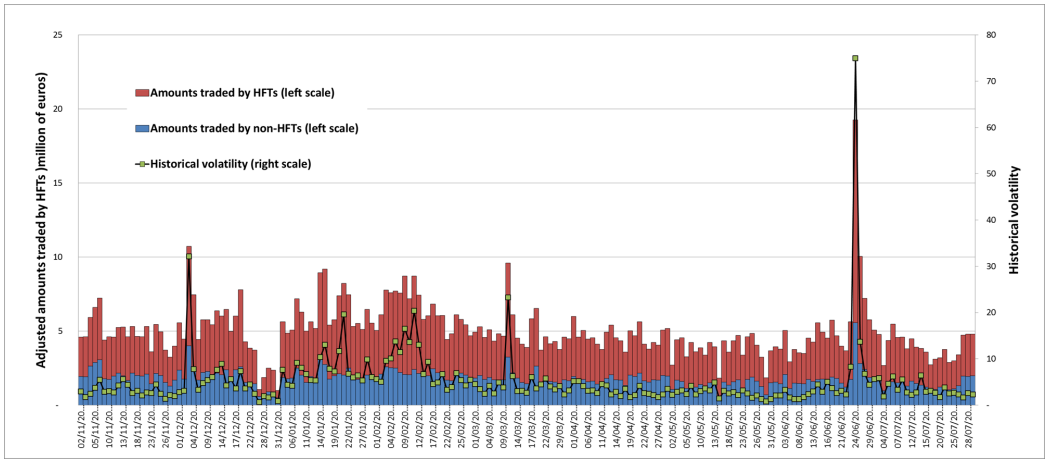


Figure I.8 – Daily evolution of historical volatility and amounts traded by HFTs and non-HFTs: Days with high historical volatility coincide with large traded volumes. HFTs and non-HFTs have essentially the same behaviour in terms of amounts traded.

Figure I.8 shows as expected a positive relationship between volatility and traded volume. The variations of HFTs and non-HFTs amounts traded are essentially the same over the period.

In Figure I.9, we display HFTs share in amounts traded together with the aggressive/passive ratio of HFTs over the period.

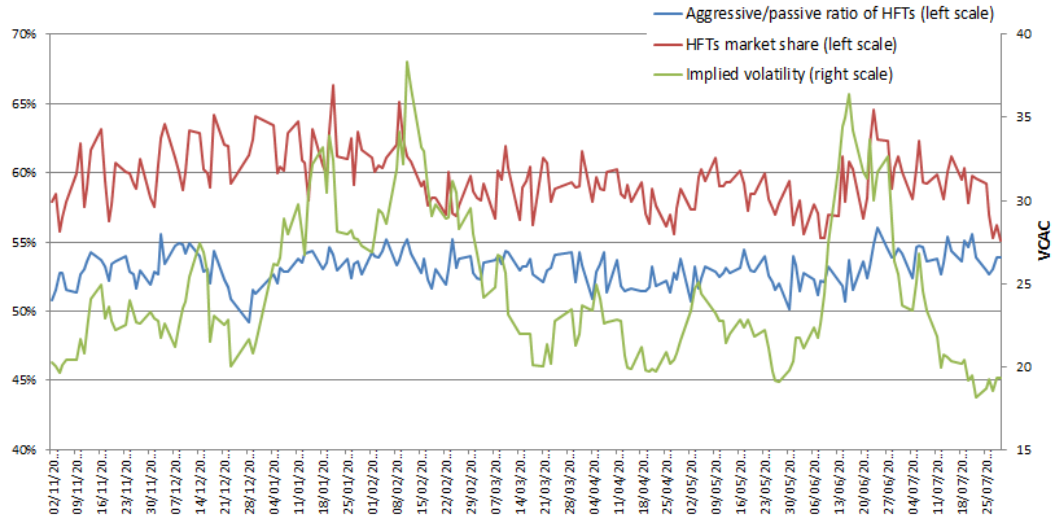


Figure I.9 – HFTs trading behaviour over the period: HFTs share in amounts traded is slightly affected by the market stress, in average close to 60%. Contrariwise, the aggressive/passive ratio is relatively stable throughout the period, close to 53%.

Figure I.9 shows that HFTs share in amounts traded is slightly dependent on the market stress, and oscillates around 60% (auctions excluded). In the beginning of the period, till February 9, the market is stressed, and at the same time, the HFTs market share is on average higher than 60%. From mid February till June, the market stress decreases and is stable. During this period, the HFTs market share decreases and is in general less than 60%. In June, the market stress increases again. Concomitantly, we can see a slight increase in the HFTs market share, which becomes higher than 60%. Though, their aggressiveness is quite stable throughout the period: on average, they consume more liquidity than they provide, with an aggressive/passive ratio around 53%.

It is clear from Figure I.8 that spikes in historical volatility coincide with sharp increases in amounts traded. However, the activity of HFTs does not seem to be largely affected by variations in implied volatility over the period: their share in amounts traded is moderately affected by the market stress and their aggressive/passive ratio remains quite constant for the entire period under study, see Figure I.9.

4.5 Intraday analysis

In Figure I.10, we consider the intraday evolution of amounts traded and historical volatility.

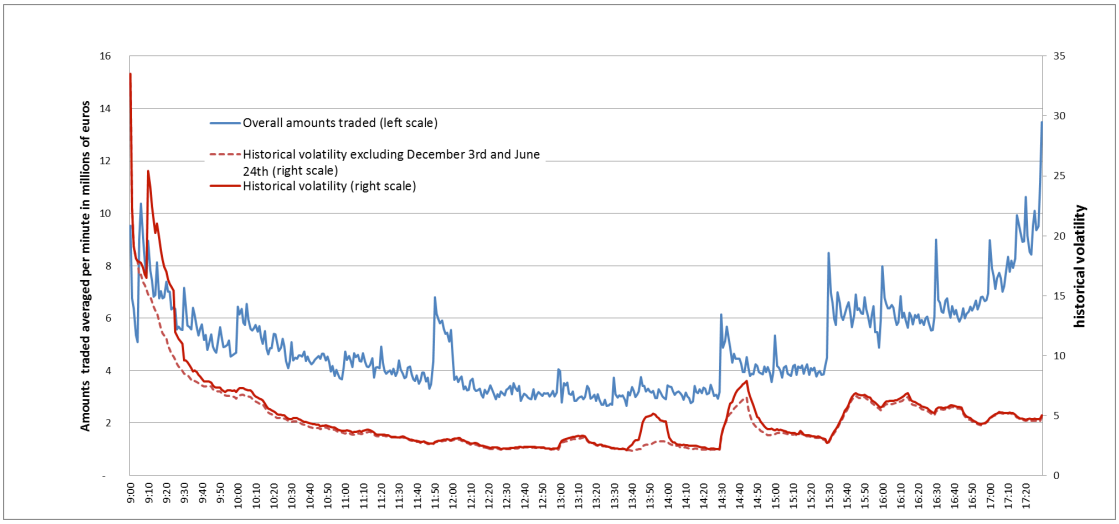


Figure I.10 – Intraday evolution of historical volatility and amounts traded: they are positively correlated, except at the very end of the day when volatility decreases as amounts traded increase.

Intraday analysis shows the same relationship between volatility and amounts traded as that mentioned above. Indeed, we see in Figure I.10 that the greater the price moves (high historical volatility), the larger the amounts traded. However, at the very end of the day, this relationship no longer holds: amounts traded increase sharply while historical volatility decreases.

In Figure I.11, we show the intraday evolution of HFTs amounts traded and HFTs share in terms of buy and sell trades.

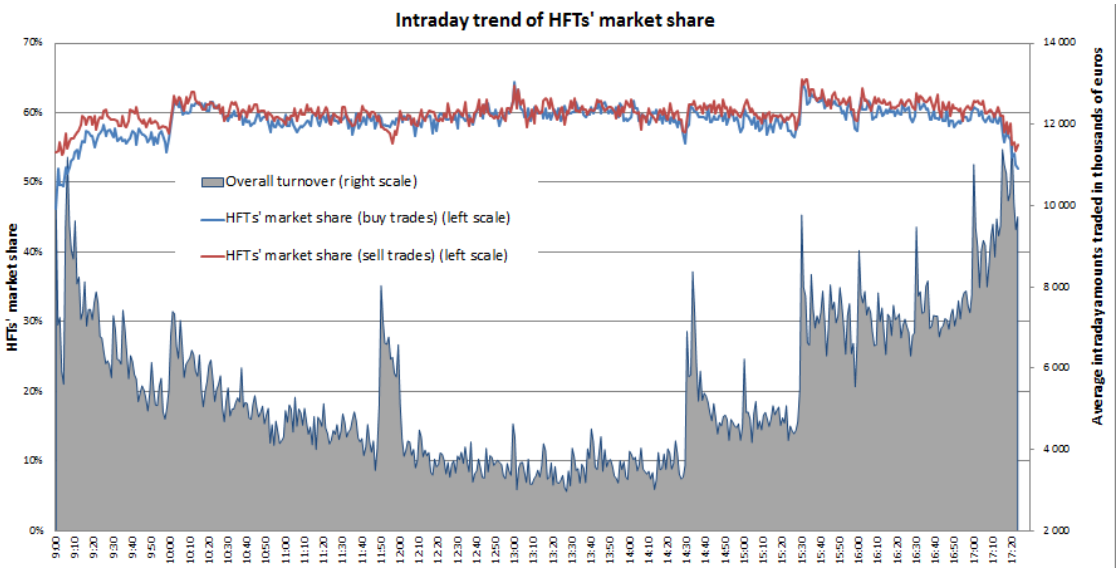


Figure I.11 – Intraday evolution of HFTs share in amounts traded: it increases at the beginning of the day and then stabilises at about 60% until the U.S. market opening, after which it rises to 65%. Amounts traded are much larger at the beginning and end of the day. The intraday HFTs share in sell trades is similar to that in buy trades except at the beginning of the day where HFTs share in sell trades is slightly larger.

In Figure I.12, we display the HFTs market share in terms of market depth (already studied in Section 3.4), the aggressive/passive ratio $R_{A/P}$ and the HFTs share in amounts traded MS_{HFT}.

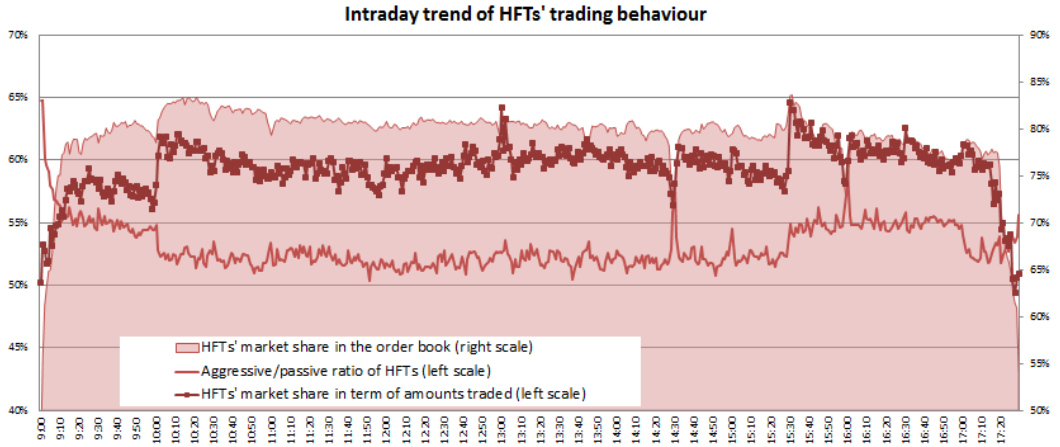


Figure I.12 – Intraday trading behaviour of HFTs: HFTs are slightly more aggressive after the U.S. market opening (the aggressive/passive ratio $R_{A/P}$ is equal to 55% versus 52% the rest of time) while their share in amounts traded MS_{HFT} increases at the same time. At the beginning and end of the day, their aggressive/passive ratio is higher, which is actually due to a decrease in their contribution to market depth. Their share in terms of amounts traded is also lower during these periods.

Figures I.11 shows that amounts traded (purchase and sale amounts follow similar trends, so no distinction is made in the sequel) by the entire market decrease in the first part of the day, then increase during the CAC 40 and EUROSTOX 50 futures price auctions (between 11.45 a.m and 12 p.m.), on the first important announcements time, at 2:30 p.m., and in particular after the U.S. market opening at 3:30 p.m. In the last hour, volumes swell until the close.

4.5.1 Intraday behaviour of HFTs

A closer examination of Figures I.11 and I.12 reveals four distinct phases:

- At the beginning of the day, HFTs' share in amounts traded increases gradually from 50% to 58%. At the same time, their aggressive/passive ratio falls from 65% to 55%. A decrease of the aggressive/passive ratio results either from a decrease in aggressive trades and/or the addition of passive orders. In this case, not only do HFTs reduce their aggressive orders, but they also increase their passive contribution to the order book.
- Before the U.S. market opening, amounts traded by the market as a whole are relatively stable, as is HFTs' share in amounts traded (60%) and their aggressive/passive ratio (52%).
- The U.S. market opening leads to an increase in activity which, in the case of HFTs, is particularly pronounced. Their share in amounts traded rises from 58% to 65%. Their

aggressive/passive ratio also increases from 52% to 55%. This surge in aggressiveness is most likely due to the appearance of arbitrage opportunities. HFTs therefore consume more liquidity than they provide.

- At the end of the day, HFTs gradually withdraw and their share in amounts traded decreases from 60% to 55%. Their aggressiveness increases due to the decrease in their market depth and probably their desire to unwind positions before the close, which tends to incite aggressive behaviour. It should be noted that at the end of the day, the decrease in the HFTs' share in amounts traded is compounded by a change in the behaviour of non-HFTs who dramatically increase their amounts traded.

4.5.2 Specific behaviour around scheduled announcements

From Figure I.12, it appears that the behaviour of HFTs is particularly influenced by announcements and times of high historical volatility (for example after the opening of the U.S. market). A specific analysis of announcement periods shows a temporary change in the behaviour of HFTs at these times.

- Before an announcement, as already seen in Section 3.4, HFTs sharply reduce their presence in the order book, much more than other participants, and their passive trades decrease accordingly. This explains the fall in their share in amounts traded before the 2:30 p.m. announcement, which declines from 60% to 56%, as well as the increase in their aggressive/passive ratio which rises from 52% to 57%. Here the aggressive/passive ratio of HFTs increases not because they increase their aggressive flows, but because they decrease their passive liquidity provision in the order book.

- After an announcement, prices are often severely affected. This leads to a sudden increase in aggressive flows from HFTs (as is the case after the U.S. market opening). HFTs share in amounts traded therefore rises from 56% to 62% at 2.30 p.m. and they are temporarily more aggressive, although their passive provision of liquidity in the order book quickly returns to its prior level, see Figures I.5 and I.6. In contrast to what happens before announcements, HFTs do not reduce their passive orders after announcements but increase their aggressive orders. When the price fluctuates widely, the aggressive/passive ratio for these participants generally exceeds 50%. This probably reflects transformation of some mean reversion strategies into short-term directional (momentum) ones, which can hardly be considered as market making.

4.5.3 Focus on the 10 most impactful announcements at 2.30 p.m.

In the spirit of what we did in Section 3.4.3, we analyse the changes in the HFTs share in amounts traded and their aggressive/passive ratio around the same 10 announcements at 2:30 p.m. that had the greatest impact (the highest historical volatility) on prices during the period under review.

In Figure I.13, we display the intraday evolution of HFTs aggressive/passive ratio, HFTs share in terms of market depth and HFTs share in amounts traded around 2:30 p.m. over all the period on the one hand, and considering only the 10 most impactful announcements at 2:30 p.m. on the other hand.

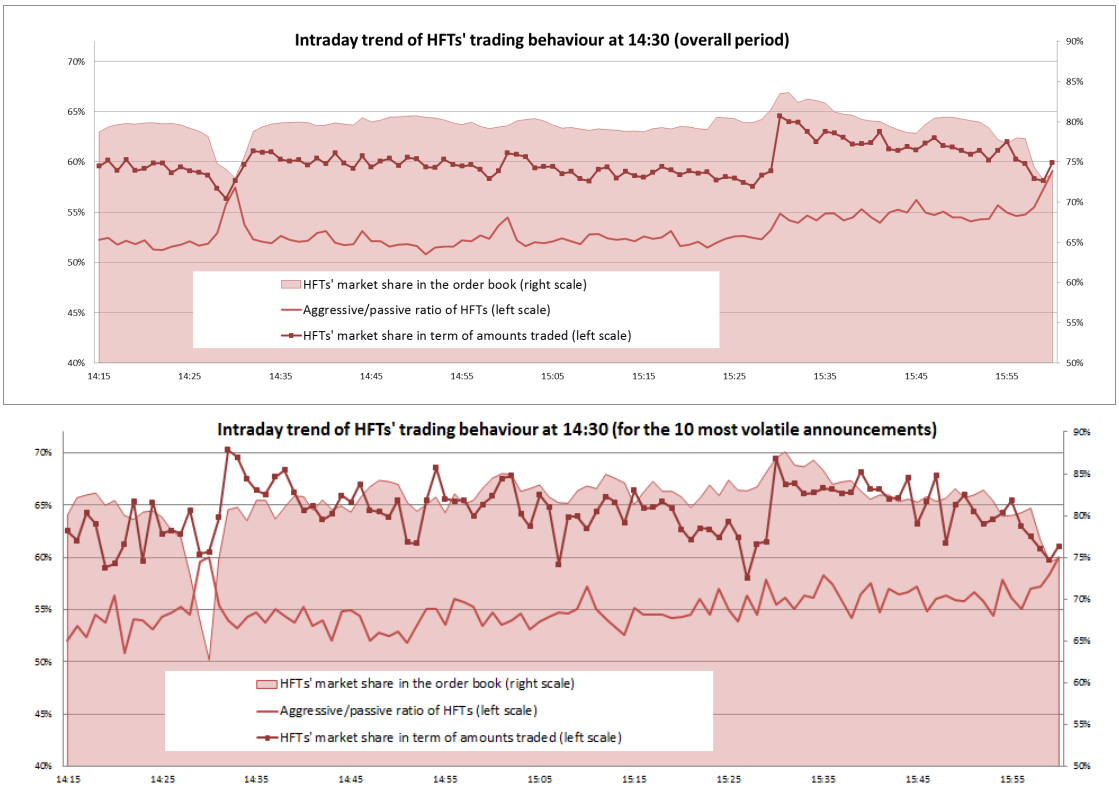


Figure I.13 – Evolution of the HFTs trading behaviour at 2:30 p.m. across all the days (top graph) and across the 10 days with the most volatile announcements (bottom graph). Announcements are accompanied by a reduction in HFTs passive orders in the order book, which results in an increase of their aggressive/passive ratio. These variations are more pronounced in periods of major announcements.

Similar patterns are observed on both graphs in Figure I.13. However, more striking effects are obtained when considering the 10 most impactful announcements only. In particular, we see that after the 10 most impactful announcements (at about 2:33 p.m.) HFTs share in amounts traded increases sharply to 71%, well above the value usually reached (62%).

5 A more quantitative analysis around the 4 p.m. announcement

In this section, we analyse HFTs behaviour by differentiating between days with announcements and days without announcement at 4 p.m.

The goal of this section is to test for a specific “announcement effect" impacting the attitude of HFTs. We take the minute at which a news is released as a dummy for this effect. Moreover, we try to control the volatility component of news. We expect more volatility during news and we want to make the difference between the usual reaction of HFTs to an increase of volatility, and an actual “scheduled announcement effect” that comes on top of this volatility effect.

It is natural to focus on the announcements happening at 2:30 p.m. and at 4 p.m. However, announcements at 2:30 p.m. do not constitute a good database for our analysis: there are 181 days with announcements versus only 10 days without announcements, which is not sufficient for a robust statistical study. This is why we focus on announcements happening at 4 p.m. We only consider news related to the U.S economy. Indeed, as already mentioned in Section 2.3.3, they represent the majority of announcements at 4 p.m. We classify the days into two groups: days with U.S announcements at 4 p.m. (140 days) and days without announcements at 4 p.m. (51 days). Then we restrict the analysis to the time interval 3:40 p.m. to 4:50 p.m.

We split time in disjoint intervals of one minute, indexed by the starting time of the interval denoted by t (interval with t =4 p.m. contains all events occurring after 4 p.m. and before 4:01 p.m.). Let B (for *Before*), D (for *During*) and A (for *After*) be dummy variables on these bins of one minute:

- If t =3:58 p.m. or t =3:59 p.m. then $B = 1$, otherwise $B = 0$.
- If t =4 p.m. then $D = 1$, otherwise $D = 0$.
- If t =4:01 p.m. or t =4:02 p.m. then $A = 1$, otherwise $A = 0$.

To account for a volatility effect (since HFTs react to volatility in general) we take here as proxy for historical volatility of stock s on day d in the time interval starting at t the price range during the considered minute[12]. This price range $\sigma^s(t,d)$ is the difference between the maximum price and the minimum price between time t and time $t+60$ seconds.

Moreover, we expect the HFTs behavioural change not to be a function of the absolute level of volatility, but of the intensity of “abnormality” of the volatility. Thus we use a “renormalised volatility" which is equal to the price range computed above, divided by the average value

[12]In the regression framework of this section, we can consider each stock separately and do not need to aggregate them. In particular, volatility is computed stock by stock in very short time intervals, hence the use of a max-min type measure.

between 3.40 p.m. and 4.50 p.m. of price range over all the days in our database for the considered stock. This new indicator is denoted by $\sigma^s_{norm}(t,d)$.

We do an Ordinary Least Squares (OLS) regression where we want to explain the following variables:

- HFTs market share in terms of market depth.
- HFTs share in amounts traded.
- HFTs aggressive/passive ratio.

Our methodology is the following. For each of these variables, a regression is first done on days without announcement. It is expected to account for a volatility effect and other phenomena such as news which are not in our database. We retain only the variables having a t-stat large enough to show a significant effect, and redo the OLS with these variables only.

Then we apply another linear model on days with announcements, controlling for the effects identified in the first regression. This enables us to isolate the impact of explanatory variables in the presence of announcements on our variables of interest.

5.1 Analysis of HFTs market share in terms of market depth

In Table I.11 in appendix, we display the regression of HFTs market share in terms of market depth during days without announcement. The coefficients of σ_{norm}, B, D and A are not significant. We give the results of the regression only in terms of significant variables in Table I.4.

Variable	Coef.	Std. err.	t	$P>\|t\|$	95% Conf. Int.
Constant	0.7866	0.003	302.564	0	[0.781, 0.792]

Table I.4 – Regression for HFTs market share in terms of market depth during days without U.S. announcement at 4 p.m., using significant explanatory variables only.

Table I.4 shows that HFTs market depth during days without announcement cannot be better explained than by a constant equal to 78.66%. There is no noticeable influence of the time slots (A, B and D) and of the renormalised volatility σ_{norm}.

In Table I.5, we consider the regression of HFTs market depth during days with announcements once the constant of Table I.4 is subtracted.

Variables	Coef.	Std. err.	t	$P>\|t\|$	95% Conf. Int.
Const.	0.011	0.002	6.302	0	[0.008, 0.015]
σ_{norm}	−0.0045	0.002	−2.664	0.008	[-0.008, -0.001]
B	−0.0520	0.004	−14.404	0	[-0.059, -0.045]
D	−0.1507	0.005	−28.941	0	[-0.161, -0.141]
A	−0.0283	0.004	−7.797	0	[-0.035, -0.021]

Table I.5 – Regression for the contribution of HFTs to market depth during days with U.S. announcements at 4 p.m., once the effect of Table I.4 removed.

Table I.5 shows that all variables are significant: announcements do have an impact on HFTs market share in terms of market depth. Moreover it says that the decrease of the market share of HFTs on liquidity provision goes beyond a simple volatility effect. HFTs do contribute less to market depth before, during and after announcements, even once a volatility effect is taken into account. Their contribution decreases by about 5% before the announcement, 15% during the announcement and 3% after the announcement. This result is consistent with the observations in Section 3.4.

5.2 Analysis of HFTs aggressive/passive ratio

In Table I.12 in appendix, we display the regression of the HFTs aggressive/passive ratio during days without announcement. The coefficients of B and A are not significant. So we redo the regression only with significant variables. Obtained results are given in Table I.6.

Variables	Coef.	Std. err.	t	$P>\|t\|$	95% Conf. Int.
Const.	0.5340	0.002	228.198	0	[0.529, 0.539]
σ_{norm}	0.0111	0.002	5.023	0	[0.007, 0.015]
D	0.0169	0.007	2.494	0.013	[0.004, 0.03]

Table I.6 – Regression for the aggressive/passive ratio of HFTs during days without U.S. announcement at 4 p.m., using significant explanatory variables only.

Table I.6 shows that the HFTs aggressive/passive ratio is essentially more than 53.4%. The ratio increases with renormalised volatility (HFTs as a whole turn to remove more liquidity when volatility increases), and increases for about 1.7% at 4 p.m. This indicates that some news other than the ones in our database may have an impact on some HFTs, or that some HFTs become more aggressive automatically at 4 p.m., whether there are news or not[13].

[13] It may be a way to be "protected" against the operational risk associated to "miss" an announcement.

In Table I.13 in appendix, we give the results of the regression of the HFTs aggressive/passive ratio during days with announcements corrected by the effects in Table I.6, denoted by $res_{R_{A/P}}$:

$$res_{R_{A/P}} = R_{A/P} - 0.5340 - 0.0111 \cdot \sigma_{norm} - 0.0169 \cdot \delta_D.$$

The coefficient of A is not significant and we consider the regression with significant variables only in Table I.7.

Variables	Coef.	Std. err.	t	$P > \|t\|$	95% Conf. Int.
Const.	0.0113	0.001	9.029	0	[0.009, 0.014]
σ_{norm}	−0.0053	0.001	−4.475	0	[-0.008, -0.003]
B	0.0184	0.003	7.116	0	[0.013, 0.023]
D	0.0268	0.004	7.237	0	[0.02, 0.034]

Table I.7 – Regression for $res_{R_{A/P}}$ during days with U.S. announcements at 4 p.m., using significant explanatory variables only.

Table I.7 shows that during days with announcements, an increase of volatility has the opposite effect than the one during days without announcement. It may be an adverse selection phenomenon. Moreover, HFTs turns to be significantly more liquidity consumers just before and during the announcements (respectively for about 2% and 2.5%). This result is consistent with the observations in Section 4.5.

5.3 Analysis of HFTs share in amounts traded

In Table I.14 in appendix, we display the regression of HFTs share in amounts traded during days without announcement. The coefficients of A, B and D are not significant. We give the results of the regression using significant variables only in Table I.8.

Variables	Coef.	Std. err.	t	$P > \|t\|$	95% Conf. Int.
Const.	0.5557	0.003	208.543	0	[0.551, 0.561]
σ_{norm}	0.0473	0.003	18.740	0	[0.042, 0.052]

Table I.8 – Regression for HFTs share in amounts traded during days without U.S. announcement at 4 p.m., using significant explanatory variables only.

Table I.8 shows that in the absence of news in our database, HFTs share in amounts traded is essentially higher than 55%, and increases with renormalised volatility.

In Table I.15 in appendix, we give the results of the regression of HFTs share in amounts traded during days with announcements, corrected by the effect in Table I.8, denoted by

res_{MS}:

$$res_{MS} = MS - 0.5557 - 0.0473 \cdot \sigma_{norm}$$

The coefficient of A is not significant. So we conduct the regression using significant explanatory variables only. Results are given in Table I.9.

Variables	Coef.	Std. err.	t	$P > \|t\|$	95% Conf. Int.
Const.	0.0097	0.001	16.799	0	[0.009, 0.011]
B	−0.0346	0.003	−10.228	0	[-0.041, -0.028]
D	−0.0469	0.005	−9.869	0	[-0.057, -0.038]

Table I.9 – Regression for res_{MS} during days with U.S. announcements at 4 p.m., using significant explanatory variables only.

Table I.9 shows that HFTs share in amounts traded decreases by 3.5% before the announcement and by 4.7% during the announcement. Recall that it is corrected from usual reaction to volatility. These results are consistent with the observations in Section 4.5.

5.4 Summary

All these regressions point out in a quantitative way that the behaviour of HFTs around announcements cannot be read as a simple reaction to associated variations of volatility. Around a scheduled announcement, on top of usual reactions to volatility, HFTs provide 15% less liquidity, are slightly more aggressive and trade less. On the contrary, when no announcement is planned, their attitude towards an increase of volatility goes in the opposite direction (trading more). We thus identify a “change of regime" in the presence of scheduled news.

6 Detailed analysis of two events

So far, we have analysed the average impact of various announcements on HFTs activity. In this section, we focus on two specific days characterised by announcements with very high impact on historical volatility (sharp price changes). The first day we consider is the 3^{rd} of December 2015 (ECB announcements) and the second one is the 24^{th} of June 2016 (Brexit day).

6.1 Focus on the 3^{rd} of December 2015

The 3^{rd} of December 2015 was one of the most volatile days of the entire period due to several ECB announcements. We first describe the three announcements that took place during this day: one misleading announcement, and two official announcements. Second, we investigate the consequences of the misleading announcement, and finally the short-term and longer-term impacts of the official announcements.

6.1.1 Description of the day

About two weeks before the 3^{rd} of December 2015 and following the publication of the ECB's minutes, markets were optimistic about the new monetary policy to be implemented by the ECB. The ECB, which was seeking to fight low inflation (0.1% in November) and stimulate the European economy, had two levers at its disposal:

- Lower its deposit rate, which was already in negative territory (-0.2%).
- Bolster its purchase programme (quantitative easing) by extending the programme's duration and/or increasing the monthly amount of asset purchases (EUR 60 billion). The market was expecting not only an extension of the programme's duration but also an increase in its monthly amount.

The ECB publishes its announcements on Thursdays at 1:45 p.m. and holds a press conference the same day at 2:30 p.m. In Figure I.14, we display the fluctuations of the CAC 40 index and the VCAC due to the various announcements on the 3^{rd} of December 2015.

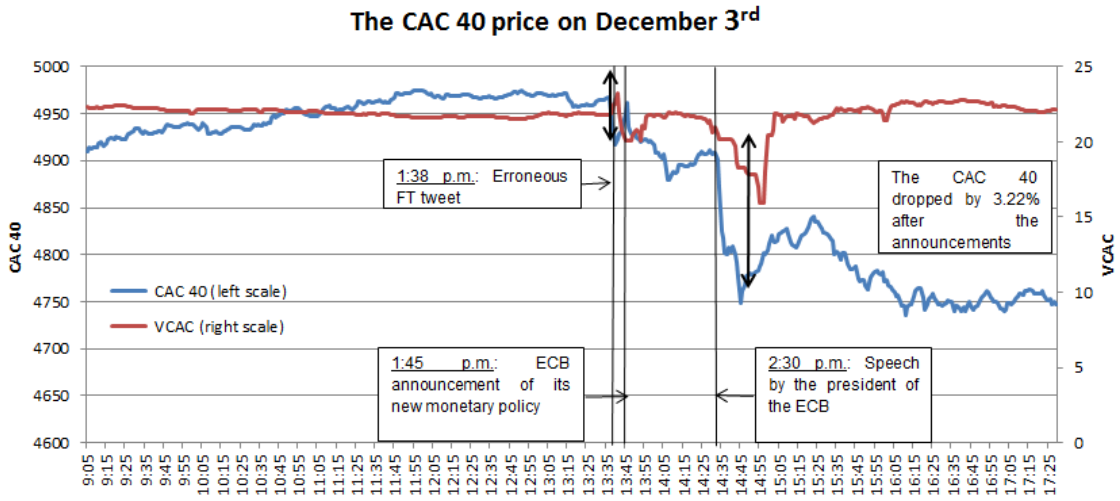


Figure I.14 – The CAC 40 index and the VCAC on the 3^{rd} of December 2015: there were three major announcements during the day. First, an information leak from the Financial Times at 1:38 p.m., followed by two official ECB statements at 1:45 p.m. and 2:30 p.m.

We see in Figure I.14 that the day was marked by three major announcements:

- At 1:38 p.m. (a few minutes before the 1:45 p.m. official announcement) the Financial Times announced on Twitter that the ECB would leave rates unchanged. This misinformation led to market distortions: at 1:38 p.m., the CAC 40 fell by 0.56% but regained its initial value just before the announcement of 1:45 p.m.

- At 1:45 p.m., the ECB announced that it was lowering its deposit rate by 10 basis points, from -0.2% to -0.3% from the 9^{th} of December, and the CAC 40 then began to decline gradually.
- However, since the market was still probably optimistic about other monetary policy measures to be communicated during the press conference held by ECB President Mario Draghi, the fall was not so significant and the index started to go back up from 2:09 p.m.
- At 2:30 p.m., Mario Draghi announced that the quantitative easing programme launched in March 2015 would be extended until at least the end of March 2017, but with no increase in the amount of purchases. This came as a severe disappointment to the market. The CAC 40 dropped significantly from 2:33 p.m. to 2:45 p.m.: in only 12 minutes, it lost 3.22% amid very heavy trading.

For the period under review, the 3^{rd} of December was the most volatile day[14]. It therefore provides an interesting case study for HFTs reactions to several consecutive major price shocks in terms of liquidity consumption and provision.

We now present a series of graphs about liquidity, amounts traded and HFTs behaviour that we analyse in details in Sections 6.1.2 and 6.1.3
In Figure I.15, we display the evolution of the bid-ask spread in ticks and HFTs market share in terms of market depth around the announcements: from 1 p.m. to 3 p.m.

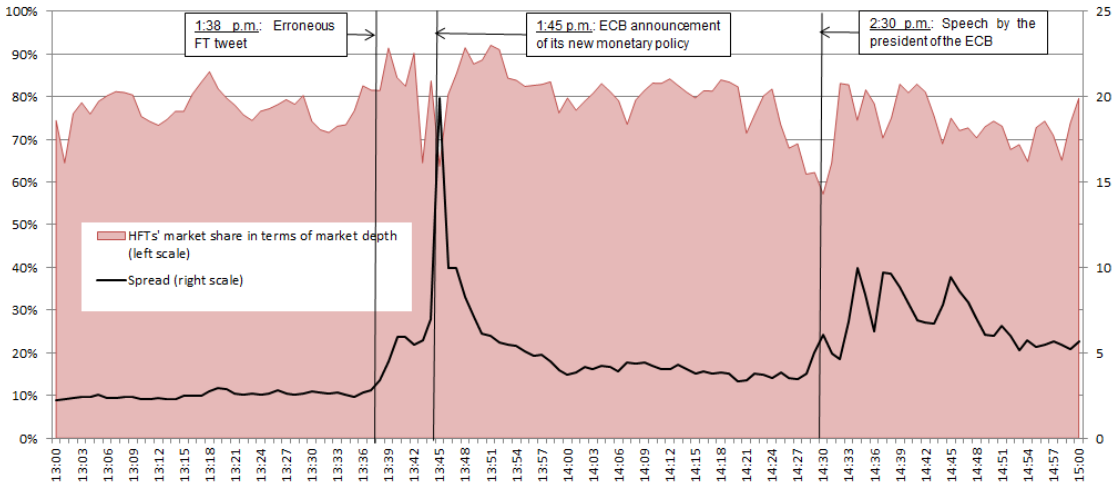


Figure I.15 – Evolution of the spread and HFTs market share in terms of market depth on the 3^{rd} of December 2015: a few minutes before the announcements, spread widened after the withdrawal of passive orders by all participants and by HFTs in particular.

[14]In fact, the day of 24^{th} of June 2016 was somehow more volatile but the volatile part of the day occurred pre-market.

In Figure I.16, we give the HFTs and non-HFTs market depth, and the HFTs market share in terms of market depth during all the day.

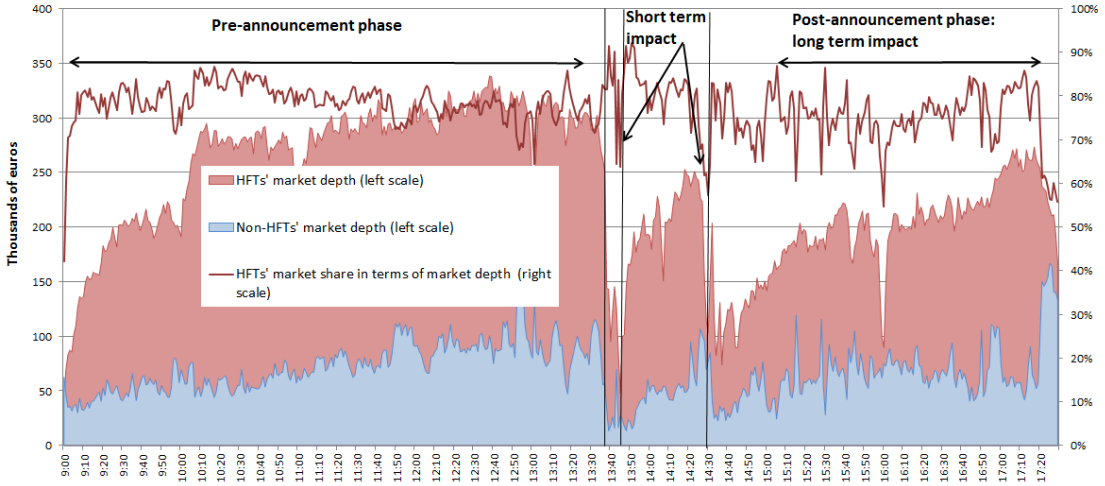


Figure I.16 – Evolution of market depth on the 3^{rd} of December 2015. The announcements had a strong immediate impact: a very sparse order book, with the share of HFTs in market depth terms falling from 82% to 57% just after the announcement at 2.30 p.m. They also had a long-term impact: less liquidity in the order book than in the pre-announcement phase.

In Figure I.17, we present the HFTs and non-HFTs aggressive flows, and the HFTs share in aggressive flows.

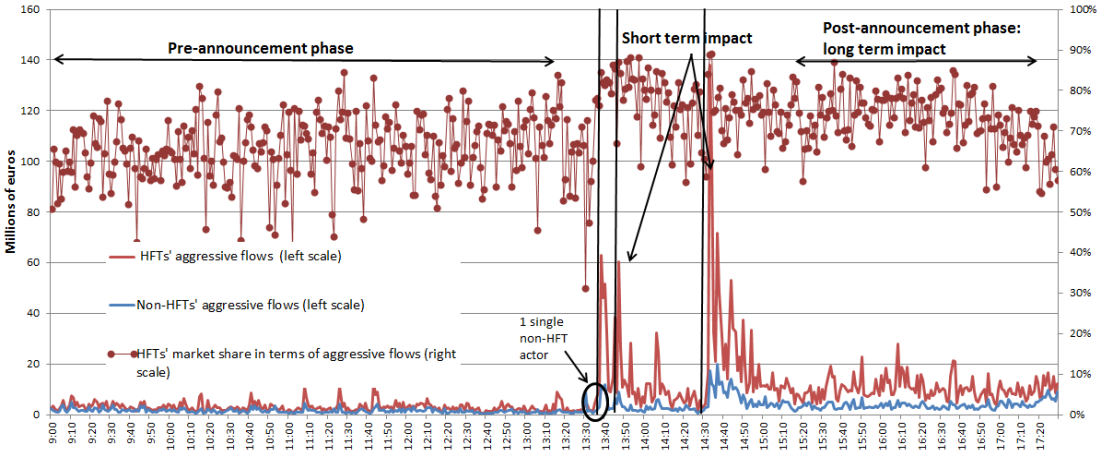


Figure I.17 – Evolution of aggressive flows on the 3^{rd} of December 2015. The announcements had a strong immediate impact: a much more active market, with the share of HFTs in aggressive trade terms increasing from 76% to 84% just after the announcement at 2.30 p.m. They also had a long-term impact: a more active market than in the pre-announcement phase, with HFTs share in aggressive trades increasing from 65% on average before the announcements to 73% after the announcements.

In Figure I.18, we consider the HFTs share in amounts traded and aggressive/passive ratio on the 3^{rd} of December.

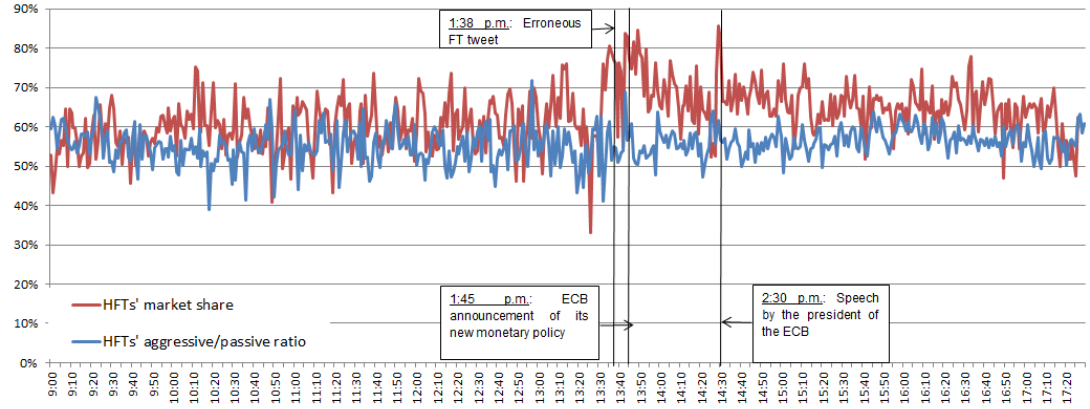


Figure I.18 – Evolution of the HFTs share in amounts traded on the 3^{rd} of December 2015. The announcements had an immediate impact on HFTs share in amounts traded, which rose significantly, and a long-term impact, as HFTs share in amounts traded remained high until the end of the day. In contrast, the aggressive/passive ratio was unaffected by these announcements.

6.1.2 Impact of the misleading announcement

At 1:38 p.m., the misleading announcement by the Financial Times was unexpected. The impact of this information leak on market participants was therefore different from that described previously for scheduled announcements. In particular:

- Before scheduled announcements, there is generally a decrease in HFTs market share in terms of market depth in the order book, as HFTs withdraw their passive orders to a greater extent than non-HFTs.
- In this specific case, reactions were different: Figure I.16 shows that after the misleading announcement at 1:38 p.m., HFTs market share in market depth increased from 80% to 90%. HFTs and non-HFTs both withdrew from the order book. However, HFTs market depth fell by about 50%, while for non-HFTs the decrease was around 75%. At 1:38 p.m., it is likely that non-HFTs could not track the price drop and hence were unable to update their orders, whereas HFTs were able to quickly move their orders and maintain part of their passive liquidity provision in the order book.

6.1.3 Impact of the official announcements

Before the announcements Before the two announcements involving accurate information (1:45 p.m. and 2:30 p.m.), market participants behaviour was in line with that described previously in Sections 3.4 and 4.5. Nevertheless, it was quite amplified.

Before each of these two announcements, market depth decreased mainly because HFTs withdrew. The share of HFTs in terms of market depth decreased from 80% to 60%, see Figure I.16. Recall that in such situation, it decreases usually on average from 80% to 73%, see Figure I.5. This withdrawal induced a very sharp widening of the bid-ask spread, see Figure I.15. The average spread on French securities reached its widest for the period under review at almost 19.9 ticks while in the three hours preceding the first announcement, it stood on average at 2.7 ticks.

The short-term impact after the official announcements Right after each of these announcements, amounts traded increased significantly. Figure I.17 shows for example that three minutes after the 2:30 p.m. announcement, they were multiplied by almost 40 compared to before the announcement. Furthermore, HFTs share in amounts traded reached 82% about two minutes after the announcement, see Figure I.18 (during this period, HFTs represented approximately 90% of aggressive trades). The order book gradually recovered approximately 30 minutes after each announcement but never reached its pre-announcement levels, in terms of market depth and spread, see Figure I.15 and Figure I.16. It is likely that some participants permanently reduced their presence in the order book after these two major shocks.

The long-term impact of the announcements: pre-announcement phase versus post-announcement phase The 3^{rd} of December 2015 announcements had a very strong impact on prices, which lasted throughout the day.

- After the announcements, overall market depth offered at the three best price limits did not return to its initial level. It decreased from EUR 700,000 on average before the announcements to EUR 450,000 on average (between 3 p.m. and 5 p.m.), see Figure I.16.
- Between 3 p.m. and 5 p.m., high traded volumes were observed despite a thinner order book, see Figure I.17. In the pre-announcement phase (before 1:38 p.m.), the amount traded on average per minute was EUR 4.6 million and HFTs represented 65% of aggressive flows. In the post-announcement phase, this amount tripled and HFTs represented 73% of aggressive orders, which led HFTs share in amounts traded to increase from 59% (pre-announcement phase) to 65% (post-announcement phase).
- It should be noted, however, that HFTs aggressive/passive ratio was unaffected by these announcements. Indeed passive liquidity provision by HFTs (including at limits in the order book beyond the third one) and aggressive flows increased simultaneously.

6.1.4 Summary

The 3^{rd} of December 2015 was marked by three announcements that had strong impact on prices. The announcements were all accompanied by a significant withdrawal from the order book by all participants and by a sharp increase in traded volumes.

Contrary to what happens in typical situations, the order book recovered slowly and only partially after each announcement. It seems that certain participants (HFTs in particular) permanently withdrew their orders from the book in an over-volatile market environment.

Despite the sharp fall in HFTs passive liquidity provision at the three best limits during this period of stress, after the announcements, their share in amounts traded was even slightly higher than usually and they were particularly aggressive (market share in aggressive trades of almost 75% after the announcements).

6.2 Focus on the 24^{th} of June 2016 (Brexit announcement)

We now analyse another day characterised by an impactful news: the Brexit announcement on the 24^{th} of June 2016. As was the case on the 3^{rd} of December 2015, the announcement caused a high historical volatility. However, a major difference between these two days is that the Brexit news occurred prior to the market opening: the day of the vote was the 23^{rd} of June 2016 and the results were announced on the 24^{th} of June 2016, before the opening of the market. Therefore, this day constitutes an interesting case study for impactful announcement occurring pre-market. First we describe what happened around the 24^{th} of June 2016, specifying what the expectations of the market were. Then we analyse the impact on liquidity and on amounts traded of the Brexit news. The study of the various metrics in this section is split into two main periods: the beginning of the day (until about 12 p.m.), which casts uncertainty among market participants and saw very high levels of volatility, and the rest of the day, where the market stabilised.

6.2.1 Description of the day

On Friday, the 24^{th} of June 2016, markets were informed that United Kingdom had decided to end 43 years of membership in the European Union (EU), causing the most volatile pre-trading day over the period under study.

Figure I.19 displays the evolution of the CAC 40 index and of the VCAC before and after the Brexit announcement, from June 23 to June 27.

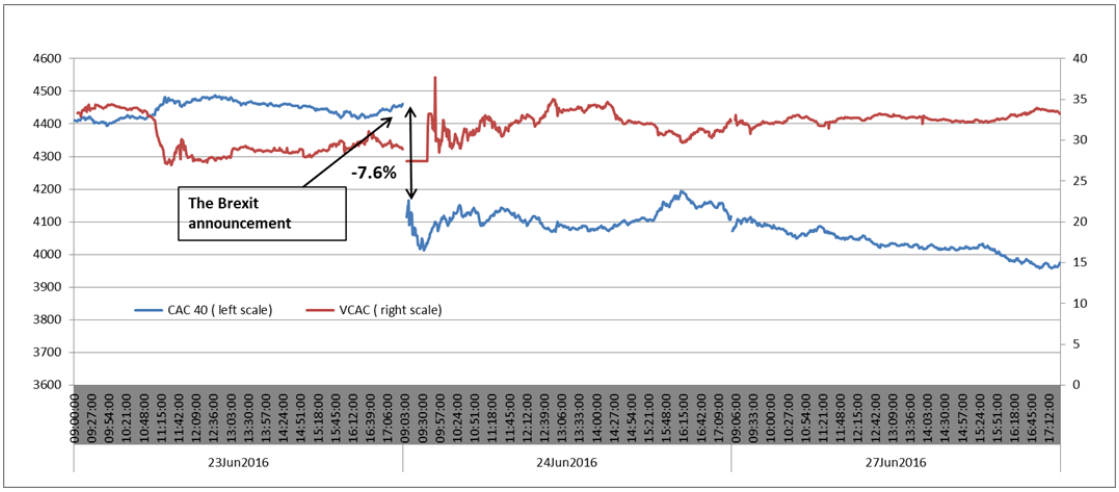


Figure I.19 – Evolution of the CAC 40 index and the implied volatility VCAC during the three days surrounding the Brexit announcement.

- Pre-Brexit: The Brexit referendum was an eagerly awaited historical decision that had been anticipated by the market several days before the actual day of the vote (June 23). Figure I.1 shows that days ahead of the vote results were characterised by high levels of market stress, starting in particular on June 7. The market saw wide price fluctuations and the indicator of implied volatility, the VCAC, began to increase from June 7 onwards (on that day, the VCAC stood at 21.73). It peaked at 36.41 on June 16. The day before the announcement, June 23, most investors and analysts expected that UK would vote to remain in the EU.

- The 24^{th} of June 2016, day of the Brexit announcement: The official vote results were published in the early morning of June 24, before European markets opening. Contrary to expectations, UK citizens decided to leave the European Union (52% of the electorate voted for Brexit). This disappointment was immediately reflected in the market: on Euronext, trading in 36 out of 40 stocks in the CAC 40 was halted at limit down at the open. The trading halt of several CAC 40 stocks lasted until 9:20 a.m.

  When the index finally opened, it was down 7.6% compared to the previous closing. It plunged by 8.04% to end the day at 4 106.73 points, after hitting a low of 4 007.97

points (-10.25%). In terms of amplitude and speed, this was one of the most severe market shocks since the 2008 crisis: it was the largest daily fall since October 2008.

- Post-Brexit: In the wake of the results announced on June 24, markets entered a period of uncertainty and instability that lasted until mid-summer.

6.2.2 Methodology of the analysis

Since the Brexit announcement occurred pre-market, similar analysis to that conducted for the 3^{rd} of December 2015, where we compared the variations of the metrics before and after the news cannot be done. Here the comparison of the metrics is made with respect to a reference period. To allow for an unbiased comparison, we choose a reference period during which the market is stressed (rise in the VCAC) pending the outcome of the referendum: from June 7 to June 23. Depths, spreads, trading amounts and aggressive/passive ratios are therefore quite comparable, not impacted by a variation in the implied volatility, but only by the announcement.

To be able to compare on an intraday basis the metrics (depths and aggressive flows) on June 24 with those in the reference period, the metrics in euros on June 24 are adjusted accounting for the value of the CAC 40. For a given metric at time t on June 24, $x_{24}(t)$, we define the adjusted metric at time t, $x_{adj}(t,24)$, by

$$x_{adj}(t,24) = x_{24}(t) \cdot \frac{P_{ref}(t)}{P_{24}(t)},$$

where $P_{24}(t)$ is the value of the CAC 40 on June 24 at time t and $P_{ref}(t)$ the average value of the CAC 40 at time t over the reference period.
After computing the adjusted metric $x_{adj}(t,24)$, we define the ratio of the metric x at time t, $r_x(t)$, by

$$r_x(t) = \frac{x_{adj}(t,24)}{x_{ref}(t)},$$

where $x_{ref}(t)$ is the metric in question averaged at time t over the reference period.

6.2.3 Impact on liquidity

Impact on the overall market depth In Figure I.20, we display the market depth during the reference period and the adjusted one on June 24.

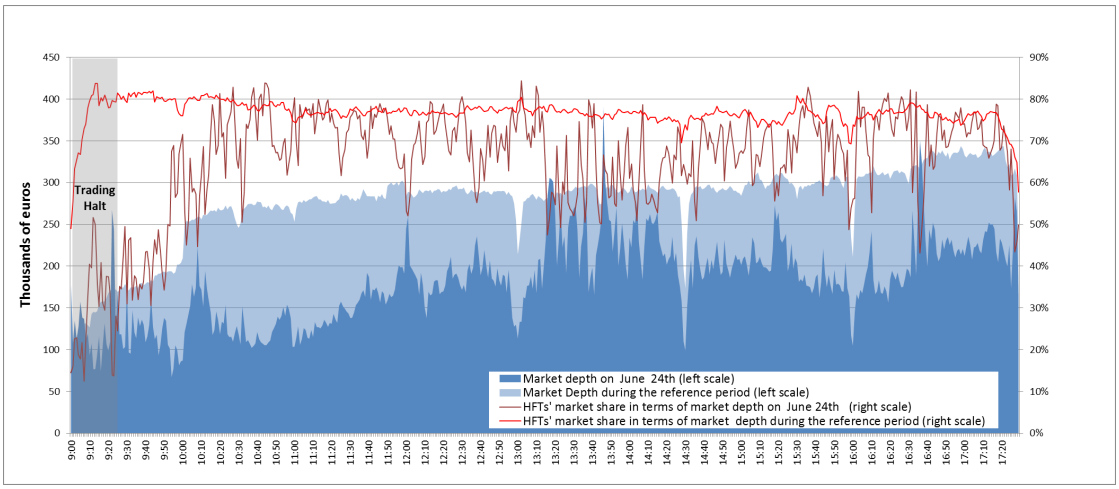


Figure I.20 – Comparison of the market depths between the reference period and June 24: the order book was not as deep on June 24 as in the reference period and HFTs showed less presence, particularly at the beginning of the day.

Figure I.20 shows that June 24 was marked by a significantly lower liquidity (HFTs and non-HFTs alike) than in the two weeks prior to the Brexit. At the beginning of the day, the uncertainty that followed the announcement led participants to sharply reduce their presence in the book: market depth offered was on 45 % lower than over the reference period. Liquidity then increased gradually during the day to stabilise around 33% lower than over the reference period.

Impact on the market depth provided by HFTs and non-HFTs: beginning of the day versus rest of the day Figure I.21 displays $r_{MD^{HFT}}$ and $r_{MD^{nHFT}}$.

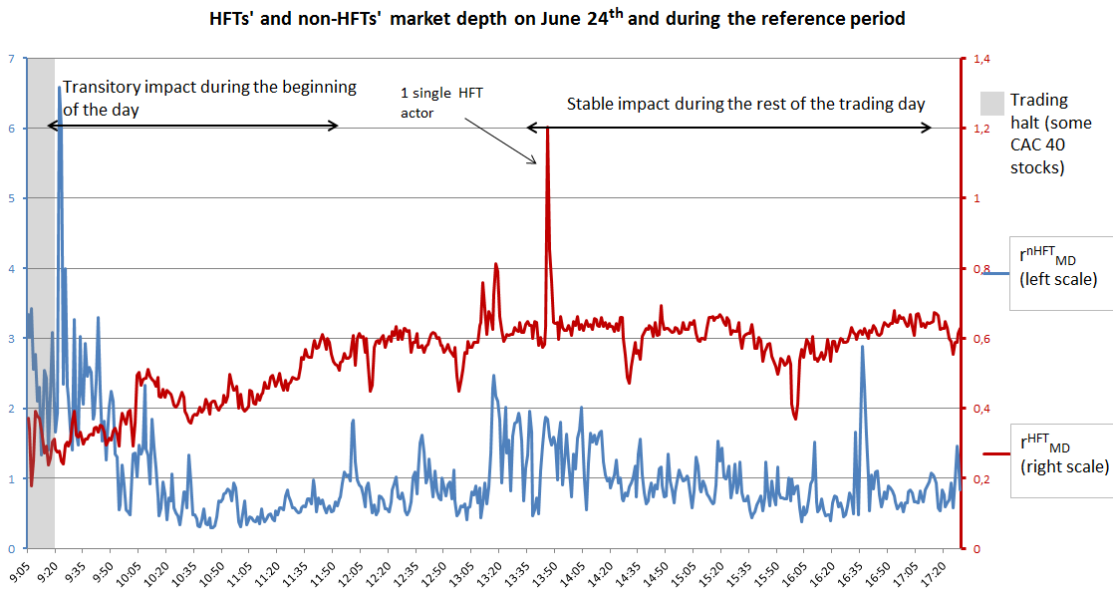


Figure I.21 – HFTs and non-HFTs market depth on June 24: ratios with respect to the reference period.

Figure I.21 shows that at the beginning of the day (transitional phase), HFTs and non-HFTs behaviours contrasted markedly:

- HFTs reduced their presence significantly in the order book: market depth offered represented one third of their usual market depth. This behaviour is no surprise since such participants avoid placing passive orders during periods of uncertainty fearing adverse selection.
- Non-HFTs partially offset HFTs reduced liquidity by providing a market depth twice larger than usually. Thus they supported liquidity, without being able to fully compensate the deficit caused by HFTs withdrawal.

  We can say here that non-HFTs became the main liquidity providers during this immediate post-Brexit period.

Over the rest of the day, HFTs gradually re-entered the order book but it was only after 11 a.m. that their presence in the order book stabilised, albeit at a level far below their usual market share in terms of market depth: they offered 40% less market depth than during normal periods and represented 65% of the order book compared with 75% during the reference period.

Non-HFTs withdrew to a much lesser degree: they offered 20% less market depth than during normal periods but represented 35% of the order book compared with 25% during the

reference period. A detailed analysis of trades by non-HFTs shows that they in particular supported liquidity by fully executing very large size buy orders between 1:15 p.m. and 2:15 p.m. In fact, studying the order book in periods when non-HFTs are very active reveals the presence of very large size best limit orders, whose quantities are often hidden and are executed quickly. These orders show that non-HFTs desire to trade supported liquidity during the session.

Impact on the spread In Figure I.22, we display the ratio for the spread r_S.

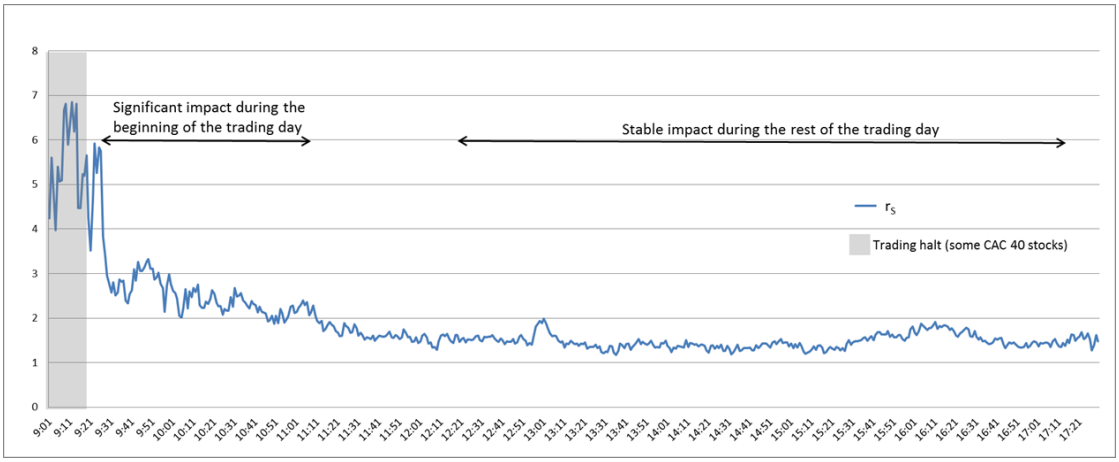


Figure I.22 – Comparison between the spread on June 24 and during the reference period: with respect to the reference period, the spread widened significantly at the beginning of the day, gradually decreased and then stabilised from 12 p.m., while remaining larger than during the reference period.

Figure I.22 shows that despite the important presence of non-HFTs in the order book at the beginning of the day, see Figure I.21, the spread was six times larger than in the weeks before the announcement. It gradually tightened, and stabilised at about 12 p.m. (in parallel with the stabilisation of the depth of HFTs in the order book) but remained 1.5 times wider than during the reference period.

6.2.4 Impact on amounts traded and aggressiveness

In Figure I.23, we consider the amounts traded on June 24 and during the reference period, the ratio of HFTs share in amounts traded and the HFTs aggressive/passive ratio.

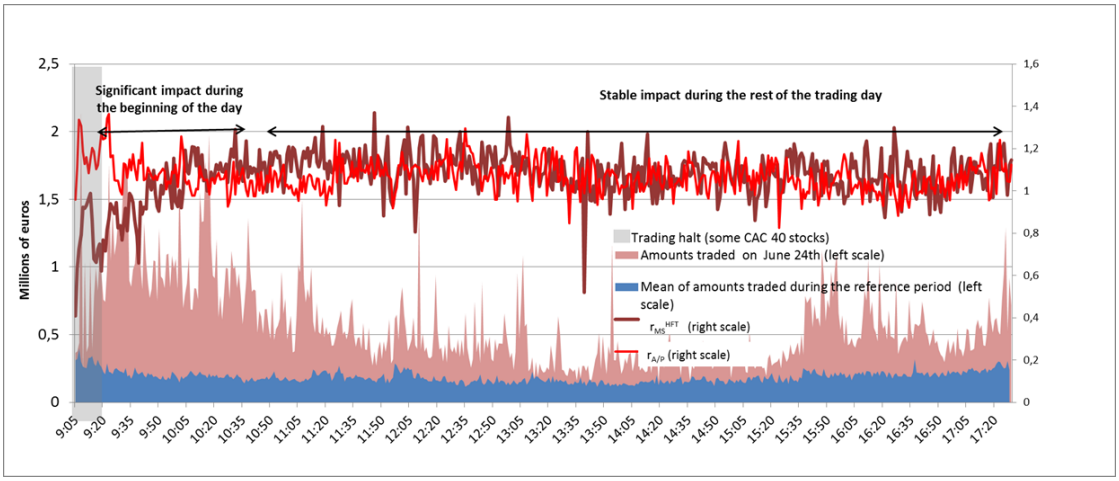


Figure I.23 – On June 24, amounts traded were higher than during the reference period and HFTs were somewhat more aggressive, particularly at the beginning of the day. HFTs share in amounts traded was also quite high, except at the beginning of the day.

Figure I.23 shows that throughout the day of June 24, the market was much more active in terms of amounts traded than during the reference period, despite low levels of order book liquidity, see Figure I.20.

At the very beginning of the day (until 9:40 a.m.), HFTs were less active on the market than during the reference period (for example at 9:20 a.m., their share in amounts traded was 20% lower than during the reference period). In fact, HFTs share in amounts traded with respect to the reference period fell because non-HFTs increased their trades more sharply than HFTs and since fewer of the HFTs limit orders were executed than those of non-HFTs (HFTs passive share in amounts traded decreased).

In addition, HFTs were more aggressive than during the reference period (for example at 9:20 a.m., their aggressive/passive ratio was 30% higher than during the reference period). Overall, they consumed more liquidity than they provided.

After 9:40 a.m., the share in amounts traded and aggressive/passive ratio of HFTs stabilised: HFTs were more active in the market (their share in amounts traded rose from 58% during the reference period to 63% on June 24), and they were slightly more aggressive.

Amounts traded fell gradually until 12 p.m.: they were five times higher than during the reference period at the beginning of the day compared with 2.5 times for the rest of the day.

6.2.5 Summary

The 24^{th} of June 2016 was characterised by a particularly sharp downturn in market quality: widening of spreads, drop in market depth at best limits and record-high amounts traded. During this time of considerable uncertainty, HFTs reduced very significantly their presence in the order book: their liquidity provision was 40% lower than usual. Market depth provided by non-HFTs at best limits supported liquidity, in particular at the beginning and in the middle of the session.

Being a little provocative, we could conclude here that when non-HFTs have time to adjust to announcements (they had the overnight in the case of the Brexit, in contrast to the situation on the 3^{rd} of December 2015), they are active as market makers. It is probably due to their capabilities in algorithmic trading, allowing them to use limit orders and market orders in a sophisticated way, see [80, Chapter 3].

7 Conclusion

The main results of this study conducted on labelled French market data from November 2015 to July 2016 are summarised in Table I.10. We established the following facts. High-frequency traders are important contributors to liquidity. They are present at the best bid or best ask prices more than 90% of time and they represent on average 80% of the market depth in the order book at the three best price limits. They constitute 60% of the total amount traded in the market. On average, they consume more liquidity than they provide with an aggressive/passive ratio equal to 53%. These relative figures do not really depend on the global level of market stresswith an exception for the HFTs market share which is slightly dependent on the market stress). Nevertheless, overall, liquidity decreases when market stress rises: HFTs withdraw their orders from the order book when implied volatility is high, the same way as non-HFTs do.

On an intraday seasonal basis, HFTs enter the order book gradually and contribute to tighten the bid-ask spread at the beginning of the day. However, they clearly reduce their presence in the order book ahead of announcements that are likely to impact prices. Moreover, their aggressiveness increases before the announcements (because they provide less liquidity) and after the announcements (because they consume more liquidity) and they trade less. On the contrary, when no announcement is planned, their reaction to volatility goes in the opposite direction (trading more). We thus identify a “change of regime" in HFTs activity in the presence of scheduled news.

Analysing two specific events (ECB announcements of the 3^{rd} of December 2015 and the Brexit), we see that when they have time to adapt, non-HFTs become the liquidity providers in place of HFTs (in the case of the Brexit, the overnight implemented a “pause” in the trading flows). But in case of a surprise (false pre-announcement in case of the 3^{rd} of December 2015), non-HFTs algorithms are not able to adjust and market participants rely on HFTs only

<table>
<tr><th></th><th colspan="6">Behaviour of HFTs</th></tr>
<tr><th></th><th rowspan="2">Average value</th><th rowspan="2">Increase in implied volatility (significant risk)</th><th rowspan="2">Increase in historical volatility (price move)</th><th colspan="2">Announcement with price change</th><th rowspan="2">Beginning of day</th><th rowspan="2">End of day</th></tr>
<tr><th>Pre-announce-ment (2 minutes)</th><th>Post-announce-ment (0-5 minutes)</th></tr>
<tr><td rowspan="2">Depth at three best limits: provision of liquidity</td><td rowspan="2">80%</td><td>Constant market share. Behaviour identical to non-HFTs.</td><td>Little impact on market share.</td><td rowspan="2">Specific withdrawal of HFTs (up to -40% at 2:30 p.m. on average and -80% for announce-ments having the highest impact).</td><td>Slight rise in or constant market share (rapid response & arbitrage opportunities).</td><td rowspan="2">Gradual increase in the quantity offered during the first hour and increase in market share.</td><td rowspan="2">Gradual withdrawal during the last 30 minutes and fall in market share.</td></tr>
<tr><td>Decrease in amounts offered.</td><td>Decrease in amounts offered.</td><td>Fall in quantity offered (80% on December 3rd), but gradual return to the pre-announcement level</td></tr>
<tr><td>Spread</td><td>Sharp impact on spread</td><td>Slightly tighter spread.</td><td>Spreads widen.</td><td colspan="2">Sharp widening of the spread in anticipation of announcements and during major price changes.</td><td>HFTs are actively involved in spread tightening</td><td>No impact (one non-HFT tightens the spreads while all HFTs withdraw their orders).</td></tr>
<tr><td rowspan="2">Amounts traded (excluding auctions)</td><td rowspan="2">60%</td><td rowspan="2">No material impact.</td><td>Increase in amounts traded.</td><td rowspan="2">No impact.</td><td rowspan="2">Increase in amounts traded (in terms of market share, up to 82% on December 3rd, and amounts).</td><td rowspan="2">Increase in market share, simultaneous with the gradual entry of HFTs in the order book. Amounts traded decrease.</td><td>Increase following U.S. market opening (more arbitrage opportunities).</td></tr>
<tr><td>Increase in market share (except at the beginning and end of the day).</td><td>Less activity at the very end of the day.</td></tr>
<tr><td>Aggressiveness (aggressive/passive ratio)</td><td>53%</td><td>No material impact.</td><td>Increase in aggressive/ passive ratio</td><td>Increase in HFTs' ratio after withdrawal of passive orders.</td><td>High degree of aggressiveness (withdrawal of passive orders & arbitrage opportunities).</td><td>After the entry of passive orders, the ratio decreases gradually to 52% at 10:00 a.m.</td><td>More aggressive behaviour due to the U.S. market opening and the increased arbitrage opportunities (52% to 55%) and at the very end of the day after the withdrawal of passive orders.</td></tr>
</table>

Table I.10 – Summary of the main findings of the study.

to make the market. However, HFTs do not seem to act as market makers in such situations: they protect themselves from potential adverse selection, and even tend to be more aggressive.

On the one hand, one could conclude that in case of a major event, more or longer trading halts could let time to non-HFTs to adjust their trading tactics, giving different instructions to their algorithms, or switching to different trading algorithms, see [80, Chapter 3] for

current electronic trading practices. However, on the other hand, one can suspect that future technological advances for trading algorithms (for instance taking the state of the liquidity or other market participants reactions into account, see [63, 32, 81]) will lead all market participants to adapt to the absence of HFTs market making-like activity.

Remark on rebates Note that the European directive MiFID II encourages rewards rather than sanctions through its act RTS8 on market making, and requires that trading venues provide additional incentives during periods of stress, see [46]. If it may be the case rebates compensate for adverse selection during "standard" market conditions, most of the results we observe in this paper show that it does not compensate for liquidity provision under market stress conditions. At least HFTs benefiting from these rebates do not provide liquidity as it could be expected under this reasoning.

I.A Raw OLS regressions of Section 5

Variables	Coef.	std. err.	t	$P>\|t\|$	95% Conf. Int.
Constant	0.7866	0.003	302.564	0	[0.781; 0.792]
σ_{norm}	−0.0031	0.002	−1.279	0.201	[-0.008; 0.002]
B	−0.0083	0.005	−1.551	0.121	[-0.019; 0.002]
D	−0.0053	0.008	−0.707	0.479	[-0.02; 0.009]
A	0.0045	0.005	0.837	0.403	[-0.006; 0.015]

Table I.11 – Regression of HFTs market share in terms of market depth during days without U.S. announcements at 4 p.m., with all the explanatory variables.

Variables	Coef.	std. err.	t	$P>\|t\|$	95% Conf. Int.
const	0.5340	0.002	228.198	0	[0.529, 0.539]
σ_{norm}	0.0111	0.002	5.023	0	[0.007, 0.015]
B	0.0023	0.005	0.47	0.638	[-0.007, 0.012]
D	0.0169	0.007	2.494	0.013	[0.004, 0.030]
A	−0.0021	0.005	−0.438	0.661	[-0.012, 0.007]

Table I.12 – Regression of the aggressive/passive ratio of HFTs during days without U.S. announcements at 4 p.m., with all the explanatory variables.

Variables	Coef.	Std. err.	t	$P>\|t\|$	95% Conf. Int.
Const.	0.0113	0.001	9.058	0	[0.009, 0.014]
σ_{norm}	−0.0055	0.001	−4.549	0	[-0.008, -0.003]
B	0.0184	0.003	7.142	0	[0.013, 0.023]
D	0.0270	0.004	7.272	0	[0.02, 0.034]
A	0.0024	0.003	0.931	0.352	[-0.003, 0.007]

Table I.13 – Regression of $res_{R_{A/P}}$ during days with U.S. announcements at 4 p.m., with all the explanatory variables.

Variables	Coef.	Std. err.	t	$P>\|t\|$	95% Conf. Int.
Const.	0.5555	0.003	207.599	0	[0.55, 0.561]
σ_{norm}	0.0476	0.003	18.768	0	[0.043, 0.053]
B	−0.006	0.006	−1.096	0.273	[-0.017, 0.005]
D	−0.0113	0.008	−1.457	0.145	[-0.027, 0.004]
A	0.0091	0.006	1.658	0.097	[-0.002, 0.02]

Table I.14 – Regression of HFTs share in amounts traded during days without U.S. announcements at 4 p.m., with all the explanatory variables.

Variables	Coef.	Std. err.	t	$P>\|t\|$	95% Conf. Int.
Const.	0.0088	0.002	5.320	0	[0.006, 0.012]
σ_{norm}	0.001	0.002	0.659	0.510	[-0.002, 0.004]
B	−0.0347	0.003	−10.242	0	[-0.041, -0.028]
D	−0.0476	0.005	−9.754	0	[-0.057, -0.038]
A	−0.0012	0.003	−0.358	0.72	[-0.008, 0.005]

Table I.15 – Regression of res_{MS} during days with U.S. announcements at 4 p.m., with all the explanatory variables.

CHAPTER II

The information content of high-frequency traders aggressive orders: recent evidences

Abstract

This empirical study uses a unique recent data set provided by the French regulator "Autorité des Marchés Financiers" and gives some evidence concerning the impact of aggressive orders on the price formation process and the information content of these orders according to the different order flow categories (high frequency traders, agency participants and proprietary participants). As expected, we find that the price impact of aggressive orders consuming exactly the quantity present at the best limit is higher than that of the ones consuming less than the quantity present at the best limit. Furthermore, the price impact is an increasing function with respect to the consumed share in percentage. We show that these price impact disparities are sustainable over time: both price impacts are permanent. On the contrary, the impact of orders consuming more than the quantity present at the best limit starts to diminish one second after the aggressive order. In contrast to previous literature, we find that the aggressive orders of HFTs are more informed than the ones of agency and proprietary members. This new finding may be an indicator of the evolution of high frequency traders activity over the years.

Keywords: High frequency trading, aggressive orders, price impact, price formation, asymmetric information, price profile, mean reversion, trend following, market microstructure.

1 Introduction

Since the emergence of High Frequency Traders (HFTs), a lot of academic research and regulatory discussions have investigated their behaviour and their impact on the markets. It is traditionnally considered that the main activity of HFTs is market making, to the extent that HFTs are described as the new market makers in [88]. Market makers are defined as market participants who provide liquidity to the market by posting simultaneously limit orders on both sides of the electronic Limit Order Book (LOB), see [29, 71]. This is why HFTs limit orders were widely studied in the literature, see for example Chapter I and [32, 35, 72, 81]. However, HFTs do send aggressive orders, even when they carry out market making strategies,

see for instance Chapter I and [71]. In the literature, it has been found that HFTs aggressive orders are not significantly more informed than those of other participants, see [17] where the authors use data going back to 2010, see also [29] for related results on date going back to 2008-2009. In this study, we investigate whether this conclusion remains true nowadays by analysing more recent data, from 2017. We furthermore aim at understanding how aggressive orders impact the price.

We first study the price impact of single aggressive orders. Note that this notion is different from the price impact widely studied in the literature: the non-linear price impact of meta-orders, that is in general proved to follow a power law, see for example [6, 93, 109]. We look at how the price impact of a single aggressive order varies according to the amount of liquidity consumed by aggressive orders. This is why we split aggressive orders in three groups: the ones consuming less than the liquidity available at the first limit (partial aggressive orders), the ones consuming exactly the liquidity available at the first limit (exact aggressive orders), and the ones consuming more than the liquidity available at the first limit (n-limit aggressive orders). We evaluate the price impact (in magnitude and durability) conditionally on these groups, by using the notion of price profile. The price profile is the evolution of the price in time around a specific market event. It is in general used to quantify the adverse selection like in [81] or to estimate the information content of orders like in [17]. We notice that the repartition of HFTs aggressive orders in these groups is not the same as the repartition of other order flow categories: they send a significant larger proportion of exact aggressive orders than the rest of the market. As expected, we find that just after the aggressive order, the impact of the exact ones is higher than that of the partial ones. For partial aggressive orders, we show that the price impact is increasing according to the consumed share in percentage: it depends on the traded volume and the quantity present at the best limit. Our first main new empirical finding is that these disparities in price impact are sustainable over time: both price impacts of partial and exact aggressive orders are permanent. On the contrary, the impact of n-limit aggressive orders attenuates gradually with time, starting one second after the aggressive order.

Second, we investigate whether HFTs have an informational advantage compared to the rest of the market. To assess their informational advantage, we compute the potential profits of the different order flow categories (HFTs, agency participants and proprietary participants) over different time horizons, and compare those of HFTs to the rest of the market. The potential profit is computed using the price profile: we look at the price variation after the aggressive order compared to the price obtained by the aggressive order. In contrast to [17], we find that HFTs are the most profitable agents. The divergence of these results may be an indicator of the evolution of HFTs activity in the market over time. Despite these different results, our third main finding is in line with [17]: HFTs typically buy after price decreases and sell after price increases, while agency members sell after price increases and buy after price decreases. Additionally, we find that partial aggressive orders are more discriminating than exact ones in terms of potential profit disparities between the different order flow categories. Furthermore, we show how dissociating flows of a given market member according to their

connectivity channels allows us to exhibit different significant "sub-behaviours".

Third, we display how aggressive orders can be used in order to classify market participants as HFTs or non-HFTs. This is of particular interest from a regulatory viewpoint. We document other statistical features, such as the autocorrelation of aggressive orders or specific price patterns related to the behaviour of certain market participants.

This paper is organised as follows. In Section 2, we present our data, describe the different order flow categories and explain our methodology to identify HFTs. We introduce in Section 3 the notion of price profile used to quantify price impacts and potential profits. In Section 4, we distinguish between three different aggressive order groups, and display the price impact of each of them. We then focus on analysing one specific group: the partial aggressive orders according to the consumed share. We measure in Section 5 the potential profit of each order flow categories. In Section 6, we shed light on some statistical features and market participants classification criteria. We even show that we can achieve a more granular and relevant classification using connectivity channels. Finally, Section 7 summarises our results.

2 Data description and HFTs identification

2.1 Data description

We recall that our data are provided by the French regulator "Autorité des Marchés Financiers"(AMF). This analysis is conducted on the CAC 40 stocks traded on Euronext Paris over a three-month period: from September 2017 to November 2017, during which the volatility on the CAC 40 was stable and reached historically low levels (see Figure II.1). Furthermore, this studied period is neither disrupted by end of year trading effects nor by MIFID II [1].

In Figure II.1, we plot the Vstoxx from 2013 till 2017. The Vstoxx is the "European VIX". It measures implied volatility of near term EuroStoxx 50 options, which are traded on the Eurex exchange.

[1] MIFID II is a legislative framework instituted by the European Union to regulate financial markets, that entered into force in January 2018.

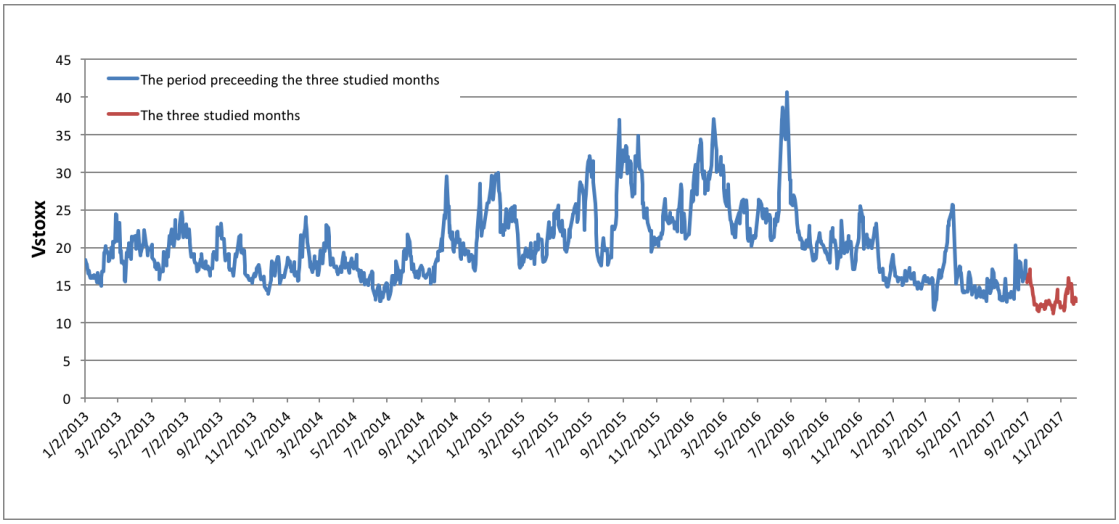


Figure II.1 – The implied volatility (Vstoxx) during the studied period varies little. The three months under study show the lowest implied volatility since 2013.

This study focuses solely on the analysis of strategies on Euronext, and does not consider other exchange platforms. Over the analysed period, we use both trade data and LOB data to describe the dynamics of the LOB accurately before and after each aggressive order. The whole data set contains approximately 8 millions aggressive orders and 423 millions events (an event can be an order insertion, an order cancellation, an order modification or a transaction). Note that we do not use market data corresponding to the initial and final twenty minutes of the trading session, as these periods usually have specific features due to the opening/closing auction phases.

2.2 The different order flow categories

The data we have gives us access to the order flow category to which each aggressive order belongs. We have four different categories of order flows:

- Agency flows: it corresponds to aggressive orders triggered by market participants acting for the account of their clients.
- Proprietary flows: it corresponds to aggressive orders triggered by market participants acting for their own account.
- Supplemental Liquidity Provider (SLP) flows: it corresponds to aggressive orders declared as part of the SLP programme to which the market participant initiating the order must belong to. The SLP programme imposes a market making activity on programme members, including order book presence time at competitive prices. In return, they get favourable pricing and rebates in the form of a maker-taker fees model directly comparable to those of the major competing platforms, see [44].

- Retail Member Organization (RMO) flows: RMO members are part of a programme offered by Euronext: the Retail Matching Facility, specialised in providing liquidity for retail participants through Retail Liquidity Provider (RLP) members. The role of RLP members is to provide liquidity to RMOs by posting buy and sell limit orders. The RMO members are eligible to trade with all market participants, while RLPs can trade only against RMO orders. In the following, the RLP are not considered since they almost never send aggressive orders.

Note that an institution can have one single or multiple member codes to access the market. In addition, using the same member code, an institution can send orders belonging to different order flow categories. For instance, a same institution with the same member code can have at the same time agency and proprietary activities. The data we have provide us with the name of the institution and the member code issuing each order. Additionally, each order is labelled by the order flow category to which it belongs (agency, proprietary, SLP or RMO).

2.3 HFTs identification

High frequency trading is a subset of algorithmic trading (MIFID II states that algorithmic trading means trading in financial instruments where a computer algorithm automatically determines individual parameters of orders such as whether to initiate the order, the timing, price or quantity of the order or how to manage the order after its submission, with limited or no human intervention) for which minimising latency is a crucial element for performance. HFTs use co-location and proximity services to minimise latency. Most of them submit large numbers of orders that are cancelled relatively shortly after submission, trade large volumes, consistently maintain a low inventory level by holding positions for very short time and turning them over rapidly, see for example Chapter I and [29].

In this work, an order is identified as belonging to the high frequency traders order flow category if it is labelled as SLP. Indeed, HFTs are now essentially the only market participants that are able to play the role of market makers on liquid stocks, see [29, 71]. This is because they are supposingly able to maintain a strong presence at best price limits and operate efficient inventory management in an increasingly fast-moving and fragmented market. Indeed, HFTs can use speed to enhance risk control by avoiding adverse selection, see [72], improving inventory management, see [5] and trading on short-lived information, see [50]. Moreover, according to the classification based on the lifetime of cancelled orders (described in details in Section 6.2) which constitutes one of the criteria used by AMF to identify HFTs, all SLP members are either classified as HFTs or mixed HFTs (investment banks with high frequency trading activity). The analysis of aggressive orders presented in Section 5.3 is thereafter generalised in Appendix II.A to all the HFTs based on this classification, and we show that the results are similar to those obtained when considering only SLPs.

3 Quantification of the price impact and the informational advantage

Let us consider a buy (resp. sell) aggressive order occurring at time t. To estimate its information content at time $t+h$, we compute the potential profit denoted by PP_{t+h} that a market participant can realise if he succeeds to unwind his position passively:

$$PP_{t+h} = \frac{BP_{t+h} - P_t}{S} * sign_t,$$

where BP_{t+h} is the best ask (resp. bid) at time $t+h$, P_t the price per share obtained by the aggressive order, $sign_t$ takes the value 1 (resp. -1) if it is a buy (resp. sell) aggressive order. The quantity S is the average spread of the asset. It is computed averaging among all the events in the data set happening between 9:20 and 17:10, and weighted by time. Finally, h varies between -17 minutes to 17 minutes. Note that when computed at time horizons before the aggressive order, this measure does not reflect the potential profit. However, it will help us to understand the strategy followed by the market participants (mean reverting or trend following), see Section 6.4. On the other hand, it will also allow us to determine whether the aggressive order has been sent at a relevant time or not. At a given point in time before the aggressive order, a negative value indicates that the participant could have obtained a better price (at least for one security). On the contrary, a positive value indicates that the participant has intervened at a convenient moment: if the aggressive order had taken place earlier, the price would have been higher.

We will also need to measure the price impact of an individual aggressive buy (resp. sell) order taking place at time t, and evaluated at time $t+h$, denoted by PI_{t+h}, and defined as follows:

$$PI_{t+h} = \frac{BP_{t+h} - BP_{t^-}}{S} * sign_t,$$

where BP_{t^-} denotes the best ask (resp. bid) one microsecond before the buy (resp. sell) aggressive order.

Note that the price impact and the potential profit coincide for aggressive orders consuming a quantity less or equal to that present at the best limit. For orders consuming a quantity larger than that present at the best limit, to quantify the price impact at time $t+h$, one can compute the difference between the potential profit at time $t+h$ and the potential profit at one microsecond before the aggressive order. This is why in the following, we focus on the potential profit measure.

Later on, orders will be merged according to the categories defined in Section 2.2. To measure the average potential profit at time $t+h$ of a given order flow A, denoted by APP^A_{t+h}, we take the average among all aggressive orders belonging to the order flow A, weighted by the quantity of each aggressive order:

$$APP^{A}_{t+h} = \frac{\sum_{i \in A} PP^{i}_{t+h} Q^{i}}{\sum_{i \in A} Q^{i}}, \tag{1}$$

where PP^{i}_{t+h} is the potential profit of the i^{th} aggressive order, and Q^{i} the quantity traded by the i^{th} aggressive order.

In our analysis, when providing average results, we merge buy and sell aggressive orders. This is because the evolution of the price at the best limit following a buy aggressive order is quite symmetric to that following a sell aggressive order.

4 Analysis of aggressive orders with respect to consumed share

We distinguish between three different groups of aggressive orders, and we show how the price impact varies according to each group. Furthermore, we emphasize that each group of aggressive orders usually takes place in specific LOB configuration. In addition to this, we obtain and interpret a relationship between the price profile 17 minutes before the aggressive order, the quantity present at the best limit just before the aggressive order, and the price impact following the aggressive order.

4.1 Three different groups of aggressive orders

We distinguish between three groups of aggressive orders:

- Partial aggressive orders: they consume less than the quantity at the best limit.
- Exact aggressive orders: they consume exactly the quantity at the best limit.
- N-limit aggressive orders: they consume more than the quantity at the best limit.

Now that we have defined these groups of aggressive orders, note that exact and n-limit aggressive orders mechanically change the price since they trigger a best price change right after the trade. This is why the price impact one microsecond after these aggressive orders is obviously significant. We aim at investigating whether these mechanical impacts are temporary or reflect a certain information content that persists over time.

4.2 Some preliminary statistics

In our database, partial and exact aggressive orders constitute the majority of aggressive orders (96%). Furthermore, HFTs send more exact aggressive orders than partial ones: 63% of the exact aggressive orders are sent by HFTs, while only 39% of partial ones are sent by them (see Table II.1).

Order flow categories	Percentage of partial aggressive orders	Percentage of exact aggressive orders
Agency	27%	16%
HFT	39%	63%
Proprietary	31%	21%
RMO	3%	0%

Table II.1 – Distribution of partial and exact aggressive orders across the different order flow categories.

On average, single partial aggressive orders consume a volume (11 k €) almost equal to that consumed by exact aggressive orders (13 k €). N-limit aggressive orders consume an amount clearly more significant than the other aggressive orders (43 k €). It is important to point out that aggressive orders, and especially exact (resp. n-limit) aggressive orders occur upon particular conditions: when the quantity at the best limit is significantly less than the average quantity at this limit over all events: 13 k € (resp. 16 k €) just before the exact (resp. n-limit) aggressive orders, versus 57 k € on average (see Table II.2).

	Average traded amount per aggressive order	Median traded amount per aggressive order	Share of aggressive orders number	Share of traded amount	Amount at the best limit just before the aggressive order
Partial aggressive orders	11 k €	6 k €	49.5%	38%	41 k €
Exact aggressive orders	13 k €	8 k €	46.5%	48%	13 k €
N-limit aggressive orders	43 k €	22 k €	4%	14%	16 k €

Table II.2 – General statistics on the different groups of aggressive orders.

This is not really surprising, and is related to the well known information content of the order book imbalance: when the quantity on one side of the book is significantly larger than that on the other side, the next aggressive order will likely hit the smallest side, see for example [107]. We now investigate more precisely the relationship between the imbalance and the aggressive order group.

4.3 Relationship between imbalance and aggressive order group

We show in this section that market participants submit exact aggressive orders when the LOB is significantly imbalanced. The imbalance at time t, just before the aggressive buy (resp. sell) order takes place, is computed as follows:

$$Imb_t = \frac{Q_t^1 - Q_t^2}{Q_t^1 + Q_t^2},$$

where Q_t^1 denotes the quantity present at the best bid (resp. ask) at time t, and Q_t^2 denotes the quantity present at the best ask (resp. bid) at time t when it is a buy (resp. sell) aggressive order.

The value of the imbalance one microsecond before the exact aggressive trades (on average equal to 27%) is significantly higher than that (on average equal to 3%) before the partial aggressive trades (see Figure II.2).

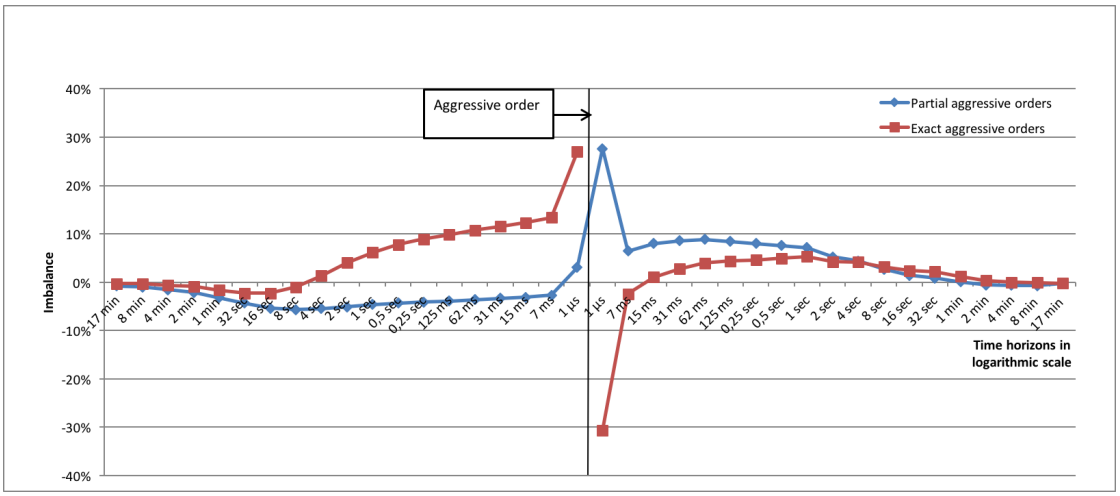


Figure II.2 – Variation of the imbalance of the LOB before and after the arrival of partial and exact aggressive orders.

4.4 Price impact according to the groups of aggressive orders

As expected, one microsecond after the aggressive order, because of the mechanical impact, the price impact due to n-limit aggressive orders is higher than that of exact ones, which is higher than that of partial ones (see the price profiles in Figure II.3, from which price impacts are obviously deduced).

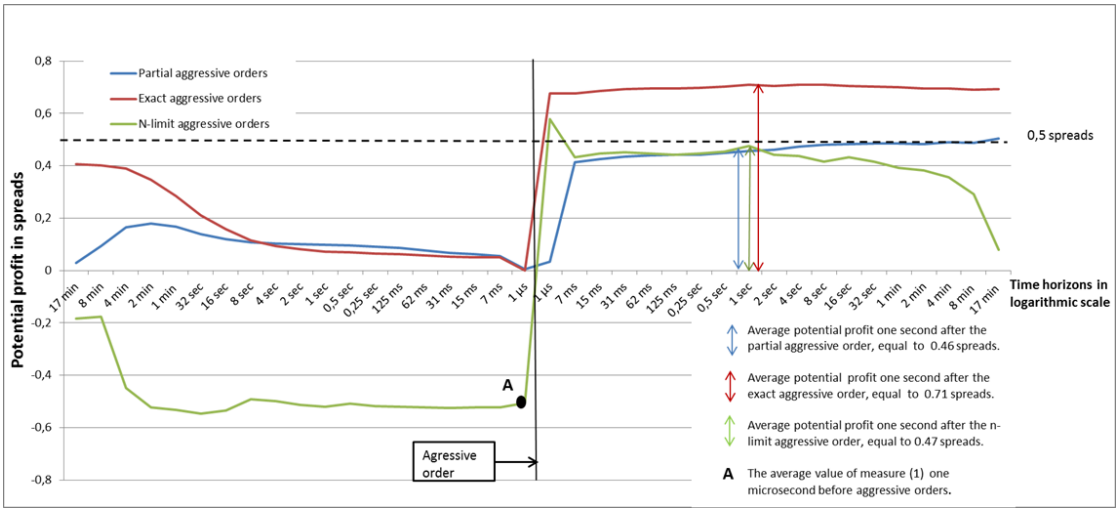


Figure II.3 – The price profile measure (1) according to the aggressive order groups.

One relevant question is whether this mechanical impact is permanent or not. Figure II.3 shows that the price impact of exact aggressive orders is permanent: it is above that of partial ones, over all time horizons, higher to two-thirds of the bid-ask spread. On the contrary, n-limit aggressive orders have a temporary component in their price impact: market participants tend to refill the LOB by submitting new orders in place of the consumed ones. Indeed, starting one second after the aggressive order, the price impact begins to attenuate. On a 17 minutes time horizon, the remaining mechanical impact of n-limit aggressive orders is quite equal to that of exact aggressive orders (recall that Figure II.3 displays the price profiles, and that the price impact of n-limit aggressive orders is deduced as the difference between the profile at time t and Point A).

4.5 Focus on partial and exact aggressive orders

We deepen our analysis by investigating the price profile of partial (consuming less than 100% of the best limit) and exact (consuming 100% of the best limit) aggressive orders according to the consumed share at the best limit in percentage. In general, the price impact is studied according to the traded volume. In this work, we choose on purpose to study it according to the consumed share in order to show that the price impact does not only depend on the traded volume but also on the quantity present at the best limit. In particular, we want to understand the relationship between historical prices evolution, consumed share and price impact.

In Table II.3, we show the proportion of aggressive orders according to the consumed share at the best limit (n-limit aggressive orders excluded).

Consumed share with respect to the total quantity at the best limit.	Proportion	Cumulated proportion
0% - 10%	10%	10%
11% - 20%	8%	18%
21% - 30%	7%	25%
31% - 40%	5%	30%
41% - 50%	5%	35%
51% - 60%	4%	39%
61% - 70%	4%	43%
71% - 80%	3%	46%
81% - 90%	3%	49%
91% - 99%	3%	52%
100%	48%	100%

Table II.3 – Distribution of partial and exact aggressive orders according to the consumed share at the best limit.

In Figure II.4, we plot the price impacts according to the consumed share.

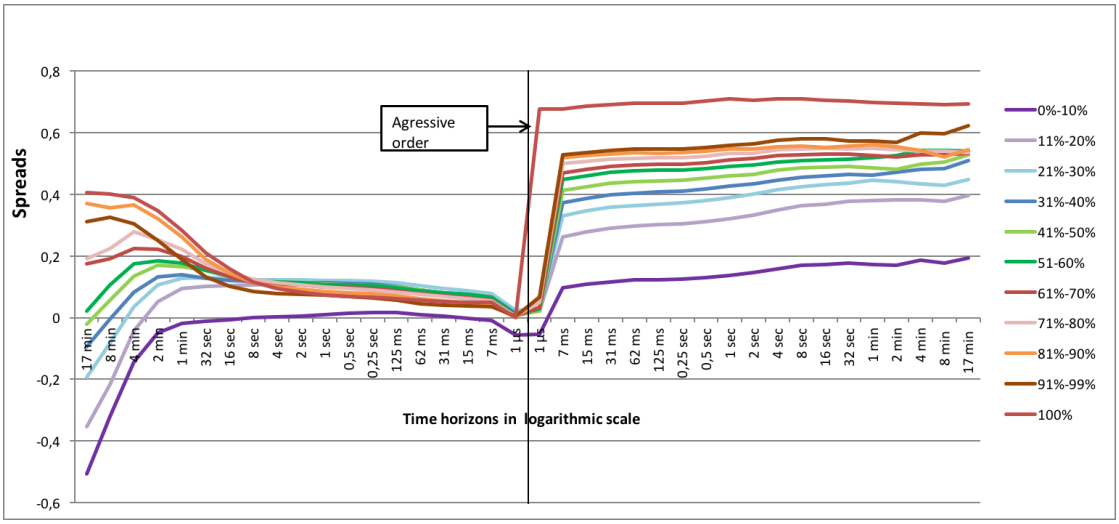


Figure II.4 – The price impact according to the consumed share at the best limit.

Figure II.4 shows that the magnitude of the price impact over all time horizons after the aggressive order is increasing with respect to the consumed share. As an example, the price impact following an aggressive order consuming 10% of the total quantity present at the best limit is significantly lower than that due to an aggressive order consuming 90% of the quantity present at the best limit. This could be interpreted by the fact that the imbalance created following an aggressive order consuming 90% of the quantity at the best limit is higher than

the one following an aggressive order consuming only 10%. As already seen, a large imbalance is likely to trigger other aggressive orders (or cancellations of limit orders).

We now investigate whether the consumed share rather depends on the quantity present at the best limit or the traded amount. Figure II.5 shows that as expected, the consumed part varies with the traded amount, but also depends significantly on the quantity present at the best limit. This means that the price impact does not depend only on the traded volume, but on the volume present at the best limit too. This is in line with [47] where the authors show that price fluctuations caused by individual market orders are essentially independent of the volume of orders and are essentially driven by liquidity fluctuations.

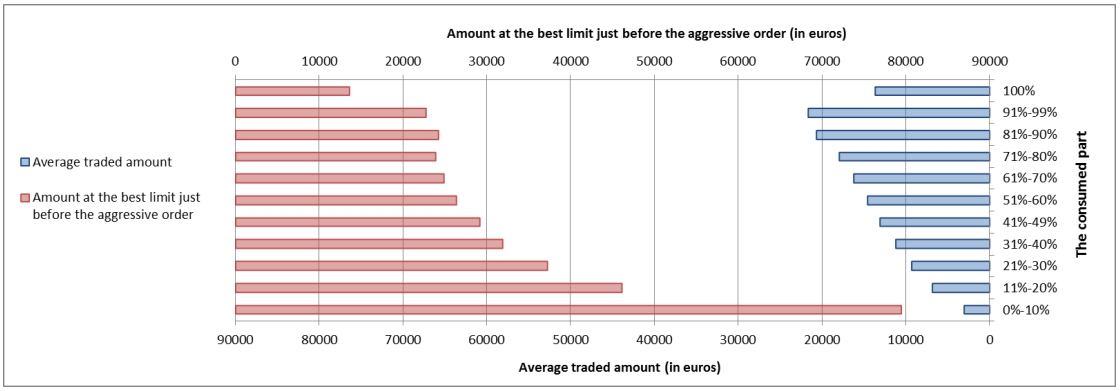


Figure II.5 – Evolution of the amount at the best limit just before the aggressive orders and the average traded amount according to the consumed part.

Another important feature appearing in Figure II.4 is that the consumed share is increasing with respect to the price profile measure (1) evaluated 17 minutes before the aggressive order. To clarify this phenomenon, we plot in Figure II.6 the measure (1) evaluated 17 minutes before the aggressive order with respect to the consumed part at the best limit.

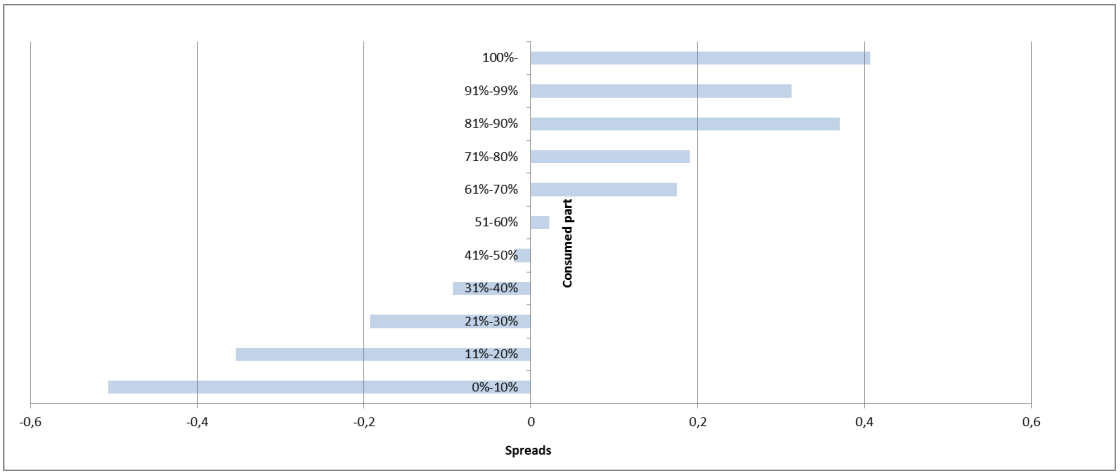


Figure II.6 – Evolution of the consumed part at the best limit according to the measure (1) evaluated 17 minutes before the aggressive order.

As previously showed, the consumed share depends on the quantity present at the best limit just before the aggressive order, which depends on the historical evolution of the price. For instance, if the price has been decreasing, the quantity present at the best ask tends do be small, since new limits are revealed. Following this logic, the quantity present at the best limit is a decreasing function with respect to the price profile 17 minutes before the aggressive order, which explains the relationship appearing in Figure II.6.

5 Potential profits according to the different order flow categories

We study in this section the potential profits according to each order flow category after partial and exact aggressive orders. We then focus on partial aggressive orders, in order to identify the potential profit disparities between market participants within the same order flow category. Finally, for a given market participant having activities belonging to different order flow categories, we check the differences in potential profit according to these different categories.

5.1 Potential profits after partial aggressive orders

The HFT flow stands out with the (significantly) highest potential profit in the case of partial aggressive orders, over all time horizons. One second after partial aggressive orders, HFTs have a potential profit 0.36 spreads higher than agency participants, and 0.29 spreads higher than proprietary participants (see Figure II.7). RMO members are the least profitable. In addition to this, we show in Section 6.1 that the aggressive orders of HFTs are the less autocorrelated, which allows us to deduce that the high potential profit of HFTs is due to an

informational advantage and not to autocorrelated orders.

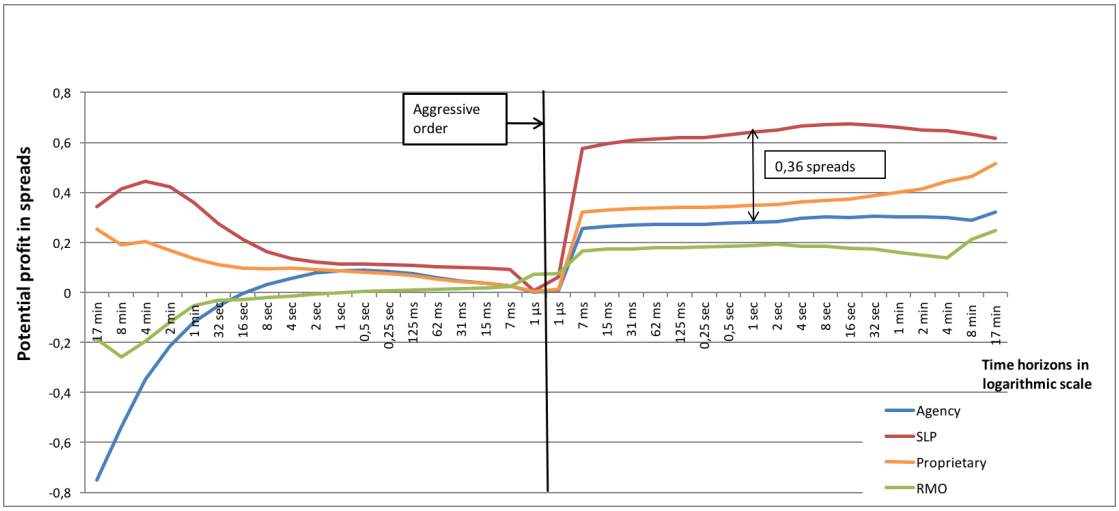


Figure II.7 – The potential profit evolution following partial aggressive orders according to the different order flow categories over different time horizons.

5.2 Potential profits after exact aggressive orders

Although HFTs still obtain a better potential profit than other market participants in the case of exact aggressive orders (see Figure II.8), the difference between the categories is not much significant: the potential profit of HFTs is only 0.04 spreads higher than that of agency members and 0.1 spreads higher than that of proprietary members (see Figure II.8).

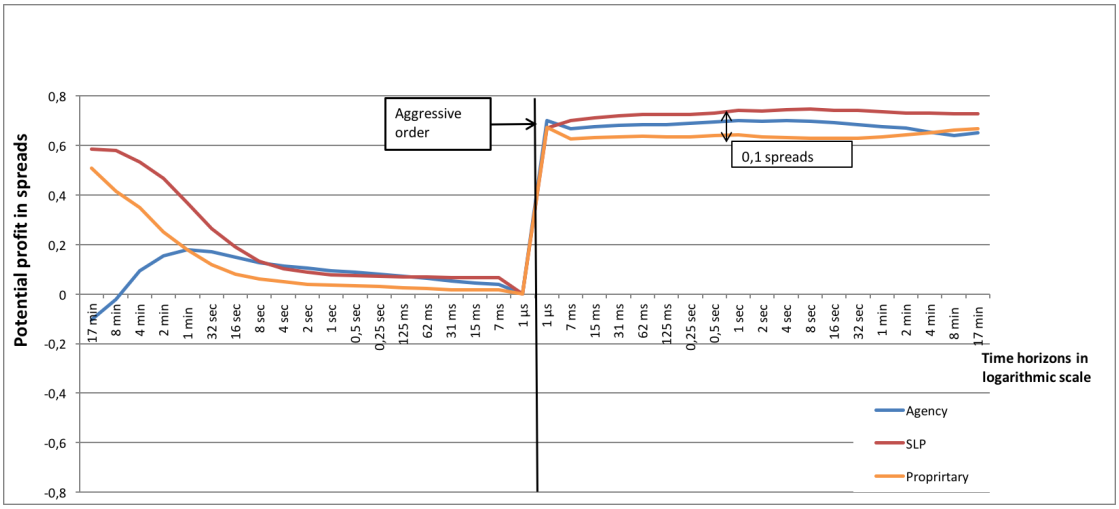


Figure II.8 – The potential profit evolution following exact aggressive orders according to the different order flow categories over different time horizons.

Furthermore, Figures II.7 and II.8 show that on average, HFTs and proprietary members are mean reverting: by this we mean that they buy when the price decreases and sell when the price increases. In contrast, agency members seem on average trend following: they buy when the price increases and sell when the price decreases. These results are in line with the finding in [17].

5.3 Does the potential profit vary among members within the same order flow category?

We now investigate the potential profit disparities between different members belonging to the same order flow category for partial aggressive orders.

Table II.1 below shows the total number of member codes according to each order flow category and the number of member codes issuing enough[2] partial aggressive orders.

[2]We consider that the number of aggressive orders is enough when there is at least one aggressive order per day and per asset.

Order flow category	Number of member codes	Number of member codes issuing enough partial aggressive orders
Agency	74	33
SLP	17	11
Proprietary	72	24
RMO	23	6
RLP	4	2

Table II.4 – Number of member codes issuing enough partial aggressive orders according to each order flow category.

We compute the proportion of member codes with a potential profit higher than the third quartile (based on the potential profits of member codes having enough partial aggressive orders) for each order flow category over different time horizons following the partial aggressive order. These proportions are plotted in Figure II.9.

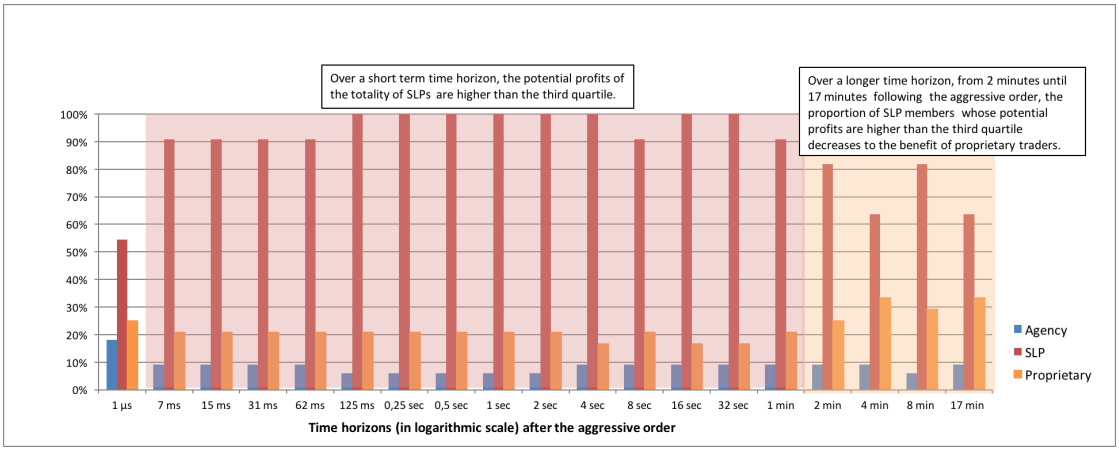


Figure II.9 – Evolution of member codes proportion with a potential profit higher than the third quartile for each order flow category over different time horizons following the partial aggressive order.

In order to understand the results in Figure II.9, we take the following example: the value relative to HFT activity 7 milliseconds after the aggressive order is equal to 90%. This means that 90% of the HFTs member codes have a potential profit higher than the third quartile. Now we can interpret the rest of the results. We find that over a short time horizon (until 2 minutes after the aggressive order), HFTs belong to the 25% market participants realising the highest short-term potential profits. Over a longer time horizon, from two minutes after the aggressive order, the proportion of HFTs with potential profit higher than the third quartile starts to decrease to the benefit of proprietary traders (see Figure II.9). This could be due to the fact that HFTs do not target long-term strategies, high frequency trading being an activity where participants typically hold positions for very short times.

5.4 Disparities in potential profits for a same member code according to the different order flow categories

We now dissociate the flows of a same member code according to the order flow categories. It allows us to identify the different potential profits generated by a member. We can notably distinguish between high frequency trading strategies targeting short-term potential profits and longer term strategies.

For members carrying out simultaneously SLP and another activity, the potential profit of the SLP flow is always higher than that of the other flows. In the majority of cases, the potential profit of the proprietary flow is higher than or equal to that of the agency flow. We illustrate these findings through two examples in Figures II.10 and II.11.

In Figure II.10, we plot the different potential profits of Member code A (having SLP and proprietary activities at the same time) according to the order flow categories.

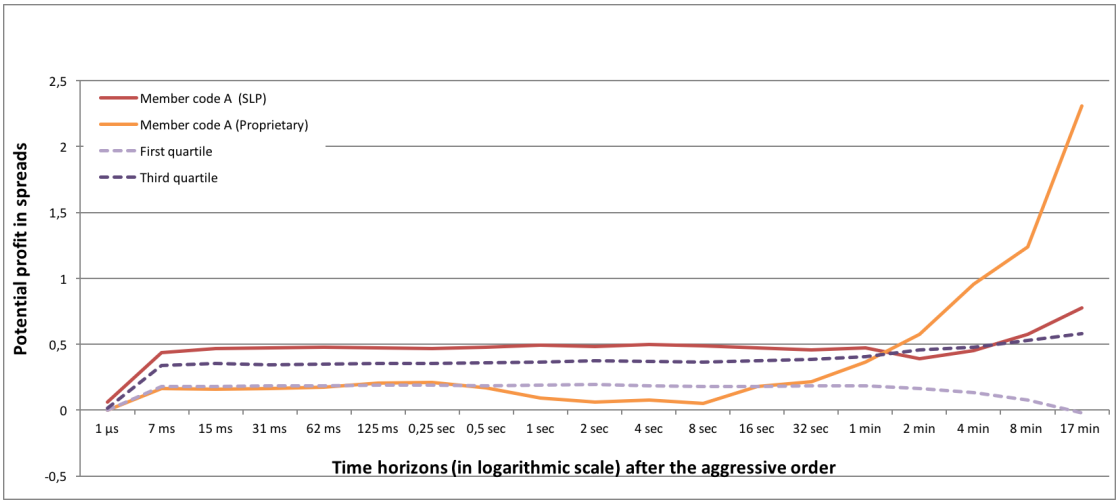


Figure II.10 – Disparities in potential profits of Member code A according to the different order flow categories issued by this member code. The quartiles are the same as those computed in Section 5.3.

Until 32 seconds after the aggressive order, the SLP flow has a potential profit higher than the third quartile, while the potential profit of the proprietary flow is equal or lower than the first quartile. The proprietary activity of Member code A seems to target a longer term strategy: 2 minutes after the aggressive order, its potential profit becomes higher, outperforming the one of SLP. On a 17 minute horizon, it is equal to 2.3 spreads, 3 times higher than the SLP flow potential profit.

In Figure II.11, we plot the different potential profits of Member code B, according to the order flow categories.

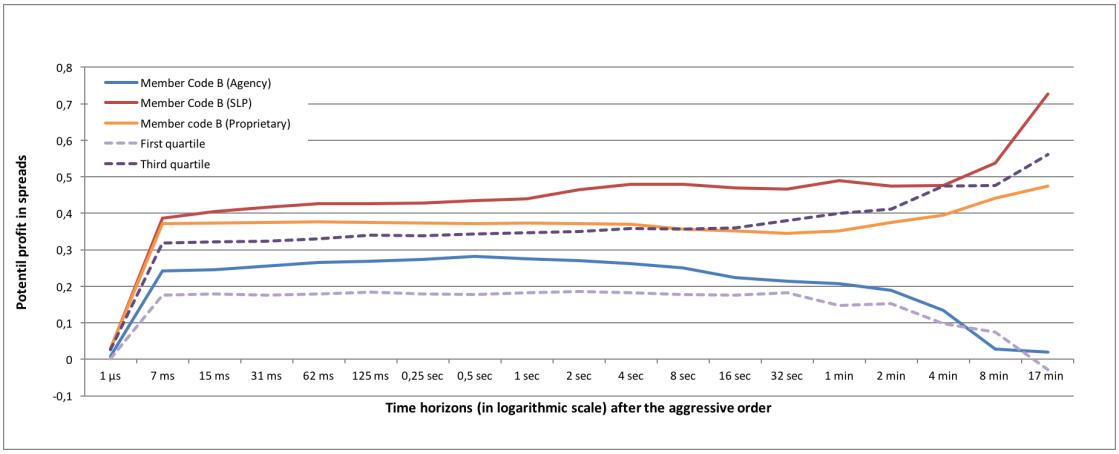


Figure II.11 – Disparities in potential profits of Member code B according to the different order flow categories issued by this member code.

Member code B has different potential profit levels, depending on the order flow category considered. The potential profit of the SLP flow is the highest at any time scale: one second after the aggressive order, the potential profit of the SLP flow (0.43 spreads) is higher than that of the proprietary one (0.37 spreads) and significantly higher than the agency flow (0.27 spreads).

6 From single aggressive orders to strategies

In this section, we show that the analysis of aggressive orders is useful to understand other features than price impact and potential profit. For instance, we study the autocorrelation of the different order flow categories. In addition to this, we propose a new classification of member codes (whose flows are segmented according to the order flow category) based on the investigation of aggressive orders. We also show that we can access to a more granular classification by segmenting member code flows according to the different connectivity channels they use. Finally, by observing the evolution of the price before the aggressive order takes place, we deduce the different strategies of member codes, such as mean reverting or trend following.

6.1 Autocorrelation of aggressive orders according to the different order flow categories

We study the autocorrelation properties of each order flow category, taken seperately. Positive autocorrelation means that once an aggressive order has been observed, the probability to observe another one in the same direction is larger than one half. Typically if an agency broker is splitting a large client's metaorder in slices, its flow exhibits a significant positive

autocorrelation, see [23, 110].

Figure II.12 shows that the autocorrelation of the first two successive aggressive orders is almost equal among all order flow categories (equal respectively to 29%, 30% and 34% for agency, proprietary and SLP order flow categories). The autocorrelation of the aggressive orders of the agency order flow category is the highest among all order flow categories (with the exception of the first two aggressive orders). The decrease in autocorrelation is the most significant for the SLP order flow category. The aggressive orders autocorrelation in this flow almost vanishes from the 4^{th} aggressive order (it is equal to 5% in this case).

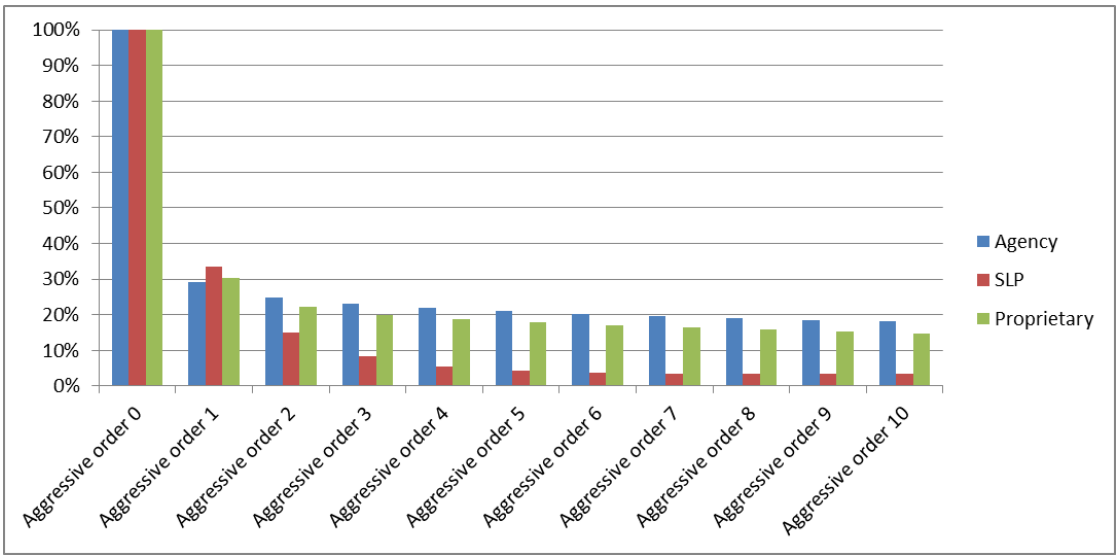


Figure II.12 – Autocorrelogram of aggressive orders according to each order flow category.

6.2 A classification tool

It is usual to consider passive orders to classify members as HFTs or non-HFTs. One of AMF classifications is based on this type of orders.

This AMF classification differentiates between three classes of market participants: HFTs, mixed HFTs and non-HFTs. It is based on the lifetime of cancelled orders and determined using two sets of conditions:

- **Condition 1 is based on a comparison with other participants**: the participant must have cancelled at least 100,000 orders during the year, and the average lifetime of his cancelled orders should be less than the average lifetime of all cancelled orders in the book.
- **Condition 2 is based on a set threshold**: the participant must have cancelled at least 500,000 orders with a lifetime of less than 0.1 second (i.e. the participant quickly

updates the orders in the limit order book) and the top percentile of the lifetime of its cancelled orders must be less than 500 microseconds (i.e. the participant regularly uses fast access to the market).

A member code is a high frequency trader if it is not an investment bank and it meets one of these conditions. An investment bank meeting one of these conditions is described as mixed HFT. Note that some members satisfy Condition 2 without satisfying Condition 1.

However, it seems also possible to classify participants by relying on aggressive order potential profits. Those realising the higher short-term potential profits (one second after the aggressive order) can be considered as HFTs, and those realising the lowest as non-HFTs. It turns out that relying on both approaches allows us to obtain a more complete classification of market participants. Three different classes can be distinguished for member codes whose flows are segmented according to the order flow category:

- Pure HFTs are characterised by a high short-term potential profit (higher than the third quartile of potential profits computed among all market participants) and a low lifetime of cancelled orders (lower than the third quartile of lifetimes of cancelled orders computed among all market participants).
- Pure non-HFTs are characterised by a high lifetime of cancelled orders and a small short-term potential profit.
- Intermediary agents are characterised by a small short-term potential profit and a low lifetime of cancelled orders.

We point out that, as expected, no member code has high short-term potential profits and high lifetime of cancelled orders (see Figure II.13). Moreover, note that all SLPs belong the pure HFT category.

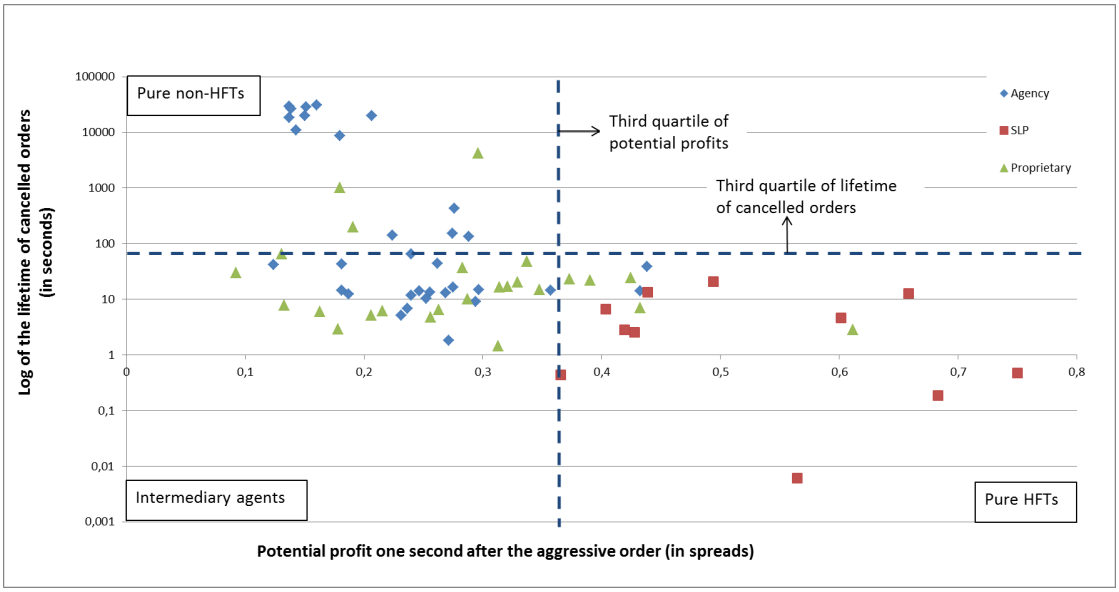


Figure II.13 – Using both methods of classification: one relying on cancellations and one relying on potential profits (the flows are grouped by member code and order flow category).

6.3 A more granular classification using the different connectivity channels

Market members connect to Euronext via connectivity channels (called "SLE", a French acronym for "Serveur Local d'Emmission", since this technology was initially developed by Canadian and French companies)[3] to convey their orders. Some member codes use their SLEs to separate their aggressive flow from their passive one, and some others to separate their different order flow categories. Dissociating the flows issued by a same member code, and belonging to the same order flow category according to SLEs can in some cases bring up new information concerning the different activities followed by this member code. The main advantage of using different SLEs is that it is easy to setup specific sets of risk limits for each of them (like the number of orders per day or per hour, the maximum traded value, etc). For instance, we dissociate the flow of Member code B who is an agency broker serving as an intermediary for a HFT (among other clients) according to the different SLEs. We plot in Figure II.14 the potential profit of each of these flows.

[3] On the three studied months and over the CAC 40 stocks, there are in total 355 SLEs. 94% of SLEs are used by only one member code, 6% are used by two member codes (belonging to the same institution). 73% of SLEs are deployed for only one order flow category (Agency, Proprietary, HFT, RMO), 26% are deployed for two different order flow categories and 1% for three different order flow categories. The number of SLEs belonging to the same institution varies between 1 and 33. There are 46 member codes using more than one SLE over 117 member codes in total.

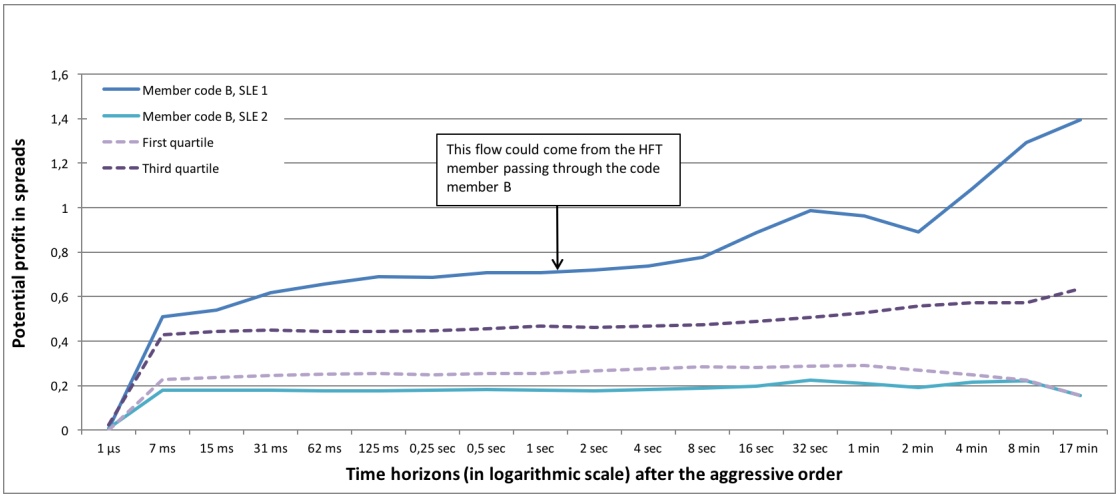


Figure II.14 – Disparities in potential profits of the same agency broker according to its different SLEs.

The difference in potential profits between the two flows, illustrated in Figure II.14 could be interpreted for example by a segmentation of these different flows allowing us to identify according to the clients' typology: one SLE is dedicated for a HFT client while another is dedicated for other type of clients.

6.4 Different strategies

By observing what happens before the aggressive order, we can distinguish between three different strategies:

- Mean reverting strategy, going against the price variations (see Figure II.15).
- Trend following strategy, following the price variations (see Figure II.16).
- Another strategy consisting in benefiting from the insertion of new orders that reduce the spread (see Figure II.17).

In Figures II.15 and II.16, we pick respectively some mean reverting and trend following SLPs and we plot their price profiles (using Measure (1)) around their aggressive orders.

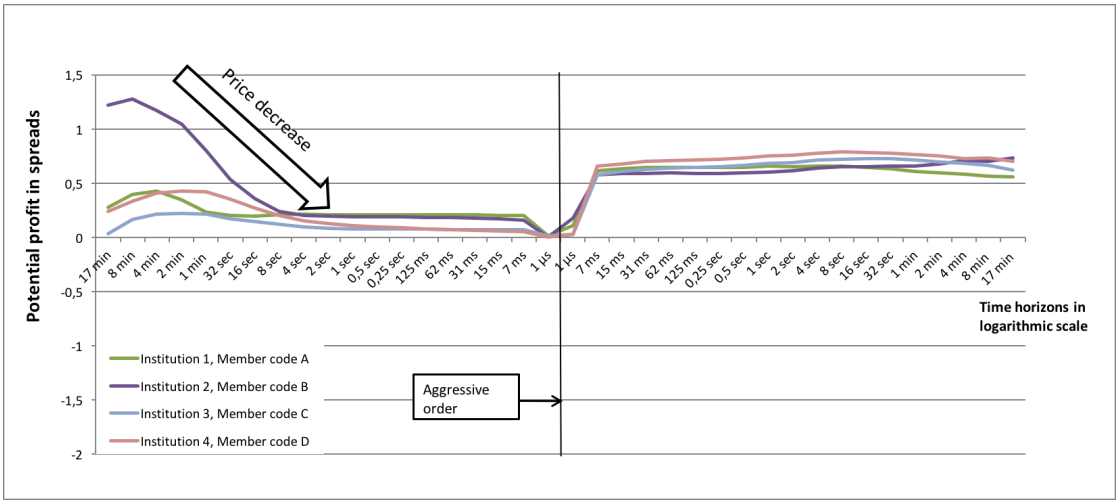


Figure II.15 – Mean reverting HFTs.

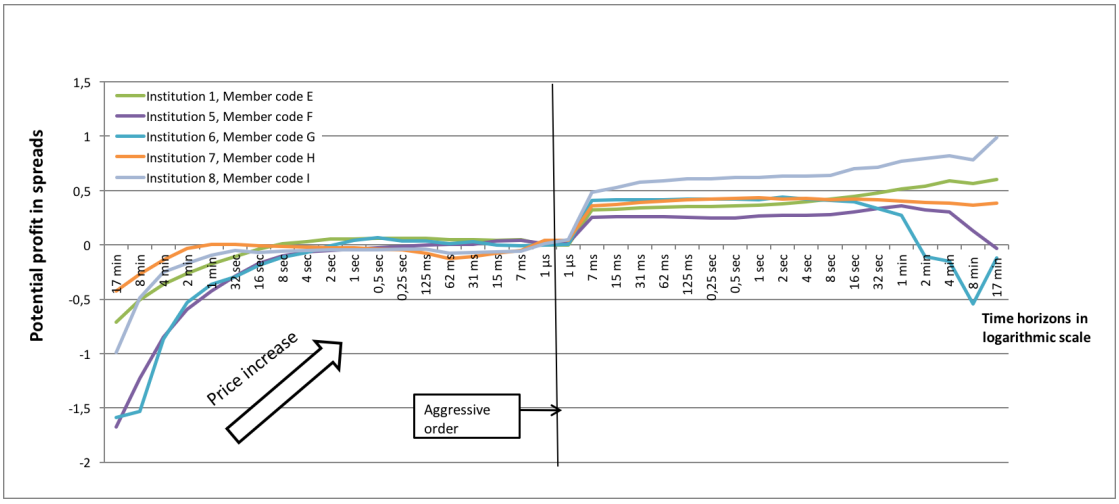


Figure II.16 – Trend following HFTs.

In Figure II.17, we plot the price profiles of three HFTs following a particular strategy consisting in seizing certain opportunities faster than other market participants (Institution 1, Institution 2 and Institution 3) and another HFT (Instituton 10) who does not follow this same particular strategy.

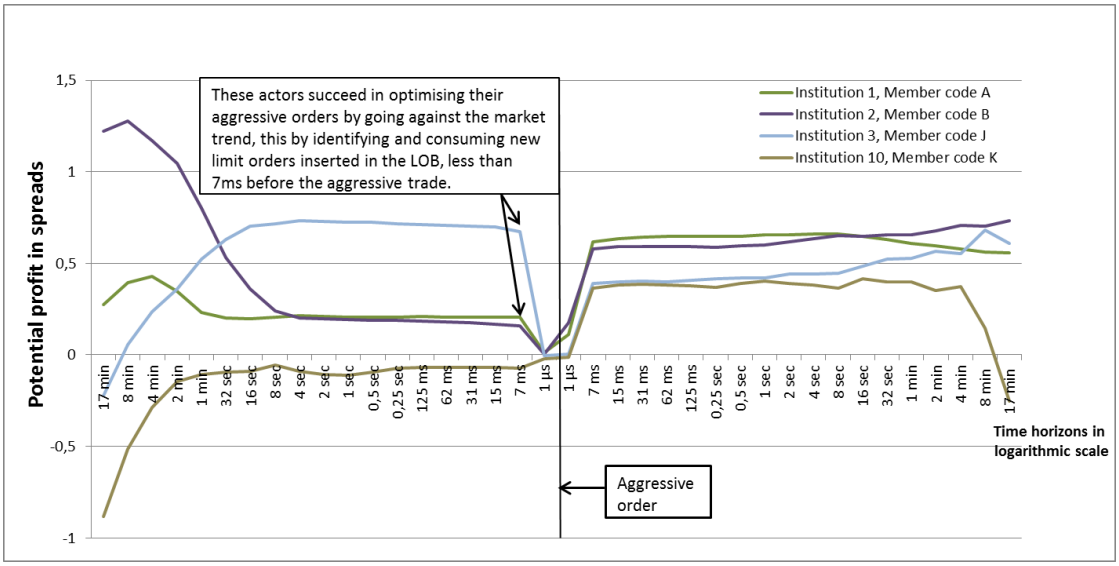


Figure II.17 – Some HFTs following a particular strategy consisting in benefiting from the insertion of new orders that reduce the spread.

The arbitrage configuration in Figure II.17 clearly stands out: the three institutions benefit from the insertion of new orders inserted in the LOB 7 ms before the aggressive order reducing the spread by 0.6 spreads on average (in the case of Institution 3) and by 0.2 spreads on average (in the case of Institutions 1 and 2). These situations are likely to occur when the spread is large (equal to several ticks). In order to check that this arbitrage configuration does not take place only at the beginning or the end of the day, we performed the same computations by excluding all the aggressive trades taking place before 11h and after 16h. The obtained results are quite unchanged. We deduce then that Institutions 1, 2 and 3 do not apply this strategy particularly at the beginning or at the end of the day. This arbitrage configuration might not be a real arbitrage configuration, but a consequence of a scenario where these institutions perform instantaneous wash trades: they insert the limit order inside the spread and then consume their own limit orders. To make sure that we are not in this kind of situation, we performed the same computations by excluding all the wash trades of all institutions. The obtained results are unchanged. We deduce that this is a real arbitrage configuration and not a wash trades one.

Some institutions carry out distinct strategies simultaneously. For instance, Institution 1 follows mean reversion and trend following strategies at the same time using different member codes. By using the same member code used for the mean reversion strategy, Institution 1 succeeds to profit from local opportunities by benefiting from the insertion of new orders that reduce the spread (it is the particular strategy showed in Figure II.17).

7 Conclusion

We show that the proportion of executed volume at the best limit does have a significant and durable impact on prices: the price impact depends on both traded volume and volume present at the best limit. In terms of strategy, when observing the price evolution before the aggressive order, it is clear that HFTs are typically mean reverting, while agency members are trend following. In addition to this, we have seen that the aggressive orders of HFTs are more informed than those of other market participants. Finally, we have validated some expected stylised facts: The aggressive orders of agency members are the most autocorrelated, while those of HFTs are the least autocorrelated. This shows that the high potential profit of HFTs is not due to the price impact that they generate but to a real informational advantage. This informational advantage is mainly due to the fact that HFTs use more sophisticated infrastructure and automation technologies than other participants, which allows them to predict the price evolution before the rest of market participants. These findings can explain why some exchanges plan to reduce the aggressive behaviour of HFTs by imposing "speed bumps" which aim is to impose a speed limit on their aggressive trades. These speed bumps that have mostly been introduced in American exchanges recently start to become more popular. This is the case for example of the Deutsche Börse Eurex platform that will test a six-month pilot project from June 3, 2019 slowing down the HFTs aggressive orders by one millisecond for trading on German and French options.

II.A Generalisation to all HFTs

We generalise the study of partial aggressive orders to all HFT members (and not to the members of the SLP programme only) that we classify according to the classification described in Section 6.2 based on the lifetime of cancelled orders. We compare their potential profit to those of other market participant classes (mixed members and non-HFTs).

The table below shows the total number of member codes and the number of member codes issuing enough partial aggressive orders (using the same criterion as in Section 5.3) relative to each market participant class. The results here are quite similar to those obtained for SLPs in the previous Section 5.3, (see Figure II.18).

Market participant class	Number of member codes	Number of member codes issuing enough partial aggressive orders
HFT	20	12
Mixed	13	13
non-HFT	85	30

Table II.5 – Number of member codes issuing enough partial aggressive orders according to each market participant class.

Out of the 12 remaining HFTs, 8 are SLPs.

Figure II.18 shows that over a short time horizon (from 7 milliseconds until two minutes approximately after the aggressive trade), the majority of HFTs (between 77% and 92% of them) display a potential profit higher than the third quartile, versus 18% on average (the average value is computed starting 7 milliseconds until 1 minute after the aggressive order) for mixed members and only 3.5% for non-HFTs. Beyond one minute, the presence of HFTs over the third quartile decreases to the benefit of other participants, in particular the mixed member codes. From 31 milliseconds to 4 seconds after the aggressive order, the proportion of HFTs having a potential profit higher than the third quartile is quite constant, equal to 92%. During this time interval, one HFT member code (whose aggressive flows constitute 7% of the total aggressive flows of this institution) only does not realise potential profits higher than the third quartile. We note that the other member codes of this institution are more profitable than 75% of the market participants over the studied time horizon.

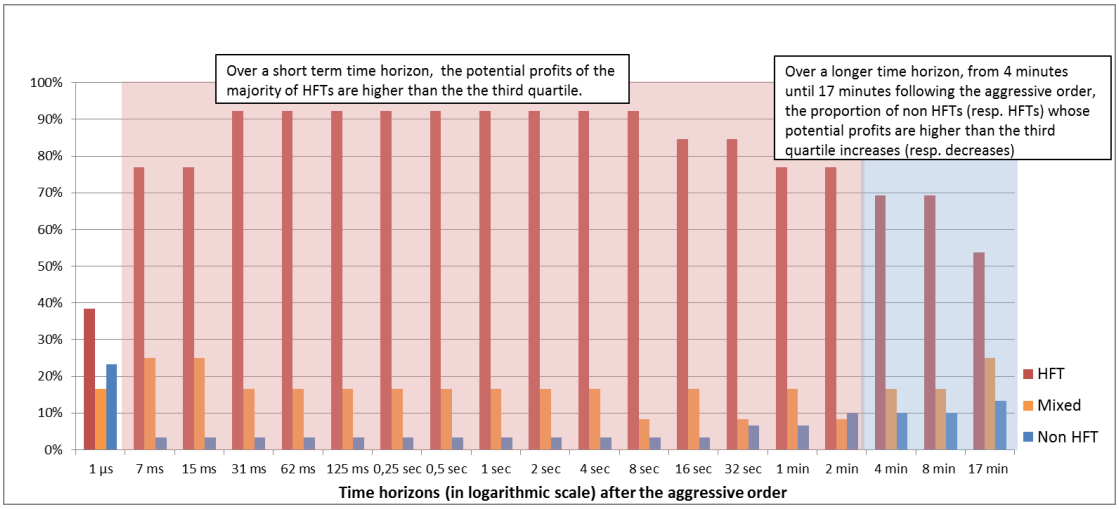


Figure II.18 – Evolution of member codes proportion with a potential profit higher than the third quartile for each market participant class over different time horizons following the partial aggressive order.

Part 2

From empirical observations to agent-based modelling

CHAPTER III

From Glosten-Milgrom to the whole limit order book and applications to financial regulation

Abstract

We build an agent-based model for the order book with three types of market participants: informed trader, noise trader and competitive market makers. Using a Glosten-Milgrom like approach, we are able to deduce the whole limit order book (bid-ask spread and volume available at each price) from the interactions between the different agents. More precisely, we obtain a link between efficient price dynamic, proportion of trades due to the noise trader, traded volume, bid-ask spread and equilibrium limit order book state. With this model, we provide a relevant tool for regulators and market platforms. We show for example that it allows us to forecast consequences of a tick size change on the microstructure of an asset. It also enables us to value quantitatively the queue position of a limit order in the book.

Keywords: Market microstructure, limit order book, bid-ask spread, adverse selection, financial regulation, tick size, queue position valuation.

1 Introduction

Limit order book (LOB) modelling has become an important research topic in quantitative finance. This is because market participants and regulators need to use LOB models for many different tasks such as optimising trading tactics, assessing the quality of the various algorithms operating on the markets, understanding the behaviours of market participants and their impact on the price formation process or designing new regulations at the microstructure level. In the literature, there are two main ways to model the LOB: statistical and equilibrium models. In statistical models, agents order flows follow suitable stochastic processes. In this type of approach, the goal is to reproduce important market stylised facts and to be useful in practice, enabling practitioners to compute relevant quantities such as trading costs, market impact or execution probabilities. Most statistical models are so-called

zero-intelligence models because order flows are driven by independent Poisson processes, see for example [1, 38, 39, 77, 106]. This assumption is relaxed in [13, 63, 65] where more realistic dynamics are obtained introducing dependencies between the state of the order book and the behaviour of market participants.

In equilibrium models, see for instance [49, 53, 96, 101], LOB dynamics arise from interactions between rational agents acting optimally: the agents choose their trading decisions as solutions of individual utility maximisation problems. For example in [96], the author investigates a simple model where traders choose the type of order to submit (market or limit order) according to market conditions, and taking into account the fact that their decisions can influence other traders. In this framework, it becomes possible to analyse accurately market equilibriums. However, the spread is exogenous and there is no asymmetric information on the fundamental value of the asset so that no adverse selection effect is considered. This is the case in the order-driven model of [101] too, where traders can also choose between market and limit orders. In this approach, all information is common knowledge and the waiting costs are the driving force. This model leads to several very relevant predictions about the links between trading flows, market impact and LOB shape.

In this paper, we introduce an equilibrium-type model. It is a simple agent-based model for the order book where we consider three types of market participants like in [76]: an informed trader, a noise trader and market makers. The informed trader receives market information such as the jumps of the efficient price, which is hidden to the noise trader. He then takes advantage of this information to gain profit by sending market orders. Market makers also receive the same information but with some delay and they place limit orders as long as the expected gain of these orders is positive (they are assumed to be risk-neutral). The informed trader and market makers represent the strategic part in the trading activity, while the random part consists in the noise trader who is assumed to send market orders according to a compound Poisson process.

Interestingly, the above simple framework allows us to deduce a link between efficient price dynamic, proportion of trades due to the noise trader, traded volume, bid-ask spread and equilibrium state for the LOB. It enables us to derive the whole order book shape (bid-ask spread and volume present at each price) from the interactions between the agents. The question of how the bid-ask spread emerges from the behaviour of market participants has been discussed in many works. It is generally accepted that the bid-ask spread is non-zero because of the existence of three types of costs: order processing costs, see [62, 101], inventory costs, see [59, 115], and adverse selection costs, see [53]. In the already mentioned paper [101], the spread is a consequence of order processing costs: to compensate their waiting costs, traders place their limit orders on different price levels (for example, a sell limit order at a higher level gets a better expected price than one at a lower level but needs longer time to be executed. Thus the case where both orders lead to the same expected utility can be considered).

In contrast, our model is inspired by [53]. Liquidity is offered by market makers only and they

face an adverse selection issue since a participant agreeing to trade at the market maker's ask or bid price may be trading because he is informed. Order processing and inventory costs are neglected and we consider the bid-ask spread as a purely informational phenomenon: limit orders are placed at different levels because liquidity providers must protect themselves from traders with superior information. In this framework, in a very similar way as in [53], the bid-ask spread emerges naturally from the fact that limit orders placed too close to the efficient price have negative expected returns when being executed: the presence of the informed trader and the potential large jumps of the efficient price prevent market makers from placing limit orders too close to the efficient price. We also find that the bid-ask spread turns out to be the sum of the tick value and of the intrinsic bid-ask spread, which corresponds to a hypothetical value of the bid-ask spread under infinitesimal tick size.

Let us emphasise that several models study the LOB assuming the presence of our three types of market participants and imposing, as we will do, a zero-profit type condition stating that limit orders can only be placed in the LOB if their expected return relative to the efficient price is non-negative. For instance, the papers [53] and [12] share multiple similarities with ours. Compared with [53], there are two main differences. First, in [53], the zero-profit assumption applies only to the two best offer limits: the bid and ask prices at each trade are set to yield zero-profit to the market maker, and time priority plays no role. In our model, we propose a generalised version of the zero-profit condition under which fast market makers can still make profits because of time priority. Second, in [53], one assumes that only unit trades can occur, which is quite restrictive. In our model we relax this assumption, which allows us to retrieve the whole LOB shape and not only the bid-ask spread. In addition to this, we also treat the case where the tick size is non-zero, whereas it is assumed to be vanishing in [53].

In [12], the authors investigate the consequences of a zero-profit condition at the level of the whole liquidity supply curve provided by each market maker. This is an intricate situation where standard equilibriums cannot be reached since a profitable deviation (from a Nash equilibrium) for any market maker is to offer the shares at a slightly higher price as explained in [16]. In this work, we rather assume that when a market maker computes his expected profit, he takes into consideration the orders submitted by other market makers. This is done so that the zero-profit condition holds only for the last order of each queue in the LOB. It in particular means that a market maker can still make positive profit. This enables us to obtain a very operational and tractable framework, where we can deduce the whole LOB shape, compute various important quantities such as priority values of limit orders, and make predictions about consequences of regulatory changes, for example on the tick size.

Note that an important point in our model is that we also consider the case where the tick size is non-zero. This allows us to analyse its role in the LOB dynamic. For instance, we derive a new and very useful relationship between the tick size and the spread. We validate this relationship on market data and show how to use it for regulatory purposes, in particular to forecast new spread values after tick size changes.

The discreteness of available price levels also enables us to value in a quantitative way the queue position of limit orders. LOBs use a priority system for limit orders submitted at the same price. Several priority rules can be employed such as price-time priority or price-size priority, see [54]. We consider here the widely used price-time mechanism which gives priority to the limit orders in a first in first out way. Therefore it encourages traders to submit limit orders early. Our model is one of the only approaches allowing to quantify with accuracy the advantage of being at the top of the queue compared to being at its end. A notable exception is the paper [91]. In this work, the authors value queue positions at the best levels for large tick assets in a queuing model taking into account price impact and some adverse selection. In our setting, we are able to compute the effects of the strategic interactions between market participants on queue position valuation. Furthermore, we are not restricted to the best levels of large tick assets. However, as will be seen in our empirical results, our findings are in line with those of [91].

Other well-known stylised facts are reproduced in our model. For instance, when the absolute value of the efficient price jumps follows a Pareto distribution, we retrieve the classical linear relationship between spread and volatility per trade proved in [86], see also [115]. This will be particularly helpful to calibrate our model so that one can use it as a market simulator for analysing regulatory measures.

The paper is organised as follows. In Section 2, we introduce our agent-based LOB model with zero tick value. Based on a greedy assumption for the informed trader's behaviour, a link is deduced between traded volume, efficient price jump distribution and LOB shape. We then add the zero-profit condition for market makers, which enables us to compute explicitly the bid-ask spread as well as the LOB shape. In Section 3, the case of non-zero tick value is considered. We show that the bid-ask spread is in fact equal to the sum of the intrinsic bid-ask spread (without the tick value constraint) and the tick value. The LOB shape under positive tick size is also deduced and we give an explicit formula for the value of the queue position of a limit order. In Section 4, based on the results of the model, we make the exercise of forecasting new spread values for the CAC 40 assets whose tick sizes have changed due to the new MiFID II directive. Section 5 is devoted to the calibration of the model and the computation of queue position values for small tick assets of the CAC 40 index. Finally, the proofs are relegated to an appendix.

2 Model and assumptions

In our model, we assume the existence of an efficient price modelled by a compound Poisson process and the presence of three different types of market participants: an informed trader, a noise trader and several market makers. In our approach, market makers choose their bid-ask quotes by computing the expected gain of potential limit orders at various price levels. This is done in a context of asymmetric information between the informed and the noise trader regarding the efficient price (the efficient price is actually used as a tool to materialize asymmetry of information). This framework enables us to obtain explicit formulas for the

spread, LOB shape and variance per trade. These quantities essentially depend on the law of the efficient price jumps, the distribution of the noise trader's orders size, and the number of price jumps compared to that of orders sent by the noise trader. Note that contrary to most LOB models which deal only with the dynamics at the best bid/ask limits, or assume that the spread is constant, see for example [38], our model allows for spread variations and applies to the whole LOB shape. We present in this section the case where the tick size is assumed to be equal to zero. The obtained results will help us to understand those in Section 3 where we consider a positive tick size.

2.1 Modelling the efficient price

We write $P(t)$ for the market underlying efficient price, whose dynamic is described as follows:

$$P(t) = P_0 + Y(t),$$

where $Y(t) = \sum_{j=1}^{N_t} B_j$ is a compound Poisson process and $P_0 > 0$. Here $\{N_t : t \geq 0\}$ is a Poisson process with intensity $\lambda^i > 0$, and the $\{B_j : j \geq 1\}$ are independent and identically distributed square integrable random variables with positive symmetric density f_ψ on $\mathbb{R}$ and cumulative distribution function F_ψ. Hence we consider that new information arrives on the market at discrete times given by a Poisson process with intensity λ^i. So we assume that at the j^{th} information arrival time, the efficient price $P(t)$ is modified by a jump of random size B_j.

Furthermore, since $\mathbb{E}[B_j] = 0$, we have that $P(t)$ is a martingale. Thus $\mathbb{E}[P(t)] = P_0$ and $\mathrm{Var}[P(t)] = \lambda^i t \mathbb{E}[B_j^2]$. We view $\lambda^i \mathbb{E}[B_j^2]$ as the macroscopic volatility of our asset. In the sequel, for sake of simplicity, we write B for B_j when no confusion is possible.

2.2 Market participants

We assume that there are three types of market participants:

- One informed trader: by this term, we mean a trader who undergoes low latency and is able to access market data and assess efficient price jumps faster than other participants, creating asymmetric information in the market. For instance, he can analyse external information or use lead-lag relationships between assets or platforms to evaluate the efficient price (for details about lead-lag see [57, 61, 67]). Therefore, we assume that the informed trader receives the value of the price jump size B (and the efficient price $P(t)$) just before it happens. He then sends his trades based on this information to gain profit. He does not send orders at other times than those of price jumps and we write Q^i for his order size that will be strategically chosen later. Note that he may not send orders at a price jump time if he considers such action would not be profitable.

- One noise trader: he sends market orders in a zero-intelligence random fashion. We assume that these trades follow a compound Poisson process with intensity λ^u. We denote by $\{Q_j^u : j \geq 1\}$ the noise trader's order sizes which are independent and identically distributed integrable random variables. We write f_{κ^u} for the density of the Q_j^u which

is positive and symmetric on $\mathbb{R}$ (a positive volume represents a buy order, while a negative volume represents a sell order) and F_{κ^u} for their cumulative distribution function. Remark that $r = \frac{\lambda^i}{\lambda^i+\lambda^u}$ corresponds to the average proportion of price jumps compared to the total number of events happening on the market (efficient price jumps and trades by the noise trader). Recall that informed trades can occur only when there is a price jump. We will assume throughout the paper that $r > 0$. We denote by Q the order size independently of the issuer of the order (noise or informed trader).

- Market makers: they receive the value of the price jump size B (and the efficient price $P(t)$) right after it happens. We assume that they are risk neutral. In practice, market makers are often high-frequency traders and considered informed too. However, contrary to our notion of informed trader, their analyses typically rely on order flows (notably through spread and imbalance) to extract the efficient price rather than on external information. This is because directional trading is not at the core of market making algorithms. We consider like in [53] that market makers know the proportion of price jumps compared to the total number of events happening on the market, that they compete with each others, and that they are free to modify their limit orders at any time after a price jump or a transaction. Market makers place their orders according to their potential profit and loss with respect to the efficient price (no inventory aspects are considered here). Thus they only send sell orders at price levels above the efficient price and buy orders at price levels below it.

We assume here that there is no tick size (this assumption will be relaxed in Section 3). The LOB is made of limit orders placed by market makers around the efficient price $P(t)$. We denote the cumulative available liquidity between $P(t)$ and $P(t)+x$ by $L(x)$[1]. When $L(x) \geq 0$ (resp. $L(x) \leq 0$), it represents the total volume of sell (resp. buy) limit orders with price smaller (resp. larger) than or equal to $P(t)+x$. This function L is called cumulative LOB shape function.

2.3 Assumptions

We do not impose any condition on the cumulative LOB shape function L which can have a singular part and discontinuities. We define its inverse L^{-1} by:

$$L^{-1}(q) = \underset{x}{\operatorname{argmin}}\{x | L(x) \geq q\}.$$

Given the function L, we now specify the behaviour of the informed trader in the next assumption. This assumption relates the traded volume of the informed trader Q^i to the LOB cumulative shape L and the size of the price jump B received by the informed trader.

Assumption 1. *Let t be a jump time of the efficient price. Based on the received value B and the cumulative LOB shape function L provided by market makers, the informed trader sends his trades*

[1]This quantity actually depends on time t but for sake of simplicity, we just write $L(x)$.

in a greedy way such that he wipes out all the available liquidity in the LOB until level $P(t)+B$. *Thus, his trade size* Q^i *satisfies:*

$$Q^i = L(B^-).$$

The informed trader computes his gain according to the future efficient price. If he knows that the price will increase (resp. decrease), which corresponds to a positive (resp. negative) jump B, he consumes all the sell (resp. buy) orders leading to positive ex-post profit. In both cases, his profit is equal to the absolute value of the difference between the future efficient price and the price per share at which he bought or sold, multiplied by the consumed quantity. Note that in the spirit of this work, the informed trader does not accumulate position intraday. What we have in mind is that he unwinds his position passively. As an illustration, if at a given moment the efficient price is equal to 10 euros and the future price jump is equal to 0.05 euros, the informed trader consumes all the sell orders at prices between 10 and 10.05 euros. He then can potentially unwind his position by submitting passive sell orders at a price equal to or higher than the new efficient price. Knowing that their latent profit is computed with respect to the efficient price, he can afford submitting them close to the new efficient price, thereby making their execution very likely.

Remark 1. *For a given order of size* Q^i *initiated by the informed trader and for a given quantity* q, *the probability that the trade size* Q^i *is less than* q *satisfies:*

$$\begin{aligned}\mathbb{P}[Q^i < q] &= \mathbb{P}[L(B^-) < q] \\ &= \mathbb{P}[B < L^{-1}(q)] \\ &= F_\psi(L^{-1}(q)).\end{aligned}$$

In the following, our goal is to compute the spread and LOB shape. We proceed in two steps. First, we derive the expected gain of potential limit orders of the market makers. Second, we consider a zero-profit assumption for market makers (due to competition). Based on these two ingredients, we show how the spread and LOB shape emerge.

2.4 Computation of the market makers expected gain

This part is the first step of our approach. We focus here on the gain of passive sell orders. The gain of passive buy orders can be readily deduced the same way.

Let L be the shape of the order book. Our goal is to compute the conditional average profit of a new infinitesimal order if submitted at price level x knowing that $Q > L(x)$ and without any information about the trade's initiator. We write $G(x)$ for this quantity[2].

We consider the profit of new orders with total volume $\varepsilon > 0$, placed between $P(t)+x-\delta p$ and $P(t)+x$ for some $x > 0$ and $\delta p > 0$, given the fact that these orders are totally executed. The volume ε submitted orders are represented by an additional cumulative LOB shape function denoted by $\hat{L}(x)$. Note that we work with orders submitted between $x-\delta p$ and x to take into

[2]Note that the gain depends on time t but we keep the notation $G(x)$ when no confusion is possible.

account two cases: $L(x)$ is continuous at x and $L(x)$ has a mass at x. The function $\hat{L}(x)$ is defined as follows:

- For $s < x - \delta p, \hat{L}(s) = 0$ and the liquidity available in the LOB up to s is equal to $L(s)$.
- For $x - \delta p \leq s \leq x$, the available liquidity is $L(s) + \hat{L}(s)$, where $\hat{L}(x - \delta p) = 0$ and $\hat{L}(x) = \varepsilon$.
- For $s \geq x$, the liquidity available in the LOB up to s is equal to $L(s) + \varepsilon$.

Furthermore, we assume that for any $s < x$, $\hat{L}(s) < \varepsilon$. Let us write:

- ν for a random variable that is equal to 1 if the trade is initiated by the informed trader and 0 if it is initiated by the noise trader.
- $G^{noise}(x - \delta p, x)$ for the gain of new orders with total volume ε submitted between $x - \delta p$ and x in case the trade is initiated by the noise trader knowing that $Q^u \geq L(x) + \hat{L}(x)$.
- $G^{inf}(x - \delta p, x)$ for the gain of new orders with total volume ε submitted between $x - \delta p$ and x in case the trade is initiated by the informed trader knowing that $Q^i \geq L(x) + \hat{L}(x)$.
- $G(x - \delta p, x)$ for the expected conditional gain of new orders with total volume ε submitted between $x - \delta p$ and x knowing that $Q \geq L(x) + \hat{L}(x)$ without any information about the trade's initiator.

The quantity $G(x - \delta p, x)$ is equal to:

$$G^{inf}(x - \delta p, x)\mathbb{P}[\nu = 1 | Q \geq L(x) + \hat{L}(x)] + G^{noise}(x - \delta p, x)\mathbb{P}[\nu = 0 | Q \geq L(x) + \hat{L}(x)].$$

Our aim being to compute the expected gain of a new infinitesimal order if submitted at price level x, we make δp and ε tend to 0. Thus we define

$$G(x) = \lim_{\varepsilon \to 0} \Big(\lim_{\delta p \to 0} \frac{G(x - \delta p, x)}{\varepsilon} \Big).$$

We have the following proposition proved in Appendix III.A.1.

Proposition 1. *For $x \geq 0$, the average profit of a new infinitesimal order if submitted at price level x satisfies:*

$$G(x) = x - \frac{r\mathbb{E}[B 1_{B > x}]}{r\mathbb{P}[B > x] + (1 - r)\mathbb{P}[Q^u > L(x)]}$$

and for $x \leq 0$

$$G(x) = -x + \frac{r\mathbb{E}[B 1_{B < x}]}{r\mathbb{P}[B < x] + (1 - r)\mathbb{P}[Q^u < L(x)]}.$$

Remark that the average profit $G(x)$ above is well defined even when $L(x) = 0$. In fact, when $L(x)$=0, $G(x)$ represents the expected gain of an infinitesimal order submitted in an empty order book at x. Note that for a given x, when $L(x)$ goes large, the expected gain of the limit orders becomes negative.

We now describe the way the LOB is built via a zero-profit type condition. Let us take the ask side of the LOB. For any point x, market makers first consider whether or not there should be liquidity between 0 and x. To do so, they compute the value $\hat{L}(x)$ which is so that we obtain $G(x) = 0$ in the expression in Proposition 1. If $\hat{L}(x)$ is positive, then competition between market makers takes place and the cumulative order book adjusts so that $L(x) = \hat{L}(x)$ in order to obtain $G(x) = 0$. If $\hat{L}(x) = 0$, then there is no liquidity between 0 and x. If $\hat{L}(x)$ is negative, we deduce that there is no liquidity between 0 and x since this liquidity should be positive. This mechanism makes sense since, as we will see in what follows, $\hat{L}(x)$ is a non-decreasing function of x, which implies two things. First, it is impossible to come across a situation where $x_1 < x_2$ and where market makers are supposed to add liquidity between $P(t)$ and $P(t) + x_1$ but not between $P(t)$ and $P(t) + x_2$. Second, the cumulative shape function for the LOB is indeed non-decreasing.

We have that $G(x) = 0$ is equivalent to:

$$x = \begin{cases} \frac{r\mathbb{E}[B\mathbf{1}_{B>x}]}{r\mathbb{P}[B>x]+(1-r)\mathbb{P}[Q^u>\hat{L}(x)]} & \text{if } x \geq 0 \\ \frac{r\mathbb{E}[B\mathbf{1}_{B<x}]}{r\mathbb{P}[B<x]+(1-r)\mathbb{P}[Q^u<\hat{L}(x)]} & \text{if } x \leq 0. \end{cases}$$

This implies:

$$\hat{L}(x) = \begin{cases} F^{-1}_{\kappa^u}\left(\frac{1}{1-r} - \frac{r}{1-r}\mathbb{E}[\max(\frac{B}{x}, 1)]\right) & \text{if } x \geq 0 \\ F^{-1}_{\kappa^u}\left(\frac{-r}{1-r} + \frac{r}{1-r}\mathbb{E}[\max(\frac{B}{x}, 1)]\right) & \text{if } x \leq 0. \end{cases}$$

The details of the computation of $\hat{L}(x)$ are given in Appendix III.A.2.

We formalise now the zero-profit assumption introduced above. It is the second step of our approach in order to eventually compute the spread and LOB shape.

Assumption 2. *For every $x > 0$ (resp. $x < 0$), market makers compute $\hat{L}(x)$. If $\hat{L}(x) \leq 0$ (resp. $\hat{L}(x) \geq 0$), market makers add no liquidity to the LOB: $L(x) = 0$. If $\hat{L}(x) > 0$ (resp. $\hat{L}(x) < 0$), because of competition, the cumulative order book adjusts so that $G(x) = 0$. We then obtain $L(x) = \hat{L}(x)$.*

The above zero-profit assumption can be seen as a generalised version of the zero-profit condition proposed in [53], in which zero-profit is only considered for the two best offer limits. It is also interesting to point out that, under this more realistic setting, those very fast market makers can still make profit as their orders are placed earlier in the LOB.

In this case where the tick size is zero, it can seem difficult to imagine how competition between different market makers takes place. One can think that every market maker specifies his own $L(x)$ (cumulative liquidity that he provides). Then Assumption 2 means that, when there is still room for future profit at x ($G(x) > 0$), other market makers will come to the market and increase the liquidity in the LOB until $G(x)$ becomes null. Note again that we consider here that market makers can insert infinitesimal quantities in the LOB. These ideas will be made clearer in Section 3 where the tick size is no longer zero.

2.5 The emergence of the bid-ask spread and LOB shape

Based on the expected gain of the market makers, see Proposition 1, and the zero-profit condition (Assumption 2), we can derive the bid-ask spread and LOB shape. We have the following theorem proved in Appendix III.A.2.

Theorem 1. *The cumulative LOB shape satisfies* $L(x) = -L(-x)$ *for any* $x \in \mathbb{R}$, $L(x) = 0$ *for* $x \in [-\mu, \mu]$ *and* L *is continuous strictly increasing for* $x > \mu$, *where* μ *is the unique solution of the following equation:*

$$\frac{1+r}{2r} = \mathbb{E}[\max(\frac{B}{\mu}, 1)]. \tag{1}$$

For $x > \mu$, $L(x) > 0$ *and*

$$L(x) = F^{-1}_{\kappa^u}\big(\frac{1}{1-r} - \frac{r}{1-r}\mathbb{E}[\max(\frac{B}{x}, 1)]\big). \tag{2}$$

For $x < -\mu$, $L(x) < 0$ *and*

$$L(x) = F^{-1}_{\kappa^u}\big(\frac{-r}{1-r} + \frac{r}{1-r}\mathbb{E}[\max(\frac{B}{x}, 1)]\big). \tag{3}$$

In particular, the bid-ask spread is equal to 2μ.

Equation (1) shows that the spread is an increasing function of r. This means that market makers are aware of the adverse selection they risk when the number of price jumps increases. As a consequence, they enlarge the spread in order to avoid this effect due to the trades issued by the informed trader just before the price jumps take place. In particular, if there is no noise trader in the market, then $r = 1$ and the spread tends to infinity. On the contrary, when the number of trades from the noise trader increases, market makers reduce the spread because they are less subject to adverse selection. All these results are consistent with the findings in [53].

Equations (2) and (3) show that the liquidity submitted by the market makers is a decreasing function of r. Indeed let us take $x > \mu$ and define $h(r) = \frac{1}{1-r} - \frac{r}{1-r}\mathbb{E}[\max(\frac{B}{x}, 1)]$. We have

$$\frac{\partial h}{\partial r}(r) = \frac{1 - \mathbb{E}[\max(\frac{B}{x}, 1)]}{(1-r)^2} \leq 0.$$

This means that h is a decreasing function of r. The function $F^{-1}_{\kappa^u}$ being increasing, we deduce that $L(x)$ is a decreasing function of r. When the number of price jumps increases, market makers reduce the quantity of submitted passive orders. In contrast, when the number of trades from the noise trader is large, the market becomes very liquid. This is in line with the empirical results in Chapter I where it is shown that just before certain announcements, in order to avoid adverse selection, market makers reduce their depth and increase their spread.

Finally, we recall that in our setting, we do not a priori impose any condition on $L(x)$. Equations (1), (2) and (3) show that the cumulative LOB we obtain is continuous and strictly

increasing beyond the spread. Remark also that $L(x)$ tends to infinity as x goes to infinity. This implies that the noise trader can always find liquidity in the LOB, whatever the size of his market order.

2.6 Variance per trade

The variance per trade is the variance of an increment of efficient price between two transactions. It can be viewed as the ratio between the cumulated variance and the number of trades over the considered period. A lot of interest has been devoted to this notion in the literature, notably because of its connection with the spread and the uncertainty zone parameter η, see [41, 86, 115].

In this work, this quantity will moreover help us estimate the law of the efficient price. Denote by τ_i the time of the i^{th} trade and by P_{τ_i} the value of the efficient price right after this transaction. We have the following result proved in Appendix III.A.3.

Theorem 2. *The variance per trade σ^2_{tr} satisfies:*

$$\sigma^2_{tr} = \mathbb{E}[(P_{\tau_{i+1}} - P_{\tau_i})^2] = \frac{\mathbb{E}[B^2]\mu}{\mathbb{E}[|B|\mathbf{1}_{|B|>\mu}]}.$$

We know that for small tick assets, we should obtain a linear relationship between the volatility per trade and the spread, with a slope coefficient between 1 and 2, see [86, 115]. If we consider such asset, we must then have:

$$\frac{\mathbb{E}[B^2]}{\mathbb{E}[|B|\mathbf{1}_{|B|>\mu}]} \sim \mu.$$

A classical choice, enabling us to satisfy the above relationship is to consider a Pareto distribution for the absolute value of the efficient price jumps with parameters k (the shape) and x_0 (the scale), with $k > 2$ in order to have a finite variance. The variance per trade is in that case equal to:

$$\sigma^2_{tr} = \begin{cases} \frac{x_0^{2-k}(k-1)\mu^k}{k-2} & \text{if } x_0 \leq \mu \\ \frac{(k-1)\mu x_0}{k-2} & \text{if } x_0 \geq \mu. \end{cases}$$

To ensure that the variance per trade is proportional to the square of the spread we should have x_0 proportional to μ. Actually, to match the constants in both cases in the above formulas, we naturally take $x_0 = \mu$. This means that at equilibrium the spread adapts to the minimal jump size or rather that market participants view modifications of the efficient price as significant only provided they are larger than half a spread. In this case, the variance per trade becomes:

$$\sigma^2_{tr} = \frac{k-1}{k-2}\mu^2.$$

Knowing that the slope coefficient between the volatility per trade and the spread lies between 1 and 2, we expect a scale parameter k larger than 2.3. Note that when k tends to infinity, the slope coefficient tends to 1. These results will be heavily used in Section 5.

3 The case of non-zero tick size

In this section, we study the effect of introducing a tick size, denoted by α, that constraints the price levels in the LOB. The same efficient price dynamic as that described in the previous section still applies, but the cumulative LOB shape becomes now a piecewise constant function. Due to price discreteness, the discontinuity points of $L(x)$ will depend on the position of the efficient price $P(t)$ with respect to the tick grid.

3.1 Notations and assumptions

Notations To deal with the discontinuity points of $L(x)$, the following notations will be used in the sequel. Let us denote by $\tilde{P}(t)$ the smallest admissible price level that is greater than or equal to the current efficient price $P(t)$, and their distance by $d := \tilde{P}(t) - P(t)$, where $d \in [0, \alpha)$. The cumulative LOB shape function $L(x)$ is now defined by $L^d(i)$:

$$L^d(i) = \begin{cases} L(d + (i-1)\alpha) & \text{for } i > 0 \\ L(d + i\alpha) & \text{for } i < 0. \end{cases} \tag{4}$$

The index $i = 1$ (resp. $i = -1$) corresponds to the closest price level that is larger (resp. smaller) than or equal to $P(t)$. When $L^d(i) > 0$ (resp. $L^d(i) < 0$), it represents the total volume of sell (resp. buy) passive orders with prices smaller (resp. larger) than or equal to the i^{th} limit.

We write $l^d(i)$ for the quantity placed at the i^{th} limit:

$$l^d(i) = \begin{cases} L^d(i) - L^d(i-1) & \text{for } i > 0 \\ L^d(i) - L^d(i+1) & \text{for } i < 0. \end{cases}$$

When $l^d(i) > 0$ (resp. $l^d(i) < 0$), it represents the volume of sell (resp. buy) limit orders placed at the i^{th} limit. Recall that $l^d(i) \geq 0$ (resp. $l^d(i) \leq 0$) for $i > 0$ (resp. $i < 0$).

Assumptions We adapt Assumption 1 to our tick size setting. We again assume that when he receives new information, the informed trader sends his trades in a greedy way such that he wipes out all the available liquidity at limits where the price is smaller than the new efficient price. This can be translated as follows.

Assumption 1. *When the informed trader sends a market order, then Q^i is equal to $L^d(i)$ for some $i \in \mathbb{Z}^*$. We have $Q^i = L^d(i)$ if and only if $B \in [d + (i-1)\alpha, d + i\alpha]$.*

Remark 1. *In practice, it is rare that a trade consumes more than one limit in the LOB. Such trade in our model should be interpreted in practice as a sequence of transactions, each of them consuming one limit.*

3.2 Computation of the market makers expected gain

As in the previous section, let us compute the conditional average profit of a new infinitesimal passive order submitted at the i^{th} limit, knowing that $Q > L^d(i)$, and without any information

about the trade's initiator. This quantity is denoted by $G^d(i)$ and defined in a similar fashion as $G(x)$ in Section 2.4. The computation of $G^d(i)$ is comparable to that of $G(x)$, and actually even easier since we now have that the volume at the i^{th} limit cannot be infinitesimal. This means that different orders can be submitted at the same price with disparities in their gain according to their position in the queue. For instance, the order placed on top of the queue has the highest expected gain, while we will impose later that the gain of a new order submitted at the rear of the queue is null. We have the following proposition proved in Appendix III.A.4.

Proposition 1. *Under Assumption 1, for $i \in \mathbb{Z}^*$, the expected gain of a new infinitesimal passive order placed at the i^{th} level, given that it is executed, satisfies:*

For $i > 0$:

$$G^d(i) = G(d + (i-1)\alpha) = d + (i-1)\alpha - \frac{r\mathbb{E}[B1_{B>d+(i-1)\alpha}]}{r\mathbb{P}[B > d + (i-1)\alpha] + (1-r)\mathbb{P}[Q^u > L^d(i)]},$$

and for $i < 0$:

$$G^d(i) = G(d + i\alpha) = d + i\alpha - \frac{r\mathbb{E}[B1_{B<d+i\alpha}]}{r\mathbb{P}[B < d + i\alpha] + (1-r)\mathbb{P}[Q^u < L^d(i)]}.$$

The quantity $G^d(i)$ can be understood as the expected gain of a newly inserted infinitesimal limit order at the i^{th} limit, under the condition that it is executed against some market order. For this situation with non-zero tick size, we follow the same reasoning as in the case with zero tick size. Indeed, for all $i \in \mathbb{Z}^*$, market makers compute $\hat{L}^d(i)$ so that $G^d(i) = 0$ in Proposition 1. The equality $G^d(i) = 0$ is equivalent to:

If $i > 0$:

$$d + (i-1)\alpha = \frac{r\mathbb{E}[B1_{B>d+(i-1)\alpha}]}{r\mathbb{P}[B > d + (i-1)\alpha] + (1-r)\mathbb{P}[Q^u > \hat{L}^d(i)]}$$

and if $i < 0$:

$$d + i\alpha = \frac{r\mathbb{E}[B1_{B<d+i\alpha}]}{r\mathbb{P}[B < d + i\alpha] + (1-r)\mathbb{P}[Q^u < \hat{L}^d(i)]}.$$

This is equivalent to:

$$\hat{L}^d(i) = \begin{cases} F^{-1}_{\kappa^u}\left(\frac{1}{1-r} - \frac{r}{1-r}\mathbb{E}[\max(\frac{B}{d+(i-1)\alpha}, 1)]\right) & \text{if } i > 0 \\ F^{-1}_{\kappa^u}\left(\frac{-r}{1-r} + \frac{r}{1-r}\mathbb{E}[\max(\frac{B}{d+i\alpha}, 1)]\right) & \text{if } i \leq 0. \end{cases}$$

As in the case without tick size, this leads to the following zero-profit assumption.

Assumption 2. *For every $i \in \mathbb{Z}^+$ (resp. $i \in \mathbb{Z}^-$), market makers compute $\hat{L}^d(i)$. If $\hat{L}^d(i) \leq 0$ (resp. $\hat{L}^d(i) \geq 0$), market makers add no liquidity to the LOB: $L^d(i) = 0$. If $\hat{L}^d(i) > 0$ (resp. $\hat{L}^d(i) < 0$), because of competition, the cumulative order book adjusts so that $G^d(i) = 0$. We then obtain then $L^d(i) = \hat{L}^d(i)$.*

The zero-profit condition applies only to a new order submitted at the bottom of the queue. The expected profit of the other orders is non-zero, maximum gain being obtained for the one on top of the queue.

3.3 Bid-ask spread and LOB formation

Based on the expected gain of the market makers, see Proposition 1, and the zero-profit condition (Assumption 2), as previously, we deduce the bid-ask spread and LOB shape. We have the following theorem proved in Appendix III.A.5.

Theorem 1. *The LOB shape function satisfies $l^d(i) = 0$ for all $-k_l^d < i < k_r^d$, where k_l^d and k_r^d are two positive integers determined by the following equations:*

$$k_r^d = 1 + \lceil \frac{\mu - d}{\alpha} \rceil, \quad k_l^d = \lceil \frac{\mu + d}{\alpha} \rceil,$$

with μ defined by (1)*, and where $\lceil x \rceil$ denotes the smallest integer that is larger than x (which can be equal to 0). Furthermore, for $i \geq k_r^d$:*

$$L^d(i) = F_{\kappa^u}^{-1}\Big(\frac{1}{1-r} - \frac{r}{1-r}\mathbb{E}[\max(\frac{B}{d + (i-1)\alpha}, 1)]\Big)$$

and for $i \leq -k_l^d$:

$$L^d(i) = F_{\kappa^u}^{-1}(\frac{-r}{1-r} + \frac{r}{1-r}\mathbb{E}[\max(\frac{B}{d + i\alpha}, 1)]).$$

For given d, the bid-ask spread ϕ_α^d satisfies:

$$\phi_\alpha^d = \alpha\big(\lceil \frac{\mu - d}{\alpha} \rceil + \lceil \frac{\mu + d}{\alpha} \rceil\big).$$

Let us consider the approximation that d is uniformly distributed on $[0, \alpha]$ (which is reasonable, see [75, 100, 112]). In this case, we obtain the following corollary proved in Appendix III.A.6.

Corollary 1. *The average spread ϕ_α satisfies:*

$$\phi_\alpha = 2\mu + \alpha. \tag{5}$$

When the tick size is vanishing, we have seen in Theorem 1 that the spread is equal to 2μ. When it is not, the spread cannot necessarily be equal to 2μ because of the tick size constraint. What is particularly interesting is that even if $\alpha \leq 2\mu$, the equilibrium spread is not 2μ. There is always a tick size processing cost leading to a spread value of $2\mu + \alpha$. Equation (5) will be used for practical applications in Section 4.

One numerical example of a limit order book For illustration, we provide now one numerical example of a limit order book. As suggested in Section 2.6, let us consider that the absolute value of the price jumps follows a Pareto distribution with shape and scale parameters respectively equal to 3 and 0.005, $r = 2/3$, $\alpha = 0.01$ and $d = 0.0075$. Moreover, we suppose that Q^u follows a standard normal distribution (here the value 1 of the standard deviation just represents a suitable unity). Under the considered parameters, the spread is equal to 2 ticks and we obtain the LOB given in Figure III.1.

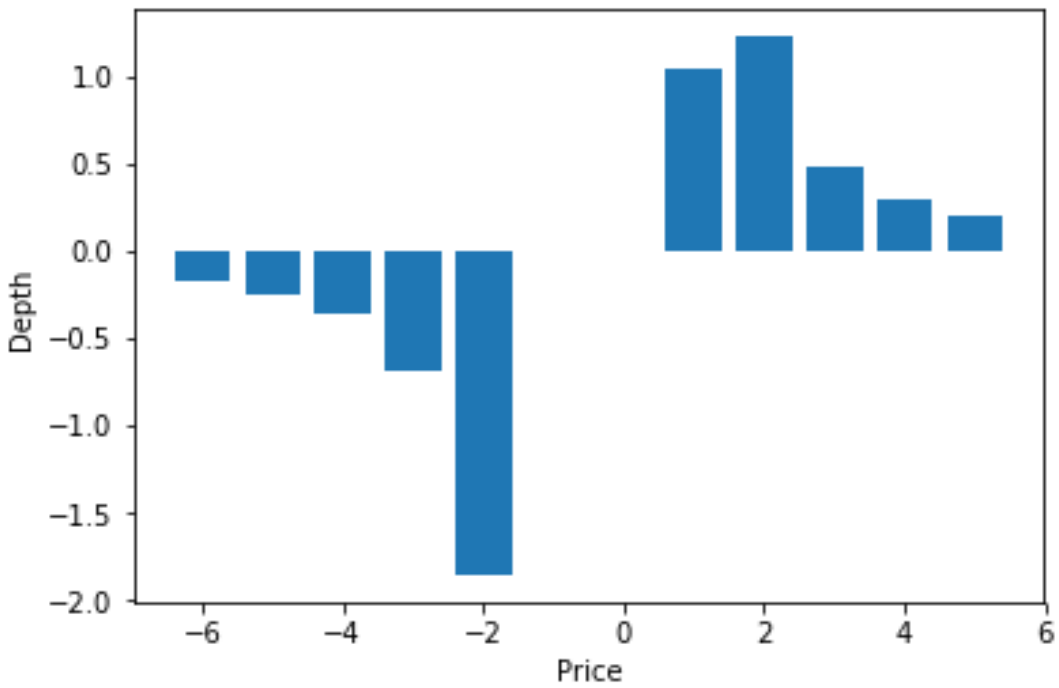


Figure III.1 – Numerical illustration of a LOB.

3.4 Variance per trade

We provide here the variance per trade (defined previously in Section 2.6) in the case where the tick size is non null. This result is obtained in a similar way as Theorem 2.

Theorem 2. *The variance per trade σ^2_{tr} satisfies:*

$$\sigma^2_{tr} = \mathbb{E}[(P_{\tau_{i+1}} - P_{\tau_i})^2] = \frac{\mathbb{E}[B^2](\mu + \alpha/2)}{\mathbb{E}[|B|\mathbf{1}_{|B|>\mu+\alpha/2}]}. \tag{6}$$

As in the zero tick size case, see Section 2.6, we consider that the absolute value of the price jumps follows a Pareto distribution and take x_0 equal to the half spread $\mu+\alpha/2$. The variance per trade becomes:

$$\sigma^2_{tr} = \frac{k-1}{k-2}(\mu + \alpha/2)^2.$$

We will estimate the scale parameter k from this relationship in Section 5.

3.5 Queue position valuation

Introducing a tick size in our modelling enables us to study the value of the position of the limit orders in the queues. We can quantify the advantage of an order placed on top of a

queue compared to another one placed at the bottom. The difference in the values of the positions in a queue is a crucial parameter for trading algorithms. It has actually lead to a technological arms race among high-frequency traders and other automated market participants to establish early (and hence advantageous) positions in the queues, see [3, 91]. Placing limit orders at the front of a queue is very valuable for different reasons. It guarantees early execution and less waiting time. In addition, it reduces adverse selection risk. In fact, as explained in [91], when a limit order is placed at the end of a queue, it is likely that it will be executed against a large trade. In contrast, a limit order placed at the front of the (best) queue will be executed against the next trade independently of the trade size. Large trades are in general sent by informed traders aiming at consuming all limit orders which will generate profit for them. In this way, a limit order submitted at the front of the queue is less likely to undergo adverse selection.

In light of this, to optimise their execution, practitioners need to place limit orders in a relevant way. This requires an estimate of the value of a limit order according to its position in the queue. This very problem is studied in [91] for the queues at the best limits for large tick assets. We complement here this nice work providing formulas valid for any queue of a large or small tick asset and taking into account strategic interactions between market participants.

Assumption 2 tells us that the expected profit of a new infinitesimal limit order placed at the bottom of a non-empty queue is equal to zero. However, under our zero-profit condition, market makers may still make profit if their orders are placed before. The value of queue position at the i^{th} level, denoted by $\tilde{G}^d(i)$, can be formulated in this model as the difference between the expected profit of the order placed on top and that of a new one that would be placed at the bottom of the i^{th} queue. Computing this quantity is very similar to deriving the equations in Proposition 1. The difference is that now we no longer consider a traded volume totally depleting the limit but a traded volume consuming all the limits before the i^{th} one. This leads to the following theorem.

Theorem 3. *For $i \geq k_r^d$, we have*

$$\tilde{G}^d(i) = d + (i-1)\alpha - \frac{r\mathbb{E}[B\mathbf{1}_{B>d+(i-1)\alpha}]}{1 - rF_\psi(d+(i-1)\alpha) - (1-r)F_{\kappa^u}\big(L^d(i-1)\big)}.$$

The formula for $i \leq -k_l^d$ is obviously deduced. We will confront the formula in Theorem 3 to data in Section 5.

4 First practical application: Spread forecasting

Our model (in particular Equation (5)) allows us to forecast the new value of the spread if the tick size is modified. In the following, we predict the spread changes due to the new tick size regime under the recent European directive MiFID II, and compare our results to the effective spread values. We expect our model to be relevant for rather liquid assets since it is based on the presence of competitive market makers. We therefore restrict ourselves to this class.

Note that there are other models in the literature enabling practitioners to forecast spreads, see notably [41] where the authors propose an approach designed for large tick assets. This methodology is applied for example in [64] on Japanese data and in [78] where spread values before and after MiFID II are compared. The advantage of our device is that it can be applied on both small and large tick assets.

4.1 The tick size issue and MiFID II directive

In the recent years, trading platforms have raced to reduce their tick sizes in order to offer better prices and gain market share. This broad trend has had adverse effects on the overall market quality: a too small tick leads to unstable LOBs and a degradation of the price formation mechanism. However, a too large tick prevents the price from moving freely according to the views of market participants. Therefore, finding suitable tick values is crucial for the fluidity of financial markets. To solve this issue, some regulators tried to use pilot programs, as was the case in Japan and in the United States, see for example [64]. This is a costly practice which does not really rely on theoretical foundations. We believe that using quantitative results such as those presented in this work could lead to a much more efficient methodology.

In Europe, MiFID II (Markets in Financial Instruments Directive II) directive introduced a harmonized tick size regime (Article 49) which is based on a two-entries table: price and liquidity (expressed in terms of number of transactions per day). Note that one of the targets for regulators was to obtain for liquid assets spreads between 1.5 and 2 ticks, see [8].

4.2 Data

Our data are provided by the French regulator *Autorité des Marchés Financiers*. We study the CAC 40 stocks over a six months time period around the implementation of MiFID II: from October 2017 to December 2017 (before the tick size changes) and from January 2018 to March 2018 (after the tick size changes). We consider assets whose tick size has changed after the implementation of MiFID II directive[3]. There are 14 stocks from the CAC 40 index then remaining. We note that for all these assets, the tick size was increased.

For each asset we compute two spreads: the first one is averaged over all events occurring in the LOB (transactions, insertions of a new order, cancellations or modifications of an existing order) over the three months before MiFID II and the second one over the three months under MiFID II.

4.3 Prediction of the spread under MiFID II and optimal tick sizes

We now forecast the new spreads of our 14 assets due to the new tick size regime, based on pre-MiFID II data. We use two different predictors. First, we consider that the spread (in euros) remains constant. Second, we compute the new value of the spread based on Equation

[3]We exclude three assets whose tick size fluctuates intraday because of price variations.

Stock	Tick size before MiF. II	Tick size under MiF. II	Average spread before MiF. II (euros)	Average spread before MiF. II (ticks)	Average spread under MiF. II (euros)	Average spread under MiF. II (ticks)	Expected spread based on our model	Relative error
Accor	0.005	0.01	0.011	2.266	0.016	1.586	0.016	3%
Bouygues	0.005	0.01	0.011	2.277	0.017	1.734	0.016	5%
Kering	0.05	0.1	0.090	1.797	0.141	1.407	0.140	1%
Legrand	0.01	0.02	0.016	1.643	0.029	1.471	0.026	10%
Publicis	0.01	0.02	0.019	1.904	0.030	1.520	0.029	4%
Safran	0.01	0.02	0.019	1.892	0.031	1.556	0.029	7%
Schneider Electric	0.01	0.02	0.016	1.579	0.025	1.235	0.026	4%
TechnipFMC	0.005	0.01	0.010	2.056	0.017	1.677	0.015	10%
Valeo	0.01	0.02	0.018	1.845	0.031	1.568	0.028	10%
Veolia Environnement	0.005	0.01	0.007	1.440	0.012	1.189	0.012	3%
Vinci	0.01	0.02	0.017	1.668	0.026	1.280	0.027	4%
Vivendi	0.005	0.01	0.007	1.408	0.012	1.162	0.012	4%

Table III.1 – Forecasting CAC 40 assets spreads under MiFID II.

(5), with μ estimated on the period from October 2017 to December 2017. We compare the accuracy of our forecasts with respect to the effective spread values in Table III.1.

The forecasts based on our model are very accurate: the average relative error is equal to 5% while it is 43% for the other predictor. Remark also that the errors obtained under our methodology are always smaller than the initial tick size, which is almost never the case if one just assumes that the spread in euros is constant.

5 Second practical application: Queue position valuation

As we have seen in Section 3.5, our approach enables us to measure quantitatively the value of queue position thanks to Theorem 3. To use this result, we need to know the distribution of B. To estimate it, we use the Pareto parametrization of Section 3.4. We will estimate the parameter k from Equation (6) and compute r using Equation (1). Values of queue positions will then be deduced.

5.1 Data

To complement the results of [91], we consider in this section the values of queue position for all small tick stocks of the CAC 40 index (that is stocks for which the average spread is larger than 2 ticks). We study this quantity under MiFID II. These assets are investigated over a three months period: from January 2018 to March 2018. This leaves us with five stocks. We compute on a daily basis the average spread over all events occurring in the LOB and the variance per trade, for each stock.

5.2 Pareto parameters estimation methodology

Under our parametrization, the variance per trade is given by

$$\sigma_{tr}^2 = \frac{(k-1)(\mu+\alpha/2)^2}{k-2}.$$

We estimate k by minimising for each stock the quadratic error:

$$\sum_j \left(\sigma_{tr}^2 - (\widetilde{\sigma_{tr}^2})_j\right)^2,$$

where σ_{tr}^2 is the variance per trade obtained in our model and $(\widetilde{\sigma_{tr}^2})_j$ is the variance per trade measured on data (realised variance based on 5 minutes price sampling divided by number of trades) on day j. We search for the optimal k between 2.001 and 20. Note that for each stock, the considered spread is equal to the average realised spread during the period under study.

5.3 Queue position valuation

We first report in Table III.2 the values of the queue position at the best ask limit according to d. We consider that d can be equal to 0.25α, 0.5α or 0.75α.

Stock	Spread (euros)	Spread (ticks)	k	r	Priority value for $d = 0.25\alpha$ (euros)	Priority value for $d = 0.25\alpha$ (spreads)	Priority value for $d = 0.5\alpha$ (euros)	Priority value for $d = 0.5\alpha$ (spreads)	Priority value for $d = 0.75\alpha$ (euros)	Priority value for $d = 0.75\alpha$ (spreads)
Airbus	0.020	2.040	3.478	71%	0.011	53%	0.012	61%	0.014	68%
Lafarge Holcim	0.026	2.438	3.316	70%	0.010	39 %	0.011	42%	0.012	48%
Renault	0.025	2.476	2.866	65%	0.010	41%	0.011	45%	0.012	50%
Saint-Gobain	0.010	2.035	4.776	79%	0.006	55%	0.007	65%	0.008	76%
Société Generale	0.010	2.012	9.910	90%	0.006	60%	0.007	73%	0.009	86%

Table III.2 – Queue position values at the best ask according to d.

We see that the values of queue position are of the same order of magnitude as the bid-ask spreads. This is in line with the findings in [91]. In addition, we get that it is increasing with d. Furthermore, remark that as expected from Section 2.6, the values of k are larger than 2.3.

We now compute in Table III.3 the values of queue position at the four best limits when $d = 0.5\alpha$.

Stock	Spread (euros)	Spread (ticks)	k	r	First limit priority value (euros)	First limit priority value (spreads)	Second limit priority value (euros)	Second limit priority value (spreads)	Third limit priority value (euros)	Third limit priority value (spreads)	Fourth limit priority value (euros)	Fourth limit priority value (spreads)
Airbus	0.020	2.040	3.478	71%	-0.005	-25%	0.012	61%	0.015	72%	0.014	67%
Lafarge Holcim	0.026	2.438	3.316	70%	-0.008	-31%	0.011	42 %	0.014	55%	0.013	52%
Renault	0.025	2.476	2.866	65%	-0.007	-30%	0.011	45%	0.013	54%	0.013	51%
Saint-Gobain	0.010	2.035	4.776	79%	-0.003	-25%	0.007	65%	0.009	84%	0.008	78%
Société Generale	0.010	2.012	9.910	90%	-0.003	-25%	0.007	73%	0.012	117%	0.013	127%

Table III.3 – Queue position values at the four best limits for $d = 0.5\alpha$.

Note that the queue position value of the first limit does not necessarily correspond to the best ask. For example, if the priority value at the first limit is negative and the one at the second

limit is positive, the best ask is the second limit. We observe that the value of queue position is increasing according to the rank of the limit, up to some level after which it decreases.

Conclusion

In this article, we introduce an agent-based model for the LOB. Inspired by the seminal paper by Glosten and Milgrom [53], we use a zero-profit condition for the market makers which enables us to derive a link between proportion of events due to the noise trader, bid-ask spread, dynamic of the efficient price and equilibrium LOB state. The effect of introducing a tick size is then discussed. We in particular show that the constrained bid-ask spread is equal to the sum of the tick value and the intrinsic bid-ask spread that corresponds to the case of a vanishing tick size. This model allows us to do spread forecasting when one modifies the tick size. Price discreteness also enables us to value queue positions in the LOB.

In our approach, market makers only are allowed to insert limit orders. In practice, the roles of informed trader and market makers are often mixed, and the informed trader also has the possibility to place passive limit orders. By doing so, he may get better prices but also leak some information to other market participants. Extending our model by taking into account accurately these intricate features is left for future work.

III.A Proofs

III.A.1 Proof of Proposition 1

We consider the gain of passive sell orders. The gain of passive buy orders can be easily deduced.

First, we compute $G^{inf}(x-\delta p, x)$. We have:

$$\begin{aligned} G^{inf}(x-\delta p, x) &= \int_{x-\delta p}^{x} (P(t)+s)\mathrm{d}\hat{L}(s) - \int_{x-\delta p}^{x} (P(t)+\mathbb{E}[B|B>x])\mathrm{d}\hat{L}(s) \\ &= \int_{x-\delta p}^{x} s\mathrm{d}\hat{L}(s) - \hat{L}(x)\mathbb{E}[B|B>x]. \end{aligned}$$

For $G^{noise}(x-\delta p, x)$ we get:

$$G^{noise}(x-\delta p, x) = \int_{x-\delta p}^{x} (P(t)+s)\mathrm{d}\hat{L}(s) - \int_{x-\delta p}^{x} P(t)\mathrm{d}\hat{L}(s) = \int_{x-\delta p}^{x} s\mathrm{d}\hat{L}(s).$$

We deduce that:

$$\begin{aligned}
G(x-\delta p,x) &= G^{inf}(x-\delta p,x)\mathbb{P}[\nu=1|Q\geq L(x)+\hat{L}(x)]+G^{noise}(x-\delta p,x)\mathbb{P}[\nu=0|Q\geq L(x)+\hat{L}(x)]\\
&= \int_{x-\delta p}^{x} s\mathrm{d}\hat{L}(s)-\mathbb{P}[\nu=1|Q\geq L(x)+\hat{L}(x)]\hat{L}(x)\mathbb{E}[B|B>x]\\
&= \int_{x-\delta p}^{x} s\mathrm{d}\hat{L}(s)-\hat{L}(x)\mathbb{E}[B|B>x]\frac{r\mathbb{P}[B>x]}{\mathbb{P}[Q\geq L(x)+\hat{L}(x)]}\\
&= \int_{x-\delta p}^{x} s\mathrm{d}\hat{L}(s)-\hat{L}(x)\frac{r\mathbb{E}[B\mathbf{1}_{B>x}]}{\mathbb{P}[Q\geq L(x)+\hat{L}(x)]}\\
&= \int_{x-\delta p}^{x} s\mathrm{d}\hat{L}(s)-\hat{L}(x)\frac{r\mathbb{E}[B\mathbf{1}_{B>x}]}{r\mathbb{P}[B>x]+(1-r)\mathbb{P}[Q^u>L(x)+\hat{L}(x)]}.
\end{aligned}$$

Integrating by part we get

$$\int_{x-\delta p}^{x} s\mathrm{d}\hat{L}(s)=\hat{L}(x)x-\int_{x-\delta p}^{x}\hat{L}(s)\mathrm{d}s=\varepsilon x-\int_{x-\delta p}^{x}\hat{L}(s)\mathrm{d}s.$$

When δp tends to 0, this tends to εx. Consequently, we have:

$$\lim_{\delta p\to 0} G(x-\delta p,x)=\varepsilon\big(x-\frac{r\mathbb{E}[B\mathbf{1}_{B>x}]}{r\mathbb{P}[B>x]+(1-r)\mathbb{P}[Q^u>L(x)+\hat{L}(x)]}\big),$$

and

$$G(x)=\lim_{\varepsilon\to 0}\big(\lim_{\delta p\to 0}\frac{G(x-\delta p,x)}{\varepsilon}\big)=x-\frac{r\mathbb{E}[B\mathbf{1}_{B>x}]}{r\mathbb{P}[B>x]+(1-r)\mathbb{P}[Q^u>L(x)]}.$$

III.A.2 Proof of Theorem 1

We consider the passive sell orders ($x>0$). We first compute $\hat{L}(x)$ which is the theoretical liquidity that market makers should add in the LOB in order to obtain $G(x)=0$. Under Proposition 1, $G(x)=0$ is equivalent to:

$$\begin{aligned}
\mathbb{P}[Q^u>L(x)] &= \frac{r}{1-r}\big(\mathbb{E}[\frac{B}{x}\mathbf{1}_{B>x}]-\mathbb{P}[B>x]\big)\\
&= \frac{r}{1-r}\big(\mathbb{E}[\frac{B}{x}\mathbf{1}_{B>x}]-1+\mathbb{P}[B<x]\big)\\
&= \frac{r}{1-r}\big(-1+\mathbb{E}[\max(\frac{B}{x},1)]\big).
\end{aligned}$$

We deduce that

$$\hat{L}(x)=F^{-1}_{\kappa^u}(\frac{1}{1-r}-\frac{r}{1-r}\mathbb{E}[\max(\frac{B}{x},1)]).$$

We now prove that the spread is positive and finite and deduce the shape of the whole LOB.

Recall that $\hat{L}(x)$ computed above is a theoretical value, and that market makers will add liquidity only when $\hat{L}(x) > 0$.

We have $\hat{L}(x) > 0$ when $(\frac{1}{1-r} - \frac{r}{1-r}\mathbb{E}[\max(\frac{B}{x}, 1)]) > \frac{1}{2}$. This holds for all x such that $\mathbb{E}[\max(\frac{B}{x}, 1)] < \frac{1+r}{2r}$. Equivalently, the inequality is satisfied for any x such that $x > \mu$, where μ is unique solution of the following equation:

$$\mathbb{E}[\max(\frac{B}{\mu}, 1)] = \frac{1+r}{2r}.$$

By Assumption 2, we deduce that for any $x \leq \mu$, $L(x) = 0$. Moreover, for any $x > \mu$,

$$L(x) = F_{\kappa^u}^{-1}\big(\frac{1}{1-r} - \frac{r}{1-r}\mathbb{E}[\max(\frac{B}{x}, 1)]\big).$$

We deduce that μ is the half spread.

The cumulative LOB we obtain is unique, continuous and strictly increasing beyond the spread (since the laws of B and Q^u have positive densities on $\mathbb{R}$).

III.A.3 Proof of Theorem 2

We denote by ν_i the random variable that is equal to 1 if the i^{th} trade is initiated by the informed trader and 0 if it is initiated by the noise trader. We write ω for the number of events[4] between two successive trades. We have:

$$\sigma_{tr}^2 = \mathbb{E}[(P_{\tau_{i+1}} - P_{\tau_i})^2] = \sum_{j=1}^{\infty} \mathbb{P}[\omega = j]\mathbb{E}[(P_{\tau_{i+1}} - P_{\tau_i})^2 | \omega = j].$$

with

$$\begin{aligned}\mathbb{E}[(P_{\tau_{i+1}} - P_{\tau_i})^2 | \omega = j] = {} & \mathbb{P}[\nu_{\tau_{i+1}} = 0 | \omega = j]\mathbb{E}[(P_{\tau_{i+1}} - P_{\tau_i})^2 | \omega = j, \nu_{\tau_{i+1}} = 0] \\ & + \mathbb{P}[\nu_{\tau_{i+1}} = 1 | \omega = j]\mathbb{E}[(P_{\tau_{i+1}} - P_{\tau_i})^2 | \omega = j, \nu_{\tau_{i+1}} = 1].\end{aligned}$$

Knowing that $\omega = j$, the j^{th} event can be a trade initiated by the noise trader or a trade initiated by the informed trader. We have

$$\begin{aligned}\mathbb{E}[(P_{\tau_{i+1}} - P_{\tau_i})^2 | \omega = j, \nu_{\tau_{i+1}} = 1] &= \mathbb{E}[(\sum_{k=1}^{j-1} B_k + B_j)^2 || B_k| < \mu, |B_j| > \mu] \\ &= (j-1)\mathbb{E}[B_k^2 || B_k| < \mu] + \mathbb{E}[B_j^2 || B_j| > \mu].\end{aligned}$$

and

$$\mathbb{E}[(P_{\tau_{i+1}} - P_{\tau_i})^2 | \omega = j, \nu_{\tau_{i+1}} = 0] = \mathbb{E}[(\sum_{k=1}^{j-1} B_k)^2 || B_k| < \mu] = (j-1)\mathbb{E}[B_k^2 || B_k| < \mu].$$

[4] An event can be either a trade sent by the noise trader (in that case it necessarily triggers a new transaction), or an information update B which may or may not trigger a trade, depending on whether or not $|B| > \mu$.

We compute the probabilities:

$$\mathbb{P}[\nu_{\tau_{i+1}} = 1 | \omega = j] = \frac{r\mathbb{P}[|B| > \mu]}{1 - r\mathbb{P}[|B| < \mu]}$$

and

$$\mathbb{P}[\nu_{\tau_{i+1}} = 0 | \omega = j] = \frac{1 - r}{1 - r\mathbb{P}[|B| < \mu]}.$$

Consequently,

$$\sigma_{tr}^2 = \mathbb{E}[B^2 | |B| \leq \mu] \sum_{j=1}^{\infty} (j-1)\mathbb{P}[\omega = j] + \mathbb{E}[B^2 | |B| > \mu] \frac{r\mathbb{P}[|B| > \mu]}{1 - r\mathbb{P}[|B| < \mu]}.$$

We have:

$$\begin{aligned}
\sum_{j=1}^{\infty} \mathbb{P}[\omega = j](j-1) &= (1 - r\mathbb{P}[|B| < \mu]) \sum_{j=1}^{\infty} (j-1)(r\mathbb{P}[|B| < \mu])^{j-1} \\
&= (1 - r\mathbb{P}[|B| < \mu]) \sum_{j=0}^{\infty} j(r\mathbb{P}[|B| < \mu])^{j} \\
&= \frac{r\mathbb{P}[|B| < \mu]}{\left(1 - r\mathbb{P}[|B| < \mu]\right)}.
\end{aligned}$$

We deduce:

$$\sigma_{tr}^2 = \frac{r\mathbb{E}[B^2 \mathbf{1}_{|B|<\mu}] + r\mathbb{E}[B^2 \mathbf{1}_{|B|>\mu}]}{1 - r\mathbb{P}[|B| \leq \mu]} = \frac{r\mathbb{E}[B^2]}{1 - r\mathbb{P}[|B| \leq \mu]}.$$

Recall that from Equation (1) :

$$\frac{1+r}{2r} = \mathbb{E}[\max(\frac{B}{\mu}, 1)] = \mathbb{E}[\frac{B}{\mu}\mathbf{1}_{B>\mu}] + \mathbb{P}[B \leq \mu].$$

this implies that

$$\frac{1+r}{r} - 1 = \mathbb{E}\left[\frac{|B|}{\mu}\mathbf{1}_{|B|>\mu}\right] + \mathbb{P}[|B| \leq \mu].$$

Thus we conclude that

$$\sigma_{tr}^2 = \frac{\mathbb{E}[B^2]\mu}{\mathbb{E}[|B|\mathbf{1}_{|B|>\mu}]}.$$

III.A.4 Proof of Proposition 1

We just give a sketch of proof here since the computations are essentially the same as for the proof of Proposition 1. In particular, we do not introduce the volume ε of limit orders and directly work in the asymptotic regime ε tending to zero. We consider the gain of passive sell orders. The gain of passive buy orders can be deduced the way.

First, we compute the gain of a new order placed at the i^{th} limit when the trade is initiated by an informed trader, knowing that $Q^i > L^d(i)$, denoted by $G^d_{inf}(i)$:

$$G^d_{inf}(i) = d + (i-1)\alpha - \mathbb{E}[B|B > d + (i-1)\alpha].$$

Second, we compute the gain of a new order placed at the i^{th} limit when the trade is initiated by a noise trade, knowing that $Q^i > L^d(i)$, denoted by $G^d_{noise}(i)$:

$$G^d_{noise}(i) = d + (i-1)\alpha.$$

Now, $G^d(i)$ satisfies:

$$\begin{aligned} G^d(i) &= G^d_{inf}(i)\mathbb{P}[v=1|Q > L^d(i)] + G^d_{noise}(i)\mathbb{P}[v=0|Q > L^d(i)] \\ &= d + (i-1)\alpha - \frac{r\mathbb{E}[B\mathbf{1}_{B>d+(i-1)\alpha}]}{\mathbb{P}[Q > L(d+(i-1)\alpha])} \\ &= d + (i-1)\alpha - \frac{r\mathbb{E}[B\mathbf{1}_{B>d+(i-1)\alpha}]}{r\mathbb{P}[B > d+(i-1)\alpha] + (1-r)\mathbb{P}[Q^u > L(d+(i-1)\alpha)]}. \end{aligned}$$

III.A.5 Proof of Theorem 1

We consider the ask side. First we show that the spread is positive and finite. Then we prove that beyond the spread, market makers insert limit orders on all possible limit prices.

We showed in the case where the tick size is null that there exists μ such that for all $x \leq \mu, L(x) = 0$ and for all $x > \mu, L(x) > 0$. The LOB being now discrete, the previous findings remain true for k^d_r instead of μ where k^d_r satisfies:

$$k^d_r = \min\{k \in \mathbb{N}^+ | d + (k-1)\alpha > \mu\}.$$

So we have:

$$k^d_r = 1 + \lceil \frac{\mu - d}{\alpha} \rceil.$$

Similarly, for the first non-empty limit at the bid side, we get:

$$k^d_l = \lceil \frac{\mu + d}{\alpha} \rceil.$$

From Equation (4), the spread is equal to $(k^d_r + k^d_l)\alpha - \alpha$. Thus the conditional constrained bid-ask spread ϕ^d_α, given the value of d, satisfies:

$$\phi^d_\alpha = \alpha(\lceil \frac{\mu - d}{\alpha} \rceil + \lceil \frac{\mu + d}{\alpha} \rceil).$$

Under Assumption 2, we have for any $i \geq k^d_r$:

$$L(d + (i-1)\alpha) = F^{-1}_{\kappa^u}(\frac{1}{1-r} - \frac{r}{1-r}\mathbb{E}[\max(\frac{B}{d+(i-1)\alpha}, 1)]).$$

We deduce that the cumulative LOB is unique and increasing beyond the spread.

III.A.6 Proof of Corollary 1

The parameter d being approximately uniformly distributed between $[0, \alpha)$, we can compute the average value of the constrained bid-ask spread by integrating ϕ_α^d:

$$\phi_\alpha = \int_0^\alpha \lceil \frac{\mu - s}{\alpha} \rceil + \lceil \frac{\mu + s}{\alpha} \rceil \mathrm{d}s.$$

Denote $u := \frac{\mu}{\alpha}$. We have:

$$\phi_u = \alpha \int_0^1 \lceil u - x \rceil + \lceil u + x \rceil \mathrm{d}x.$$

We decompose u such that $u = u_i + u_f$, where u_i represents the integer part of u. We get:

$$\phi_\alpha = \alpha \int_0^1 \lceil u_i + u_f - x \rceil + \lceil u_i + u_f + x \rceil \mathrm{d}x.$$

$$\phi_\alpha = \alpha \Big(\int_0^{u_f} (u_i + 1) \mathrm{d}x + \int_{u_f}^1 u_i \mathrm{d}x + \int_0^{1-u_f} (u_i + 1) \mathrm{d}x + \int_{(1-u_f)}^1 (u_i + 2) \mathrm{d}x \Big).$$

$$\phi_\alpha = \alpha \big(u_f (u_i + 1) + (1 - u_f) u_i + (1 - u_f)(u_i + 1) + u_f (u_i + 2) \big).$$

$$\phi_\alpha = \alpha (2u_i + 2u_f + 1) = \alpha + 2\mu = \alpha + \phi.$$

CHAPTER IV

From asymptotic properties of general point processes to the ranking of financial agents

Abstract

We propose a general non-linear order book model that is built from the individual behaviours of the agents. Our framework encompasses Markovian and Hawkes based models. Under mild assumptions, we prove original results on the ergodicity and diffusivity of such system. Then we provide closed form formulas for various quantities of interest: stationary distribution of the best bid and ask quantities, spread, liquidity fluctuations and price volatility. These formulas are expressed in terms of individual order flows of market participants. Our approach enables us to establish a ranking methodology for the market makers with respect to the quality of their trading.

Keywords: Market microstructure, limit order book, high-frequency trading, market making, queuing model, Hawkes processes, ergodic properties, volatility, regulation.

1 Introduction

In the last two decades, the development of electronic and fragmented markets has lead to a deep disruption in the landscape of market participants. In particular, traditional market making institutions have been largely replaced by high-frequency market makers. Market makers are intermediaries between buyers and sellers. In an electronic limit order book, they provide liquidity to market participants willing to trade immediately by simultaneously posting limit orders on both sides of the book. Market makers undergo different types of risk, mainly adverse selection and inventory risks. To avoid adverse selection risk, they must be able to update very frequently their quotes in response to other order submissions or cancellations. To minimise their inventory risk, they need to use smart algorithms enabling them to hold positions for very short time periods only, see for example [84].

High-frequency traders (HFTs) are now the only market participants that are indeed able to play the role of market makers on liquid stocks, see [71]. This is achieved thanks to an intense use of speed (co-location) and technology. They are supposedly capable to maintain a strong presence at best price limits and control adverse selection at the same time, see [72], while operating efficient inventory management in an increasingly fast-moving market, see [5, 11]. This is to the extent that HFTs are described as the new market makers in [88].

Since the arrival of these new market makers, academics, regulators and practitioners aim at understanding whether their activity is harmful or beneficial for markets. On the one hand, some argue that HFTs have a positive impact on markets: the competition between market makers leads to an increase in market depth, to narrower bid-ask spreads which is equivalent to reduced trading costs for other investors, see [58, 72] and to better price discovery, see [58, 98]. On the other hand, others assert that high-frequency market makers have toxic consequences. For example, they worsen market volatility during flash crashes by aggressively liquidating their long positions, see [74, 85].

One important common point in most studies analysing the behaviour of HFTs is that they try to measure how HFTs impact the market as a group, without investigating individual behavioural disparities among them. Chapters I and II shed light on the fact that all HFTs do not behave similarly, showing for example that they have very different levels of aggressiveness and liquidity provision. In this paper, we wish to participate to the debate about the role of HFTs on market quality by bringing some new quantitative elements enabling regulators and exchanges to assess the individual effects of each high-frequency market maker operating on the market. In particular, we want to be able to rank market makers according to the quality of their trading.

We use several metrics for market quality such as spread and liquidity fluctuations, but a particular focus is given to the price volatility. This idea of disentangling market participants contribution to volatility is used in [97]. In this work, the authors nicely model the interactions between the various orders of the different market participants using linear Hawkes processes. This model is very interpretable: an order of type A of Agent i raises the likelihood of an order of type B of Agent j by a certain amount. Consequently, the authors naturally define the contribution of Agent A to the volatility by the weighted sum over all possible types of orders of Agent A of the squared mean price jump triggered by each of these orders, the associated weight being the intensity of the corresponding order type.

Our focus here is on market makers. Thus one crucial element to take into account is the well-known fact that the main market driver of any market making strategy is the state of the limit order book (and not single individual orders of other market participants), see [63, 81, 91]. Therefore, in the spirit of the Queue-reactive model of [63], we assume that the state of the order book, which is a common component, affects the interactions between our high-frequency market participants. However, to get a really accurate modelling of the behaviour of the agents, we also let their individual actions depend on their own past ones and

on those of other participants, in the spirit of [97]. We allow for strong non-linearities in the dependences with the past, leading to a much generalised version of Hawkes-Queue-reactive type order book models, see [92, 114].

In this extended and non-Markovian framework, we are able to prove the ergodicity and diffusivity of our system, see [65] for inspiring ideas. Furthermore, we provide asymptotic expressions for market quantities such as spread, liquidity fluctuations or price volatility in terms of the individual order flows of market participants. This notably enables us to forecast the dynamics of the market in case one market makers leaves it. The idea is that we consider that market makers interact with the market through their algorithms which are specified for example in term of average event size or in term of relative quantities such as the imbalance. If we remove one market participant while the others do not modify their algorithms, we can for instance compute a new volatility. If it is larger (smaller) than the actual one, we can say that the considered market maker has a stabilising (destabilising) effect on the market. This eventually leads us to a ranking of market makers with respect to the quality of their trading.

Let us now give a brief description of our model. Let n be a positive integer representing the index of the n-th order book event e_n. Each event e_n happens at time T_n and is characterised by a variable X_n that encodes all the needed information to describe e_n. For example, X_n contains the order size, the type of the order (limit order, liquidity consuming order such as market order or cancellation), the order posting price and the identity of the agent. A detailed description of the sequences $(T_n)_{n\geq 1}$ and $(X_n)_{n\geq 1}$ is given in Section 2.2. The order book state is modelled by the process $U_n = \big(Q_n^1, Q_n^2, S_n\big)$ with Q_n^1 the available quantity at the best bid, Q_n^2 the available quantity at the best ask and S_n the spread at time T_n. For a detailed description of the dynamic of U_n, see Equation (1). Here we focus on the first limits to reduce the dimension of the state space and keep a tractable model[1]. Finally, we use a general approach to infer the behaviour of the price process from that of (U_n), in the spirit of [65, 82], see Section 4 for the detailed formulation. We define the non-linear Hawkes-Markovian arrival rate $\lambda_t(e)$ of an order book event e (e containing the identity of the involved agent) at time $t \in \mathbb{R}_+$ as follows:

$$\lambda_t(e) = \psi\big(e, U_{t^-}, t, \sum_{T_i<t} \phi(e, U_{t^-}, t - T_i, X_i)\big),$$

where ψ is a non-linear function, U_{t^-} is the order book state relative to the last event before t and ϕ is the Hawkes kernel representing the influence of past events. The functions ϕ and ψ are both $\mathbb{R}_+$-valued. In absence of the kernel ϕ, the function ψ leads to a classical Markovian approach since the arrival rate of an event e depends essentially on the order book state U_{t^-}. When ϕ is non-zero, ψ controls the interactions between the past events and the current order book state. Note that we allow ψ to have a polynomial growth while in the literature, it has at most a linear growth, see [26]. Additionally, we do not impose ψ and ϕ to be continuous, which means that a sudden change of regime in the order book dynamic is also incorporated in our modelling. Finally, we propose an agent-based model since market

[1]However, we can model deeper limits by enlarging the dimension of the state space.

participant identities are contained in the order book events e through the variables $(X_i)_{i\geq 1}$.

Our framework is a generalised order book model where the arrival rate of the events follows a non-linear Hawkes-type dynamic that depends on the order book state. This approach covers most existing bid-ask order book models. It is a natural extension of the Poisson intensity models, see [1, 106], the Markovian Queue-reactive model introduced in [63] and the Hawkes based models such as [2, 92, 97]. In this setting, under mild assumptions, we provide new ergodic results and limit theorems, expressing all the limiting quantities in terms of the individual flows of market participants. Furthermore, we build an estimation methodology for the intensity functions which turns out to be similar to the one used in the Queue-reactive case, see [65], although the model here is much more general and non-Markovian. These theoretical results for our point processes, which largely extend classical ergodicity properties limited to the Markov case, are the basis for the assessment of the role of the different market participants on market quality as explained above.

The paper is organised as follows. First, we introduce in Section 2 our order book model and describe how to recover market dynamics from the individual behaviours of each agent. Then, we prove the ergodicity of our system in Section 3 and its diffusivity in Section 4. In Section 5, we provide the needed formulas to compute the order book stationary distribution, the price volatility and the liquidity fluctuations. Finally, numerical results and ranking of market makers on several assets are provided in Section 6. Proofs and additional results are relegated to an appendix.

2 Market modelling

In this section, we describe the order book model and show how to recover the market dynamics given the agents individual behaviours.

2.1 Introduction to the model

In the order book mechanism buyers and sellers send their orders to a continuous-time double auction system. Market participants orders have a specific size that is measured in average event size (AES)[2] and the orders can be sent to different price levels that are separated by a minimum distance which is the tick size. In our model, we only consider the price levels between the best bid and ask prices to reduce the dimension of the state space. Additionally, we assume that the agents can take three elementary decisions:

- Insert a limit order of a specific size at the best bid or ask price, hoping to get an execution.
- Insert a buying or selling limit order of a specific size within the spread.

[2] AES is the average size of events observed in the limit order book.

- Send a liquidity consuming order of a specific size at the best bid or ask price. Cancellation and market orders have the same effect on liquidity. Thus, they are aggregated to constitute the liquidity consumption orders.

The size of the orders is not constant in the model. Finally, the mid price moves in a fixed grid separated by the tick. A simple example is to consider the case where the mid price decreases (resp. increases) by one tick when the best bid (resp. ask) is totally depleted. Here, the mid price jumps size may be larger than one tick. In the rest of the article, we take the mid price as our reference price for simplification. The dynamic of the model is illustrated in Figure IV.1.

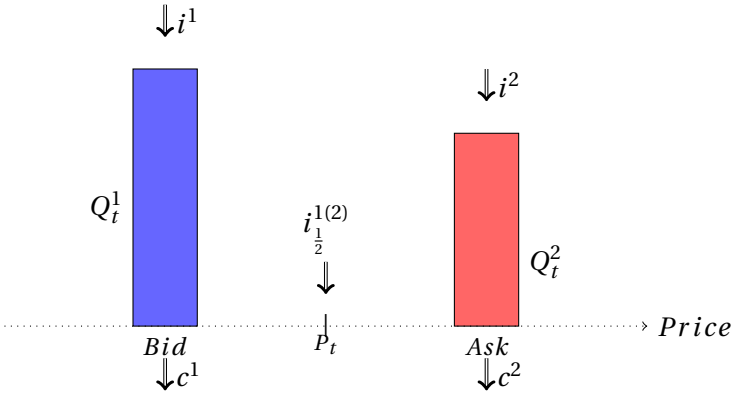


Figure IV.1 – Diagram of flows affecting our order book model. The quantity i^1 (resp. i^2) represents the insertion of limit orders at the best bid (resp. ask). The quantity $i^1_{\frac{1}{2}}$ (resp. $i^2_{\frac{1}{2}}$) is associated to buying (resp. selling) limit orders within the spread. The quantities c^1 and c^2 refer to the orders that consume respectively the liquidity at the best bid and ask.

Notations. We consider the following notations:

- The current physical time is t.
- The mid price is P_t, the best best bid price is P_t^1 and the best ask price is P_t^2.
- The spread is $S_t = \dfrac{P_t^2 - P_t^1}{2}$ and α_0 is the tick size.
- The available quantity at the best bid (resp. ask) is Q_t^1 (resp. Q_t^2).

2.2 Order book dynamic

Let $(\Omega, \mathscr{F})$ be a measurable space and $(T_n)_{n\geq 1}$ a non-decreasing sequence of random variables such that $T_n < T_{n+1}$ on the event $\{T_n < \infty\}$. We associate to each T_n a random variable X_n taking its value on a measurable space $(E, \mathscr{E})$. In our case, T_n are the times when events

happen in the order book and X_n are variables describing each event. We endow Ω with the filtration $(\mathscr{F}_t)_{t\geq 0}$ defined such that $\mathscr{F}_t = \sigma(\{T_n \in C\} \times \{X_n \in B\}, C \in \mathscr{B}(\mathbb{R}) \cap (-\infty, t], B \in \mathscr{E})$. Each event is characterised by:

- **The size of the order**: is an integer representing the minimum quantity that can be inserted in the order book[3].
- **The price of the order**: equals to $k \in \mathbb{N}$ when the order is inserted at the price $P^1 + k\alpha_0$.
- **The direction of the order**: equals to $+1$ if it provides liquidity and -1 when liquidity is removed.
- **The type of the order**: equals to 1 (resp. 2) when it modifies the bid (resp. ask)[4] side.
- **The identity of the agent**: is valued in $\{1, \dots, N\}$ since the market consists in N agents.

Since we track only the first limits, we add the following variables to describe the new order book state when one of these limits is depleted: $\tilde{Q}^1$ (resp. $\tilde{Q}^2$) the new bid (resp. ask) queue and $\tilde{S}$ the new spread after a depletion. Note that when there is no depletion, the random vector $(\tilde{Q}^1, \tilde{Q}^2, \tilde{S})$ is arbitrary[5] and its values are not used. Finally, we record the order book state after an event to add a dependence between the arrival rate of the events and the past order book states. The order book dynamic is described below. Hence, we consider the following form for $E = \bar{\mathbb{N}} \times \mathbb{T} \times \mathbb{S} \times \mathbb{B} \times \tilde{\mathbb{U}} \times \mathbb{U} \times \mathbb{A}$ with:

- $\bar{\mathbb{N}} = \mathbb{N}^*$: the set where the orders size is valued.
- $\mathbb{T} = \mathbb{N}$: the set where the price levels are valued.
- $\mathbb{S} = \{+1, -1\}$: the set where the orders direction is valued.
- $\mathbb{B} = \{1, 2\}$: the set where the orders type is valued.
- $\tilde{\mathbb{U}} = \left\{\mathbb{N}^2 \times \alpha_0\mathbb{N}\right\} \setminus \mathbb{U}^0$: the set where the order book states after a depletion are valued [6].
- $\mathbb{U} = \left\{\mathbb{N}^2 \times \alpha_0\mathbb{N}\right\} \setminus \mathbb{U}^0$: the set where the order book states after an event are valued.
- $\mathbb{A} = \{1, \dots, N\}$: the set where the agents identity is valued.
- $\mathbb{U}^0 = \{0\}^2 \times \alpha_0\mathbb{N}$: the set of unreachable order book states.

Example 1. *We place ourselves in the case where the minimum order size is a quarter of the AES and* $(\tilde{Q}^1, \tilde{Q}^2, \tilde{S}) = c$ *when there is no depletion with* c *is a fixed constant. Thus, a buy limit order of size* 0.5 *AES inserted at the best bid price* $+1$ *tick by the agent* 5 *when the best bid size is* $Q_i^1 = 1$ *AES, the best ask size is* $Q_i^2 = 3$ *AES and the spread* $S = 2$ *ticks is represented by the event* $e = (2, 1, +1, 1, c, u, 5)$ *with* $u = (2, 12, 1)$.

[3] In practice, the minimum quantity can be taken as a quarter of the the average event size (AES).

[4] A buy (sell) limit order within the spread, a liquidity consumption at the bid (resp. ask) or a limit order at the bid (resp. ask) modify the bid side first.

[5] To fix the ideas we can take $(\tilde{Q}^1, \tilde{Q}^2, \tilde{S}) = c$ with c a fixed constant when there is no depletion.

[6] The state where the best bid or ask size is zero is fictitious state that allow us to model the price changes, see Remark 18.

Order book dynamic. The order book state is modelled by the process $U_t = \big(Q_t^1, Q_t^2, S_t\big)$ where Q_t^1 (resp. Q_t^2) is the best bid (resp. ask) quantity and S_t is the spread. The dynamic of the reference price is going to be deduced from the one of the process $(U_t)_{t\geq 0}$, see Section 4. The process U_t is defined in the following way:

$$U_t = \sum_{T_i<t} \Delta U_i, \quad \Delta U_i = U_i - U_{i-1},$$

with $U_i = (Q_i^1, Q_i^2, S_i) \in \mathbb{U}$ the order book state after the i-th event (we write U_i for U_{T_i} when no confusion is possible). Thus, we only need to describe the variables $(U_i)_{i\geq 1}$. Let $i \geq 1$ and $X_i = (n_i, t_i, s_i, b_i, \tilde{U}_i, U_i, a_i) \in E$ with $n_i \in \bar{\mathbb{N}}$, $t_i \in \mathbb{T}$, $s_i \in \mathbb{S}$, $b_i \in \mathbb{B}$, $\tilde{U}_i = (\tilde{Q}_i^1, \tilde{Q}_i^2, \tilde{S}_i) \in \mathbb{U}$, $U_i = (Q_i^1, Q_i^2, S_i) \in \mathbb{U}$ and $a_i \in \mathbb{A}$. The variable U_i satisfies

$$\begin{aligned} S_i &= \mathbf{1}_{\epsilon_i=0} S_{i-1} - (t_i^1 + t_i^2) + \mathbf{1}_{\epsilon_i=1} \tilde{S}_i, \\ Q_i^1 &= \mathbf{1}_{\epsilon_i=0} Q_{i-1}^1 + (n_i^{1,+} - n_i^{1,-} + n_i^{1,\frac{1}{2}}) + \mathbf{1}_{\epsilon_i=1} \tilde{Q}_i^1, \\ Q_i^2 &= \mathbf{1}_{\epsilon_i=0} Q_{i-1}^2 + (n_i^{2,+} - n_i^{2,-} + n_i^{2,\frac{1}{2}}) + \mathbf{1}_{\epsilon_i=1} \tilde{Q}_i^2, \end{aligned} \tag{1}$$

where ϵ_i is a price move indicator (i.e $\epsilon = 0$ when there is no depletion and $\epsilon = 1$ otherwise), the variable t_i^1 (resp. t_i^2) is the spread variation when a buy (resp. sell) limit order is inserted within the spread. The variables $n_i^{1,+}$ (resp. $n_i^{2,+}$), $n_i^{1,-}$ (resp. $n_i^{2,+}$) and $n_i^{1,\frac{1}{2}}$ (resp. $n_i^{2,\frac{1}{2}}$) are respectively the best bid (resp. ask) increments when a buy limit order is inserted at the best bid (resp. ask), when a consumption order is sent at the best bid (resp. ask) and when a buy (resp. sell) limit order is inserted within the spread. We now explain how the previous quantities can be written in terms of the state variables:

$$\begin{aligned} \epsilon_i &= \mathbf{1}_{\{s_i=-1\}\cap\left(\{b_i=1, n_i\geq Q_{i-1}^1\}\cup\{b_i=2, n_i\geq Q_{i-1}^2\}\right)}, \\ t_i^1 &= \min(t_i\alpha_0, S_{i-1} - \alpha_0)\mathbf{1}_{\{b_i=1,\, t_i\neq 0\}}, \\ t_i^2 &= (S_{i-1} - t_i\alpha_0)_+ \mathbf{1}_{\{b_i=2,\, t_i\neq \frac{S_{i-1}}{\alpha_0}\}}, \\ n_i^{1(2),+} &= n_i \mathbf{1}_{\{s_i=+1, t_i=0(\frac{S_{i-1}}{\alpha_0}), b_i=1(2)\}}, \\ n_i^{1(2),-} &= n_i \mathbf{1}_{\{s_i=-1, t_i=0(\frac{S_{i-1}}{\alpha_0}), b_i=1(2), n_i<Q_{i-1}^{1(2)}\}}, \\ n_i^{1(2),1/2} &= n_i \mathbf{1}_{\{s_i=+1, t_i\notin\{0, \frac{S_{i-1}}{\alpha_0}\}, b_i=1(2)\}}. \end{aligned}$$

We denote by λ_t the intensity of the point process (T_n, X_n). For $e \in E$, $\lambda_t(e)$ corresponds to the arrival rate of an event of type e conditional on the past history of the process and it is defined as

$$\lambda_t(e) = \lim_{\delta t\to 0} \frac{\mathbb{P}\big[\#\{T_n \in (t, t+\delta t], X_n = e\} \geq 1 | \mathscr{F}_t\big]}{\delta t},$$

with $\#A$ is the cardinality of the set A. We consider the following expression for the intensity:

$$\lambda_t(e) = \psi\big(e, U_{t^-}, t, \sum_{T_i<t} \phi(e, U_{t^-}, t - T_i, X_i)\big), \tag{2}$$

where ψ and ϕ are $\mathbb{R}_+$-valued functions. The individual behaviour of each agent is encoded in the functions ψ and ϕ through e and $(X_i)_{i\geq 1}$, see Equation (2).

Note that we can recover the full definition of the intensity of the process $N = (T_n, X_n)$ using the following proposition:

Proposition 1. *For any $B \in \mathscr{E}$ and $t \in \mathbb{R}_+$, we have*

$$\lim_{\delta t \to 0} \frac{\mathbb{P}\left[\#\{T_n \in (t, t+\delta t], X_n \in B\} \geq 1 | \mathscr{F}_t\right]}{\delta t} = \sum_{e \in B} \lambda_t(e). \tag{3}$$

The proof of Proposition 1 is given in Appendix IV.A. The existence and the uniqueness of a probability measure $\mathbb{P}$ on the filtered probability space $(\Omega, \mathscr{F}, \mathscr{F}_t)$ such that (3) is satisfied and λ_t verifies Equation (2) is ensured as soon as $\sum_{e \in E} \lambda_t(e)$ is locally integrable, see [68]. We prove that $\sum_{e \in E} \lambda_t(e)$ is locally integrable in Appendix IV.C.

2.3 Market reconstitution

We can recover the market intensity λ_t^M using the corollary below.

Corollary 1. *When λ_t verifies Equation (2), the market intensity $\lambda_t^M(e')$ of an event e' (e' does not contain the identity of the agent) in the exchange is given by*

$$\lambda_t^M(e') = \lim_{\delta t \to 0} \frac{\mathbb{P}\left[\#\{T_n \in (t, t+\delta t], X_n \in (e', \mathbb{A})\} \geq 1 | \mathscr{F}_t\right]}{\delta t} = \sum_{a \in \mathbb{A}} \lambda_t((e', a)), \tag{4}$$

for any $e' \in E' = \bar{\mathbb{N}} \times \mathbb{T} \times \mathbb{S} \times \mathbb{B} \times \tilde{\mathbb{U}} \times \mathbb{U}$.

The proof of Corollary 1 is a consequence of Proposition 1.

2.4 Some specific models

Poisson intensity. We introduce here a simple version of the Poisson intensity model where the variable $X_n = (n_n, t_n^o, s_n, b_n, \tilde{U}_n, U_n, a_n)$ with $U_n = (Q_n^1, Q_n^2, S_n)$ satisfies

- the order size $n_n = 1$: all the events have the same size 1 AES.
- the price level $t_n^o \in \{0, \frac{S_n}{\alpha_0}\}$: orders are inserted at the best bid or ask.
- the law of $\tilde{U}_n$ is unchanged: when one limit is depleted, the new state is drawn from the stationary distribution of the order book.

For any $e = (n, t^o, s, b, \tilde{u}, u, a) \in E$ with $u = (Q^1, Q^2, S)$, we can recover Poisson models by taking the following choice of the parameters:

$$\psi(e, u, t, z) = \tilde{h}(s, b, a) \mathbf{1}_{n=1, t^o \in \{0, \frac{S}{\alpha_0}\}}, \quad \forall z, t \in \mathbb{R}_+,$$

with $\tilde{h}$ a deterministic function valued on $\mathbb{R}_+$. Thus, the expression of the intensity becomes

$$\lambda_t(e) = \tilde{h}(s, b, a) \mathbf{1}_{n=1, t^o \in \{0, \frac{S}{\alpha_0}\}}.$$

Such modelling was introduced in [1, 38, 106].

Queue-reactive intensity. In the Queue-reactive model, the arrival rate of the events depends only on the current order book state. For any $e \in E$ and $u \in U$, we take

$$\psi(e, u, t, z) = \tilde{h}(e, u), \quad \forall z, t \in \mathbb{R}_+,$$

to reproduce the Queue-reactive dynamic with $\tilde{h}$ a deterministic function valued on $\mathbb{R}_+$. Hence, the intensity reads

$$\lambda_t(e) = \tilde{h}(e, u).$$

Such modelling was studied in [63, 65].

Hawkes Queue-reactive intensity. In the Hawkes framework, the arrival rate of each event depends fully on all the past market events. For any $e \in E$ and $u \in U$, we generate the Hawkes Queue-reactive dynamic by taking

$$\psi(e, u, t, z) = h(e, u, t) + z, \quad \forall z, t \in \mathbb{R}_+.$$

Thus intensity has the following expression

$$\lambda_t(e) = h(e, U_{t^-}, t) + \sum_{T_i < t} \phi(e, U_{t^-}, t - T_i, X_i).$$

Close modelling was used [2, 10, 69, 92, 97].

Quadratic Hawkes process. The quadratic Hawkes processes generalise the linear Hawkes processes by adding an interaction term between the pairs of past events. In the classical one-dimensional case, the intensity function of a quadratic Hawkes process reads

$$\lambda_t(e) = h(t) + \sum_{T_i < t} \phi(t - T_i) + \sum_{T_i, T_i' < t} K(t - T_i, t - T_i'),$$

with $K : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ the quadratic kernel. We can recover a simple case of the quadratic Hawkes models when K is separable (i.e $K(t, s) = k(t)k(s)$ with k a non negative function) by taking ψ of the following form:

$$\psi(e, u, t, z) = h(e, u, t) + z^2, \quad \forall z, t \in \mathbb{R}_+.$$

Hence, the expression of the intensity becomes

$$\lambda_t(e) = h(e, U_{t^-}, t) + \sum_{T_i < t} \phi^2(e, U_{t^-}, t - T_i, X_i) + \sum_{T_i, T_i' < t} \phi(e, U_{t^-}, t - T_i, X_i)\phi(e, U_{t^-}, t - T_i', X_i').$$

Quadratic Hawkes models were introduced in [22, 95].

Remark 1. *In our modelling, the linear term is necessarily ϕ^2. However, to overcome this limitation we can add a new argument to the function ψ which differentiates the linear kernel from the quadratic one. This will not modify the proofs.*

3 Ergodicity

3.1 Notations and definitions

Let Z_t be a process defined on the probability space $(\Omega, \mathscr{F}, \mathscr{F}_t, \mathbb{P})$ and valued in $(W_0, \mathscr{W}_0)$. We consider another process V_t defined on $(W_0, \mathscr{W}_0)$ and valued in $(X, \mathscr{X})$ and we denote by $P_t(x,.)$ the probability distribution of $V_t^{0,x}$ starting at 0 with the initial condition $x \in W_0$. For any measure μ defined on $(W_0, \mathscr{W}_0)$ viewed as a random starting condition, we denote by $P_t(\mu,.) = \int_{W_0} P_t(x,.)\mu(dx)$.

Definition 1 (Invariant distribution). *The measure μ is invariant if the probability distribution $P_t(\mu,.)$ does not depend on the time t.*

This definition is consistent with the one given in [26, 55, 90]. The process V_t starting with the initial distribution μ is stationary if and only if μ is invariant. We define the total variation norm between two measures π and π' such that $||\pi - \pi'||_{TV} = \sup_{A \in \mathscr{X}} |\pi(A) - \pi'(A)|$.

Definition 2 (Ergodicity). *Let $C \in \mathscr{W}_0$. The process V_t is C-ergodic if for any $x \in C$ there exists an invariant measure μ such that $P_t(x,.) \underset{t\to\infty}{\to} P_0(\mu,.)$ in total variation.*

Remark 2. *This definition is consistent with the one given in [90]. Ergodicity is interesting since it ensures the convergence of the order book process U_t towards an invariant probability distribution. Thus the stylized facts observed on market data can be explained by a law of large numbers type phenomenon for this invariant distribution.*

Remark 3. *In this Section, we work with a continuous time processes Z_t and V_t with $t \in \mathbb{R}_+$. However, all the definitions are similar for a discrete time processes Z_n and V_n with $n \in \mathbb{N}$. We just have to replace t by n in the definitions above.*

The space Ω and the filtration $\mathscr{F}_t$ considered here are defined in Section 2.2, $\mathscr{F} = \mathscr{F}_\infty$, the filtered space W_0 is the space of sequences indexed by $\mathbb{N}^-$ and valued on $\mathbb{R}_+ \times E$, $X = \mathbb{U} \times (\mathbb{R}_+)^E$ and $\mathscr{X} = \mathscr{U} \times \mathscr{B}(\mathbb{R}_+)^{\otimes E}$ with $\mathscr{U}$ the σ-algebra generated by the discrete topology on $\mathbb{U}$, $\mathscr{B}(\mathbb{R}_+)^{\otimes E}$ the cylinder σ-algebra for $(\mathbb{R}_+)^E$, $\mathscr{B}(\mathbb{R}_+)$ the borel σ-algebra of $\mathbb{R}_+$ and $\mathscr{W}_0 = \left(\mathscr{B}(\mathbb{R}_+) \times \mathscr{E}\right)^{\otimes \mathbb{N}^-}$ with $\mathscr{E}$ the σ-algebra generated by the discrete topology on E. We need to work on the functional space W_0 since the dynamic of the process depend on its whole past.

3.2 Ergodicity

In this section, we provide under general assumptions a theoretical result on the ergodicity of the process $\bar{U}_t = (Q_t^1, Q_t^2, S_t, \lambda_t)$ with λ_t the intensity defined by (2).

We denote by $\lambda_Q^{i,+}$ (resp. $\lambda_Q^{i,-}$) and λ_S^+ (resp. λ_S^-) the arrival rate of the events that respectively increase (resp. decrease) the limit Q^i and the spread S for any $i \in \mathbb{B}$. Let $U_t = (Q_t^1, Q_t^2, S_t)$ be

the order book process and $e \in E$ be a market event, the quantities $\lambda_Q^{i,\pm}$ and $\lambda_S^{\pm}$ are defined by the following formulas:

$$\lambda_Q^{i,\pm}(U_{t^-}, n) = \sum_{e \in E_Q^{i,\pm}(U_{t^-}, n)} \lambda_t(e), \quad \lambda_S^{\pm}(U_{t^-}, k) = \sum_{e' \in E_S^{\pm}(U_{t^-}, k)} \lambda_t(e), \tag{5}$$

with $n \in \mathbb{N}$, $k \in \mathbb{N}$ and

$$\begin{array}{rcl} E_Q^{i,\pm}(U_{t^-}, n) & = & \{e \in E;\ s.t \quad \Delta Q_t^i = \pm n\}, \\ E_S^{\pm}(U_{t^-}, k) & = & \{e \in E;\ s.t \quad \Delta S_t = \pm k\}, \end{array} \tag{6}$$

with $\Delta X_t = X_t - X_{t^-}$ for any process X_t. For simplicity and since there is no ambiguity, we do not write the dependence of $\lambda_Q^{i,\pm}$ and $\lambda_S^{\pm}$ on the current time t. For any $n \in \mathbb{N}^*$, we write

$$\mathscr{P}(n) = \{\mathbf{k}_m = \{k_1, \dots, k_m\} \in (\mathbb{N}^*)^m;\quad s.t \quad k_1 + \dots + k_m = n, \quad m \in \mathbb{N}^*\},$$

for the set containing all the partitions of n.

Assumption 1 (ψ growth). *We assume that there exist $c \geq 0$, $d \geq 0$ and $n_\psi \in \mathbb{N}$ such that*

$$\begin{array}{l} \tilde{\psi}(e, z) \leq c(e) + d(e) z^{n_\psi}, \\ \sup_{e \in E} \left\{ d(e) \sum_{\mathbf{k}_m \in \mathscr{P}(n_\psi)} \binom{n_\psi}{\mathbf{k}_m} \int_{\mathbb{R}_+^m} \prod_{i=1}^m \phi^{*\,k_i}(e, s_i)\, ds_i \right\} < 1, \end{array}$$

with $\tilde{\psi}(e, z) = \sup_{(u,t) \in \mathbb{U} \times \mathbb{R}_+} \psi(e, u, t, z)$, $\phi^*(e, s) = \sup_{u \in \mathbb{U}} \sum_{x \in E} \phi(e, u, s, x)$ and $\binom{n_\psi}{\mathbf{k}_m} = \binom{n_\psi}{k_1, \dots, k_m} = \frac{n_\psi!}{k_1! \dots k_m!}$.

Assumption 1 is natural. To see this, we take a 1-d stationary non-linear Hawkes process N_t with an intensity λ_t that verifies

$$\lambda_t = c + d\Big(\sum_{T_i < t} \phi(t - T_i)\Big)^{n_\psi} = c + d\Big(\int_{-\infty}^t \phi(t - s) dN_s\Big)^{n_\psi}, \quad \forall\, t \in \mathbb{R}_+.$$

By stationarity, we have

$$\begin{aligned} \bar{\lambda} = \mathbb{E}[\lambda_t] &= c + d\mathbb{E}\Big[\Big(\int_{-\infty}^t \phi(t - s) dN_s\Big)^{n_\psi}\Big] \\ &= c + d\left\{ \sum_{\mathbf{k}_m \in \mathscr{P}(n_\psi)} \binom{n_\psi}{\mathbf{k}_m} \int_{(-\infty, t)^m} \prod_{i=1}^m \phi^{k_i}(t - s_i)\, \mathbb{E}[dN_{s_1} \dots dN_{s_m}] \right\}, \end{aligned}$$

with $\binom{n_\psi}{\mathbf{k}_m}$ an enumeration factor. In fact, if we have n_ψ possible events divided in m groups such that the j-th group is composed of k_j events, then the quantity $\binom{n_\psi}{\mathbf{k}_m}$ counts the number of possible groups. Here each group represents the jumps that happen at the same time. Since the jumps have a unit size, the Brascamp-Lieb inequality ensures that $\mathbb{E}[dN_{s_1} \dots dN_{s_m}] \leq \prod_{i=1}^m \mathbb{E}[dN_{s_i}^m]^{1/m} = \prod_{i=1}^m \mathbb{E}[dN_{s_i}]^{1/m} = \prod_{i=1}^m \mathbb{E}[\lambda_{s_i}]^{1/m} = \bar{\lambda}$ which leads to

$$\bar{\lambda} \leq c + q\bar{\lambda},$$

with $q = d \sum_{\mathbf{k}_m \in \mathscr{P}(n_\psi)} \binom{n_\psi}{\mathbf{k}_m} \int_{(\mathbb{R}_+)^m} \prod_{i=1}^m \phi^{k_i}(e, s_i)\, ds_i$. The condition $q < 1$ of Assumption 1 guarantees that $\bar{\lambda}$ is finite.

Remark 4. *Non linear Hawkes process are studied mainly when the function ψ admits at most a linear growth (i.e $n_\psi \leq 1$). When $n_\psi = 1$, we recover the classical condition*

$$\sup_{e\in E} d(e)\left\{\int_{\mathbb{R}_+} \phi^*(e,s)ds\right\} < 1.$$

When $n_\psi = 2$, Assumption 1 becomes

$$\sup_{e\in E} d(e)\left\{\Big(\int_{\mathbb{R}_+} \phi^*(e,s)ds\Big)^2 + \int_{\mathbb{R}_+} \phi^*(e,s)^2 ds\right\} < 1.$$

Assumption 2 (Negative drift). *There exist positive constants C_{bound}, $z_0 > 1$ and δ such that*

$$\begin{array}{ll} \sum_{n\geq 0}(z_0^n - 1)\big(\lambda_Q^{i,+}(U_{t^-},n) - \lambda_Q^{i,-}(U_{t^-},n)\frac{1}{z_0^n}\big) \leq -\delta, & a.s \quad \text{when } Q^i_{t^-} \geq C_{bound}, \\ \sum_{k\geq 0}(z_0^{\alpha_0 k} - 1)\big(\lambda_S^{+}(U_{t^-},k) - \lambda_S^{-}(U_{t^-},k)\frac{1}{z_0^{\alpha_0 k}}\big) \leq -\delta, & a.s \quad \text{when } S_{t^-} \geq C_{bound}, \end{array} \tag{7}$$

for any $i \in \mathbb{B}$ and $U_t = (Q_t^1, Q_t^2, S_t) \in \mathbb{U}$ where α_0 is the tick size.

Assumption 2 ensures that both the size of the first limits and the spread tend to decrease when they become too large. Same kind of hypothesis are used in [63, 82] but when the order book dynamic is Markov.

Remark 5. *In practice, Assumption 2 is verified when the following conditions are satisfied:*

$$\begin{array}{ll} \sum_{n\geq 0}(z_0^n - 1)\big(\psi_Q^{i,+}(u,n,t,z) - \psi_Q^{i,-}(u,n,t,z)\frac{1}{z_0^n}\big) \leq -\delta, & \text{when } q^i \geq C_{bound}, \\ \sum_{n\geq 0}(z_0^{\alpha_0 k} - 1)\big(\psi_S^{+}(u,k,t,z) - \psi_S^{-}(u,k,t,z)\frac{1}{z_0^{\alpha_0 k}}\big) \leq -\delta, & \text{when } s^i \geq C_{bound}, \\ \phi_Q^{i,+}(u,n,t,x) \leq \phi_Q^{i,-}(u,n,t,x), & \text{when } q^i \geq C_{bound}, \\ \phi_S^{+}(u,k,t,x) \leq \phi_S^{-}(u,k,t,x), & \text{when } s^i \geq C_{bound}, \\ \psi(e,u,t,z), \quad \textit{is non-decreasing in } z, & \text{when } q^i \geq C_{bound}, \\ \psi(e,u,t,z), \quad \textit{is non-decreasing in } z, & \text{when } s^i \geq C_{bound}, \end{array} \tag{8}$$

where $u = (q^1, q^2, s) \in \mathbb{U}$, $i \in \mathbb{B}$ and $\psi_Q^{i,\pm}$, $\psi_S^{\pm}$, $\phi_Q^{i,\pm}$ and $\phi_S^{\pm}$ are functions defined such that

$$\begin{array}{ll} \psi_Q^{i,\pm}(u,n,t,z) = \sum_{e\in E_Q^{i,\pm}(u,n)} \psi(e,u,t,z), & \phi_{Q(S)}^{i,+}(u,n,t,x) = \sup_{e\in E_{Q(S)}^{i,+}(u,n)} \phi(e,u,t,x), \\ \psi_S^{\pm}(u,k,t,z) = \sum_{e\in E_S^{\pm}(u,k)} \psi(e,u,t,z), & \phi_{Q(S)}^{-}(u,k,t,x) = \inf_{e\in E_{Q(S)}^{-}(u,k)} \phi(e,u,t,x), \end{array}$$

with $(n,k,t,z) \in \mathbb{N}^2 \times \mathbb{R}_+^2$. Although Inequalities (7) and (8) are not equivalent, there is a large panel of functions that satisfy (8). A proof of this result is given Appendix IV.B.

Assumption 3 (Bound on the overall flow). *We assume that there exist $z_1 > 1$, M and $\underline{\psi} > 0$ satisfying*

$$\begin{array}{lll} c^* = \sum_{e\in E} c(e) < \infty, & & \\ \lambda^* = \sum_{e\in E, \mathbf{k}_m \in \mathscr{P}(n_\psi)} d(e)\binom{n_\psi}{\mathbf{k}_m} \int_{\mathbb{R}_+^m} \prod_{j=1}^m \phi^{*\,k_j}(e,s_j)\,ds_j < \infty, & & \\ Q^i_\infty = \sum_{n\in\mathbb{N}}(z_1^n - 1)\mathbb{E}_{\mathbf{x}}\big[\lambda_Q^{i,+}(u,n) - \frac{\lambda_Q^{i,-}(u,n)}{z_1^n}\big] < M, & & \text{when } q^i \leq C_{bound}, \\ S_\infty = \sum_{k\in\mathbb{N}}(z_1^k - 1)\mathbb{E}_{\mathbf{x}}\big[\lambda_S^{+}(u) - \frac{\lambda_S^{-}(u,n)}{z_1^k}\big] < M, & & \text{when } s \leq C_{bound}, \\ \underline{\lambda}_t(e) = \sum_{e\in E} \lambda_t(e) \geq \underline{\psi}, & & a.s. \end{array}$$

with $c(e)$, $d(e)$ and ϕ^* defined in Assumption 1, $i \in \mathbb{B}$, $\mathbf{x} \in W_0$ and C_{bound} defined in Assumption (2). Similar assumptions are considered in [63, 82] in the Markov case.

Assumption 3 ensures no explosion in the system since it forces the arrival rate of orders, the size of the limits and the spread to stay bounded.

Remark 6. *In practice, we can find path-wise conditions similar to those used in Remark 5 such that the inequalities $Q^i_\infty < M$, $S_\infty < M$ and $\underline{\lambda}_t(e) \geq \bar{\psi}$, a.s are satisfied.*

Theorem 1 (Existence). *Under Assumptions 1, 2 and 3, the process $\bar{U}_t = (Q^1_t, Q^2_t, S_t, \lambda_t)$ admits an invariant distribution.*

The proof of this result is given in Appendix IV.C.

Assumption 4 (Regularity). *We assume that ψ is a càdlàg function continuous with respect to z, ϕ is a positive càdlàg function and there exist $\bar{\psi} : \mathbb{R}_+ \to \mathbb{R}_+$ and $n_1 \in \mathbb{N}$ such that*

$$|\psi(e,u,s,x) - \psi(e,u,s,y)| \leq |\bar{\psi}(x) - \bar{\psi}(y)|, \quad \forall (e,u,s,x,y) \in E \times \mathbb{U} \times \mathbb{R}^3_+,$$

and

$$|\bar{\psi}(x) - \bar{\psi}(y)| \leq K|x-y||1 + x^{n_1} + y^{n_1}|, \quad \forall (x,y) \in \mathbb{R}^2_+,$$

with K a positive constant.

Remark 7. *Assumption 4 is satisfied in the special case where $\bar{\psi}$ is a polynomial.*

We have the following result.

Theorem 2 (Ergodicity). *Under Assumptions 1, 2, 3 and 4, the process $\bar{U}_t$ is $\mathcal{W}_0$-ergodic, which means that there exists an invariant measure μ, see Definition 1, that satisfies*

$$\lim_{t\to\infty} P_t(x,A) = P_0(\mu,A), \quad \forall x \in W_0,\ A \in \mathcal{X},$$

where $P_t(x,A)$ is the probability that $\bar{U}_t \in A$ starting from the initial condition x. Additionally, we have the following speed of convergence:

$$||P_t(x,.) - P_0(\mu,.)||_{TV} \leq K_1 e^{-K_2 t}, \quad \forall x \in W_0,$$

with K_1, K_2 are positive constants and $||.||_{TV}$ the total variation norm.

The proof of this result is given in Appendix IV.D. We can construct pathwise the point process $N = (T_n, X_n)$ defined in Section 2 using the following algorithm.

Remark 8 (Pathwise construction of N). *Using the thinning algorithm proposed by Lewis in [83] and Ogata in [95], the point process $N = (T_n, X_n)$ defined in Section 2 satisfies $N = \lim_{m\to\infty} N^m$ where N^m is defined as follows*

$$\begin{array}{lcl}
\lambda^{m+1}_t(e) & = & \psi\big(e, U^m_{t^-}, t, \sum_{T^m<t} \phi(e, U^m_{t^-}, t - T^m, X^m)\big)\mathbf{1}_{T^m \leq t < T^{m+1}} + \lambda^m_t(e)\mathbf{1}_{t<T^m}, \\
N^{m+1}((0,t]\times B) & = & \int_{(T^m, T^{m+1}]\times B} N^*(dt \times (0, \lambda^{m+1}_t(e)] \times de)\mathbf{1}_{t>T^m} + N^m((0, t\wedge T^m] \times B), \\
T^{m+1} & = & \sup\{t > T^m; \quad \int_{(T^m,t]\times\mathcal{E}} N^*(dt \times (0, \lambda^m_t(e)] \times de) = 0\},
\end{array}$$

with U^m the order book process generated by N^m and described in (1), $N^* = (T^*_n, R^*_n, X^*_n)$ a Poisson process valued on $\mathbb{R}^2_+ \times E$ which admits $\mathrm{d}t\mathrm{d}z\nu(\mathrm{d}e)$ as an $\mathcal{F}^{N^*}_t$ intensity and $\nu = \sum_{e\in E}\delta_e$.

This is a well known result that were used in many contexts, see [26, 40, 79, 83, 95]. The proof of Theorem 1 ensures that the above algorithm is well defined.

4 Limit theorems

Let n be the index of the n-th jump, $(\eta_n)_{n\geq 0}$ be a process satisfying $\eta_n = f((U_i)_{i\leq n},(Y_i)_{i\leq n})$ with f a measurable function valued on $(\mathbb{R},\mathscr{B}(\mathbb{R}))$, $(Y_i)_{i\geq n}$ is a geometrically ergodic sequence, see 15.7 in [89], independent of $(U_i)_{i\geq n}$. Here, we write μ for the invariant measure of the joint process (U,Y), $V_n = \sum_{k=1}^n \eta_k$ and $S_n = \sum_{k=1}^n (\eta_k - \mathbb{E}_\mu[\eta_k])$. We denote by

$$X_n(t) = \frac{S_{\lfloor nt\rfloor}}{\sqrt{n}}, \qquad \forall\, t \geq 0.$$

Assumption 5. *Under the invariant measure μ, the sequence $(\eta_i)_{i\geq 0}$ is stationary and $\mathbb{E}_\mu[|\eta_0|] < \infty$.*

Assumption 6. *Under the invariant measure μ, we have $\mathbb{E}_\mu[(\eta_0 - \mathbb{E}_\mu[\eta_0])^2] < 1$.*

Proposition 2. *Under Assumption 5, we have*

$$\frac{V_n}{n} \underset{n\to\infty}{\longrightarrow} \mathbb{E}_\mu[\eta_0], \quad a.s. \tag{9}$$

Moreover when both Assumptions 5 and 6 are verified, the quantity $X_n(t)$ satisfies

$$X_n(t) \xrightarrow{\mathscr{L}} \sigma W_t, \tag{10}$$

with $\sigma^2 = \mathbb{E}_\mu[\eta_0^2] + 2\sum_{k\geq 1}\mathbb{E}_\mu[\eta_0\eta_k]$ and μ the invariant measure of (U_i,Y_i) and W_t a standard brownian motion.

Note that $\sigma^2 < \infty$ under Assumption 6. The proof of this result is given in Appendix IV.E.

Remark 9. *The leading term in the expression of σ^2 is $\mathbb{E}_\mu[\eta_0^2]$. Numerically, it can be computed as soon as we have an estimate of the stationary distribution of η_0, see Proposition 4.*

Proposition 2 ensures that the large scale limit of S in event time is a brownian motion. However, it is more relevant to study the large scale limit of the process S in calendar time. Thus we now consider the process

$$\tilde{X}_n(t) = \frac{S_{N(nt)}}{\sqrt{n}}, \qquad \forall\, t \geq 0.$$

The following proposition provides the large scale limit of the process $S_{N(nt)}$.

Proposition 3. *Under Assumption 5, we have*

$$\frac{V_{N(nt)}}{n} \underset{n\to\infty}{\longrightarrow} \frac{\mathbb{E}_\mu[\eta_0]}{\mathbb{E}_\mu[\Delta T_1]}, \quad a.s. \tag{11}$$

Moreover when both Assumptions 5 and 6 are verified, the quantity $\tilde{X}_n(t)$ satisfies

$$\tilde{X}_n(t) \xrightarrow{\mathscr{L}} \frac{\sigma}{\sqrt{\mathbb{E}_\mu[\Delta T_1]}} W_t, \tag{12}$$

with $\sigma^2 = \mathbb{E}_\mu[\eta_0^2] + 2\sum_{k\geq 1}\mathbb{E}_\mu[\eta_0\eta_k]$, μ the invariant measure of (U_i, Y_i), $\Delta T_n = T_n - T_{n-1}$ the inter-arrival time between the n-th and $(n-1)$-th jump and W_t a standard brownian motion.

The proof of this result is given in Appendix IV.E.

Remark 10. *The mid price after n jumps P_n satisfies $P_n = P_0 + \sum_{i=1}^n \Delta P_i$ with $\Delta P_i = (P_i - P_{i-1}) = \eta_i$. When $(\eta_i)_{i\geq 0}$ verifies Assumptions 5 and 6, the rescaled price process $\tilde{P}_n(t) = \frac{P_{N(nt)}}{\sqrt{n}}$ converges towards a Brownian diffusion.*

5 Formulas

In this section, we provide a calibration methodology for the intensities and computation formulas for the quantities of interest: the stationary distribution of the order book, the price volatility and the fluctuations of liquidity.

5.1 Stationary probability computation

In this section, we denote by μ the invariant measure of $\bar{U} = (Q^1, Q^2, S, \lambda)$ defined on $(W_0, \mathscr{W}_0)$. Let $\zeta_t = f((U_i)_{T_i\leq t})$ be a stationary process under μ with f a measurable function valued in $(Z, \mathscr{Z})$, Z a countable space and π the stationary distribution of ζ_t. The proposition below provides a fixed point formula satisfied by π.

Proposition 4. *The stationary distribution π satisfies*

$$\begin{aligned} \pi Q &= 0 \\ \pi \mathbf{1} &= 1. \end{aligned} \tag{13}$$

where the infinite dimensional matrix Q verifies

$$Q(z, z') = \sum_{e\in E(z,z')} \mathbb{E}_\mu[\lambda(e)|\zeta_0 = z], \tag{14}$$

with $E(z, z')$ the set of events directly leading to z' from z.

The proof of this result is provided in Appendix IV.F.

Remark 11. *When $\zeta_t = U_t = (Q_t^1, Q_t^2, S_t)$, Proposition 4 provides a fixed point equation for the computation of the stationary distribution π of the order book.*

Remark 12. *The operator Q is the infinitesimal generator of the process ζ defined such that $Q(z, z') = \lim_{\delta\to 0} \frac{\mathbb{P}_\mu[\zeta_\delta = z'|\zeta_0 = z]}{\delta}$ for any $z \neq z'$. The proof of this result is given in Equation (61) of Appendix IV.F.*

5.1.1 Markov framework

In the Markov case, it is a well known result that Q satisfies (13), see [94]. In this case, the coefficients of Q are parameters of the model and can be estimated using (15).

5.1.2 General case

Let us take z and z' two states such that $z \neq z'$, $N_t^{z,z'} = \sum_{T_i<t} \delta^i_{z,z'}$ with $\delta^i_{z,z'} = \mathbf{1}_{\{\zeta_{T_{i-1}}=z, \zeta_{T_i}=z'\}}$ and $t^z = \sum_{T_i<t} \Delta T_i \mathbf{1}_{\{\zeta_{T_{i-1}}=z\}}$ with $\Delta T_i = T_i - T_{i-1}$. We have the following results:

Proposition 5. *When $(\delta^i_{z,z'})_{i\geq 1}$ satisfies Assumption 5, we have*

$$\hat{Q}(z,z') = \frac{N_t^{z,z'}}{t^z} \underset{t\to\infty}{\to} Q(z,z'), \quad a.s. \tag{15}$$

The proof of this result is given in Appendix IV.G.

Remark 13 (Confidence interval). *We can compute a confidence interval for the estimator $\hat{Q}(z,z')$, see Appendix IV.G for the details.*

Remark 14. *When $\zeta_t = U_t = (Q_t^1, Q_t^2, S_t)$, Proposition 5 provides an estimator for the operator $Q(u,u')$ with $u,u' \in \mathbb{U}$ and $u \neq u'$.*

Remark 15. *In the Markov case and $\zeta_t = U_t$, see [65], the authors used the estimator presented in Proposition 5 to evaluate $Q(u,u')$.*

Remark 16. *Let $(z,z') \in \mathbb{U}^2$ such that $z \neq z'$ and $a \in \mathbb{A}$, we consider the quantity $Q(z,z',a) = \sum_{e\in E(z,z')\cap E(a)} \mathbb{E}[\lambda(e)|\zeta_0 = z]$ with $E(a)$ the set of events generated by the agent a. This quantity represents the infinitesimal probability that agent a sends an order that moves ζ from z to z'. It can be estimated by $\hat{Q}(z,z',a) = \frac{N_t^{z,z',a}}{t^z}$ which satisfies*

$$\hat{Q}(z,z',a) = \frac{N_t^{z,z',a}}{t^z} \underset{t\to\infty}{\to} Q(z,z',a), \quad a.s, \tag{16}$$

with $N_t^{z,z',a} = \sum_{T_i<t} \delta^i_{z,z',a}$, $\delta^i_{z,z',a} = \mathbf{1}_{\{\zeta_{T_{i-1}}=z, \zeta_{T_i}=z', A_i=a\}}$ where A_i is the identity of the agent causing the i-th event. The quantity $Q(z,z',a)$ allows us to infer the market dynamic (i.e the operator Q) for a specific combination of the agents, see Equation (14).

5.2 Spread computation

Since the process U_t is ergodic the spread S_t has a stationary distribution. Then, we can compute $\mathbb{E}_\pi[S_\infty]$ where π is the stationary distribution of U. The computation formula for π is detailed in Proposition 4 and the estimation methodology of Q is described in Proposition 5.

5.3 Price volatility computation

We place ourselves in the case of Remark 10 and assume that the mid price moves $(\eta_i)_{i\geq 0}$ are valued in $\zeta = \alpha_0\mathbb{Z}$ with α_0 the tick size. In such situation, the limit theorem of Section 4 ensures the convergence of $\bar{P}_n(t)$ towards

$$\bar{P}_n(t) \xrightarrow{\mathscr{L}} \sigma W_t,$$

with $\sigma^2 = \mathbb{E}_\mu[\eta_0^2] + 2\sum_{k\geq 1}\mathbb{E}_\mu[\eta_0\eta_k]$ and μ the invariant measure of $\bar{U}$. The quantity of interest is σ^2. To compute σ^2, we need to evaluate $\mathbb{E}_\mu[\eta_0\eta_k]$ for all $k \geq 0$. We have

$$\begin{array}{l} \mathbb{E}_\mu[\eta_0^2] = \sum_{\eta\in\zeta}\pi_{\eta_0}(\eta)\eta^2, \\ \mathbb{E}_\mu[\eta_0\eta_k] = \sum_{\eta\in\zeta}\pi_{\eta_0}(\eta)\eta\mathbb{E}_\mu[\eta_k|\eta_0=\eta], \quad \forall k \geq 1, \end{array} \tag{17}$$

with $\pi_{\eta_0}(\eta) = \mathbb{P}_\mu[\eta_0 = \eta]$. Thus we need to estimate π_{η_0} and $\mathbb{E}_\mu[\eta_k|\eta_0=\eta]$ to evaluate σ^2. The computation of the leading term $\mathbb{E}_\mu[\eta_0^2]$ requires only the knowledge of the stationary distribution π_{η_0}. The latter is evaluated using Proposition 4. To estimate $\mathbb{E}_\mu[\eta_k|\eta_0=\eta]$ with $k \geq 1$, we use the following proposition.

Proposition 6. *Let us take $k \geq 1$, $\eta \in \zeta$, $N_n^{\eta,(k)} = \sum_{j\leq n}\eta_j\delta_\eta^{j(k)}$ with $\delta_\eta^{j(k)} = \mathbf{1}_{\{\eta_{j-k}=\eta\}}$ and $n^\eta = \sum_{j\leq n}\delta_\eta^{j(k)}$. When both $(\eta_i\delta_\eta^{i(k)})_{i\geq 1}$ and $(\delta_\eta^{i(k)})_{i\geq 1}$ satisfy Assumption 5, we have*

$$\hat{E}(\eta_0, k) = \frac{N_n^{\eta_0(k)}}{n^\eta} \underset{n\to\infty}{\to} \mathbb{E}_\mu[\eta_k|\eta_0=\eta], \quad a.s. \tag{18}$$

The proof of this result is similar to the one of Proposition 5.

Remark 17 (Markov case). *When the dynamic of U is Markov and $\eta_i = f_0(U_i)$ for any $i \geq 0$ with f_0 a deterministic function, see Remark 18. We have*

$$\mathbb{E}_\pi[\eta_0\eta_k] = \sum_{u\in\mathbb{U}}\pi(u)\eta_0(u)\mathbb{E}_u[\eta_k], \tag{19}$$

where π is the stationary distribution of U that can be computed using Proposition 4 and $\mathbb{E}_u[\eta_k] = (P^k * \eta_0)_u = \sum_{u'\in\mathbb{U}}P^k_{u,u'}\eta_0(u')$ with P^k the k-th power of the Markov chain P associated to the process U and which satisfies

$$\begin{array}{l} P_{u,u'} = \left\{ \begin{array}{ll} -Q_{u,u'}/Q_{u,u} & \text{if } u \neq u' \text{ and } Q_{u,u} \neq 0, \\ 0 & \text{if } u \neq u' \text{ and } Q_{u,u} = 0, \end{array} \right. \\ P_{u,u} = \left\{ \begin{array}{ll} 0 & \text{if } Q_{u,u} \neq 0, \\ 1 & \text{if } Q_{u,u} = 0, \end{array} \right. \end{array} \tag{20}$$

where the quantity $P_{u,u'}$ represents $P_{u,u'} = \mathbb{P}[U_1 = u'|U_0 = u]$ with U_1 the state of the order book after one jump.

Remark 18. *In Section 6, for any $u = (q^1, q^2, s)$, we consider the following function:*

$$f_0(u) = \begin{cases} -1 & \text{if } q^1 = 0 \text{ and } q^2 > 0, \\ +1 & \text{if } q^2 = 0 \text{ and } q^1 > 0, \\ 0 & \text{otherwise}, \end{cases}$$

for the numerical simulations. Note that the states where $q^1 = 0$ or $q^2 = 0$ are fictitious states that are not observable in practice. These states are introduced to handle the price changes. Indeed, the states where $q^1 = 0$ (resp. $q^2 = 0$) correspond to a price decrease (resp. increase) by one tick and the states where both $q^1 = 0$ and $q^2 = 0$ are unreachable.

5.4 An alternative measure of market stability

Another way to look at market stability is to investigate the behaviour of the disequilibrium between offer and demand. This equilibrium can be for example measured through the cumulative imbalance $N_t = V_t^b - V_t^a$ where V_t^b (resp. V_t^a) is the net number of inserted limit orders at the bid (resp. ask). From no arbitrage argument, we know that the dynamic of N_t is closely related to that of the price [69, 73]. Consequently, it is natural to view the long term volatility of this object as an alternative measure of market stability.

In this section, we follow the same methodology of Section 5.3. The cumulative imbalance after n jumps N_n satisfies $N_n = N_0 + \sum_{i=1}^n \Delta N_i$ where $\Delta N_i = N_i - N_{i-1} = n_i$. Hence, when $(n_i)_{i\geq 0}$ satisfies Assumptions 5 and 6, we have the following convergence result:

$$X_n^N = \frac{\sum_{k=1}^n (n_k - \mathbb{E}_u[n_k])}{\sqrt{n}} \xrightarrow{\mathcal{L}} \tilde{\sigma} W_t,$$

with $\tilde{\sigma}^2 = \mathbb{E}_\mu[n_0^2] + 2\sum_{k\geq 1} \mathbb{E}_\mu[n_0 n_k]$ and μ the stationary distribution of $\bar{U}$ given by proposition 4. The quantity $\mathbb{E}_\mu[n_0 n_k]$ can be computed using the same methodology of Section 5.3.

6 Numerical experiments

In this section, we propose a ranking of the market makers for four different assets, based on their impact on volatility. For each asset, we compute first the liquidity provision and consumption intensities relative to the whole market using Equation (15)[7]. Then, we estimate the stationary measure of the order book, see Equation (13), and use it to compute the two following estimators of the market volatility:

$$\begin{aligned} \sigma^{2,G} &= \mathbb{E}_\mu[\eta_0^2], \\ \sigma_k^{2,M} &= \mathbb{E}_\pi[\eta_0^2] + 2\sum_{j=1}^k \mathbb{E}_\pi[\eta_0 \eta_j], \end{aligned}$$

where μ is the invariant measure of $\bar{U}$ given by Theorem 2, π is the stationary distribution of U when both the order book dynamic is Markov and $\eta_i = f_0(U_i)$ with f_0 defined in Remark

[7] A liquidity provision (resp. consumption) event is assimilated to an increase (resp. decrease) of the best bid or ask size by 1 unit. To fix the ideas, one (AES) is our unit here.

18. The estimator $\sigma^{2,G}$ is computed by applying Equation (17) and $\sigma_k^{2,M}$ is evaluated using Remark 17. Thereafter, for each market maker, we compute its own intensities using Equation (16). After that, we estimate the new market intensities in a situation where we suppose that he withdraws from the exchange by subtracting the agent intensity from the market one, see Corollary 1. We finally compute the new market volatility estimators $\sigma^{2,G}$ and $\sigma_k^{2,M}$ corresponding to this new scenario using Equation (17) and Remark 17 again.

Remark 19. *In the simple case where the order book dynamic is Markov and the queues are independent, see Section 2.3.3 in [63], minimizing the first order approximation of the price volatility* $\sigma^2 \sim \mathbb{E}_\pi[\eta_0^2]$ *is similar to selecting the agent with the highest ratio insertion/consumption* $\frac{\lambda_Q^{1(2),+}}{\lambda_Q^{1(2),-}}$. *This condition is a well-known result which means that the new agent needs to have an insertion/consumption ratio greater than the one of the market. The proof of this result is given in Section IV.H.*

Remark 20. *The reconstruction methodology of the market assumes that other participants will not modify their behaviours when an agent leaves the market. In practice, this assumption is satisfied since agents react to global variables such as the imbalance and not to a specific agent-based information. Additionally, when an agent leaves the market, the other participants do not have enough order flow history to calibrate all the parameters of their models.*

Remark 21. *The reconstruction methodology of the market takes into account the volume exchanged by each agent since this information is included in the estimated intensities. Indeed, the intensity of an agent who trades a large volume is high because he either interacts frequently with the market or generates significant changes in the order book state.*

6.1 Database description.

We study four large tick European stocks: Air Liquid, EssilorLuxottica, Michelin and Orange, on Euronext, over a year period: from January 2017 till December 2017. The data under study are provided by the French Regulator Autorité des marchés financiers. For each of these assets, we have access to the trades and orders data. Using both data, we rebuild the Limit Order book (LOB) up to the first limit of both sides, whenever an event (an order insertion, an order cancellation or an aggressive order) happens on one of these limits. Note that we remove market data corresponding to the first and last hour of trading, as these periods have usually specific features because of the opening/closing auction phases. We present in Table IV.1 some preliminary statistics on the different considered assets.

Asset	Number of insertion orders (in millions of orders)	Number of cancellation orders (in millions of orders)	Number of aggressive orders (in millions of orders)	Ratio of cancellation orders number over aggressive orders number	Average spread (in ticks)
Air Liquide	2.36	2.40	0.21	11.4	1.07
EssilorLuxottica	3.90	3.96	0.34	11.6	1.11
Michelin	3.81	4.01	0.32	12.5	1.14
Orange	6.60	6.66	0.47	14.1	1.14

Table IV.1 – Preliminary statistics on the assets.

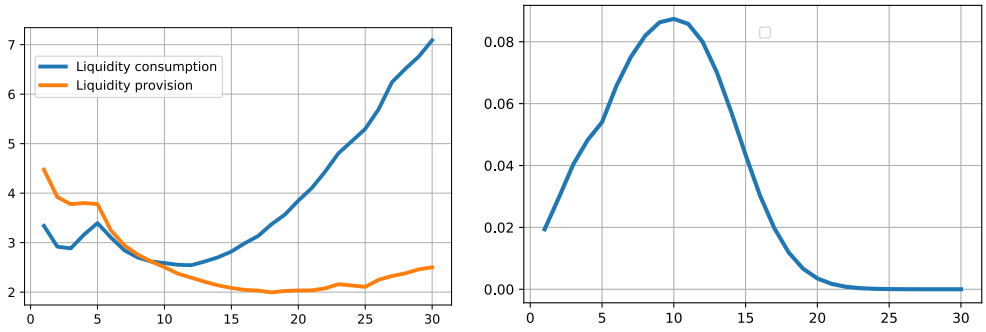
Table IV.1 shows that the number of insertion orders is lower than that of cancellation orders. A priori, this seems contradictory, but what happens in practice is that some agents insert orders that they cancel partially and progressively at a later stage by sending multiple cancellation orders, which leads to a number of cancellation orders higher than that of insertion orders.

The considered market makers, that we aim at ranking, are the Supplemental Liquidity Providers (SLP) members. The SLP programme imposes a market making activity on programme members, including order book presence time at competitive prices. In return, they get favorable pricing and rebates in the form of a maker-taker fees model directly comparable to those of the major competing platforms. This programme includes 9 members. Some of them have at the same time SLP activity and other activities, such like proprietary or agency activity. In our analysis, we only analyse the SLP flow of these members. We denote the market makers by MM1 to MM9.

6.2 Computation of the intensities and the stationary measure

We compute the liquidity consumption and provision intensities at the first limit relative to the whole market according to the queue size, the corresponding stationary measure and the long term volatility for Air Liquide. Results relative to EssilorLuxottica, Michelin and Orange are relegated to Appendix IV.I. The estimation methodology of the intensities is based on Proposition 5. To apply this proposition, we record, for every event occurring in the LOB at the best limits (best ask and bid), the type of this order (insertion or consumption), the waiting time (in number of seconds) between this event and the preceding one occurring at the same limit and the queue size before the event. The queue size is then approximated by the smaller integer that is larger than or equal to the volume available at the queue, divided by the stock average event size (AES) computed for each limit on a daily basis. In practice, the spread cannot be equal to one tick all the time. This is why we exclude from our analysis all the events that occur when the spread is higher than one tick.

(a) Intensity of the market (b) Stationary measure Q^1



Long term price volatility $\sigma^{2,G} = 0.035$, $\sigma^{2,M}_{10} = 0.227$.

Figure IV.2 – (a) Liquidity insertion and consumption intensities (in orders per second) with respect to the queue size (in AES) and (b) the corresponding stationary distribution of Q^1 with respect to the queue size (in AES), proper to Air Liquide.

We can see that for all these assets, the liquidity provision intensity is approximately a decreasing function of the queue size. This result reveals a quite common strategy used in practice: posting orders when the queue is small to seize priority (for further details about the priority value, see Chapter III). For all assets, the consumption intensity is an increasing function when the queue size is large. For small queue sizes, we notice a slight decrease of this intensity, see Figure IV.2. Indeed, the increasing aspect corresponding to large queue sizes is explained by market participants waiting for better price when liquidity is abundant. The decreasing aspect associated to small queue sizes is due to aggressive orders sent by agents to get the last remaining quantities available at the first limits: market participants rushing for liquidity when it is rare. The lower the ratio of cancellation orders number over aggressive orders number is, the clearer the decreasing shape for small queue sizes stands out, see Table IV.1 and Figures IV.2, IV.4, IV.5 and IV.6.

6.3 Ranking of the market makers

For each of the assets and for each one of the market makers, we compute the liquidity consumption and provision intensities, and the corresponding price volatility $\sigma^{2,M}_{10}$ that we would obtain in a situation where the studied market maker withdraws from the market. Since the estimators $\sigma^{2,G}$ and $\sigma^{2,M}_{10}$ give the same ranking, we choose to show the values for $\sigma^{2,M}_{10}$ alone. We show next the results relative to Air Liquide; those of EssilorLuxottica, Michelin and Orange are relegated to Appendix IV.I.

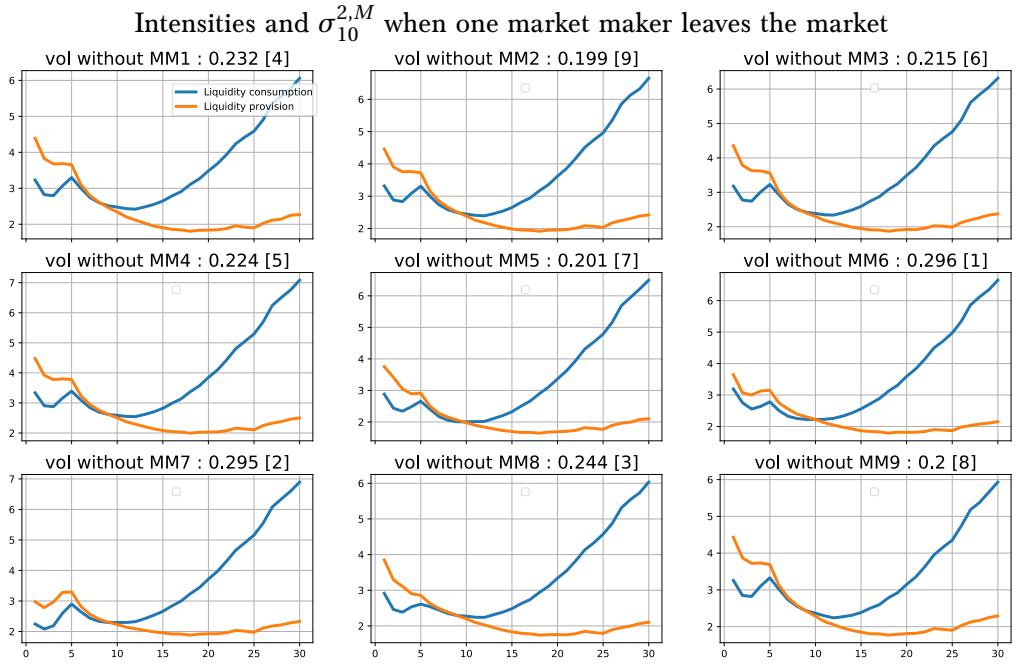


Figure IV.3 – Liquidity insertion and consumption intensities (in orders per second) with respect to the queue size (in AES) and $\sigma_{10}^{2,M}$ when one market maker is ejected from the market for the stock Air Liquide.

Based on the previous results, we carry out for each asset the ranking of the different market makers according to their contribution to volatility. To do so, we compare the expected volatility when removing each market maker from the market to the actual one when all the market makers in the market: if the expected volatility is higher (resp. lower) than the actual one, this means that the market maker into question decreases (resp. increases) market volatility. The market maker who decreases[8] (resp. increases[9]) volatility the most is ranked first (resp. last). In the following table, we add a star next to market makers deceasing volatility: a zero star (resp. a four stars) means that the market maker increases (resp. decreases) the market volatility of the 4 studied assets.

[8]The expected volatility of the new market without this market maker is the highest.
[9]The expected volatility of the new market without this market maker is the lowest.

Market maker	Ranking Air Liquide	Market share Air Liquide	Ranking Exilor-Luxottica	Market share Exilor-Luxottica	Ranking Michelin	Market share Michelin	Ranking Orange	Market share Orange
MM1***	4	4%	3	3%	3	4%	3	3%
MM2	9	1%	9	1%	9	1%	7	1%
MM3	6	5%	6	5%	7	4%	5	4%
MM4	5	1%	4	1%	4	0%	4	1%
MM5	7	5%	8	5%	8	5%	9	5%
MM6****	1	3%	2	3%	1	3%	1	4%
MM7****	2	7%	1	12%	2	9%	2	7%
MM8*	3	9%	5	5%	5	5%	6	4%
MM9	8	2%	7	2%	6	2%	8	2%

Table IV.2 – Market share and ranking of markets makers

IV.A Market reconstitution

Proof of Proposition 1. Let $t \geq 0$ be the current time. For any $B \in \mathscr{E}$, we denote by $T^{t,e}$ the first time greater than t when an event $e \in B$ happens given $\mathscr{F}_t$ and $T^{t,B} = \min_{e \in B} T^{t,e}$ the next market event. Thus, we have

$$\begin{aligned}\lambda_t(B) &= \lim_{\delta t \to 0} \frac{\mathbb{P}\left[\#\{T_n \in (t, t+\delta t], X_n \in B\} \geq 1 | \mathscr{F}_t\right]}{\delta t} \\ &= \lim_{\delta t \to 0} \frac{\mathbb{P}\left[\{T^{t,B} \in (t, t+\delta t]\} | \mathscr{F}_t\right]}{\delta t}.\end{aligned}$$

We write $f^{t,e}$ for the density function of $T^{t,e}$ and $F_B^{t,e}(s) = \mathbb{P}[\left(\min_{\tilde{e} \in B \setminus \{e\}} T^{t,\tilde{e}}\right) \geq s | T^{t,e} \leq s]$ for any $s \geq 0$. Using the monotone convergence theorem, we have

$$\begin{aligned}\lim_{\delta t \to 0} \frac{\mathbb{P}\left[\{T^{t,B} \in (t, t+\delta t]\} | \mathscr{F}_t\right]}{\delta t} &= \lim_{\delta t \to 0} \frac{\sum_{e \in B} \int_t^{t+\delta t} f^{t,e}(s) F_B^{t,e}(s)\, ds}{\delta t} \\ &= \sum_{e \in B} \lim_{\delta t \to 0} \frac{\int_t^{t+\delta t} f^{t,e}(s) F_B^{t,e}(s)\, ds}{\delta t} \\ &= \sum_{e \in B} f^{t,e}(t) F_B^{t,e}(t) = \sum_{e \in B} \lambda_t(e),\end{aligned}$$

since $f_a^{t,e'}(t) = \lambda_t((e', a))$ using Equation (2) and $F_a^{t,e'}(t) = 1$ by definition. This completes the proof. □

IV.B Proof of Remark 5

Proof of Remark 5. Let $N = (T_n, X_n)$ be the point process defined in Section 2 and $i \in \mathbb{B} = \{1, 2\}$. We define $\phi_Q^{i,\pm,n}$ in the following way:

$$\begin{aligned}\phi_Q^{i,+,n} &= \sup_{e\in E_Q^{i+}(u,n)} \sum_{T_i<t} \phi(e, U_{t^-}, n, t - T_i, X_i),\\ \phi_Q^{i,-,n} &= \inf_{e\in E_Q^{i-}(u,n)} \sum_{T_i<t} \phi(e, U_{t^-}, n, t - T_i, X_i),\end{aligned}$$

with $U_t = (Q_t^1, Q_t^2, S_t)$. When $Q_{t^-}^i \geq C_{bound}$, using that ψ is non-decreasing in z, we have

$$\begin{aligned}\sum_{n\geq 0}(z_0^n - 1)\big(\lambda_Q^{i,+}(U_{t^-}, n) - \lambda_Q^{i,-}(U_{t^-}, n)\frac{1}{z_0^n}\big) &\leq \sum_{n\geq 0}(z_0^n - 1)\big(\psi_Q^{i,+}(U_{t^-}, n, t, \phi_Q^{i,+,n}) - \lambda_Q^{i,-}(U_{t^-}, n)\frac{1}{z_0^n}\big)\\ &= \sum_{n\geq 0}(z_0^n - 1)\big(\psi_Q^{i,+}(U_{t^-}, n, t, \phi_Q^{i,+,n}) - \psi_Q^{i,-}(U_{t^-}, n, t, \phi_Q^{i,+,n})\frac{1}{z_0^n}\big)\\ &+ \sum_{n\geq 0}(1 - \frac{1}{z_0^n})\big(\psi_Q^{i,-}(U_{t^-}, n, t, \phi_Q^{i,+,n}) - \lambda_Q^{i,-}(U_{t^-}, n)\big) = (i) + (ii).\end{aligned}$$

Using Equation (8), we have

$$(i) = \sum_{n\geq 0}(z_0^n - 1)\big(\psi_Q^{i,+}(U_{t^-}, n, t, \phi_Q^{i,+,n}) - \psi_Q^{i,-}(U_{t^-}, n, t, \phi_Q^{i,+,n})\frac{1}{z_0^n}\big) \leq -\delta, \quad a.s, \tag{21}$$

when $Q_{t^-}^i \geq C_{bound}$. Moreover, using that ψ is non-decreasing in z, we have

$$\begin{aligned}(ii) &= \sum_{n\geq 0}(1 - \frac{1}{z_0^n}) \sum_{e\in E_Q^{i-}(u,n)} \big(\psi(e, U_{t^-}, n, t, \phi^{i,+,n}) - \psi(e, U_{t^-}, n, t, \sum_{T_i<t}\phi(e, U_{t^-}, n, t - T_i, X_i))\big)\\ &\leq \sum_{n\geq 0}(1 - \frac{1}{z_0^n}) \sum_{e\in E_Q^{i-}(u,n)} \big(\psi(e, U_{t^-}, n, t, \phi^{i,+,n}) - \psi(e, U_{t^-}, n, t, \phi^{i,-,n})\big), \quad a.s,\end{aligned}$$

when $Q_{t^-}^i \geq C_{bound}$. Since Equation (8) ensures that $\phi^{i,+,n} \leq \phi^{i,-,n}$, $a.s$ and ψ is non-decreasing in z, we deduce that

$$(ii) = \sum_{n\geq 0}(1 - \frac{1}{z_0^n})\big(\psi_Q^{i,-}(U_{t^-}, n, t, \phi_Q^{i,+,n}) - \lambda_Q^{i,-}(U_{t^-}, n)\big) \leq 0, \quad a.s, \tag{22}$$

when $Q_{t^-}^i \geq C_{bound}$. Using Equations (21) and (22), we get

$$\sum_{n\geq 0}(z_0^n - 1)\big(\lambda_Q^{i,+}(U_{t^-}, n) - \lambda_Q^{i,-}(U_{t^-}, n)\frac{1}{z_0^n}\big) \leq -\delta \quad a.s,$$

when $Q_{t^-}^i \geq C_{bound}$. By following the same methodology, we also get

$$\sum_{n\geq 0}(z_0^n - 1)\big(\lambda_S^{i,+}(U_{t^-}, n) - \lambda_S^{i,-}(U_{t^-}, n)\frac{1}{z_0^n}\big) \leq -\delta, \quad a.s,$$

when $S_{t^-} \geq C_{bound}$. This completes the proof. □

IV.C Proof of Theorem 1

IV.C.1 Preliminary results

For any $k \geq 1$, we denote by $T_{n+1}(e)$, $T^{i\pm}_{Q_{n+1}}(k)$ and $T^{i\pm}_{S_{n+1}}(k)$ respectively the arrival time of the first event e, $e^{i\pm}_Q(k) \in E^{i,\pm}_Q$ and $e^{\pm}_S(k) \in E^{\pm}_S$ greater than T_n. The sets $E^{i,\pm}_Q$ and $E^{\pm}_S$ are defined in Equation (6). They contain the events that increase or decrease the best bid, best ask and spread by k.

Lemma 1. *Let $n \geq 0$ and $i \in \mathbb{B}$. The order book increments satisfy the following formulas:*

$$\mathbb{P}[\Delta Q^i_{n+1} = \pm k] = \mathbb{E}\Big[\int_{\mathbf{R}_+} \lambda^{i,\pm}_{Q_n}(t,k) Z_n(t)\, dt\Big],$$
$$\mathbb{P}[\Delta S_{n+1} = \pm k] = \mathbb{E}\Big[\int_{\mathbf{R}_+} \lambda^{\pm}_{S_n}(t,k) Z_n(t)\, dt\Big],$$

with $\Delta Q^i_{n+1} = Q^i_{n+1} - Q^i_n$, $\Delta S_{n+1} = S_{n+1} - S_n$ and

$$Z_n(t) = e^{-[\sum_e \int_0^t \lambda_n(e, s+T_n)\, ds]}, \qquad \lambda^{i,\pm}_{Q_n}(t,k) = \sum_{e \in E^{i\pm}_Q(k)} \lambda_n(e, t+T_n),$$
$$\lambda^{\pm}_{S_n}(t,k) = \sum_{e \in E^{i\pm}_S(k)} \lambda_n(e, t+T_n), \qquad \lambda_n(e,t) = \psi\big(e, U_{T_n}, t, \sum_{T_i \leq T_n} \phi(e, U_{T_n}, t-T_i, X_i)\big), \quad \forall t \geq 0.$$

Proof of Lemma 1. We write $\Delta T_{n+1}(e) = T_{n+1}(e) - T_n$ for any event $e \in E$ and $\Delta T^{i\pm}_{Q_{n+1}}(k) = T^{i\pm}_{Q_{n+1}}(k) - T_n$. Using Remark 8, the increments $(\Delta T_{n+1})_{n\geq 0}$ are independent given $\mathscr{F}_n$ and $\Delta T_{n+1}(e)|\mathscr{F}_n$ follows a non homogeneous exponential distribution with an intensity $\lambda_n(e,.)$. Thus, we have

$$\begin{aligned}
\mathbb{P}[\Delta Q^i_{n+1} = \pm k] &= \mathbb{E}[\mathbb{P}[\Delta T^{i\pm}_{Q_{n+1}}(k) < \Delta T_{n+1}(e), \quad \forall e \notin E^{i\pm}_Q(k) | \mathscr{F}_n]] \\
&= \mathbb{E}\Big[\int_{\mathbf{R}_+} \lambda^{i,\pm}_{Q_n}(t,k)\, e^{-\int_0^t \lambda^{i,\pm}_{Q_n}(s,k)\, ds}\, dt \prod_{e \notin E^{i\pm}_Q(k)} \Big(\int_{\mathbb{R}_+} \mathbf{1}_{t<t_e} \lambda_n(e,t_e) e^{-\int_0^{t_e} \lambda_n(e,s)\, ds} dt_e\Big)\Big] \\
&= \mathbb{E}\Big[\int_{\mathbf{R}_+} \lambda^{i,\pm}_{Q_n}(t,k) e^{-[\sum_e \int_0^t \lambda_n(e,s)\, ds]}\, dt\Big] = \mathbb{E}\Big[\int_{\mathbf{R}_+} \lambda^{i,\pm}_{Q_n}(t,k) Z_n(t)\, dt\Big]. \qquad (23)
\end{aligned}$$

By following the same methodology used in Equation (23), we get

$$\mathbb{P}[\Delta S_{n+1} = \pm k] = \mathbb{E}\Big[\int_{\mathbf{R}_+} \lambda^{\pm}_{S_n}(t,k) Z_n(t)\, dt\Big],$$

which completes the proof. □

Let $\tau_{\mathscr{O}}$ be the first entrance period of $N^i = (T_{i+j}, X_{i+j})_{j\leq 0}$ to the set $\mathscr{O} \in \mathscr{W}_0$, C_{bound} defined in Assumption 2 and $1 < z \leq \min(z_0, z_1)$ with z_0 and z_1 are respectively defined in Assumptions 2 and 3.

Lemma 2 (Drift condition). *Under Assumptions 2 and 3, the process $U_n = (Q_n^1, Q_n^2, S_n)$ satisfies the following drift condition:*

$$\begin{aligned}
&\mathbb{E}\big[z^{Q_{n+1}^i - C_{bound}} \mathbf{1}_{\tau_{\mathscr{O}} \geq n+1}\big] \leq \lambda \mathbb{E}\big[z^{Q_n^i - C_{bound}} \mathbf{1}_{\tau_{\mathscr{O}} \geq n+1}\big] + B\mathbb{E}\big[\mathbf{1}_{\tau_{\mathscr{O}} \geq n+1}\big], \\
&\mathbb{E}\big[z^{S_{n+1} - C_{bound}} \mathbf{1}_{\tau_{\mathscr{O}} \geq n+1}\big] \leq \lambda \mathbb{E}\big[z^{S_n - C_{bound}} \mathbf{1}_{\tau_{\mathscr{O}} \geq n+1}\big] + B\mathbb{E}\big[\mathbf{1}_{\tau_{\mathscr{O}} \geq n+1}\big], \qquad \forall n \in \mathbb{N}, \forall i \in \mathbb{B},
\end{aligned}$$

with $\lambda < 1$ and B two constants.

Remark 22. *We define*

$$V_{C_{bound}}(u) = \sum_{i \in \{1,2\}} z^{q^i - C_{bound}} + z^{s - C_{bound}}, \quad \forall u \in \mathbb{U}. \tag{24}$$

Using Lemma 2, we deduce that

$$\mathbb{E}\big[V_{C_{bound}}(U_{n+1}) \mathbf{1}_{\tau_{\mathscr{O}} \geq n+1}\big] \leq \lambda \mathbb{E}\big[V_{C_{bound}}(U_n) \mathbf{1}_{\tau_{\mathscr{O}} \geq n+1}\big] + 3B\mathbb{E}\big[\mathbf{1}_{\tau_{\mathscr{O}} \geq n+1}\big], \quad \forall n \in \mathbb{N},$$

Proof of Lemma 2. We write $\tilde{\mathbb{E}}\big[X\big] = \mathbb{E}\big[X \mathbf{1}_{\tau_{\mathscr{O}} \geq n+1}\big]$ for any random variable X to simplify the notations and V instead of $V_{C_{bound}}$ since there is no possible confusion. We have

$$\tilde{\mathbb{E}}\big[z^{Q_{n+1}^i} | \mathscr{F}_n\big] = \tilde{\mathbb{E}}\big[z^{Q_n^i} | \mathscr{F}_n\big] + \sum_{u' \neq U_n} \tilde{\mathbb{P}}[Q_{n+1}^i = q' | \mathscr{F}_n] \left[z^{q'} - z^{Q_n^i}\right].$$

Using Lemma 1, we get

$$\mathbb{P}[\Delta Q_{n+1}^i = \pm k] = \mathbb{E}\big[\int_{\mathbf{R}_+} \lambda_{Q_n}^{i,\pm}(t,k) Z_n(t)\, dt\big],$$

which leads to

$$\begin{aligned}
\tilde{\mathbb{E}}\big[z^{Q_{n+1}^i}\big] &= \tilde{\mathbb{E}}\big[z^{Q_n^i}\big] + \tilde{\mathbb{E}}\left[\int_{\mathbb{R}_+} Z_n(t) \left\{\sum_{k \geq 1} \lambda_{Q_n}^{i,+}(t,k) \left[z^{Q_n^i + k} - z^{Q_n^i}\right] + \sum_{k \geq 1} \lambda_{Q_n}^{i,-}(t,k) \left[z^{Q_n^i - k} - z^{Q_n^i}\right]\right\} dt\right], \\
&= \tilde{\mathbb{E}}\big[z^{Q_n^i}\big] + \tilde{\mathbb{E}}\big[\int_{\mathbb{R}_+} Z_n(t) \{\mathscr{Q}_u(t, U_n)\}\, dt\big],
\end{aligned} \tag{25}$$

with $\mathscr{Q}_u(t,U_n) = \sum_{k \geq 1} \lambda_{Q_n}^{i,+}(t,k) \left[z^{Q_n^i + k} - z^{Q_n^i}\right] + \sum_{k \geq 1} \lambda_{Q_n}^{i,-}(t,k) \left[z^{Q_n^i - k} - z^{Q_n^i}\right]$. By rearranging the above terms, we get

$$\mathscr{Q}_u(t,U_n) = z^{Q_n^i - C_{bound}} \sum_{1 \leq k} (z^k - 1) \left[\lambda_{Q_n}^{i,+}(t,k) - \lambda_{Q_n}^{i,-}(t,k) \frac{1}{z^k}\right].$$

We write $\tilde{\mathbb{E}}\big[\int_{\mathbb{R}_+} Z_n(t) \{\mathscr{Q}_u(t,U_n)\}\, dt\big] = T_1 + T_2$ with

$$\begin{aligned}
T_1 &= \tilde{\mathbb{E}}\big[\int_{\mathbb{R}_+} Z_n(t) \mathbf{1}_{Q_n^i \leq C^{bound}} \{\mathscr{Q}_u(t,U_n)\}\, dt\big], \\
T_2 &= \tilde{\mathbb{E}}\big[\int_{\mathbb{R}_+} Z_n(t) \mathbf{1}_{Q_n^i > C^{bound}} \{\mathscr{Q}_u(t,U_n)\}\, dt\big].
\end{aligned}$$

We first handle the term T_1. When $Q_n^i \leq C^{bound}$, the quantity $z^{Q_n^i - C_{bound}} < 1$ is bounded. Additionally, we have $\sum_{e\in E}\lambda_n(e, s+T_n) \geq \underline{\psi} > 0$ under Assumption 3. This ensures that $Z_n(t) \leq e^{-\psi t}$, *a.s.* Thus, there exist $c^1 > 0$ and $d^1 > 0$ such that

$$T_1 \leq \int_{\mathbb{R}_+} e^{-\psi t}\tilde{\mathbb{E}}\big[\{\mathcal{Q}_u(t, U_n)\}\, dt\big] \leq -c^1\tilde{\mathbb{E}}\Big[z^{Q_n^i - C_{bound}}\mathbf{1}_{Q_n^i \leq C^{bound}}\Big] + d^1. \tag{26}$$

In the last inequality we used Assumption 3 again. For the term T_2, we use Assumption 2 and $Z_n(t) \leq e^{-\psi t}$, *a.s*, to deduce that

$$T_2 \leq -\frac{\delta}{\psi}\tilde{\mathbb{E}}\Big[z^{Q_n^i - C_{bound}}\mathbf{1}_{Q_n^i > C^{bound}}\Big]. \tag{27}$$

By combining Inequalities (26) and (27), we have

$$\tilde{\mathbb{E}}\big[\int_{\mathbb{R}_+} Z_n(t)\{\mathcal{Q}_u(t, U_n)\}\, dt\big] \leq -c\tilde{\mathbb{E}}\big[z^{Q_n^i - C_{bound}}\big] + \tilde{\mathbb{E}}\big[d\big],$$

with $c = \min(c^1, \frac{\delta}{\psi})$ and $d = d^1$ which proves the first inequality of Lemma 2. By following the same steps, we also prove the second inequality. This completes the proof. □

IV.C.2 Outline of the proof

To prove the existence of an invariant distribution, we first construct N as a limiting process of the sequence N^m defined in Remark 8. This construction is based on the thinning algorithm. After that, we show, in Steps (ii) and (iii), that N is well defined. Then, we introduce the process $\bar{U}^\infty = \operatorname{ess\,sup}_{t\geq 0}\bar{U}_t$ which dominates $\bar{U}_t$ and prove that is does not explode in Step (iv). This ensures the tightness of the family $\cup_{t\geq 0}\bar{U}_t$. Additionally, the process $\bar{U}$ satisfies the Feller property since E is a countable space and $\mathbb{E}[\|\bar{U}_t\|]$ is uniformly bounded. Thus, we deduce that $\bar{U}$ admits an invariant distribution and complete the proof.

IV.C.3 Proof

Proof of Theorem 1. Let us take N^* and U^* the processes described in Remark 8 with $\nu = \sum_{e\in E}\delta_e$. For clarity, we forget the dependence of $\mathbb{E}_{\mathbf{x}}[.]$ on the initial condition $\mathbf{x} \in W_0$.

Step (i): In this step, we prove that the process N, defined by Equation (3), exists as a limiting process of the sequence N^m. To do so, we first introduce some notations. We define recursively the processes λ^m and N^m as in Remark 8. Note that $U^m = (Q^{m1}, Q^{m2}, S^m)$ can be decomposed in the following way:

$$Q_t^{mi} = Q_t^{mi,+} - Q_t^{mi,-}, \quad S_t^m = S_t^{m+} - S_t^{m-}, \tag{28}$$

with

$$Q_t^{mi,+} = \sum_{T^m < t}\Delta Q_t^{mi}\mathbf{1}_{\Delta Q_t^{mi}>0}, \quad Q_t^{mi,-} = \sum_{T^m<t}\Delta Q_t^{mi}\mathbf{1}_{\Delta Q_t^{mi}<0},$$

$$S^{m+}{}_t = \sum_{T^m<t}\Delta S^m{}_t\mathbf{1}_{\Delta S^m{}_t>0}, \qquad S^{m-}{}_t = \sum_{T^m<t}\Delta S^m{}_t\mathbf{1}_{\Delta S^m{}_t<0},$$

with $i \in \mathbb{B}$ and $\Delta Z_t = Z_t - Z_{t^-}$ for any process Z. For all $\omega \in \Omega$, each one of the processes N^m, λ^m, $Q^{m\,i,\pm}$ and $S^{m\,\pm}$ is non decreasing with m by induction. Hence, they admit limiting processes N, λ, $Q^{1(2),\pm}$ and $S^{\pm}$. This implies that U^m converges towards U. To ensure that N admits λ as an intensity, we need to prove that $\sum_{e\in E}\lambda_t(e)$ and U are both finite $a.s$, see Steps (ii)-(iii).

Step (ii): In this step, we prove by induction on m that $\sup_t \mathbb{E}[\sum_{e\in E}\lambda_t^m(e)]$ is uniformly bounded which ensures that $\sup_t \mathbb{E}[\sum_{e\in E}\lambda_t(e)]$ is finite and that $\sum_{e\in E}\lambda_t(e)$ does not explode. We write $\lambda_n^m(e,t) = \lambda_t^m(e)\mathbf{1}_{T_n^m<t\le T_{n+1}^m}$. For $m=0$, we have $\mathbb{E}[\lambda_n^m(t,e)] = 0$ since $\lambda_t^m(e) = 0$ for any $t \ge 0$. We have by construction

$$\begin{aligned} &\mathbb{E}[\lambda_n^{m+1}(e,t)] = \mathbb{E}[\lambda_n^m(e,t)], \quad &&\text{when } n \le m,\\ &\mathbb{E}[\lambda_n^{m+1}(e,t)] = 0, &&\text{when } n > m+1. \end{aligned}$$

for any $t \ge 0$. Thus, we only need to study the case $n = m+1$. Using Remark 8 and Assumption 1, we have

$$\begin{aligned} \sup_t \mathbb{E}[\lambda_{m+1}^{m+1}(e,t)] &\le c(e) + d(e)\sup_t \mathbb{E}[\Big(\sum_{T_i^m<t}\bar{\phi}(e,t-T_i^m,X_i^m)\Big)^{n_\psi}]\\ &= c(e) + d(e)\sup_t \sum_{\{\mathbf{k}_m\}\in\mathscr{P}(n_\psi)}\binom{n_\psi}{\mathbf{k}_k}\sum_{\mathbf{x}\in E^k}\int_{(-\infty,t)^k}\prod_{i=1}^k \bar{\phi}^{k_i}(e,t-s_i,x_i)\mathbb{E}[dN_{s_1}^m\dots dN_{s_k}^m], \end{aligned}$$

with $\bar{\phi}(e,t,x) = \sup_{u\in\mathbb{U}}\phi(e,u,t,x)$ and $\binom{n_\psi}{\mathbf{k}_k} = \frac{n_\psi!}{k_1!\dots,k_k!}$. Using the above equation and the Brascamp-Lieb inequality, we have

$$\begin{aligned} \sup_t \mathbb{E}[\lambda_{m+1}^{m+1}(e,t)] &\le c(e) + d(e)\sup_t \sum_{\mathbf{k}_k\in\mathscr{P}(n_\psi)}\binom{n_\psi}{\mathbf{k}_k}\sum_{\mathbf{x}\in E^k}\int_{(-\infty,t)^k}\prod_{i=1}^k \bar{\phi}^{k_i}(e,t-s_i,x_i)(\sup_{t,n}\mathbb{E}[\lambda_n^m(x_i,t)])^{1/k}\,ds_i,\\ &= c(e) + \bar{\lambda}^m d(e)\sum_{\mathbf{k}_k\in\mathscr{P}(n_\psi)}\binom{n_\psi}{\mathbf{k}_k}\sum_{\mathbf{x}\in E^k}\int_{\mathbb{R}_+^k}\prod_{i=1}^k \bar{\phi}^{k_i}(e,s_i,x_i)\,ds_i,\\ &\le c(e) + q\bar{\lambda}^m, \end{aligned} \tag{29}$$

with $\bar{\lambda}^m = \sup_{t,e,n}\mathbb{E}[\lambda_n^m(e,t)]$ and $q = \sup_e\{d(e)\sum_{\mathbf{k}_k\in\mathscr{P}(n_\psi)}\binom{n_\psi}{\mathbf{k}_k}\int_{\mathbb{R}_+^k}\prod_{i=1}^k \phi^{*\,k_i}(e,s_i)\,ds_i\}$ where ϕ^* is defined in Assumption 1. Using (29), we deduce that

$$\bar{\lambda}^{m+1} \le \frac{c}{1-q} + q^{m+1}\bar{\lambda}^0 = \bar{x}.$$

Since $q<1$ under Assumption 1, it ensures that $\bar{\lambda} = \sup_m \bar{\lambda}^m$ is finite. To complete the proof, we use (29) and Assumption 3, to get the following inequality:

$$\sup_t \mathbb{E}[\sum_{e\in E}\lambda_t^m(e)] \le c^* + \bar{\lambda}\sum_{e\in E,\,\mathbf{k}_k\in\mathscr{P}(n_\psi)} d(e)\binom{n_\psi}{\mathbf{k}_k}\sum_{\mathbf{x}\in E^k}\int_{\mathbb{R}_+^k}\prod_{i=1}^k \bar{\phi}^{k_i}(e,t-s_i,x_i)\,ds_i < \infty.$$

Step (iii): We write $U_n^m = (Q_n^{m\,1}, Q_n^{m\,2}, S_n^m) = U_{T_n}^m$. We prove here that $\mathbb{E}[Q_n^{m\,i}]$ and $\mathbb{E}[S_n^m]$ are uniformly bounded for all $m \geq 0$ and $n \geq 0$ to ensure that S and Q^i do not explode. Let us prove that

$$\mathbb{E}\big[z^{Q_{n+1}^{m\,i}}\big] \leq \lambda \mathbb{E}\big[z^{Q_n^{m\,i}}\big] + B, \quad \forall n \leq m-1,\, m \geq 1. \tag{30}$$

with $z \leq \min(z_0, z_1)$ and z_0 and z_1 are respectively defined in Assumption 2 and 3, $\lambda < 1$ and $B \geq 0$. Let $m \geq 1$, we have by construction

$$\mathbb{E}[z^{Q_{n+1}^{m+1\,i}}] = \mathbb{E}[z^{Q_{n+1}^{m\,i}}], \quad \text{when } n \leq m-1.$$

Thus, we only need to investigate the case $n = m$. This is proved in Lemma 2. Using Inequality (30), we get

$$\mathbb{E}\big[z^{Q_n^{m\,i}}\big] \leq \frac{B}{1-\lambda} + \lambda^n z^{Q_0^{m\,i}}, \quad \forall n \leq m, \tag{31}$$

with $z^{Q_0^{m\,i}}$ fixed. Thus, $\mathbb{E}\big[Q_n^{m\,i}\big]$ is uniformly bounded. Using similar lines of argument, we also have $\mathbb{E}[S_n^m]$ uniformly bounded. Hence, the limiting processes U does not explode.

Step (iv): First, note that the process N is well defined since λ_t is locally integrable, see Step (ii)-(iii) and [68]. Additionally, we can construct it pathwise using the thinning algorithm, see Remark 8.

Let $\bar{U}_s$ be the process described in Theorem 1 and for which we just proved the existence. This process is dominated by the process $\bar{U}^\infty = (U^\infty, \lambda^\infty) = \operatorname{ess\,sup}_{s\geq 0} \bar{U}_s$. In this part, we prove that both $\mathbb{E}[U^\infty]$ and $\mathbb{E}[\lambda^\infty]$ are finite.

First, we prove that $\mathbb{E}[U^\infty] < \infty$. Let $\lambda < \rho < \rho^1 < 1$, $C > 0$, $\mathscr{S}$ the set $\mathscr{S} = \{u \in \mathbb{U};\, u > C,\, c.w.\}$ where $c.w$ means component-wise and $\mathcal{S}$ a set $\mathcal{S} \in \mathscr{U} \subset \mathscr{S}$. Since $U_n \mathbf{1}_{U_n \in \mathcal{S}^c}$ is bounded $a.s$, we only need to show $\mathbb{E}[U^{\infty,\mathcal{S}}]$ is finite with $U^{\infty,\mathcal{S}} = \operatorname{ess\,sup}_{n\in\mathbb{N}} U_n^{\mathcal{S}}$ and $U_n^{\mathcal{S}} = U_n \mathbf{1}_{U_n \in \mathcal{S}}$. Using the Doob's decomposition, we have $U_n^{\mathcal{S}} = M_n^{\mathcal{S}} + A_n^{\mathcal{S}}$ with $M_n^{\mathcal{S}}$ a martingale and $A_n^{\mathcal{S}} = \sum_{k=1}^n \big(\mathbb{E}[U_k^{\mathcal{S}} | \mathscr{F}_{k-1}] - U_{k-1}^{\mathcal{S}}\big)$ a predictable process. Thus, we get

$$\mathbb{E}[U^{\infty,\mathcal{S}}] \leq \mathbb{E}[\operatorname*{ess\,sup}_{n\geq 0} M_n^{\mathcal{S}}] + \mathbb{E}[\operatorname*{ess\,sup}_{n\geq 0} A_n^{\mathcal{S}}], \quad c.w.$$

The Doob's inequality and Fatou's Lemma ensure that $\mathbb{E}[\sup_{n\geq 0} M_n^{\mathcal{S}}] \leq 2 \lim_{n\to\infty} \mathbb{E}[M_n^{\mathcal{S}\,2}]^{\frac{1}{2}}$, $c.w.$ Using the martingale property of $M_n^{\mathcal{S}}$ and the Doob's decomposition of $U_n^{\mathcal{S}}$, we find

$$\mathbb{E}[(M_n^{\mathcal{S}})^2] - \mathbb{E}[(M_0^{\mathcal{S}})^2] = \sum_{k=1}^n \mathbb{E}[(M_k^{\mathcal{S}} - M_{k-1}^{\mathcal{S}})^2], \qquad M_k^{\mathcal{S}} - M_{k-1}^{\mathcal{S}} = U_k^{\mathcal{S}} - \mathbb{E}[U_k^{\mathcal{S}} | \mathscr{F}_{k-1}],\, c.w.$$

We have

$$\mathbb{E}[(M_k^{\mathcal{S}} - M_{k-1}^{\mathcal{S}})^2] = \mathbb{E}[(U_k^{\mathcal{S}} - \mathbb{E}[U_k^{\mathcal{S}} | \mathscr{F}_{k-1}])^2] \leq 2\big(\mathbb{E}[(U_k^{\mathcal{S}})^2] + \mathbb{E}[\mathbb{E}[(U_k^{\mathcal{S}})^2 | \mathscr{F}_{k-1}]]\big) \leq 4\mathbb{E}[(U_k^{\mathcal{S}})^2],\, c.w.$$

Let us prove that $\sum_{k\geq 0}\mathbb{E}[(U_k^{\mathcal{S}})^2]<\infty$. Using Lemma 2 and by taking $\mathcal{O}=\{(T_j,X_j)_{j\leq 0}\in W_0;\, X_j=(n_j,t_j,b_j,\tilde{u}_j,u_j,a_j)\in E \text{ and } u_0\geq C,\, c.w.\}$, we have

$$\mathbb{E}\big[V_C(U_{n+1})\mathbf{1}_{U_{n+1}\in\mathcal{S},U_n\in\mathcal{S}}\big]\leq\mathbb{E}\big[V_C(U_{n+1})\mathbf{1}_{U_n\in\mathcal{S}}\big]\leq\lambda\mathbb{E}\big[V_C(U_n)\mathbf{1}_{U_n\in\mathcal{S}}\big]+\mathbb{E}\big[B\mathbf{1}_{U_n\in\mathcal{S}}\big]. \tag{32}$$

By following the same lines of arguments used to prove (25) in Lemma 2 and basic approximations, we have the following inequality:

$$\mathbb{E}\big[z^{Q^i_{n+1}-C}\mathbf{1}_{\{U_{n+1}\in\mathcal{S},U_n\in\mathcal{S}^c\}}\big]\leq\mathbb{E}\big[z^{Q^i_n-C}\mathbf{1}_{\{U_{n+1}\in\mathcal{S},U_n\in\mathcal{S}^c\}}\big]+\mathbb{E}\big[\int_{\mathbb{R}_+}Z_n(t)\mathbf{1}_{\{U_{n+1}\in\mathcal{S},U_n\in\mathcal{S}^c\}}\{\mathcal{Q}_u(t,U_n)\}\,dt\big],$$

In the set $\{U_n\in\mathcal{S}^c\}$, we have $Q^i_n\leq C$ which implies $z^{Q^i_n-C}<1$. Additionally, we have $\sum_{e\in E}\lambda_n(e,s+T_n)\geq\underline{\psi}>0$ under Assumption 3. This ensures that $Z_n(t)\leq e^{-\psi t}$, *a.s.* Thus, using Assumption 3, there exists B^1 such that

$$\mathbb{E}\big[z^{Q^i_n-C}\mathbf{1}_{\{U_{n+1}\in\mathcal{S},U_n\in\mathcal{S}^c\}}\big]+\mathbb{E}\big[\int_{\mathbb{R}_+}Z_n(t)\mathbf{1}_{\{U_{n+1}\in\mathcal{S},U_n\in\mathcal{S}^c\}}\{\mathcal{Q}_u(t,U_n)\}\,dt\big]\leq\mathbb{E}[B^1\mathbf{1}_{\{U_{n+1}\in\mathcal{S},U_n\in\mathcal{S}^c\}}], \tag{33}$$

We take $C\geq C^*=\max(\log(\frac{2B}{\rho-\lambda}+1),\log(\frac{B^1}{1-\rho^1}),C_{bound})$ to ensure that

$$\begin{cases}\big[B-(\rho-\lambda)V_C(U_n)\big]\mathbf{1}_{U_n\in\mathcal{S}}<0, & a.s.\\ \big[B^1-(1-\rho^1)V_C(U_n+1)\big]\mathbf{1}_{U_{n+1}\in\mathcal{S}}<0, & a.s.\end{cases}$$

By combining Inequalities (32) and (33) and taking $C\geq C^*$, we deduce that

$$\begin{aligned}\rho^1\mathbb{E}\big[V_C(U_{n+1})\mathbf{1}_{U_{n+1}\in\mathcal{S}}\big]&\leq\rho\mathbb{E}\big[V_C(U_n)\mathbf{1}_{U_n\in\mathcal{S}}\big]+\mathbb{E}\big[\big(B-(\rho-\lambda)V_C(U_n)\big)\mathbf{1}_{U_n\in\mathcal{S}}\big]\\&\quad+\mathbb{E}\big[\big(B^1-(1-\rho^1)V_C(U_{n+1})\big)\mathbf{1}_{U_{n+1}\in\mathcal{S}}\big],\\&\leq\rho\mathbb{E}\big[V_C(U_n)\mathbf{1}_{U_n\in\mathcal{S}}\big],\end{aligned}$$

which ensures that $\mathbb{E}\big[V_C(U_{n+1})\mathbf{1}_{U_{n+1}\in\mathcal{S}}\big]\leq r\mathbb{E}\big[V_C(U_n)\mathbf{1}_{U_n\in\mathcal{S}}\big]$ with $r=\frac{\rho}{\rho^1}<1$. Since $(U_k^{\mathcal{S}})^2\leq c_1V_C(U_k^{\mathcal{S}})$, this proves that $\sum_{k\geq 0}\mathbb{E}[(U_k^{\mathcal{S}})^2]<c_1\sum_{k\geq 0}\mathbb{E}[V_C(U_k^{\mathcal{S}})]\leq\frac{c_1}{1-\rho}<\infty$. Hence, we get $\mathbb{E}[\operatorname{ess\,sup}_{n\geq 0}M_n^{\mathcal{S}}]\leq\big(\frac{c_1}{1-\rho}\big)^{\frac{1}{2}}$, *c.w* .
We also have

$$A_n^{\mathcal{S}}\leq\tilde{A}_n^{\mathcal{S}}=\sum_{k=1}^n|\mathbb{E}[U_k^{\mathcal{S}}|\mathcal{F}_{k-1}]-U_{k-1}^{\mathcal{S}}|\leq 2\sum_{k=1}^n\mathbb{E}[|U_k^{\mathcal{S}}|],\ c.w,$$

with $\tilde{A}_n^{\mathcal{S}}$ a component-wise non-decreasing process. Since $\mathbb{E}[|U_k^{\mathcal{S}}|]\leq\big(\mathbb{E}[(U_k^{\mathcal{S}})^2]\big)^{\frac{1}{2}}$, we get $\mathbb{E}[\tilde{A}_n]\leq\big(\frac{c_1}{1-\rho}\big)^{\frac{1}{2}}$. Hence, we deduce that $\mathbb{E}[\operatorname{ess\,sup}_{n\geq 0}A_n^{\mathcal{S}}]\leq\big(\frac{c_1}{1-\rho}\big)^{\frac{1}{2}}$, *c.w* which ensures that $\mathbb{E}[U^{\infty,\mathcal{S}}]<\infty$.
Second, we prove that $\mathbb{E}[\lambda^\infty]$ is finite. Let $t\geq 0$ and $\mathcal{T}=\{t_0=0<t_1<\ldots<t_n=t\}$ be a partition of $[0,t]$. Using the monotone convergence theorem, we have

$$\mathbb{E}[\sum_{k=1}^n|\lambda_{t_k}-\lambda_{t_{k-1}}|]\leq\mathbb{E}[\frac{\sum_{k=1}^n(t_k-t_{k-1})}{t}|\tilde{\lambda}_{t_k}-\tilde{\lambda}_{t_{k-1}}|]=\mathbb{E}[\frac{\int_0^t f^{\mathcal{T}}\,\mathrm{d}s}{t}]\leq\frac{\int_0^t\mathbb{E}[f^{\mathcal{T}}]\,\mathrm{d}s}{t},$$

with $f^{\mathcal{T}} = \sum_{k=1}^{n} |\lambda_{t_k} - \lambda_{t_{k-1}}| \mathbf{1}_{t_{k-1} \le t < t_k}$. Since $\mathbb{E}[|\lambda_{t_k} - \lambda_{t_{k-1}}|] \le 2 \sup_t \mathbb{E}[|\lambda_t|] \le \frac{c}{1-q} < \infty$, we get

$$\mathbb{E}[\sum_{k=1}^{n} |\lambda_{t_k} - \lambda_{t_{k-1}}|] \le \frac{c}{1-q} < \infty.$$

We can then apply Bichteler-Dellacherie theorem to write $\lambda_t = M_s + A_s$ with M_s a martingale and A_s a predictable process with almost surely finite variation over finite time intervals such that

$$\mathbb{E}[\mathbf{var}_t(\lambda)] = \mathbb{E}[\mathbf{var}_t(M)] + \mathbb{E}[\mathbf{var}_t(A)],$$

where $\mathbf{var}_t(Z)$ is the variation of the process Z over the interval $[0, t]$. Since

$$\mathbb{E}[\lambda^{\infty}] \le \mathbb{E}[\operatorname*{ess\,sup}_s M_s] + \mathbb{E}[\operatorname*{ess\,sup}_s A_s], \quad \operatorname*{ess\,sup}_{s \le t} M_s \le \mathbf{var}_t(M), \quad \operatorname*{ess\,sup}_{s \le t} A_s \le \mathbf{var}_t(A),$$

and $\sup_t \mathbb{E}[\mathbf{var}_t(\lambda)] < \infty$, we deduce that $\mathbb{E}[\lambda^{\infty}] < \infty$. Finally, we have $\mathbb{E}[\|\bar{U}_t\|] \le \mathbb{E}[\|\bar{U}^{\infty}\|] < \infty$, for all $t \ge 0$. Thus, the family $\cup_{t \ge 0} \bar{U}_t$ is tight. Moreover, the process $\bar{U}_t$ satisfies the Feller property since $\mathbb{U}$ and E are countable states and $\mathbb{E}[\|\bar{U}_t\|]$ is uniformly bounded. Thus the process $\bar{U}$ admits an invariant distribution μ which completes the proof. □

IV.D Proof of Theorem 2

IV.D.1 Preliminary result

Lemma 3. *Let $(\mathcal{F}_n)_{n \ge 0}$ be a sequence of σ-algebras such that $\mathcal{F}_n \underset{n \to \infty}{\to} \mathcal{F}_\infty$ with $\mathcal{F}_\infty$ a σ-algebra and $(X_n)_{n \ge 0}$ be a sequence of random variables valued in $\mathbb{R}$ such that $X_n \underset{n \to \infty}{\to} X$, a.s, X_n is $\mathcal{F}_\infty$-measurable, X is $\mathcal{F}_\infty$-measurable and $\sup_n \mathbb{E}[X_n^2] < \infty$. Then, we have*

$$\mathbb{E}\big[X_n | \mathcal{F}_n\big] \underset{n \to \infty}{\to} X, \quad a.s.$$

Remark 23. *In the above Lemma 3, we can replace the condition $\sup_n \mathbb{E}[X_n^2] < \infty$ by the condition $\mathbb{E}[\sup_n X_n] < \infty$ and recover the same result.*

Proof of Lemma 3. Let m and n be two positive integers. We write $X_n^m = \mathbb{E}[X_m | \mathcal{F}_n]$.

Step (i): Since $\sup_n \mathbb{E}[X_n^2] < \infty$, we can apply a conditional dominated convergence theorem to show that $X_n^m \underset{m \to \infty}{\to} X_n = \mathbb{E}[X | \mathcal{F}_n]$, *a.s.*

Step (ii): Since $\mathcal{F}_\infty = \lim_{n \to \infty} \mathcal{F}_n$, there exists a sequence $(A_n)_{n \ge 0}$ such that $A_n \in \mathcal{F}_n$ and $A_n \underset{n \to \infty}{\to} A$. By definition, we have

$$\mathbb{E}[X_n \mathbf{1}_{A_n}] = \mathbb{E}[X \mathbf{1}_{A_n}].$$

Note that the family $(X_n)_{n\geq 0}$ is tight. Indeed, using Doob's and Jensen's inequalities, we have

$$\mathbb{E}\big[\sup_{i\leq n}|X_i|\big] \leq \mathbb{E}\big[\big(\sup_{i\leq n}|X_i|\big)^2\big]^{\frac{1}{2}} = \mathbb{E}\big[\sup_{i\leq n}X_i^2\big]^{\frac{1}{2}} \leq 2\mathbb{E}\big[X_n^2\big]^{\frac{1}{2}}.$$

Then, using Fatou's Lemma, we get $\mathbb{E}\big[\sup_{i\leq n}X_i\big] \leq 2(\sup_n \mathbb{E}\big[X_n^2\big])^{\frac{1}{2}} < \infty$ which ensures that $(X_n)_{n\geq 0}$ is tight. Thus, we can extract a sub sequence $(X_{n_k})_{k\geq 0}$ such that $X_{n_k} \underset{k\to\infty}{\to} Z$ *a.s.* Since $\sup_n \mathbb{E}[X_n^2] < \infty$, we can use the dominated convergence theorem to get

$$\mathbb{E}[Z\mathbf{1}_A] = \lim_{k\to\infty}\mathbb{E}[X_{n_k}\mathbf{1}_{A_{n_k}}] = \lim_{k\to\infty}\mathbb{E}[X\mathbf{1}_{A_{n_k}}] = \mathbb{E}[X\mathbf{1}_A].$$

Thus, we have $Z = X$, $\mathscr{F}_\infty - a.s.$ Since all the variables X_k are $\mathscr{F}_\infty$-measurable, the variable Z is also $\mathscr{F}_\infty$-measurable for any $n \geq 0$. Given that Z and X are both $\mathscr{F}_\infty$-measurable, we deduce that every accumulation point Z of $(X_n)_{n\geq 0}$ satisfies $Z = X$, *a.s.* Finally, we get $\lim_{\substack{m\to\infty\\ n\to\infty}} X_n^m = X$, *a.s.* and we can use a composition argument, to deduce that $\mathbb{E}\big[X_n|\mathscr{F}_n\big] \underset{n\to\infty}{\to} X$, *a.s.* □

We borrow the following definition from [26].

Definition 3 (Coupling). *Two point processes N and N' couple if and only if*

$$\lim_{t\to\infty}\mathbb{P}\big[N_s = N'_s, \quad \forall s \in (t,\infty)\big] = 1.$$

Lemma 4. *Let N be a point process and λ its intensity. We have*

$$\mathbb{P}[N_s - N_t = 0, \quad \forall s \in (t,\infty)|\mathscr{F}_t] = \mathbb{E}[e^{-\int_t^\infty \lambda_u \mathbf{1}_{A_u}\,ds}|\mathscr{F}_t],$$

with $A_u = \{N_u - N_t = 0\}$ for all $u \geq t$.

Proof. See Lemma 1 in [26]. □

Lemma 5. *Two point processes N and N' which admit respectively λ and λ' as intensities couple if and only if*

$$\int_0^\infty \sup_{e\in E}\mathbb{E}\big[|\lambda_s(e) - \lambda'_s(e)|\big]\,ds < \infty.$$

Proof. Let $\mathscr{F}_t = \mathscr{F}_t^N \vee \mathscr{F}_t^{N'}$. Using the canonical coupling, the point process $|N - N'|$ admits $|\lambda_t - \lambda'_t|$ as an $\mathscr{F}_t$-intensity. Using Lemma (4) and Jensen's Inequality, we have

$$\mathbb{P}\big[\sup_e |N_s(e) - N'_s(e)| = 0, \quad \forall s \in (t,\infty)\big] \geq \mathbb{E}[e^{-\int_t^\infty \sup_e |\lambda_s(e)-\lambda'_s(e)|\,ds}] \geq e^{-\int_t^\infty \sup_e \mathbb{E}[|\lambda_s(e)-\lambda'_s(e)|]\,ds}.$$

Since $\int_0^\infty \sup_e \mathbb{E}\big[|\lambda_s(e) - \lambda'_s(e)|\big]\,ds < \infty$, we have $\int_t^\infty \sup_e \mathbb{E}\big[|\lambda_s(e) - \lambda'_s(e)|\big]\,ds \underset{t\to\infty}{\to} 0$ which implies that

$$\mathbb{P}\big[\sup_e |N_s(e) - N'_s(e)| = 0, \quad \forall s \in (t,\infty)\big] \underset{t\to\infty}{\to} 1.$$

This completes the proof. □

IV.D.2 Uniqueness

IV.D.2.1 Outline of the proof

Let $N^\infty = (T_i^\infty, X_i^\infty)$ be the stationary process constructed in Theorem 1 and $N = (T_i, X_i)$ be a point process whose intensity satisfies (2). We write λ (resp. λ^∞) for the intensity of N (resp. N^∞). To prove the uniqueness of the invariant distribution, we only need to show that $\int_0^\infty \sup_{e\in E} \mathbb{E}\big[|\lambda_s(e) - \lambda_s^\infty(e)|\big]\, ds < \infty$, see Lemma 5. To do so, we first show that $(U_n)_{n\geq 0}$ is f-geometrically ergodic, see Lemma 8. The proof of this result requires Lemmas 6 and 7. Using this result, we prove, in Lemma 9, that $f(t) = \sup_e \mathbb{E}\big[|\lambda_t(e) - \lambda_t^\infty(e)|\big]$ satisfies the following inequality:

$$f(t) \leq u(t) + c_3 G\Big(\int_0^t \bar{h}(t-s) f(s)\, ds\Big),$$

with $u(t) = c_2\mathbb{E}[|||U_t - U_t^\infty|||] + c_1\mathbb{E}\big[\|U_t - U_t^\infty\|^{\beta p}\big]^{\frac{1}{\beta p}}$, $G(t) = t^{\frac{1}{\beta}}$ and $\bar{h}(t) = \sup_{e,u,x} \phi\big(e,u,t,x\big)$ with c_1, c_2, c_3, $\beta > 1$ and $p > 1$ positive constants. Then, we use Theorem 3 in [14] and the above inequality, to show that $\int_{\mathbb{R}_+} f(t)\, dt < \infty$ which ensures the uniqueness.

IV.D.2.2 Proof

Let $\lambda < 1$ given by Lemma 2 and $\lambda < \rho < 1$. We denote by $s = \{(T_j, X_j)_{j\leq 0} \in W_0; X_j = (n_j, t_j, b_j, \tilde{u}_j, u_j, a_j) \in E$ and $V(u_0) \leq \frac{2B}{\rho-\lambda} + 1\}$ and by α a set $\alpha \in \mathscr{W}_0 \subset s$. We have the following lemma.

Lemma 6. *Under Assumptions 2 and 3, the function $f = V + 1$ with V defined in Equation (24) and $r > 1$ such that*

$$\sup_{\mathbf{x}\in W} \mathbb{E}_{\mathbf{x}}\Big[\sum_{n=1}^{\tau_\alpha} f(U_n) r^n\Big] < \infty.$$

Proof. The proof is similar to Theorem 6.3 in [89]. □

Let $\mathscr{F}_n$ and $\mathscr{F}_{l\leq j\leq n}$ be respectively defined in the following way $\mathscr{F}_n = \sigma\big(T_j \times X_j, \forall\, j \leq n\big)$, $\mathscr{F}_{l\leq j\leq n} = \sigma\big(T_j \times X_j, \forall\, l \leq j \leq n\big)$. We also write p_k^n as follows:

$$p_k^n(u) = |\mathbb{P}[U_n = u|\mathscr{F}_{k\leq j\leq n-1}\big] - \mathbb{P}[U_n = u|\mathscr{F}_{j\leq n-1}]|, \quad \forall\, n \in \mathbb{N}, \forall\, k \leq n-1, \forall\, u \in \mathbb{U}.$$

Lemma 7. *Under Assumptions 1, 3 and 4, we have*

$$p_k^n = \sup_{u\in\mathbb{U}} p_k^n(u) \underset{k\to\infty}{\to} 0, \quad a.s. \tag{34}$$

Proof. Using Lemma 1, we have

$$\begin{aligned} p_k^n(u) &= |\mathbb{E}[\int_{\mathbb{R}_+} Z_n(t)\lambda_n(u,t)\, dt|\mathscr{F}_{k\leq j\leq n-1}] - \mathbb{E}[\int_{\mathbb{R}_+} Z_n(t)\lambda_n(u,t)\, dt|\mathscr{F}_{j\leq n-1}]| \\ &= |\mathbb{E}[\int_{\mathbb{R}_+} Z_n(t)\lambda_n(u,t)\, dt|\mathscr{F}_{k\leq j\leq n-1}] - \int_{\mathbb{R}_+} Z_n(t)\lambda_n(u,t)\, dt|, \end{aligned}$$

with $\lambda_n(u,t) = \sum_{e\in E(U_{n-1},u)} \lambda_n(e,t)$, $\lambda_n(e,t) = \psi(e, U_{n-1}, t+T_{n-1}, r_n(t))$, $r_n(t) = \sum_{j\le n-1} \phi(e, U_{n-1}, t+T_{n-1}-T_j, X_j)$ and $Z_n(t) = e^{-\left[\int_0^t \sum_e \lambda_n(e,s)\,ds\right]}$.
Since $p_k = \sup_{u\in\mathbb{U}} p_k^n(u)$, we can construct a sequence $(u_j)_{j\ge 0}$ such that $p_k^n(u_j) \underset{j\to\infty}{\to} p_k^n$, *a.s.* We write $u_j = (q_j^1, q_j^2, s_j)$. Without loss of generality, we can consider that $(q_j^1)_{j\ge 0}$ is monotonic by taking a sub-sequence of $(q_j^1)_{j\ge 0}$. Hence, there exists a limiting process q_∞^1 such that $q_j^1 \underset{j\to\infty}{\to} q_\infty^1$, *a.s.* By repeating this argument several times, we can always construct (u_j) such that

$$p_k^n(u_j) \underset{j\to\infty}{\to} p_k, \quad u_j \underset{j\to\infty}{\to} u_\infty, \qquad a.s.$$

Let us prove that $\lambda_n(u_j,t) \underset{j\to\infty}{\to} \lambda_n(u_\infty,t)$, *a.s.* To do so, we distinguish two sets $A_1 = \{w\in\Omega;\, u_\infty(w) < \infty\}$ and $A_2 = \{w\in\Omega;\, u_\infty(w) = \infty\}$. When $u_\infty < \infty$, we have $u_j = u_\infty$ for j large enough since $\mathbb{U}$ is countable. This ensures that $E(U_{n-1}, u_j) = E(U_{n-1}, u_\infty)$, *a.s* for j large enough. Thus, we get

$$\lambda_n(u_j,t)\mathbf{1}_{A_1} \underset{j\to\infty}{\to} \sum_{e\in E(U_{n-1},u_\infty)} \psi(e, U_{n-1}, t+T_{n-1}, r_n(t))\mathbf{1}_{A_1}, \quad a.s.$$

When $u_\infty = \infty$, we have $\sum_{e\in E(U_{n-1},u_\infty)} \lambda_n(e,t) = 0$ since $E(U_{n_\infty-1}, u_\infty) = \varnothing$. Using $\sum_{e\in E} \lambda_{n_j}(e,t) < \infty$, *a.s*, see Step (ii) in the proof of Theorem 1, we deduce that $\sum_{e\in E(U_{n_j-1},C^c)} \lambda_{n_j}(e,t) \underset{c\to\infty}{\to} 0$, *a.s* with $C^c = \{u\in\mathbb{U};\, u > c,\, c.w\}$, $c > 0$ and $c.w$ means component-wise. Since $E(U_{n_j-1}, u_j) \subset E(U_{n_j-1}, C^c)$ for j large enough, we get $\sum_{e\in E(U_{n_j-1},u_j)} \lambda_{n_j}(e,t) \underset{j\to\infty}{\to} 0$, *a.s* which means that

$$\lambda_n(u_j,t)\mathbf{1}_{A_2} \underset{j\to\infty}{\to} \sum_{e\in E(U_{n-1},u_\infty)} \psi(e, U_{n-1}, t+T_{n-1}, r_n(t))\mathbf{1}_{A_2} = 0, \quad a.s,$$

and proves $\lambda_n(u_j,t) \underset{j\to\infty}{\to} \lambda_n(u_\infty,t)$, *a.s.*
Additionally, we have $\mathbb{E}[\sup_{n,s}\sum_e \lambda_n(e,s)] < \infty$, see Step (iv) in the proof of Theorem 1. Thus, we get $\mathbb{E}[\sup_{n,u,s} \lambda_n(u,s)] < \infty$. Since $\sum_e \lambda_n(e,s) \ge \underline{\psi}$ under Assumption 3, we have $Z_n(t) \le e^{-\underline{\psi} t}$, *a.s.* Then, we can apply the dominated convergence theorem to show that

$$\int_{\mathbb{R}_+} Z_n(t)\lambda_n(t,u_j)\,dt \underset{j\to\infty}{\to} \int_{\mathbb{R}_+} Z_n(t)\lambda_n(t,u_\infty)\,dt, \quad a.s.$$

Furthermore, we have

$$\mathbb{E}\big[\sup_j \int_{\mathbb{R}_+} Z_n(t)\lambda_n(u_j,t)\,dt\big] \le \mathbb{E}\big[\int_{\mathbb{R}_+} \sup_j e^{-\underline{\psi} t}\lambda_n(u_j,t)\,dt\big] \overset{\text{Fubini}}{=} \int_{\mathbb{R}_+} e^{-\underline{\psi} t}\mathbb{E}\big[\sup_j \lambda_n(u_j,t)\big]\,dt,$$

with $\mathbb{E}\big[\sup_j \lambda_n(u_j,t)\big] < \infty$. Hence, we can use the conditional dominated convergence to show

$$\mathbb{E}\big[\int_{\mathbb{R}_+} Z_n(t)\lambda_n(u_j,t)\,dt|\mathscr{F}_{k\le r\le n-1}\big] \underset{j\to\infty}{\to} \mathbb{E}\big[\int_{\mathbb{R}_+} Z_n(t)\lambda_n(u_\infty,t)\,dt|\mathscr{F}_{k\le r\le n-1}\big].$$

Finally, since $\mathscr{F}_{k\le r\le n-1} \underset{k\to\infty}{\to} \mathscr{F}_{r\le n-1}$, we can apply Lemma 3 to deduce that

$$p_k^n \underset{k\to\infty}{\to} 0, \ a.s.$$

This completes the proof. □

Let $\Delta T_n = T_n - T_{n-1}$ be the inter-arrival time between n-th jump and the $n-1$-th jump with T_n the time of the n-th event. Let $N^\infty = (T_i^\infty, X_i^\infty)$ be the stationary process constructed in Lemma 1 and $N = (T_i, X_i)$ be a point process whose intensity satisfies (2). We write $U^\infty = (Q^{1^\infty}, Q^{2^\infty}, S^\infty)$ (resp. $U = (Q^1, Q^2, S)$) for the order book state associated to N^∞ (resp. N). We denote by λ^∞ (resp. λ) the intensity of N^∞ (resp. N). We have the following result.

Lemma 8. *Under Assumptions 1, 2, 3 and 4, the process $(U_n)_{n\ge0}$ is f-geometrically ergodic, see 15.7 in [90], in the sense that there exists $r>1$ such that*

$$\sup_{\mathbf{x}\in W_0}\sum_{n\ge1}\mathbb{E}_{\mathbf{x}}\big[\|f(U_n)-f(U_n^\infty)\|\, r^n\big]<\infty.$$

Proof. Let $P^n(\mathbf{x},A)$ be the probability of being in the set $A=\{(t_k,x_k)_{k\le0}\in W_0;\, x_k=(n_k,t_k,b_k,s_k,\tilde{u}_k,u_k,a_k),\, u_0\in a\}$, $a\in\mathscr{U}$, with $\mathscr{U}$ the σ-algebra generated by the discrete topology on $\mathbb{U}$, after n jumps conditional on $\mathbf{x}=(t_k,x_k)_{k\le0}\in W_0=(\mathbb{R}_+\times E)^{\mathbb{N}^-}$. Let $\mathbf{y}\in W_0$. We write π for the stationary distribution of the process $U_n^\infty=(Q^{1^\infty}{}_n, Q^{2^\infty}{}_n, S_n^\infty)$ and τ_{α^k} for the first entrance time of U to the set $\alpha^k=\{\mathbf{z}\in W_0;\, \mathbf{z}_{-k+1\le j\le0}=\mathbf{y}_{-k+1\le j\le0}\}$. Using the first-entrance last-exit decomposition of $P^n(\mathbf{x},A)$, see Section 8.2 in [90], we have

$$\begin{aligned}
P^n(\mathbf{x},A) &= {}_{\alpha^k}P^n(\mathbf{x},A)+\sum_{j=1}^n\sum_{i=1}^j\Big[\int_{\mathbb{U}^{j-i}_{\mathbf{u},\alpha^k}}\int_{\mathbb{U}^i_{\mathbf{x},\alpha^k}} {}_{\alpha^k}P^i(\mathbf{x},du)P^{j-i}(\mathbf{u},dv)\,{}_{\alpha^k}P^{n-j}(\mathbf{v},A)\Big]\\
&= {}_{\alpha^k}P^n(\mathbf{x},A)+\sum_{j=1}^n\sum_{i=1}^j\Big[\int_{\mathbb{U}^{j-i}_{\mathbf{u},\alpha^k}}\int_{\mathbb{U}^i_{\mathbf{x},\alpha^k}} {}_{\alpha^k}P^i(\mathbf{x},du)P^{j-i}(\mathbf{u},dv)\,{}_{\alpha^k}P^{n-j}(\mathbf{y},A)\Big]\\
&+\sum_{j=1}^n\sum_{i=1}^j\Big[\int_{\mathbb{U}^{j-i}_{\mathbf{u},\alpha^k}}\int_{\mathbb{U}^i_{\mathbf{x},\alpha^k}} {}_{\alpha^k}P^i(\mathbf{x},du)P^{j-i}(\mathbf{u},dv)\,|{}_{\alpha^k}P^{n-j}(\mathbf{v},A)-{}_{\alpha^k}P^{n-j}(\mathbf{y},A)|\Big]. \qquad (35)
\end{aligned}$$

with ${}_{\alpha^k}P^n(\mathbf{x},A)=\mathbb{P}[(T_k,X_k)_{k\le0}=\mathbf{x},\, U_n\in A,\, \tau_{\alpha^k}\ge n]$ and $\mathbb{U}^i_{\mathbf{x},\alpha^k}=\{\mathbf{z}\in\alpha^k;\,(\mathbf{z}_k)_{k\le i}=\mathbf{x}\}$. Using $\mathbb{E}_{\mathbf{x}}[\tau_{\alpha^k}]<\infty$ for all $\mathbf{x}\in S$ and the arguments used in the proof of Theorem 10.2.1 in [90], we deduce that the stationary distribution admits the following representation:

$$\pi(A)=\mathbb{E}_{\mathbf{y}}[\tau_{\alpha^k}]^{-1}\mathbb{E}_{\mathbf{y}}[\sum_{j=1}^{\tau_{\alpha^k}}\mathbf{1}_{\bar{U}_n\in A}]=\pi(\alpha^k)\sum_{j=1}^\infty {}_{\alpha^k}P^j(\mathbf{y},A). \qquad (36)$$

By combining (35) and (36), we get

$$\begin{aligned}
P(\mathbf{x},A)-\pi(A) &= {}_{\alpha^k}P^n(\mathbf{x},A)+\Big[\big({}_{\alpha^k}P(\mathbf{x})*P(\alpha^k)-\pi(\alpha^k)\big)*{}_{\alpha^k}P(\mathbf{y})\Big]_n(A)+\pi(\alpha^k)\sum_{j\ge n+1}{}_{\alpha^k}P^j(\mathbf{y},A)\\
&+\big({}_{\alpha^k}P(\mathbf{x})*P(\alpha^k)\big)*\big({}_{\alpha^k}P(\alpha^k)-{}_{\alpha^k}P(\mathbf{y})\big)_n(A). \qquad (37)
\end{aligned}$$

with $*$ the integrated Cauchy product between two sequences which is defined as follows:

$$[u(B) * v(C)]_n(A) = \sum_{i=1}^{n} \int_{\mathbb{U}^i_{B,C}} u_i(B, du)\, v_{n-i}(\mathbf{u}, A), \quad \forall (B, C, A) \in (\mathscr{W}_0)^3$$

with $(u_n)_{n\geq 0}$ and $(v_n)_{n\geq 0}$ two sequences such that $u_n, v_n : (\mathscr{W}_0)^2 \to \mathbb{R}$. Let f be the function defined in Lemma 6, $\pi(f) = \int_{\mathscr{U}} \pi(du) f(u) < \infty$, $\mathbb{E}_{\mathbf{x}}[f(U_n)] = \int_{\mathscr{U}} P^n(\mathbf{x}, du) f(u)$ and $|P^n(\mathbf{x}, .) - \pi|_f = |\mathbb{E}_{\mathbf{x}}[f(U_n)] - \pi(f)|$. Using (37), we have

$$\begin{aligned} |P^n(\mathbf{x}, .) - \pi|_f \leq \mathbb{E}_{\mathbf{x}}[f(U_n)\mathbf{1}_{\tau_{\alpha^k} \geq n}] + [{}_{\alpha^k}P(\mathbf{x}) * P(\alpha^k) - \pi(\alpha^k)] * t_n^f \\ + \pi(\alpha^k) \sum_{j \geq n+1} t_j^f + |{}_{\alpha^k}P(\mathbf{x}) * P(\alpha^k) * \Delta t_n^f|, \end{aligned} \tag{38}$$

with $t_n^f = \mathbb{E}_{\mathbf{y}}[f(U_n)\mathbf{1}_{\tau_{\alpha^k} \geq n}]$ and $\Delta t_n^f(\mathbf{v}) = (\mathbb{E}_{\mathbf{v}}[f(U_n)\mathbf{1}_{\tau_{\alpha^k} \geq n}] - t_n^f)$. To prove geometric ergodicity we have to show

$$\sup_{\mathbf{x}} \sum_{n \geq 1} |P^n(\mathbf{x}, .) - \pi|_f r^n < \infty, \tag{39}$$

with $r > 1$. Let us take $\bar{n} \in \mathbb{N}^*$ and the delay $k(\bar{n}) \in \mathbb{N}$ associated to α^k depending on $\bar{n}$. Using (38), we have

$$\begin{aligned} \sum_{n \geq 1}^{\bar{n}} |P^n(\mathbf{x}, .) - \pi|_f r^n \leq \sum_{n \geq 1}^{\bar{n}} \mathbb{E}_{\mathbf{x}}[f(U_n)\mathbf{1}_{\tau_{\alpha^k} \geq n}] r^n + \sum_{n \geq 1}^{\bar{n}} |\left({}_{\alpha^k}P(\mathbf{x}) * P(\alpha^k) - \pi(\alpha^k)\right) * t_n^f r^n| \\ + \pi(\alpha^k) \sum_{n \geq 1}^{\bar{n}} \sum_{j \geq n+1} t_j^f r^n + \sum_{n \geq 1}^{\bar{n}} [{}_{\alpha^k}P(\mathbf{x}) * P(\alpha^k)] * \Delta t_n^f r^n = \text{(i)} + \text{(ii)} + \text{(iii)} + \text{(iv)}. \end{aligned}$$

The error term (i) can be dominated by

$$\sum_{n \geq 1}^{\bar{n}} \mathbb{E}_{\mathbf{x}}[f(U_n)\mathbf{1}_{\tau_{\alpha^k} \geq n}] r^n \leq \sum_{n \geq 1} \mathbb{E}_{\mathbf{x}}[f(U_n)\mathbf{1}_{\tau_{\alpha^k} \geq n}] r^n = \mathbb{E}_{\mathbf{x}}\Big[\sum_{n=1}^{\tau_{\alpha^k}} f(U_n) r^n\Big]. \tag{40}$$

The error term (iii) can be bounded by

$$\pi(\alpha^k) \sum_{n \geq 1}^{\bar{n}} \sum_{j \geq n+1} t_j^f r^n \leq \pi(\alpha^k) \sum_{n \geq 1} \sum_{j \geq n+1} t_j^f r^n \leq \frac{\pi(\alpha^k)}{r-1} \sup_{\mathbf{v}} \mathbb{E}_{\mathbf{v}}\Big[\sum_{n=1}^{\tau_{\alpha^k}} f(U_n) r^n\Big]. \tag{41}$$

Now we move to the error term (iv). We have

$$\left({}_{\alpha^k}P(\mathbf{x}) * P(\alpha^k)\right) * \Delta t_n^f \leq \sum_{j \leq n,\, i \leq j} [\int_{\mathbb{U}^{j-i}_{\mathbf{u},\alpha^k} \times \mathbb{U}^i_{\mathbf{x},\alpha^k} \times W_0} {}_{\alpha^k}P^i(\mathbf{x}, du) P^{j-i}(\mathbf{u}, dv)\big]\, \Delta_{\alpha^k} P^{n-j}(dw) f(w),$$

with $\Delta_{\alpha^k} P^{n-j}(dw) = |{}_{\alpha^k}P^{n-j}(\mathbf{v}, dw) - {}_{\alpha^k}P^{n-j}(\mathbf{y}, dw)|$. Using Equations (35) and (36), we get

$$\begin{cases} \Delta_{\alpha^k} P^{n-j}(dw) \leq {}_{\alpha^k}P^{n-j}(\mathbf{v}, dw) + {}_{\alpha^k}P^{n-j}(\mathbf{y}, dw), \\ \sum_{j \leq n,\, i \leq j} [\int_{\mathbb{U}^{j-i}_{\mathbf{u},\alpha^k} \times \mathbb{U}^i_{\mathbf{x},\alpha^k} \times \mathscr{W}} {}_{\alpha^k}P^i(\mathbf{x}, du) P^{j-i}(\mathbf{u}, dv)\big]\, {}_{\alpha^k}P^{n-j}(\mathbf{v}, dw) f(w) \leq \mathbb{E}_{\mathbf{x}}\big[f(U_n) r^n\big] < \infty, \\ \sum_{j \leq n,\, i \leq j} [\int_{\mathbb{U}^{j-i}_{\mathbf{u},\alpha^k} \times \mathbb{U}^i_{\mathbf{x},\alpha^k} \times \mathscr{W}} {}_{\alpha^k}P^i(\mathbf{x}, du) P^{j-i}(\mathbf{u}, dv)\big]\, {}_{\alpha^k}P^{n-j}(\mathbf{y}, dw) f(w) \leq \mathbb{E}_{\mathbf{y}}\big[f(U_{n-j}) r^n\big] < \infty, \end{cases}$$

Since $\Delta_{\alpha^k} P^{n-j} \underset{k\to\infty}{\to} 0$, see Lemma 7, the dominated convergence theorem ensures that

$$\big({}_{\alpha^k}P(\mathbf{x}) * P(\alpha^k)\big) * \Delta t_n^f \underset{k\to\infty}{\to} 0.$$

Thus, there exists $\bar{k}(\bar{n})$ such that $\big({}_{\alpha}P(\mathbf{x}) * P(\alpha)\big) * \Delta t_n^f \leq \epsilon(\bar{n})$ for any $k \geq \bar{k}(\bar{n})$. Hence the error term (iv) can be majorated by

$$\sum_{n\geq 1}^{\bar{n}} [{}_{\alpha^k}P(\mathbf{x}) * P(\alpha^k)] * \Delta t_n^f r^n \leq \epsilon(\bar{n}) \frac{r^{\bar{n}+1}-1}{r-1}, \tag{42}$$

which means that we have to choose $\epsilon(\bar{n}) < c_1 \frac{r-1}{r^{\bar{n}+1}-1}$ with c_1 a positive constant. Finally, using the property

$$\lim_{n\to\infty} (u * v)_n = \lim_{n\to\infty} u_n \times \lim_{n\to\infty} v_n, \tag{43}$$

we dominate the error term (ii) by

$$|\sum_{n\geq 1}^{\bar{n}} \big({}_{\alpha^k}P(\mathbf{x}) * P(\alpha^k) - \pi(\alpha^k)\big) * t_n^f r^n| \leq \big(\sum_{n\geq 1} |\big[{}_{\alpha^k}P(\mathbf{x}) * P(\alpha^k)\big]_n(\alpha^k) - \pi(\alpha^k)| r^n\big) \sup_{\mathbf{v}} \mathbb{E}_{\mathbf{v}}\big[\sum_{n=1}^{\tau_{\alpha^k}} f(U_n) r^n\big]. \tag{44}$$

Additionally, we have

$$\begin{aligned}
|\big[{}_{\alpha^k}P(\mathbf{x}) * P(\alpha^k)\big]_n(\alpha^k) - \pi(\alpha^k)| &= |\big[{}_{\alpha^k}P(\mathbf{x}) * (P(\alpha^k) - \pi(\alpha^k))\big]_n(\alpha^k) - \pi(\alpha^k) \sum_{i\geq n+1} {}_{\alpha^k}P^i(\mathbf{x}, \alpha^k)| \\
&= |\big[{}_{\alpha^k}P(\mathbf{x}) * (P(\alpha^k) - P(\mathbf{y}))\big]_n(\alpha^k) + \big[{}_{\alpha^k}P(\mathbf{x}) * (P(\mathbf{y}) - \pi(\alpha^k))\big]_n(\alpha^k) \\
&\quad - \pi(\alpha^k) \sum_{i\geq n+1} {}_{\alpha^k}P^i(\mathbf{x}, \alpha^k)| \\
&\leq \big[{}_{\alpha^k}P(\mathbf{x}) * |P(\mathbf{y}) - P(\alpha^k)|\big]_n(\alpha^k) + \big[{}_{\alpha^k}P(\mathbf{x}) * |P(\mathbf{y}) - \pi(\alpha^k)|\big]_n(\alpha^k) \\
&\quad + \pi(\alpha^k) \sum_{i\geq n+1} {}_{\alpha^k}P^i(\mathbf{x}^u, \alpha^k),
\end{aligned}$$

for any $n \in \mathbb{N}$. Using Equation (43), we get

$$\begin{aligned}
\sum_{n\geq 1}^{\bar{n}} \big[|{}_{\alpha^k}P(\mathbf{x}) * P(\alpha^k) - \pi(\alpha^k)|\big]_n(\alpha^k) r^n &\leq \sum_{n\geq 1}^{\bar{n}} \big[{}_{\alpha^k}P(\mathbf{x}) * |P^{\mathbf{y}}(\alpha^k) - P(\alpha^k)|\big]_n(\alpha^k) r^n \\
&\quad + \sum_{n\geq 1}^{\bar{n}} \big[{}_{\alpha^k}P(\mathbf{x}) * |P(\mathbf{y}) - \pi(\alpha^k)|\big]_n(\alpha^k) r^n + \pi(\alpha^k) \sum_{\substack{n\geq 1 \\ i\geq n+1}} {}_{\alpha^k}P^i(\mathbf{x}, \alpha^k) r^n \\
&\leq \big(\sum_{n\geq 1} {}_{\alpha^k}P^n(\mathbf{x}, \alpha^k) r^n\big)\big(\sum_{n\geq 1} \sup_{\mathbf{w}\in\alpha^k} |P^n(\mathbf{y}, \alpha^k) - P^n(\mathbf{w}, \alpha^k)| r^n\big) \\
&\quad + \big(\sum_{n\geq 1} {}_{\alpha^k}P^n(\mathbf{x}, \alpha^k) r^n\big)\big(\sum_{n\geq 1} |P^n(\mathbf{y}, \alpha^k) - \pi(\alpha^k)| r^n\big) \\
&\quad + \pi(\alpha^k) \sum_{n\geq 1} \sum_{i\geq n+1} {}_{\alpha^k}P^i(\mathbf{x}, \alpha^k) r^n = (1) + (2) + (3).
\end{aligned}$$

The term (2) is bounded by

$$(2) \leq \mathbb{E}_{\mathbf{x}}\big[r^{\tau_\alpha}\big] \sup_{\mathbf{y}} \Big(\sum_{n\geq 1} |P^n(\mathbf{y},\alpha) - \pi(\alpha)| r^n \Big).$$

Since the Kendall theorem ensures that $E_{\mathbf{x}}\big[r^{\tau_{\alpha^k}}\big] < \infty$ and $\sum_{n\geq 1} |P^n(\mathbf{x},\alpha^k) - \pi(\alpha^k)| r^n < \infty$ are equivalent, the quantity (1) is finite if and only if $\sup_{\mathbf{v}} E_{\mathbf{v}}\big[r^{\tau_{\alpha^k}}\big] < \infty$. The term (1) is majorated by

$$(1) \leq \mathbb{E}_{\mathbf{x}}\big[r^{\tau_{\alpha^k}}\big] \Big(\sum_{n\geq 1} \sup_{\mathbf{w}\in\alpha^k} |P^n(\mathbf{w},\alpha^k) - P^n(\mathbf{y},\alpha^k)| r^n \Big).$$

To ensure that the sequence $v(\bar{n}) = \sum_{n\geq 1}^{\bar{n}} \sup_{\mathbf{w}\in\alpha^k} |P^n(\mathbf{w},\alpha^k) - P^n(\mathbf{y},\alpha^k)| r^n$ is bounded, the put a dependence k and $\bar{n}$. Let $\epsilon^1(\bar{n}) > 0$. By following the same arguments used in the proof of Inequality (42), there exists $\bar{k}^1(\bar{n})$ such that for any $k \geq \bar{k}^1(\bar{n})$, we have

$$\sum_{n\geq 1}^{\bar{n}} \sup_{\mathbf{w}\in\alpha^k} |P^n(\mathbf{y},\alpha^k) - P^n(\mathbf{w},\alpha^k)| r^n \leq \epsilon^1(\bar{n}) \frac{r^{\bar{n}+1} - 1}{r - 1}.$$

By taking $\epsilon^1(\bar{n}) \leq c_1 \frac{r-1}{r^{\bar{n}+1}-1}$, we get $(1) \leq c_1 \sup_{\mathbf{x}} \mathbb{E}_{\mathbf{x}}\big[r^{\tau_{\alpha^k}}\big]$. Furthermore, the term (3) can be dominated by $(3) \leq \mathbb{E}_{\mathbf{x}}\big[r^{\tau_{\alpha^k}}\big]$. Thus, we deduce that

$$(ii) \leq c_1 \mathbb{E}_{\mathbf{x}}\big[r^{\tau_\alpha}\big] (1 + \sup_{\mathbf{v}} \sum_{n\geq 1} |P^n(\mathbf{v},\alpha) - \pi(\alpha)| r^n) \sup_{\mathbf{v}} \mathbb{E}_{\mathbf{v}}\Big[\sum_{n=1}^{\tau_{\alpha^k}} f(U_n) r^n \Big]. \tag{45}$$

By combining Inequalities (40), (41), (44) and (45), we have (39) when $\sup_{\mathbf{x}} E_{\mathbf{x}}\big[\sum_{n=1}^{\tau_{\alpha^k}} f(U_n) r^n\big]$ and $\sup_{\mathbf{x}} E_{\mathbf{x}}\big[r^{\tau_{\alpha^k}}\big]$ are both finite. Since $E_{\mathbf{x}}\big[\sum_{n=1}^{\tau_{\alpha^k}} f(U_n) r^n\big] < \infty$ implies $E_{\mathbf{x}}\big[r^{\tau_{\alpha^k}}\big] < \infty$, we only need to prove

$$\mathbb{E}_{\mathbf{x}}\Big[\sum_{n=1}^{\tau_{\alpha^k}} f(U_n) r^n \Big] < \infty.$$

This last inequality is satisfied thanks to Lemma 6. □

Lemma 9. *Under Assumptions 1, 3 and 4, the process $\bar{U}$ is ergodic.*

Proof of Lemma 9. For simplicity, we write c_1, c_2 and c_3 for positive constants and forget the dependence of $\mathbb{E}_{\mathbf{x}}[X]$ on the initial state $\mathbf{x}$ for any random variable X. Let $N^\infty = (T_i^\infty, X_i^\infty)$ be the stationary process constructed in Lemma 1 and $N = (T_i, X_i)$ be a point process whose intensity satisfies (2). We write $U^\infty = (Q^{1\infty}, Q^{2\infty}, S^\infty)$ (resp. $U = (Q^1, Q^2, S)$) for the order book state associated to N^∞ (resp. N). We denote by λ^∞ (resp. λ) the intensity of N^∞ (resp. N). To prove the uniqueness, we need to show that N and N^∞ couple which is satisfied when

$$\int_0^\infty \sup_e \mathbb{E}\big[|\lambda_t(e) - \lambda_t^\infty(e)|\big]\, dt < \infty,$$

thanks to Lemma 5. We write $f(t) = \sup_e \mathbb{E}\big[|\lambda_t(e) - \lambda_t^\infty(e)|\big]$ for any $t \geq 0$.

Step (i): For any $\gamma = \frac{p}{q} > 1$ with $p, q \in \mathbb{N}^*$ and β such that $\frac{1}{\beta} + \frac{1}{\gamma} = 1$. Let us first prove that

$$f(t) \leq u(t) + g_1(t) G\Big(\int_0^t \bar{h}(t-s) f(s)\, ds \Big), \tag{46}$$

with $u(t) = c_3 \mathbb{E}[||U_t - U_t^\infty||] + c_1 \mathbb{E}\big[\|U_t - U_t^\infty\|^{\beta p}\big]^{\frac{1}{\beta p}} [1 + 2B(t)]$, $g_1(t) = c_2(1 + 2B(t))$, $G(t) = t^{\frac{1}{\beta}}$, $\bar{h}(t) = \sup_{e,u,x} \phi\big(e, u, t, x\big)$ and $B(t) = \sup_{0 \leq k \leq n_\psi - 1} \Big[B_k(t)\Big]^{\frac{1}{p\gamma}}$ with $B_k(t)$ defined in Equation (50). The quantities c_1, c_2 and c_3 are positive constants. We have

$$\begin{aligned} f(t) &= \mathbb{E}\big[|\psi(e, U_t, t, r_t) - \psi(e, U_t^\infty, t, r_t^\infty)|\big] \\ &\leq \mathbb{E}\big[|\psi(e, U_t, t, r_t) - \psi(e, U_t^\infty, t, r_t)|\big] + \mathbb{E}\big[|\psi(e, U_t^\infty, t, r_t) - \psi(e, U_t^\infty, t, r_t^\infty)|\big] \\ &= (1) + (2), \end{aligned}$$

with $r_t = \int_0^t \phi(e, U_t, t-s, X_s)\, dN_s$ and $r_t^\infty = \int_0^t \phi(e, U_t^\infty, t-s, X_s^\infty)\, dN_s^\infty$. Let us first handle the term (2). Using Assumption 4, we have

$$\begin{aligned} \mathbb{E}\big[|\psi(e, U_t^\infty, t, r_t) - \psi(e, U_t^\infty, t, r_t^\infty)|\big] &\leq \mathbb{E}\big[|\bar{\psi}(r_t) - \bar{\psi}(r_t^\infty)|\big] \\ &\leq K \mathbb{E}\big[|r_t - r_t^\infty||1 + r_t^{n_1} + r_t^{\infty^{n_1}}|\big] \\ &\leq K \overbrace{\mathbb{E}\big[|r_t - r_t^\infty|^\beta\big]^{\frac{1}{\beta}}}^{(i)} \overbrace{\mathbb{E}\big[|1 + r_t^{n_1} + r_t^{\infty^{n_1}}|^\gamma\big]^{\frac{1}{\gamma}}}^{(ii)}. \end{aligned}$$

The term (i) can be dominated by

$$\begin{aligned} \mathbb{E}\big[|r_t - r_t^\infty|^\beta\big]^{\frac{1}{\beta}} &\leq \mathbb{E}\big[|\int_0^t \phi(e, U_t, t-s, X_s) dN_s - \phi(e, U_t^\infty, t-s, X_s) dN_s^\infty|^\beta\big]^{\frac{1}{\beta}} \\ &\leq 2^{\frac{\beta-1}{\beta}} \mathbb{E}\big[|\int_0^t \phi(e, U_t, t-s, X_s) dN_s - \phi(e, U_t^\infty, t-s, X_s) dN_s|^\beta\big]^{\frac{1}{\beta}} \\ &+ 2^{\frac{\beta-1}{\beta}} \mathbb{E}\Big[\Big(\int_0^t \bar{h}(t-s) |dN_s - dN_s^\infty|\Big)^\beta\Big]^{\frac{1}{\beta}} \\ &\leq 2^{\frac{\beta-1}{\beta}} \mathbb{E}\big[\|U_t - U_t^\infty\|^\beta |\int_0^t \tilde{h}(e, t-s, X_s) dN_s|^\beta\big]^{\frac{1}{\beta}} + 2^{\frac{\beta-1}{\beta}} \big[\int_0^t \bar{h}(t-s) f(s)\, ds\big]^{\frac{1}{\beta}} \\ &\leq 2^{\frac{\beta-1}{\beta}} \mathbb{E}\big[\|U_t - U_t^\infty\|^{\beta p}\big]^{\frac{1}{\beta p}} \mathbb{E}\big[|\int_0^t \tilde{h}(e, t-s, X_s) dN_s|^{\beta q}\big]^{\frac{1}{\beta q}} + 2^{\frac{\beta-1}{\beta}} \big[\int_0^t \bar{h}(t-s) f(s)\, ds\big]^{\frac{1}{\beta}} \\ &= c_1 \mathbb{E}\big[\|U_t - U_t^\infty\|^{\beta p}\big]^{\frac{1}{\beta p}} + c_2 \big[\int_0^t \bar{h}(t-s) f(s)\, ds\big]^{\frac{1}{\beta}}, \end{aligned} \tag{47}$$

with $\bar{h}(s) = \sup_{e,u,x} \phi(e, u, s, x)$, $\tilde{h}(e, s, x) = \frac{2}{\min(\alpha_0, 1)} \sup_u \phi(e, u, s, x)$ and $\min(\alpha_0, 1)$ represents the minimum distance between two elements in the countable space $\mathbb{U}$. The quantity

$\mathbb{E}\big[|\int_0^t \tilde{h}(e,t-s,X_s)dN_s|^{\beta q}\big]$ is bounded since

$$\begin{aligned}\mathbb{E}\big[|\int_0^t \tilde{h}(e,t-s,X_s)dN_s|^{\beta q}\big] &\leq \mathbb{E}\big[|\int_0^t \tilde{h}(e,t-s,X_s)dN_s|^{q}\big]^{\frac{1}{\beta}}\\ &\leq \Big\{\sum_{\mathbf{k}_m\in\mathscr{P}(q)}\sum_{\tilde{x}\in E^m}\binom{q}{\mathbf{k}_m}\times\int_{(-\infty,t)^m}\mathbb{E}\big[\prod_{i=1}^m \tilde{h}(e,t-s_i,x_i)dN_{s_i}\big]\Big\}^{\frac{1}{\beta}}\\ &\leq \Big\{\sum_{\mathbf{k}_m\in\mathscr{P}(q)}\sum_{\tilde{x}\in E^m}\binom{q}{\mathbf{k}_m}\times\int_{(-\infty,t)^m}\prod_{i=1}^m \tilde{h}(e,t-s_i,x_i)\mathbb{E}\big[\lambda_{s_i}\big]ds_i\Big\}^{\frac{1}{q}}<\infty.\end{aligned}$$

The term (ii) satisfies

$$\mathbb{E}\big[|1+r_t^{n_1}+r^{\infty^{n_1}}|^{\gamma}\big]^{\frac{1}{\gamma}}\leq 3^{\frac{\gamma-1}{\gamma}}\big(1+\mathbb{E}[|r_t^{n_1}|^{\gamma}]^{\frac{1}{\gamma}}+\mathbb{E}[|r_t^{\infty^{n_1}}|^{\gamma}]^{\frac{1}{\gamma}}\big),\tag{48}$$

with $\gamma=\frac{p}{q}$ and $p,q\in\mathbb{N}^*$. We have

$$\begin{aligned}\mathbb{E}[|r_t^{n_1}|^{\frac{p}{q}}] &\leq \mathbb{E}[|r_t^{n_1}|^{p}]^{\frac{1}{q}}\\ &= \mathbb{E}[\big(\int_0^t \phi(e,U_t,t-s,X_s)dN_s\big)^{n_1 p}]^{\frac{1}{q}}\\ &= \Big\{\sum_{\mathbf{k}_m\in\mathscr{P}(\bar{p})}\sum_{\tilde{x}\in E^m}\binom{\bar{p}}{\mathbf{k}_m}\times\int_{(-\infty,t)^m}\mathbb{E}\Big[\prod_{i=1}^m\bar{\phi}(t-s_i,x_i)dN_{s_i}\Big]\Big\}^{\frac{1}{q}},\end{aligned}\tag{49}$$

with $\bar{\phi}(t,x)=\sup_{e,u}\phi(e,u,t,x)$ and $\bar{p}=n_1 p$. Using (49) and the Brascamp-Lieb inequality, we have

$$\begin{aligned}\mathbb{E}[|r_t^{n_1}|^{\frac{p}{q}}] &\leq \Big[\sum_{\mathbf{k}_m\in\mathscr{P}(\bar{p})}\sum_{\tilde{x}\in E^m}\binom{\bar{p}}{\mathbf{k}_m}\times\int_{(-\infty,t)^m}\prod_{i=1}^m\bar{\phi}(t-s_i,x_i)\mathbb{E}\big[\lambda_{s_i}\big]^{\frac{1}{m}}ds_i\Big]^{\frac{1}{q}}\\ &= \Big[\sum_{\mathbf{k}_m\in\mathscr{P}(\bar{p})}\sum_{\tilde{x}\in E^m}\binom{\bar{p}}{\mathbf{k}_m}R_m(t)\Big]^{\frac{1}{q}}=B_k(t)^{\frac{1}{q}},\end{aligned}\tag{50}$$

with $R_m(t)=\int_{(-\infty,t)^m}\prod_{i=1}^m\bar{\phi}(t-s_i,x_i)\mathbb{E}\big[\lambda_{s_i}\big]^{\frac{1}{m+m'}}ds_i$ and $B_k(t)=\sum_{\mathbf{k}_m\in\mathscr{P}(\bar{p})}\sum_{\tilde{x}\in E^m}\binom{\bar{p}}{\mathbf{k}_m}R_m(t)$. Similarly, we also have

$$\mathbb{E}[|r_t^{\infty^{n_1}}|^{\frac{p}{q}}]\leq B_k(t)^{\frac{1}{q}}.\tag{51}$$

Using Inequalities (48) and (50), we deduce that (ii) verifies

$$\mathbb{E}\big[|1+r_t^{n_1}+r^{\infty^{n_1}}|^{\gamma}\big]^{\frac{1}{\gamma}}\leq 3^{\frac{\gamma-1}{\gamma}}(1+2\sup_{0\leq k\leq n_\psi-1}\Big[B_k(t)\Big]^{\frac{1}{q\gamma}}).\tag{52}$$

By combining inequalities (47) and (52), we deduce that

$$(2)\leq 3^{\frac{\gamma-1}{\gamma}}\Big[c_1\mathbb{E}\big[\|U_t-U_t^{\infty}\|^{\beta p}\big]^{\frac{1}{\beta p}}+c_2\big[\int_0^t g(t-s)f(s)\,ds\big]^{\frac{1}{\beta}}\Big]\Big[1+2\sup_{0\leq k\leq n_\psi-1}\Big[B_k(t)\Big]^{\frac{1}{q\gamma}}\Big].\tag{53}$$

Using Theorem 1, we have $\sup_{e,t}\mathbb{E}[\sup_u\psi(e,u,t,r_t)]$ is finite. Thus, there exists K such that

$$(1)\leq c_3\mathbb{E}[||U_t-U_t^{\infty}||].\tag{54}$$

Thus using Equations (53) and (54), we prove (46).

Step (ii): By a density argument, there exist continous sequences of functions $(u^p)_{p\geq 1}$, $(g_1^p)_{p\geq 1}$ and $(\bar{h}^p)_{p\geq 1}$ such that $u^p(t) \underset{p\to\infty}{\to} u(t)$ and $u \leq u^p$, $g_1^p(t) \underset{p\to\infty}{\to} g_1(t)$ and $g_1 \leq g_1^p$ and $\bar{h}^p \underset{p\to\infty}{\overset{L^1}{\to}} \bar{h}$ and $\bar{h} \leq \bar{h}^p$. Thus, we have

$$f(t) \leq u^p(t) + g_1^p(t) G\Big(\int_0^t \bar{h}^p(s) f(s)\, ds\Big).$$

Using a density argument again, we can find a sequence of functions $(f^k)_{k\geq 1}$ converges uniformly towards f. By affording ourselves to use sub-sequences, we can always consider that

$$f^p(t) \leq \tilde{u}^p(t) + g_1^p(t) G\Big(\int_0^t \bar{h}^p(s) f^p(s)\, ds\Big),$$

with $\tilde{u}^p(t) = u^p(t) + |f - f^p|_\infty$. Using Theorem 3 in [14] and Inequality (46), we have

$$f^p(t) \leq v^p(t) F^p(t) \left\{1 + G\Big[H^{-1}\Big(\int_0^t \bar{h}^p(s) g_1^p(s)\, ds\Big)\Big]\right\},$$

with $H(s) = \int_0^s \frac{dt}{1+G(t)}$, $v^p(t) = \max(G_1(\tilde{u}^p)(t), 1)$, $F^p(t) = \max(G_1(g_1^p)(t), 1)$ and

$$G_1(w)(t) = w(t)\left(1 + \int_0^t w(s)\bar{h}^p(s) e^{\int_s^t \bar{h}^p g_1^p\, du}\, ds\right).$$

By sending p to infinity, we deduce that

$$f(t) \leq v(t) F(t) \left\{1 + G\Big[H^{-1}\Big(\int_0^t \bar{h}(s) g_1(s)\, ds\Big)\Big]\right\}, \tag{55}$$

with $v(t) = \max(G_1(u)(t), 1)$ and $F(t) = \max(G_1(g_1)(t), 1)$.

Step (iii): Let us prove that $\int_{\mathbb{R}_+} u(t)\, dt < \infty$. Since $B(t)$ is uniformly bounded, we only need to prove that

$$\begin{cases} \int_{\mathbb{R}_+} \mathbb{E}\big[\|U_t - U_t^\infty\|\big]\, dt < \infty \\ \int_{\mathbb{R}_+} \mathbb{E}\big[\|U_t - U_t^\infty\|^{\beta p}\big]^{\frac{1}{\beta p}}\, dt < \infty. \end{cases}$$

Since $0 < \underline{\psi} = \inf_{u,t,r} \sup_e \psi(e,u,t,r) \leq \lambda_n$, we have

$$\begin{aligned} \mathbb{E}\Big[\int_{\mathbb{R}_+} \|U_t - U_t^\infty\|\, dt\Big] &= \leq \mathbb{E}\Big[\sum_{n\geq 0} \|U_n - U_n^\infty\| \int_{T_n}^{T_{n+1}} dt\Big] \\ &= \mathbb{E}\Big[\sum_{n\geq 0} \|U_n - U_n^\infty\| \mathbb{E}[T_{n+1} - T_n | \mathscr{F}_n]\Big] \\ &\leq \mathbb{E}\Big[\sum_{n\geq 0} \|U_n - U_n^\infty\| \frac{1}{\underline{\psi}}\Big]. \end{aligned}$$

Using Lemma 8, we have $\mathbb{E}\big[\sum_{n\geq 0} \|U_n - U_n^\infty\|\big] < \infty$ which ensures that $\mathbb{E}\big[\int_{\mathbb{R}_+} \|U_t - U_t^\infty\|\, dt\big] < \infty$. By using a similar methodology and the fact that $\sum_{n\geq 0} \mathbb{E}\big[\|U_n - U_n^\infty\|^{\beta p}\big] r^n < \infty$ with $r > 1$, see Lemma 8, we also have $\int_{\mathbb{R}_+} \mathbb{E}\big[\|U_t - U_t^\infty\|^{\beta p}\big]^{\frac{1}{\beta p}}\, dt < \infty$.

Step (iv): Since g_1 is bounded and $\int_0^t \bar{h}(s)\,ds < \infty$, the functions $F(t)$ and $\left\{1+G\Big[H^{-1}\big(\int_0^t \bar{h}(s)g_1(s)\,ds\big)\Big]\right\}$ are bounded as well. Moreover, $\int_{\mathbb{R}_+} u(t)\,dt < \infty$ thanks to the previous step. Thus, by applying Inequality (55), we have that $\int_{\mathbb{R}_+} f(t)\,dt < \infty$ which completes the proof. □

IV.D.3 Speed of convergence

Lemma 10. *We have the following error estimate:*

$$||P_t(w,.) - \bar{\pi}||_{TV} \leq K_1 e^{-K_2 t}, \quad \forall w \in W,$$

with $K_3 > 0$ *and* $K_2 > 0$.

Proof of Lemma 10. We forget the dependence of $\mathbb{E}_{\mathbf{x}}[X]$ on the initial state $\mathbf{x}$ for any random variable X. We have

$$\begin{aligned} ||P_t(w,.) - \bar{\pi}||_{TV} &\leq \mathbb{P}[\sup_e |N_s - N_s^\infty| \neq 0, \quad \forall s \in (t,\infty)] \\ &= \left(1 - \mathbb{P}[\sup_e N_s = N_s^\infty, \quad \forall s \in (t,\infty)]\right) = (i). \end{aligned}$$

Using Lemma 4 and Jensen's Inequality, we have

$$(i) \leq 1 - e^{-\int_t^\infty f(s)\,ds},$$

with $f(t) = \sup_e \mathbb{E}\big[|\lambda_t(e) - \lambda_t^\infty(e)|\big]$ for any $t \geq 0$. Using Inequality (55) and the boundedness of F and $\left\{1+G\Big[H^{-1}\big(\int_0^t \bar{h}(s)g_1(s)\,ds\big)\Big]\right\}$, we have

$$(i) \leq c_1 \int_t^\infty u(t)\,dt, \tag{56}$$

with c_1 a positive constant. Let us now prove that

$$u(t) \leq c_1 e^{-\alpha t}, \tag{57}$$

with α a positive constant. We have

$$\mathbb{E}[\|U_t - U_t^\infty\|] = \mathbb{E}[\|U_{N(t)} - U_{N^\infty(t)}^\infty\|] \leq \mathbb{E}[\|U_{N(t)} - U_{N(t)}^\infty\|] + \mathbb{E}[\|U_{N(t)}^\infty - U_{N^\infty(t)}^\infty\|].$$

Using the fact that $\sum_{n\geq 1} \mathbb{E}[\|U_n - U_n^\infty\|] r^n < \infty$, there exists $\alpha > 0$ such that $\mathbb{E}[\|U_n - U_n^\infty\|] \leq Ae^{-\alpha n}$. Let us denote by $U_t^{\infty,\delta}$ the δ-translated process defined such that $U_t^{\infty,\delta} = U_{t+\delta}^\infty$. By applying Lemma 9 to the process $U^{\infty,\delta}$, we also have $\sup_\delta \big(\sum_{n\geq 1} \mathbb{E}[\|U_n^{\infty,\delta} - U_n^\infty\|] r^n\big) < \infty$ which ensures that $\mathbb{E}[\|U_n^{\infty,\delta} - U_n^\infty\|] \leq Ae^{-\alpha n}$. Using Lemma 11 below and the uniqueness of the stationary distribution, we have $\frac{N(t)}{t} \underset{t\to}{\to} \frac{1}{\mathbb{E}_\pi[\Delta T_1]}$ and $\frac{N^\infty(t)}{t} \underset{t\to}{\to} \frac{1}{\mathbb{E}_\pi[\Delta T_1]}$, *a.s.* Thus, we deduce that

$$\mathbb{E}[\|U_t - U_t^\infty\|] \leq c_1 e^{-\alpha t}. \tag{58}$$

Using the same lines of argument, we also have

$$\mathbb{E}[\|U_t - U_t^{\infty}\|^{\beta p}]^{\frac{1}{\beta p}} \leq c_1 e^{-\alpha t}. \tag{59}$$

By combining Inequalities (58) and (59) and using the expression of $u(t)$, we recover Inequality (57) which ensures that

$$(i) \leq c_1 e^{-\alpha t}.$$

This completes the proof. □

Lemma 11. *For any initial state $u \in \mathbb{U}$, the process ΔT_n satisfies*

$$\frac{\sum_{i=1}^{n} \Delta T_i}{n} \underset{n\to\infty}{\to} \mathbb{E}_{\mu}[\Delta T_1] \qquad a.s,$$

with μ the unique stationary distribution of the point process N.

Proof. Since there exists $\underline{\lambda} > 0$ such that $\inf_{t,u,r} \sum_{e\in E} \lambda_t(e,u,r) > \underline{\lambda}$, we have $\mathbb{E}[\Delta T_n] \leq \frac{1}{\underline{\lambda}}$ for any $n \geq 1$. Thus, ΔT_n admits a finite stationary distribution. Using the Theorem 17.1.2 in [90], we complete the proof. □

IV.E Proof of Propositions 2 and 3

Proof of Proposition 2. The proof of Equation (9) is a direct application of Theorem 2 in [42]. Since (U_n) is f-geometrically ergodic, see Lemma 8, (Y_n) is g-geometrically ergodic and U_n and Y_n are independent, the process (U_n, Y_n) is $\tilde{f}$-geometrically ergodic with $\tilde{f}(u,y) = f(u) + g(y)$. Let g and h be two functions such that $g^2, h^2 \leq \tilde{f}$, μ the stationary distribution of (U,Y) and $\bar{v} = v - \mathbb{E}_{\mu}[v]$ for any function v. By following the same lines of argument of Lemma 16.1.5 in [90], we have

$$|\mathbb{E}_{\pi}[\bar{h}(Z_n)\bar{g}(Z_{n+k})]| \leq R\mathbb{E}_{\pi}[\bar{\tilde{f}}(Z_0)]r^k,$$

with $Z_n = (U_n, Y_n)$, $r < 1$ and R a positive constant. The quantity $\mathbb{E}_{\pi}[\bar{\tilde{f}}(Z_0)]$ is bounded by Lemma 2. Thus Z is a geometric mixing and Theorems 19.1 and 19.2 in [21] give the result. □

Proof of Proposition 3. Using Lemma 11 and Proposition 3, the proof of this result is analagous to the proof of Theorem 4.2 in [65]. □

IV.F Stationary distribution computation

Proof of Proposition 4. Let $z \in Z$ and $z' \in Z$ such that $z \neq z'$. Since ζ is stationary under μ, we have

$$\sum_{z'\in Z} \int_{A^{z'}} \mu(dw) P_t(w, A^z) = \int_{W_0} \mu(dw) P_t(w, A^z) = \mu(A^z), \quad \forall\, t \geq 0, \tag{60}$$

with $P_t(w,.)$ the probability distribution of $\zeta_t^{0,w}$ starting from the initial condition w and $A^z = \{(w_s)_{s\leq 0} \in W_0; \zeta_0^{0,w} = z\}$. Since $\int_{A^{z'}} \mu(dw)P_t(w,A^z) = \mathbb{P}_\mu[\zeta_t = z, \zeta_0 = z'] = \mathbb{P}_\mu[\zeta_0 = z']\mathbb{P}_\mu[\zeta_t = z|\zeta_0 = z']$ and $\mu(A^z) = \mathbb{P}_\mu[\zeta_0 = z]$, the quantity $\pi(z) = \mu(A^z)$ defined in Section 5.1 satisfies

$$\sum_{z'\in Z} \pi(z')\mathbb{P}_\mu[\zeta_t = z|\zeta_0 = z'] = \pi(z), \quad \forall t \geq 0,$$

which also leads to the following equation:

$$\sum_{z'\in Z} \pi(z')\tilde{Q}(z,z') = 0, \quad \sum_{z'\in Z} \pi(z') = 1,$$

with $\tilde{Q}(z,z') = \lim_{\delta\to 0} \frac{\mathbb{P}_\mu[\zeta_\delta = z'|\zeta_0 = z]}{\delta}$. The quantity $\tilde{Q}(z,z')$ satisfies

$$\begin{aligned}
\tilde{Q}(z,z') = \lim_{\delta\to 0} \frac{\mathbb{P}_\mu[U_\delta = z'|U_0 = z]}{\delta} &= \lim_{\delta\to 0} \frac{\mathbb{P}_\mu[\{T_1 \leq \delta,\, e_1 \in E(z,z')\}|\zeta_0 = z] + \epsilon}{\delta} \\
&= \lim_{\delta\to 0} \frac{\mathbb{E}_\mu[\mathbb{P}[\{T_1 \leq \delta,\, e_1 \in E(z,z')\}|\mathscr{F}_0]|\zeta_0 = z] + \epsilon}{\delta} \\
&= \mathbb{E}_\mu[\lim_{\delta\to 0} \frac{\mathbb{P}[\{T_1 \leq \delta,\, e_1 \in E(z,z')\}|\mathscr{F}_0]}{\delta}|\zeta_0 = z] + \lim_{\delta\to 0}\frac{\epsilon}{\delta} \\
&= \mathbb{E}_\mu[\sum_{e_1\in E(z,z')} \lambda_0(e_1)|\zeta_0 = z] + \lim_{\delta\to 0}\frac{\epsilon}{\delta} \\
&= \sum_{e_1\in E(z,z')} \mathbb{E}_\mu[\lambda_0(e_1)|\zeta_0 = z] + \lim_{\delta\to 0}\frac{\epsilon}{\delta},
\end{aligned}$$

where ϵ is an error term associated to the cases when at least two events happen in the interval $[0,\delta]$. Since $\sum_{e_1\in E}\mathbb{E}_\mu[\lambda_0(e_1)]$ is finite, we have $\epsilon \leq c_1\delta^2$ with c_1 a positive constant. We deduce that

$$\tilde{Q}(z,z') = \sum_{e_1\in E(z,z')} \mathbb{E}_\mu[\lambda_0(e_1,u)] = Q(z,z'). \tag{61}$$

This completes the proof. □

IV.G Proof of Proposition 5

Proof of Proposition 5. We write $\lambda_s^{u,u'} = \sum_{e\in E(u,u')} \lambda_s(e)$ and $E(u,u')$ the set of events that moves the order book from the state u to u'. We have

$$\frac{N_t^{u,u'}}{t} = \frac{\int_0^t \lambda_s \delta^s_{u,u'}\, ds}{t} + \Big(\frac{N_t^{u,u'} - \int_0^t \lambda_s \delta^s_{u,u'}\, ds}{t}\Big). \tag{62}$$

Since $(\lambda_s)_{s\geq 0}$ is stationary under $\bar{\pi}$ and $\mathbb{E}_{\bar{\pi}}[\lambda_s] < \infty$, the Theorem 2.1-chapter X in [42] ensures that

$$\begin{aligned}\frac{\int_0^t \lambda_s \delta^s_{u,u'}\, ds}{t} \underset{t\to\infty}{\to} \mathbb{E}_{\bar{\pi}}[\lambda_0 \delta^0_{u,u'}] = \sum_{e\in E(u,u')} \mathbb{E}_{\bar{\pi}}[\lambda_0(e)\delta^0_{u,u'}] &= \sum_{e\in E(u,u')} \mathbb{E}_{\bar{\pi}}[\lambda_0(e)\delta^0_{u,u'}|U_0 = u]\mathbb{P}_{\bar{\pi}}[U_0 = u] \\ &= \mathbb{P}_{\bar{\pi}}[U_0 = u] \sum_{e\in E(u,u')} \mathbb{E}_{\bar{\pi}}[\lambda_0(e)|U_0 = u] \\ &= \mathbb{P}_{\bar{\pi}}[U_0 = u]Q(u,u'), \quad a.s. \end{aligned} \tag{63}$$

Moreover, since $N^{u,u'}_t - \int_0^t \lambda^{u,u'}_s\, ds$ is a martingale and $\sup_{s\geq 0,u,u'} \mathbb{E}[\lambda^{u,u'}_s] < \infty$, we have

$$\frac{N^{u,u'}_t - \int_0^t \lambda_s \delta^s_{u,u'}\, ds}{t} \underset{t\to\infty}{\to} 0, \quad a.s. \tag{64}$$

Hence, by combining (62), (63) and (64), we prove $\dfrac{N^{u,u'}_t}{t} \underset{t\to\infty}{\to} \mathbb{P}_{\bar{\pi}}[U_0 = u]Q(u,u')$, *a.s.* On the other hand, we have

$$\frac{t^u}{t} = \frac{\int_0^t \delta^s_u\, ds}{t}. \tag{65}$$

Since $(U_s)_{s\geq 0}$ is stationary under $\bar{\pi}$ and $\mathbb{E}_{\bar{\pi}}[\delta^s_u] < \infty$, the Theorem 2.1-chapter X in [42] ensures that

$$\frac{\int_0^t \delta^s_u\, ds}{t} \underset{t\to\infty}{\to} \mathbb{E}_{\bar{\pi}}[\delta^0_u] = \mathbb{P}_{\bar{\pi}}[U_0 = u] \quad a.s. \tag{66}$$

Thus, we deduce that

$$\frac{N^{u,u'}_t}{t^u} = \frac{\dfrac{N^{u,u'}_t}{t}}{\dfrac{t^u}{t}} \underset{t\to\infty}{\to} Q(u,u'), \quad a.s,$$

which completes the proof. □

Proof of confidence interval computation. By applying Theorem 2 to the sequence of $\eta_s = \lambda_s \delta^s_{u,u'}$ and use basic inequalities to approximate t by its integer part $\lfloor t \rfloor$, we have

$$\sqrt{t}\Big(\frac{N^{u,u'}_t}{t} - \mathbb{P}_{\bar{\pi}}[U_0 = u]Q(u,u')\Big) \xrightarrow{\mathscr{L}} \sigma^1 W_1, \tag{67}$$

with $\sigma^2_1 = \mathbb{E}_\mu[(\lambda_0\delta^0_{u,u'})^2] + 2\sum_{k\geq 1} \mathbb{E}_\mu[\lambda_0\delta^0_{u,u'}\lambda_k\delta^k_{u,u'}]$ and W_t a standard brownian motion. Similarly, by using the same arguments, we also have

$$\sqrt{t}\Big(\frac{t^u}{t} - \mathbb{P}_{\bar{\pi}}[U_0 = u]\Big) \xrightarrow{\mathscr{L}} \sigma^2 W_1, \tag{68}$$

with $\sigma_2^2 = \mathbb{E}_\mu[(\delta_u^0)^2] + 2\sum_{k\geq 1}\mathbb{E}_\mu[\delta_u^0\delta_u^k]$. Using (67) and (68), we have with asymptotic probability 95% that

$$\begin{aligned}\mathbb{P}_{\bar{\pi}}[U_0 = u]Q(u,u') &\in [\frac{N_t^{u,u'}}{t} + \frac{1.96\sigma_1}{\sqrt{t}}, \frac{N_t^{u,u'}}{t} - \frac{1.96\sigma_1}{\sqrt{t}}] \\ \mathbb{P}_{\bar{\pi}}[U_0 = u]^{-1} &\in [\frac{t}{t^u} + \frac{1.96\sigma_2}{\sqrt{t}} \times \frac{t}{t^u}, \frac{t}{t^u} - \frac{1.96\sigma_2}{\sqrt{t}} \times \frac{t}{t^u}].\end{aligned} \tag{69}$$

Equation (69) ensures that we have with probability 90%

$$Q(u,u') \in [(\frac{N_t^{u,u'}}{t} + \frac{1.96\sigma_1}{\sqrt{t}})(\frac{t}{t^u} + \frac{1.96\sigma_2}{\sqrt{t}} \times \frac{t}{t^u}), (\frac{N_t^{u,u'}}{t} - \frac{1.96\sigma_1}{\sqrt{t}})(\frac{t}{t^u} - \frac{1.96\sigma_2}{\sqrt{t}} \times \frac{t}{t^u})].$$

□

IV.H Proof of Remark 19

Proof. We assume that the insertion (resp. consumption) intensity λ^+ (resp. λ^-) is constant and focus on the best bid limit Q^1. The stationary distribution π^{old} of Q^1 verifies

$$\pi^{old}(q) = \pi^{old}(0)(\rho^{old})^q, \quad \pi^{old}(0) = (1 + \sum_{q=1}^{\infty}(\rho^{old})^q)^{-1}, \quad \rho^{old} = \frac{\lambda^+}{\lambda^-}, \tag{70}$$

with $q \geq 1$ the size of Q^1. We add to the market a new agent whose insertion (resp. consumption) intensity $\lambda^{+,a}$ (resp. $\lambda^{-,a}$) is also constant. The stationary distribution π^{new} of Q^1 in the new market satisfies

$$\pi^{new}(q) = \pi^{new}(0)(\rho^{new})^q, \quad \pi^{new}(0) = (1 + \sum_{q=1}^{\infty}(\rho^{new})^q)^{-1}, \quad \rho^{new} = \frac{\lambda^+ + \lambda^{+,a}}{\lambda^- + \lambda^{-,a}}, \tag{71}$$

with $q \geq 1$ the size of Q^1. Using Equations (70) and (71), we can write

$$\rho^{new} = \rho^{old}(1 + R(\lambda, \lambda^a)), \quad \pi^{new}(0) = \Big(1 + \sum_{q=1}^{\infty}(\rho^{old})^q(1 + R(\lambda,\lambda^a))^q\Big)^{-1}, \tag{72}$$

with $\lambda = (\lambda^+, \lambda^-)$, $\lambda^a = (\lambda^{+,a}, \lambda^{-,a})$ and $R(\lambda, \lambda^a) = (1 + \frac{\lambda^{+,a}}{\lambda^+})/(1 + \frac{\lambda^{-,a}}{\lambda^-}) - 1$. We want the new introduced agent to reduce the volatility of the old market which at the first order reads

$$\mathbb{E}_{\pi^{new}}[\eta_0^2] \leq \mathbb{E}_{\pi^{old}}[\eta_0^2]. \tag{73}$$

Using Equation (72), we can reformulate Inequality (73) in the following way:

$$\sum_q \frac{(\rho^{old})^q\Big(1 + R(\lambda,\lambda^a)\Big)^q}{\Big(1 + \sum_{j=1}^{\infty}(\rho^{old})^j(1 + R(\lambda,\lambda^a))^j\Big)}\eta_0^2(q) \leq \sum_q \frac{(\rho^{old})^q}{\Big(1 + \sum_{j=1}^{\infty}(\rho^{old})^j\Big)}\eta_0^2(q), \tag{74}$$

for any function η_0. To satisify Inequality (74) we need $R(\lambda, \lambda^a) \geq 0$ which leads to

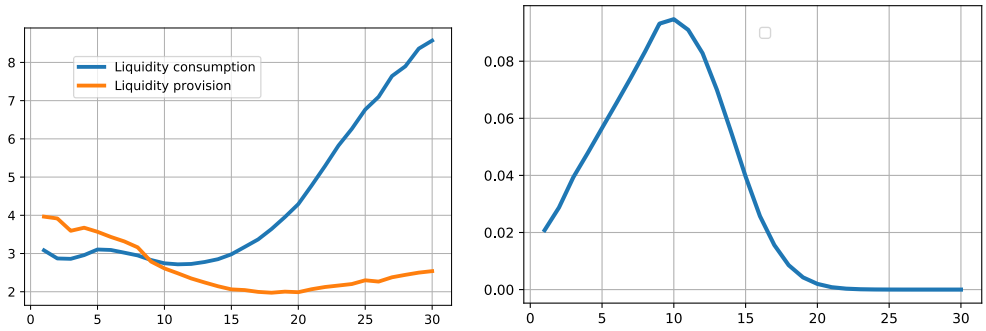
$$\frac{\lambda^{+,a}}{\lambda^{-,a}} \geq \frac{\lambda^+}{\lambda^-},$$

This condition is a well-known result which ensures that the new agent needs to have an insertion/consumption ratio greater than the one of the market. □

IV.I Supplementary numerical results

The three next figures show the liquidity consumption and provision intensities at the first limit relative to the whole market according to the queue size, the corresponding stationary measure and the long term volatility, respectively for EssilorLuxottica, Michelin and Orange.

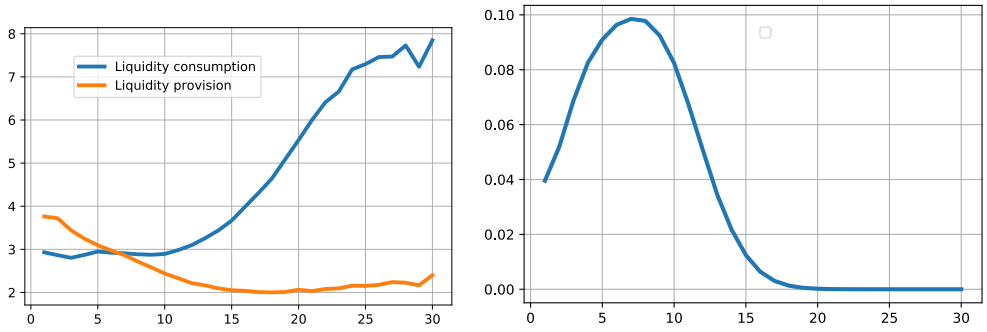
(a) Intensity of the market (b) Stationary measure Q^1



Long term price volatility $\sigma^{2,G} = 0.038$, $\sigma^{2,M}_{10} = 0.26$.

Figure IV.4 – (a) Liquidity insertion and consumption intensities (in orders per second) with respect to the queue size (in AES) and (b) the corresponding stationary distribution of Q^1 with respect to the queue size (in AES), proper to ExilorLuxottica.

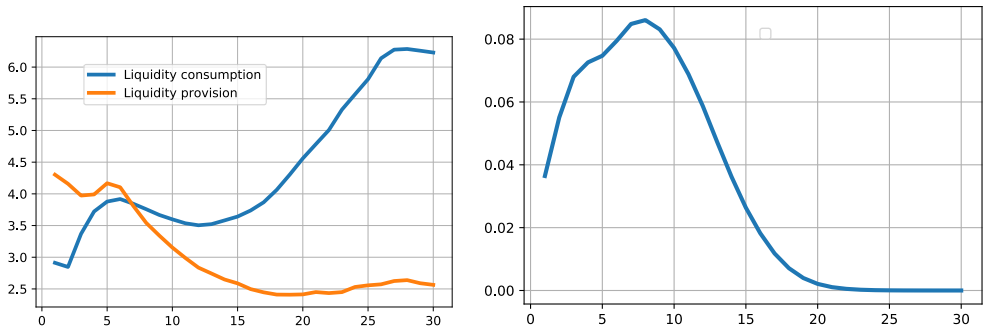
(a) Intensity of the market (b) Stationary measure Q^1



Long term price volatility $\sigma^{2,G} = 0.075$, $\sigma^{2,M}_{10} = 0.490$.

Figure IV.5 – (a) Liquidity insertion and consumption intensities (in orders per second) with respect to the queue size (in AES) and (b) the corresponding stationary distribution of Q^1 with respect to the queue size (in AES), proper to Michelin.

(a) Intensity of the market (b) Stationary measure Q^1



Long term price volatility $\sigma^{2,G} = 0.065$, $\sigma^{2,M}_{10} = 0.453$.

Figure IV.6 – (a) Liquidity insertion and consumption intensities (in orders per second) with respect to the queue size (in AES) and (b) the corresponding stationary distribution of Q^1 with respect to the queue size (in AES), proper to Orange.

For each of the market makers, we compute the liquidity consumption and provision intensities, and the corresponding stationary measure that we would obtain in a situation where the studied market maker withdraws from the market and the other market participants do not change their behaviour. We show respectively the results relative to EssilorLuxottica, Michelin and Orange.

Intensities and $\sigma_{10}^{2,M}$ when one market maker leaves the market : stock EssilorLuxottica

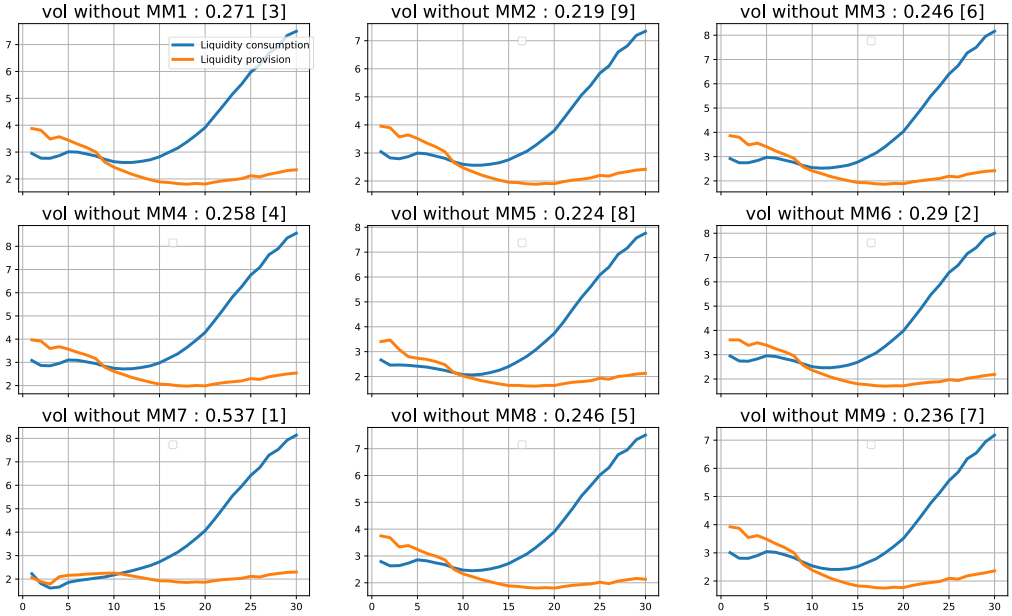


Figure IV.7 – Liquidity insertion and consumption intensities (in orders per second) with respect to the queue size (in AES) and $\sigma_{10}^{2,M}$ when one market maker is ejected from the market for the stock EssilorLuxottica.

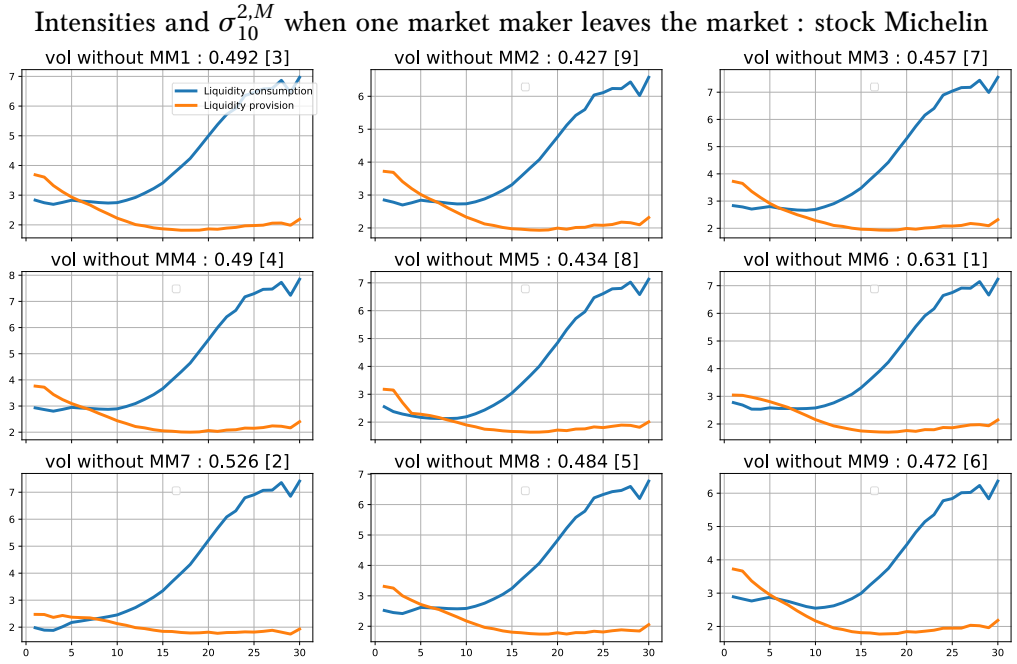


Figure IV.8 – Liquidity insertion and consumption intensities (in orders per second) with respect to the queue size (in AES) and $\sigma_{10}^{2,M}$ when one market maker is ejected from the market for the stock Michelin.

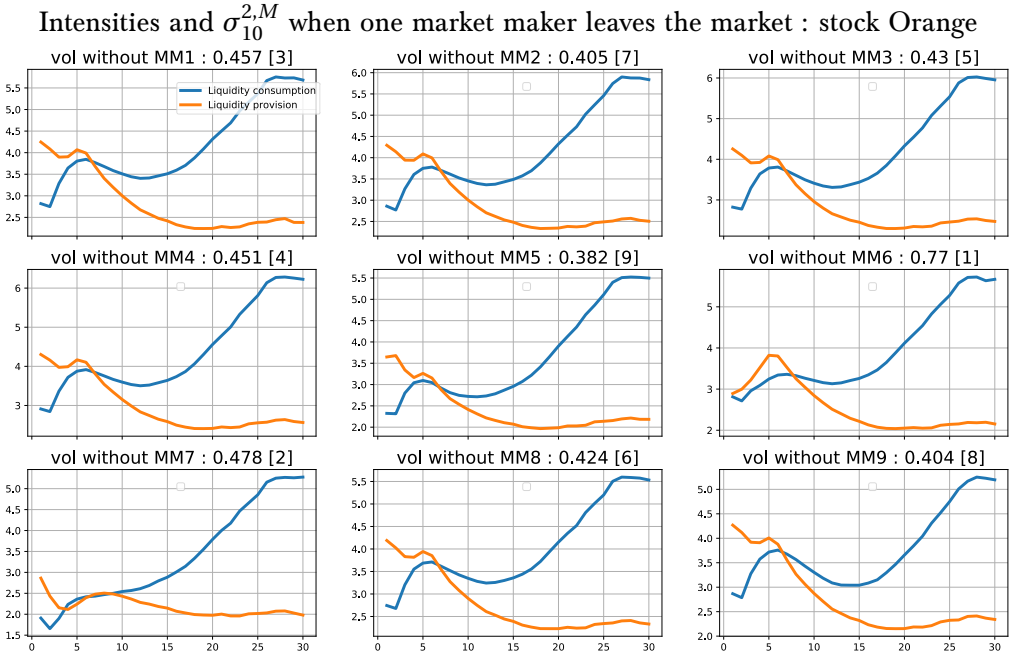


Figure IV.9 – Liquidity insertion and consumption intensities (in orders per second) with respect to the queue size (in AES) and $\sigma_{10}^{2,M}$ when one market maker is ejected from the market for the stock Orange.

CHAPTER V

Market makers inventories and price pressure: theory and multi-platform empirical evidences

Abstract

We are interested in the link between price pressure and market makers inventories. We extend the literature establishing a negative correlation between both quantities to the multi-agents framework. We then investigate this relationship empirically thanks to a unique cross-platform trades dataset of European stocks with participant level information. Using these data, we are able to provide evidences that high-frequency market makers perform cross-market inventory netting and show how inventory management induces specific price pressures.

Keywords: Price pressure, market making, inventory, price impact, market fragmentation, multi-platform analysis, regulation.

1 Introduction

With market electronification and fragmentation, high-frequency traders (HFTs) started to emerge and became major actors in the markets. Since then, several studies have analysed their behaviour and impact on markets. Several studies focus on the reaction of HFTs with respect to market features considered exogenous such as volatility, see [116], macroeconomic announcements, see Chapter I and [34], flash crashes, see [74], or new regulation like the tick size changes, see [9]. In this work, we rather aim at understanding the endogenous impact of high-frequency market makers on market quality in the spirit of Chapter IV and [97]. In these papers, the authors investigate the contribution of HFTs to volatility via their order flows. Here, our main variable of interest is high-frequency market makers inventory. Inventory is indeed a key component of market making algorithms: market makers typically wish to carry limited inventory risk and end their day without holding any position. This is why we want to analyse how inventory management strategies affect market dynamics. More precisely, we

are interested in the relationship between individual high-frequency market maker inventories and price pressure, that is the local price impact due to submitted market and limit orders.

From a theoretical point of view, there are several models that analyse the determinants of the best bid and ask. Some of them focus on adverse selection costs, others on inventory costs. In models studying adverse selection costs, see Chapter III and [53, 76], bid and ask quotes are optimally chosen by the market maker based on the probability that incoming orders are informed. Models dealing with inventory costs, see [4, 5, 7, 108], suppose that market makers set quotes in order to maintain their inventory around a target level. The majority of inventory costs models consider a monopolistic market maker, see [4]. Some papers like [31] work with Cournot competitive risk-averse market makers. An important innovation of our paper is that we extend the results of [31] to the case of market makers who are heterogeneous in their initial holdings. We show that the negative relationship between price pressure and market maker inventory holds in all cases: positive inventories lead to negative price pressure and conversely.

From an empirical viewpoint, this negative relationship between inventory and price pressure has never really been investigated accurately. This is probably because of the lack of multi-platform data with participant level information. Indeed, having such data is crucial when studying market making activity since market makers act simultaneously on many platforms. Some papers analyse this relationship empirically but assume there is only one monopolistic market maker and use approximations to compute its inventory, see [4]. These assumptions are quite far from reality. In practice, there are multiple market makers, and from a regulatory viewpoint, it is important to assess this relationship at the individual level for each market maker to understand its own contribution to price pressure. For instance, a market maker who impacts prices significantly when his inventory is large intensifies the risk of a flash crash during periods of stress, see [4]. To our knowledge, our paper is the first multi-platform study of high-frequency market making activity with participant level information. Using this unique data set, we can identify different agents who act as market makers. Furthermore, we provide original empirical evidences that market makers perform cross-market netting of inventories. We are indeed able to compute the actual inventories of market makers, instead of estimating them from the cumulative net volume as typically done in the literature.

We first identify on our data the market participants that can be viewed as high-frequency market makers by analysing their end-of-day inventories. We then display how their trades are distributed across multiple platforms. In our theoretical model, like in most approaches in the literature, we consider that market makers are purely passive. However, we show that in practice market makers do use aggressive orders in significant proportions confirming the results of Chapter II. In contrast to the classical idea suggesting that market makers become more aggressive when their inventories are large, we find that their aggressiveness level stays quite constant. Still, they send more buy (resp. sell) than sell (resp. buy) aggressive orders when their inventory is highly negative (resp. positive). Thus when studying empirically the relationship between market makers inventory and price pressure, we decide to take into ac-

count both aggressive and passive flows.

The paper is organised as follows. In Section 2, we introduce our model for the behaviour of buyers and sellers and the optimisation problem of market makers. The case where market makers are identical is considered in Section 3 while we treat in Section 4 the situation where they have different initial inventories. We describe our dataset in Section 5, explain how to identify market makers and give some preliminary statistics. We investigate in Section 6 the role of aggressive orders in inventory management. Finally, for each of our market makers, we display the relationship between his inventory and the price pressure in Section 7 .

2 Model

2.1 Buy and sell investors

Buy and sell investors arrive in the market at random times. We denote the cumulative number of buyers and sellers arriving in time interval $[0,t]$ respectively by N_t^B and N_t^S, and assume that N_t^B and N_t^S are independent Poisson processes with identical arrival intensity $\lambda > 0$. We assume that the fundamental price of the asset follows a Brownian motion with volatility $\sigma > 0$:

$$dS_t = \sigma dB_t, \quad t \geq 0,$$

where $(B_t)_{t\geq 0}$ is a standard Brownian motion, independent of the Poisson processes N^B and N^S, and the initial value $S_0 > 0$ is a positive constant.

We denote by $S_t + \tilde{a}_t$ and $S_t + \tilde{b}_t$, respectively the ask and bid quotes offered by the competitive market makers. That is, $\tilde{a}_t$ and $\tilde{b}_t$ measure the deviation of ask and bid quotes from the fundamental price of the asset. We refer to $\tilde{a}_t$ ($\tilde{b}_t$, resp.) as the ask (bid, resp.) price pressures. If a buy or sell investor arrives at time t, then he instantaneously trades with the entire body of market makers, and then leaves the market. The demand and supply functions of the buy/sell investors are given by

$$Q^B(\tilde{a}_t) = c(\tilde{p} - \tilde{a}_t), \quad Q^S(\tilde{b}_t) = c(\tilde{b}_t + \tilde{p}), \tag{1}$$

where the demand slope $c > 0$ is constant, $\tilde{p}$ ($-\tilde{p}$, resp.) is the reservation price of the buyer (seller, resp.) relative to the fundamental price, and $Q^B(\tilde{a}_t)$ ($Q^S(\tilde{b}_t)$, resp.) is the amount of shares the buyer (seller, resp.) trades with the market makers at the price pressure $\tilde{a}_t$ ($\tilde{b}_t$, resp.) determined at time t.

2.2 Multiple market makers

We describe the control problem faced by the N market makers with $N \geq 1$. For any $l = 1,2,\ldots,N$, Market Maker l chooses a predictable control strategy $x^l \equiv (x_t^{l,a}, x_t^{l,b})_{t\in[0,\infty)}$ specifying the size of buy and sell orders that will be executed against the randomly arriving buy/sell investors. Because marker makers are the only counterparty available for trade when end investors arrive, the total liquidity demanded by buy end investors needs to be absorbed

by market makers, and symmetrically, the total liquidity supplied by market makers needs to be absorbed by sell end investors. Using (1), this implies that the ask and bid price pressures at time t are given by

$$\tilde{a}_t = \tilde{p} - \frac{1}{c}\sum_{l=1}^{N} x_t^{l,a}, \qquad \tilde{b}_t = -\tilde{p} + \frac{1}{c}\sum_{l=1}^{N} x_t^{l,b}. \tag{2}$$

The l-th market maker maximises expected utility from wealth, minus a quadratic penalty for holding inventory. His value function is given by

$$\bar{v}_l(i) = \sup_{x^l} \mathbb{E}\left[\int_0^\infty e^{-\beta t}(dW_s^{l,(x^l,x^{-l})} - \Theta(I_s^{l,x^l})^2 ds) | I_t^{l,x^l} = i\right],$$

where $\beta > 0$ and $\Theta > 0$ are positive constants.

The wealth of Market Maker l at time t, controlled by his trading quantities $x = (x_s^{l,a}, x_s^{l,b})_{s<t}$ and denoted by $W_t^{l,x}$, is given by the cumulative proceeds from sales to buy end investors, net of purchases from the sell end investors. Purchases and sales from end investors change the inventory. Hence, for any $l = 1\ldots, N$, the inventory I_t^{l,x^l} of the l-th market maker evolves according to:

$$dI_t^{l,x^l} = -x_t^{l,a} dN_t^B + x_t^{l,b} dN_t^S. \tag{3}$$

The market maker measures his trading revenue relative to the fundamental price. Using the expression for the ask and bid prices in (2), we obtain that the (controlled) dynamics of Market Maker l's wealth is given by

$$dW_t^{l,x} = \Big(\tilde{p} - \frac{1}{c}\sum_{n=1}^{N} x_t^{i,a}\Big) x_t^{l,a} dN_t^B - \Big(-\tilde{p} + \frac{1}{c}\sum_{n=1}^{N} x_t^{i,b}\Big) x_t^{l,b} dN_t^S. \tag{4}$$

Equations (3) and (4) describe the (controlled) dynamics of Market Maker l's inventory and wealth.

3 Case of identical market makers

We consider now that market makers are identical: their costs for holding inventory and their initial inventories are equal. The computations is this section are based on [31]. Using the dynamic programming principle, the value function $\bar{v}_l(i)$ of the control problem solved by the l-th market maker is the solution to the Bellman equation:

$$\begin{aligned}\Theta i^2 + \beta \bar{v}_l(i) = \lambda \sup_{x^{l,a}, x^{l,b}} \Bigg[&\left(\tilde{p} - \frac{1}{c}\sum_{n=1}^{N} x^{n,a}\right) x^{l,a} + \bar{v}_l(i - x^{l,a}) - \bar{v}_l(i) \\ &- \left(\frac{1}{c}\sum_{n=1}^{N} x^{n,b} - \tilde{p}\right) x^{l,b} + \bar{v}_l(i + x^{l,b}) - \bar{v}_l(i)\Bigg]\Big|_{x^n = x^{l,\star} \text{ for all } n \neq l},\end{aligned} \tag{5}$$

where $x^{l,\star}$ is the optimizer of Hamiltonian above, in which we set $x^n = x^{l,\star}$ for all $n \neq l$ because we are considering a symmetric Markov perfect equilibrium. We make the ansatz that the value function $\bar{v}_l(i)$ is quadratic and concave in i. Moreover, since all market makers are identical, value function and strategies are the same for all market makers, i.e.,

$$\bar{v}_l(i) = \bar{v}(i) = -Ai^2 + Bi + C,$$

for some constant $A > 0$. It then follows that the optimal control strategy for each Market maker is given by

$$\begin{cases} x^a = \dfrac{\tilde{p} + 2Ai - B}{\frac{N+1}{c} + 2A}, \\ x^b = \dfrac{\tilde{p} - 2Ai + B}{\frac{N+1}{c} + 2A}, \end{cases} \tag{6}$$

Using the expressions above, we can then rewrite (5) as

$$\frac{1}{\lambda}[(\Theta - \beta A)i^2 + \beta B i + \beta C] = \frac{8A^2c(1+cA)}{(N+1+2cA)^2}i^2 + \frac{8cA(1+cA)}{(N+1+2cA)^2}Bi + \frac{2c(1+cA)(B^2+(\tilde{p})^2)}{(N+1+2cA)^2}.$$

By matching the coefficients of i^2, i and imposing that the constant term in the above equation is zero, we obtain that (A, B, C) must satisfy the following equations

$$\frac{1}{\lambda}(\Theta - \beta A) = \frac{8cA^2(1+cA)}{(N+1+2cA)^2}, \tag{7}$$

$$\frac{\beta}{\lambda}B = \frac{8cA(1+cA)}{(N+1+2cA)^2}B, \tag{8}$$

$$\frac{\beta}{\lambda}C = \frac{2c(1+cA)(B^2+(\tilde{p})^2)}{(N+1+2cA)^2}. \tag{9}$$

Hence, $A^\star$ must be the unique positive solution to (7), and $B^\star = 0$ otherwise (8) would not hold. Define

$$C^\star = \frac{\lambda}{\beta}\frac{2c(1+cA^\star)(\tilde{p})^2}{(N+1+2cA^\star)^2} = \frac{\frac{\delta}{\beta}-1}{4A^*}(\tilde{p})^2. \tag{10}$$

Then we obtain that a solution to (7)-(9) is given by $(A^\star, 0, C^\star)$. Hence, the value function is given by

$$\bar{v}(i) = -A^\star i^2 + C^\star.$$

Using (2) and (6), we obtain that the ask and bid price pressure policy functions are time-homogeneous, and given by

$$\begin{cases} \tilde{a}(i) = \dfrac{\tilde{p}(1+2cA^\star) - 2NA^\star i}{N+1+2cA^\star}, \\ \tilde{b}(i) = \dfrac{-\tilde{p}(1+2cA^\star) - 2NA^\star i}{N+1+2cA^\star}. \end{cases}$$

Notice that because $A^\star > 0$, we have that both ask and bid price pressures are negatively related to the inventory i.

The bid-ask spread is given by

$$\tilde{a}(i) - \tilde{b}(i) = \frac{2(1+2cA^\star)}{N+1+2cA^\star}\tilde{p},$$

The spread is independent of the inventory and strictly increasing with respect to the holding cost Θ. The positive relationship between Θ and the spread means that when Θ is large, market makers set wide spreads to face their inventory risk by having a large per-unit trading profit. Furthermore, the fact that market makers exercise price pressures while conserving constant spread is in line with the empirical observations of [113]. In this paper, it is found that market makers maintain tight bid-ask spreads but exercise an implicit price pressure by reducing the quantity of orders they supply.

4 Case of heterogeneous market makers

We now consider two market makers who are heterogeneous: their costs for holding inventory are the same but their initial inventories are different ($i_1 \neq i_2$). This case represents for example the scenario where some market makers end their trading day with a null inventory, while others end their trading day with a non null inventory that they liquidate the following day. Each market maker knows the inventory of the other one. Note that some of the computations in this section are based on approximations.

The value functions $\bar{v}_1$ and $\bar{v}_2$ are given by

$$\begin{aligned}
\bar{v}_1(i_1,i_2) &= \sup_{x^1}\mathbb{E}\left[\int_0^\infty e^{-\beta t}(dW_s^{1,(x^1,x^2)} - \Theta(I_s^{1,x^1})^2 ds)|I_0^{1,x^1} = i_1, I_0^{2,x^2} = i_2\right], \\
\bar{v}_2(i_1,i_2) &= \sup_{x^2}\mathbb{E}\left[\int_0^\infty e^{-\beta t}(dW_s^{2,(x^1,x^2)} - \Theta_(I_s^{2,x^2})^2 ds)|I_0^{1,x^1} = i_1, I_0^{2,x^2} = i_2\right],
\end{aligned}$$

where we recall that $x^1 \equiv (x_t^{1,a}, x_t^{1,b})_{t\in[0,\infty)}$ and $x^2 \equiv (x_t^{2,a}, x_t^{2,b})_{t\in[0,\infty)}$. These value functions are the solutions of the following HJB equations:

$$\begin{aligned}
\Theta i_1^2 + \beta\bar{v}_1(i_1,i_2) = \lambda \sup_{x^{1,a},x^{1,b}}\Big[&\left(\tilde{p} - \frac{1}{c}(x^{1,a}+x^{2,a})\right)x^{1,a} + \bar{v}_1(i_1 - x^{1,a}, i_2 - x^{2,a}) - \bar{v}_1(i_1,i_2) \\
&- \left(\frac{1}{c}(x^{1,b}+x^{2,b}) - \tilde{p}\right)x^{1,b} + \bar{v}_1(i_1 + x^{1,b}, i_2 + x^{2,b}) - \bar{v}_1(i_1,i_2)\Big]\Big|_{x^2 = x^{1,\star}}, \quad (11)
\end{aligned}$$

and

$$\Theta i_2^2+\beta\bar{v}_2(i_1,i_2)=\lambda\sup_{x^{2,a},x^{2,b}}\Big[\Big(\tilde{p}-\frac{1}{c}(x^{1,a}+x^{2,a})\Big)x^{2,a}+\bar{v}_2(i_1-x^{1,a},,i_2-x^{2,a})-\bar{v}_2(i_1,i_2)$$
$$-\Big(\frac{1}{c}(x^{1,b}+x^{2,b})-\tilde{p}\Big)x^{2,b}+\bar{v}_2(i_1+x^{1,b},i_2+x^{2,b})-\bar{v}_2(i_1,i_2)\Big]\Big|_{x^1=x^{2,\star}}$$

Since both market makers have same holding cost parameter, then they should have symmetric value functions.Consider quadratic ansatz:

$$\bar{v}_1(i_1,i_2)=-Ai_1^2+2Bi_1i_2+Ci_2^2+Di_1+Ei_2+F,$$
$$\bar{v}_2(i_1,i_2)=Ci_1^2+2Bi_1i_2-Ai_2^2+Ei_1+Di_2+F.$$

We speculate that $A,C>0$. The rationale is that each market maker should prefer a relatively small position to, say, a huge position in either side. However, if the other market maker has a huge position, then he likely sacrifices the trading revenue in exchange for a correction of his inventory level, and such scenario may benefit the market maker whose inventory is at the right level.

Given that $A>0$, the optimal selling strategy given i_1,i_2 is given by

$$\begin{cases}x_{1,a}=c\dfrac{(1+2c(A+B))(\tilde{p}-D)+2(2A+B+2c(A^2-B^2))i_1-2(A+2B)i_2}{(3+2c(A-B))(1+2c(A+B))},\\ x_{2,a}=c\dfrac{(1+2c(A+B))(\tilde{p}-D)+2(2A+B+2c(A^2-B^2))i_2-2(A+2B)i_1}{(3+2c(A-B))(1+2c(A+B))}.\end{cases}\tag{12}$$

The associated ask price pressure is

$$\frac{(1+2c(A-B))\tilde{p}-2(A-B)(i_1+i_2)}{3+2c(A-B)}.$$

The optimal buying strategy given i_1,i_2 is given by

$$\begin{cases}x_{1,b}=c\dfrac{(1+2c(A+B))(\tilde{p}+D)-2(2A+B+2c(A^2-B^2))i_1+2(A+2B)i_2}{(3+2c(A-B))(1+2c(A+B))},\\ x_{2,b}=c\dfrac{(1+2c(A+B))(\tilde{p}+D)-2(2A+B+2c(A^2-B^2))i_2+2(A+2B)i_1}{(3+2c(A-B))(1+2c(A+B))}.\end{cases}\tag{13}$$

The associated bid price pressure is

$$\frac{-(1+2c(A-B))\tilde{p}-2(A-B)(i_1+i_2)}{3+2c(A-B)}.$$

Below we analyse the inventory change after a buy or sell: suppose a buyer comes, then Market Maker 1 has a new inventory of $i_1-x_{1,a}$, and Market Maker 2 has a new inventory of $i_2-x_{2,a}$. The difference between these two new inventory levels, is given by

$$\frac{i_1-i_2}{1+2c(A+B)}.$$

Similarly, suppose a seller comes, then Market Maker 1 has a new inventory of $i_1 + x_{1,b}$, and Market Maker 2 has a new inventory of $i_2 + x_{2,b}$. The difference between these two new inventory levels, is given by

$$\frac{i_1 - i_2}{1 + 2c(A+B)}.$$

Therefore, if two market makers begin with the same inventory, then they have the same inventory at all time. However, if $i_1 \neq i_2$, then their inventory is not the same forever. We expect that the difference between their inventory levels converges to zero as time goes to infinity. Hence, we speculate that $A + B > 0$. In fact, we look for A, B, C such that

$$A > 0, -A < B, C > 0. \tag{14}$$

In the following we assume $i_1 \neq i_2$, so the subsequent result does not reduce to those in [31].

Let us assume for the moment that we already solved A, B, C. Plugging (12) and (13) into (11), and setting the coefficients of i_1, i_2 to zero, we obtain a 2-by-2 linear system of D and E. Solving the system yields

$$D = 0, E = 0.$$

Similarly, plugging (12) and (13) into (11) with $D = E = 0$, and setting the constant term to zero, we obtain that

$$F = \frac{2c\lambda}{\beta} \frac{(1 + c(A+C))(\tilde{p})^2}{(3 + 2c(A-B))^2}. \tag{15}$$

Therefore, the problem is solved once we fix A, B, C. To that end, plugging (12) and (13) into (11), and setting the coefficients of $i_1^2, i_1 i_2, i_2^2$ to zero, we obtain a 3-by-3 non-linear system of A, B, C. We write the system as

$$f_1(\Theta, A, B, C) = f_2(\Theta, A, B, C) = f_3(\Theta, A, B, C) = 0. \tag{16}$$

Explicit expressions for $f_i, i = 1, 2, 3$ are tedious but can be determined with Mathematica. Since we already know that $D = E = 0$, if (14) holds, then Market Maker 1's best inventory level, which maximises his value, is $i_1 = \frac{B i_2}{A}$. Intuitively, if $i_2 >> 0$ then Market Maker 2 wants to sell a lot, Market Maker 1 would benefit from this selling motive if his inventory happens to be negative. Thus, we conjecture that

$$B \leq 0. \tag{17}$$

Thus, the objective of the subsequent analysis is, for $\Theta > 0$, solve A, B, C that solves system (16), subject to (14) and (17).

To begin, we notice that $(\Theta, A, B, C) = (0, 0, 0, 0)$ is a solution to the system. The Jacobian of the system is

$$\begin{pmatrix} -2 & 2\beta & 0 & 0 \\ 0 & 0 & 0 & -2\beta \\ 0 & 0 & -2\beta & 0 \end{pmatrix}$$

So for any $\Theta > 0$ sufficiently small, there exists at least one solution to the system. By analytical implicit function theorem, we may consider Taylor expansions of A, B, C in terms of Θ around 0:

$$\begin{aligned} A &= a_1\Theta + a_2\Theta^2 + a_3\Theta^3 + \dots \\ B &= b_1\Theta + b_2\Theta^2 + b_3\Theta^3 + \dots \\ C &= c_1\Theta + c_2\Theta^2 + c_3\Theta^3 + \dots \end{aligned}$$

and pluging into (16), and setting the coefficients of $\Theta, \Theta^2, \Theta^3$ in the Θ-expansions at 0 to 0, we obtain that

$$\begin{aligned} A &= \frac{1}{\beta}\Theta - \frac{32c\lambda}{9\beta^3}\Theta^2 + \frac{32c^2\lambda(21\beta + 92\lambda)}{81\beta^5}\Theta^3 \dots \\ B &= -\frac{16c\lambda}{9\beta^3}\Theta^2 + \frac{32c^2\lambda(15\beta + 67\lambda)}{81\beta^5}\Theta^3 + \dots \\ C &= \frac{8c\lambda}{9\beta^3}\Theta^2 - \frac{8c^2\lambda(39\beta + 176\lambda)}{81\beta^5}\Theta^3 + \dots \end{aligned}$$

The resulting expansions indicate that for sufficiently small $\Theta > 0$, there is a unique tuple (A, B, C) that solves (16), subject to (14) and (17). This means that the negative relationship between inventory and price pressure that holds in the case of identical market makers extends to the case of heterogeneous market makers. We verify this empirically in Section 7.3.5.

4.1 Comparison with the case where market makers are homogeneous ($i_1 = i_2$)

In this case, we know that both market makers have the same inventory at all time. In this case, we have $v(i) = -A^\star i^2 + C^\star$, where $A^\star > 0$ solves (7) with $N = 2$. Then $A^\star$ has Taylor expansion:

$$A^\star = \frac{1}{\beta}\Theta - \frac{8c\lambda}{9\beta^3}\Theta^2 + \frac{8c^2\lambda(3\beta + 16\lambda)}{81\beta^5}\Theta^3 + \dots$$

So i^2's coefficient is

$$-\frac{1}{\beta}\Theta + \frac{8c\lambda}{9\beta^3}\Theta^2 - \frac{8c^2\lambda(3\beta + 16\lambda)}{81\beta^5}\Theta^3 + \dots \tag{18}$$

In contrast, if we were summing up $-A, 2B, C$ which we obtained for $i_1 \neq i_2$,[1] we obtain the total coefficient of $i_1^2, i_1 i_2, i_2^2$ is

$$-A + 2B + C = -\frac{1}{\beta}\Theta + \frac{8c\lambda}{9\beta^3}\Theta^2 - \frac{8c^2\lambda(3\beta + 8\lambda)}{81\beta^5}\Theta^3 + \dots$$

which is larger than the coefficient in (18) by $\frac{64c^2\lambda^2}{81\beta^5}\Theta^3 + \dots$ In other words, different initial inventory reduces the concavity of the value function. This again confirms that the problem with different initial inventory levels cannot be reduced to that with identical inventory levels.

[1] Although we know that $i_1 \neq i_2$ forever, let's assume for the moment that we could get $i_1 = i_2$ at some point and compute the coefficient of the quadratic term.

Moreover, from (10) and (15) we get

$$C^\star - F = \frac{32c^3\lambda^2(\tilde{p})^2}{243\beta^4}\Theta^2 + ...$$

So value of market makers is higher when they begin the same inventory at 0, than slightly different inventory around 0.

5 Data description and preliminary statistics

5.1 Data description

The data under study are provided by the French Regulator: Autorité des Marchés Financiers (AMF). They include cross-platform transactions of two assets: Société Général and Renault during June 2017 (with 21 trading days) with participant-level information and millisecond time granularity. We choose these assets because they are the two smallest tick assets among those of CAC 40: the average spreads of Renault and Société Générale are equal to 2.67 ticks and 2.63 ticks respectively on Euronext and their ticks are equal to 0.01 euros and 0.005 euros respectively. In fact, price pressures are clearer on a small than a large tick asset, bid and ask variations being not restrained by the tick value.

The collection of these data is possible thanks to the Transaction Reporting Exchange Mechanism (TREM). Under TREM, all regulators must collect transactions data from their national entities and send it to the competent regulator. This reporting mechanism allows AMF to collect all transactions involving any European investment firm on any French financial instrument[2]. TREM data do not provide information about whether the buyer or seller triggers the transaction, which is a relevant information for us. However, merging these data with limit order books data provided by Thomson Reuters Tick History, also with millisecond granularity, allows us to deduce the trigger of the transactions. For each transaction, we compare the price at which the trade takes place to the values of the best bid and ask just before the transaction. Hence, we deduce which member submitted the aggressive order. We are able to do this reconstitution on multiple platforms: Aquis Exchange, Bats, CHI-X, Equiduct, Turquoise and Euronext. We consider all high-frequency market makers OTC trades to be passive orders. These data enable us to conduct a cross-platform study, unique in the literature.

We illustrate market fragmentation in Tables V.1 and V.2 by displaying the repartition of the traded volume and number of trades for the assets Société Générale and Renault across different platforms.

[2]A French instrument is an instrument on which AMF is the relevant competent authority with regard to the transactions reporting according to MiFID directive. The determination of relevant competent authorities is based on the most relevant market in terms of liquidity.

Exchange	OTC	EURONEXT	CHI-X	TURQUOISE	BATS	XUBS	Other
Market share in traded volume	52%	31%	6%	4%	2%	1%	4%
Market share in number of trades	6%	52%	18%	12%	4%	3%	5%

Table V.1 – Market fragmentation (Société Générale).

Exchange	OTC	EURONEXT	CHI-X	TURQUOISE	BATS	XUBS	Other
Market share in traded volume	59%	23%	7%	4%	3%	1%	3%
Market share in number of trades	8%	45%	23%	14%	5%	2%	3%

Table V.2 – Market fragmentation (Renault).

Tables V.1 and V.2 show that Euronext has the highest market share among lit platforms in the case of both assets. Furthermore, these results shed light on OTC trades that constitute a small market share in number of trades but the highest in volume. This shows that the average size of OTC trades is significantly higher than that of the other platforms.

5.2 Market makers activity and identification

Market makers are in general supposed to have a flat position at the end of the day. This feature is not observed when focusing on one platform only. Using cross-platforms trades allows us to track down the flat position of market makers at the end of the day. For instance, we can see in Figure V.1 the difference between the intraday inventory evolution on the 1^{st} of June of one specific market maker on the asset Société Générale when considering Euronext only and when taking into account all the platforms.

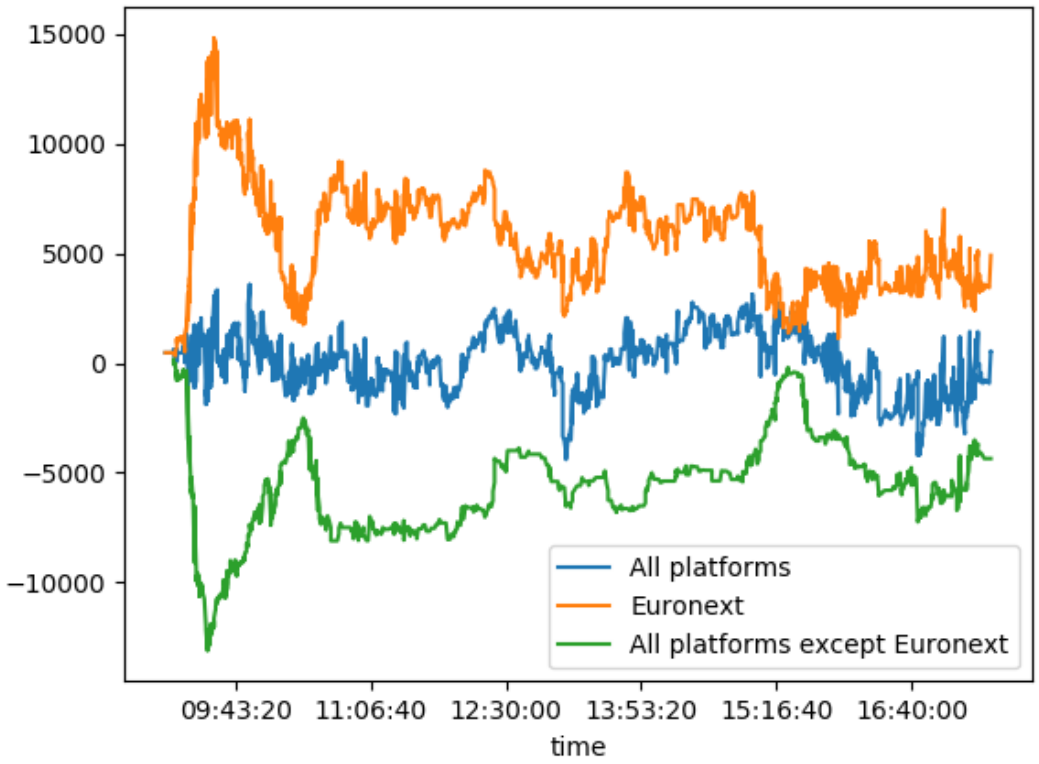


Figure V.1 – The intra-day inventory (in shares) on the 1^{st} of June of a specific market maker on Euronext vs other platforms (Société Générale).

Figure V.1 shows that when looking at the activity of the considered market maker only on Euronext, he seems to be long all over the day. In opposite, he is short on the rest of the platforms cumulated. When merging his trades across all platforms, it is clear that his inventory fluctuates around zero during the trading day and that he ends his day with a null inventory.

In the following, we want to identify traders playing the role of market makers on the studied assets. It is well-known that in nowadays markets, more and more fragmented and fast, only high-frequency traders succeed in playing the role of market makers. This is why, for each asset, we select all HFTs (that we identify based on an AMF classification) participating in transactions as potential market makers. Then we compute their end-of-day position. A given HFT is considered to be a market maker if he satisfies all the following criteria:

- He participates at least to half of the trading days of the period.
- He ends at least half of the trading days of the period with a position less than 100 shares in absolute value.

- His average number of trades per trading day of the period is higher than 10.

In total, we study the end-of-day inventory of 20 HFTs enumerated from MM0 to MM20. Based on the criteria above, the same five market makers stand out for Société Générale and Renault: MM0, MM3, MM4 and MM5. We show the repartition of their trading days according to their end-of-day inventories in Tables V.3 and V.4.

HFT	Number of days ending with a null inventory	Number of days ending with an inventory between 1 and 10 shares (in absolute value)	Number of days ending with an inventory between 11 and 100 shares (in absolute value)	Number of days ending with an inventory between 101 and 1000 shares (in absolute value)	Number of days ending with more than 1000 shares (in absolute value)	Total number of trades
MM0	10	0	1	10	0	75086
MM3	9	10	1	1	0	170621
MM4	21	0	0	0	0	161868
MM5	1	5	15	0	0	14044

Table V.3 – Repartition of trading days according to the end-of-day position of the identified market makers (Société Générale).

HFT	Number of days ending with a null inventory	Number of days ending with an inventory between 1 and 10 shares (in absolute value)	Number of days ending with an inventory between 11 and 100 shares (in absolute value)	Number of days ending with an inventory between 101 and 1000 shares (in absolute value)	Number of days ending with more than 1000 shares (in absolute value)	Total number of trades
MM0	10	0	4	7	0	61665
MM3	10	7	1	3	0	88684
MM4	21	0	0	0	0	110659
MM5	0	12	9	0	0	9309

Table V.4 – Repartition of trading days according to the end-of-day position of the identified market makers (Renault).

Tables V.3 and V.4 show that none of the market makers end his trading day with an inventory higher than 1000 shares. Furthermore, for both assets, MM4 ends the totality of his trading days with a null inventory.

We now show the market share (in number of trades and in traded volume) of each identified market maker, OTC included and OTC excluded for Société Générale and Renault in Tables V.5 and V.6 respectively.

Market maker	Market share in number of trades, OTC included	Market share in traded volume, OTC included	Market share in number of trades, OTC excluded	Market share in traded volume, OTC excluded
MM0	4%	1%	9%	10%
MM3	10%	2%	21%	17%
MM4	9%	2%	20%	17%
MM5	1%	0%	1%	1%

Table V.5 – Market share of each high-frequency market maker in volume (Société Générale).

Market maker	Market share in number of trades, OTC included	Market share in traded volume, OTC included	Market share in number of trades, OTC excluded	Market share in traded volume, OTC excluded
MM0	5%	1%	14%	15%
MM3	8%	2%	20%	18%
MM4	10%	2%	25%	20%
MM5	1%	0%	1%	1%

Table V.6 – Market share of each high-frequency market maker in volume (Renault).

Tables V.5 and V.6 show that high-frequency market makers do not trade significantly OTC. When taking into account OTC trades, their market share is equal to 24% in the case of both assets. When excluding these trades, their market share becomes almost double and is equal to 51% for Société Générale and 60% for Renault.

We illustrate in Tables V.7 and V.8 the market fragmentation by describing the repartition of the traded volume of each market maker across the different platforms respectively for Société Générale and Renault.

Market maker	OTC	EURONEXT	CHI-X	TURQUOISE	BATS	XUBS	Other
MM0	0%	54%	22%	21%	1%	0%	2%
MM3	0%	66%	11%	4%	0%	14%	5%
MM4	0%	65%	16%	13%	3%	0%	3%
MM5	54%	0%	30%	3%	0%	4%	9%

Table V.7 – Repartition of the trading activity of each market maker across different platforms (Société Générale).

Market maker	OTC	EURONEXT	CHI-X	TURQUOISE	BATS	XUBS	Other
MM0	0%	55%	33%	6%	5%	0%	1%
MM3	0%	58%	32%	0%	0%	7%	3%
MM4	0%	52%	19%	18%	7%	0%	4%
MM5	57%	0%	30%	2%	0%	2%	10%

Table V.8 – Repartition of the trading activity of each market maker across different platforms (Renault).

Tables V.7 and V.8 show that the activity of high-frequency market makers is mainly concentrated on lit platforms with the exception of MM5 who has more than half of his activity OTC.

6 Market makers aggressiveness

6.1 Preliminary statistics on market makers aggressiveness

We compute here for each market maker the percentage of his aggressive, passive and undefined orders (orders for which we are not able to find the trigger of the transaction based on the merge of our databases). We present the results for Société Générale and Renault in Tables V.9 and V.10 respectively.

Market maker	Aggressive flow share	Passive flow share	Unknown flow share
MM0	23%	58%	18%
MM3	33%	42%	25%
MM4	73%	15%	12%
MM5	2%	81%	17%

Table V.9 – Market makers aggressiveness (Société Générale).

Market maker	Aggressive flow share	Passive flow share	Unknown flow share
MM0	35%	50%	15%
MM3	50%	34%	16%
MM4	77%	9%	14%
MM5	3%	83%	14%

Table V.10 – Market makers aggressiveness (Renault).

In our model, we consider that market makers are purely passive, which is not the case in practice, as we can see in Tables V.9 and V.10. In the following, we study whether market makers use aggressive orders for arbitrage opportunities only or also for inventory management.

6.2 Market makers aggressiveness according to the inventory

For each market maker, we compute his inventory just before he sends an aggressive order. The inventory is normalised by the average trade size (ATS) of all identified market makers, and grouped into bins. For instance, all inventories with values larger than 1 ATS and lower than 2 ATS are grouped in one inventory bin corresponding to 1 ATS. For each inventory bin, we compute the average sign of the aggressive orders (+1 for a buy aggressive order and -1 for a sell aggressive order) weighted by the traded volume. We only represent inventory bins for which we have at least 200 aggressive orders. MM5 is not represented because of his reduced number of aggressive orders. We show the results obtained for each market maker in the case of Société Générale and Renault in Figures V.2 and V.3 respectively .

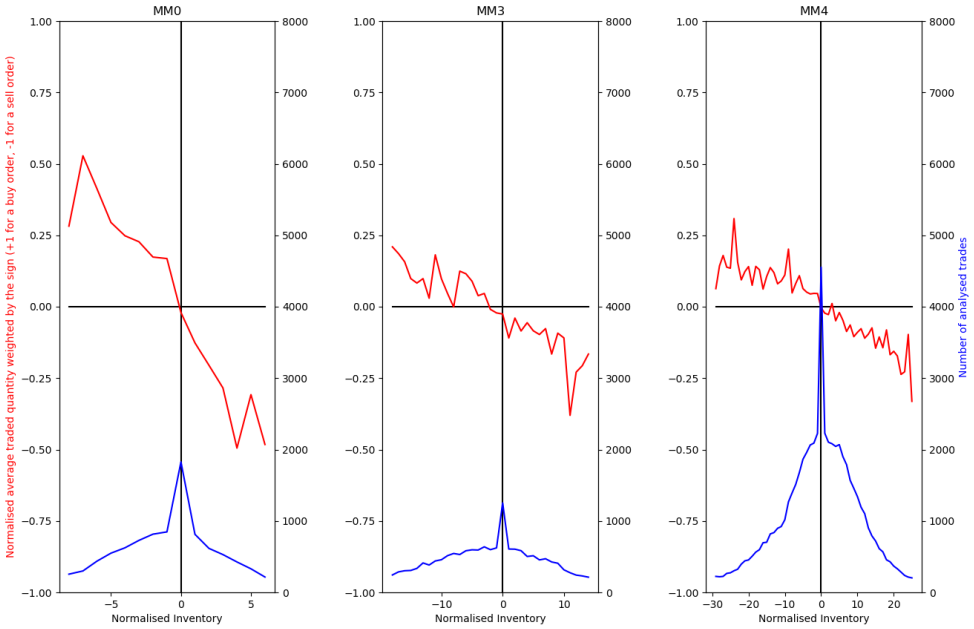


Figure V.2 – Aggressive orders used for inventory management (Société Générale). The red plot corresponds to the average sign of aggressive orders weighted by the traded volume (left y-scale) and the blue plot correspond to the analysed number of orders (right y-scale).

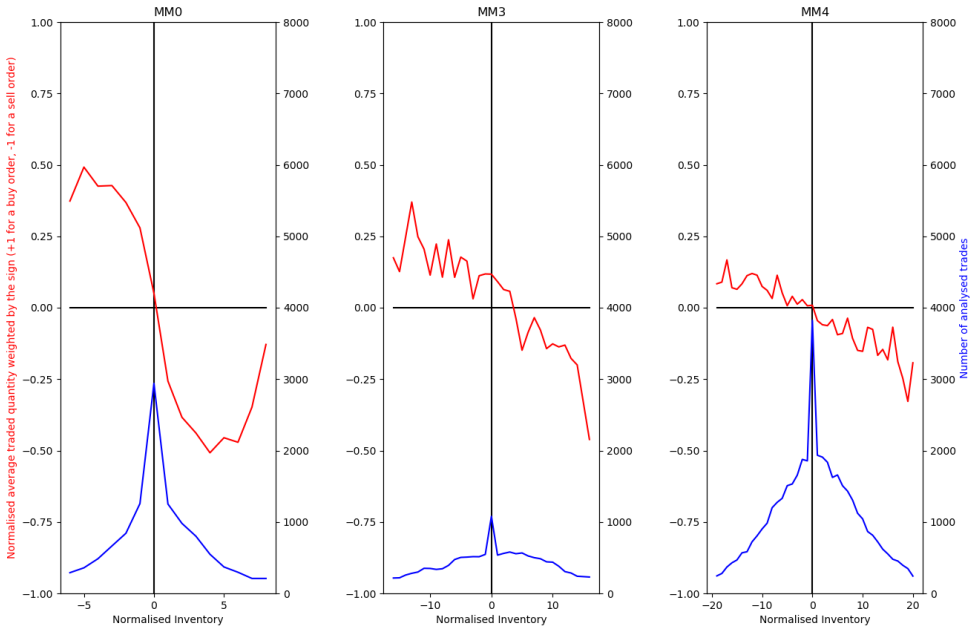


Figure V.3 – Aggressive orders used for inventory management (Renault). The red plot corresponds to the average sign of aggressive orders weighted by the traded volume (left y-scale) and the blue plot correspond to the analysed number of orders (right y-scale).

Note that in Figures V.2 and V.3, the peak of the number of aggressive orders when the inventory is null is due to the fact that this value gathers all inventories between -1 ATS and 1 ATS. For both assets, and for all studied market makers, the average sign of aggressive orders weighted by the traded volume seems to be a decreasing function of the inventory: when the inventory is negative (resp. positive), market makers send more aggressive buy (resp. sell) orders than aggressive sell (resp. buy) orders. When the inventory is null, they buy aggressively as much as they sell aggressively. These observations show that aggressive orders are not used for arbitrage opportunities only, but for inventory management reason too. For this reason, when analysing the relationship between individual inventory and price pressure in Section 7, we take into account aggressive orders too.

It is important to remark that despite the use of aggressive orders by market makers for inventory management, in contrast to the belief that market makers are more aggressive when their inventory is large, we find that their aggressive passive ratio (which is equal to the volume traded aggressively over the volume traded aggressively and passively) remains approximately constant with respect to their inventory, see Figures V.4 and V.5.

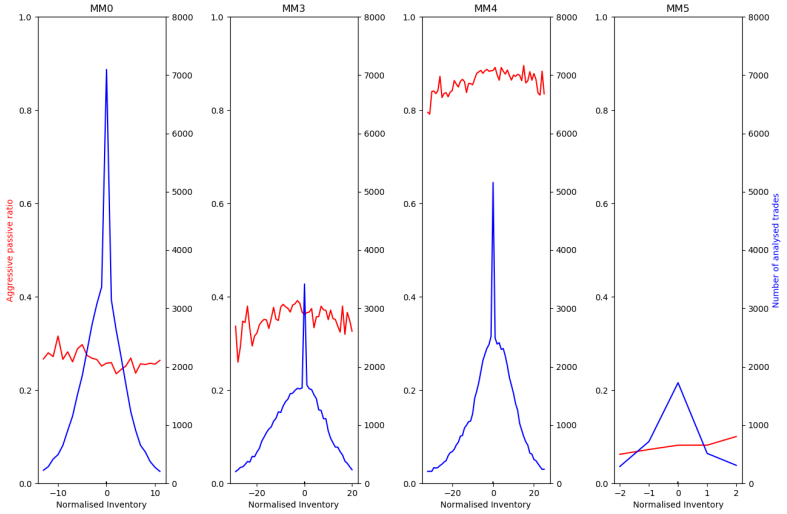


Figure V.4 – In red, the evolution of the aggressive passive ratio of market makers according to the inventory (Société Générale).

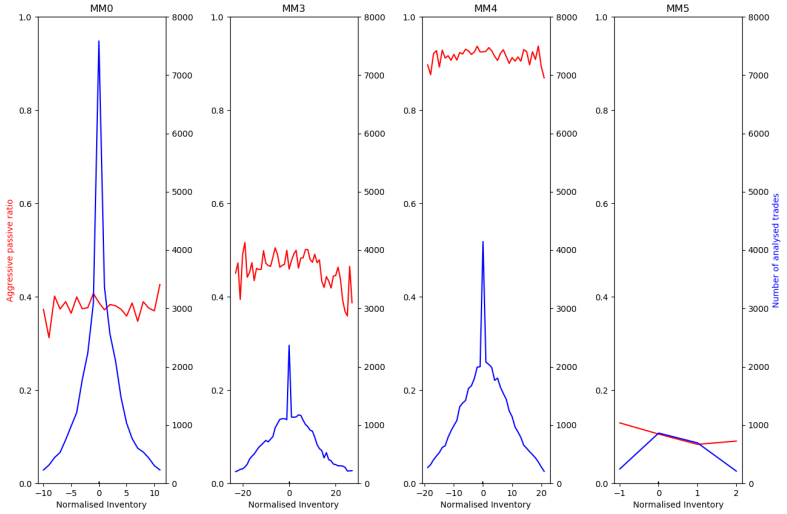


Figure V.5 – In red, the evolution of the aggressive passive ratio of market makers according to the inventory (Renault).

In Figures V.4 and V.5, we represent MM5 too because when passive orders are taken into account, the number of analysed orders becomes sufficient. The aggressive passive ratio is approximately constant for all studied market makers, for both assets, and it even tends to decrease with respect to the inventory in some cases (see MM3 and MM4 in Figure V.4 and MM3 in Figure V.5). This finding is in contrast to the classical idea suggesting that market makers become more aggressive when their inventories are large.

7 Price pressure and market makers inventories

7.1 Measure of the price pressure

Since we have shown that aggressive orders are also used for inventory management, we now study the price pressure of both aggressive and passive orders according to the inventory. The ask (resp. bid) pressure is considered when the market maker sends an aggressive buy (resp. sell) order or a sell (resp. buy) passive order. We define in the following our measures to quantify the ask pressure in the case of a buy aggressive order and a sell passive order. Our measures for the bid pressure are obviously deduced.

When a market maker sends a buy aggressive order, the price can be impacted after the aggressive order is executed. On a given platform, the ask pressure due to the n^{th} buy aggressive order, denoted by AP_n^{ag}, is measured as follows:

$$AP_n^{ag} = A_{n+1}^{ag} - A_n^{ag},$$

where A_n^{ag} is the best ask value on the considered platform just before the n^{th} buy aggressive order.

When a market maker manages his inventory passively, he inserts new limit orders at the ask and bid sides and the price can be impacted as soon as the order is inserted. To avoid spurious price pressure, for example when a market maker inserts a limit order and cancels it immediately, we compute the price pressure of executed passive orders only. On a given platform, the ask pressure due to the n^{th} (executed) sell passive order, denoted by AP_n^{pass}, is measured as follows:

$$AP_n^{pass} = A_n^{pass} - A_{n-1}^{pass},$$

where A_n^{pass} is the best ask value on the considered platform just before the n^{th} executed passive order is executed.

In the following, we are interested in the price pressure due to the set of aggressive and passive orders sent by each high-frequency market maker.

7.2 Empirical analysis of the impact of market makers inventories on prices

We now study the impact of individual inventories on price pressure. Note that we merge here pressures on the bid and ask sides and that both aggressive and passive price pressures

are taken into account (similar patterns are obtained when considering aggressive and passive pressures separately).

We display the endogenous impact of market makers inventories on the price for our two assets in Figures V.6 and V.7. Note that the price pressure cannot be measured for orders taking place OTC since the order book is not transparent and we do not have access to the best bid and ask. This is why the activity of MM5 is only partially studied in the sequel: more than 50% of his activity takes place OTC, as shown in Tables V.7 and V.8.

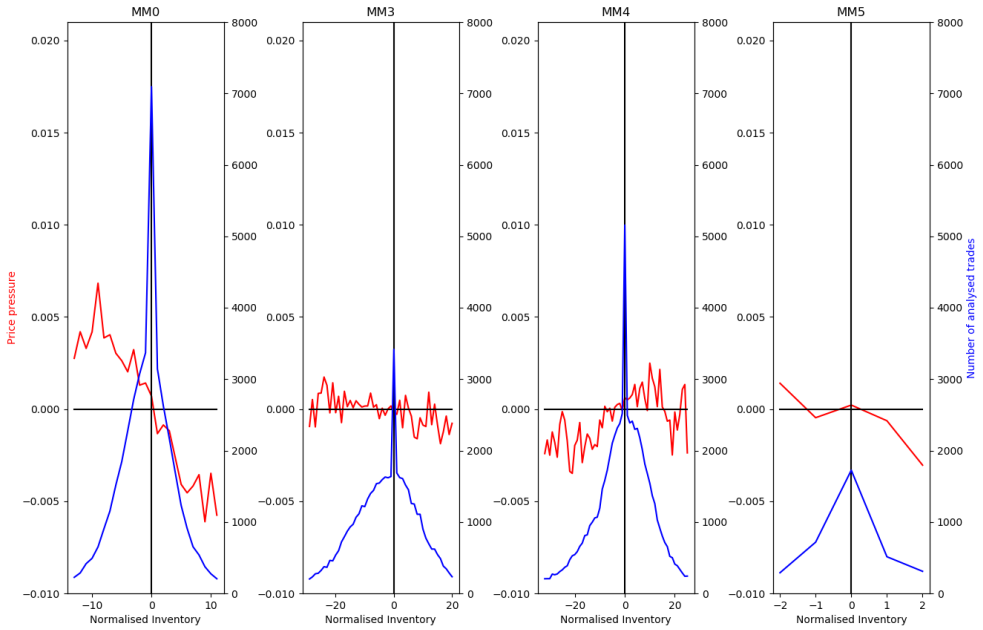


Figure V.6 – The price pressure of market makers according to the inventory (Société Générale). The red curve represents the average price pressure in euros according to the inventory, and the blue one is the number of analysed aggressive and passive orders.

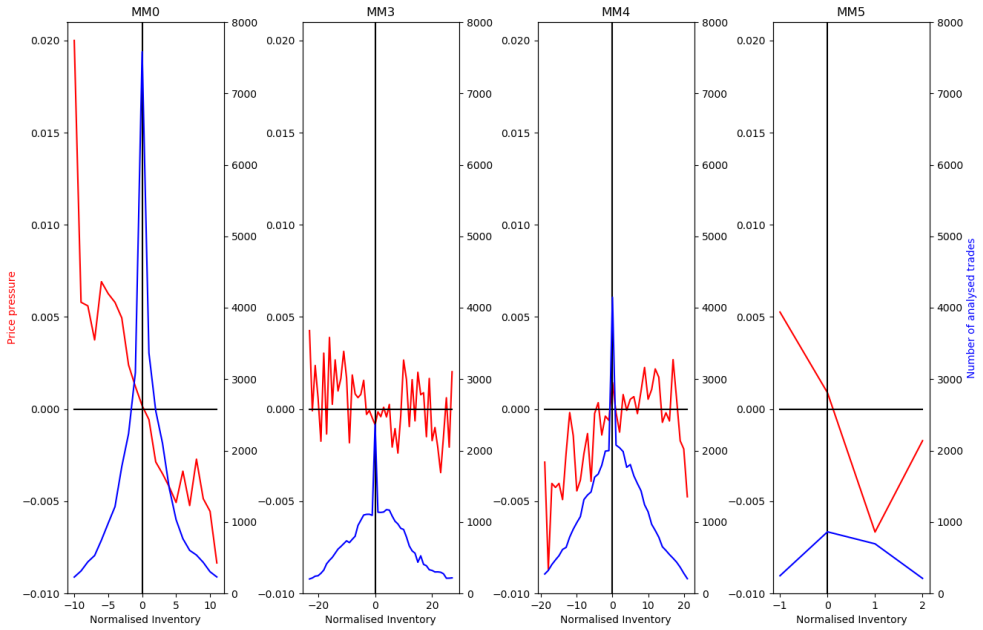


Figure V.7 – The price pressure of market makers according to the inventory (Renault). The red curve represents the average price pressure in euros according to the inventory, and the blue one is the number of analysed aggressive and passive orders.

Figures V.6 and V.7 clearly confirm that the negative relationship between inventory and price pressure found in our theoretical model holds for MM0. This result is less clear for MM3 and MM5. However, not in contradiction with the theory. In contrast, it does not hold for MM4. MM4 seems to have a strategy which is completely different from those of the rest of the agents: we find a positive relationship between his inventory and his price pressure. This shows that MM4 does not follow the behaviour of a typical market maker: his main strategy is probably short-term arbitrage and not pure market making.

The previous graphs are a way to illustrate empirically the impact of the inventory on price pressure. Now, in order to draw more solid conclusions, we use a regression analysis.

7.3 Regression analysis

For each market maker, we consider all the price pressures corresponding to their aggressive and passive orders. We compute linear regressions to explain price pressures by individual market maker inventories (as previously, inventory bins are normalised by the ATS), for the assets Société Générale and Renault. The results for the significant explanatory variables for

each market maker are displayed in Tables V.11 to V.14. We recall that the activity of MM5 is only partially studied.

7.3.1 MM0

Asset	Variable	Coefficient	Std err	t-stat	p-value	[0.025 0.975]	Number of observations
Société Générale	MM0 inventory	-0.0004	$2.25e^{-5}$	-19.503	0.000	[-0.000 -0.000]	41771
Renault	Constant	0.0007	0.000	2.919	0.004	[0.000 0.001]	33771
Renault	MM0 inventory	-0.001	$4.95e^{-5}$	-21.038	0.000	[-0.001 -0.001]	33771

Table V.11 – Regression on MM0 price pressure according to his inventory for Société Générale and Renault.

Table V.11 shows that the negative relationship between inventory and price pressure holds for MM0 in the case of both assets.

7.3.2 MM3

Asset	Variable	Coefficient	Std err	t-stat	p-value	[0.025 0.975]	Number of observations
Société Générale	MM3 inventory	$-2.47e^{-5}$	$6.1e^{-6}$	-4.054	0.000	$[-3.67e^{-5}$ $-1.28e^{-5}]$	51195
Renault	MM3 inventory	$-3.609e^{-5}$	$1.17e^{-5}$	-3.086	0.002	$[-5.9e^{-5}$ $-1.32e^{-5}]$	37369

Table V.12 – Regression on MM3 price pressure according to his inventory for Société Générale and Renault.

Table V.12 shows that as for MM0, the negative relationship between inventory and price pressure holds for MM3 in the case of both assets.

7.3.3 MM4

Asset	Variable	Coefficient	Std err	t-stat	p-value	[0.025 0.975]	Number of observations
Société Générale	MM4 inventory	$4.98e^{-5}$	$6.17e^{-6}$	8.071	0.000	$[3.77e^{-5}$ $6.19e^{-5}]$	71658
Renault	Constant	-0.0007	0.000	-3.329	0.001	[-0.001 -0.000]	49346
Renault	MM4 inventory	0.0002	$2.03e^{-5}$	7.793	0.000	[0.000 0.000]	49346

Table V.13 – Regression on MM4 price pressure according to his inventory for Société Générale and Renault.

We see in Table V.13 that the negative relationship does not hold for MM4: his price pressure and inventory are positively correlated. As explained above, the main strategy of MM4 is probably short-term arbitrage and not pure market making: His aggressive passive ratio higher than 80% (see Figures V.4 and V.5) supports this interpretation, knowing also that HFTs aggressive orders are in general more informed than those of the rest of the market, allowing them to benefit from arbitrage opportunities, see Chapter II.

7.3.4 MM5

Asset	Variable	Coefficient	Std err	t-stat	p-value	[0.025 0.975]	Number of observations
Renault	MM5 inventory	-0.0022	0.001	-4.096	0.000	[-0.003 0.001]	2305

Table V.14 – Regression on MM5 price pressure according to his inventory for Société Générale and Renault.

Table V.14 shows that the negative relationship holds for MM5 in the case of Renault, but no significant results are obtained in the case of Société Générale. Again, this can be explained by the fact that we study partially the activity of MM5 since more than half of his activity takes place OTC.

To sum up, the behaviours of MM0 and MM3 are perfectly in line with the theoretical model. The negative relationship holds for MM5 in the case of Renault but not in the case of Société Générale. This can be due to the fact that his activity is also only partially studied because of his important OTC flow. Finally, the theoretical model does not fit for MM4: the relationship between price pressure and inventory is positive, in opposite to the negative correlation we expect. A possible explanation is that his strategy could be more arbitrage driven than market making driven. Finally, our results suggest that MM0, MM3 and MM5 could accentuate flash crashes in a situation where they detain a large positive inventory and a sudden stress occurs in the market.

Until now, we studied the negative relationship for each market maker individually. In the following, we assess the theoretical negative relationship between price pressure and inventory in the case of multiple market makers. Based on our previous results, MM0, MM3 and MM5 seem to fit the behaviour of a typical market maker. In the following, we consider these three market makers.

7.3.5 Multiple market makers inventories: MM0, MM3 and MM5

We compute now a regression to explain price pressure due to the three considered market makers together by their cumulated inventory. According to Tables V.3 and V.4, these market makers represent the case of heterogeneous market makers since they hold different inventories. The results of the regressions are displayed in Table V.15[3].

[3]The number of observations is not equal to the sum of those of the three studied market makers because the trades issued simultaneously (to the nearest millisecond) by these three different market participants are now considered as one observation.

Asset	Variable	Coefficient	Std err	t-stat	p-value	[0.025 0.975]	Number of observations
Société Générale	Constant	-0.0002	$8.18e^{-5}$	-2.181	0.029	[-0.000 $-1.81e^{-5}$]	81960
Société Générale	Sum of inventories	$-8.314e^{-5}$	$5.55e^{-6}$	-14.989	0.000	[$-9.4e^{-5}$ $-7.23e^{-5}$]	81960
Renault	Sum of inventories	-0.0001	$1.04e^{-5}$	-11.185	0.000	[-0.000 $-9.59e^{-5}$]	61280

Table V.15 – Regression on price pressures due to the three market makers together according to their cumulated inventory for Société Générale and Renault.

The negative link between price pressure and market makers inventory is once again illustrated in Table V.15, here in the case of heterogeneous market makers. This is in line with the theoretical results in Section 4.

Bibliography

[1] Frédéric Abergel and Aymen Jedidi. A mathematical approach to order book modeling. *International Journal of Theoretical and Applied Finance*, 16(05):1350025, 2013.

[2] Frédéric Abergel and Aymen Jedidi. Long-time behavior of a Hawkes process–based limit order book. *SIAM Journal on Financial Mathematics*, 6(1):1026–1043, 2015.

[3] Frédéric Abergel, Charles-Albert Lehalle, and Mathieu Rosenbaum. Understanding the stakes of high-frequency trading. *The Journal of Trading*, 9(4):49–73, 2014.

[4] Tobias Adrian, Agostino Capponi, Erik Vogt, and Hongzhong Zhang. Intraday market making with overnight inventory costs. *Available at SSRN 2844881*, 2016.

[5] Yacine Ait-Sahalia and Mehmet Sağlam. High frequency market making: Optimal quoting. *Available at SSRN 2331613*, 2017.

[6] Robert Almgren, Chee Thum, Emmanuel Hauptmann, and Hong Li. Direct estimation of equity market impact. *Risk*, 18(7):58–62, 2005.

[7] Yakov Amihud and Haim Mendelson. Dealership market: Market-making with inventory. *Journal of Financial Economics*, 8(1):31–53, 1980.

[8] Autorité des Marchés Financiers. Study of the behaviour of high-frequency traders on Euronext Paris. *Risk and Trend Mapping*, 2017.

[9] Autorité des Marchés Financiers. MIFID II: Impact of the new tick size regime. Technical report, AMF, 2018.

[10] Emmanuel Bacry, Adrian Iuga, Matthieu Lasnier, and Charles-Albert Lehalle. Market impacts and the life cycle of investors orders. *Market Microstructure and Liquidity*, 1(02):1550009, 2015.

[11] Matthew Baron, Jonathan Brogaard, Björn Hagströmer, and Andrei Kirilenko. Risk and return in high-frequency trading. *Journal of Financial and Quantitative Analysis*, 54:993–1024, 2019.

[12] Shmuel Baruch and Lawrence R Glosten. Tail expectation, imperfect competition, and the phenomenon of flickering quotes in limit order book markets. *Unpublished Manuscript, David Eccles School of Business, University of Utah*, 2017.

[13] Christian Bayer, Ulrich Horst, and Jinniao Qiu. A functional limit theorem for limit order books with state dependent price dynamics. *The Annals of Applied Probability*, 27(5):2753–2806, 2017.

[14] Paul R Beesack. On some Gronwall-type integral inequalities in n independent variables. *Journal of Mathematical Analysis and Applications*, 100(2):393–408, 1984.

[15] Mario Bellia, Loriana Pelizzon, Marti G. Subrahmanyam, Jun Uno, and Darya Yuferova. Low-latency trading and price discovery: Evidence from the Tokyo stock exchange in the pre-opening and opening periods. *Available at SSRN 2841242*, 2017.

[16] Dan Bernhardt and Eric Hughson. Splitting orders. *The Review of Financial Studies*, 10(1):69–101, 1997.

[17] Bruno Biais, Fany Declerck, and Sophie Moinas. Who supplies liquidity, how and when? *Available at SSRN 2789736*, 2016.

[18] Bruno Biais and Thierry Foucault. HFT and market quality. *Bankers, Markets & Investors*, 128(1):5–19, 2014.

[19] Bruno Biais, Thierry Foucault, and Sophie Moinas. Equilibrium fast trading. *Journal of Financial Economics*, 116(2):292–313, 2015.

[20] Bruno Biais, David Martimort, and Jean-Charles Rochet. Competing mechanisms in a common value environment. *Econometrica*, 68(4):799–837, 2000.

[21] Patrick Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

[22] Pierre Blanc, Jonathan Donier, and Jean-Philippe Bouchaud. Quadratic Hawkes processes for financial prices. *Quantitative Finance*, 17(2):171–188, 2017.

[23] Jean-Philippe Bouchaud, Doyne J Farmer, and Fabrizio Lillo. How markets slowly digest changes in supply and demand. In *Handbook of financial markets: dynamics and evolution*, pages 57–160. Elsevier, 2009.

[24] Selma Boussetta, Laurence Lescourret, and Sophie Moinas. The role of pre-opening mechanisms in fragmented markets. *Available at SSRN 2939502*, 2017.

[25] Antoine Bouveret, Cyrille Guillaumie, Carlos Aparicio Roqueiro, Christian Winkler, and Steffen Nauhaus. High-frequency trading activity in EU equity markets. *ESMA Report on Trends, Risks and Vulnerabilities*, 1:41–47, 2014.

[26] Pierre Brémaud and Laurent Massoulié. Stability of nonlinear Hawkes processes. *The Annals of Probability*, pages 1563–1588, 1996.

[27] Jonathan Brogaard. High frequency trading and its impact on market quality. Technical report, Northwestern University Kellogg School of Management Working Paper, 2010.

[28] Jonathan Brogaard, Allen Carrion, Thibaut Moyaert, Ryan Riordan, Andriy Shkilko, and Konstantin Sokolov. High frequency trading and extreme price movements. *Journal of Financial Economics*, 128(2):253–265, 2018.

[29] Jonathan Brogaard, Terrence Hendershott, and Ryan Riordan. High-frequency trading and price discovery. *The Review of Financial Studies*, 27(8):2267–2306, 2014.

[30] Eric Budish, Peter Cramton, and John Shim. The high-frequency trading arms race: frequent batch auctions as a market design response. *The Quarterly Journal of Economics*, 130(4):1547–1621, 2015.

[31] Agostino Capponi, Albert J Menkveld, and Hongzhong Zhang. Large orders in small markets: On optimal execution with endogenous liquidity supply. *Available at SSRN 3326313*, 2019.

[32] Pierre Cardaliaguet and Charles-Albert Lehalle. Mean field game of controls and an application to trade crowding. *Mathematics and Financial Economics*, 12:1–29, 2017.

[33] Alain Chaboud, Sergey Chernenko, and Jonathan Wright. Trading activity and macroeconomic announcements in high-frequency exchange rate data. *Journal of the European Economic Association*, 6(2-3):589–596, 2008.

[34] Alain Chaboud, Benjamin Chiquoine, Erik Hjalmarsson, and Clara Vega. Rise of the machines: Algorithmic trading in the foreign exchange market. *The Journal of Finance*, 69(5):2045–2084, 2014.

[35] Tarun Chordia, Richard Roll, and Avanidhar Subrahmanyam. Market liquidity and trading activity. *The Journal of Finance*, 56(2):501–530, 2001.

[36] Commission Européenne. Règlement délégué de la commission du 25.4.2016 complétant la directive 2014/65/ue, 2016.

[37] Markets Committee. High-frequency trading in the foreign exchange market. *Bank for International Settlements, Markets Committee Publications*, (5), 2011.

[38] Rama Cont and Adrien De Larrard. Price dynamics in a Markovian limit order market. *SIAM Journal on Financial Mathematics*, 4(1):1–25, 2013.

[39] Rama Cont, Sasha Stoikov, and Rishi Talreja. A stochastic model for order book dynamics. *Operations Research*, 58(3):549–563, 2010.

[40] Daryl J Daley and David Vere-Jones. *An introduction to the theory of point processes: volume II: general theory and structure.* Springer Science & Business Media, 2007.

[41] Khalil Dayri and Mathieu Rosenbaum. Large tick assets: implicit spread and optimal tick size. *Market Microstructure and Liquidity*, 1(01):1550003, 2015.

[42] Joseph L Doob. *Stochastic processes.* Wiley Classics Library. John Wiley & Sons Inc., New York, 1990.

[43] Euronext. Launch of a new service: Retail matching facility, 2012.

[44] Euronext. Revision of the supplemental liquidity provider programme, 2016.

[45] Euronext. Volatility indices rule book version 16-01, 2016.

[46] European parliament. Supplementing directive 2014/65/eu of the European parliament and of the council on markets in financial instruments with regard to regulatory technical standards specifying the requirements on market making agreements and schemes, 2016.

[47] Doyne J Farmer, Laszlo Gillemot, Fabrizio Lillo, Szabolcs Mike, and Anindya Sen. What really causes large price changes? *Quantitative Finance*, 4(4):383–397, 2004.

[48] Fidessa. Fidessa online fragmentation index, 2017.

[49] Thierry Foucault. Order flow composition and trading costs in a dynamic limit order market. *Journal of Financial Markets*, 2(2):99–134, 1999.

[50] Thierry Foucault, Johan Hombert, and Ioanid Roşu. News trading and speed. *The Journal of Finance*, 71(1):335–382, 2016.

[51] Thierry Foucault, Roman Kozhan, and Wing Wah Tham. Toxic arbitrage. *The Review of Financial Studies*, 30(4):1053–1094, 2017.

[52] Lawrence R Glosten. Is the electronic open limit order book inevitable? *The Journal of Finance*, 49(4):1127–1161, 1994.

[53] Lawrence R Glosten and Paul R Milgrom. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics*, 14(1):71–100, 1985.

[54] Martin D Gould, Mason A Porter, Stacy Williams, Mark McDonald, Daniel J Fenn, and Sam D Howison. Limit order books. *Quantitative Finance*, 13(11):1709–1742, 2013.

[55] Martin Hairer. Ergodic properties of a class of non-markovian processes trends in stochastic analysis (London Mathematical Society Lecture note series vol 353), 2009.

[56] Larry Harris. What to do about high-frequency trading. Technical report, CFA Institute, 2013.

[57] Takaki Hayashi and Yuta Koike. Wavelet-based methods for high-frequency lead-lag analysis. *SIAM Journal on Financial Mathematics*, 9(4):1208–1248, 2018.

[58] Terrence Hendershott, Charles M Jones, and Albert J Menkveld. Does algorithmic trading improve liquidity? *The Journal of Finance*, 66(1):1–33, 2011.

[59] Thomas S Ho and Hans R Stoll. Optimal dealer pricing under transactions and return uncertainty. *Journal of Financial Economics*, 9(1):47–73, 1981.

[60] Thomas S Ho and Hans R Stoll. The dynamics of dealer markets under competition. *The Journal of Finance*, 38(4):1053–1074, 1983.

[61] Marc Hoffmann, Mathieu Rosenbaum, and Nakahiro Yoshida. Estimation of the lead-lag parameter from non-synchronous data. *Bernoulli*, 19(2):426–461, 2013.

[62] Roger D Huang and Hans R Stoll. The components of the bid-ask spread: A general approach. *The Review of Financial Studies*, 10(4):995–1034, 1997.

[63] Weibing Huang, Charles-Albert Lehalle, and Mathieu Rosenbaum. Simulating and analyzing order book data: The queue-reactive model. *Journal of the American Statistical Association*, 110(509):107–122, 2015.

[64] Weibing Huang, Charles-Albert Lehalle, and Mathieu Rosenbaum. How to predict the consequences of a tick value change? Evidence from the Tokyo Stock Exchange pilot program. *Market Microstructure and Liquidity*, 2(03n04):1750001, 2016.

[65] Weibing Huang and Mathieu Rosenbaum. Ergodicity and diffusivity of Markovian order book models: a general framework. *SIAM Journal on Financial Mathematics*, 8(1):874–900, 2017.

[66] Weibing Huang, Mathieu Rosenbaum, and Pamela Saliba. From Glosten-Milgrom to the whole limit order book and applications to financial regulation. *Available at SSRN 3343779*, 2019.

[67] Nicolas Huth and Frédéric Abergel. High frequency lead/lag relationships: Empirical facts. *Journal of Empirical Finance*, 26:41–58, 2014.

[68] Jean Jacod. Multivariate point processes: predictable projection, Radon-Nikodym derivatives, representation of martingales. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 31(3):235–253, 1975.

[69] Thibault Jaisson. Market impact as anticipation of the order flow imbalance. *Quantitative Finance*, 15(7):1123–1135, 2015.

[70] George J Jiang, Ingrid Lo, and Giorgio Valente. High-frequency trading around macroeconomic news announcements: Evidence from the us treasury market. Technical report, Bank of Canada Working Paper, 2014.

[71] Charles M Jones. What do we know about high-frequency trading? *Available at SSRN 2236201*, 2013.

[72] Boyan Jovanovic and Albert J Menkveld. Middlemen in limit order markets. *Available at SSRN 1624329*, 2016.

[73] Paul Jusselin and Mathieu Rosenbaum. No-arbitrage implies power-law market impact and rough volatility. *arXiv preprint arXiv:1805.07134*, 2018.

[74] Andrei Kirilenko, Albert S Kyle, Mehrdad Samadi, and Tugkan Tuzun. The flash crash: high frequency trading in an electronic market. *The Journal of Finance*, 72(3):967––998, 2017.

[75] P Kosulajeff. Sur la répartition de la partie fractionnaire d'une variable. *Matematiceskij Sbornik*, 2(5):1017–1019, 1937.

[76] Albert S Kyle. Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, 53:1315–1335, 1985.

[77] Aimé Lachapelle, Jean-Michel Lasry, Charles-Albert Lehalle, and Pierre-Louis Lions. Efficiency of the price formation process in presence of high frequency participants: a mean field game analysis. *Mathematics and Financial Economics*, 10(3):223–262, 2016.

[78] Sophie Laruelle, Mathieu Rosenbaum, and Emel Savku. Assessing MiFID II regulation on tick sizes: A transaction costs analysis viewpoint. *Available at SSRN 3256453*, 2018.

[79] Günter Last. On dependent marking and thinning of point processes. *Stochastic processes and their applications*, 45(1):73–94, 1993.

[80] Charles-Albert Lehalle and Sophie Laruelle. *Market Microstructure in Practice*. World Scientific publishing, 2013.

[81] Charles-Albert Lehalle and Othmane Mounjid. Limit order strategic placement with adverse selection risk and the role of latency. *Market Microstructure and Liquidity*, 3(01):1750009, 2017.

[82] Charles-Albert Lehalle, Othmane Mounjid, and Mathieu Rosenbaum. Optimal liquidity-based trading tactics. *arXiv preprint arXiv:1803.05690*, 2018.

[83] Peter A W Lewis and Gerald S Shedler. Simulation of nonhomogeneous Poisson processes by thinning. *Naval research logistics quarterly*, 26(3):403–413, 1979.

[84] Ananth Madhavan. Market microstructure: A survey. *Journal of Financial Markets*, 3(3):205–258, 2000.

[85] Ananth Madhavan. Exchange-traded funds, market structure, and the flash crash. *Financial Analysts Journal*, 68(4):20–35, 2012.

[86] Ananth Madhavan, Matthew Richardson, and Mark Roomans. Why do security prices change? a transaction-level analysis of nyse stocks. *Review of Financial Studies*, 10(4):1035–1064, 1997.

[87] Nicolas Megarbane, Pamela Saliba, Charles-Albert Lehalle, and Mathieu Rosenbaum. The behavior of high frequency traders under different market stress scenarios. *Market Microstructure and Liquidity*, 3(03n04):1850005, 2017.

[88] Albert J Menkveld. High frequency trading and the new market makers. *Journal of Financial Markets*, 16(4):712–740, 2013.

[89] Sean P Meyn and Richard L Tweedie. Stability of Markovian processes i: Criteria for discrete-time chains. *Advances in Applied Probability*, 24(3):542–574, 1992.

[90] Sean P Meyn and Richard L Tweedie. *Markov chains and stochastic stability*. Springer Science & Business Media, 2012.

[91] Ciamac C Moallemi and Kai Yuan. A model for queue position valuation in a limit order book. *Working paper*, 2016.

[92] Maxime Morariu-Patrichi and Mikko S Pakkanen. State-dependent hawkes processes and their application to limit order book modelling. *arXiv preprint arXiv:1809.08060*, 2018.

[93] Esteban Moro, Javier Vicente, Luis G Moyano, Austin Gerig, Doyne J Farmer, Gabriella Vaglica, Fabrizio Lillo, and Rosario N Mantegna. Market impact and trading profile of hidden orders in stock markets. *Physical Review E*, 80(6):066102, 2009.

[94] James R. Norris. *Markov Chains (Cambridge Series in Statistical and Probabilistic Mathematics)*. Cambridge University Press, July 1998.

[95] Yosihiko Ogata. On Lewis' simulation method for point processes. *IEEE Transactions on Information Theory*, 27(1):23–31, 1981.

[96] Christine A Parlour. Price dynamics in limit order markets. *The Review of Financial Studies*, 11(4):789–816, 1998.

[97] Marcello Rambaldi, Emmanuel Bacry, and Jean-François Muzy. Disentangling and quantifying market participant volatility contributions. *arXiv preprint arXiv:1807.07036*, 2018.

[98] Ryan Riordan and Andreas Storkenmaier. Latency, liquidity and price discovery. *Journal of Financial Markets*, 15(4):416–437, 2012.

[99] Kevin Rock. The specialist's order book and price anomalies. *Review of Financial Studies*, 9:1–20, 1996.

[100] Mathieu Rosenbaum. Integrated volatility and round-off error. *Bernoulli*, 15(3):687–720, 2009.

[101] Ioanid Roşu. A dynamic model of the limit order book. *The Review of Financial Studies*, 22(11):4601–4641, 2009.

[102] Pamela Saliba. The information content of high frequency traders aggressive orders: recent evidences. *Available at SSRN*, 2019.

[103] Martin Scholtus, Dick van Dijk, and Bart Frijns. Speed, algorithmic trading, and market quality around macroeconomic news announcements. *Journal of Banking & Finance*, 38:89–105, 2014.

[104] SEC. Concept release on equity market structure. Technical report, Release No. 34-61358; File No. S7-02-10, 2010.

[105] Duane J Seppi. Liquidity provision with limit orders and a strategic specialist. *Review of Financial Studies*, 10(1):103–150, 1997.

[106] Eric Smith, Doyne J Farmer, Laszlo Gillemot, and Supriya Krishnamurthy. Statistical theory of the continuous double auction. *Quantitative finance*, 3(6):481–514, 2003.

[107] Sasha Stoikov. The micro-price: a high frequency estimator of future prices. *Quantitative Finance*, 18(12):1959–1966, 2018.

[108] Hans R Stoll. The supply of dealer services in securities markets. *The Journal of Finance*, 33(4):1133–1151, 1978.

[109] Bence Tóth, Yves Lemperiere, Cyril Deremble, Joachim De Lataillade, Julien Kockelkoren, and Jean-Philippe Bouchaud. Anomalous price impact and the critical nature of liquidity in financial markets. *Physical Review X*, 1(2):021006, 2011.

[110] Bence Toth, Imon Palit, Fabrizio Lillo, and Doyne J Farmer. Why is equity order flow so persistent? *Journal of Economic Dynamics and Control*, 51:218–239, 2015.

[111] Trésor-Economics. Impact of falling oil prices on the major emerging economies. Technical Report 157, Ministère des Finances et des Comptes Publics Ministère de l'Économie de l'Industrie et du Numérique, November 2016.

[112] John Wilder Tukey. On the distribution of the fractional part of a statistical variable. *Matematiceskij Sbornik*, 4(3):561–562, 1938.

[113] Federal Reserve Bank of New York U.S. Securities U.S. Department of the Treasury, Board of Governors of the Federal Reserve System and Exchange Commission U.S. Commodity Futures Trading Commission. The U.S. treasury market on October 15, 2014, 2015.

[114] Peng Wu, Marcello Rambaldi, Jean-François Muzy, and Emmanuel Bacry. Queue-reactive hawkes models for the order flow. *arXiv preprint arXiv:1901.08938*, 2019.

[115] Matthieu Wyart, Jean-Philippe Bouchaud, Julien Kockelkoren, Marc Potters, and Michele Vettorazzo. Relation between bid–ask spread, impact and volatility in order-driven markets. *Quantitative Finance*, 8(1):41–57, 2008.

[116] Frank Zhang. High-frequency trading, stock volatility, and price discovery. *Available at SSRN 1691679*, 2010.

université PARIS-SACLAY | École doctorale de mathématiques Hadamard (EDMH)

Titre: Trading haute fréquence: Analyse statistique, modélisation et régulation
Mots-clés: bid-ask spread, carnet d'ordres, microstructure de marché, modèle multi-agents, pas de cotation, régulation financière, sélection adverse, tenue de marché, trading haute fréquence, volatilité.

Résumé: Cette thèse est constituée de deux parties liées l'une à l'autre. Dans la première, nous étudions empiriquement le comportement des traders haute fréquence sur les marchés financiers européens. Nous utilisons les résultats obtenus afin de construire dans la seconde partie de nouveaux modèles multi-agents. L'objectif principal de ces modèles est de fournir aux régulateurs et plateformes de négociation des outils innovants leur permettant de mettre en place des règles pertinentes pour la microstructure et de quantifier l'impact des divers participants sur la qualité du marché. Dans la première partie, nous effectuons deux études empiriques sur des données uniques fournies par le régulateur français. Nous avons accès à l'ensemble des ordres et transactions des actifs du CAC 40, à l'échelle de la microseconde, avec par ailleurs les identités des acteurs impliqués. Nous commençons par comparer le comportement des traders haute fréquence à celui des autres intervenants, notamment pendant les périodes de stress, en termes de provision de liquidité et d'activité de négociation. Nous approfondissons ensuite notre analyse en nous focalisant sur les ordres consommant la liquidité. Nous étudions leur impact sur le processus de formation des prix et leur contenu informationnel selon les différentes catégories de flux. Dans la seconde partie, nous proposons trois modèles multi-agents. À l'aide d'une approche à la Glosten-Milgrom, nous parvenons avec notre premier modèle à construire l'ensemble du carnet d'ordres (spread et volume disponible à chaque prix) à partir des interactions entre trois types d'agents: un agent informé, un agent non informé et des teneurs de marché. Ce modèle nous permet par ailleurs de développer une méthodologie de prédiction du spread en cas de modification du pas de cotation et de quantifier la valeur de la priorité dans la file d'attente. Afin de se concentrer sur une échelle individuelle, nous proposons une deuxième approche où les dynamiques spécifiques des agents sont modélisées par des processus de type Hawkes non linéaires et dépendants de l'état du carnet d'ordres. Dans ce cadre, nous sommes en mesure de calculer en fonction des flux individuels plusieurs indicateurs pertinents relatifs à la microstructure. Il est notamment possible de classer les teneurs de marché selon leur contribution propre à la volatilité. Enfin, nous introduisons un modèle où les fournisseurs de liquidité optimisent leurs meilleurs prix à l'achat et à la vente en fonction du profit qu'ils peuvent générer et du risque d'inventaire auquel ils sont confrontés. Nous mettons alors en évidence théoriquement une corrélation négative entre inventaire des teneurs de marché et pression exercée sur les prix. Nous confirmons ce résultat empiriquement en étudiant les inventaires individuels des teneurs de marché dans un contexte multi-plateformes.

Title : High-frequency tading : Statistical analysis, modelling and regulation
Keywords : adverse selection, agent-based model, bid-ask spread, financial regulation, high frequency trading, limit order book, market making, market microstructure, queue position valuation, volatility, tick size.

Abstract : This thesis is made of two related parts. In the first one, we study the empirical behaviour of high-frequency traders on European financial markets. We use the obtained results to build in the second part new agent-based models for market dynamics. The main purpose of these models is to provide innovative tools for regulators and exchanges allowing them to design suitable rules at the microstructure level and to assess the impact of the various participants on market quality. In the first part, we conduct two empirical studies on unique data sets provided by the French regulator. It covers the trades and orders of the CAC 40 securities, with microseconds accuracy and labelled by the market participants identities. We begin by investigating the behaviour of high-frequency traders compared to the rest of the market, notably during periods of stress, in terms of liquidity provision and trading activity. We work both at the day-to-day scale and at the intraday level. We then deepen our analysis by focusing on liquidity consuming orders. We give some evidence concerning their impact on the price formation process and their information content according to the different order flow categories: high-frequency traders, agency participants and proprietary ones. In the second part, we propose three different agent-based models. Using a Glosten-Milgrom type approach, the first model enables us to deduce the whole limit order book (bid-ask spread and volume available at each price) from the interactions between three kinds of agents: an informed trader, a noise trader and several market makers. It also allows us to build a spread forecasting methodology in case of a tick size change and to quantify the queue priority value. To work at the individual agent level, we propose a second approach where market participants specific dynamics are modelled by non-linear and state dependent Hawkes type processes. In this setting, we are able to compute several relevant microstructural indicators in terms of the individual flows. It is notably possible to rank market makers according to their own contribution to volatility. Finally, we introduce a model where market makers optimise their best bid and ask according to the profit they can generate and the inventory risk they face. We then establish theoretically a negative relationship between market makers inventories and price pressure. We confirm this result studying empirically individual market makers inventories in a multi-platform framework.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France