



**HAL**  
open science

# Analysis and modeling of high frequency dynamics in financial markets

Peng Wu

► **To cite this version:**

Peng Wu. Analysis and modeling of high frequency dynamics in financial markets. General Mathematics [math.GM]. Université Paris sciences et lettres, 2022. English. NNT : 2022UPSLD008 . tel-03721215

**HAL Id: tel-03721215**

**<https://theses.hal.science/tel-03721215>**

Submitted on 12 Jul 2022

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

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



**THÈSE DE DOCTORAT**  
**DE L'UNIVERSITÉ PSL**

Préparée à Université Paris-Dauphine

**Modélisation de flux d'ordres à haute fréquence et de la  
dynamique des prix sur les marchés financiers**

Soutenue par

**Peng Wu**

Le 16 Juin 2022

École doctorale n°543

**Ecole Doctorale SDOSE**

Spécialité

**Mathématiques**

Composition du jury :

M. Jean-Philippe Bouchaud Directeur de recherche, Capital Fund Management	<i>Rapporteur</i>
M. Damien Challet Professeur, CentraleSupélec	<i>Rapporteur</i>
M. Fabrizio Lillo Professeur, University of Bologna	<i>Membre du jury</i>
Mme. Sophie Laruelle Maître de Conférences, Université Paris-Est Créteil	<i>Membre du jury</i>
M. Mathieu Rosenbaum Professeur, Ecole Polytechnique	<i>Président du jury</i>
M. Emmanuel Bacry Directeur de recherche CNRS, Univer- sité Paris-Dauphine	<i>Directeur de thèse</i>
M. Jean-François Muzy Directeur de recherche CNRS, Univer- sité de Corse	<i>Directeur de thèse</i>



The universe is amazingly broad, but  
life is even broader.

---

Liu Cixin, science fiction writer

## Remerciements

Je tiens en premier lieu à remercier mes directeurs de thèse Emmanuel Bacry et Jean-François Muzy. Leur enthousiasme et leur vaste connaissance en mathématique m'ont constamment guidé sur le bon chemin. Merci pour tout ce que vous m'avez fait sous votre direction !

Je remercie Jean-Philippe Bouchaud et Damien Challet pour l'intérêt qu'ils ont porté à mon travail en acceptant de rapporter ma thèse. Je suis également très honoré de la présence de Fabrizio Lillo, Mathieu Rosenbaum et de Sophie Laruelle dans mon jury de thèse.

Je suis très reconnaissant envers mon co-auteur Marcello Rambaldi. Pour moi, il est un fiable ami et aussi un mentor. J'ai beaucoup appris de lui à avancer dans mes recherches.

Je remercie ensuite tous mes amis, en particulier les doctorants, ingénieurs et jeunes chercheurs j'ai reconnu pendant ma thèse. J'ai bénéficié des séminaires et des nombreuses discussions avec eux, qui ont tous devenu mon souvenir précieux. L'ambiance conviviale du CEREMADE m'as constamment inspiré et encouragé. Je tiens à remercier : Philip Deagen, Shuoqing Deng, Martin Bompaire, Ruihua Ruan, Maryan Morel et Yating Liu. Je tiens à remercier spécialement Long Meng et Changqing Fu, pour les moments de la randonnée et du basketball ou on s'échappait du bureau.

Je remercie également à Marie Belle, Isabelle Bellier et Cesar Faivre pour leur chaleureux support. Je remercie aussi Gilles Bares et Thomas Duleu.

A mon parrain Robert Pelletier et sa femme Catherine Pelletier. Vous êtes devenu ma deuxième famille depuis mon arrivée en France.

Mes derniers remerciements sont réservés à mes parents et à l'ensemble des membres de ma famille. Ils ont toujours soutenu pendant ces années. Je suis infiniment reconnaissant pour leurs encouragements précieux et leur confiance en moi.

Après tout, c'est un voyage magique auquel je ne m'attendais pas. Pour conclure, je voudrais vous présenter un ancien proverbe chinois, que l'on peut traduire par : « Tout est au-delà des mots ».

## Résumé

Cette thèse est consacrée à l'étude de la dynamique des prix sur les marchés électroniques. Notre premier objectif est de construire un modèle fiable pour les flux d'ordres qui arrivent aux meilleurs niveaux du carnet d'ordres. La question sous-jacente à laquelle nous tentons de répondre est de savoir comment caractériser les interactions entre les flux d'ordres. L'état du carnet d'ordres influence-t-il également cette interaction ?

Le deuxième sujet de cette thèse est l'analyse de la volatilité multifractale. Nous étudions tout d'abord les similarités entre les propriétés d'invariance d'échelle de deux classes populaires de modèles de volatilité stochastique, à savoir le modèle de "Rough Fractional Stochastic Volatility" (RFSV) et le modèle de "Multifractal Random Walk" (MRW). Les similitudes suggèrent-elles des liens plus profonds? Quelle est la meilleure méthode pour mesurer la rugosité de la volatilité à partir de données empiriques ? Pour répondre à ces questions, nous commençons par analyser la limite  $H \rightarrow 0$  dans une variante du modèle RFSV.

Enfin, nous considérons un sujet plus ambitieux qui consiste à modéliser des dynamiques plus complexes comme la covariance des prix avec des processus ponctuels. Le défi consiste à quantifier la contribution marginale de chaque participant du marché à la covariance. Existe-t-il un lien entre la contribution de la covariance et les profils statistiques des participants du marché ?

Tous les modèles décrits dans cette thèse sont examinés avec divers ensembles de données. Des résultats empiriques sont rapportés, avec quelques faits stylisés commentés.

Le manuscrit est divisé en cinq chapitres. Dans le Chapitre 2, nous introduisons le modèle *Single Queue Reactive Hawkes* (SQRH). Nos résultats suggèrent que pour construire un modèle fiable de flux d'ordres, l'état actuel de LOB ainsi que l'effet d'auto-excitation de flux d'ordres précédents sont importants.

Dans le Chapitre 3, nous étendons le SQRH introduit dans le chapitre précédent sur les deux meilleurs côtés du LOB. Ce modèle permet de révéler la dynamique conjointe de flux d'ordres arrivant sur les deux côtés du LOB. Nous constatons notamment que notre variable d'état ainsi définie est fortement liée au déséquilibre des volumes.

Dans le Chapitre 4, nous introduisons une famille de mesures aléatoires log-normales paramétrées et construisons le modèle dit S-fBM. Ce modèle nous permet de considérer les deux classes populaires de modèles de volatilité dans le même cadre : le modèle multifractal (caractérisée par un paramètre de Hurst  $H = 0$ ) et le modèle volatilité rugueuse (caractérisée par un paramètre de Hurst  $H > 0$ ). Nous montrons notamment que l'estimation de ce paramètre à partir des propriétés d'échelle peut conduire à une forte surestimation. Nous proposons une méthode d'estimation basée sur GMM qui, lorsqu'elle est appliquée à un large ensemble de données empiriques sur la volatilité, conduit à des valeurs de  $H$  très proches de 0 pour les actions tandis que  $H$  est significativement plus grand pour les indices.

Dans le Chapitre 5, nous introduisons un modèle basé sur les processus de Hawkes pour reconstruire la covariance des prix à partir des événements du carnet d'ordres. En introduisant des relations approchées au premier ordre de la norme de la matrice des noyaux, nous arrivons à quantifier la contribution marginale de chaque participant du marché à la covariance des prix. Les résultats empiriques suggèrent que la contribution des participants du marché à la covariance des prix est indépendante des actifs aux lesquels interviennent ces acteurs. Nous montrons que leur contribution est fortement liée à leur profil statistique. Ce chapitre ne contient que des travaux préliminaires avec des résultats illustratifs, qui pourraient être étendus dans les recherches en future.

## Abstract

This thesis is devoted to the study of price dynamics in electronic markets. The first objective of this thesis is to construct a reliable model for order flows that arrive at the best levels of the Limit Order Book. The underlying question we try to answer is how to characterize interactions between order flows? Does the state of the limit order book influence such interaction?

The second topic of this thesis is the analysis of multifractal volatility. We start by examining the similarity in the so-called scaling behaviors between two popular classes of stochastic volatility models, namely the “Rough Fractional Stochastic Volatility” (RFSV) model, and the “Multifractal Random Walk” (MRW) models. Does such similarity suggest more profound links? Also, is there a better method to measure the roughness of volatility from empirical data? To answer such questions, we proceed by considering the  $H \rightarrow 0$  limit in the RFSV model.

Finally, a more ambitious topic consists of modeling complex dynamics, such as the covariance of prices, with point processes. The challenge is to quantify each market participant’s marginal contribution to the covariance. Meanwhile, is there a link between the covariance contribution and the statistical profiles of market participants?

All models outlined in this thesis are examined with various data sets. Empirical findings are reported, together with some stylized facts commented.

This thesis is divided into five chapters. In the introduction chapter, we define terms and present preceding works that inspired this thesis. In Chapter 2, we introduce the *Single Queue Reactive Hawkes* (SQRH) model. Our results suggest that to construct a reliable order flow model, both the current state of LOB and the self-exciting effect between order flows are important.

In Chapter 3, we extend the SQRH introduced in the previous chapter to both the best ask and best bid side of the LOB. This model reveals the joint dynamics of order flows on both sides of the LOB. We notably find that our so-defined state variable is closely linked to the imbalance of volume.

In Chapter 4, we introduce a family of parametrized log-normal random measures and construct the so-called S-fBM model. This model allows us to consider the two popular classes of multifractal volatility (features with a Hurst parameter  $H = 0$ ) and rough volatility (features with a Hurst parameter  $H > 0$ ) models within the same framework. We notably show that estimating the Hurst parameter  $H$  from the scaling properties can lead to strong overestimation. We propose a GMM-based estimation method that, when applied to a large set of empirical volatility data, leads to values of  $H$  very close to 0.

In Chapter 5, we introduce a multivariate Hawkes process based model to retrieve covariance of price (as per return) from Limit Order Book events. By introducing approximate relations to the first order of the kernel norm matrix, we quantified each market participant’s contribution toward the price variance. Empirical results suggest

that market participants' contribution toward covariance is asset-independent. It is also closely linked with the statistical profile of market participants. This chapter only contains preliminary works with illustrative results, which could be extended as future research.

# Contents

<b>1</b>	<b>Introduction</b>	<b>11</b>
1.1	Motivations and thesis axis . . . . .	11
1.2	Basics of financial markets . . . . .	14
1.2.1	Electronic Markets . . . . .	14
1.2.2	Limit order book . . . . .	14
1.2.3	Market participants . . . . .	16
1.3	Modelling financial markets . . . . .	17
1.3.1	Price modeling . . . . .	17
1.3.2	Hawkes process . . . . .	20
1.3.3	The Hawkes model for order flows in LOB . . . . .	23
1.4	Rough and multifractal volatility . . . . .	23
1.4.1	Multifractal volatility . . . . .	25
1.4.2	Rough volatility . . . . .	27
1.4.3	The $H \rightarrow 0$ limit in rough volatility models . . . . .	28
1.4.4	Generalized method of moments (GMM) . . . . .	29
1.5	Data presentation . . . . .	31
1.6	Résultats de la thèse . . . . .	32
1.6.1	Modèle de Hawkes réactif à file d'attente pour le flux d'ordres . . . . .	33
1.6.2	Un modèle de Hawkes réactif en file d'attente pour les meilleures limites du carnet d'ordres . . . . .	37
1.6.3	De la volatilité rugueuse à la volatilité multifractale . . . . .	40
1.6.4	Dissocier les contributions à la covariance des prix entre différents actifs par des participants au marché . . . . .	45
<b>2</b>	<b>A single queue reactive Hawkes models for the order flow</b>	<b>49</b>
2.1	Introduction . . . . .	49
2.2	A Single Queue Reactive model with memory . . . . .	52
2.2.1	Model and assumptions . . . . .	52
2.2.2	Calibration results . . . . .	55
2.3	Summary and prospects . . . . .	64
2.4	Acknowledgements . . . . .	65

2.5	Appendix . . . . .	65
2.5.1	Proof of the existence of invariant distribution . . . . .	65
2.5.2	Calculation of the log-likelihood function of SQRH model and its gradient . . . . .	68
2.5.3	Choice of parameters $U$ and $\beta_m$ . . . . .	70
<b>3</b>	<b>A Queue Reactive Hawkes model for the best limits of the order book</b>	<b>71</b>
3.1	Introduction . . . . .	71
3.2	The Queue Reactive Hawkes (QRH) model . . . . .	72
3.2.1	Introduction to the QRH model . . . . .	72
3.2.2	From log-likelihood estimation to least square estimation . . . . .	73
3.2.3	Fitting results and comments . . . . .	76
3.3	Summary and prospects . . . . .	82
3.4	Appendix . . . . .	83
3.4.1	Calculating the least squares function of QRH model and its gradient	83
<b>4</b>	<b>From Rough to Multifractal volatility: the log S-fBM model</b>	<b>87</b>
4.1	Introduction . . . . .	87
4.2	The log Stationary fractional Brownian Motion (log S-fBM) stochastic volatility model . . . . .	89
4.2.1	Multifractal and rough volatility models . . . . .	89
4.2.2	The log S-fBM random measure : a common framework for RFSV and MRM models . . . . .	91
4.3	Second order properties of $M_{H,T}([0, t])$ and its logarithm . . . . .	93
4.3.1	Integrated variance correlation function . . . . .	94
4.3.2	Small $\lambda^2$ approximation of the logarithm integrated variance moments . . . . .	95
4.4	Estimation . . . . .	97
4.4.1	Bias of the moment scaling method proposed in Ref. [65] . . . . .	97
4.4.2	Low versus high frequency regime for GMM estimations . . . . .	99
4.4.3	Defining two GMM estimators for $H$ and $\lambda^2$ . . . . .	103
4.4.4	Numerical illustrations and empirical performances of the GMM methods . . . . .	103
4.5	Application to realized volatility of asset returns . . . . .	107
4.6	Conclusion . . . . .	110
4.7	Appendix . . . . .	111
4.7.1	Construction of the S-fBM process . . . . .	111
4.7.2	Constructing fBM process from Takenaka field and proof of Proposition 3 . . . . .	114
4.7.3	The case $H \rightarrow 0$ : Convergence towards the MRM log-normal measure . . . . .	116

4.7.4	Proof of Eqs. (4.3.21) and (4.3.22) . . . . .	118
4.7.5	Proof of proposition 8 . . . . .	119
<b>5</b>	<b>Disentangling market participants contributions toward covariance of price between different assets</b>	<b>121</b>
5.1	Introduction . . . . .	121
5.2	Modeling covariance of price by multivariate Hawkes process . . . . .	122
5.2.1	Endogenous covariance and exogenous covariance . . . . .	124
5.2.2	Measuring market participants' marginal contribution toward the price covariance . . . . .	125
5.3	Empirical results . . . . .	128
5.3.1	Calibration approach . . . . .	128
5.3.2	Interpretation of numerical results . . . . .	132
5.4	Summary and prospects . . . . .	138
5.5	Appendix . . . . .	139
5.5.1	Example of calibration error by Bootstrapping . . . . .	139



# Chapter 1

## Introduction

### 1.1 Motivations and thesis axis

This thesis was motivated by a desire to analyze the price dynamics in financial markets. Historically, stochastic calculus has been developed for price modeling and risk management purposes, where the price of assets is implicitly supposed to be a diffusion process.

From a practical point of view, such characterizations are no longer realistic when examined at the scale where the price is formed. At this scale, it is indispensable to incorporate fundamental structures in financial markets into the modeling, such as the Limit Order Book and tick size. In the meantime, new mathematical techniques from statistical inference and signal processing have been developed in the past decades, which allow one to calibrate and test complicated models against observations of large volumes.

In this thesis, the objective is to build more realistic models at the price formation level. Notably, we are interested in quantifying the dynamics of various types of events in financial markets and describing their interactions. Some of the underlying questions we try to answer are the followings: **What is the influence of past events on the financial market? How to describe the interactions between order flows? Could the modeling of order flows be extended to quantify more complicated quantities, such as the volatility and the covariance of the price? Besides, could statistical inference techniques bring new insights by revealing connections between known stylized facts?**

To answer these questions, the first step is to examine historical data which records various manifestations of financial markets. We identify and comment on some stylized facts observed in our empirical data sets. Inspired by some intuitive findings, we proceed by formalizing these ideas within rigorous frameworks of financial mathematics, which finally leads to the models presented in this thesis. In fact, these models not only capture the desired stylized facts but also provide a satisfying interpretation of other related behaviors observed on the market.

From the perspective of modeling, ideally, the ultimate model should be able to describe the dynamics observed across different scales. However, models presented in this thesis adapt themselves to either a specific time scale where the stylized fact is observed or to a specific quantity whose behavior is modeled across different time scales. For instance, the order flows models presented in Chapter 2 and Chapter 3 yield satisfactory results at the tick level, however, the setup of these models is too granular to reproduce empirical behaviors on a coarser scale. In contrast, a scale-invariant volatility model is proposed in Chapter 4 to account for recently discovered stylized facts.

In the following, the main motivations behind each chapter are presented. We comment on our findings and present new insights brought by our models. For a more detailed presentation of each work, please refer to Section 1.6.

The work presented in Chapter 2 focuses on modeling the stochastic time evolution of a single queue (best bid queue or best ask queue) of a limit order book. We noticed that the so-called queue-reactive (QR) model presented in [87] properly described the intuitive idea that the arrival of order flows depends on the current state of the limit order book. Meanwhile, Hawkes processes offer a good framework to describe the mutual exciting effect between order flows, see [22] for example. The challenge is, **can we improve the queue reactive model by incorporating the self-exciting property described by Hawkes processes? Or equivalently, can we bring the queue-reactive property into previous models based on Hawkes processes?** With this motivation, we introduce the model named “Single Queue Reactive Hawkes” (SQRH) as a multivariate Hawkes process with an explicit dependency on queue size. We provide an explicit way to calibrate this model with a Maximum-Likelihood method. Empirical results show that this model improves the description of the order flow properties and the shape of the queue distributions. This chapter will be published as an article in “Market microstructure and liquidity”, co-authored with Marcello Rambaldi, Emmanuel Bacry and Jean-Francois Muzy.

Chapter 3 is the continuation of the previous chapter. We were driven by the motivation of **extending the SQRH model to both the best ask and best bid side of the Limit Orderbook. Our objective is to reveal the joint dynamics of order flows on both sides of the LOB.** The state dependency is set to queue sizes at both the best bid and the best ask prices. Empirical results suggest that our so-defined state variable is closely linked to the imbalance of volume. This chapter is an unpublished work with Emmanuel Bacry and Jean-Francois Muzy.

In Chapter 4, we turn our attention to the modeling of the volatility of the price. We are inspired by the scale-invariant properties in multifractal random walk models, see [20]. By noticing the inspiring similarities between rough volatility models and multifractal volatility models, **we try to address the question of the unification**

of the self-similar multifractal measure defined in [118, 12] with the **Rough Fractional Stochastic Volatility (RFSV) model** under the same framework. **Meanwhile, we propose a more reliable approach to calibrate the roughness of volatility from empirical data.** To start, we define a new model named S-fBM which allows properly defining the  $H \rightarrow 0$  limit of the RFSV model. We find that by taking the limit  $H \rightarrow 0$ , we retrieved the celebrated log-normal multifractal random measure (MRM), which allows us to consider, the multifractal ( $H = 0$ ) volatility models and rough volatility ( $0 < H < 1/2$ ) models within the same framework. For estimation issues, we notably show that the direct estimation of  $H$  from the so-called scaling properties can strongly overestimate the value of  $H$ . We propose a better GMM estimation method that is shown to be valid in the high-frequency asymptotic regime. When applied to a large set of empirical volatility data, empirical evidence suggests that stock indices have values around  $H \simeq 0.1$ . We also bring evidence that the estimation of the so-called “intermittency coefficient”  $\lambda^2$ , which is the product of  $\nu^2$  (the variance parameter of the volatility process) and the Hurst exponent  $H$ , appears to be far more reliable, leading to values that seem to be universal for respectively all individual stocks and all stock indices. This chapter is submitted as an article toward “Physica A: Statistical mechanics and its applications”, co-authored with Emmanuel Bacry and Jean-Francois Muzy.

In Chapter 5, we shift to the topic of modeling the correlated price changes at the order book level. **This chapter is motivated by the idea of reconstructing the joint mid-price changes of a basket of assets from corresponding Limit Order Book events.** We introduced a general framework based on the multivariate Hawkes process to express correlated price changes as cumulative consequences of order flows in correlated markets. With some reasonable assumptions, we continue by quantifying each market participant’s contribution to price covariance as their marginal influence over the kernel norm matrix of the underlying Hawkes process. Empirical results suggest that our model successfully reproduces the covariance of price. Besides, a detailed investigation suggests that active market participants adopt multi-asset strategies, and their contribution toward price correlation is uniform over different pairs of assets. From the result obtained in this Chapter, some possible improvement includes characterizing market participants with more various statistical features and establishing quantitative relations between their covariance contribution and their statistical profile, which remains for future research. This chapter is a preliminary work that is yet unpublished.

Let us now introduce the terms we use and the frameworks we work with throughout this thesis. First, let us introduce some general background for the topics discussed in this thesis. In Section 1.2, we present basic concepts of the continuous auction and several important structures in electronic markets. In Section 1.3, we walk over

previous works about modeling price dynamics on the microscopic level. We also cover the basics of multivariate Hawkes processes. In Section 1.4, we present the basics of stochastic volatility and a brief history of the development of multifractal volatility and rough volatility models. In Section 1.5, we present the data we use for numerical studies. In Section 1.6, we present a selection of the main results of this thesis.

## 1.2 Basics of financial markets

### 1.2.1 Electronic Markets

A financial market is a place where assets are traded. The nature of assets traded on financial markets could be roughly categorized as fixed income, equity, or commodity. Assets traded on financial markets could often be derivatives, which are contracts that derive their value from the performance of underlying entities. Some common types of derivatives are futures, swaps, options, and exotic derivatives.

In this thesis, we focus on electronic markets. Electronic markets are virtual markets based on modern communications networks. Electronic markets make it possible for market participants in different locations to trade live in the same market. Orders sent to electronic markets are processed by high-speed computers. Typically market participants in the electronic market remain anonymous to each other.

In a specific electronic market, market participants announce the quantity and the desired price they intend to buy or sell. Their offers are known as **bid** offer and **ask** offer. A **trade** is settled by the two parties if the bid price of one market participant matches the ask price of another market participant.

If the ask price is lower than the bid price, the way to determine the effective traded price is specified by market rules, which may differ according to markets. Unless otherwise specified, the volume traded is usually the smaller amount in the bid offer and ask offer.

If there is no match between the bid price and the ask price, the maximum bid price is smaller than the minimum ask price, which is defined respectively as the **best bid** price and the **best ask** price. The difference between the best ask price and the best bid price is known as the **bid-ask spread**. In practice, we usually refer to the simple arithmetic average of the best bid price and the best ask price as the **mid-price**.

### 1.2.2 Limit order book

In an **electronic market**, market participants' actions are transmitted and processed electronically. We refer to the organizer of an electronic market as an **Exchange**, whose powerful infrastructure allows to process market participants' operations at a very high speed.

The order matching process in an electronic market is a continuous auction. In this process, new arriving orders compete with existing orders simultaneously. The overall

result of such competition will determine the execution price. The most recent execution price is known as **quote price**.

In practice, an **Exchange** keeps track of details of orders emitted into the market, which is known as the **Incoming queue**. Records in the incoming queue contain the timestamp and specification of orders, such as price, volume and other properties. Unless fully matched or canceled, incoming events are kept track in the queue for further processing.

An important structure is the **Limit order book (LOB)**. The limit order book contains all active orders that could be executed in the forthcoming auction. The limit order book is monitored and lively updated alongside new events, such as the arrival of new orders, modification or cancellation of existing orders, etc. LOB contains the information as the aggregated volume of all bid/ask offers presented at all possible prices, which is usually made visible to all market participants. LOB also keeps track of other kinds of information. For example, LOB records the priority of orders emitted by different market participants at a fixed price.

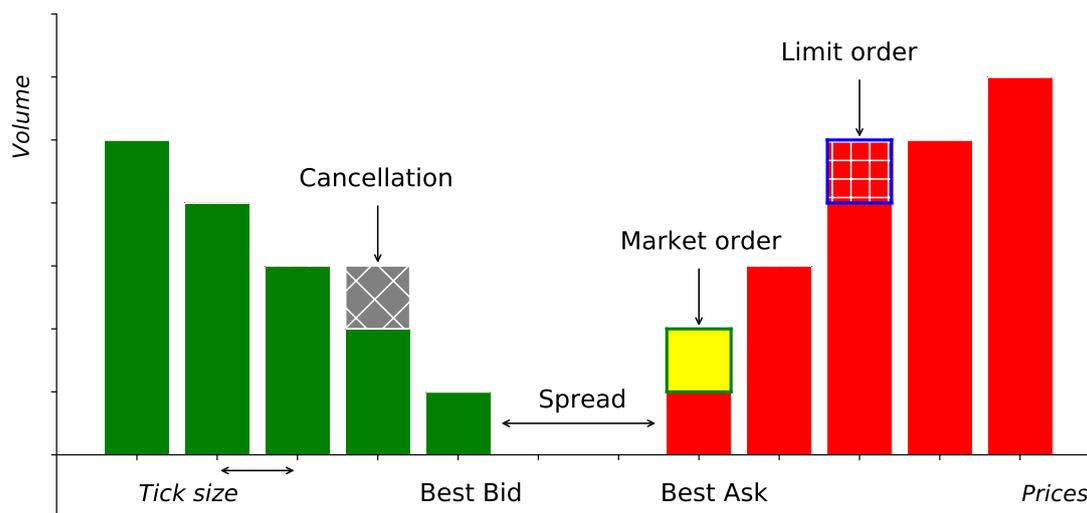


Figure 1.1: Illustration of a Limit order book

**Tick size** In an electronic market, the order sent to the LOB could only be put on a fixed grid of prices. This grid is usually uniform. The smallest difference between two consecutive prices is known as **tick size**.

Tick size plays an important role in the electronic market. On the one hand, since the bid-ask spread is at least one tick, an overlarge tick size will inhibit potential trades since the prices on which market participants can trade could largely deviate from their desired price.

On the other hand, an ultra-thin tick size is also harmful. To avoid a potential loss, market participants are obliged to frequently adjust the price of their limit orders with

the flow of new information. Such actions consume considerable electronic resources of the Exchange. It leads to an unstable limit order book whose state is rapidly changing.

In most markets, the tick size is chosen to adapt to the price of an asset. For example, stocks traded on the Euronext Exchange with prices smaller than 9.999 have a tick size of 0.001, while all stocks with prices above 10 have a tick size of 0.005. For stocks with prices between 50 and 99.99, the tick size is 0.01. Finally, for those above 100, the tick size is 0.05. Similar rules are also observed in other stock exchanges. Previous studies such as [49] have proposed different strategies to determine the optimal tick size.

### 1.2.3 Market participants

Depending on their motivation and the nature of different markets, market participants adopt various categories of strategies. The heterogeneity between agents is so widespread that a complete characterisation exceeds the scope of this study. Nevertheless, we refer readers to recent literatures as [96], [111], [1], [84], [31], [109] and [75] as detailed surveys. These studies proposed various statistical indicators to characterise market participants' behaviours in different aspects.

Analysis based on these features allows speculating on the nature of market participants. We roughly identify the following types of market participants according to their motivation and business focus :

- **Market makers** Market makers exhibit special importance in the electronic market. They place orders at both sides of the LOB with a spread, with the intention of making a profit by indirectly matching slower buyers and sellers asynchronously. Market makers provide precious liquidity in the market, which helps orders be matched in a shorter time and at better prices. For this reason, exchanges could sometimes reward market makers with a special liquidity provider scheme.

While holding positions, market makers are exposed to risks related to potential price changes. Price changes are constantly driven by external information that may move in a favorable or unfavorable direction. Hence, market makers must react agilely to avoid being taken advantage of by others. For this reason, market makers usually adopt high-frequency trading techniques. In fact, lots of high-frequency firms are market makers de facto.

Another striking feature of market makers is that they are not directional traders. Despite the fact that they place a massive volume of limit orders in the LOB, their inventory at the end of the day is usually flat. As they avoid holding excessive positions, they will change the prices of their offers and increase the bid-ask spread when the price change is not in their favor.

- **Hedge fund and proprietary trading firm** Hedge funds and proprietary trading firms are professional investors in the market. They look for arbitrage opportunities in the market to maximize the return of capital under their management.

Their strategies could be based on fundamental analysis or based on technical analysis as well.

Since the nature of these strategies could be different, there does not exist a typical profile of hedge funds and prop trading firms. As for the presence in the LOB and trading speed, they could possibly adopt high-frequency trading techniques but also focus on low- and mid-frequency strategies.

In the late decade, hedge funds exhibit increasing similarity with market makers on statistical features, as they start to focus on optimal execution and adopt “counter market-making” techniques, see [73] for a recent summary of optimal executing strategies.

One of the most distinguished characterisation of hedge funds and proprietary trading firms is that they hold positions for a relatively longer period than market makers, which could range from several weeks to decades.

- **Investment Banks** Banks are big players in the financial industry. They feature a significant amount of capital and derivatives in their inventory. Since the financial crisis of 2007-2008, increasing regulatory demands definitively reshaped banks’ business lines.

In the last decade, investment banks have been more considered financial service providers. They make a profit by charging margins in trades settled with other market participants. Banks offer a wide range of derivative products and usually play a counter-party role in these trades.

Today, banks are also engaged in market making and brokerage business. Banks also apply optimal execution and statistical arbitrage strategies as well, which has become an increasingly important business segment.

- **Brokers** Brokers are institutions that place orders on behalf of their clients, who usually don’t have direct access to financial markets. Their clients could be institutions that are not professionals in finance or individual investors. For example, a company may seek assistance from a broker to sell stocks in large quantities or exchange a large amount of money for another currency. Also, individual investors may use brokers to invest in stocks and other assets.

## 1.3 Modelling financial markets

### 1.3.1 Price modeling

Modeling the dynamics of financial markets, especially modeling the evolution of price, has always been a fundamental challenge for financial mathematics. It not only allows market participants to maximize their profits but also to mitigate risks of different natures effectively, which is pursued by regulators as well. Failing to identify and regulate

risky behaviors in the financial market would induce systematic risk, which leads to chaotic market behaviors and even financial crisis.

According to the scales at which price dynamics are observed and modeled, we roughly classify existing models into two categories:

**Macroscopic level models** In the past development of financial mathematics, theories based on Itô calculus were developed to model dynamic trading in continuous time. Derivatives are represented as expectations with respect to the risk-neutral probability measure of the derivatives discounted payoffs. This setting successfully enables the computation of no-arbitrage prices and hedging strategies.

In this approach, assets' prices are considered as continuous semi-martingales. They are driven by diffusion processes such as Brownian motion or geometric Brownian motion. Stochastic integration with respect to a Brownian motion and Itô calculus are developed to model dynamic trading in continuous time.

**Microscopic level models** In this thesis, we focus on the microscopic level models as they offer a more realistic description of the price formation mechanism. The microscopic level is the ultimate level where the price (usually taken as the last traded price or simply the midprice) is determined. Microscopic level models are not necessarily under the high-frequency context. In practice, the frequency at which the price dynamics are modeled/sampled could vary from a few microseconds to several minutes.

We mention several streams of literature for microscopic level models. The first stream of literature focuses on the overall statistical properties of LOBs, like the dynamic of spread and the distribution of lengths of intervals between consecutive executions. The earliest work of this stream could be tracked back to [37] and [50], with follow-up works as [132]. In the works mentioned above, simplified dynamics for the order flow are assumed to build mathematically tractable models that can reproduce, at least partially, some of these observed properties.

In the second stream of literature, models developed by the economics community focus on the behavior of rational agents that act strategically to optimize their utility function (see [125] for example).

As the last stream of literature, we especially mention the **Zero Intelligence** model firstly introduced in the pioneering work [132]. In this model, the order book is seen as a pure stochastic system with order flows governed by independent Poisson processes, indicating that market participants send orders without strategies. This is the origin of its name "Zero Intelligence".

Although the assumption that order flows are driven by Poisson processes is inconsistent with empirical observations, the zero-intelligence model quickly became an influential and successful model. In [56], the authors examined the zero intelligence model with empirical data and concluded that important statistical properties, such as the size of the bid-ask spread and even the market impact of orders are satisfactorily

reproduced. The zero intelligence model quickly became influential due to its simplicity. Relying only on some basic hypotheses, the zero intelligence model usually serves as a baseline model, which could be further improved to reproduce specific empirical phenomenons. It also inspires follow-up works that suppose more realistic dynamics of order flows. My thesis contributes to the latter category and builds itself on previous works.

This stream of literature features the framework where an asset's price is purely driven by limit order book events. These events could be, for example, matched trades between two market participants or events that emptied all volumes presented at the best bid/ best ask position. Since these events occur discretely over time, their dynamic is described by point processes. Prices are discrete values, and they change discretely over time. This is different from most macroscopic level models, which describe the price as a continuous diffusion process.

Under this framework, limit order book events are usually simplified into three types. Events that bring volume into all possible price levels are considered as often collectively classified as **Limit orders**. The partial or total removal of previously emitted limit orders is classified as **Cancellations** (short for cancellation orders). Events that immediately trigger trades between two market participants are named **Market orders**. Usually, market orders consume available volume at the best bid/ best ask prices. Under this setting, the price is determined by order flows, modeled as birth-death processes where limit orders bring new volumes, and cancellation/market orders consume existing volumes.

Continuing along with this stream of literature, the work [45] is one of the first papers to clearly state the problem of LOB modeling in the context of queuing theory and Markov chains. By leveraging the properties of Markov chains, the authors successfully derived conditional probabilities between events, such as the conditional probability that a market order arrives on the LOB within a specific time interval before the mid-price changes.

Recently, the authors of [4] kept the same assumption of independent queues driven by Poisson processes and proved that using the theory of infinitesimal generators and Lyapunov stability criteria, cancellations are crucial to ensure the stability of the LOB. They also show that under their model, the process of mid-price converges to a Wiener Process.

In order to take the strategies practiced by market participants into account, a possible improvement from the framework based on independent Poisson processes is to model the order flows as co-dependent point processes. We mention Hewlett in [85], and Large in [98] as pioneers who firstly describes the arrival of order flows by multivariate Hawkes processes. In the following section, we present the basics of the multivariate Hawkes process and discuss microscopic level modeling with the Hawkes process.

### 1.3.2 Hawkes process

The multivariate Hawkes process is a family of point processes firstly introduced in [80] and [81] by A.G Hawkes. The most distinguishing feature of the Hawkes process compared to the Poisson process is that the arrival rate of events depends explicitly on past events.

Hawkes processes offer a flexible but robust framework for modeling clustered events. It is widely used in various disciplines, such as seismology ([122]), Neuroscience ([93]), and machine learning ([110]). We refer readers to [16] as a general survey written by A.G Hawkes.

As summarized in a recent survey [82], Hawkes processes are also widely applied in quantitative finance. One of the major distinctions of the Hawkes process is that, as a self-exciting process, it successfully captures the empirical fact that current events (such as transactions) will trigger future events. Its self-exciting property is particularly pertinent in the context of modeling microscopic level events. In his book [133], Soros even argues that most orders on the market are merely responses to previous orders (See also [39]). Without being exhaustive, we list some recent literature: high-frequency events modeling ([11], [27]), analysis of market impact ([17], [18]), volatility modeling ([127]), modeling of correlated price changes ([38], [128]). For limit order book modeling we invite readers to Chapter 2 and Chapter 3 for more details.

**Definition** A multivariate Hawkes process of dimension  $d$  is a  $d$ -dimension counting process  $N(t) = (N^1(t), N^2(t), \dots, N^d(t))$ , for  $t \geq 0$ . Each component  $N^i$  accounts for the number of events of the  $i$ -th type during the interval  $[0, t]$ . We note by  $\mathcal{F}_t$  the filtration generated by the process  $N(t)$  and  $\mathbb{P}$  the associated probability measure.

Under reasonable assumptions (see Daley and Vera-Jones book [47]), the counting process  $N(t)$  admits a conditional intensity  $\lambda(t) = (\lambda^1(t), \lambda^2(t), \dots, \lambda^d(t))$ , for  $t \geq 0$ , which has the following interpretation:

$$\mathbb{P}(N^i \text{ jumped in the interval } [t, t + dt] | \mathcal{F}_t) = \lambda^i(t) dt, \quad \forall 1 \leq i \leq d. \quad (1.3.1)$$

According to [89], the quantity

$$N_i(t) - \int_0^t \lambda_i(s) ds$$

is a  $\mathcal{F}_t$ -martingale. The intensity function  $\lambda(t)$  is a combination of realized events, which takes the form

$$\lambda^i(t) = g_i\left(\sum_{j=1}^d \int \phi^{ij}(t-s) dN_s^j\right), \quad \forall 1 \leq i \leq d, \quad (1.3.2)$$

where the  $\phi^{ij}(\cdot)$  are causal functions, i.e.  $\text{supp } \phi^{ij}(\cdot) \subset \mathbb{R}^+$ . In most cases, the linear form  $g_i(x) = \mu^i + x$  is considered, with the restriction that  $\mu^i \geq 0$ . The intensity function  $\lambda(t)$  could be then written in the form:

$$\lambda^i(t) = \mu^i + \sum_{j=1}^d \int_{-\infty}^t \phi^{ij}(t-s) dN_s^j. \quad (1.3.3)$$

It is worth mentioning that the Poisson process is the degenerate case where all  $\phi^{ij} = 0$ .  $\phi^{ij}$  are mentioned as **self-exciting kernels**, or simply **kernels**. In the following, it is assumed that all kernels  $\phi^{ij}(t)$  satisfy

$$\phi^{ij}(s) \geq 0, \quad \forall t > 0. \quad (1.3.4)$$

In the rest of this study, we note  $\Phi(t)$  the matrix of kernel  $\phi^{ij}(t)$ . We note  $\Phi^{ij}$  the integral of  $\phi^{ij}(t)$ :

$$\Phi^{ij} = \int_0^{\infty} \phi^{ij}(s) ds, \quad (1.3.5)$$

and we note  $\Phi$  as the matrix of  $\Phi^{ij}$ . From [47] and [16], the multivariate Hawkes process  $N_t$  has asymptotically stationary increments and  $\lambda(t)$  is asymptotically stationary if  $\Phi$  satisfies the following condition:

**Condition 1.** (*Stationary condition*) : For the Hawkes process defined as above to be stationary, a sufficient condition is that the spectral radius of the matrix  $\Phi$  is smaller than 1.

In the following, we assume the condition 1 is always satisfied.

**Cluster representation** The Hawkes process defined above admits a clustering representation, which is also referred to as the branching representation. The Hawkes process in Eq. (1.3.3) describes the total population in the following immigration-birth process: the arrival of immigrants is governed by a Poisson process with intensity  $\mu$ . From its arrival, each individual (including immigrants and their descendants) starts a new generation: it gives birth to new individuals at the rate  $\Phi(\cdot)$ . All individuals share the same fertility function  $\Phi(\cdot)$ . We refer readers to [10] for more details about the Galton-Watson branching process.

As it describes the arrival of immigrants, the  $\mu$  in the intensity function Eq. (1.3.3) is often mentioned as the **exogenous intensity**. The term with kernel  $\phi^{ij}$  in Eq. (1.3.3) governs the new birth brought by existing individuals. For this reason, it is named as **endogenous intensity**.

A direct interpretation of  $\Phi^{ij}$  is the average number of events of type  $j$  that an event of type  $i$  is given birth to. For this reason, the matrix  $\Phi$  is referred to as the branching ratio matrix in some literature.

Let's continue with the interpretation of integrated cumulants of the multivariate Hawkes process. Consider the convolution of kernels:

$$R(t) = \sum_{n=0}^{\infty} \Phi^{*n}(t), \quad (1.3.6)$$

where the notation  $*n$  stands for  $\Phi * \Phi * \dots * \Phi$  with  $\Phi$  repeating for  $n$  times. The operator  $*$  stands for convolution between two matrices in the following way: for two square matrices  $A(t)$  and  $B(t)$  whose elements are functions,  $A * B(t)$  yields a square matrix whose  $i, j$ -th element is:

$$\int \sum_k A^{ik}(s) B^{kj}(t-s) ds. \quad (1.3.7)$$

By convention, for  $n = 0$ ,  $\Phi^0(t) = \mathbf{I}_d$ . In references [47] and [16] it is pointed out that the so-defined matrix  $\mathbf{R}(t)$  is the solution of the following equation:

$$\Phi(t) + \mathbf{R}(t) * \Phi(t) = \mathbf{R}(t). \quad (1.3.8)$$

Similar to the definition of  $\Phi^{ij}$ , if we define

$$\mathbf{R}^{ij} = \int_0^\infty \mathbf{R}^{ij}(s) ds \quad (1.3.9)$$

and note  $\mathbf{R}$  as the matrix of  $\mathbf{R}^{ij}$ , we have the following relation which links  $\Phi$  and  $\mathbf{R}$ :

$$\mathbf{R} = (\mathbf{I}_d - \Phi)^{-1}. \quad (1.3.10)$$

The  $\mathbf{R}^{ij}$  defined above is the average number of events of the  $i$ th component triggered (directly or indirectly) by an exogenous event of the  $j$ th component. The matrices  $\Phi$  and  $\mathbf{R}$  can be read as Granger causality relationships (see [72]) between different types of events. Such interpretation allows one to quantify the influence between entangled flows of events. We refer readers to [8] where a dedicated study about the interpretation of integrated cumulants of multivariate Hawkes process under the context of Granger causality is provided.

**Simulation and statistical inference of Hawkes process** Regarding the simulation and effective calibration of the Hawkes process, we mentioned the earliest works as [123] by Ogata and [124] by Osaki. In their works, the thinning algorithm is proposed to effectively simulate point processes. They also discussed the maximum likelihood estimator for some widely used kernels as exponential kernels and power-law kernels. For recent works on the effective simulation of Hawkes process, we mention [119], [120] and [68].

For non-parametric estimation, we mention [23] and [18] where Bacry et al. proposed a non-parametric estimation method for multivariate Hawkes processes based on second-order analysis. They are followed by [22] where the non-parametric Hawkes kernel estimation procedure is improved and carefully examined with empirical data set. We also give special credit to the open-source statistical referencing library *Tick*, see [21], which offers efficient implementation of the algorithms mentioned above.

### 1.3.3 The Hawkes model for order flows in LOB

Let's continue with the thread of modeling limit order book dynamics at the microscopic level. In the stream of literature following [132], the LOB is seen as a high-dimensional queuing system with orders arriving randomly, which focuses on the overall statistical properties of LOBs and assumes certain simplified dynamics for the order flow. We especially mention two works that inspired our work in Section 2.2. In [3], the authors drop the assumption of uncorrelated order flow and introduce a memory effect by choosing to model the rates of limit and market order arrivals (respectively  $\lambda^L$  and  $\lambda^M$ ) by a Hawkes process:

$$\lambda^\ell(t) = \mu^\ell + \sum_{m \in \{L, M\}} \int \phi^{\ell m}(t-s) dN_s^m, \text{ with } \ell \in \{L, M\}. \quad (1.3.11)$$

In [87], the authors focus instead on the influence of the current state of the LOB on trading decisions. They propose a simple Markov model, the so-called *Queue-Reactive* (QR) model, where the order flow arrival at a given price level is modeled as an inhomogeneous Poisson process with an intensity that depends only on the current state of the order book through the available volume :

$$\lambda^\ell(t) = \mu^\ell(q(t)), \quad (1.3.12)$$

for any  $\ell \in \{L, M, C\}$ .

Both works have established the conditions under which their model possesses ergodic properties, making it possible to reproduce the empirical LOB queue size distributions as the invariant distribution of a Markov process. More recently, [102] extends the model of [87] by allowing the order book dynamics to depend also on the type of the order that led to a complete depletion of a level (i.e., a market or cancel order) and also by taking into account the order size. Thus [102] departs slightly from the pure Markovian framework. Optimal market-making strategies are also discussed, and performance is assessed on real data.

We mention [114] as an interesting reference as well. This work also describes the dynamics of LOB by a state-dependent Hawkes process. However, the state of LOB is determined differently from Chapter 2 and Chapter 3. For other recent works about LOB modeling via Hawkes process, we refer readers to [4], [136] and [5].

## 1.4 Rough and multifractal volatility

In this section, we start by recalling basic definitions and properties of stochastic volatility. Then we enter into the domain of rough volatility and notably introduce two influential models, namely the Multifractal Random Walk model (MRW in short) and the Rough Fractional Stochastic Volatility model (RFSV in short). The  $H \rightarrow 0$  limit of the RFSV model is carefully examined in this thesis. Under this context, we discuss

some inspiring similarities between rough and multifractal volatility models under this context. We also discuss the parallel topic of the reliability of parameter estimation in rough volatility models. In the end, we refresh readers with the basics of the generalized method of moments (GMM) estimation.

**Notations for volatility process** Since the epoch-making work [35] by Black and [88] by Hull and White, volatility modeling plays a vital role in risk management and derivative pricing. Let the stochastic process  $P(t)$  stand for an asset's price, and  $X(t) = \log P(t)$  stands for the stochastic process of log price. From the moment  $t$  to  $t + \tau$ , the difference of the log-price  $X(t)$  approximates the return of asset:

$$\delta_\tau X(t) = X(t + \tau) - X(t) \approx \frac{\delta_\tau P(t)}{P(t)}. \quad (1.4.13)$$

Although it is known since Mandelbrot and Fama [105, 55] that returns exhibit excess kurtosis, returns are generally considered as *i.i.d* gaussian distributions, see [9, 78].

We define the squared volatility of price  $P(t)$  over the interval  $[t, t + \tau]$  as the quadratic variation of the log-price  $X(t)$ , which is noted  $\sigma_\tau^2(t)$  :

$$\sigma_\tau^2(t) = [\delta_\tau X(t)]. \quad (1.4.14)$$

Under certain circumstances, we assume that there exists a random measure  $M(\cdot)$  and  $\sigma_\tau^2(t)$  is simply the length of interval  $[t, t + \tau]$  under  $M(\cdot)$ :

$$\sigma_\tau^2(t) = M[t, t + \tau]. \quad (1.4.15)$$

If  $M(\cdot)$  further admits a density process denoted as  $\sigma^2(s)$ , we can further write.

$$\sigma_\tau^2(t) = \int_t^{t+\tau} dM(s) = \int_t^{t+\tau} \sigma^2(s) ds. \quad (1.4.16)$$

However, we emphasize that not every random measure  $M(\cdot)$  admits a density process, such as the case with MRM. Under this context, from the stationarity of the return process, we retrieve

$$\mathbb{E}[\sigma_\tau^2(t)] = Ct. \quad (1.4.17)$$

The equation above is in line with the famous *squarerootoftime* rule, which states that when high-frequency volatility is scaled to a lower frequency, it should be multiplied by the square root of the ratio between time scales. Precisely, we have

$$\mathbb{E}[\sigma_T^2(t)] = \frac{T}{\tau} \mathbb{E}[\sigma_\tau^2(t)]. \quad (1.4.18)$$

### 1.4.1 Multifractal volatility

**Multifractality** The term fractal was first introduced by Benoit Mandelbrot in 1974, see his famous book [106]. The application of multifractality in financial mathematics started with work [118], [107] and [54]. Since then, the volatility process has been known to be much less regular than a Brownian motion. Multifractal models have been designed in order to reproduce this irregular behavior along with the scale-invariant properties, see [41] by Calvet, [26] by Barral, and [20] by Bacry et al. Generally speaking, a continuous multifractal function  $f : \mathbb{R} \rightarrow \mathbb{R}$  is defined by the fact that around any point  $x$ , the local increment  $f(x+h) - f(x)$  is described by a local power law:

$$f(x+\delta) - f(x) \sim \delta^{h(x)}, \quad (1.4.19)$$

where the exponent  $h(x) > 0$  is called the **singularity exponent** (or the Hölder exponent), as it describes the local degree of singularity around the point  $x$ . For  $h \in \mathbb{R}^+ \cup \{+\infty\}$ , we note

$$E_f(h) := \{x | h(x) = h\} \quad (1.4.20)$$

the set of points with the same singularity exponent  $h$ . The **singularity spectrum** of  $f$ , noted  $D(h)$ , is the mapping from  $h$  to the Hausdorff dimension of the set  $E_f(h)$ .

Often, instead of directly characterizing the singularity spectrum, the multifractal behavior is introduced by the following definition :

**Definition 1.4.1.** A process  $X(t)$  with stationary increments is scale-invariant if  $\delta_\tau X(t) = X(t+\tau) - X(t)$  follows a power law:

$$\mathbb{E}[|\delta_\tau X(t)|^q] \sim C_q \tau^{\zeta(q)}, \text{ when } \tau \rightarrow 0. \quad (1.4.21)$$

We further distinguish the process  $X(t)$  as monofractal or multifractal according to the form of the multifractal exponent  $\zeta(q)$ . The process  $X(t)$  is **monofractal** if  $\zeta(q)$  is a linear function of  $q$ . If  $\zeta(q)$  is a nonlinear function of  $q$ , the process  $X(t)$  is called **multifractal**.

The so-called multifractal formalism (see [117], [19], [90] and [91]) established an exact correspondence between the singularity spectrum  $D(h)$  and the multifractal exponents  $\zeta(q)$  through a Legendre transform under reasonable assumptions. Such results paved the way for applying multifractal processes for volatility modeling since in practice,  $\zeta(q)$  is much easier to measure.

We mention the fractional Brownian motion as an example of monofractal processes with  $\zeta(q) = qH$ . We also refer readers to dedicated studies of multifractal processes as [97], which further proved the concavity of the multifractal exponent :

**Proposition 1.** *If the scaling law Eq. (1.4.21) holds true when  $\tau \rightarrow 0$ , i.e.*

$$\mathbb{E}[|\delta_\tau X(t)|^q] \underset{\tau \rightarrow 0}{\sim} C_q \tau^{\zeta(q)}, \quad (1.4.22)$$

then the multifractal scaling component  $\zeta(q)$  is a concave function of  $q$ .

**Multifractal random walk** The “Multifractal Random Walk” (MRW) was first introduced by Bacry et al. in [118] and [12] as a model for asset prices in order to account for their multifractal properties, i.e., the fact observed by various authors (see [67, 66, 104]) that empirical moments of asset return obey non-trivial scaling properties. The MRW model relies on a multifractal stochastic volatility model, namely the “Multifractal Random Measure” (MRM) model (see [118, 12]), in which the log-volatility is provided by a log-correlated Gaussian field. Such a class of processes, also referred to as Gaussian multiplicative chaos, has been at the heart of many studies in a large variety of applications [129]. In [115, 20], Gaussian multiplicative chaos and the associated log-normal random cascades have been extended to any infinitely divisible distribution by Bacry and Muzy. In the MRW model, the log price  $X_T(t)$  is :

$$X_T(t) = B(\widetilde{M}_T[0, t]), \quad (1.4.23)$$

where  $B_t$  is a standard Brownian motion and  $\widetilde{M}_T$  is a multifractal random measure (MRM) obtained as the following weak limit :

$$\widetilde{M}_{\ell, T}(dt) \xrightarrow[\ell \rightarrow 0]{w} \widetilde{M}_T(dt), \quad (1.4.24)$$

where  $\widetilde{M}_{\ell, T}(dt)$  is defined by

$$\widetilde{M}_{\ell, T}(dt) = e^{\omega_{\ell, T}(t)} dt \quad (1.4.25)$$

and the notation  $\xrightarrow{w}$  stands for the weak convergence. The process  $\omega_{\ell, T}(t)$  is Gaussian and stationary with a logarithmic covariance vanishing for lags greater than  $T$ , see Eq. (4.7.75).

Since such a logarithmic decreasing function can be interpreted using random multiplicative cascades in the limit when the preferred scale ratio goes to 1, one often refers to such a model as continuous cascade [14] models. In [115, 20], MRM measures have been extended from log-normal statistics to any log-infinitely divisible law which satisfies :

$$\mathbb{E}[|\delta_\tau M_T(t)|^q] = \mathbb{E}[|M_T(t + \tau) - M_T(t)|^q] = C_q \tau^{\zeta(q)}, \quad \forall t < T. \quad (1.4.26)$$

The MRW model serves as a concise yet powerful stochastic volatility model where the log volatility is provided by a log-correlated Gaussian field, which is also referred to as Gaussian multiplicative chaos. The Gaussian multiplicative chaos has been at the heart of many studies in a large variety of applications, see [129]. The model presented in Chapter 4 could be regarded as a specific construction of such log-normal infinitely divisible distribution.

### 1.4.2 Rough volatility

The pioneering work [65] by Gatheral et al. introduced a new class of model (rough volatility model) in which the log volatility is modeled by a fractional Brownian motion with a small Hurst parameter  $H$ , meaning that the log volatility is less regular than a standard Brownian motion (see also [32, 33]). It is worth mentioning that the multifractal volatility models (including the MRM model) involve behaviors of the log volatility that are less regular than any fractional Brownian motion. This point will be further illustrated in Chapter 4.

The empirical evidence reported in [65] (followed-up by later studies such as [101] and [62]) suggests that the logarithm of the asset price stochastic variance can be represented by a fractional Brownian motion (fBM) of Hurst exponent  $H$  close to  $H \simeq 0.1 < 1/2$ . More recent studies based either on quasi-likelihood approach [62] or GMM-approach [36], consistently suggest that  $H$  can be close to 0, i.e.,  $H \lesssim 0.05$  for a large panel of equity data.

Rough volatility has quickly become a trending topic in the community of quantitative finance. New profound insights on stochastic volatility models have been obtained in following up studies. We especially mention [61], which argues that volatility has to be rough to match the power-law form of volatility skew observed in empirical data. More recent work like [94] further justifies the roughness of volatility as a corollary of the non-arbitrage principle.

The popularity of rough volatility models is not limited to successfully capturing the empirical properties of realized volatility. When roughness in volatility is incorporated into option pricing, a very good fit of option prices, and notably, the power-law behavior of ATM skew close to maturity is obtained, see [101, 29, 61]. Here we give more references about option pricing with rough volatility. In [86], a toy model is proposed to price option while supposing the underlying volatility process is rough. Following-up studies for option pricing under rough volatility are covered in a general survey [95].

Recently, rough volatility has been incorporated into term structure models. This is an exciting advance because term structure models, such as the Heston model and Bergomi model, are used as industry standards for pricing Fixed Income Currencies Commodities (FICC) products. In [29], the rough Bergomi model is firstly introduced. In a series of recent works [51], [52] and [53] by El Euch et al., a rough version of the Heston model is introduced.

Literature about rough volatility is not limited to hedging and pricing purposes. In [69], a trading strategy based on the roughness of the price process is proposed. In [6], approximation problems of rough volatility models are firstly investigated. In [30] and [60], strict mathematical tools to analyze regularity structures of rough volatility models are introduced.

**RFSV model** The Rough fractional stochastic volatility(RFSV) model proposed in [65] by Gatheral et al. assumes that the volatility  $\sigma_\tau^2(t)$  on a given interval  $[t, t + \tau]$  is provided by integrating a stationary log-normal process:

$$\sigma_\tau^2(t) = \int_t^{t+\tau} v(s) ds = \int_t^{t+\tau} e^{o_{H,T}(s)} ds, \quad (1.4.27)$$

where  $o_{H,T}(t)$  is a fractional Ornstein-Uhlenbeck (fOU) process that satisfies, for some  $0 < H < 1/2$ , the equation

$$do_{H,T}(t) = \nu dB_t^H - \alpha(o_{H,T}(t) - m) dt. \quad (1.4.28)$$

In the equation above,  $B^H$  is a fractional Brownian motion with Hurst parameter  $H$ . The parameter  $T := \alpha^{-1}$  accounts for the time scale of mean-reversion. In [65], it is argued that when  $T \rightarrow \infty$ ,  $o_{H,T}(t)$  behaves locally as a fractional Brownian motion  $B_H(t)$  in the following sense:

$$\mathbb{E}[\sup_{t \in [0, t_0]} |o_{H,T}(t) - o_{H,T}(0) - \nu B_t^H|] \rightarrow 0, \forall t_0 > 0. \quad (1.4.29)$$

For  $\tau > 0$  small enough, the authors calculate covariance function of  $o_{H,T}(t)$  from the covariance of  $B_H(t)$ :

$$\text{Cov}[o_{H,T}(t), o_{H,T}(t + \tau)] \simeq \frac{\nu^2}{2} \left( T^{2H} \Gamma(2H + 1) - \tau^{2H} \right), \quad (1.4.30)$$

where  $\Gamma(\cdot)$  is the Gamma function.

### 1.4.3 The $H \rightarrow 0$ limit in rough volatility models

Since [65], more recent studies based either on quasi-likelihood approach ([62]) or GMM-approach ([36]) consistently suggest that the Hurst parameter  $H$  is close to 0, with  $H \lesssim 0.05$  measured for a large panel of equity data. Based on empirical evidence, it is natural to consider the limit  $H \rightarrow 0$  in the rough process driving the volatility logarithm. Even if one cannot plug  $H = 0$  in the power-law expression of the fractional Brownian motion covariance, formally, it corresponds to a logarithmic behavior.

Such behavior is precisely the one that characterizes the so-called continuous random cascade models introduced two decades ago by Bacry et al. in [118, 12] and presented in Section 1.4.1. Recovering a multifractal volatility model as the limit  $H \rightarrow 0$  of a rough volatility model or, from a more general perspective, defining a meaningful limit  $H \rightarrow 0$  of a fractional Brownian motion and one of its variants has been the subject of various recent studies. In [63], the authors build a  $H = 0$  - fBM by considering a regularisation from the harmonizable representation of fBM's while in [121, 74] a  $H = 0$  limiting process is obtained using a peculiar normalisation and centering of the fBM. In [59] (see also [58]), the authors consider the limit  $H \rightarrow 0$  of the exponential of a rescaled Riemann-Liouville fBM and its relationship with Gaussian multiplicative chaos. Finally,

in [28], Bayer et al. propose a new class of rough models that consists in modulating the Riemann-Liouville fBM power-law kernel by a logarithmic factor. The so-obtained “super-rough” stochastic volatility remains well-defined as a continuous process when  $H = 0$ .

In this thesis, our goal is to introduce a new version of rough volatility models. The Hurst parameter  $H$  is explicitly specified in this new model. Meanwhile, the new model should be a stationary replicant of the RFSV model, which is, de facto, driven by non-stationary fBM processes. This indicates that the covariance function in the new model should be exactly the one obtained when considering the small-time approximation of the correlation of the fractional Ornstein-Uhlenbeck process considered in [65]. The question we want to answer is, in the  $H \rightarrow 0$  limit, can we extract the exact self-similar multifractal measure defined in [118, 12] from the new model?

#### 1.4.4 Generalized method of moments (GMM)

In the context of multifractal models, the parameters are related to the moments of the increments of the volatility or its logarithm. It’s therefore reasonable to consider the **Generalized Method of Moment (GMM)** for estimating the parameters.

The GMM was firstly proposed by Hansen in [76], but the basic idea could be traced back to Sargan’s work [131]. For readers’ reference, here we mention [42, 43] by Calvet and Fisher, and [15] by Bacry et al. as applications of the GMM in multifractal models.

Consider a multi-dimension stochastic process  $Y_t$  which is defined by the parameters set  $\Theta^*$  of dimension  $p$ . In most cases, the process  $Y_t$  is supposed to be a weakly stationary ergodic stochastic process.

Let us now consider a sample  $Y_t$  composed of  $n$  observations. To estimate the parameter set  $\Theta$ , GMM consists of choosing a vector-valued function  $g(Y, \Theta)$  of dimension  $q > p$  therefore the following equation remains true only with the authentic parameter  $\Theta^*$  :

$$m(\Theta_0) := \mathbb{E}[g(Y_t, \Theta^*)] = 0. \quad (1.4.31)$$

Meanwhile, for all other parameters set  $\Theta \neq \Theta^*$ ,  $m(\Theta)$  should be non-trivial. Eq. (1.4.31) is known as the **moment condition**. In practice,  $m(\Theta_0)$  is approximated by its empirical counterpart:

$$g_N(\Theta) := \frac{1}{n} \sum_{k=1}^n g(Y_k, \Theta). \quad (1.4.32)$$

**Definition 1.4.2.** The GMM estimator  $\hat{\Theta}$  of  $\Theta^*$  is defined as

$$\hat{\Theta} = \arg \min_{\Theta} (g_N^T W_N g_N), \quad (1.4.33)$$

where  $W_N$  is a series of weighted matrices which converge to  $W_\infty$  when  $N \rightarrow +\infty$ .

The validity of GMM is based on the following theorem:

**Theorem 1.4.3.** (Hansen) *If*

(i) *The process  $Y_t$  is ergodic*

(ii) *The series  $\{f(Y_{\Theta^*}[k], \theta)\}_k$  satisfies the central limit theorem, i.e.:*

$$\frac{1}{\sqrt{N}} \sum_{k=1}^N f(Y_{\Theta^*}[k], \theta) \rightarrow \mathcal{N}(0, V_{\Theta}) \quad (1.4.34)$$

where the matrix  $V_{\Theta}$  is defined by

$$V_{\Theta} = \lim_{k \rightarrow +\infty} \sum_{k=-n}^n \mathbb{E}[f(Y_{\Theta^*}[k], \Theta) f(Y_{\Theta^*}[k], \Theta)^{\top}]. \quad (1.4.35)$$

(iii) *The matrix  $D_{g_N} = \frac{\partial g_N}{\partial \Theta}$  (which is of dimension  $r \times p$ ) is of rank  $p$ . When  $N \rightarrow +\infty$ , it converges to*

$$Df = \mathbb{E}\left[\frac{\partial g(Y_{\Theta^*}[k], \Theta)}{\partial \Theta}\right]. \quad (1.4.36)$$

Then, the GMM estimator  $\hat{\Theta}$  is consistent and

$$\sqrt{N}(\hat{\Theta} - \Theta) \rightarrow \mathcal{N}(0, \Sigma) \quad (1.4.37)$$

where

$$\Sigma = (Df^{\top} W_{\infty} Df)^{-1} Df^{\top} W_{\infty} V_{\Theta^*} W_{\infty} Df (Df^{\top} W_{\infty} Df)^{-1}. \quad (1.4.38)$$

We recommend readers to refer to [76] for the complete proof, which also comes with the following proposition :

**Proposition 1.4.4.** *The GMM estimator is optimal if  $W_{\infty}$  equals to  $V_{\Theta^*}^{-1}$ . In this case, the covariance matrix of the GMM estimator is*

$$\Sigma_{opt} = (Df^{\top} V_{\Theta^*}^{-1} Df)^{-1}. \quad (1.4.39)$$

In practice, since the true parameter set  $\Theta_0$  is unknown,  $W_{\infty}$  couldn't be chosen as  $V_{\Theta^*}^{-1}$  a priori. One common solution is the so-called **iterated GMM**, which consists of the following recursive procedure :

1. Initialise the matrix  $W_N$  with arbitrary values, such as  $W_N = \mathbb{I}_d$  or any *a priori* estimation of  $\Theta^*$ .
2. Calculate  $\hat{\Theta}$  with the current  $W_N$  as per Eq. (1.4.33).
3. Update  $W_N = V_{\hat{\Theta}}^{-1}$  with the estimator  $\hat{\Theta}$  obtained in the previous step.
4. Repeat steps 2 and 3 until the series of  $\hat{\Theta}$  has sufficiently converged.

## 1.5 Data presentation

The reliability of conclusions established in this thesis is supported by empirical results obtained with various data sets, which record price variations and different types of limit order book events during the double auction process on electronic markets. Since these data sets differ in assets and specifications of information recorded, we provide a separate description for each of them.

### Bund future and DAX future

The DAX (Deutscher Aktienindex as the German stock index) is a stock index consisting of the 30 major German companies trading on the Frankfurt Stock Exchange. The Bund is the sovereign debt instrument issued by Germany's federal government to finance outgoing expenditures. In our studies, we exploit futures on the DAX index (DAX in the following) and futures on German government debt: 10-years Euro-Bund (Bund in the following).

In finance, a futures contract is a standardized legal agreement to trade an asset at a predetermined price at a specified future moment. Futures are traded on an exchange. Similar contracts traded over the counter (OTC) are called forwards, which also allow for private customizations.

Unlike equities, futures contracts have expiration dates where physical delivery of the underlying is expected. For the same underlying, futures with different maturities (usually in different months) are traded over the market. In the context of futures, The month currently closest to delivery is called **nearby month**, which is also referred to as **front month** or **spot month**. A future for the nearby month is the shortest contract that an investor can purchase. Trading is usually most active in the nearby month. For this reason, the nearby month exhibits the most volatility.

At maturity, market participants need to reposition their portfolios to avoid delivery. This procedure is called roll-over. It consists of selling the contract that one currently holds to buy the same contract with deferred maturity (usually the next maturity). Due to liquidity reasons, the roll-over is performed on the last day or the last week of the month preceding the delivery. In our studies, we always refer to the future of the nearby month and the adjusted prices when rolling over takes place.

Our data set contains tick by tick level L1 data of Bund future and DAX index future traded on the Eurex electronic future market. The data was recorded from October 1st, 2013 to September 30th, 2014. The dataset consists of events happening at the first level of the order book, each with a timestamp indicating the record time with microsecond precision that provides prices and outstanding quantities. This data set is used in Chapter 2, Chapter 3 and Chapter 5.

### Stocks in the CAC40 index

The CAC40 (Cotation Assistée en Continu) is the principal stock index in France.

The index is composed of the 40 most significant stocks among the 100 largest market caps on the Euronext Paris (formerly the Paris Bourse), weighted by capital. It is one of the main European indices alongside Brussels' BEL20, London's FTSE100, and Amsterdam's AEX index.

The components of the CAC40 originate from companies in different sectors to reflect the global trend of the economy. Based on recent changes in the capital, the list of components stocks is regularly updated.

Our data set is provided by EURONEXT Exchange, which records limit order book events for all stocks in CAC40, spanning from April 2017 to March 2018. During this period, one stock is removed and then replaced by a new stock in the CAC40 index. This data set is the raw record of all events arriving on (at least) the first five best levels of the limit order book. Events' associated timestamps are recorded at microsecond precision.

This extremely rich data set makes it possible to calculate statistical profiles for market participants, such as changes in inventories and presence in the best positions of the LOB. It is used in Chapter 5.

### **Oxford-Man realized volatility database**

The Oxford-Man realized library contains daily close-to-close returns together with daily non-parametric measures of realized volatility for a variety of assets. Readers can refer to [83] and [25] for detailed descriptions of their data cleaning, non-parametric volatility calculation methodology, together with the complete list of assets covered in this data set.

The Oxford-Man realized library is of particular interest not only because of its comprehensive coverage of a wide range of assets. It is also the most common reference data set in the community rough volatility. In Chapter 4, the Oxford-Man realized library is exploited as a part of empirical studies.

### ***Yahoo Finance***

Yahoo Finance<sup>1</sup> provides financial news and historical data, including stock quotes and financial reports. In Chapter 5, the historical daily open, highest, lowest and close price time-series of 296 individual stocks are collected from Yahoo Finance. Stocks were taken from either the S&P 500 index (historical data from 1985-01-01 to 2021-12-31) or the CAC 40 index from 2000-01-01 to 2021-12-31.

## **1.6 Résultats de la thèse**

Dans cette section, les résultats obtenus dans cette thèse sont résumés à un haut niveau avec des illustrations, des conclusions et des remarques sélectionnées de chaque travail.

---

<sup>1</sup><http://YahooFinance.com>

Dans le chapitre 2, nous introduisons le modèle SQRH (Single Queue Reactive Hawkes). Nos résultats suggèrent que pour construire un modèle de flux d'ordres fiable, il faut tenir compte à la fois de l'état actuel de la LOB, de l'effet d'auto-excitation et de l'excitation croisée de flux de ordres précédents.

Dans le chapitre 3, nous étendons le modèle SQRH introduit dans le chapitre 2 à la fois à la modélisation de la meilleure demande et de la meilleure offre de la LOB. Ce modèle nous permet de révéler la dynamique conjointe de flux d'ordres des deux côtés de la LOB. Nous constatons notamment que notre variable d'état ainsi définie (Voir le chapitre 3 pour plus de détails) est liée de manière unique au déséquilibre du volume.

Dans le chapitre 4, nous introduisons une famille de mesures aléatoires log-normales paramétrées et nous construisons le modèle S-fBM. Ce modèle nous permet de considérer les deux classes populaires de modèles de volatilité multifractale et rugueuse dans le même cadre. Nous montrons notamment que l'estimation directe du paramètre de Hurst  $H$  à partir des propriétés d'échelle peut conduire à une forte surestimation. Nous proposons une méthode d'estimation basée sur le GMM qui, lorsqu'elle est appliquée à un large ensemble de données empiriques de volatilité, conduit à des valeurs de  $H$  très proches de 0 pour les prix des actions.

Dans le chapitre 5, nous introduisons un modèle multivarié basé sur le processus de Hawkes pour extraire la covariance du prix (en fonction du rendement) des événements du carnet d'ordres à cours limité. En introduisant des relations approximatives au premier ordre de la matrice de la norme du noyau, nous quantifions la contribution de chaque participant du marché à la variance du prix. Les résultats empiriques suggèrent que la contribution des participants du marché à la covariance est uniforme entre les différents marchés et étroitement liée à leur profil statistique à haute fréquence.

### 1.6.1 Modèle de Hawkes réactif à file d'attente pour le flux d'ordres

Le premier projet de ma thèse est de construire un modèle fiable pour décrire les flux d'ordres arrivant dans la LOB. Avant de commencer, nous citons d'abord un modèle appelé *Queue Reactive* (QR) proposé dans [87], où le taux d'arrivée de flux d'ordres est modélisé comme un processus de Poisson inhomogène, dont l'intensité dépend de l'état actuel du LOB.

**Modèle réactif de file d'attente avec processus de Hawkes** La notion de réactivité de la file d'attente a été introduite pour la première fois par Huang, Lehalle et Rosenbaum dans [87]. Comme mentionné ci-dessus, ils ont proposé dans cette étude un

modèle où l'arrivée du flux d'ordres à un niveau de prix donné est modélisée comme un processus de Poisson inhomogène, avec une intensité qui ne dépend que de l'état actuel du carnet d'ordres, c'est-à-dire de la taille des files d'attente. Ils nomment cette propriété *Queue Reactive* (QR).

Nous ajoutons un terme de Hawkes au modèle SQR, afin que le nouveau modèle tienne compte non seulement des dépendances de la taille de la file d'attente mais aussi des effets de mémoire dans les flux d'ordres. Pour cette raison, nous appelons ce nouveau modèle le modèle SQRH (Single Queue Reactive Hawkes). Formellement, le modèle SQRH associe le modèle SQR à un processus de Hawkes multivarié.

Dans la suite,  $N_t^L$ ,  $N_t^C$ ,  $N_t^M$ ,  $\lambda^L(t)$ ,  $\lambda^C(t)$  et  $\lambda^M(t)$  désigneront les processus de comptage et leurs intensités associées. Ils représentent respectivement les arrivées d'ordres limite, d'ordres annulation et d'ordres au marché à la meilleure file d'attente. Pour  $\ell \in \{L, M, C\}$  (respectivement pour les ordres de limite, de marché et d'annulation), le modèle SQRH définit  $\lambda^\ell(t)$  comme :

$$\lambda^\ell(t) = \left( \mu^\ell(q(t^-)) + \sum_{m \in \{L, M, C\}} \int_0^t \phi^{\ell m}(t-s) dN_s^m \right) \mathbb{1}_{(\ell=L) \vee (q(t^-) > 0)} \quad (SQRH), \quad (1.6.40)$$

avec chaque fonction noyau de forme somme-exponentielle,  $\phi^{\ell m}(t) = \alpha^{\ell m} e^{-\beta t}$ , et la taille de la file d'attente  $q(t)$  est simplement donnée par  $q(t) = q(0) + N_t^L - N_t^M - N_t^C$  et le facteur  $\mathbb{1}_{(\ell=L) \vee (q(t^-) > 0)}$  garantit qu'aucun ordre de marché ou d'annulation ne peut se produire si la file d'attente est vide. Les intensités de base  $\{\mu^\ell(q)\}$  dépendent de l'état de la file d'attente  $q$  tandis que les noyaux de Hawkes  $\phi^{\ell m}(t)$  tiennent compte de l'effet des ordres passés de type  $m$  sur l'intensité actuelle  $\lambda^\ell(t)$ .

### Propriété de Markov dans le modèle SQRH

Nous prouvons la propriété de Markov pour le modèle SQRH en considérant un cadre simplifié, où tous les flux d'ordres entrants ont un volume identique. Sous cette hypothèse, exprimons d'abord la taille de la file d'attente  $q(t)$  comme la somme des processus de comptage  $N(\cdot)$  :

$$q(t) = N_t^L - \sum_{\ell \neq L} N_t^\ell. \quad (1.6.41)$$

Considérons ensuite la quantité

$$o_{\ell mu}(t) := \int_0^t \alpha_u^{\ell m} e^{-\beta u(t-s)} dN_s^m, \quad (1.6.42)$$

et en désignant par  $\vec{o}(t)$  le vecteur obtenu par empilement vertical de  $o_{\ell mu}(t)$  ( $u \in \{1, \dots, U\}$ ,  $\ell, m \in \{L, M, C\}$ ),  $(\frac{q(t)}{\vec{o}(t)})$  est un processus de Markov vectoriel. De plus, si on suppose qu'il existe deux constantes positives  $c^-$  et  $c^+$  telles que

$$\sum_{\ell \neq L} \mu^\ell(q) \geq c^- q; \text{ and } ; \mu^L(q) \leq c^+ ; ,$$

Bund				
	$L$	AIC	BIC	# parameters
SQR	$2.046 \times 10^7$	$-4.093 \times 10^7$	$-4.092 \times 10^7$	450
SQRH	$2.055 \times 10^8$	$-4.110 \times 10^8$	$-4.110 \times 10^8$	477
Hawkes	$2.052 \times 10^8$	$-4.105 \times 10^8$	$-4.105 \times 10^8$	30
DAX				
	$L$	AIC	BIC	# parameters
SQR	$7.268 \times 10^5$	$-1.453 \times 10^6$	$-1.452 \times 10^6$	75
SQRH	$9.506 \times 10^6$	$-1.901 \times 10^7$	$-1.901 \times 10^7$	102
Hawkes	$9.386 \times 10^6$	$-1.877 \times 10^7$	$-1.877 \times 10^7$	30

Table 1.1: Valeurs d’log-vraisemblance, AIC et BIC pour les trois modèles considérés pour les données Bund et DAX.

nous montrons, en annexe 2.5, à l’aide de l’approche des fonctions de Lyapunov selon la même ligne que dans [4, 3], que le processus  $(\frac{q(t)}{\sigma(t)})$  est uniformément ergodique ce qui signifie notamment que  $q(t)$  admet une distribution invariante et que cet équilibre est atteint exponentiellement vite.

**La qualité de l’ajustement** En termes de qualité d’ajustement, nous comparons le modèle SQRH avec le modèle SQR par le critère de vraisemblance. Nous ajoutons le modèle standard de Hawkes dans la comparaison. Pour tenir compte de la différence dans le nombre de paramètres, le critère d’information d’Akaike (AIC) et le score du critère d’information de Schwartz (BIC) sont également rapportés dans le tableau 1.1. Notre résultat suggère que le modèle SQRH surpasse les modèles concurrents en termes d’AIC et de BIC.

**La distribution invariante** Nous donnons d’abord dans la section 2.2.2 un argument empirique selon lequel, la distribution invariante est atteinte avant que le meilleur prix d’achat/de vente ne change pour l’actif étudié (Bund future et DAX future).

La distribution empirique de la taille de la file d’attente est mesurée en prenant des instantanés du carnet toutes les 30 secondes. Pour le modèle SQRH, nous estimons la distribution invariante de  $q(t)$  en effectuant une simulation sur une longue période. La mesure invariante du modèle SQR peut être directement déduite des estimations de  $\mu^\ell(q)$ .

La distribution invariante pour les futures Bund et DAX est présentée dans la figure 1.2. Nous observons que le modèle SQRH fournit, dans les deux cas, le meilleur ajustement des données empiriques, notamment dans la région de la queue. Nos résultats montrent que la prise en compte de l’auto-interaction par le processus de Hawkes

dans un modèle réactif de file d'attente est importante non seulement pour décrire correctement la dynamique du flux d'ordres mais aussi pour fournir un meilleur modèle pour les distributions de la taille des files d'attente.

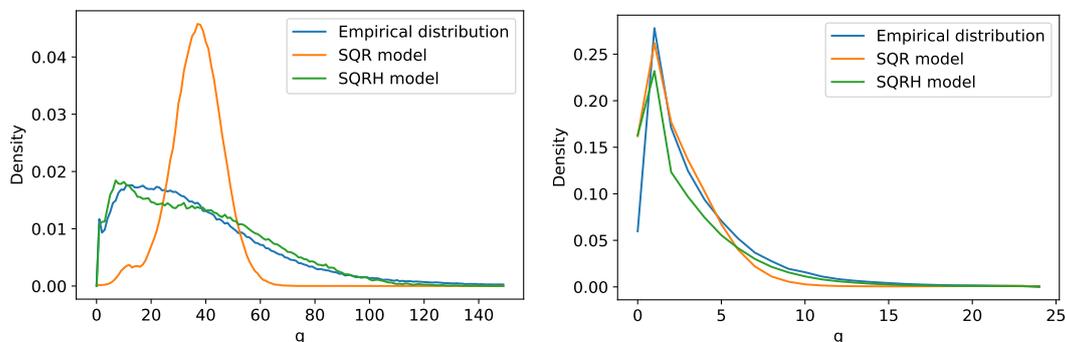


Figure 1.2: Comparaison des distributions invariantes par le modèle QR et le modèle QRH avec la distribution empirique. Futur du Bund à gauche, Futur du DAX à droite.

**La dépendance à l'égard des États** Dans la Figure 2.3 nous présentons les paramètres estimés  $\mu(q)$  pour le modèle SQR, tandis que dans la Figure 2.4 nous présentons les quantités analogues pour le modèle SQRH 1.6.40.

Nous faisons deux remarques générales en comparant ces graphiques. Tout d'abord, nous constatons que la dépendance à la taille de la file d'attente saisie par les deux modèles est à peu près concordante, dans la mesure où les fonctions  $\mu^\ell(q)$  ont des formes similaires dans les deux modèles. Nous remarquons également plusieurs différences entre les résultats du Bund et du DAX, très probablement dues à la dynamique différente du carnet d'ordres des actifs à grand et petit tick.

Comme dans [87], nous observons un taux décroissant d'arrivées d'ordres de marché à mesure que la taille de la file d'attente augmente. Cela peut s'expliquer par le fait que les agents ont tendance à consommer la liquidité plus rapidement lorsque celle-ci devient rare.

Nous constatons également que lorsque  $q(t)$  est suffisamment grand, l'intensité de l'annulation est une fonction croissante de la taille de la file d'attente. Il s'agit d'une caractéristique attendue et supposée dans la plupart des anciens modèles LOB (voir par exemple, [132, 45]) puisque les annulations sont plus susceptibles de se produire lorsqu'il y a beaucoup d'ordres à cours limité actifs.

Enfin, contrairement au comportement observé dans [87] sur des actions spécifiques, nous n'observons pas que l'intensité de l'insertion des ordres à cours limité est indépendante de la taille de la file d'attente. Elle est plutôt une fonction décroissante de la taille de la file d'attente, reflétant probablement une moindre recherche de priorité lorsque  $q$  est grand.

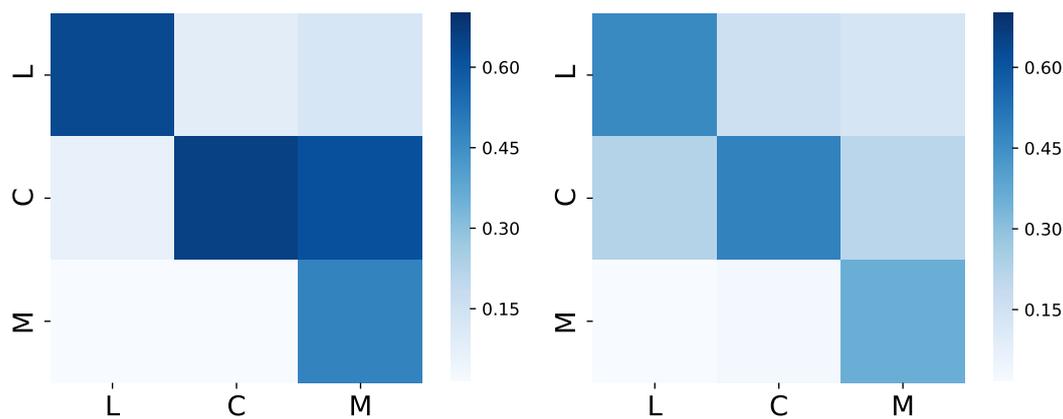


Figure 1.3: Matrices de normes de noyau pour Bund future (à gauche) et DAX future (à droite).

**L'interprétation de la matrice de Hawkes** Dans la figure 1.3, nous présentons dans une carte thermique les normes du noyau de Hawkes  $|\phi^{\ell m}| = \int_0^\infty \phi^{\ell m}(t) dt$ . Notons que dans notre cadre, ces quantités sont simplement données par  $|\phi^{\ell m}| = \sum_{u=1}^U \alpha_u^{\ell m}$ .

Malgré l'ajout du terme dépendant de la file d'attente, nous retrouvons plusieurs des caractéristiques déjà observées dans les études précédentes, comme la forte composante diagonale correspondant à l'auto-excitation, résultat probable de la corrélation dans le flux d'ordres induite par les stratégies de fractionnement des ordres.

Nous confirmons également que les ordres de marché influencent la liquidité beaucoup plus que l'effet inverse. En particulier, puisque nous étudions ici une interaction du même côté du carnet, nous constatons qu'en moyenne, les ordres au marché ont un effet stimulant sur les annulations.

### 1.6.2 Un modèle de Hawkes réactif en file d'attente pour les meilleures limites du carnet d'ordres

Ce modèle peut être considéré comme un développement ultérieur du modèle SQRH présenté au chapitre 2. Le modèle SQRH est limité à une seule (meilleure) file d'attente afin d'obtenir la propriété d'ergodicité, qui est réinitialisée chaque fois que la position de cette meilleure limite change. Pour cette raison, SQRH ne tient pas compte de la dynamique croisée entre les deux meilleures limites (best bid/best ask), ni des changements de leur prix correspondant. Le modèle présenté dans ce chapitre est consacré à la lutte contre ces inconvénients avec l'ambition de construire un modèle pour la LOB, où les deux dépendances sur le flux d'ordres passés et l'état de la LOB sont présents. Ce modèle sera appelé le "Queue Reactive Hawkes" (modèle QRH).

Comme nous l'avons fait pour le modèle SQRH, le modèle QRH implique la combinaison d'une composante "Queue Reactive" (comme dans [87]), c'est-à-dire une dépendance aux tailles des deux meilleures de file d'attente à l'instant courant, et d'une com-

posante Hawkes prenant en compte la dépendance de flux d'ordres passés sur ces deux files d'attente. De plus, afin de tenir compte de la variation des prix, la réinitialisation du modèle ne se fera qu'au moment de la clôture du marché.

**Dynamique du modèle** Nous considérons pour le modèle QRH huit types d'événements aux files d'attente de la meilleure demande de la meilleure offre d'une LPP, à savoir

$P^+$  ( $P^-$ ) pour les événements qui déplacent le midprice <sup>2</sup> vers le haut (vers le bas) indépendamment de la taille de ce déplacement,

$L^a$  ( $L^b$ ) pour les ordres à cours limité au meilleur ask (bid) qui ne modifient pas le midprice,

$C^a$  ( $C^b$ ) pour les annulations au meilleur ask (bid) qui ne modifient pas le midprice,

$M^a$  ( $M^b$ ) pour les ordres au marché au meilleur ask (bid) qui ne modifient pas le midprice.

Pour chaque  $\ell \in \{P^+, P^-, L^a, L^b, C^a, C^b, M^a, M^b\}$ , définissons  $N_t^\ell$  comme le processus de comptage associé aux événements de type  $\ell$  et  $\lambda^\ell(t)$  l'intensité conditionnelle associée.

Le modèle QRH correspond à la possibilité la plus simple où la partie exogène et la partie auto-excitante de la fonction d'intensité partagent la même dépendance multiplicative sur les états :

$$\lambda^\ell(t) = f^\ell(q_a(t), q_b(t)) \left( \mu^\ell + \sum_m \int_0^t \phi^{\ell m}(t-s) dN_s^m \right)^+, \quad (1.6.43)$$

où l'opérateur  $(\cdot)^+ = \max(\cdot, 0)$ , les fonctions  $f^\ell$  qui codent la dépendance aux états du carnet d'ordre, modulent non seulement l'intensité exogène mais aussi le terme de Hawkes. Nous choisissons une forme paramétrique pour les noyaux  $\phi^{\ell m}$  et en particulier nous adoptons la même spécification somme-exponentielle :

$$\phi^{\ell m}(t) = \sum_{u=1}^U \alpha_u^{\ell m} \beta_u e^{-\beta_u(t-s)}. \quad (1.6.44)$$

Contrairement au modèle SQRH, le modèle QRH est principalement un modèle pour le flux d'ordres. Nous ne tenons pas compte de l'influence du flux d'ordres sur la taille des files d'attente, et nous considérons ces dernières comme des variables exogènes.

**Fonction de perte des moindres carrés** Le modèle QRH défini dans l'équation (1.6.43) n'est pas linéaire en les paramètres, donc l'ELM ne s'agit plus d'un problème convexe. Comme expliqué dans [77], le problème de calibration est bien formé dans un certain sens lorsque la perte des moindres carrés est adoptée.

<sup>2</sup>Nous rappelons que le midprice correspond à la moyenne du meilleur cours vendeur avec le meilleur cours acheteur.

Signalons également que forcer les noyaux à n'avoir que des valeurs positives (c'est-à-dire forcer tous les  $\alpha_u^{lm}$  à être positifs) conduira a priori à des valeurs fortement biaisées pour les noyaux à valeurs négatives, puisque l'estimation réalisée est une estimation conjointe impliquant des relations complexes entre différents types d'événements. Nous invitons les lecteurs à consulter [127] comme exemple d'estimation par les moindres carrés avec des noyaux à valeurs négatives, ainsi que [18], qui donne plusieurs exemples de tels biais.

Nous adoptons donc l'estimation basée sur les moindres carrés au lieu d'utiliser l'ELM. Les détails de l'estimation par les moindres carrés se trouvent dans l'annexe 2.5.3, qui consiste à minimiser la perte suivante  $R(\theta)$  :

$$R(\theta) = \sum_{\ell=1}^D R^\ell(\theta), \quad \text{with} \quad R^\ell(\theta) = \int_0^T \lambda_\ell^2(t; \theta | \mathcal{F}_t) dt - \sum_{k=1}^{N^\ell} \lambda_\ell(t_k^\ell; \theta | \mathcal{F}_{t_k^\ell}) \quad (1.6.45)$$

où  $\lambda_\ell$  est donné par Eq. (1.6.43)<sup>3</sup>.

Le  $R(\cdot)$  ainsi défini est une fonction convexe du  $\alpha_u^{lm}$  (voir Eq. (3.2.3)) et du  $\mu^\ell$ . L'existence d'un optimum global est donc garantie. Comme le montre l'annexe 2.5.3, cette paramétrisation permet un calcul efficace de la fonction de perte au carré  $R$  et de son gradient.

**Valeurs négatives dans la matrice de la norme du noyau** Dans la Figure 1.4, nous montrons les normes de noyau estimées  $\{\int \phi^{lm}(t) dt\}_{\ell m}$  pour le Bund et le DAX. La matrice fournit des informations sur les interactions entre les différents types d'événements lorsque la dépendance de la file d'attente n'est pas prise en compte.

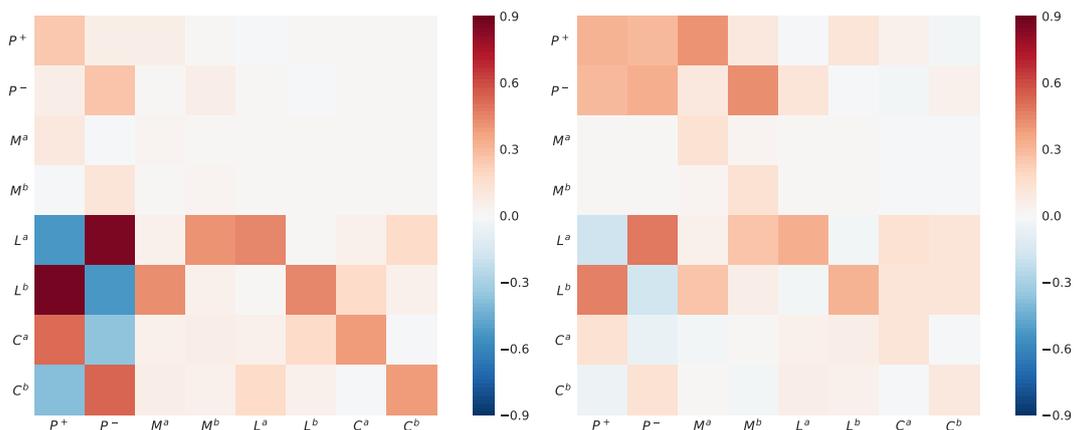


Figure 1.4: Les normes matricielles estimées  $\int \phi^{lm}(t) dt$  en utilisant l'estimation par les moindres carrés du modèle QRH. Futur du Bund à gauche et futur du DAX à droite.

Nous retrouvons de nombreuses caractéristiques mises en évidence dans [13], comme les fortes composantes diagonales pour les ordres à cours limité, les ordres au marché

<sup>3</sup>ou alternativement par Eq 3.2.6 dans le sens donné dans [77]

et les ordres d'annulation (une signature du fractionnement des ordres), ainsi que le fait que les ordres au marché et le mouvement des prix semblent influencer beaucoup plus les ordres à cours limité et les ordres d'annulation que l'inverse. Nous soutenons également que le taux d'arrivée de flux d'ordres après un événement de changement de prix est fortement influencé par la taille du tick de l'actif et donnons une interprétation raisonnable. Voir le chapitre 3.2.3 pour une discussion complète.

**Variable d'état et le déséquilibre de la LOB** Nous observons une dépendance claire des taux d'arrivée des ordres à la fois sur  $q_a^i$  et  $q_b^j$ , indiquant que l'état de la LOB a une influence claire sur les taux d'arrivée des ordres.

A partir du résultat empirique rapporté dans les Figures 3.2 et Figures 3.3, les observations suivantes sont formulées : premièrement, la variation du déséquilibre sur le carnet d'ordres capture la plupart des variations des paramètres intensifs  $f^\ell(q_a, q_b)$  ; ceci est plus évident sur l'actif à grand tick (Bund) pour lequel  $f$  peut couvrir presque trois types de magnitudes (pour  $P$ , et  $M$  événements) alors que  $I$  varie de  $-1$  à  $+1$ . Les variations de  $f^\ell(q_a, q_b)$  sont plus faibles pour l'actif à petit tick (DAX), de sorte que l'effet du déséquilibre est moins prononcé bien que toujours visible pour les événements  $P$  et, dans une moindre mesure,  $L$ . Deuxièmement, les variations d'intensité qui ne sont pas capturées par le déséquilibre sont principalement situées autour de  $I = 0$  où les tailles des files d'attente individuelles ( $q_a$  ou  $q_b$ , qui sont presque égales) semblent avoir un impact important. Nous invitons les lecteurs à consulter le chapitre 3 pour des discussions plus détaillées sur le lien entre  $f^\ell(q_a, q_b)$  et le déséquilibre des tailles de file d'attente.

### 1.6.3 De la volatilité rugueuse à la volatilité multifractale

Bien que la construction du modèle RFSV semble assez différente du processus log-normal MRW, les calculs dans les annexes 4.7.1, 4.7.2 et 4.7.3 suggèrent une similarité assez profonde entre eux. Nous remarquons d'abord que lorsque  $H \simeq 0$  dans la RFSV, les fonctions d'auto-covariance de la volatilité logarithmique semblent être logarithmiques, c'est-à-dire comme sa contrepartie MRW. Intuitivement, le MRW pourrait donc être considéré comme le cas "super rugueux" (par exemple,  $H = 0$ ) dans le modèle RFSV. C'est le premier sujet que nous examinons dans le chapitre 4.

**Retrouver RFSV et MRW à partir de processus S-fBM** Dans l'appendice 4.7.1 du chapitre 4, nous construisons une famille de processus aléatoires, que l'on nomme mouvement brownien fractionnaire stationnaire (S-fBM) à partir d'un champ aléatoire gaussien. Cette construction peut être rapprochée de l'approche originale proposée par Takenaka pour construire des champs corrélés, notamment le mouvement brownien fractionnaire (voir [130] et Appendix 4.7.2). Nous montrons qu'à travers une variation continue de l'ensemble des paramètres, nous pouvons retrouver à la fois RFSV et MRW

comme des cas particuliers de processus S-fBM.

Puisque le fOU traité et le logarithme de la fonction de densité impliquée dans le MRM sont tous deux des processus gaussiens stationnaires, ils sont caractérisés par leur fonction de covariance. La fonction de covariance des processus S-fBM (notée  $\omega_{H,T}(t)$  dans ce qui suit) est :

$$\text{Cov}[\omega_{H,T}(t), \omega_{H,T}(t + \tau)] = \begin{cases} \frac{\lambda^2}{2H(1-2H)} [T^{2H} - \tau^{2H}] & \tau < T \\ 0 & \tau \geq T. \end{cases} \quad (1.6.46)$$

La similarité entre les processus S-fBM et MRW est également révélée par la mesure stochastique S-fBM, qui est définie comme suit :

$$M_{H,T}(dt) = \sigma^2 e^{\omega_{H,T}(t)} dt. \quad (1.6.47)$$

Nous montrons en annexe 4.7.3, que la mesure MRM peut être récupérée à partir de la limite du log S-fBm lorsque  $H \rightarrow 0$ . Plus précisément, nous avons

**Proposition 1.6.5.** *Soit  $M_{H,T}(t) = M_{H,T}([0, t])$  le processus log S-fbm défini à partir de 4.2.13. Alors, lorsque  $H \rightarrow 0$ , on a*

$$M_{H,T}(dt) \Rightarrow \widetilde{M}_T(dt), \quad (1.6.48)$$

où  $\Rightarrow$  représente la convergence faible et  $\widetilde{M}_T$  est un MRM log-normal avec le même coefficient d'intermittence et la même échelle intégrale  $T$  que  $M_{H,T}$ .

La preuve est fournie dans l'annexe 4.7.3, qui indique que le MRM peut être considéré comme un cas limite d'un log S-fBM et, par conséquent, pourrait être considéré comme la limite "super rugueuse".

**Validité de l'estimation des paramètres via les propriétés d'échelle** Nous discutons de la fiabilité de l'approche de calibration des paramètres, à savoir  $H, \nu^2$  et  $T$  à travers la soi-disant "propriété d'échelle" du processus de volatilité. Une telle méthode est utilisée dans [65], où le paramètre de Hurst  $H$  est signalé comme étant autour de 0,15. Précisément, la calibration de  $H$  est basée sur la relation suivante :

$$\mathbb{E}[|\delta_\tau \omega_{H,T}(t)|^q] = C_q \tau^{qH}. \quad (1.6.49)$$

Cependant, comme  $\omega_{H,T}(t)$  ne peut pas être directement observé, les auteurs considèrent comme un proxy de  $\mathbb{E}[|\delta_\tau \omega_{H,T}(t)|^q]$ , les moments observables :

$$m(q, H, \tau, \Delta) = C_q \tau^{qH}. \quad (1.6.50)$$

où

$$m(q, H, \tau, \Delta) = \mathbb{E}\left(|\ln M_{H,T,\Delta}(t + \tau) - \ln M_{H,T,\Delta}(t)|^q\right) \quad (1.6.51)$$

et  $M_{H,T,\Delta}(t)$  est la variance dite intégrée sur un intervalle de taille  $\Delta$  :

$$M_{H,T,\Delta}(t) = \sigma^2 \int_t^{t+\Delta} e^{\omega_{H,T}(s)} ds. \quad (1.6.52)$$

Pour récupérer l'Eq. (1.6.49) à partir de l'Eq. (1.6.50), deux hypothèses sont implicitement faites. Nous soulignons que la première hypothèse est

$$\int_t^{t+h} \omega_{H,T}(s) ds \approx \log\left(\int_t^{t+h} e^{\omega_{H,T}(s)} ds\right). \quad (1.6.53)$$

La deuxième hypothèse est

$$\omega_{H,T}(t) \approx \frac{1}{h} \int_t^{t+h} \omega_{H,T}(s) ds. \quad (1.6.54)$$

Bien que ces deux hypothèses semblent raisonnables, elles ne sont jamais justifiées. Dans le paragraphe suivant, nous montrons que la première hypothèse conduirait à un biais important dans l'estimation du paramètre de Hurst  $H$ .

**Expansion de la petite intermittence pour les moments des logarithmes des mesures** Nous commençons par justifier l'hypothèse dans l'Eq. (1.6.54), qui est également mentionnée sous le nom de *small intermittency*, c'est-à-dire l'expansion asymptotique avec  $\lambda^2 \ll 1$  dans [14]. Il est facile de vérifier que toutes les preuves et tous les résultats établis dans [14] sous la limite  $\lambda^2 \rightarrow 0$  pour la mesure MRM log-normale  $\widetilde{M}_T$  restent valables pour  $M_{H,T,\Delta}$  pour  $H > 0$ , ce qui conduit au résultat suivant :

**Proposition 2.** *Soient  $t_1, \dots, t_n$  des  $n$  temps arbitraires. Les moments généralisés du logarithme de  $\Delta^{-1}M_{H,T,\Delta}(t)$  admettent le développement en série de Taylor suivant autour de  $\lambda^2 = 0$  :*

$$\mathbb{E}\left[\ln\left(\frac{M_{H,T,\Delta}(t_1)}{\Delta}\right) \dots \ln\left(\frac{M_{H,T,\Delta}(t_n)}{\Delta}\right)\right] = \lambda^n \Delta^{-n} \mathbb{E}\left[\Omega_{H,T,\Delta}(t_1) \dots \Omega_{H,T,\Delta}(t_n)\right] + o(\lambda^n), \quad (1.6.55)$$

où  $\Omega_{H,T,\Delta}(t)$  est le processus gaussien défini par

$$\Omega_{H,T,\Delta}(t) = \frac{1}{\lambda} \int_t^{t+\Delta} \left(\omega_{H,T}(u) - \mathbb{E}(\omega_{H,T}(u))\right) du. \quad (1.6.56)$$

A notre connaissance, c'est la première fois que l'Eq. (1.6.54) est strictement justifiée.

**L'effet de lissage** Pour examiner la validité de l'équation (1.6.53), nous avons calculé la forme exacte de la variance de  $\ln M_{H,T,\Delta}(t+\tau) - \ln M_{H,T,\Delta}(t)$ , notée  $V(H, \tau, \Delta)$ . Pour  $\tau < T$ ,

$$V(H, \tau, \Delta) = 2\text{Var}(Z_\Delta) - 2C_Z(\Delta, \tau) \quad (1.6.57)$$

$$= \frac{\lambda^2}{H(1-2H)} \left( \frac{(\tau + \Delta)^{2H+2} + |\tau - \Delta|^{2H+2} - 2\tau^{2H+2}}{\Delta^2(2H+1)(2H+2)} - \frac{2\Delta^{2H+2}}{\Delta^2(2H+1)(2H+2)} \right) \quad (1.6.58)$$

$$= \lambda^2 \tau^{2H} g_H\left(\frac{\Delta}{\tau}\right) \quad (1.6.59)$$

avec

$$g_H(z) = \frac{|1+z|^{2H+2} + |1-z|^{2H+2} - 2|z|^{2H+2} - 2}{z^2 H(1-2H)(2H+1)(2H+2)}. \quad (1.6.60)$$

L'expression finale pour les moments des incréments du logarithme de mesure est la suivante :

$$m(q, H, \tau, \Delta) \stackrel{\nu}{=} 2^{\frac{q}{2}} \pi^{-1/2} \Gamma\left(\frac{q+1}{2}\right) \lambda^q \tau^{qH} \left[ g_H\left(\frac{\Delta}{\tau}\right) \right]^{q/2} \quad (1.6.61)$$

À cause du terme  $g_H(\cdot)$ , l'estimation de  $H$  basée sur l'équation (1.6.49) peut être biaisée. Une analyse détaillée de ce biais est présentée dans la section 4.4.1.

**Estimateurs GMM pour  $H$  et  $\lambda^2$**  Nous proposons deux estimateurs GMM basés sur les moments du second ordre du processus log S-fBM  $M_{H,T}$  ou de son logarithme en considérant respectivement  $C_M(\Delta, \tau)$  la fonction de corrélation de  $M_{H,T,\Delta}$  et  $C_{\ln M}(\Delta, \tau)$ , la fonction de covariance de  $\ln M_{H,T,\Delta}$ .

Si  $L$  désigne la taille globale de l'intervalle où les données empiriques sont disponibles à l'échelle  $\Delta$ , on peut mesurer  $M_{H,T,\Delta}(k\Delta)$  (ou de manière équivalente  $\ln M_{H,T,\Delta}(k\Delta)$ ) pour  $k = 1 \dots N$  où  $N = \frac{L}{\Delta}$  et les estimateurs des fonctions de corrélation précédentes se lisent :

$$\widehat{C}_M(\Delta, k\Delta) = N^{-1} \sum_{j=1}^{N-k} M_{H,T,\Delta}(j\Delta) M_{H,T,\Delta}((j+k)\Delta), \quad (1.6.62)$$

$$\widehat{C}_{\ln M}(\Delta, k\Delta) = N^{-1} \sum_{j=1}^{N-k} \left( \ln M_{H,T,\Delta}(j\Delta) - \widehat{\mu}_\Delta \right) \left( \ln M_{H,T,\Delta}((j+k)\Delta) - \widehat{\mu}_\Delta \right), \quad (1.6.63)$$

$$\widehat{\mu}_\Delta = \frac{1}{N} \sum_{k=1}^N \ln M_{H,T,\Delta}(k\Delta). \quad (1.6.64)$$

Dans la section 4.4.2, nous montrons que dans le régime de haute fréquence, les estimations de  $T$  ou de  $\sigma^2$  sont inatteignables. Nous considérons exclusivement le problème de l'estimation des valeurs des paramètres  $H$  et  $\lambda^2$  (ou alternativement  $\nu^2$ ) en utilisant l'un des deux ensembles de moments suivants :

$$\text{GMM}_M : \mathcal{M}_1 = \left( \widehat{C}_M(1, j_1) - \widetilde{C}_M(1, j_1), \dots, \widehat{C}_M(1, j_Q) - \widetilde{C}_M(1, j_Q) \right),$$

$$\text{GMM}_{\ln M} : \mathcal{M}_2 = \left( \widehat{C}_{\ln M}(1, j_1) - \widetilde{C}_{\ln M}(1, j_1), \dots, \widehat{C}_{\ln M}(1, j_Q) - \widetilde{C}_{\ln M}(1, j_Q) \right)$$

où  $Q$  est le nombre de moments,  $j_1, j_2, \dots, j_Q$  sont différents indices de temps,  $\widehat{C}_M$  et  $\widehat{C}_{\ln M}$  sont les estimateurs empiriques de respectivement  $C_M$  et  $C_{\ln M}$  :

$$\widetilde{C}_{\ln M}(1, n) = K_1 + \widetilde{D}_{\ln M}(n) + V_1 \delta_n, \quad (1.6.65)$$

$$\widetilde{C}_M(1, n) = K_2 \widetilde{R}_M(n) \quad (1.6.66)$$

où  $K_1, K_2$  et  $V_1$  sont 3 constantes positives aléatoires et  $\delta_n$  représente la fonction de Kronecker.

**Résultats empiriques** Nous appliquons l'estimateur  $\text{GMM}_{\ln M}$  pour calibrer l'exposant de rugosité  $H$  et le coefficient d'intermittence  $\lambda^2$  à partir de la volatilité empirique réalisée de divers actifs. Notre étude se base sur la bibliothèque réalisée par l'Oxford-Man Institute of Quantitative Finance (OIQFRL) et la base de données Yahoo-Finance (YF), qui concernent respectivement les indices boursiers et les actions individuelles.

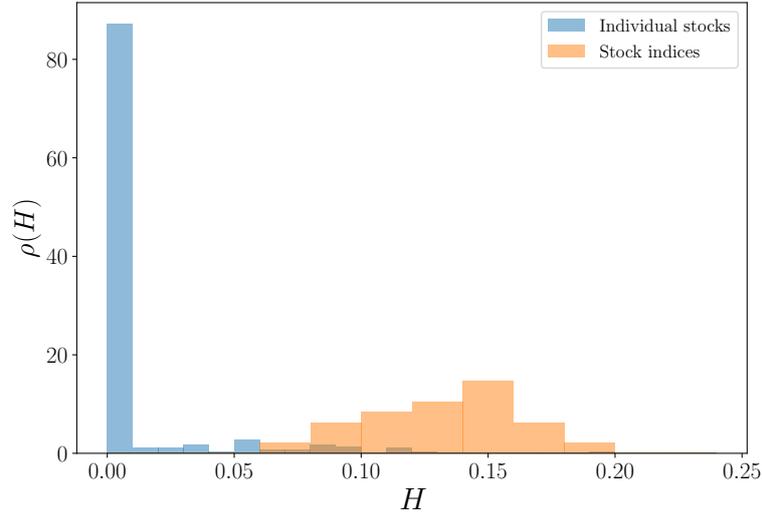


Figure 1.5: Estimation de la distribution de densité de probabilité de l'estimation de l'exposant de Hurst  $\hat{H}$  pour les 296 actions individuelles (histogramme bleu) de la base YF et pour les 24 indices boursiers (histogramme orange) de la base OIQFRL.

Les estimations ainsi obtenues pour l'ensemble des 24 indices boursiers de la base de données OIQFRL et les 296 actions individuelles de la base de données YF sont résumées dans la Figure. 1.5, où nous avons reporté les histogrammes normalisés de l'estimation  $\hat{H}$  pour les exposants de Hurst des indices boursiers et des actions individuelles des deux ensembles de données.

Les exposants de Hurst des indices boursiers se situent autour de  $H \simeq 0.13$  avec une dispersion assez importante. La distribution des valeurs  $H$  des actions individuelles culmine principalement autour d'une très petite valeur moyenne de  $H \simeq 0,01$ . Il apparaît que la log-volatilité des indices boursiers est beaucoup plus régulière que la log-volatilité des actions individuelles, qui s'avère être bien décrite par un modèle multifractal caractérisé par  $H = 0$ .

Parallèlement, les résultats des simulations numériques de la section 4.4.4 suggèrent que le coefficient d'intermittence défini par  $\lambda^2 = H(1 - 2H)\nu^2$  semble être une quantité beaucoup plus fiable que la variance  $\nu^2$  de la log-volatilité. Nous avons reporté dans la Fig. 1.6(a) les valeurs estimées  $\hat{\lambda}^2$  pour les 296 actions individuelles (points bleus) et les 24 indices boursiers (triangles orange) en fonction du logarithme de l'exposant estimé  $\hat{H}$ . Nous pouvons voir que tous les points sont distribués autour de la valeur  $\lambda^2 \simeq 0.07$  pour les actions et  $\lambda^2 \simeq 0.05$  pour les indices. En revanche, si l'on estime le paramètre de variance  $\nu^2$ , on observe une très grande dispersion de ses valeurs.

Un tel résultat favorise le coefficient d'intermittence  $\lambda^2$  comme paramètre pertinent pour rendre compte des fluctuations de la volatilité. Cette quantité semble également être “presque universelle” avec une valeur  $\lambda^2 \simeq 0.07$  pour les actions et 0.05 pour les indices.

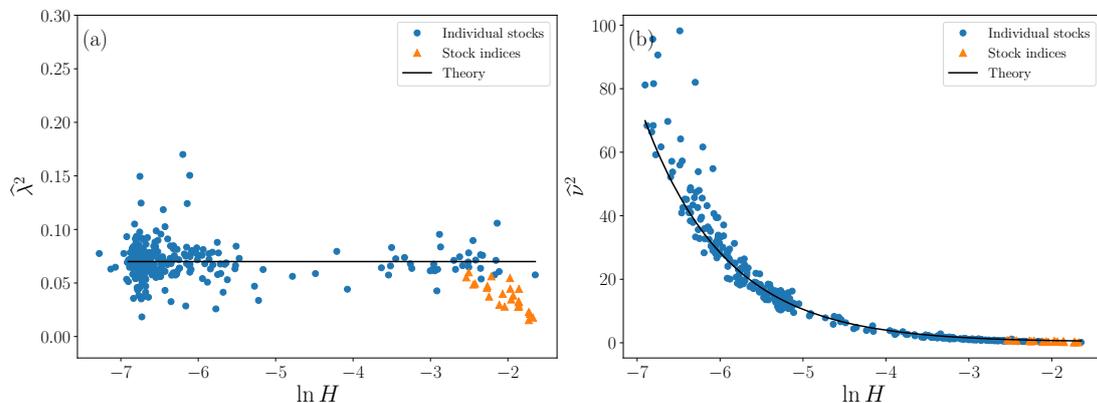


Figure 1.6: Estimation des paramètres d'intermittence et de variance en fonction de l'exposant de Hurst estimé. (a) Coefficient d'intermittence estimé  $\hat{\lambda}^2$  en fonction du logarithme de l'exposant de Hurst estimé  $H$ . La ligne continue représente la valeur constante  $\lambda^2 = 0,07$  représentant le meilleur ajustement des données des stocks individuels. (b) Coefficient de variance estimé  $\hat{\nu}^2$  en fonction du logarithme de l'exposant de Hurst estimé  $H$ . La ligne continue représente l'expression logarithmique S-fBM (4.4.54). Dans (a) et (b) les points bleus représentent les données des actions individuelles tandis que les triangles oranges sont associés aux estimations des séries temporelles des indices.

#### 1.6.4 Dissocier les contributions à la covariance des prix entre différents actifs par des participants au marché

Ce chapitre met en évidence la modélisation de la covariance des prix sur les marchés électroniques. Nous avons présenté un modèle plus élaboré, qui peut être considéré comme un développement ultérieur de [127]. A notre connaissance, nous proposons pour la première fois un modèle qui permet de mesurer la contribution marginale de chaque participant du marché à la covariance des variations de prix de deux actifs.

Un cadre de processus de Hawkes multivarié est introduit pour modéliser les sauts de prix comme le résultat cumulatif de flux d'ordres. La covariance est reconstruite à partir de l'intensité moyenne de flux d'ordres. Nous introduisons une relation approximative avec le premier ordre de la matrice du noyau. Sous des hypothèses raisonnables, nous tenons compte de la contribution marginale de chaque participant du marché à la covariance.

Les résultats empiriques obtenus avec les actions composites de l'indice CAC40 sont présentés. Une méthode d'estimation non paramétrique appelée loi conditionnelle im-

plémentée dans la bibliothèque *Tick* (voir [21]) est réalisée. Notre résultat suggère que notre cadre basé sur le processus de Hawkes est capable de reproduire la covariance du prix (par rapport au rendement) entre différents actifs. De plus, l'évidence empirique suggère que la contribution des participants du marché à la covariance ne dépend pas de l'actif.

### Modélisation de la covariance du prix par un processus de Hawkes multivarié

Considérons les flux d'ordres de deux actifs notés  $S_1$  et  $S_2$ , dont les prix moyens sont notés de manière correspondante  $P_1(t)$  et  $P_2(t)$ , où la souscription numérique désigne des actifs différents. On note par  $\mathcal{N}_1$  tous les types de flux d'ordres de l'actif 1 et  $\mathcal{N}_2$  tous les types de flux d'ordres de l'actif 2 avec  $\#\{\mathcal{N}_1\} = \#\{\mathcal{N}_2\} = N$ . De plus, nous notons en particulier  $\mathcal{M}_1 \subset \mathcal{N}_1$  ( $\mathcal{M}_2 \subset \mathcal{N}_2$ ) les types d'événements qui pourraient changer le prix moyen de l'actif 1 (correspondant à l'actif 2). Nous supposons que les flux d'ordres de deux actifs sont régis par un processus de Hawkes multivarié dont l'intensité, pour le type d'événement  $i$ , prend la forme suivante

$$\lambda_i(t) = \mu^i + \sum_{j=1}^{2N} \int \phi^{ij}(t-s) dN_s^j. \quad (1.6.67)$$

Nous définissons la covariance entre  $P_1(t)$  et  $P_2(t)$  dans l'intervalle  $[t, t + \tau]$  comme la variation corrélée des prix :

$$\text{Cov}_\tau[S_1, S_2](t) := \text{Cov}[\Delta_\tau P_1(t), \Delta_\tau P_2(t)]. \quad (1.6.68)$$

Pour le processus de Hawkes associé, nous notons sa matrice noyau par  $\Phi$  et la convolution de la matrice noyau par  $\mathbf{R}$ . Sous certaines hypothèses raisonnables, nous montrons que la covariance des prix peut être représentée à l'aide du vecteur d'intensité moyenne  $\Lambda := \mathbb{E}[\lambda]$  et de la matrice  $\mathbf{R}$  :

$$\begin{aligned} \text{Cov}[S_1, S_2] &= \sum_{i \in \mathcal{M}_1} \sum_{j \in \mathcal{M}_2} \delta_i \delta_j \sum_{k \in \mathcal{N}} \Lambda^k \mathbf{R}^{ik} \mathbf{R}^{jk} \\ &= \sum_{k \in \mathcal{N}} \Lambda^k \left( \sum_{i \in \mathcal{M}_1} \delta_i \mathbf{R}^{ik} \right) \left( \sum_{j \in \mathcal{M}_2} \delta_j \mathbf{R}^{jk} \right) \end{aligned} \quad (1.6.69)$$

où  $\delta_i$  est le changement du prix moyen résultant d'un événement de type  $i$ .

**Approximation de premier ordre de la covariance** Les preuves empiriques suggèrent que sur les marchés électroniques, la plupart de flux d'ordres sont excités par les flux d'ordres précédents au sein du même actif. L'effet d'excitation croisée pour les flux d'ordres entre différents actifs est beaucoup plus faible que l'effet d'auto-excitation. Avec quelques hypothèses raisonnables, nous élaborons une formule approximative pour la covariance :

$$\begin{aligned}
\text{Cov}[S_1, S_2] &= \sum_{k \in \mathcal{N}} \Lambda^k \left( \sum_{i \in \mathcal{M}_1} \delta_i R^{ik} \right) \left( \sum_{j \in \mathcal{M}_2} \delta_j R^{jk} \right) \\
&= \Lambda^1 (\delta_1 R^{1,1}) (\delta_2 R^{2,1}) + \Lambda^2 (\delta_1 R^{1,2}) (\delta_2 R^{2,2}) \\
&= \Lambda^1 (\delta_1 R^{1,1}) (\delta_2 R^{2,2} \Phi^{2,1} R^{1,1}) + \Lambda^2 (\delta_1 R^{1,1} \Phi^{1,2} R^{2,2}) (\delta_2 R^{2,2}).
\end{aligned} \tag{1.6.70}$$

Sous la forme scalaire,

$$\begin{aligned}
\text{Cov}[S_1, S_2] &= \sum_{k \in \mathcal{N}_1} \Lambda_k \Delta R_D^{1k} \Delta R_I^{2k} + \sum_{k \in \mathcal{N}_2} \Lambda_k \Delta R_D^{2k} \Delta R_I^{1k} \\
&= \sum_{k \in \mathcal{N}_1} \Lambda_k \Delta R_1^k \Delta R_2^m \phi_{2,1}^{m,j} R_1^{j,k} + \sum_{k \in \mathcal{N}_2} \Lambda_k \Delta R_2^k \Delta R_1^m \phi_{1,2}^{m,j} R_2^{j,k},
\end{aligned} \tag{1.6.71}$$

où  $\phi_{2,1}^{m,j}$  (resp.  $\phi_{1,2}^{m,j}$ ) est le  $m, j$ -ème élément de la matrice  $\Phi^{2,1}$  (resp.  $\Phi^{1,2}$ ). On définit  $\Delta R_1^k$  de la manière suivante : Si nous notons  $u_1$  (resp.  $u_2$ ) et  $d_1$  (resp.  $d_2$ ) le type d'événement qui déplace le prix moyen de  $S_1$  (resp.  $S_2$ ) vers le haut et vers le bas, alors nous pouvons écrire

$$\Delta R_1^k = R_1^{u_1, k} - R_1^{d_1, k} \tag{1.6.72}$$

et de la même manière

$$\Delta R_2^k = R_2^{u_2, k} - R_2^{d_2, k}. \tag{1.6.73}$$

À la représentation en cascade du processus de Hawkes multivarié, l'équation (1.6.70) a une interprétation plutôt spontanée. En prenant le premier terme comme exemple, un événement de type  $k \in \mathcal{N}_1$  déclenchera des événements de changement de prix de  $S_1$ , qui changeront finalement  $P_1$  par  $\Delta R_1^k R_1^{j,k}$ . Elle excitera également les flux d'ordres suivants sur le marché de  $S_2$ . Si la cascade de flux d'ordres est limitée à la première génération, tous les flux d'ordres déclenchés par le second effet finiront par modifier le prix de  $S_2$  par  $\Delta R_2^m \phi_{2,1}^{m,j}$ .

**Contribution marginale de covariance par des participants au marché** Nous quantifions la contribution marginale de chaque participant au marché à la covariance des prix entre une paire d'actifs. La norme de la matrice  $\Phi$  quantifie la somme des activités de tous les participants au marché. Comme chaque ordre est généré par un participant de marché spécifique, pour le participant de marché  $\alpha$ , sa contribution à la covariance est mesurée par la contribution de ses activités à la norme de la matrice  $\Phi$ , notée  $\Phi_\alpha$ . Nous définissons donc sa contribution à la covariance comme suit

$$\text{Cov}_\alpha = \sum_{k \in \mathcal{N}_1} \Lambda_k \Delta R_1^k \Delta R_2^m \phi_{2,1}^{m(\alpha), j} R_1^{j,k} + \sum_{k \in \mathcal{N}_2} \Lambda_k \Delta R_2^k \Delta R_1^m \phi_{1,2}^{m(\alpha), j} R_2^{j,k} \tag{1.6.74}$$

où le symbole  $m(\alpha)$  dans  $\phi^{m(\alpha), j}$  désigne les flux d'ordres de type  $m$  émis par le participant  $\alpha$ .

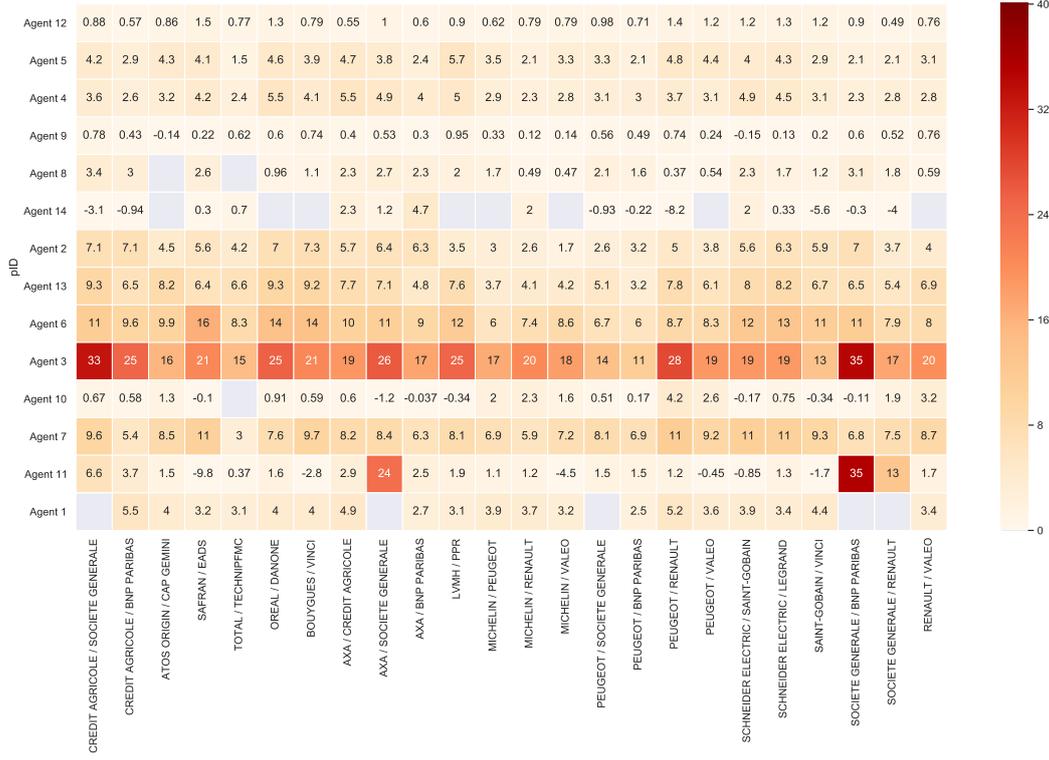


Figure 1.7: Carte thermique de la contribution de l'agent à la corrélation, en pourcentage

**Contribution uniforme de la covariance entre les différents actifs** Nous présentons la contribution des participants au marché à la covariance sous la forme de  $\text{Cov}_\alpha$  définie dans l'équation (1.6.74).

Nous notons la volatilité réalisée du stock 1 et du stock 2 comme étant  $\sigma_1^2$  et  $\sigma_2^2$ . Nous notons  $\text{Cov}$  comme la covariance réalisée et  $\text{Corr}$  comme le coefficient de corrélation. La contribution du participant au marché  $\alpha$  à la corrélation est définie comme suit :

$$\text{Corr}_\alpha := \frac{\text{Cov}_\alpha}{\sigma_1 \sigma_2} = \frac{\text{Cov}_\alpha}{\text{Cov}} \frac{\text{Cov}}{\sigma_1 \sigma_2} = \frac{\text{Cov}_\alpha}{\text{Cov}} \text{Corr}. \quad (1.6.75)$$

Selon cette définition, nous n'avons pas nécessairement  $\sum_\alpha \text{Corr}_\alpha = \text{Corr}$ . Dans la figure 1.7, nous présentons la contribution à la covariance de chaque participant au marché sous la forme de  $\text{Corr}_\alpha$ , sous forme de carte thermique.

Les résultats empiriques suggèrent que notre cadre est capable de reproduire la covariance des prix. En outre, les hypothèses et les approximations que nous avons faites sont validées par les résultats empiriques. Nous observons que la contribution des acteurs du marché à la corrélation des prix est uniforme sur différentes paires d'actifs. De plus, une étude détaillée de la contribution de chaque agent suggère que certains participants actifs du marché peuvent adopter des stratégies multi-actifs. Enfin, notre résultat suggère que la contribution des participants au marché à la covariance est étroitement liée au volume total qu'ils émettent.

## Chapter 2

# A single queue reactive Hawkes models for the order flow

### 2.1 Introduction

Building faithful models for the Limit Order Book (LOB) is a longstanding issue on which many efforts have been invested in the quantitative finance community. A rich literature of theoretical and empirical studies of limit order books has emerged in the last decade (see, e.g., [71] and [2] for a recent review). Modeling the LOB is a challenging task due to its intricate dependence structure. Indeed, the configuration of the limit order book is determined by the arrival of multiple types of orders: limit, cancel and market orders in the simplest setting, and the way these orders arrive on the market is non-trivial. For example, it is well known that order arrival inter-event times present strong and persistent autocorrelation (see e.g. [44]), implying that past order flow influences the current state of the book. At the same time, anecdotal as well as empirical evidence ([99]) suggest that market participants look at the state of the order book in order to make their trading decisions.

Models for the LOB can be roughly divided into two main classes. On one side are the models developed by the economics community where the focus is on the behavior of rational agents that act strategically to optimize their utility function (see e.g. [125]). On the other side, a stream of literature, beginning notably with [132], focuses instead on the overall statistical properties of LOBs and assumes a certain simplified dynamics for the order flow in order to build mathematically tractable models that can reproduce, at least partially, some of these observed properties. The present work contributes to the latter and builds on previous works in this field.

As stated above, in the pioneering work [132], the order book is seen as a purely stochastic system - a so-called *zero intelligence* model (all orders arrive randomly) - that allows one to make testable predictions based on measurable input. The work in [45] is one of the first papers to clearly frame the problem of LOB modeling in the context of queuing

theory and Markov chains. By leveraging the properties of Markov chains, the authors are able to derive several conditional probabilities such as the likelihood of a mid-price move or the probability of a limit order execution before a price change. The authors of [4] keep the same assumption of Poisson-driven independent queues and prove, using the theory of infinitesimal generators and Lyapunov stability criteria, the importance of the cancellation structure to ensure the stability of the LOB distribution and also show that under their model the price process converges to a Wiener Process. Although the hypothesis made by these models is in disagreement with some major empirical facts, they present the advantage of being very tractable and allowing the derivation of many useful quantities analytically. In [3], the authors drop the assumption of uncorrelated order flow and introduce a memory effect by choosing to model the rates of limit and market order arrivals (respectively  $\lambda^L$  and  $\lambda^M$ ), by a Hawkes process ([79]):

$$\lambda^\ell(t) = \mu^\ell + \sum_{m \in \{L, M\}} \int \phi^{\ell m}(t-s) dN_s^m, \text{ with } \ell \in \{L, M\}. \quad (2.1.1)$$

By setting each kernel function to exponential form,  $\phi^{\ell m}(t) = \alpha^{\ell m} e^{-\beta t}$ , the process  $(\lambda^L, \lambda^M, N^L, N^M)$  has the Markov property and thus the authors are able to use a similar machinery to [4] in order to study the limiting behavior of their model. In [17], [13] and [126], the authors also use multivariate Hawkes processes to analyze the order flow interactions at the first level of the order book. Their model is calibrated without any assumption on the Hawkes kernel shapes using a non-parametric method.

In [87], the authors focus instead on the influence of the current state of the LOB on trading decisions. They propose a simple Markov model, the so-called *Queue-Reactive* (QR) model, where the order flow arrival intensity of limit, market or cancel orders depends only on the current state of LOB through the available volume :

$$\lambda^\ell(t) = \mu^\ell(q(t)), \quad (2.1.2)$$

for any  $\ell \in \{L, M, C\}$ . They also establish the conditions under which their model possesses ergodic properties, making it possible to reproduce the empirical LOB queue size distributions as the invariant distribution of a Markov process. We also mentioned that in an earlier work [64] by Garèche et al., explicitly dependence of trading decisions on the current state of the limit order book is incorporated in a Fokker-Planck framework. More recently, [102] extends the model of [87] by allowing the order book dynamics to depend also on the type of the order that led to a complete depletion of a level (i.e. a market or cancel order) and also by taking into account the order size. [102] thus departs slightly from the pure Markovian framework. Optimal market-making strategies are also discussed and performance is assessed on real data.

In this paper, we aim at contributing to this stream of literature by building on the work of [87] on one side, and on [17], [3] and [13] as the other side by presenting a single best queue (best ask queue or best bid queue) model taking into account both the dependence on past order flows (through a Hawkes kernel) and the dependence on

the state of the queue. This can be simply done by combining the Hawkes approach of (2.1.1), i.e., a Single Queue Hawkes (SQH) model, and the QR approach of (2.1.2), i.e., a Single Queue Reactive (SQR) model. This leads to the model we name Single Queue Reactive Hawkes (SQRH) model:

$$\lambda^\ell(t) = \mu^\ell(q(t)) + \sum_{m \in \{L, M, C\}} \int \phi^{\ell m}(t-s) dN_s^m, \quad (2.1.3)$$

for  $\ell \in \{L, M, C\}$ . Let us notice that the issue of considering both dependencies has also been considered a few years ago (during the completion of the first version of the present work) by Morariu-Patrichi and Pakkanen [114]. These authors proposed a general framework called “state-dependent Hawkes process” where the Hawkes kernels  $\phi^{\ell m}$  are functions of some state process  $X(t)$  that can take a finite number of values and that switches from one state to another one when an event of the Hawkes process occurs and according to a transition rule that depends on the type of this event. The specific application of this framework to LOB modeling proposed in [114] mainly consists in considering either the volume imbalance or the spread as the state variable. Let us mention another very recent and related work in the paper of Daw and Pender [48] that defined and studied a Markov process constructed a pair of inter-dependent processes  $(N_t, Q_t)$ , where  $N_t$  is a counting process and  $Q_t$  a queuing process.

By calibrating the model (2.1.3) using high-frequency data from Eurex future markets, we show that it achieves a significantly better fit of the data than the pure Single Queue Reactive (SQR) model defined by (2.1.2) (which basically amounts to dropping the Hawkes kernel term in (2.1.3)) or than a pure Single Queue Hawkes (SQH) model (which basically amounts in considering a constant exogenous intensity  $\mu^\ell$  in (2.1.3)).

This chapter is organized as follows. In Section 2.2 we define precisely the SQRH model that consists in adding an order flow dependence as provided by a multivariate Hawkes process to a single queue version (SQR) of the Queue Reactive (QR) model introduced in [87]. We show how such a model can be calibrated using a maximum likelihood approach and prove that, very much like the QR model, under some reasonable assumption, the queue size admits an invariant distribution. We then compare the likelihood of the SQRH model with both a standard Hawkes model with no state-dependence and with the SQR model of [87] on real data from the Eurex exchange. Comparisons with empirical data show that the SQRH model represents an important improvement of the SQR model not only with respect to the inter-event time statistics but also regarding the predicted shape of the equilibrium queue size distribution. Concluding remarks and prospects for future research are provided in Section 2.3. Technical results like the proof of the ergodicity of the SQRH model and the model calibration issues by Maximum Likelihood are given in the Appendices.

## 2.2 A Single Queue Reactive model with memory

### 2.2.1 Model and assumptions

As mentioned in the introduction, Huang, Lehalle and Rosenbaum present, in [87], a model where the order flow arrival at a given price level is modeled as an inhomogeneous Poisson process with an intensity that depends only on the current state of the order book, i.e., on the queue sizes. They name this property *Queue Reactive* (QR). The main purpose of our paper is not to build a realistic model of the full order book but simply to show that such a model could be improved greatly by accounting for the correlated nature of the order flows. For that purpose and for the sake of simplicity, we chose to focus on only modeling a single best queue (best ask or best bid queue) and not trying to model the entire LOB.

In [87], the authors define a reference price  $p_{\text{ref}}$  that separates bid and ask sides and which is equal to the midprice  $p_{\text{mid}}$  (i.e., the mean value of the best available bid and best available ask prices with non zero quantities) if the spread is an odd multiple of the tick size and equal to  $p_{\text{mid}} \pm \frac{\delta}{2}$ , whichever is closer to the previous  $p_{\text{ref}}$ , if the spread is an even multiple of the tick size. The best queue we consider can be either the queue corresponding to the best ask price above this reference (i.e., the best ask queue) or the queue corresponding to the best bid price below this reference (i.e., the best bid queue). The size of this queue at time  $t$  is referred to as  $q(t)$  and can be modified by the arrival of either limit (L), market (M) or cancel (C) orders. Following the QR framework in [87], we will assume all these orders to be of unitary volume (corresponding to the average event size), so a limit order adds one unit to the queue, while a market or a cancel order subtracts one unit. The arrival intensities at time  $t$  of respectively limit, market and cancel orders on the considered queue of size  $q(t)$  will be denoted by  $\lambda^L(t)$ ,  $\lambda^M(t)$  and  $\lambda^C(t)$ .

A single queue version of the QR model defined in [87], referred to as Single Queue Reactive (SQR) model (not that it is referred to as “model I” in [87]), assumes that these rates are only functions of the queue size  $q(t^-)$  just before a potential order arrival at time  $t$  :

$$\begin{aligned}\lambda^L(t) &= \mu^L(q(t^-)) \\ \lambda^C(t) &= \mu^C(q(t^-)) \\ \lambda^M(t) &= \mu^M(q(t^-))\end{aligned}\tag{2.2.4}$$

where the functions  $\{\mu^\ell(q)\}_\ell$  are the parameters of the model. They correspond to the rates of a birth-death Markov process and can easily be estimated via maximum likelihood which, in this case, amounts to the computation of simple conditional empirical means of intensities. As the labeling of a price level is relative to the reference price, when  $p_{\text{ref}}$  changes, the level labels also change. Hence, the estimation is performed on intervals where  $p_{\text{ref}}$  is constant and each period is regarded as an independent realization

of the process. As shown in [87], the SQR model is an ergodic continuous-time jump Markov process provided the limit order rate is bounded for large queue sizes, and the rate at which orders are removed is larger than the rate which increases the queues. In this respect, the SQR model represents a simple and parsimonious Markov model that allows one to account for the state-dependent nature of the dynamics on the best queue of a LOB.

As emphasized above, our goal is to consider an extension of the SQR model that accounts not only for the queue size dependencies but also for the memory effects in the order flows. For that purpose, the SQRH model associates the SQR approach with a multivariate Hawkes process. In the following,  $N_t^L$ ,  $N_t^C$ ,  $N_t^M$ ,  $\lambda^L(t)$ ,  $\lambda^C(t)$  and  $\lambda^M(t)$  will denote the counting processes and their associated intensities defined by the arrivals of respectively limit, cancel and market orders at the best queue. For  $\ell \in \{L, M, C\}$  (for respectively limit, market and cancel orders), the SQRH model thus defines  $\lambda^\ell(t)$  as:

$$\lambda^\ell(t) = \left( \mu^\ell(q(t^-)) + \sum_{m \in \{L, M, C\}} \int_0^t \phi^{\ell m}(t-s) dN_s^m \right) \mathbb{1}_{(\ell=L) \vee (q(t^-) > 0)} \quad (SQRH), \quad (2.2.5)$$

where the queue size  $q(t)$  is simply given by  $q(t) = q(0) + N_t^L - N_t^M - N_t^C$  and the factor  $\mathbb{1}_{(\ell=L) \vee (q(t^-) > 0)}$  ensures that no market or cancel order can occur if the queue size is empty. The baseline intensities  $\{\mu^\ell(q)\}$  depend on the queue state  $q$  while the Hawkes kernels  $\phi^{\ell m}(t)$  account for the effect of past orders of type  $m$  occurrence on the current intensity  $\lambda^\ell(t)$ . In full rigor, to complete the model definition, one should specify the law of the initial queue size  $q(0)$ . Since, as shown below, we will consider a situation where the queue process is a component of an ergodic vector Markov process, the choice of the law of  $q(0)$  is not essential and we simply choose  $q(0) = 0$ .

Using the same formalism, let us note that the previously introduced SQR model can be written as

$$\lambda^\ell(t) = \mu^\ell(q(t^-)) \mathbb{1}_{(\ell=L) \vee (q(t^-) > 0)}. \quad (SQR) \quad (2.2.6)$$

It can be seen as a special case of the SQRH model with zero kernels  $\phi^{\ell m} = 0$ . In the same way, the Single Queue Hawkes (SQH) model can be defined as a SQRH model with constant exogenous intensities:

$$\lambda^\ell(t) = \left( \mu^\ell + \sum_{m \in \{L, M, C\}} \int_0^t \phi^{\ell m}(t-s) dN_s^m \right) \mathbb{1}_{(\ell=L) \vee (q(t^-) > 0)} \quad (SQH). \quad (2.2.7)$$

We proceed as in [87] and we assume that these 3 models (SQR, SQH, SQRH) hold in periods when the reference price is constant, and furthermore that such periods can be considered as independent realizations. Note that by doing so, we reset the Hawkes memory every time there is a change in the reference price. We will discuss this point below when analyzing the empirical results and the order flows simulated by the SQRH model.

The model parameters can be estimated using the maximum likelihood method. The log-likelihood  $L$  of a  $D$ -dimensional point process where the components do not share any parameters has the following general form (see [46], page 21)

$$L(\theta) = \sum_{\ell=1}^D L_{\ell}(\theta), \quad \text{with} \quad L_{\ell}(\theta) = \int_0^T \log \lambda^{\ell}(t; \theta | \mathcal{F}_t) dN_t^{\ell} - \int_0^T \lambda^{\ell}(t; \theta | \mathcal{F}_t) dt \quad (2.2.8)$$

where  $\theta$  denotes the parameter set and  $\lambda^{\ell}$  is the intensity function of the  $\ell$ -th component. To use the method in practice, a parametric form must be specified for the interaction kernels  $\phi^{\ell m}$  in Eq. (2.2.5). A standard choice is to consider that  $\phi^{\ell m}$  can be written as sum of exponential kernels:

$$\phi^{\ell m}(t) = \sum_{u=1}^U \alpha_u^{\ell m} \beta_u e^{-\beta_u(t-s)} \quad (2.2.9)$$

where  $\alpha_u^{\ell m}$  are parameters of the model and  $\beta_u$ ,  $U$  are hyper-parameters suitably chosen using an iterative grid-search as described in Appendix 2.5.3. Expression (3.2.3) also presents the important advantage that the resulting log-likelihood is a convex function of the model parameters  $\{\alpha_u^{\ell m}\}_{\ell, m, u}$  and  $\{\mu^{\ell}\}_{\ell}$ . To facilitate the notations, we use  $\vec{\mu}$  and  $\vec{\alpha}$  to represent all  $\mu^{\ell}$  and  $\alpha_u^{\ell m}$ . With such parametrization,  $\theta = (\vec{\mu}, \vec{\alpha})$ . The log-likelihood of the SQRH model (2.2.5) thus reads:

$$\begin{aligned} L(\theta) = & \sum_{\ell \in \{L, M, C\}} \sum_{k=1}^{N^{\ell}} \log(\mu^{\ell}(q(t_k))) + \sum_{m \in \{L, M, C\}} \sum_{u=1}^U \alpha_u^{\ell m} \beta_u \int_0^t e^{-\beta_u(t-s)} dN_s^m \\ & - \sum_{\ell \in \{L, M, C\}} \int_0^T (\mu^{\ell}(q(s)) + \sum_{m \in \{L, M, C\}} \sum_{u=1}^U \alpha_u^{\ell m} \beta_u \int_0^s e^{-\beta_u(s-v)} dN_v^m) ds \end{aligned} \quad (2.2.10)$$

As we show in Appendix 2.5.2, the specific choice of a sum of exponential functions (Eq. (3.2.3)) allows for a computationally efficient calculation of the log-likelihood and of its gradient.

Another important advantage of the parametrization (3.2.3) is that it allows us to work within the framework of Markov processes. Let us first remark that the queue size  $q(t)$  simply corresponds to:

$$q(t) = N_t^L - \sum_{\ell \neq L} N_t^{\ell}. \quad (2.2.11)$$

If one defines

$$o_{\ell m u}(t) = \int_0^t \alpha_u^{\ell m} e^{-\beta_u(t-s)} dN_s^m, \quad (2.2.12)$$

and by denoting  $\vec{o}(t)$  the vector obtained by a vertical stacking of  $o_{\ell m u}(t)$  ( $u \in \{1, \dots, U\}$ ,  $\ell, m \in \{L, M, C\}$ ), then  $(\frac{q(t)}{\vec{o}(t)})$  is a vector Markov process. This property can be proved exactly along the same lines as in Proposition 2.2 of [92]. Moreover, if one assumes that there exist two positive constants  $c^-$  and  $c^+$  such that

$$\sum_{\ell \neq L} \mu^{\ell}(q) \geq c^- q \quad \text{and} \quad \mu^L(q) \leq c^+,$$

we show, in Appendix 2.5.1, with the help of Lyapunov functions approach along the same line as in [4, 3], that the process  $(\frac{q(t)}{\sigma(t)})$  is V-uniformly ergodic which notably means that  $q(t)$  admits an invariant distribution and that this equilibrium is reached exponentially fast.

### 2.2.2 Calibration results

**Data** In this study, we use tick-by-tick level L1 data of Bund future and DAX index future traded on the Eurex electronic future market. The data span the period from October 1st 2013 to September 30th 2014. The dataset consists of snapshots of the first level of the order book, each with a timestamp indicating the record time with microsecond precision, that provides prices and outstanding quantities. Every time a trade occurs a specific line is added to the dataset, thus allowing to precisely determine the type of order (i.e. limit order, cancellation, or market order<sup>1</sup>) that lead to a change in the LOB. The Eurex future market is open from 8 a.m. to 10 p.m., Frankfurt time, however, throughout this paper we only consider the time slot from 9 a.m. to 9 p.m. in order to capture the most active period. In Table 2.1, we report some descriptive statistics of our datasets. We note that, from the microstructural point of view, the Bund future can be considered as a large tick asset, with an average spread very close to one, whereas the DAX has a considerably smaller perceived tick size compared to the Bund. This difference at the market microstructure level will reflect also in the queue dynamics and in the result of our model. Indeed, the queues on the Bund are often large as the midprice stays constant for relatively long periods of time while orders accumulate at the best quotes. On the other hand, for the DAX the midprice changes more frequently, resulting in slimmer queues at the best quotes. Further details on the datasets as well as a more detailed description of the inter-event time distribution can be found in [126].

Let us remind that our simplified framework considers the order book as a collection of independent queues with a strict bid-ask symmetry so all the results presented in the following are obtained by averaging Level I estimations on the bid and ask sides.

**Estimation and goodness-of-fit analysis** In order to estimate the parameters of our model, for each day in our sample, we first compute the reference price  $p_{\text{ref}}$  as specified at the beginning of this section. Then we determine the queue sizes, as in [87], we assume that the order size for all events is a constant corresponding to the average event size (AES), defined as the average volume of all types of orders arriving at the best bid or ask side. We therefore measure the queue size  $q(t)$  in units of AES as

$$q(t) = \lceil \frac{v(t)}{\text{AES}} \rceil \quad (2.2.13)$$

---

<sup>1</sup>With a slight abuse of language, in this paper we use the term “market order” to denote any order that immediately gives rise to trade, regardless to whether or not it has a limit price.

	# $L$	# $C$	# $M$	Avg. spread	Med. spread	AES	Med. inter-event time
Bund	$5.41 \times 10^7$	$4.67 \times 10^7$	$6.29 \times 10^6$	1.012	1.0	6.34	$4.89 \times 10^{-4}$
DAX	$5.46 \times 10^6$	$5.62 \times 10^6$	$6.68 \times 10^5$	1.591	2.0	1.30	$1.73 \times 10^{-3}$

Table 2.1: Descriptive statistics of our dataset. Average number of limit, cancel and market order at the best quotes per day. Average and median spread are measured in ticks, average order sizes are expressed in contracts while median inter-event times are in seconds.

where  $\lceil \cdot \rceil$  is the ceiling function and  $v(t)$  is the volume available at time  $t$  in the queue. In Figure 2.1 we show the empirical distribution of the so-defined  $q(t)$  for Bund (left panel) and DAX (right panel). These distributions are obtained by sampling the book state every 30s over the whole time period. We have kept the same convention as in [87], that the state  $q(t)$  is set to zero if and only if  $v(t) = 0$ . At this moment there is no standing volume on this queue. However, since the midprice is not yet shifted, we still refer to it as the best bid/ask queue. We observe that the Bund future presents a broader and smoother distribution as compared to the one of the DAX, which is on the contrary more concentrated on small queue sizes. This is a direct consequence of the different perceived tick sizes as observed above.

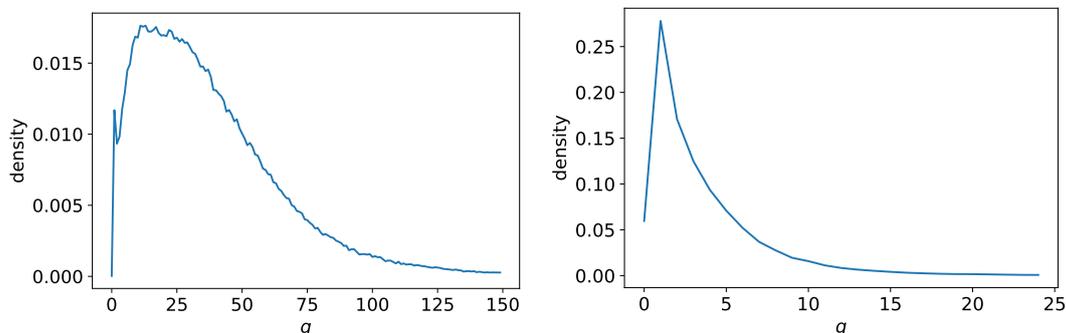


Figure 2.1: Empirical distribution of the queue states (measured in unit of AES) as defined in Eq. (2.2.13). Left: Bund future. Right: DAX future.

Once  $p_{\text{ref}}$  and  $q(t)$  is determined for each side, we divide each day in periods where  $p_{\text{ref}}$  is constant. Then, each period is considered as an independent sample and we determine the parameters of our model by numerically optimizing the joint log-likelihood over all the so-obtained independent samples. Notice that for statistical estimation purposes, in the results reported below we considered only samples that contain at least a total of 20 events and disregarded the other ones. In total, we have 1,043,876 periods for the Bund and 190,126 for the DAX under this criteria. <sup>2</sup>

<sup>2</sup>For Bund future, the average number of events per sample is 103.59. The 10% and 90% percentiles

We observe that, if  $s_k$  stands for the length of the realization  $k$  and  $k_t$  the index of the realization located around time  $t$ , the quantity

$$\tau_m = \frac{\mathbb{E}s_k^2}{\mathbb{E}s_k}$$

represents the average length of the realization  $k_t$  if one chooses  $t$  at random (i.e. with a uniform probability). This quantity is pertinent when performing averages over a fixed grid of times  $\{t_j\}_j$ . For the Bund we have  $\tau_m \simeq 100$  s while for the DAX we estimated  $\tau_m \simeq 16$  s.

As we pointed out above, in order to estimate our model, we need to fix the number of the exponential decays  $U$  as well as the values of the decays  $\beta$  themselves. According to the methodology described in Appendix 2.5.3, we found that  $U = 3$  with  $\beta_1 = 60s^{-1}$ ,  $\beta_2 = 1500s^{-1}$  and  $\beta_3 = 5500s^{-1}$  for the Bund and  $\beta_1 = 40s^{-1}$ ,  $\beta_2 = 2100s^{-1}$ ,  $\beta_3 = 5200s^{-1}$  for the DAX, represent a good compromise between the total number of parameters to estimate (the number of parameters  $\alpha$  grows linearly with  $U$ ) and the model goodness of fit as measured by (penalised) log-likelihood.

The maximum likelihood allows us to perform a quantitative comparison of the SQR and SQRH models in terms of goodness of fit, which is one of the central results of this paper. For the sake of completeness, we also consider a standard Hawkes model, i.e. with no dependence on the queue state. In Table 2.2 we report the log-likelihood values for the three models as well as the Akaike Information Criterion (AIC) score

$$\text{AIC} = 2k - 2L \quad (2.2.14)$$

and the Schwartz information criterion (BIC) score

$$\text{BIC} = k \log N - 2L \quad (2.2.15)$$

where  $k$  is the number of parameters,  $L$  is the log-likelihood and  $N$  is the total sample size (number of events in our case). These scores allow one to compare nested models according to their likelihood while taking into account the different number of parameters (a lower score is better). By looking at the values reported in Table 2.2, we observe that the SQRH model has better scores in terms of AIC and BIC for both assets. We can also use the likelihood ratio test in order to compare the models. Indeed the SQRH model reduces to the SQR model when all the  $\alpha$  are set to zero. Likewise, the SQRH model reduces to a standard Hawkes model when,  $\forall \ell$ ,  $\mu^\ell(q) = \mu^\ell$ , i.e the dependence on the queue state is dropped. We report the test statistics

$$\text{LR} = 2(L(\hat{\theta}_1) - L(\hat{\theta}_0)) \quad (2.2.16)$$

---

are respectively 30 and 207 events. For DAX future the average number of events per sample is 31.87 events while 10% and 90% percentiles are 22 and 46 respectively. We have checked, using 100.000 samples of 20 events obtained from numerical simulations of the SQRH model with parameters close to the ones observed empirically, that our MLE approach provides a reliable estimation of parameters  $\alpha$  and  $\mu$ .

Bund				
	$L$	AIC	BIC	# parameters
SQR	$2.046 \times 10^7$	$-4.093 \times 10^7$	$-4.092 \times 10^7$	450
SQH	$2.052 \times 10^8$	$-4.105 \times 10^8$	$-4.105 \times 10^8$	30
SQRH	$2.055 \times 10^8$	$-4.110 \times 10^8$	$-4.110 \times 10^8$	477
DAX				
	$L$	AIC	BIC	# parameters
SQR	$7.268 \times 10^5$	$-1.453 \times 10^6$	$-1.452 \times 10^6$	75
SQH	$9.386 \times 10^6$	$-1.877 \times 10^7$	$-1.877 \times 10^7$	30
SQRH	$9.506 \times 10^6$	$-1.901 \times 10^7$	$-1.901 \times 10^7$	102

Table 2.2: Log-likelihood, AIC, and BIC values for the three considered models for Bund and DAX data.

where  $\hat{\theta}_1$  and  $\hat{\theta}_0$  are the maximum likelihood estimates for the null and for the alternative model respectively, and  $p$ -values for the likelihood ratio test in Table 2.3. We note that both the SQR and the SQH model are rejected with a very high degree of significance when compared to the SQRH model.

To complete the goodness-of-fit comparison of the models, we look at the inter-event time distribution. In particular, in Figure 2.2 we compare by means of a quantile-quantile plot the empirical inter-event time's distribution with the ones produced by simulations of the calibrated SQR and SQRH models. It is clear from the figure that the SQRH model reproduces strikingly better the empirical inter-event distribution, indicating that including the dependence on the past event is crucial in order to build a good model for the order flow fluctuations.

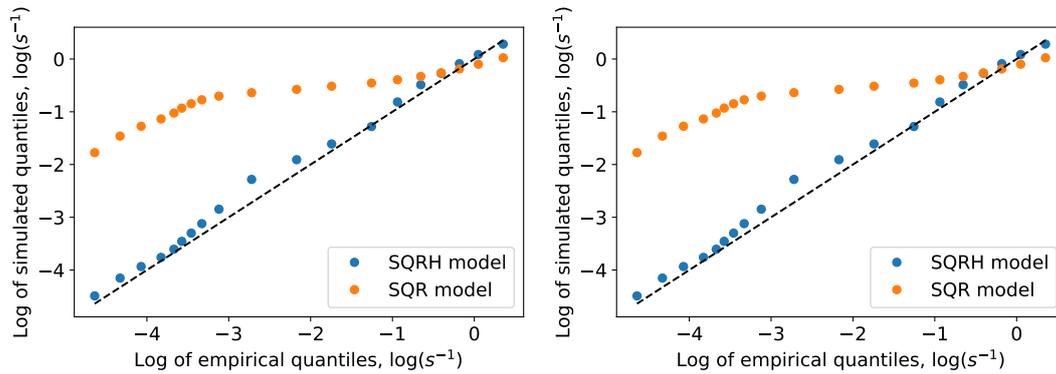


Figure 2.2: Log qq-plot of inter-event times. Log of quantiles of inter-events times simulated by model (horizontal) is plotted against log of empirical quantiles (vertical). Left: Bund future. Right: DAX future. Note that, since qq-plots of the SQH model can be hardly distinguished from those of SQRH model, we choose to not display it.

Bund			
	Difference of log-likelihood	df	$p$ -value
$H_0 = \text{SQR}, H_1 = \text{SQRH}$	$3.7 \cdot 10^8$	27	$< 10^{-16}$
$H_0 = \text{SQH}, H_1 = \text{SQRH}$	$6.0 \cdot 10^5$	447	$< 10^{-16}$
DAX			
	Difference of log-likelihood	df	$p$ -value
$H_0 = \text{SQR}, H_1 = \text{SQRH}$	$1.8 \cdot 10^7$	27	$< 10^{-16}$
$H_0 = \text{SQH}, H_1 = \text{SQRH}$	$2.4 \cdot 10^5$	72	$< 10^{-16}$

Table 2.3: Likelihood ratio test statistic and  $p$ -values for the case where the null hypothesis is the SQR model and for the case where the null hypothesis is a SQH model. The “Degrees of freedom” (“df”) value indicates the difference in the number of parameters between the two models.

The results presented in this section suggest that both LOB-state (i.e., queue size) dependence and memory effects due to correlation in the order flow are relevant variables that need to be taken into account in order to build a faithful model for the order book dynamics. Crucially, adding an order flow dependence in the form of a Hawkes term dramatically increases the model likelihood as well as its capability of reproducing the observed inter-event time distribution.

**State dependency and Hawkes matrix empirical estimations** In Figure 2.3 we report the estimated parameters  $\mu(q)$  for the SQR model, while in Figure 2.4 we plot the analogous quantities for the SQRH model (2.2.5). Note that since, as shown in Fig. 2.1, the number of events rapidly drops as the queue size  $q$  increases, the calibration results of  $\mu(q)$  are not reliable for large  $q$ . For that reason, we only show  $\mu(q)$  estimations over the first half of the full available range of  $q$ . Within this domain, given the observed number of events, estimation errors can be evaluated from numerical simulations or using the inverse of the Hessian matrix of the log-likelihood. Both methods provide a magnitude of the relative estimation error around a few percent. We can make two general remarks while comparing these plots. First, we note that the dependence on the queue size captured by the two models is roughly concordant, in that the functions  $\mu^\ell(q)$  have similar shapes in both models. However let us remark that the values estimated within the SQRH model are much smaller, indicating that a large part of the intensity is now explained by the self- and cross-exciting Hawkes components (see the discussion below). As in [87], we observe a decreasing rate of market order arrivals as the queue size increases. This can be explained by the fact that agents tend to consume liquidity faster as this liquidity becomes rare. We also find that, when  $q(t)$  is large enough, the rate of cancellation is an increasing function of the queue size. This is an expected

feature assumed in most former LOB models (see e.g., [132, 45]), since cancellations are more likely to occur when they are many active limit orders. As shown in Appendix 2.5, this behavior ensures the ergodicity of the queue process. Let us finally notice, that unlike the observed behavior in [87] on specific stocks, we don't observe that the intensity of limit order insertion is almost independent of the queue size. It is rather a decreasing function of the queue size probably reflecting a lesser quest for priority when  $q$  is large.

It is also interesting to look at the quantity

$$e^\ell(q) = 1 - \frac{\mu^\ell(q)}{\Lambda^\ell(q)} \quad (2.2.17)$$

where

$$\Lambda^\ell(q) = \mathbb{E}\lambda^\ell(t)|q(t^-) = q \quad (2.2.18)$$

is the average intensity in a given state  $q$ .  $e^\ell(q)$  corresponds to the fraction of the total average intensity explained by the endogenous self- and cross-exciting mechanism as a function of the queue size  $q$ . While  $\Lambda^\ell(q)$ , in the case of the SQR model, is directly provided by the parameter  $\mu^\ell(q)$ , for the SQRH model it is given by the contribution of both the baseline intensity  $\mu^\ell(q)$  and the Hawkes interactions. Unlike standard multivariate Hawkes processes, the SQRH model does not admit a closed form formula of  $\Lambda^\ell(q)$  from its parameters. Therefore, while  $e^\ell(q)$  is trivially zero for the SQR model, we resort to numerical computation of  $\Lambda^\ell(q)$  in order to compute  $e^\ell(q)$  for the SQRH model.

The results are shown in Figure 2.5, where we have plotted the estimated  $e^\ell(q)$  for all types of orders and for both the Bund (top panels) and the DAX (bottom panels) futures. Overall we see that a large part, from 60% to 80% of the total average intensity, is explained by the self- and cross-exciting effect. We note that for cancel and market orders, the intensity is maximally explained by the Hawkes term when the queue is small. This is likely the result of persistence in the order flow, captured by the self-exciting term in the Hawkes model, which for market and cancel orders leads to a depletion of the queue. This explanation is corroborated by the observation that the opposite effect is found for limit orders, namely a higher endogeneity for higher values of  $q$ .

To complete the analysis of the QRH model results, in Figure 2.6 we plot in a color map the Hawkes kernel norms  $|\phi^{\ell m}| = \int_0^\infty \phi^{\ell m}(t) dt$ . Note that in our setting these quantities are simply given by  $|\phi^{\ell m}| = \sum_{u=1}^U \alpha_u^{\ell m}$ . As discussed in [16], these quantities represent the average direct effect of an event of type  $m$  (columns) over the intensity of type  $\ell$  (rows) events. Hawkes kernel matrices of order book events have been extensively studied in [13, 126]. Here, we note that despite the addition of the queue-dependent term, we recover many of the features already observed in previous studies, such as the strong diagonal component corresponding to self-excitation, likely the result of correlation in the order flow induced by order splitting strategies. We also confirm that market orders influence liquidity much more than the opposite effect. In particular, since here we

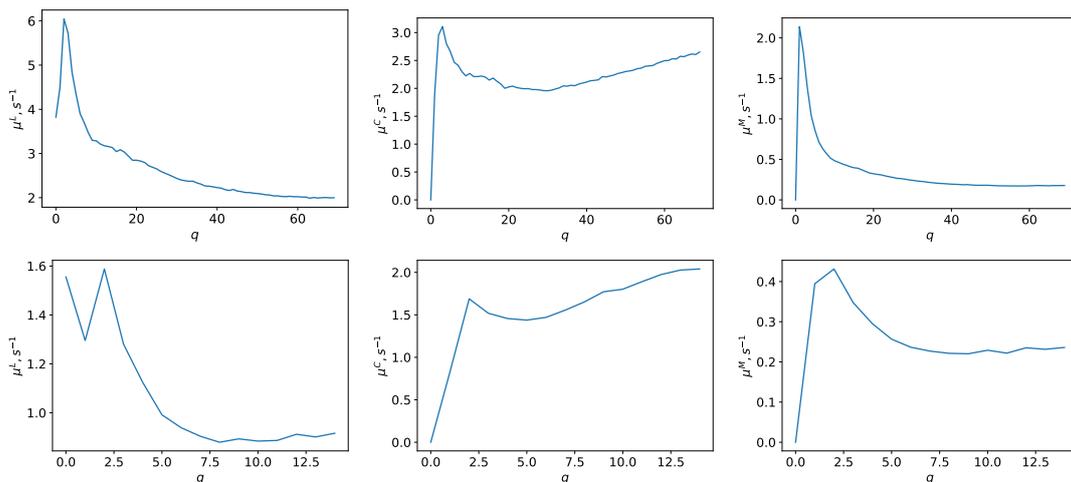


Figure 2.3: From left to right: estimated values  $\mu_q$  for limit order insertion, limit order cancellation and market orders, SQR model. Top row: Bund future. Bottom row: DAX future.

look at interactions on the same side of the book, we note that market orders have on average an exciting effect on cancellations. As observed in the aforementioned studies, flows of market orders at a given price signal that the “true” price is closer to that side and therefore liquidity adapts, with outstanding orders being canceled in order not to be adversely selected.

**Equilibrium and empirical queue size distributions** In [87], the authors emphasized that the QR model provides a simple framework to account for the observed queue size distributions in the order book. For that purpose, they have shown that the model invariant distribution fits quite well the empirical laws notably at the first bid/ask levels. In Appendix 2.5.1, we show that, under some conditions that appear to be empirically fulfilled, the SQRH model is also an ergodic process and the queue size can thus be described by its invariant distribution. Before comparing the performances of SQR and SQRH models with respect to their prediction of the equilibrium queue size distribution, let us emphasize that some caution is needed when addressing this issue. Indeed, this distribution, even if reached exponentially fast, does not necessarily correspond to the empirically observed queue law since when the queue is empty, the reference price has a non-vanishing probability to change. This directly implies that for small values of the queue size, the invariant distribution is not supposed to account for the observed values from snapshots of the empirical book state. Moreover, since the initial queue size has no reason to be drawn with the invariant distribution, this law is pertinent only after a short delay that has to be compared to the length of each realization, i.e., the time period between two changes of the reference price. The exponential rate involved in the ergodic theorem is however hard to estimate. One can use a proxy

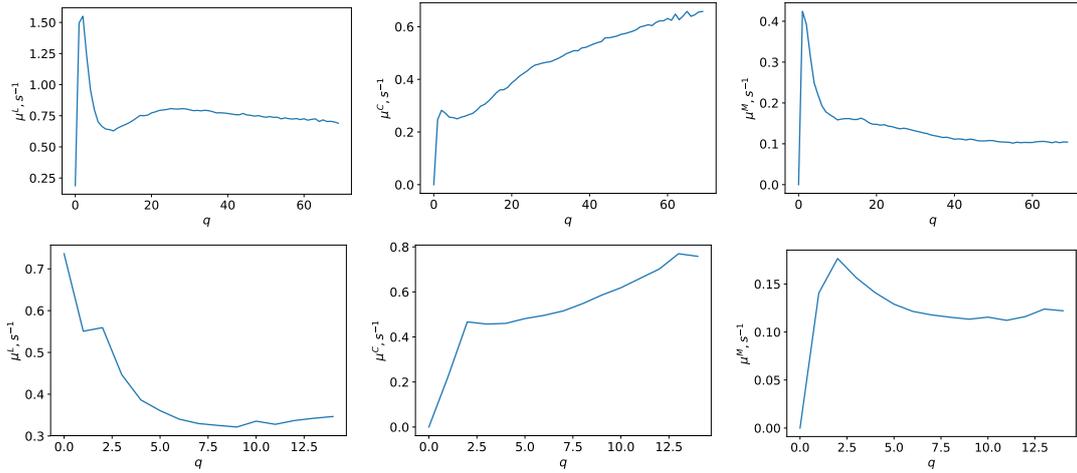


Figure 2.4: From left to right: estimated values  $\mu_q$  for limit order insertion, limit order cancellation and market orders, SQRH model. Top row: Bund future. Bottom row: DAX future.

as given for example by the exponential decay of empirical queue size autocorrelation function. That is an alternative measure of a mixing coefficient that can be, under some conditions, related to the distribution relaxation time [40]. If one assumes that the decay of autocorrelation of the queue size takes the form  $\rho(t) = a \exp^{-t/\tau_c}$ , we find empirically that  $\tau_c \simeq 15$  s for the Bund future and  $\tau_c \simeq 2$  s for the DAX. For both assets, these correlation characteristic scales have to be compared with the average realization length, namely  $\tau_m \simeq 100$  s for the Bund and  $\tau_m \simeq 16$  s for the DAX. Since in both cases we have  $\tau_c \ll \tau_m$ , it is likely that the invariant distribution is pertinent to account for the queue size distribution as observed at randomly chosen times.

With previous observations in mind, we now proceed with the comparison of the empirical queue distribution, measured by taking snapshots of the book every 30 s and the invariant distributions produced by the SQR and SQRH models. Since we do not have any explicit formula for the SQRH model, we estimate the invariant distribution of  $q(t)$  by performing a simulation over a long time period. The invariant measure of the SQR model can be directly deduced from the estimations of  $\mu^\ell(q)$  in Figure 2.3 using the analytical formula in Sec. 2.3.3 of [87]. The plots the both SQR and SQRH invariant measures together with the empirical queue size distribution for both Bund and DAX are reported in Figure 2.7. First of all, we observe that the SQRH model provides in both cases a better fit of empirical data, notably in the tail region, than the SQR model. The latter is particularly far from the observed distribution in the large tick case of the Bund data. Its performance for the smaller tick asset (DAX) is closer to the results reported in [87] for stock data. Beyond the fact that this striking difference between large and small tick assets is hard to explain (though the analytical formula in [87] shows that the overall shape of the distribution can vary quite drastically when on

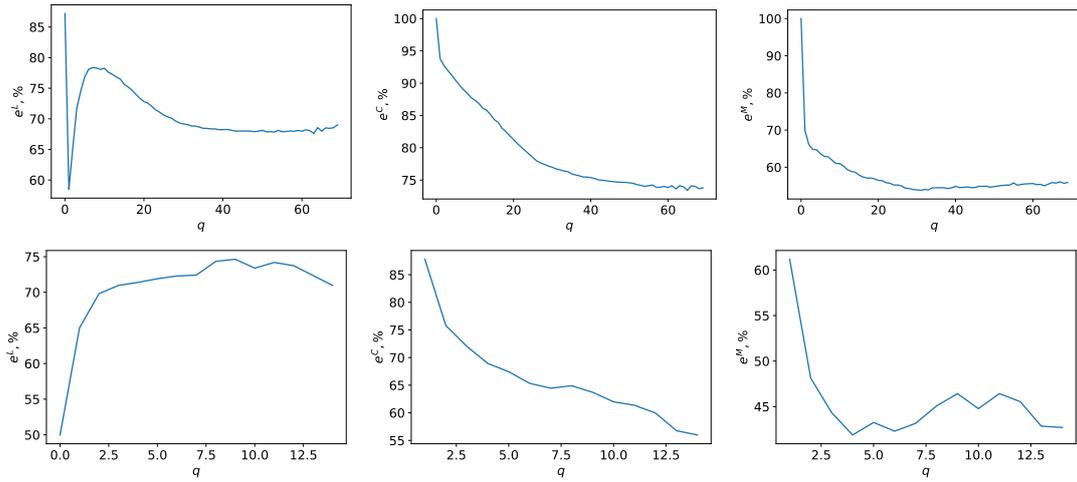


Figure 2.5: From left to right: Endogenous fraction for limit order insertion,  $e^{\ell}(q)$  as defined in Eq. (2.2.17), for limit ( $\ell = L$ ), cancellation ( $\ell = C$ ) and market ( $\ell = M$ ) orders by SQRH model. Top row: Bund future. Bottom row: DAX future.

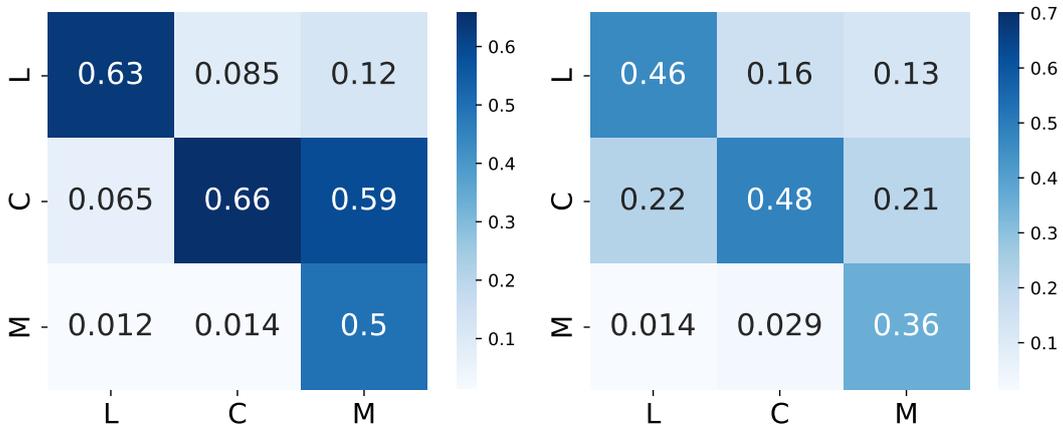


Figure 2.6: Kernel norm matrices for Bund future (left) and DAX future (right).

changes the respective behavior of the  $\mu^{\ell}(q)$  functions) our findings show that accounting for the Hawkes self-interaction within a queue reactive model is important not only to describe correctly the order flow dynamics but also to provide a better model for the queue size distributions.

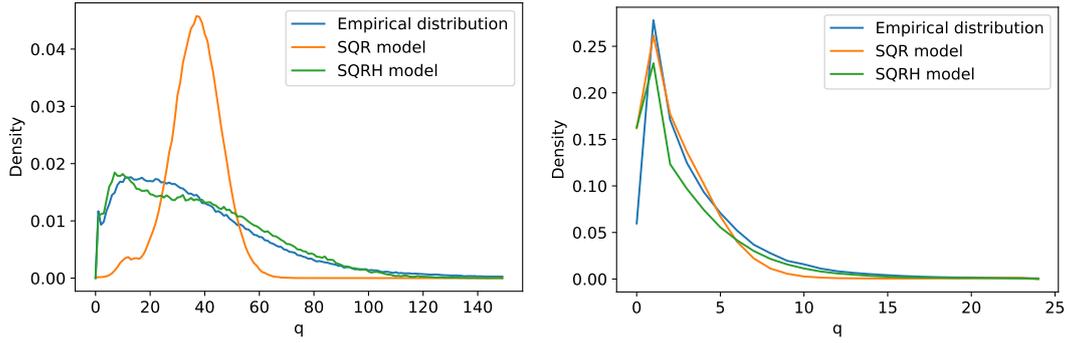


Figure 2.7: Comparison of the invariant distributions of the SQR and SQRH models with the empirical one. Left: Bund future. Right: DAX future.

We notice furthermore that the distribution of queue size simulated by the SQRH model deviates from the empirical distribution, especially around states where the queue is small. As discussed previously, when the queue is empty, the reference price has a large probability to change and therefore the corresponding empirical sample stops. This results in a statistical bias with respect to the “theoretical model” where such event type does not exist since when the queues are empty nothing happens until a new limit order arrives. The states with low queue values are therefore more likely to be visited in the model than in empirical observations.

## 2.3 Summary and prospects

In this paper, we introduced a “Single Queue Reactive Hawkes model” with the ambition to improve respectively the approach of Huang et al. [87] on the queue reactive nature of the LOB dynamics and the model of Bacry et al. [13]. We show that such models can be easily calibrated within a parametric approach. Our empirical findings on two different future assets from Eurex, namely Bund and DAX order book data, suggest that both queue reactive and past order flow dependencies are relevant to account for the occurrence likelihood of future order book events. The SQRH model outperforms a pure Hawkes model as well as a pure queue-reactive one in terms of goodness of fit. As far as the SQRH model is concerned, our framework allows one to remain within the framework of Markov processes that has ergodic properties so, along the same line as in the Huang et al. approach, we can define and estimate the queue size distribution associated with the invariant measure of the model. The SQRH also leads us to refine Bacry et al. findings [13] by accounting for the states of the LOB. Our results suggest that by putting the queue-dependency property in the intensity function, a better description of order flows is achieved. Furthermore, the ergodicity property of this model can be proved under reasonable assumptions, which allows the SQRH model to be adopted for simulation purposes. Besides considering various applications of these models to design and optimize high-frequency trading strategies, one fundamental challenge is to

extend our approach in order to account for several queue interactions and describe the book dynamics up to a given depth. It could be also interesting to generalize definition (2.1.3) to cases where both exogenous intensities and Hawkes kernel may depend on the queue sizes. From a mathematical point of view, a deeper understanding of the stability and stationarity conditions for queue-dependent Hawkes models remains to be developed. Meanwhile, some substantial simplifications we made in this study could also be removed in order to have an even more realistic model such as, for instance, dropping the assumption of unitary order sizes by adopting a similar approach to [102]. More fundamentally, a clear understanding of the observed shapes of the exogenous and endogenous intensities and the nature of queue dependency property in terms of the (rational) behavior of various market participants remain open questions.

## 2.4 Acknowledgements

This work was funded in part by the French government under the management of Agence Nationale de la Recherche as part of the “Investissements d’avenir” program, reference ANR-19-P3IA-0001 (PRAIRIE 3IA Institute).

## 2.5 Appendix

### 2.5.1 Proof of the existence of invariant distribution

In this section, we prove the existence of the invariant distribution of the SQRH model. The ergodicity allows for approximating the empirical distribution of queue size by simulating the SQRH model for a sufficiently long time. The proof is made via constructing a Lyapunov function. First, we define

$$o_{\ell mu}(t) = \int_0^t \alpha_u^{\ell m} e^{-\beta_u(t-s)} dN_s^m. \quad (2.5.19)$$

Then by combining (2.2.5) and (3.2.3), the intensity function could be written in the form

$$\lambda^\ell(t) = \left( \mu^\ell(q(t^-)) + \sum_{m \in \{L, M, C\}} \sum_{u=1}^U o_{\ell mu}(t) \right) \mathbb{1}_{(\ell=L) \vee (q(t^-) > 0)}. \quad (2.5.20)$$

We take two empirically fulfilled assumption introduced in [4, 3], that there exist two positive constants  $c_-$  and  $c^+$  satisfying

$$\sum_{\ell \neq L} \mu^\ell(q) \geq c_- q \text{ and } \mu^L \leq c^+.$$

These assumptions limit the arrival rate of limit order while guaranteeing that existing volume is consumed at a rate which is at least proportional to queue size. Under

such assumptions,  $\lim_{t \rightarrow \infty} \mathbb{E}[q(t) < \infty]$ . At moment  $t$ , the queue size is determined by the sum of past order flows:

$$q(t) = \sum_{\ell=L} N_\ell - \sum_{\ell \neq L} N_\ell. \quad (2.5.21)$$

We note  $\vec{o}(t)$  be the vector obtained as a vertical stacking of the components  $o_{\ell mu}(t)$  for all  $(\ell, m, u) \in \{L, M, C\}^2 \times \{1, \dots, U\}$ . Then it is not difficult to verify that the vector process  $(q, \vec{o})^T$  is Markovian. The rest part is about constructing a Lyapunov function for  $(q, \vec{o})^T$ . We first construct two Lyapunov functions separately for  $q$  and  $\vec{o}$ , then we construct the Lyapunov functions for  $(q, \vec{o})^T$  by combining the two.

### Lyapunov function for $\vec{o}$

We first prepare the differential form of  $\vec{o}(t)$ :

$$do_{\ell mu}(t) = -\beta_u o_{\ell mu}(t) dt + \alpha_u^{\ell m} dN_t^m. \quad (2.5.22)$$

For any arbitrary suitable function  $F$  which maps  $\mathbb{R}^{2D+U}$  to  $\mathbb{R}$ , the infinitesimal generator is:

$$\mathcal{L}F(\vec{o}) = \sum_m \lambda_m (F(\vec{o} + \Delta_m(\vec{o})) - F(\vec{o})) - \sum_{\ell, m, u} \beta_u o_{\ell mu} \frac{\partial F}{\partial o_{\ell mu}}, \quad (2.5.23)$$

where  $\lambda_m$  is the probability for an event to happen in dimension  $m$  of the point process  $\vec{N}$ , and  $\Delta_m(\vec{o})$  is the jump of  $\vec{o}$  caused by this event. We then define the matrix  $A$  as:

$$A_{\ell m} = \sum_u \frac{\alpha_u^{\ell m}}{\beta_u}. \quad (2.5.24)$$

We further assume that if  $f(q)$  is set to constant value, SQRH model corresponds to a stable Hawkes process, i.e., the maximal eigenvalue  $\kappa$  of  $A$  satisfies  $0 < \kappa < 1$ . Notice that, according to Perron-Frobenius theorem, the associated eigenvector  $\vec{\epsilon}$  of  $\kappa$  satisfies  $\forall l, \epsilon_l > 0$ .

We then note

$$\delta_{\ell mu} := \delta_{\ell u} = \frac{\epsilon_\ell}{\beta_u}, \quad (2.5.25)$$

and we choose function  $V_1$  of  $\vec{o}$  as:

$$V_1(\vec{o}) = \sum_{\ell, m, u} \delta_{\ell mu} o_{\ell mu}. \quad (2.5.26)$$

With the preparations above, we could verify that  $\mathcal{L}V_1(\vec{o})$  is a Lyapunov function

for  $\vec{o}$ , i.e.

$$\begin{aligned}
\mathcal{L}V_1(\vec{o}) &= \sum_m (\mu_m + \sum_{p,q} o_{mpq}) \mathbb{1}_{m=0 \vee q(t) > 0} \sum_{\ell,u} \delta_{\ell mu} \alpha_u^{\ell m} - \sum_{\ell,m,u} \beta_u o_{\ell mu} \delta_{\ell mu} \\
&\leq \sum_m (\mu_m + \sum_{p,q} o_{mpq}) \sum_{\ell,u} \delta_{\ell mu} \alpha_u^{\ell m} - \sum_{\ell,m,u} \beta_u o_{\ell mu} \delta_{\ell mu} \\
&= \sum_{\ell,m,u} \mu_m \delta_{\ell mu} \alpha_u^{\ell m} + \sum_m (\sum_{p,q} o_{mpq}) (\sum_{\ell,u} \frac{\epsilon_\ell}{\beta_u} \alpha_u^{\ell m}) - \sum_{\ell,m,u} \beta_u o_{\ell mu} \delta_{\ell mu} \\
&= C_1 + \sum_m (\sum_{p,q} o_{mpq}) (\sum_\ell \epsilon_\ell \sum_u \frac{\alpha_u^{\ell m}}{\beta_u}) - \sum_{\ell,m,u} \beta_u o_{\ell mu} \delta_{\ell mu} \\
&= C_1 + \sum_m (\sum_{p,q} o_{mpq}) \epsilon_m \kappa - \sum_{\ell,m,u} \epsilon_\ell o_{\ell mu} \\
&= C_1 - (1 - \kappa) \sum_{\ell,m,u} \epsilon_\ell o_{\ell mu} \\
&= C_1 - (1 - \kappa) \sum_{\ell,m,u} \beta_u \delta_{\ell mu} o_{\ell mu} \\
&\leq -\rho_1 V_1 + C_1,
\end{aligned} \tag{2.5.27}$$

where the constant  $\rho_1$  is chosen as

$$\rho_1 = (1 - \kappa) \inf \beta_u. \tag{2.5.28}$$

### Lyapunov function for $q$

Similarly, we first write the differential form of  $q(t)$ :

$$dq(t) = \sum_{m \in J} dN_t^m - \sum_{\ell \in I} dN_t^\ell. \tag{2.5.29}$$

We keep the same notation of  $\Delta$  as in Eq 2.5.23. For any arbitrary suitable function  $F$  who maps  $\mathbb{R}^+$  to  $\mathbb{R}$ , the infinitesimal generator is:

$$\mathcal{L}F(q) = \sum_m \lambda_m (F(q + \Delta_m(q)) - F(q)). \tag{2.5.30}$$

Next, it could be easily verified that  $V_2(q) := q$  is a Lyapunov function for  $q$ :

$$\begin{aligned}
\mathcal{L}V_2(q) &= \sum_{m \in J} \lambda_m(q) - \sum_{\ell \in I} \lambda_\ell(q) \\
&\leq \sum_{m \in J} \mu_m(q) - \sum_{\ell \in I} \mu_\ell(q) + \sum_{\ell,m,u} o_{\ell mu} \\
&\leq c^J - c^I q + \sum_{\ell,m,u} o_{\ell mu} \\
&\leq -\rho_2 q + C_2.
\end{aligned} \tag{2.5.31}$$

In the equation  $c^I, c^J$  are previously defined constants, and  $\rho$  is taken as  $\rho_2 = c^I$ . Also, it must be noted that the constant  $C_2$  depends on  $\vec{o}$  as well:

$$C_2 := c^J + \sum_{\ell,m,u} o_{\ell mu}. \tag{2.5.32}$$

### Lyapunov function for $(q, \vec{\sigma})$

The final step is to construct a Lyapunov function  $V$  for both  $q$  and  $\vec{\sigma}$ . We choose the following:

$$V(q, \vec{\sigma}) = V_2(q) + \frac{1}{\eta} V_1(\vec{\sigma}). \quad (2.5.33)$$

By combining inequalities established for  $V_1$  and  $V_2$ , and the expression 2.5.32 with  $C_2$ , direct calculation shows that:

$$\begin{aligned} \mathcal{L}V(q, \vec{\sigma}) &= \mathcal{L}V_2(q) + \frac{1}{\eta} \mathcal{L}V_1(\vec{\sigma}) \\ &\leq -\rho_2 V_2 + C_2 - \frac{\rho_1}{\eta} V_1 + \frac{C_1}{\eta} \\ &\leq -\rho_2 V_2 + \sum_{\ell, m, u} o_{\ell mu} - \frac{\rho_1}{\eta} V_1 + C' \end{aligned} \quad (2.5.34)$$

Here we can choose  $C' = c^J + \frac{C_1}{\eta}$ . Notice that since the coefficient  $\delta_{\ell mu}$ ,  $o_{\ell mu}$ , and  $\rho_1$  are all positive, there must exist an  $\eta$  who satisfies

$$\sum_{\ell, m, u} o_{\ell mu} < \frac{\rho_1}{2\eta} V_1(\vec{\sigma}) \quad (2.5.35)$$

By substituting the previous inequality into Eq 2.5.34, we can then show that  $V$  is a Lyapunov function for  $(q, \vec{\sigma})$ :

$$\begin{aligned} \mathcal{L}V(q, \vec{\sigma}) &\leq -\rho_2 V_2 - \frac{\rho_1}{2\eta} V_1 + C' \\ &\leq -\min(\rho_2, \frac{\rho_1}{2}) V + C'. \end{aligned} \quad (2.5.36)$$

Given the existence of the Lyapunov function and the geometric drift condition above, together with the assumption that the spectral radius of  $A$  is smaller than one, Theorem 6.1 in [112] guarantees that the process  $q(t)$  is ergodic. Also, it converges exponentially fast towards its unique stationary distribution.

### 2.5.2 Calculation of the log-likelihood function of SQRH model and its gradient

For SQRH model, the log-likelihood is a function of  $\mu$  and  $\alpha$ . Let's note  $t_k^\ell$  the timestamp of the  $k$ th event of type  $\ell$ , and  $N^\ell$  the total number of event of type  $\ell$ . With such notation, the log-likelihood could be expressed as:

$$\begin{aligned} L(\vec{\alpha}, \vec{\mu}) &= \sum_{\ell \in \{L, M, C\}} \sum_{k=1}^{N^\ell} \log \left( \mu^\ell(q(t_k^{\ell,-})) + \sum_{m \in \{L, M, C\}} \sum_{u=1}^U \alpha_u^{\ell m} \beta_u \int_0^{t_k} e^{-\beta_u(t-s)} dN_s^m \right) \\ &\quad - \sum_{\ell \in \{L, M, C\}} \int_0^T \left( \mu^\ell(q(t)) + \sum_{m \in \{L, M, C\}} \sum_{u=1}^U \alpha_u^{\ell m} \beta_u \int_0^s e^{-\beta_u(s-v)} dN_v^m \right) ds. \end{aligned} \quad (2.5.37)$$

Next, we introduce two auxiliary functions  $g(t)$  and  $G(t)$ , whose value don't depend on the parameters  $\vec{\mu}$  and  $\vec{\alpha}$ :

$$g_u^m(t) = \sum_{t_k^m < t} \beta_u e^{-\beta_u(t-t_k^m)}, \quad (2.5.38)$$

and

$$G_u^m(t) = \int_0^t g_u^m(s) ds = \int_0^t \int_0^s \beta_u e^{-\beta_u(s-v)} dN_v^m ds. \quad (2.5.39)$$

We note that both  $g(\cdot)$  and  $G(\cdot)$  only need to be evaluated over a grid of timestamps  $\{t_k\}_k$ , when new orders arrived. Their value is calculated using the following recurrence relation:

$$g_u^m(t_k) = \sum_{t_{k'}^m < t_k} \beta_u e^{-\beta_u(t_k-t_{k'}^m)} \quad (2.5.40)$$

$$= \sum_{t_{k'}^m < t_{k-1}} \beta_u e^{-\beta_u(t_{k-1}-t_{k'}^m)} e^{-\beta_u(t_k-t_{k-1})} + \sum_{t_{k-1} \leq t_{k'}^m < t_k} \beta_u e^{-\beta_u(t_k-t_{k'}^m)} \quad (2.5.41)$$

$$= e^{-\beta_u(t_k-t_{k-1})} g_u^m(t_{k-1}) + \beta_u e^{-\beta_u(t_k-t_{k-1})} \mathbb{1}_{type(t_{k-1}^+) = m}, \quad (2.5.42)$$

and

$$G_u^m(t_k) - G_u^m(t_{k-1}) = \int_{t_{k-1}}^{t_k} g_u^m(s) ds \quad (2.5.43)$$

$$= \frac{1 - e^{-\beta_u(t_k-t_{k-1})}}{\beta_u} g_u^m(t_{k-1}) + \left(1 - e^{-\beta_u(t_k-t_{k-1})}\right) \mathbb{1}_{type(t_{k-1}^+) = m}. \quad (2.5.44)$$

### Log-likelihood function

With the definition of  $g$  and  $G$ , the log-likelihood function  $L$  could be rewritten in the following way:

$$\begin{aligned} L(\vec{\alpha}, \vec{\mu}) &= \sum_{\ell \in \{L, M, C\}} \sum_{k=1}^{N^\ell} \log \left( \mu^\ell(q(t_k^{\ell, -})) \right) + \sum_{m \in \{L, M, C\}} \sum_{u=1}^U \alpha_u^{\ell m} g_u^m(t_k) \\ &- \sum_{\ell \in \{L, M, C\}} \sum_{k=1}^N \mu^\ell(q(t_k^-))(t_k - t_{k-1}) - \sum_{\ell \in \{L, M, C\}} \mu_{q(t_N^+)}^\ell (T - t_N) \\ &- \sum_{\ell \in \{L, M, C\}} \sum_{k=1}^N \sum_{m \in \{L, M, C\}} \sum_{u=1}^U \alpha_u^{\ell m} (G_u^m(t_k) - G_u^m(t_{k-1})) - \sum_{\ell \in \{L, M, C\}} \sum_{m \in \{L, M, C\}} \sum_{u=1}^U \alpha_u^{\ell m} (G_u^m(T) - G_u^m(t_{k-1})) \end{aligned} \quad (2.5.45)$$

By deriving the equation above against  $\mu$  and  $\alpha$ , a direct calculation gives:

$$\frac{\partial L}{\partial \mu^\ell(q)} = \sum_{k=1}^{N^\ell} \frac{\mathbb{1}_{q(t_k^{\ell,-})=q}}{\mu^\ell(q(t_k^{\ell,-})) + \sum_{m=1}^D \sum_{u=1}^U \alpha_u^{\ell m} g_u^m(t_k^\ell)} - \sum_{k=1}^N \mathbb{1}_{q(t_k^-)=q}(t_k - t_{k-1}) - \mathbb{1}_{q(t_N^+)=q}(T - t_N), \quad (2.5.46)$$

and

$$\frac{\partial L}{\partial \alpha_u^{\ell m}} = \sum_{k=1}^{N^\ell} \frac{g_u^m(t_k^\ell)}{\mu^\ell(q(t_k^{\ell,-})) + \sum_{m=1}^D \sum_{u=1}^U \alpha_u^{\ell m} g_u^m(t_k^\ell)} - \sum_{k=1}^N (G_u^m(t_k) - G_u^m(t_{k-1})) - (G_u^m(T) - G_u^m(t_N)). \quad (2.5.47)$$

### 2.5.3 Choice of parameters $U$ and $\beta_m$

In SQRH model, both  $U$  and  $\beta = \{\beta_1, \beta_2, \dots, \beta_U\}$  are treated as hyper-parameters. With pre-determined  $\beta$ , the log-likelihood of the SQRH model is convex and could be hence optimized. The ideal way to determine the optimal  $\beta$  with  $U$  components is to search over a  $U$ -dimensional grid while maximizing the log-likelihood for increasing values  $U = 1, 2, 3, \dots$ . However, a search of this kind could be very costly in computation time. Therefore, we opted to use the following iterative approach instead:

1. We search  $\beta_1$  over a logarithmic grid to maximize the log-likelihood of the SQRH model over a given data set and take  $\beta = \{\beta_1\}$ .
2. While keeping  $\beta_1$  fixed, we search for a second component  $\beta_2$  over a logarithmic grid and maximise the log-likelihood of SQRH model with  $\beta = \{\beta_1, \beta_2\}$ .
3. We then fix  $\beta_2$  and adjust  $\beta_1$  again to further maximize the log-likelihood of SQRH model. Then we iteratively adjust  $\beta_2$  (resp.  $\beta_1$ ) while keeping other components in  $\beta$  fixed. The log-likelihood converges after several iterations. Then we take  $\beta = \{\beta_1, \beta_2\}$ .
4. Similarly, we repeat step 2 and step 3 to add a third component as  $\beta_3$  into  $\beta$ . This process could be repeated to obtain a  $\beta$  with an arbitrary number of components.

We point out that adding more components in  $\beta$  improves the log-likelihood. However, it also increases the risk of over-fitting. Meanwhile, computation time increases dramatically with the number of components. Our numerical experiment suggested that adopting  $U = 3$  components in  $\beta$  generates sufficiently good performance in terms of log-likelihood while keeping the calibration time under a reasonable scale.

## Chapter 3

# A Queue Reactive Hawkes model for the best limits of the order book

### 3.1 Introduction

The model presented in this chapter can be seen as a further development from the SQRH model presented in Chapter 2, which has been developed with the same ambition: developing a model for the Limit Order Book (LOB) where dependencies on both past order flow and the state of the LOB are present. We will not reiterate the motivation to do so, which would be largely redundant with the introduction of the previous chapter.

Let us recall that in Chapter 2 we introduced a “Single Queue Reactive Hawkes” (SQRH) model with the ambition to improve respectively the approach of Huang et al. [87] on the queue reactive nature of the LOB dynamics and the model of Bacry et al. [13]. We restricted this model to a single (best) queue modeling in order to be able to prove the existence of an invariant distribution. Like the model in [87], whenever the price corresponding to this best limit changed, the model was reset. For this reason, SQRH does not account for the cross dynamic between the two best limits (e.g., the best bid and the best ask queue), nor does it account for the changes in their corresponding price. This chapter is devoted to a model that tackles these drawbacks.

In this chapter, we present a more sophisticated model than SQRH. It considers the order flows on the two best limit queues, their interactions, and the changes in the prices of the best bid and the best bid queue. This model will be referred to as the “Queue Reactive Hawkes” (QRH model).

As we did for the SQRH model, the QRH model will involve a combination of a “Queue Reactive” component (as in [87]), i.e., a dependence on the two best queue sizes at the current time, and a Hawkes component taking into account the dependency on the past order flows on these two queues. Moreover, in order to take into account the

change in prices, the resetting of the model will only occur at market close time.

## 3.2 The Queue Reactive Hawkes (QRH) model

### 3.2.1 Introduction to the QRH model

As explained in the introduction, the QRH model is intended to capture the dynamics of the two best limit queues (best ask and best bid) of the LOB: including both the dynamics of the queue sizes and the dynamics of the midprice, where the midprice is defined as the average of the best bid price and the best ask price.

Let us point out that, contrary to the SQRH model presented in the previous chapter and the QR model in [87], we define the best ask (resp. bid) queue as the queue corresponding to the lowest (resp. highest) available price (i.e., non-empty queue) on the ask (resp. bid) side of the LOB. Thus, following this definition, the size of the best ask (resp. bid) queue can never reach 0. Under this setting, the QRH model does not make any use of the referenced price  $p_{\text{ref}}$  as defined in the previous chapter or as in [87]. As presented below, we track the changes in the midprice by monitoring particular events. Moreover, this model no longer needs to be reset regularly. The resetting only occurs when the market closes.

For the QRH model, we consider eight event types at the best ask of best bid queues of a LOB, namely

$P^+$  ( $P^-$ ) for events that move the midprice up (down),

$L^a$  ( $L^b$ ) for limit orders at the best ask (bid) that do not change the midprice,

$C^a$  ( $C^b$ ) for cancellations at the best ask (bid) that do not change the midprice,

$M^a$  ( $M^b$ ) for market orders at the best ask (bid) that do not change the midprice.

For each  $\ell \in \{P^+, P^-, L^a, L^b, C^a, C^b, M^a, M^b\}$ , let us define  $N_t^\ell$  as the counting process associated with events of type  $\ell$  and  $\lambda^\ell(t)$  the associated conditional intensity.

The authors in [13] consider a multivariate Hawkes model where each event type can influence, and be influenced by, the others so that the conditional intensities read:

$$\lambda^\ell(t) = \mu^\ell + \sum_m \int_0^t \phi^{\ell m}(t-s) dN_s^m, \quad (3.2.1)$$

where  $\ell$  and  $m$  can take any value in  $\{P^+, P^-, L^a, L^b, C^a, C^b, M^a, M^b\}$ . In their work, the kernels  $\phi^{\ell m}$  are estimated by the non-parametric estimation method first described in [18]. This model allows the authors to highlight the rich influence structure between events in a limit order book, including the high-frequency midprice reversion and the persistent autocorrelation in the order flow determined by order splitting strategies, but also some more refined market-maker induced dynamics (see [13]).

In order to add queue-dependency to this Hawkes process based approach, we proceed in much the same way as we did in the previous chapter, except that the order book state is no longer represented by a single queue size  $q$  (that represented either the best ask or the best bid quantity) but by both the best bid and the best ask queue sizes  $(q_a, q_b)$ . Let us point out that now the queue sizes  $q_a$  and  $q_b$ , by definition, never reach zero. Moreover, each process  $dN^\ell$  encodes the full history of all the events of type  $\ell$  arriving at the corresponding queue, which is independent of changes in the best ask/bid prices.

So adding queue-dependency to such a model calls for a mechanism that can modulate (as a function of the order book state) not only the exogenous intensity  $\mu^\ell$  (as in the previous model) but also the convolutional part of the intensity. To illustrate this point, we can mention a situation where the Hawkes kernel matrix should explicitly depend on the state of queues. For instance,  $P^+$  events are supposed very unlikely to occur (for a large tick asset) when the size of the ask queue is large, and the size of the bid queue is small.

The QRH model corresponds to the simplest possibility where both the exogenous and the self-exciting part of the intensity share the same multiplicative dependence on the states. We introduce the following model,

$$\lambda^\ell(t) = f^\ell(q_a(t), q_b(t)) \left( \mu^\ell + \sum_m \int_0^t \phi^{\ell m}(t-s) dN_s^m \right), \quad (QRH) \quad (3.2.2)$$

where the functions  $f^\ell$  that encode the dependence on the order book states, modulate not only the exogenous intensity but also the Hawkes term. Let us notice that, unlike the SQRH model, the QRH model is mainly a model for order flows. We disregard the effect of the order flow on the queue sizes and consider queue sizes as (observable) exogenous variables.

As for the SQRH model, we choose a parametric form for the kernels  $\phi^{\ell m}$  and in particular we adopt the same exponential-sum specification

$$\phi^{\ell m}(t) = \sum_{u=1}^U \alpha_u^{\ell m} \beta_u e^{-\beta_u(t-s)} \quad (3.2.3)$$

where  $\alpha_u^{\ell m}$  are parameters of the model and  $\beta_u, U$  are suitably chosen hyper-parameters. To facilitate the notation, we use  $\vec{\mu}$  and  $\vec{\alpha}$  to represent all  $\mu^\ell$  and  $\alpha_u^{\ell m}$ . With such parametrization, we further note  $\theta = (\vec{\mu}, \vec{\alpha})$ .

### 3.2.2 From log-likelihood estimation to least square estimation

Once the parametric form (3.2.3) for the kernels has been specified, the model can be estimated using MLE. Although the QRH model differs from the SQRH model, the calculation of its log-likelihood function and its gradients follows the same track presented in Appendix 3.4.1, with only some trivial modifications. Moreover, thanks to the chosen

parametrization, the log-likelihood is again a convex function of the parameters, thus guaranteeing the existence of a global optimum.

Since the number of states of the LOB grows quadratically fast with the number of states considered per side, considering every possible state (i.e., every possible queue size) is unrealistic. To limit the number of configurations, we consider the limit order book to be only in one of the states, i.e.  $(q_a, q_b) = (i, j)$  with a limited number of  $i$  and  $j$ . This setting leads to a piecewise constant function  $f^\ell$  noted by  $f(q_a^i, q_b^j)$ . Finally, we choose the normalization (note that any other normalization would be equivalent up to a rescaling of the kernels and of the  $\mu^\ell$  in Eq. (3.2.2)) that

$$f^\ell(q_a^1, q_b^1) = 1. \quad (3.2.4)$$

In the following context, we adopt a specified partition scheme for  $i$  and  $j$ . For both the ask side and the bid side of the asset,  $i$  and  $j$  are determined in the same way. Let's take ask side for example. First, there are in total  $N$  states corresponding to  $N$  intervals, each is of length  $m$ .  $N$  and  $m$  are integers, and they are chosen as power of 2 in a way adapted to asset.

$$I_m = \begin{cases} [m * s, (m + 1) * s) & 0 \leq m < N - 1 \\ [m * s, \infty) & m = N - 1 \end{cases} \quad (3.2.5)$$

If the queue size satisfies  $q_a \in I_m$ , we take  $i = m$ . For Bund future, we choose  $N = 16$  and  $m = 64$ . For DAX future, we choose  $N = 8$  and  $m = 2$ . It is worth mentioning that our experiments suggest that finer binning with large  $N$  will improve fitting performance measured by the loss function. However, we are limited by computation resources to solve such calibration problems.

We calibrate the so-obtained model using MLE with the same dataset used in the previous section. As shown in Tables 3.1 and 3.2, our model outperforms the pure-Hawkes model introduced in [13] in terms of goodness of fit. In particular, a likelihood ratio test rejects the null hypothesis of a pure-Hawkes model with a  $p$ -value  $< 10^{-16}$ .

Bund				
	$L$	AIC	BIC	# parameters
QRH	$5.348 \times 10^8$	$-1.070 \times 10^9$	$-1.070 \times 10^9$	400
Hawkes	$5.200 \times 10^8$	$-1.040 \times 10^9$	$-1.040 \times 10^9$	200
DAX				
	$L$	AIC	BIC	# parameters
QRH	$4.626 \times 10^8$	$-9.253 \times 10^8$	$-9.253 \times 10^8$	400
Hawkes	$4.488 \times 10^8$	$-8.976 \times 10^8$	$-8.976 \times 10^8$	200

Table 3.1: Log-likelihood, AIC, and BIC values for the QRH model (defined by Eq.(3.2.2)) and the Hawkes model (defined in [13]) for Bund and DAX data.

Bund			
	LR	df	$p$ -value
$H_0 = \text{Hawkes}, H_1 = \text{QRH}$	$2.9 \cdot 10^7$	200	$< 10^{-16}$
DAX			
	LR	df	$p$ -value
$H_0 = \text{Hawkes}, H_1 = \text{QRH}$	$2.8 \cdot 10^7$	200	$< 10^{-16}$

Table 3.2: Likelihood ratio test statistic and  $p$ -values for the case where the null hypothesis is the QRH model (defined by Eq.(3.2.2)) and for the case where the null hypothesis is the Hawkes model (defined in [13]).

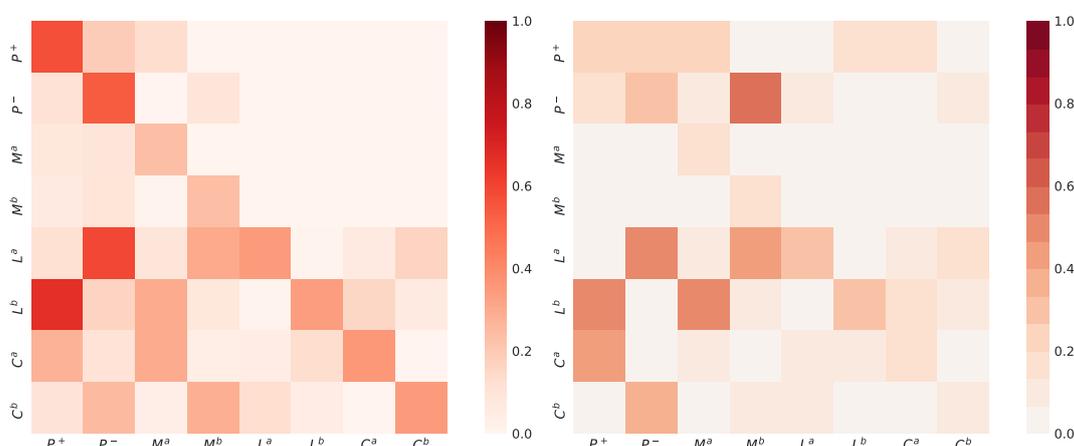


Figure 3.1: MLE Kernel norm matrices  $\int \phi^{\ell m}(t) dt$  for Bund future (left) and DAX future (right) for the QRH model (defined by Eq.(3.2.2)).

A detailed discussion on the so-obtained MLE results is in Section 3.2.3. For simplicity, let us first put apart the effect of the queue dependency and only look at the kernel estimation.

Figure 3.1 represents the so-obtained kernel norm matrices  $\{\int \phi^{\ell m}(t) dt\}_{\ell m}$  given by our model. This figure has to be compared with the matrix obtained in Figure 4 of [13] for which the same model (without queue dependency) has been used. Let us point out that in order to obtain this matrix, [13] performed a non-parametric estimation which allows negative values for the kernels and consequently negative values for some elements of the matrix  $\{\int \phi^{\ell m}(t) dt\}_{\ell m}$ . These negative values account for inhibition dynamics, i.e., decreasing the intensity of a given type of event. One could show that negative values for a kernel could lead at a finite time (and with a non zero probability) to some situation where the sum (3.2.2) is negative leading to a negative intensity which, of course, does not make any sense. In order to circumvent this difficulty, it is common

to replace equation (3.2.2) by a non-linear equation of the form

$$\lambda^\ell(t) = f^\ell(q_a(t), q_b(t)) \left( \mu^\ell + \sum_m \int_0^t \phi^{\ell m}(t-s) dN_s^m \right)^+, \quad (3.2.6)$$

where the operator  $(\cdot)^+ = \max(\cdot, 0)$ . However, in the framework of this nonlinear Hawkes model, MLE becomes intractable since it is no longer a convex problem. As explained in [77], this problem is tractable in some sense when considering the least square approach (see also [127] as an example of least square estimation with negatively valued kernels).

Thus, the least square estimations allow negatively valued kernels whereas, in the MLE framework as presented above, we forced all the kernels (i.e., all the  $\alpha_u^{\ell m}$ 's in Eq. (3.2.3)) to be positively valued. This mainly explains the differences found between Figure 4 in [13] and our Figure 3.1.

For the sake of just naming one striking difference, in [13], the kernel integral  $\int \phi^{L^b P^-}$  (resp.  $\int \phi^{L^a P^+}$ ) is found to be strongly negative. Actually, as seen in Appendix B.6 in [13], the kernel itself  $\phi^{L^b P^-}$  (resp.  $\phi^{L^a P^+}$ ) is mostly negative at all time scales. This fact can be seen as a natural dynamic induced by market makers: when the price goes down, the efficient price is closer to the best bid price; thus, fewer limit orders are placed on the bid size (the gain is small compared to sending an aggressive order) and more limit orders are placed on the ask side.

Let us point out that forcing the kernels to have only positive values (i.e., forcing all the  $\alpha_u^{\ell m}$  to be positive as assumed when performing MLE) will a priori not only lead to highly biased values for kernels with negative values but is likely also to induce high bias for kernels with only positive values since the estimation performed is a joint estimation of all the kernels involving intricate relationships between these kernels (see [18] for examples of such biases).

Consequently, it appears that one should use least square based estimation rather than MLE. Details about least square estimation can be found in Appendix 3.4.1. It consists in minimizing  $R(\theta)$  as defined by

$$R(\theta) = \sum_{\ell=1}^D R^\ell(\theta), \quad \text{with} \quad R^\ell(\theta) = \int_0^T \lambda_\ell^2(t; \theta | \mathcal{F}_t) dt - \sum_{k=1}^{N^\ell} \lambda_\ell(t_k^\ell; \theta | \mathcal{F}_{t_k^\ell}) \quad (3.2.7)$$

where  $\lambda_\ell$  is given by Eq. (3.2.2)<sup>1</sup>. One could easily verify that this problem is convex as a function of the  $\alpha_u^{\ell m}$  (see Eq. (3.2.3)) and  $\mu^\ell$ . Thus the existence of a global optimum is guaranteed. As shown in Appendix 2.5.3, this parametrization allows a computationally efficient calculation of the squares loss function  $R$  together with its gradient.

### 3.2.3 Fitting results and comments

In this section, we present and comment on the results obtained through the least square estimation of the QRH model as defined by Eq. (3.2.2).

<sup>1</sup>or alternatively by Eq (3.2.6) in the sense given in [77]

**Estimation of the  $f^\ell(q_a, q_b)$ .** In our results, we document a clear dependence of the order arrival rates on both  $q_a^i$  and  $q_b^j$ , indicating that the state of the LOB has a clear influence on the order arrival rates. Let us point out that previous works (e.g., [126] and [99]) suggest that this dependence is dependent on the imbalance of the queue sizes as defined by

$$I(t) = \frac{v_b(t) - v_a(t)}{v_b(t) + v_a(t)} = \frac{q_b(t) - q_a(t)}{q_b(t) + q_a(t)} \quad (3.2.8)$$

where  $v_{a/b}(t)$  denote the volume available at time  $t$  at best ask/bid prices and we have assumed that orders have a constant volume corresponding to the AES as defined previously. The imbalance represents the simplest proxy to account for the instantaneous buying pressure. In this respect, in Figures 3.2 and 3.3, we plot  $f^\ell(q_a^i, q_b^j)$  in logarithmic scale for each order type  $\ell$  as a function of the imbalance  $I$  calculated as the median imbalance in the state interval associated with quantiles  $(q_a^i, q_b^j)$ . Let us note that in these plots, the dot sizes indicate the sizes of the corresponding  $q_b^j$ .

By looking at Figures 3.2 and 3.3 we can make the following observations: First, the variation of the imbalance on the order book captures most of the variations of the intensive parameters  $f^\ell(q_a, q_b)$ , this is more evident on the large tick asset (Bund) for which  $f$  can span almost three order of magnitudes (for  $P$ , and  $M$  events) as  $I$  ranges from  $-1$  to  $+1$ . The variations of  $f^\ell(q_a, q_b)$  are smaller for the small tick asset (DAX), so the effect of the imbalance is less pronounced albeit still visible for  $P$  and, to a lesser extent,  $L$  events. This is in line with the observation that the imbalance is a very good predictor for midprice changes for large tick assets while its predictive power is less marked for small tick ones (see, e.g. [70]). One aspect to keep in mind while analyzing these figures is that, especially for a small tick asset, there is also a considerable amount of information in the deeper levels of the book that is not taken into account here.

In Figure 3.2 corresponding to the Bund results, we observe that for midprice changes ( $P$  events), there seems to be a kind of threshold effect, in that for imbalance values below  $I = -1/2$ , upwards price increments are dramatically inhibited, while large positive imbalance only marginally increases the likelihood of upwards price changes. This is rather natural since an imbalance smaller than  $-1/2$  corresponds to the case the ask size is at least three times bigger than the bid size, making an upwards movement of the midprice very unlikely.

More surprisingly, the agents seem to strongly condition their decision to use aggressive orders ( $M$  events) on the state of the order book. This hints that agents are rushing to get the last available liquidity before an upwards midprice move (indeed,  $f$  gets very large for  $M$  events when the imbalance is large, i.e., when there is hardly anything left on the ask size). On the other hand, limit and cancel orders appear to be much less sensitive to the state of the book. For the DAX, as we already pointed out, the dependence on the imbalance is much less pronounced (see Figure 3.3) and, except for the mid-price changes, all factors  $f^\ell(q_a, q_b)$  are weakly varying with  $I$ . For the market ( $M$ ) and cancel ( $C$ ) orders, we observe a regime-switching around  $I = 1/2$ .

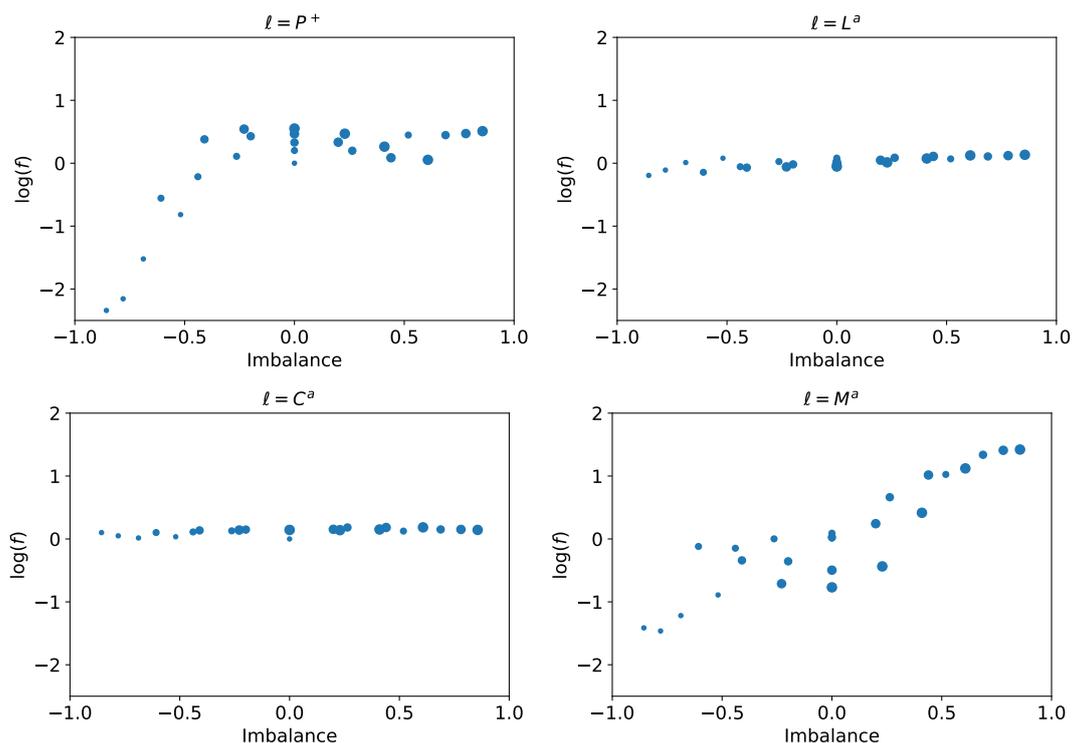


Figure 3.2: From left to right, upside to downside,  $\log_{10}(f^l(q_a^i, q_b^j))$  for  $l = P^+, L^a, C^a, M^a$  of QRH model as a function of the imbalance 3.2.8, Bund future. The quantiles are the same for bid and ask sides and correspond to  $q_a^1 = q_b^1 = ]0, 80]$ ,  $q_a^2 = q_b^2 = ]80, 165]$ ,  $q_a^3 = q_b^3 = ]165, 258]$ ,  $q_a^4 = q_b^4 = ]258, 386]$  and  $q_a^5 = q_b^5 = ]386, +\infty[$ .

These regimes correspond respectively to extremely large and small ask queue sizes. In the latter case (that turns out to occur when the spread equals one tick), the occurrence of the market and cancel orders which do not change the midprice is very unlikely.

Last but not least, let us remark that the intensity variations that are not captured by the imbalance are mainly located around  $I = 0$  where the individual queue sizes ( $q_a$  or  $q_b$ , which are almost equal) seem to have an important impact. Thus, for instance, for the Bund,  $f^{P^+}(q, q)$  seems to increase with  $q$ . This can be explained by a more thorough analysis that shows that the average spread is increasing with  $q$  and is very close to 2 when  $q$  is large. The same kind of remarks could be done for market orders. As this example illustrates, other microstructural variables, notably the spread, should be considered for an even more complete model. This, however, is outside the scope of the present work.

**Comparison of estimated and empirical intensity.** To further validate the QRH model, we test its ability to reproduce the empirical intensity. The MLE of the averaged

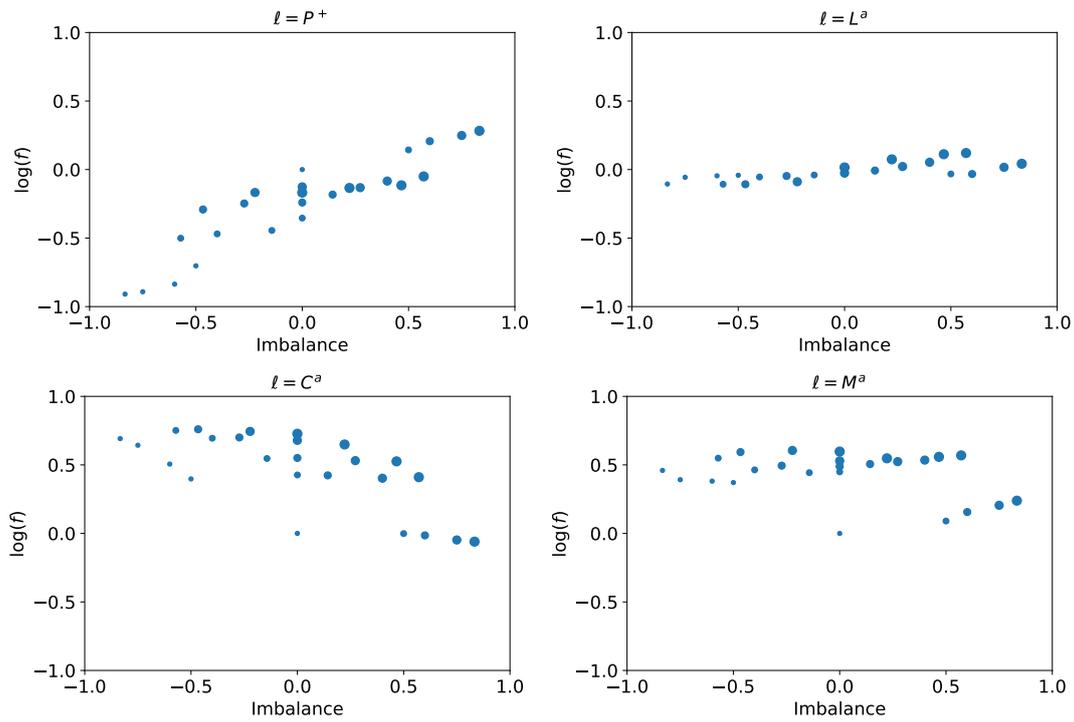


Figure 3.3: From left to right, upside to downside,  $\log_{10}(f^l(q_a^i, q_b^j))$  for  $l = P^+, L^a, C^a, M^a$  of QRH model as a function of the imbalance 3.2.8, DAX index future. The quantiles are the same for bid and ask sides and correspond to  $q_a^1 = q_b^1 = ]0, 2]$ ,  $q_a^2 = q_b^2 = ]2, 3]$ ,  $q_a^3 = q_b^3 = ]3, 5]$ ,  $q_a^4 = q_b^4 = ]5, 8]$  and  $q_a^5 = q_b^5 = [8, +\infty[$ .

intensity conditioned by the state  $q = (q_a, q_b)$  writes (in a non parametric framework)

$$\hat{\Lambda}^\ell(q) = \hat{\Lambda}(q) \frac{\sum_{t_k} \mathbb{1}_{q(t_k^-)=q}}{\sum_{t_k} \mathbb{1}_{q(t_k^-)=q}} \quad (3.2.9)$$

$$\text{with } \hat{\Lambda}(q) = \text{mean}(t_k - t_{k-1} | q(t_k^-) = q)^{-1},$$

where the operator  $\text{mean}(\cdot)$  corresponds to the empirical mean. This estimation is simply based on the observations of the process  $\{N_t^l\}_l$ . One can compute a corresponding estimator  $\hat{\Lambda}_{\text{QRH}}^\ell(q)$  using the parametric form of  $\hat{\lambda}_{\text{QRH}}^\ell(t_k^{\ell,-} | q)$ , which leads to

$$\hat{\Lambda}_{\text{QRH}}^\ell(q) = \text{mean}(\hat{\lambda}_{\text{QRH}}^\ell(t_k^{\ell,-} | q(t_k^{\ell,-}) = q)). \quad (3.2.10)$$

In order to synthesize the so-obtained results, we choose not to present the comparison of  $\hat{\Lambda}_{\text{QRH}}^\ell(q)$  with  $\hat{\Lambda}^\ell(q)$  for all types of orders and for all states  $q$ . Instead, for each type of order, we report the weighted relative error  $\Delta^\ell$  defined as:

$$\Delta^\ell = \frac{\sum_q |\hat{\Lambda}^\ell(q) - \hat{\Lambda}_{\text{QRH}}^\ell(q)| N^\ell(q)}{\sum_q \hat{\Lambda}^\ell(q) N^\ell(q)}. \quad (3.2.11)$$

The weighted relative error for Bund and DAX is presented in Table 3.3. We observe that  $\Delta^\ell$  are at the order of 10%, which provides a satisfactory match to the empirically observed intensity.

	$P^+$	$P^-$	$L^a$	$L^b$	$C^a$	$C^b$	$M^a$	$M^b$
Bund	14.2%	10.5%	6.7%	6.0%	7.4%	8.1%	4.0%	12.0%
DAX	8.2%	5.9%	0.5%	4.7%	7.1%	1.1%	1.6%	5.9%

Table 3.3: Error of average intensities by order type

**Analysis of the kernel norm matrices.** Finally, to complete the analysis of our results, in Figure 3.4 we display the matrices of the estimated norms  $\{\int \phi^{\ell m}(t) dt\}_{\ell m}$  for both the Bund and the DAX. These matrices reveal the average interactions between different event types when queue dependence is disregarded. As such they are the counterpart of the kernel norm matrices shown in Figure 4 of [13].

We recover a lot of the features highlighted in [13], such as the strong diagonal components for limit, market, and cancel orders (a signature of order splitting), as well as the fact that market orders and price movement appear to influence much more limit and cancel than the other way round.

Notably, for the Bund, when the midprice moves up (resp. down), the rate of limit orders on the ask (resp. bid) side decreases (resp. increases) since the efficient price is close to the best ask, and there is no gain to send a limit versus a market. This also explains why the rate of cancel orders sent on the best ask (resp. bid) side increases (resp. decreases). The same effect can be seen (but attenuated a lot) on the DAX. It is

attenuated because the tick size is small on the DAX, so the efficient price argument is not as strong. Let us point out that the exact same argument can be used to explain the effects of market orders on the limit and cancel orders. The influence of market order over price change is mainly because, under our settings, market order consumes the liquidity at the best prices, which could eventually create a new price. Since DAX is a small tick asset and queue size at the best prices is smaller than Bund, market orders are more likely to generate new prices. So the influence of market order over price change is more visible. We can see that for the DAX future, limit orders and cancellations also have a stronger influence on price changes for the same reason.

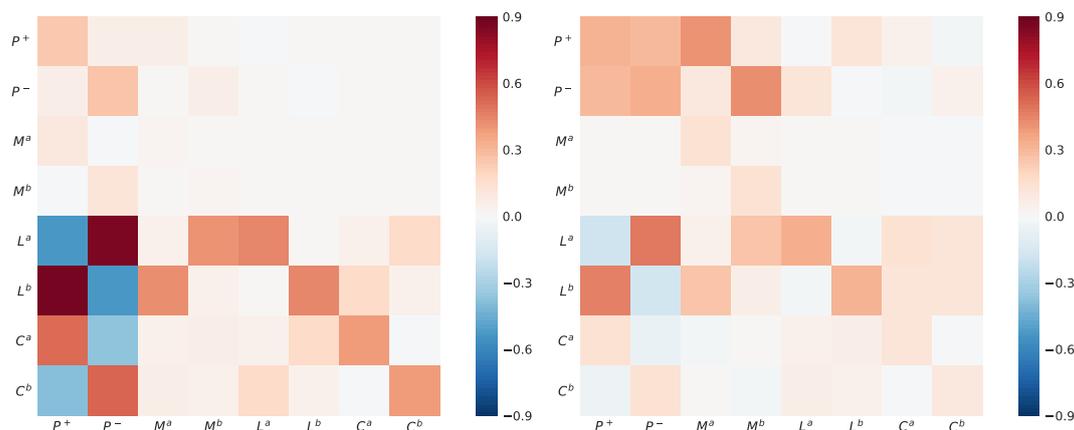


Figure 3.4: The estimated matrix norms  $\int \phi^{lm}(t)dt$  using least square estimation QRH model. Bund future on the left and DAX future on the right.

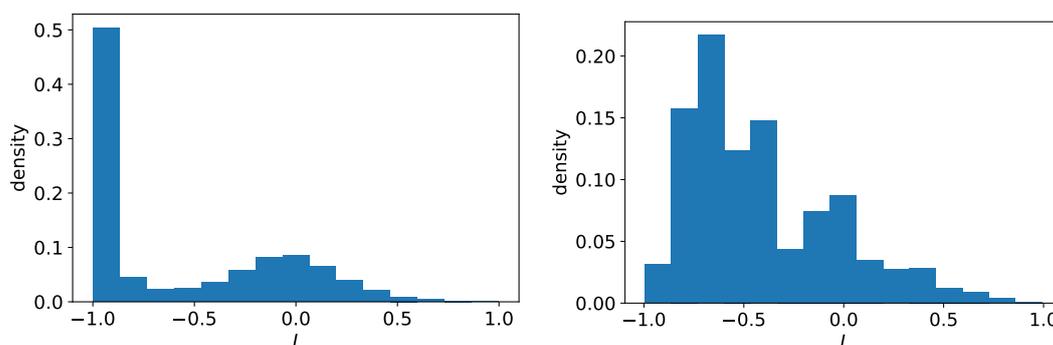


Figure 3.5: Empirical frequencies of observed imbalance values right after a  $P^+$  event for the Bund (left panel) and the DAX (right panel). Notice the large components for negative imbalance values in both cases.

One can also see that, apart from the self-exciting effect due to the splitting of orders, for the Bund, limit orders on one side are coupled with cancellations on the other side. This can be seen as a simple market-making strategy (rebalancing the position). It does not show on the DAX because, since the tick size is much smaller, the same rebalancing

strategy does not affect necessary both best sides. We refer the reader to [13], Section 4 for further details on the interpretation of these features.

Rate of mean-reversion	
Bund	0.65
DAX	0.56

Table 3.4: Measure of mean-reversion of price: Empirical probabilities than two successive midprice change events have opposite directions.

Though when compared in most features, QRH model (Figure 3.4) and the pure Hawkes model (Figure 4 in [13]) are similar, we can however observe a striking difference in the  $P \rightarrow P$  and  $P \rightarrow T$  submatrices. We notice that the QRH model indicates an absence of a strong excitation between  $P^+$  to  $P^-$  and  $P^-$  to  $P^+$  (top left  $2 \times 2$  submatrices in Figure 3.4), which should be the signature of the high-frequency price mean reversion (as explained in [13]). For a pure Hawkes processes based model, as in work [13], the mean-reversion of price is reflected on the strong anti-diagonal terms of the  $P \rightarrow P$  kernel norms submatrix (i.e., strong  $P^+ \rightarrow P^-$  and  $P^- \rightarrow P^+$  terms). This results from the fact that (on average) a  $P^+$  event will generate more  $P^-$  events than  $P^+$  events and vice versa. As shown in Table 3.4, the actual midprice series are strongly mean-reverting, so it is likely that within the QRH model, this feature is explained by the queue-reactive function  $f(q_a, q_b)$ . Indeed, as illustrated in Figure 3.5, after a mid-price change, the evolution of the imbalance is in favor of a price move in the opposite direction. For instance, an upward price jump  $P^+$  leads, most of the time, to a negative imbalance either because of a refill of the best ask queue  $q_a$  or simply because a single limit order is sent within the spread. According to Figure 3.2 and Figure 3.3, in such situations we have  $f^{P^+}(q_a, q_b) \ll 1$  and  $f^{P^-}(q_a, q_b) > 1$ . Conversely, after a downward price change,  $P^-$ , we will have  $f^{P^-}(q_a, q_b) \ll 1$  and  $f^{P^+}(q_a, q_b) > 1$ . The  $2 \times 2$  submatrix  $P \rightarrow P$  estimated within a pure Hawkes model is then likely to correspond to the QRH Hawkes submatrix multiplied by a large factor on its anti-diagonal and a small factor on its diagonal. The same kind of argument based on imbalance impact can be invoked to explain the high diagonal values of the  $P \rightarrow T$  submatrix while the highest values in [13] were rather observed on the anti-diagonal.

### 3.3 Summary and prospects

This chapter is a continuation of Chapter 2, where the “Single Queue Reactive Hawkes (SQRH) model” is extended to the best limit or both sides of the LOB. It shares the same ambition to improve respectively the approach of Huang et al. [87] on the queue reactive nature of the LOB dynamics and the model of Bacry et al. [13].

We show that the QRH model can be calibrated by parametric approaches. We

extend the loss function from log-likelihood to least square to accommodate negative values in the convolutional kernel. The QRH model leads us to refine Bacry et al.'s findings in [13] by accounting for the states of the best bid and best ask queues. Our results suggest that the imbalance of volume presented on two sides of the LOB allows us to explain a large part of the queue dependence.

The QRH model could be further improved by accounting for the interactions between the queue sizes (and thus the imbalance) and the order flow while introducing an explicit dependence on the spread for small tick assets. Some substantial simplifications we made could also be removed to obtain an even more realistic model, such as dropping the assumption of unitary order sizes by adopting a similar setup as in [102].

Besides considering various applications of these models to design and optimize high-frequency trading strategies, one fundamental challenge is to adapt these models to the general case of Eq. (2.1.3). That is, setting the queue dependence of the exogenous and Hawkes part arbitrary, and considering all order book levels up to a given depth simultaneously. From a mathematical point of view, a deeper understanding of the stability and stationarity conditions for queue dependent Hawkes model remains to be developed. More fundamentally, a clear understanding of the observed shapes of the exogenous intensities and the imbalance of the order flow arrival rates in terms of various market participants' (rational) behavior remain open questions.

## 3.4 Appendix

### 3.4.1 Calculating the least squares function of QRH model and its gradient

Let's note  $q$  the state of the LOB by combining the state of both the ask side and the bid side.

$$q = q_a \times q_b \quad f(q(t)) := f(q_a(t^-), q_b(t^-)). \quad (3.4.12)$$

Then at time  $t$ , the intensity function of dimension  $\ell$  is:

$$\lambda_t^\ell = f^\ell(q(t)) \left( \mu^\ell + \sum_m \sum_u \alpha_u^{\ell m} \int_0^t \beta_u e^{-\beta_u(t-s)} dN_s^m \right). \quad (3.4.13)$$

Similar to the calculation of the log-likelihood function and its gradients, we first define some intermediate variables whose value doesn't depend on  $\mu$ ,  $\alpha$  or  $f$ . They only need to be calculated once in the pre-processing stage. Then they could be reused during the calculation of the least-square loss function and its gradient.

$$g_u^\ell(t) = \sum_{t_k^\ell < t} \beta_u e^{-\beta_u(t-t_k^\ell)}, \quad (3.4.14)$$

$$G_u^m(q) = \int_0^T g_u^m(s) \mathbf{1}_{type(q(t))=q} dt, \quad (3.4.15)$$

$$H_{uu'}^{mm'}(q) = \int_0^T \mathbb{1}_{\text{type}(q(t))=q} g_u^m(t) g_{u'}^{m'}(t) dt, \quad (3.4.16)$$

$$D(q) = \int_0^T \mathbb{1}_{\text{type}(q(t))=q} dt, \quad (3.4.17)$$

$$C^m(q) = \sum_{k=1}^{N^m} \mathbb{1}_{\text{type}(t_k^{m,-})=q}. \quad (3.4.18)$$

### Computation of $g$

$$g_u^m(t_k) = \sum_{t_{k'}^m < t_k} \beta_u e^{-\beta_u(t_k - t_{k'}^m)} \quad (3.4.19)$$

$$= \sum_{t_{k'}^m < t_{k-1}} \beta_u e^{-\beta_u(t_{k-1} - t_{k'}^m)} e^{-\beta_u(t_k - t_{k-1})} + \sum_{t_{k-1} \leq t_{k'}^m < t_k} \beta_u e^{-\beta_u(t_k - t_{k'}^m)} \quad (3.4.20)$$

$$= e^{-\beta_u(t_k - t_{k-1})} g^m(t_{k-1}) + \beta_u e^{-\beta_u(t_k - t_{k-1})} \mathbb{1}_{\text{type}(t_k^-)=m}. \quad (3.4.21)$$

### Computation of $G$

$$G_u^m(t_k) - G_u^m(t_{k-1}) = \int_{t_{k-1}}^{t_k} g_u^m(s) ds \quad (3.4.22)$$

$$= \frac{1 - e^{-\beta_u(t_k - t_{k-1})}}{\beta_u} g_u^m(t_{k-1}) + \left(1 - e^{-\beta_u(t_k - t_{k-1})}\right) \mathbb{1}_{\text{type}(t_{k-1}^+)=m}. \quad (3.4.23)$$

### Computation of $H$

$$H_{uu'}^{mm'}(q) = \sum_k \int_{t_{k-1}}^{t_k} \mathbb{1}_{\text{type}(q(t))=q} g_u^m(t) g_{u'}^{m'}(t) dt \quad (3.4.24)$$

$$= \sum_k \mathbb{1}_{\text{type}(q(t_k^-))=q} g_u^m(t_{k-1}) g_{u'}^{m'}(t_{k-1}) e^{-(\beta_u + \beta_{u'})(t_k - t_{k-1})}. \quad (3.4.25)$$

### Least squares function

With the intermediate variables defined in the previous part, we can rewrite the least squares function defined in 3.2.7 in a more efficient form for calculation. In dimension  $\ell$ ,

$$\begin{aligned} R^\ell(\theta) &= \int_0^T \lambda_\ell^2(t; \theta | \mathcal{F}_t) dt - 2 \sum_{k=1}^{N^\ell} \lambda_\ell(t_k^\ell; \theta | \mathcal{F}_{t_k^\ell}) \\ &= \int_0^T f^\ell(q(t))^2 \left( \mu^\ell + \sum_m \sum_u \alpha_u^{\ell m} \int_0^t \beta_u e^{-\beta_u(t-s)} dN_s^m \right)^2 dt \\ &\quad - 2 \sum_{k=1}^{N^\ell} f^\ell(q(t_k^\ell)) \left( \mu^\ell + \sum_m \sum_u \alpha_u^{\ell m} \int_0^{t_k^\ell} \beta_u e^{-\beta_u(t_k^\ell - s)} dN_s^m \right). \end{aligned} \quad (3.4.26)$$

Let's name the first item and the second item in the formula above *term I* and *term II* separately. We could further decompose *term I* into new items *term I.1*, *term I.2* and etc:

$$\begin{aligned}
\text{term I} &= \int_0^T f^\ell(q(t))^2 \left( \mu^\ell + \sum_m \sum_u \alpha_u^{\ell m} \int_0^t \beta_u e^{-\beta_u(t-s)} dN_s^m \right)^2 dt \\
&= \int_0^T f^\ell(q(t))^2 \mu^{\ell^2} dt \\
&\quad + \int_0^T 2f^\ell(q(t))^2 \mu^\ell \left( \sum_m \sum_u \alpha_u^{\ell m} \int_0^t \beta_u e^{-\beta_u(t-s)} dN_s^m \right) dt \\
&\quad + \int_0^T f^\ell(q(t))^2 \left( \sum_m \sum_u \alpha_u^{\ell m} \int_0^t \beta_u e^{-\beta_u(t-s)} dN_s^m \right)^2 dt.
\end{aligned} \tag{3.4.27}$$

For *term I.1*, it could be calculated from the intermediate variables defined above:

$$\int_0^T f^\ell(q(t))^2 \mu^{\ell^2} dt = \mu^{\ell^2} \sum_q D(q) f^\ell(q)^2. \tag{3.4.28}$$

For *term I.2*,

$$\begin{aligned}
&\int_0^T 2f^\ell(q(t))^2 \mu^\ell \left( \sum_m \sum_u \alpha_u^{\ell m} \int_0^t \beta_u e^{-\beta_u(t-s)} dN_s^m \right) dt \\
&= \int_0^T 2f^\ell(q(t))^2 \mu^\ell \left( \sum_m \sum_u \alpha_u^{\ell m} g_u^m(t) \right) dt \\
&= 2\mu^\ell \sum_q f^\ell(q)^2 \left( \sum_m \sum_u \alpha_u^{\ell m} G_u^m(q) \right).
\end{aligned} \tag{3.4.29}$$

And for *term I.3*,

$$\begin{aligned}
&\int_0^T f^\ell(q(t))^2 \left( \sum_m \sum_u \alpha_u^{\ell m} \int_0^t \beta_u e^{-\beta_u(t-s)} dN_s^m \right)^2 dt \\
&= \int_0^T f^\ell(q(t))^2 \left( \sum_m \sum_{m'} \sum_u \sum_{u'} \alpha_u^{\ell m} \alpha_{u'}^{\ell m'} g_u^m(t) g_{u'}^{m'}(t) \right) dt \\
&= \sum_m \sum_{m'} \sum_u \sum_{u'} \alpha_u^{\ell m} \alpha_{u'}^{\ell m'} \left( \sum_q f^\ell(q)^2 H_{uu'}^{mm'}(q) \right).
\end{aligned} \tag{3.4.30}$$

For the convenience of notation, we flip the sign of *term II*. Then we decompose it into *term II.1* and *term II.2*:

$$\begin{aligned}
II &= \sum_{k=1}^{N^\ell} f^\ell(q(t_k^\ell)) \left( \mu^\ell + \sum_m \sum_u \alpha_u^{\ell m} \int_0^{t_k^\ell} \beta_u e^{-\beta_u(t_k^\ell-s)} dN_s^m \right) \\
&= \sum_{k=1}^{N^\ell} f^\ell(q(t_k^\ell)) \mu^\ell + \sum_{k=1}^{N^\ell} f^\ell(q(t_k^\ell)) \left( \sum_m \sum_u \alpha_u^{\ell m} \int_0^{t_k^\ell} \beta_u e^{-\beta_u(t_k^\ell-s)} dN_s^m \right).
\end{aligned} \tag{3.4.31}$$

For *term II.1*,

$$\sum_{k=1}^{N^\ell} f^\ell(q(t_k^\ell)) \mu^\ell = \mu^\ell \sum_q f(q) C^\ell(q). \tag{3.4.32}$$

For term II.2,

$$\sum_{k=1}^{N^\ell} f^\ell(q(t_k^\ell)) \left( \sum_m \sum_u \alpha_u^{\ell m} \int_0^{t_k^m} \beta_u e^{-\beta_u(t_k^m - s)} dN_s^m \right) = \sum_{k=1}^{N^\ell} f^\ell(q(t_k^\ell)) \left( \sum_m \sum_u \alpha_u^{\ell m} g_u^m(t_k^\ell) \right). \quad (3.4.33)$$

The least-square function defined in 3.2.7 could be obtained by summing up all these terms.

### Gradients

Using the intermediate variables and results presented above, direct calculation shows that:

$$\begin{aligned} \frac{\partial R}{\partial \mu^\ell} &= 2\mu^\ell \sum_q D(q) f^\ell(q)^2 \\ &\quad + 2 \sum_q f^\ell(q)^2 \left( \sum_m \sum_u \alpha_u^{\ell m} G_u^m(q) \right) \\ &\quad - 2 \sum_q f(q) C^\ell(q), \end{aligned} \quad (3.4.34)$$

$$\begin{aligned} \frac{\partial R}{\partial \alpha_u^{\ell m}} &= 2\mu^\ell \sum_q f^\ell(q)^2 G_u^m(q) \\ &\quad + 2 \sum_{m'} \sum_{u'} \alpha_{u'}^{\ell m'} \left( \sum_q f^\ell(q)^2 H_{uu'}^{mm'}(q) \right) \\ &\quad - 2 \sum_{k=1}^{N^\ell} f^\ell(q(t_k^\ell)) g_u^m(t_k^\ell), \end{aligned} \quad (3.4.35)$$

$$\begin{aligned} \frac{\partial R}{\partial f^\ell(q)} &= \mu^{\ell^2} \sum_q 2f^\ell(q) D(q) + 4\mu^\ell f^\ell(q) \left( \sum_m \sum_u \alpha_u^{\ell m} G_u^m(q) \right) \\ &\quad + 2 \sum_m \sum_{m'} \sum_u \sum_{u'} \alpha_u^{\ell m} \alpha_{u'}^{\ell m'} f^\ell(q) H_{uu'}^{mm'}(q) \\ &\quad - \mu^\ell C^\ell(q) - 2 \sum_{k=1}^{N^\ell} \mathbb{1}_{\text{type}(t_k^\ell) = q} \sum_m \sum_u \alpha_u^{\ell m} g_u^m(t_k^\ell). \end{aligned} \quad (3.4.36)$$

## Chapter 4

# From Rough to Multifractal volatility: the log S-fBM model

### 4.1 Introduction

During the past few years, new insights on stochastic volatility models have been obtained after the observation by Gatheral et al. [65] (see also [32, 33]) that the logarithm of the realized volatility is rough, i.e., is less regular than a standard Brownian motion. Rough volatility models have become very popular not only because they allow one to account for main empirical realized volatility properties but also because, when they are considered in asset price models, they provide a very good fit of option prices, and notably, their ATM skew power-law behavior close to maturity [101, 29, 61]. The first empirical evidence reported in [65] suggests that the logarithm of the asset price stochastic variance can be represented by a fractional Brownian motion (fBM) of Hurst exponent  $H$  close to  $H \simeq 0.1 < 1/2$ . More recent studies based either on quasi-likelihood approach [62] or GMM-approach [36], consistently suggest the  $H$  is even closer to  $H = 0$ , i.e.,  $H \lesssim 0.05$  for a large panel of equity data. In that respect, it is natural to consider the limit  $H \rightarrow 0$  in the rough process driving the volatility logarithm. Even if one cannot plug  $H = 0$  in the power-law expression of the fractional Brownian motion covariance, formally, it corresponds to a logarithmic behavior.

Such a logarithmic behavior is precisely the one that characterizes the so-called continuous random cascade models introduced two decades ago by Bacry et al. [118, 12]. Indeed, in 2000, these authors proposed the “Multifractal Random Walk” (MRW) as a model for asset prices in order to account for their multifractal properties, i.e., the fact observed by various authors (see e.g. [67, 66, 104]) that asset return empirical moments obey non-trivial scaling properties. The MRW model relies on a multifractal stochastic volatility model, namely the “Multifractal Random Measure” (MRM) model [118, 12], in which the log-volatility is provided by a log-correlated Gaussian field. Such a class of processes, also referred to as Gaussian multiplicative chaos, has been at the heart

of many studies in a large variety of applications [129]. Gaussian multiplicative chaos and the associated log-normal random cascades have been extended to any infinitely divisible distribution by Bacry and Muzy in [115, 20].

Recovering a multifractal volatility model as the limit  $H \rightarrow 0$  of a rough volatility model or, from a more general perspective, defining a meaningful limit  $H \rightarrow 0$  of a fractional Brownian motion and one of its variants has been the subject of various recent studies. In [63], the authors build a  $H=0$  - fBM by considering a regularisation from the harmonizable representation of fBM's while in [121, 74] a  $H = 0$  limiting process is obtained using a peculiar normalisation and centering of the fBM. In [59] (see also [58]), the authors consider the limit  $H \rightarrow 0$  of the exponential of a rescaled Riemann-Liouville fBM and its relationship with Gaussian multiplicative chaos. Finally, in [28], Bayer et al. propose a new class of rough models that consists in modulating the Riemann-Liouville fBM power-law kernel by a logarithmic factor. The so-obtained "super-rough" stochastic volatility remains well-defined as a continuous process when  $H = 0$ .

In this paper, our goal is to add a contribution to this problem by introducing a new version of rough volatility models based on the so-called "stationary" fBM (S-fBM). S-fBM is a variant of fBM whose covariance function is exactly the one obtained when considering the small-time approximation of the correlation of the fractional Ornstein-Uhlenbeck process considered in [65]. We prove that when  $H \rightarrow 0$ , one recovers the exact self-similar multifractal measure defined in [118, 12]. Our construction is based on the same approach proposed in [115, 20] where the log-volatility is obtained from the integration of a 2D Gaussian white noise over a triangular domain in a time-scale plane. It turns out such an approach corresponds to the same method defined by Takenaka to build the fractional Brownian motion [130]. Our model, therefore, provides a unified framework to consider both rough and multifractal stochastic volatility models. Beyond defining the main statistical properties of the model, we aim at estimating its parameters on a large panel of market data. For that purpose, we extend the GMM method proposed in [14] that is based on a "small intermittency" expansion of the moments of the measure logarithms.

The paper is organized as follows: in section 4.2, after recalling the basic notions underlying usual rough volatility models and the definition of the multifractal random measure (MRM), we introduce the log S-fBM random measure  $M_{H,T}(dt)$  as the exponential of the S-fBM random process which is nothing but a "stationary" version of the fractional Brownian motion of Hurst parameter  $0 < H < 1$ . We show that one recovers the celebrated Mandelbrot-Van Ness fBM when  $T$ , the correlation parameter of our model, tends to infinity. In this section, we also show that the log S-fBM converges, when  $H \rightarrow 0$ , towards a Multifractal Random Measure, consequently leading to a unified framework for rough volatility models ( $M_{H,T}(dt)$ , for  $H \neq 0$ ) and multifractal volatility models (by extension,  $M_{H=0,T}(dt)$ ). In section 4.3, we establish, within this unified framework, analytical expressions for the second-order moments of respectively

$M_{H,T}([t, t + \Delta])$  and its logarithm, while in section 4.4 we define two GMM parameter estimation methods based on these expressions. Our approach is illustrated in various numerical examples. Application to empirical data, namely the daily volatility of many individual stocks as well as market indices is provided in section 4.5. Section 4.6 summarizes our findings while technical material and mathematical proofs are provided in Appendices.

## 4.2 The log Stationary fractional Brownian Motion (log S-fBM) stochastic volatility model

### 4.2.1 Multifractal and rough volatility models

Before introducing our new model of stochastic volatility measure (log S-fBM), let us briefly walk through the two popular former classes of stochastic volatility models it is notably designed to unify, namely the Rough Fractional Stochastic Volatility (RFSV) model and the Multifractal Random Walk (MRW) or Multifractal Random Measure (MRM) models.

**The MRM/MRW models** The MRW was firstly introduced in 2001 by Bacry et al. [118, 12] as a model for log-prices  $X(t)$  that has exact (log-normal) multifractal properties, i.e., such that the moment of price returns  $\delta_\tau X(t) = X(t + \tau) - X(t)$  obeys exact scaling properties:

$$\mathbb{E}[|\delta_\tau X(t)|^q] = \mathbb{E}[|X(t + \tau) - X(t)|^q] \sim C_q \tau^{\zeta(\frac{q}{2})} \quad (4.2.1)$$

where the multifractal scaling spectrum  $\zeta(q)$  is a non-linear (namely parabolic) concave function that only depends on a single positive parameter  $\lambda^2$  (which quantifies the level of non linearity of  $\zeta(q)$ ) and such that  $\zeta(1) = 1$ . Let us point out that the parameter  $\lambda^2$  is generally referred to as the *intermittency coefficient* since it governs the degree of multifractality of the model, i.e., the range of the Hölder exponents that characterize the paths  $X(t)$ . It consequently controls the degree of appearance of volatility bursts. When  $\lambda = 0$ , the model is said to be *monofractal*,  $X$  then simply corresponds to a Brownian motion which is almost everywhere of Hölder regularity  $H = 1/2$ .

The MRW model involves a log-normal stochastic volatility, that is a multifractal random measure (MRM)  $\widetilde{M}_T$ , obtained as the weak limit

$$\widetilde{M}_{\ell,T}(dt) \xrightarrow[\ell \rightarrow 0]{w} \widetilde{M}_T(dt), \quad (4.2.2)$$

where  $\widetilde{M}_{\ell,T}(dt)$  is defined by

$$\widetilde{M}_{\ell,T}(dt) = e^{\omega_{\ell,T}(t)} dt,$$

where  $\xrightarrow{w}$  stands for the weak convergence and the process  $\omega_{\ell,T}(t)$  is Gaussian and stationary with a logarithmic covariance vanishing for lags greater than  $T$  (see Eq.

(4.7.75)). Let us point out that multifractality of the limit process is obtained in the Gaussian multiplicative chaos context [129] which implies that, at the same time  $\ell$  goes to 0, the mean (resp. variance) of  $\omega_{\ell,T}$  has to go to  $-\infty$  (resp.  $+\infty$ ). Thus though the stochastic measure  $e^{\omega_{\ell,T}(t)} dt$  has a weak limit, the Gaussian process  $\omega_{\ell,T}(t)$  does not have a limit. We refer the reader to the beginning of Appendix 4.7.3 for detailed construction of the log-normal MRM.

Since such a logarithmic decreasing covariance can be interpreted using random multiplicative cascades as the limit case where the scale ratio goes to 1, one often refers to such a model as “continuous cascade” [14] models. In [115, 20], MRM measures have been extended from log-normal statistics to any log-infinitely divisible law so that they obey the exact scaling law:

$$\mathbb{E}[|\delta_\tau \widetilde{M}_T(t)|^q] = \mathbb{E}[|\widetilde{M}_T(t+\tau) - \widetilde{M}_T(t)|^q] \sim C_q \tau^{\zeta(q)}, \quad (4.2.3)$$

where  $\zeta_q$  is the cumulant generating index of the infinitely divisible law (let us point out that it is parabolic only in the Gaussian case).

The MRM process has been used in various works since 2001 for volatility modeling. Not only does it have stationary increments but it reproduces a lot of volatility stylized facts (including scale invariance and self-similarity properties). Moreover, it also benefits from a concise geometric construction, which allows one to easily obtain the auto-covariance function in the desired form.

**The original RFSV model.** In 2018, Gatheral et al. [65] introduced a new (but related) class of models called “rough” fractional stochastic volatility (RFSV) models. Instead of focusing on the scaling properties of price increments, Gatheral et al. examined the regularity properties of the log-volatility and observed (as the case for a multifractal model) that volatility appears to be far less regular than a Brownian motion. RFSV model quickly became a popular model. Within the RFSV framework, the volatility measure  $V_{H,T}([t, t+\tau])$  of some given interval  $[t, t+\tau]$  is supposed to be provided by a density measure  $v(t)$  corresponding to a log-normal stationary process:

$$V_{H,T}([t, t+\tau]) = \int_t^{t+\tau} v(s) ds = \int_t^{t+\tau} e^{o_{H,T}(s)} ds, \quad (4.2.4)$$

where  $o_{H,T}(t)$  is a fractional Ornstein-Uhlenbeck (fOU) process that satisfies, for some  $0 < H < 1/2$ , the equation

$$do_{H,T}(t) = \nu dB_t^H - \alpha(o_H(t) - m) dt, \quad (4.2.5)$$

where  $B_t^H$  is a fractional Brownian motion with Hurst parameter  $H$ . The parameter  $\nu^2$  (resp.  $m$ ) is the variance (resp. mean) of  $o_{H,T}(t)$  and  $\alpha = \frac{1}{T}$ , where  $T$  represents a characteristic correlation time that accounts for the typical mean reversion length of the process. Indeed, Gatheral et al. show that, for  $\tau > 0$  small enough, the covariance

function of  $o_{H,T}(t)$  can be approximated as:

$$\text{Cov}[o_{H,T}(t), o_{H,T}(t + \tau)] \simeq \frac{\nu^2}{2} \left( T^{2H} \Gamma(2H + 1) - \tau^{2H} \right), \quad (4.2.6)$$

where  $\Gamma$  represents the Gamma function. In [65], it is also shown that when  $T \rightarrow \infty$ ,  $o_{H,T}(t)$  behaves locally as a fractional Brownian motion  $B_t^H$  in the sense that,  $\forall t_0$ :

$$\mathbb{E} \left[ \sup_{t \in [0, t_0]} |o_{H,T}(t) - o_{H,T}(0) - \nu B_t^H| \right] \rightarrow 0. \quad (4.2.7)$$

This result can be of practical importance for application in finance since empirically it appears that  $T$  is very large and consequently  $\nu B_t^H$  can be used as a volatility model instead of the associated fOU process  $o_{H,T}(t)$  as long as  $t \ll T$ .

Let us point out that, since the original work [65], many other versions of RFSV models have been introduced in the literature, each of them serving some specific purposes (making some explicit computations or estimations simpler) while keeping the main feature of the original RFSV model, i.e., the "roughness" of the volatility modeled using an fBM-like process. In the next section, we will introduce a new version that will enable us to unify in the same framework an RFSV model and the MRM framework.

#### 4.2.2 The log S-fBM random measure : a common framework for RFSV and MRM models

In this section, we build the main model of this paper. This model allows us to define a common framework for RFSV and MRM models. It is built in three steps. First we introduce a stationary version of a fractional Brownian motion, namely the S-fBM process  $\{\omega_{H,T}(t)\}_t$  for  $H > 0$ . Then using this S-fBM process, we define the log S-fBM stochastic measure  $M_{H,T}$  ( $H > 0$ ) which can be seen as a new version of a rough volatility model (RFSV). Finally, we prove that this process converges when  $H$  goes to 0 to a measure that we will refer to as  $M_{0,T}$ , which is shown to be an MRM.

##### Step 1/3 : Defining the S-fBM process $\{\omega_{H,T}(t)\}_t$ for $H > 0$

The S-fBM process  $\{\omega_{H,T}(t)\}_t$  is a stationary Gaussian process and can thus be defined by its mean and its covariance function. In Appendix 4.7.1, we provide the details of its construction by following the one proposed by Bacry & Muzy ([115, 20]) for building log-infinitely divisible Multifractal Random Measures (MRM). Let us point out that such a construction can also be related to the original approach proposed by Takenaka to build correlated fields (see [130] and Appendix 4.7.2). Thus, following the construction detailed in Appendix 4.7.1, the S-fBM process  $\{\omega_{H,T}(t)\}_t$  is defined for  $H > 0$  as a stationary Gaussian process whose covariance function is:

$$C_\omega(\tau) = \text{Cov}[\omega_{H,T}(t), \omega_{H,T}(t + \tau)] = \begin{cases} \frac{\nu^2}{2} [T^{2H} - \tau^{2H}], & \text{when } |\tau| < T \\ 0, & \text{when } |\tau| \geq T \end{cases} \quad (4.2.8)$$

The parameter  $H$  is analog to the Hurst parameter of the fBM process since it controls the “roughness” of the model. The variance parameter  $\nu$  controls the average amplitude of the process and the constant  $T$  is a large time scale that corresponds to the correlation scale. Let us point out that the approximated covariance provided by Eq. (4.2.6) of the fOU process  $\omega_{H,T}$ , involved in the construction of the RFSV model, holds exactly for the S-fBM  $\omega_{H,T}$  (up to a rescaling of  $T$ ) for lags smaller than  $\tau$ . Both the S-fBM process and fOU could then be regarded as stationary versions of an fBM process but, unlike the fOU process, the correlation function of S-fBM exactly vanishes for lags greater than  $T$ , i.e., the S-fBM values at different timestamps are independent when the distance between timestamps is large enough (i.e., greater than  $T$ ).

It is noteworthy that, when  $T \rightarrow \infty$ , one recovers the original Takenaka construction of the fBM [130] by proving that  $\omega_{H,T}(t) - \omega_{H,T}(0) \rightarrow B_H(t)$ . More precisely, in Appendix 4.7.2 we show that, when  $T \rightarrow \infty$ , the analog of Eq. (4.2.7) holds for  $\omega_{H,T}$ :

**Proposition 3.** *There exists  $B_H(t)$  a fractional Brownian motion of Hurst index  $H$  and unit variance at  $t = 1$  such that,  $\forall t_0 > 0$ , one has:*

$$\lim_{T \rightarrow \infty} \mathbb{E} \left[ \sup_{t \in [0, t_0]} |\omega_{H,T}(t) - \omega_{H,T}(0) - \nu B_H(t)| \right] = 0. \quad (4.2.9)$$

A direct result from the similarity in auto-covariance function is that S-fBM has the same scaling property as RFSV. According to Appendix 4.7.1, for  $\frac{\tau}{2} < T$ ,

$$\delta_\tau \omega_{H,T}(t) = \omega_{H,T}(t + \tau) - \omega_{H,T}(t) \sim \mathcal{N}(0, \nu^2 \tau^{2H}). \quad (4.2.10)$$

It leads to the following scaling property of generalized moments,  $\forall q > 0$ :

$$\mathbb{E}[|\delta_\tau \omega_{H,T}(t)|^q] = \nu^q \frac{2^{q/2} \Gamma(\frac{q+1}{2})}{\sqrt{\pi}} \tau^{qH} = C_q \tau^{qH}. \quad (4.2.11)$$

This means that  $\log(\mathbb{E}[|\delta_\tau \omega_{H,T}(t)|^q])$  is linear against  $\log(\tau)$  with slope  $qH$ . From Kolmogorov continuity theorem it results that the paths of  $\omega_{H,T}$  are continuous functions. More precisely,  $\omega_{H,T}(t)$  is  $\alpha$ -Hölder continuous for all regularity exponents  $\alpha < H$ . We especially point out that the calibration of the Hurst parameter  $H$  of the log-volatility process in [65] is based on the obtained scaling behavior (4.2.11).

**Step 2/3 : Defining the log S-fBM stochastic measure  $\{M_{H,T}(t)\}_t$  for  $H > 0$**

The log S-fBM stochastic measure is then defined as:

$$M_{H,T}(dt) = e^{\omega_{H,T}(t)} dt. \quad (4.2.12)$$

Then for any interval  $I$ , one has:

$$M_{H,T}(I) = \int_I e^{\omega_{H,T}(t)} dt. \quad (4.2.13)$$

Under this setting, we retrieve the so-called stationarity of volatility process, i.e.:

$$\mathbb{E}[M_{H,T}(I)] = \sigma^2 |I| \quad (4.2.14)$$

with

$$\sigma^2 = e^{m + \frac{\nu^2}{2}},$$

where  $m$  (resp.  $\nu^2$ ) is the mean (resp. variance) of  $\omega_{H,T}$ . The quantity  $\sigma^2$  can be regarded as the variance of the price fluctuations on a unit-time interval.

**Step 3/3 : Convergence of  $\{M_{H,T}(t)\}_t$  towards an MRM when  $H$  goes to 0**

As shown in Appendix 4.7.3, the MRM measure can be recovered from the log S-fBM by taking the limit  $H \rightarrow 0$ . More precisely, the following proposition holds true:

**Proposition 4.** *Let  $M_{H,T}(t) = M_{H,T}([0, t])$  be the log S-fbm process defined by (4.2.13) and define the intermittency coefficient,*

$$\lambda^2 = H(1 - 2H)\nu^2. \quad (4.2.15)$$

*Considering both  $\lambda^2$  and the variance of the price fluctuations*

$$\sigma^2 = e^{m + \nu^2/2}$$

*are fixed, then, when  $H \rightarrow 0$  (and consequently,  $\nu^2 \rightarrow +\infty$  and  $m \rightarrow -\infty$ ), one has*

$$M_{H,T}(dt) \xrightarrow{w} \widetilde{M}_T(dt) \quad (4.2.16)$$

*where  $\xrightarrow{w}$  stands for the weak convergence and  $\widetilde{M}_T$  is a log-normal MRM (as defined by (4.2.2)) with the intermittency coefficient  $\lambda^2$  and integral scale  $T$ .*

The proof is provided in Appendix 4.7.3. This result indicates that the MRM can be considered as a limit case of a log S-fBM and therefore could be regarded as an "extremely rough" case.

**Conclusion and notations for the remaining of the paper**

For the sake of simplicity, in the following, the MRM  $\widetilde{M}_T(dt)$  will be referred to as  $M_{0,T}(dt)$ . Thus, we can consider that we have built a class of models  $M_{H,T}(dt)$ , which correspond for  $H > 0$  to an RFSV model and for  $H = 0$  to an MRM model.

### 4.3 Second order properties of $M_{H,T}([0, t])$ and its logarithm

In section 4.4, we will consider the problem of estimating the parameters of the S-fBM, namely  $H, \nu^2$  (or equivalently  $\lambda^2$ ) and  $T$  through the expression of various "statistical moments" of the process. Among these moments, the correlation function of  $M_{H,T}$  or of  $Z_{H,T} = \ln M_{H,T}$  are particularly interesting since, as emphasized below, they can be approximated by simple analytical expressions.

Let us first remark that in [65], Gatheral et al. proposed to estimate the roughness exponent  $H$  of the RFSV model (equivalently  $H > 0$  in the log-SfBM model) by

considering the scaling of the increments of  $\omega_{H,T}$  as in Eq. (4.2.11). However, since  $\omega_{H,T}(t)$  cannot be directly observable, they consider as a proxy of  $\mathbb{E}[|\delta_\tau \omega_{H,T}(t)|^q]$ , the observable moments:

$$m(q, H, \tau, \Delta) = \mathbb{E}\left(|\ln M_{H,T,\Delta}(t + \tau) - \ln M_{H,T,\Delta}(t)|^q\right), \quad (4.3.17)$$

where  $M_{H,T,\Delta}(t)$  is the so-called integrated variance over an interval of size  $\Delta$ :

$$M_{H,T,\Delta}(t) = \sigma^2 \int_t^{t+\Delta} e^{\omega_{H,T}(s)} ds. \quad (4.3.18)$$

Thus, the exponent  $H$  is measured from the scaling behavior in  $\tau$  of this proxy of  $\mathbb{E}[|\delta_\tau \omega_{H,T}(t)|^q]$  using the Eq. (4.2.11).

However, as emphasized below (see Section 4.4.1), the estimation of  $H$  based on Eq. (4.2.11) can be highly biased. In order to get an unbiased estimation of  $H$  in the framework of an RFSV model, Ref. [36] introduces a totally different framework. The authors provide a GMM method that is based on the correlation function of  $M_{H,T,\Delta}$ :

$$C_M(\Delta, \tau) = \mathbb{E}[M_{H,T,\Delta}(t)M_{H,T,\Delta}(t + \tau)]. \quad (4.3.19)$$

More precisely, they show that under peculiar conditions, its asymptotic behavior when  $\tau \gg \Delta$  can be obtained and then a GMM formula can be derived. Within the framework of various RFSV models (namely the one involving an fBM or its Riemann-Liouville variant), the authors advocate the use of this GMM method and show that it provides reliable estimates for both the roughness parameter  $H$  and the variance parameter  $\nu^2$ .

Following this latter path, in this work, we aim at defining a GMM method for the log-SfBM framework, that works for both  $H > 0$  (the RFSV case) and  $H = 0$  (the MRM case). We thus need to establish exact or good approximations of correlation function  $C_M(\Delta, \tau)$  of  $M_{H,T,\Delta}$ . This is the purpose of the next section (Section 4.3.1).

Moreover, as we will see, the process  $\ln M_{H,T,\Delta}(t)$  is, in some sense, close to be a Gaussian process, consequently it is also natural to operate the GMM not on the process  $M_{H,T,\Delta}$  itself but on its logarithm  $\ln M_{H,T,\Delta}(t)$ . We therefore also need to establish exact or good approximations of the correlation function of  $\ln M_{H,T,\Delta}$ , which is defined by :

$$C_{\ln M}(\Delta, \tau) = \text{Cov}[\ln M_{H,T,\Delta}(t), \ln M_{H,T,\Delta}(t + \tau)]. \quad (4.3.20)$$

This is the purpose of Section 4.3.2.

### 4.3.1 Integrated variance correlation function

In Appendix 4.7.4 we prove the following Proposition that gives an explicit analytic formula for  $C_M(\Delta, \tau)$ :

**Proposition 5.** *For any  $\tau \leq T$ , one has*

$$C_M(\Delta, \tau) = K_1(F(\tau + \Delta) + F(\tau - \Delta) - 2F(\tau)) \quad (4.3.21)$$

with

$$F(z) = \begin{cases} zK_2^{-\frac{1}{2H}}\gamma(\frac{1}{2H}, K_2z^{2H}) - K_2^{-\frac{1}{H}}\gamma(\frac{1}{H}, K_2z^{2H}), & \text{when } H > 0 \\ \frac{z^{2-\lambda^2}}{(2-\lambda^2)(1-\lambda^2)}, & \text{when } H = 0 \end{cases} \quad (4.3.22)$$

where  $\gamma(a, z)$  stands for the (lower) incomplete Gamma function,

$$\gamma(a, z) = \int_0^z t^{a-1} e^{-t} dt,$$

and where we have denoted

$$\begin{aligned} K_1 &= \frac{\sigma^4 e^{K_2 T^{2H}}}{2H}, \quad \text{if } H > 0 \\ &= \sigma^4 T^{\lambda^2}, \quad \text{if } H = 0 \\ K_2 &= \frac{\nu^2}{2} = \frac{\lambda^2}{2H(1-2H)}. \end{aligned}$$

Let us notice that when  $\tau > T$ , since  $M_{H,T,\Delta}(t)$  and  $M_{H,T,\Delta}(t + \tau)$  are independent, one has:

$$C_M(\Delta, \tau) = \mathbb{E}[M(\Delta, \tau)]^2 = \sigma^4 \Delta^2.$$

Moreover, using the equality:

$$\gamma(s, z) = s^{-1} z^s e^{-z} U(1, s+1, z),$$

where  $U(1, s, z)$  is the Kummer's confluent hypergeometric function, the function  $F(z)$  can be simply rewritten as:

$$F(z) = \sigma^4 z^2 e^{C_\omega(z)} \left( U(1, 1 + \frac{1}{2H}, K_2 z^{2H}) - \frac{1}{2} U(1, 1 + \frac{1}{H}, K_2 z^{2H}) \right), \quad (4.3.23)$$

where  $C_\omega(z)$  is the covariance of  $\omega_{H,T}$  provided by Eq. (4.2.8). Since, when  $|b| \rightarrow \infty$ ,  $U(1, b, z) \simeq 1 + \frac{z}{b}$ , when  $H \ll 1$ , one finally gets the following approximation for  $F(z)$  :

$$F(z) \simeq \sigma^4 z^2 e^{C_\omega(z)} \left( \frac{1}{2} + \frac{3H}{2} z^{2H} \right). \quad (4.3.24)$$

### 4.3.2 Small $\lambda^2$ approximation of the logarithm integrated variance moments

In this section, our goal is to obtain analytical expressions for the moments of  $\ln M([t, t + \Delta])$  instead of  $M([t, t + \Delta])$ . In [14] a GMM method to estimate the parameters of the MRM  $\widetilde{M}_T$  has been proposed relying on the expression of such logarithmic moments that were obtained within a *small intermittency*, i.e.  $\lambda^2 \ll 1$ , asymptotic behavior. In fact, it is straightforward to check that all proofs and results established in [14] in the limit  $\lambda^2 \rightarrow 0$  for the log-normal MRM measure  $\widetilde{M}_T$  remain valid for  $M_{H,T,\Delta}$  for  $H > 0$ , i.e. in the log S-fBM framework introduced in this paper. Indeed, in particular by simply checking that all conditions required for MRM also hold for the log S-fBM measure  $M_{H,T,\Delta}$ , a direct consequence of Proposition 13 in [14] is the following result:

**Proposition 6.** *Let  $t_1, \dots, t_n$  be  $n$  arbitrary times. The generalized moments of the logarithm of  $\Delta^{-1}M_{H,T,\Delta}(t)$  admit the following Taylor series expansion around  $\lambda^2 = 0$ :*

$$\mathbb{E}\left[\ln\left(\frac{M_{H,T,\Delta}(t_1)}{\Delta}\right) \cdots \ln\left(\frac{M_{H,T,\Delta}(t_n)}{\Delta}\right)\right] = \lambda^n \Delta^{-n} \mathbb{E}\left[\Omega_{H,T,\Delta}(t_1) \cdots \Omega_{H,T,\Delta}(t_n)\right] + o(\lambda^n), \quad (4.3.25)$$

where  $\Omega_{H,T,\Delta}(t)$  is the Gaussian process defined by

$$\Omega_{H,T,\Delta}(t) = \frac{1}{\lambda} \int_t^{t+\Delta} (\omega_{H,T}(u) - \mathbb{E}(\omega_{H,T}(u))) du. \quad (4.3.26)$$

Within this approximation, one can directly compute  $C_{\ln M}(\Delta, \tau)$  the correlation function of  $\ln M_{H,T,\Delta}(t)$  as defined in (4.3.20). From the definition of  $\Omega_{H,T,\Delta}$  and the expression (4.2.8) for the covariance of  $\omega_{H,T}(t)$ , it results:

**Proposition 7.** *To the first order in  $\lambda^2 \ll 1$ , the covariance function of  $\ln M_{H,T,\Delta}$ , reads:*

$$\begin{aligned} C_{\ln M}(\Delta, \tau) &= \frac{\lambda^2}{2H(1-2H)} \Delta^{-2} \int_0^\Delta du \int_\tau^{\tau+\Delta} (T^{2H} - |u-v|^{2H}) dv \quad (4.3.27) \\ &= \frac{\lambda^2}{2H(1-2H)} \left( T^{2H} - \frac{(\tau+\Delta)^{2H+2} + |\tau-\Delta|^{2H+2} - 2\tau^{2H+2}}{\Delta^2(2H+1)(2H+2)} \right) \quad (4.3.28) \end{aligned}$$

Let us first start with two direct consequences of these propositions

- When  $\Delta \rightarrow 0$  one gets  $C_{\ln M}(\Delta, \tau) \simeq \frac{\lambda^2}{2H(1-2H)} (T^{2H} - \tau^{2H})$  which is nothing but the covariance of  $\omega_{H,T}(t)$ .
- When  $H \rightarrow 0$ , one recovers the expression in Proposition 10 of [14] in the MRM case.
- Proposition 6 leads to approximating  $\ln M_{H,T,\Delta}(t)$  by a Gaussian process.

Using this last consequence, Proposition 6 can also be used to get an approximation to the first order in  $\lambda^2$  of the moments defined in (4.3.17). Indeed, if one supposes that  $\ln M_{H,T,\Delta}(t+\tau) - \ln M_{H,T,\Delta}(t)$  is a Gaussian random variable of variance  $V(H, \tau, \Delta)$ , then,

$$m(q, H, \tau, \Delta) \stackrel{\nu}{=} \pi^{-1/2} 2^{\frac{q}{2}} \Gamma\left(\frac{q+1}{2}\right) (V(H, \tau, \Delta))^{\frac{q}{2}} \quad (4.3.29)$$

in which  $\stackrel{\nu}{=}$  indicates that equality holds in the first order of  $\lambda^2$ .

From expression (4.3.28), one has when  $\tau < T$ ,

$$\begin{aligned} V(H, \tau, \Delta) &= 2\text{Var}(\ln M_{H,T,\Delta}(t)) - 2C_{\ln M}(\Delta, \tau) \\ &\stackrel{\nu}{=} \frac{\lambda^2}{H(1-2H)} \left( \frac{(\tau+\Delta)^{2H+2} + |\tau-\Delta|^{2H+2} - 2\tau^{2H+2}}{\Delta^2(2H+1)(2H+2)} - \frac{2\Delta^{2H+2}}{\Delta^2(2H+1)(2H+2)} \right) \\ &\stackrel{\nu}{=} \lambda^2 \tau^{2H} g_H\left(\frac{\Delta}{\tau}\right) \end{aligned}$$

with

$$g_H(z) = \frac{|1+z|^{2H+2} + |1-z|^{2H+2} - 2|z|^{2H+2} - 2}{z^2 H(1-2H)(2H+1)(2H+2)}. \quad (4.3.30)$$

The final expression for the moments of the increments of the measure logarithm  $\ln M_{H,T,\Delta}$  reads, in the first order in  $\lambda^2$ ,

$$m(q, H, \tau, \Delta) \stackrel{\nu}{=} 2^{\frac{q}{2}} \pi^{-1/2} \Gamma\left(\frac{q+1}{2}\right) \lambda^q \tau^{qH} \left[ g_H\left(\frac{\Delta}{\tau}\right) \right]^{q/2}. \quad (4.3.31)$$

Let us remark that we have the following asymptotic relation:

$$g_H(z) \sim \begin{cases} \frac{1}{H(1-2H)} + O(z^{2H}) & \text{when } z \rightarrow 0 \\ z^{-2} \left( (1+z)^2 \ln(1+z) + (1-z)^2 \ln(1-z) - 2z^2 \ln(z) \right) + O(H) & \text{when } H \rightarrow 0 \end{cases} \quad (4.3.32)$$

and when  $H > 0$ , one recovers that when  $\Delta \rightarrow 0$  one gets, in the first order in  $\lambda^2$ ,

$$m(q, H, \tau, \Delta) \stackrel{\nu}{=} C \tau^{qH} + O((\Delta/\tau)^{2H}), \quad (4.3.33)$$

which is the expression used to estimate  $H$  in [65] where  $\ln M_{H,T,\Delta}(t)$  corresponds to the logarithm of the (daily) realized volatility.

## 4.4 Estimation

This section is devoted to the estimation of  $H$  in the framework of log-SfBM. We first show (in Section 4.4.1) that if  $H$  is measured from the scaling behavior of  $\mathbb{E}[|\delta_\tau \omega_{H,T}(t)|^q]$  against  $\tau$  using Eq. (4.2.11) (as advocated in [65]), the estimation of  $H$  can be highly biased.

In order to get an unbiased estimation of  $H$ , we construct two GMM based estimators in Sections 4.4.2 and 4.4.3. The first one is based on the use of the moments of the log-SfBM process itself mainly relying on the explicit covariance formula in Eq. (4.3.21). The second one is based on the use of moments of the logarithm of the log-SfBM process and involves the explicit covariance provided by Eq. (4.3.28).

We show that both estimators are expected to be reliable even in the "high-frequency regime" when data are only available over an interval that is smaller than the overall correlation scale  $T$ , i.e. in a regime when one does not expect any ergodic hypothesis to hold.

### 4.4.1 Bias of the moment scaling method proposed in Ref. [65]

In [65], the parameter  $H$  is estimated from the scaling behavior of  $\mathbb{E}[|\delta_\tau \omega_{H,T}(t)|^q]$  against  $\tau$  as described in Eq. (4.2.11). More precisely, the unobservable quantity  $\mathbb{E}[|\delta_\tau \omega_{H,T}(t)|^q]$  is substituted by its observable proxy  $m(q, H, \tau, \Delta)$  as defined in Eq. (4.3.17), whose explicit form is worked out in Eq. (4.3.31). Then, a linear regression of  $\ln m(q, H, \tau, \Delta)$  against  $\ln(\tau)$  is performed in order to estimate  $H$ .

We can show see that this approach can lead to a significantly biased estimation of  $H$ . Indeed, by taking logarithm on both sides of Eq. (4.3.31), one has (using again the notation  $\stackrel{\nu}{\simeq}$  for equality up to the first order of  $\lambda^2$ )

$$\ln(m(q, H, \tau, \Delta)) \stackrel{\nu}{\simeq} C(q, \nu) + qH \ln(\tau) + \frac{q}{2} \ln \left( g_H \left( \frac{\Delta}{\tau} \right) \right). \quad (4.4.34)$$

where the expression of  $g_H(z)$  is provided in Eq. (4.3.30). Since the term  $\ln(g_H(\frac{\Delta}{\tau}))$  also depends on  $\ln(\tau)$ , assuming, on a given range of  $\tau$ , that

$$\ln(g_H(\frac{\Delta}{\tau})) \simeq B_H \ln(\tau/\Delta) + C, \quad (4.4.35)$$

the measured slope  $\hat{H}$  in the relation  $\ln m_q(H, \tau, \Delta)$  against  $\ln(\tau)$  is biased as:

$$\hat{H} = H + \frac{B_H}{2}, \quad (4.4.36)$$

and the bias depends on both the considered range of  $\tau$  and the value of  $H$ .

Let us illustrate this phenomenon on some numerical simulations. For that purpose, let us consider an arbitrary value  $\Delta = 1$  and  $\tau \in [1, 500]$ . For the specific value  $H = 0.002$ , Fig. 4.1 plots  $\ln(g_H(\frac{\Delta}{\tau}))$  as a function of  $\ln(\frac{\tau}{\Delta})$ . We note that the behavior is, to a first approximation, assumed to be linear in the range when  $\tau$  is sufficiently large in front of  $\Delta$ . A linear regression leads to a slope value of  $B \simeq 0.160$  which dominates the (highly biased) estimation of  $H$  :  $\hat{H} \simeq 0.08$ . Using the same procedure, we checked that for different values of  $H$  in the range  $[0.0, 0.15]$ , one systematically overestimates  $H$  with a bias that decreases from 0.08 to 0.03. It is noteworthy that the same kind of bias analysis has been considered by the author of Ref. [65] themselves (see Appendix C).

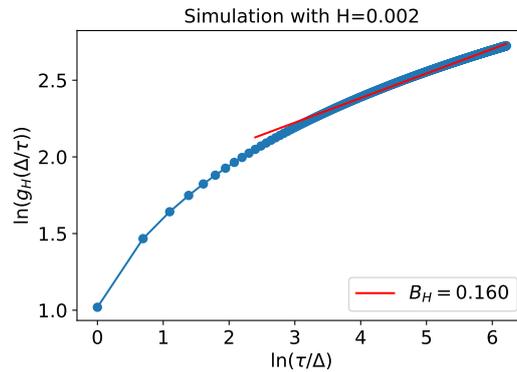


Figure 4.1: Estimation of the bias term  $B$  (Eq. (4.4.35)) involved in the estimation of  $H$  using linear regression on Eq. (4.4.34). Following Eq. (4.4.35),  $\ln(g_H(\frac{\Delta}{\tau}))$  is displayed against  $\ln(\tau/\Delta)$  (where  $\Delta = 1$  and  $\tau$  varies from 1 to 500).  $B_H$  is estimated by linear regression on this curve over the range  $\tau \in [10, 500]$

### 4.4.2 Low versus high frequency regime for GMM estimations

As already explained, our purpose is to build two GMM estimators based on the second order moments of the log-SfBM process  $M_{H,T}$  or its logarithm. More precisely, we will consider respectively  $C_M(\Delta, \tau)$  the correlation function of  $M_{H,T,\Delta}$  (using the explicit covariance formula (4.3.21)) and  $C_{\ln M}(\Delta, \tau)$ , the covariance function of  $\ln M_{H,T,\Delta}$  (using the explicit covariance formula (4.3.28)). If  $L$  denotes the overall size of the interval where the empirical data are available at scale  $\Delta$ , one can measure  $M_{H,T,\Delta}(k\Delta)$  (or equivalently  $\ln M_{H,T,\Delta}(k\Delta)$ ) for  $k = 1 \dots N$  where  $N = \frac{L}{\Delta}$  and the estimators of previous correlation functions read:

$$\widehat{C}_M(\Delta, k\Delta) = N^{-1} \sum_{j=1}^{N-k} M_{H,T,\Delta}(j\Delta) M_{H,T,\Delta}((j+k)\Delta) \quad (4.4.37)$$

$$\widehat{C}_{\ln M}(\Delta, k\Delta) = N^{-1} \sum_{j=1}^{N-k} \left( \ln M_{H,T,\Delta}(j\Delta) - \widehat{\mu}_\Delta \right) \left( \ln M_{H,T,\Delta}((j+k)\Delta) - \widehat{\mu}_\Delta \right) \quad (4.4.38)$$

$$\widehat{\mu}_\Delta = \frac{1}{N} \sum_{k=1}^N \ln M_{H,T,\Delta}(k\Delta) \quad (4.4.39)$$

In general, GMM methods rely on some ergodic hypothesis that ensures the convergence of previous empirical means towards the expected values. As advocated in [14] or in [36], these approaches allow one to build efficient parameter estimator in the limit  $N = \frac{L}{\Delta} \rightarrow \infty$ , which, when  $\Delta$  is kept fixed, corresponds to  $L \rightarrow \infty$ . When  $L \gg T$  (recall that  $T$  is the correlation length of  $M_{H,T,\Delta}(t)$ ), this ergodicity assumption can be proven to hold. We refer to such a situation as the “low-frequency regime”.

However, as first remarked in [14], one can alternatively consider the asymptotic regime  $N \rightarrow \infty$  when  $\Delta \rightarrow 0$ , while  $L = \mathcal{O}(T)$  is fixed. This is the “high-frequency regime”. Thus, whereas the low-frequency regime corresponds to  $\Delta < T \ll L$ , the second one corresponds to  $\Delta \ll L = \mathcal{O}(T)$ .

Let us point out that, as emphasized in [14] and motivated by the empirical results reported in [116] (see also Sec. 4.5 below), in many practical situations and notably for financial time series, the high-frequency regime appears to fit more precisely the empirical conditions. Notably, it appears that the correlation scale  $T$  of the realized volatility always seems to be larger than the observation size  $L$ . For instance, in Fig. 6(b) of [116], the authors plotted the logarithm of Dow-Jones realized daily volatility from 1928 to 2011 and observed deviations far from the “mean value” that are lasting for decades. The same kind of observation can be done in Fig. 4.6(a) below. In [116], it is also observed that the estimated correlation scale increases linearly with the observation size  $L$  from a few days to several years in agreement with the hypothesis that the true correlation scale is extremely large. In such a situation, assuming that the low-frequency regime  $L \gg T$  is reachable and consequently that the ergodic hypothesis holds, is clearly unrealistic.

These remarks call for developing GMM estimations in the high-frequency regime

$\Delta \ll L \leq T$ . Let us first start by noticing that from the expression of the covariance of  $\omega_{H,T}$  (Eq. (4.2.8)) one gets that for  $t \in I$  (where  $I$  is any interval such that  $|I| < L$ ), one has

$$\{\omega_{H,T}(t)\}_{t \in I} \stackrel{\mathcal{L}}{=} \{\Omega + \omega_{H,L}(t)\}_{t \in I} \quad (4.4.40)$$

where  $\stackrel{\mathcal{L}}{=}$  means an equality of all finite dimensional distributions and  $\Omega$  is a Gaussian random variable independent of  $\omega_{H,L}$  and of variance  $\frac{\nu^2}{2}(T^{2H} - L^{2H})$ . It thus results that we have, in any interval  $I$  of size  $|I| < L$ ,

$$\{M_{H,T,\Delta}\}_{t \in I} \stackrel{\mathcal{L}}{=} \{e^{\Omega} M_{H,L,\Delta}(t)\}_{t \in I} \quad \text{and} \quad (4.4.41)$$

$$\{\ln M_{H,T,\Delta}\}_{t \in I} \stackrel{\mathcal{L}}{=} \{\Omega + \ln M_{H,L,\Delta}(t)\}_{t \in I}. \quad (4.4.42)$$

As already discussed in [14] for the case of the MRM measure  $M_{H=0,L,\Delta}$ , these properties show that one cannot measure the parameters  $T$  and  $\sigma^2$  over an interval of size  $L < T$  since by redefining  $\sigma^2$  as  $\sigma^2 e^{\Omega}$ , one can always assume that  $T = L$ . It can also be seen on expression (4.3.22) that, when  $\tau < T$ , the large correlation scale  $T$  can be absorbed in a redefinition of the variance parameter  $\sigma^2$ .

If one seeks to consider correlation function based GMM estimators in the high frequency regime, one thus needs to study the behavior of respectively the estimators,  $\widehat{C}_{\ln M}(\Delta, n\Delta)$ ,  $\widehat{C}_M(\Delta, n\Delta)$  in the limit  $\Delta \rightarrow 0$ . A rigorous study of this problem is beyond the scope of the present paper, but we can refer to Theorem 10 of [14] where the authors proved that, in the multifractal case ( $H = 0$ ), the behavior of  $\widehat{C}_{\ln M}(\Delta, n\Delta)$  can be used to build an asymptotically unbiased and consistent estimator of  $\lambda^2$  in the high frequency regime. In the present paper, we just give a sketch of proof that one can build moments functions with vanishing fluctuations in the limit  $\Delta \rightarrow 0$ .

First, let us notice that, without loss of generality, one can always perform an overall change of scale,  $\Delta \rightarrow 1$ ,  $L \rightarrow \frac{L}{\Delta}$ ,  $T \rightarrow \frac{T}{\Delta}$ . This amounts to assume that  $\Delta = 1$  while the limit  $\Delta \rightarrow 0$  becomes  $L, T \rightarrow \infty$  and  $L = CT$  with  $C = \mathcal{O}(1)$ .

Then, Appendix 4.7.5 provides an heuristic proof of the following result:

**Proposition 8.** *Suppose that  $H < \frac{1}{2}$ . Then, for of any  $n_{\max} < \infty$ ,  $C \leq 1$ , when  $L = CT \rightarrow \infty$ , then, to the first order in  $\lambda^2$ , one has for all  $n \leq n_{\max}$ :*

$$D(n) \stackrel{\text{def}}{=} \widehat{C}_{\ln M}(1, n) - \widehat{C}_{\ln M}(1, 0) \xrightarrow{P} \widetilde{D}_{\ln M}(n) \quad (4.4.43)$$

where  $\xrightarrow{P}$  means that the convergence holds “in probability” and where

$$\widetilde{D}_{\ln M}(n) = -\nu^2 \frac{|n+1|^{2H+2} + |n-1|^{2H+2} - 2n^{2H+2}}{2(2H+1)(2H+2)}. \quad (4.4.44)$$

Moreover, numerical experiments (see Fig. 4.2 below) also suggest that an equivalent result holds for  $\widehat{C}_M$ , i.e.,

$$R(n) = \frac{\widehat{C}_M(1, n)}{\widehat{C}_M(1, 0)} \xrightarrow{P} \widetilde{R}_M(n) \quad (4.4.45)$$

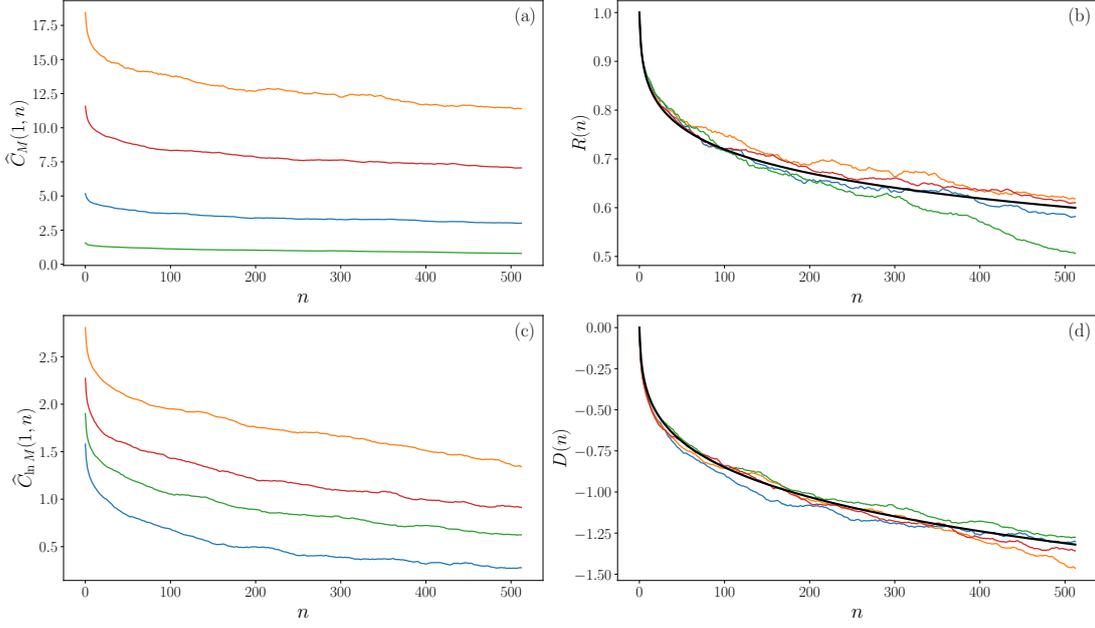


Figure 4.2: Estimation of the correlation functions of  $M_{H,T,\Delta}$  and  $\ln M_{H,T,\Delta}$  over an interval of size  $L$  such that  $\Delta = 1 \ll L < T$ . (a)  $\widehat{C}_M(\Delta = 1, n)$  as estimated from 4 independent realisations of  $M_{H=0.1, T=2^{17}, \Delta=1}$  ( $\lambda^2 = 0.03$ ) over an interval of size  $L = 2^{14}$ . Each estimation appears to be multiplied by an arbitrary random factor. (b) the ratio  $R(n) = \frac{\widehat{C}_M(1, n)}{\widehat{C}_M(1, 0)}$  vs.  $n$ . With such normalisation, all curves are superimposed and appear to be well fitted by the analytical expression  $\widetilde{R}(n)$  represented by the bold black curve. (c)  $\widehat{C}_{\ln M}(\Delta = 1, n)$  as estimated from 4 independent realisations of  $M_{H=0.1, T=2^{17}, \Delta=1}$  ( $\lambda^2 = 0.08$ ) over an interval of size  $L = 2^{14}$ . Each estimation appears to be shifted by a random term. (d) the difference  $D(n) = \widehat{C}_{\ln M}(1, n) - \widehat{C}_{\ln M}(1, 0)$  vs.  $n$ . When shifting all curves in such a way, they are superimposed and appear to be well fitted by the analytical expression  $\widetilde{D}(n)$  represented by the bold black curve.

with

$$\widetilde{R}_M(n) = F(n+1) + F(n-1) - 2F(n) \quad (4.4.46)$$

where  $F(z)$  is defined in Eq. (4.3.22).

The consequence of Eqs. (4.4.43), (4.4.45) is that, for  $L$  large enough, there exist two positive random variables  $K_1$  and  $K_2$  such that, in the first order of  $\lambda^2$ ):

$$\widehat{C}_{\ln M}(1, n) \simeq K_1 + \widetilde{D}_{\ln M}(n), \quad (4.4.47)$$

$$\widehat{C}_M(1, n) \simeq K_2 \widetilde{R}_M(n). \quad (4.4.48)$$

Numerical illustrations of these relations are given in Figs. 4.2 and 4.3.

In Fig. 4.2, we have displayed the estimated correlation functions  $\widehat{C}_M(\Delta, n\Delta)$  and  $\widehat{C}_{\ln M}(\Delta, n\Delta)$  for 2 sets of 4 realisations of  $M_{H,T,\Delta}(t)$  over an interval of size  $L = 2^{14}$

with  $\Delta = 1$ ,  $H = 0.1$ ,  $T = 2^{17}$  and  $\lambda^2 = 0.03$  (for  $\widehat{C}_M$ ) or  $\lambda^2 = 0.08$  (for  $\widehat{C}_{\ln M}$ ). One clearly sees in Fig. 4.2(a) that each estimate  $\widehat{C}_M$  seems to differ from the other one by a significant geometric random factor while estimates of  $\widehat{C}_{\ln M}$  appear to be randomly shifted in Fig. 4.2(c). In order to check these assertions, we have plotted respectively the ratios  $R(n)$  and the differences  $D(n)$  in Figs 4.2(b) and 4.2(d). As expected, all the curves appear to collapse to a single curve that is well described by analytical expressions obtained from respectively Eq. (4.4.44) and (4.4.46) (represented by bold curves).

The asymptotic convergence of Proposition 8 is illustrated in Fig. 4.3 where we have plotted  $D(n)$  as defined in Eq. (4.4.43) as obtained from random samples of  $M_{H,T,\Delta}$  with  $\Delta = 1$ ,  $H = 0.1$ ,  $\lambda^2 = 0.08$ ,  $T = 2L$  and  $L = 2^{12}, 2^{14}, 2^{16}, 2^{18}$ . All the curves are shifted by an arbitrary small constant for clarity purpose. As predicted by Eq. (4.4.43), one sees that, as  $L$  increases, the empirical curves become less and less noisy and increasingly close to the analytical expectation (4.4.44) (black curve).

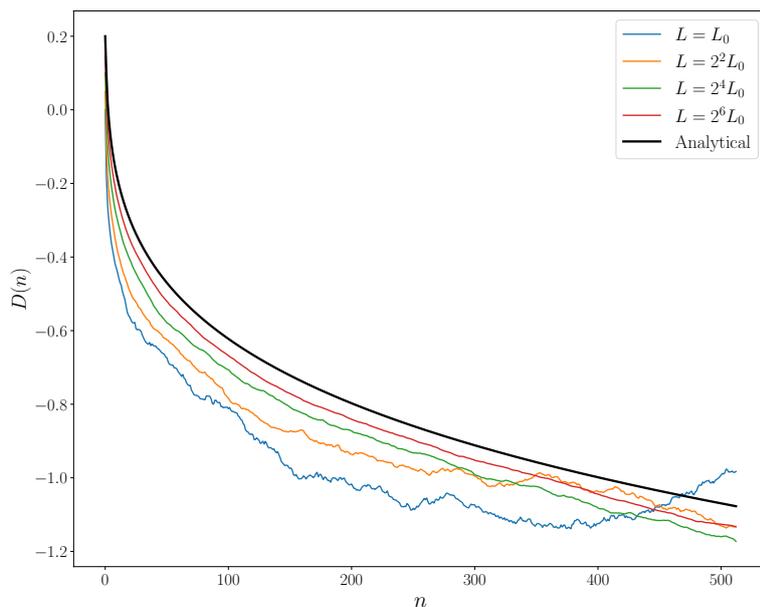


Figure 4.3: Estimation of the correlation function difference  $D(n)$  as defined in Proposition 8.  $D(n)$  from observations over intervals of increasing size  $L = L_0$ ,  $L = 4L_0$ ,  $L = 16L_0$  and  $L = 64L_0$  with  $L_0 = 2^{12}$ . The remaining parameters are  $\Delta = 1$ ,  $H = 0.1$ ,  $\lambda^2 = 0.08$  and  $T = 2L$ . One sees that as  $L \rightarrow \infty$ , the empirical fluctuations become smaller and smaller and the empirical estimations appear to converge towards to the theoretical expression (4.4.44) (black bold curve). The curves have been shifted by an increasing constant for the sake of clarity.

### 4.4.3 Defining two GMM estimators for $H$ and $\lambda^2$

We are now ready for defining two GMM estimators for  $H$  based respectively on the moments of  $M_{H,T,\Delta}$  or its logarithm in the high-frequency limit. By using the previously established expressions (4.4.44) and (4.4.46) for the empirical correlation function  $\widehat{C}_{\ln M}(1, n)$  and  $\widehat{C}_M(1, n)$ , one can devise two GMM methods along the same line as the methods proposed respectively in [14] and [36].

As explained in the previous section (Section 4.4.2), in the high-frequency regime, estimations of  $T$  or  $\sigma^2$  are unreachable. Thus, hereafter, we consider exclusively the problem of estimating the values of the parameters  $H$  and  $\lambda^2$  (or alternatively  $\nu^2$ ) using one of the following two sets of moments:

$$\text{GMM}_M : \mathcal{M}_1 = \left( \widehat{C}_M(1, j_1) - \widetilde{C}_M(1, j_1), \dots, \widehat{C}_M(1, j_Q) - \widetilde{C}_M(1, j_Q) \right),$$

$$\text{GMM}_{\ln M} : \mathcal{M}_2 = \left( \widehat{C}_{\ln M}(1, j_1) - \widetilde{C}_{\ln M}(1, j_1), \dots, \widehat{C}_{\ln M}(1, j_Q) - \widetilde{C}_{\ln M}(1, j_Q) \right),$$

where  $Q$  is the number of moments,  $j_1, j_2, \dots, j_Q$  different time indices,  $\widehat{C}_M$  and  $\widehat{C}_{\ln M}$  are the empirical estimators of respectively  $C_M$  and  $C_{\ln M}$  and  $\widetilde{C}_M, \widetilde{C}_{\ln M}$  the following analytical expressions:

$$\widetilde{C}_{\ln M}(1, n) = K_1 + \widetilde{D}_{\ln M}(n) + V_1 \delta_n \quad (4.4.49)$$

$$\widetilde{C}_M(1, n) = K_2 \widetilde{R}_M(n) \quad (4.4.50)$$

where  $K_1, K_2$  and  $V_1$  are 3 random positive constants and  $\delta_n$  stands for the Kronecker function. Notice that the term  $V_1 \delta_n$  allows one to account for the eventual presence of a white noise (of variance  $V_1$  superimposed to  $\ln M_{L,T,1}(t)$ ) as described in ref. [36].

### 4.4.4 Numerical illustrations and empirical performances of the GMM methods

In order to verify our approach and compare the performances of  $\text{GMM}_M$  and  $\text{GMM}_{\ln M}$ , we have carried out various numerical experiments. However, since historical volatility is not directly observable in financial markets, in order to consider a more realistic scenario, we decided to run the experiments directly on a price model. We consider that a “price”  $X_t$  is modelled by a Brownian motion whose variance is a log S-fBM measure  $dM$ , i.e.,

$$dX_t = e^{\omega_{H,T}(t)} dB_t = \frac{M_{H,T}(dt)}{dt} dB_t, \quad (4.4.51)$$

where  $M_{H,T}$  is the log-fBM defined in (4.2.12) while  $B_t$  is a Brownian motion independent of  $M_{H,T}$ . Let us notice that, when  $H = 0$ ,  $X_t$  is precisely the MRW process introduced in [118, 12].

Alternatively, an equivalent definition of  $X_t$  can be obtained using a time-warp of the Brownian motion:

$$X_t = B(M_{H,T}(t)). \quad (4.4.52)$$

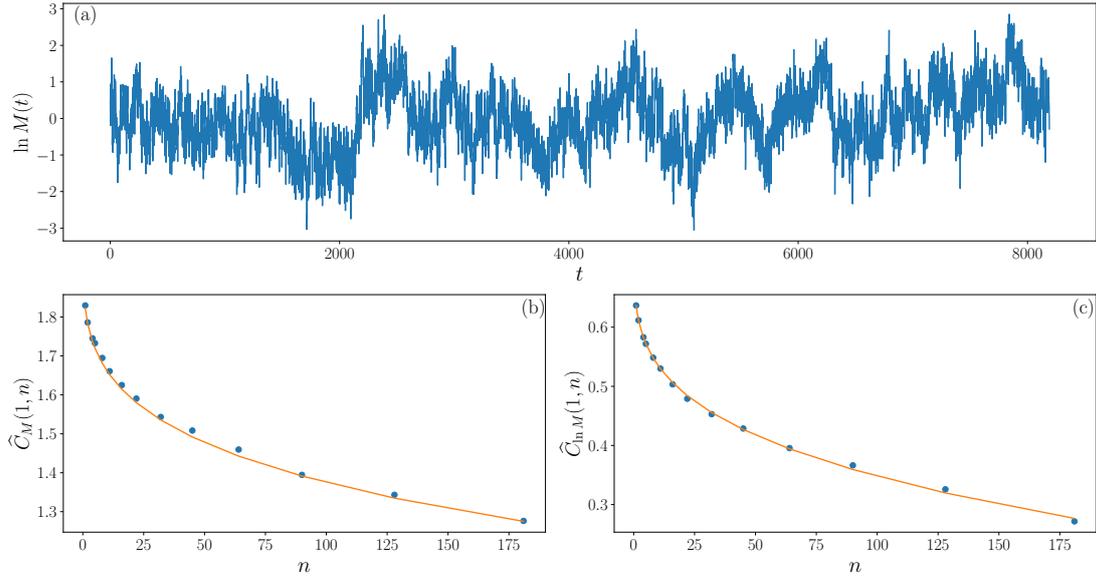


Figure 4.4: GMM estimations of  $H$  and  $\lambda^2$  (a) A sample of length  $L = 2^{14}$  of the “log-volatility”,  $\ln \widehat{M}_{H,T,\Delta}(t)$ , with  $n = 32$ ,  $\Delta = 1$ ,  $H = 0.08$ ,  $\lambda^2 = 0.1$  and  $T = 2^{17}$ . (b) Best GMM fit of  $\hat{C}_M(\tau)$  as a function of  $\tau$ . (c) Best GMM fit of  $\hat{C}_{\ln M}(\tau)$  as a function of  $\tau$ .

Within this framework,  $M_{H,T}(t)$  is called the (stochastic) volatility of  $X_t$ . If one does not observe directly  $M_{H,T}(dt)$  but only the process  $X_t$ , as emphasized notably in [24], a proxy of the integrated volatility over an interval of size  $\Delta$  is provided by an estimation of the quadratic variation of  $X$ :

$$\widehat{M}_{H,T,\Delta}(t) = \sum_{i=1}^n \left( X_{t+\frac{i\Delta}{n}} - X_{t+\frac{(i-1)\Delta}{n}} \right)^2. \quad (4.4.53)$$

As shown in [24] (see also [36]), as  $n \rightarrow \infty$ , under mild conditions,  $\widehat{M}_{H,T,\Delta} \rightarrow M_{H,T,\Delta}$  while even for moderate  $n$ ,  $\widehat{M}_{H,T,\Delta}(t)$  and  $\ln \widehat{M}_{H,T,\Delta}(t)$  provide excellent approximations of the integrated volatility and its logarithm. For the purpose of this paper, we have checked that  $n = 32$  is sufficient to disregard any significant difference between  $\widehat{M}$  and  $M$ .

We simulated independent samples of S-fBM processes and the associated processes  $X_t$  with  $H = 0.02$ ,  $H = 0.08$  and  $H = 0.15$ , with 2 different values of  $\lambda^2$ , namely 0.02 and 0.1. We chose  $T = 2^{17}$ ,  $L = 2^{14}$  and fixed arbitrary  $\sigma^2 = 1$ .

For all these parameters we run both  $\text{GMM}_M$  and  $\text{GMM}_{\ln M}$  estimators with  $Q = 19$  and  $\{\tau_k\}_{k=0,\dots,18} = \lfloor \sqrt{2^k} \rfloor$ . Our GMM implementations closely follow the one detailed in [36] and notably the error covariance is estimated using the Newey-West HAC type estimator with a lag  $L^{1/3}$  and the initialization is performed using the scaling estimator provided in [65]. We used the *L-BFGS-B* minimization algorithm as provided by *scipy.optimize* library in Python but we find similar results using alternative methods.

In Fig. 4.4, are displayed a fit of respectively  $\widehat{C}_M(1, n)$  and  $\widehat{C}_{\ln M}(1, n)$  using expressions (4.4.44) and (4.4.46) with the estimated GMM parameters for a sample of length  $L = 16384$  with  $H = 0.08$  and  $\lambda^2 = 0.02$  or  $\lambda^2 = 0.1$ . Our estimation results are summarised in Table 4.1 where we reported the obtained mean values and standard deviation of estimated  $H$  and  $\lambda^2$  for each set of parameters.

$\lambda^2 = 0.02$	$H = 0$	$H = 0.02$	$H = 0.08$	$H = 0.15$
$\widehat{H}$ (GMM <sub>M</sub> )	0.010 (0.01)	0.007 (0.015)	0.077 (0.033)	0.146 (0.05)
$\widehat{H}$ (GMM <sub>lnM</sub> )	0.010 (0.01)	0.018 (0.015)	0.082 (0.02)	0.153 (0.02)
$\widehat{\lambda}^2$ (GMM <sub>M</sub> )	0.010 (0.01)	0.010 (0.01)	0.018 (0.006)	0.021 (0.005)
$\widehat{\lambda}^2$ (GMM <sub>lnM</sub> )	0.019 (0.001)	0.020 (0.001)	0.019 (0.002)	0.020 (0.002)
$\lambda^2 = 0.1$	$H = 0$	$H = 0.02$	$H = 0.08$	$H = 0.15$
$\widehat{H}$ (GMM <sub>M</sub> )	0.010 (0.02)	0.018 (0.02)	0.11 (0.22)	0.16 (0.26)
$\widehat{H}$ (GMM <sub>lnM</sub> )	0.010 (0.01)	0.02 (0.01)	0.078 (0.02)	0.16 (0.02)
$\widehat{\lambda}^2$ (GMM <sub>M</sub> )	0.08 (0.03)	0.08 (0.02)	0.09 (0.045)	0.08 (0.07)
$\widehat{\lambda}^2$ (GMM <sub>lnM</sub> )	0.095 (0.001)	0.10 (0.005)	0.10 (0.008)	0.10 (0.008)

Table 4.1: Summary of GMM<sub>M</sub> & GMM<sub>lnM</sub> estimation performances. For each parameter set, we report the mean values and standard deviations as obtained from estimations realized on 50 independent samples of length  $L = 2^{14}$  of log S-fBM stochastic volatility model.

We clearly see that the GMM<sub>lnM</sub> method relies on logarithms of integrated volatilities outperforms the GMM<sub>M</sub> method built on integrated volatilities. This latter approach appears to have significantly larger bias and variance errors notably for very small  $H$  values. GMM<sub>lnM</sub> method provides more reliable estimates and in particular one sees that the errors on  $\lambda^2$  is very small for all sets of parameters though it slightly depends on  $\lambda^2$  (one has roughly  $\frac{\sigma_{\lambda^2}}{\lambda^2} \simeq 10^{-1}$ ) but not on  $H$ .

Let us emphasize that the reported estimations were obtained by estimating  $H$  and the variance parameter  $\nu^2$  from which  $\lambda^2$  is estimated using Eq. (4.2.15). We checked that estimating directly  $\lambda^2$  instead of deriving it from  $\nu^2$ , provides the same results. However, we observed that the errors on  $\nu^2$  are much larger than the errors on  $\lambda^2$ . More precisely, it appears that, for a fixed  $\lambda^2$ , the measured bias is strongly related to  $\widehat{H}$  as precisely expected from :

$$\widehat{\nu}^2 \simeq \frac{\lambda^2}{\widehat{H}(1 - 2\widehat{H})}. \quad (4.4.54)$$

This is illustrated in Fig. 4.5 in which two experiments were run with  $H = 0.02$ . For the first one we chose  $\lambda^2 = 0.02$  and for the other one we chose  $\lambda^2 = 0.1$ .

For each sample, we have reported  $\widehat{\nu}^2$  as a function of  $\frac{1}{\widehat{H}(1 - 2\widehat{H})}$  estimated by GMM<sub>lnM</sub> method. One can easily see that in each case ( $\lambda^2 = 0.02$  or  $\lambda^2 = 0.1$ ), one gets a very large dispersion on  $\widehat{\nu}^2$  (whose expected values should be respectively  $\nu^2 = 1.04$  and

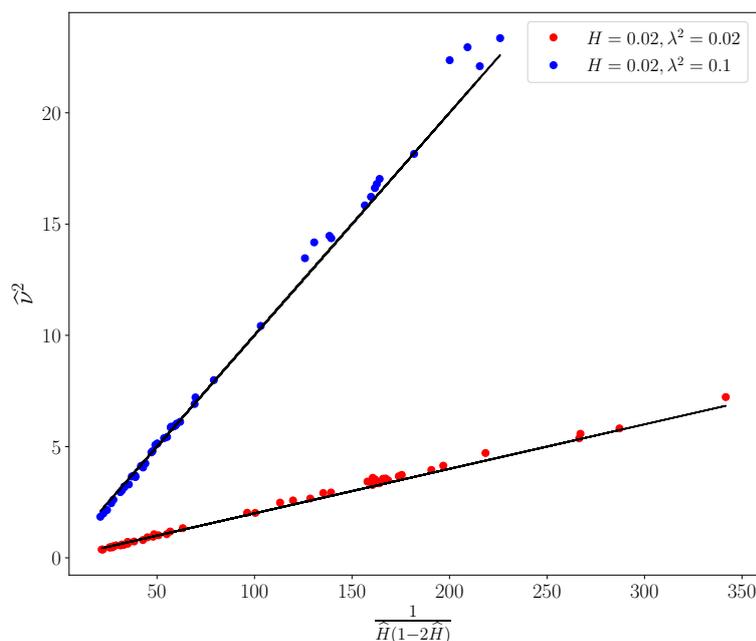


Figure 4.5:  $\text{GMM}_{\ln M}$  estimation of  $\nu^2$  as a function of  $\frac{1}{\widehat{H}(1-2\widehat{H})}$ . Each point corresponds to one estimation on a sample of length  $L = 2^{14}$ . The parameters are  $H = 0.02$  and  $\lambda^2 = 0.02$  (red symbols) or  $\lambda^2 = 0.1$  (blue symbols). The straight lines represent the fit provided by Eq. (4.4.54).

$\nu^2 = 5.2$ ) that, however, strikingly appears to be proportional to  $\frac{1}{\widehat{H}(1-2\widehat{H})}$  (which, when  $H$  is very small, has a large dispersion). As shown by the linear fits predicted by Eq (4.4.54) (continuous line in Fig. 4.5), the proportionality constant is precisely the value of the intermittency coefficient  $\lambda^2$  for which the estimation is quite accurate. These observations suggest that while  $\lambda^2$  can be estimated with a very small error, this is not at all the case of  $\nu^2$ , when  $H \ll 1$ . The intermittency coefficient  $\lambda^2$  appears to be a much more reliable quantity than the variance  $\nu^2$  of the log-volatility. This can be easily explained by the fact that, in order for the S-fBM measure to converge when  $H \rightarrow 0$  (towards the MRM  $\tilde{M}$ ), one has to choose a variance proportional to  $1/H$ . Therefore, in the moment estimation method, in order to match the empirical covariance values when the estimated  $H$  is very small, the parameter  $\nu^2$  must scale as  $H^{-1}$ .

Finally, let us mention that, by considering, besides  $L = 2^{14}$ , various sample lengths ( $L = 2^{10}, 2^{11}, 2^{12}, 2^{13}$ ), we checked that, as predicted by Prop. 8, the estimation errors vanishes when  $L$  increases. The prediction of Appendix 4.7.5, that the error behaves as  $L^{H-1/2}$  appears to agree with numerical experiments.

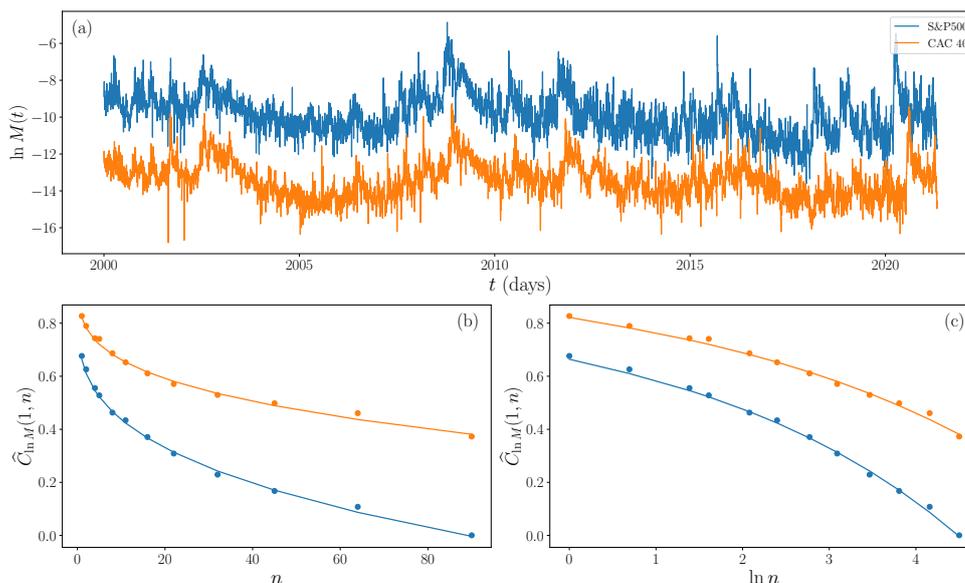


Figure 4.6:  $\text{GMM}_{\ln M}$  estimation of daily volatility of S&P500 and CAC40 indices from Oxford-Man dataset (a) log-realized bipower-variation from January 2000 to March 2021. (b) (resp. (c)) : the dots represent the estimations  $\hat{C}_{\ln M}(\tau, \Delta)$  of the corresponding correlation function  $\hat{C}_{\ln M}(\tau, \Delta)$  for each index as a function of  $\tau$  (resp.  $\ln \tau$ ). The plain lines correspond to the  $\text{GMM}_{\ln M}$  fits. The so-obtained estimated values of  $H$  are respectively  $H \simeq 0.14$  (for S&P) and  $H \simeq 0.13$  (for CAC40). CAC40 curves in Figures (a),(b) and (c) have been arbitrary shifted for the sake of clarity.

## 4.5 Application to realized volatility of asset returns

In this section we consider the application of the estimator of the former section to characterize the roughness exponent  $H$  and the intermittency coefficient  $\lambda^2$  of realized volatility associated with various assets. Section 4.4 suggests that the GMM estimator  $\text{GMM}_{\ln M}$  outperforms the other candidate  $\text{GMM}_M$ . This is why we exclusively consider the  $\text{GMM}_{\ln M}$  applied to various empirical daily volatility data. Our study is based on 2 datasets containing respectively stock market indices and individual stock prices:

### *Oxford-Man Institute of Quantitative Finance Realized Library (OIQFRL)*

The Oxford-Man Institute's Realized Library<sup>1</sup>, contains historical records of various estimators of daily realized volatility of several stock indices. This dataset is widely used in various empirical studies and in particular, it was used as a benchmark database in many former studies on rough volatility (see e.g. [65, 36]). So we apply  $\text{GMM}_{\ln M}$  estimator to analyze the daily volatility time series associated with 24 major stock market indices considered in [36]. Following this latter work, in the following, we only report obtained results when using bipower variation volatility estimator but we have

<sup>1</sup><http://realized.oxford-man.ox.ac.uk/data>

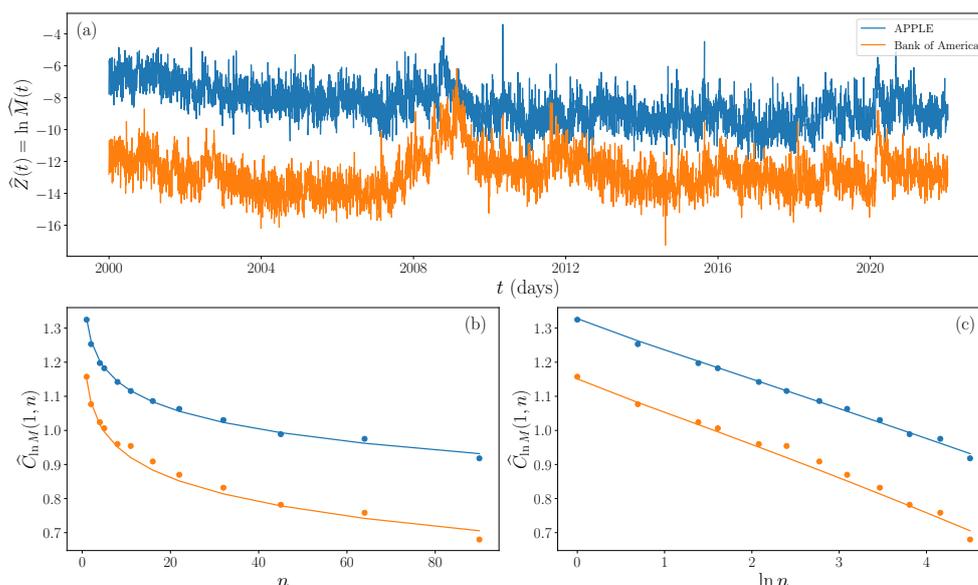


Figure 4.7:  $\text{GMM}_{\ln M}$  estimation of daily volatility of Apple and Bank of America from Yahoo Finance dataset. (a) log-realized Garman-Klass estimation of volatility [108] from January 2000 to December 2021. (b) (resp. (c)) : the dots represent the estimations  $\hat{C}_{\ln M}(\tau, \Delta)$  of the corresponding correlation function  $\hat{C}_{\ln M}(\Delta, \tau)$  for each index as a function of  $\tau$  (resp.  $\ln \tau$ ). The plain lines correspond to the  $\text{GMM}_{\ln M}$  fits. The so-obtained estimated values of  $H$  are respectively  $H \simeq 0.01$  (for Apple) and  $H \simeq 0.022$  (for Bank of Am.). Apple curves in Figures (a),(b) and (c) have been arbitrary shifted for the sake of clarity.

checked that the same results are obtained when using realized variance estimators at scale 5 min or 10 min. Two  $\text{GMM}_{\ln M}$  estimations are illustrated in Fig. 4.6 : one on CAC40 data and one on S&P500 data. The corresponding daily historical volatilities (using bipower-variation estimator) are illustrated in Fig. 4.6(a). We observe that over the 20-years period, the volatilities of S&P 500 and CAC40 are strongly correlated. One can also notice that some of the correlated departures from the mean value are lasting several years. This observation seriously questions any ergodic hypothesis that would result from short-term correlations as assumed in many papers (see, e.g., [65, 36]). Figs 4.6(b) (resp. (c)) displays the corresponding estimated correlation functions  $\hat{C}_{\ln M}(\Delta, \tau)$  as a function of  $\tau$  (resp.  $\ln \tau$ ) and their  $\text{GMM}_{\ln M}$  fits. The so-obtained estimations for  $H$  are  $H \simeq 0.14$  (for S&P) and  $H \simeq 0.13$  (for CAC40). For both indices one gets  $\lambda^2 \simeq 0.05$ .

### ***Yahoo Finance database (YF)***

We collected historical daily open, high, low close price time-series of 296 individual stocks (from S&P 500 and CAC 40 indices) from Yahoo Finance<sup>2</sup>. Stocks were taken

<sup>2</sup><http://YahooFinance.com>

from either the S&P 500 index (historical data from 1985-01-01 to 2021-12-31) or the CAC 40 index ((historical data from 2000-01-01 to 2021-12-31). For each stock, we constructed a proxy of the daily volatility using the Garman-Klass (GK) estimator described in [108]. We checked, using Oxford-Man data indices, that such a GK estimator provides results that are consistent with the intraday estimators of realized volatility. For all stocks, we performed a  $\text{GMM}_{\ln M}$  estimation of  $H$  and  $\lambda^2$ . In Fig. 4.7, following the exact same structure as Fig. 4.6, we illustrated such experiments with two examples, namely Apple and Bank of America realized volatility. Again we observe that stock volatility fluctuations seem to be long-term correlated. For the selected two stocks, the estimated values of  $H$  are respectively 0.01 and 0.02.

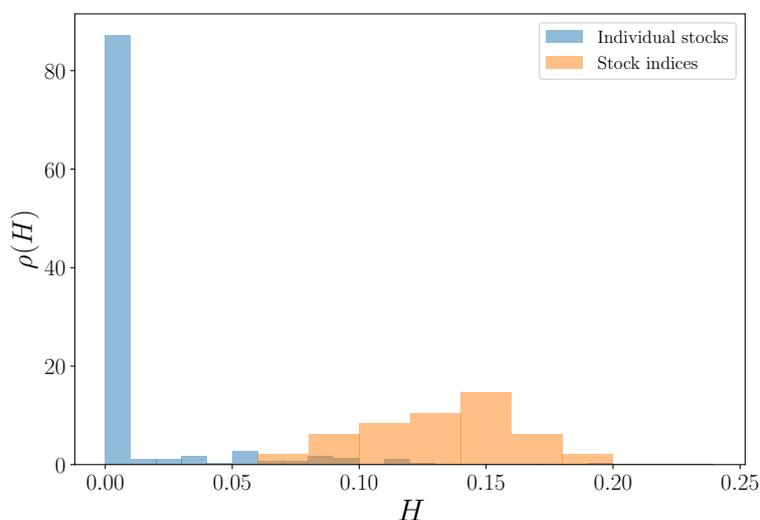


Figure 4.8: Estimation of the probability density distribution of Hurst exponent estimation  $\hat{H}$  for the 296 individual stocks (blue histogram) of the YF database and for the 24 stock market indices (orange histogram) of the OIQFRL database.

The so-obtained estimations on all the 24 stock indices of the OIQFRL database and the 296 individual stocks of the YF database are summarised in Figures 4.8 and 4.9. In Fig. 4.8, we have reported the normalized histograms of the estimation  $\hat{H}$  for the Hurst exponents of the stock indices and the individual stocks of the two datasets. We can observe that the two distributions are quite different: while the Hurst exponents of the stock market indices are spread around  $H \simeq 0.13$  with a rather large dispersion (corresponding to an RMS of 0.03), the distribution of  $H$  values of individual stocks is mainly peaked around a very small average value  $H \simeq 0.01$  (with an RMS of 0.015). Therefore, it clearly appears that the log-volatility of stock indices is much more regular than the log-volatility of individual stocks, which is well described by a multifractal model characterized by  $H = 0$ . Moreover, in agreement with the findings of [65] (and in contrast with the results reported in [36]), Stock indices are confirmed to be well

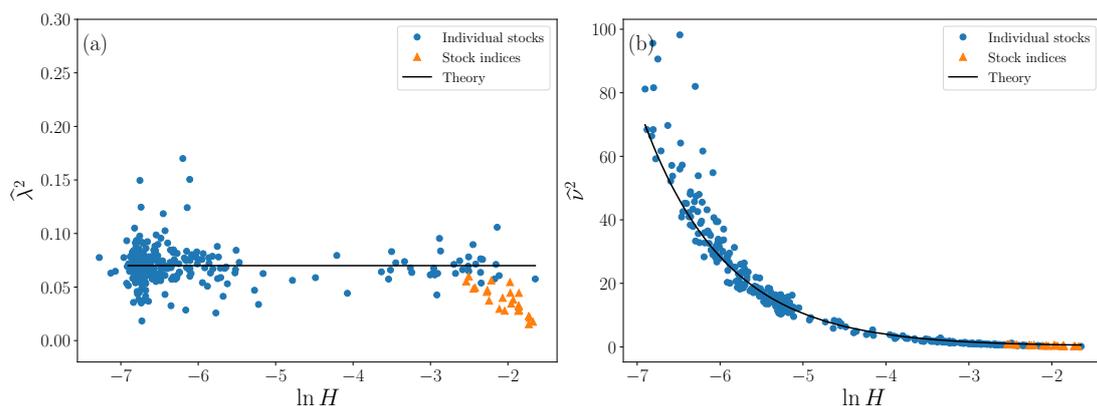


Figure 4.9: Estimation of the intermittency and variance parameters as a function of the estimated Hurst exponent. (a) Estimated intermittency coefficient  $\hat{\lambda}^2$  as a function of the logarithm of the estimated Hurst exponent  $H$ . The solid line represents the constant value  $\lambda^2 = 0.07$  representing the best fit of individual stock data. (b) Estimated variance coefficient  $\hat{\nu}^2$  as a function of the logarithm of the estimated Hurst exponent  $H$ . The solid line represents the log S-fBM expression (4.4.54). In (a) and (b) blue dots represent the individual stock data while orange triangle are associated with estimations from indices times series.

described by a "rough volatility" model with a typical value of the Hurst exponent close to  $H = 0.1$ .

As far as the intermittency coefficient  $\lambda^2$  is concerned, we reported in Fig. 4.9(a) the estimated values  $\hat{\lambda}^2$  for the 296 individual stocks (blue bullets) and the 24 stock indices (orange triangles) as a function of the logarithm of the estimated exponent  $\hat{H}$ . We can see that all the points are distributed around the value  $\lambda^2 \simeq 0.07$  for stocks and  $\lambda^2 \simeq 0.05$  for indices. In contrast, if one estimates the variance parameter  $\nu^2$ , one observes a very large dispersion of its values. Actually, as it can be checked in Fig. 4.9(b), the data closely follow the curve  $\nu^2 = \frac{0.07}{H(1-2H)}$  as represented by the solid line. Whether  $\hat{H}$  varies because  $H$  itself is varying or because of estimation errors, it appears that  $\nu^2$  is related to  $\hat{H}$  through the relationship (4.4.54). This suggests that the intermittency coefficient  $\lambda^2$  is more likely to be the pertinent parameter to account for volatility fluctuations. Moreover, this latter quantity appears to be "almost universal" with a value  $\lambda^2 \simeq 0.07$  for stocks and 0.05 for indices.

## 4.6 Conclusion

We have introduced the log S-fBM, a class of log-normal "rough" random measures  $M_{H,T}(dt)$  that converge, when  $H \rightarrow 0$ , to the log-normal multifractal random measure. This model allows us to consider, within the same framework, the two popular classes of multifractal ( $H = 0$ ) and rough volatility ( $0 < H < 1/2$ ) models. Besides the roughness exponent  $H$ , the model involves 3 supplementary parameters:  $\sigma^2 = \frac{\mathbb{E}(M_{H,T}(dt))}{dt}$  that

provides the mean value of  $M_{H,T}([t, t'])$ , the intermittency coefficient  $\lambda^2$  which is related to the variance of  $\ln M_{H,T}$  and the correlation length  $T$  (also referred to as the “integral scale” in the multifractal literature) above which the process values are independent. The second-order properties are studied and notably, we have computed the correlation function of  $\ln M_{H,T}$  to the first order in  $\lambda^2$ . By studying the self-similarity properties of  $M_{H,T}(dt)$  when one changes the correlation length  $T$ , it appears that one cannot estimate  $T$  and  $\sigma^2$  in the “high-frequency” estimation regime, i.e., if one observes, at a small scale  $\Delta \ll T$ , a single sample of  $M_{H,T}$  over an interval of length  $L = \mathcal{O}(T)$ .

We design two efficient GMM estimation methods,  $\text{GMM}_M$  and  $\text{GMM}_{\ln M}$  based on the expressions of respectively  $M_{H,T}$  and  $\ln M_{H,T}$  correlation functions. We provide theoretical arguments and numerical evidence showing that very much like the method introduced in [14],  $\text{GMM}_{\ln M}$  provides an efficient estimation of  $H$  and  $\lambda^2$  even in the high-frequency asymptotic regime.

We illustrate on various numerical examples that, when  $H < 1/2$ , the most pertinent parameter for accounting for volatility fluctuations is not, as it is always used in the rough volatility literature [65, 36], the variance parameter  $\nu^2 = \frac{\lambda^2}{H(1-2H)}$ , but the intermittency parameter  $\lambda^2$ . Indeed the estimation of the variance parameter is shown to fluctuate a lot and to strongly depend on the estimation error on  $H$ .

Finally, when calibrating the log S-fBM model on a large set of empirical daily volatility data, we observe that stock market indices have values around  $H = 0.1$  (close to a rough volatility behavior) whereas individual stocks are characterized by values of  $H$  that can be very close to 0 (close to a multifractal volatility behavior). Moreover, not surprisingly, the estimations of the intermittency coefficient  $\lambda^2$  are much more robust than the ones of the variance parameter  $\nu^2$ , and its value seems to be quite universal and spread around  $\lambda^2 = 0.07$  for stocks and  $\lambda^2 = 0.05$  for stock market indices in agreement with the values formerly reported for the multifractal model [14].

## 4.7 Appendix

### 4.7.1 Construction of the S-fBM process

In this Appendix, we explain in every details how the S-fBM process  $\omega_{H,T}(t)$  is defined. It depends on three parameters :

- the (Hurst) parameter  $H \in ]0, 1[$ ,
- the decorrelation time scale  $T > 0$
- and the coefficient  $\lambda^2 > 0$  which is linked to the variance parameter  $\nu^2$  by

$$\nu^2 = \frac{\lambda^2}{H(1-2H)}.$$

This parameter will be referred to as the intermittency parameter since it controls the intensity of intermittent “bursts” observed in  $M_{H,T}$  and it is the name given to that quantity in the framework of MRM.

### Construction of the S-fBM process $\omega_{H,T}(t)$

In the upper half-plane  $(t, h) \in \mathcal{S} = \mathbb{R} \times \mathbb{R}^*$ , we first consider the area  $C_{\ell,T}(t^*)$  illustrated in Fig. 4.10 which is defined as:

$$C_{\ell,T}(t^*) = \{(t, s) | h > \ell, |t - t^*| < \frac{1}{2} \min(h, T)\}. \quad (4.7.55)$$

For  $\ell = 0$ , we will use the notation  $C_T(t^*) = C_{0,T}(t^*)$ .

We then consider in  $\mathcal{S}$  a non homogeneous Gaussian white noise  $dG_H(t, h)$  of variance:

$$dp_H(t, h) = \mathbb{E} \left( dG_H(t, h)^2 \right) = \lambda^2 h^{2H-2} dh dt. \quad (4.7.56)$$

We will see below that  $H$  is the analog of the Hurst parameter of the fBM process.

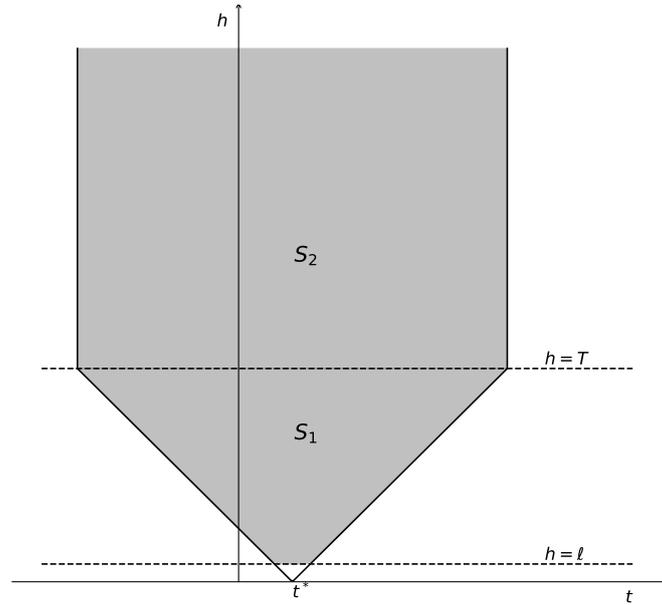


Figure 4.10: Definition of time-scale domain  $C_{\ell,T}(t)$

We then define the Gaussian process  $\omega_{H,T}(t)$  as:

$$\omega_{H,T}(t) = \mu_{H,T}(t) + \int_{C_T(t)} dG_H, \quad (4.7.57)$$

where  $\mu_{H,T}(t)$  is a normalising constant such that

$$\mathbb{E}[e^{\omega_{H,T}(t)}] = 1. \quad (4.7.58)$$

### Covariance function of the S-fBM process $\omega_{H,T}(t)$

As a Gaussian process, the S-fBM is mainly characterized by its covariance function.

This covariance can be directly calculated as the variance of integral of the random measure  $dG_H(t, h)$  on the overlapping area of  $C_T(t_1)$  and  $C_T(t_2)$  displayed in Fig. 4.11:

$$\text{Cov}(\omega_{H,T}(t_1), \omega_{H,T}(t_2)) = \text{Var} \left( \int_{C_{\ell,T}(t_1) \cap C_{\ell,T}(t_2)} dG_H \right). \quad (4.7.59)$$

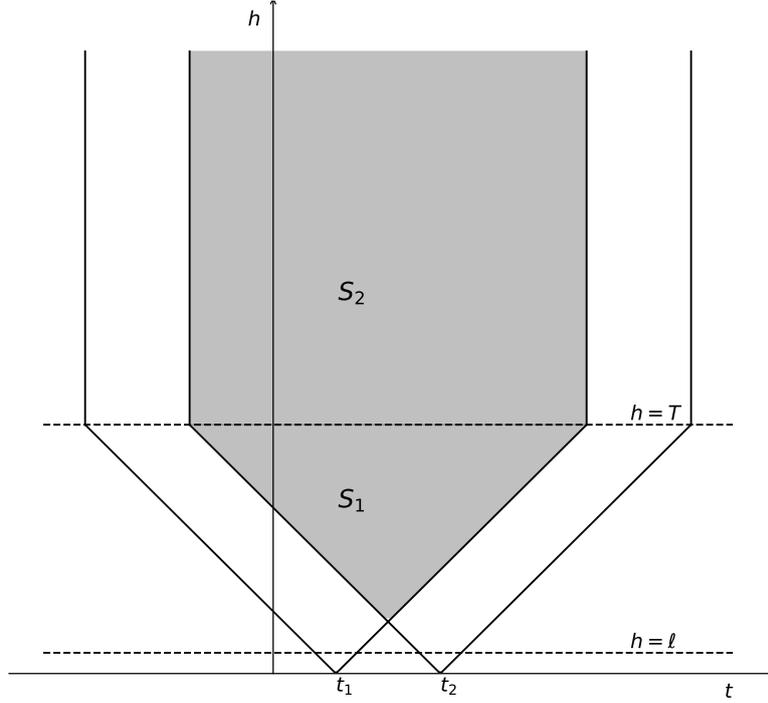


Figure 4.11: The overlapping area

Let us assume, without loss of generality, that  $t_2 > t_1$  and denote  $\tau = t_2 - t_1$ . When  $\tau > T$ ,  $C_T(t_1) \cap C_T(t_2) = \emptyset$  and thus  $\text{Cov}(\omega_{H,T}(t_1), \omega_{H,T}(t_2)) = 0$ . For  $\tau < T$ , we have, using the notations of Fig. 4.11,

$$\begin{aligned} \text{Cov}(\omega_{H,T}(t_1), \omega_{H,T}(t_2)) &= \int_{S_1 \cup S_2} dp_H(h, t) \\ &= \int_{S_1} dp_H(t, h) + \int_{S_2} dp_H(h, t). \end{aligned} \quad (4.7.60)$$

Using (4.7.56), we have for the first term,

$$\begin{aligned} \int_{S_1} dp_H(h, t) &= \lambda^2 \int_{t_2-t_1}^T h^{2H-2} dh \int_{t_2-h/2}^{t_1+h/2} dt \\ &= \lambda^2 \int_{\tau}^T h^{2H-2} (h - \tau) \\ &= \frac{\lambda^2}{2H} (T^{2H} - \tau^{2H}) - \frac{\lambda^2}{2H-1} \tau (T^{2H-1} - \tau^{2H-1}). \end{aligned} \quad (4.7.61)$$

For the second term,

$$\begin{aligned} \int_{S_2} p(h, t) dh dt &= \lambda^2 (T - \tau) \int_T^\infty h^{2H-2} dh \\ &= -\frac{\lambda^2}{2H-1} (T - \tau) T^{2H-1}. \end{aligned} \quad (4.7.62)$$

By composing the results above,

$$\text{Cov}(\omega_{H,T}(t_1), \omega_{H,T}(t_2)) = \frac{\lambda^2}{2H(1-2H)} [T^{2H} - \tau^{2H}]. \quad (4.7.63)$$

Similarly, if we consider a strictly positive  $\ell$  and  $\tau < \ell$ , direct calculation shows:

$$\text{Cov}(\omega_{H,T}(t_1), \omega_{H,T}(t_2)) = \lambda^2 \left[ \frac{1}{2H(1-2H)} (T^{2H} - \ell^{2H}) + \frac{\ell^{2H}}{1-2H} \left(1 - \frac{\tau}{\ell}\right) \right]. \quad (4.7.64)$$

### 4.7.2 Constructing fBM process from Takenaka field and proof of Proposition 3

Let us denote by  $C(t_0)$  the full cone obtained by considering  $T \rightarrow \infty$  in  $C_T(t)$ :

$$C(t_0) = \{(t, h) | h > 0, |t - t^*| < \frac{h}{2}\} \quad (4.7.65)$$

and consider the domain:

$$D(t) = C(t) \Delta C(0) = D_+(t) \cup D_-(t), \quad (4.7.66)$$

where  $\Delta$  stands for the symmetric difference between two sets and  $D_+(t)$ ,  $D_-(t)$  are the two disjoint sets:

$$D_+(t) = C(t) - C(0) \quad \text{and} \quad D_-(t) = C(0) - C(t).$$

Along the the same line as definition (4.7.57), let us define the Gaussian processes:

$$\omega_\pm(t) = \int_{D_\pm(t)} dG_H \quad \text{and} \quad B_H(t) = \omega_+(t) + \omega_-(t). \quad (4.7.67)$$

Notice that  $\omega_\pm(0) = 0$  and therefore  $B_H(0) = 0$ . It is easy to show that, after a little algebra that, for  $0 \leq t_1 \leq t_2$ :

$$\begin{aligned} \mathbb{E}[\omega_-(t_1) \omega_-(t_2)] &= \frac{\lambda^2}{2H(1-2H)} t_1^{2H}, \\ \mathbb{E}[\omega_+(t_1) \omega_+(t_2)] &= \frac{\lambda^2}{2H(1-2H)} (t_2^{2H} - |t_2 - t_1|^{2H}), \\ \mathbb{E}[\omega_\pm(t_1) \omega_\mp(t_2)] &= 0. \end{aligned}$$

It directly results that:

$$\gamma(t_1, t_2) = \mathbb{E}[B_H(t_1) B_H(t_2)] = \frac{\nu^2}{2} (t_1^{2H} + t_2^{2H} - |t_1 - t_2|^{2H}) \quad (4.7.68)$$

with  $\nu^2 = \frac{\lambda^2}{H(1-2H)}$ . Since  $B_H(0) = 0$ , we see that  $B_H(t)$  is nothing but a fractional Brownian motion. This construction corresponds to the 1D version of Takenaka fractional Brownian fields as discussed in [130].

In order to prove Proposition 3, let us first work out

$$R_T(t, s) = \mathbb{E} [(\omega_{H,T}(t) - \omega_{H,T}(0))B_H(s)] .$$

This amounts to compute the "areas" of the intersections of  $D_{\pm}(s)$  with  $C_T(t)$  and  $C_T(0)$  respectively. After a little algebra, one gets, for any  $0 \leq t, s \leq T$ :

$$\begin{aligned} R_T(t, s) &= \gamma(t, s) + \gamma'(t, s) \text{ with} \\ \gamma'(t, s) &= B \left( \left( \frac{T}{2} \right)^{2H} + \left( |t-s| + \frac{T}{2} \right)^{2H} - \left( t + \frac{T}{2} \right)^{2H} - \left( s + \frac{T}{2} \right)^{2H} \right) \end{aligned}$$

where  $B$  is a positive constant depending on  $H$  and  $\lambda^2$ . Similarly, if

$$S_T(t, s) = \mathbb{E} [(\omega_{H,T}(t) - \omega_{H,T}(0))(\omega_{H,T}(s) - \omega_{H,T}(0))] ,$$

one has if  $0 \leq s, t \leq T$ :

$$S_T(t, s) = \gamma(t, s) . \quad (4.7.69)$$

Let

$$Z_{H,T}(t) = \omega_{H,T}(t) - \omega_{H,T}(0) - B_H(t)$$

and

$$d(t, s) = \left[ \mathbb{E} (Z_{H,T}(t) - Z_{H,T}(s))^2 \right]^{1/2} .$$

By expanding the square  $(Z_{H,T}(t) - Z_{H,T}(s))^2$  one directly obtains:

$$\begin{aligned} d(t, s)^2 &= 2\gamma(t, t) - 2R_T(t, t) + 2\gamma(s, s) - 2R_T(s, s) - 4\gamma(t, s) + 4R_T(t, s) \\ &= -2\gamma'(t, t) - 2\gamma'(s, s) + 4\gamma'(t, s) \\ &= 4B \left( \left( |t-s| + \frac{T}{2} \right)^{2H} - \left( \frac{T}{2} \right)^{2H} \right) . \end{aligned}$$

Therefore,  $\forall t_0, t, s < t_0$ , we have when  $T \rightarrow \infty$ :

$$d(t, s) = CT^{H-\frac{1}{2}}|t-s|^{1/2} + o(T^{H-\frac{1}{2}}) . \quad (4.7.70)$$

One can consider  $d(t, s)$  as a metric and define  $N(t_0, \varepsilon)$  as the number of boxes of  $d$ -radius  $\varepsilon$  need to cover the set  $[0, t_0]$ . Let  $D = \sup_{t, s \in [0, t_0]} d(t, s)$ . Then, according to Dudley inequality [113], there exists a positive universal constant  $K$  such that:

$$\mathbb{E} \left( \sup_{t \in [0, t_0]} |Z_H(t)| \right) \leq K \int_0^D \sqrt{\log N(t_0, \varepsilon)} d\varepsilon . \quad (4.7.71)$$

From Eq. (4.7.70), one has  $D \sim CT^{H-\frac{1}{2}}t_0^{\frac{1}{2}}$ . Moreover, one has

$$N(t_0, \varepsilon) \simeq 1 + \left\lfloor \frac{t_0 T^{2H-1} C^2}{\varepsilon^2} \right\rfloor = 1 + \left\lfloor \frac{D^2}{\varepsilon^2} \right\rfloor ,$$

where  $\lfloor x \rfloor$  stands for the largest integer not greater than  $x$ . We then have when  $\varepsilon$  is small with respect to  $D$ :

$$\log N(t_0, \varepsilon) \simeq 2 \ln(D) - 2 \ln(\varepsilon).$$

Thus

$$\int_0^D \sqrt{\log N(t_0, \varepsilon)} d\varepsilon \sim \int_0^D \sqrt{2 \ln(D) - 2 \ln(\varepsilon)} d\varepsilon = \sqrt{\frac{\pi}{2}} D,$$

since for  $H < 1/2$ ,  $D \rightarrow 0$  when  $T \rightarrow \infty$ . Proposition 3 follows directly from inequality (4.7.71).

### 4.7.3 The case $H \rightarrow 0$ : Convergence towards the MRM log-normal measure

We now examine the case  $H = 0$  in the geometric construction above. The definition of  $C_{\ell, T}$  remains unchanged and we consider the Gaussian random noise when  $H = 0$ ,  $dG_0(t, h)$  of variance:

$$dp_0(t, h) = \lambda^2 h^{-2} dh dt \quad (4.7.72)$$

Then we define a random process  $\omega_{\ell, T}$  as previously

$$\omega_{\ell, T}(t) = \mu_{\ell, T} + \int_{C_{\ell, T}(t)} dG_0. \quad (4.7.73)$$

As proven in [20], provided  $\mu_{\ell, T}$  is chosen such that  $\mathbb{E}e^{\omega_{\ell, T}(t)} = 1$ , when  $\ell \rightarrow 0$  we have

$$\widetilde{M}_{\ell, T}(dt) = e^{\omega_{\ell, T}(t)} dt \xrightarrow[\ell \rightarrow 0]{w} \widetilde{M}_T(dt), \quad (4.7.74)$$

where  $\xrightarrow{w}$  stands for the weak convergence and where  $\widetilde{M}_T(dt)$  is the so-called log-normal "Multifractal Random Measure" (MRW), a non trivial singular continuous random measure with exact multifractal properties [115, 20, 14].

In [115, 20], it is also shown that the covariance of  $\omega_{\ell, T}$  reads (for  $\tau \geq 0$ ):

$$\text{Cov}(\omega_{\ell, T}(t), \omega_{\ell, T}(t + \tau)) = \begin{cases} \lambda^2 \ln\left(\frac{T}{\tau}\right) & \text{if } \ell \leq \tau \leq T \\ \lambda^2 \left(\ln\left(\frac{T}{\ell}\right) + 1 - \frac{\tau}{\ell}\right) & \text{if } \tau \leq \ell \\ 0, & \text{otherwise} \end{cases} \quad (4.7.75)$$

We can remark that this expression of the covariance of  $\omega_{\ell, T}$  in the range  $\tau \geq \ell$ , can be recovered from Eq. (4.7.63), (4.7.64) when  $H \rightarrow 0$ .

Let us show a strong mean square convergence of S-fBM to MRM when  $H \rightarrow 0$  as claimed in Proposition 3.

Since  $\widetilde{M}(t) = \widetilde{M}([0, t])$  is regular enough, in order to establish the weak convergence we just have to prove that  $\forall t$ ,

$$\lim_{H \rightarrow 0} \mathbb{E}[(M_{H, T}[0, t] - \widetilde{M}_T[0, t])^2] = \lim_{H \rightarrow 0} \lim_{\ell \rightarrow 0} \mathbb{E}[(M_{H, T}[0, t] - \widetilde{M}_{\ell, T}[0, t])^2] = 0. \quad (4.7.76)$$

Before starting, it is useful to calculate the covariance between  $\omega_{H,T} = \omega_{\ell=0,H,T}$  and  $\omega_{\ell,T} = \omega_{\ell,H=0,T}$ . Following similar computation as in Appendix 4.7.1,

$$\text{Cov}(\omega_{H,T}(t_1), \omega_{\ell,T}(t_2)) = \begin{cases} \frac{\lambda^2}{H(1-H)}(T^H - \tau^H) & \text{if } \ell \leq \tau \leq T \\ \lambda^2 \left( \frac{1}{H(1-H)}(T^H - \ell^H) + \frac{\ell^H}{1-H}(1 - \frac{\tau}{\ell}) \right) & \text{if } \tau < \ell \\ 0 & \text{otherwise} \end{cases} \quad (4.7.77)$$

By expanding the square in Eq. (4.7.76), one gets:

$$\mathbb{E}[(M_{H,T}[0, t] - \widetilde{M}_{\ell,T}[0, t])^2] = \int_0^t \int_0^t \left( e^{\text{Cov}[\omega_{H,T}(u), \omega_{H,T}(v)]} + e^{\text{Cov}[\omega_{\ell,T}(u), \omega_{\ell,T}(v)]} - 2e^{\text{Cov}[\omega_{H,T}(u), \omega_{\ell,T}(v)]} \right) du dv. \quad (4.7.78)$$

Since, for a symmetric function  $f$ , one has:

$$\int_0^t \int_0^t f(u - v) du dv = 2 \int_0^t (t - z) f(z) dz.$$

Then the previous expression becomes:

$$\mathbb{E}[(M_{H,T}[0, t] - \widetilde{M}_{\ell,T}[0, t])^2] = 2 \int_0^t dz (t - z) \left( e^{\text{Cov}[\omega_{H,T}(0), \omega_{H,T}(z)]} + e^{\text{Cov}[\omega_{\ell,T}(0), \omega_{\ell,T}(z)]} - 2e^{\text{Cov}[\omega_{H,T}(0), \omega_{\ell,T}(z)]} \right). \quad (4.7.79)$$

Let us split this integral as a sum of two integrals,  $I_1$  and  $I_2$ , according to whether one considers the integration domains  $z > \ell$  and  $z \leq \ell$  respectively. In the first case, by replacing the covariance with their expressions, one gets:

$$I_1 = \int_{\ell}^t (t - z) \left( e^{\frac{\lambda^2}{2H(1-2H)}(T^{2H} - z^{2H})} + \left( \frac{T}{z} \right)^{\lambda^2} - 2e^{\frac{\lambda^2}{H(1-H)}(T^H - z^H)} \right) dz. \quad (4.7.80)$$

Since  $\lambda^2 < 1$ , one can safely take  $\ell \rightarrow 0$  in the lower integral bound and then, thanks to dominated convergence theorem, observe that  $I_1$  converges to 0 when  $H \rightarrow 0$  since the expression inside the integral vanishes in this limit. The second integral, when  $z \leq \ell$  is:

$$I_2 = \int_0^{\ell} (t - z) \left( e^{\frac{\lambda^2}{2H(1-2H)}(T^{2H} - \tau^{2H})} + e^{\lambda^2 \left( \frac{T}{\ell} + 1 - \frac{z}{\ell} \right)} - 2e^{\lambda^2 \left( \frac{T^H - \ell^H}{H(1-H)} + \frac{\ell^H}{1-H} \left( 1 - \frac{z}{\ell} \right) \right)} \right) dz. \quad (4.7.81)$$

For  $0 \leq z \leq \ell$ , the first and last terms inside the integral can be bounded by a constant that does not depend on  $\ell$  while the second term can be bounded by  $C\ell^{-\lambda^2}$ . Therefore we can see that, if  $\lambda^2 < 1$ ,  $I_2 \rightarrow 0$  when  $\ell \rightarrow 0$ . This concludes the proof.

#### 4.7.4 Proof of Eqs. (4.3.21) and (4.3.22)

Let us compute the analytical expression of  $C_M(\tau, \Delta) = \mathbb{E}[M_{H,T,\Delta}(t)M_{H,T,\Delta}(t + \tau)]$  and establish expressions (4.3.21) and (4.3.22). For that purpose, let us first remark that from the definition (4.3.18) of  $M_{H,T,\Delta}(t)$  and from the expression (4.2.8) of the covariance of  $\omega_{H,T}(t)$ , we have (when  $\tau < T$ ):

$$C_M(\tau, \Delta) = K_1 \int_0^\Delta du \int_\tau^{\tau+\Delta} dv e^{-K_2|u-v|^{2H}} \quad (4.7.82)$$

with  $K_1 = e^{K_2 T^{2H}}$  and  $K_2 = \frac{\lambda^2}{2H(1-2H)}$ .

Moreover, let us prove that, if  $f(z)$  is a symmetric function, then

$$\int_0^\Delta du \int_\tau^{\tau+\Delta} dv f(u-v) = \int_0^\Delta dz (\Delta - z) (f(z + \tau) + f(z - \tau)). \quad (4.7.83)$$

Indeed, as shown in [127], we have, when  $\tau = 0$ :

$$\int_0^\Delta du \int_0^\Delta dv f(u-v) = 2 \int_0^\Delta dz (\Delta - z) f(z).$$

In the l.h.s. of (4.7.83), let us set  $v' = v - \tau$  and use respectively symmetry argument and previous expression to obtain

$$\begin{aligned} \text{l.h.s.} &= \int_0^\Delta du \int_0^\Delta dv f(\tau + u - v) \\ &= \int_0^\Delta du \int_0^u dv f(\tau + |u - v|) + \int_0^\Delta du \int_u^\Delta dv f(\tau - |u - v|) \\ &= \frac{1}{2} \left( \int_0^\Delta du \int_0^\Delta dv f(\tau + |u - v|) + \int_0^\Delta du \int_0^\Delta dv f(\tau - |u - v|) \right) \\ &= \int_0^\Delta dz (\Delta - z) (f(z + \tau) + f(z - \tau)). \end{aligned}$$

By using (4.7.83) in (4.7.82), we get:

$$\begin{aligned} C_M(\tau, \Delta) &= K_1 \int_0^\Delta dz (\Delta - z) \left( e^{-K_2|\tau+z|^{2H}} e^{-K_2|\tau-z|^{2H}} \right) \\ &= K_1 \int_\tau^{\tau+\Delta} dz (\Delta + \tau - z) e^{-K_2|z|^{2H}} + K_1 \int_{\tau-\Delta}^\tau dz (\Delta - \tau + z) e^{-K_2|z|^{2H}} \\ &= F(\tau + \Delta) + F(\tau - \Delta) - 2F(\tau), \end{aligned}$$

where we have denoted

$$F(x) = K_1 \left( x \int_0^x dz e^{-K_2|z|^{2H}} - \int_0^x dz z e^{-K_2|z|^{2H}} \right).$$

If one considers the lower-incomplete Gamma function  $\gamma(a, z)$ ,

$$\gamma(a, x) = \int_0^x t^{a-1} e^{-t} dt$$

and makes the change of variable  $t = K_2|z|^{2H}$  in previous integrals, one obtains the following exact expression for  $F(x)$ :

$$F(x) = \frac{K_1}{2H} \left( \frac{x}{K_2^{\frac{1}{2H}}} \gamma\left(\frac{1}{2H}, K_2 x^{2H}\right) - \frac{1}{K_2^{\frac{1}{H}}} \gamma\left(\frac{1}{H}, K_2 x^{2H}\right) \right),$$

which corresponds to Eq. (4.3.22). When  $H = 0$ , i.e. for  $\widetilde{M}_T$ , one can show that the former expression reduces to:

$$F(x) = \frac{T^{\lambda^2}}{(2 - \lambda^2)(1 - \lambda^2)} x^{2-\lambda^2}.$$

#### 4.7.5 Proof of proposition 8

In this section we provide a proof of Proposition 8 based on small intermittency approximation of Proposition 6. Let  $\Delta = 1$  and  $N = \frac{L}{\Delta} = L$  be the number of samples  $M_{H,T,1}$  in the interval  $[0, L]$ . We will suppose that  $L \rightarrow \infty$  with  $T = CL$ , so that we are in the high frequency regime. Let us consider the empirical mean:

$$\widehat{\mu}_N = \frac{1}{N} \sum_{k=1}^N \ln M_{H,T,\Delta=1}(k) \quad (4.7.84)$$

and define the ‘‘centered’’ random variable:

$$Z(k) = \ln M_{H,T,1}(k) - \widehat{\mu}_N. \quad (4.7.85)$$

If  $C_Z(k) = \text{Cov}[Z(j), Z(j+k)]$ , one has obviously:

$$C_Z(k) = C_{\ln M}(1, k) - \text{Var}[\widehat{\mu}_N]. \quad (4.7.86)$$

One can use Proposition 6 to compute, to the first order in  $\lambda^2$ , all terms in Eq. (4.7.86). Indeed, the expression of  $C_{\ln M}(k)$  is provided by Proposition 7 (Eq. (4.3.28)) and order to compute  $\text{Var}[\widehat{\mu}_{\Delta,N}]$ , one can use Prop. 6 to show that, to the first order in  $\lambda^2$ ,

$$\text{Var}[\widehat{\mu}_N] = \text{Var}[\ln M_{H,T,\Delta=L}]$$

and therefore, from expression (4.3.28), one has:

$$\text{Var}[\widehat{\mu}_{\Delta,N}] = \frac{\lambda^2}{2H(1-2H)} \left( T^{2H} - \frac{L^{2H}}{(2H+1)(H+1)} \right). \quad (4.7.87)$$

It thus results that:

$$C_Z(k) = \frac{\lambda^2}{2H(1-2H)} \left( \frac{L^{2H}}{(1+2H)(1+H)} - \frac{|k+1|^{2H+2} + |k-1|^{2H+2} - 2|k|^{2H+2}}{(2H+1)(2H+2)} \right). \quad (4.7.88)$$

Let us consider the empirical covariance:

$$\widehat{C}_{\ln M}(1, k) = N^{-1} \sum_{k=1}^{N-k} \left( \ln M_{H,T,1}(j) - \widehat{\mu}_N \right) \left( \ln M_{H,T,1}((j+k)) - \widehat{\mu}_N \right), \quad (4.7.89)$$

Since one has  $\mathbb{E}[D(n)] = \mathbb{E}[\widehat{C}_{\ln M}(1, n) - \widehat{C}_{\ln M}(1, 0)] = C_Z(k) - C_Z(0) = \widetilde{D}(n)$  (as defined in Eq. (4.4.44)), in order to prove Eq. (4.4.43), it is sufficient to show that

$$\lim_{N \rightarrow \infty} \text{Var}[D(n)] = 0$$

To that end, remark that, from the definition of  $D(n)$ ,

$$\begin{aligned} \text{Var}[D(n)] &= N^{-2} \sum_{i=1}^{N-n} \sum_{j=1}^{N-n} \text{Cov}[Z(i)Y(i), Z(j)Y(j)] & (4.7.90) \\ &= N^{-2} \sum_{i=1}^{N-n} \text{Var}[Z(i)Y(i)] + 2N^{-2} \sum_{i=1}^{N-n} \sum_{j=i+1}^{N-n} \text{Cov}[Z(i)Y(i), Z(j)Y(j)] \end{aligned}$$

where we have denoted

$$Y(i) = Z(i+n) - Z(i).$$

From proposition 6, because  $\Omega_{T,H,\Delta}(t)$  is a Gaussian process, we have, to the first order in  $\lambda^2$ ,

$$\begin{aligned} \text{Cov}[Z(i)Y(i), Z(j)Y(j)] &= \text{Cov}(Z(i), Z(j)) \text{Cov}(Y(i), Y(j)) + \text{Cov}(Z(i), Y(j)) \text{Cov}(Z(j), Y(i)) \\ &= C_Z(j-i) \left( 2C_Z(j-i) - C_Z(i-j+n) - C_Z(j-i-n) \right) \\ &\quad + \left( C_Z(j+n-i) - C_Z(j-i) \right) \left( C_Z(j-i-n) - C_Z(j-i) \right). \end{aligned}$$

Thereby, from the expression (4.7.88) of  $C_Z(k)$ , after a little algebra, one can show that there exists a constant  $C$  such that

$$\text{Cov}[Z(i)Y(i), Z(j)Y(j)] \leq CN^{2H}(1 + |i-j|)^{2H-2}.$$

Then, Eq. (4.7.91) gives:

$$\text{Var}[D(n)] \leq CN^{2H-1} + 2CN^{2H-2} \sum_{i=1}^N \int_i^N x^{2H-2} dx \leq C'N^{2H-1} \quad (4.7.92)$$

and thus, if  $H < \frac{1}{2}$ ,

$$\lim_{N \rightarrow \infty} \text{Var}[D(n)] = 0.$$

This concludes the proof of Eq. (4.4.43).

## Acknowledgement

This research is partially supported by the Agence Nationale de la Recherche as part of the Investissements d'avenir program (reference ANR-19-P3IA-0001; PRAIRIE 3IA Institute).

## Chapter 5

# Disentangling market participants contributions toward covariance of price between different assets

### 5.1 Introduction

Modeling the covariance between price processes of different assets remained an interesting yet practical subject. For example, liquidity providers need to monitor correlated price changes in different markets and update their offer prices. In this chapter, we highlight modeling price covariance in electronic markets.

Literature deviating from various origins has appeared over the past ten years, which could be roughly divided into two categories. From a physics background, a linear propagator model is proposed in [34]. This model comes with the condition that the market is free of arbitrage and price manipulation. We also mention [135] and [134], where the author modeled correlated price changes as responses to previous transactions.

Another stream of literature models correlated price variations as consequences of limit order book events. We mention [57], where a Hawkes process based framework is proposed to capture the lead-lag relationship between the order flows of different assets. The authors argue that it is impossible to capture their so-called lead-lag relationship at low frequency because such an effect is only visible on the high-frequency scale. A similar approach is taken in [38], where multivariate Hawkes processes are applied to model co-jumps of different assets' prices. The authors argued that there exist a large number of high-frequency co-jumps which leads to non-trivial covariance between price processes.

In this chapter, we present a more elaborated model which can be seen as a further development from [127]. To our best knowledge, we propose for the first time a model which allows measuring each market participant's marginal contribution toward the covariance.

This chapter is organized as follows. In Section 5.2 we present our multivariate Hawkes process framework to model price jumps as the cumulative result of order flows, see Eq. (5.2.20), where the covariance is reconstructed from the average intensity of order flows. According to the origin of order flows, we further distinguish the price covariance as exogenous or endogenous in Eq. (5.2.22). In Section 5.2.2, we introduce an approximate relation to the first order of the norm of the kernel matrix. Under reasonable assumptions, we justify Eq. (5.2.35) and Eq. (5.2.38) in the context of modeling high-frequency order flows. Finally, we account for each market participant's marginal contribution toward the covariance in Eq. (5.2.38).

In Section 5.3, we present our dataset and our metrics of empirical covariance. We discuss calibration issues and present our method to reduce the dimension of the calibration problem. We refer readers to [100] as a general survey of measuring realized volatility, whose method is extended to measure realized covariance in this chapter. We also mention [103] and [7], which argue that when multivariate Hawkes processes are applied for estimating order flows, non-parametric fitting is necessary. We apply the non-parametric estimation approach based on the conditional law, which is elaborated in [22] and implemented in the library *Tick* (see [21]).

Empirical results obtained with the composite stocks of the CAC40 index are reported in section 5.3.2. Our result suggests that our Hawkes process based framework is capable of reproducing the price covariance between different assets. Besides, empirical evidence suggests that market participants' contribution toward the covariance is uniform among different pairs of assets.

## 5.2 Modeling covariance of price by multivariate Hawkes process

A multivariate Hawkes process of dimension  $D$  is a counting processes  $N_t := \{N_t^i\}_{i=1}^D$  whose intensity vector depends on past events. The  $i$ -th component of the intensity vector takes the form

$$\lambda_i(t) = \mu^i + \sum_{j=1}^D \int \phi^{ij}(t-s) dN_s^j. \quad (5.2.1)$$

The quantity  $\{\mu^i\}_{i=1}^D$  reads the vector of exogenous intensity. The matrix  $\Phi(t) := \{\phi^{ij}(t)\}_{i,j=1}^D$  reads the kernel matrix. We then define the matrix  $R(t)$  as the sum of convolution of  $\Phi(t)$ :

$$R(t) = \sum_{n=0}^{\infty} \Phi^{*n}(t). \quad (5.2.2)$$

We have shown in the introduction section, that under reasonable assumptions,

$$R(t) = (\mathbf{I}_d - \Phi(t))^{-1}. \quad (5.2.3)$$

We note  $\Phi^{ij}$  and  $R^{ij}$  as integrals of  $\Phi^{ij}(\cdot)$  and  $R^{ij}(\cdot)$  from 0 to  $\infty$ , i.e

$$\Phi^{ij} := \int_0^\infty \phi^{ij}(t) dt, \quad (5.2.4)$$

$$R^{ij} := \int_0^\infty R^{ij}(t) dt. \quad (5.2.5)$$

and we note  $\Phi$  and  $R$  as the matrix of  $\Phi^{ij}$  and  $R^{ij}$ . We now consider order flows of a single asset. We note  $P(t)$  the midprice observed on the LOB at  $t$ . We note by  $\mathcal{N}$  all types of events that we consider and especially  $\mathcal{M} \subset \mathcal{N}$  the types of events that change the midprice. We suppose that order flows are driven by a multivariate Hawkes process whose intensity is described by Eq.(5.2.1). According to the introduction chapter, the average intensity of events of type  $i$  could be expressed as:

$$\Lambda^i = \sum_{k \in \mathcal{N}} R^{ik} \mu^k. \quad (5.2.6)$$

Considering  $i$  and  $j$  as event types, the infinitesimal covariance matrix is defined by

$$C^{ij}(t) dt ds = \mathbb{E}[dN_s^i dN_{s+t}^j] - \Lambda^i \Lambda^j dt ds, \quad (5.2.7)$$

and  $C^{ij}$  could be obtained from  $R(t)$  and  $\Lambda$ :

$$C^{ij}(t) = \sum_{k \in \mathcal{N}} \Lambda^k \int R^{ik}(u) R^{jk}(t+u) du. \quad (5.2.8)$$

The price change over the period  $[t, t + \tau]$  is determined by all orders flows with type  $\mathcal{M}$ :

$$\Delta_\tau P(t) = P(t + \tau) - P(t) = \sum_{i \in \mathcal{M}} \delta_i \int_t^{t+\tau} dN_s^i, \quad (5.2.9)$$

where  $\delta_i$  is the average amount of mid price change caused by an event of type  $i$ . We further assume the no-trend condition, that for all  $\tau$ ,

$$\mathbb{E}[\Delta_\tau P(t)] = 0. \quad (5.2.10)$$

Putting this condition into Eq. (5.2.9),

$$\sum_{i \in \mathcal{M}} \delta_i \Lambda^i = 0. \quad (5.2.11)$$

We now consider two assets denoted by  $S_1$  and  $S_2$ . We note their midprice correspondingly as  $P_1(t)$  and  $P_2(t)$ , where the numerical subscription denotes different assets. We use similar notations as  $\mathcal{N}_1$  (resp.  $\mathcal{N}_2$ ) and  $\mathcal{M}_1$  (resp.  $\mathcal{M}_2$ ) for different types of event of  $S_1$  (resp.  $S_2$ ). Over the period  $[t, t + \tau]$ , we define the covariance between  $P_1(t)$  and  $P_2(t)$  as the correlated price change :

$$\text{Cov}_\tau[S_1, S_2](t) := \text{Cov}[\Delta_\tau P_1(t), \Delta_\tau P_2(t)]. \quad (5.2.12)$$

With the non-arbitrage condition (5.2.10), it leads to

$$\text{Cov}_\tau[S_1, S_2](t) = \mathbb{E}[\Delta_\tau P_1(t) \Delta_\tau P_2(t)], \quad (5.2.13)$$

then

$$\text{Cov}_\tau[S_1, S_2](t) = \mathbb{E}\left[\sum_{i \in \mathcal{M}_1} \sum_{j \in \mathcal{M}_2} \int_t^{t+\tau} \delta_i dN_u^i \int_t^{t+\tau} \delta_j dN_v^j\right]. \quad (5.2.14)$$

Using (5.2.7), the covariance reads

$$\text{Cov}_\tau[S_1, S_2](t) = \sum_{i \in \mathcal{M}_1} \sum_{j \in \mathcal{M}_2} \delta_i \delta_j \int_t^{t+\tau} \int_t^{t+\tau} (C^{ij}(v-u) + \Lambda^i \Lambda^j) dv du, \quad (5.2.15)$$

then using Eq.(5.2.11),

$$\text{Cov}_\tau[S_1, S_2](t) = \sum_{i \in \mathcal{M}_1} \sum_{j \in \mathcal{M}_2} \delta_i \delta_j \int_t^{t+\tau} \int_t^{t+\tau} C^{ij}(v-u) dv du. \quad (5.2.16)$$

By the change of variable  $y = v - u$ ,

$$\text{Cov}_\tau[S_1, S_2](t) = \sum_{i \in \mathcal{M}_1} \sum_{j \in \mathcal{M}_2} \delta_i \delta_j \left[ \int_{-\tau}^{\tau} (\tau - |y|) C^{ij}(y) dy \right]. \quad (5.2.17)$$

Then by considering the case  $\tau \rightarrow \infty$ , we simplify the notation by

$$\text{Cov}[S_1, S_2] := \lim_{\tau \rightarrow \infty} \frac{\text{Cov}_\tau[S_1, S_2]}{\tau}, \quad (5.2.18)$$

and by making the assumption that  $C(y)$  is decreasing fast enough for large lags (when  $y \rightarrow \pm\infty$ ) so that  $yC^{ij}(y) \in L^1$  (in fact it is sufficient to have  $C^{ij}(y) \sim \frac{1}{y^{2+\epsilon}}$ ), we get

$$\text{Cov}[S_1, S_2] = \sum_{i \in \mathcal{M}_1} \sum_{j \in \mathcal{M}_2} \delta_i \delta_j \int_{-\infty}^{\infty} C^{ij}(y) dy. \quad (5.2.19)$$

Replacing  $C^{ij}$  by Eq. (5.2.8) leads to

$$\begin{aligned} \text{Cov}[S_1, S_2] &= \sum_{i \in \mathcal{M}_1} \sum_{j \in \mathcal{M}_2} \delta_i \delta_j \sum_{k \in \mathcal{N}} \Lambda^k \mathbf{R}^{ik} \mathbf{R}^{jk} \\ &= \sum_{k \in \mathcal{N}} \Lambda^k \left( \sum_{i \in \mathcal{M}_1} \delta_i \mathbf{R}^{ik} \right) \left( \sum_{j \in \mathcal{M}_2} \delta_j \mathbf{R}^{jk} \right). \end{aligned} \quad (5.2.20)$$

### 5.2.1 Endogenous covariance and exogenous covariance

Under the framework of multivariate Hawkes processes, the self-excitation and cross-excitation effect between order flows is described by the kernel matrix  $\Phi(\cdot)$ , which is assumed to be constant over time. For order flows of different assets, correlated changes in the exogenous intensity  $\mu$  could also contribute to  $\text{Cov}[S_1, S_2]$ , which is independent of the self-excitation and cross-excitation effect. However, this phenomenon is not well illustrated by Eq. (5.2.20).

In this study, we distinguish two types of covariance: “exogenous” covariance and “endogenous” covariance. Under the framework of multivariate Hawkes processes, the exogenous covariance is a direct consequence of the variation of exogenous intensities, while the endogenous covariance is purely triggered by the self-exciting effect between order flows. We consider the variation of exogenous intensities varies over a daily scale, which is noted by the superscript  $d$ . For each day  $d$ , we suppose that the intensity of the multivariate Hawkes process takes the form:

$$\lambda^{i,d}(t) = \mu^{i,d} + \sum_{j \in \mathcal{N}} \int_0^t \phi^{ij}(t-s) dN_s^j. \quad (5.2.21)$$

We illustrate the difference between exogenous covariance and endogenous covariance with the following decomposition:

$$\begin{aligned} \text{Cov}[S_1, S_2] &= \mathbb{E}[\Delta P_1 \Delta P_2] - \mathbb{E}[\Delta P_1] \mathbb{E}[\Delta P_2] \\ &= \mathbb{E}[\mathbb{E}[\Delta P_1 \Delta P_2 | d]] - \mathbb{E}[\Delta P_1] \mathbb{E}[\Delta P_2] \\ &= \mathbb{E}[\text{Cov}[P_1, P_2 | d]] + \mathbb{E}[\mathbb{E}[\Delta P_1 | d] \mathbb{E}[\Delta P_2 | d]] - \mathbb{E}[\Delta P_1] \mathbb{E}[\Delta P_2]. \end{aligned} \quad (5.2.22)$$

The endogenous covariance is defined as

$$\text{Cov}^{Endo} := \mathbb{E}[\text{Cov}[P_1, P_2 | d]] = \sum_{k \in \mathcal{N}_1 \cup \mathcal{N}_2} \Lambda^{k,d} \left( \sum_{i \in \mathcal{M}_1} \delta_i R^{ik} \right) \left( \sum_{j \in \mathcal{M}_2} \delta_j R^{jk} \right), \quad (5.2.23)$$

while the exogenous covariance is the remaining part:

$$\begin{aligned} \text{Cov}^{Exo} &:= \text{Cov}[P_1, P_2] - \text{Cov}^{Endo} \\ &= \mathbb{E}[\mathbb{E}[\Delta P_1 | d] \mathbb{E}[\Delta P_2 | d]] - \mathbb{E}[\Delta P_1] \mathbb{E}[\Delta P_2] \\ &= \mathbb{E}[\mathbb{E}[\Delta P_1 | d] \mathbb{E}[\Delta P_2 | d]]. \end{aligned} \quad (5.2.24)$$

The exogenous covariance is zero only if and only if the daily variations of two assets are independent, e.g.

$$\mathbb{E}[\Delta P_1 | d] \perp\!\!\!\perp \mathbb{E}[\Delta P_2 | d]. \quad (5.2.25)$$

### 5.2.2 Measuring market participants’ marginal contribution toward the price covariance

This section aims to quantify each market participant’s marginal contribution to the price covariance between a pair of assets. We consider one pair of assets denoted by the subscript 1 and 2. We note  $\mathcal{A}_1$  the set of active market participants who trade  $S_1$  and  $\mathcal{A}_2$  for the set of active market participants who trade  $S_2$ . In this study, we only examine market participants who actively engaged in trading both assets, which are denoted by  $\mathcal{A} = \mathcal{A}_1 \cap \mathcal{A}_2$ .

We consider same types of order flows of asset 1 and asset 2, which we note by  $\mathcal{N}_1$  and  $\mathcal{N}_2$  with  $\#\{\mathcal{N}_1\} = \#\{\mathcal{N}_2\} = N$ . We then consider a multivariate Hawkes process of dimension  $2N$  for all types of orders of both assets. We keep the notation  $\Phi$  for the kernel matrix of the so-obtained Hawkes process, and  $R$  the convolution of  $\Phi$  as in Eq.(5.2.3). Then consider the decomposition of blocks by asset:

$$R = \begin{pmatrix} R^{1,1} & 0 \\ 0 & R^{2,2} \end{pmatrix} + \begin{pmatrix} 0 & R^{1,2} \\ R^{2,1} & 0 \end{pmatrix} := R_S + R_C \quad (5.2.26)$$

and

$$\Phi = \begin{pmatrix} \Phi^{1,1} & 0 \\ 0 & \Phi^{2,2} \end{pmatrix} + \begin{pmatrix} 0 & \Phi^{1,2} \\ \Phi^{2,1} & 0 \end{pmatrix} := \Phi_S + \Phi_C, \quad (5.2.27)$$

where  $R^{i,j}$  and  $\Phi^{i,j}$  are matrix of dimension  $N \times N$ . The superscript  $(i, j)$  indicates different assets. In the equation above,  $S$  (abbreviation for self) stands for the influences between order flows of the same asset, while  $C$  (abbreviation for cross) stands for the influence of order flows between different assets.

The matrix  $R_C$  and  $\Phi_C$  characterize the self-exciting and the cross-exciting effect between different markets. We start by examining the following phenomenon: market participants observe order flows in one market and respond by emitting order flows in another market. Since the covariance is a direct result of correlated price change, market participants' contribution to covariance could be quantified by their marginal contribution to the matrix  $\Phi_C$ .

Empirical evidence suggests that in electronic markets, most order flows are excited by previous order flows within the same asset. The cross-exciting effect for order flows between different assets is much weaker than the self-exciting effect. We hence consider limiting the cross-exciting effect between order flows in different markets to the first level, i.e., in the cascade representation of Hawkes processes, we only consider the first generation of order flows of one asset triggered by order flows of another asset. This idea could be transformed into the following condition:

$$\epsilon := \frac{\|\Phi_C\|}{\|\Phi_S\|} \ll 1. \quad (5.2.28)$$

Under this condition, direct calculation shows that

$$R_S = (Id - \Phi_S)^{-1} + o(\epsilon) \quad (5.2.29)$$

and

$$R_C = R_S \Phi_C R_S + o(\epsilon). \quad (5.2.30)$$

Then

$$\begin{aligned} R &= (Id - \Phi)^{-1} = (Id - \Phi_S - \Phi_C)^{-1} \\ &= (Id - R_S \Phi_C)^{-1} R_S \\ &= R_S + R_S \Phi_C R_S + O(\epsilon^2). \end{aligned} \quad (5.2.31)$$

According to Eq. (5.2.30) and Eq. (5.2.29), we have explicitly:

$$\mathbf{R}^{1,2} = \mathbf{R}^{1,1}\Phi^{1,2}\mathbf{R}^{2,2}, \quad (5.2.32)$$

$$\mathbf{R}^{2,1} = \mathbf{R}^{2,2}\Phi^{2,1}\mathbf{R}^{1,1}. \quad (5.2.33)$$

We now revisit the expression of covariance in Eq. (5.2.20) while carrying on with the condition in Eq. (5.2.28). To better illustrate market participant's contribution toward the total covariance, we plug  $\mathbf{R}^{1,2}$  and  $\mathbf{R}^{2,1}$  inside Eq. (5.2.20), which leads to:

$$\begin{aligned} \text{Cov}[S_1, S_2] &= \sum_{k \in \mathcal{N}} \Lambda^k \left( \sum_{i \in \mathcal{M}_1} \delta_i \mathbf{R}^{ik} \right) \left( \sum_{j \in \mathcal{M}_2} \delta_j \mathbf{R}^{jk} \right) \\ &= \Lambda^1 (\delta_1 \mathbf{R}^{1,1}) (\delta_2 \mathbf{R}^{2,1}) + \Lambda^2 (\delta_1 \mathbf{R}^{1,2}) (\delta_2 \mathbf{R}^{2,2}) \\ &= \Lambda^1 (\delta_1 \mathbf{R}^{1,1}) (\delta_2 \mathbf{R}^{2,2} \Phi^{2,1} \mathbf{R}^{1,1}) + \Lambda^2 (\delta_1 \mathbf{R}^{1,1} \Phi^{1,2} \mathbf{R}^{2,2}) (\delta_2 \mathbf{R}^{2,2}). \end{aligned} \quad (5.2.34)$$

Or in the scalar form,

$$\begin{aligned} \text{Cov}[S_1, S_2] &= \sum_{k \in \mathcal{N}_1} \Lambda_k \Delta \mathbf{R}_D^{1k} \Delta \mathbf{R}_I^{2k} + \sum_{k \in \mathcal{N}_2} \Lambda_k \Delta \mathbf{R}_D^{2k} \Delta \mathbf{R}_I^{1k} \\ &= \sum_{k \in \mathcal{N}_1} \Lambda_k \Delta \mathbf{R}_1^k \Delta \mathbf{R}_2^m \phi_{2,1}^{m,j} \mathbf{R}_1^{j,k} + \sum_{k \in \mathcal{N}_2} \Lambda_k \Delta \mathbf{R}_2^k \Delta \mathbf{R}_1^m \phi_{1,2}^{m,j} \mathbf{R}_2^{j,k}, \end{aligned} \quad (5.2.35)$$

where  $\phi_{2,1}^{m,j}$  (resp.  $\phi_{1,2}^{m,j}$ ) is the  $m, j$ -th element of the matrix  $\Phi^{2,1}$  (resp.  $\Phi^{1,2}$ ).  $\Delta \mathbf{R}_1^k$  is defined in the following way: If we note  $u_1$  (resp.  $u_2$ ) and  $d_1$  (resp.  $d_2$ ) the type of event that move the mid-price of  $S_1$  (resp.  $S_2$ ) up and down, then we can write

$$\Delta \mathbf{R}_1^k = R_1^{u_1, k} - R_1^{d_1, k} \quad (5.2.36)$$

and similarly

$$\Delta \mathbf{R}_2^k = R_2^{u_2, k} - R_2^{d_2, k}. \quad (5.2.37)$$

From the cascading representation of the multivariate Hawkes process, Eq. (5.2.35) has a rather spontaneous interpretation. Taking the first term as an example, an event of type  $k \in \mathcal{N}_1$  will trigger price-changing events of  $S_1$ , which will eventually change  $P_1$  by  $\Delta \mathbf{R}_1^k \mathbf{R}_1^{j,k}$ . It will also excite following up order flows in the market of  $S_2$ . If the cascade of order flows is limited to the first generation, all triggered order flows by the second effect will eventually change the price of  $S_2$  by  $\Delta \mathbf{R}_2^m \phi_{2,1}^{m,j}$ .

According to this interpretation, covariance originates from the self-exciting and the cross-exciting effect between order flows between different markets, which is characterized by the  $\Phi_C$  matrix. Market participants could then be considered as ‘‘bridges’’ between two markets. As the norm of the  $\Phi$  matrix is the expected number of triggered events, it is the common result of all market participants' activities, where each event could be attributed to a specific market participant. For a specific market participant  $\alpha$ , its covariance contribution could be measured by :

$$\text{Cov}_\alpha := \Lambda^1(\delta_1 R^{1,1})(\delta_2 R^{2,2} \Phi_\alpha^{2,1} R^{1,1}) + \Lambda^2(\delta_1 R^{1,1} \Phi_\alpha^{1,2} R^{2,2})(\delta_2 R^{2,2}). \quad (5.2.38)$$

where  $\Phi_\alpha$  is the amount of  $\Phi$  that is related to its activities. Equivalently, we can write in the scalar form that

$$\text{Cov}_\alpha = \sum_{k \in \mathcal{N}_1} \Lambda_k \Delta R_1^k \Delta R_2^m \phi_{2,1}^{m(\alpha),j} R_1^{j,k} + \sum_{k \in \mathcal{N}_2} \Lambda_k \Delta R_2^k \Delta R_1^m \phi_{1,2}^{m(\alpha),j} R_2^{j,k} \quad (5.2.39)$$

where the symbol  $m(\alpha)$  in  $\phi^{m(\alpha),j}$  denotes order flows of type  $m$  emitted by market participant  $\alpha$ .

## 5.3 Empirical results

### 5.3.1 Calibration approach

**Data** This study is based on tick level data of composition stocks of the CAC40 index stocks provided by Quanthouse. This dataset records all events in the limit order book from April 2017 to March 2018. In this dataset, each LOB event is anonymously labeled with its creator.

During this period, one stock is removed from the CAC40 index. We consider only the resting 39 stocks, which have complete records during the whole year. The dataset is then pre-treated. In this study, we only consider orders that arrive at the best bid/best ask prices of the LOB. We consider four types of orders at the best limit of both the bid side and ask side, i.e.,  $\mathcal{N} = \{P^+, P^-, L^a, L^b, C^a, C^b, M^a, M^b\}$ , representing orders that move the midprice up/down, followed by limit orders, cancellations, and market orders.

**Methodology of covariance estimation** We measure the realized volatility and the realized covariance over a daily scale. As advocated by [100], we use the average of 5-min realized variance (RV5) and covariance (RCov5) as the proxy of daily volatility and covariance:

$$RV_k = \sqrt{\sum_i (P_{t_i}^k - P_{t_{i-1}}^k)^2}, \quad (5.3.40)$$

$$RCov = \sum_i (P_{t_i}^1 - P_{t_{i-1}}^1)(P_{t_i}^2 - P_{t_{i-1}}^2), \quad (5.3.41)$$

where  $k = 1, 2$  denotes the asset,  $t_i - t_{i-1} = 5$  minutes, and the starting point  $t_0$  is chosen randomly in the first 5 minutes of the trading day. We specially note that the variation of price  $P_{t_i}^k - P_{t_{i-1}}^k$  is measured in tick size, hence  $RV$  and  $RCov$  are dimensionless.

The estimator above may fluctuate with the selection of  $t_0$ . To mitigate this issue, we perform 10 times of subsampling of  $t_0$  and take the average of estimated price variation

as  $RV$  and  $RCov$ . The daily correlation is calculated from the realized covariance and realized volatility :

$$RCorr = \frac{RCov}{RV_1 RV_2}. \quad (5.3.42)$$

**Selection of assets and market participants** Another necessary pre-treat is performed on the data set, that we restrained the time range from 10 a.m. to 4 p.m to avoid the fluctuations at the opening and closing of the market. We also exclude orders emitted with several special schemes provided by the exchange, where orders are matched independently from the best quotes of the limit order book.

We rank all possible pairings of stock by the measured  $RCorr$ . As limited by the calculation resource, we analyze the 14 most correlated pairs of stocks. Since the nonparametric estimation requires a minimum number of events, we only select active market participants (also mentioned as agents in the following context). For a pair of stocks, we selected agents who sent at least 2000 orders with a minimum of 200 orders per order type per day for both stocks. For all the 39 stocks presented in the index during the whole year, a total of 14 active agents over the selected stocks is chosen.

**Model calibration** In the practice of model calibration, fitting order flows by all selected agents leads to a significant rise in the dimension of our model. This difficulty could be mitigated by the following observation: since agents can only observe the current state of the LOB, the creator of order flows remains anonymous. Therefore, the influence of other agents' order flow over the agent  $\alpha$  is independent of its creator.

With this simplification, the origin calibration problem with  $n$  agents could be broken down into  $n$  calibration problems with smaller scales, each with the agent  $\alpha$  and all other market participants excluding  $\alpha$ , whose order flows could be combined. We note the rest of the market as  $\alpha^\top$ . It is worth mentioning that since  $\alpha^\top$  regroups all other market participants other than  $\alpha$ , therefore all order flows on both markets are included in our model setup.

As an alleviation of notation, we note the intensity of the obtained Hawkes process as:

$$\lambda_t^{i,\alpha} = \mu^{i,\alpha} + \sum_{\beta \in \{\alpha, \alpha^\top\}} \sum_{j \in \mathcal{N}} \int_0^t \phi^{i,\alpha;j,\beta}(s) dN_s^{j,\beta}, \quad (5.3.43)$$

and

$$\lambda_t^{i,\alpha^\top} = \mu^{i,\alpha^\top} + \sum_{\beta \in \{\alpha, \alpha^\top\}} \sum_{j \in \mathcal{N}} \int_0^t \phi^{i,\alpha^\top;j,\beta}(s) dN_s^{j,\beta}. \quad (5.3.44)$$

This simplification leads to a multivariate Hawkes process of dimension  $8 \times 2 \times 2$ , accounting for 8 types of order flows of  $S_1$  and  $S_2$ , emitted by  $\alpha$  and  $\alpha^\top$ . The so-defined multivariate Hawkes process is calibrated with a non-parametric estimation described

in [22]. For other calibration methodologies, we refer readers to [7] and [103], where parametrical calibration with various forms of kernel is adopted.

We especially point out that due to the limitation of data size, we can not provide reliable calibration of the exogenous intensity  $\mu$  for each day. Therefore the  $\mu$  is calibrated as a constant vector over the whole time span. Under this setting, the covariance reproduced by the multivariate Hawkes process is considered as a proxy of  $\text{Cov}^{Endo}$ . The difference between the realized covariance and  $\text{Cov}^{Endo}$  is then considered as the exogenous covariance  $\text{Cov}^{Exo}$ .

The reliability of our calibration approach is justified by bootstrapping. In Appendix 5.5.1, we provide an illustrative example of the 90% confidential interval of the calibrated parameters over a selected pair of stocks.

**Statistical profile of market participants** Before presenting the calibration results and commenting on market participants' contribution to the covariance, we first provide a statistical characterization of the active market participants identified in the previous context.

Since market participants are engaged in trading for different purposes, their behaviors differ in various aspects. Such differences are revealed with statistical features measuring for example the trading speed, the average inventory held by the end of the day, and the presence in the best queues of the LOB.

The features listed in Table 5.1 are inspired by [96], [111], [84], [109] and [75]. We also refer readers to recent works such as [31] and [1] for a summary of frequently adopted statistical features.

Feature Name	Description
Total order flow	Total volume provided, including all types of orders
Total traded volume	Total traded volumes
Fraction of trade	Fraction of trade orders in all orders
Aggressive volume	Volume traded by aggressive orders
Passive volume	Volume traded by passive orders
Aggressive fraction	Number of aggressive trades against the total number of trades
Trade asymmetry	Asymmetry caused by market orders
Total asymmetry	Asymmetry caused by all types of orders
Order Lifetime	Median time between limit order insertion and cancellation/modification.
Inter Event Time	Median time between two different orders by the same agent.
Presence in L1	Fraction of time presented on the best quotes
Presence on both sides	Fraction of time presenting on both sides of the LOB

Table 5.1: List of statistical features.

Since the value of statistical features presented above may vary significantly over

different stocks, they can't be simply averaged. We perform the following normalization procedure: for each feature, we first calculate the median value of all selected agents over all considered stocks. Then each agent's relative deviation from the median value averaged over stocks is reported. We especially note that features like *Total asymmetry* and *Trade asymmetry* could be negative. For those features, the symbol only indicates the direction of order flows (i.e. bid and ask). We only account for the use of their absolute value for normalization. Also, *Aggressive fraction Presence in L1* and *Presence in both sides* are highly stable over selected stocks. For this reason, we report the average value of these features over all selected stocks.

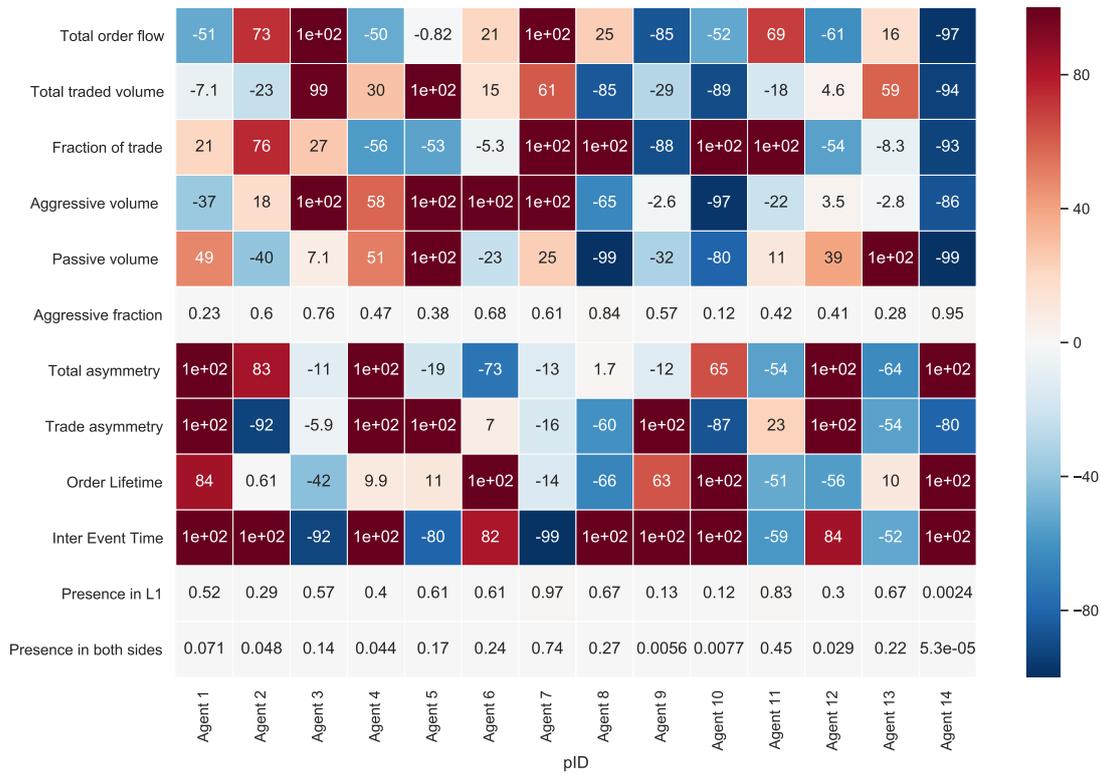


Figure 5.1: Visualisation of normalised features over all selected stocks. Deviations bigger than  $\pm 100\%$  are limited by  $\pm 100\%$ .

The normalized features are summarized in Figure 5.1 in the form of a heat map. The reported statistical features distinguished market participants in various dimensions. In this study, we particularly focus on features such as the direction of trading, the speed of trading, and the presence in the best level of the limit order book. These informations is encoded in features like *Total asymmetry*, *Trade asymmetry*, *Inter Event Time* and *Presence in both sides*. By examining these features, we identify agents 3, 6, 7, 11 and 13 as market makers. These agents feature flat closing positions and frequent presence at the best limit queues.

### 5.3.2 Interpretation of numerical results

**Covariance reproduction by Hawkes process** As a sanity check, we first compare the empirical volatility measured by RV with the volatility reproduced according to Eq. (5.2.20) over selected stocks. We point out that this could be seen as a special case of our model with  $S_1 = S_2$ . The result is presented in Table 5.2. In summary, our multivariate Hawkes process based covariance model reproduces well the empirical volatility.

Asset	RV	Vol_Hawkes	Vol_Hawkes/RV
ATOS ORIGIN(IT)	508.945500	470.056430	0.923589
AXA(Finance)	1869.151915	1822.583323	0.971565
BNP PARIBAS(Finance)	6055.273273	4139.952818	0.683622
BOUYGUES(Undefined)	2983.215706	2765.219510	0.926926
CAP GEMINI(IT)	4829.973477	3291.174249	0.681406
CREDIT AGRICOLE(Finance)	778.805576	788.846234	1.012892
DANONE(Food)	2529.891058	2055.162209	0.812352
EADS(Industry)	5472.442949	4648.714742	0.849477
LEGRAND(Electric)	1488.569545	1308.988175	0.879360
LVMH(Fashion)	1260.628091	1246.642910	0.988906
MICHELIN(Car)	573.399182	436.593885	0.761414
OREAL(Chemistry)	655.176545	664.901227	1.014843
PEUGEOT(Car)	1649.808136	1611.593794	0.976837
PPR(Fashion)	2266.783364	2040.326803	0.900098
RENAULT(Car)	5894.956296	5602.597923	0.950242
SAFRAN(Industry)	2947.968773	2644.221831	0.896964
SAINT-GOBAIN(Industry)	5425.193030	4315.534925	0.795462
SCHNEIDER ELECTRIC(Electric)	3018.097545	2192.108071	0.726321
SOCIETE GENERALE(Finance)	9318.011953	8653.800389	0.928717
TECHNIPFMC(Energy)	4223.617701	3269.536862	0.774108
TOTAL(Energy)	6104.395803	4686.902088	0.767791
VALEO(Car)	2752.624727	2409.368392	0.875299
VINCI(Undefined)	3099.234108	2273.032807	0.733418

Table 5.2: Reproduction of realized volatility by Hawkes process. All price variation is measured in tick size.

Our result suggests that compared to the endogenous volatility, the exogenous volatility is much smaller. Our findings agree well with the empirical fact that most order flows are endogenous as a response to previous order flows, rather than a genuine intention to trade, which leads to a small exogenous fraction in the intensity, which is

argued in [133] and [39].

Next, we present the empirical covariance and the covariance reproduced by the multivariate Hawkes process over the same data set in Table.5.3. We report the realized covariance (RCov) measured in tick size and the endogenous covariance defined in Eq.(5.2.23).

Asset 1	Asset 2	RCov	Cov_endo	Cov_Endo/RCov
CREDIT AGRICOLE(Finance)	SOCIETE GENERALE(Finance)	2050.463595	1866.574294	0.910318
CREDIT AGRICOLE(Finance)	BNP PARIBAS(Finance)	1298.020364	1315.751568	1.013660
ATOS ORIGIN(IT)	CAP GEMINI(IT)	870.366727	549.732802	0.631611
SAFRAN(Industry)	EADS(Industry)	2444.066117	1758.283834	0.719409
TOTAL(Energy)	TECHNIPFMC(Energy)	1913.176095	1647.644057	0.861209
BOUYGUES(Undefined)	VINCI(Undefined)	1890.806413	1317.765804	0.696933
AXA(Finance)	CREDIT AGRICOLE(Finance)	717.189364	700.992632	0.977416
AXA(Finance)	SOCIETE GENERALE(Finance)	2605.304516	2129.782377	0.817479
AXA(Finance)	BNP PARIBAS(Finance)	1634.497864	1605.125970	0.982030
LVMH(Fashion)	PPR(Fashion)	1069.291273	862.890058	0.806974
MICHELIN(Car)	PEUGEOT(Car)	391.864545	374.165451	0.954834
MICHELIN(Car)	RENAULT(Car)	800.511636	742.314410	0.927300
MICHELIN(Car)	VALEO(Car)	562.806364	552.866560	0.982339
PEUGEOT(Car)	SOCIETE GENERALE(Finance)	1587.317564	1183.819806	0.745799
PEUGEOT(Car)	BNP PARIBAS(Finance)	1105.056409	841.893755	0.761856
PEUGEOT(Car)	RENAULT(Car)	2070.644591	1903.019389	0.919047
PEUGEOT(Car)	VALEO(Car)	1062.397212	1075.715105	1.012536
SCHNEIDER ELECTRIC(Electric)	SAINT-GOBAIN(Industry)	2345.217909	1670.294244	0.712213
SCHNEIDER ELECTRIC(Electric)	LEGRAND(Electric)	1245.159182	909.744049	0.730625
SAINT-GOBAIN(Industry)	VINCI(Undefined)	2065.217472	1480.337879	0.716795
SOCIETE GENERALE(Finance)	BNP PARIBAS(Finance)	4927.157938	4405.775994	0.894182
SOCIETE GENERALE(Finance)	RENAULT(Car)	3178.333303	2229.184608	0.701369
RENAULT(Car)	VALEO(Car)	1987.905818	1957.767949	0.984839

Table 5.3: Reproduction of covariance by Eq. (5.2.20) and endogenous covariance by Eq. (5.2.23). Price variation is measured in tick size.

From Table 5.3, we remark that similar to the reproduction of volatility, the majority of covariance could be explained by the endogenous covariance. However, we also emphasize an outstanding observation that for a pair of stocks, the fraction of endogenous covariance decreases significantly when the business of the two companies overlaps.

Here we give one possible explanation that companies sharing similar businesses are more likely to be influenced by the same flow of information. For example, in the sector of finance or energy, stocks of different companies are mutually influenced by trending news. This effect could be interpreted by a more correlated daily variation of the exogenous intensity term  $\mu^d$  in Eq. (5.2.23).

**First order approximation of covariance** We examine the validity of Eq. (5.2.35). We compare the covariance reproduced by Eq. (5.2.20) (noted as the endogenous covariance with the covariance reproduced by Eq. (5.2.35) (noted as the 1st order approximated covariance). Their relative difference is reported in Table.5.4.

	Asset1	Asset2	Relative difference, in %
0	RENAULT(Car)	VALEO(Car)	-10.727614
1	AXA(Finance)	SOCIETE GENERALE(Finance)	-9.952873
2	OREAL(Chemistry)	DANONE(Food)	-9.194710
3	SAFRAN(Industry)	EADS(Industry)	-9.773529
4	MICHELIN(Car)	VALEO(Car)	-10.942049
5	PEUGEOT(Car)	BNP PARIBAS(Finance)	-3.372851
6	BOUYGUES(Undefined)	VINCI(Undefined)	-12.962753
7	SOCIETE GENERALE(Finance)	BNP PARIBAS(Finance)	-31.120657
8	SOCIETE GENERALE(Finance)	RENAULT(Car)	-3.079902
9	CREDIT AGRICOLE(Finance)	BNP PARIBAS(Finance)	-25.913793
10	PEUGEOT(Car)	VALEO(Car)	-12.181595
11	LVMH(Fashion)	PPR(Fashion)	-10.288422
12	ATOS ORIGIN(IT)	CAP GEMINI(IT)	-9.569091
13	CREDIT AGRICOLE(Finance)	SOCIETE GENERALE(Finance)	-19.452585
14	SAINT-GOBAIN(Industry)	VINCI(Undefined)	-7.344969
15	SCHNEIDER ELECTRIC(Electric)	SAINT-GOBAIN(Industry)	-12.098872
16	AXA(Finance)	CREDIT AGRICOLE(Finance)	-8.202498
17	TOTAL(Energy)	TECHNIPFMC(Energy)	-6.077109
18	AXA(Finance)	BNP PARIBAS(Finance)	-11.928216
19	PEUGEOT(Car)	RENAULT(Car)	-15.668649
20	PEUGEOT(Car)	SOCIETE GENERALE(Finance)	-2.730582
21	MICHELIN(Car)	PEUGEOT(Car)	-7.365318
22	SCHNEIDER ELECTRIC(Electric)	LEGRAND(Electric)	-11.816030
23	MICHELIN(Car)	RENAULT(Car)	-10.030083

Table 5.4: Relative difference between endogenous covariance and the first order approximated covariance. Price variation is measured in tick size.

Overall, the covariance approximated by Eq. (5.2.35) reproduces well the endogenous covariance with a median difference of around 10%. Our result suggests that in the cascading representation of the multivariate Hawkes process, the majority of self-exciting and cross-exciting effects between order flows is limited to the first level. We also observe in Table.5.4 that the relative difference is notably higher when companies share similar businesses. Especially, a noticeable difference exists for the pair Société Générale - BNP Paribas, followed by Crédit Agricole - BNP Paribas and then Crédit Agricole - Société Générale. A similar phenomenon is also observed for companies in other sectors. We mentioned Renault and Peugeot as an example.

Such evidence suggests that highly correlated assets may be traded as proxies. Market participants may adopt multi-asset strategies and trade stocks as proxies for different

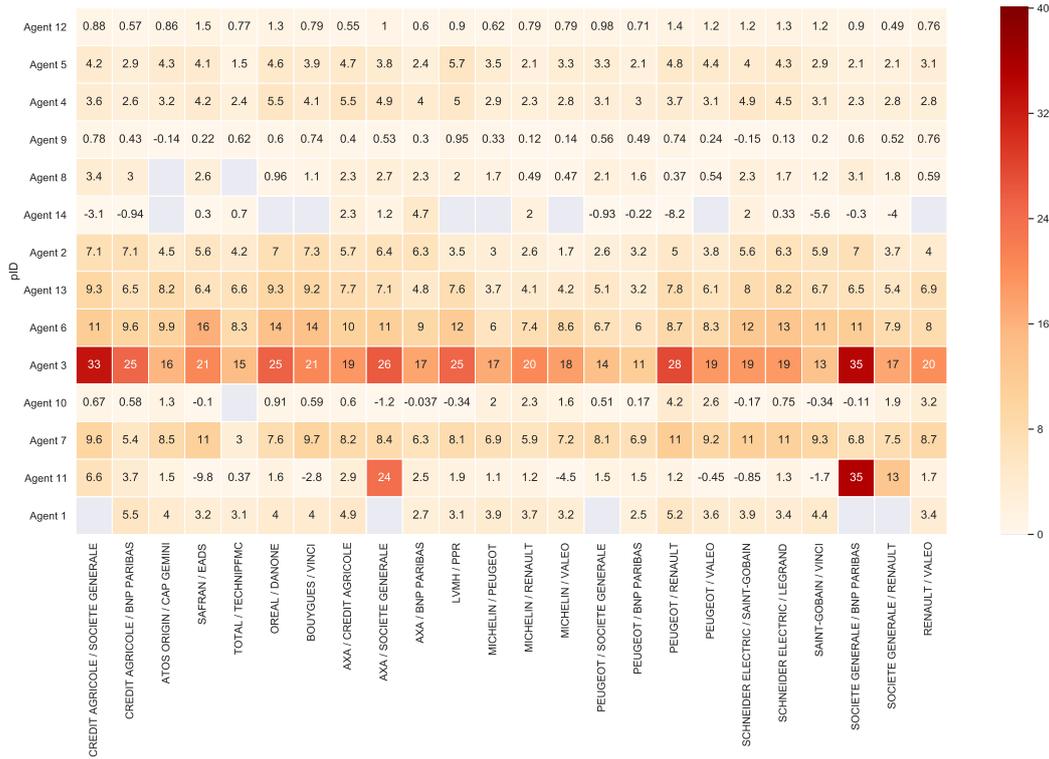


Figure 5.2: Heatmap of the agent’s correlation contribution, in percentage.

purposes, such as hedging or seeking immediate liquidity. In this case, our model still reproduces the majority of covariance. However, in the cascading representation of the multivariate Hawkes process, the influence between order flows should not be only limited to the first level.

**Measuring market participants’ covariance contribution** As the last piece of obtained empirical results, we report market participants’ contribution to the covariance measured as  $\text{Cov}_\alpha$  in Eq. (5.2.38). To facilitate the comparison between different pairs of stocks, we first normalize the obtained covariance by the realized volatility of corresponding stocks.

We note the realized volatility of stock 1 and stock 2 as  $\sigma_1$  and  $\sigma_2$ . We note by  $\text{Cov}$  as the realized covariance and  $\text{Corr}$  as the correlation coefficient. Agent  $\alpha$ ’s contribution to the correlation is defined as

$$\text{Corr}_\alpha := \frac{\text{Cov}_\alpha}{\sigma_1 \sigma_2} = \frac{\text{Cov}_\alpha}{\text{Cov}} \frac{\text{Cov}}{\sigma_1 \sigma_2} = \frac{\text{Cov}_\alpha}{\text{Cov}} \text{Corr}. \quad (5.3.45)$$

Under this definition we do not necessarily have  $\sum_\alpha \text{Corr}_\alpha = \text{Corr}$ . In Figure 5.2, we present agents’ contribution as  $\text{Corr}_\alpha$  in the form of a heat map.

Our first remark is that agents’ contribution to the correlation is rather uniform over all pairs of assets. Especially several agents showed a strong correlation contribution, indicating that their order flows led to correlated price changes of different assets.

We also find that, unlike in [127], no simple relation could be established between agents' statistical features as reported in previous sections and agents' contribution to correlation. Neither market participants' contribution to correlation nor the normalized contribution by volume correlates with the statistical features. Indeed, the mechanism of the generation of covariance is more complicated than the mechanism of the generation of volatility. Features measured for single stock fail to capture joint dynamics of order flows between two assets. Especially, the statistical features we introduced fail to capture how order flows are propagated to different markets via agents' activities.

### Correlation contribution by order type

We present each agent's contribution to correlation by event type in Figure 5.3 and their contribution to correlation per order in Figure 5.4. The following is obtained by normalizing the covariance contribution in Figure 5.3 with the total number of orders of each type.

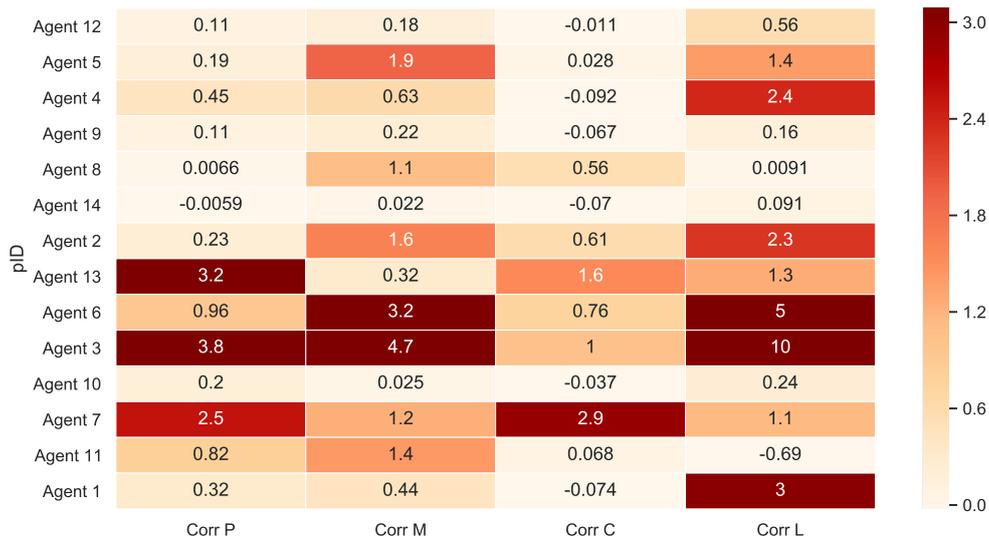


Figure 5.3: Contribution to correlation by event type. Averaged over all selected pairs, in percentage.

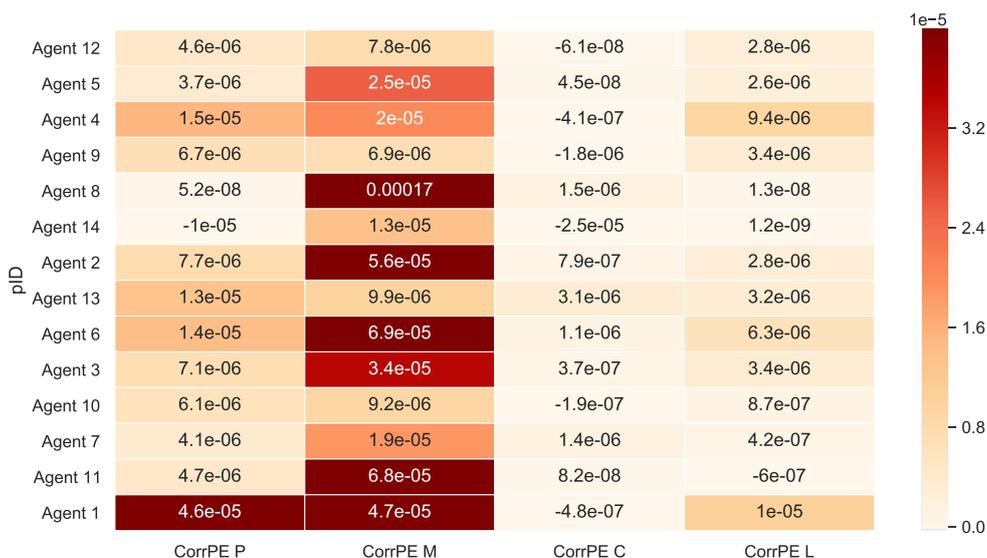


Figure 5.4: Contribution to correlation per order by event type. Averaged over all selected pairs, in percentage.

We start by examining Figure 5.4. First, it is not surprising that price-changing events and market orders exhibit a larger average correlation contribution than cancellations and limit orders. Secondly, agents 3, 6, 7, 11 and 13 are identified as market makers in the previous section. They are also agents with strong average correlation contributions per price-changing order and per market order. We also notice that market makers have a positive correlation contribution per cancellation order. One possible explanation is that cancellations originated by market makers quickly emptied the liquidity, which may lead to further price moves and contributes to correlations. This could be used as another indicator to identify market makers.

We continue by grouping agents' contribution to correlation by order type as in Figure 5.3, where agents' cumulative contribution to correlation exhibits a noticeable difference pattern from that of Figure 5.4. For price-changing events and market orders, the difference between agents should be mainly explained by the difference in the total volume, since their covariance contribution per event is rather uniform. This is the case for agents 3, 7 and 13, where large covariance contribution could be explained by the large volume of price-changing orders they emitted. Another example could be agent 3 and agent 6. Despite that agent 6 has a higher correlation contribution per event for all types of events, the large volume emitted by agent 3 leads to a larger correlation contribution. Such observation suggests that the total volume plays an important role in market participants' correlation contributions. For cancellations, we notice that market makers emit a large number of cancellations, which leads to a significant cumulative correlation contribution. As for limit orders, the situation is similar: the difference in cumulative correlation contribution is mostly explained by the volume of limit orders emitted.

## 5.4 Summary and prospects

In this chapter, we introduced a general framework based on multivariate Hawkes processes to capture price covariance in correlated markets. In this framework, the correlated price change is a natural result of correlated order flows.

We then distinguish the covariance as exogenous and endogenous. The previous type is due to the fluctuation of exogenous intensity over time, while the second type is due to the self-exciting and cross-exciting effect between order flows. Due to the limitation of the data scale, only the endogenous covariance could be calibrated from empirical data. Under some reasonable assumption, we quantify for the first time each market participant's contribution to the covariance measured by their marginal contribution in the kernel norm matrix.

Empirical results suggest that our framework can reproduce the majority of price covariance. Besides, the assumption and approximations we made are validated with empirical results. We observe that market participants' contribution toward price correlation is uniform over different pairs of assets. Moreover, a detailed investigation into each agent's contribution suggests that some active market participants may adopt multi-asset strategies. Finally, our result suggests that market participants' covariance contribution is closely linked to the total volume they emit.

From the result obtained in this chapter, possible future improvements include characterizing market participants with statistical features and establishing quantitative relations between their covariance contribution and statistical profiles, which remains open questions.

## 5.5 Appendix

### 5.5.1 Example of calibration error by Bootstrapping

Agent	std of estimated $\text{Cov}_\alpha$ , %	std of estimated $\text{Corr}_\alpha$ , %
Agent 9	0.162358	2.300237
Agent 10	0.167564	0.645140
Agent 13	0.345652	0.426668
Agent 2	0.519005	0.497563
Agent 8	0.416504	0.500643
Agent 12	0.129820	0.488825
Agent 4	0.117903	0.454265
Agent 3	0.790632	0.173280
Agent 5	0.237159	0.309868
Agent 11	8.646219	8.361573
Agent 6	0.642228	0.698290
Agent 7	0.247414	0.063884

Table 5.5: Standard error of correlation contribution (measured as  $\text{Corr}_\alpha$ ) by active agents on the pair Société Générale - BNP Paribas. Calculated from 20 times of bootstrapping each with 80% of data randomly chosen from the original data set.



# Bibliographie

- [1] High-frequency trading activity in eu equity markets. *ESMA Economic Report*, 1, 2014.
- [2] F. Abergel, M. Anane, A. Chakraborti, A. Jedidi, and I. Muni Toke. *Limit order books*. Cambridge University Press, 2016.
- [3] F. Abergel and A. Jedidi. Long time behaviour of a hawkes process-based limit order book. *SIAM Journal on Financial Mathematics*, 6:1026–1043, 2015.
- [4] Frederic Abergel and Aymen Jedidi. A Mathematical Approach to Order Book Modeling. March 2013.
- [5] 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, November 2015. Publisher: Society for Industrial and Applied Mathematics.
- [6] Eduardo Abi Jaber and Omar El Euch. Multifactor approximation of rough volatility models. *SIAM Journal on Financial Mathematics*, 10(2):309–349, 2019.
- [7] M. Achab, E. Bacry, J. F. Muzy, and M. Rambaldi. Analysis of order book flows using a non-parametric estimation of the branching ratio matrix. *Quantitative Finance*, 18(2):199212, 2017.
- [8] Massil Achab, Emmanuel Bacry, Stéphane Gaïffas, Iacopo Mastromatteo, and Jean-François Muzy. Uncovering causality from multivariate hawkes integrated cumulants. *J. Mach. Learn. Res.*, 18(1):69987025, January 2017.
- [9] Dong-Hyun Ahn, Jacob Boudoukh, Matthew Richardson, and Robert F. Whitelaw. Optimal Risk Management Using Options. *The Journal of Finance*, 54(1):359–375, feb 1999.
- [10] Krishna B. Athreya and Peter Ney. *Branching processes*. Springer-Verlag Berlin, New York, 1972.
- [11] E. Bacry, S. Delattre, M. Hoffmann, and J. F. Muzy. Modelling microstructure noise with mutually exciting point processes. *Quantitative Finance*, 13(1):65–77, January 2013.

- [12] E. Bacry, J. Delour, and J. F. Muzy. Multifractal random walk. *Phys. Rev. E*, 64:026103, Jul 2001.
- [13] E. Bacry, T. Jaisson, and J.-F. Muzy. Estimation of slowly decreasing hawkes kernels: Application to high-frequency order book dynamics. *Quantitative Finance*, 16(8):1179–1201, 2016.
- [14] E. Bacry, A. Kozhemyak, and J. F. Muzy. Log-normal continuous cascade model of asset returns: aggregation properties and estimation. *Quantitative Finance*, 13(5):795818, 2013.
- [15] E. Bacry, A. Kozhemyak, and Jean-François Muzy. Continuous cascade models for asset returns. *Journal of Economic Dynamics and Control*, 32(1):156–199, 2008. Applications of statistical physics in economics and finance.
- [16] E. Bacry, I. Mastromatteo, and J.-F. Muzy. Hawkes processes in finance. *Market Microstructure and Liquidity*, 01(01):1550005, 2015.
- [17] E. Bacry and J.-F. Muzy. Hawkes model for price and trades high-frequency dynamics. *Quantitative Finance*, 14(7):1147–1166, 2014.
- [18] E. Bacry and J.-F. Muzy. First-and second-order statistics characterization of Hawkes processes and non-parametric estimation. *IEEE Transactions on Information Theory*, 62(4):2184–2202, 2016.
- [19] E. Bacry, J. F. Muzy, and A. Arnodo. Singularity spectrum of fractal signals from wavelet analysis: Exact results. *Journal of Statistical Physics*, 70(3-4):635–674, February 1993.
- [20] E. Bacry and J.f. Muzy. Log-infinitely divisible multifractal processes. *Communications in Mathematical Physics*, 236(3):449475, 2003.
- [21] Emmanuel Bacry, Martin Bompaire, Stéphane Gaïffas, and Soren Poulsen. tick: a python library for statistical learning, with a particular emphasis on time-dependent modeling, Jul 2017.
- [22] Jaisson T. & Muzy J. Bacry, E. Estimation of slowly decreasing hawkes kernels: application to high-frequency order book dynamics. *Quantitative Finance*, 16:1179–1201, 2016.
- [23] Bacry, E., Dayri, K., and Muzy, J.F. Non-parametric kernel estimation for symmetric hawkes processes. application to high frequency financial data. *Eur. Phys. J. B*, 85(5):157, 2012.
- [24] O.E. Barndorff-Nielsen and N. Shephard. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *J.R. Statist. Soc. B*, 64, 2002.

- [25] Ole E. Barndorff-Nielsen, Silvia Kinnebroek, and Neil Shephard. *Measuring downside risk: realised semivariance*, pages 117–136. Oxford University Press, (edited by t. bollerslev, j. russell and m. watson) edition, 2010.
- [26] Julien Barral and Benoit B. Mandelbrot. Multifractal products of cylindrical pulses. *Probability Theory and Related Fields*, 124:409–430, 2001.
- [27] Luc Bauwens and Nikolaus Hautsch. *Modelling Financial High Frequency Data Using Point Processes*. Springer Berlin Heidelberg, 2009.
- [28] C. Bayer, F. Harang, and P. Pigato. Log-modulated rough stochastic volatility models. arXiv:2008.03204, 2020.
- [29] Christian Bayer, Peter Friz, and Jim Gatheral. Pricing under rough volatility. *Quantitative Finance*, 16(6):887–904, 2016.
- [30] Christian Bayer, Peter K. Friz, Paul Gassiat, Jorg Martin, and Benjamin Stemper. A regularity structure for rough volatility. *Mathematical Finance*, 30(3):782–832, 2020.
- [31] Mario Bellia. High-frequency market making: Liquidity provision, adverse selection, and competition. *SSRN Electronic Journal*, 2017.
- [32] Mikkel Bennedsen. Semiparametric estimation and inference on the fractal index of gaussian and conditionally gaussian time series data. *Econometric Reviews*, 39(9):875903, Feb 2020.
- [33] Mikkel Bennedsen, Asger Lunde, and Mikko S. Pakkanen. Decoupling the short- and long-term behavior of stochastic volatility. arXiv:1610.00332, 2021.
- [34] M Benzaquen, I Mastromatteo, Z Eisler, and J-P Bouchaud. Dissecting cross-impact on stock markets: an empirical analysis. *Journal of Statistical Mechanics: Theory and Experiment*, 2017(2):023406, 2017.
- [35] Fischer Black and Myron Scholes. The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, 81(3):637–654, May 1973.
- [36] Anine E. Bolko, Kim Christensen, Mikko S. Pakkanen, and Bezirgen Veliyev. Roughness in spot variance? a gmm approach for estimation of fractional log-normal stochastic volatility models using realized measures. 2020.
- [37] Tim Bollerslev, Ian Domowitz, and Jianxin Wang. Order flow and the bid-ask spread: An empirical probability model of screen-based trading. *Journal of Economic Dynamics and Control*, 21(8-9):1471–1491, June 1997.
- [38] Giacomo Borretti, Lucio Maria Calcagnile, Michele Treccani, Fulvio Corsi, Stefano Marmi, and Fabrizio Lillo. Modelling systemic price cojumps with hawkes factor models. *Quantitative Finance*, 15(7):11371156, 2015.

- [39] Jean-Philippe Bouchaud and Julius Bonart. *Trades, quotes and prices: financial markets under the microscope*. Cambridge University Press, 2018.
- [40] R. C. Bradley. Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys*, 2:107–144, 2005.
- [41] Laurent Calvet and Adlai Fisher. *Multifractal Volatility Theory, Forecasting and Pricing*. Elsevier Science & Technology Books, San Diego, CA, USA, 2000.
- [42] Laurent Calvet and Adlai Fisher. Forecasting multifractal volatility. *Journal of Econometrics*, 105(1):27–58, November 2001.
- [43] Laurent E. Calvet and Adlai J. Fisher. Regime-Switching and the Estimation of Multifractal Processes. *SSRN Electronic Journal*, 2003.
- [44] A. Chakraborti, I. Muni Toke, M. Patriarca, and F. Abergel. Econophysics review: I. Empirical facts. *Quantitative Finance*, 11(7):991–1012, 2011.
- [45] R. Cont, S. Stoikov, and R. Talreja. A stochastic model for order book dynamics. *Operations Research*, 58:549–563, 2010.
- [46] D. J. Daley. *An introduction to the theory of point processes: Elementary theory and methods*, volume 1. Springer, 2008.
- [47] D.J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure*. Probability and Its Applications. Springer New York, 2007.
- [48] A. Daw and J. Pender. The Queue-Hawkes Process: Ephemeral Self-Excitement. *arXiv e-prints*, page arXiv:1811.04282, 2018.
- [49] Khalil Dayri and Mathieu Rosenbaum. Large tick assets: Implicit spread and optimal tick size. *Market Microstructure and Liquidity*, 01(01):1550003, 2015.
- [50] Ian Domowitz and Jianxin Wang. Auctions as algorithms: Computerized trade execution and price discovery. *Journal of Economic Dynamics and Control*, 18(1):29–60, 1994. Special Issue on Computer Science and Economics.
- [51] Omar El Euch and Mathieu Rosenbaum. Perfect hedging in rough heston models. *Ann. Appl. Probab.*, 28(6):3813–3856, Dec 2018.
- [52] Omar El Euch and Mathieu Rosenbaum. The characteristic function of rough heston models. *Mathematical Finance*, 29(1):3–38, 2019.
- [53] Omar El Euch, Jim Gatheral, Rado Radoii, and Mathieu Rosenbaum. The zumbach effect under rough heston. *Quantitative Finance*, 20(2):235–241, 2020.
- [54] Carl J.G. Evertsz. *Fractal Geometry Of Financial Time Series*, volume 03. 1995.

- [55] Eugene F. Fama. The Behavior of Stock-Market Prices. *The Journal of Business*, 38(1):34, January 1965.
- [56] J. D. Farmer, P. Patelli, and I. I. Zovko. The predictive power of zero intelligence in financial markets. *Proceedings of the National Academy of Sciences*, 102(6):2254–2259, 2005.
- [57] José Da Fonseca and Riadh Zaatour. Correlation and lead-lag relationships in a Hawkes microstructure model. *Journal of Futures Markets*, 37(3):260285, 2016.
- [58] M. Forde, M. Fukasawa, S. Gerhold, and B. Smith. The rough Bergomi model as  $h \rightarrow 0$  skew flattening/blow up and non-gaussian rough volatility, 2020.
- [59] Martin Forde and Benjamin Smith. The Riemann-Liouville field as a  $h \rightarrow 0$  limit-sub, critical and super critical GMC, decompositions and explicit spectral expansions, 2020.
- [60] Peter K. Friz, Paul Gassiat, and Paolo Pigato. Short dated smile under rough volatility: asymptotics and numerics, 2020.
- [61] Masaaki Fukasawa. Volatility has to be rough. *Quantitative Finance*, 21(1):1–8, 2021.
- [62] Masaaki Fukasawa, Tetsuya Takabatake, and Rebecca Westphal. Is volatility rough? 2019.
- [63] Y. V. Fyodorov, B. A. Khoruzhenko, and N. J. Simm. Fractional Brownian motion with Hurst index  $H = 0$  and the Gaussian Unitary Ensemble. *The Annals of Probability*, 44(4):2980 – 3031, 2016.
- [64] A. Garèche, G. Disdier, Julien Kockelkoren, and Jean-Philippe Bouchaud. Fokker-Planck description for the queue dynamics of large tick stocks. *Physical Review E, Statistical, nonlinear, and soft matter physics*, 88:032809, 09 2013.
- [65] Jim Gatheral, Thibault Jaisson, and Mathieu Rosenbaum. Volatility is rough. *Quantitative Finance*, 18(6):933–949, 2018.
- [66] S. Ghoshghaie, W. Breymann, J. Peinke, and P. Talkner. Turbulence and financial markets. In S. Gavrilakis, L. Machiels, and P. A. Monkewitz, editors, *Advances in Turbulence VI*, pages 167–170, Dordrecht, 1996. Springer Netherlands.
- [67] S. Ghoshghaie, W. Breymann, J. Peinke, P. Talkner, and Y. Dodge. Turbulent cascades in foreign exchange markets. *Nature*, 381:767–770, 1996.
- [68] K. Giesecke, H. Kakavand, and M. Mousavi. Exact simulation of point processes with stochastic intensities. *Operations Research*, 59(5):1233–1245, 2011.

- [69] Paul Glasserman and Pu He. Buy rough, sell smooth. *Quantitative Finance*, 20(3):363–378, 2020.
- [70] M. D. Gould and J. Bonart. Queue imbalance as a one-tick-ahead price predictor in a limit order book. *Market Microstructure and Liquidity*, 2(02):1650006, 2016.
- [71] M. D. Gould, M. A. Porter, S. Williams, M. McDonald, D. J. Fenn, and S. D. Howison. Limit order books. *Quantitative Finance*, 13(11):1709–1742, 2013.
- [72] C. W. J. Granger. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, 37(3):424, August 1969.
- [73] Olivier Gueant. *The financial mathematics of market liquidity: from optimal execution to market making*. Chapman & Hall/CRC financial mathematics series. CRC Press, Taylor & Francis Group, Boca Raton, 2016.
- [74] Paul Hager and Eyal Neuman. The multiplicative chaos of  $h = 0$  fractional brownian fields. *Papers*, 2020.
- [75] Björn Hagströmer and Lars L. Norden. The diversity of high frequency traders. *SSRN Electronic Journal*, 2012.
- [76] Lars Peter Hansen. Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, 50(4):1029, July 1982.
- [77] N. R. Hansen, P. Reynaud-Bouret, and V. Rivoirard. Lasso and probabilistic inequalities for multivariate point-processes. *Bernoulli*, 21(1):83–143, 2015.
- [78] Deborah J. Harper and Larry V. St.Louis. Regional Economic Diversification and Efficiency: Baumols Likely Lower Confidence Limit Measure of Risk. *Review of Regional Studies*, May 1999.
- [79] A. G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58:83–90, 1971.
- [80] Alan G. Hawkes. Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, 33(3):438443, 1971.
- [81] Alan G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971.
- [82] Alan G. Hawkes. Hawkes processes and their applications to finance: a review. *Quantitative Finance*, 18(2):193–198, 2018.
- [83] Lunde A. Shephard N. Heber, G. and K. Sheppard. Oxford-man institute of quantitative finance. realized library.

- [84] Terrence Hendershott and Ryan Riordan. High frequency trading and price discovery. *SSRN Electronic Journal*, 2012.
- [85] Patrick Hewlett. Clustering of order arrivals, price impact and trade path optimisation. 2006.
- [86] Blanka Horvath, Antoine Jacquier, and Peter Tankov. Volatility options in rough volatility models. *SIAM Journal on Financial Mathematics*, 11(2):437–469, 2020.
- [87] W. Huang, C.-A. Lehalle, and M. Rosenbaum. Simulating and analyzing order book data: The queue-reactive model. *Journal of the American Statistical Association*, 110(509):107–122, 2015.
- [88] John Hull and Alan White. The Pricing of Options on Assets with Stochastic Volatilities. *The Journal of Finance*, 42(2):281–300, June 1987.
- [89] Jean Jacod. Multivariate point processes: predictable projection, Radon-Nikodym derivatives, representation of martingales. 31(3):235–253, 1975.
- [90] S. Jaffard. Multifractal Formalism for Functions Part I: Results Valid For All Functions. *SIAM Journal on Mathematical Analysis*, 28(4):944–970, July 1997.
- [91] S. Jaffard. Multifractal Formalism for Functions Part II: Self-Similar Functions. *SIAM Journal on Mathematical Analysis*, 28(4):971–998, July 1997.
- [92] A. Jedidi and F. Abergel. On the stability and price scaling limit of a hawkes process-based order book model. *SSRN Electronic Journal*, pages 1–22, 05 2013.
- [93] Don H. Johnson. Point process models of single-neuron discharges. *Journal of Computational Neuroscience*, 3:275–299, 2004.
- [94] Paul Jusselin and Mathieu Rosenbaum. No-arbitrage implies power-law market impact and rough volatility. *Mathematical Finance*, 30(4):1309–1336, 2020.
- [95] Martin Keller-Ressel, Martin Larsson, and Sergio Pulido. Affine rough models, 2018.
- [96] 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):967998, 2017.
- [97] Alexey Kozhemyak. Modélisation de séries financières à l’aide de processus invariants d’échelle. application à la prédiction du risque. 2006.
- [98] Jeremy Large. Measuring the resiliency of an electronic limit order book. *Journal of Financial Markets*, 10(1):1–25, 2007.

- [99] C.-A. Lehalle, O. Mounjid, and M. Rosenbaum. Optimal liquidity-based trading tactics. pages 1–39, 2018.
- [100] Lily Y. Liu, Andrew J. Patton, and Kevin Sheppard. Does anything beat 5-minute rv? a comparison of realized measures across multiple asset classes. *Journal of Econometrics*, 187(1):293 – 311, 2015.
- [101] Giulia Livieri, Saad Mouti, Andrea Pallavicini, and Mathieu Rosenbaum. Rough volatility: Evidence from option prices. *IISE Transactions*, 50(9):767–776, 2018.
- [102] X. Lu and F. Abergel. Order-book modelling and market making strategies. *Market Microstructure and Liquidity*, 04:1950003, 2018.
- [103] Xiaofei Lu and Frédéric Abergel. High-dimensional hawkes processes for limit order books: modelling, empirical analysis and numerical calibration. *Quantitative Finance*, 18(2):249264, 2018.
- [104] B. B. Mandelbrot, A. J. Fisher, and L. E. Calvet. A multifractal model of asset returns. *Cowles Foundation Discussion Paper No. 1164, Sauder School of Business Working Paper, Available at SSRN*, sep 1997.
- [105] Benoit Mandelbrot. The Variation of Certain Speculative Prices. *The Journal of Business*, 36(4):394, January 1963.
- [106] Benoit B. Mandelbrot. *The fractal geometry of nature*. W.H. Freeman, San Francisco, 1982.
- [107] Benoit B. Mandelbrot. *Fractals and Scaling in Finance*. Springer New York, New York, NY, 1997.
- [108] Garman M.B and Klass M.J. On the estimation of security price volatility from historical data. *The Journal of Business*, 53:67–78, 1980.
- [109] 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*, 03(03n04):1850005, 2017.
- [110] Hongyuan Mei and Jason Eisner. The neural hawkes process: A neurally self-modulating multivariate point process. In *NIPS*, 2017.
- [111] Albert J. Menkveld. High frequency trading and the new market makers. *Journal of Financial Markets*, 16(4):712740, 2013.
- [112] S. P. Meyn and R. L. Tweedie. Stability of markovian processes III: Foster-lyapunov criteria for continuous-time processes. *Advances in Applied Probability*, 25:518548, 1993.

- [113] Yuliya Mishura and Mounir Zili. Gaussian processes. In Yuliya Mishura and Mounir Zili, editors, *Stochastic Analysis of Mixed Fractional Gaussian Processes*, pages 1–29. Elsevier, 2018.
- [114] M. Morariu-Patrichi and M. S. Pakkanen. State-dependent hawkes processes and their application to limit order book modelling. *arXiv e-prints*, page arXiv:1809.08060, September 2018.
- [115] Jean-François Muzy and Emmanuel Bacry. Multifractal stationary random measures and multifractal random walks with log infinitely divisible scaling laws. *Physical Review E*, 66(5), Nov 2002.
- [116] Jean-François Muzy, Rachel Baïle, and Emmanuel Bacry. Random cascade model in the limit of infinite integral scale as the exponential of a nonstationary 1/f noise: Application to volatility fluctuations in stock markets. *Physical Review E*, 87(4), 2013.
- [117] J.F. Muzy, E. Bacry, and A. Arneodo. The Multifractal Formalism Revisited with Wavelets. *International Journal of Bifurcation and Chaos*, 04(02):245–302, April 1994.
- [118] Muzy, J. F., Delour, J., and Bacry, E. Modelling fluctuations of financial time series: from cascade process to stochastic volatility model. *Eur. Phys. J. B*, 17(3):537–548, 2000.
- [119] Jesper Møller and Jakob G. Rasmussen. Perfect simulation of hawkes processes. *Advances in Applied Probability*, 37(3):629646, 2005.
- [120] Jesper Møller and Jakob G. Rasmussen. Approximate Simulation of Hawkes Processes. *Methodology and Computing in Applied Probability*, 8(1):53–64, March 2006.
- [121] Eyal Neuman and Mathieu Rosenbaum. Fractional brownian motion with zero hurst parameter: a rough volatility viewpoint. *Electronic Communications in Probability*, 23, 2018.
- [122] Y. Ogata. Seismicity analysis through point-process modeling: A review. *pure and applied geophysics*, 155:471–507, 1999.
- [123] Yoshihiko Ogata and Hirotugu Akaike. On linear intensity models for mixed doubly stochastic poisson and self- exciting point processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 44(1):102–107, 1982.
- [124] Tohru Ozaki. Maximum likelihood estimation of hawkes’ self-exciting point processes. *Annals of the Institute of Statistical Mathematics*, 31:145–155, 1979.

- [125] C.-A. Parlour and D. J. Seppi. Limit order markets: A survey. *Handbook of financial intermediation and banking*, 5:63–95, 2008.
- [126] M. Rambaldi, E. Bacry, and F. Lillo. The role of volume in order book dynamics: a multivariate Hawkes process analysis. *Quantitative Finance*, 17(7):999–1020, 2017.
- [127] Marcello Rambaldi, Emmanuel Bacry, and Jean-François Muzy. Disentangling and quantifying market participant volatility contributions. *Quantitative Finance*, 19(10):1613–1625, 2019.
- [128] Marcello Rambaldi, Paris Pennesi, and Fabrizio Lillo. Modeling fx market activity around macroeconomic news: a hawkes process approach. *Research Papers in Economics*, 2014.
- [129] Rémi Rhodes and Vincent Vargas. Gaussian multiplicative chaos and applications: A review. *Probability Surveys*, 11:315–392, 05 2013.
- [130] G. Samorodnitsky and M.S. Taqqu. *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance: Stochastic Modeling*. 1994.
- [131] J. D. Sargan. The Estimation of Economic Relationships using Instrumental Variables. *Econometrica*, 26(3):393, July 1958.
- [132] E. Smith, J. D. Farmer, L. Gillemot, and S. Krishnamurthy. Statistical theory of the continuous double auction. *Quantitative Finance*, 6:481–514, 2003.
- [133] George Soros. *The Alchemy of Finance*. John Wiley & Sons, 2003.
- [134] Shanshan Wang, Rudi Schäfer, and Thomas Guhr. Average cross-responses in correlated financial markets. *The European Physical Journal B*, 89(9), 2016.
- [135] Shanshan Wang, Rudi Schäfer, and Thomas Guhr. Cross-response in correlated financial markets: individual stocks. *The European Physical Journal B*, 89(4), 2016.
- [136] Ban Zheng, François Roueff, and Frédéric Abergel. Ergodicity and scaling limit of a constrained multivariate Hawkes process. February 2014.



## RÉSUMÉ

---

Cette thèse est consacrée à l'étude de la microstructure du marché et de la dynamique des prix sur le marché électronique. Nous commençons par construire un modèle de flux d'ordres dans le cadre du processus ponctuel. Nous combinons le processus de Hawkes avec la propriété dite "queue reactive" introduite pour la première fois dans [87]. Dans notre modèle, l'intensité du taux d'arrivée des commandes dépend explicitement de l'état actuel du carnet de commandes limité et également de flux de commandes passés. Nous prouvons l'ergodicité dans ce modèle, ce qui permet de l'appliquer à des fins de simulation. La seconde partie est consacrée à l'analyse de la volatilité rugueuse. À partir d'un champ aléatoire gaussien, nous construisons une famille de processus aléatoires paramétrés. Notre approche unifie deux modèles de volatilité rugueuse célèbres, comme le RFSV et la marche aléatoire multifractale (MRW), sous le même cadre. Nous avons également proposé un estimateur GMM plus fiable pour calibrer le paramètre de Hurst  $H$ . La dernière partie de cette thèse porte sur l'analyse de la covariance des prix. Dans le cadre du processus de Hawkes, nous exprimons la covariance du prix comme une conséquence de flux d'ordres arrivant sur le LOB. Cette représentation permet d'identifier l'influence de chaque participant du marché sur la covariance des prix.

## MOTS CLÉS

---

Microstructure de marché, carnet d'ordres, processus de points, volatilité multifractale

## ABSTRACT

---

This thesis is dedicated to the study of market microstructure and price dynamics in the electronic market. We start by constructing an order flows model under the framework of point processes. We combine multivariate Hawkes processes with the so-called "queue reactive" property firstly introduced in [87]. In our model, the intensity of order flows depends explicitly on the current state of the Limit Order Book and also on past order flows. Ergodicity is proven in this model, which allows one to apply it for simulation purposes. The second part is dedicated to the analysis of rough volatility. From the Gaussian random field, we construct a family of parametrized random processes. Our approach unifies two famous volatility models, the rough fractional stochastic volatility (RFSV) model and the multifractal random walk (MRW), under the same framework. We also proposed a more reliable GMM estimator to calibrate the Hurst parameter  $H$ . The last part of this thesis highlights the analysis of price covariance. Under the framework of multivariate Hawkes processes, we express the covariance of price as a consequence of cascading order flows arriving on the LOB. This representation allows identifying each market participant's influence over the price covariance.

## KEYWORDS

---

Market microstructure, limit order book, point process, multifractal volatility