Unraveling Financial Market Quakes : Exploring Endogenous Volatility Dynamics in Interconnected Markets

Cécilia Aubrun

INSTITUT POLYTECHNIQUE DE PARIS

NNT : 2024IPPAX066

Thèse de doctorat

ÉCOLE POLYTECHNIQUE

IP PARIS

Unraveling Financial Market Quakes: Exploring Endogenous Volatility Dynamics in Interconnected Markets

Thèse de doctorat de l'Institut Polytechnique de Paris
préparée à l'Ecole Polytechnique

École doctorale n°126 Ecole Doctorale de l'Institut polytechnique de Paris (ED IP Paris)
Spécialité de doctorat : Physique

Thèse présentée et soutenue à Palaiseau, le 18 octobre 2024, par

CECILIA AUBRUN

Composition du Jury :

Johannes Muhle-Karbe Imperial College London	Président
Aurélien Alfonsi Ecole des Ponts	Rapporteur
Fabrizio Lillo Università di Bologna	Rapporteur
Sophie Laruelle University Paris-Est Créteil	Examinateur
Mathieu Rosenbaum Ecole Polytechnique	Examinateur
Jean-Philippe Bouchaud Capital Fund Management & Académie des sciences	Directeur de thèse
Michael Benzaquen CNRS & École polytechnique	Directeur de thèse
Antoine (Jack) Jacquier Imperial College London	Invité

Declaration

The work contained in this thesis is my own work unless otherwise stated.

Remerciements

Many individuals have contributed to the successful completion of this PhD, and I am deeply grateful to all of them for their support and participation in my academic journey.

First and foremost, I would like to express my deepest gratitude to my PhD advisor, Professor Jean-Philippe Bouchaud. Your unwavering support, expert guidance and patience have been the the keystone of my doctoral journey. It is an immense privilege to learn from you. I am also profoundly thankful to my co-PhD advisor, Professor Michael Benzaquen, for his trust and invaluable time.

Additionally, I am deeply grateful to the reviewers, Aurélien Alfonsi and Fabrizzio Lillo, for their extremely valuable feedback. It is an honor to have had their insights and to have them take the time to evaluate my work. Moreover, I would like to extend my gratitude to all the members of the defense jury for their time and interest in my research – Sophie Laruelle, Mathieu Rosenbaum, Johannes Muhle-Karbe and Jack Jacquier. In particular, my thanks go to Johannes Muhle-Karbe and Jack Jacquier, who, four years ago, encouraged me to pursue this PhD path–a path I might not have considered otherwise–and who have been very supportive throughout this journey. Furthermore, I am thankful to my thesis committee members, Mathieu and Iacopo, for their time, guidance, and continued support.

I also wish to extend my thanks to the administrative and technical staff at Polytechnique–Lutz, Melanie, Sandrine, Magdalena, Christophe, Toaï, Delphine and the psychologist Sylvie Coussot. Despite their busy schedules, they always found the time to support my project and help find solutions.

One of the notable advantages of pursuing a PhD with the EconophysiX chair is the opportunity to learn from not one but two teams of dedicated and skilled individuals. Thus, I would like to thank CFM's teams for their support, particularly Jean-David, Lamine, Kenji, Julien, Géraldine, Christophe, Riccardo, Mehdi, Maria, Adam, Remy, Emmanuel, and Marcello. I am honored and excited to continue my professional journey alongside them.

I am grateful to the present and former students of the EconophysiX chair–Nirbhay, Pierre, Theo, Guillaume, Antoine, Antoine-Cyrus, Swann, Ruben, Victor, Elia. A special thank you goes to Rudy for undertaking a challenging project with me and seeing it through to publication. I am also thankful to Natascha and Jutta for their commitment to our common projects. Additionally, I would like to extend a special thank you to the colleagues whom I now consider friends–Salma, Samy, and Jérôme–for their support, engaging discussions (both scientific and otherwise), and open-mindedness. This PhD experience would not have been the same without you.

Je voudrais également exprimer ma gratitude envers mes amis et ma famille.

Je tiens d'abord à remercier mes parents de m'avoir toujours soutenue moralement, financièrement et logistiquement. Leur engagement a largement contribué à la réussite de mes projets et je me sais très chanceuse de les avoir à mes côtés. Je souhaite également remercier mon petit frère, Lucas, pour son soutien et nos conversations enrichissantes. Ton intelligence sans effort m'épatera toujours ! Je tiens à remercier chaleureusement ma petite soeur Tina, non seulement pour avoir partager son quotidien avec moi pendant un temps à Londres mais surtout pour son soutien et son aide dans la rédaction de mon introduction. Tu es un exemple inspirant de tenacité, d'engagement et d'ambition. Je te souhaite plein de réussite et de bonheur dans ton doctorat !

Je remercie mes grands-parents–Andrée, Daniel, Anne et Michel–pour leur soutien inconditionnel au long de cette thèse. Je remercie également mes cousins, Simon et Camille, pour avoir tracé la voie, pour être de véritables modèles, et pour avoir démystifié le parcours des études interminables.

Je tiens à remercier mes ami.e.s parisien.e.s (ou presque) pour leur soutien et leur curiosité pendant ces 4 années–Max, Lucie, Carl, Aurélie, François, Océane, Baptou, Paul, Manon, Benoît, Inès, Lauren, Zar, Reff, Baka, Kichot, Tokx, Dassy, Modjo, Trig. Je remercie particulièrement

Bubu pour avoir sauvé mes dents et mon sourire et m'avoir épargné beaucoup de stress lors de la fin de cette thèse.

Enfin, je tiens à adresser mes plus sincères remerciements à mon futur mari, Florian, pour son soutien, sa compréhension, sa patience, pour toutes nos conversations, pour m'avoir aidée à préparer mes présentations et pour me faire rire dans toutes les situations. Je t'aime !

Thank you all for your contributions to this significant milestone.

Résumé

Le rôle des mécanismes de rétroaction dans la création d'instabilités sur des marchés financiers a largement été étudié dans la littérature. L'endogénéité de la volatilité et de l'activité des marchés a effectivement conduit à plusieurs krachs notoires. Le plus connu, le flash crash du 6 mai 2010, illustre comment les instabilités du marché découlent de ses caractéristiques intrinsèques. En effet, c'est une exécution excessivement rapide d'ordres de vente qui a déclenché la chute rapide suivie d'une remontée du S&Pmini en moins d'une heure. De plus, les instabilités des marchés sont accentuées par leur nature multidimensionnelle et leur connectivité, comme le montre la propagation de la volatilité à travers divers actifs financiers lors d'événements tels que le flash crash mentionné ci-dessus. En effet, le 6 mai 2010, le flash crash du S&Pmini a affecté 300 autres actifs. Cette thèse étudie empiriquement et théoriquement l'endogénéité des mouvements de prix en multi-dimension.

Dans un premier temps, nous cherchons à caractériser les sauts de prix empiriques. Basés sur des recherches interdisciplinaires montrant que l'asymétrie temporelle peut être utilisée pour classer les sauts d'activité comme exogènes ou endogènes, nous développons une nouvelle méthode, non supervisée, basée sur des coefficients d'ondelettes (particulièrement adaptés pour refléter l'asymétrie temporelle) afin de mesurer la réflexivité des sauts de prix univariés. Par ailleurs, notre représentation a révélé que le retour à la moyenne et l'alignement avec la tendance sont deux caractéristiques supplémentaires, permettant d'identifier de nouvelles classes de sauts. Enfin, cette représentation permet d'étudier les propriétés réflexives des co-sauts, définis par des sauts de prix simultanés (au cours de la même minute) de plusieurs actifs. Il apparaît qu'une fraction significative des co-sauts résulte d'un mécanisme de contagion endogène. Ainsi, l'événement du 6 mai n'était pas un incident isolé, et des dynamiques endogènes couplées avec une forte connectivité contribuent aux instabilités des marchés. Parallèlement, nous étudions le modèle du Hawkes quadratique (QHawkes), utilisé pour décrire la volatilité à haute fréquence. Les processus QHawkes sont des processus de Poisson qui, par l'expression de leur intensité, décrivent l'influence du passé sur la probabilité de l'activité future. Des travaux antérieurs ont montré que le modèle QHawkes univarié reproduit plusieurs caractéristiques empiriques : les queues épaisses des distributions, le regroupement temporel de la volatilité et l'asymétrie temporelle (effets de levier et effet Zumbach). En outre, des résultats supplémentaires sur la stabilité des processus de QHawkes sont discutés, montrant que la rétroaction quadratique peut induire des événements extrêmes tout en restant stable grâce à l'équilibre les réalisations inhibitrices et excitatrices.

Pour étudier le caractère multidimensionnel des marchés, nous étendons le modèle QHawkes à plusieurs dimensions, en considérant plusieurs actifs et leurs interactions croisées. Un cadre multi-actifs nécessite la prise en compte de faits stylisés supplémentaires, tels que la prévalence des co-sauts et les effets d'asymétrie temporelle croisés. En effet, ce travail met en lumière les effets de levier et de Zumbach croisés. En développant deux modèles, nous montrons que le modèle QHawkes multivarié (MQHawkes) peut reproduire les faits empiriques observés sur les marchés financiers. La calibration du modèle sur des paires d'actifs confirme que les marchés sont au bord de l'instabilité.

Pour compléter, un autre modèle multivarié de volatilité dépendante de la trajectoire est étudié : le Nested Factor Model avec des processus log-SfBM comme volatilités. Ce modèle permet de réconcilier les différences de rugosité entre les indices et les actions, offrant de nouvelles perspectives sur les dynamiques de la volatilité multivariée.

Abstract

Past research has highlighted that feedback mechanisms underlie many financial markets instabilities. Endogenous dynamics of markets volatility and activity have indeed led to various notable crashes. Case in point: the events of May 6th, 2010, commonly referred to as the 2010 flash crash, exemplify how market instabilities stem from intrinsic features of financial markets. As evidence, an excessively rapid execution of sell orders triggered the rapid decline and subsequent recovery of the S&Pmini within the span of an hour. Moreover, market instabilities are compounded by their multidimensional nature and interconnectedness, as demonstrated by the propagation of volatility across diverse financial assets during events like the aforementioned flashcrash. Indeed, on May 6th, 2010, the S&Pmini flash crash affected 300 other assets alongside the S&Pmini.

This thesis presents both a data-driven approach and a theoretical approach to investigate the endogenous nature of price movements within a multivariate framework. Our data-driven approach aims to characterize empirical price jumps. Leveraging interdisciplinary research suggesting that the time-asymmetry of activity can be used to classify bursts of activity as exogenous or endogenous, we develop a new unsupervised method based on wavelet coefficients (particularly suitable to reflect time asymmetry) to measure reflexivity of univariate price jumps. On top of that, our wavelet-based representation revealed that mean-reversion and trend are two additional key features, permitting identification of new classes of jumps. Furthermore, this representation allows to investigate the reflexive properties of co-jumps, defined by multiple stocks experiencing price jumps within the same minute. We argue that a significant fraction of co-jumps results from an endogenous contagion mechanism. Thus, May 6th event was not an isolated incident, and the interplay of endogenous dynamics alongside high levels of interconnectedness contributes to the instabilities observed within markets.

Concomitantly, our theoretical inquiry focuses on the quadratic Hawkes (QHawkes) framework, originally introduced to describe volatility dynamics at tick-by-tick level. QHawkes processes are Poisson processes, which, through the expression of their intensity, depict the influence of the past on the probability of future activity. Previous work has proved that the univariate QHawkes model replicates several empirical features of financial time series, including fat tails of the returns' distribution, volatility clustering and the time asymmetry effects (leverage and Zumbach effects). Indeed, the supplementary quadratic and leverage feedback allow to overcome the limitations of the original (linear) Hawkes framework. Besides, additional results on the stability of QHawkes processes are discussed, showing that the quadratic feedback can induce extreme events while staying stable by balancing inhibitory and excitatory realizations. To explore market interconnectedness, we extend QHawkes processes into multidimensional settings, encompassing several assets and their cross-interactions. A multi-assets framework necessitates consideration of additional stylized facts, such as the prevalence of co-jumps and cross time asymmetry effects. Indeed, this work sheds light on the cross leverage and cross Zumbach effects. Developing two frameworks, we show that the multivariate QHawkes (MQHawkes) can reproduce the empirical facts observed in financial markets. Calibrating the model on asset pairs further confirms that markets operate on the brink of instability.

To be thorough, another multivariate, path-dependent volatility model is studied: the nested factor model with log-SfBM processes as volatilities. Our findings suggest that this framework reconciles differences in roughness between indices and stocks, offering further insights into the dynamics of multivariate volatility.

Contents

Introduction

> To succeed in a spectacular fashion you had to be spectacularly unusual.
>
> Michael Lewis, *The Big Short*

Market crashes have been an enduring feature of financial markets. Early instances of financial crises date back as early as 33 CE[1] with the Ancient Rome credit crunch [1], while the "Black Thursday" crash on October 24, 1929, set a precedent for the subsequent century as one of the earliest major modern financial crises. Since then, financial markets have experienced significant downturns such as the "Black Monday" on October 19, 1987, the subprime mortgage crisis marked by the collapse of Lehman Brothers on September 15, 2008, and the 2020 COVID-19 crisis. These events underscore the importance of understanding why markets are prone to such extreme price movements. More broadly, this raises the fundamental question:

What drives financial price movements?

One school of thought supports the **Efficient Market Hypothesis**. Introduced by Eugene Fama in the 70s [2], this theory posits that market prices incorporate all available public information and therefore reflect the "true value" of an asset. Consequently, price movements are caused solely by changes in the asset fundamental value, which are aligned with new public information.

Several empirical observations challenge the assumption that markets are efficient. A striking one is Shiller's excess volatility puzzle [3] which demonstrates that price fluctuations are too significant to be accounted for solely by changes in fundamental value. Moreover, comparing the timing of significant price jumps with relevant news release times, numerous studies revealed that many price jumps do not align with external news events [5, 6, 7, 4]. Notably, Joulin *et al.* (2008) [7] found that 95% of their sample of price jumps (comprising over 800 US stocks from 2004 to 2006) were not news related. To illustrate this phenomenon, Figure 1 compares the occurrence times of co-jumps (indicating simultaneous price jumps in multiple assets, to be further explored in Chapter 5) with the times of news releases. The abundance of green points, representing co-jumps that are not news-related, underscores that many extreme price movements are not induced by external news events.

If news is not the main catalyst for prices,

What drives financial price movements then?

With external news excluded, it is logical to inquire into internal mechanisms to elucidate financial price changes.

Previous literature has indeed demonstrated that markets exhibit strong endogeneity. The concept of "market reflexivity", introduced by Soros in 1994 [8] which posits that price movements predominantly arise from feedback mechanisms within markets, has been widely studied [9, 10, 6, 4, 11], with empirical evidence supporting its validity. Notably, measuring the markets reflexivity, Hardiman *et al.* (2013) showed that 80% of market volatility[2] is endogenous [10]. This result was subsequently corroborated by the findings of Chicheportiche *et al.* (2014) and Wehrli *et al.* (2022) [12, 11].

[1] Common Era

[2] **Volatility** is a complex process which accounts for the liability of prices to change rapidly and significantly. For example, an asset with high volatility is more likely to see its price suddenly drop or skyrocket (see Chap. 1 for more details).

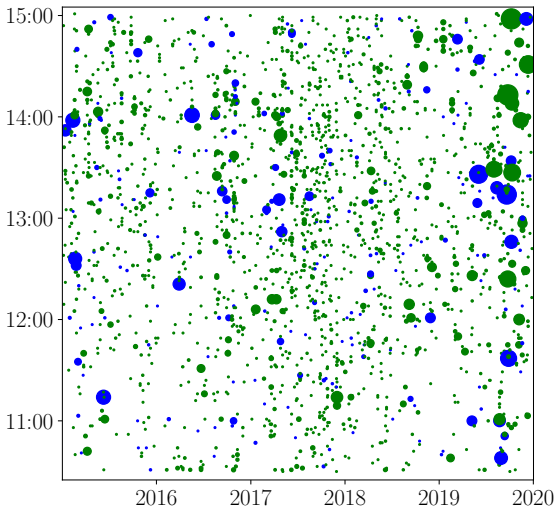


Figure 1: Visualization of our co-jump dataset (295 US stocks over 8 years) (from [4]). The horizontal axis corresponds to the day in the sample and the vertical axis gives the time of day. The size of the circle encodes the number of stocks simultaneously jumping in a given minute (the sizes of the co-jumps span from 2 to 245 assets). The circle's color signifies whether a co-jump coincided with news related to one of the involved stocks within a 3-minute window (blue), or in the absence of reported news (green). Days with FED announcements were excluded (see Figure 3.1 in Chapter 3, which includes these days).

More broadly, in complex systems, extreme events, such as significant price jumps or crashes in financial markets, often arise due to endogenous dynamics. In fact, endogenous occurrences of extreme events and cascades of events are typical features not only of financial markets [13, 14] but also of seismologic activity (earthquakes) [15], biological neural networks [16, 17], crime rates [18], riot propagation [19], book sales shocks [20, 21], daily views of YouTube videos [22], etc. Understanding how these phenomena form is essential to stabilizing their dynamics. Specifically, to illustrate how feedback mechanisms can lead to instabilities in financial markets, we outline several concrete examples extracted from *Trades, Quotes and Prices* by Bouchaud *et al.* (2008) [23].

- The first feedback mechanism which leads to instabilities we review is the so-called **trend following** behavior. Indeed, market participants who consider that past price behavior contains information aim to join a trending market as early as possible. For instance, if they observe a price trending downward, they need to join by selling as early as possible. Consequently, this behavior accelerates the trend since selling further accelerates the price drop.

- A trending market can lead to another form of instability: **Market makers**[3] **panic**. Observing the emergence of a trend, market makers might assume that other participants possess private information[4], prompting them to avoid trading to prevent being "picked off". Consequently, they would cancel their orders on the opposite side of the trend, accelerating the trend and potentially leading to a liquidity[5] crisis, where one side of the market becomes too scared to trade and thus vanishes. It is noteworthy that this scenario led to the infamous 2010 flash crash (see below and Figure 2).

[3]**Market makers** are particular market participants who facilitate trading in financial instruments by providing liquidity to the market (see a subsequent footnote or the glossary for the definition of liquidity). Market makers maintain an inventory of securities and are willing to buy or sell them at best quoted prices, thereby providing liquidity to the market. Their primary role is to ensure that there is enough trading activity in a particular security or asset so that buyers and sellers can transact at any time. Some exchanges also pay them to ensure enough liquidity. Market makers gain money by selling high and buying low, faster than other participants.

[4]Note that this is called bearing the *adverse selection* risk.

[5]**Liquidity** is the ease of converting an asset or security into cash. A liquidity crisis is characterized by the vanishing of market participants, when they are too afraid to trade, at any price.

Cover of LIFE issue of June 8, 1962, following the market crash of May 1962, which ran with the headline: *"What went wrong in the wild stock market and what it means for the U.S."*.

- Trending markets can also give rise to **deleveraging spirals**. Some trading strategies have thresholds on the asset prices they trade (such as a minimum price). Once the price falls below this threshold, traders begin closing their positions (selling), which accelerates the (downward) trend. This, in turn, prompts other traders with similar strategies to follow suit, further amplifying the phenomenon.

- Finally, instabilities also stem from markets interconnectiveness, often leading to **contagion** effects. For instance, a price drop can spread across assets and self-reinforce. If assets A and B influence each other, a price change in asset A affects asset B, which in turn influences asset A's price again, and so on. Contagion mechanisms were notably at play during the 2010 flash crash: the crash originated on the E-mini[6] index but affected more than 300 assets. Moreover, some portfolios aim to replicate indices (such as the S&P500 index), which are combinations of multiple stocks. Consequently, when one liquidates its entire portfolio, it impacts all the stocks comprising the replicated index. This scenario occurred during the quant crunch of August 2007 in ETF index trading.

It is worth noting that all the instabilities mentioned above are self-excited and arise from dynamics that are inherent to markets. Concrete manifestations of these instabilities are numerous throughout the last century. Besides, among the most striking and intriguing realisations of market endogeneity are **flash crashes**.

A typical realisation of markets endogeneity: flash crashes

> At 2:45 on May 6, 2010, for no obvious reason, the market fell six hundred points in a few minutes. A few minutes later, like a drunk trying to pretend he hadn't just knocked over the fishbowl and killed the pet goldfish, it bounced right back up to where it was before. If you weren't watching closely you could have missed the entire event—unless, of course, you had placed orders in the market to buy or sell certain stocks.
>
> Michael Lewis, *Flash boys*

The most famous flash crash, illustrated on the left figure of Figure 2, happened on May 6, 2010 in the US equity market.

On this day, the E-mini value dropped by nearly 9% and bounced back in less than 30 minutes. Upon scrutiny, it appears that at 2.32pm, a large market participant initiated a sell order with a high execution rate. At first, market makers (who trade at high frequency) absorbed the liquidity

[6]The **E-mini index** is the future contract on the S&P500 index which tracks the stock value of the 500 largest companies listed on stock exchanges in the United States.

and bought the orders. However, as more sell orders were submitted, market makers suspected that the sellers had private information. Due to the risk of adverse selection (see above), they began liquidating their positions, exacerbating the selling pressure on the market and accelerating the price drop. At 2.45pm, trading on the E-mini was paused by the Chicago stock exchange (CME), and resumed 5 seconds later, after which prices stabilised. At 3.08pm, prices had returned to their pre-crash levels. This recovery was driven by lower-frequency traders who, recognizing that the price was significantly below its "real" value, saw an opportunity for profit and began submitting buy orders, replenishing the buy side and pushing prices back to their pre-crash levels. The crash propagated to more than 300 other assets, mainly linked to the S&P500 index.
This episode has been extensively studied in academic literature [24, 25, 26] and investigated by regulators such as the Securities and Exchange Commission (SEC) [27] and the Bank of England, which subsequently provided an official definition of flash crashes.

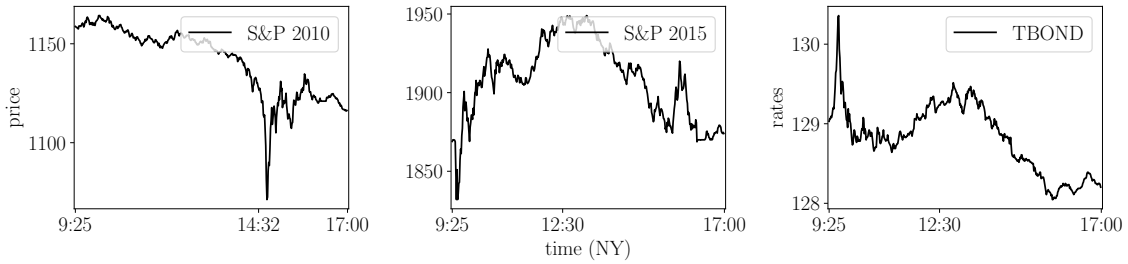


Figure 2: Left figure: Time series of the E-mini during the flash crash of May 6, 2010 (the crash happened around 2.30pm). Middle figure: Time series of the E-mini during the flash crash of August 24, 2015 (the crash happened around 9.30am, markets opening). Right figure: Time series of the TBOND[7] future during the flash crash of October 15, 2014 (the crash happened around 9.32am, right after markets opening).

Definition 0.0.1 (flash crash, *Bank of England, July 2019 Financial Stability Report*)**.** Large and rapid change in the price of an asset **that does not coincide with** – or in some cases substantially overshoots – **changes in economic fundamentals**, before typically retracing those moves shortly afterwards.

The role of high frequency traders in the crash has been a key point of inquiry. More generally, it is often suggested that high frequency trading is responsible for markets instabilities. Notwithstanding high-frequency traders' contribution to accelerating the crash, History shows that crashes caused by internal market mechanisms did not require the presence of high-frequency trading to occur. Indeed, other endogenous crashes happened prior to the era of high frequency trading, as illustrated by the following examples.

- October 24, 1929 - "Black Thursday": on this day, the market lost 11% of its value. This date was followed by "Black Monday", "Black Tuesday" and the Great Depression.
- May 28, 1962 - S&P500 flash crash: LIFE reported on the 1962 flash crash "*Then suddenly, around lunchtime on Monday, May 28, the sell-off swelled to an avalanche*".
- October 19, 1987 - "Black Monday": on this single day, the market fell more than 20%. This crash is attributed to the widely used models which underestimated the risk of large negative returns (these models will be outlined in Chapter 1).

Despite investigations by the SEC and efforts to prevent a recurrence of the events of May 2010, flash crashes persist. Indeed, recently, several notable flash crashes have occurred. We review some of them subsequently.

- October 15, 2014 - US Treasury bond flash crash: in 12 minutes time at market opening, the US Treasury bond market fell off and recovered from an extraordinary 1.6% movement, for no apparent reason. The right figure of Figure 2 illustrates this bond flash crash by plotting the time series of the rates of the "TBOND"[7] at the moment of the crash. Note that rates and prices are inversely proportional which explain why the price drop is represented by a rate spike in Figure 2.

[7]The TBOND is the future contract on US 10 years Treasury bond.

- August 24, 2015 - E-mini flash crash (middle figure of Figure 2): at market opening the E-mini went down by 7%.
- October 7, 2016 - British Pound flash crash: on this day, the British currency fell by 9%. Although there is no certainty as to the reasons for the crash, two leads are predominant: the presence of a fat finger (like in 2010 on the E-mini) or concerns linked to the Brexit negotiations.
- October 23, 2019 - Bitcoin flash crash: The crypto-money fell by 7% in less than 4 minutes without explanation, but it is bitcoin... What would you expect?

In addition to the high-profile flash crashes that garner media attention, financial markets also experience numerous instances of mini flash crashes. These events are characterized by sudden sharp movements in prices in one direction, which are quickly reverted. The timescale of such events is typically less than 1.5 seconds. Mini flash crashes' detection [24, 28, 29] as well as their origins [29, 30, 25] or their link with regulation and high frequency trading [26, 29, 30, 31, 32] have been widely studied in the literature. Notably, Johnson *et al.* (2012) [29] revealed that there were more than 18,000 mini flash crashes in the US equity market between 2006 and 2011. More recently, Christensen *et al.* (2022) [28] found that mini flash crashes occur approximately every other week on the E-mini index, and they are three times more frequent in the Crude Oil futures and Gold futures markets. Additional examples of (mini) flash crashes can be found in [33]. The frequent instances of (mini) flash crashes underscore the pervasive endogeneity of markets.

Outline of the manuscript

Given these observations and inquiries, we conducted this thesis with the aim of gaining deeper insights into the feedback mechanisms inherent to financial markets. Specifically, drawing from existing research, we focused on scrutinizing the endogenous drivers of price movements and volatility, within a multi-asset environment.

This manuscript is organised as follows. Chapter 1 serves as a literature review and presents traditional volatility models while challenging their ability to replicate stylised facts inherent to financial time series, which we need to incorporate.

Chapters 2, 3 and 4 are dedicated to our model of interest: QHawkes process. QHawkes processes come in really handful to delineate the drivers of activity and thus are a relevant choice for our study. Chapter 2 presents the model in one dimension and outlines its main properties. Notably, it discusses the link between the self-excitation of the process (endogeneity level), its stationarity and its explosiveness. Chapter 3 extends the QHawkes in multiple dimensions that is, encompassing several assets and defining feedback loops that influence assets' activities. Finally, Chapter 4 implements a general method of moments to estimate the parameters of the model on real data.

Chapter 5 conducts a more data driven analysis. It aims at investigating the nature of price jumps and co-jumps. It presents a systematic method, which does not depend on labels, to assess the endogeneity of price jumps and co-jumps and thus to evaluate the reflexivity of markets.

Finally, Chapter 6 investigates multivariate volatility with a different model: the Nested Factor model, incorporating log-Stationnary Brownian Motions. The goal is to show that this model conciliates all stylized facts of financial markets, and notably, the differences of roughness between the log-volatility of stocks and that of indices.

Kindly note that a glossary and a notation table are provided in Appendix. When terms from the glossary are first introduced in the text, they are defined in a footnote.

Chapter 1

Modelling financial price changes

Background and Literature Review

> The models don't have any idea of what this world has become...
>
> Michael Lewis, *The Big Short*

In this initial chapter, we begin by introducing the concept of volatility and the methodologies employed to assess it. We then review some of the most widely used frameworks for modeling price changes. By highlighting empirical characteristics of financial time series that they fail to replicate, we challenge those models. Acknowledging these model limitations is important as the use of such models may contribute to instabilities in markets.

A subsequent chapter will elaborate on an additional model, the (Q)Hawkes process. For it was the main focus of my PhD study, Chapters 2, 3 and 4 are dedicated to it and its extension to multiple dimensions.

1.1 What is volatility?

Volatility is a complex process which represents the likelihood and amplitude of price fluctuations. High volatility typically indicates the potential for significant price changes. Volatility is commonly associated with the variance or the standard deviation of price movements.

1.1.1 Price changes, or returns

Within a time window between t and $t + \mathrm{d}t$ (where $\mathrm{d}t$ represents one day for daily data, or, for instance, one minute for intraday data), the price of a financial asset changes from its value at the beginning of the time window, P_t^{open}, to its value at the end of the time window, P_t^{close}. The price change within the time window is then given by:

$$\mathrm{d}P_t = P_t^{\text{close}} - P_t^{\text{open}}$$

Hereafter, we denote X_t the value of the variable X related to the time window $[t; t + \mathrm{d}t]$.

Price changes measure the performance of an asset over time, and are commonly referred to as *returns*. Measuring the relative performance is often favored over using additive price returns, as defined above, notably because, at large time scales, price changes are proportional to asset prices. Thus, log-returns, defined as

$$\mathrm{d}P_t = \log(\frac{P_t^{\text{close}}}{P_t^{\text{open}}}),$$

are frequently preferred to additive ones.

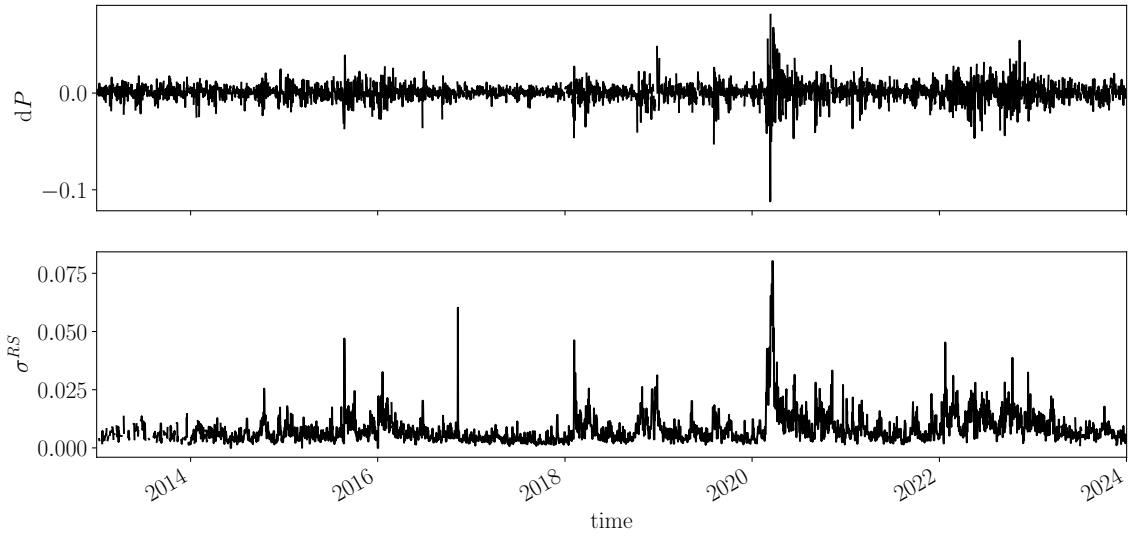


Figure 1.1: Times series of daily financial log-returns (top) and Roger Satchell volatility from equation (1.1.2) (bottom) of the S&Pmini, between 2013 and 2023. Crisis periods, like COVID (2020) are characterised by high volatility and extreme returns. Note that these periods of instabilities are clustered in time.

1.1.2 Estimating volatility

The main challenge with the volatility process is that it is not directly observable; we can only work with estimates. Various proxies of volatility exist (see [34] for a review). This section presents the estimators we use in the subsequent empirical analysis. Some of these estimators require "intra-bin" information, such as the highest and lowest value of the price within each bin, for estimating volatility. Notwithstanding the potential difficulty in obtaining such data, those estimators often provide a more robust valuation of volatility.

Definition 1.1.1 (Garman-Klass Volatility)**.** The Garman-Klass volatility σ^{GK} in a time bin $[t; t + \mathrm{d}t]$ is defined as follow,

$$(\sigma_t^{GK})^2 = 0.5 \log(\frac{P_t^{\text{high}}}{P_t^{\text{low}}})^2 - (2\log(2) - 1) \log(\frac{P_t^{\text{close}}}{P_t^{\text{open}}})^2, \tag{1.1.1}$$

where P_t^{high}, respectively P_t^{low}, is the highest, respectively lowest, price reached in the time bin $[t; t+\mathrm{d}t]$, and P_t^{open}, respectively P_t^{close}, is the first, respectively last, value of the price in the time bin $[t; t + \mathrm{d}t]$.

The above definition does capture the idea that the volatility is null ($(\sigma_t^{GK})^2 = 0$) if there is no price movement within the time bin $[t; t + \mathrm{d}t]$ (i.e., $P_t^{\text{low}} = P_t^{\text{high}} = P_t^{\text{open}} = P_t^{\text{close}}$).

The second estimator we introduce is the Roger Satchell estimator. The particularity of this estimator is that it removes the drift from the price dynamics. The estimator is defined as follow.

Definition 1.1.2 (Rogers Satchell Volatility)**.** The Roger-Satchell volatility σ^{RS} in a time bin $[t; t + \mathrm{d}t]$ is given by

$$(\sigma_t^{RS})^2 = \log(\frac{P_t^{\text{high}}}{P_t^{\text{close}}}) \log(\frac{P_t^{\text{high}}}{P_t^{\text{open}}}) + \log(\frac{P_t^{\text{low}}}{P_t^{\text{close}}}) \log(\frac{P_t^{\text{low}}}{P_t^{\text{open}}}), \tag{1.1.2}$$

where P_t^{high}, respectively P_t^{low}, is the highest, respectively lowest, price reached in the time bin $[t; t+\mathrm{d}t]$, and P_t^{open}, respectively P_t^{close}, is the first, respectively last, value of the price in the time bin $[t; t + \mathrm{d}t]$. Figure 1.1 depicts the time series of the Roger-Satchell volatility in the S&P500 index, along with its log-returns, from 2013 to 2023.

Note that, if the price goes on a direct line without oscillations (following a drift), from P_t^{open} to P_t^{close}, increasingly for instance, then $P_t^{\text{open}} = P_t^{\text{low}}$ and $P_t^{\text{close}} = P_t^{\text{high}}$ rendering $\sigma_t^{RS} = 0$ to signify that fluctuations were null even if the price changed.

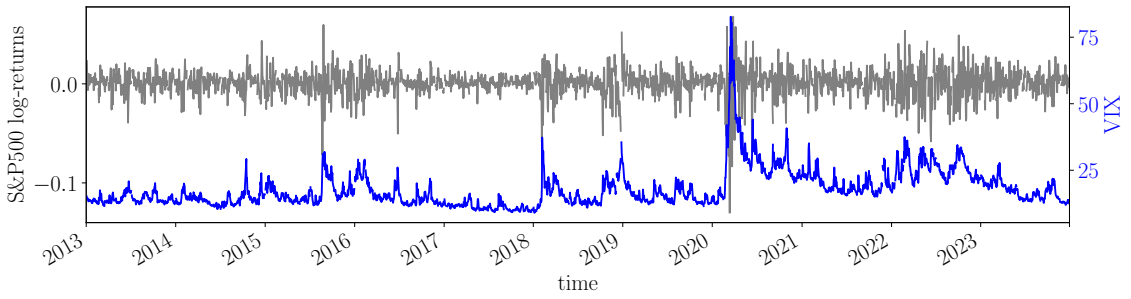


Figure 1.2: Time series of the S&P500 index log returns (grey, left axis) and of the VIX index (blue, right axis) from 2013 to 2023.

An alternative proxy can be derived from Bachelier's thesis [35]. Bachelier demonstrated that for a random walk, the difference between the highest and the lowest values within a bin averages twice the difference between the initial and final values of the bin. Therewith we can derive the following estimator.

Definition 1.1.3 (Bachelier volatility)**.** We define the Bachelier volatility $\sigma^{\mathscr{B}}$ in a time bin $[t; t+\mathrm{d}t]$ as follows:

$$(\sigma_t^{\mathscr{B}})^2 = \frac{1}{3}(P_t^{\text{high}} - P_t^{\text{low}}) + \frac{2}{3}|P_t^{\text{close}} - P_t^{\text{open}}| \tag{1.1.3}$$

where P_t^{high}, respectively P_t^{low}, is the highest, respectively lowest, price reached in the time bin $[t; t+\mathrm{d}t]$, and P_t^{open}, respectively P_t^{close}, is the first, respectively last, value of the price in the time bin $[t; t+\mathrm{d}t]$.

As delineated earlier, "intra-bin" estimates require accessing "intra-bin" values, specifically the highest and lowest values within each bin, to gauge volatility accurately. Obtaining such intra-bin data can pose challenges, particularly with high-frequency data. Consequently, the sum of squared returns is often used as a practical alternative. We define it below.

Definition 1.1.4 (Volatility with sum of square returns)**.** The volatility in a time window $[t; t+\mathrm{d}t]$ (for example the volatility of one day) is the sum of squared returns within that bin (for example, the sum of all the 5min squared returns of the day):

$$(\sigma_t^{SSR})^2 = \sum_{t_i \in [t; t+\mathrm{d}t]} (\mathrm{d}P_{t_i})^2. \tag{1.1.4}$$

Lastly, a notable proxy of volatility is the VIX index, which can be seen as the volatility of the S&P500 index, and is occasionally utilized as an empirical indicator of market volatility. The VIX index is commonly referred to as the "fear index", as it represents the confidence of market participants. Figure 1.2 depicts the VIX index from 2013 to 2023 alongside the log-returns of the S&P500 index. It is worth noting that the dynamics of the VIX closely resemble those of the Roger-Satchell volatility illustrated in Figure 1.1, and that, both Figures 1.1 and 1.2 demonstrate that high volatility is associated with extreme return values.

Throughout this thesis, we will consistently employ these estimators, particularly in the empirical calibration of the models under investigation.

The subsequent sections are dedicated to presenting and scrutinizing various volatility models.

1.2 Volatility models with constant volatility

The first models we review, Bachelier and Black-Scholes models, consider constant volatility dynamics.

1.2.1 Bachelier model

In his PhD thesis, *Théorie de la spéculation*, published in 1900 [35], Louis Bachelier introduced his pioneering model for price changes. The concept posits that the change in price, $\mathrm{d}P_t$, within a time window of size $\mathrm{d}t$, follows an arithmetic Brownian Motion[1] W_t, i.e.

$$\mathrm{d}P_t = \sigma \mathrm{d}W_t,$$

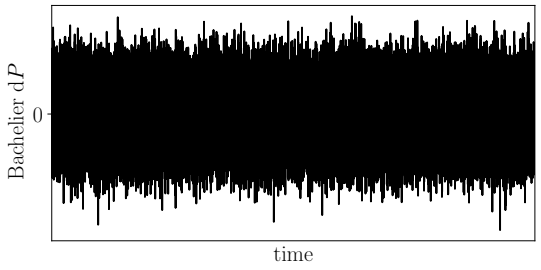


Figure 1.3: Example of $\mathrm{d}P_t$ time series in Bachelier framework.

where σ is constant and represents the volatility of the process $(P_t)_t$. An example of such dynamics is illustrated in Figure 1.3. This model laid the ground for price dynamics theory, and, in particular, for the Black-Scholes model which we present subsequently.

1.2.2 Log-normal price (Black-Scholes model)

As in the Bachelier model, the Black Scholes model assumes constant the volatility, σ, and expresses price changes, $\mathrm{d}P_t$, as related to a Brownian motion, W_t. However, within this framework, prices adhere to a geometric Brownian motion, specifically:

$$\mathrm{d}P_t = \sigma P_t \mathrm{d}W_t.$$

This model was vastly employed by practitioners in the financial industry despite its simplifying assumptions and is believed to be at the origin of many financial crises.

Fun fact: in 1997, R. C. Merton and M. S. Scholes obtained the Nobel price for their model (in collaboration with F. Black), and a year later, their Hedge Fund, Long-Term Capital Management, had to be bailed out after a huge loss.

1.2.3 Empirical observation: volatility is not constant

Both, Bachelier and Black-Scholes frameworks portray volatility as a constant. Nevertheless, numerous empirical observations challenge this assumption.

Firstly, **volatility is clustered in time**. As illustrated in Figure 1.1, portraying the log-returns of the S&P500 index from 2013 to 2024, certain periods display heightened agitation, characterized by more elevated volatility and extreme return values, notably during events such as the COVID-19 crisis or the Ukraine invasion. In contrast, returns generated by the Bachelier model, in Figure 1.3, demonstrate monotonic behavior attributable to the model's assumption of constant volatility.

Secondly, upon examination at intraday level, another empirical evidence against constant volatility emerges: the **intraday pattern of volatility**. Estimating intraday S&Pmini Roger-Satchell volatility (Equation (1.1.2)) on 1-minute intervals over multiple days (utilizing 5 years of data in this instance) and averaging across days yields Figure 1.4b. This figure illustrates the intraday volatility pattern commonly referred to as the U-shape. In fact, there is typically more activity/volatility during the market opening, attributed to reactions to overnight news, and preceding market closure, driven by traders seeking to close positions before the end of day, often after having waited for optimal timing throughout the day. The remainder of the trading day tends to be relatively tranquil, except for a surge around 2 p.m., often prompted by news releases such as FED announcements. This observation further contradicts the idea of constant volatility.

[1] a **Brownian motion** is a way to describe the seemingly random movement of particles or variables over time. It follows a Normal distribution.

It is noteworthy that the crash of 1987 is largely attributed to the assumption of constant volatility within continuous time frameworks. In response to these limitations, more sophisticated models incorporating time-varying and asset price-dependent volatility have been developed. We delve into several of these models in the subsequent sections.

1.3 Models with time-varying and asset price-dependent volatility

1.3.1 Stochastic volatility models

Stochastic volatility models define volatility as a function of time and asset price, $\sigma = \sigma(t, P_t)$.

One of the most used stochastic volatility model is the Heston model [36], named after Steven L. Heston. It describes the price of an asset P_t, and its squared volatility ν_t, via the following stochastic processes:

$$\begin{cases} \mathrm{d}P_t = \sqrt{\nu_t} P_t \mathrm{d}W_t \\ \mathrm{d}\nu_t = \kappa(\theta - \nu_t)\mathrm{d}t + \sqrt{\nu_t}\xi \mathrm{d}B_t \end{cases},$$

where W_t and B_t are Brownian motions with correlation ρ and θ is the long time variance (the value towards which ν_t is driven at a rate κ).

As for Bachelier and Black-Scholes models, the price dynamics within stochastic volatility models are driven by Brownian motions. Consequently, the Gaussian properties inherent in Brownian motions also govern the dynamics of these models. Specifically, the thin-tailed distributions and the time-reversal symmetry of Gaussian processes will manifest in the generated times series. We discuss these two features in subsequent sections.

1.3.2 Empirical observation: price changes are not Gaussian

> The market often underestimated the likelihood of extreme moves in prices
>
> Michael Lewis, *The Big Short*

Given that Bachelier, Black-Scholes, and stochastic volatility models are predicated on Brownian motion dynamics, their resultant return time series adhere to Gaussian distributions[2]. However, a comparison between the empirical distribution of the S&Pmini intraday returns, spanning from 2013 to 2023, and a theoretical Gaussian distribution, as depicted in Figure 1.4a, reveals that the Gaussian distribution fails to encompass all empirical returns. Notably, all empirical returns lying beyond the solid line represent values with a negligible probability of being generated by a Gaussian model. Furthermore, neglecting these extreme values, signifying the most significant price changes that occurred between 2013 and 2023, in a model implies disregarding risky movements. The presence of fat tails in the empirical financial returns distribution stands as one of the most prominent stylized facts observed in financial time series, underscoring the imperative for a robust statistical model to account for them.

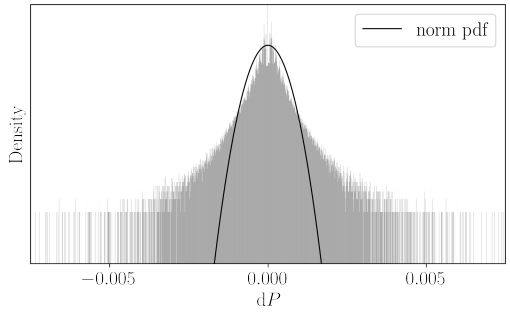
Remark 1.3.1. Stochastic volatility models can yield time series exhibiting fat tails when specific parameter conditions are met [37]. However, these particular parameter regimes are not reflective of the conditions observed in empirical financial time series.

Furthermore, the fat-tailed distribution derived from the stochastic volatility model presented in [37] is based on a geometric representation of returns. This representation is deemed less realistic at the intraday scale when compared to an additive representation [38].

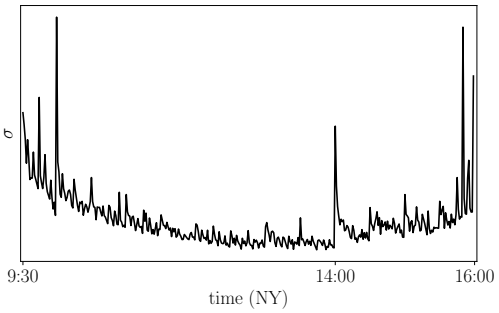
1.3.3 Empirical observation: time-reversal asymmetry of financial time series

Additionally, Gaussian time series also exhibit time-reversal symmetry. However, empirical financial time series manifest time-reversal **a**symmetry. Time asymmetry implies that traversing the

[2]The **probability distribution** is a statistical tool which represents the probability of occurrence of the different outcomes.



(a) Histogram of intraday returns, $\mathrm{d}P$, on 1min bins of the S&Pmini between 2013 and 2023. The plain line represents the probability distribution of a Gaussian. All empirical returns outside the Gaussian line, would not be reproduced by a Gaussian model.



(b) Average 1min bin volatility on the S&P500 futures (computed with Equation (1.1.3)) over all trading days between 2013 and 2018. The intraday pattern of volatility clearly appears as a U-shape with a peak at 2pm due to market news delivery (FED announcements for example).

Figure 1.4: Stylised facts of financial markets: fat tails of returns distribution and U-shape of intraday volatility

time series from past to future is not equivalent to traversing it from future to past, as illustrated in Figure 1.1. This time asymmetry is characterized by two phenomena: the Leverage effect and the Zumbach effect.

Leverage effect

The leverage effect accounts for the one-way correlation between volatility and returns.

Definition 1.3.2 (Leverage effect)**.** The leverage effect is defined, for $\tau > 0$, as

$$|\mathrm{cov}(\mathrm{d}P_t, \sigma^2_{t-\tau})| < |\mathrm{cov}(\mathrm{d}P_{t-\tau}, \sigma^2_t)|, \tag{1.3.1}$$

where $\mathrm{cov}(\mathrm{d}P_{t-\tau}, \sigma^2_t)$ is the covariance between past returns and future volatility (lagged by τ) and, conversely, $\mathrm{cov}(\mathrm{d}P_t, \sigma^2_{t-\tau})$ is the covariance between past volatility and future returns.

Intuitively, the leverage effect can be understood by recognizing that large negative returns induce market panic, thereby increasing future volatility, while, high volatility does not necessarily lead to large future returns. This first asymmetry effect is illustrated for the S&Pmini and the TBOND in Figure 1.5. It is evident that for both the S&Pmini and the TBOND, the correlation between past returns and future volatility (for $\tau < 0$ in Figure 1.5) is (negatively) stronger than the correlation between past volatility and future returns (for $\tau > 0$ in Figure 1.5).

While stochastic volatility models, as discussed in Section 1.3.1, may capture the leverage effect, particularly through the correlation parameter ρ, they are unable to account for all forms of time asymmetry. Specifically, they fail to reproduce the Zumbach effect, which we describe in the following section.

Zumbach Effect

The Zumbach effect [39, 40] accounts for the asymmetry in the correlation between high frequency volatility, σ_t, and low frequency volatility, $(\mathrm{d}P_t)^2$.

Definition 1.3.3 ((Self-)Zumbach effect)**.** To highlight the Zumbach effect, we utilize aggregated returns of day t defined as $R_{t,\tau} := P^{\mathrm{close}}_{t-1}/P^{\mathrm{open}}_{t-1+\tau} - 1$ for $\tau < 0$ and $R_{t,\tau} := P^{\mathrm{close}}_{t+1+\tau}/P^{\mathrm{open}}_{t+1} - 1$ for $\tau > 0$. Low frequency volatility is then represented by $R^2_{t,\tau}$. Further, we note high frequency volatility of day t, σ_t (which can be approximated by the sum of 5min squared returns of day t). For $\tau > 0$, the Zumbach effect is expressed by

$$\mathrm{cov}(\sigma^2_t, (R_{t,\tau})^2) < \mathrm{cov}(\sigma^2_t, (R_{t,-\tau})^2), \tag{1.3.2}$$

that is, past low frequency volatility has more influence on future high frequency volatility than conversely.

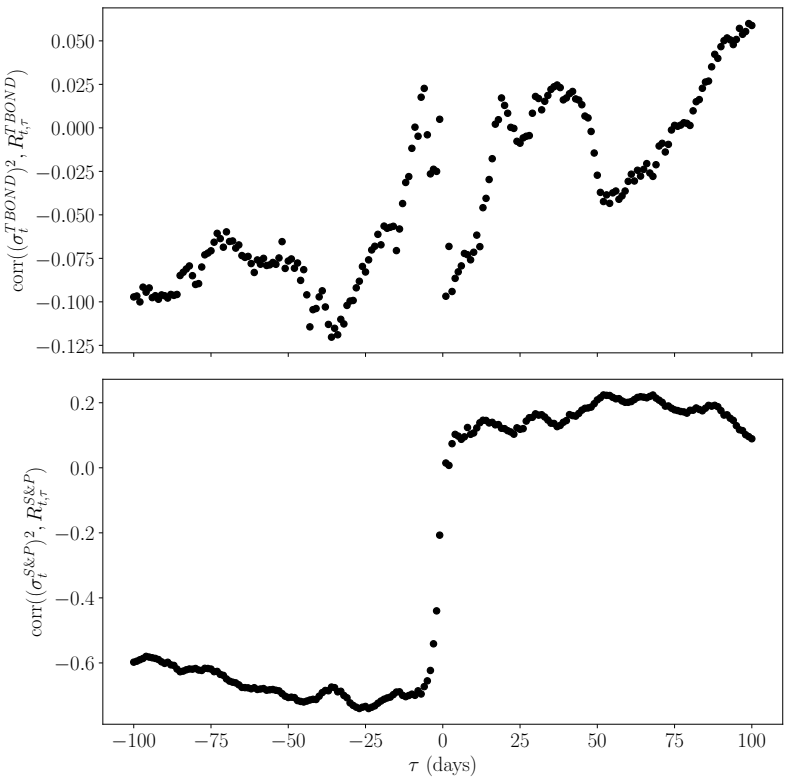


Figure 1.5: Leverage effect for the US 10-years Treasury bond futures ("TBOND") (top) and the S&P500 futures ("S&Pmini") (bottom). The correlation between the high-frequency daily volatility σ_t^2 and the past and future returns $R_{t,\tau}$ on scale τ, shows a clear asymmetry between past and future, specially from past S&Pmini volatility towards S&Pmini returns. The daily volatility σ^2 is computed as the mean of the square returns on 5 min windows over one day (overnight excluded). Aggregated returns are defined as $R_{t,\tau} := P_{t-1}^{\text{close}}/P_{t-1+\tau}^{\text{open}} - 1$ for $\tau < 0$ and $R_{t,\tau} := P_{t+1+\tau}^{\text{close}}/P_{t+1}^{\text{open}} - 1$ for $\tau > 0$. The lag τ is in days. The data covers the period 2013-2018.

Intuitively, the Zumbach effect captures the notion that a trending market (characterized by an increase of R_t^2 over several days) heightens market participants' uncertainty, leading to increased volatility, while increased volatility does not necessarily result in a trending market.

The Zumbach asymmetry for the S&Pmini and the TBOND is illustrated in Figure 1.6. The correlation from past trend to future volatility ($\tau < 0$) is indeed stronger than conversely ($\tau > 0$).

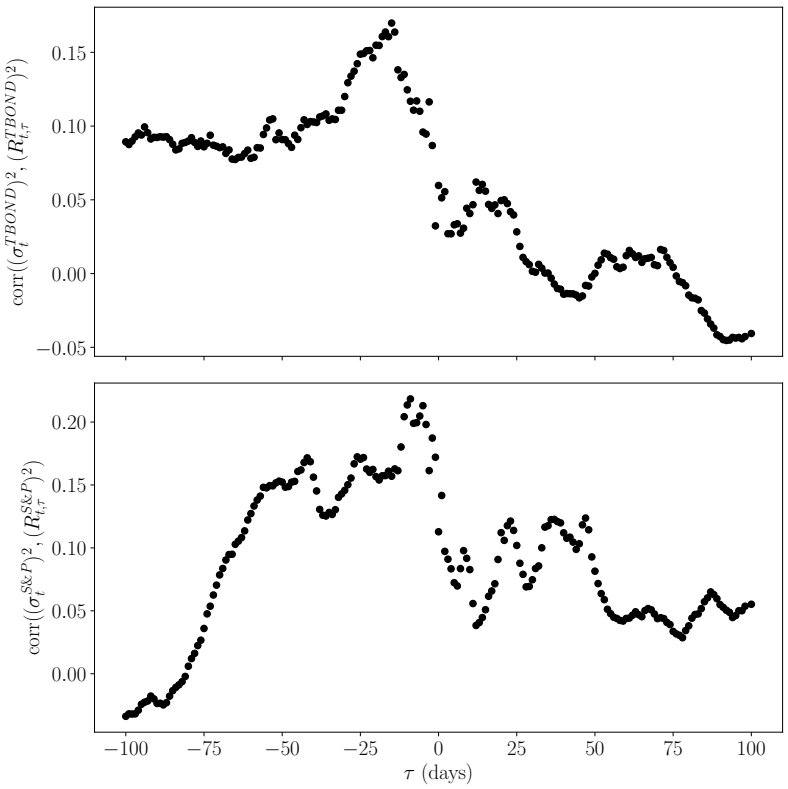


Figure 1.6: Zumbach effect for the US 10-years Treasury bond futures ("TBOND") (top) and the S&P500 futures ("S&Pmini") (bottom) (after factoring in the dominant leverage effect). The correlation between the volatility σ_t^2 and the squared returns $R_{t,\tau}^2$ on scale τ, shows a clear asymmetry between past and future, specially from past volatility towards future trends. The daily volatility σ^2 is computed as the mean of the square returns on 5 min windows over one day (overnight excluded). Aggregated returns are defined as $R_{t,\tau} := P_{t-1}^{\text{close}}/P_{t-1+\tau}^{\text{open}} - 1$ for $\tau < 0$ and $R_{t,\tau} := P_{t+1+\tau}^{\text{close}}/P_{t+1}^{\text{open}} - 1$ for $\tau > 0$. The lag τ is in days. The data covers the period 2013-2018.

Despite their limitations, mostly due to their Gaussian properties, stochastic volatility models are widely used by practitioners and researchers due to their analytical tractability, which allows fast pricing and calibration. Besides, to account for recent empirical findings–specifically, the observation that volatility paths are rougher than Brownian motions [41]–extensions such as the rough Heston model have been developed, opening the way to more comprehensive models [42].

Moreover, there exists another class of time-varying and asset price-dependent volatility models that, unlike stochastic volatility models, generates time-reversal asymmetry by constructing volatility trajectories based on past values. These so-called path-dependent models are presented in the subsequent section.

1.3.4 Path-dependent volatility models

Path-dependent models explicitly describe the dependence of volatility on its past values. Those models are particularly well-suited for capturing the long-range memory of volatility and volatility clustering. In this section, we present two general path-dependent models: the GARCH process and the Multifractal Random Walk.

GARCH model

The generalized autoregressive conditional heteroskedasticity (GARCH) model is widely favored among practitioners and researchers because it replicates many of the stylized facts outlined previously, although it fails to accurately capture the fat tails observed in the distributions of returns. In fact, it tends to overstate the extent of fat tails compared to empirical observations. The model is defined as follows.

Definition 1.3.4 (GARCH)**.** A price returns process $(\mathrm{d}P_t)_t$, driven by a GARCH(p,q), is a centered Gaussian process with volatility σ_t (i.e., $\mathrm{d}P_t \sim \mathscr{N}(0,\sigma_t^2)$), where the dynamics of σ_t are determined by

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i (\mathrm{d}P_{t-i})^2 + \sum_{i=1}^{p} \beta_i (\sigma_{t-i})^2,$$

where α_i and β_j are positive constants.

From the definition it becomes apparent that the influence of the past is directly embedded in the coefficients α_i, β_j, allowing for the generation of activity clustering and the emergence of time asymmetry.

Many versions of GARCH-(or ARCH-)type models have been introduced since the first version of the ARCH/GARCH by Engle in 1982 [43] and Bollerslev (1986) [44]; for a detailed survey see [45].

Multifractal Random Walk

In the 2000s, Bacry *et al.* introduced another path-dependent model: the Multifractal Random Walk (MRW) [46, 47, 48, 49]. The MRW characterises price changes $\mathrm{d}P_t$ such that

$$\mathrm{d}P_t = \exp\left(\omega_T(t)\right) \mathrm{d}W_t, \tag{1.3.3}$$

where $(W_t)_t$ is a Brownian motion and ω_T is an autocorrelated Gaussian process defined with its covariance structure:

$$\mathrm{cov}\left(\omega_T(t), \omega_T(t+\tau)\right) = \begin{cases} -\lambda^2 \ln(\frac{\tau}{T}), & \text{if} \quad \tau < T \\ 0, & \text{otherwise} \end{cases}.$$

T is the time range of correlation and λ is the so-called intermittent parameter. Thus, the path-dependency lies in the autocorrelation structure of the log-volatility ω_T.

The MRW was introduced to capture the multifractal properties of financial time series [50, 51]. Fractals, introduced by Mandelbrot, *are geometric shapes that can be separated into parts, each of which is a reduced-scale version of the whole* [52]. In the context of financial time series, fractal properties imply that they exhibit similar characteristics across different time- or price- scales. For instance, dynamics observed at daily or weekly time-scale are often similar, as illustrated by

Figure 1.7 which depicts the S&Pmini log-returns series for various frequencies. The *multi*-fractal property then captures the different regimes of the market: the volatility clustering. Intuitively, it comes down to stretching or squeezing the time axis [52].

In 2022, Wu *et al.* [53] introduced the stationnary fractional Brownian Motion (SfBM) which serves as a bridge between the rough (stochastic) volatility models and the MRW. Chapter 6 demonstrates how SfBM processes, within a specific framework and with appropriately chosen parameters allow to account for the observed roughness of volatility [41, 53]. For further details on the SfBM, see also [53].

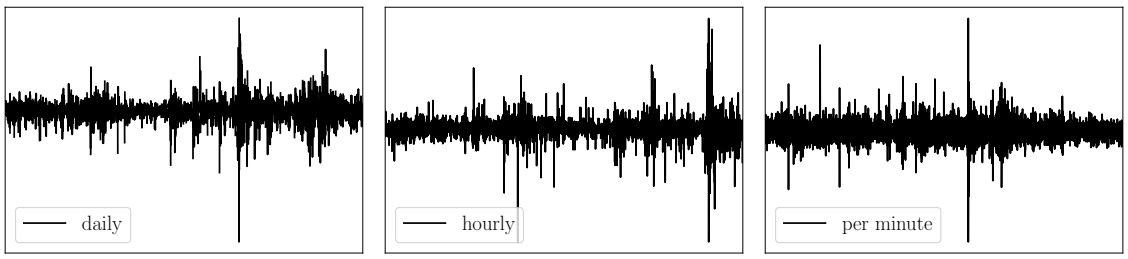


Figure 1.7: S&Pmini log-returns times series for various frequencies. Left figure: daily log-returns from 2013 to 2023. Middle figure: hourly log-returns from 2013 to 2020. Right figure: 1-min log-returns from 2013 to 2017.

Remark 1.3.5. Definition 1.3.4 and Equation (1.3.3) show that both the GARCH model and the MRW are related to Brownian motion. Although the Gaussian properties mentioned above fail to replicate the fat-tailed returns distribution and time-reversal asymmetry of financial time series, they do not limit the GARCH or MRW models. In fact, the "explosive" volatility structure (exponential) and path-dependency of these models allow for the generation of extreme fluctuations (fat tails) and time-reversal asymmetry.

Thus far, we have introduced frameworks and empirical characteristics for describing the price dynamics of individual assets. However, the complexity of markets also stems from their multidimensionality and the interactions among a multitude of traded assets. Given that asset managers construct portfolios comprising multiple assets and that instabilities can propagate across assets, as discussed in the Introduction, it is imperative to consider a multi-asset framework. The following section is dedicated to introducing the additional considerations necessary for modeling volatility across multiple assets.

1.4 Volatility dynamics of several assets

We now consider an environment encompassing multiple financial assets. From the previous sections, we already know that each individual asset exhibits the following characteristics within its financial time series:

- volatility clustering
- fat tails of returns distribution
- multifractal properties
- time-reversal asymmetry.

But what empirical features arise in a multi-asset framework? This section introduces two key features unique to multi-asset environments: co-jump occurrences and cross-time-reversal asymmetry.

1.4.1 Co-jumps

Co-jumps refer to the simultaneous occurrence of jumps in the prices of multiple assets. It is noteworthy that if price processes were driven by independent jump processes, the probability of

simultaneous jumps would be negligible; however, such phenomena are observed frequently.

For instance, upon analyzing the returns time series of 300 US stocks between 2013 and 2022 at a 1-minute time-scale, we identify price jumps (defined as when the value of returns deviates by 4 standard deviations from its mean value) and visualize co-jumps (instances where at least two assets experience a price jump in the same time window) in Figure 1.8. The size and color of the points encode the number of assets experiencing simultaneous jumps within the same minute. As in [54, 7], outcomes of Figure 1.8 are manifolds.

Firstly, note that individual price jumps occur frequently, as evidenced by the abundance of data points and the rolling sum of the number of jumps over a 30-day period, depicted by the solid line in the inset of Figure 1.8. In a Gaussian framework, as in Bachelier, Black-Scholes and stochastic volatility models, such extreme movements (defined as deviations of 4 standard deviations from the mean) would have a negligible probability of occurrence. Thus, Figure 1.8 further refutes Gaussian-based models.

Secondly, co-jumps also occur frequently, with some involving nearly the entire market. Specifically, 2019 was very prone to large co-jumps attributed to the trading war between two world leaders, D. Trump and Xi Jinping. However, notwithstanding those jumps and upon excluding days when FED made announcements (those days were removed for Figure 1.8), no discernible pattern emerges regarding the occurrence of co-jumps. It remains unclear whether co-jumps arise due to external news events or through endogenous propagation. This question will be further explored in Chapter 5.

Lastly, Figure 1.8 reveals that co-jumps are clustered in time, as evidenced by sequences of successive co-jumps. This clustering might be the result of a propagation phenomenon through time and across stocks.

Given the prevalence of co-jumps, it is imperative to acknowledge their existence when working within a multi-asset environment.

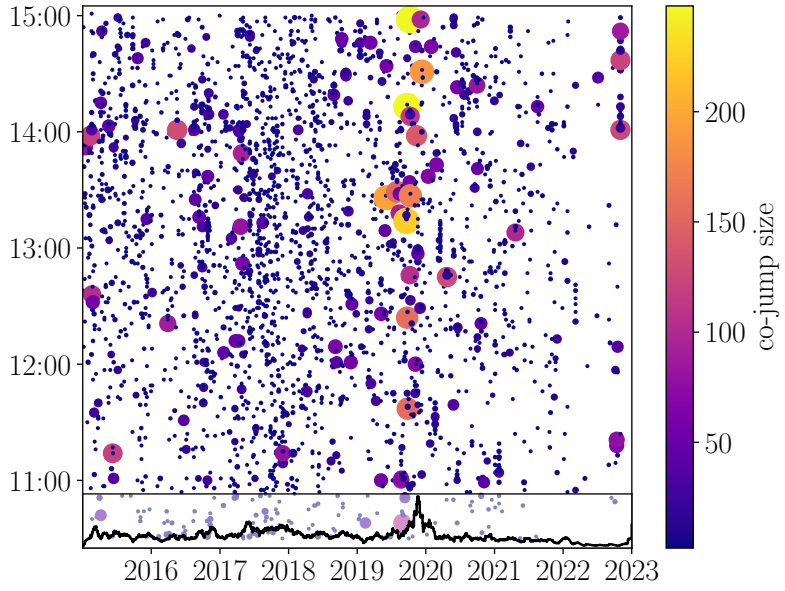


Figure 1.8: Visualization of the co-jumps dataset (295 US stocks, 8 years) (as in [54, 55, 56]). The horizontal axis corresponds to the day of the co-jump and the vertical axis gives the time of day. The size and color of the circle encode the number of stocks jumping simultaneously (in the same minute). Inset: number of jumps on a rolling window of 30 days.

1.4.2 Cross-Time Asymmetry

The time asymmetry effects presented in Part 1.3.3, which were observed for individual assets, can also be investigated across assets. Specifically, we investigate the cross-leverage and cross-Zumbach effects. The cross-leverage effect reveals the asymmetry in the correlation between the returns of an

asset (A) and the volatility of another asset (B). The upper plots of Figure 1.9 illustrate the cross-leverage effect for the pair TBOND and S&Pmini. Furthermore, the cross-Zumbach effect, depicted in the lower figures of Figure 1.9 for the same pair, describes the asymmetry in the correlation between the volatility of an asset (A) and the trend of another asset (B). Again, the asymmetry in the correlations but also across assets is evident from Figure 1.9. For instance, as demonstrated in the lower right plot of Figure 1.9, the past trend of the TBOND exerts more influence on the future volatility of the S&Pmini than vice versa.

Chapter 3 is devoted to the multivariate QHawkes model, which explicitly characterizes asymmetry effects and directly addresses the modeling of co-jumps.

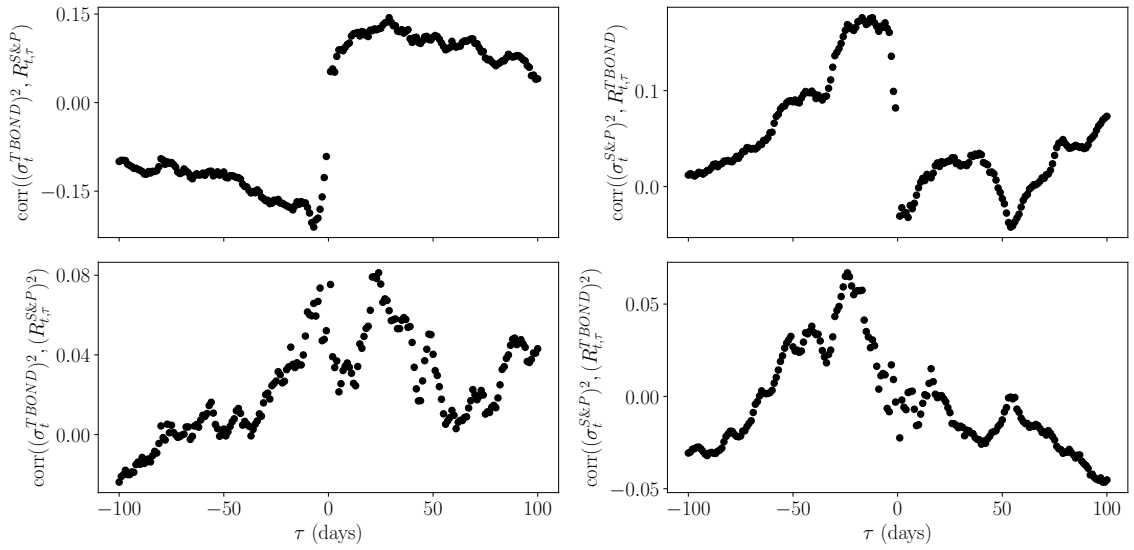


Figure 1.9: Top: Cross asset leverage effect between the S&P500 futures ("S&Pmini") and the US 10 years Treasury bond futures ("TBOND"). We plot the correlation between the high-frequency daily volatility σ_t^2 and the past and future returns $R_{t,\tau}$ on scale τ. Bottom: Cross-Zumbach effect between the S&Pmini and the TBOND (after factoring in the dominant leverage effect). We plot the correlation between the volatility σ_t^2 and the squared returns $R_{t,\tau}^2$ on scale τ, showing a clear asymmetry between past and future, specially from past TBOND trends towards S&Pmini volatility. The daily volatility σ^2 is computed as the mean of the square returns on 5 min windows over one day (overnight excluded). Aggregated returns are defined as $R_{t,\tau} := P_{t-1}^{\text{close}}/P_{t-1+\tau}^{\text{open}} - 1$ for $\tau < 0$ and $R_{t,\tau} := P_{t+1+\tau}^{\text{close}}/P_{t+1}^{\text{open}} - 1$ for $\tau > 0$. The lag τ is in days. The data covers the period 2013-2018.

1.4.3 Additional stylised facts in a multi-asset framework

As co-jumps manifest, financial stocks exhibit joint dynamics, which are also connected to those of indices, often referred to as the "market" directly. Here, we review some of the stylized facts observed in previous research concerning the interconnected dynamics of assets.

Lillo *et al.* [57] observed that the skewness of stock returns distribution is dependent on market (index) returns. Specifically, on days when the market has negative, respectively positive, returns, then the distribution of all stocks returns tends to exhibit negative, respectively positive, skewness.

Furthermore, Cizeau *et al.* [58] and later Allez *et al.* [59] showed that not only does the returns distribution depend on the market, but the idiosyncratic contributions of stocks–those components remaining once the market dynamics are accounted for–also depend on the behavior of the indices.

Lastly, Chicheportiche *et al.* observed that the *joint distribution of stock returns is not elliptical*[3] [60].

[3]**Elliptic models** describe return dynamics as $\mathrm{d}P^i = \sigma_i \epsilon_i$, with ϵ_i a random variable, allowing for the analytical derivation of various properties.

Introduced in [61], the Nested Factor model (NFM) establishes a multi-asset framework that addresses these stylised facts and the fat-tailed distribution of returns. In Chapter 6, we revisit the Nested Factor Model to show that it also accounts for the log-volatility roughness of stocks and indices as observed in [53].

Conclusion

Modeling price changes poses numerous challenges due to the complex nature of financial time series and the intricate dynamics of markets as complex systems. Extending beyond traditional models such as Bachelier, Black-Scholes or stochastic volatility models, recent models have grown increasingly sophisticated, progressively incorporating additional stylized facts of financial time series. Of particular interest are GARCH-type models, which effectively capture volatility clustering, fat-tailed return distributions, and time asymmetry effects. In the following chapter, we introduce (Q)Hawkes processes, which also delineate the influence of past activity on future activity while providing a microstructure description of price dynamics.

The multi-dimensional nature of financial markets significantly contributes to the complexity of the system, introducing new empirical characteristics to consider. Specifically, previous works have highlighted the intricate relationship between stocks and indices [61, 58, 57, 59, 60]. Additionally, we shed light on two empirical characteristics of multi-asset framework: co-jumps and cross-time asymmetry.

Finally, it is worth mentioning that what most volatility models, including both the aforementioned traditional models and the (Q)Hawkes processes, which will be employed in the subsequent chapters, fundamentally lack is a microstructure foundation wherein the resulting stochastic dynamics emerge from the aggregation of behaviors of a large number of agents.

Take Home Message

- Volatility is a complex time process which represents the likelihood and amplitude of a price change.
- Volatility cannot be observed and needs to be estimated.
- The main empirical characteristics of financial time series are:
  - Fat tails of the returns distribution
  - Volatility clustering
  - Volatility intraday patterns
  - Time asymmetries: Leverage & Zumbach effects.
- Bachelier and Black-Scholes models were pionneers in modelling price dynamics. Their representations assume a constant volatility. Empirical observations show that this assumption is wrong: volatility is clustered in time and it presents intraday patterns (U- or J- shape).
- Stochastic volatility models build a stochastic, i.e. time dependent, volatility. Their Gaussian properties, also presented by Bachelier and Black-Scholes models, neither account for the time asymmetry, nor for the fat tails of financial time series.
- The interconnectivity and multidimensionality of markets make it necessary to extend models in multiple dimensions and thus, to consider other stylised facts as co-jumps and cross-time asymmetry.
- None of the volatility models presented above are micro-founded.

Chapter 2

(Q)Hawkes processes

> Time moves in one direction memory another.
>
> William Gibson, *Distrust That Particular Flavor*

This chapter is partly based on [62, 63], written in collaboration with Michael Benzaquen and Jean-Philippe Bouchaud.

Introduction

Hawkes processes have been used in various fields to model endogenous dynamics, where past activity triggers more activity in the future. Indeed, Hawkes processes were found to be relevant to capture the self-excited nature of the dynamics in biological neural networks [16, 17], in financial markets [13, 14], in seismologic activity (earthquakes) [64, 65], and also in crime rates or riot propagation [18, 19]. Standard linear Hawkes processes are basically akin to a branching process, where each event generates on average n_H "child" events. The process cannot be stable when $n_H > 1$, as events proliferate exponentially with time, and no stationary state can ever be reached. When $n_H < 1$, on the other hand, the average event rate reaches a finite constant that diverges as $(1-n_H)^{-1}$ as $n_H \to 1$[1]. Therefore, for standard Hawkes processes, the stability criterion coincides with the condition that the event rate remains finite.

As argued by Kanazawa and Sornette in [68, 69], non-linear Hawkes processes allow one to combine both excitatory and inhibitory effects, and can describe an even broader range of phenomena. One special class of such non-linear processes, called Quadratic Hawkes Processes (QHawkes), were introduced and studied in [70]. On top of the standard Hawkes feedback, a signed process (price changes in the context of [70]) also contributes to the current activity rate, in a quadratic way (see below for a more precise definition). On top of the n_H child events triggered by the Hawkes feedback, the new quadratic feedback contributes to n_Q extra child events. What was shown in [70] is that the average event rate now diverges when the *total* endogeneity ratio $n_H + n_Q$ reaches unity. From this result, it was concluded that the QHawkes is only stationary when $n_H + n_Q < 1$.

The aim of this chapter is to introduce Hawkes processes and their quadratic extension and discuss their stationarity properties. Specifically, we claim that whenever $n_H < 1$ the QHawkes is always stationary, albeit with a *diverging mean intensity* in the case $n_H + n_Q > 1$. More precisely, the distribution density of the local intensity decays asymptotically as a power-law, with an exponent that becomes smaller than 2 whenever $n_H + n_Q > 1$, such that the average intensity diverges. Stationary processes with infinite mean intensity also appeared very recently in the context of non-linear Hawkes processes in Refs. [68, 69]. Intuitively, the mixed excitatory/inhibitory effects encoded by QHawkes allows one to avoid the exponential run-away of Hawkes processes when $n_H > 1$, while describing a highly intermittent process with divergent mean intensity.

[1] The case $n_H = 1$ is special and can also reach a stationary state with finite mean intensity when the immigration rate is zero, see [66], or infinite mean intensity but finite typical (or median) intensity, as shown in [67].

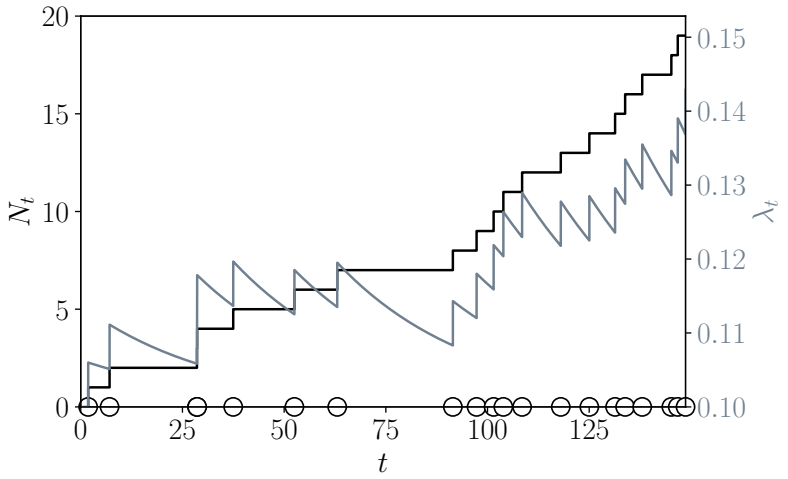


Figure 2.1: Illustration of a Hawkes process with exponential kernel. The circles represent the time of the jumps ($\mathrm{d}N_t = 1$), the black plain line (steps) (with y-axis on the left axis) represents the count of events through time (N_t). The grey line (with y-axis on the right axis) represents the intensity (λ_t), in this illustration the baseline is $\lambda_\infty = 0.1$. At each event (each circle), intensity increases and then exponentially releases and the number of events is incremented by 1. Note the clustering of events, around 100 for instance.

2.1 Hawkes Processes: Definition

2.1.1 Linear Hawkes Process

Definition 2.1.1 (Hawkes processes)**.** A Hawkes process $(N_t)_{t\geq 0}$ is an inhomogenous Poisson process (meaning that its intensity is time dependent), the intensity of which is defined with the past realisations of the process according to the following equation:

$$\lambda_t = \lambda_\infty + \int_{-\infty}^{t} \phi(t-u)\mathrm{d}N_u, \tag{2.1.1}$$

where λ_t is the local intensity, i.e. probability that $\mathrm{d}N_t$ is equal to 1 is $\lambda_t \mathrm{d}t$; λ_∞ is called the baseline intensity and $\phi(\cdot)$ the influence kernel, from which one obtains the endogeneity ratio n_H as the norm of ϕ:

$$n_H = ||\phi|| = \int_{\mathbb{R}} \phi(u)\mathrm{d}u < +\infty. \tag{2.1.2}$$

For an univariate linear Hawkes process to be stable, one needs $n_H < 1$. In fact, this is the necessary condition for the mean intensity $\bar{\lambda} := \mathbb{E}(\lambda_t)$ to be finite. Figure 2.1 gives an illustration of Hawkes process.

The usual framework to describe financial markets using Hawkes processes is to use two processes $(N_t^+, N_t^-)_t$ to describe a change in the price P_t [13]. N^+ counting the events "price is going up" and N^- the event "price is going down". The price process can then be described with $P_t = \psi(N_t^+ - N_t^-)$, where ψ is the tick size[2].

One of the main limitations of Hawkes processes for financial applications is that the stationary distribution of local intensities, $P(\lambda)$, has "thin tails" [71] that cannot reproduce the fat tailed distribution of activity/volatility observed in most financial time series (see Figure 1.4a). Furthermore, as noted in [70], linear Hawkes processes cannot reproduce the violations of time reversal symmetry observed in financial time series [39] (see leverage and Zumbach effects, in Figures 1.5 and 1.6). Figure 2.2 gives intuitions on the stationary distribution and on the time continuous limit of Hawkes process.

Hawkes process have then been developed quadratically in [70].

[2]The **tick size** is minimum increment between two price changes. It depends on the traded asset.

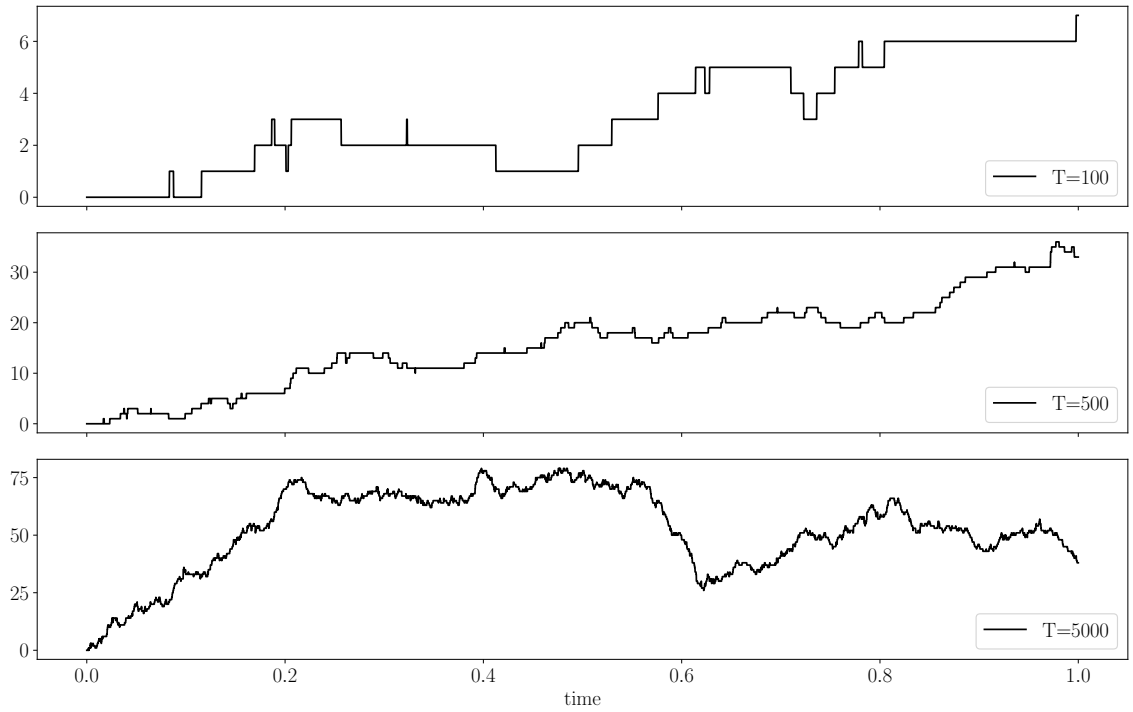


Figure 2.2: Based on [13], the figure plots the process $(N_t^+ - N_t^-)$ at different timescales $T = 100, 500, 5000$, where (N_t^+, N_t^-) are two independent linear Hawkes processes with exponential kernels. It gives some intuition on the continuous time limit of Hawkes processes. Discrete jumps are visible at small timescale, but becomes irrelevant in the scaling limit $T \to +\infty$. Time continuous limit of Hawkes process is a fractional CIR Heston process [71], which is characterized by thin tails and time reversal symmetry.

Remark 2.1.2. The time continuous limit of a discrete process can be seen as if we run the process for a long time, we first "dezoom" from the discrete realisation, and we look at it as if it was a realisation of a continuous processes, which has then a value at every time t.

2.1.2 Quadratic Hawkes

To overcome the limitations of linear Hawkes processes, Blanc *et al.* [70] introduced a quadratic extension of Hawkes processes for financial returns.

Definition 2.1.3 (QHawkes processes (QHawkes))**.** A quadratic Hawkes process ($(N_t)_{t>0}$) is an inhomogenous Poisson process. Its intensity is dependent on both past activity ($\mathrm{d}N_t$) and past price returns defined as $\mathrm{d}P_t := \epsilon_t \psi \mathrm{d}N_t$, where $\epsilon_t = \pm 1$ is an unbiased random sign, independently chosen at each price change, and ψ is the tick size (size of elementary price changes). The Quadratic Hawkes process intensity is now defined as:

$$\lambda_t = \lambda_\infty + \frac{1}{\psi}\int_{-\infty}^{t} L(t-s)\mathrm{d}P_s + \frac{1}{\psi^2}\int_{-\infty}^{t}\int_{-\infty}^{t} Q(t-s,t-u)\mathrm{d}P_s\mathrm{d}P_u, \tag{2.1.3}$$

where $L(\cdot)$ is called the leverage kernel (breaking the $\mathrm{d}P_t \to -\mathrm{d}P_t$ symmetry) and $Q(\cdot,\cdot)$ the quadratic kernel. Note that L and Q must be such that the quadratic form in $\mathrm{d}P_s$ is positive definite, see [70].

Since $\mathrm{d}P_t^2 = \psi^2 \mathrm{d}N_t$, a purely diagonal Q (to wit: $Q(t-s,t-u) = \phi(t-s)\delta(s-u)$) recovers the standard Hawkes kernel. New effects arise when considering non-diagonal contributions to Q, see below.

Taking the expectation of Eq. (2.1.3) provides an exact equation for the mean intensity $\bar{\lambda}$, provided it exists. Noting that $\mathbb{E}[\epsilon_s] = 0$ and $\mathbb{E}[\epsilon_s \epsilon_u] = \delta(s-u)$, one readily obtains:

$$\bar{\lambda} = \lambda_\infty + n\bar{\lambda}; \qquad n := \int_0^{+\infty} Q(s,s)\mathrm{d}s. \tag{2.1.4}$$

Hence, $\bar{\lambda} = \lambda_\infty/(1-n)$ is positive and finite whenever $n < 1$, but becomes formally negative when $n > 1$ which was interpreted in [70] as a regime where the QHawkes becomes non-stationary, in analogy with what happens in the case of standard linear Hawkes processes. As we shall see below, this conclusion is not always warranted.

2.1.3 ZHawkes

As an interesting special case that captures the Zumbach effect (i.e. the correlation between future volatility and past trends [39, 40]), Blanc *et al.* [70] proposed the following ZHawkes specification:

$$L(s) \equiv 0; \qquad Q(s,u) = \phi(s)\delta(s-u) + z(s)z(u), \tag{2.1.5}$$

i.e. a quadratic kernel that is diagonal plus a rank-one contribution. In this case, the intensity of the QHawkes becomes:

$$\lambda_t = \lambda_\infty + H_t + Z_t^2 \tag{2.1.6}$$

where

$$H_t = \int_{-\infty}^{t} \phi(t-s)\mathrm{d}N_s$$

represents the Hawkes component of the intensity whereas

$$Z_t = \int_{-\infty}^{t} z(t-s)\mathrm{d}P_s$$

represents the trend-induced (Zumbach) component. Correspondingly, one can then obtain the endogeneity ratio n (defined in Eq. (2.1.4)) as the sum of the two terms: the Hawkes endogeneity ratio $n_H := ||\phi||$ and the Zumbach endogeneity ratio $n_Z = ||z^2||$. Naively, the stability of the ZHawkes process should read:

$$n = n_H + n_Z < 1. \tag{2.1.7}$$

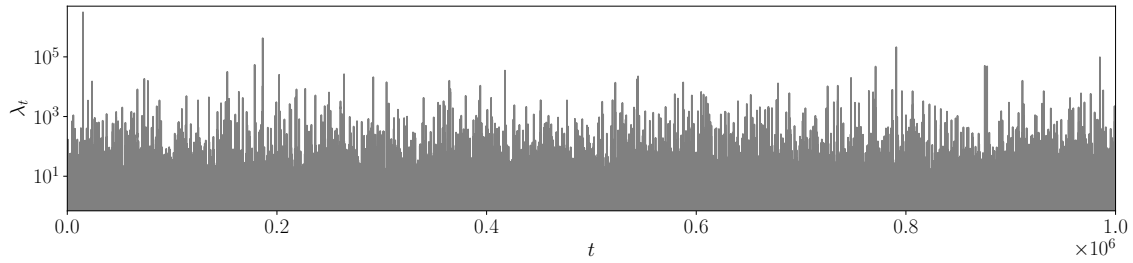


Figure 2.3: Time Series of the Intensity of a simulated ZHP - Parameters: $n_Z = 2$, $T = 10^6$, $\omega = 0.03$, $n_H = 0$ $\lambda_\infty = 0.5$. Note the log-scale on the y-axis.

However, our simulations show that one can have a non explosive process when $n \geq 1$ provided $n_H < 1$. In this case, the QHawkes is stationary but with an infinite mean intensity $\bar{\lambda}$. The stability of the QHawkes for $n_Z + n_H > 1$ arises from the fact that the inhibiting realisations of Z_t (corresponding to locally mean-reverting behaviour of the price) are compensating the exciting ones (corresponding to local trends), see the detailed discussion in [68].

In the following, we show the results of our simulations of an univariate ZHawkes process (ZHP) with $n_H = 0$ and $n_Z > 1$, and an exponential Zumbach kernel $z(\cdot)$. We also simulate the corresponding continuous limit of the ZHawkes process, as worked out in [70], and reinterpret the analytical results of Blanc et al. in the context of the present discussion.

2.2 Univariate ZHP: Numerical Results

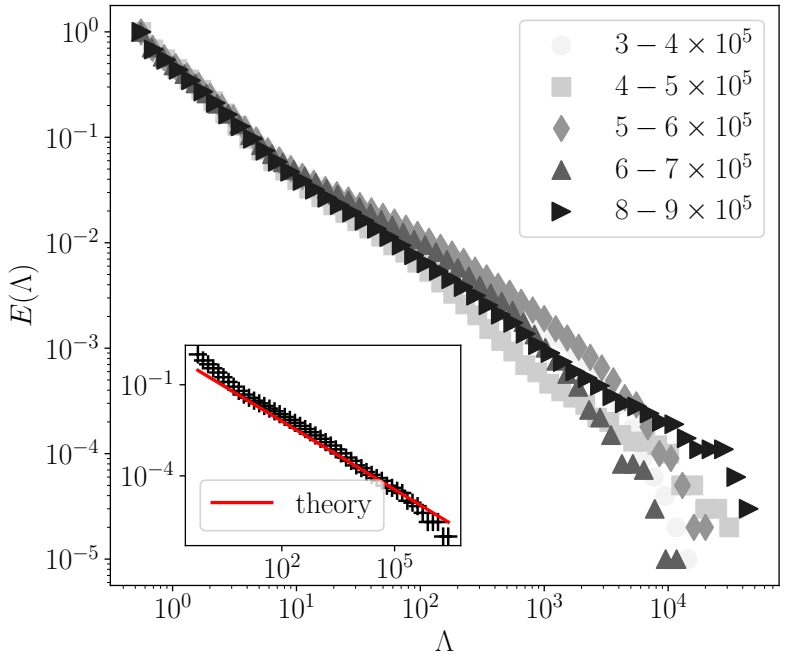


Figure 2.4: The main figure shows the survival function $E(\Lambda) = \mathbb{P}[\lambda > \Lambda]$ in log-log, for several subperiods $t \in (n \times 10^5, (n+1) \times 10^5)$ of the total simulation, with n indicated in the legend. The results fluctuate somewhat, but there is no systematic trend towards a fatter tail at longer times. The inset shows (in log-log) the survival function for the whole period $10^5 \leq t \leq T = 10^6$, together with the theoretical prediction for the tail of the distribution (red line), as given by Eq. (2.2.1). Parameters are: $n_Z = 2$, $T = 10^6$, $\omega = 0.03$, $n_H = 0$ and $\lambda_\infty = 0.5$.

In this section, we provide evidence that the intensity of a simulated ZHawkes Process with $n_Z > 1$ is stationary. In order to simulate a ZHP process, we adapt the thinning algorithm presented by Ogata (1981) in [72]. The Zumbach kernel $z(\cdot)$ is chosen to be a pure exponential,

such that Z_t can be interpreted as an exponential moving average of past returns, i.e. a proxy for the recent trend in prices. More precisely, we set

$$z(s) = \gamma\, e^{-\omega s}, \qquad \gamma := \sqrt{2 n_Z \omega}$$

with $n_Z = 2$, $\omega = 0.03$ and a total simulation time of $T = 10^6$. We also set $n_H = 0$, i.e. no Hawkes feedback, and choose the baseline rate to be $\lambda_\infty = 0.5$.

Figure 2.3 represents the whole time series of the simulated intensity, which shows that the process does not explode and looks stationary. More precisely, we find that the survival function of the process $E(\Lambda) := \mathbb{P}[\lambda \geq \Lambda]$ does not significantly evolve with time, see Figure 2.4. In particular, the distribution does not become significantly "fatter" as time increases, as would be expected if the process was on an explosive path. (Note that a formal Kolmogorov-Smirnov test of this statement is not straightforward because the λ_t's are correlated in time, see [73]). Finally, and most importantly, the empirical distribution function very nicely matches the theoretical prediction of [70], namely:

$$E(\Lambda) \underset{\Lambda \gg \lambda_\infty}{\propto} \Lambda^{-\frac{1}{2}(1+\frac{1}{n_Z})}, \qquad (n_H = 0): \tag{2.2.1}$$

see the inset in Figure 2.4. The expected slope $-3/4$ for $n_Z = 2$ is indeed very close to the fitted slope -0.77 in the range $\Lambda \in [10^2, 10^5]$, beyond which finite size effects become visible.

2.3 Continuous Time Limit of the ZHawkes process

We now further assume that the Hawkes kernel $\phi(\cdot)$ is also exponential and reads

$$\phi(s) = n_H \beta\, e^{-\beta s}.$$

When the parameters ω, β in the kernels $z(\cdot)$ and $\phi(\cdot)$ are sufficiently small, a continuous time limit of the ZHP was derived in [70]. The corresponding two-dimensional SDE reads

$$\begin{cases} \mathrm{d}H_t &= \beta\big[-(1-n_H)H_t + n_H(\lambda_\infty + (Z_t)^2)\big]\mathrm{d}t \\ \mathrm{d}Z_t &= -\omega Z_t \mathrm{d}t + \gamma\sqrt{\lambda_\infty + H_t + (Z_t)^2}\mathrm{d}W_t, \end{cases} \tag{2.3.1}$$

where $\mathrm{d}W_t$ is a Wiener noise. Instead of simulating the original ZHP using the thinning method of the previous section, one can simulate the above SDE, with results shown in Figure 2.5, this time with a non zero Hawkes parameter $n_H = 0.2$, and with $\omega = 0.1$, $\beta = 1$, $\lambda_\infty = 0.5$, $n_Z = 1.5$; such that $n = 1.7 > 1$ but $n_H < 1$. The resulting time series of the process Z_t and the intensity λ_t are presented in Figure 2.5 and look, again, perfectly stationary, even though the criterion ensuring that $\mathbb{E}[H] < +\infty$ and $\mathbb{E}[Z^2] < +\infty$, namely $n_H + n_Z < 1$, is violated here.

In fact, the stationary pdf of the two-dimensional process Eq. (2.3.1) was studied in [70]. It was shown that the tail of the distribution of the intensity λ_t is a power-law, given by an extension of Eq. (2.2.1):

$$E(\Lambda) \underset{\Lambda \gg \lambda_\infty}{\propto} \Lambda^{-\frac{1}{2}(1+\frac{1}{n_Z(1+a)})}, \tag{2.3.2}$$

where a can only be computed in some limits:

$$a \approx \frac{n_H}{1-n_H}\left[1 - \chi\frac{1-n_H-n_Z}{(1-n_H)^2}\right] \quad (\chi = \frac{2\omega}{\beta} \to 0), \tag{2.3.3}$$

and

$$a \approx \frac{n_H}{\chi(1-n_Z)} \quad (\chi = \frac{2\omega}{\beta} \to \infty). \tag{2.3.4}$$

Note that the latter expression is in fact valid for arbitrary χ when $n_H \to 0$, provided $n_H \ll \chi(1-n_Z)$. In particular, $a = 0$ when $n_H = 0$, recovering Eq. (2.2.1).

The conclusion is that the ZHP, at least in the continuum limit, always reaches a stationary distribution when $n_H < 1$, albeit the tail of the distribution of λ_t is a power-law, with an exponent that becomes smaller than unity when $n_H + n_Z > 1$, i.e. leads in that case to a divergent mean intensity.

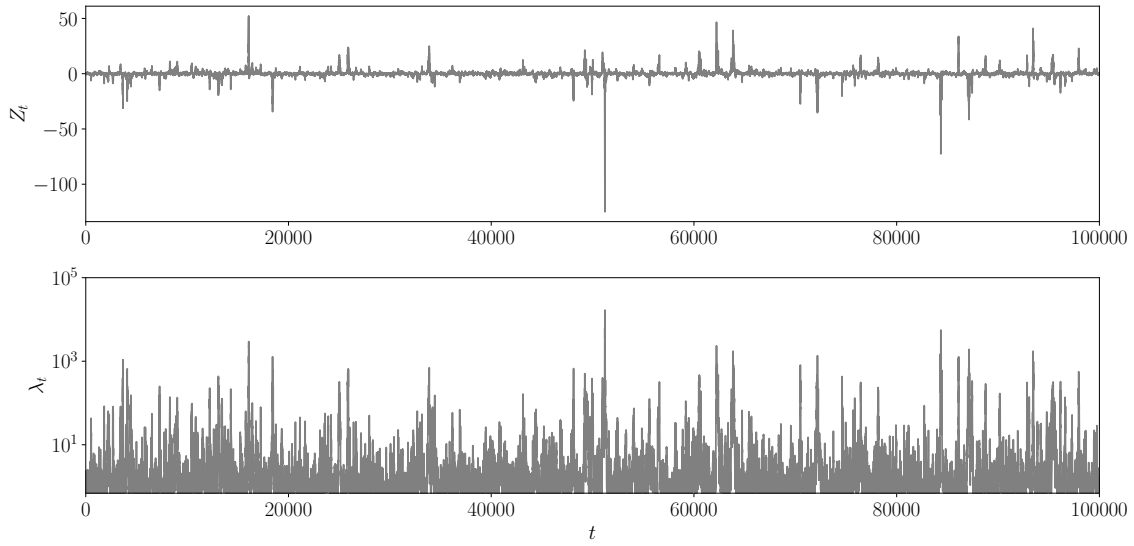


Figure 2.5: Times Series of the Continuous Time Limit of a ZHawkes process according to Eq. (2.3.1). Top: $(Z_t)_t$; Bottom: $\lambda_t = \lambda_\infty + H_t + Z_t^2$ - Parameters: $N = 10^5$ $\omega = 0.1$ $\beta = 1$, $\lambda_\infty = 0.5$ $n_H = 0.2$ $n_Z = 1.5$. Note the log scale on the y-axis for λ_t.

Conclusion

In this chapter, we introduced Hawkes processes and their quadratic extension. Both models offer a microstructure view of the price movements, i.e. the framework is discrete, events are described at a tick-by-tick scale and it is not a time continuous process.

(Q)Hawkes processes are self-exciting processes for which the influence of the past is directly encoded in the intensity definition. The main idea behind these processes is that the more past events occurred, the more likely future events are going to happen. Hawkes processes are used in various fields: biological neural networks [16, 17], financial markets [13, 14], seismologic activity (earthquakes) [64, 65], crime rates or riot propagation [18, 19]... However, the Gaussian properties of Hawkes' time continuous limit make them unfit to describe financial time series. Consequently, a quadratic extension of the process QHawkes, has been developed [70]. The latter allows to reproduce the fat tails of financial time series and the leverage and Zumbach effects.

This chapter also revisited the properties of Quadratic Hawkes processes in the strong feedback regime. Based on numerical simulations and analytical results, we have argued that a new regime exists, where the process reaches a stationary state with an infinite mean intensity. Such a regime is absent for standard (linear) Hawkes processes: the stability of the process requires the mean intensity to be finite, except in the critical case [67]. As argued by Kanazawa & Sornette [68, 69], non-linear Hawkes processes allow for a rich phenomenology, with inhibitory and excitatory effects that can balance each other in a subtle way, resulting in a highly fluctuating, but non-explosive process, for which they provide several other examples.

QHawkes processes naturally lead to a power-law tail distribution for the local intensity, with an exponent that can become less than unity, in which case the mean intensity diverges. The resulting price process then converges to a Lévy stable process with an infinite variance, in spite of the fact that elementary price changes are strictly bounded (and, in our example, equal to $\pm\psi$). As such, this regime is not directly relevant to financial markets, since price returns, although fat-tailed, have a finite variance. Notwithstanding, we believe that the possibility of creating Lévy stable random walks based on a self-exciting mechanism is interesting in itself, and may have applications in other fields.

Finally, our analysis is far from mathematically rigorous. Although the results of [70] for the continuous time version of the Z-Hawkes process are suggestive (see section 2.3), a more formal proof of the stationarity of Quadratic Hawkes processes with general kernels and in the infinite mean intensity regime would be welcome.

Take Home Message

- Originally developed for modeling earthquakes activity, Hawkes processes encode the influence of past events on future activity and are applied in various fields. The main idea of their dynamics is that the occurrence of more events increases the likelihood of future events. A stationary linear Hawkes process is characterised by an endogeneity ratio strictly less than 1, meaning that each event, on average, triggers fewer than one subsequent event.
- Jaisson and Rosenbaum [71] demonstrated that the continuous-time limit of a (near-critical) Hawkes process is a fractional CIR Heston process. This stochastic volatility model generates time series with distributions that have thinner tails than empirical returns distributions and exhibit time-reversal symmetry, in contrast to empirical financial time series.
- To alleviate the limitations of the linear Hawkes for financial prices, Donier, Blanc and Bouchaud introduced the class of "Quadratic" Hawkes (QHawkes) processes [70], which encodes a feedback between past price *trends* and future volatility, directly delineating the time-asymmetry Zumbach effect [39]. Additionally, the QHawkes quadratic feedback naturally generates extreme events and thus replicates the fat-tailed distributions of financial time series (even in the continuous-time limit).
- The ZHawkes process is a particular QHawkes process where the quadratic kernel is a sum of a diagonal and a rank-one component. Thus, the ZHawkes feedback is characterised by the sum of a linear Hawkes feedback and a quadratic feedback. ZHawkes processes have two stationary regimes:
  - when the total endogeneity ratio is strictly less than 1 ($n < 1$), then the process has a finite mean intensity.
  - when the total endogeneity ratio is greater than 1 ($n > 1$) but the Hawkes component endogeneity ratio is strictly less than 1 ($n_H < 1$), then the process has an infinite mean intensity.

  Our results (see also [69]) suggest that the quadratic feedback produces both inhibitory and excitatory outcomes, ensuring stability of the process even when the endogeneity ratio exceeds 1.

Chapter 3

Multivariate Quadratic Hawkes processes

> Someone's sitting in the shade today because someone planted a tree a long time ago.
>
> Warren Buffett

This chapter is largely based on [63], written in collaboration with Michael Benzaquen and Jean-Philippe Bouchaud.

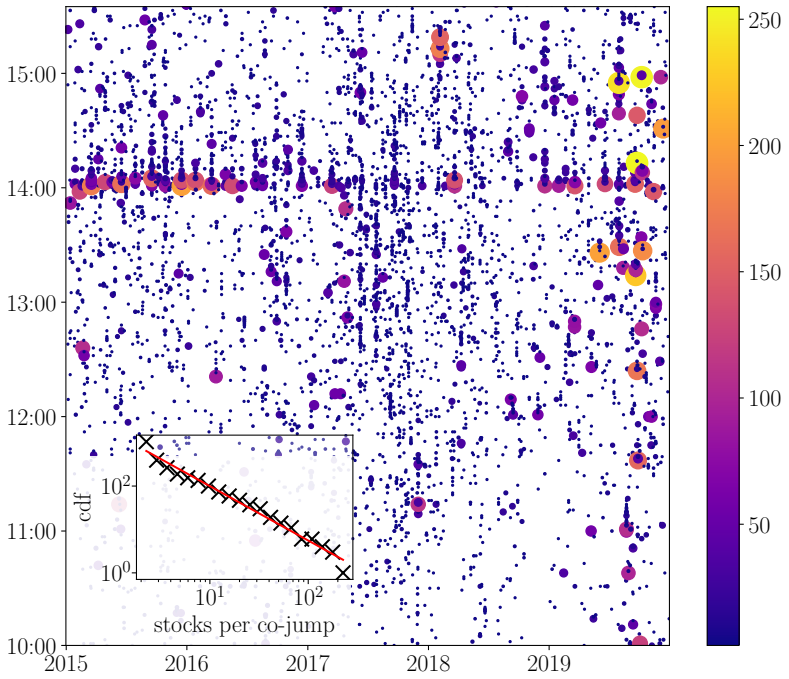


Figure 3.1: Time series of 1 minute bins co-jumps in a set of 295 US stocks over 5 years. The horizontal axis corresponds to the day in the sample and the vertical axis gives the time of day. The size and the color of the circle encodes the number of stocks simultaneously jumping in a given minute (see color bar). The inset shows the cumulative distribution function of the number of stocks in co-jumps for jumps that were classified as endogenous according to the methodology developed in [6], the slope is -1.25.

Introduction

QHawkes processes, introduced in [70] and presented in Chapter 2, provide an interesting alternative to more traditional models, while clearly highlighting the feedback loop at the origin of the stylized facts characterizing financial prices.

In this chapter, we generalize the monovariate QHawkes process of Ref. [70] to the multivariate case, allowing one to capture how the trend on one asset can impact the volatility of another asset. Like for the univariate case, we expect the multivariate expansion to capture some stylised facts that are not accounted for within a linear Hawkes framework. In order to motivate our study, we illustrated in Figure 1.9 the existence of a *cross-asset* leverage and Zumbach effects, i.e. the fact that past trends on the E-mini (resp. TBOND) increases volatility of the TBOND (resp. E-mini), in a way that is asymmetric between past and future, much as the usual "self" Zumbach effect. Additionally, Figure 3.1 (also discussed in Figure 1 and Chapter 5) reveals high activity episodes across the whole stock market, with the number of simultaneous co-jumps being distributed as a power-law. This suggests the existence of a propagation phenomenon similar to a branching process, which requires a specific non-linear cross-asset interaction of the type considered in this work.

3.1 Model presentation

3.1.1 MQHawkes Processes

Let us now consider N financial assets with prices $(P_t^i)_{t\geq 0, i=1,\dots,N}$. We associate such price processes with jump processes $(N_t^i)_{t\geq 0}$, with:

$$\mathrm{d}P_t^i = \epsilon_i \psi_i \mathrm{d}N_t^i, \tag{3.1.1}$$

where we set henceforth $\psi_i \equiv 1$, $\forall i$, without loss of generality. Each jump process $(N_t^i)_{t\geq 0}$ is a conditionally independent[1] Quadratic Hawkes process with intensity $(\lambda_{i,t})_{t\geq 0}$

$$\lambda_{i,t} = \lambda_{i,\infty} + \sum_{j=1}^{N} \int_{-\infty}^{t} L_j^i(t-s)\,\mathrm{d}P_s^j + \sum_{j\leq k}^{N} \iint_{-\infty}^{t} K_{jk}^i(t-s, t-u)\,\mathrm{d}P_s^j\,\mathrm{d}P_u^k. \tag{3.1.2}$$

Note that the superscript in the kernels indicates which asset is affected by the feedback, whereas subscripts indicate which assets are responsible for it.

In the following we shall generically call $\mathbb{K}_\mathrm{d}$ the "diagonal" feedback terms, i.e. the kernels K_{jk}^i with $j = k$. These terms describe the quadratic feedback from two price changes of the same asset j onto the activity of asset i. Similarly, $\boldsymbol{K}_\times$ describes cross effects, i.e. K_{jk}^i with $j < k$, from two different assets $j < k$ onto the activity of asset i. In order to guarantee that intensities λ_t^i remain positive at all times, kernels $L, \mathbb{K}_\mathrm{d}$ and $\mathbf{K}_\times$ need to verify some conditions, which are detailed in Appendix A.1.

Although all these terms could in principle play a role, in the present paper we shall restrict to cross terms K_{jk}^i such that either j or k are equal to i. The intuitive meaning of such terms will become clear below in the context of ZHawkes processes; in particular, we shall see why terms with $j \neq i$ and $k \neq i$ are not expected to play a large role in practice.

For the sake of clarity we will mainly focus on the bivariate case $N = 2$, for which the leverage kernel and the diagonal quadratic kernel are 2×2 matrices:

$$\mathbb{L} := \begin{pmatrix} L_1^1 & L_2^1 \\ L_1^2 & L_2^2 \end{pmatrix}; \quad \mathbb{K}_\mathrm{d} := \begin{pmatrix} K_{11}^1 & K_{22}^1 \\ K_{11}^2 & K_{22}^2 \end{pmatrix},$$

whereas the cross quadratic kernel is a vector

$$\boldsymbol{K}_\times := \begin{pmatrix} K_{12}^1 \\ K_{12}^2 \end{pmatrix}.$$

In the following 2×2 matrices and 2D vectors are respectively noted $\mathbb{A}$ and $\boldsymbol{A}$, for example $\boldsymbol{\lambda} := (\lambda_1, \lambda_2)$. We will also need to distinguish the "time diagonal" of a matrix $\mathbb{A}(\tau_1, \tau_2)$, by which we mean $\mathbb{A}(\tau, \tau)$, from the "diagonal" of $\mathbb{A}$, which refers to the diagonal components in asset space $\mathbb{A}_{ii}$.

[1] This means that for given intensities $\lambda_{i,t}$, the inhomogeneous Poisson processes $\mathrm{d}N_t^i$ are independent. See further down for the case of correlated jump processes.

3.1.2 ZHawkes Model

The multivariate generalisation of the ZHawkes model amounts to assuming that the quadratic kernels $\mathbb{K}_{\mathrm{d}}$ and $\boldsymbol{K}_{\times}$ can be written as arbitrary singular time-diagonal and time-rank-one contributions, to wit:

$$K^i_{jj}(s,u) = (\phi_{\mathrm{d}})^i_{jj}(s)\,\delta(s-u) + k^i_{jj}(s)k^i_{jj}(u)$$
$$K^i_{ij}(s,u) = (\phi_{\times})^i_{ij}(s)\,\delta(s-u) + k^i_{ji}(s)k^i_{ij}(u) \quad (i<j).$$

However, as long as we consider independent Poisson processes, we can set $\boldsymbol{\phi}_{\times} = \mathbf{0}$. This is because the two processes will almost never jump simultaneously, such that for all u, $\mathrm{d}P^i_u \mathrm{d}P^j_u = 0$ when $i \neq j$ (see however Eq. (3.1.6) below when co-jumps are taken into account). Actually in the so-called Thinning Algorithm ([72] and Appendix B.1) for multivariate processes, commonly used to simulate inhomogenous Poisson process, at most one process can be jumping at each time step.

3.1.3 Endogeneity Ratio and Stationarity Condition

The endogeneity ratios indicate by how much, on average, the feedback loop contributes to the future of the process, and thus, whether the process is stationary or not. To define them for the MQHawkes we use analogies with univariate QHawkes and multivariate linear Hawkes.

Mean Intensity

Since price changes are centred and processes are assumed to be independent, we have $\mathbb{E}(\mathrm{d}\boldsymbol{P}) = 0$ and $\mathbb{E}(\mathrm{d}P^i_s \mathrm{d}P^j_u) = \delta_{ij}\delta_{s,u}\bar{\lambda}^i \mathrm{d}s$. Using this, we find that the vector of mean intensities $\bar{\boldsymbol{\lambda}}$, when finite, writes:

$$\bar{\boldsymbol{\lambda}} = \left(\mathbb{I} - \int_0^\infty \mathbb{K}_{\mathrm{d}}(s,s)\mathrm{d}s\right)^{-1} \boldsymbol{\lambda_\infty}. \tag{3.1.3}$$

This expression must be interpreted with care when $\mathbb{K}_{\mathrm{d}}(s,u)$ contains a singular time-diagonal contribution $\Phi_{\mathrm{d}}(s)\,\delta(s-u)$. In such a case, and throughout this paper, we will interpret $\mathbb{K}_{\mathrm{d}}(s,s)$ as $\mathbb{K}_{\mathrm{d,reg.}}(s,s) + \Phi_{\mathrm{d}}(s)$, where $\mathbb{K}_{\mathrm{d,reg.}}$ is the regular part of the diagonal quadratic feedback. Equation (3.1.3) shows that the mean intensity diverges when the spectral radius[2] of the matrix $\int_0^\infty \mathbb{K}_{\mathrm{d}}(s,s)\mathrm{d}s$ tends to one from below.

Endogeneity Ratio

In the multidimensional case, the endogeneity ratio of a standard linear Hawkes process is defined by the spectral radius of the kernel matrix Φ involved in the expression of $\bar{\boldsymbol{\lambda}}$ [13]. More precisely, introducing

$$\|f\| := \int_0^{+\infty} f(s)\mathrm{d}s$$

one constructs the matrix $\|(\phi)^i_j\|$ and determines its eigenvalue with largest modulus, which in turn defines the dominant endogeneity ratio n. Thus, $n = \rho_{\mathrm{spectral}}(\|\Phi\|)$, where ρ_{spectral} is the spectral radius and $\|\Phi\|$ stands for the matrix $\|(\phi)^i_j\|$.

For a general univariate QHawkes model, one can always decompose the endogeneity ratio n as the sum of the Hawkes endogeneity ratio n_H associated with the singular time-diagonal contribution to $K(\cdot,\cdot)$, and a regular contribution n_Q:

$$n = \int_0^{+\infty} \phi(s)\mathrm{d}s + \int_0^{+\infty} K_{\mathrm{reg.}}(s,s)\mathrm{d}s := n_H + n_Q. \tag{3.1.4}$$

In the special case of a ZHawkes quadratic kernel, the regular contribution reads:

$$n_Q = n_Z = \int_0^{+\infty} k^2(s)\mathrm{d}s. \tag{3.1.5}$$

[2]The spectral radius of a square matrix is the maximum of the absolute values of its eigenvalues.

For a general MQHawkes, the endogeneity ratio is then defined by the spectral radius of $\int_0^\infty \mathbb{K}_{\mathrm{d}}(s,s)\mathrm{d}s$, $n = \rho_{\mathrm{spectral}}(\|\mathbb{K}_{\mathrm{d}}\|)$. However, when decomposing the kernel as a singular time-diagonal contribution and a regular contribution, one must be careful about the fact that the two matrices $\mathbb{K}_{\mathrm{d,reg.}}(s,s)$ and Φ_{d} do *not* commute in general. Hence the spectral radius defining the endogeneity ratio cannot be simply expressed as the sum of a Hawkes contribution n_H (i.e. the spectral radius of $\|\Phi_{\mathrm{d}}\|$) and of a regular or Zumbach contribution (i.e. the spectral radius of $\|\mathbb{K}_{\mathrm{d,reg.}}\|$).

Stationarity Condition

For linear Hawkes processes to be stationary, the endogeneity ratio n needs to be strictly less than one. Hence, only processes with finite mean intensity $\bar{\boldsymbol{\lambda}}$ can be stationary. In the presence of a quadratic feedback, the situation is more intricate. In a previous communication [63] we found that for a univariate Z-Hawkes process to be stationary one needs $n_H < 1$, but not necessarily $n = n_H + n_Q < 1$. When $n_H < 1$ and $n > 1$ one has a stationary process with an infinite mean. In the present multivariate framework, we conjecture that a similar situation holds, with the following definitions: let n_H be the spectral radius of $\|\Phi_{\mathrm{d}}\|$, and $n = \rho_{\mathrm{spectral}}(\|\mathbb{K}_{\mathrm{d}}\|)$, with in general $n \neq n_H + n_Q$. Then, there exists a value $n^\star$ such that:

- if $n_H < 1$ and $n < n^\star$, the process is stationary with finite mean intensity;
- if $n_H < 1$ and $n > n^\star$ the process is stationary with infinite mean intensity.
- if $n_H > 1$, the process explodes and no stationary state can be reached.

We have performed numerical simulations (not shown) that support this conjecture. The value of the critical point $n^\star$ is however difficult to compute in the general case, but in the specific case of a weakly anisotropic two dimensional ZHawkes model, some progress can be made, see Section 3.3.4 and Eq. (3.3.11). In particular, in the isotropic case, the univariate result $n^\star = 1$ is recovered.

3.1.4 Correlated Poisson Processes

One may wonder if the hypothesis of independence of the $\mathrm{d}N_t^i$ for different $i = 1, \ldots, N$ is not too strong to faithfully account for events happening in financial markets. In fact, as in [54], we find that *co-jumps* (i.e. simultaneous jumps in the price of different assets within 1 minute bins) occur fairly frequently, adding a new stylised fact to consider in this multivariate framework. In order to investigate co-jumps empirically, we use a jump detection method (see [6]) on 295 large cap. US stock prices from January 2015 to December 2020. We then count how many stocks display anomalous price jumps in a given minute, and represented such counts in Figure 3.1. Co-jumps are clearly seen to occur. On average, there are 4.5 co-jumps per day, and up to 68 co-jumps in one day. Using [6] to classify each jump as endogenous or exogenous, we compute the cumulative density function of number of stocks included in endogenous co-jumps which displays a power law of slope -1.25 (see insert in Fig. 3.1).

Co-jumps may be due to either a common exogenous shock (like an external piece of news affecting several stocks), or to some endogenous instability triggering a jump for one given stock, which propagates to other stocks. The very interesting question of the exogenous/endogenous nature of co-jumps clearly needs a more refined investigation, in the spirit of [6], and is left for future work.

In [54], the authors show that multivariate linear Hawkes models with independent realisations of the Poisson process do not satisfactorily reproduce co-jumps. Here we propose a way to enforce correlations between Poisson processes, and allow one to account for co-jumps within bivariate QHawkes processes.

Bivariate Poisson Processes

We focus on the bivariate case $N = 2$. The extension to $N > 2$ is also possible, although beyond the scope of the present paper. In order to allow for the possibility of co-jumps, i.e. such that $\mathrm{d}P_t^1 \mathrm{d}P_t^2 \neq 0$, we consider three independent QHawkes counting processes $(N_t^1, N_t^2, N_t^c)_{t\geq 0}$ with intensities $(\mu_{1,t}, \mu_{2,t}, \mu_{\mathrm{c},t})$ defined from past returns:

$$\mu_{a,t} = \mu_{a,\infty} + \sum_{j\in\{1,2\}} \int_{-\infty}^t L_j^a(t-s)\mathrm{d}P_s^j + \sum_{j\leq k\in\{1,2\}} \iint_{-\infty}^t Q_{jk}^a(t-s, t-u)\mathrm{d}P_s^j \mathrm{d}P_u^k, \tag{3.1.6}$$

with $a = 1, 2, \mathrm{c}$, and we model price moves as:

$$\begin{aligned} \mathrm{d}P_t^1 &= \epsilon_t^1(\mathrm{d}N_t^1 + \mathrm{d}N_t^{\mathrm{c}}) \\ \mathrm{d}P_t^2 &= \epsilon_t^2(\mathrm{d}N_t^2 + \mathrm{d}N_t^{\mathrm{c}}), \end{aligned} \tag{3.1.7}$$

where $\epsilon_t^i = \pm 1$, with $\mathbb{E}(\epsilon_t^i) = 0$ and $\mathbb{E}(\epsilon_t^1 \epsilon_t^2) = \rho_t$. While the sign correlation ρ_t could indeed be time dependent, here we will assume that $\rho_t = \rho$ is independent of time. Thus, price moves have both an idiosyncratic part, represented by $\mathrm{d}N^i$, $i \in \{1, 2\}$, and a common part $\mathrm{d}N^{\mathrm{c}}$, which make co-jumps possible[3] (hence the subscript c). In fact, because $\mathrm{d}N_t^{\mathrm{c}} \in \{0, 1\}$, then $(\mathrm{d}N_t^{\mathrm{c}})^2 = \mathrm{d}N_t^{\mathrm{c}}$, one now has:

$$\mathrm{d}P_t^1 \mathrm{d}P_t^2 = \epsilon_t^1 \epsilon_t^2 \mathrm{d}N_t^{\mathrm{c}}. \tag{3.1.8}$$

We can then define the total intensities $(\lambda_{1,t}, \lambda_{2,t})_{t \geq 0}$ of the price jumps $(P_t^1, P_t^2)_{t \geq 0}$, according to the definition of Poisson processes:

$$\begin{cases} \lambda_{1,t} := \mu_{1,t} + \mu_{\mathrm{c},t} \\ \lambda_{2,t} := \mu_{2,t} + \mu_{\mathrm{c},t}, \end{cases} \tag{3.1.9}$$

and notice that the structure of the equations governing the dynamics of intensities is conveniently the same as in the independent case, with the following identifications:

$$\begin{aligned} L_1^i &\leftarrow L_1^i + L_1^{\mathrm{c}} & \qquad L_2^i &\leftarrow L_2^i + L_2^{\mathrm{c}} \\ K_{11}^i &\leftarrow Q_{11}^i + Q_{11}^{\mathrm{c}} & \qquad K_{22}^i &\leftarrow Q_{22}^i + Q_{22}^{\mathrm{c}} \\ K_{12}^i &\leftarrow Q_{12}^i + Q_{12}^{\mathrm{c}} & \qquad \lambda_\infty^i &\leftarrow \mu_\infty^i + \mu_\infty^{\mathrm{c}}. \end{aligned}$$

Mean Intensity

Now correlations between instantaneous price changes are non-zero, i.e. $\mathbb{E}(\mathrm{d}P_s^1 \mathrm{d}P_s^2) = \rho \overline{\mu^{\mathrm{c}}} \mathrm{d}s \neq 0$ (see Appendix A.2), the expression for the mean intensity $\bar{\boldsymbol{\lambda}}$ is modified and reads, for regular kernels:

$$\bar{\boldsymbol{\lambda}} = \Big(\mathbb{I} - \int_0^\infty \mathbb{K}_{\mathrm{d}}(s, s) \mathrm{d}s \Big)^{-1} \Big(\boldsymbol{\lambda_\infty} + \rho \overline{\mu^{\mathrm{c}}} \int_0^\infty \mathbf{K}_\times(s, s) \mathrm{d}s \Big), \tag{3.1.10}$$

with:

$$\overline{\mu^{\mathrm{c}}} = \frac{\mu_\infty^{\mathrm{c}} + \bar{\boldsymbol{\lambda}}^\top \cdot \int_0^\infty \boldsymbol{Q}^{\mathrm{c}}(s, s) \mathrm{d}s}{1 - \rho \int_0^\infty Q_{12}^{\mathrm{c}}(s, s) \mathrm{d}s},$$

and

$$\boldsymbol{Q}^{\mathrm{c}} = \begin{pmatrix} Q_{11}^{\mathrm{c}} \\ Q_{22}^{\mathrm{c}} \end{pmatrix}.$$

In the presence of a singular contribution to the K's, one should again take it into account by substituting all $K(s, s)$ by the corresponding $K_{\mathrm{reg}}(s, s) + \phi(s)$, and similarly for Q's. The stationarity conditions are now that:

1. The spectral radius of the Hawkes kernel $||\Phi_{\mathrm{d}}||$ is strictly less than 1

2. The time-diagonal of the co-jump kernel must be such that:

$$\left| \rho \int_0^\infty Q_{12}^{\mathrm{c}}(s, s) \mathrm{d}s \right| < 1.$$

[3]Note that if we relax the assumption that $\mathrm{d}P$ can only be equal to one tick, one can build a more general model

$$\begin{aligned} \mathrm{d}P_t^1 &= \epsilon_t^1(\mathrm{d}N_t^1 + C_1 \mathrm{d}N_t^c) \\ \mathrm{d}P_t^2 &= \epsilon_t^2(\mathrm{d}N_t^1 + C_2 \mathrm{d}N_t^c) \end{aligned}$$

which would lead to a richer correlation structure. Such a model would mean that co-jumps have a different size than usual price changes.

3.2 Covariance Structure & Yule-Walker Equations

Of course, the feedback kernels L and K cannot be directly observed in data. However, as we now show, they can be computed from covariance functions, which can easily be estimated from empirical data. Here we introduce the covariance structures of a multivariate QHawkes process, thereby establishing the matrix Yule-Walker equations (which link covariance structures and QHawkes kernels).

In order to fully characterise the kernels $\mathbb{K}_{\mathrm{d}}$ and $\mathbf{K}$, we need covariance structures containing information on both the diagonal part of the kernels ($(\tau_1 = \tau_2)$ and their non diagonal part $\tau_1 \neq \tau_2$. When time is discretized on a grid up to lag q, the number of unknowns is, for N assets, $(q + q(q-1)/2) \times N^2$ for $\mathbb{K}_{\mathrm{d}}$ and $q(q-1) \times N(N-1)$ unknowns for $\mathbf{K}$ (with $i = j$ or $i = k$ as considered in this paper and without explicit co-jumps).

Now, Equation (3.2.5) below on two-point correlations can only provide $q \times N(N+1)/2$ equations; three-point correlations are thus also needed to fully determine these kernels (see [12, 70] for the corresponding univariate case).

3.2.1 Two- and Three-point Covariances

The first quantity of interest is the covariance of the activities of the process. For all $\tau \neq 0$:

$$\mathbb{C}_{ij}(\tau) := \mathbb{E}\left(\frac{\mathrm{d}N_t^i}{\mathrm{d}t}\frac{\mathrm{d}N_{t-\tau}^j}{\mathrm{d}t}\right) - \bar{\lambda}_i\bar{\lambda}_j. \tag{3.2.1}$$

As for the univariate QHawkes, $\mathbb{C}$ is even, and its extension to $\tau = 0$ can be worked out following [15]. Thus, for $i = j$, the extension will be the same as in [70], with $\mathbb{C}_{ii}^\star(\tau) := \mathbb{C}_{ii}(\tau) + \delta_{\tau,0}\bar{\lambda}_i$. (Note indeed that $\mathbb{E}((\mathrm{d}N^i)^2) = \mathbb{E}(\mathrm{d}N^i) = \bar{\lambda}_i \mathrm{d}t$, if we consider that events cannot overlap). For $i \neq j$, the extension must account for co-jumps and now reads $\mathbb{C}_{ij}^\star(\tau) := \mathbb{C}_{ij}(\tau) + \delta_{\tau,0}\bar{\mu}_{\mathrm{c}}$. Without co-jumps, one has $\mathbb{C}_{ij}^\star(\tau) := \mathbb{C}_{ij}(\tau)$ for $i \neq j$.

We also define a relevant three-point correlation structure $\mathcal{D}$ as the following tensor:

$$\mathcal{D}_{ijk}(\tau_1, \tau_2) = \mathbb{E}\left[\left(\frac{\mathrm{d}N_t^i}{\mathrm{d}t} - \bar{\lambda}_i\right)\frac{\mathrm{d}P_{t-\tau_1}^j}{\mathrm{d}t}\frac{\mathrm{d}P_{t-\tau_2}^k}{\mathrm{d}t}\right]. \tag{3.2.2}$$

Since price returns are assumed to be of martingales, $\mathcal{D}_{ijk}(\tau_1, \tau_2)$ is only nonzero when $\tau_1 > 0$ and $\tau_2 > 0$. Note that when $\tau_1 = \tau_2$ and $j = k$, one has $\mathcal{D}_{ijj}(\tau, \tau) = \mathbb{C}_{ij}(\tau)$.

In the bivariate case, $\mathcal{D}$ defines again two types of 2×2 matrices: "diagonal" ($j = k$) and "cross" ($j \neq k$). We shall consistently use the following notations to distinguish them:

$$\mathbb{D}_{\mathrm{d}}(\tau_1, \tau_2) := \begin{pmatrix} \mathcal{D}_{111}(\tau_1, \tau_2) & \mathcal{D}_{122}(\tau_1, \tau_2) \\ \mathcal{D}_{211}(\tau_1, \tau_2) & \mathcal{D}_{222}(\tau_1, \tau_2) \end{pmatrix}, \tag{3.2.3}$$

and

$$\mathbb{D}_{\times}(\tau_1, \tau_2) = \begin{pmatrix} \mathcal{D}_{112}(\tau_1, \tau_2) & \mathcal{D}_{121}(\tau_1, \tau_2) \\ \mathcal{D}_{212}(\tau_1, \tau_2) & \mathcal{D}_{221}(\tau_1, \tau_2) \end{pmatrix}. \tag{3.2.4}$$

3.2.2 Two-point Yule-Walker Equations

In order to deduce kernels from empirical correlations, direct relations must be determined. The method we use to find such relations is very similar to that used in Appendix 1 of [70]. Assuming that prices are martingales,[4] and without considering co-jumps, we find the following matrix equation for $\mathbb{C}$:

$$\begin{aligned}
\mathbb{C}(\tau) &= \mathbb{K}_{\mathrm{d}}(\tau)\,\overline{\mathbb{\Lambda}} + \int_{0^+}^{+\infty} \mathbb{K}_{\mathrm{d}}(u,u)\mathbb{C}(\tau - u)\mathrm{d}u \\
&+ 2\int_{0^+}^{+\infty}\int_{u^+}^{+\infty} \mathbb{K}_{\mathrm{d}}(\tau + u, \tau + r)\mathbb{D}_{\mathrm{d}}(u, r)\mathrm{d}r\mathrm{d}u \\
&+ \int_{0^+}^{+\infty}\int_{u^+}^{+\infty} \mathbf{K}(\tau + u, \tau + v)(\mathbf{D})^\top(u, v)\mathrm{d}v\mathrm{d}u \\
&+ \int_{0^+}^{+\infty}\int_{v^+}^{+\infty} \mathbf{K}(\tau + u, \tau + v)(\mathbf{D})^\top(u, v)\mathrm{d}u\mathrm{d}v,
\end{aligned} \tag{3.2.5}$$

[4] This assumption is often violated at high frequencies, and some amendments will need to be introduced when calibrating the model on actual HF data – see Chapter 4.

where $\overline{\mathbb{\lambda}}$ is a 2×2 matrix defined as

$$\overline{\mathbb{\lambda}} := \begin{pmatrix} \bar{\lambda}_1 & 0 \\ 0 & \bar{\lambda}_2 \end{pmatrix}. \tag{3.2.6}$$

and where

$$\mathbf{D} := \begin{pmatrix} \mathcal{D}_{112} \\ \mathcal{D}_{212} \end{pmatrix} \tag{3.2.7}$$

This Yule-Walker equation boils down to Eq. (8) of Ref. [70] in the univariate case.

The Yule-Walker equation for $\mathbb{C}$ accounting for co-jumps can be found in Appendix A.3.

3.2.3 Three-point Yule-Walker Equations

The full three-point Yule-Walker equations for the tensor $\mathcal{D}$ are quite intricate. In order to give a simplified version of these equations we restrict here to the case with no co-jumps, i.e. $\mu^{\mathrm{c}} \equiv 0$. We find that, for $\tau_1 > \tau_2 > 0$:

$$\begin{aligned} \mathbb{D}_{\mathrm{d}}(\tau_1, \tau_2) &= 2\mathbb{K}_{\mathrm{d}}(\tau_1, \tau_2)\overline{\mathbb{C}}_{\mathrm{d}}(\tau_2 - \tau_1) \\ &+ \int_{\tau_1^+}^{+\infty} \mathbb{K}_{\mathrm{d}}(u)\mathbb{D}_{\mathrm{d}}(\tau_1 - u, \tau_2 - u)\,\mathrm{d}u \\ &+ 2\int_{\tau_1^+}^{+\infty} \mathbb{K}_{\mathrm{d}}(u, \tau_1)\mathbb{D}^0_{\mathrm{d}}(u - \tau_1, \tau_2 - \tau_1)\,\mathrm{d}u \\ &+ \int_{\tau_1^+}^{+\infty} \mathbf{K}(\tau_1, u)(\mathbf{D}^1)^\top(u - \tau_1, \tau_2 - \tau_1)\mathrm{d}u \\ &+ \int_{\tau_1^+}^{+\infty} \mathbf{K}(u, \tau_1)(\mathbf{D}^2)^\top(u - \tau_1, \tau_2 - \tau_1)\mathrm{d}u, \end{aligned} \tag{3.2.8}$$

where we have introduced the two following diagonal matrices:

$$\overline{\mathbb{C}}_{\mathrm{d}}(\tau) := \begin{pmatrix} \mathbb{C}_{11}(\tau) + (\bar{\lambda}_1)^2 & 0 \\ 0 & \mathbb{C}_{22}(\tau) + (\bar{\lambda}_2)^2 \end{pmatrix}, \tag{3.2.9}$$

$$\mathbb{D}^0_{\mathrm{d}}(\tau_1, \tau_2) := \begin{pmatrix} \mathcal{D}_{111}(\tau_1, \tau_2) & 0 \\ 0 & \mathcal{D}_{222}(\tau_1, \tau_2) \end{pmatrix}, \tag{3.2.10}$$

as well as the notation $\mathbf{D}$ for the 2-vectors:

$$\mathbf{D}^1 := \begin{pmatrix} \mathcal{D}_{121} \\ 0 \end{pmatrix}, \quad \mathbf{D}^2 := \begin{pmatrix} 0 \\ \mathcal{D}_{212} \end{pmatrix}.$$

For the corresponding Yule-Walker equation for $\mathbb{D}$ see Appendix A.3.

3.2.4 Asymptotic Behavior with Power Law Kernels

An interesting special case for which the Yule-Walker equations can be asymptotically solved is when kernels are decaying as power-laws, as considered in [70]. Of special interest is the relationships between exponents governing the correlation functions and the kernel exponents when $\tau \to +\infty$.

The general analysis is quite cumbersome and is relegated to Appendix A.4.1. In the non-critical case ($n < 1$), the calculations are not particularly difficult and generalise the results of [70] to the multivariate case.

The critical case ($n = 1$), on the other hand, is much more subtle, in particular for the QHawkes processes.[5] In order to treat this case, we have generalised the method introduced by Hawkes in [75], which combines the Yule-Walker equations in the Fourier domain with Liouville's Theorem

[5] For the linear version, one can refer to [12] or to [74] where Bacry, Jaisson and Muzy use convolution to link the Fourier transform of $\mathbb{C}$ with that of $\mathbb{K}$.

to find a direct relationship between exponents. This method is recalled in Appendix A.4.3, from which the value of the exponents in the critical case can be derived, see Tables in Appendix A.5. Some of the exponent values reported in [70] turn out to be incorrect; the correct values can be inferred from our new results. Note that in the critical case the decay of the correlation functions cannot be faster than τ^{-1}, as in the case of linear Hawkes models without ancestors (see the work of Brémaud & Massoulié [66]).

3.3 Power-Law Tails of the Volatility Distribution

Here we investigate the distribution of volatility of a MQHawkes process. We adapt the methodology of Ref. [70] to the multivariate setting, restricting for simplicity to the two assets case. Our main goal is to establish that MQHawkes lead to fat (power-law) tails for the empirical intensity distribution, which translates into fat tails in the distribution of returns [70]. We limit our study to the ZHawkes specification with exponentially decaying kernels, that allow one to construct a tractable continuous time limit.

3.3.1 ZHawkes Model with Exponential Kernels

We neglect the leverage feedback and set $\mathbb{L} = 0$. We also neglect the possible presence of co-jumps, i.e. set $\mu_c \equiv 0$ hereafter. Within the ZHawkes specification, we can rewrite the intensities as follows:

$$\begin{cases} \lambda_1 &= H_1^1 + H_2^1 + (Z_1^1)^2 + (Z_2^1)^2 + Y_2^1 \\ \lambda_2 &= H_1^2 + H_2^2 + (Z_1^2)^2 + (Z_2^2)^2 + Y_1^2 \end{cases}$$

with

$$H_j^i := \int_{-\infty}^t (\phi_d)_{jj}^i (t-s) \mathrm{d}N_s^j$$

and

$$\begin{aligned} Z_j^i &= \int_{-\infty}^t k_{jj}^i (t-s) \mathrm{d}P_s^j \\ Y_j^i &= \left(\int_{-\infty}^t k_{ji}^i (t-s) \mathrm{d}P_s^i \right) \left(\int_{-\infty}^t k_{ij}^i (t-u) \mathrm{d}P_u^j \right). \end{aligned}$$

To keep things tractable, we choose all kernels to be exponentials and consider that only four "features" are important to describe all feedback effects, namely:

$$\begin{cases} h_1(t) = \beta_1 \int_{-\infty}^t e^{-\beta_1 (t-s)} \mathrm{d}N_s^1 \\ h_2(t) = \beta_2 \int_{-\infty}^t e^{-\beta_2 (t-s)} \mathrm{d}N_s^2, \end{cases} \tag{3.3.1}$$

for activity feedback, and

$$\begin{cases} z_1(t) = \omega_1 \int_{-\infty}^t e^{-\omega_1 (t-s)} \mathrm{d}P_s^1 \\ z_2(t) = \omega_2 \int_{-\infty}^t e^{-\omega_2 (t-s)} \mathrm{d}P_s^2, \end{cases} \tag{3.3.2}$$

for trend feedback, with ω_i's and β_i's positive constants. From such features, we construct the quantities H, Z and Y as

$$\begin{aligned} H_1^i &= n_{H,1}^i h_1(t) & H_2^i &= n_{H,2}^i h_2(t) \\ Z_1^i &= a_{Z,1}^i z_1(t) & Z_2^i &= a_{Z,2}^i z_2(t) \\ Y_2^1 &= a_{Z,}^1 z_1(t) z_2(t) & Y_1^2 &= a_{Z,}^2 z_1(t) z_2(t). \end{aligned}$$

Under such hypotheses, the intensities write:

$$\begin{aligned} \lambda_{1,t} =& \lambda_{1,\infty} + n_{H,1}^1 h_1(t) + n_{H,2}^1 h_2(t) + (a_{Z,1}^1 z_1(t))^2 + (a_{Z,2}^1 z_2(t))^2 + a_{Z,}^1 z_2(t) z_1(t) \\ \lambda_{2,t} =& \lambda_{2,\infty} + n_{H,1}^2 h_1(t) + n_{H,2}^2 h_2(t) + (a_{Z,1}^2 z_1(t))^2 + (a_{Z,2}^2 z_2(t))^2 + a_{Z,}^2 z_2(t) z_1(t). \end{aligned} \tag{3.3.3}$$

3.3.2 Endogeneity Ratios

As mentioned earlier, the Hawkes endogeneity ratio n_H is obtained as the spectral radius of the 2×2 matrix of Hawkes coefficients, namely

$$\mathbb{N}_H := \begin{pmatrix} n_{H,1}^1 & n_{H,2}^1 \\ n_{H,1}^2 & n_{H,2}^2 \end{pmatrix}. \tag{3.3.4}$$

The Zumbach endogeneity coefficient n_Z is the spectral radius of the trend feedback kernel. In the monovariate case, one finds $n_Z = \int_0^{+\infty} k^2(s)\mathrm{d}s$. From the features defined above, we now have $k_{jj}^i(t) = a_{Z,j}^i \omega_j e^{-\omega_j t}$. Noting that

$$\omega^2 \int_0^\infty e^{-2\omega s}\,\mathrm{d}s = \frac{\omega}{2},$$

we simply need, in the present case, to find the top eigenvalue of the following 2×2 matrix:

$$\mathbb{N}_Z := \frac{1}{2}\begin{pmatrix} (a_{Z,1}^1)^2\omega_1 & (a_{Z,2}^1)^2\omega_2 \\ (a_{Z,1}^2)^2\omega_1 & (a_{Z,2}^2)^2\omega_2 \end{pmatrix}. \tag{3.3.5}$$

In the simplest case when all coefficients and decay rates are equal, one finds that n_H is equal to any of the coefficients $n_{H,j}^i$ (which are all equal), and $n_Z = a_Z^2\omega$, where $a_Z := a_{Z,j}^i$ (again all equal).

Note that in the general case, one needs to diagonalize the matrix $\mathbb{N}_H + \mathbb{N}_Z$, which leads to a total endogeneity ratio that is in general different from $n_H + n_Z$.

3.3.3 Fokker-Planck Equation

As in [70], we consider the process on a time scale $T > 0$, that shall eventually tend to $+\infty$, and simultaneously take all decay rates $\beta, \omega \to 0$, but in such a way that ωT and βT remain finite. For the Zumbach feedback not to disappear in this limit one needs to simultaneously scale up both $(a_{Z,j}^i)_{i,j\in\{1,2\}}$ as $\omega^{-1/2}$ and $(a_{Z,}^i)_{i\in\{1,2\}}$ as ω^{-1}. In this scaling regime, one can establish the following Fokker-Planck equation for the time dependent probability density p_t of (h_1, h_2, z_1, z_2) (see Appendix A.6):

$$\begin{aligned} \partial_t p_t =& \beta_1 \partial_{h_1}\big((h_1 - \lambda_{1,t})p_t\big) + \beta_2 \partial_{h_2}\big((h_2 - \lambda_{2,t})p_t\big) \\ &+ \omega_1 \partial_{z_1}\big(z_1 p_t\big) + \omega_2 \partial_{z_2}\big(z_2 p_t\big) \\ &+ \frac{\omega_1^2}{2}\partial_{z_1 z_1}^2\big(\lambda_{1,t} p_t\big) + \frac{\omega_2^2}{2}\partial_{z_2 z_2}^2\big(\lambda_{2,t} p_t\big), \end{aligned} \tag{3.3.6}$$

where p_t is a shorthand for $p_t(h_1, h_2, z_1, z_2)$, $\lambda_{1,t}$ and $\lambda_{2,t}$ are given by Eq. (3.3.3), and where we have disregarded co-jumps and direct correlations between the returns of asset 1 and asset 2, meaning that $\mathbb{E}(\mathrm{d}P^1\mathrm{d}P^2) = 0$. The inclusion of such correlations can be considered and adds further cross-derivative terms $\partial_{z_1 z_2}^2$ in the Fokker-Planck equation.

Solving for the stationary distribution p_∞ of the Fokker Planck equation allows one to determine the tail behaviour of the distribution of intensities (which translate into volatilities since $\mathbb{E}[(\mathrm{d}P_t)^2] = \lambda_t \mathrm{d}t$). In the monovariate case, Ref. [70] established that p_∞ decays as a power-law, with an exponent α that depends on both n_H and n_Z. The general expression for α is however not available in closed form, although asymptotic results in various regimes could be worked out, in particular when $n_H \to 0$. The most important conclusion is that $\alpha \to \infty$ when $n_Z \to 0$, i.e. power-law tails disappear in the absence of a quadratic Zumbach coupling. Interestingly, the coupling between the Hawkes feedback with $n_H \sim 1$ and even a small Zumbach effect ($n_Z \ll 1$) was shown to generate an exponent compatible with empirical data.

3.3.4 ZHawkes without Hawkes ($n_H = 0$)

To make further analytical progress, we now consider the case in which the Hawkes coupling is absent ($n_H = 0$), that is when h and z decouple, leading to a tractable two-dimensional Fokker-Planck for z_1 and z_2. In the stationary regime, the Fokker Planck equation (3.3.6) becomes:

$$\begin{aligned} 0 =& \omega_1 \partial_{z_1}\big(z_1 p_\infty(z_1, z_2)\big) + \omega_2 \partial_{z_2}\big(z_2 p_\infty(z_1, z_2)\big) \\ &+ \frac{\omega_1^2}{2}\partial_{z_1 z_1}^2\big(\lambda_1 p_\infty(z_1, z_2)\big) + \frac{\omega_2^2}{2}\partial_{z_2 z_2}^2\big(\lambda_2 p_\infty(z_1, z_2)\big). \end{aligned} \tag{3.3.7}$$

This equation describes the stationary measure of the stochastic path of the bivariate process $(Z_1, Z_2)_t$ defined as:

$$\begin{aligned} \mathrm{d}Z_1 &= -\omega_1 Z_1 \mathrm{d}t + \omega_1 \sqrt{\lambda_{1,t}} \mathrm{d}W_t^1 \\ \mathrm{d}Z_2 &= -\omega_2 Z_2 \mathrm{d}t + \omega_2 \sqrt{\lambda_{2,t}} \mathrm{d}W_t^2, \end{aligned}$$

with

$$\begin{aligned} \lambda_{1,t} &= \lambda_{1,\infty} + a^1_{Z,} Z_2 Z_1 + (a^1_{Z,1} Z_1)^2 + (a^1_{Z,2} Z_2)^2 \\ \lambda_{2,t} &= \lambda_{2,\infty} + a^2_{Z,} Z_2 Z_1 + (a^2_{Z,2} Z_2)^2 + (a^2_{Z,1} Z_1)^2. \end{aligned} \tag{3.3.8}$$

These equations allow one to simulate paths $(Z_1, Z_2)_t$ numerically, from which an empirical determination of $p_\infty(z_1, z_2)$ can be confronted to our analytical solution of Eq. (3.3.7). Note that the coefficients $(a^i_{Z,j})_{i,j\in\{1,2\}}$ and $(a^i_{Z,})_{i\in\{1,2\}}$ must satisfy the following inequalities for $\lambda_{1,t}$ and $\lambda_{2,t}$ to remain positive at all times:

$$4(a^i_{Z,1} a^i_{Z,2})^2 \geq (a^i_{Z,})^2, \qquad i = 1, 2.$$

How can one determine the tail exponent for such a two dimensional process? We first introduce polar coordinates (r, θ), such that $z_1 = r\cos\theta$ and $z_2 = r\sin\theta$. We then surmise that when $r^2 \gg \max(\lambda_{1,\infty}, \lambda_{2,\infty})$ the stationary distribution $p_\infty(r, \theta)$ factorizes into an angular component $F(\theta)$ and a power-law contribution, to wit: $p_\infty(r, \theta) \underset{r\to\infty}{\approx} F(\theta) r^{-\alpha}$, where α is the tail exponent. Note that α should be strictly larger than 2 for p_∞ to be normalisable, and strictly larger than 3 for the mean intensity to be non divergent. Injecting our factorized guess into Eq. (3.3.7) and taking the limit $r^2 \gg \lambda^\infty$, we find a second order ODE on the function F, where α appears as a parameter (see Appendix A.7, Eq. (A.7.1)).

The value of α is selected by the analogue of energy quantification in quantum mechanics: only for some special values of α can one find a solution F of the above ODE that satisfies the correct boundary conditions compatible with symmetries of the problem. Clearly, F must be everywhere non-negative and, because of the symmetry $z_1 \to -z_1$ and $z_2 \to -z_2$, one must have $F(\theta + \pi) = F(\theta)$ (see below for explicit examples). In principle, there can be more than one value of α that allows one to find an acceptable solution F. This is the analogue of the energy spectrum in quantum mechanics. The value of α that governs the tail behaviour of $p_\infty(r, \theta)$ is then the *smallest* of all such acceptable values. Once the asymptotic tail behaviour of $p_\infty(r, \theta)$ is known, it is easy to derive the tail behaviour of the marginals $p_\infty(z_1)$ and $p_\infty(z_2)$, which both behave as $z^{-1-\mu}$ with $\mu = \alpha - 2$. The volatility distribution then has the same tail behaviour.

In order to illustrate this general procedure on a simple example, we focus in the following on the case where $a^j_{Z,i} = \sqrt{2n_Z/\omega_i}$, $\forall i, j = 1, 2$, such that the two eigenvalues of $\mathbb{N}_Z$ are equal to n_Z (the Zumbach endogeneity ratio) and 0. We also set $a^i_{Z,} = 2\gamma n_Z / \sqrt{\omega_1 \omega_2}$, where γ is an arbitrary coefficient $\in (-2, 2)$ (such that λ_1 and λ_2 are always positive). When $\omega_1 = \omega_2$, Eq. (A.7.1) considerably simplifies and reads:

$$\left[(\alpha - 2)(\alpha - \alpha_0) + (\alpha - 2)^2 \gamma \cos(\theta)\sin(\theta)\right] F(\theta) + \left[(1 + \gamma\cos(\theta)\sin(\theta))F(\theta)\right]'' = 0, \tag{3.3.9}$$

where we have defined

$$\alpha_0 := 2 + \frac{1}{n_Z}.$$

Note that for a given value of α this equation is invariant under the simultaneous change $\gamma \to -\gamma$, $\theta \to -\theta$. Hence the value of α can only depend on $|\gamma|$.

The isotropic case $\gamma = 0$

When cross terms $a^i_{Z,}$ are absent and $\omega_1 = \omega_2$, the problem becomes isotropic in the sense that the dynamics of $r^2 = z_1^2 + z_2^2$ decouple from that of θ. The problem then boils down to the univariate ZHawkes model without Hawkes coupling, for which the value of α is known, and given by α_0.

Furthermore, the evolution of θ is that of a free Brownian motion on the unit circle, leading to a uniform distribution $F(\theta) = F_0$, which is indeed a solution of Eq. (3.3.9) in this case. Note that other periodic solutions exist whenever

$$(\alpha - 2)(\alpha - \alpha_0) = 4\ell^2, \qquad \ell = 0, 1, 2, \ldots$$

but lead to larger values of α when $\ell > 0$.

The case $\gamma \neq 0$

In order to make progress, we posit that α and F can be expanded as power series of γ, namely

$$\alpha = \alpha_0 + \alpha_1\gamma + \alpha_2\gamma^2 + \dots$$
$$F(\theta) = F_0 + F_1(\theta)\gamma + F_2(\theta)\gamma^2 + \dots$$

where $\alpha_0 = 2 + 1/n_Z$ is the solution for $\gamma = 0$ and F_0 is the constant solution found above. The coefficient α_1 must be zero for symmetry reasons.

Inserting this expansion in Eq. (3.3.9) and imposing that all $F_n(\cdot)$ remain π-periodic, the identification of terms of order γ^n finally leads to

$$\alpha = \alpha_0 + \frac{\gamma^2}{32}\left(\frac{4}{n_Z} - \frac{1}{n_Z^3}\right) + O(\gamma^4). \quad (3.3.10)$$

Figure 3.2 displays the numerical values of α as a function of γ for $n_Z = 0.4$ and $n_Z = 0.6$ in with the corresponding theoretical parabolas, see Eq. (3.3.10). Note that the γ^2 correction changes sign when $n_Z = 1/2$.

In this case ($n_Z = 1/2$), finding an exact solution of the associated Schrodinger solution is possible [76], and leads to $\alpha = 4$ for all values of γ.[6] The corresponding solution for F is also known in that case and is a constant independent of θ, as can be directly checked from Eq. (3.3.9). An expansion around the special point $n_Z = 1/2$ can in fact be performed and leads to first order to

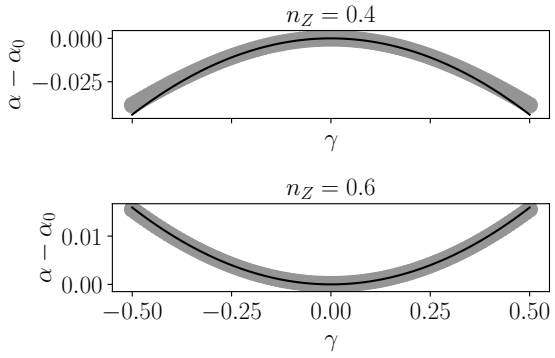


Figure 3.2: Plot of $\alpha - \alpha_0$ as function of γ for $n_Z = 0.4$ (above) and $n_Z = 0.6$ (below). The α are found numerically by looking for the smallest α for a π-periodic and positive solution F. The solid black line represents the theoretical prediction, Eq. (3.3.10), that is $\alpha \approx \alpha_0 - 0.18\gamma^2$ for $n_Z = 0.4$ and $\alpha \approx \alpha_0 + 0.06\gamma^2$ for $n_Z = 0.6$.

$$\alpha = 4 + \frac{16}{4+\gamma^2}\left(\frac{1}{2} - n_Z\right) + o\left(\frac{1}{2} - n_Z\right).$$

Finally, note that the condition $\alpha > 3$ (ensuring that the mean activity is finite) reads, to first order in γ:

$$n_Z < n^\star \approx 1 + \frac{\gamma^2}{8} + O(\gamma^4). \quad (3.3.11)$$

The isotropic case $\gamma = 0$ boils down to the univariate ZHawkes model, for which $n^\star = 1$, see [63].

The case $n_Z \to +\infty$

Equation (3.3.9) can be easily analysed when $n_Z \to +\infty$, in which case α tends to the smallest possible value, 2, corresponding to a maximally "fat" distribution. The only periodic solution to Eq. (3.3.9) in that limit is

$$F_\infty(\theta) = \frac{C}{2 + \gamma\sin(2\theta)}, \quad (3.3.12)$$

where C is a constant. When n_Z is very large but not infinite, we posit that the solution writes

$$F(\theta) = F_\infty(\theta) + \frac{1}{n_Z^2}G(\theta) + O(\frac{1}{n_Z^4}), \quad (3.3.13)$$

together with $\alpha = 2 + \zeta/n_Z$, with $G(\cdot)$ and ζ to be determined. Inserting Eq. (3.3.13) in the ODE on F Eq. (3.3.9), one obtains the ODE for G:

$$[(2 + \gamma\sin(2\theta))G(\theta)]'' = -C\frac{2\zeta(\zeta-1) + \gamma\zeta^2\sin(2\theta)}{2 + \gamma\sin(2\theta)}. \quad (3.3.14)$$

[6] The other exact solutions found in [76] unfortunately correspond to sub-dominant, larger values of α.

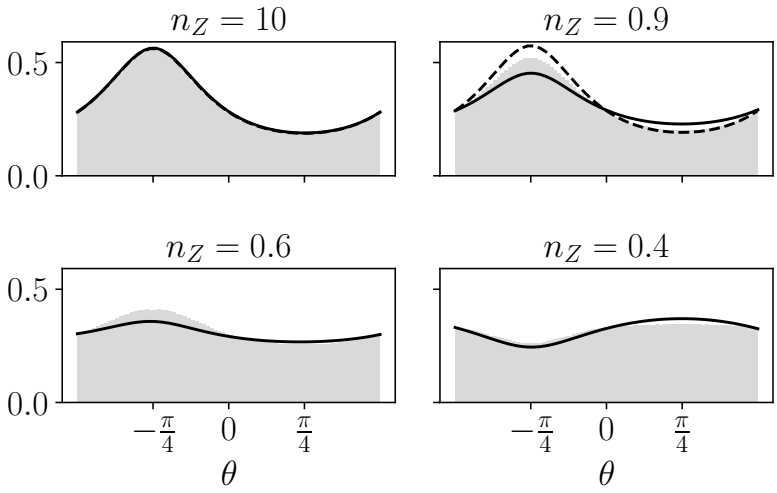


Figure 3.4: Comparison between the numerical solution F of Eq. (3.3.9) (black lines) and the results of simulating Eq. (3.3.8) (grey histogram) for several values of n_Z, and $\gamma = 1$. The solid line shows the numerical solution F found by looking for a solution couple (F, α) of a positive π-periodic solution F of Eq. (3.3.9) and a value α. The the dashed line for $n_Z = 10$ and $n_Z = 0.9$ shows the solution F_∞ for $\alpha = 2$ ($n_Z \to +\infty$). The upper right figure, obtained for $n_Z = 10$, matches very well the solution obtained when $n_Z \to +\infty$ since corrections are $O(n_Z^{-2})$. Note that when $n_Z = 1/2$, the exact solution $F(\theta)$ is independent of θ for all values of γ.

Imposing $G(\cdot)$ to be π-periodic, that is imposing that the right hand side of Eq. (3.3.14) integrates to zero between $\theta = 0$ and $\theta = \pi$, sets the value of ζ:

$$\zeta = \left(1 - \frac{\gamma^2}{4}\right)^{-\frac{1}{2}}, \tag{3.3.15}$$

recovering $\zeta = 1$ when $\gamma = 0$ and the small γ expansion result above, see Eq. (3.3.10). The function $G(\cdot)$ is plotted in Fig. 3.3 for $\gamma = 1$.

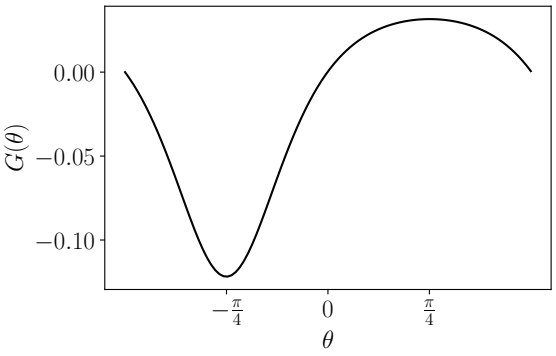


Figure 3.3: Plot of $G(\theta)$ for $\gamma = 1$, $C = 0.6$, see Eq. (3.3.14).

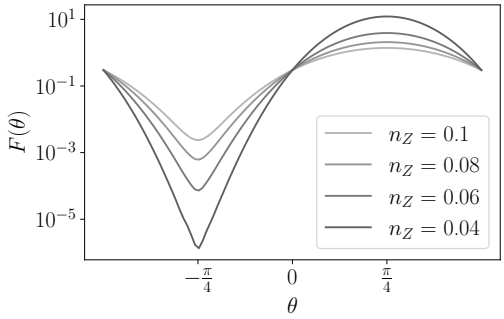
When comparing the solution above with the histogram of simulated θ for $n_Z = 10$ and $\gamma = 1$, we find an excellent match with F_∞, without need of any correction, see Fig. 3.4. This is expected since the correction term $G(\theta)/n_Z^2$ is of the order of 1% in that case. When n_Z decreases, corrections become more pronounced. The numerically computed F is in good agreement with the angular distribution obtained from a direct numerical simulation of the two dimensional stochastic process (Eq. (3.3.8)).

The case $n_Z \to 0$

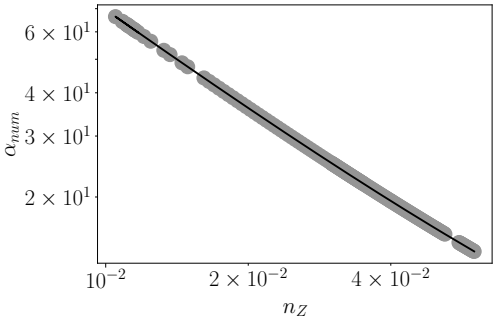
When n_Z tends to zero, we expect that the exponent α of the power-law tail diverges. Looking again for α of the form $\alpha = 2 + \zeta/n_Z$, we find that Eq. (3.3.9) reads:

$$\big(2\zeta(\zeta - 1) + \zeta^2\gamma\sin(2\theta)\big)F(\theta) + n_Z^2\left[(2 + \gamma\sin(2\theta))F(\theta)\right]'' = 0. \tag{3.3.16}$$

When $n_Z \to 0$, this equation looks self-contradictory because the remaining term can only be zero if $F(\theta) = 0$. However, the second derivative term is a singular perturbation, so it must be treated with care. The idea is to look for a solution F which is zero nearly everywhere, except very close to some special values of θ where the second derivative diverges. It turns out that all the action takes place close to $\theta = \pi/4$ when $\gamma > 0$ and $\theta = -\pi/4$ when $\gamma < 0$.



(a) Solution F of Eq. (3.3.9) when $n_Z \leq 0.1$. Note the logarithm scale on the y-axis, such that the Gaussian solution for $\Psi(u)$, Eq. (3.3.17), is a parabola around $\theta = \pi/4$.



(b) Tail exponent α as a function of n_Z for $n_Z \leq 0.06$. Circles: numerical determination of α such that F is a π-periodic solution of equation (3.3.9). Solid black line: theoretical value $\alpha = 2 + \zeta/n_Z$ with ζ given by Eq. (3.3.18)

Figure 3.5: Results when $n_Z \to 0$

Choosing $\gamma > 0$, $\theta = \pi/4 + u\sqrt{n_Z}$, and $n_Z \to 0$, Eq. (3.3.16) becomes the Schrodinger equation of a harmonic oscillator, up to terms $O(n_Z^2)$:

$$\left[-(2+\gamma)\frac{\mathrm{d}^2\Psi}{\mathrm{d}u^2} + 2\zeta^2\gamma u^2\Psi\right] = \frac{2\zeta(\zeta-1)+\zeta^2\gamma}{n_Z}\Psi,$$

with

$$\Psi(u) := F\left(\frac{\pi}{4} + u\sqrt{n_Z}\right).$$

The smallest α solution (or "ground state") is

$$\Psi(u) = C' e^{-\varkappa u^2} + O(n_Z), \qquad \varkappa = \zeta\sqrt{\frac{\gamma}{4+2\gamma}}, \tag{3.3.17}$$

where C' is another constant, together with

$$\zeta = \frac{2}{2+\gamma} + \sqrt{\frac{2\gamma}{2+\gamma}}n_Z + O(n_Z^2). \tag{3.3.18}$$

This solution is only accurate in a region of width $\sim \sqrt{n_Z}$ around $\pi/4$, beyond which it quickly goes to zero. This is in perfect agreement with the numerical solution of Eq. (3.3.9) for small n_Z, shown in Fig. 3.5a. The parabolic shape in a semi-log plot around $\theta = \pi/4$ agrees with the predictions of Eq. (3.3.17). Moreover, the value of $\alpha = 2 + \zeta/n_Z$, with ζ given in Eq. (3.3.18), also perfectly matches the numerical values reported in Fig. 3.5b. Note that for $\gamma < 0$, the same results hold with $|\gamma|$ replacing γ in the above equations.

3.3.5 The General Case

In the previous section, we have shown how to compute the power-law tail exponent in the case where only the quadratic "ZHawkes" kernel is present. We also restricted to simple cases where the frequencies ω and coupling constants a_Z are symmetric. Although analytically more challenging, the method outlined above can be implemented more generally, and amounts to solving a problem akin to the quantification condition for the Schrödinger equation. Similarly to the monovariate case, the volatility distribution develops a power-law tail for all values of the Hawkes feedback n_H, as long as some amount of Zumbach feedback n_Z is present. When $n_H \neq 0$, the equation setting the tail in the bivariate case is a three dimensional partial derivative equation generalizing the equation written in the appendix of Ref. [70]. We expect that even a tiny amount of Zumbach feedback coupled to the standard Hawkes effect brings the exponent α into the empirical range, as found in the monovariate case [70].

Conclusion

Let us summarise what we have achieved in this study. Building on the work of Blanc *et al.* [70], we extended the Quadratic Hawkes model to a multivariate framework (MQHawkes). We emphasized that in the multivariate case, both idiosyncratic and common jumps must be considered in the general case, leading to a quite complex framework that we only detailed in the bivariate case.

We defined the endogeneity ratio of MQHawkes, as well as the associated conditions for the process to be stationary. Within the ZHawkes approximation – where quadratic kernels write as a sum of a time diagonal component, reproducing a linear Hawkes feedback, and a rank one component – we gave a deeper understanding of the roles of the different feedback terms in the stationarity condition. We found that, in particular, the spectral radius of the Hawkes component needs to be strictly inferior to 1, as for the 1D case. The rank one component contains both exciting and inhibiting realisations and is not involved in the condition, although such a contribution can lead to a divergence of the average intensity of the process, see [63].

We further defined the covariance structures for MQHawkes processed and established the associated Yule-Walker equations. The latter allow one to fully determine the quadratic kernels from data, and thereby pave the way for their empirical calibration.

Finally, we studied the volatility distribution of a 2D MQHawkes process. Restricting our study to ZHawkes without Hawkes ($n_H = 0$), with exponential kernels and in symmetric cases, we were able to characterize exactly the tail of the joint probability density function of the ZHawkes intensity terms. We found that it displays a power law behavior, in line with the observed fat tails of financial returns. Note that, interestingly, the coupling between assets imposes that the exponent α is the same for all assets – a mechanism that may explain the apparent universality of the power-law tail observed in financial markets.

In the forthcoming Chapter (Chapter 4), we shall calibrate the model on empirical data. On the theoretical side, while expected to be quite heavy, it would be interesting to further develop the analysis in the presence of co-jumps and correlations. Note that for the sake of clarity most of our explicit expressions are given in the 2D case. Expanding them further in the N-dimensional case will be necessary in order to calibrate the model on a large number of assets, say a pool of stocks.

Also, in the present analysis of the volatility distribution, we restricted to symmetric coefficients and exponential kernels. Confirming that our conclusions are qualitatively robust against changes in symmetry and kernel functionnals, and studying its quantitative implications would also be of interest. Furthermore, we have focused on the specific case of the quadratic Hawkes model because of its ability to reproduce multiple financial stylised facts, but it would also be interesting to consider the multivariate generalisation of other non-linear Hawkes models, following the interesting insights and methods presented in Ref. [69].

Take Home Message

- Modeling multivariate volatility requires accounting for cross-time-reversal asymmetry effects and co-jumps occurrences, in addition to the characteristics of univariate financial time series.
- The QHawkes model is a suitable choice for modeling multivariate volatility as it inherently generates the characteristics of univariate financial time series. Additionally, its ability to explicitly delineate the influence of the past on future activity is particularly relevant for understanding cross-time-reversal asymmetry effects.
- To account for the frequent occurrence of co-jumps, two frameworks employing the QHawkes model can be considered:
  - Independent QHawkes processes:

$$\begin{cases} \mathrm{d}P_t^a = & \pm \mathrm{d}N_t^a \\ \mathrm{d}P_t^b = & \pm \mathrm{d}N_t^b \end{cases},$$

  - Bivariate QHawkes processes – where price dynamics are driven by both idiosyncratic components and a common component, modeled using QHawkes processes:

$$\begin{cases} \mathrm{d}P_t^a = & \epsilon_t^a(\mathrm{d}N_t^a + \mathrm{d}N_t^c) \\ \mathrm{d}P_t^b = & \epsilon_t^b(\mathrm{d}N_t^b + \mathrm{d}N_t^c) \end{cases}, \text{ where } \epsilon^{a,b} = \pm 1.$$

  In both cases, intensity processes governing the volatilities of P^a and P^b are influenced by feedback from the activities of all QHawkes processes.
- Multivariate QHawkes processes seem to consistently extend the relevant features of the univariate QHawkes:
  - Fat tails of the returns distribution
  - Volatility clustering
  - (Cross)-time-reversal asymmetry

Chapter 4

Calibration of Quadratic Hawkes on Empirical Data

> Modern portfolio theory poses a danger to those who believe in it too strongly and is a powerful challenge for the theoretician.
>
> Benoît Mandelbrot, *A Multifractal Walk down Wall Street*

This chapter builds upon the work of the previous chapter by attempting to calibrate the MQHawkes model on real data. The research presented herein is the result of collaboration with Jean-Philippe Bouchaud. We extend our gratitude to Michael Benzaquen and Marcello Rambaldi for insightful discussions.

Introduction

The increasing amount of high frequency data in financial markets has made it possible to study and calibrate microstructure models. An increasingly renowned one is the Hawkes process. Initially introduced for modeling seismic activity, its adaptability and interpretability have rendered it highly appealing for financial data, as it distinctly emphasizes the feedback loop underlying the stylized facts that characterize financial prices. In the previous chapter (Chapter 3), we introduced a multivariate version of the quadratic Hawkes. The present chapter aims at introducing a non-parametric calibration method of this multivariate QHawkes (MQHawkes) on empirical data.

Whether using binned events in a regular time grid or successive time events, many methods exist to calibrate Hawkes processes and their extensions on empirical data. A good review is presented in Appendix C of [13]. Methods fall into two categories: parametric and non-parametric.

Among the frequently employed parametric approaches, the maximum likelihood [77, 78] and Expectation Maximisation (EM) method [79] stand out as prominent examples.

Non-parametric methods offer the advantage of agnostic kernel forms, as they do not impose specific shapes. The primary non-parametric approaches include the Expectation Maximisation-based method [80, 81], the minimisation of contrast function [82, 83], and the method of moments or Wiener-Hopf-based method [74, 84, 85, 14, 86]. Additionally, the *method of cumulants*, a non parametric method [87, 88], allows to assess the average endogeneity of the process without having to fully estimate the feedback kernels.

In this study, to remain agnostic regarding the kernel's shape, we implement the method of moments on binned data for several pairs of assets. The outline of the chapter is as follows. The first section describes the data and the pre-processing methods we apply to it. The second section presents the key ingredients for the calibration: the approximation in discrete time, the covariance structures and the calibration method. Finally, the last sections are dedicated to our calibration results on empirical data.

For the sake of simplicity and numerical stability, we work with a 2-dimensions framework as it is enough to characterise all the feedback contributions of Chapter 3's model. Along the chapter, we use the pair E-mini VS TBOND as an illustrative example for the method. Calibration on synthetic data is also demonstrated in Appendix B.8. We would like to point out that, although a lot of

work has already been done and we are well advanced, the calibration proposed here is not the final version, as there are still some points to be reworked, as discussed in Section B.8.2 of Appendix B.8.

Notations stay consistent with the ones introduced previously and are presented in the notations table.

4.1 Data

For all of the assets pairs considered hereafter, we applied the following methodology prior to calibration, unless stated otherwise.

4.1.1 Data description

For each studied asset i, we collect the opening (o), closing (c), highest (h) and lowest (l) prices of each 1-minute interval on every trading day from 2013 to 2023. We only consider the bins between 10 am and 3 pm to exclude the high activity at market openings (as participants react to overnight news) and at market closings (as participants close positions before the market session ends).

From these price time series, we compute the two quantities needed for the calibration for each 1-minute bin: the log-returns $r^i_{t,d}$ and the volatility $\sigma^i_{t,d}$ as follows:

$$r^i_{t,d} = \log\left(\frac{\mathrm{c}^i_{t,d}}{\mathrm{o}^i_{t,d}}\right) \tag{4.1.1}$$

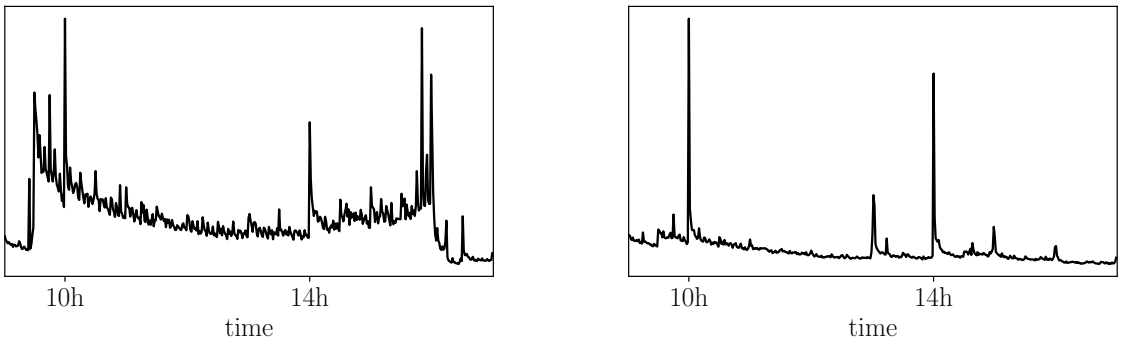
$$\sigma^i_{t,d} = \frac{1}{3}\frac{\mathrm{h}^i_{t,d} - \mathrm{l}^i_{t,d}}{\mathrm{o}^i_{t,d}} + \frac{2}{3}\frac{|\mathrm{c}^i_{t,d} - \mathrm{o}^i_{t,d}|}{\mathrm{o}^i_{t,d}} \tag{4.1.2}$$

where d accounts for the day and t for the 1-minute bin within day d.

4.1.2 Data pre-processing

After collecting the log-returns $(r^i_{t,d})_{t,d,i}$ and the volatility $(\sigma^i_{t,d})_{t,d,i}$ time series, we proceed with two steps of data pre-processing: the normalisation and the "martingalisation". The first step removes the intraday pattern of the volatility and standardises the volatility across the entire studied period. The "martingalisation" process allows us to transform the returns to martingales, meaning that $\mathbb{E}(r_t r_{t-\tau}) = 0$. We detail the two steps subsequently.

Normalisation



(a) E-mini intraday squared volatility profile

(b) TBOND intraday squared volatility profile

Figure 4.1: Average intraday profiles of the squared volatility Bachelier estimates, as defined by Equation (4.1.2), for a period spanning from 2013 to 2023 for the E-mini (left panel) and the TBOND (right panel).

The objectives of these normalization procedures are twofold: first, to make the time series stationary over time, and second, to standardise the activity throughout the day.

To achieve stationarity, each intraday value, that is, the value of each 1-minute bin, is divided by its average value over the past 100 days as described in Equations (4.1.3) and (4.1.4):

$$r^i_{t,d} \leftarrow \frac{r^i_{t,d}}{\sqrt{\frac{1}{100}\sum_{k=d-100}^{d}(r^i_{t,k})^2}}, \tag{4.1.3}$$

$$\sigma^i_{t,d} \leftarrow \frac{\sigma^i_{t,d}}{\sqrt{\frac{1}{100}\sum_{k=d-100}^{d}(\sigma^i_{t,k})^2}}. \tag{4.1.4}$$

To normalize the activity throughout the day, each intraday value, that is again, the value of each 1-minute bin, is divided by its average value over all days, as described in Equations (4.1.5) and (4.1.6). Without this normalization, the daily activity profile exhibits a U-shape or even a J-shape, as it is the case for the E-mini and the TBOND as demonstrated in Figure 4.1 (see also [23]).

$$\sigma^i_{t,d} \leftarrow \frac{\sigma^i_{t,d}}{\sqrt{\langle(\sigma^i_{t,d})^2\rangle_t}} \tag{4.1.5}$$

$$r^i_{t,d} \leftarrow \frac{r^i_{t,d}}{\langle(\sigma^i_{t,d})^2\rangle_t} \tag{4.1.6}$$

"Martingalisation"

Upon examining empirical data, it appears that at the 1-minute time scale, returns are not uncorrelated. Specifically the correlation between r_t and r_{t-1} is revealed to be non zero. However, in the QHawkes framework presented in Chapter 2, price returns are assumed to be martingales, and other properties stem from this assumption. Therefore, we perform a "martingalisation" of the returns time series, meaning that the component of the returns at time t that is predictable from time $t-1$ is removed to isolate the return's "surprise" component, as follows (see also [14] and Appendix B.4 for more details):

$$r^i_{t,d} \leftarrow r^i_{t,d} - \mathbb{E}(r^i_{t,d}/\mathscr{F}_{t-1,d}).$$

The data pre-processing is completed by centering and normalising the returns by their standard deviation (which accounts for considering $\psi = 1$ in Chapter 3).

The subsequent section describes the calibration method to be applied on the returns and volatility time series.

4.2 Calibration method

This section outlines the calibration method we implemented. For the sake of clarity, we present here the main elements and Appendices B.5, B.7 and B.6 provide more comprehensive details.

To stay agnostic on the shape of the kernels, we chose to implement a method of moments. This method relies on linear relationships between the kernels we want to recover and quantities that we can compute from data, that are covariance structures. These linear relationships, the so-called Yule-Walker equations, were introduced in Equations (3.2.5) and (3.2.8) for the MQHawkes model as defined in Chapter 3. The discretization of data imposes these equations to be adapted, however they are still a key ingredient to the calibration. A second key ingredient is the set of covariance structures, indeed they compose the linear system required to recover the kernels. In Chapter 3, we introduced $\mathbb{C}$ and $\mathcal{D}$ in Equations (3.2.1) and (3.2.2). As for the Yule-Walker system of equations, the covariance structures need to be adapted to account for the discretization of data. Since it obviously has a lot of consequences on the calibration, this chapter starts by discussing the data binning and the modifications that follow.

4.2.1 Approximation for discrete data – toward the MQGARCH model

The advantage of the method of moments is that it is non-parametric, unlike the Maximum Likelihood method for instance, hence it allows to remain agnostic on the kernel's shape. The method of moments relies on the computation of covariance structures, such as those introduced Equations (3.2.1) and (3.2.2) in Chapter 3. The estimation of these covariance structures can be done directly using the times of events, that are the times of price changes in the QHawkes model, which necessitates working with high frequency data. However, high frequency data is sometimes difficult to obtain or heavily impacted by microstructure noise [89]. Furthermore, high frequency price changes strongly reflect the bid-ask bounce effect (strong mean reversion), which is not of interest in this study. Therefore, we choose to work with aggregated data.

Specifically, relying on the work of Blanc *et al.* [70] on the univariate QHawkes, the QHawkes intensity λ_t can be approximated by the squared volatility as follows:

$$\frac{\left(\sigma_t^{(\mathrm{d}t)}\right)^2}{\mathrm{d}t} \xrightarrow[\mathrm{d}t\to 0^+]{} \lambda_t, \tag{4.2.1}$$

where $\sigma_t^{(\mathrm{d}t)}$ is a volatility estimate over the time bin of size $\mathrm{d}t$.

Thus, the Hawkes continuous point process $\mathrm{d}N$ can be approximated by the discrete volatility process $\sigma^{(\mathrm{d}t)}$ over a time bin of size $\mathrm{d}t$. This approximation critically depends on the choice of the time bin size $\mathrm{d}t$, which then significantly influences the calibration of the underlying QHawkes model.

Indeed within this limit, the QHawkes relation between the event and the price change, $\mathrm{d}P_t = \pm \mathrm{d}N_t$, might no longer valid, particularly if $\mathrm{d}t \gg 1/\bar{\lambda}$. Consequently, some of the terms in Yule-Walker equations would need to be modified and the ZHawkes model would no longer be properly defined. Indeed, if we decompose the quadratic kernel into a diagonal component and a regular one ($\mathbb{K}_{\mathrm{d}}(s,s) = \mathbb{K}_{\mathrm{d,reg.}}(s,s) + \mathbb{\Phi}_{\mathrm{d}}(s)$), the diagonal component would no longer replicate the linear Hawkes feedback, but includes a feedback loop on $(\mathrm{d}P)^2$.

Additionally, the choice of $\mathrm{d}t$ highly influences the estimation of the covariance values that determine the Yule-Walker equations. For a small $\mathrm{d}t$, say $\mathrm{d}t \ll 1/\bar{\lambda}$, consecutive events are rare potentially leading to an underestimation of correlations. Conversely, with a large $\mathrm{d}t$ ($\mathrm{d}t \gg 1/\bar{\lambda}$), numerous consecutive events occur, potentially resulting in an overestimation of correlations. Appendix B.8-Section B.8.1 discusses this aspect in details (see also Appendix B.2 which discusses data binning).

Following [70], we alleviate these difficulties by approximating the MQHawkes model by a MQGARCH model. Relying on the QGARCH framework introduced in [12, 90], we defined the MQGARCH model for 2 assets, A and B, which defines the squared volatility of asset j, σ_j^2, as follows (note that thereafter, we swap (A, B) and (1, 2) whichever makes the writing clearer):

$$\begin{aligned}
\sigma_{j,t}^2 =& \sigma_{j,\infty}^2 \\
&+ \sum_{i=1}^{2}\sum_{\tau=1}^{+\infty} L_i^j(\tau) r_{i,t-\tau} \\
&+ \sum_{i=1}^{2}\sum_{\tau_1=1}^{+\infty}\sum_{\tau_2=1}^{+\infty} K_i^j(\tau_1,\tau_2) r_{i,t-\tau_1} r_{i,t-\tau_2} \\
&+ \sum_{\tau_1=1}^{+\infty}\sum_{\tau_2=1}^{+\infty} K_\times^j(\tau_1,\tau_2) r_{A,t-\tau_1} r_{B,t-\tau_2}.
\end{aligned} \tag{4.2.2}$$

For better interpretability, the model can be rewritten by decomposing the kernels K and $K_\times$ into their time-diagonal and off-time-diagonal components as follows

$$\begin{aligned}\sigma_{j,t}^2 =&\sigma_{j,\infty}^2\\ &+\sum_{i=1}^{2}\sum_{\tau=1}^{+\infty} L_i^j(\tau) r_{i,t-\tau}\\ &+\sum_{i=1}^{2}\sum_{\tau=1}^{+\infty} \phi_i^j(\tau) r_{i,t-\tau}^2 + 2\sum_{i=1}^{2}\sum_{\tau_1=1}^{+\infty}\sum_{\tau_2=\tau_1+1}^{+\infty} K_i^j(\tau_1,\tau_2) r_{i,t-\tau_1} r_{i,t-\tau_2}\\ &+\sum_{\tau=1}^{+\infty} \phi_\times^j(\tau)(r_{A,t-\tau} r_{B,t-\tau} - \overline{r_A r_B})\\ &+\sum_{\tau_1=2}^{+\infty}\sum_{\tau_2=1}^{\tau_1-1} K_\times^j(\tau_1,\tau_2) r_{A,t-\tau_1} r_{B,t-\tau_2} + \sum_{\tau_1=1}^{+\infty}\sum_{\tau_2=\tau_1+1}^{+\infty} K_\times^j(\tau_1,\tau_2) r_{A,t-\tau_1} r_{B,t-\tau_2}.\end{aligned} \tag{4.2.3}$$

In this framework, the returns $(r_{i,t})_{t>0, i\in\{A,B\}}$ are not defined at microscale anymore, and are related to the volatility such that $r_{i,t} = \sigma_{i,t}\xi_{i,t}$ where $\xi_{i,t} \sim \mathcal{N}(0,1)$. It is interesting to notice that the MQGARCH framework is more consistent with the data format presented in Section 4.1, since it is discrete quantities. Under this formulation of the (MQ)GARCH model, kernels and covariance structures are now independent of the parameter dt and only depends on the discretized quantities.

Additionally, the mean squared volatility is then determined through Equation (4.2.4):

$$\begin{pmatrix}\overline{\sigma_{A,\infty}{}^2}\\ \overline{\sigma_{B,\infty}{}^2}\end{pmatrix} = \left(I_2 - \left(\sum_{\tau=1}^{+\infty}\begin{pmatrix}\phi_1^1 & \phi_2^1\\ \phi_1^2 & \phi_2^2\end{pmatrix}(\tau)\right)\right)\begin{pmatrix}\overline{\sigma_A{}^2}\\ \overline{\sigma_B{}^2}\end{pmatrix}. \tag{4.2.4}$$

The subsequent section introduces the "adapted" covariance structures needed to characterise the kernels of Equation (4.2.3), that are $(L_j^i)_{i,j\in\{1,2\}}$, $(\phi_j^i)_{i,j\in\{1,2\}}$, $(K_j^i)_{i,j\in\{1,2\}}$, $(\phi_\times^i)_{i\in\{1,2\}}$ and $(K_\times^i)_{i\in\{1,2\}}$.

4.2.2 Covariance structures

The calibration aims at determining the feedback kernels $(L_j^i)_{i,j\in\{1,2\}}$, $(\phi_j^i)_{i,j\in\{1,2\}}$, $(K_j^i)_{i,j\in\{1,2\}}$, $(\phi_\times^i)_{i\in\{1,2\}}$ and $(K_\times^i)_{i\in\{1,2\}}$ and baseline values $(\sigma_{i,\infty})_{i\in\{1,2\}}$ of Equation (4.2.3) which cannot be directly observed in data. The method of moments relies on the fact that kernels can be deduced from covariance functions, which can easily be estimated from empirical data. Given the new model delineated in Equation (4.2.3), we must consider adapting the covariance structures and the corresponding Yule-Walker system.

Let us detail the number of unknowns to determine in this model. Considering a time grid up to lag q,

- $(L_j^i(\tau))_{i,j\in\{1,2\},1\le\tau\le q}$ and $(\phi_j^i(\tau))_{i,j\in\{1,2\},1\le\tau\le q}$ bring each $q\times 2\times 2$ unknowns
- $(K_j^i(\tau_1,\tau_2))_{i,j\in\{1,2\},1\le\tau_1<\tau_2\le q}$ brings $\frac{q(q-1)}{2}\times 2\times 2$ unknowns. Indeed, since $(K_j^i)_{i,j\in\{1,2\}}$ are time-symmetric, it is sufficient to only determine the upper triangle entries.
- $(\phi_\times^i(\tau))_{i\in\{1,2\},1\le\tau\le q}$ and $(K_\times^i(\tau_1,\tau_2))_{i\in\{1,2\},1\le\tau_1,\tau_2\le q,\tau_1\neq\tau_2}$ bring together $q\times q\times 2\times 1$ unknowns.

Additionally, given the form of the model, we need covariance structures containing information on both the time-diagonal of the kernels ($\tau_1=\tau_2$) and their off-time-diagonal $\tau_1\neq\tau_2$. The first quantity of interest is the covariance of the activities of the process, for all $\tau>0$, for $i,j\in\{1,2\}$,

$$\mathcal{C}_{ij}(\tau) := \mathbb{E}(\sigma_{i,t}^2 r_{j,t-\tau}^2) - \overline{\sigma_i^2}\,\overline{r_j^2}. \tag{4.2.5}$$

$(\mathcal{C}_{ij}(\tau))_{i,j\in\{1,2\},1\le\tau\le q}$ allow to fix $q\times(2\times 2)$ equations and predominantly shape the kernels $(\phi_i^j)_{i,j\in\{1,2\}}$.

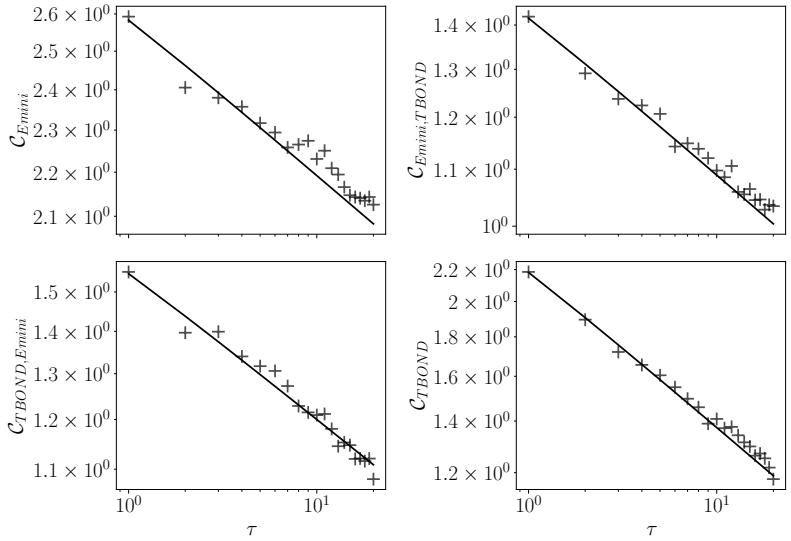


Figure 4.2: 2-points correlation $\mathcal{C}_{ij}(\tau)$, as defined in Equation (4.2.5), calculated for the pair E-mini vs TBOND. The scatter points represent the empirical points, whereas the solid black line represents the fit used for the calibration. The subplots are in log-log scale.

Consistently, we also define two relevant three-point correlation structures, for $0 < \tau_1 < \tau_2$,

$$\mathcal{D}_{ij}(\tau_1, \tau_2) := \mathbb{E}\left(\left(\sigma^2_{i,t} - \langle \sigma^2_{i,t} \rangle_t\right) r_{j,t-\tau_1} r_{j,t-\tau_2}\right) \tag{4.2.6}$$

and, for $0 < \tau_1, \tau_2$,

$$\mathcal{D}_{\times j}(\tau_1, \tau_2) := \mathbb{E}\left(\left(\sigma^2_{j,t} - \langle \sigma^2_{j,t} \rangle_t\right) r_{A,t-\tau_1} r_{B,t-\tau_2}\right). \tag{4.2.7}$$

$\mathcal{D}_{ij}$ is time-symmetric and allows to fix $\frac{q(q-1)}{2} \times (2 \times 2)$ equations. It mainly shapes the kernels $(K^i_j)_{i,j\in\{1,2\}}$. $\mathcal{D}_{\times j}$ fixes $q \times q \times (2 \times 1)$ equations, and mainly forms $(\phi^i_\times)_{i\in\{1,2\}}$ and $(K^i_\times)_{i\in\{1,2\}}$.

The above covariance structures enable to characterise the kernels $(\phi^i_j)_{i,j\in\{1,2\}}$, $(K^i_j)_{i,j\in\{1,2\}}$, $(\phi^i_\times)_{i\in\{1,2\}}$ and $(K^i_\times)_{i\in\{1,2\}}$ up to a lag q, after which the covariance structures and kernels are considered negligible. Moreover, Equation (4.2.4) enables to retrieve the baseline values from $(\phi^i_j)_{i,j\in\{1,2\}}$.

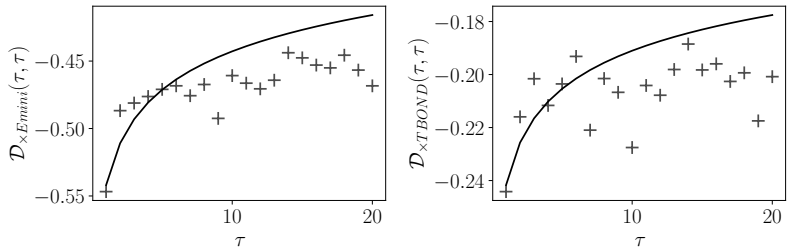


Figure 4.3: 3-points correlation $(\mathcal{D}_{\times j})$ for the pair E-mini vs TBOND, as defined by Equation (4.2.7). The scatter points represent the empirical points, whereas the solid black line represents the fit used for the calibration.

Finally, to recover the leverage feedback which weight the influence of past returns on future activity, we define an additional two-point correlation:

$$\mathcal{V}_{ij}(\tau) := \mathbb{E}\left(\sigma^2_{i,t} r_{j,t-\tau}\right). \tag{4.2.8}$$

$(\mathcal{V}_{ij}(\tau))_{i,j\in\{1,2\},1\leq\tau\leq q}$ allows to fix $q \times (2 \times 2)$ equations and predominantly shapes the kernels $(L^j_i)_{i,j\in\{1,2\}}$.

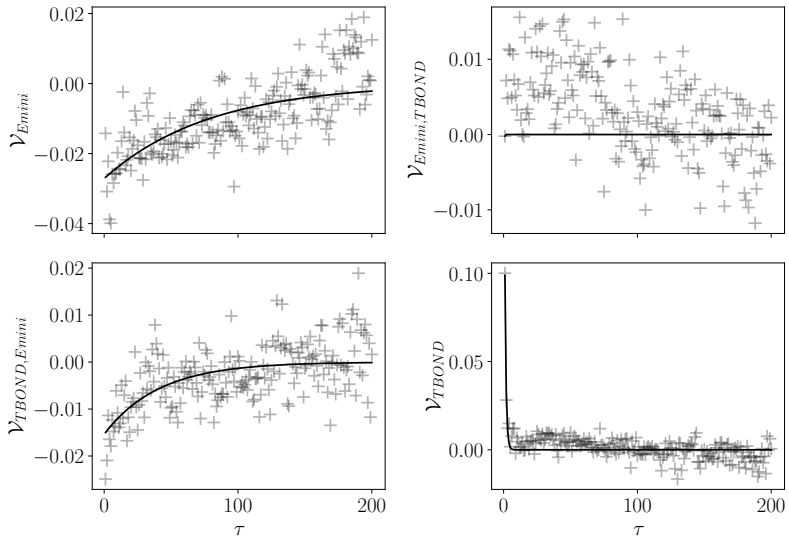


Figure 4.4: 2-points correlation $(\mathcal{V}_{ij})_{i,j\in\{1,2\}}$, as defined in Equation (4.2.8), calculated for the pair E-mini vs TBOND. The scatter plots represent the empirical points, whereas the solid black line represents the fit used for the calibration.

Let us note that some of the aforementioned covariance structures involve the fourth moments of financial returns which, due to their fat-tailed distribution, might be undefined and could lead to numerical instability. An interesting alternative would be to use mixed moments as employed in Chicheportiche *et al.* (2014) [12].

So as not to confuse or weigh down the explanation, the adapted Yule-Walker equations linking the feedback kernels to these covariance structures are presented and demonstrated in Appendix B.5.

As the calibration relies on these four covariance structures, to eliminate the noise, the covariance functions computed on empirical data are approximated ("smoothed") the following way:

- $(\mathcal{C}_{ij})_{i,j\in\{1,2\}}$ and the time-diagonal of $(\mathcal{D}_{ij})_{i,j\in\{1,2\}}$ and $(\mathcal{D}_{\times j})_{j\in\{1,2\}}$, i.e. $\tau \to (\mathcal{D}_{ij}(\tau,\tau))_{i,j\in\{1,2\}}$ and $\tau \to (\mathcal{D}_{\times j}(\tau,\tau))_{j\in\{1,2\}}$, are replaced by their best fit of $\tau \to \frac{n\exp(-\beta\tau)}{(1+\gamma\tau)^{\alpha}}$.

  Figure 4.2 and Figure 4.3 illustrate the approximation of $(\mathcal{C}_{ij})_{i,j\in\{1,2\}}$ and $(\mathcal{D}_{\times j}(\tau,\tau))_{j\in\{1,2\}}$, respectively, for the pair E-mini VS TBOND (note that the time-diagonal of $(\mathcal{D}_{ij})_{i,j\in\{1,2\}}$ is directly $(\mathcal{C}_{ij})_{i,j\in\{1,2\}}$). Additionally, Figure 4.2 and Figure 4.3 show that $(\mathcal{C}_{ij})_{i,j\in\{1,2\}}$ and the time-diagonal of $(\mathcal{D}_{\times j})_{j\in\{1,2\}}$ exhibit long-range correlations (note the log-log scale in Figure 4.2). This observation justifies the use of a power law function to smooth the empirical values; it also means that volatility movements that happened early in the day still have influence on the end-of-day volatility. Moreover, Figure 4.3 demonstrates that the diagonals of $(\mathcal{D}_{\times,j})_{j\in\{\text{E-mini, TBOND}\}}$ are negative for the pair E-mini VS TBOND which could potentially indicate that, since the returns of the E-mini and the TBOND are negatively correlated for the studied period (correlation coefficient is -0.15), when past returns covariance is, in absolute value, greater than its average value, then it tends to increase future volatility, and conversely, when the returns of the E-mini and the TBOND become less negatively correlated, i.e. when their covariance tends to zero, then it tends to decrease the asset's volatility. However, the time-diagonals of $(\mathcal{D}_{\times j})_{j\in\{A,B\}}$ also incorporate volatility effects which could explain such negative values as we know that volatility begets volatility. In fact, an increase variance in either the E-mini or the TBOND returns, would increase, in absolute value, the returns covariance, and be associated with an increase future volatility.

- Observing that, for all $i,j\in\{1,2\}$, the largest eigenvalue of the off-time-diagonal of $\mathcal{D}_{ij}$ and $(\mathcal{D}_{\times j})$, i.e. the off-diagonal of the matrices $(\mathcal{D}_{ij}(\tau_1,\tau_2))_{1\le\tau_1,\tau_2\le q}$ and $(\mathcal{D}_{\times j}(\tau_1,\tau_2))_{1\le\tau_1,\tau_2\le q}$, is a lot more significant that the subsequent ones, we approximate, for all $i,j\in\{1,2\}$, $(\mathcal{D}_{ij}(\tau_1,\tau_2))_{1\le\tau_1,\tau_2\le q,\tau_1\neq\tau_2}$ and $(\mathcal{D}_{\times j}(\tau_1,\tau_2))_{1\le\tau_1,\tau_2\le q,\tau_1\neq\tau_2}$ by their rank-one

approximation, where the first eigenvector is replaced by its best fit of $\tau \to \exp(-\beta\tau)$. For instance, Figure 4.5 presents the first eigenvector and its fit of the 3-points correlation $\mathcal{D}_{ij}$ for the pair E-mini vs TBOND. Let us note that, the "martingalisation" process described in Section 4.1.2 facilitates this exponential fit. Indeed, without applying the "martingalisation" process, the first eigenvector of the 3-point correlation matrix $\mathcal{D}_{ij}$ would reflect the strong mean-reversion observed at this scale–a property not pertinent to the current analysis–characterized by the first two points having opposite signs. At this stage, we also observe that the off-time-diagonal of $(\mathcal{D}_{\times j})_{j\in\{1,2\}}$, that is $(\mathcal{D}_{\times j}(\tau_1,\tau_2))_{1\leq\tau_1,\tau_2\leq q,\tau_1\neq\tau_2,j\in\{1,2\}}$, is mostly noise, hovering around zero, giving us an hint on the off-time-diagonal of the kernels $(K^j_\times)_{j\in\{A,B\}}$.

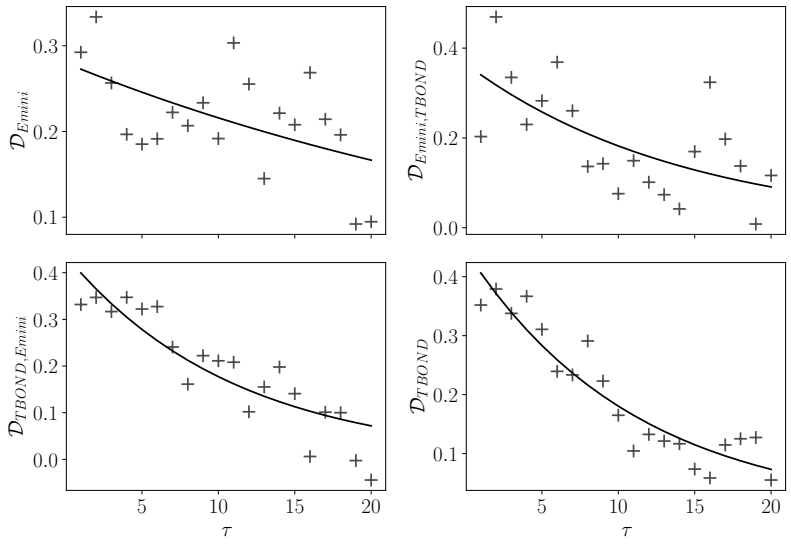


Figure 4.5: First eigenvector of the 3-points correlation $\mathcal{D}_{ij}$, as defined in Equation (4.2.6), for the pair E-mini vs TBOND. The scatter plots are the empirical values and the solid black line is the fit used in the rank-one approximation of $\mathcal{D}_{ij}$.

- $(\mathcal{V}_{ij})_{i,j\in\{1,2\}}$ is approximated by the best fit of $\tau \to a\exp(-b\tau)$. Figure 4.4 presents the approximation of $(\mathcal{V}_{ij})_{i,j\in\{1,2\}}$ for the pair E-mini VS TBOND. Figure 4.4 also shows that the feedback effect of the TBOND returns on the E-mini volatility (upper right panel) is very noisy and close to zero, conversely to feedback effects from the E-mini on its own volatility and on that of the TBOND (left panels). Moreover, the lower panels show that E-mini negative past returns and, consistent with the negative correlation between the E-mini and TBOND returns, TBOND positive past returns tend to increase TBOND future volatility. It is also notable that the TBOND self-leverage feedback (lower right panel) seems to diminish very fast.

4.2.3 4-stage calibration with the method of moments

This section aims at delineating the calibration method and Appendices B.7 and B.5 provide further details.

In practice, the calibration is performed within a two-dimensional (2D) framework, encompassing two assets, and is decomposed into several steps. This approach is chosen to maintain numerical stability and is sufficient to capture all leverage, linear, quadratic, and cross feedback effects. All steps rely on linear relationships between the covariance and the kernels, the so-called Yule-Walker system (Appendix B.5 develops such system for the model delineated in Equation (4.2.3)). Solving these linear systems of equations is achieved by considering that after a lag q the kernels and the covariance structures are negligible (see Appendix B.7 for details on the implementation of the Yule-Walker systems for the 2D-QGARCH of Equation (4.2.3)). We now outline the main steps of the calibration.

To accurately distinguish leverage feedback from quadratic and cross feedback effects, we initially estimate the linear, quadratic, and cross kernels using symmetric data. Subsequently, we estimate the leverage kernels on the original dataset, excluding from $\mathcal{V}$ the contributions accounted for by the linear feedback.

Specifically, to completely eliminate the leverage effect, we transform our datasets as follows

$$\begin{cases} r^* & \leftarrow [r, -r] \\ \sigma^* & \leftarrow [\sigma, \sigma] \end{cases}$$

Thus, the leverage effect, characterised by, for $\tau > 0$, $|\mathbb{E}(\sigma_t r_{t-\tau})| > \mathbb{E}(\sigma_t r_{t+\tau}) \approx 0$, is no longer present since $\mathbb{E}(\sigma_t^* r_{t-\tau}^*) = \mathbb{E}(\sigma_t r_{t-\tau}) - \mathbb{E}(\sigma_t r_{t-\tau}) = 0$. The linear, quadratic and cross kernels are then estimated from (r^*, σ^*) in a three-step procedure and a fourth step is implemented to determine the leverage kernels. Specifically, the steps are structured as follows.

1. Since it is numerically stable, as shown in Appendix B.8, the univariate QHawkes feedback kernels of each asset are estimated up to a certain lag q. Thus, from this first step, we obtain $(\phi_i^i(\tau))_{i\in\{1,2\},1\leq\tau\leq q}$ and $(K_i^i(\tau_1,\tau_2))_{i\in\{1,2\},1\leq\tau_1<\tau_2\leq q}$.
2. Considering the cross contributions null ($\phi_\times = k_\times = 0$), we estimate the cross linear and quadratic contributions characterised by the values $(\phi_j^i(\tau))_{i,j\in\{1,2\},i\neq j,1\leq\tau\leq q}$ and $(K_j^i(\tau_1,\tau_2))_{i,j\in\{1,2\},i\neq j,1\leq\tau_1<\tau_2\leq q}$. After this second step, the estimation of the baseline activity $\sigma_{A,\infty}^2$ and $\sigma_{B,\infty}^2$ can be achieved using Equation (4.2.4).
3. Accounting for the contributions of the kernels obtained in steps 1 and 2, we estimate the cross feedback components up to a certain lag q, i.e., $(\phi_\times^i(\tau))_{i\in\{1,2\},1\leq\tau\leq q}$ and $(K_\times^i(\tau_1,\tau_2))_{i\in\{1,2\},1\leq\tau_1,\tau_2\leq q,\tau_1\neq\tau_2}$.
4. Once these three steps are completed, the leverage kernels, $(L_i^j(\tau))_{i,j\in\{1,2\},1\leq\tau\leq q}$, can then be estimated up to a lag q on the original dataset (r,σ) while removing the contributions of the quadratic kernels (we refer to Section B.7.7 of Appendix B.7 for further details).

At this stage, it is worth noticing some empirical observations.

First, observing the empirical covariance structures and calibrating the 2D-QGARCH on various pairs of assets demonstrated that $\mathcal{D}_{\times j}(\tau_1,\tau_2)$ is mostly null for $\tau_1 \neq \tau_2$. Notably, the correlation between the returns of two assets is significant only when computed for returns occurring simultaneously; otherwise, it is null. This observation, which justifies further the martingale hypothesis, leads to $K_\times^j(\tau_1,\tau_2) = 0$ when $\tau_1 \neq \tau_2$. Consequently, only the time-diagonal $\phi_\times$ is to be characterised hereafter.

Second, consistently with the empirical findings of [70], the calibration of the off-time-diagonal of K_i^j reveals that these kernels can then be represented by a rank-one approximation k_i^j, such that $K_i^j(\tau_1,\tau_2)_{1\leq\tau_1,\tau_2\leq q,\tau_1\neq\tau_2} = k_i^j(\tau_1)k_i^j(\tau_2)$. Indeed, the first eigenvalue of the matrices $(K_i^j(\tau_1,\tau_2)_{1\leq\tau_1,\tau_2\leq q})_{i,j\in\{1,2\}}$ with a diagonal of zeros, is larger than the subsequent ones. Besides, the time-diagonal of these kernels is directly represented by $(\phi_i^j(\tau))_{1\leq\tau\leq q}$.

Remark 4.2.1. Note that the calibration by Maximum Likelihood estimation, whose implementation is detailed in Appendix B.3, has also been implemented and provides similar results as the method of moments. Notably, the Maximum Likelihood estimation is particularly efficient when the GMM outcome is used as an initial guess.

4.2.4 Calibration of the 2D-QGARCH for the pair E-mini VS TBOND

Implementing the above procedure on the time series of the E-mini and the TBOND resulted in the kernels presented Figure 4.6. For clearer representation, Figure 4.6b presents the rank-one approximation of the off-time-diagonal of the quadratic kernels $(K_i^j)_{i,j\in\{\text{E-mini,TBOND}\}}$. Let us comment

on the results.

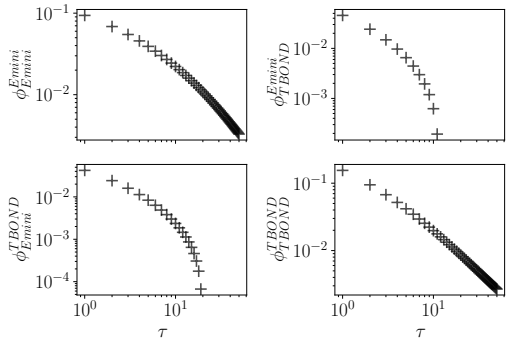
It appears that the covariance structures forms, $\mathcal{C}$, $\mathcal{D}$, $\mathcal{D}_\times$ and $\mathcal{V}$ (see Figures 4.2, 4.5, 4.3 and 4.4, respectively), shape the resulting kernels. Notably, $(\mathcal{C}_{ij})_{i,j\in\{\text{E-mini,TBOND}\}}$'s long-range correlations are reflected in the kernels $(\phi_i^j)_{i,j\in\{\text{E-mini,TBOND}\}}$ shown in Figure 4.6a.

Moreover, the left panels of Figure 4.6c show that the negative exponential shape of the E-mini's past returns influence on its volatility and that of the TBOND (see the left panels of Figure 4.4) is well-captured by the kernels $(L^i_{\text{E-mini}})_{i\in\{\text{E-mini,TBOND}\}}$. Additionally, the leverage kernel capturing the TBOND's returns feedback on its own volatility, depicted in the lower right panel of Figure 4.6c, exhibits a positive shape that rapidly decays to zero. This observation is consistent with $\mathcal{V}_{\text{TBOND}}$'s shape in Figure 4.4, and with the negative correlation between the E-mini and the TBOND returns. Consistently with the observation of Figure 4.4, the leverage kernel representing the feedback of the TBOND returns on the E-mini volatility mainly consists of noise around zero and, therefore, is not represented in Figure 4.6. Stronger leverage signal is expected at larger time scale, as the leverage effect predominantly manifests at the daily scale, as demonstrated by Figure 1.5.

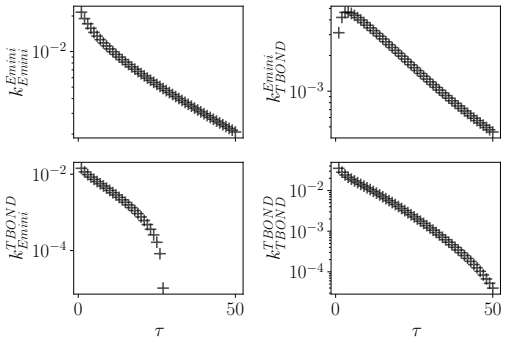
Furthermore, the left panels of Figure 4.6d delineate the profiles of the returns covariance feedback kernels $(\phi^j_\times)_{j\in\{\text{E-mini,TBOND}\}}$ whose values indicate that when the absolute value of the covariance between the TBOND and the E-mini, which is negative for the studied period, becomes stronger, in absolute value, it tends to decrease the TBOND's volatility while it increases the E-mini volatility. Note that, this is not exactly the effect that was previously discussed for $(\mathcal{D}_{\times j})_{j\in\{\text{E-mini,TBOND}\}}$ in Figure 4.3. In fact, the kernels $(\phi^j_\times)_{j\in\{\text{E-mini,TBOND}\}}$ reveal solely the effect of past returns covariance, while $(\mathcal{D}_{\times j})_{j\in\{\text{E-mini,TBOND}\}}$ also reflect feedback effects from the quadratic kernels.

Finally, the exponential shape of the rank-one approximation of the off-time-diagonal of $(K_i^j)_{i,j\in\{\text{E-mini,TBOND}\}}$ is inherent to the exponential smoothing of the covariance structure $\mathcal{D}$.

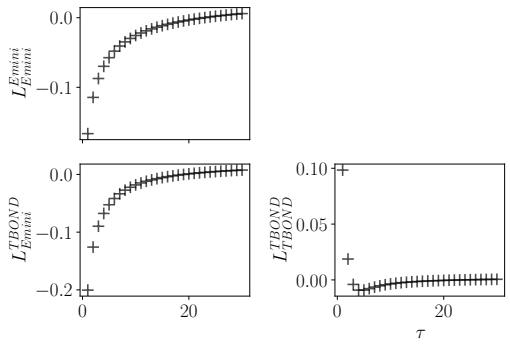
The subsequent section presents the calibration of the 2D-QGARCH obtained with this method on various pairs of assets.



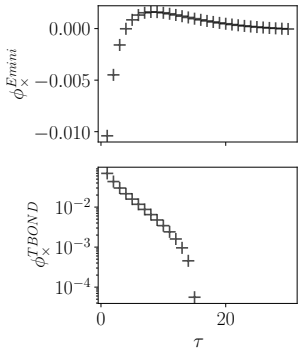
(a) Linear kernels $(\phi_i^j)_{i,j\in\text{E-mini, TBOND}}$ up to a lag $q = 50$, in log-log scale.



(b) Rank one approximation of the off-time-diagonal of the quadratic kernel $(K_i^j)_{i,j\in\text{E-mini, TBOND}}$ up to a lag $q = 50$. The y-axis is in log-scale.



(c) Left panels: Leverage kernels of the E-mini feedback $(L^j_{\text{E-mini}})_{j\in\text{E-mini, TBOND}}$ up to a lag $q = 30$. Right panels: Self-leverage kernel of the TBOND feedback $L^{\text{TBOND}}_{\text{TBOND}}$ up to a lag $q = 30$.



(d) Returns past covariance kernels $(\phi^j_\times)_{j\in\text{E-mini, TBOND}}$ up to a log $q = 30$. For the lower panel, the y-axis is in log-scale.

Figure 4.6: Results of the 2D QGARCH calibration on the pair E-mini VS TBOND.

4.3 2D-QGARCH on futures on indices

Following the methodology explained in Sections 4.1 and 4.2, we calibrate 2D-QGARCH on all the pairs composed by the following futures on indices: E-mini, E-mini-3 (3months futures), futures contracts on NASDAQ, futures contracts on DOW JONES, futures contracts on CRUDE OIL and TBOND. Note that, with the exception of the futures contracts on CRUDE OIL and the TBOND which are futures on a commodities-index and futures on a bond-index respectively, all of these futures contracts are futures on stock-indices.

Figure 4.7 presents the norm-L1 of the resulting kernels and Appendix B.9 presents their full profiles.

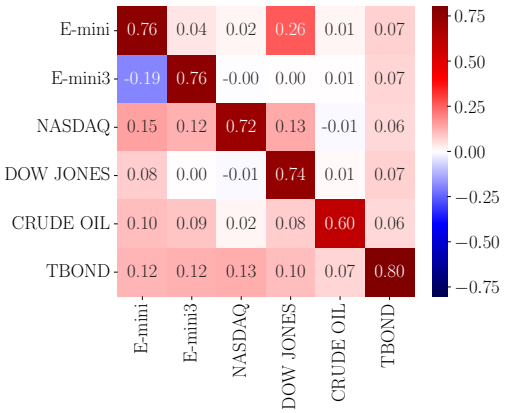
Volatility feedback kernels ϕ_i^j

The primary observation from Figure 4.7 concerns the kernels characterising the influence of past volatility on future volatility, $(\phi_i^j)_{i,j\in\{1,2\}}$, shown in Figure 4.7a. The strong diagonal of Figure 4.7a indicates that most of the system endogeneity originates from the feedback of the asset's own volatility on itself. Additionally, Figure 4.7a demonstrates that the system operates near criticality, as the diagonal values are all above 0.6 and the largest eigenvalues of the pariwise matrices are between 0.73 and 0.90.

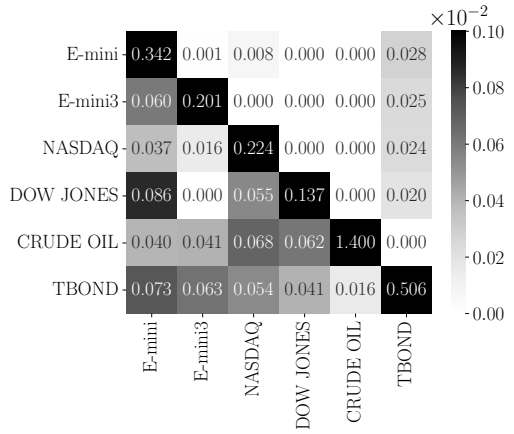
Subsidiary observations regarding the other kernels are worth noting.

Quadratic trend feedback kernels k_i^j

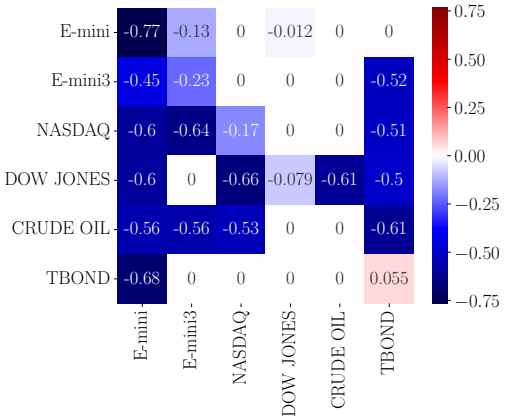
The norms of the rank-one approximations k_i^j of the off-time-diagonal of the the quadratic



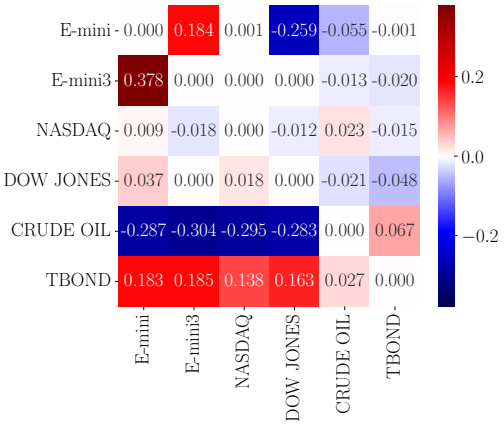
(a) Norm of the linear kernels $||\phi_i^j||_{i,j}$.



(b) Norm of the rank-one approximation of the off-diagonal of the quadratic kernels $||(k_i^j)^2||_{i,j}$.



(c) Norm of the leverage kernels $||L_i^j||_{i,j}$.



(d) Norm of the cross kernels $||\phi_\times^j||_j$.

Figure 4.7: Norms of the kernels from the calibration of 2D-QGARCHs on pairs of futures on indices. For the four sub-figures, the x-labels determine the index providing feedback and the y-labels determine the index receiving the feedback, i.e., for a kernel K_i^j, i is labelled on the x-axis while j is labelled on the y-axis.

kernels K_i^j, in Figure 4.7b, are consistent with their profiles in Appendix B.9. The strong diagonal of Figure 4.7b shows that the trend self-quadratic feedback is quite strong compared to the other quadratic feedback, meaning that the futures contracts past trends tend to increase their own volatility. Furthermore, some futures contracts present particular trend quadratic feedback. Notably, the first column of Figure 4.7b demonstrates that the quadratic feedback of the E-mini on futures on indices is quite strong compared to the other cross quadratic feedback, meaning that the S&P500 index past trends predominantly increase the volatility of the other futures on indices. Similarly, the last column of Figure 4.7b shows that the quadratic feedback of the TBOND on futures on stock-indices is quite strong and consistent across contracts, meaning that the trend of the futures contracts on the US-10Y-Treasury-bond increases uniformly the volatility of the futures on stock-indices. It is also interesting to note that futures contracts on the CRUDE OIL and on the 10Y-Treasury bond appear to be the most impacted by these trend quadratic feedback, as demonstrated by the two last rows of Figure 4.7b.

Leverage feedback kernels L_i^j

It appears that leverage kernels are not always relevant, as evidenced by the many zeros in Figure 4.7c and exemplified by the pair E-mini VS TBOND. When present, except in the case of the TBOND as discussed above, the leverage feedback is negative, consistent with the idea that

large negative returns trigger increased volatility. Notably, the DOW JONES futures[1] seems to be particularly affected by the leverage effect from both itself and other futures on indices. We also point out that the model is calibrated intraday, whereas the leverage effect predominantly appears on a daily scale.

Returns covariance feedback kernels $\phi^j_\times$

Figure 4.7d reports on the influence of the past covariance of the returns of the two assets involved in each pair on the volatility of both assets in the pair.

The pair E-mini VS E-mini-3 exhibits a particularly strong returns covariance feedback compared to the other pairs (see the upper left 2×2 matrix of Figure 4.7d), which signifies that an increase past covariance between the returns of the E-mini and those of the E-mini-3 tends to increase the volatility of both futures contracts.

Additionally, the volatility of the futures on bond-indices (TBOND) and on commodities-indices (CRUDE OIL) appears to be significantly impacted by the past covariance of their returns with the returns of futures on stock-indices, as demonstrated by the last two rows of Figure 4.7d. Specifically, for all the pairs combining the CRUDE OIL futures with futures on stock-indices, the returns covariance feedback kernels, $\phi_\times$, exhibit a negative norm, while the pairs combining a future on stock-index with the TBOND, display a positive returns covariance feedback. During the studied period, the CRUDE OIL futures contracts and futures on stock-indices were positively correlated while stocks and bonds were negatively correlated. Therefore, the sign of the norm of the returns covariance feedback kernels $(||\phi^j_\times||)_j$ for pairs combining futures on stock-index with the TBOND and for pairs combining futures on stock-index with the CRUDE OIL conveys the same meaning in both cases: when the absolute value of the past returns covariance increases, it tends to decrease future volatility. Conversely, when the two involved assets, generally in phase (or in opposite phase in the case of the TBOND), decouple, it tends to increase future volatility.
While this effect appear to be symmetric, albeit slightly down, for pairs combining future on stock-index with the CRUDE OIL futures, meaning that an increase past covariance between the CRUDE OIL returns and those of a stock-index tends to decrease the stock-index volatility. It is, nevertheless, different for pairs combining future on stock-index with the TBOND. Notably, an increase, in absolute value, covariance between the TBOND returns and a stock-index returns tends to increase the stock-index volatility. As for the pair CRUDE OIL futures VS TBOND itself, an increase, in absolute value, past returns covariance tends to decrease TBOND volatility while it increases CRUDE OIL volatility.

[1] US index.

4.4 MQGARCH×factor-model calibration on pairs of type Index × stocks

To calibrate on a large scope of stocks, while accommodating for the significant feedback effect from indices, specifically from the S&P500 index, we calibrate the 2D-QGARCH model coupled with a factor model. The framework is outlined below.

4.4.1 MQGARCH×factor − Model framework

The starting point of this framework is a 1-factor model. Specifically, as described by Equation (4.4.1), the returns of stock i are decomposed into a sum of their market exposure, represented by the factor component f_0, with an exposure coefficient β_i, and a residual component $e_{i,t}$, which is idiosyncratic to the stock's dynamics.

$$r_t^i = \beta_i f_{0,t} + e_{i,t}. \tag{4.4.1}$$

Empirically, f_0 represents the market mode and is directly sourced from the E-mini returns while $(\beta_i)_i$ is estimated with the correlations between the returns of stock i and f_0.

The particularity of such framework, and our main interest, lies in the dynamics of f_0 and $(e_i)_i$. Specifically, we consider here that f_0 dynamics are those of a 1D-QGARCH, while $(e_i)_i$ follow a 2D-QGARCH model with self-feedback and market feedback to account for the empirical observations that the stock idiosyncratic dynamics depends on those of the market [57, 58, 59].

Thus, the volatility of f_0 is driven by

$$\sigma_{f_0,t}^2 = \sigma_{f_0,\infty}^2 + \sum_{k=1}^{+\infty} L_0(k) f_{0,t-k} + \sum_{k_1=1}^{+\infty}\sum_{k_2=1}^{+\infty} K_0(k_1,k_2) f_{0,t-k_1} f_{0,t-k_2},$$

while the volatility dynamics of $(e_i)_i$ are characterised by

$$\begin{aligned}\sigma_{i,t}^2 =& \sigma_{i,\infty}^2 + \sum_{k=1}^{+\infty} L_i(k) e_{i,t-k} + \sum_{k=1}^{+\infty} L_\times^i(k) f_{0,t-k} \\ &+ \sum_{k_1=1}^{+\infty}\sum_{k_2=1}^{+\infty} K_i(k_1,k_2) e_{i,t-k_1} e_{i,t-k_2} + \sum_{k_1=1}^{+\infty}\sum_{k_2=1}^{+\infty} K_\times^i(k_1,k_2) f_{0,t-k_1} f_{0,t-k_2}.\end{aligned} \tag{4.4.2}$$

Scrutinizing this later expression, we see that the idiosyncratic term's volatility, $(\sigma_{i,t})_{i,t}$, is influenced by the past returns of both the factor and the residual, characterised by the leverage kernels $L_\times$ and L_i respectively, and by the past trends of both the factor and the residual, characterised by the quadratic kernels $K_\times$ and K_i respectively. As before, we can decompose the quadratic kernels to highlight the diagonal feedback (feedback of the activity), which primarily characterizes the endogeneity of the system, and the off-diagonal feedback (feedback of the trends) as follows:

$$\begin{aligned}\sigma_{i,t}^2 =& \sigma_{i,\infty}^2 + \sum_{k=1}^{+\infty} L_i(k) e_{i,t-k} + \sum_{k=1}^{+\infty} L_\times^i(k) f_{0,t-k} \\ &+ \sum_{k=1}^{+\infty} \phi_i(k) e_{i,t-k}^2 + 2\sum_{k_1=1}^{+\infty}\sum_{k_2=k_1+1}^{+\infty} K_i(k_1,k_2) e_{i,t-k_1} e_{i,t-k_2} \\ &+ \sum_{k=1}^{+\infty} \phi_\times^i(k) f_{0,t-k}^2 + \sum_{k_1=1}^{+\infty}\sum_{k_2=k_1+1}^{+\infty} K_\times^i(k_1,k_2) f_{0,t-k_1} f_{0,t-k_2}.\end{aligned} \tag{4.4.3}$$

Note that in this framework, relying on empirical observations, cross-trend feedback are no longer present.

The goal of this section is to investigate the values of $(\sigma_{i,\infty})_i$ and the features of the various feedback mechanisms on the residual's future activity for US stocks. Notably, we want to characterise the kernels of the residual and index leverage feedback ($(L_i)_i$ and $(L_\times^i)_i$), the kernels weighting the influence of the residual and index past volatility ($(\phi_i)_i$ and $(\phi_\times^i)_i$) and the kernels

outlining the influence of the past trends of the residual and of the index (the off-diagonal of $(K_i)_i$ and $(K^i_\times)_i$).

The subsequent section is dedicated to presenting the data employed in this study, followed by a discussion of the results obtained.

4.4.2 Data

Our initial dataset contains 1-minute returns $(r^i)_{i=1,\dots,317}$ of 317 US stocks spanning from 2013 to 2023 and the 1-minute E-mini returns, f_0, for the same period. We consider only stocks which belonged to the S&P500 index for the whole studied period.

This initial dataset is then used to deduce the time series $(e_i)_i$ according to Equation (4.4.1) and taking, for all i, β_i as the covariance coefficients between the returns of stock i and the returns of the E-mini. The squared volatility time series, $\sigma^2_{i,t}$ and $\sigma^2_{f_0,t}$, are then approximated by the squared 1-minute returns, $e^2_{i,t}$ and $f^2_{0,t}$ respectively.

At this stage, we thus end up with two time series for the E-mini and for each stock of our 317 US stocks sample: one characterising the returns and the second one representing the squared volatility. These times series are then processed as before by the steps described in Section 4.1.2.

The method of moments is then used to calibrate the 1D-QGARCH on the E-mini, resulting in the values of $\sigma^2_{f_0,\infty}$, $(L_0(k))_{1\leq k\leq q}$ and $(K_0(k_1,k_2))_{1\leq k_1,k_2\leq q}$, where q is the lag after which the feedback is considered negligible.

Finally, the calibration of the 2D-QGARCH on the 317 stocks residuals, adapting the method described in Section 4.2 and Appendix B.7, is performed and the results are presented in the subsequent section.

4.4.3 MQGARCH×factor – Calibration Results

The kernels L_0, ϕ_0 and k_0, where $||\phi_0||$ and k_0 are the time-diagonal and the rank-one approximation of the off-time-diagonal of K_0 respectively, result from the calibration of a 1D-QGARCH on the E-mini, and are therefore identical to those shown in Figure 4.6.

Figure 4.8 presents the average profiles of the quadratic kernels shaping the volatility of the stock idiosyncratic dynamics, being the volatility feedback kernels $(\phi_i)_i$ and $(\phi^i_\times)_i$, and the rank-one approximation of the off-time-diagonal of the feedback kernels $(K_i)_i$ and $(K^i_\times)_i$.

Interestingly, on average, the norm of the residual's volatility feedback kernel is close to one ($\langle||\phi_i||\rangle_i = 0.94$), supporting the aforementioned results that suggest endogeneity primarily stems from the self-volatility feedback. Additionally, the profiles of the kernels characterising the residual volatility feedback on itself exhibit a long-range scale, as demonstrated by the log-log scale of the left panel of Figure 4.8, meaning that the idiosyncratic activity that occurred early in the day, continue to influence the idiosyncratic component's volatility throughout the entire day. Besides, these profiles are quite uniform across our stock sample as delineated by the thin grey area on the left panel of Figure 4.8.

Moreover, the average endogeneity arising from the market contribution is not negligible, reaching up to $||\phi^i_\times||_{\max} = 0.12$, with values hovering around $\langle||\phi^i_\times||\rangle_i = 0.02$. Notably, the past volatility of the market appears to predominantly influence the residual volatility of stocks in the Financials, Real Estate, and Industrials sectors. Additionally, the log-log scale of the second panel from the left of Figure 4.8 seems to exhibit long-range index influence on some stocks idiosyncratic activity.

As previously found, the off-time-diagonal of the trend feedback kernels, characterised by their rank-one approximation, have lower values than the diagonal feedback kernels with a norm averaging at $\langle||k^2_i||\rangle_i = 0.012$, which is consistent with the results of the calibration of the QHawkes on stocks in [70]. The average profile of k_is does not reflect the differences between stocks as demonstrated by the large grey area in the second panel from the right of Figure 4.8. Specifically, we find that for all stocks, the rank-one approximation k_i exhibits a decaying exponential shape, whose characteristic time and amplitude depend on the specific stock. Notably, the higher the residual activity, the longer the range of the residual trend feedback. Additionally, the Real Estate and Utilities sectors, which show the highest average levels of idiosyncratic activity, consistently present trend feedback kernels with the highest norms and longer ranges on average, specifically, $\langle||k^2_i||\rangle_{i\in\text{real estate}} = 0.017$ and $\langle||k^2_i||\rangle_{i\in\text{utilities}} = 0.018$. The kernels characterising the market's

past trends contribution to future residual's volatility exhibit smaller values, with a norm averaging $\langle ||(k^i_\times)^2|| \rangle_i = 4.7\ 10^{-4}$, but they still demonstrate a discernible signal, as shown by the profile in the right panel of Figure 4.8. Once again, the Real Estate and Utilities sectors stand out as having the most significant trend index feedback.

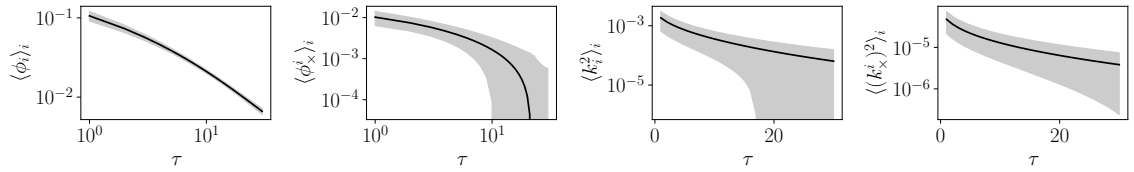


Figure 4.8: Average profiles of kernels for the dynamics of the residuals e_i. The solid black line represents the average profile and the grey area indicates the range of one standard deviation across the stocks.

Furthermore, we found highly disparate results across our sample of stocks for the self-leverage kernels, which characterise the influence of past residual returns on the residual's future volatility. In fact, some stocks demonstrate a positive leverage feedback that is difficult to interpret. These results are somewhat expected, as leverage effect predominantly manifests at the daily scale.
Nevertheless, the leverage feedback of the index on the volatility of the idiosyncratic term of stocks is quite stable across stocks and its average profile is represented in Figure 4.9. The negative exponential shape of the index leverage feedback on the stock idiosyncratic component's volatility indicates that E-mini negative past returns increase stock residual volatility. This result appears to hold for most the stocks of our sample as demonstrated by the thin grey area of Figure 4.9.

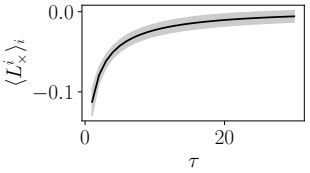


Figure 4.9: Average profile of the leverage kernel from the E-mini on the residual $L^i_\times$

Finally, the endogeneity ratios of these 2D-QGARCH×factor systems[2] are very close to 1, with all values exceeding 0.955 and an average of 0.96, indicating that the systems operate near criticality. Correspondingly the baseline values of the stock idiosyncratic activity average at $\langle \sigma^2_{i,\infty} \rangle_i = 0.04$.

To recap, this study yields two main results. First, most of the system's endogeneity stems from the long-range feedback of the idiosyncratic term's volatility, consistent with previous findings on the calibration of futures on indices; Second, the past volatility and negative returns of the index impact the stock's residual activity, supporting empirical findings that stock's idiosyncratic dynamics depend on those of the market [57, 58, 59].

A natural extension of this framework is to investigate a factor model with more than one factor, thereby characterising the influence of the index not only on the stock's idiosyncratic component but also on the various industrial sectors, whose dynamics would be accounted for by the additional factors.

Conclusion

Let us summarise the findings of this chapter.

Building upon the work of Blanc *et al.* [70] and the theoretical work presented in Chapter 3, we presented a non-parametric calibration method to characterise the feedback mechanisms that influence the dynamics of future volatility. To overcome the challenges associated with estimating an underlying MQHawkes process from aggregated data, the micro-scale QHawkes framework is approximated by a discrete QGARCH framework. The key to this transformation is assimilating the Hawkes intensity as the squared volatility divided by a time constant characteristic of the discretisation time scale.

We introduced the key ingredients and steps required to calibrate the 2D-MQGARCH framework using the method of moments, employing the pair E-mini VS TBOND as a pedagogical example.

[2]The endogeneity ratio of the 2D-QGARCH×factor model for stock i is the largest eigenvalue of $\begin{pmatrix} ||\phi_0|| & 0 \\ ||\phi^i_\times|| & ||\phi_i|| \end{pmatrix}$, and is then directly $\max(||\phi_0||, ||\phi_i||)$.

To recap, the primarily idea of the calibration is to build a linear system of equations using the covariance structures, which are directly observable in real data, to deduce the kernel functions.

Utilising this calibration method on pairs of futures on indices revealed that the strongest feedback effect on future volatility originates from the past volatility itself. Moreover, this analysis demonstrated that the past trends and returns of the E-mini significantly contribute to the activity of other futures contracts, compared to other cross-feedback effects.

Additionally, to investigate the feedback mechanisms on stock volatility while distinguishing market dynamics from stock idiosyncratic dynamics, we combined the 2D-QGARCH framework with a 1-factor model. Thus, the studied feedback mechanisms directly report on the influence of the past events on the stock idiosyncratic activity. Calibrating the 2D-QGARCH$\times$1-factor model on 317 US stocks, over a period spanning from 2013 to 2023, and sourcing the factor from the E-mini, proved to be quite stable across stocks. Notably, this calibration demonstrated that while most of the system endogeneity stems from the influence of past residual activity on itself, the past returns and past volatility of the factor still impact the future volatility of the stock idiosyncratic component. In the present analysis, we noted some disparities among industrial sectors, and it would be interesting to further develop a MQGARCH$\times$multi-factor framework to investigate the feedback loops between the market, the industrial sectors and the stock idiosyncratic component. We leave such study for future research.

Finally, it would be interesting to undertake a more detailed study of the calibration method to find an implementation that would allow the underlying QHawkes process at the microstructure level to be reconstructed from the aggregated data.

Take Home Message

- Approximating the (M)QHawkes process with the (M)QGARCH model provides a framework compatible with aggregated data and less prone to the microstructural noise of financial time series.
- Using a non-parametric method, such as the method of moments, to calibrate the model enables to remain agnostic on the kernel's shape.
- The method of moments relies on a linear system of equations, built from covariance structures that are empirically observable, and yields the kernel functions up to a fixed lag q.
- The calibration consistently demonstrates that most of the endogeneity stems from the asset's own volatility.
- Empirical endogeneity ratios exhibit values close to 1, suggesting that market activity operates near criticality.
- The past trends and negative past returns of the E-mini, futures contracts on the S&P500 index, significantly contribute to the increased activity of other assets, including both stocks and futures on indices.
- Coupling the MQGARCH with a factor model representation enables the study of the feedback mechanisms influencing the stock volatility while removing most of the market dynamics. Investigating the feedback kernels in this 2D-QGARCH$\times$factor model framework demonstrated that, although most of the index dynamics were filtered out from the stock idiosyncratic dynamics, the index's negative past returns and past volatility still tend to increase the volatility of the stock idiosyncratic component.

Chapter 5

Riding Wavelets: A Method to Discover New Classes of Price Jumps

It is as if price changes themselves are the main sources of news

J.-P Bouchaud, J. Bonart, J. Donier, M. Gould, *Trades, Quotes and Prices*

This chapter is largely based on [4], written in collaboration with Rudy Morel, Michael Benzaquen and Jean-Philippe Bouchaud.

Introduction

Extreme events and cascades of events are widespread occurrences in both natural and social systems [91]. Examples include earthquakes, volcanic eruptions, hurricanes, epileptic crises [16, 17], epidemic spread, financial crashes [9, 10, 87], economic crises [92, 93], book sales shocks [20, 21], riot propagation [18, 19] or failures in socio-technical systems [94]. Understanding the origin of such events is essential for forecasting and possibly stabilizing their dynamics.

A widely studied question is the reflexive nature of those shocks – the concept of financial market reflexivity was introduced by Soros in [8], to describe the idea that price dynamics are mostly endogenous and arise from internal feedback mechanisms, as was first surmised by Cutler, Poterba and Summers in 1988 [5] (see also [95]). Extreme events, in particular, are considered to be endogenous when they arise from feedback mechanisms within the system's structure [96, 97, 91]. Quantifying the extent of reflexivity in a complex system and distinguishing events caused by external shocks from those provoked endogenously, and more generally identifying different classes of events, are crucial questions.

Prior research has proposed to differentiate between endogenous and exogenous dynamics by analyzing the profile of activity around the shock [98, 20, 21, 22], in particular in the context of financial markets [99, 6]. It has been observed that endogenous shocks are preceded by a growth phase mirroring the post event power-law relaxation, in contrast to exogenous shocks that are strongly asymmetric, as Figure 5.1 exemplifies. The universality of this result is quite intriguing as they have been observed in various contexts: intraday book sales on Amazon [20, 21], daily views of YouTube videos [22] and intraday financial market volatility and price jumps [99, 6]. Besides, Wu *et al.* [100] differentiate exogenous and endogenous bursts of comment posting on social media using the analysis of collective emotion dynamics and time-series distributions of comment arrivals.

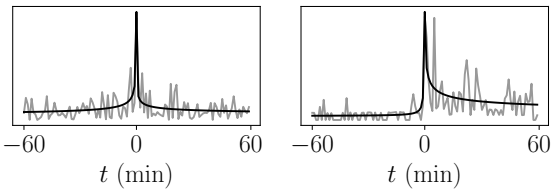


Figure 5.1: Examples of different asymmetry profiles in price jumps. Plain black lines are power law fits from [6] described in Equation (5.1.1).

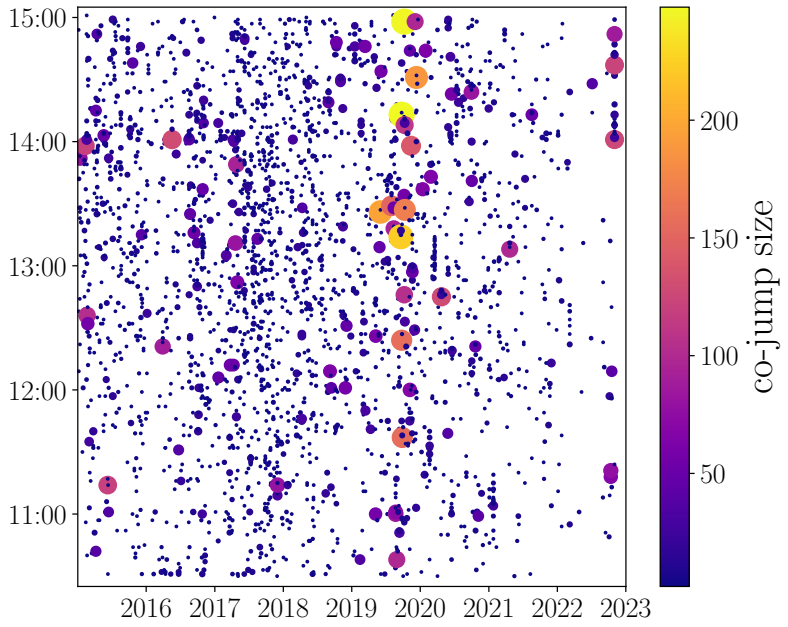


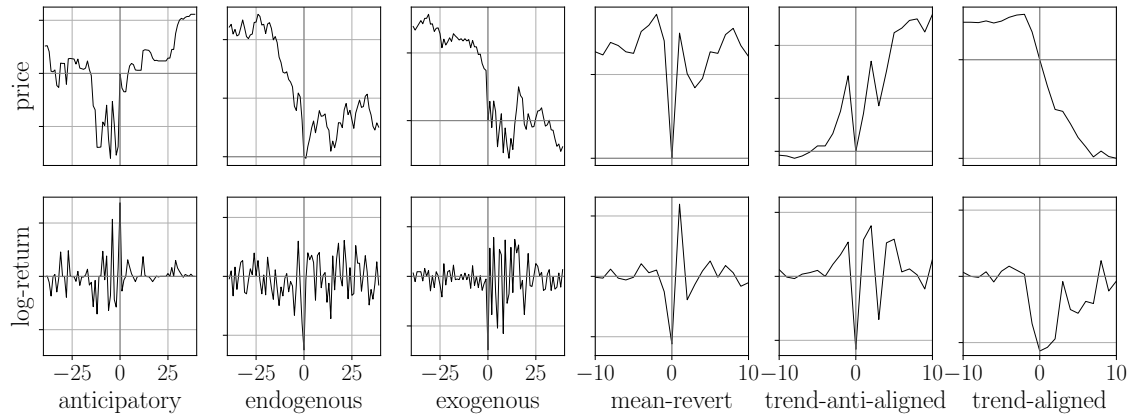
Figure 5.2: Visualization of our co-jumps dataset (295 US stocks, 8 years) (as in [54, 55, 56]). The horizontal axis corresponds to the day of the co-jump and the vertical axis gives the time of day. The size and color of the circle encode the number of stocks jumping simultaneously (in the same minute). Inset: number of jumps on a rolling window of 30 days.

Furthermore, in complex systems, occurrences can propagate along two directions: temporally and towards the other elements of the system. Financial markets offer an attractive setting for studying multi-dimensional shocks due to the abundance of available data, the frequent occurrence of financial shocks and price jumps and the inter-connectivity of markets. In fact, a recent study by Lillo *et al.* [54, 55] demonstrates the frequent occurrence of "co-jumps", defined as simultaneous jumps of multiple stocks (as illustrated in Figure 5.2) and establishes a correlation between their prevalence and the inter-connectivity of different markets.

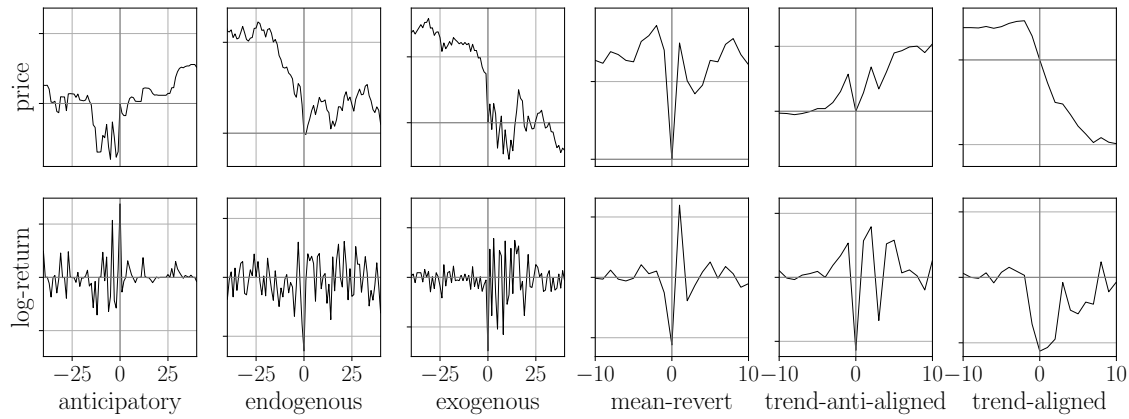
In this paper, we address the problem of classifying financial price jumps (and co-jumps), in particular measuring their reflexivity, by analyzing their time-series using wavelets. We introduce an unsupervised classification based on an embedding $\Phi(x)$ of each jump time-series of returns $x(t)$ into a low dimensional-space more appropriate to clustering. Such embedding, composed of wavelet scattering coefficients (see [101] and below), relies on wavelet coefficients $Wx(0)$ of the time-series at the time of the jump $t = 0$ and wavelet coefficients of volatility $W|Wx|(0)$. Such coefficients are particularly suitable to characterize (among other properties) the asymmetry of time-series at multiple scales.

Through a Principal Component Analysis we retrieve the fact that time-asymmetry of volatility indeed plays an important role for classification. However, our analysis identifies two further crucial features for characterizing the nature of price jumps: mean-reversion and trend. Specifically, mean-reverting jumps are such that pre-jump and post-jump returns are of opposite signs, whereas trend-aligned and trend-anti-aligned jumps occur on a sequence of returns of same sign before and after the jump, but either aligned with the jump itself, or of opposite sign.

For each jump, our analysis provides a measure of the volatility asymmetry, the mean-reversion and the trend. We propose a visualization of our dataset of price jumps in the form of *two* 2D projections. For both projections, one direction characterizes price jumps based on volatility asymmetry, or "reflexivity level". The second direction characterizes jumps either in terms of mean-reversion, or in terms of alignement with the local trend behavior. One can then measure the endogeneity of price co-jumps, revealing that many jumps/co-jumps are *not* related to news and arise only due to endogenous dynamics. This is consistent with the observed power-law distribution of the number of firms affected by a co-jump, indeed predicted by a simple branching (or contagion) process.



(a) Synthetic examples



(b) Observed examples

Figure 5.3: Classes of price jumps: synthetic (top) and observed (bottom) examples. Each column shows an example of a class of jumps (price and log-return time-series). The three first classes (anticipatory, endogenous, exogenous) are separated by measuring volatility asymmetry. The three last classes (mean-reverting, trend-anti-aligned, trend-aligned) are identified by analyzing the signed returns around the jump.

Surprisingly, we uncover that a significant number of large co-jumps (affecting a large number of stocks), which might have been assumed to be caused by a common factor and thus share analogous dynamics, actually have uncorrelated returns both pre- and post-jump. This again suggests that such jumps are mostly of endogenous origin.

The outline of our paper is as follows. Section 5.1 describes our dataset of price jumps resulting from Marcaccioli *et al.* [6], reviews their supervised classification method based on news labels, and investigates its limitation. Section 5.2 presents our unsupervised classification of univariate jump time-series based on wavelet coefficients. Such classification identifies three main directions in the dataset, the time-asymmetry, the mean-reversion and the trend. Finally, section 5.3 is devoted to the characterization of the endogeneity of co-jumps.

5.1 Supervised classification through reflexivity

Prior work has identified *reflexivity* as an important feature for the classification of jumps in financial markets [6]. Given the time-series of a jump, the main challenge is to efficiently measure such reflexivity.

One can for example look at contemporaneous news labels to determine whether a jump is exogenous. Indeed, news labels may serve as ground truth to learn a classification model on the activity profile around a shock. To exemplify, Figure 5.1, from the work of Marcaccioli *et al.* [6], illustrates the time asymmetry difference between endogenous and exogenous jumps.

In this section, we first introduce the jump detection method, which allows us to build our

dataset. Then, we present the supervised classification based on news labels introduced in [6] and show its limitations. This will motivate an alternative approach in section 5.2.

5.1.1 Jump detection

We refer to [99, 6, 102] for a detailed description of the method to detect price jumps. The detection relies on an estimator of "jump-score" $x(t) = r(t)/(f(t)\sigma(t))$, which is the ratio of 1-minute returns time-series $r(t)$ and de-seasonalized local volatility $f(t)\sigma(t)$ where $\sigma(t)$ is an estimator of local volatility and $f(t)$ an estimator of the intraday periodicity (the so-called "U-shape"). Throughout this paper, our statistical analyses will focus on $x(t)$, or on its "jump-aligned" version $\overline{x}(t) := x(t)\text{sign}(x(0))$, where $x(0)$ is the return corresponding to the jump. In other words, $\overline{x}(t)$ is the rescaled return profile in the direction of the jump.

Under the null hypothesis of Gaussian residuals (no jump hypothesis) $|x(t)|$ converges towards a Gumbel distribution. A statistical test then allows us to reject the null hypothesis. The resulting method comes down to detecting price movements where the z-score deviates by more than 4-sigma from their average value (here equal to zero).

The jump detection is performed on time-series describing individual stocks dynamics but also on averaged time-series across stocks belonging to the same sector. Hence, we obtain price jumps of individual stocks but also sectoral price jumps.

Similarly to Marcaccioli *et al.* [6], we find that price jumps are clustered in time. We assume that jumps taking place within the same "time-cluster" subsequent to an initial jump are merely replicas of the initial jump. They are likely to be either of the same dynamics (as they occurred for the same reason) or endogenously induced by the first jump of the cluster. We thus discard all the jumps that follows an initial jump. This leads to the same detection method as in [6] which allows to retrieve an exponential distribution for the inter time between two consecutive initial jumps (see part II.D of [6]).

From such a collection of price jumps, we can then extract "co-jumps". A co-jump is simply defined as a set of jumps occurring in the same minute. Here we avoid tackling the question of lagged jumps and consider only simultaneous jumps (up to the minute resolution).

The price behavior before and after a jump can be used to classify the jump. In light of Marcaccioli *et al.*'s findings [6], which indicate that volatility can begin to rise up to 75 minutes prior to the jump, we adopt a time window of 2 hours centered around the jump occurrence at time $t = 0$. Consequently, for each jump we extract a time-series of 119 rescaled returns $x(t)$, corresponding to 1 hour preceding the jump and 1 hour following the jump.

We implement such detection on 301 US stocks from January 2015 to December 2022, considering only what happened between 10:30 and 15:00 in order to avoid special jumps due to the high activity at the beginning (due to people reacting to the overnight news/movement) and at the end of the day (due to market closing). In order to discard major market shocks, we also remove all co-jumps involving more than 250 stocks, and days on which the FED made an announcement (1 per month[1]). We end up with 37 452 jumps, of which 16 127 belong to one of the 2534 co-jumps, and the remainder (21 325) are single jumps.

5.1.2 Classification based on news labels

In an attempt to characterize the reflexivity of a jump, one can gather the date and time of news associated to each stocks we consider[2] and of the main US announcements[3]. According to such news labels, we might label as "news-related" a jump which happened within 3 minutes of a news and label as "non-news-related" any other jump. That would lead to a puny $\approx 4.3\%$ of the jumps being classified as "news-related" and is illustrated in Figure 1. Hence, as previously argued in [99, 54, 6], it appears that individual price jumps and more surprisingly co-jumps are often *not* related to news announcements.

However, it is clear that some news may affect a whole economic sector and lead to a co-jump without appearing in our considered set of news. An example would be an OPEC announcement that affects oil prices and in turn ricochets onto stocks prices, without any of them explicitly showing up in the news feed. Another vivid example is the impeachment of the US president D.

[1] see FOMC Calendars
[2] source: Bloomberg
[3] source: economic-calendar

Trump in September 2019[4]. Our "news-related" label is blind to such events. One objective of our study will be to propose a possible classification of co-jumps that does not rely on the news feed, see section 5.3.

5.1.3 Classification based on the volatility profile

In [6], Marcaccioli *et al.* built a supervised classification of univariate jumps into exogenous and endogenous classes. The classification relies on parameters derived from fitting $|x(t)|$ to the following functional form [21]:

$$|x(t)| = \mathbf{1}_{t<t_c} \frac{N_<}{|t-t_c|^{p_<}} + \mathbf{1}_{t>t_c} \frac{N_>}{|t-t_c|^{p_>}} + d \tag{5.1.1}$$

and on a measure of the asymmetry of the jump, defined as:

$$\mathscr{A}_{\text{jump}} = \frac{\mathscr{A}_> - \mathscr{A}_<}{\mathscr{A}_> + \mathscr{A}_<} \tag{5.1.2}$$

where $\mathscr{A}_{</>} := \sum_{t<0/t>0} |x(t) - \min_{t<0/t>0}(x(t))|$. Such an indicator means that when the activity is stronger before (resp. after) the jump, one has $\mathscr{A}_{\text{jump}} < 0$ (resp. $\mathscr{A}_{\text{jump}} > 0$). The classification is then obtained as a logistic regression of the news label (endogenous/exogenous) by the parameters $(\mathscr{A}_{\text{jump}}, p_<, p_>, N_<, N_>, t_c)$. Exogenous jumps appear as strongly asymmetric jumps with little activity ahead of the jump, i.e. $\mathscr{A}_{\text{jump}} > 0$, whereas self-exciting endogenous jumps are much more symmetric with $\mathscr{A}_{\text{jump}} \approx 0$ [6].

The above approach, based on news labels, presents several limitations:

- The classification partly relies on the goodness of fit of a power-law function (5.1.1), which is not assured. As a consequence, Marccacioli et al. [6] restrict their study to only ~ 5000 jumps out of the ~ 37000 in the dataset, for which such a fit is acceptable.

- As discussed above, news labels might miss some relevant economic news, so the resulting price jumps might be wrongly labeled as "non news-related".

- Exogenous jumps could have two types of dynamics: if the exogenous shock is a complete surprise, there should indeed be no activity before the jump. However, if the announcement is planned or if there was some news leakage, there might be a growth of activity before the jump. In this case, one would wrongly classify a news-related jump as endogenous based on its approximately symmetric activity profile.

In light of such limitations and in order to uncover new classes of jumps, beyond the sole study of their reflexivity, we opt in the rest of the paper for an unsupervised classification which significantly improves upon the method of [6] while still leaving open some ambiguities, as we will see below.

In the following, although news labels do not reveal the whole truth about the reflexive nature of a jump, we will still call "news-related" jumps that occurred within 3 minutes of a news present in our database and "non news-related" all the others.

5.2 Classification of single jumps using wavelets

The rescaled return time-series around a jump $x(t) \in \mathbb{R}^T$ is inherently noisy. Relevant features $\Phi(x) \in \mathbb{R}^q$ must be extracted to effectively distinguish different classes of jumps. Such features should be selected carefully, in particular, they should include time-asymmetry measures. Indeed, authors in [98, 20, 21, 22, 6] show that the jumps mostly differ in their time-asymmetry: endogenous jumps tend to be more symmetric around the jump than exogenous ones. But what are the other possibly relevant features? In this section, we embrace a signal processing approach to discover important features of univariate jumps and unveil new classes of jumps that are prevalent in the data.

[4]For example, the largest co-jump is related to Nancy Pelosi announcement of a formal impeachment inquiry into US President Donald Trump. On 2019-09-24, at 14:13, 248 stocks saw their price jump in the same minute.

5.2.1 Some intuition first

Before delving into the technical formulation, we aim to provide some intuition on why wavelets are particularly well-suited for capturing the asymmetry of a signal. Figure 5.4 offers a simplified representation of the computation of wavelet coefficients, with the imaginary part of the Battle-Lemarié wavelet (in blue), for two types of jumps (whose volatility profiles are presented in black): a symmetric jump (left) and an asymmetric jump (right).

The computation of a wavelet coefficient relies on the convolution between the jump and the wavelet. Specifically, the wavelet coefficients at time $t = 0$ are computed by multiplying the jump time-series and the wavelet centered at $t = 0$ and by summing the result.

Given the form of the Battle-Lemarié wavelet, the areas which contribute the most to the integration are those around $t = 0$. Focusing on these areas, Figure 5.4 highlights the positive contributions in green and the negative contributions in red. For a symmetric jump, the wavelet coefficient will be zero, as the positive and negative contributions cancel each other out. In contrast, for an asymmetric jump, the positive contributions will significantly outweigh the negative contributions, resulting in a positive wavelet coefficient.

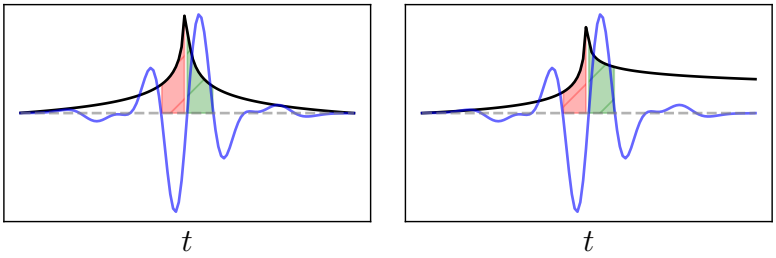


Figure 5.4: Illustration of reflexivity score computed with wavelets. The black plain lines represent the volatility profile of a jump (symmetric in the left figure and asymmetric in the right figure). The blue line is the imaginary Battle-Lemarié wavelet (see Figure 5.5). The green (red) hashed areas represent the positive (negative) contributions to the convolution between the wavelet and the jump at time $t = 0$. Adding the red and green areas shows that the imaginary part of the wavelet coefficient detects time-asymmetry around $t = 0$.

We now give a formal presentation of the concept of wavelet coefficients.

5.2.2 Wavelet and scattering coefficients

Wavelet filters have been used to analyze and classify transient events, see e.g. [103, 104, 105, 106]. A complex wavelet filter $\psi(t)$ is a filter whose Fourier transform $\widehat{\psi}(\omega) = \int \psi(t)\, e^{-i\omega t}\, \mathrm{d}t$, is real. It is localized both in time and Fourier domains, see Figure 5.5.

It has a fast decay away from $t = 0$ and a zero-average $\int \psi(t)\, \mathrm{d}t = 0$. We write $\psi(t) = \mathrm{Re}\,\psi(t) + i\,\mathrm{Im}\,\psi(t)$ where $\mathrm{Re}\,\psi(t)$ and $\mathrm{Im}\,\psi(t)$ are its real and imaginary parts. They are respectively even and odd functions:

$$\mathrm{Re}\,\psi(-t) = \mathrm{Re}\,\psi(t) \quad \text{and} \quad \mathrm{Im}\,\psi(-t) = -\mathrm{Im}\,\psi(t). \tag{5.2.1}$$

The wavelet coefficients $W_j x(t)$ compute the variations of the signal x around t at scale 2^j, for $j = 1, \ldots, J$ with

$$W_j x(t) := x \star \psi_j(t) \quad \text{where} \quad \psi_j(t) = \psi(2^{-j}t). \tag{5.2.2}$$

where $\star$ denotes the convolution: $x \star y(t) := \int x(t-u)y(-u)\, \mathrm{d}u$.

The sign of the jump $\mathrm{sign}(x(0))$ and its amplitude $|x(0)|$ vary, but they are not necessarily informative for their classification. To remove this source of variability we consider the *jump-aligned* time-series

$$\overline{x}(t) = \mathrm{sign}(x(0))\, x(t) \tag{5.2.3}$$

and we further normalize the wavelet coefficients (5.2.2) by the corresponding "volatility" σ_j of the full time-series, defined as $\sigma_j^2 = \big\langle |x \star \psi_j(t)|^2 \big\rangle_t$, where $\langle \cdot \rangle_t$ denotes the empirical average over time t.

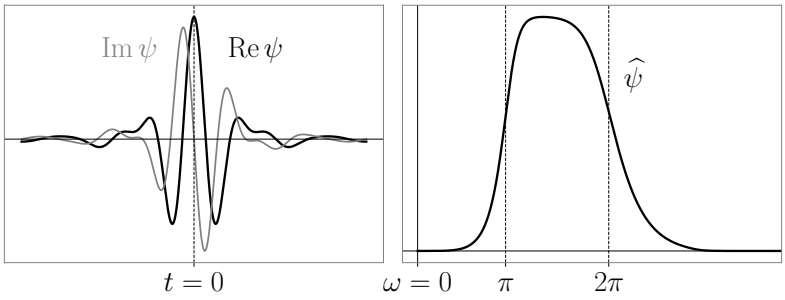


Figure 5.5: Filter used to analyze jump time-series. Left: complex Battle-Lemarié wavelet $\psi(t)$ as a function of t. Right: Fourier transform $\widehat{\psi}(\omega)$ as a function of ω.

From Equation (5.2.1), one can see that if x is an even signal i.e. $x(-t) = x(t)$ then $\operatorname{Im} Wx(t, j) \equiv 0$. This property is key to detect asymmetry of a signal at different scales.

Volatility information can be extracted by taking a modulus. The time-series $|W_j x(t)|$ provides the volatility of the signal at scale 2^j. This volatility can be asymmetrical in $t = 0$. In order to quantify it, we again consider the wavelet coefficients at $t = 0$

$$W_{j_2}|W_{j_1}x|(t) := |x \star \psi_{j_1}| \star \psi_{j_2}(t). \tag{5.2.4}$$

Our representation for univariate jumps in this paper is thus composed of wavelet coefficients (5.2.2) at $t = 0$ and scattering coefficients (5.2.4) at $t = 0$

$$\Phi(x) = \Big(W_j \overline{x}(0)\,,\, W_{j_2}|W_{j_1}x|(0) \Big). \tag{5.2.5}$$

For a time-series of size T, it contains less than $(\log_2 T)^2/2$ coefficients which represents few coefficients. In our case, $T = 119$ and we chose $J = 6$, which yields 42 coefficients (21 real parts and 21 imaginary parts). The normalized scattering features $\Phi(x)$ (Equation (5.2.5)) are invariant to sign change and to dilation

$$\Phi(-x) = \Phi(x) \quad \text{and} \quad \Phi(\lambda x) = \Phi(x).$$

which means we do not aim at discriminating jumps neither based on their sign nor on their amplitude.

In order to classify price jumps, we are interested in Principal Component directions of the 42-dimensional vector $\Phi(x)$ in the dataset. This method, called kernel PCA [107], relies on the linear separation power of our scattering coefficients $\Phi(x)$. We considered several directions, i.e. combinations of scattering coefficients, and found three salient features: the time-asymmetry of the volatility, the mean-reversion and the trend behavior of the price around the jump.

5.2.3 First Direction D_1: Volatility asymmetry

Three types of jumps

The first PCA direction (called D_1 henceforth) is a linear combination of the 15 coefficients $\operatorname{Im} W_{j_2}|W_{j_1}x|(0)$ in Equation (5.2.5), which characterizes time-asymmetry of the *volatility profile* at multiple scales 2^{j_2}, confirming previous analysis that postulated this asymmetry to be relevant. Such a linear combination allows one to embed each jump time-series into a one dimensional space, which quantifies the reflexive nature of each jump. In fact, Figures 5.6 and 5.13 display average profiles $|x(t)|$ along the "reflexive direction" D_1. One can visually verify that such a representation discriminates jumps according to the asymmetry of their profiles as measured by $\mathscr{A}_{\text{jump}}$ (Equation (5.1.2)): the D_1 direction continuously separates asymmetric jumps with dominant activity before the shock from asymmetric jumps with dominant activity after the shock; see Figs. 5.6, 5.13 and 5.17.

From this analysis, three types of jumps can thus be defined:

- Asymmetric jumps with dominant activity *before* the shock. This type of jumps, which we call "anticipatory", was quite unexpected and was not discussed in [6].

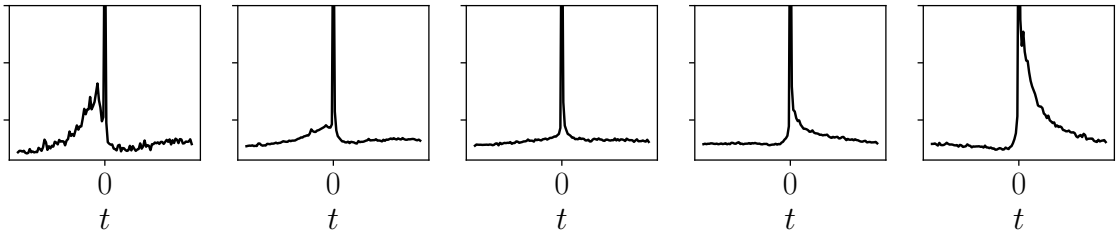


Figure 5.6: Average absolute profiles $|x(t)|$ of jumps along direction D_1 (sliced into five bins, delimited by quantiles 0.1, 0.25, 0.35, 0.9). From left to right: anticipatory jumps, endogenous jumps and exogenous jumps.

- Symmetric jumps, with an pre-shock excitation activity that approximately mirrors the post-shock relaxation activity. These were called "endogenous jumps" in [6]: increased activity before the jump is in fact responsible for the jump itself, with some decay of activity thereafter. The symmetry of the profile for endogenous jumps is in fact predicted by a Hawkes process description of the self-exciting mechanism, see [21, 6].
- Asymmetric jumps with dominant activity after the shock. These were called "exogenous jumps" in [6]: the market reacts *after* unexpected news, but not before.

In order to validate the above analysis, we created synthetic time-series with volatility profiles of varying time-asymmetry and applied our classification method. Results of this benchmark case are shown in Appendix C.1, and fully confirm that the D_1 direction indeed separates jumps according to their asymmetry $\mathscr{A}_{\text{jump}}$.

Discussion

Using the above classification, we find that a large proportion ($\sim 50\%$) of our sample exhibit positive asymmetry and should naively be considered as exogenous jumps. This seems in contradiction with the results of [6], where exogenous jumps were found to be a minority, and with a fraction of jumps associated to a news found to be 4.3%, as already quoted above. Several arguments can explain such a difference.

- The main one is the fact that our analysis includes all jumps involved in a sector jump (corresponding to 24% of all jumps) whereas those jumps were discarded in [6]. Sector jumps are such that many stocks of the same industry jump simultaneously. While some of these jumps are likely due to major exogenous shocks – like macro-economic announcements – that affect a whole economic sector or even the whole market, we argue in section 5.3 that these jumps can actually be induced by a jump of one particular stock of the sector, which is deemed as "news" in and by itself. In any case, taking these sector jumps into account mechanically increases the count of jumps with a positive D_1 score. In the present study, we chose to keep these co-jumps and study their statistics, to which we will specifically turn in section 5.3.
- As already noted above, the classification of single jump profiles in [6] relies on the goodness of fit of power law function (5.1.1), and as such, was only conducted on a smaller sample for which such a fit is acceptable (~ 5000 jumps out of ~ 37000 jumps).

The appearance of "anticipatory jumps", where the asymmetry parameter $\mathscr{A}_{\text{jump}}$ (see Equation (5.1.2)) is negative, came somewhat as a surprise to us. One possible interpretation is that these jumps are in fact also endogenous, with a pre-shock self-exciting dynamics and very little "after-shocks". Indeed, if such jumps are immediately deemed endogenous by the market, it might make sense that activity quickly reverts back to normal. This would simply mean that the Hawkes framework predicting a symmetric profile is not adapted to describe all endogenous shocks.

Another possibility is that such events correspond to news/exogenous events whose *timing* is expected by the market, which leads to increased activity before the actual release time. But if the actual news content turns out to be insignificant, it would again make sense that the market activity quickly wanes off. We in fact find a very small fraction of news-related jumps with $D_1 < 0$, see in Figure 5.10, bottom graph.

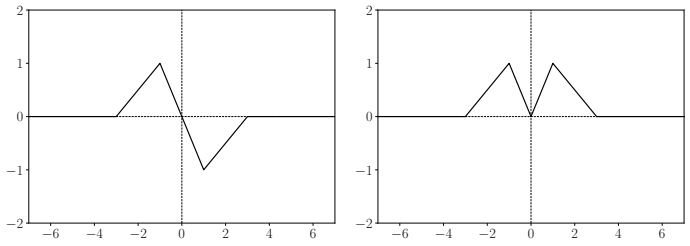
5.2.4 Second Direction D_2: Mean-Reversion

Capturing Mean-Reversion

We observed that coefficients $\operatorname{Im} W_{j_1}\overline{x}(0)$ (5.2.5) for fine scales, i.e. small j_1, are consistently chosen by the leading PCA directions. They amount to multiplying the jump-aligned time-series $\overline{x}(t)$ by the imaginary filter $\operatorname{Im}\psi_1(t)$ (see Figure 5.5) and averaging over t. Such coefficients capture the asymmetry of the return profile shortly before and shortly after the jump, and define what we will call below direction D_2.

A typical time-series that maximizes this coefficient is thus characterized by a positive value of $\overline{x}(-1)$ and a negative value of $\overline{x}(1)$. In other words, large positive values along the D_2 coordinate capture mean-reverting return profiles, i.e. positive (resp. negative) returns before a positive (resp. negative) jump that become negative (resp. positive) immediately after the jump.

Large negative values along the D_2 coordinate, on the other hand, also capture mean-reverting return profiles, but in this case mean-reversion starts with (or is triggered by?) the jump itself, and not after the jump.



(a) ψ_{MR} (b) ψ_{TR}

Figure 5.7: Handcrafted filters for measuring the mean-reversion (filter ψ_{MR}) or the trend (filter ψ_{TR}) character of a jump. Average profiles along resulting mean-reversion and trend directions are shown in Figure 5.8 and Figure 5.14.

Now that we identified a potentially discriminating direction using PCA, we transition to a simpler filter tailored to capture short time mean-reversion, depicted in Figure 5.7. This filter is then applied to the jump-aligned time-series $\overline{x}(t)$

$$\widetilde{D}_2(x) := \overline{x} \star \psi_{\mathrm{MR}}(0), \tag{5.2.6}$$

where the tilde indicates that we have simplified the true second PCA direction and only retained the component spanned by ψ_{MR}.

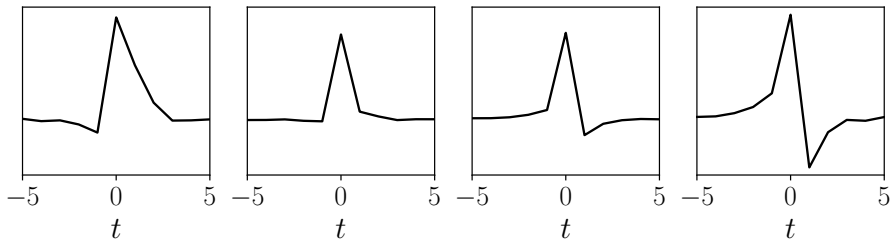
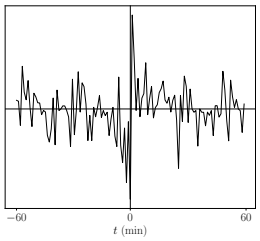


Figure 5.8: Mean-reverting profiles. Average jump-aligned return profiles $\overline{x}(t) = \operatorname{sign}(x(0))\, x(t)$ along the mean-reverting direction $\widetilde{D}_2$ (sliced into four bins, delimited by quantiles 0.1, 0.5 & 0.9). Left-most graph: price jumps mean-revert on previous trends. Right-most graph: prices mean-revert after the jump.

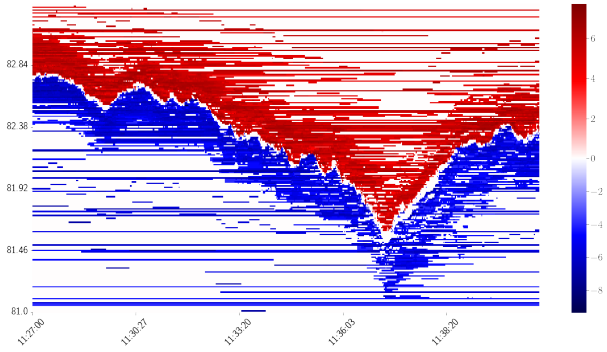
In order to illustrate the discriminating power of such coefficient, Figure 5.8 displays the average profiles of $\overline{x}(t)$ along the $\widetilde{D}_2$ axis. One can see that jumps with a high coefficient $\widetilde{D}_2$ (rightmost graph) are characterized by a strong pre-jump trend aligned with the jump, followed by a change of sign in the next minute *after* the jump.

The leftmost graph, on the other hand, shows relatively mild pre-jump trends opposite to the jump, followed by stronger trends in the direction of the jump, not very different from the cases

corresponding to quantiles between 0.1 and 0.5. In our dataset, 60% of the jumps have a positive mean-reversion score $D_2 > 0$; we refer to Figure 5.15 for the full distribution of D_2.



(a) Jump-aligned return profile $\overline{x}(t) = \text{sign}(x(0))\, x(t)$.



(b) Respective LOB illustration.

Figure 5.9: LOB illustration of the jump which occurred on stock HCA (Hospital Corporation of America), on the 2017-3-21 at 11h37. mean-reverting score: 2.61.

To confirm this observation and ascertain that it is not attributable to spurious effects in the data processing, we looked deeper into these jumps. To get a better understanding of the mechanisms at play, we investigated what happens at tick-by-tick scale in the Limit Order Book. Figures 5.9 shows an illustrative example, where the right figure depicts the profile of the jump, and the left one the respective Limit Order Book illustration (for the left plot, each colored square represents an order in the LOB whose price is referred on the y-axis. The x-axis describes the time. Red is for the ask side, blue for the bid side. The color bar depicts the size of the order and is in log scale). We again observe, at a different time resolution, a strong mean-reversion behavior induced by order placement.

Note finally that mean-reversion is characterized by a V-shape price profile (see Figure 5.9), which has recently been used as a criterion to detect price jumps in time-series [24].

A 2D representation of jumps

Based on the first volatility asymmetry direction D_1 and the mean-reversion direction $\widetilde{D}_2$, we are in a position to propose the 2D representation of jumps shown in Figure 5.10 (top), in which the horizontal axis corresponds to D_1 and the vertical axis corresponds to the mean-reversion index $\widetilde{D}_2$. Average profiles of $\bar{x}(t)$ and $|x(t)|$, through the grid are shown in Figure 5.12 and, respectfully, Figure 5.13.

Visually, news-related jumps are mostly to the right of the projection, corresponding to increased volatility after the jump, as expected.

Note that both exogenous, or endogenous jumps can have such mean reverting behavior, as clear from the 2D representation Figure 5.10. In fact, a mean reverting behavior can be expected both following an exaggerated response to a news release, or after a self-initiated jump with no discernible catalyst. This is confirmed by Figure 5.11 which shows positive average values of $\widetilde{D}_2$ for all levels of reflexivity D_1, except for strongly exogenous jumps (large values of $D_1 > 0$), where the mean-reversion disappears ($\widetilde{D}_2 \approx 0$).

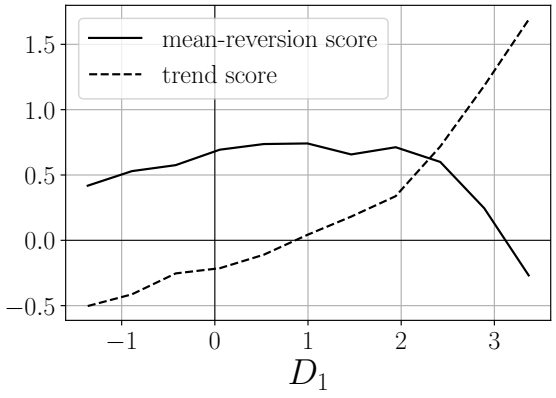


Figure 5.11: Mean-reversion and trend scores along reflexivity of a jump.

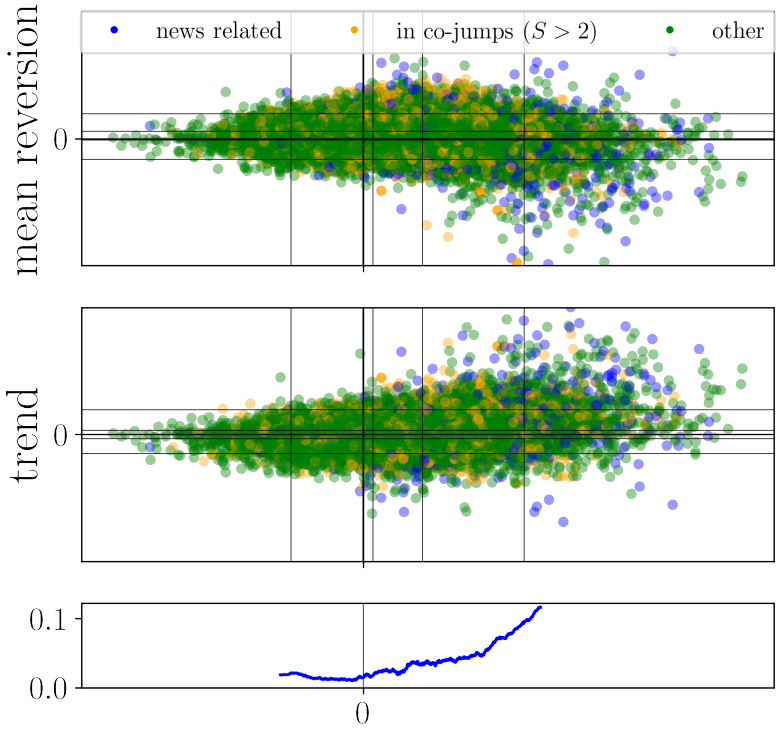


Figure 5.10: Top graph: Projection of jumps in our dataset onto the reflexive direction D_1 (horizontal axis) and mean-reverting direction $\widetilde{D}_2$ (vertical axis) Middle graph: Projection of our dataset on the reflexive direction D_1 (horizontal axis) and trend direction $\widetilde{D}_3$ (vertical axis). Each point represents a jump, the blue color corresponds to news-related jumps according to the classification of Section 5.1.2, the oranges are jumps involved in a co-jump of size greater than 2 and non news related and the greens are all the other jumps. The vertical and horizontal lines represent the following quantiles: 0.05, 0.35, 0.65, 0.95. Bottom graph: ratio of "news-related" jumps along the reflexive direction D_1, based on a direct classification using the news feed (rolling ratio every 2000 jumps).

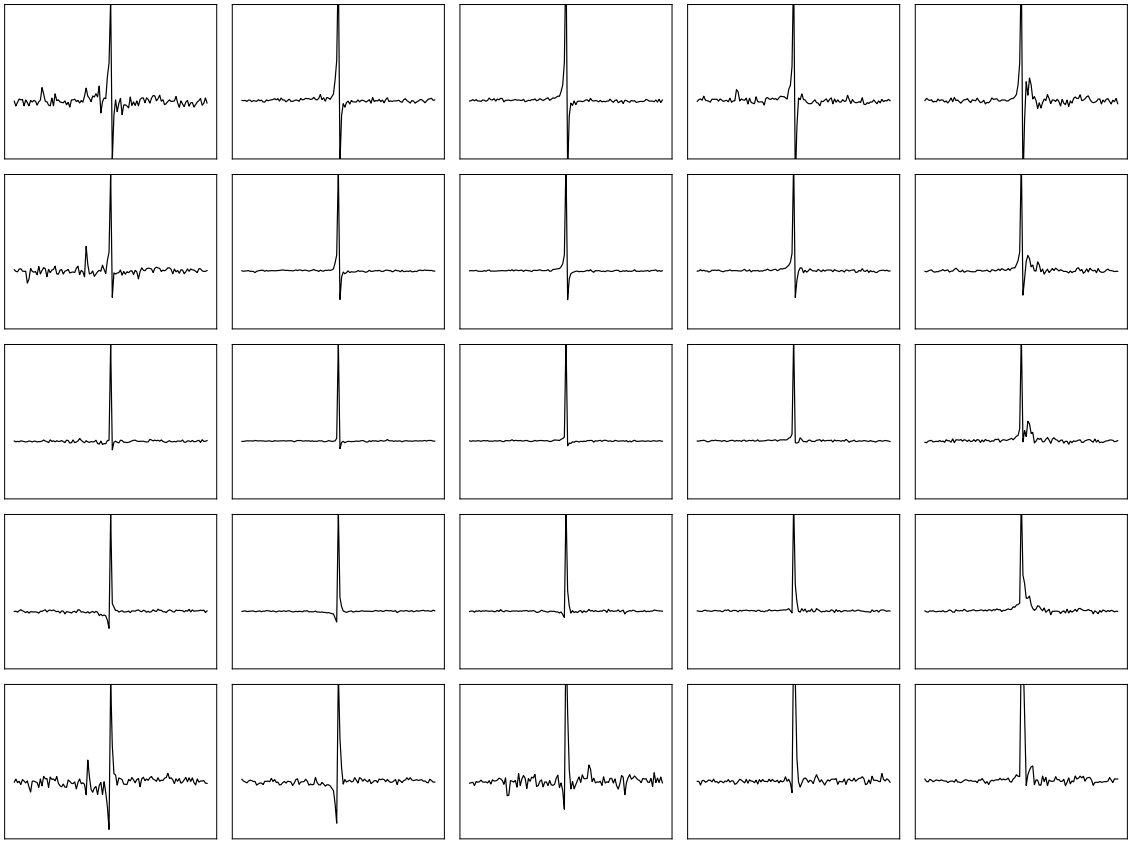


Figure 5.12: Average jump-aligned return profiles $\overline{x}(t) = \text{sign}(x(0))\, x(t)$. Each plot represents the average over the jumps whose 2D projection falls in the respective box in the upper figure in Figure 5.10.

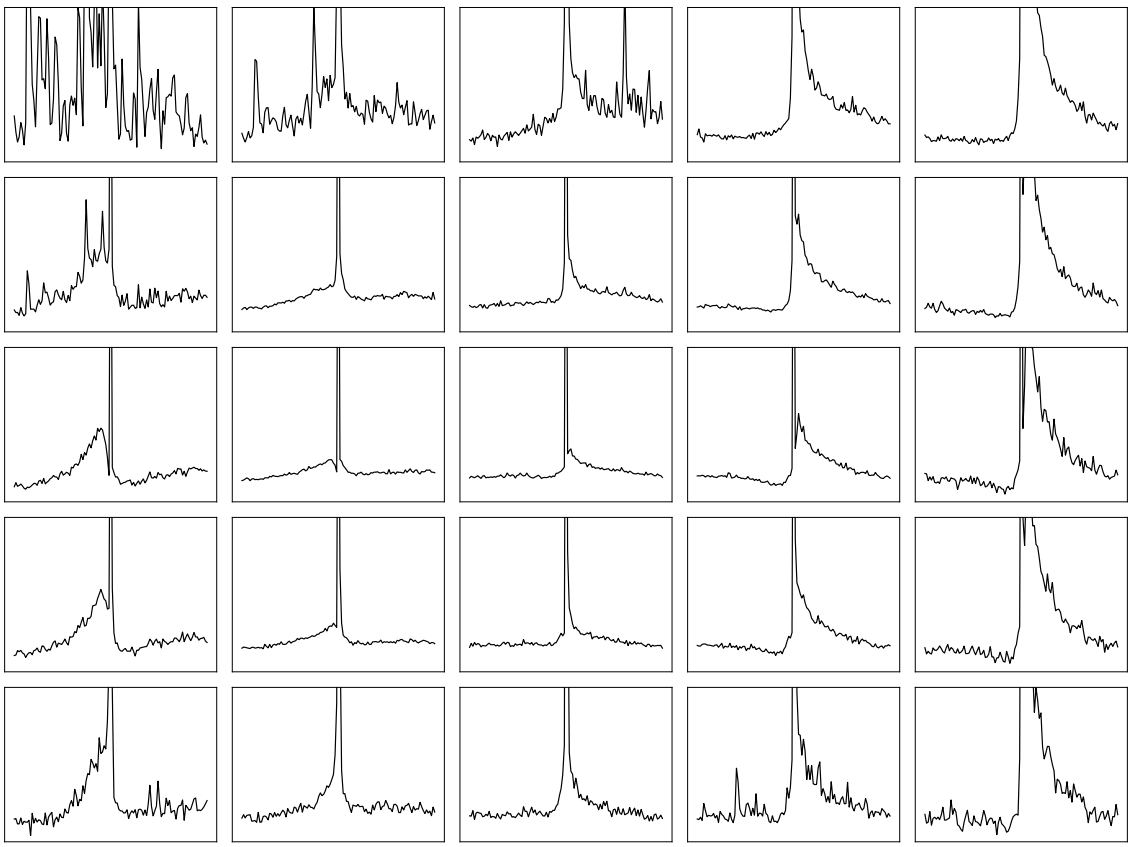


Figure 5.13: Average absolute profiles $|x(t)|$. Each plot represents the average over the jumps whose 2D projection falls in the respective box in the upper figure of Figure 5.10.

5.2.5 Third Direction D_3: Trend

In the previous section, we have defined a filter ψ_{MR} that detects mean-reversion, but is by construction orthogonal to trends, i.e. post-jump returns continuing in the same direction as pre-jump returns. This feature can be naturally captured by the trend filter ψ_{TR} shown in Figure 5.7, which is orthogonal to the mean-reversion filter ψ_{MR}. This filter is then applied to the jump-aligned profile $\overline{x}(t)$ to get the following trend score

$$\widetilde{D}_3(x) := \overline{x} \star \psi_{\text{TR}}(0). \tag{5.2.7}$$

A large positive value of $\widetilde{D}_3(x)$ therefore describes a persistent trend aligned with the direction of the jump. If such jumps exist, we refer to them as "trend-aligned" jumps. A large negative value of $\widetilde{D}_3(x)$ indicates that the jump goes against the pre- and post-jump trend. If such jumps exist, we refer to them as "trend-anti-aligned" jumps.

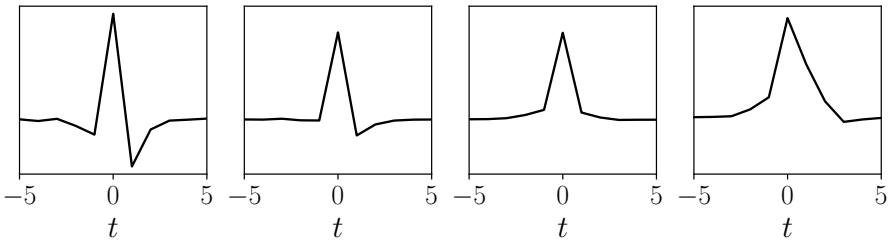
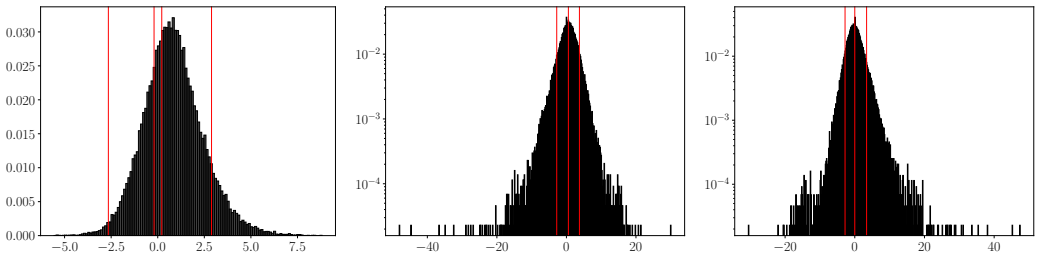


Figure 5.14: Trending profiles. Average jump-aligned return profiles $\overline{x}(t) = \text{sign}(x(0))\, x(t)$ along the trend direction $\widetilde{D}_3$ (again sliced into four bins, delimited by quantiles 0.1, 0.5 & 0.9). Left-most graph: anti-aligned trends. Right-most graph: aligned trends.

Figure 5.14 shows that both classes of jumps do indeed exist: the average profiles in the first and last quantiles in Figure 5.14 do conform to expectations. Furthermore we directly observe many stylized examples such as the one reported in Figure 5.3b. As for the mean-reversion indicator, we can represent all jumps in 2D plane based on D_1 and $\widetilde{D}_3$ (see the bottom graph in Figure 5.10). Visually, trending news-related jumps appear to be mostly aligned with the jump (top-right corner), although anti-aligned trends can also be spotted for moderate values of D_1. Average profiles of $\overline{x}(t)$ through to the grid of Figure 5.10 are shown in Figure 5.16.



(a) Reflexivity (b) Mean-reversion (c) Trend

Figure 5.15: Distribution of reflexivity $D_1(x)$, mean-reversion $\widetilde{D}_2$ and trend $\widetilde{D}_3$ scores used in this paper to identify classes of jumps. The red vertical lines indicate the quantiles used to delimit the zones for the jumps taken into account when computing the average profiles in Figs. 5.6,5.8,5.14.

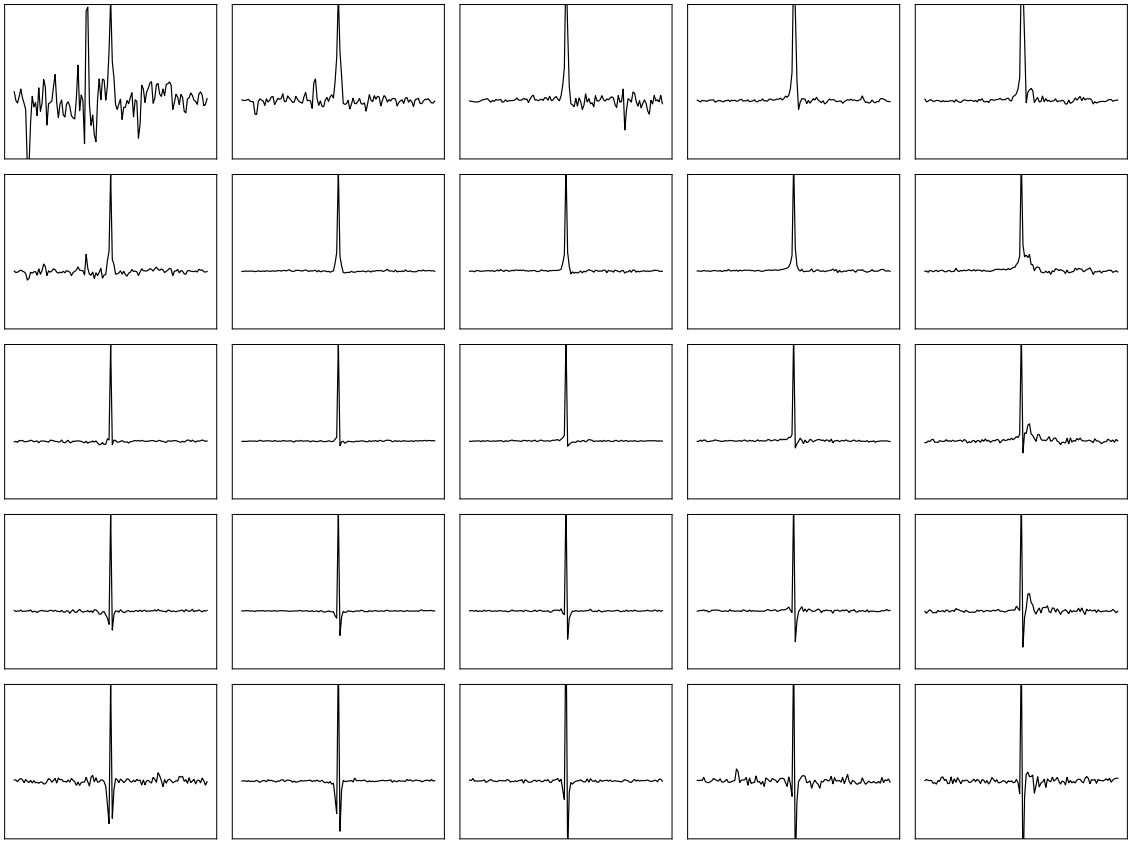


Figure 5.16: Average jump-aligned return profiles $\overline{x}(t) = \text{sign}(x(0))\, x(t)$. Each plot represents the average over the jumps whose 2D projection falls in the respective box in the lower figure of Figure 5.10.

5.2.6 Preliminary Conclusions

Let us summarize the results obtained by our unsupervised approach so far. First, our proposed 2D projections provide an embedding of a jump according to three meaningful, intuitive properties: its self-reflexive nature (along horizontal axis), its mean-reversion character or its trend character (along vertical axis). On top of the separation between exogenous and endogenous jumps, our clustering method revealed new classes of jumps, some of which we did not expect a priori: anticipatory jumps, mean-reverting jumps, trend-aligned and trend-anti-aligned jumps. Identifying additional interpretable classes of jumps might be possible by considering more expressive wavelet-based embeddings such as Scattering Spectra recently used in the context of financial time series [108, 109]. However, our attempts so far seemed to mostly recover directions which overlap with the volatility time-asymmetry and mean-reverting directions.

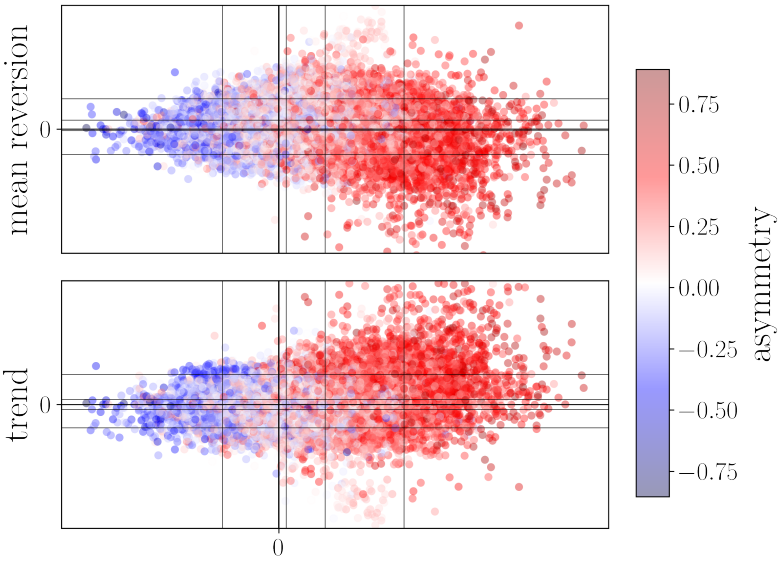


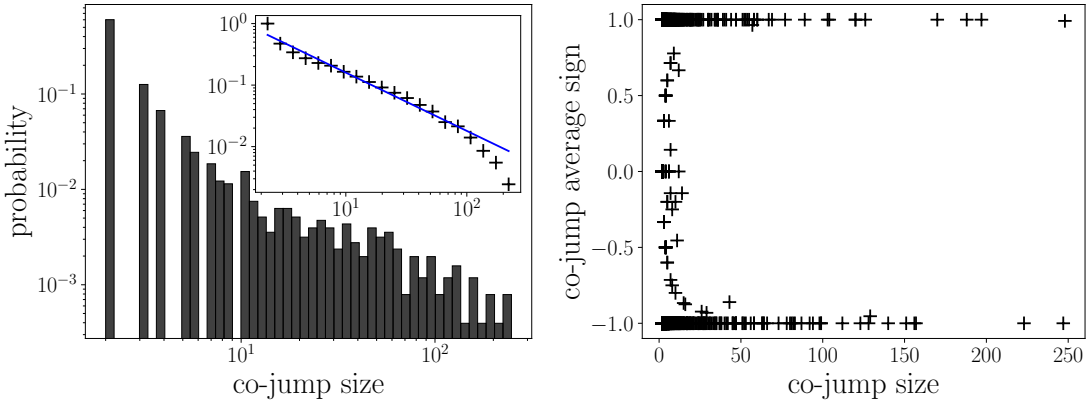
Figure 5.17: The 2D projections where the color represents the asymmetry of the jump computed with Equation (5.1.2).

5.3 Classification of co-jumps

A "co-jump" is defined as a collection of jumps across several stocks, occurring in the same minute. The number S of assets involved in the co-jump is referred to as the "size" of the co-jump. Co-jumps reveal inter-connectivity and contagion in financial markets [110, 54, 55]. As such, studying them – in particular their possible reflexive nature – is a crucial question for investors and regulators alike. This section aims at investigating whether co-jumps are created through endogenous dynamics or exogenous shocks.

To assemble our co-jump dataset we consider the same dataset of jumps as in the previous section. We end up with 2534 co-jumps, the size of which varies from 2 stocks to 248 stocks. The co-jumps cumulative size distribution, restricted to endogenous jumps, is shown in Figure 5.18a, inset. Quite remarkably, the tail of this distribution is well fitted by a power-law $S^{-\tau}$ with exponent $\tau \approx 1$, with a cut-off for $S \gtrsim 100$. As we discuss in Appendix C.2, such a value for τ can be rationalized within the framework of critical branching processes [111], as if co-jumps were the result of a contagion mechanism. Such a power-law behaviour was already noted in previous works: in Ref. [99] on a US data set from 2004 to 2006, in [55] from 2001 to 2013 and in [56] from 2013 to 2018.

The signs of the jumps involved in a co-jump are, most of the time, all aligned, i.e. different stocks jump in the same direction, as shown in Figure 5.18b.



(a) Co-jump size

(b) Co-jump sign

Figure 5.18: Statistics on co-jumps. (a) Main: Distribution of co-jumps size i.e. number of stocks involved in a co-jump. Inset: Cumulative distribution of co-jumps size for co-jumps with $\min(D_1) < 0$ and $\min(D_1) < \overline{D_1} - 1\sigma$, defining the LL and LR regions in Figure 5.22. The slope of the fit in log-log coordinate (plain line in blue) is $-\tau = -0.95$. Notice that the data bends down faster for large S. (b) Average sign of jumps involved in a co-jump, showing that most co-jumps are composed of jumps in the same direction.

The first stage of co-jump characterization is to classify jumps according to their reflexivity coordinate along the D_1 direction. In Figure 5.19, we highlight the coordinates of three particular co-jumps in the 2D projections introduced in the previous section. Each color point is a stock involved in one of the three co-jumps. Let us comment on each of these three cases in turn:

- The purple co-jump, with 29 stocks involved, has most of its elements in the right side of the 2D projection, suggesting an exogenous, news driven shock. However, one of the jump is below the 0.35 quantile and therefore appears endogenous. This might be a mis-classification because of the inherent noise in our D_1 reflexivity score. An alternative interpretation might however be that this particular stock jumped for no particular reason and this created a surprise to which other stocks reacted.

- The pink co-jump, with 19 stocks involved, staunchly belongs to the anticipatory class – which we believe to be of endogenous nature, as explained above. Co-jumps with a negative or positive but moderate maximum value of the D_1 score can thus be deemed endogenous.

- The yellow co-jump, with 9 stocks involved, has most of its elements in the intermediate "endogenous" region, except one which is classified as exogenous. This might be either again

a mis-classification because of the inherent noise in our D_1 reflexivity score, or else a stock that was not part of the anomalous pre-jump activity but is drawn into the jump through contagion.

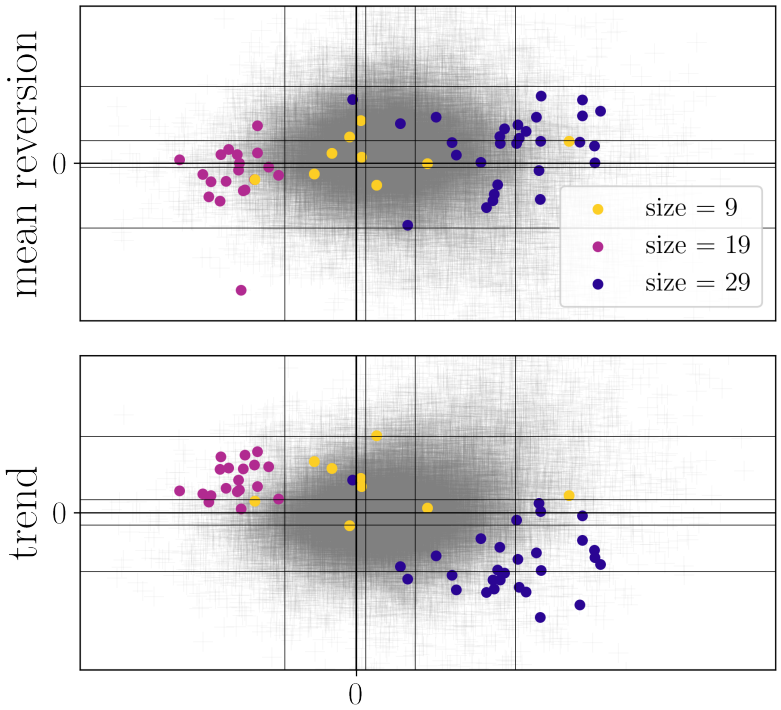


Figure 5.19: Projections of 3 co-jumps along our 2D projections. Yellow co-jump: one jump is exogenous and the others are more endogenous. Pink co-jump: all jumps of the co-jumps are endogenous and are trend-aligned. Purple co-jump: Most jumps appear to be exogenous except one. Those jumps are also trend-anti-aligned.

From these cursory observations, one may propose three natural indicators for classifying co-jumps:

1. The average value of the individual reflexivity score $\overline{D_1}$ over all jumps belonging to a given co-jump, see Figure 5.20.
2. The maximum value of the individual reflexivity score D_1 over all jumps belonging to a given co-jump: if the most exogenous jump is still deemed endogenous, the whole co-jump is classified as endogenous (see distribution in Figure 5.21b).
3. The minimum value of the individual reflexivity score D_1 over all jumps belonging to a given co-jump: if the most endogenous jump is still deemed exogenous, the whole co-jump is classified as exogenous (see distribution in Figure 5.21a).

Figure 5.22 represents the normalized minimum value of reflexivity score D_1 over all jumps of a given co-jump as a function of the normalized average value of reflexivity score D_1 over all jumps of a given co-jump (co-jump indicator 3 as a function of co-jump indicator 1). The normalization is such that Figure 5.22 can be read in units of standard deviation of the reflexivity score D_1 for co-jumps of same size, i.e. σ is the average of the standard deviation of the score D_1 over co-jumps with same size. The size and color of a point depict the size of the co-jump. The gray shaded region represents jumps with insignificant differences between the mean and the minimum value of the D_1 score.

Co-jumps with negative minimum and average values of reflexivity score D_1 (lower left quadrant of Figure 5.22, LL) can be deemed endogenous, whereas co-jumps with positive minimum and average values of reflexivity score D_1 (upper right quadrant of Figure 5.22, UR) can be deemed exogenous.

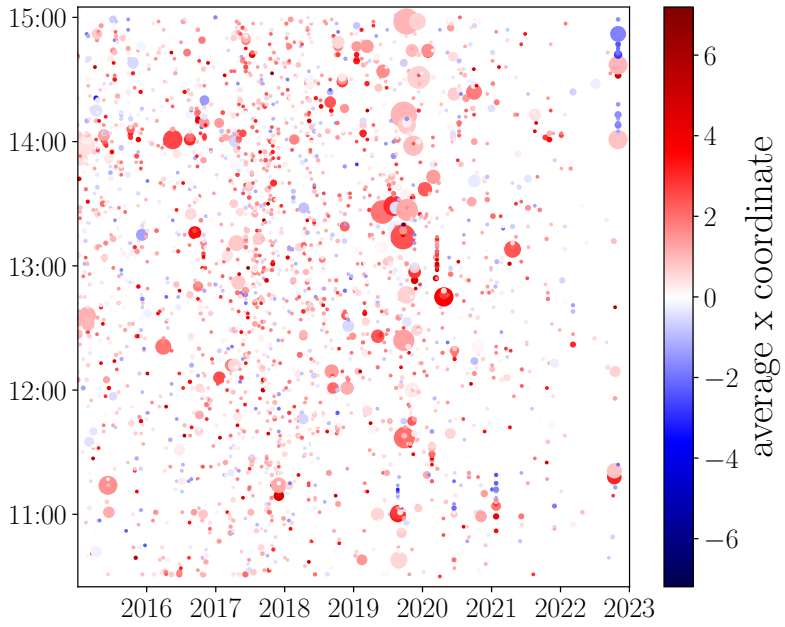
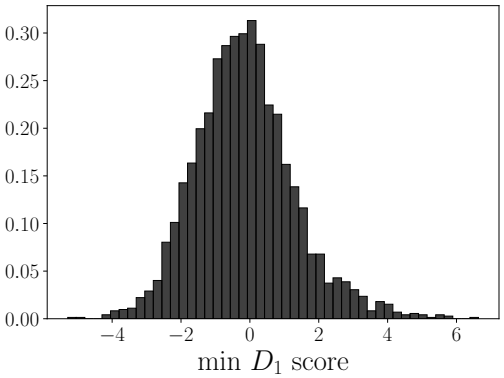
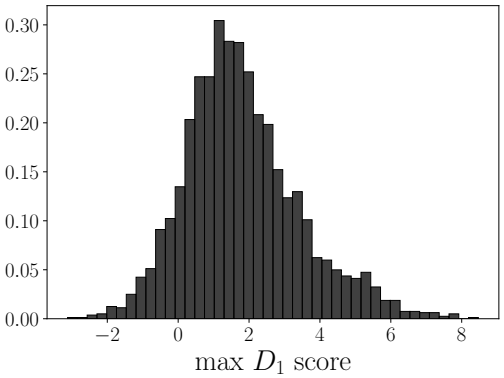


Figure 5.20: Reflexive score $\overline{D_1}$ of co-jumps in our dataset, obtained by averaging the reflexive score D_1 of each jump involved in a co-jump. Large co-jumps tend to have a higher average score (in red) but, surprisingly, there many large co-jumps with pale color that would be classified as endogenous. See discussion in the text.



(a) Distribution of the D_1 score of the most endogenous jumps of each co-jump (the leftmost jump in our 2D projections (see Figure 5.10) of all jumps belonging to a same co-jumps).



(b) Distribution of the D_1 score of the most exogenous jumps of each co-jump (the rightmost jump in our 2D projections (see Figure 5.10) of all jumps belonging to a same co-jumps).

Figure 5.21: Statistics on co-jumps from the reflexive direction on co-jumps.

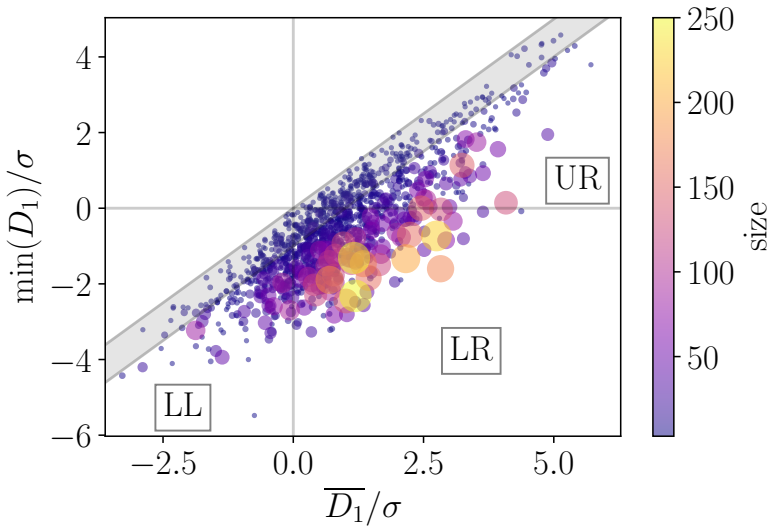


Figure 5.22: Minimum value of reflexivity score D_1 over all jumps of a given co-jump as a function of the average value $\overline{D_1}$ of reflexivity score D_1 over all jumps of a given co-jump (co-jump indicator 3 as a function of co-jump indicator 1). Both indicators have been normalized by the average standard deviation of the reflexivity score D_1 of co-jumps with the same size σ. The size and color of a point depict the size of the co-jump. The grey area represents the zone between $\min(D_1) = \overline{D_1}$ and $\min(D_1) = \overline{D_1} - 1$, corresponding to co-jumps where the difference between the minimum and the average D_1 score is less than 1σ. Here, we only consider co-jumps with a size strictly greater than 2. LL, LR & UR stand for lower left, lower right and upper right.

The lower right quadrant (LR) represent more intriguing co-jumps. Indeed, according to their average score D_1 those co-jumps should naively be classified as exogenous, however they contain at least one strongly endogenous co-jump. It might be that those endogenous jumps, whose pre-activity starts while most other stocks are still quiet, are interpreted in and by themselves as news. This surprise triggers all other jumps – which therefore appear as exogenous, with no special pre-jump activity but without being related to any news!

Note that the largest co-jumps are in the LR region; our interpretation in terms of a contagion mechanism would then naturally explain the power-law distribution of size $S^{-\tau}$ shown in Figure 5.18a.

There are obviously also large sector wide co-jumps that are truly news-related – upper-right quadrant of Figure 5.22. For instance, the significant co-jumps highlighting the year 2019 mostly exhibit a negative average (exogenous) and are related to the announcements during the US vs China trade war.

Conversely, some co-jumps (20% of our sample) involve only jumps exhibiting a symmetric or anticipatory profile (LL region of Figure 5.22). Those co-jumps are usually $S = 2$ stocks co-jumps (76%), but their size can go up to $S = 87$ stocks.

Hence, the most striking conclusion of this section is that many large co-jumps are in fact explained by endogenous dynamics and propagate across stocks, rather than being due to impactful external news. A (in)famous example of such propagation is the flash crash of May 6th 2010, where the S&Pmini crashed in less than 30min, due to a sell algorithm set with an excessively high execution rate. This crash triggered a price drop in other US stocks. Here, our results suggest that this synchronization phenomenon is not such a rare event and actually happens quite often [110, 54].

This finding is further supported by examining the correlation of the individual jump time-series composing a co-jump. Naively, one would expect large co-jumps to be exogenous, i.e. induced by news. As a result, the stocks involved in the co-jump should all share the same profile around the jump, as in Figure C.3a for example. In fact, Figure C.2 shows that there remain many co-jumps whose constituting univariate jump profiles are weakly correlated (see Appendix C.3 for more details).

A word on "sector jumps"

As discussed in Section 5.2.3, "sector jumps" are such that many stocks of the same industry jump simultaneously. Hence, all "sector jumps" are co-jumps. To be thorough, we also analyzed the sectoral price jumps through the scope of our projection. Those "sector jumps" are obtained by averaging all time series of stocks belonging to the same sector and detecting "sector jumps" using the method presented in Section 5.1.1. One can then compute a reflexive score for each of those "sector jumps" (along with a mean reversion score and a trend score). We present the results in Figure 5.23. The gray points represent the 2D projections of individual jumps of individual stocks. The blue points represent the projection of the "sector jumps" time-series. It appears that the sector jumps are indeed not all exogenous, as discussed in Section 5.2.3.

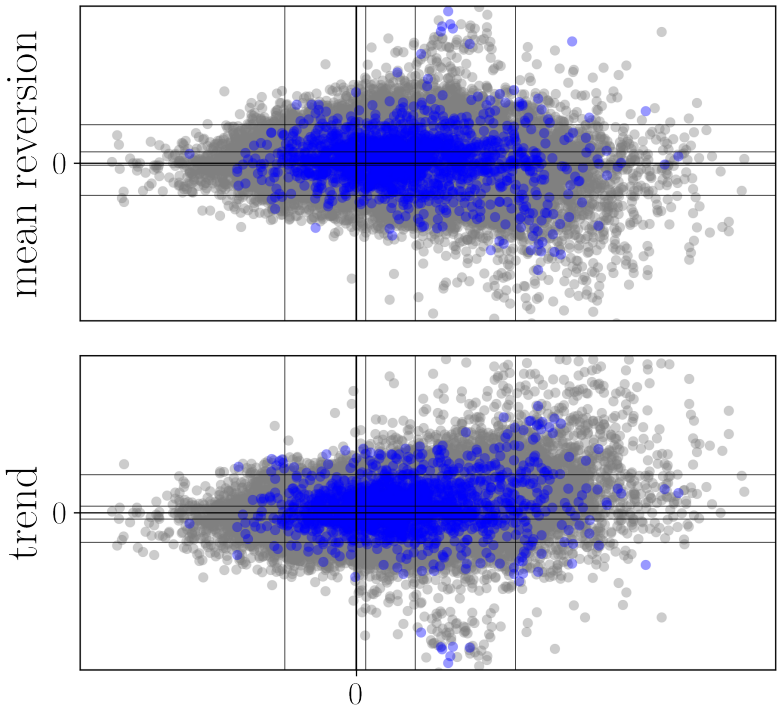


Figure 5.23: Projection of sector jumps in our 2D projections. The gray points represent the 2D projections of individual jumps of individual stocks. The "sector jumps" are obtained by averaging time-series across socks of a same sector. The blue points represent the projection of those sector jumps time-series.

Conclusion

Thanks to an unsupervised approach based on wavelet scattering coefficients, we have identified three main directions along which price jumps can be classified. The first, well-known direction relates to the time-asymmetry of the volatility of the price around the jump and results in three classes of jumps, endogenous, exogenous and anticipatory.

We also evidenced that mean-reversion and trend are important features for classification. This allowed us to identify three additional classes of jumps, "mean-reverting", "trend-aligned" and "trend-anti-aligned" which concerns a significant portion of the dataset. Thanks to this classification we have shown that a large portion of the jumps are endogenous or anticipatory jumps, confirming – but also making much more precise – the main conclusions of [99, 6].

Extending our analysis to co-jumps, we have gathered several pieces of evidence that a large proportion of these co-jumps should also, quite surprisingly, be classified as endogenous in the sense that they seem to originate from the contagion of one single endogenous jump triggering the jump of possibly many others. One striking signature of such a scenario is the power-law distribution of co-jump sizes, which is indeed close to that predicted by a critical branching (contagion) process. Such a broad, power-law distribution of co-jump sizes was noted previously for different datasets in [99, 55, 56]. Further work should focus on higher frequency data that would allow one to dissect more precisely the contagion mechanism and ascertain that many large co-jumps are indeed *not* triggered by exogenous news, but related to the close-knit nature of financial markets that brings them close to critical fragility, as argued many times in the past, see e.g. [97, 112, 113, 94] and refs. therein.

Unlike parametric fit of the time-series, the wavelet scattering embedding is defined and can be computed for any time-series. As such, our study could be transposed to other fields as well.

Take Home Message

- The reflexive nature of price jumps, or more generally of bursts of activity, can be measured by the degree of time asymmetry around the burst. This time asymmetry can be quantified using wavelet coefficients.
- The empirical study of price returns of 300 US stocks over an 8 years period yielded many results:
  - additionally to the jump reflexive nature, we uncovered two additional features to characterise price jumps: mean-reversion strength and trend alignment;
  - we retrieved that not all price jumps are news related (see also [5, 7, 6]);
  - we identified a new class of financial price jumps: *anticipatory jumps*, characterized by the majority of activity occurring before the jump;
  - co-jumps, even market-wide ones, are not always news related and seem to occur because of contagion mechanisms across stocks.
- **Technical outcomes**: unsupervised classification method to characterised bursts of activity based on their level of reflexivity, their mean-reversion strength and their alignment with the trend.

Chapter 6

Why is stock volatility so much rougher than index volatility?

> It is a rough road that leads to the heights of greatness.
>
> Lucius Annaeus Seneca

This chapter delves into the exploration of multivariate volatility from a different perspective, shedding light on the complex interplay between indices and individual stocks. The research presented herein is the result of collaboration with Othmane Zarhali, Jean-François Muzy, Emmanuel Bacry, and Jean-Philippe Bouchaud. We extend our gratitude to Rudy Morel and Samy Lakhal for their valuable assistance with numerical simulations.

Introduction

To properly introduce this chapter, we would like to first give some intuitions on the concept of signal roughness/smoothness. This characteristic is quantified by the Hurst exponent (H) of a signal, which ranges between 0 and 1. To illustrate the influence of H, Figure 6.1 depicts fractional Brownian motions for various values of H. It is evident from the visualization that lower values of H correspond to signals exhibiting greater roughness.

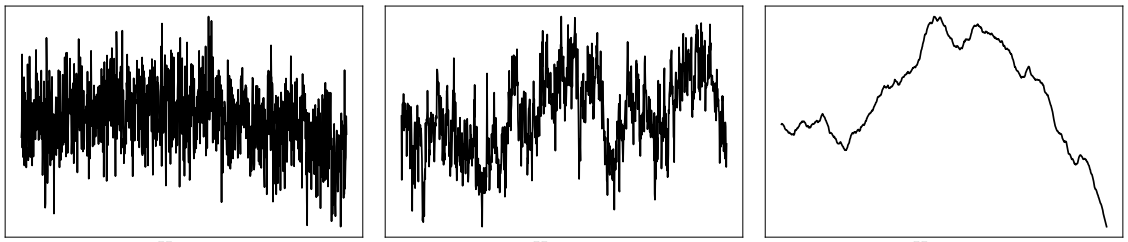


Figure 6.1: Some intuitions on the characterisation of roughness with the Hurst exponent. Realisations of fractional Brownian motions (Equation (6.0.1)) for different Hurst exponents: from left to right $H = 0.01, H = 0.2, H = 0.9$ (see Appendix D.3 for simulation method).

In particular, a H-fractional Brownian motion W^H, where H denotes the Hurst exponent, is characterised by its correlation structure:

$$\mathbb{E}\left(W_t^H W_s^H\right) = \frac{1}{2}\left(t^{2H} + s^{2H} - (t-s)^{2H}\right) \tag{6.0.1}$$

Of notable interest is the case when $H = \frac{1}{2}$, which yields the classic Brownian motion encountered in Black-Scholes and stochastic volatility models (as discussed in Chapter 1). Rougher ($H < \frac{1}{2}$) or smoother ($H > \frac{1}{2}$) dynamics than Brownian motion can enhance volatility models by introducing new features. For instance, pioneering work by Comte and Renaud in 1998 [114] proposed a smooth volatility process with $H > \frac{1}{2}$ to account for the long-range memory of volatility.

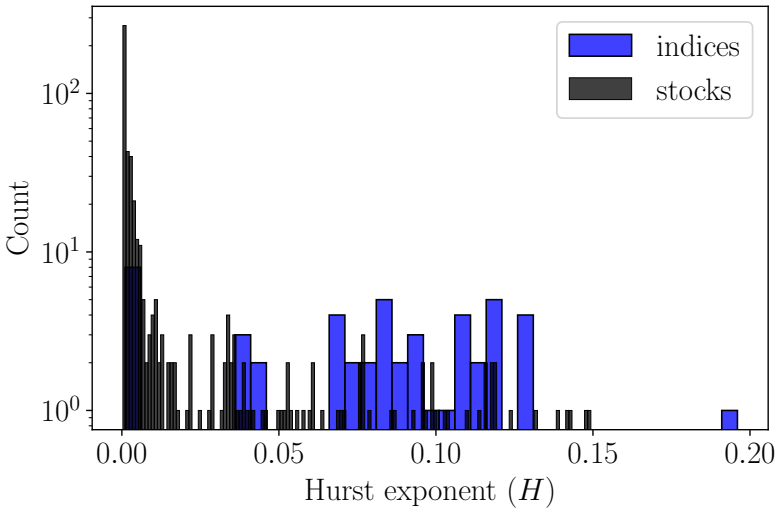


Figure 6.2: Hurst exponents estimation of 503 stocks (in black) and 49 indices (in blue) using the calibration method of Wu *et al.* [53] on daily Garman-Klass estimation of the log-volatility from 2013 to 2023 (see Garman-Klass estimator Definition 1.1.1).

In 2016, empirical work on the roughness of log-volatility in financial market by Gatheral *et al.* in [41] (see also [115]) demonstrated that *volatility is rough* (less smooth than a Brownian motion). Indeed measuring the roughness of the log volatility of major indices such as the S&P500 and the DAX[1], they find Hurst components (H) ranging between 0.06 and 0.2. Since then, rough volatility models have gather momentum [41, 116, 117, 118].

Furthermore, recent empirical investigations into volatility roughness have revealed that log volatility may exhibit even greater roughness than previously observed by Gatheral *et al.*, with equity indices showing a Hurst exponent below 0.05 [119, 120]. These findings are particularly interesting given that the behavior of rough volatility when $H \to 0$ aligns with the multifractal random walk (MRW) models introduced in 2000 in [121, 48, 47] and designed to capture the multifractal properties of asset prices [50, 51]. Although the MRW was actually the first model to replicate the empirical roughness of volatility, it was only after the 2016 work of Gatheral *et al.* that this topic gained significant traction among academics.

Seeking to build a framework that interpolates between MRW and rough volatility models (meaning that it conciliates the observed Hurst exponent of $H = 0.14$ with the asymptotic behavior where $H \to 0$), P. Wu, E. Bacry and J.-F Muzy introduced the logarithmic Stationary Brownian Motion (log-SfBM) in their 2022 work [53] (see also [122, 123, 124, 125] for alternative approaches). Calibrating their model on real world data, they found intriguing results. Specifically, their research shows that while individual stocks exhibit greater roughness ($H \approx 0$), indices, despite being composed of these stocks, tend to display a lower level of roughness ($H \approx 0.1$), as illustrated in Figure 6.2 (and Figure 10 of [53]).

This chapter is dedicated to elucidating the captivating empirical findings of Wu *et al.* which highlights disparities in the roughness of log-volatility between individual stocks and indices.

Our reasoning is derived from the intuition that the roughness exhibited in a composite signal, composed of several signals with varying degrees of roughness, tends to reflect the roughest component present (see Figure 6.1 for more intuition). Consequently, given empirical observations of Figure 6.2 showing that indices log-volatility is less rough than that of stocks, exceedingly rough signals are unlikely to contribute significantly to index dynamics, or at the very least, would need to be "smoothed out" within the averaging process of stock dynamics. The rougher component, characterized by $H \approx 0.$, is thus expected to be associated with an idiosyncratic component of individual stocks. Factor models offer a suitable framework for investigating such a hypothesis, as they delineate stock returns as the summation of a factor component (common to all stocks) and an idiosyncratic component (specific to each stock).

In particular, we opt to revisit the Nested Factor model proposed by Chicheportiche *et al.* [61] to accommodate additional stylized facts of financial volatility. Notably, the Nested Factor model

[1]**DAX**: German stock index.

holds significant relevance in describing the volatility dynamics of stocks, as it replicates several empirical characteristics observed in financial time series. It not only captures the presence of fat tails in the distribution of returns but also overcomes the limitations of elliptical models by reproducing the anomalous copula structure among stocks (see [60]). Moreover, it explicitly formulates the shared dynamics between the market mode and the residual components of individual stocks [58, 59].

In this chapter, we will see how the factor-residuals framework inherent in the Nested Factor model accommodates the empirical findings of Wu *et al.*: the log-volatility of individual stocks can indeed be rougher than that of indices, although indices are made of stocks.

The outline of the chapter is as follows. Sections 6.1 and 6.2 are dedicated to the presentation the models of interest: the log-SfBM of Wu *et al.* [53] and the Nested Factor model of Chicheportiche *et al.* [61]. Section 6.3 exposes our intuitions on how the Nested Factor model can reconcile the difference of roughness between the log volatility of stocks and of indices and is supported by Section 6.4 which presents numerical experiments. Section 6.5 presents our analytical analysis and Section 6.6 our empirical results.

6.1 Log Stationary fractional Brownian Motion model

In an attempt to reconcile empirical observations of Hurst exponents greater than 0.1 for indices [41] and tending towards 0 for stocks [119, 120], Wu *et al.* introduced the Stationary fractional Brownian Motion model (SfBM) in 2022. We outline the main characteristics of this model and provide key information for its calibration on empirical data. For comprehensive details, we refer to [53].

6.1.1 Model

Definition 6.1.1 (Stationary fractional Brownian Motion (S-fBM))**.** Introduced in [53], for Hurst exponent $H > 0$, the Stationary fractional Brownian Motion (S-fBM) $(\omega_{H,T}(t))_t$ is a stationary Gaussian process defined by its mean and covariance function:

$$\mathcal{C}_{\omega_{H,T}}(\tau) = \mathrm{cov}(\omega_{H,T}(t), \omega_{H,T}(t+\tau)) = \begin{cases} \frac{\nu^2}{2}\left(1 - (\frac{\tau}{T})^{2H}\right), & \text{when } |\tau| < T \\ 0 \quad \text{otherwise} \end{cases}$$

where ν is a dimensionless parameter that controls the amplitude of the process, and T is the time range of correlations.

Still for $H > 0$, Wu *et al.* then define the log S-fBm random measure $(M_{H,T}(t))_t$ as

$$M_{H,T}(\mathrm{d}t) = \exp(\omega_{H,T}(t))\mathrm{d}t \tag{6.1.1}$$

Interestingly, Wu *et al.* show that, when H goes to 0, $(M_{H,T}(t))_t$ converges toward a Multifractal Random Measure (MRM), previously introduced in [121, 49, 48, 47]. Specifically, for a Multifractal Random Walk (MRW) $(\omega_T(t))_t$, one has:

$$\mathcal{C}_{\omega_T}(\tau) = \mathrm{cov}(\omega_T(t), \omega_T(t+\tau)) = \begin{cases} -\lambda^2 \ln(\frac{\tau}{T}), & \text{when } |\tau| < T \\ 0 \quad \text{otherwise} \end{cases}$$

where λ^2 is the intermittency coefficient that is a dimensionless positive parameter and is related to ν^2 with $\lambda^2 = H(1-2H)\nu^2$. Hence, this framework harmonizes the cases where $H > 0$ and where $H \to 0$, which were both observed in real data. Wu *et al.* complete their model by proposing financial price dynamics to be described with the log-SfBM, such that returns $\mathrm{d}P_t^i$ of stock i write:

$$\mathrm{d}P_t^i = \exp(\omega_{H,T}^i(t))\mathrm{d}W_t^i = \frac{M_{H,T}^i(\mathrm{d}t)}{\mathrm{d}t}\mathrm{d}W_t^i$$

where $(\omega_{H,T}^i(t))_{t>0}$ is a SfBM of Hurst exponent H, and $(W_t^i)_{i,t}$ is a Brownian motion.

For more intuitions on these models, we refer to simulation algorithms in Appendix D.3 and references herein. The next section discusses the empirical estimation of the roughness on financial times series.

6.1.2 Empirical Estimation of the roughness (estimation of H)

In light of the discoveries by Gatheral *et al.*, numerous estimators for the roughness of log-volatility, namely the Hurst exponent, have been introduced in the literature [41, 115, 119, 120]. Here, we briefly outline two methodologies which will be employed in the empirical and numerical investigations. The first method, as employed in Gatheral *et al.* [41], relies on analyzing the increments of log-volatility, while the second method, introduced by Wu *et al.* [53], involves calibrating log-SfBM to estimate the Hurst exponent H alongside other parameters.

To estimate the Hurst exponent, Gatheral *et al.* utilize the observation that the moments of the increments of log-volatility σ, denoted as m_σ hereafter, follow a power law of qH, i.e.

$$m_\sigma(q,\tau) = \mathbb{E}(|\log(\sigma_{t+\tau}) - \log(\sigma_t)|^q) = K_q \tau^{qH} \tag{6.1.2}$$

where K_q is a constant. Consequently, measuring the increments of the log-volatility $m_\sigma(q,\tau)$ for a fixed q (e.g., $q=2$), one can estimate H by regressing $\log\left(m_\sigma(2,\tau)\right))$ against $2\log(\tau)$. Although this method has proven effective with high-frequency data, notwithstanding noise, it has demonstrated less robustness with larger time scales (e.g., daily data), prompting the development of alternative estimators [115, 119, 120, 117].

To calibrate the log-SfBM using empirical data, Wu *et al.* introduced two General Methods of Moments. These methods rely on estimating the autocovariance of $M_{H,T}$ and $\ln\left(M_{H,T}\right)$ at a fixed set of lags, where $M_{H,T}$ is the Multifractal Random Measure defined in Equation (6.1.1). Thus, one needs to compute $\mathbb{E}(M_{H,T}(t)M_{H,T}(t+\tau))$ or $\mathbb{E}\left(\ln\left(M_{H,T}(t)\right)\left(M_{H,T}(t+\tau)\right)\right)$ ([53], Proposition 6 provides detailed information). By applying their calibration method to daily volatility Garman-Klass estimates from 2013 to 2023 for 503 US stocks and 49 indices, results depicted in Figure 6.2 were obtained (also see Figure 10 of Wu *et al.*, 2022). Notably, the discrepancy in the order of magnitude between the Hurst exponents of indices (in blue) and individual stocks (in black) is puzzling. Indeed, stocks log-volatility are rougher ($H \approx 0.$) than that of indices ($H \approx 0.1$), although indices are made of stocks. To investigate such result, we structure the log-volatility dynamics combining the log-SfBM and the Nested factor model, which we delineate in the subsequent section.

6.2 Nested Factor model

Introduced in [61], the Nested Factor model seeks to replicate the dynamics of stocks returns and their interplay with market dynamics. This section aims at describing the model and its advantages.

6.2.1 Motivations

The development of the Nested Factor model stemmed from the observation that *joint distribution of stock returns is not elliptical* [60]. Elliptic models describe return dynamics as $\mathrm{d}P^i = \sigma_i \epsilon_i$, with ϵ_i a random variable, allowing for the analytical derivation of various properties. Specifically, in elliptical models, the linear correlation coefficient ρ between two stock returns is given by $\rho = \cos(2\pi C(\frac{1}{2},\frac{1}{2}))$, where $C(\frac{1}{2},\frac{1}{2})$ denotes the median point of bivariate copulas. By comparing empirical measures with the expected outcomes of an elliptic model, Chicheportiche *et al.*, in [60], demonstrated that $\rho \to -\rho - \cos(2\pi C(\frac{1}{2},\frac{1}{2}))$ is actually non zero for stocks, thereby refuting the suitability of elliptic models for describing the common dynamics of stock returns.

Additionally, the Nested Factor model intends to explicitly delineate the intriguing link between the distribution of stock returns and market (indices) returns. Previous research [57] has observed that the skewness of stock returns distribution was dependent of market returns: on average, on days when the market has negative, respectively positive, returns, then the distribution of all stocks returns has negative, respectively positive, skewness. As demonstrated in Cizeau *et al.* [58], this dependency is not captured by a "simple" factor model describing returns of stock i at time t, $\mathrm{d}P_t^i$ as:

$$\mathrm{d}P_t^i = \beta_i f_{0,t} + \epsilon_{i,t} \tag{6.2.1}$$

where $f_{0,t}$ represents the market returns (indices returns), $\epsilon_{i,t}$ is a residual component independent of $f_{0,t}$ and β_i signifies the exposure of the stock i to the market ($f_{0,t}$). Furthermore, Cizeau

et al. [58] and later Allez *et al.* [59], showed that not only does the returns distribution depend on the market, but the idiosyncratic term, ϵ from Equation (6.2.1), also depends on the market mode (f_0 in Equation (6.2.1)).

In addition to the correlation structures of stocks and the relationship between stock returns distribution and the market mode, Chicheportiche *et al.* [61] constructed their model in such a way as to generate fat-tailed return series. The subsequent section provides an introduction to the Nested Factor model as presented in Chicheportiche *et al.* (2015) [61].

6.2.2 Model

In light of these observations, Chicheportiche *et al.* established the Nested Factor model such that, returns $\mathrm{d}P^i$ of stock i are composed of two components: a factor and a residual part. The innovation lies in the structure of the log-volatilities which consist of both a common and a residual component, as elucidated below.

Definition 6.2.1 (Nested Factor model)**.** Nested Factor model with M factors describes the returns $\mathrm{d}P_t^i$ of stock i at time t as

$$\mathrm{d}P_t^i = \sum_{k=0}^{M-1} \beta_k^i f_{k,t} + e_{i,t} \quad \text{with} \quad \begin{cases} f_{k,t} = & \eta_{k,t} \exp(A_{0k}\Omega_0(t) + A_{1k}\Omega_1(t) + \omega_k(t)) \\ e_{i,t} = & \xi_{i,t} \exp(B_{0i}\Omega_0(t) + B_{1i}\Omega_1(t) + \tilde{\omega}_i(t)) \end{cases} \tag{6.2.2}$$

where

- the M time series $(f_{k,t})_{k\in[\![0,M-1]\!],t>0}$ are the factors, serving as "benchmarks" derived from the market. The series $f_{0,t}$ represents the market itself, and is frequently materialized as the returns of the S&P500 index. Conversely, for $k > 0$, $f_{k,t}$ may depict returns from a specific industrial sector. These time series $(f_{k,t})_{k\in[\![0,M-1]\!],t>0}$ are uncorrelated;
- β_j^i represents the exposure of asset i to a factor f_j;
- the time series $(e_{i,t})_{i,t>0}$ represent the residuals of each asset's returns, signifying the portion of the asset's returns not accounted for by market movements. These residuals are uncorrelated with the factors and are idiosyncratic to the asset itself;
- Ω_0, Ω_1 are log-volatility stochastic *factors* involved in the dynamics of both the factors $(f_{k,t})_{k\in[\![0,M-1]\!],t>0}$ and the residuals $(e_{i,t})_{i,t>0}$. ω_k denotes the stochastic *idiosyncratic* log-volatility of factor f_k, while $\tilde{\omega}_i$ is the stochastic *idiosyncratic* log-volatility of residual e_i;
- parameters As and Bs quantify the contribution of each factor volatility mode. They can be estimated using correlation functions (denoted C^{ff} and C^{rr} in [61], note that, their dominant eigenvector provide a reliable initial estimate);
- finally, η and ξ are Brownian motions, independent of all the other time series.

For the sake of simplicity and to reduce the number of parameters, in the numerical and empirical investigations, we will use simplified versions of the Nested Factor model, which we introduce subsequently, keeping the same notations.

6.2.3 Simplified versions

One common mode of log volatility and M factors

For the empirical investigation (Section 6.6), which involves calibrating the NFM on real data, we use the Nested Factor Model with one dominant volatility mode Ω_0, such that:

$$\mathrm{d}P_t^i = \sum_{k=0}^{M-1} \beta_{i,k} f_{k,t} + e_{i,t} \quad \text{with} \quad \begin{cases} f_{k,t} = \eta_{k,t} \exp(A_{k0}\Omega_0(t) + \omega_k(t)) \\ e_{i,t} = \epsilon_{i,t} \exp(B_{i0}\Omega_0(t) + \tilde{\omega}_i(t)) \end{cases}, \tag{6.2.3}$$

which comes down to set Ω_1 to zero.

One common mode of log volatility and a unique factor (the market mode)

To simplify further for our numerical simulations (Section 6.4), we implement the Nested Factor Model with one factor and with one dominant volatility mode, such that:

$$\mathrm{d}P_t^i = \beta_i f_{0,t} + e_{i,t} \quad \text{with} \quad \begin{cases} f_{0,t} = \eta_t \exp(A_0\Omega_0(t) + \omega_0(t)) \\ e_{i,t} = \epsilon_{i,t} \exp(B_{i0}\Omega_0(t) + \tilde{\omega}_i(t)) \end{cases}, \tag{6.2.4}$$

which comes down to set Ω_1 to zero and M to 1. The only factor f_0 represents the market mode and, for US stocks for instance, can directly be estimated using the returns of the S&P500 index.

In the rest of the chapter, we revisit the Nested Factor model by considering that the log-volatilities $(\Omega_0, \omega_k, \tilde{\omega}_i)$ follow a Stationary fractional Brownian Motion (SfBM). Subsequently, we investigate their roughness to elucidate the empirical findings of Wu *et al.* (2022) [53].

6.3 Conjectures & Investigation

The primary aim of this chapter is to demonstrate how the Nested Factor model can elucidate the discrepancy in magnitudes between $H_{\text{indices}}(\approx 0.14)$ and $H_{\text{stocks}}(\approx 0.)$, despite indices being composed of stocks. Specifically, we inquire into 2 conjectures:

- **Subsidiary conjecture** ("Minimal Roughness Dominance conjecture"): We postulate that the amalgamation of two signals characterized by differing levels of roughness yields a signal whose roughness closely resembles that of the signal with the higher roughness level (lower value of Hurst exponent H). This can be visualized by referring to Figure 6.1: when summing the signal with the lowest roughness and the one with the highest roughness, the resulting signal would primarily reflect the roughness of the latter.

- **Main conjecture** ("Log S-fBM Nested factor model"): We propose a Nested Factor model with SfBM as log-volatilities (i.e., Ω_0, $(\omega_k)_k$ and $(\tilde{\omega}_i)_i$ are SfBM), with well chosen parameters. We postulate that such model could reconcile the empirical findings of Wu *et al.* ([53] and Figure 6.2). Specifically, we anticipate that selecting parameters satisfying:

  - $H_{\Omega_0} \gtrsim 0.1$,
  - $H_{\omega_0} \gtrsim 0.1$,
  - $H_{\tilde{\omega}_i} \approx 0$ for all stocks i,

  could replicate the empirical observations of Figure 6.2.

  This second conjecture stems from the notion that, within the Nested Factor model, the roughness of the log-volatility of an individual stock i is contingent upon the roughness of Ω_0, $(\omega_k)_k$, and $\tilde{\omega}_i$, whereas the roughness of indices is determined by Ω_0 and ω_0. This occurs as idiosyncratic components offset each other when averaging the financial returns $\mathrm{d}P^i$ to derive the financial returns of an index (market mode f_0). From the "Minimal Roughness Dominance conjecture", we expect that the roughest component of each time series, for both stocks and indices, will prominently manifest.

  Given that the roughness of stock log-volatility exceeds that of indices ($H_{\text{stocks}} \approx 0 < H_{\text{indices}} \approx 0.1$), we infer that the roughest component must originate from the idiosyncratic part of the stock $\tilde{\omega}_i$, while indices' log-volatilities, Ω_0 and ω_0, are presumed to exhibit lesser roughness ($\gtrsim 0.1$).

We explore these conjectures along three dimensions: numerical simulations allow to assess the aforementioned hypotheses; analytical study establishes the validity of our hypothesis within a specific regime; and finally, empirical analysis through the calibration of the Nested Factor Model (NFM) using real-world data and the investigation of the log-volatilities roughness confirms our main conjecture.

6.4 Investigation via Numerical Simulation

To test our conjectures from the previous section, we implement numerical simulations. The method to simulate log-SfBM, and more generally Gaussian processes defined through their auto-correlation structure, is described in Appendix D.3.

6.4.1 Test of the subsidiary conjecture: Minimal Roughness Dominance hypothesis

Simulation

In this first part, we test by numerical simulations our subsidiary hypothesis which postulates (see Section 6.3) that the amalgamation of two signals characterized by differing levels of roughness yields a signal whose roughness closely resembles that of the signal with the higher roughness level. For this first hypothesis, we implement 500 simulations of the following framework.

We consider 100 stocks/trajectories whose returns dynamics are defined by:

$$\mathrm{d}P_t^i = \sigma_f \exp(\omega_H(t))\mathrm{d}W_t^{\beta_i} + \tilde{\sigma}\exp(\tilde{\omega}_i(t))\mathrm{d}B_t^i. \tag{6.4.1}$$

The log-volatility ω_H is a shared volatility mode across all trajectories. Its dynamics conform to that of a SfBM with intermittent parameter $\lambda = 0.079$ and Hurst exponent $H_{\omega_H} = 0.14$. Additionally, each trajectory i is characterised by an idiosyncratic log-volatility $\tilde{\omega}_i$ which is a rough SfBMs (MRW) with intermittent parameter $\lambda = 0.079$ and Hurst exponent $H_{\tilde{\omega}_i} = 0$. Moreover the values of σ_f and $\tilde{\sigma}$ are determined such that the variance of the factor term $(\mathbb{V}(\sigma_f \exp(\omega_H(t))\mathrm{d}W_t^{\beta_i}))$ approximately accounts for 30% of the total variance, as observed empirically for the S&P500. $(W_t^{\beta_i})_{t>0, i\in[\![0,99]\!]}$ are Brownian motions, satisfying $\mathbb{E}(\mathrm{d}W_t^{\beta_i}\mathrm{d}W_t^{\beta_j}) = \beta^i\beta^j\mathrm{d}t$, where $(\beta_i)_{i\in[\![0,99]\!]}$ are normally distributed, as observed empirically. Finally, $(B_t^i)_{t>0, i\in[\![0,99]\!]}$ are independent Brownian motions, also independent of $(W_t^{\beta_i})_{t>0, i\in[\![0,99]\!]}$. Section 6.5.1 provides an analytical study of an analogous framework.

Results

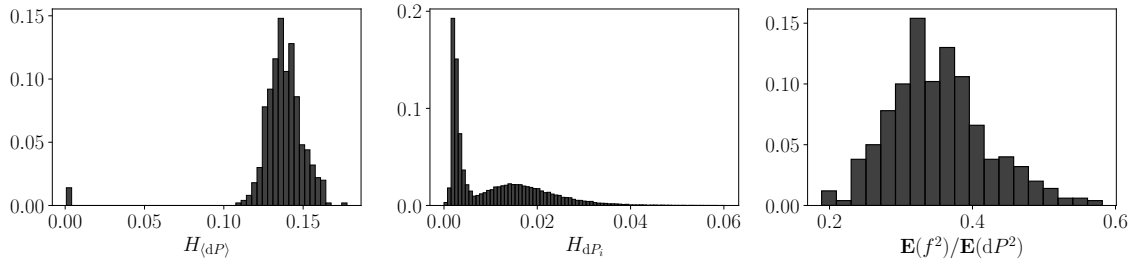


Figure 6.3: Results derived from the simulation of Equation (6.4.1). Left histogram: Roughness estimation of average path $\langle \mathrm{d}P\rangle = \frac{1}{100}\sum_{i=1}^{100}(\mathrm{d}P^i)$ across the 500 simulations. Middle histogram: Roughness estimation of individual log-volatility of each of the 100 stock returns $\mathrm{d}P^i$ for the 500 simulations. Right histogram: variance proportion of the factor term $f = \sigma_f \exp(\omega_H(t))\mathrm{d}W_t^{\beta_i}$ on total variance across the 500 simulations.

We commence by confirming that the factor component indeed accounts for approximately 30% of the overall variance, aligning with empirical observations. The histogram on the right-hand side of Figure 6.3 demonstrates the consistency of this assertion within our simulation set.

For each simulation, we then corroborate our secondary conjecture by estimating the roughness of each individual trajectory $(\mathrm{d}P_t^i)_{t>0, i\in[\![0,99]\!]}$ using the General Method of Moments proposed by [53]. The outcomes are depicted in the middle histogram of Figure 6.3. It is evident that all individual trajectories exhibit highly rough paths ($H_{\tilde{\omega}_i} < 0.06$ for all i). Hence, the roughness of each trajectory $(\mathrm{d}P_t^i)$ is primarily influenced by its roughest component $\tilde{\omega}_i$, thereby substantiating our subsidiary hypothesis. This holds true even when the factor term explains a substantial proportion of the variance. As indicated in the rightmost figure, the factor term consistently explains at least 20% of the total variance and sometimes up to 50%. However, even in the

latter scenario, the roughness of the resulting signal remains below 0.06. This finding is further corroborated by the analytical results presented in Section 6.5.1.

To be thorough, we also assess the roughness of the average trajectory $\langle \mathrm{d}P \rangle = \frac{1}{100} \sum_{i=1}^{100} (\mathrm{d}P^i)$. The results of this estimation, across our 500 simulations, are depicted in the leftmost histogram of Figure 6.3. Our simulations reveal that when the individual trajectories are averaged, as would be done to reconstruct the index trajectory, the individual idiosyncrasies vanish, and the resulting averaged trajectory exhibits the same Hurst exponent as the factor term ($\widehat{H}_{\langle \mathrm{d}P \rangle} \approx 0.14 = H_{\omega_H}$). Consequently, at this stage, the simple factor model outlined in Equation (6.2.1) already aligns with the empirical findings of [53]. However, this version of the model would not adequately account for other stylised facts as the relationship between the volatility and skewness of stock returns and the market returns [59, 58, 57]. Therefore, we proceed to examine a more sophisticated model, the Nested Factor model.

6.4.2 Main conjecture: log SfBM Nested Factor Model to reconcile rough stock and less rough indices

Section 6.4.1 demonstrated that a "simple" factor model, where the factor term and the idiosyncratic term each have their own volatility mode ($\Omega_0 = \Omega_1 = 0$ in the NFM), replicates the empirical results reported by Wu *et al.* [53]. This second numerical investigation seeks to examine whether the Nested Factor model with a common volatility mode ($\Omega_0 \neq 0$) as described in Equation (6.2.4) can also reproduce the differing levels of roughness between stocks and indices observed by Wu *et al.* [53].

Simulation

The implementation involves simulating, as detailed in Appendix D.3, the Nested Factor model with 1 common volatility mode, 1 factor and 100 trajectories/stocks i (as outlined in Equation (6.2.4)). The common volatility component Ω_0 follows a SfBM with intermittent parameter $\lambda = 0.079$ and Hurst exponent $H_{\Omega_0} = 0.11$. The residual log-volatility of the factor, ω_0, is also a SfBM of intermittent parameter $\lambda = 0.079$ and Hurst exponent $H_{\omega_0} = 0.14$. Thus, we expect the combination of these two trajectories, Ω_0 and ω_0, to replicate the roughness observed in indices ($H_{\text{indices}} \approx 0.1$).

The residual log-volatilities $\tilde{\omega}_i$ associated with the idiosyncratic term e_i are rough SfBMs (MRWs) with intermittent parameter $\lambda = 0.079$ and Hurst exponent $H_{\tilde{\omega}_i} = 0$. We anticipate that this rough component will dominate in the individual trajectories, thereby mirroring the behavior of stocks log-volatility ($H_{\text{stocks}} \approx 0$).

Coefficients A_0 and B_0 are selected in accordance with the values documented in [61]. Similarly, as in the preceding simulation, the coefficients $(\beta_i)_{i \in [\![0,99]\!]}$ are normally distributed. Additionally, to maintain consistency with empirical observations, we restrict our simulations set to instances where the factor term explain approximately 30% of the total variance, as illustrated in the rightmost histogram of Figure 6.4. This criterion yields 68 simulations comprising 100 trajectories each.

Results

In each simulation, we employ the General Method of Moments introduced by [53] to estimate the roughness of the 100 individual trajectories. The resulting outcomes are illustrated in the middle histogram of Figure 6.4. It is evident that all individual trajectories, representing the returns of individual stocks, manifest highly rough paths, with all Hurst exponents falling below 0.07 and averaging 0.04. Consequently, the roughness of the individual signal remains predominantly influenced by the roughest component $\tilde{\omega}_i$, which characterizes the idiosyncratic dynamics of each trajectory. This initial finding supports with our main hypothesis, suggesting that the Nested Factor model, when the residual log-volatility of the idiosyncratic term is highly rough, can reproduce the notably rough paths observed in individual stocks.

Furthermore, we estimate the roughness of the log-volatility of the simulation's average path, $\langle \mathrm{d}P \rangle = \frac{1}{100} \sum_{i=1}^{100} (\mathrm{d}P^i)$, which would reconstruct the index trajectory. The leftmost histogram of Figure 6.4 displays the resulting Hurst exponent values. It clearly appears that the average trajectory exhibits a less rough path compared to individual trajectories, with Hurst exponents hovering around 0.12. Thus, as anticipated, the idiosyncratic terms offset each other during averaging, and

the roughness of the average path is mostly driven by the log-volatility components of the factor term, Ω_0 and ω_0.

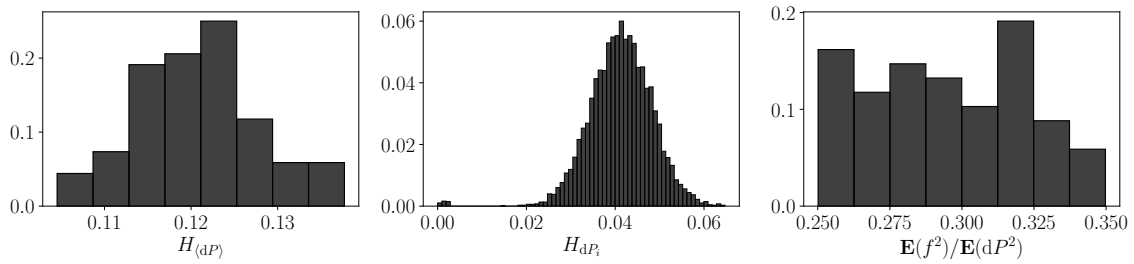


Figure 6.4: Results derived from the simulation of Equation (6.2.4). Left histogram: Roughness estimation of log volatility of the average path $\langle dP \rangle = \frac{1}{100}\sum_{i=1}^{100}(dP^i)$ across our set of simulations. Middle histogram: Roughness estimation of individual log-volatility of each of the 100 stock returns i for all the simulations. Right histogram: variance proportion of the factor term f_0 on total volatility across our set of simulations.

Those results confirm that the NFM enables the replication of the empirical observations from Wu *et al.* ([53] and Figure 6.2). This conclusion is further supported by the analytical results presented in the subsequent section.

6.5 Analytical Investigation

To validate our conjectures, we also conduct an analytical investigation. We study two combinations of SfBMs with differing degrees of roughness. First combination consists in a factor model with 2 independent log-SfBMs following the traditional form of factors models. The second combination is a log-SfBM with 2 volatility modes, i.e. two SfBMs.

6.5.1 Sum of two log-SfBMs

First, we investigate the roughness of the following signal:

$$dP_t = \exp\left(\Omega\left(t\right)\right) dW_t + \exp\left(\tilde{\omega}\left(t\right)\right) dB_t, \tag{6.5.1}$$

where Ω and $\tilde{\omega}$ are independent SfBM defined by their Hurst exponents, respectively, H and $\tilde{h}$, such that $\tilde{h} \ll H$, and intermittent parameters respectively, ν and $\tilde{\nu}$. W_t and B_t are independent Brownian motions.

Based on the characterisation in Equation (6.1.2), to inquire into the roughness of the resulting signal, we want to find $\hat{h}$ such that

$$\begin{aligned} m(2,\Delta) =& \mathbb{E}\left(\left|\log\left(\exp\left(2\Omega(\Delta)\right) + \exp\left(2\tilde{\omega}(\Delta)\right)\right) - \log\left(\exp\left(2\Omega(0)\right) + \exp\left(2\tilde{\omega}(0)\right)\right)\right|^2\right) \\ =& K_2\Delta^{2\hat{h}} \quad \text{with } K_2 \text{ constant} \end{aligned}$$

Leveraging the replica trick [126] and some combinatory analysis (see Appendix D.1), we obtain (up to order $\Delta^{2\tilde{h}}$):

$$m(2,\Delta) = 4\tilde{\nu}^2\mathbb{E}\left(\exp(2\tilde{\omega}_t)^2\right)\mathbb{E}\left(\left(\exp(2\Omega) + \exp(2\tilde{\omega}_t)\right)^{-2}\right)\Delta^{2\tilde{h}} + o(\Delta^{2\tilde{h}})$$

and, if we develop up to order Δ^{2H}:

$$m(2,\Delta) = 4\mathbb{E}\left(\left(\exp(2\Omega) + \exp(2\tilde{\omega}_t)\right)^{-2}\right)\left(\tilde{\nu}^2\Delta^{2\tilde{h}}\mathbb{E}\left(\exp(2\tilde{\omega}_t)^2\right) + \nu^2\Delta^{2H}\mathbb{E}\left(\exp(2\Omega)^2\right)\right) + o(\Delta^{2H})$$

We see that asymptotically, when $\Delta \to 0$, the roughest signal (with the smaller Hurst exponent H) is predominant in the resulting signal.

6.5.2 log-SfBM with 2 modes of volatility

Secondly, we investigate the roughness of the log-volatility of a signal defined by:

$$\mathrm{d}P_t = \exp\left(A_0\Omega(t) + \tilde{\omega}(t)\right)\mathrm{d}W_t$$

where Ω and $\tilde{\omega}$ are two SfBMs characterised by their Hurst exponents, respectively H and $\tilde{h}$ with $\tilde{h} \ll H$, and their intermittent parameter ν and $\tilde{\nu}$. $(W_t)_{t>0}$ is a Brownian motion. As before, to characterize the roughness of such signal, one's aim is to find $\hat{h}$ such that

$$\begin{aligned} m(2,\Delta) =& \mathbb{E}\left(\left|\log\left(\exp\left(2A_0\Omega(\Delta) + 2\tilde{\omega}(\Delta)\right)\right) - \log\left(\exp\left(2A_0\Omega(0) + 2\tilde{\omega}(0)\right)\right)\right|^2\right) \\ =& K_2\Delta^{2\hat{h}} \quad \text{with } K_2 \text{ constant} \end{aligned}$$

We obtain (see Appendix D.2 for details):

$$m(2,\Delta) = 4\left(A_0^2K_2\nu^2\Delta^{2H} + k_2\tilde{\nu}^2\Delta^{2\tilde{h}}\right)$$

We see that asymptotically, when $\Delta \to 0$, the roughest signal (with the smaller Hurst exponent, $\tilde{h}$) is again predominant in the resulting sum.

Although partial, analytical results tend to confirm our conjectures. We now intend to empirically measure the roughness of the Nested Factor model volatility components.

6.6 Investigation of roughness of empirical NFM components

To test our conjecture from Section 6.3, we calibrate the Nested Factor Model with a single common mode of volatility and 10 factors ($M = 10$ in Equation (6.2.3)) on empirical data. Using several estimation methods, we investigate the roughness of the log-volatility components.

6.6.1 Data description

We calibrate the Nested Factor model using data from 323 US stocks spanning the period from 2013 to 2023. Our selection is limited to stocks that are constituents of the S&P500 throughout the entire study period. We analyze both daily and intraday data, as roughness estimation tends to be more effective with higher frequency data. The datasets consist of close-to-close returns that are standardized by centering and normalization. This dataset is sufficient for calibrating the Nested Factor model and obtaining time series for Ω_0, ω_0, $(\omega_k)_{k\in[\![1,9]\!]}$, and $(\tilde{\omega}_i)_{i\in[\![1,323]\!]}$. The calibration procedure of the Nested Factor model using real-world data is delineated below (for comprehensive details, see [61]).

6.6.2 Calibration of the NFM on data

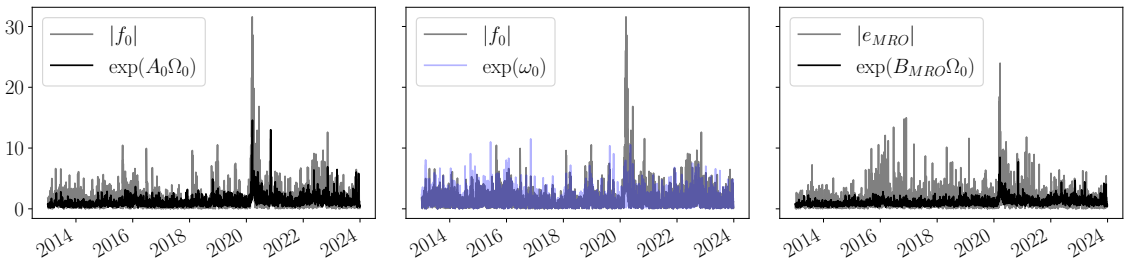


Figure 6.5: Illustration of the NFM calibration on empirical data, with 10 factors on 323 stocks from 2013 to 2023. Left plot: the market mode $|f_0|$ and common volatility mode $\exp(A_0\Omega_0)$ time series. Middle plot: the market mode $|f_0|$ and its residual volatility $\exp(\omega_0)$ time series. Right plot: the idiosyncratic time series of Marathon Oil stock $|e_{MRO}|$ and the common volatility mode $\exp(B_{MRO}\Omega_0)$ time series (with the relevant coefficient B_i).

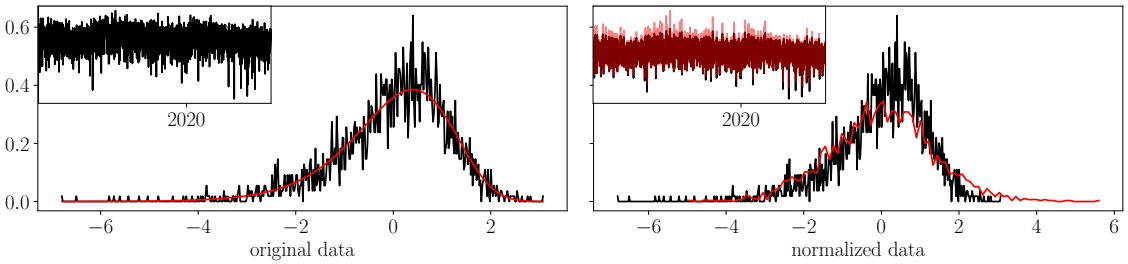


Figure 6.6: Illustration of the "Gaussianization" procedure applied to the time series $(\tilde{\omega}_{MRO}(t))_t$. Left figure: distribution of the values of the original time series before the "Gaussianization". The best theoretical distribution to describe $(\tilde{\omega}_{MRO}(t))_t$ is a log-gamma distribution, plotted in red with appropriately chosen parameters. The inset shows the original time series as a function of time. Right figure: distribution of the values of the "Gaussianized" time series in red superimposed with the distribution of the original time series values in black. The inset displays both the original and "Gaussianized" time series as functions of time.

Guided by [61], we calibrate the Nested Factor model depicted in Equation (6.2.3), for $M = 10$ with 323 stocks. The calibration relies on the following steps:

1. The exposure coefficients $(\beta_{i,k})$ are obtained by minimizing the off diagonal of $||X^\top X - \beta^\top \beta||$ for β, where $X^\top X$ is the correlation matrix of stock returns $(\mathrm{d}P_t^i)_{i\in[\![1,323]\!],t>0}$.

2. Then, $f_{k,t}$ and $e_{i,t}$ are obtained by regressing $\mathrm{d}P_t^i$ on $(\beta_{i,k})$.

3. A method of moments allow then to obtain $(A_{k0})_k$ and $(B_{i0})_i$ (see the correlation structures in Appendix B of [61]).

4. Finally, regressing $\log(|f_{k,t}|) - \langle \log(|f_{k,t}|)\rangle$ and $\log(|e_{i,t}|) - \langle \log(|e_{i,t}|)\rangle$ on A_{k0} and B_{i0} permit to obtain the time series of $\Omega_{0,t}$, $(\omega_{k,t})_{k\in[\![0,9]\!]}$ and $(\tilde{\omega}_{i,t})_{i\in[\![1,323]\!]}$. Figure 6.5 illustrates the Nested Factor model decomposition into factors f, residuals e and common volatility mode Ω_0 by juxtaposing those time series.

5. *(mostly for daily data)* Before estimating the roughness of $\Omega_{0,t}$, $(\omega_{k,t})_{k\in[\![0,9]\!]}$ and $(\tilde{\omega}_{i,t})_{i\in[\![1,323]\!]}$, it is prudent to address the pronounced skewness evident in the idiosyncratic time series $(\tilde{\omega}_{i,t})_{i\in[\![1,323]\!]}$. This skewness arises from the logarithmic transformation, which tends to amplify the impact of lower return values. Such skewness subsequently propagates into the log-volatility components, contravening the Gaussian assumption of the SfBM model and adversely affecting the estimation of the Hurst exponent, particularly with daily data. To adhere to this assumption, we undertake a "Gaussianization" procedure for the time series $(\tilde{\omega}_{i,t})_{i\in[\![1,323]\!]}$ for the daily dataset. Note that this procedure could also be done before step 4, on the time series $e_{i,t}$. The "Gaussianization" procedure, illustrated Figure 6.6, involves two main steps. The first step entails identifying the best theoretical distribution and its parameters to describe the values of the time series $(\tilde{\omega}_t)_t$. For our data sample, the residual log-volatility modes of stock idiosyncratic terms are best described by the log-gamma distribution. For example, for Marathon Oil stock, the log-gamma fit is illustrated in red in the left panel of Figure 6.6. Given the most appropriate log-Gamma distribution with scale γ and cumulative distribution function F_Γ, the "Gaussianization" is then achieved using the transformation $F_{\mathcal{N}}^{(-1)} \circ F_\Gamma$ where $F_{\mathcal{N}}^{(-1)}$ is the inverse cumulative function of the Gaussian distribution with mean 0 and scale γ. The right panel of Figure 6.6 displays the distributions of the data before (in black) and after the procedure (in red) for $(\tilde{\omega}_{MRO}(t))_t$, demonstrating that the skewness is corrected.

The resulting time series of Ω_0 and ω_0 are illustrated Figure 6.7 (left and middle figures respectively). As an illustrative example, the residual log-volatility of the idiosyncratic component for the Marathon Oil stock is depicted in the leftmost plot of Figure 6.7. Note that the residual log-volatility of the idiosyncratic component is less skewed than the residual log-volatility of the factor component, ω_0, thanks to the "Gaussianization" procedure.

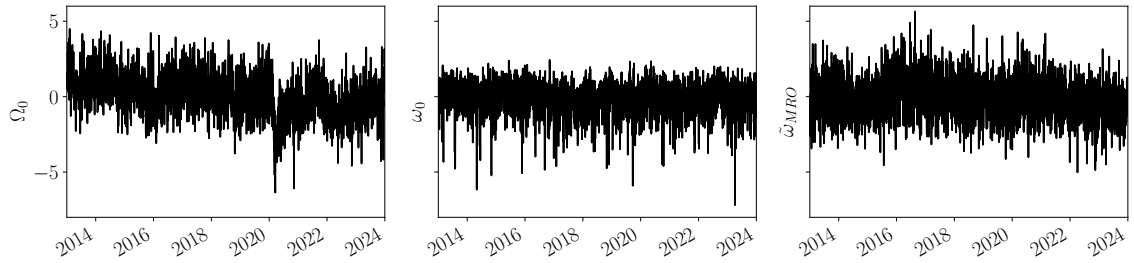


Figure 6.7: Resulting daily log-volatilities from calibration of NFM with 10 factors on 323 stocks from 2013 to 2023. Left plot: common log-volatility mode Ω_0 time series. Middle plot: market residual log-volatility ω_0 time series. Right plot: idiosyncratic residual log-volatility of the Marathon Oil stock $\tilde{\omega}_{MRO}$.

6.6.3 Roughness estimation of the empirical NFM components

We proceed with the empirical estimation of the roughness for various log-volatility components within the model. The goal is to empirically validate our main conjecture by estimating the roughness of the following components: the common log-volatility component Ω_0, the residual log-volatility of the market mode ω_0 and more generally, the residual log-volatility modes of the factors $(\omega_k)_{k\in[\![1,9]\!]}$ and of the idiosyncratic terms $(\tilde{\omega}_i)_{i\in[\![1,323]\!]}$.

As a reminder, our subsidiary conjecture posits that when two signals with differing levels of roughness are combined, the resulting signal's roughness is predominantly influenced by the roughest component. Hence, observing that stocks log-volatility is rougher than that of indices, our main conjecture postulates that the heightened roughness stems from the idiosyncratic component of the stocks. This idiosyncratic roughness vanishes when averaging stock returns to reconstruct index trajectories, consequently imposing a lower roughness on index log-volatility ($H_{\text{indices}} \gtrsim 0.1$).

Hence, for the Nested Factor model to align with the empirical results of Wu *et al.* (see Figure 6.2), we expect to find log-volatility components with the following Hurst exponents:

- $H_{\Omega_0} \gtrsim 0.1$
- $H_{\omega_0} \gtrsim 0.1$
- for all stocks i, $H_{\tilde{\omega}_i} \approx 0$

Estimation of the roughness on daily data

To assess the roughness of the log-volatility components within the NFM, we analyze the log-volatility times series of Ω_0, ω_0, $(\omega_k)_{k\in[\![1,9]\!]}$ and $(\tilde{\omega}_i)_{i\in[\![1,323]\!]}$, derived from the calibration of the Nested Factor model, by calibrating SfBMs on these time series with the GMM from Wu *et al.* [53]. Results are presented in Figure 6.8.

The analysis of the idiosyncratic residual log-volatility roughness, $H_{\tilde{\omega}_i}$, depicted in the leftmost histogram of Figure 6.8, reveals that these components tend to be highly rough ($H_{\tilde{\omega}_i} \approx 0$), which is consistent with our conjecture. Some of them, however, seem to demonstrate a greater Hurst exponent ($H_{\tilde{\omega}_i} \gtrsim 0.2$). This discrepancy might stem from noise present in our data, as the intraday estimation does not yield such values (see below). Nonetheless, this finding remains consistent with the general Hurst exponent patterns depicted in Figure 6.2, where higher H values (greater than 0.1, for instance) are observed for certain individual stocks, predominantly from the energy and real estate sectors.

Additionally, the factor residual log-volatilities, $(\omega_k)_{k\in[\![0,9]\!]}$, whose estimated Hurst exponents are depicted in the right histogram of Figure 6.8, exhibit trajectories with less roughness, as their Hurst exponent values mostly exceed 0.2. In particular, the residual log-volatility of the market mode ω_0 demonstrates a Hurst exponent of $H_{\omega_0} = 0.28$. Since ω_0, along with the common log-volatility mode Ω_0, determines the roughness of indices log-volatility, the value $H_{\omega_0} = 0.28 > 0.1$ is consistent with our conjecture. Some factor residuals, however, appear to exhibit very rough trajectories. Specifically, residual modes associated with the Information Technology, Consumer Goods, and Materials sectors have Hurst exponents close to 0 (see the right histogram of Figure 6.8). Despite some indices in Figure 6.2 exhibiting high roughness, which could explain these

low Hurst values for certain factor residuals, future research could explore the roughness in relation to industrial sectors.

Finally, applying the GMM calibration of [53] on the common log-volatility mode, Ω_0, we obtain $H_{\Omega_0} = 0.09$ (which is close to 0.1). This result further aligns with our main conjecture.

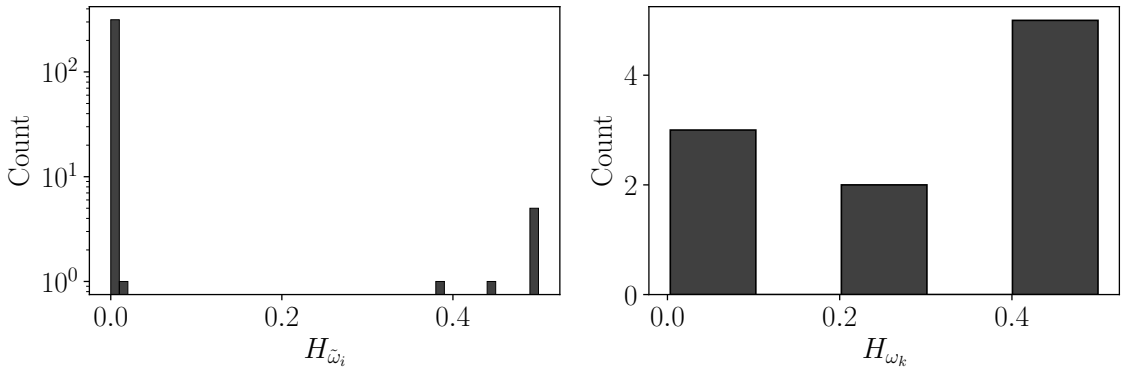


Figure 6.8: Hurst exponent estimation using GMM calibration from Wu *et al.* [53] on the times series of the idiosyncratic residuals $(\tilde{\omega}_i)_{i\in[\![1,323]\!]}$ (left plot) and of the factor residuals $(\omega_k)_{k\in[\![0,9]\!]}$ (right plot) derived from thecalibration of NFM with 10 factors on daily data, on 323 US stocks, from 2013 to 2023.

Sourcing f_0 as the S&P500

To mitigate uncertainties and noise from the calibration process, one can leverage the known roughness of S&P500 from prior research [41] by specifying that the market factor f_0 is directly sourced from the S&P500 returns time series. It is worth noticing that the calibration procedure, explained and implemented above, yields a time series f_0 that exhibits a high correlation with S&P500 returns (with a correlation coefficient of 0.93). Additionally, the time series of Ω_0 and ω_0 derived from the S&P500 returns are also correlated with the time series obtained from the calibration, with correlation coefficients of 0.78 and 0.61, respectively. The roughness of Ω_0 and ω_0 derived from the S&P500 is then similar to those estimated above. Indeed, using the GMM from Wu *et al.*, we find $H_{\Omega_0^{S\&P}} = 0.16$ and $H_{\omega_0^{S\&P}} = 0.35$ which remain consistent with our main conjecture since both values are greater than 0.1.

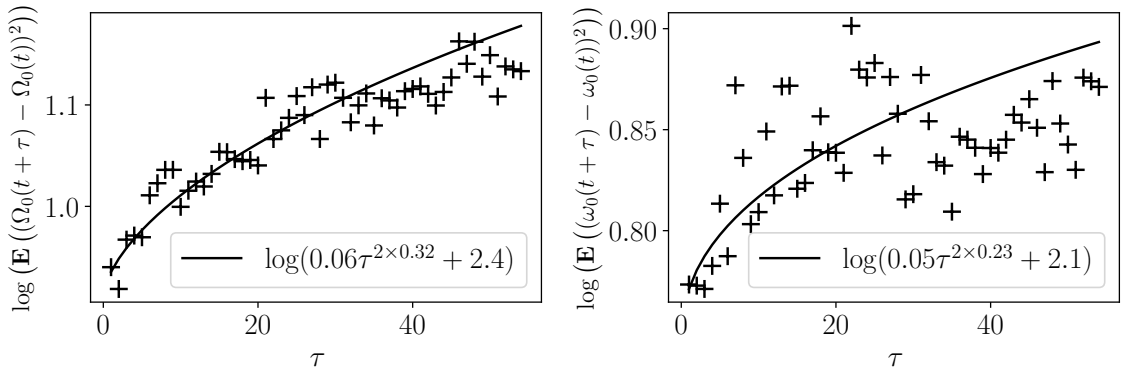


Figure 6.9: Estimation of the roughness of the market log-volatility modes Ω_0 and ω_0 using the increments second moments (see Equation (6.6.1)). The scatter plots depict the log-volatility increments second moments $(\tau \to \log\left(\mathbb{E}(|\log(\sigma_{t+\tau}) - \log(\sigma_t)|^2)\right))$ for daily estimation of the log-volatility modes $\sigma = \Omega_0$ (left plot) and $\sigma = \omega_0$ (right plot). Ω_0 and ω_0 were derived from the NFM calibration on daily data, imposing f_0 to be the S&P500 returns. The plain line is the fit of Equation (6.6.1), $\tau \to \log(A\tau^{2H} + 2C)$, where H is the Hurst exponent reporting on the roughness.

Given that these time series derived from the S&P500 exhibit less noise, we also attempted to estimate the roughness following the method of Gatheral *et al.*. As explained in Section 6.1.2, this

method estimates the roughness of a log-volatility σ by regressing linearly $\log\left(\mathbb{E}(|\log(\sigma_{t+\tau}) - \log(\sigma_t)|^q)\right)$ against $\log(\tau)$. Considering the persistent noise in our data, we opted to fit the parameters (A, H, C) to the increments of the log-volatility using the following model:

$$\log\left(\mathbb{E}(|\log(\sigma_{t+\tau}) - \log(\sigma_t)|^2)\right) = \log(A\tau^{2H} + 2C) \tag{6.6.1}$$

where H is the Hurst exponent reporting on the roughness of log-volatility σ. The estimation of H_{Ω_0} and H_{ω_0}, illustrated in Figure 6.9, yields $H_{\Omega_0} = 0.3$ and $H_{\omega_0} = 0.2$. These findings align with our main conjecture and are in agreement with the results obtained using the GMM, albeit slightly higher. Using Gatheral *et al.* method, we also estimated the roughness of the factor residuals $(\omega_k)_{k\in[\![0,9]\!]}$ and the idiosyncratic residuals $(\tilde{\omega}_i)_{i\in[\![1,323]\!]}$. The resulting histograms are shown in Figure 6.10. Similar to previous observations, the idiosyncratic residuals log-volatility, $(\tilde{\omega}_i)_{i\in[\![1,323]\!]}$, generally appear to be highly rough, although some instances exhibit higher Hurst exponents. Regarding the factor residuals $(\omega_k)_{k\in[\![0,9]\!]}$, except for the residual log-volatility of the market mode ω_0, their roughness estimation leads to very low Hurst exponent. We attribute these results to the fact that while some factor residuals are indeed highly rough, the lack of intraday insights in daily data may hinder a robust estimation. In the subsequent section, we assess log-volatility roughness using intraday data to potentially address this limitation.

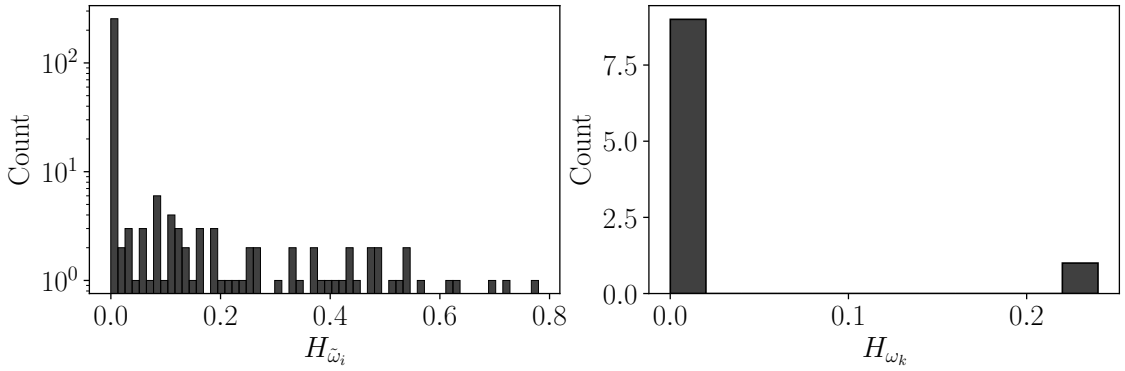


Figure 6.10: Hurst exponent estimation using the calibration method of Equation (6.6.1) on the time series of the idiosyncratic residuals $(\tilde{\omega}_i)_{i\in[\![1,323]\!]}$ (left plot) and of the factor residuals $(\omega_k)_{k\in[\![0,9]\!]}$ (right plot) derived from the calibration of NFM on daily data, on 323 US stocks, from 2013 to 2023, imposing f_0 to be the S&P500 returns.

Estimation of the roughness on intraday data with GMM from Wu *et al.*

The assessment of log-volatility roughness often necessitates intraday evaluations of volatility, which can be facilitated by estimators like the Roger Satchell estimator (see Equation (1.1.2)) or the Garman-Klass estimator (see Equation (1.1.1)). For example, employing Wu *et al.* GMM, absolute daily returns of the S&P500 yield a Hurst exponent of $H^{|r|}_{S\&P500} = 0.04$, whereas both the Roger-Satchell and Garman-Klass estimators yield $H^{RS/GK}_{S\&P500} = 0.1$, aligning with previous findings ($H_{\text{S\&P500}} \sim 0.14$ as reported in [41]). To be thorough, we thus calibrated the Nested Factor model on intraday data, at the minute time scale. Subsequently, we estimated, using the General Method of Moments from Wu *et al.* [53], the roughness of the resulting log-volatility components. Especially, as before, we examined the roughness of the common log volatility component Ω_0, the residual log-volatility of the market mode ω_0, the factor residual log-volatility modes $(\omega_k)_{k\in[\![1,9]\!]}$ and the idiosyncratic residual log-volatility modes $(\tilde{\omega}_i)_{i\in[\![1,323]\!]}$.

The Hurst exponent estimation of the factor residual log-volatility modes, $(\omega_k)_{k\in[\![0,9]\!]}$, and of the idiosyncratic residuals, $(\tilde{\omega}_i)_{i\in[\![1,323]\!]}$, are presented in Figure 6.11. The analysis reveals that the Hurst exponent of the residual log-volatilities stemming from the idiosyncratic component is nearly 0, with all values falling below 0.05. This result is in line with our conjecture and the observation that individual stocks exhibit high roughness. Furthermore, certain factor residual log-volatility modes also demonstrate considerable roughness. Specifically, with the exception of the modes linked to the market, the Communication sector and the spread between the Financials and Energy sectors, all factor modes exhibit Hurst exponents below 0.008. These findings are

still consistent with our overarching investigation, as we expect only the market modes Ω_0 and ω_0 to exhibit Hurst exponents around or above 0.1. We defer the exploration of the relationship between industrial sectors and roughness to future research. Besides, notably, the roughness of the residual log-volatility of the market mode is found to be $H_{\omega_0} = 0.13 > 0.1$, aligning with our conjecture. Additionally, we assess the roughness of the common log-volatility mode and find $H_{\Omega_0} = 0.3 > 0.1$. Consequently, conversely to the previous estimations, the roughness of the index appears to be attributed to the residual log-volatility of the market mode, as it is rougher than the common volatility mode. This observation adds further evidence to the difficulty of estimating Hurst exponents.

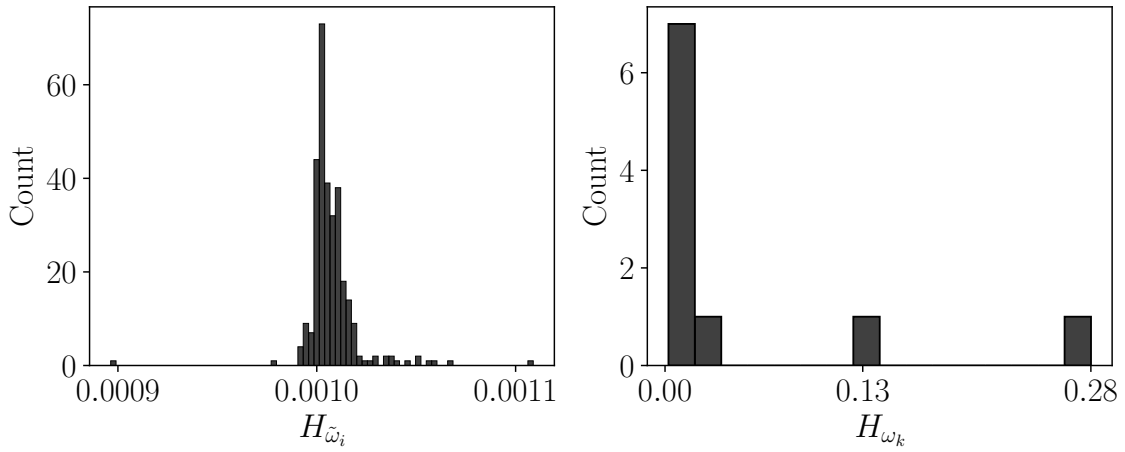


Figure 6.11: Hurst exponent estimation using the GMM of Wu *et al.* [53] on the time series of the idiosyncratic residuals $(\tilde{\omega}_i)_{i\in[\![1,323]\!]}$ (left plot) and of the factors residuals $(\omega_k)_{k\in[\![0,9]\!]}$ (right plot) derived from the NFM calibration on intraday data (1min bin), on 323 US stocks, from 2013 to 2023.

Discussion

Let us summarize the key findings obtained from the empirical calibration of the Nested Factor model incorporating SfBM components.

Our investigation indicates that the residual log-volatility modes of the stock idiosyncratic terms generally demonstrate a highly rough trajectory, characterized by a Hurst exponent close to 0. This observation remains consistent across both daily and intraday data. This finding supports the conjecture that the idiosyncratic residual mode largely contributes to the roughness for individual stock trajectories. However, it is noteworthy that our analysis also identifies specific idiosyncratic residual modes with a Hurst exponent exceeding 0.1, predominantly from the energy and real estate sectors. Nonetheless, these findings remain consistent with the empirical results depicted in Figure 6.2 which shows that not all stocks exhibit $H \approx 0$.

Estimating the roughness of the factor residual modes proves to be particularly challenging, displaying notable variability across different estimations or data frequencies. However, a key finding emerges regarding the roughness of the market mode residual, ω_0. Across all our estimations, this component consistently demonstrates a Hurst exponent greater than 0.1. Additionally, to provide a comprehensive analysis of the market mode, we estimated the roughness of the common log-volatility mode Ω_0, which consistently manifests a Hurst exponent greater than 0.09. Thus, our estimation the roughness of the market log-volatility modes align with both our main conjecture and with empirical findings of Wu *et al.*.

In Table 6.1, we compile the various Hurst exponent values for the market mode components: H_{Ω_0} and H_{ω_0}. It is evident that, in all our experiments, both market modes exhibit a Hurst exponent exceeding 0.09.

Data sample	Estimation method	H_{Ω_0}	H_{ω_0}
Daily data	GMM from Wu *et al.* [53]	0.09	0.28
Daily data sourcing f_0 from S&P500	GMM from Wu *et al.* [53]	0.16	0.35
Daily data sourcing f_0 from S&P500	Log-volatility increments moments (Equation (6.6.1))	0.3	0.2
Intraday data	GMM from Wu *et al.* [53]	0.3	0.13

Table 6.1: Estimation of the Hurst exponent of the common log volatility component Ω_0 and of the residual log-volatility of the market mode ω_0 for different methods and different data sample.

Conclusion

Starting from the intriguing observation that, although indices are made of stocks, indices log-volatility are less rough than stocks log-volatility and building on the idea that the mixture of several signals with different roughness would report on the roughest component, we postulate that the roughness of the stocks has to come from an idiosyncratic component.

Revisiting the Nested Factor model with Stationary fractional Brownian Motion as log-volatilities, we investigated our conjecture concomitantly via analytical study, numerical simulations and empirical calibration.

Already proven to reproduce many of empirical characteristics of joint stock returns, and notably the dependence of stock returns volatility on market returns, the Nested Factor model also proves here that it reconciles the difference of roughness between the log-volatility of indices and that of stocks.

The calibration of the model on empirical data and the estimation of the roughness of each log-volatility component of the model corroborate that the roughness of the log-volatility of stocks primarily originates from the residual log-volatility of the stock returns idiosyncratic component. Specifically, we observe that the log-volatility components driving the market mode are less rough ($H \gtrsim 0.1$) compared to the residual log-volatility modes of the stock idiosyncratic terms ($H \approx 0$). Consequently, the empirical roughness of indices is well-accounted for by the components of the market mode ($H \gtrsim 0.1$), as the rougher idiosyncratic contribution diminishes when averaging across stocks to reconstruct the indices trajectory.

Take Home Message

- Empirical findings indicate that the log-volatility of individual stocks is rougher ($H \approx 0.$) compared to that of indices ($H \gtrsim 0.1$) (see also Wu *et al.* (2022)).
- **Initial intuition**: the roughness of the combination of two signals with differing levels of roughness reflects the rougher signal.
- **Main conjecture**: factor models in which the idiosyncratic component exhibits significant roughness, while the factor terms are comparatively smoother, could replicate the empirical findings of Wu *et al.* (2022). Specifically, we conjecture that the rougher idiosyncratic component would manifest in individual trajectories but would be averaged out in the index trajectory, which would then predominantly reflect the factor's roughness.
- Due to its established advantages, we chose to investigate the ability of the Nested Factor model, incorporating SfBM as log-volatility modes, to replicate the empirical observations of differing levels of roughness between stocks and indices.
- Investigation, which involved concomitantly numerical simulations, analytical developments and empirical studies, consistently supports our main conjecture.
- Notably, empirical calibration shows
  - market mode roughness H_{Ω_0} and H_{ω_0} greater than 0.09
  - idiosyncratic residual roughness (generally) $H_{\tilde{\omega}_i} \approx 0.$

Conclusion

To succeed, jump as quickly at opportunities as you do at conclusions.

Benjamin Franklin

Overview of the results

This thesis aimed to advance the field of multivariate volatility modeling by examining the stylized facts that emerge when multiple assets are considered and by proposing frameworks to replicate these phenomena. To achieve this, we conducted concurrent empirical and theoretical work. Despite the fact that our approach involved going back and forth between data and model development, this section delineates the main empirical findings and the key theoretical contributions separately for enhanced clarity.

In addition to the goal of replicating, in a multidimensional context, the well-established stylized facts of financial time series –such as fat-tailed distributions, volatility clustering, and univariate time reversal asymmetry– our work was further motivated by the identification of two additional empirical characteristics: the time-reversal asymmetry between pairs of assets (discussed in Chapter 1) and the intriguing nature of price (co-)jumps (outlined in Chapter 5).

Indeed, our study elucidated the cross-leverage and cross-Zumbach effects observed between assets. These cross-time-reversal asymmetry effects further prompted the adoption of a path-dependent model to delineate cross-asset feedback.

Additionally, our data-driven study of financial price jumps of 300 US stocks over an 8-year period, employing an unsupervised classification method based on wavelet coefficients that we developed, yielded multifaceted results. Specifically, our method enables investigation into the reflexive nature, the mean-reversion strength and the trend-aligned behavior of price jumps and co-jumps. Consistently with prior research [5, 7, 6], we found that not all price jumps are related to news events. In fact, we demonstrated that a majority of price jumps and co-jumps, including significant instances, originate from endogenous mechanisms and contagion effects. The temporal clustering of co-jumps, observed in Chapters 1 and 5, further supports the notion of propagation across stocks.

These empirical findings further motivated the development of volatility models in multiple dimensions. Given their ability to replicate the majority of stylized facts observed in single-asset financial time series, QHawkes models were a relevant choice for investigating multivariate volatility and thus constituted the main focus of our analysis.

Chapter 2 provided a comprehensive overview of the model and of its principal properties. Notably, the analysis of the stationary properties in one dimension revealed that the quadratic feedback loop of the QHawkes model achieves a balance between inhibiting and exciting realisations, thereby generating extreme events while maintaining stability (see also [69, 63]).

Subsequently, we extended the QHawkes process to multiple dimensions, encompassing several assets. We specifically outlined the feedback mechanisms wherein activity in one asset triggers increased activity in others. These feedback mechanisms were formulated to replicate the cross-time-reversal asymmetry observed in Chapter 1. Moreover, considering the significant occurrence

of co-jumps and their often endogenous nature, we developed two frameworks: a multivariate independent QHawkes process, which remains tractable but may have limitations in replicating co-jump occurrences, and a bivariate QHawkes process that ensures multiple processes move simultaneously at the cost of increased complexity. We compiled our theoretical results in Chapter 3, demonstrating that our frameworks enable the replication of empirical characteristics of financial time series, particularly the fat-tailed returns distribution.

The empirical calibration of the model, detailed in Chapter 4, focused on characterizing the linear and quadratic kernels of the QHawkes. The findings consistently indicated that endogeneity primarily stems from the long-range self-linear feedback. In the case of a factor model encompassing MQHawkes, our calibration demonstrated that, although most of the index dynamics were filtered out from the stock idiosyncratic dynamics, the index's past negative returns and past volatility still tend to increase the stock's idiosyncratic activity.

Finally, to be thorough, in Chapter 6, we delved into multivariate volatility through the lens of a specific factor model: the Nested Factor Model. This model offered a suitable framework to comprehend the empirical observations by Wu, Bacry and Muzy [53], that stock log-volatility tends to exhibit more roughness compared to that of indices. Our analysis of the Nested Factor Model, employing Stationary Fractional Brownian Motions as log-volatilities, through theoretical, numerical, and empirical scrutiny, revealed that the roughness of stocks emanates from the residual log-volatility of the idiosyncratic term of returns. The roughness of the idiosyncratic log-volatility of the stock is apparent when it is juxtaposed with smoother paths, like indices contributions, but vanishes when averaged across stock paths to construct that of the index (market). Correspondingly, the log-volatility modes of "market" (indices) exhibit smoother paths, a characteristic that emerges only when the rough idiosyncratic components of individual stocks are averaged out.

Main Results

- **Main empirical results**
  - Time-reversal asymmetry effects manifest not only within individual asset, but also across assets.
  - Many co-jumps, even large ones, seem to arise due to contagion mechanisms and endogenous feedback.
- **Main theoretical results**
  - The quadratic extension of the Hawkes processes allows for the generation of extreme events while maintaining the stability properties of the linear Hawkes model, even in regimes where the endogeneity ratio exceeds 1.
  - In multi-dimensions, the QHawkes model preserves the notable properties of the univariate model that are, the volatility clustering, the generation of extreme events (fat-tailed distribution) and the (cross)-time reversal asymmetry.
  - The calibration of the MQHawkes revealed that memory kernels exhibit long-range memory and that a significant portion of endogeneity originates from the linear feedback of the asset's volatility on itself.
  - The Nested Factor Model, incorporating Stationary fractional Brownian motions as log-volatility modes, offers a well-suited framework for investigating multivariate volatility, particularly, the intricate relationship between stocks and indices. Indeed, this model replicates all empirical characteristics of multivariate financial returns, including the differing levels of roughness observed between stocks and indices.
- **Technical outcomes**: We introduced an unsupervised approach to investigate the nature of bursts of activity. While it was applied to financial price jumps in this context, it can be adapted to study bursts of activity in other fields (for more comprehensive details, refer to Morel's GitHub tutorial [127]).

Future directions

Throughout the chapters, we have identified several questions and limitations that remain unresolved. This final section highlights the most intriguing remaining inquiries.

This thesis investigated the reflexive nature of co-jumps. However, scrutinizing at higher frequency co-jumps identified as endogenous would enhance our understanding of these dynamics and facilitate the identification of pivotal stocks driving contagion. Additionally, investigating whether multivariate processes with independent QHawkes processes can accurately replicate the observed occurrence of co-jumps and their clustering would provide further insights into the micro-structural origins of co-jumps. Future research endeavors could address this question by replicating the experiments conducted by Lillo *et al.* [54]. This investigation would determine whether independent jump processes are sufficient or if a dedicated "co-jumps" process is necessary.

Besides from a technical standpoint, implementing the calibration of MQHawkes has proven to be particularly challenging and non-robust, especially when attempting to calibrate for more than two assets simultaneously. In this thesis, we focused on pairs of assets, as this approach still allowed us to characterize all the main feedback mechanisms. Nevertheless, the development of a robust calibration method (potentially using Fourier domain techniques) would undoubtedly enhance the applicability of this model which exhibits intriguing properties.

Main Remaining Questions

- **Understand further the microstructure of co-jumps**
  - How do they originate at the microstructural level, on an event-by-event scale?
  - Can independent QHawkes processes replicate the empirical abundance co-jumps?
- **Technical improvements can be engineered**
  Implementation of a robust calibration method for the (M)QHawkes model, enabling recovery of the underlying (M)QHawkes dynamics from aggregated data.

Bibliography

[1] Charles Bartlett. The financial crisis, then and now: Ancient rome and 2008 ce, 2018.

[2] Eugene F Fama. Efficient capital markets. *Journal of finance*, 25(2):383–417, 1970.

[3] Robert J Shiller et al. Do stock prices move too much to be justified by subsequent changes in dividends? 1981.

[4] Cecilia Aubrun, Rudy Morel, Michael Benzaquen, and Jean-Philippe Bouchaud. Riding wavelets: A method to discover new classes of price jumps. *arXiv preprint arXiv:2404.16467*, 2024.

[5] David M Cutler, James M Poterba, and Lawrence H Summers. *What moves stock prices?*, volume 487. National Bureau of Economic Research Cambridge, Massachusetts, 1988.

[6] Riccardo Marcaccioli, Jean-Philippe Bouchaud, and Michael Benzaquen. Exogenous and endogenous price jumps belong to different dynamical classes. *Journal of Statistical Mechanics: Theory and Experiment*, 2022(2):023403, 2022.

[7] Armand Joulin, Augustin Lefevre, Daniel Grunberg, and Jean-Philippe Bouchaud. Stock price jumps: news and volume play a minor role. *arXiv preprint arXiv:0803.1769*, 2008.

[8] George Soros. The alchemy of finance: Reading the mind of the market by george soros (1994-05-06).

[9] Vladimir Filimonov and Didier Sornette. Quantifying reflexivity in financial markets: Toward a prediction of flash crashes. *Physical Review E*, 85(5):056108, 2012.

[10] Stephen J Hardiman, Nicolas Bercot, and Jean-Philippe Bouchaud. Critical reflexivity in financial markets: a hawkes process analysis. *The European Physical Journal B*, 86(10):1–9, 2013.

[11] Alexander Wehrli and Didier Sornette. The excess volatility puzzle explained by financial noise amplification from endogenous feedbacks. *Scientific reports*, 12(1):18895, 2022.

[12] Rémy Chicheportiche and Jean-Philippe Bouchaud. The fine-structure of volatility feedback i: Multi-scale self-reflexivity. *Physica A: Statistical Mechanics and its Applications*, 410:174–195, 2014.

[13] Emmanuel Bacry, Iacopo Mastromatteo, and Jean-François Muzy. Hawkes processes in finance. *Market Microstructure and Liquidity*, 1(01):1550005, 2015.

[14] Antoine Fosset, Jean-Philippe Bouchaud, and Michael Benzaquen. Non-parametric estimation of quadratic hawkes processes for order book events. *The European Journal of Finance*, pages 1–16, 2021.

[15] Alan G Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971.

[16] Ivan Osorio, Mark G Frei, Didier Sornette, John Milton, and Ying-Cheng Lai. Epileptic seizures: quakes of the brain? *Physical Review E*, 82(2):021919, 2010.

[17] Didier Sornette and Ivan Osorio. Prediction. *chapter in "Epilepsy: The Intersection of Neurosciences, Biology, Mathematics, Physics and Engineering", Editors: Osorio I., Zaveri H.P., Frei M.G., Arthurs S., CRC Press, Taylor & Francis Group, pp. 203-237*, 2010.

[18] George O Mohler, Martin B Short, P Jeffrey Brantingham, Frederic Paik Schoenberg, and George E Tita. Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, 106(493):100–108, 2011.

[19] Laurent Bonnasse-Gahot, Henri Berestycki, Marie-Aude Depuiset, Mirta B Gordon, Sebastian Roché, Nancy Rodriguez, and Jean-Pierre Nadal. Epidemiological modelling of the 2005 french riots: a spreading wave and the role of contagion. *Scientific reports*, 8(1):1–20, 2018.

[20] Didier Sornette, Fabrice Deschâtres, Thomas Gilbert, and Yann Ageon. Endogenous versus exogenous shocks in complex networks: An empirical test using book sale rankings. *Physical Review Letters*, 93(22):228701, 2004.

[21] Fabrice Deschatres and Didier Sornette. Dynamics of book sales: Endogenous versus exogenous shocks in complex networks. *Physical Review E*, 72(1):016112, 2005.

[22] Riley Crane and Didier Sornette. Robust dynamic classes revealed by measuring the response function of a social system. *Proceedings of the National Academy of Sciences*, 105(41):15649–15653, 2008.

[23] Jean-Philippe Bouchaud, Julius Bonart, Jonathan Donier, and Martin Gould. *Trades, quotes and prices: financial markets under the microscope*. Cambridge University Press, 2018.

[24] Maria Flora and Roberto Renò. V-shapes. In *Proceedings of the EUROFIDAI-ESSEC Paris December Finance Meeting*, 2022.

[25] Kang Gao, Perukrishnen Vytelingum, Stephen Weston, Wayne Luk, and Ce Guo. High-frequency financial market simulation and flash crash scenarios analysis: an agent-based modelling approach. *arXiv preprint arXiv:2208.13654*, 2022.

[26] Andrei Kirilenko, Albert S Kyle, Mehrdad Samadi, and Tugkan Tuzun. The flash crash: High-frequency trading in an electronic market. *The Journal of Finance*, 72(3):967–998, 2017.

[27] CFTC SEC. Findings regarding the market events of may 6, 2010. *Washington DC*, 2010.

[28] Kim Christensen, Roel Oomen, and Roberto Renò. The drift burst hypothesis. *Journal of Econometrics*, 227(2):461–497, 2022.

[29] Neil Johnson, Guannan Zhao, Eric Hunsader, Jing Meng, Amith Ravindar, Spencer Carran, and Brian Tivnan. Financial black swans driven by ultrafast machine ecology. *arXiv preprint arXiv:1202.1448*, 2012.

[30] Neil Johnson and Brian Tivnan. Flash crash phenomena and a taxonomy of extreme behaviors. *Eur. Phys. J. Special Topics*, 205:65–78, 2012.

[31] Anton Golub, John Keane, and Ser-Huang Poon. High frequency trading and mini flash crashes. *arXiv preprint arXiv:1211.6667*, 2012.

[32] Bruno Biais, Thierry Foucault, et al. Hft and market quality. *Bankers, Markets & Investors*, 128(1):5–19, 2014.

[33] Nanex LLC. Ongoing research - market events and phenomena, 2012.

[34] Christos Floros. Modelling volatility using high, low, open and closing prices: evidence from four s&p indices. *International Research Journal of Finance and Economics*, 28:198–206, 2009.

[35] Louis Bachelier. Théorie de la spéculation. In *Annales scientifiques de l'École normale supérieure*, volume 17, pages 21–86, 1900.

[36] Steven L Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The review of financial studies*, 6(2):327–343, 1993.

[37] Leif BG Andersen and Vladimir V Piterbarg. Moment explosions in stochastic volatility models. *Finance and Stochastics*, 11(1):29–50, 2007.

[38] Jean-Philippe Bouchaud, Andrew Matacz, and Marc Potters. Leverage effect in financial markets: The retarded volatility model. *Physical review letters*, 87(22):228701, 2001.

[39] Gilles Zumbach. Time reversal invariance in finance. *Quantitative Finance*, 9(5):505–515, 2009.

[40] Gilles Zumbach. Volatility conditional on price trends. *Quantitative Finance*, 10(4):431–442, 2010.

[41] Jim Gatheral, Thibault Jaisson, and Mathieu Rosenbaum. Volatility is rough. *Quantitative finance*, 18(6):933–949, 2018.

[42] Omar El Euch and Mathieu Rosenbaum. The characteristic function of rough heston models. *Mathematical Finance*, 29(1):3–38, 2019.

[43] Robert F Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the econometric society*, pages 987–1007, 1982.

[44] Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.

[45] Christian Francq and Jean-Michel Zakoian. *GARCH models: structure, statistical inference and financial applications*. John Wiley & Sons, 2019.

[46] Emmanuel Bacry and Jean François Muzy. Log-infinitely divisible multifractal processes. *Communications in Mathematical Physics*, 236:449–475, 2003.

[47] 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):056121, 2002.

[48] Jean-François Muzy, Jean Delour, and Emmanuel Bacry. Modelling fluctuations of financial time series: from cascade process to stochastic volatility model. *The European Physical Journal B-Condensed Matter and Complex Systems*, 17:537–548, 2000.

[49] Emmanuel Bacry, Jean Delour, and Jean-François Muzy. Modelling financial time series using multifractal random walks. *Physica A: statistical mechanics and its applications*, 299(1-2):84–92, 2001.

[50] Shoaleh Ghashghaie, Wolfgang Breymann, Joachim Peinke, Peter Talkner, and Yadollah Dodge. Turbulent cascades in foreign exchange markets. *Nature*, 381(6585):767–770, 1996.

[51] Benoit B Mandelbrot, Adlai J Fisher, and Laurent E Calvet. A multifractal model of asset returns. 1997.

[52] Benoit B Mandelbrot. A multifractal walk down wall street. *Scientific American*, 280(2):70–73, 1999.

[53] Peng Wu, Jean-François Muzy, and Emmanuel Bacry. From rough to multifractal volatility: The log s-fbm model. *Physica A: Statistical Mechanics and its Applications*, 604:127919, 2022.

[54] Giacomo Bormetti, Lucio Maria Calcagnile, Michele Treccani, Fulvio Corsi, Stefano Marmi, and Fabrizio Lillo. Modelling systemic price cojumps with hawkes factor models. *Quantitative Finance*, 15(7):1137–1156, 2015.

[55] Lucio Maria Calcagnile, Giacomo Bormetti, Michele Treccani, Stefano Marmi, and Fabrizio Lillo. Collective synchronization and high frequency systemic instabilities in financial markets. *Quantitative Finance*, 18(2):237–247, 2018.

[56] Cécilia Aubrun, Michael Benzaquen, and Jean-Philippe Bouchaud. Multivariate quadratic hawkes processes—part i: theoretical analysis. *Quantitative Finance*, 23(5):741–758, 2023.

[57] Fabrizio Lillo and Rosario N Mantegna. Symmetry alteration of ensemble return distribution in crash and rally days of financial markets. *The European Physical Journal B-Condensed Matter and Complex Systems*, 15:603–606, 2000.

[58] Pierre Cizeau, Marc Potters, and Jean-Philippe Bouchaud. Correlation structure of extreme stock returns. 2001.

[59] Romain Allez and Jean-Philippe Bouchaud. Individual and collective stock dynamics: intra-day seasonalities. *New Journal of Physics*, 13(2):025010, 2011.

[60] Rémy Chicheportiche and Jean-Philippe Bouchaud. The joint distribution of stock returns is not elliptical. *International Journal of Theoretical and Applied Finance*, 15(03):1250019, 2012.

[61] Rémy Chicheportiche and J-P Bouchaud. A nested factor model for non-linear dependencies in stock returns. *Quantitative Finance*, 15(11):1789–1804, 2015.

[62] Cecilia Aubrun, Michael Benzaquen, and Jean-Philippe Bouchaud. On hawkes processes with infinite mean intensity. *arXiv preprint arXiv:2112.14161*, 2021.

[63] Cecilia Aubrun, Michael Benzaquen, and Jean-Philippe Bouchaud. Hawkes processes with infinite mean intensity. *Physical Review E*, 105(3):L032101, 2022.

[64] Misha Koshelev. Testing earthquake damage prevention methods with simulated fractal earthquakes. *physical Research*, 2853:2862, 1981.

[65] Jiancang Zhuang, David S Harte, Maximilian J Werner, Sebastian Hainzl, and Shiyong Zhou. Basic models of seismicity: Temporal models. 2012.

[66] Pierre Brémaud and Laurent Massoulié. Hawkes branching point processes without ancestors. *Journal of applied probability*, 38(1):122–135, 2001.

[67] A Saichev and Didier Sornette. Superlinear scaling of offspring at criticality in branching processes. *Physical Review E*, 89(1):012104, 2014.

[68] Kiyoshi Kanazawa and Didier Sornette. Exact and robust asymptotic solutions to nonlinear hawkes processes: power law exponents of intensity distributions and methods. *arXiv preprint arXiv:2110.01523*, 2021.

[69] Kiyoshi Kanazawa and Didier Sornette. Ubiquitous power law scaling in nonlinear self-excited hawkes processes. *Physical review letters*, 127(18):188301, 2021.

[70] Pierre Blanc, Jonathan Donier, and J-P Bouchaud. Quadratic hawkes processes for financial prices. *Quantitative Finance*, 17(2):171–188, 2017.

[71] Thibault Jaisson, Mathieu Rosenbaum, et al. Limit theorems for nearly unstable hawkes processes. *Annals of Applied Probability*, 25(2):600–631, 2015.

[72] Yosihiko Ogata. On lewis' simulation method for point processes. *IEEE transactions on information theory*, 27(1):23–31, 1981.

[73] Rémy Chicheportiche and Jean-Philippe Bouchaud. Goodness-of-fit tests with dependent observations. *Journal of Statistical Mechanics: Theory and Experiment*, 2011(09):P09003, 2011.

[74] Emmanuel Bacry, Thibault Jaisson, and Jean-François Muzy. Estimation of slowly decreasing hawkes kernels: application to high-frequency order book dynamics. *Quantitative Finance*, 16(8):1179–1201, 2016.

[75] Alan G Hawkes. Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, 33(3):438–443, 1971.

[76] Qiong-Tao Xie. New quasi-exactly solvable periodic potentials. *Journal of Physics A: Mathematical and Theoretical*, 44(28):285302, 2011.

[77] Yosihiko Ogata. Estimators for stationary point processes. *Ann. Inst. Statist. Math*, 30(Part A):243–261, 1978.

[78] Yosihiko Ogata and Hirotugu Akaike. On linear intensity models for mixed doubly stochastic poisson and self-exciting point processes. *Selected Papers of Hirotugu Akaike*, pages 269–274, 1998.

[79] Alejandro Veen and Frederic P Schoenberg. Estimation of space–time branching process models in seismology using an em–type algorithm. *Journal of the American Statistical Association*, 103(482):614–624, 2008.

[80] Erik Lewis and George Mohler. A nonparametric em algorithm for multiscale hawkes processes. *Journal of nonparametric statistics*, 1(1):1–20, 2011.

[81] David Marsan and Olivier Lengline. Extending earthquakes' reach through cascading. *Science*, 319(5866):1076–1079, 2008.

[82] Niels Richard Hansen, Patricia Reynaud-Bouret, and Vincent Rivoirard. Lasso and probabilistic inequalities for multivariate point processes. 2015.

[83] Patricia Reynaud-Bouret and Sophie Schbath. Adaptive estimation for hawkes processes; application to genome analysis. 2010.

[84] Emmanuel Bacry and Jean-François Muzy. Hawkes model for price and trades high-frequency dynamics. *Quantitative Finance*, 14(7):1147–1166, 2014.

[85] Emmanuel Bacry and Jean-Francois Muzy. Second order statistics characterization of hawkes processes and non-parametric estimation. *arXiv preprint arXiv:1401.0903*, 2014.

[86] Emmanuel Bacry, Khalil Dayri, and Jean-François Muzy. Non-parametric kernel estimation for symmetric hawkes processes. application to high frequency financial data. *The European Physical Journal B*, 85:1–12, 2012.

[87] Stephen J Hardiman and Jean-Philippe Bouchaud. Branching-ratio approximation for the self-exciting hawkes process. *Physical Review E*, 90(6):062807, 2014.

[88] Massil Achab, Emmanuel Bacry, Stéphane Gaıffas, Iacopo Mastromatteo, and Jean-François Muzy. Uncovering causality from multivariate hawkes integrated cumulants. pages 1–10, 2017.

[89] Emmanuel Bacry, Sylvain Delattre, Marc Hoffmann, and Jean-François Muzy. Modelling microstructure noise with mutually exciting point processes. *Quantitative finance*, 13(1):65–77, 2013.

[90] Pierre Blanc, Rémy Chicheportiche, and Jean-Philippe Bouchaud. The fine structure of volatility feedback ii: Overnight and intra-day effects. *Physica A: Statistical Mechanics and its Applications*, 402:58–75, 2014.

[91] Didier Sornette. Endogenous versus exogenous origins of crises. *Extreme events in nature and society*, pages 95–119, 2006.

[92] Per Bak, Kan Chen, José Scheinkman, and Michael Woodford. Aggregate fluctuations from independent sectoral shocks: self-organized criticality in a model of production and inventory dynamics. *Ricerche economiche*, 47(1):3–30, 1993.

[93] José Moran and Jean-Philippe Bouchaud. May's instability in large economies. *Physical Review E*, 100(3):032307, 2019.

[94] José Moran, Frank P Pijpers, Utz Weitzel, Jean-Philippe Bouchaud, and Debabrata Panja. Temporal criticality in socio-technical systems. *arXiv preprint arXiv:2307.03546*, 2023.

[95] Ray C Fair. Events that shook the market. *The Journal of Business*, 75(4):713–731, 2002.

[96] P Bäk. How nature works: The science of self-organized criticality (copernicus, new york). 1996.

[97] Per Bak and Maya Paczuski. Complexity, contingency, and criticality. *Proceedings of the National Academy of Sciences*, 92(15):6689–6696, 1995.

[98] Didier Sornette and Agnès Helmstetter. Endogenous versus exogenous shocks in systems with memory. *Physica A: Statistical Mechanics and its Applications*, 318(3-4):577–591, 2003.

[99] Armand Joulin, Augustin Lefevre, Daniel Grunberg, and Jean-Philippe Bouchaud. Stock price jumps: news and volume play a minor role. *Wilmott Magazine*, 46, 2008.

[100] Qianyun Wu, Yukie Sano, Hideki Takayasu, and Misako Takayasu. Classification of endogenous and exogenous bursts in collective emotions based on weibo comments during covid-19. *Scientific Reports*, 12(1):3120, 2022.

[101] Joan Bruna and Stéphane Mallat. Invariant scattering convolution networks. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1872–1886, 2013.

[102] Kris Boudt, Christophe Croux, and Sébastien Laurent. Robust estimation of intraweek periodicity in volatility and jump detection. *Journal of Empirical Finance*, 18(2):353–367, 2011.

[103] SA Probert and YH Song. Detection and classification of high frequency transients using wavelet analysis. In *IEEE Power Engineering Society Summer Meeting,*, volume 2, pages 801–806. IEEE, 2002.

[104] Do-In Kim, Tae Yoon Chun, Sung-Hwa Yoon, Gyul Lee, and Yong-June Shin. Wavelet-based event detection method using pmu data. *IEEE Transactions on Smart grid*, 8(3):1154–1162, 2015.

[105] Luis Rueda, Alben Cardenas, Sousso Kelouwani, and Kodjo Agbossou. Transient event classification based on wavelet neuronal network and matched filters. In *IECON 2018-44th Annual Conference of the IEEE Industrial Electronics Society*, pages 832–837. IEEE, 2018.

[106] Elena Cuoco, Massimiliano Razzano, and Andrei Utina. Wavelet-based classification of transient signals for gravitational wave detectors. In *2018 26th European Signal Processing Conference (EUSIPCO)*, pages 2648–2652. IEEE, 2018.

[107] Bernhard Schölkopf, Alexander Smola, and Klaus-Robert Müller. Kernel principal component analysis. In *International conference on artificial neural networks*, pages 583–588. Springer, 1997.

[108] Rudy Morel, Gaspar Rochette, Roberto Leonarduzzi, Jean-Philippe Bouchaud, and Stéphane Mallat. Scale dependencies and self-similarity through wavelet scattering covariance. *arXiv preprint arXiv:2204.10177*, 2022.

[109] Rudy Morel, Stéphane Mallat, and Jean-Philippe Bouchaud. Path shadowing monte-carlo. *arXiv preprint arXiv:2308.01486*, 2023.

[110] Austin Gerig. High-frequency trading synchronizes prices in financial markets. *arXiv preprint arXiv:1211.1919*, 2012.

[111] Theodore Edward Harris et al. *The theory of branching processes*, volume 6. Springer Berlin, 1963.

[112] Jean-Philippe Bouchaud. The endogenous dynamics of markets: Price impact, feedback loops and instabilities. *Lessons from the credit crisis*, pages 345–74, 2011.

[113] Antoine Fosset, Jean-Philippe Bouchaud, and Michael Benzaquen. Endogenous liquidity crises. *Journal of Statistical Mechanics: Theory and Experiment*, 2020(6):063401, 2020.

[114] Fabienne Comte and Eric Renault. Long memory in continuous-time stochastic volatility models. *Mathematical finance*, 8(4):291–323, 1998.

[115] Mikkel Bennedsen, Asger Lunde, Pakkanen, and Mikko S. Decoupling the short-and long-term behavior of stochastic volatility. *Journal of Financial Econometrics*, 20(5):961–1006, 2022.

[116] Christian Bayer, Peter Friz, and Jim Gatheral. Pricing under rough volatility. *Quantitative Finance*, 16(6):887–904, 2016.

[117] Masaaki Fukasawa. Volatility has to be rough. *Quantitative finance*, 21(1):1–8, 2021.

[118] Giulia Livieri, Saad Mouti, Andrea Pallavicini, and Mathieu Rosenbaum. Rough volatility: evidence from option prices. *IISE transactions*, 50(9):767–776, 2018.

[119] Anine Eg Bolko, Kim Christensen, Bezirgen Veliyev, and Mikko Pakkanen. Roughness in spot variance? a gmm approach for estimation of fractional log-normal stochastic volatility models using realized measures. *A GMM Approach for Estimation of Fractional Log-Normal Stochastic Volatility Models Using Realized Measures (October 9, 2020)*, 2020.

[120] Masaaki Fukasawa, Tetsuya Takabatake, and Rebecca Westphal. Is volatility rough? *arXiv preprint arXiv:1905.04852*, 2019.

[121] Emmanuel Bacry, Jean Delour, and Jean-François Muzy. Multifractal random walk. *Physical Review E*, 64(2):026103, 2001.

[122] Martin Forde, Masaaki Fukasawa, Stefan Gerhold, and Benjamin Smith. The riemann–liouville field and its gmc as h→ 0, and skew flattening for the rough bergomi model. *Statistics & Probability Letters*, 181:109265, 2022.

[123] Martin Forde, Masaaki Fukasawa, Stefan Gerhold, and Benjamin Smith. The rough bergomi model as h→ 0–skew flattening/blow up and non-gaussian rough volatility. *preprint*, 2020.

[124] YV Fyodorov, BA Khoruzhenko, and NJ3531684 Simm. Fractional brownian motion with hurst index h=0 and the gaussian unitary ensemble. 2016.

[125] Paul Hager and Eyal Neuman. The multiplicative chaos of h= 0 fractional brownian fields. *The Annals of Applied Probability*, 32(3):2139–2179, 2022.

[126] Marc Mézard, Giorgio Parisi, and Miguel Angel Virasoro. *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*, volume 9. World Scientific Publishing Company, 1987.

[127] Rudy Morel. Rudy morel's github, scattering spectra, jumps, 2012.

[128] Thierry Bochud and Damien Challet. Optimal approximations of power laws with exponentials: application to volatility models with long memory. *Quantitative Finance*, 7(6):585–589, 2007.

[129] Claude R Dietrich and Garry N Newsam. Fast and exact simulation of stationary gaussian processes through circulant embedding of the covariance matrix. *SIAM Journal on Scientific Computing*, 18(4):1088–1107, 1997.

[130] Circulant-matrices. `https://web.mit.edu/18.06/www/Spring17/Circulant-Matrices.pdf`. Accessed: 2022-11-30.

Glossary

TBOND is the 10 years future contract on US Treasury bond. 9

Brownian motion is a way to describe the seemingly random movement of particles or variables over time. It follows a Normal distribution. 14

E-mini index is the future contract on the S&P500 index which tracks the stock value of the 500 largest companies listed on stock exchanges in the United States. 8

Elliptic models describe return dynamics as $\mathrm{d}P^i = \sigma_i \epsilon_i$, with ϵ_i a random variable, allowing for the analytical derivation of various properties. 22

Liquidity is the ease of converting an asset or security into cash. A liquidity crisis is characterized by the vanishing of market participants, when they are to afraid to trade, at any price. 7

Market makers are particular market participants who facilitate trading in financial instruments by providing liquidity to the market (see a subsequent footnote or the glossary for the definition of liquidity). Market makers maintain an inventory of securities and are willing to buy or sell them at best quoted prices, thereby providing liquidity to the market. Their primary role is to ensure that there is enough trading activity in a particular security or asset so that buyers and sellers can transact at any time. Some exchanges also pay them to ensure enough liquidity. Market makers gain money by selling high and buying low, faster than other participants. 7

SEC is the Securities Exchange Commission. 9

tick size is minimum increment between two price changes. It depends on the traded asset. 26

Notations

	For all chapters
$\mathbb{A}$	Matrix
$\mathbb{A}^\top$	Matrix transpose
$\mathbb{A}(\tau,\tau)$	Matrix time diagonal
$\mathbb{A}_{ii}$	Matrix diagonal
$\mathbf{X}$	Vector
$(N_t^i)_{t\geq 0}$	Quadratic Hawkes process describing the time of the price changes of asset i
$(N_t^c)_{t\geq 0}$	Quadratic Hawkes process describing the times of co-jumps for 2 price processes in the correlated case
$(P_t^i)_{t\geq 0}$	Price process of the asset i at time t
$(\mathrm{d}P_t^i)_{t\geq 0}$	Price returns process of the asset i at time t
σ_t	Volatility process at time t
σ_t^{RS}	Roger Satchell estimation of the volatility process at time t
σ_t^{GK}	Garman Klass estimation of the volatility process at time t
$\sigma_t^{\mathscr{B}}$	Bachelier estimation of the volatility process at time t
ψ	Tick size
$\mathcal{C}$	two-point covariance
W, B	fractional Brownian motions of Hurst exponent $H = 1/2$
L_j^i	Leverage kernel of intensity i reflecting the feedback of the price changes of asset j
K_{jk}^i	Quadratic kernel of intensity i reflecting the feedback of the price changes of assets j and k
K_j^i	$\equiv K_{jj}^i$
k_{jk}^i	Rank one contribution of K_{jk}^i
k_j^i	Rank one contribution of K_{jj}^i
	For Chapters 1, 2 and 3
$\lambda_{i,t}$	Intensity of the QHawkes process N^i at time t
$\lambda_{i,\infty}$	Baseline intensity of the QHawkes process N^i
$\mu_{i,t}$	Intensity of the QHawkes process N^i at time t in the correlated case
$\mu_{i,\infty}$	Baseline intensity of the QHawkes process N^i in the correlated case
$n/n_H/n_Q/n_Z$	Endogeneity ratio of the process/Hawkes contribution/regular contribution/Zumbach contribution
ϵ_t	Sign of the price change at time t, $\epsilon_t \in \{-1, 1\}$
Q_{jk}^i	Same as K_{jk}^i for the correlated case
ρ	Correlation of the signs of the price changes $\mathrm{d}P^1$ and $\mathrm{d}P^2$ in the correlated case
$\mathbb{K}_\mathrm{d}$	Kernel matrix such that $\mathbb{K}_{\mathrm{d}\,ij} = K_{jj}^i$
$\mathbf{K}_\times$	Kernel vector representing cross feedback, composed by K_{jk}^i with $j < k$
$\bar{\lambda}$	Mean Intensity
$\delta_{\cdot,\cdot}$	Kronecker delta: $\delta_{x,y} = 1,$ when $x = y$ and $\delta_{x,y} = 0$ otherwise
$\delta(\cdot)$	Dirac mass
$\mathbb{C}_{ij}$	Covariance $\mathbb{C}_{ij}(\tau) := \mathbb{E}\left(\frac{\mathrm{d}N_t^i}{\mathrm{d}t}\frac{\mathrm{d}N_{t-\tau}^j}{\mathrm{d}t}\right) - \bar{\lambda}_i\bar{\lambda}_j$

$\mathcal{D}_{ijk}$	3 points correlation structure $\mathcal{D}_{ijk}(\tau_1,\tau_2) := \mathbb{E}\left(\frac{\mathrm{d}N_t^i}{\mathrm{d}t}\frac{\mathrm{d}P_{t-\tau_1}^j}{\mathrm{d}t}\frac{\mathrm{d}P_{t-\tau_2}^k}{\mathrm{d}t}\right)$
$\mathbb{D}_\times$	Square (2x2) correlation matrix such that $\mathbb{D}_{\times\, i} = \mathcal{D}_{iij}$
$\mathbb{D}_\mathrm{d}$	Square (2x2) correlation matrix such that $\mathbb{D}_{\times\, ij} = \mathcal{D}_{ijk}$ with $j \neq k$
$(\phi_\mathrm{d})_{jj}^i$	Time diagonal contribution of $(\mathbb{K}_\mathrm{d})_{jj}^i$
ϕ	Kernel from the linear feedback loop of (Q)Hawkes process, in 1D
H_j^i	Hawkes component of intensity i from feedback j; $H_j^i = n_{H,j}^i h_j(t)$
$n_{H,j}^i$	Endogeneity ratio of Hawkes component H_j^i
Z_j^i	Quadratic component of intensity i from feedback j
$a_{Z,j}^i$	Amplitude of the quadratic component Z_j^i; $a_{Z,i}^j = \sqrt{2n_Z/\omega_i}$
Y_j^i	Cross component of intensity i; $Y_j^i = a_{Z,}^i z_1 z_2$
$a_{Z,}^i$	Amplitude of the cross component of intensity i
γ	Anisotropy coefficient $\gamma \in (-2,2)$, such that $a_{Z,}^i = 2\gamma n_Z/\sqrt{\omega_1\omega_2}$
$p_\infty(r,\theta) \underset{r\to\infty}{\approx} F(\theta) r^{-\alpha}$	Joint pdf of (z_1,z_2) in polar coordinates in the case ZHawkes without Hawkes with exponential kernels.
	For Chapter 4
$r_{t,d}^i$	Returns of asset i of the minute t of day d
$\sigma_{t,d}^i$	Volatility of asset i of the minute t of day d
$r_{i,t}$	Returns of asset i at time t
$\sigma_{i,t}$	Volatility of asset i at time t
$\sigma_{j,\infty}$	Baseline volatiltiy of asset j
$K_\times^j$	In a 2D-QGARCH model, kernel reflecting the cross correlation of the two returns feedback on future volatility of asset j
$\phi_\times^j$	Time-diagonal of $K_\times^j$
ϕ_i^j	Time-diagonal of K_i^j
$\mathcal{C}_{ij}$	Covariance $\mathcal{C}_{ij}(\tau) = \mathbb{E}(\sigma_{i,t}^2 r_{j,t-\tau}^2) - \overline{\sigma_i^2}\,\overline{r_j^2}$
$\mathcal{D}_{ij}$	Three-point correlation structures $\mathcal{D}_{ij}(\tau_1,\tau_2) := \mathbb{E}\left(\left(\sigma_{i,t}^2 - \langle\sigma_{i,t}^2\rangle_t\right) r_{j,t-\tau_1} r_{j,t-\tau_2}\right)$
$\mathcal{D}_{\times j}$	Three-point correlation structure $\mathcal{D}_{\times j}(\tau_1,\tau_2) := \mathbb{E}\left(\left(\sigma_{j,t}^2 - \langle\sigma_{j,t}^2\rangle_t\right) r_{A,t-\tau_1} r_{B,t-\tau_2}\right)$
$\mathcal{V}_{ij}$	Covariance $\mathcal{V}_{ij}(\tau) := \mathbb{E}\left(\sigma_{i,t}^2 r_{j,t-\tau}\right)$
β_i	Exposure of stock i to the market
$f_{0,t}$	Market (E-mini) returns at time t
$e_{i,t}$	Idiosyncratic returns component of stock i at time t
$\sigma_{f_0,t}$	Volatility of the market (E-mini) at time t
$\sigma_{i,t}$	Volatility of the idiosyncratic returns component of stock i at time t
$\sigma_{f_0,\infty}$	Baseline volatility of the market (E-mini) at time t
$\sigma_{i,\infty}$	Baseline volatility of the idiosyncratic returns component of stock i at time t
L_0	Leverage kernel of index volatility reflecting the feedback of the index returns
K_0	Quadratic kernel of index volatility reflecting the feedback of the index returns
L_i	Leverage kernel of the volatility of the stock idiosyncratic component reflecting the feedback of the residual returns
K_i	Quadratic kernel of the volatility of the stock idiosyncratic component reflecting the feedback of the residual returns
ϕ_i	Time-diagonal of K_i
k_i	Rank-one approximation of the off-time-diagonal of K_i
$L_\times^i$	Leverage kernel of the volatility of the stock idiosyncratic component reflecting the feedback of the index returns
$K_\times^i$	Quadratic kernel of the volatility of the stock idiosyncratic component reflecting the feedback of the index returns

$\phi^i_\times$	Time-diagonal of $K^i_\times$
$k^i_\times$	Rank-one approximation of the off-time-diagonal of $K^i_\times$
	Additionally, for Chapter 6
β_i	Exposure of stock i to the market
$f_{0,t}$	Market (or index) returns at time t
$f_{k,t}$	Returns of factor term k at time t
$\epsilon_{i,t}$	Idiosyncratic returns component of stock i at time t
$\Omega_{0,1}(t)$	Common volatility modes between factors and idiosyncratic terms, at time t
$\omega_k(t)$	Residual volatility mode of factor k, at time t
$\tilde{\omega}_i(t)$	Residual volatility mode of stock i, at time t
H_ω	Hurst exponent of the log-volatility ω

Appendix A

Appendix of Chapter 3: MQHawkes definition

A.1 Condition to Guarantee Positive Sign of QHawkes Intensities in 2D Case

Inspired by the method in [90], we detail here sufficient conditions for the intensity of the process i to be positive, when considering N assets. We first consider the simple case $\mathbb{L} \equiv 0$. First, considering that the kernels are negligible up to a value q, and using a discrete approximation of the integration, we rewrite the intensity associated with asset i:

$$\lambda_{i,t} = \lambda_{i,\infty} + \sum_{j \leq k}^{N} \iint_{-\infty}^{t} K_{jk}^{i}(t-s, t-u)\, \mathrm{d}P_s^j\, \mathrm{d}P_u^k \simeq \lambda_{i,\infty} + \sum_{j \leq k}^{N} \sum_{u=1}^{q} \sum_{s=1}^{q} \mathrm{d}P_{t-u}^j K_{jk}^{i}(s,u) \mathrm{d}P_{t-s}^k$$

So we can write:

$$\lambda_{i,t} = \lambda_{i,\infty} + \boldsymbol{r_t}^\top \mathbb{K}_i \boldsymbol{r_t}$$

where

$$\mathbb{K}_i = \begin{pmatrix} K_{11}^i(1,1) & \dots & K_{11}^i(1,q) & \dots & \frac{1}{2}K_{1N}^i(1,1) & \dots & \frac{1}{2}K_{1N}^i(1,q) \\ \vdots & \ddots & \vdots & & \vdots & \ddots & \vdots \\ K_{11}^i(q,1) & \dots & K_{11}^i(q,q) & \dots & \frac{1}{2}K_{1N}^i(q,1) & \dots & \frac{1}{2}K_{1N}^i(q,q) \\ & \vdots & & \ddots & & \vdots & \\ \frac{1}{2}K_{1N}^i(1,1) & \dots & \frac{1}{2}K_{1N}^i(1,q) & \dots & K_{NN}^i(1,1) & \dots & K_{NN}^i(1,q) \\ \vdots & \ddots & \vdots & & \vdots & \ddots & \vdots \\ \frac{1}{2}K_{1N}^i(q,1) & \dots & \frac{1}{2}K_{1N}^i(q,q) & \dots & K_{NN}^i(q,1) & \dots & K_{NN}^i(q,q) \end{pmatrix}, \qquad \boldsymbol{r_t} = \begin{pmatrix} \mathrm{d}P_{t-1}^1 \\ \vdots \\ \mathrm{d}P_{t-q}^1 \\ \vdots \\ \mathrm{d}P_{t-1}^N \\ \vdots \\ \mathrm{d}P_{t-q}^N \end{pmatrix}$$

So $\mathbb{K}_i$ is a symmetric blocs matrix where the bloc j of the diagonal is $\mathrm{bloc}_{jj} = \begin{pmatrix} K_{jj}^i(1,1) & \dots & K_{jj}^i(1,q) \\ \vdots & \ddots & \vdots \\ K_{jj}^i(q,1) & \dots & K_{jj}^i(q,q) \end{pmatrix}$ and the bloc (k,j) when $k < j$ is $\mathrm{bloc}_{kj} = \frac{1}{2}\begin{pmatrix} K_{kj}^i(1,1) & \dots & K_{kj}^i(1,q) \\ \vdots & \ddots & \vdots \\ K_{kj}^i(q,1) & \dots & K_{kj}^i(q,q) \end{pmatrix}$ (when $j < k$ the cross kernels is K_{jk}^i).

Thus, in this case, where $\mathbb{L} \equiv 0$, sufficient condition to keep the intensity positive is to have $\mathbb{K}_i$ positive semi-definite.

If we now consider the case where $\mathbb{L} \neq 0$, following the same method, we can write:

$$\lambda_{i,t} = \lambda_{i,\infty} + \mathbf{L}_i \boldsymbol{r_t} + \boldsymbol{r_t}^\top \mathbb{K}_i \boldsymbol{r_t}$$

where $\boldsymbol{r_t}$ and $\mathbb{K}_i$ are the same as before and $\mathbf{L}_i$ is

$$\mathbf{L}_i = \begin{pmatrix} L_1^i(1) & \dots & L_1^i(q) & \dots & L_N^i(1) & \dots & L_N^i(q) \end{pmatrix}^\top$$

Assuming $\mathbb{K}_i$ is invertible, one can complete the square by writing:

$$\begin{aligned} \lambda_{i,t} &= \lambda_{i,\infty} + \mathbf{L}_i \boldsymbol{r_t} + \boldsymbol{r_t}^\top \mathbb{K}_i \boldsymbol{r_t} \\ &= \lambda_{i,\infty} + \left(\boldsymbol{r_t} + \frac{1}{2} \mathbb{K}_i^{-1} \mathbf{L}_i \right)^\top \mathbb{K}_i \left(\boldsymbol{r_t} + \frac{1}{2} \mathbb{K}_i^{-1} \mathbf{L}_i \right) - \frac{1}{4} \mathbf{L}_i^\top \mathbb{K}_i^{-1} \mathbf{L}_i \end{aligned}$$

In conclusion, a sufficient condition for intensities to stay positive when $\mathbb{L} \neq 0$ is to have all kernels $\mathbb{K}_i$ positive definite and

$$\lambda_{i,\infty} \geq \frac{1}{4} \mathbf{L}_i^\top \mathbb{K}_i^{-1} \mathbf{L}_i, \qquad \forall i$$

A.2 Stationary Condition in Bivariate Case

In the bivariate case, we no longer have $\mathbb{E}(\mathrm{d}P_u^i \mathrm{d}P_u^j) = 0$ when $i \neq j$. Thus, when computing $\mathbb{E}(\boldsymbol{\lambda_t})$ to find the mean intensity and thus the stationary condition, we need to compute the mean of cross feedback.

We first introduce $m(dt, \mathrm{d}\xi^1, \mathrm{d}\xi^2)$ the joint Punctual Poisson Measure associate to the jump processes of the prices (P^1, P^2). So it is a pure jump process with i.i.d jump sizes $(\xi^1\xi^2)$ of common law p on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. We assume $\int_{\mathbb{R}} \xi^1\xi^2 p(\mathrm{d}\xi^1, \mathrm{d}\xi^2) = \psi_1\psi_2$.

$$
\begin{aligned}
\frac{1}{\psi_1\psi_2}\mathbb{E}\Big(\int_{-\infty}^t \int_{-\infty}^t K_{12}(t-s, t-u)\mathrm{d}P_s^1 \mathrm{d}P_u^2\Big) &= \frac{1}{\psi_1\psi_2}\mathbb{E}\Big(\int_{-\infty}^t K_{12}(t-u, t-u)\mathrm{d}P_u^1 \mathrm{d}P_u^2\Big) \\
&= \frac{1}{\psi_1\psi_2}\mathbb{E}\Big(\int_{\mathbb{R}}\int_{-\infty}^t K_{12}(t-u, t-u) s_u^1 s_u^2 \xi_u^1 \xi_u^2 m(\mathrm{d}u, \mathrm{d}\xi^1, \mathrm{d}\xi^2)\Big) \\
&= \frac{1}{\psi_1\psi_2}\mathbb{E}\Big(\int_{\mathbb{R}}\int_{-\infty}^t K_{12}(t-u, t-u) s_u^1 s_u^2 \xi_u^1 \xi_u^2 \mu_u^{\mathrm{c}} \mathrm{d}u p(\mathrm{d}\xi^1, \mathrm{d}\xi^2)\Big) \\
&= \frac{1}{\psi_1\psi_2}\mathbb{E}\Big(\int_{-\infty}^t K_{12}(t-u, t-u) s_u^1 s_u^2 \mu_u^{\mathrm{c}} \mathrm{d}u\Big)\mathbb{E}\Big(\int_{\mathbb{R}} \xi_u^1 \xi_u^2 p(\mathrm{d}\xi^1, \mathrm{d}\xi^2)\Big) \\
&= \int_{-\infty}^t K_{12}(t-u, t-u)\mathbb{E}(s_u^1 s_u^2)\mathbb{E}(\mu_u^{\mathrm{c}})\mathrm{d}u \\
&= \rho\overline{\mu^{\mathrm{c}}}\int_{-\infty}^t K_{12}(t-u, t-u)\mathrm{d}u
\end{aligned}
$$

A.3 Yule-Walker Equations: Complement

We write the Yule-Walker equations for the matrix and one for $\mathbb{D}_\times$, the calculations can be made as in Appendix 1 of [70]:

$$\begin{aligned}\mathbb{D}_\times(\tau_1,\tau_2) = &\int_{\tau_1^+}^{+\infty} \mathbb{K}_{\mathrm{d}}(u)\mathbb{D}_\times(\tau_1-u,\tau_2-u)\mathrm{d}u \\ &+ 2\int_{\tau_1^+}^{+\infty} \mathbb{K}_{\mathrm{d}}(u,\tau_1)\begin{pmatrix}\mathcal{D}_{112}(u-\tau_1,\tau_2-\tau_1) & 0\\ 0 & \mathcal{D}_{221}(u-\tau_1,\tau_2-\tau_1)\end{pmatrix}\mathrm{d}u \\ &+ \int_{\tau_1^+}^{+\infty} \mathbf{K}_\times(u,\tau_1)\begin{pmatrix}0\\ \mathcal{D}_{211}(u-\tau_1,\tau_2-\tau_1)\end{pmatrix}^\top \mathrm{d}u + \int_{\tau_1^+}^{+\infty} \mathbf{K}_\times(\tau_1,u)\begin{pmatrix}\mathcal{D}_{122}(u-\tau_1,\tau_2-\tau_1)\\ 0\end{pmatrix}^\top \mathrm{d}u\end{aligned} \tag{A.3.1}$$

In the bivariate case, we need to consider $\mathrm{d}N^c$. Thus, $\mathbb{C}$ becomes a 3×3 matrix, and the three-point correlation now also considers the components $\mathbb{D}_{cij}$. With this in mind, we can calculate the Yule-Walker equation for $\mathbb{C}$ in the bivariate case:

$$\begin{aligned}\mathbb{C}(\tau) =& \bar{\mathbb{\lambda}}\mathbb{Q}_{\mathrm{d}}(\tau) + \int_0^{+\infty} \mathbb{Q}_{\mathrm{d}}(u,u)\mathbb{C}(\tau-u)\mathrm{d}u + 2\int_0^{+\infty}\int_{u^+}^{+\infty} \mathbb{Q}_{\mathrm{d}}(\tau+u,\tau+v)\mathbb{D}_{\mathrm{d}}(u,v)\mathrm{d}v\mathrm{d}u \\ &+ \int_0^{+\infty}\int_{u^+}^{+\infty} \mathbf{Q}_\times(\tau+u,\tau+v)\begin{pmatrix}\mathcal{D}_{112}\\ \mathcal{D}_{212}\\ \mathcal{D}_{c12}\end{pmatrix}^\top (u,v)\mathrm{d}v\mathrm{d}u + \int_0^{+\infty}\int_{u^+}^{+\infty} \mathbf{Q}_\times(\tau+u,\tau+v)\begin{pmatrix}\mathcal{D}_{112}\\ \mathcal{D}_{212}\\ \mathcal{D}_{c12}\end{pmatrix}^\top (u,v)\mathrm{d}u\mathrm{d}v \\ &+ \rho\int_0^{+\infty}\begin{pmatrix}Q_{12}^1\\ Q_{12}^2\\ Q_{12}^c\end{pmatrix}(u)\begin{pmatrix}\mathcal{C}_{c1}\\ \mathcal{C}_{c2}\\ \mathcal{C}_{c3}\end{pmatrix}^\top (\tau-u)\mathrm{d}u + \int_0^{+\infty}\begin{pmatrix}Q_{11}^1\\ Q_{22}^2\\ Q_{cc}^c\end{pmatrix}(u)\begin{pmatrix}\mathcal{C}_{c1}\\ \mathcal{C}_{c2}\\ \mathcal{C}_{c3}\end{pmatrix}^T (\tau-u)\mathrm{d}u\end{aligned}$$

where $\bar{\mathbb{\lambda}}$ is a 2×2 matrix defined as

$$\bar{\mathbb{\lambda}} := \begin{pmatrix}\bar{\lambda}_1 & \bar{\mu}_{\mathrm{c}}\\ \bar{\mu}_{\mathrm{c}} & \bar{\lambda}_2\end{pmatrix}.$$

A.4 Asymptotic Behavior of Decaying Power Law Kernels

A.4.1 Asymptotic Forms

We consider decaying power law kernels, such that:

$$\mathbb{K}_{\mathrm{d}}(\tau) \underset{\tau\to+\infty}{\sim} \begin{pmatrix} k_{d1}\tau^{-1-\epsilon_{\mathrm{d}}} & k_{12}\tau^{-1-\epsilon_{\mathrm{o}}} \\ k_{21}\tau^{-1-\epsilon_{\mathrm{o}}} & k_{d2}\tau^{-1-\epsilon_{\mathrm{d}}} \end{pmatrix} \quad \mathbb{K}_{\mathrm{d}}(\tau v_1, \tau v_2) \underset{\tau\to+\infty}{\sim} \begin{pmatrix} \tilde{K}_{11}(v_1,v_2)\tau^{-2\rho_{\mathrm{d}}} & \tilde{K}_{12}(v_1,v_2)\tau^{-2\rho_{\mathrm{o}}} \\ \tilde{K}_{21}(v_1,v_2)\tau^{-2\rho_{\mathrm{o}}} & \tilde{K}_{22}(v_1,v_2)\tau^{-2\rho_{\mathrm{d}}} \end{pmatrix}$$

$$\mathbb{K}_{\times}(\tau) \underset{\tau\to+\infty}{\sim} \begin{pmatrix} k_{x1}\tau^{-1-\epsilon_{\times}} \\ k_{x2}\tau^{-1-\epsilon_{\times}} \end{pmatrix} \quad \mathbb{K}_{\times}(\tau v_1, \tau v_2) \underset{\tau\to+\infty}{\sim} \begin{pmatrix} \tilde{K}^1_{12}(v_1,v_2)\tau^{-2\rho_{\times}} \\ \tilde{K}^2_{12}(v_1,v_2)\tau^{-2\rho_{\times}} \end{pmatrix}$$

Where $\tilde{K}_{11}$, $\tilde{K}_{12}$, $\tilde{K}_{21}$, $\tilde{K}_{22}$, $\tilde{K}^1_{12}$ and $\tilde{K}^2_{12}$ are bounded. Given the Yule-Walker equations of Section 3.2.2, we expect the correlation structure to have a similar form. So we look for them as decaying power law functions with parameters defined such as:

$$\mathbb{C}(\tau) \underset{\tau\to+\infty}{\sim} \begin{pmatrix} c_{d1}\tau^{-\beta_{\mathrm{d}}} & c_{12}\tau^{-\beta_{\mathrm{o}}} \\ c_{21}\tau^{-\beta_{\mathrm{o}}} & c_{d2}\tau^{-\beta_{\mathrm{d}}} \end{pmatrix} \quad \mathbb{D}_{\mathrm{d}}(\tau v_1, \tau v_2) \underset{\tau\to+\infty}{\sim} \begin{pmatrix} \tilde{\mathcal{D}}_{111}(v_1,v_2)\tau^{-2\delta_{\mathrm{d}}} & \tilde{\mathcal{D}}_{122}(v_1,v_2)\tau^{-2\delta_{\mathrm{o}}} \\ \tilde{\mathcal{D}}_{211}(v_1,v_2)\tau^{-2\delta_{\mathrm{o}}} & \tilde{\mathcal{D}}_{222}(v_1,v_2)\tau^{-2\delta_{\mathrm{d}}} \end{pmatrix}$$

$$\mathbb{D}_{\times}(\tau) \underset{\tau\to+\infty}{\sim} \begin{pmatrix} d_{112}\tau^{-\beta^{\times}_{\mathrm{d}}} & d_{121}\tau^{-\beta^{\times}_{\mathrm{o}}} \\ d_{212}\tau^{-\beta^{\times}_{\mathrm{o}}} & d_{221}\tau^{-\beta^{\times}_{\mathrm{d}}} \end{pmatrix} \quad \mathbb{D}_{\times}(\tau v_1, \tau v_2) \underset{\tau\to+\infty}{\sim} \begin{pmatrix} \tilde{\mathcal{D}}_{112}(v_1,v_2)\tau^{-2\delta_{\times}} & \tilde{\mathcal{D}}_{121}(v_1,v_2)\tau^{-2\delta_{\times}} \\ \tilde{\mathcal{D}}_{212}(v_1,v_2)\tau^{-2\delta_{\times}} & \tilde{\mathcal{D}}_{221}(v_1,v_2)\tau^{-2\delta_{\times}} \end{pmatrix}$$

As in [70], we make the following hypothesis on the exponents: $\rho_{\mathrm{d}} > \frac{1}{2}$, $\rho_{\mathrm{o}} > \frac{1}{2}$, $\rho_{\times} > \frac{1}{2}$, so the first and second moments are finite. Moreover, we focus on the cases:

- Non critical case ($n_H < 1$), where we assume $0 < \epsilon_{\mathrm{d}} < 1$, $0 < \epsilon_{\mathrm{o}} < 1$,
- Critical case ($n_H = 1$), where we assume $0 < \epsilon_{\mathrm{d}} < \frac{1}{2}$, $0 < \epsilon_o < \frac{1}{2}$,

In the non-critical case ($n < 1$), the method consists in replacing $\mathbb{K}_{\mathrm{d}}$, $\mathbb{K}_{\times}$, $\mathbb{C}$, $\mathbb{D}_{\mathrm{d}}$ and $\mathbb{D}_{\times}$ by their asymptotic expressions presented above in the Yule-Walker equations (Equations. (3.2.5), (3.2.8) and (A.3.1)) and study the limit $\tau \to +\infty$.

The rest of this section is organised as follow. Section A.4.2 serves as a pedagogical tool and introduce, for a univariate linear Hawkes, the methods to derive the relationship between the exponents of the correlation structures and the exponents of the kernels in both the non critical and the critical case. Section A.4.3 describes the extension of the methods for a quadratic Hawkes in the critical case (non critical case being more natural to derive from the linear one). Finally, Appendix A.5 gives the resulting relationships for both the critical and non critical case for a quadratic Hawkes.

A.4.2 Asymptotic behavior of auto-covariance structures - Methods for critical and non critical linear Hawkes

As a pedagogical tool, we derive here the method to study asymptotic behavior of the auto-covariance structure in the simple case of univariate linear Hawkes. Those methods are then derived to obtain the results for the multivariate quadratic case, one of which is detailed in Appendix A.4.3 for the univariate quadratic case.

Let's consider a univariate Hawkes process with kernel ϕ, baseline intensity λ_∞ and define the covariance $\mathcal{C}$ as in Equation (3.2.1). In the univariate lienar case, we assume the following asymptotic forms:

$$\phi(\tau) \underset{\tau\to+\infty}{\sim} c_0\tau^{-1-\epsilon} \quad \text{and} \quad \mathcal{C}(\tau) \underset{\tau\to+\infty}{\sim} c_1\tau^{-\beta}$$

where

- in the non critical case ($n_H < 1$), $0 < \epsilon < 1$
- in the critical case ($n_H = 1$), $0 < \epsilon < \frac{1}{2}$

Non Critical case - Time domain

We recall the Yule-Walker equation for linear Hawkes,

$$\begin{aligned} \mathcal{C}(\tau) &= \bar{\lambda}\phi(\tau) + \int_0^{+\infty} \phi(u)\mathcal{C}(\tau-u)\mathrm{d}u \\ \mathcal{C}(\tau) &= \mathcal{C}_1(\tau) + \mathcal{C}_2(\tau) \\ \text{with:}\quad \mathcal{C}_1(\tau) &= \bar{\lambda}\phi(\tau) \\ \mathcal{C}_2(\tau) &= \int_0^{+\infty} \phi(u)\mathcal{C}(\tau-u)\mathrm{d}u \end{aligned}$$

So, we have:

$$\mathcal{C}_1(\tau) = \bar{\lambda}\phi(\tau) = \bar{\lambda}c_0\tau^{-1-\epsilon} + o(\tau^{-1-\epsilon})$$

Then, the second term, we assume the integral is dominated by u with big values ($u \sim \tau$), thus:

$$\begin{aligned} \mathcal{C}_2(\tau) = \int_0^{\infty} \phi(u)\mathcal{C}(\tau-u)\mathrm{d}u &= \int_0^{\infty} \phi(u)c_1(\tau-u)^{-\beta}\mathrm{d}u \\ &= \int_0^{\infty} \phi(u)c_1(1-\frac{u}{\tau})^{-\beta}\tau^{-\beta}\mathrm{d}u \\ &\underset{\tau\to\infty}{\to} \tau^{-\beta}c_1\int_0^{\infty}\phi(u)\mathrm{d}u \end{aligned}$$

We know $\int_0^{\infty} K(u,u)\mathrm{d}u$ is finite, so we did extract the asymptotic behavior.
In the end, we have:

$$\begin{aligned} c_1\tau^{-\beta} + o(\tau^{-\beta}) = \kappa\bar{\lambda}c_0\tau^{-1-\epsilon} &+ o(\tau^{-1-\epsilon}) \\ &+ c_1\tau^{-\beta}\int_0^{+\infty}\phi(u,u)\mathrm{d}u + o(\tau^{-\beta}), \end{aligned}$$

from which we directly read the result $\boxed{\beta = 1+\epsilon}$.

Critical case with Fourier relation from Hawkes-1971-*Point Spectra of Some Mutually Exciting Point Processes*

In critical case, when $n_H = 1$, the relationship between auto-correlation structures exponents and kernel exponents can not be determined by looking at the limit $\tau \to +\infty$. Thus, to overcome this difficulty, we use a second method to investigate asymptotic behavior using Fourier-domain.

We define the Fourier transform of a function f such as:

$$\hat{f}(\omega) = \int_{\mathbb{R}} f(t)e^{-i\omega t}\mathrm{d}t$$

Step 1: Find the regularity of the Yule-Walker terms

Let's recall the Yule-Walker equation for the covariance $\mathcal{C}$ defined in Equation (3.2.1) for a linear Hawkes process with kernel ϕ:

$$\mathcal{C}(\tau) = \phi(\tau)\bar{\lambda} + \int_{-\infty}^{+\infty}\phi(\tau-u)\mathcal{C}(u)\mathrm{d}u. \tag{A.4.1}$$

As in [15], $\mathcal{C}$'s extension in 0, in Fourier domain, gives:

$$\hat{\mathcal{C}}^*(\omega) = \bar{\lambda} + \hat{\mathcal{C}}(\omega). \tag{A.4.2}$$

We introduce $\hat{B}$ such that:

$$\hat{\mathcal{C}}(\omega) = (1-\hat{\phi}(\omega))^{-1}(\hat{\phi}(\omega)\bar{\lambda} - \hat{B}(\omega)),$$

thus,

$$\hat{\mathcal{C}}(\omega) = (1 - \hat{\phi}(\omega))^{-1}(\hat{\phi}(\omega)\bar{\lambda} - \hat{B}(\omega)). \tag{A.4.3}$$

By definition, $B(\tau) = 0$ for $\tau > 0$ so that the Fourier transforms $\hat{B}(\omega)$ is regular[1] in the upper half plane $\mathrm{Im}(\omega) > 0$. Similarly from the condition $\phi(u) = 0$ for $u < 0$ it follows that $\hat{\phi}(\omega)$ is regular in the lower half-plane $\mathrm{Im}(\omega) < 0$.

Using $\hat{\mathcal{C}}^\top(\omega) = \hat{\mathcal{C}}(-\omega)$, we obtain:

$$(\hat{\phi}(\omega)\bar{\lambda} - \hat{B}(\omega))^\top(1 - \hat{\phi}^\top(\omega))^{-1} = (1 - \hat{\phi}(-\omega))^{-1}(\hat{\phi}(-\omega)\bar{\lambda} - \hat{B}(-\omega)),$$

which gives:

$$(1 - \hat{K}(-\omega))\hat{B}^\top(\omega) + \hat{K}(-\omega)D = \hat{B}(-\omega)(1 - \hat{K}^\top(\omega)) + \hat{K}^\top(\omega)D.$$

On the left side we have $(1-\hat{K}(-\omega))\hat{B}^\top(\omega)+\hat{K}(-\omega)D$, which is regular on $\mathrm{Im}(\omega) > 0$, whereas on the right side we have $\hat{B}(-\omega)(1 - \hat{K}^\top(\omega)) + \hat{K}^\top(\omega)D$ which is regular on $Im(\omega) < 0$. So, as in [75], we can say that both side are equal to zero and find $\hat{B}(\omega) = -\hat{K}^\top(-\omega)D(1 - \hat{K}(-\omega))^{-1}$. Plugging the later in Equation (A.4.3), one eventually gets:

$$\begin{aligned}
\hat{\mathcal{C}}(\omega) =&(1 - \hat{\phi}(\omega))^{-1}(\hat{\phi}(\omega)\bar{\lambda} - \hat{B}(\omega)) \\
=&(1 - \hat{\phi}(\omega))^{-1}(\hat{\phi}(\omega)\bar{\lambda} + \hat{\phi}^\top(-\omega)\bar{\lambda}(1 - \hat{\phi}(-\omega))^{-1}) \\
=&(1 - \hat{\phi}(\omega))^{-1}\hat{\phi}(\omega)\bar{\lambda} + (1 - \hat{\phi}(\omega))^{-1}\hat{\phi}^\top(-\omega)D(1 - \hat{\phi}(-\omega))^{-1} \\
=&(1 - \hat{\phi}(\omega))^{-1}\hat{\phi}(\omega)\bar{\lambda}(1 - \hat{\phi}(-\omega))(1 - \hat{\phi}(-\omega))^{-1} + (1 - \hat{\phi}(\omega))^{-1}\hat{\phi}^\top(-\omega)\bar{\lambda}(1 - \hat{\phi}(-\omega))^{-1} \\
=&(1 - \hat{\phi}(\omega))^{-1}\hat{\phi}(\omega)\bar{\lambda}(1 - \hat{\phi}(-\omega))(1 - \hat{\phi}(-\omega))^{-1} + (1 - \hat{\phi}(\omega))^{-1}\hat{\phi}^\top(-\omega)\bar{\lambda}(1 - \hat{\phi}(-\omega))^{-1} \\
=&(1 - \hat{\phi}(\omega))^{-1}\hat{\phi}(\omega)D(1 - \hat{\phi}(-\omega))^{-1} - (1 - \hat{\phi}(\omega))^{-1}\hat{\phi}(\omega)D\hat{\phi}(-\omega)(1 - \hat{\phi}(-\omega))^{-1} \\
&+ (1 - \hat{\phi}(\omega))^{-1}\hat{\phi}^\top(-\omega)D(1 - \hat{\phi}(-\omega))^{-1} \\
=&(1 - \hat{\phi}(\omega))^{-1}(\hat{\phi}(\omega)\bar{\lambda} + \bar{\lambda}\hat{\phi}^\top(-\omega) - \hat{\phi}(\omega)\bar{\lambda}\hat{\phi}^\top(-\omega))(1 - \hat{\phi}(-\omega))^{-1}.
\end{aligned}$$

Equation (A.4.6) allows to obtain:

$$\boxed{\hat{\mathcal{C}^*}(\omega) = (1 - \hat{\phi}(\omega))^{-1}\bar{\lambda}(1 - \hat{\phi}(-\omega))^{-1}}. \tag{A.4.4}$$

We then develop the expressions of $\hat{\phi}$ and $\hat{\mathcal{C}^*}$.

Step 2: Fourier transform of the kernel $\hat{\phi}$

$$\begin{aligned}
\hat{\phi}(\omega) =& \int_{-\infty}^{+\infty} e^{-i\omega\tau}\phi(\tau)\mathrm{d}\tau \\
=& \int_{-\infty}^{+\infty} \phi(\tau)\mathrm{d}\tau - \int_0^{+\infty}(1 - e^{-i\omega\tau})\phi(\tau)\mathrm{d}\tau \\
=&n_H - \int_0^{+\infty}(1 - e^{-i\omega\tau})\phi(\tau)\mathrm{d}\tau \\
&\text{assuming the integral is dominated by } \tau \text{ large,} \\
=&n_H - \int_0^{+\infty}(1 - e^{-i\omega\tau})c_0\tau^{-1-\epsilon}\mathrm{d}\tau \\
&\text{with the change of variable} x = \omega\tau \\
=&n_H - c_0\omega^\epsilon\int_0^{+\infty}(1 - e^{-ix})x^{-1-\epsilon}\mathrm{d}x \\
&\text{with an integration by parts:}
\end{aligned}$$

[1] We say a function f is regular in a space if, when considering $\omega = \omega_R + i\omega_{im}$, $\hat{f} = \int_{\mathbb{R}} f(t)e^{\omega_{im}t}e^{-i\omega_R t}\mathrm{d}t$ is well defined. For instance, for $\omega_{im} > 0$, one needs to pay attention to the limit $\lim_{t\to+\infty} e^{\omega_{im}t} = +\infty$.

$$
\begin{aligned}
&= n_H - c_0\omega^\epsilon\big([(1-e^{-ix})\frac{x^{-\epsilon}}{-\epsilon}]_0^{+\infty} + \frac{i}{\epsilon}\int_0^{+\infty} e^{-ix}x^{-\epsilon}\mathrm{d}x\big)\\
&= n_H - \frac{c_0\omega^\epsilon i}{\epsilon}\int_0^{+\infty} e^{-ix}x^{-\epsilon}\mathrm{d}x
\end{aligned}
$$

Then, we study $\int_0^{+\infty} e^{-ix}x^{-\epsilon}\mathrm{d}x$. Since $\epsilon < 1$, the integral is well defined in 0, and in the limit $+\infty$, we have $\int_1^{+\infty} e^{-ix}x^{-\epsilon}\mathrm{d}x = [\frac{e^{-ix}}{-i}x^{-\epsilon}]_1^{+\infty} - \frac{1}{i}\int_1^{+\infty} e^{-ix}x^{-\epsilon-1}\mathrm{d}x$ which is finite. Introducing, $k^- = \frac{c_0 i}{\epsilon}\int_0^{+\infty} e^{-ix}x^{-\epsilon}\mathrm{d}x$, one then as $\boxed{\hat{K}(\omega) = n_H - \omega^\epsilon k^-}$.

Step 3: Fourier transform of the covariance $\hat{\mathcal{C}}(\tau)$
Similarly, we have

$$
\begin{aligned}
\hat{\mathcal{C}}^*(\omega) &= \int_{-\infty}^{+\infty} e^{-i\omega\tau}\mathcal{C}(\tau)\mathrm{d}\tau\\
&= \int_{-\infty}^{0} e^{-i\omega\tau}\mathcal{C}(\tau)\mathrm{d}\tau + \int_0^{+\infty} e^{-i\omega\tau}\mathcal{C}(\tau)\mathrm{d}\tau\\
&\quad \mathcal{C} \text{ being even,}\\
&= 2\int_0^{+\infty} \cos(\omega\tau)\mathcal{C}(\tau)\mathrm{d}\tau\\
&\quad \text{assuming the integral is dominated by } \tau \text{ large,}\\
&= 2\int_0^{+\infty} \cos(\omega\tau)c_1\tau^{-\beta}\mathrm{d}\tau\\
&\quad \text{the change of variable } x = \omega\tau \text{ gives}\\
&= 2c_1\omega^{\beta-1}\int_0^{+\infty} \cos(x)x^{-\beta}\mathrm{d}x
\end{aligned}
$$

Introducing $A = \int_0^{+\infty} \cos(x)x^{-\beta}\mathrm{d}x$, one obtains $\boxed{\hat{\mathcal{C}}^*(\omega) = 2c_1 A\omega^{\beta-1}}$.

Conclusion
We can finally conclude, with Equation (A.4.4), when $n_H \to 1$, $\boxed{\beta = 1 - 2\epsilon}$. (Note that this result with the hypothesis $0 < \epsilon < \frac{1}{2}$ also garantee the well definition of $\int_0^{+\infty} \cos(x)x^{-\beta}\mathrm{d}x$).

A.4.3 Method for the Asymptotic Study in the Critical Case of QHawkes process

We present now the method for the critical quadratic Hawkes process. For the sake of simplicity we limit ourselves to the 1D case here, the multivariate case can be worked out similarly.

The definitions of the autocorrelation structures in the 1D case can be found in [70] and are substantially similar to those used here.

As above, we define the Fourier transform of a function f such as:

$$\hat{f}(\omega) = \int_{\mathbb{R}} f(t)e^{-i\omega t}\mathrm{d}t$$

Step 1: Find the regularity of the Yule-Walker terms
As for the linear case, we start from the Yule-Walker equation on $\mathcal{C}$ for $\tau \neq 0$ (see equation (9) in [70]):

$$\mathcal{C}(\tau) = K(\tau)\bar{\lambda} + \int_0^{+\infty} K(\tau-u)\mathcal{C}(u)\mathrm{d}u + 2\int_{0+}^{+\infty}\int_{u+}^{+\infty} K(\tau+u,\tau+r)\mathcal{D}(u,r)\mathrm{d}r\mathrm{d}u, \quad \text{(A.4.5)}$$

and its extension in 0 in Fourier domain gives:

$$\hat{\mathcal{C}}^*(\omega) = \bar{\lambda} + \hat{\mathcal{C}}(\omega). \quad \text{(A.4.6)}$$

To use the Fourier transform, we need to extend K and $\mathcal{D}$ for $\tau, \tau_1, \tau_2 < 0$. Thus, we consider the function K defined on $\mathbb{R}$ with $K(\tau) = 0$ for $\tau < 0$ similarly for $\mathcal{D}$. Thus, we expect to have $\hat{K}(\omega)$ and $\hat{\mathcal{D}}(\omega)$ regular in the half plan $\mathrm{Im}(\omega) < 0$.

The 2 first terms of Eq. (A.4.5) are the same as in the Yule-Walker equation for the linear case. Hence, the regularity arguments are the same as in [75].

However, we need to study the last term of Eq. (A.4.5), $C_3(\tau) = 2\int_{0^+}^{+\infty}\int_{u^+}^{+\infty} K(\tau+u,\tau+r)\mathcal{D}(u,r)\mathrm{d}r\mathrm{d}u$.

In order to study the regularity we decompose ω into $\omega = \omega_R + i\omega_I$. Then, switching integration order and using Chasles relation we obtain:

$$\begin{aligned}
\hat{C}_3(\omega) =& 2\int_0^{+\infty}\int_{u^+}^{+\infty}\int_{-\infty}^{-u} e^{-i\omega_R\tau}e^{\omega_I\tau}K(\tau+u,\tau+r)\mathcal{D}(u,r)d\tau\mathrm{d}r\mathrm{d}u \quad \Big\} \ \hat{C}_{3_1}(\omega)\\
&+ 2\int_0^{+\infty}\int_{u^+}^{+\infty}\int_{-u}^{0} e^{-i\omega_R\tau}e^{\omega_I\tau}K(\tau+u,\tau+r)\mathcal{D}(u,r)d\tau\mathrm{d}r\mathrm{d}u \quad \Big\} \ \hat{C}_{3_2}(\omega)\\
&+ 2\int_0^{+\infty}\int_{u^+}^{+\infty}\int_{0}^{+\infty} e^{-i\omega_R\tau}e^{\omega_I\tau}K(\tau+u,\tau+r)\mathcal{D}(u,r)d\tau\mathrm{d}r\mathrm{d}u \quad \Big\} \ \hat{C}_{3_3}(\omega)
\end{aligned}$$

We study each term $\hat{C}_{3_1}$, $\hat{C}_{3_2}$ and $\hat{C}_{3_3}$ separately. We first have:

$$\hat{C}_{3_1}(\omega) = 2\int_0^{+\infty}\int_{u^+}^{+\infty}\int_{-\infty}^{-u} e^{-i\omega_R\tau}e^{\omega_I\tau}K(\tau+u,\tau+r)\mathcal{D}(u,r)\mathrm{d}\tau\mathrm{d}r\mathrm{d}u$$

Here $\tau < -u$ so $\tau + u < 0$, so $K(\tau+u,\tau+r) = 0$, so $\hat{C}_{3_1}(\omega) = 0$.

For the second term, we switch integration order, and we obtain:

$$\hat{C}_{3_2}(\omega) = 2\int_{-\infty}^{0}\int_{-\tau}^{+\infty}\int_{u^+}^{+\infty} e^{-i\omega_R\tau}e^{\omega_I\tau}K(\tau+u,\tau+r)\mathcal{D}(u,r)\mathrm{d}r\mathrm{d}u\mathrm{d}\tau$$

In this case, we have $\tau \geq -u$, so $\tau + u \geq 0$, and $r \geq u^+ \geq -\tau$, so $K(\tau+u,\tau+r) > 0$. We are only interested in $\tau \to -\infty$, so $\hat{C}_{3_2}(\omega)$ is regular in the half plan $\omega_I > 0$.

For $\hat{C_{3_3}}(\omega)$, we have $\tau + u > 0$ and $\tau + r > 0$, so $K(\tau+u,\tau+r) \geq 0$. Since we are only worried about $\tau \to +\infty$. $\hat{C_{3_3}}(\omega)$ is regular in half plan $\omega_I < 0$.

With those results, we now introduce the Yule-Walker equation A.4.5 in Fourier domain and, as in [75], the function $\hat{B}$:

$$\begin{cases} \hat{\mathcal{C}}(\omega) &= \hat{K}(\omega)D + \hat{K}(\omega)\hat{\mathcal{C}}(\omega) + \hat{C_{3_2}}(\omega) + \hat{C_{3_3}}(\omega) \\ \hat{B}(\omega) &= \hat{K}(\omega)D + \hat{K}(\omega)\hat{\mathcal{C}}(\omega) + \hat{C_{3_2}}(\omega) + \hat{C_{3_3}}(\omega) - \hat{\mathcal{C}}^*(\omega) \end{cases} \tag{A.4.7}$$

Thus,

$$\hat{\mathcal{C}}(\omega) = (1-\hat{K}(\omega))^{-1}(\hat{K}(\omega)D - \hat{B}(\omega) + \hat{C_{3_2}}(\omega) + \hat{C_{3_3}}(\omega)) \tag{A.4.8}$$

Since $\mathcal{C}$ is even, we have $(\hat{\mathcal{C}}^\top(\omega) = \hat{\mathcal{C}}(-\omega))$ and Eq. (A.4.8) becomes:

$$(\hat{K}(\omega)D + \hat{C_{3_2}}(\omega) + \hat{C_{3_3}}(\omega) - \hat{B}(\omega))^\top(1-\hat{K}^\top(\omega))^{-1} = (1-\hat{K}(-\omega))^{-1}(\hat{K}(-\omega)D + \hat{C_{3_2}}(-\omega) + \hat{C_{3_3}}(-\omega) - \hat{B}(-\omega))$$

Multiplying by $1-\hat{K}^\top(\omega)$ on the right side and by $1-\hat{K}(-\omega)$ on the left side, we then develop the expression and mark the regularity propriety below each term.

$$\underbrace{D\hat{K}^\top(\omega)}_{\mathrm{Im}(\omega)<0}+\underbrace{\hat{C_{3_2}}^\top(\omega)}_{\mathrm{Im}(\omega)>0}+\underbrace{\hat{C_{3_3}}^\top(\omega)}_{\mathrm{Im}(\omega)<0}-\underbrace{\hat{B}^\top(\omega)}_{\mathrm{Im}(\omega)>0}-\underbrace{\hat{K}(-\omega)\hat{C_{3_2}}^\top(\omega)}_{\mathrm{Im}(\omega)>0}-\hat{K}(-\omega)\hat{C_{3_3}}^\top(\omega)+\underbrace{\hat{K}(-\omega)\hat{B}^\top(\omega)}_{\mathrm{Im}(\omega)>0}=$$
$$\underbrace{\hat{K}(-\omega)D}_{\mathrm{Im}(\omega)>0}+\underbrace{\hat{C_{3_2}}(-\omega)}_{\mathrm{Im}(\omega)<0}+\underbrace{\hat{C_{3_3}}(-\omega)}_{\mathrm{Im}(\omega)>0}-\underbrace{\hat{B}(-\omega)}_{\mathrm{Im}(\omega)<0}-\underbrace{\hat{C_{3_2}}(-\omega)\hat{K}^\top(\omega)}_{\mathrm{Im}(\omega)<0}-\hat{C_{3_3}}(-\omega)\hat{K}^\top(\omega)+\underbrace{\hat{B}(-\omega)\hat{K}^\top(\omega)}_{\mathrm{Im}(\omega)<0}$$

We notice we have a problem with $\hat{C_{3_3}}(-\omega)\hat{K}^\top(\omega)$ and $\hat{K}(-\omega)\hat{C_{3_3}}^\top(\omega)$. In fact, for each of them one factor is regular for $\mathrm{Im}(\omega)>0$ and the other one for $\mathrm{Im}(\omega)<0$... If we still reorder the terms to have the left side of the equality regular in the half plan $\mathrm{Im}(\omega)<0$ and the right side of the equality regular in the half plan $\mathrm{Im}(\omega)>0$, without defining the regularity plan for the two problematic terms, we obtain the following:

$$\begin{aligned}&\underbrace{D\hat{K}^\top(\omega)+\hat{C_{3_3}}^\top(\omega)-\hat{C_{3_2}}(-\omega)+\hat{B}(-\omega)+\hat{C_{3_2}}(-\omega)\hat{K}^\top(\omega)-\hat{B}(-\omega)\hat{K}^\top(\omega)}_{\mathrm{Im}(\omega)<0}-\hat{K}(-\omega)\hat{C_{3_3}}^\top(\omega)=\\&\underbrace{\hat{K}(-\omega)D+\hat{C_{3_3}}(-\omega)-\hat{C_{3_2}}^\top(\omega)+\hat{B}^\top(\omega)+\hat{K}(-\omega)\hat{C_{3_2}}^\top(\omega)-\hat{K}(-\omega)\hat{B}^\top(\omega)}_{\mathrm{Im}(\omega)>0}-\hat{C_{3_3}}(-\omega)\hat{K}^\top(\omega)\end{aligned}\tag{A.4.9}$$

We now need to study in details the regularity of the problematic terms.

Step 2: Expand $\hat{C_{3_3}}(-\omega)\hat{K}^\top(\omega)$ and $\hat{K}(-\omega)\hat{C_{3_3}}^\top(\omega)$

We have

$$\hat{K}(-\omega)\hat{C_{3_3}}^\top(\omega)=2\int_0^{+\infty}\int_{u+}^{+\infty}\int_0^{+\infty}\Big(\hat{K}(-\omega)e^{-i\omega\tau}\Big)\mathcal{D}^\top(u,r)K^\top(\tau+u,\tau+r)\mathrm{d}\tau\mathrm{d}r\mathrm{d}u$$

and

$$\hat{C_{3_3}}(-\omega)\hat{K}^\top(\omega)=2\int_0^{+\infty}\int_{u+}^{+\infty}\int_0^{+\infty}K(\tau+u,\tau+r)\mathcal{D}(u,r)\Big(e^{i\omega\tau}\hat{K}^\top(\omega)\Big)\mathrm{d}\tau\mathrm{d}r\mathrm{d}u$$

We look at $\hat{K}(-\omega)e^{-i\omega\tau}$ with the change of variables $\phi(t)=t+\tau$ and the Chasles relation:

$$\hat{K}(-\omega)e^{-i\omega\tau}=\int_0^{+\infty}\int_{-\infty}^{+\infty}e^{i\omega t}K(t+\tau,t+\tau)dtd\tau=\int_0^{+\infty}\int_{-\infty}^{-\tau}e^{i\omega t}K(t+\tau,t+\tau)\mathrm{d}t\mathrm{d}\tau\tag{A.4.10}$$
$$+\int_0^{+\infty}\int_{-\tau}^{0}e^{i\omega t}K(t+\tau,t+\tau)\mathrm{d}t\mathrm{d}\tau\tag{A.4.11}$$
$$+\int_0^{+\infty}\int_{0}^{+\infty}e^{i\omega t}K(t+\tau,t+\tau)\mathrm{d}t\mathrm{d}\tau\tag{A.4.12}$$

The first term, (A.4.10), $\int_0^{+\infty}\int_{-\infty}^{-\tau}e^{i\omega t}K(t+\tau,t+\tau)\mathrm{d}t\mathrm{d}\tau$ is null, because $K(t+\tau)$ is null for $t\leq-\tau$. If we switch the integration order in the second term, (A.4.11), we have $\int_0^{+\infty}\int_{-\tau}^{0}e^{i\omega t}K(t+\tau,t+\tau)\mathrm{d}t\mathrm{d}\tau=\int_{-\infty}^{0}\int_{-t}^{+\infty}e^{i\omega t}K(t+\tau,t+\tau)\mathrm{d}\tau\mathrm{d}t$ and then, we need to have a convergent exponential $e^{i\omega t}$ when $t\to-\infty$, so the second term is regular in half plan $\mathrm{Im}(\omega)<0$. And the last term, (A.4.12), $\int_0^{+\infty}\int_0^{+\infty}e^{i\omega t}K(t+\tau,t+\tau)\mathrm{d}t\mathrm{d}\tau$ is regular in half plan $\mathrm{Im}(\omega)>0$.

We now need to consider the term $e^{i\omega\tau}\hat{K}^\top(\omega)$, similarly, we have:

$$e^{i\omega\tau}\hat{K}^\top(\omega) = \int_0^{+\infty}\int_{-\infty}^{+\infty} e^{-i\omega t}K^\top(t+\tau,t+\tau)\mathrm{d}t\mathrm{d}\tau = \int_0^{+\infty}\int_{-\infty}^{-\tau} e^{-i\omega t}K^\top(t+\tau,t+\tau)\mathrm{d}t\mathrm{d}\tau \tag{A.4.13}$$

$$+\int_0^{+\infty}\int_{-\tau}^{0} e^{-i\omega t}K^\top(t+\tau,t+\tau)\mathrm{d}t\mathrm{d}\tau \tag{A.4.14}$$

$$+\int_0^{+\infty}\int_{0}^{+\infty} e^{-i\omega t}K^\top(t+\tau,t+\tau)\mathrm{d}t\mathrm{d}\tau \tag{A.4.15}$$

For same reasons as before, first term, (A.4.13), is null, second term, (A.4.14), is regular in half plan $\mathrm{Im}(\omega) > 0$, and last term, (A.4.15), is regular in half plan $\mathrm{Im}(\omega) < 0$.

Wrapping up those results, we have

$$\begin{aligned}\hat{K}(-\omega)\hat{C}_{3_3}^\top(\omega) =& 2\int_0^{+\infty}\int_{u+}^{+\infty}\int_0^{+\infty}\int_{-\tau}^{0} e^{i\omega t}K(t+\tau,t+\tau)\mathrm{d}t\mathcal{D}^\top(u,r)K^\top(\tau+u,\tau+r)\mathrm{d}\tau\mathrm{d}r\mathrm{d}u \quad \Big\} \; F_1(\omega)\\ &+ 2\int_0^{+\infty}\int_{u+}^{+\infty}\int_0^{+\infty}\int_{0}^{+\infty} e^{i\omega t}K(t+\tau,t+\tau)\mathrm{d}t\mathcal{D}^\top(u,r)K^\top(\tau+u,\tau+r)\mathrm{d}\tau\mathrm{d}r\mathrm{d}u \quad \Big\} \; F_2(\omega)\\ =& F_1(\omega) + F_2(\omega)\end{aligned}$$

with $F_1(\omega)$ being regular for $\mathrm{Im}(\omega) < 0$ and $F_2(\omega)$ for $\mathrm{Im}(\omega) > 0$.

Similarly,

$$\begin{aligned}\hat{C_{3_3}}(-\omega)\hat{K}^\top(\omega) =& 2\int_0^{+\infty}\int_{u+}^{+\infty}\int_0^{+\infty} K(\tau+u,\tau+r)\mathcal{D}(u,r)\int_{-\tau}^{0} e^{-i\omega t}K^\top(t+\tau,t+\tau)\mathrm{d}t\mathrm{d}\tau\mathrm{d}r\mathrm{d}u \quad \Big\} E_1(-\omega)\\ &+ 2\int_0^{+\infty}\int_{u+}^{+\infty}\int_0^{+\infty} K(\tau+u,\tau+r)\mathcal{D}(u,r)\int_{0}^{+\infty} e^{-i\omega t}K^\top(t+\tau,t+\tau)\mathrm{d}t\mathrm{d}\tau\mathrm{d}r\mathrm{d}u \quad \Big\} E_2(-\omega)\\ =& E_1(-\omega) + E_2(-\omega)\end{aligned}$$

with $E_1(-\omega)$ being regular for $\mathrm{Im}(\omega) > 0$ and $E_2(-\omega)$ for $\mathrm{Im}(\omega) < 0$.

Step 3: Wrap up

We can now go back to Eq. (A.4.9) which becomes:

$$\begin{aligned}&D\hat{K}^\top(\omega) + \hat{C_{3_3}}^\top(\omega) - \hat{C_{3_2}}(-\omega) + \hat{B}(-\omega) + \hat{C_{3_2}}(-\omega)\hat{K}^\top(\omega) - \hat{B}(-\omega)\hat{K}^\top(\omega) - F_1(\omega) + E_2(-\omega) =\\ &\hat{K}(-\omega)D + \hat{C_{3_3}}(-\omega) - \hat{C_{3_2}}^\top(\omega) + \hat{B}^\top(\omega) + \hat{K}(-\omega)\hat{C_{3_2}}^\top(\omega) - \hat{K}(-\omega)\hat{B}^\top(\omega) - E_1(-\omega) + F_2(\omega)\end{aligned}$$

We then have the left side regular in half plan $\mathrm{Im}(\omega) < 0$ and the right one in half plan $\mathrm{Im}(\omega) > 0$. So as in [75], we can say that the lower side is null, and obtain an expression for $\hat{B}$:

$$\begin{aligned}&0 = \hat{K}(-\omega)D + \hat{C_{3_3}}(-\omega) - \hat{C_{3_2}}^\top(\omega) + \hat{B}^\top(\omega) + \hat{K}(-\omega)\hat{C_{3_2}}^\top(\omega) - \hat{K}(-\omega)\hat{B}^\top(\omega) - E_1(-\omega) + F_2(\omega)\\ &\Longrightarrow \begin{cases}\hat{B}^\top(\omega) &= -\Big(\hat{K}(-\omega)D + \hat{C_{3_3}}(-\omega) - \hat{C_{3_2}}^\top(\omega) + \hat{K}(-\omega)\hat{C_{3_2}}^\top(\omega) - E_1(-\omega) + F_2(\omega)\Big)^\top(1-\hat{K}^\top(-\omega))^{-1}\\ \hat{B}(\omega) &= -\Big(D\hat{K}^\top(-\omega) + \hat{C}_{3_3}^\top(-\omega) - \hat{C}_{3_2}(\omega) + \hat{C}_{3_2}(\omega)\hat{K}^\top(-\omega) - E_1^\top(-\omega) + F_2^\top(\omega)\Big)(1-\hat{K}^\top(-\omega))^{-1}\end{cases}\end{aligned}$$

Then we inject the expression of $\hat{B}$ in $\hat{\mathcal{C}}^*$, and after some manipulations where we use $\hat{C_{3_3}}(\omega)\hat{K}^\top(-\omega) = E_1(\omega) + E_2(\omega)$:

$$\begin{aligned}\hat{\mathcal{C}}(\omega) =& (1-\hat{K}(\omega))^{-1}\Big(\hat{K}(\omega)D + \hat{C_{3_3}}(\omega) - \hat{K}(\omega)D\hat{K}^\top(-\omega) - E_1(\omega) - E_2(\omega)\\ &+ D\hat{K}^\top(-\omega) + \hat{C}_{3_3}^\top(-\omega) - E_1^\top(-\omega) + F_2^\top(\omega)\Big)(1-\hat{K}^\top(-\omega))^{-1}.\end{aligned}$$

Note that in 1D, the transpose and commutative operations do not really matter. However, in order to use this method in the multivariate case we presented the method paying attention to it.

In 1D, we have $F_1(\omega) = E_1(\omega)$ and $F_2(\omega) = E_2(\omega)$, the transpose are equal the element itself, and the products are commutative, thus:

$$\hat{\mathcal{C}}(\omega) =(1 - \hat{K}(\omega))^{-1}\Big(\hat{K}(\omega)D + \hat{C_{3_3}}(\omega) - \hat{K}(\omega)D\hat{K}(-\omega) - E_1(\omega) + \hat{K}(-\omega)D + \hat{C_{3_3}}(-\omega) - E_1(-\omega)\Big)(1 - \hat{K}(-\omega))^{-1}.$$

Finally, using Eq. (A.4.6), we obtain for QHawkes in 1D:

$$\hat{\mathcal{C}^\star}(\omega) = (1 - \hat{K}(\omega))^{-1}\Big(D + \hat{C_{3_3}}(\omega) + \hat{C_{3_3}}(-\omega) - E_1(\omega) - E_1(-\omega)\Big)(1 - \hat{K}(-\omega))^{-1}.$$

For the multivariate case the matrices we integrate in C_{3_3}, E_1, E_2, F_1 and F_2, will be a mix of $\mathbb{D}_\mathrm{d}$, $\mathbb{D}$, $\mathbb{K}_\mathrm{d}$ and $\mathbf{K}$.

A.5 Asymptotic Study- Results Tables

Based on the notations in Appendix A.4.1, we give here the results of the asymptotic behavior of covariance structures when considering power law decaying kernels.

In both non critical and critical case, we have the same result for the exponent of non time diagonal parts $\delta = \rho + \delta_o - \frac{1}{2}$ and $\delta_{\mathrm{o}} = \rho_{\mathrm{o}}$. The exponent $\delta_{\mathrm{d}} = \rho_{\mathrm{d}}$ can take 2 values:

1. $\delta_{\mathrm{d}} = \rho_{\mathrm{d}}$
2. $\delta_{\mathrm{d}} = 2\rho + \delta_{\mathrm{o}} - \frac{1}{2}$

From empirical observations, we see that it makes sense if the diagonal terms persist longer in time than the non diagonal terms. Hence, we expect $\delta_{\mathrm{d}} < \delta_{\mathrm{o}}$. Considering this hypothesis, we would only keep the first case, $\delta_{\mathrm{d}} = \rho_{\mathrm{d}} < \rho_{\mathrm{o}}$.

For the time diagonal exponent (β_{d},β_{o},β_{d}, β_{o}), we need to make a difference between the non critical case ($n_H < 1$), and the critical case ($n_H = 1$). We call $\epsilon = \min(\epsilon_{\mathrm{d}}, \epsilon_{\mathrm{o}})$ and $\rho = \frac{1}{2}\min(\delta_{\mathrm{d}} + \rho_{\mathrm{d}}, \delta_{\mathrm{o}} + \rho_{\mathrm{o}}, \delta_{\mathrm{d}} + \rho_{\mathrm{o}}, \delta_{\mathrm{o}} + \rho_{\mathrm{d}}, \delta + \rho) = \frac{1}{2}\min(\rho_{\mathrm{d}} + \rho_{\mathrm{o}}, 2\rho_{\mathrm{o}}, 2\rho_2, \delta + \rho)$.

Non Critical		
	if $\frac{3+\epsilon_{\mathrm{d}}}{4} < \min(\rho_{\mathrm{d}}, \rho + \frac{\rho_{\mathrm{o}}}{2} + \frac{1}{4})$	if $\frac{3+\epsilon_{\mathrm{d}}}{4} > \min(\rho_{\mathrm{d}}, \rho + \frac{\rho_{\mathrm{o}}}{2} + \frac{1}{4})$
β_{d}	$\beta_{\mathrm{d}} = 1 + \epsilon_{\mathrm{d}}$	$\beta_{\mathrm{d}} = \min(4\rho + 2\rho_{\mathrm{o}} - 1, 4\rho_{\mathrm{d}} - 2)$
	if $\frac{3+\epsilon_{\mathrm{o}}}{4} < \min(\frac{\rho_{\mathrm{d}}+\rho_{\mathrm{o}}}{2}, \rho + \frac{\rho_{\mathrm{o}}}{2} + \frac{1}{4})$	if $\frac{3+\epsilon_{\mathrm{o}}}{4} > \min(\frac{\rho_{\mathrm{d}}+\rho_{\mathrm{o}}}{2}, \rho + \frac{\rho_{\mathrm{o}}}{2} + \frac{1}{4})$
β_{o}	$\beta_{\mathrm{o}} = 1 + \epsilon_{\mathrm{o}}$	$\beta_{\mathrm{o}} = \min(4\rho + 2\rho_{\mathrm{o}} - 1, 2(\rho_{\mathrm{d}} + \rho_{\mathrm{o}}) - 2)$
Critical		
	if $\rho < \frac{3}{2}$	if $\rho > \frac{3}{2}$
$\beta_{\mathrm{d}} = \beta_{\mathrm{o}} = \beta$	$\beta = 4\rho - 2 - 2\epsilon$	$\beta = 1 - 2\epsilon$

Table A.1: $\mathbb{C}$ exponents for both non critical and critical case, when using the following notations $\epsilon = \min(\epsilon_{\mathrm{d}}, \epsilon_{\mathrm{o}})$ and $\rho = \frac{1}{2}\min(\delta_{\mathrm{d}} + \rho_{\mathrm{d}}, \delta_{\mathrm{o}} + \rho_{\mathrm{o}}, \delta_{\mathrm{d}} + \rho_{\mathrm{o}}, \delta_{\mathrm{o}} + \rho_{\mathrm{d}}, \delta + \rho)$

Non Critical		
	if $\rho < 2\rho_{\mathrm{d}} - \frac{1}{2}$	if $\rho > 2\rho_{\mathrm{d}} - \frac{1}{2}$
β_{d}	$\beta_{\mathrm{d}} = 2\rho + \rho_{\mathrm{o}} - 1$	$\beta_{\mathrm{d}} = 2\rho_{\mathrm{d}} + \delta - 1$
	if $\rho < 2\rho_{\mathrm{o}} - \frac{1}{2}$	if $\rho > 2\rho_{\mathrm{d}} - \frac{1}{2}$
β_{o}	$\beta_{\mathrm{o}} = 2\rho + \rho_{\mathrm{o}} - 1$	$\beta_{\mathrm{o}} = 2\rho_{\mathrm{o}} + \delta - 1$
Critical		
	if $\frac{2\rho+\delta_o}{3} > \frac{2}{3}$ and $\frac{2\rho_d+\delta_\times}{3} > \frac{2}{3}$	else
β_{d}	$\beta_{\mathrm{d}} = 1 - \epsilon$	$\beta_{\mathrm{d}} = \min(2\rho_{\mathrm{d}} + \delta - \epsilon, 2\rho + \delta_{\mathrm{o}} - \epsilon)$
	if $\frac{2\rho+\delta_o}{3} > \frac{2}{3}$ and $\frac{2\rho_o+\delta_\times}{3} > \frac{2}{3}$	else
β_{o}	$\beta_{\mathrm{o}} = 1 - \epsilon$	$\beta_{\mathrm{o}} = \min(2\rho_{\mathrm{o}} + \delta - \epsilon, 2\rho + \delta_{\mathrm{o}} - \epsilon)$

Table A.2: $\mathbb{D}$ "time diagonal" exponents for both non critical and critical case, when using the following notations $\epsilon = \min(\epsilon_{\mathrm{d}}, \epsilon_{\mathrm{o}})$ and $\rho = \frac{1}{2}\min(\delta_{\mathrm{d}} + \rho_{\mathrm{d}}, \delta_{\mathrm{o}} + \rho_{\mathrm{o}}, \delta_{\mathrm{d}} + \rho_{\mathrm{o}}, \delta_{\mathrm{o}} + \rho_{\mathrm{d}}, \delta + \rho)$

A.6 Computation of the Infinitesimal Generator

A.6.1 For independent processes

We consider the processes, introducing a time scale $T > 0$ that will eventually diverge, we define the processes $\bar{h}_1^T(t) = h_1(tT)$, $\bar{h}_2^T(t) = h_2(tT)$, $\bar{z}_1^T(t) = z_1(tT)$ and $\bar{z}_2^T(t) = z_2(tT)$, $\bar{N}_t^{1T} = N_{tT}^1$, $\bar{N}_t^{2T} = N_{tT}^2$ with the parameters (β_{1T}, β_{2T}), $(\omega_{1T}, \omega_{2T})$ and $(\gamma_{1T}, \gamma_{2T})$ that may depend on T. In the decaying exponential case, we have

$$\begin{cases} \mathrm{d}(h_1)_t &= \beta_1(-(h_1)_t\mathrm{d}t + \mathrm{d}N_t^1) \\ \mathrm{d}(h_2)_t &= \beta_2(-(h_2)_t\mathrm{d}t + \mathrm{d}N_t^2) \\ \mathrm{d}(z_1)_t &= -\omega_1(z_1)_t\mathrm{d}t + \gamma_1\mathrm{d}P_t^1 \\ \mathrm{d}(z_2)_t &= -\omega_2(z_2)_t\mathrm{d}t + \gamma_2\mathrm{d}P_t^2 \end{cases} \Longrightarrow \begin{cases} \mathrm{d}(\bar{h}_1^T)_t &= -\beta_{1T}((\bar{h}_1^T)_tT\mathrm{d}t + n_H\mathrm{d}N_t^{1T}) \\ \mathrm{d}(\bar{h}_2^T)_t &= -\beta_{2T}((\bar{h}_2^T)_tT\mathrm{d}t + n_H\mathrm{d}N_t^{2T}) \\ \mathrm{d}(\bar{z}_1^T)_t &= -\omega_{1T}(\bar{z}_1^T)_tT\mathrm{d}t + \gamma_{1T}\mathrm{d}P_t^{1T} \\ \mathrm{d}(\bar{z}_2^T)_t &= -\omega_{2T}(\bar{z}_2^T)_tT\mathrm{d}t + \gamma_{2T}\mathrm{d}P_t^{2T} \end{cases}$$

The intensity of, respectively, the first and second process are given by $T[\lambda_\infty^1 + n_{H,1}^1\bar{h}_1^T + n_{H,2}^1\bar{h}_2^T + (a_{Z,1}^1\bar{z}_1^T)^2 + (a_{Z,2}^1\bar{z}_2^T)^2 + a_{Z,\times}^1\bar{z}_1^T\bar{z}_2^T]$ and $T[\lambda_\infty^2 + n_{H,1}^2\bar{h}_1^T + n_{H,2}^2\bar{h}_2^T + (a_{Z,1}^2\bar{z}_1^T)^2 + (a_{Z,2}^2\bar{z}_2^T)^2 + a_{Z,\times}^2\bar{z}_1^T\bar{z}_2^T]$. Moreover, the price processes can either go up or down with same probability $\frac{1}{2}$.

Thus, the infinitesimal generator can be written using *Theorem 1.22 of Oksendal and Sulem*:

$$\begin{aligned} Af^T(h_1,h_2,z_1,z_2) =& -T\beta_{1T}h_1\partial_{h_1}f - T\beta_{2T}h_2\partial_{h_2}f - \omega_{1T}Tz_1\partial_{z_1}f - \omega_{2T}Tz_2\partial_{z_2}f \\ &+ T[\lambda_\infty^1 + n_{H,1}^1h_1 + n_{H,2}^1h_2 + (a_{Z,1}^1z_1)^2 + (a_{Z,2}^1z_2)^2 + a_{Z,\times}^1z_2z_1]\Big(\\ &\frac{1}{2}f(h_1+\beta_{T1},h_2,z_1+\omega_{1T},z_2) + \frac{1}{2}f(h_1+\beta_{T1},h_2,z_1-\gamma_{1T},z_2) - f(h_1,h_2,z_1,z_2)\Big) \\ &+ T[\lambda_\infty^2 + n_{H,1}^2h_1 + n_{H,2}^2h_2 + (a_{Z,1}^2z_1)^2 + (a_{Z,2}^2z_2)^2 + a_{Z,\times}^2z_2z_1]\Big(\\ &\frac{1}{2}f(h_1,h_2+\beta_{T2},z_1,z_2-\omega_{2T}) + \frac{1}{2}f(h_1,h_2+\beta_{T2},z_1,z_2+\gamma_{2T}) - f(h_1,h_2,z_1,z_2)\Big) \end{aligned}$$

As in [70], we use the scaling $\beta_{1T} = \frac{\bar{\beta}_1}{T}$, $\beta_{2T} = \frac{\bar{\beta}_2}{T}$, $\omega_{1T} = \frac{\bar{\omega}_1}{T}$, $\omega_{2T} = \frac{\bar{\omega}_2}{T}$, $\gamma_{1T} = \frac{\bar{\gamma}_1}{T}$, $\gamma_{2T} = \frac{\bar{\gamma}_2}{T}$ and the Taylor development when $T \to +\infty$ which result in:

$$\begin{aligned} Af^T(h_1,h_2,z_1,z_2) =& -\bar{\beta}_1h_1\partial_{h_1}f - \bar{\beta}_2h_2\partial_{h_2}f - \bar{\omega}_1z_1\partial_{z_1}f - \bar{\omega}_2z_2\partial_{z_2}f \\ &+ T[\lambda_\infty^1 + n_{H,1}^1h_1 + n_{H,2}^1h_2 + (a_{Z,1}^1z_1)^2 + (a_{Z,2}^1z_2)^2 + a_{Z,\times}^1z_2z_1]\Big(\frac{\bar{\beta}_1}{T}\partial_{h_1}f + \frac{(\bar{\gamma}_1)^2}{2T}\partial_{z_1z_1}^2f + o(\frac{1}{T})\Big) \\ &+ T[\lambda_\infty^2 + n_{H,1}^2h_1 + n_{H,2}^2h_2 + (a_{Z,1}^2z_1)^2 + (a_{Z,2}^2z_2)^2 + a_{Z,\times}^2z_2z_1]\Big(\frac{\bar{\beta}_2}{T}\partial_{h_2}f + \frac{(\bar{\gamma}_2)^2}{2T}\partial_{z_2z_2}^2f + o(\frac{1}{T})\Big) \end{aligned}$$

Thus, the infinitesimal generator is given by:

$$\begin{aligned} Af^\infty(h_1,h_2,z_1,z_2) =& -\bar{\beta}_1h_1\partial_{h_1}f - \bar{\beta}_2h_2\partial_{h_2}f - \bar{\omega}_1z_1\partial_{z_1}f - \bar{\omega}_2z_2\partial_{z_2}f \\ &+ [\lambda_\infty^1 + n_{H,1}^1h_1 + n_{H,2}^1h_2 + (a_{Z,1}^1z_1)^2 + (a_{Z,2}^1z_2)^2 + a_{Z,\times}^1z_2z_1]\Big(n_H\bar{\beta}_1\partial_{h_1}f + \frac{(\bar{\gamma}_1)^2}{2}\partial_{z_1z_1}^2f\Big) \\ &+ [\lambda_\infty^2 + n_{H,1}^2h_1 + n_{H,2}^2h_2 + (a_{Z,1}^2z_1)^2 + (a_{Z,2}^2z_2)^2 + a_{Z,\times}^2z_2z_1]\Big(n_H\bar{\beta}_2\partial_{h_2}f + \frac{(\bar{\gamma}_2)^2}{2}\partial_{z_2z_2}^2f\Big) \end{aligned}$$

One can also include co-jumps, which would change the coefficients above and add a term of the form $\rho\partial_{z_1z_2}^2f$, where ρ is the correlation between the Poisson processes driving 1 and 2.

When considering only ZHawkes without Hawkes ($n_H = 0$), we have $h_1 = h_2 = 0$, thus the infinitesimal generator results in:

$$\begin{aligned} Af^\infty(z_1,z_2) =& -\bar{\omega}_1z_1\partial_{z_1}f - \bar{\omega}_2z_2\partial_{z_2}f \\ &+ [\lambda_\infty^1 + (a_{Z,1}^1z_1)^2 + (a_{Z,2}^1z_2)^2 + a_{Z,\times}^1z_2z_1]\frac{(\bar{\gamma}_1)^2}{2}\partial_{z_1z_1}^2f \\ &+ [\lambda_\infty^2 + (a_{Z,1}^2z_1)^2 + (a_{Z,2}^2z_2)^2 + a_{Z,\times}^2z_2z_1]\frac{(\bar{\gamma}_2)^2}{2}\partial_{z_2z_2}^2f \end{aligned}$$

A.6.2 In the bivariate case

In this case, both processes have the same sign with probability ρ, and they are centered. So, their behavior can be describe such that:

- both go down with probability $\frac{1+\rho}{4}$
- both go up with probability $\frac{1+\rho}{4}$
- first one goes up and second one goes down with probability $\frac{1-\rho}{4}$
- first one goes down and second one goes up with probability $\frac{1-\rho}{4}$

Thus, the infinitesimal generator can be written, we use *Theorem 1.22 of Oksendal and Sulem*:

$$
\begin{aligned}
Af^T(h_1,h_2,z_1,z_2) = &- T\beta_{1T}h_1\partial_{h_1}f - T\beta_{2T}h_2\partial_{h_2}f - \omega_{1T}Tz_1\partial_{z_1}f - \omega_{2T}Tz_2\partial_{z_2}f \\
&+ T[\lambda^1_\infty + n^1_{H,1}h_1 + n^1_{H,2}h_2 + (a^1_{Z,1}z_1)^2 + (a^1_{Z,2}z_2)^2 + a^1_{Z,\times}z_2z_1]\Big(\\
&\frac{1+\rho}{4}f(h_1+n_H\beta_{T1}, h_2+n_H\beta_{T2}, z_1-\gamma_{1T}, z_2-\gamma_{2T}) \\
&+ \frac{1+\rho}{4}f(h_1+n_H\beta_{T1}, h_2+n_H\beta_{T2}, z_1+\gamma_{1T}, z_2+\gamma_{2T}) \\
&+ \frac{1-\rho}{4}f(h_1+n_H\beta_{T1}, h_2+n_H\beta_{T2}, z_1+\gamma_{1T}, z_2-\gamma_{2T}) \\
&+ \frac{1-\rho}{4}f(h_1+n_H\beta_{T1}, h_2+n_H\beta_{T2}, z_1-\gamma_{1T}, z_2+\gamma_{2T}) \\
&- f(h_1,h_2,z_1,z_2)\Big) \\
&+ [\lambda^2_\infty + n^2_{H,1}h_1 + n^2_{H,2}h_2 + (a^2_{Z,1}z_1)^2 + (a^2_{Z,2}z_2)^2 + a^2_{Z,\times}z_2z_1]\Big(\\
&\frac{1+\rho}{4}f(h_1+n_H\beta_{T1}, h_2+n_H\beta_{T2}, z_1-\gamma_{1T}, z_2-\gamma_{2T}) \\
&+ \frac{1+\rho}{4}f(h_1+n_H\beta_{T1}, h_2+n_H\beta_{T2}, z_1+\gamma_{1T}, z_2+\gamma_{2T}) \\
&+ \frac{1-\rho}{4}f(h_1+n_H\beta_{T1}, h_2+n_H\beta_{T2}, z_1+\gamma_{1T}, z_2-\gamma_{2T}) \\
&+ \frac{1-\rho}{4}f(h_1+n_H\beta_{T1}, h_2+n_H\beta_{T2}, z_1-\gamma_{1T}, z_2+\gamma_{2T}) \\
&- f(h_1,h_2,z_1,z_2)\Big)
\end{aligned}
$$

As before, we use the scaling $\beta_{1T} = \frac{\bar\beta_1}{T}$, $\beta_{2T} = \frac{\bar\beta_2}{T}$, $\omega_{1T} = \frac{\bar\omega_1}{T}$, $\omega_{2T} = \frac{\bar\omega_2}{T}$, $\gamma_{1T} = \frac{\bar\gamma_1}{T}$, $\gamma_{2T} = \frac{\bar\gamma_2}{T}$ and the Taylor development when $T \to +\infty$ which result in:

$$
\begin{aligned}
Af^T(h_1,h_2,z_1,z_2) = &- \bar\beta_1h_1\partial_{h_1}f - \bar\beta_2h_2\partial_{h_2}f - \bar\omega_1z_1\partial_{z_1}f - \bar\omega_2z_2\partial_{z_2}f \\
&+ T[\lambda^1_\infty + n^1_{H,1}h_1 + n^1_{H,2}h_2 + (a^1_{Z,1}z_1)^2 + (a^1_{Z,2}z_2)^2 + a^1_{Z,\times}z_2z_1]\Big(\\
&n_H\frac{\bar\beta_1}{T}\partial_{h_1}f + n_H\frac{\bar\beta_2}{T}\partial_{h_2}f + \frac{(\bar\gamma_1)^2}{2T}\partial^2_{z_1z_1}f + \frac{(\bar\gamma_2)^2}{2T}\partial^2_{z_2z_2}f + \rho\frac{\bar\gamma_2\bar\gamma_1}{T}\partial_{z_1z_2}f + o(\frac{1}{T})\Big) \\
&+ T[\lambda^2_\infty + n^2_{H,1}h_1 + n^2_{H,2}h_2 + (a^2_{Z,1}z_1)^2 + (a^2_{Z,2}z_2)^2 + a^2_{Z,\times}z_2z_1]\Big(\\
&n_H\frac{\bar\beta_1}{T}\partial_{h_1}f + n_H\frac{\bar\beta_2}{T}\partial_{h_2}f + \frac{(\gamma_1)^2}{2T}\partial^2_{z_1z_1}f + \frac{(\bar\gamma_2)^2}{2T}\partial^2_{z_2z_2}f + \rho\frac{\bar\gamma_2\bar\gamma_1}{T}\partial_{z_1z_2}f + o(\frac{1}{T})\Big)
\end{aligned}
$$

Thus, the infinitesimal generator is given by:

$$\begin{aligned}
Af^{\infty}(h_1,h_2,z_1,z_2) = &-\bar{\beta}_1 h_1 \partial_{h_1} f - \bar{\beta}_2 h_2 \partial_{h_2} f - \bar{\omega}_1 z_1 \partial_{z_1} f - \bar{\omega}_2 z_2 \partial_{z_2} f \\
&+ [\lambda^1_{\infty} + n^1_{H,1} h_1 + n^1_{H,2} h_2 + (a^1_{Z,1} z_1)^2 + (a^1_{Z,2} z_2)^2 + a^1_{Z,\times} z_2 z_1] \Big(\\
&n_H \bar{\beta}_1 \partial_{h_1} f + n_H \beta_2 \partial_{h_2} f + \frac{(\bar{\gamma}_1)^2}{2} \partial^2_{z_1 z_1} f + \frac{(\bar{\gamma}_2)^2}{2} \partial^2_{z_2 z_2} f + \rho \bar{\gamma}_2 \bar{\gamma}_1 \partial_{z_1 z_2} f \Big) \\
&+ [\lambda^2_{\infty} + n^2_{H,1} h_1 + n^2_{H,2} h_2 + (a^2_{Z,1} z_1)^2 + (a^2_{Z,2} z_2)^2 + a^2_{Z,\times} z_2 z_1] \Big(\\
&n_H \bar{\beta}_1 \partial_{h_1} f + n_H \bar{\beta}_2 \partial_{h_2} f + \frac{(\gamma_1)^2}{2} \partial^2_{z_1 z_1} f + \frac{(\bar{\gamma}_2)^2}{2} \partial^2_{z_2 z_2} f + \rho \bar{\gamma}_2 \bar{\gamma}_1 \partial_{z_1 z_2} f \Big)
\end{aligned}$$

When considering only ZHawkes without Hawkes ($n_H = 0$), we have $h_1 = h_2 = 0$, thus the infinitesimal generator results in:

$$\begin{aligned}
Af^{\infty}(z_1,z_2) = &-\bar{\omega}_1 z_1 \partial_{z_1} f - \bar{\omega}_2 z_2 \partial_{z_2} f \\
&+ [\lambda^1_{\infty} + (a^1_{Z,1} z_1)^2 + (a^1_{Z,2} z_2)^2 + a^1_{Z,\times} z_2 z_1] \Big(\frac{(\bar{\gamma}_1)^2}{2} \partial^2_{z_1 z_1} f + \frac{(\bar{\gamma}_2)^2}{2} \partial^2_{z_2 z_2} f + \rho \bar{\gamma}_2 \bar{\gamma}_1 \partial_{z_1 z_2} f \Big) \\
&+ [\lambda^2_{\infty} + (a^2_{Z,1} z_1)^2 + (a^2_{Z,2} z_2)^2 + a^2_{Z,\times} z_2 z_1] \Big(\frac{(\bar{\gamma}_1)^2}{2} \partial^2_{z_1 z_1} f + \frac{(\bar{\gamma}_2)^2}{2} \partial^2_{z_2 z_2} f + \rho \bar{\gamma}_2 \bar{\gamma}_1 \partial_{z_1 z_2} f \Big)
\end{aligned}$$

A.7 General ODE for $F(\theta)$

We write here the general ODE on $F(\theta)$ for any parameters a_Z's and ω's:

$$
\begin{aligned}
&\Big(\omega_1+\omega_2+(a^1_{Z,1}\omega_1)^2+(a^2_{Z,2}\omega_2)^2-\alpha(\omega_1+2(a^1_{Z,1}\omega_1)^2)\cos^2(\theta)-\alpha(\omega_2+2(a^2_{Z,2}\omega_2)^2)\sin^2(\theta)\\
&-(a^2_{Z,}(\omega_2)^2+a^1_{Z,}(\omega_1)^2)\alpha\cos\theta\sin\theta\\
&+\frac{1}{8}\alpha(\alpha+(2+\alpha)\cos(2\theta))(\omega_1)^2\left((a^1_{Z,1})^2+(a^1_{Z,2})^2+((a^1_{Z,1})^2-(a^1_{Z,2})^2)\cos(2\theta)+a^1_{Z,}\sin(2\theta)\right)\\
&+\frac{1}{8}\alpha(\alpha-(2+\alpha)\cos(2\theta))(\omega_2)^2\left((a^2_{Z,1})^2+(a^2_{Z,2})^2+((a^2_{Z,1})^2-(a^2_{Z,2})^2)\cos(2\theta)+a^2_{Z,}\sin(2\theta)\right)\Big)F(\theta)\\
&+\Big(a^2_{Z,}(\omega_2)^2\cos^2(\theta)-a^1_{Z,}(\omega_1)^2\sin^2(\theta)-\cos\theta\sin\theta\left(\omega_1-\omega_2+2(a^1_{Z,1}\omega_1)^2-2(a^2_{Z,2}\omega_2)^2\right)\\
&+\frac{1}{4}(1+\alpha)(\omega_1)^2\sin(2\theta)((a^1_{Z,1})^2+(a^1_{Z,2})^2+((a^1_{Z,1})^2-(a^1_{Z,2})^2)\cos(2\theta)+a^1_{Z,}\sin(2\theta))\\
&-\frac{1}{4}(1+\alpha)(\omega_2)^2\sin(2\theta)((a^2_{Z,1})^2+(a^2_{Z,2})^2+((a^2_{Z,1})^2-(a^2_{Z,2})^2)\cos(2\theta)+a^2_{Z,}\sin(2\theta))\Big)F'(\theta)\\
&+\frac{1}{4}\Big((\omega_1)^2\sin^2(\theta)((a^1_{Z,1})^2+(a^1_{Z,2})^2+((a^1_{Z,1})^2-(a^1_{Z,2})^2)\cos(2\theta)+a^1_{Z,}\sin(2\theta))\\
&+(\omega_2)^2\cos^2(\theta)((a^2_{Z,1})^2+(a^2_{Z,2})^2+((a^2_{Z,1})^2-(a^2_{Z,2})^2)\cos(2\theta)+a^2_{Z,}\sin(2\theta))\Big)F''(\theta)=0
\end{aligned}
\tag{A.7.1}
$$

Appendix B

Appendix of Chapter 4: MQHawkes Calibration

B.1 Simulation of MQHawkes

This section aims at detailing the (M)QHawkes simulation. The first part outlines the thinning algorithms used for the generation of (M)QHawkes processes, as introduced by Ogata (1981) [72]. The second part presents various results and methods that are useful when studying synthetic (Q)Hawkes processes.

B.1.1 Generate synthetic MZHawkes – Thinning algorithm

The section specifies the thinning algorithm for (M)ZHawkes processes. Firstly, we detail the steps for generating an univariate ZHawkes with exponential kernels. Secondly, we highlight the modifications needed for generating multivariate processes.

Generate synthetic univariate ZHawkes with exponential kernels

We consider the ZHawkes process as defined in [70] and presented in Chapter 2. Hence, the intensity λ of the ZHawkes process N with marks $\mathrm{d}P$ is defined as

$$\lambda_t = \lambda_\infty + \int_{-\infty}^{t} h(t-s)\mathrm{d}N_s + \left(\int_{-\infty}^{t} k(t-s)\mathrm{d}P_s\right)^2,$$
$$\text{with } h(t) = n_H \beta \exp(-\beta t) \text{ and } k(t) = \sqrt{2 n_Z \omega}\exp(-\omega t).$$

In the following, the Hawkes feedback loop is denoted H and the quadratic feedback loop Z, i.e.

$$H_t = \int_{-\infty}^{t} h(t-s)\mathrm{d}N_s, \quad Z_t = \int_{-\infty}^{t} k(t-s)\mathrm{d}P_s.$$

Generating a ZHawkes process requires tracking event times and marks to compute the feedback loops H and Z at each instance. For kernels with general form, it is usually necessary to recalculate H and Z from the beginning at each step to determine the probability of a new jump, resulting in significant time consumption during the generation process. Exponential kernels offer the advantage[1] of rendering the feedback loops Markovian (see Remark B.1.1 below). Therefore, computing the probability of a new jump comes down to updating, from the previous jump, the feedback loops H and Z, as illustrated subsequently and explained in Remark B.1.1.

We adopt the following notations in the algorithms below:

- t is the **list** of event times, i.e times t_i such that $\mathrm{d}N_{t_i \in t} = 1$;
- M is the **list** of event marks, also denoted $\mathrm{d}P_{t_i \in t}$, at each event;

[1] Note that this advantage persists with power-law kernels, which better approximate empirically observed kernels, as a power law can be approximated through a sum of exponentials as demonstrated in [128].

To initiate the generation process, we assume no jumps have occurred yet. Therefore, the algorithm begins with initializing the intensity λ to its baseline value λ_∞, and setting the feedback loops H and Z to 0. Moreover, the lists of events and marks are initialized empty.

Algorithm 1 Simulation of ZHawkes 1D - Initialisation

$\lambda^* \leftarrow \lambda_\infty$
$t \leftarrow []$ ▷ List of time of jumps
$M \leftarrow []$ ▷ List of marks
$H \leftarrow 0$ ▷ Hawkes component
$Z \leftarrow 0$ ▷ quadratic component

Next, a specific step is dedicated to generating the first jump, which follows a Poisson process with a baseline intensity of λ_∞. This initial jump initiates the feedback loops H and Z and modifies the intensity for future events. For each event, including this initial one, the marks are randomly assigned as either plus or minus 1 with equal probability.

Algorithm 2 Simulation of ZHawkes 1D - First jump

$\lambda^* \leftarrow \lambda_\infty$
Generate $u \sim U([0,1])$
$s \leftarrow -\frac{1}{\lambda^*}\ln(u)$
Ensure: $s < T$
t append (s)
Generate $m = \pm 1$
M append m
$H \leftarrow n_H\beta$
$Z \leftarrow \sqrt{2n_Z\omega} \times m$
$\lambda \leftarrow \lambda_\infty + H + Z^2$
$\lambda^* \leftarrow \lambda + n_H\beta + 2n_Z\omega$

After the generation of the first event, the feedback loops become active and contribute to the intensity, thereby influencing the probability of subsequent event occurrences. At each iteration, a new event time is proposed for consideration, and its acceptance depends on the intensity with and without the proposed event. The generation process yields a list of event times (t) along with their corresponding marks (M).

Remark B.1.1. Exponential kernels have the advantage to render the feedback loops Markovian, accelerating the generation of such point process. To understand the Markovian property let us consider the feedback loop H at some event time t_n:

$$\begin{aligned}
\frac{1}{n_H\beta}H_{t_n} &= \sum_{t_i, i\in 1,n} \exp(-\beta(t_n - t_i)) \\
\frac{1}{n_H\beta}H_{t_n} &= 1 + \exp(-\beta(t_n - t_{n-1})) \sum_{t_i, i\in 1,n-1} \exp(-\beta(t_{n-1} - t_i))) \\
\frac{1}{n_H\beta}H_{t_n} &= 1 + \exp(-\beta(t_n - t_{n-1}))\frac{1}{n_H\beta}H_{t_{n-1}} \\
H_{t_n} &= n_H\beta + \exp(-\beta(t_n - t_{n-1}))H_{t_{n-1}}.
\end{aligned}$$

Generate synthetic multivariate ZHawkes with exponential kernels

In multivariate environment, the algorithm operates similarly to generate each time jump, with three notable exceptions:

- As discussed in Chapter 3, independent processes have a zero probability of jumping at the exact same time. Therefore, for each "proposed" jump time, it is necessary to determine which process, if any, is the more likely to jump. Below, we outline how Algorithms 2 and 3 can be modified to accommodate this requirement.

Algorithm 3 Simulation of ZHawkes 1D - other jumps

while $s < T$ **do**
  Generate $u \sim U([0,1])$
  $v \leftarrow -\frac{\log(u)}{\lambda^*}$
  $s \leftarrow s + v$
  $H_{tmp} \leftarrow H \exp(-\beta(s - t[-1]))$
  $Z_{tmp} \leftarrow Z \exp(-2\omega(s - t[-1]))$
  $\lambda_{tmp} \leftarrow \lambda_\infty + H_{tmp} + Z_{tmp}^2$
  Generate $u \sim U([0,1])$
  if $u \leq \frac{\lambda_{tmp}}{\lambda^*}$ **then**
    t append (s)
    Generate $m = \pm 1$
    M append m
    $H \leftarrow H_{tmp} + n_H \beta$
    $Z \leftarrow Z_{tmp} + \sqrt{2 n_Z \omega} \times m$
    $\lambda \leftarrow \lambda_\infty + H + Z^2$
    $\lambda^* \leftarrow \lambda + n_H \beta + 2 n_Z \omega$
  else
    $\lambda^* \leftarrow \lambda_{tmp}$
  end if
end while
Return t,M

- Since we deal with several assets, λ, H and Z are not scalars anymore but arrays. In particular, in the case of 2 dimensions, λ is a 2×1 array (representing the intensities of the two processes), and H and Z are 2×2 arrays (each intensity having both self-feedback and cross feedback component). Additionally, it is necessary to track which process is jumping at each event time. This is managed using the array *jump* which records the index of the jumping processes for each event time in the list t.

- A new variable I^K defined such that $I^K(t) = \sum_{i=1}^K \lambda^i(t)$, which represents the sum the K first intensities at time t, is necessary for the extension.

We now proceed to describe the modifications made to the algorithm steps to accommodate a multivariate process of size M, with intensities denoted as $\lambda^i_{i \in [\![1,M]\!]}$.

The initialisation of the algorithm for generating MZHawkes is similar to that of the univariate case, with the addition of the variable *jump*. As previously mentionned, *jump* records the indices of the jumping processes for each event time in the list t. Since no jump have occurred at this stage, *jump*, as t and M, is empty.

Algorithm 4 Simulation of MZHawkes – Initialisation

$I^* \leftarrow \sum_{i=1}^M \lambda^i_\infty$
$t \leftarrow []$ ▷ List of time of jumps
$M \leftarrow []$ ▷ List of marks
$jump \leftarrow []$ ▷ List of index of the jumping process

At each step, the occurrence of a jump depends on the comparison between the sum of intensities with (I^*) and without (I^M) a new jump. Subsequently, determining which process, if any, is affected by the jump is managed by finding n_0 such that $\frac{I^{n_0-1}(s)}{I^*} < w < \frac{I^{n_0}(s)}{I^*}$, where w is a random uniform variable. Thus, the generation of the first jump proceeds as follows.

At each event, the feedback loops H and Z are updated in line with the event that has just occurred. It is important to note that not all components of H and Z undergo the same update process at each step: the feedback loops of the jumping process will increase, while that of other processes will decay exponentially. For instance, if process 1 jumps, only the feedback loops H_1 and Z_1 will be increased by a jump amplitude (which corresponded to $n_H \beta$ and $\sqrt{2 n_Z \omega}$ in the 1D

Algorithm 5 Simulation of MZHawkes – First jump

Generate $u \sim U([0,1])$
$s \leftarrow -\frac{1}{\lambda^*}\ln(u)$
Ensure: $s < T$
t append (s)
M append (± 1 (generated with equal probability))
Generate $w \sim U([0,1])$
find n_0 such that $\frac{I^{n_0-1}(0)}{I^*} < w < \frac{I^{n_0}(0)}{I^*}$
$jump$ append n_0

case), while the H_0 and Z_0 will decay exponentially. This method holds for each event, including the first one.

Algorithm 6 Simulation of MZHawkes – other jumps

while $s < T$ **do**
  Generate $u \sim U([0,1])$
  $v \leftarrow -\frac{\log(u)}{\lambda^*}$
  $s \leftarrow s + v$
  $I^* \leftarrow I^M(t[-1]) + \sum_{i=1}^M n_H^i \beta^i + 2\sum_{i=1}^M n_Z^i \omega^i$
  Generate $w \sim U([0,1])$
  if $w \leq \frac{I^M(s)}{I^*}$ **then**
    t append (s)
    M append (± 1 (generated with equal probability))
    find n_0 such that $\frac{I^{n_0-1}(s)}{I^*} < w < \frac{I^{n_0}(s)}{I^*}$
    jump append n_0
  else
    $I^* \leftarrow I^M(s)$
  end if
end while
Return t,M,$jump$

The generation algorithm in the multivariate case results in 3 arrays t, M and $jump$, which denote the time, mark of events and which process is affected by the jumps.

Example
Let's consider the following output

- $t = [0.1, 0.5, 1.0, 2.1]$
- $M = [1, -1, 1, 1]$
- $jump = [0, 0, 1, 2]$

Then, it means that the process $i = 0$ jumped at time 0.1 and 0.5 with marks 1 and -1, whereas the process $i = 1$ jumped once at time 1.0 with mark 1 and the process $i = 2$ jumped once at time 2.1 with mark 1.

B.2 Computation of 1$^{\text{rst}}$ order moments from linear Hawkes simulation

This section addresses the computation of moments derived from the simulation of Hawkes processes and their extensions ((M)QHawkes). Accurate computation of moments is crucial for implementing non-parametric calibration of the model, making it pertinent to dedicate attention to this aspect.

Specifically, to maintain clarity, we focus on the first-order moment defined as

$$\mathcal{C}(\tau) = \frac{1}{\bar{\lambda}^2}\left(\mathbb{E}\left(\frac{\mathrm{d}N_t}{\mathrm{d}t}\frac{\mathrm{d}N_{t-\tau}}{\mathrm{d}t}\right) - \bar{\lambda}^2\right).$$

As evident from the definition of $\mathcal{C}$, the computation of moments depends on the parameter $\mathrm{d}t$, which signifies the timescale for computing covariances. The selection of $\mathrm{d}t$ influences moment calculations significantly. Specifically, for a small $\mathrm{d}t$ ($\mathrm{d}t \ll 1/\bar{\lambda}$), consecutive events are rare, potentially leading to underestimated correlations. Conversely, with a large $\mathrm{d}t$ ($\mathrm{d}t \gg 1/\bar{\lambda}$), numerous consecutive events occur, potentially resulting in overestimated correlations.

Furthermore, the introduction of a timescale discretizes a continuous process, potentially affecting some analytical results. In particular, below we derive the analytical form of $\mathcal{C}$ for both continuous and discrete processes.

B.2.1 Correlation function in continuous time

For a linear Hawkes process, the Yule-Walker equation relating the first order moment $\mathcal{C}$ to the feedback kernel ϕ is given by, for $\tau > 0$,

$$\mathcal{C}(\tau) = \frac{1}{\bar{\lambda}}\phi(\tau) + \int_0^{+\infty}\phi(u)\mathcal{C}(\tau-u)\mathrm{d}u. \tag{B.2.1}$$

For an exponential kernel of the form $\phi(u) = n_h\beta\exp(-\beta u)$, the covariance structure $\mathcal{C}$ is expressed as $\mathcal{C}(\tau > 0) = \frac{\beta n_h(2-n_h)}{2\phi_0}\exp(-(1-n_h)\beta t)$ [23]. The demonstration of this result is given below.

Proof. We look for $\mathcal{C}$ with the form ($\mathcal{C}(t>0) = a\exp(-bt)$), then the Yule Walker equation (B.2.1) gives:

$$\begin{aligned}
\mathcal{C}(\tau) =& \bar{\lambda}\phi(\tau) + \int_0^{+\infty}\phi(\tau+u)\mathcal{C}(u)\mathrm{d}u + \int_0^{\tau}\phi(\tau-u)\mathcal{C}(u)\mathrm{d}u \\
a\exp(-b\tau) =& \bar{\lambda}n_h\beta\exp(-\beta\tau) + \int_0^{+\infty} n_h\beta a\exp(-\beta(\tau+u)-bu)\mathrm{d}u + \int_0^{\tau} n_h\beta a\exp(-\beta(\tau-u)-bu)\mathrm{d}u \\
a\exp(-b\tau) =& \bar{\lambda}n_h\beta\exp(-\beta\tau) + \frac{n_h\beta a}{\beta+b}\exp(-\beta\tau) + \frac{n_h\beta a}{b-\beta}(\exp(-\beta\tau)-\exp(-b\tau)).
\end{aligned}$$

Thus, by writing down the equations on $\exp(-\beta\tau)$ and on $\exp(-b\tau)$, one obtains

$$\begin{cases} 0 = \bar{\lambda}n_h\beta + \frac{an_h\beta}{\beta+b} + \frac{an_h\beta}{b-\beta} \\ a = -\frac{an_h\beta}{b-\beta} \end{cases} \iff \begin{cases} a = \frac{\Phi_0}{1-n_h}\frac{\beta n_h(2-n_h)}{2(1-n_h)} \\ b = \beta(1-n_h) \end{cases}.$$

□

B.2.2 Approximation of the continuous exponential in discrete time framework

Certain combination of parameters pose difficulties in recovering $\mathcal{C}$ in its continuous form as stated above, primarily due to process discretization. Therefore, we derive the theoretical expression of the autocorrelation structure $\mathcal{C}$ within a discrete framework. Specifically, we acknowledge the approximation

$$\begin{aligned}
\lim_{n\to+\infty}(1+\frac{x}{n})^n =& \exp(x), \\
\text{then, } \Lambda^\tau =& \exp(\tau\ln(\Lambda)) \underset{\Lambda\sim 1}{\sim} \exp(\tau(\Lambda-1)).
\end{aligned}$$

We can then write the discrete version of the exponential kernel as $\phi(\tau) = k_0\Lambda^\tau$. The link with the continuous case, where the kernel is an exponential such that $\phi_{\text{continuous}}(\tau) = n_h\beta\exp(-\beta\tau)$, is then directly $\Lambda - 1 \sim -\beta$ and $k_0 \sim n_h\beta$.

As before, the form of the covariance structure is directly defined by that of the kernel. Hence, in the discretized framework, $\mathcal{C}$ is fully characterised by parameters a, μ and $\bar{\lambda}$ such that $\mathcal{C}(\tau > 0) = a\mu^\tau$ and $\mathcal{C}(0) = \frac{1}{\bar{\lambda}}$, and where a and μ depends on k_0, Λ and $\bar{\lambda}$. Solving the discretized Yule-Walker equation results in $\mu = \Lambda(k_0 + 1)$ and $a = \frac{1}{\bar{\lambda}}\big(\frac{(1-\Lambda\mu)(1-\frac{\Lambda}{\mu})}{1-2\Lambda\mu+\Lambda^2}\big)$, as demonstrated below.

Proof. As before, we start from the Yule-Walker equation. The discretised version of Eq. (B.2.1) writes

$$\mathcal{C}(\tau) = \sum_{u=1}^{+\infty} \Phi(u)\mathcal{C}(\tau - u).$$

Now we can develop the computation:

$$\begin{aligned}
a\mu^\tau =& \sum_{u=1}^{\tau-1} \Phi(u)\mathcal{C}(\tau-u) + \Phi(\tau)\mathcal{C}(0) + \sum_{u=\tau+1}^{+\infty} \Phi(u)\mathcal{C}(u-\tau)\\
a\mu^\tau =& \sum_{u=1}^{\tau-1} k_0\Lambda^u a\mu^{\tau-u} + k_0\Lambda^\tau\frac{1}{\bar{\lambda}} + \sum_{u=\tau+1}^{+\infty} k_0\Lambda^u a\mu^{u-\tau}\\
a\mu^\tau =& k_0 a(\mu)^\tau \sum_{u=1}^{\tau-1}(\frac{\Lambda}{\mu})^u + k_0\Lambda^\tau\frac{1}{\bar{\lambda}} + k_0 a(\frac{1}{\mu})^\tau \sum_{u=\tau+1}^{+\infty}(\Lambda\mu)^u\\
a\mu^\tau =& k_0 a(\mu)^\tau\big(\frac{\frac{\Lambda}{\mu} - (\frac{\Lambda}{\mu})^\tau}{1-\frac{\Lambda}{\mu}}\big) + k_0\Lambda^\tau\frac{1}{\bar{\lambda}} + k_0 a(\frac{1}{\mu})^\tau\big(\frac{1}{1-\Lambda\mu} - \frac{1-(\Lambda\mu)^{\tau+1}}{1-\Lambda\mu}\big)\\
a\mu^\tau =& k_0 a\big(\frac{\Lambda\mu^\tau}{\mu-\Lambda}\big) - k_0 a\big(\frac{\Lambda^\tau}{1-\frac{\Lambda}{\mu}}\big) + k_0\Lambda^\tau\frac{1}{\bar{\lambda}} + k_0 a\frac{\Lambda^{\tau+1}\mu}{1-\Lambda\mu}.
\end{aligned}$$

By writing down the equations on Λ^τ and μ^τ, we obtain

$$\begin{cases} a = k_0 a\frac{\Lambda}{\mu-\Lambda} \\ 0 = -k_0 a\big(\frac{1}{1-\frac{\Lambda}{\mu}}\big) + k_0\frac{1}{\bar{\lambda}} + k_0 a\frac{\Lambda\mu}{1-\Lambda\mu} \end{cases}$$
$$\begin{cases} \mu = \Lambda(k_0+1) \\ 0 = a\big(\frac{-1}{1-\frac{\Lambda}{\mu}} + \frac{\Lambda\mu}{1-\Lambda\mu}\big) + \frac{1}{\bar{\lambda}} \end{cases}$$
$$\begin{cases} \mu = \Lambda(k_0+1) \\ a = \frac{1}{\bar{\lambda}}\big(\frac{(1-\Lambda\mu)(1-\frac{\Lambda}{\mu})}{1-\Lambda\mu-\Lambda\mu(1-\frac{\Lambda}{\mu})}\big) \end{cases}$$
$$\begin{cases} \mu = \Lambda(k_0+1) \\ a = \frac{1}{\bar{\lambda}}\big(\frac{(1-\Lambda\mu)(1-\frac{\Lambda}{\mu})}{1-2\Lambda\mu+\Lambda^2}\big) \end{cases}.$$

We check that these results are consistent in the continuous limit

$$\begin{cases} \mu \sim (-\beta+1)(n_h\beta+1) \\ a \sim \frac{1}{\bar{\lambda}}\big(\frac{\beta(2-n_h)\beta}{1-2(1-\beta(2-n_h))+(1-\beta)^2}\big) \end{cases}$$
$$\begin{cases} \mu = -n_h\beta^2 + \beta(n_h-1) + 1 \\ a \sim \frac{1}{\bar{\lambda}}\big(\frac{\beta(2-n_h)\beta}{1-2(1-\beta(2-n_h))+(1-\beta)^2}\big) \end{cases} \iff \begin{cases} \mu \sim \beta(n_h-1)+1 \\ a \sim \frac{1}{\bar{\lambda}}\frac{\beta n_h(2-n_h)}{2(1-n_h)} \end{cases}.$$

□

B.2.3 Aggregate the data

In practice, obtaining data at a microscale with precise event times is rare or computationally intensive. Often, aggregated data is used instead. Therefore, to consistently evaluate calibration procedures on simulations, aggregation methods should be applied beforehand on the synthetic processes.

In the case of QHawkes, where simulation results consist in lists of price changes with their respective timestamps, one can reconstruct the price time series, bin them, and estimate the volatility time series using one of the proxies detailed in Chapter 1.

Additionally, there are more simplistic aggregation methods, such as the approach outlined in [86], or directly summing the events within each bin. As previously discussed, the bin size significantly impacts the results and should be selected with care. Specifically, Bacry et al. [86] recommend setting $\mathrm{d}t = 1/\bar{\lambda}$.

B.3 Parametric Calibration - Maximum Likelihood

In this appendix, we describe a widely used parametric calibration method for (Q)Hawkes processes, with intensity λ. The method relies on maximising the likelihood of jump times. We present here two applications:

- maximum likelihood when exact jump times are available. Note, that it is the method used in the package `tick` of `Python` to calibrate (multidimensional) linear Hawkes.
- maximum likelihood with proxies, when event-by-event data is not available.

B.3.1 With exact time of events

Log likelihood for point process

For this first method of likelihood maximisation, one needs to have access to the event times $(t_i)_i$. The likelihood is then defined as the joint distribution of those time of events. Considering a Hawkes process of intensity λ, we first give the probability of event occurring between time t_i and $t_i + \mathrm{d}t$, without any event happening between time t_{i-1} and t_i.

Definition B.3.1 (probability of event between time t_i and $t_i + \mathrm{d}t$ but no event between t_{i-1} and t_i)**.**

$$\mathbb{P}\left(\text{one event in } [t_i, t_i + \mathrm{d}t] \text{ but no event in } [t_{i-1}, t_i]\right) = \exp\left(-\int_{t_{i-1}}^{t_i} \mathrm{d}u\lambda(u|\theta)\right)\lambda(t_i|\theta)\mathrm{d}t$$

From definition B.3.1, one can then write the joint distribution of N events happening in $[0, T)$ at arrival times $(t_i)_i$, and deduce the log-likelihood.

Definition B.3.2 (Log Likelihood of 1D point process)**.** For N events in $[0, T)$ at arrival times $(t_i)_i$, then, the log-likelihood is:

$$\ln(\mathbf{L}(t_1, t_2, ..., t_N, \theta)) = \sum_{i=1}^{N} \ln(\lambda(t_i/\theta)) - \int_0^T \mathrm{d}u\lambda(u/\theta) \tag{B.3.1}$$

where λ is the intensity and θ the parameters we are looking to estimate.

Proof. From definition B.3.1, we write the joint distribution of N independent events happening in $[0, T)$ at arrival times $(t_i)_i$

$$\begin{aligned}\mathbf{L}(t_1, t_2, ..., t_N, \theta) &= \left(\prod_{i=1}^{N} \exp\left(-\int_{t_{i-1}}^{t_i} \mathrm{d}u\lambda(u|\theta)\right)\lambda(t_i|\theta)\right)\exp\left(-\int_{t_N}^{T} \mathrm{d}u\lambda(u|\theta)\right)\\ &= \left(\prod_{i=1}^{N} \lambda(t_i|\theta)\right)\exp\left(-\int_0^T \mathrm{d}u\lambda(u|\theta)\right).\end{aligned}$$

Taking the log gives the result.

□

Log likelihood for linear Hawkes process

For linear Hawkes processes, the intensity is defined as

$$\lambda(t/\theta) = \lambda_\infty + \int_{-\infty}^{t} h_\theta(t-s)\mathrm{d}N_s = \lambda_\infty + \sum_{t_i<t} h(t-t_i)$$

where $(t_i)_i$ are the times of the jumps.

Particularly, with exponential kernels, we have $\theta = (\lambda_\infty, n_h, \beta)$, $h_\theta(t) = n_h\beta\exp(-\beta t)$. Then the expression of the log-likelihood evolves to:

$$\ln(\mathbf{L}(t_1, t_2, ..., t_N, \theta)) = -T\lambda_\infty + \sum_{i=1}^{N}\left(\ln(\phi_0 + n_H\beta H_i) - n_H(\exp(-\beta(T-t_i)) - 1)\right)$$

with

$$H_i = \sum_{k=1}^{i-1} \exp(-\beta(t_i - t_k))$$

Indeed, we have $\lambda(t_i/\theta) = \lambda_\infty + n_H \beta H_i$. And

$$\begin{aligned}
\int_0^T \mathrm{d}u \lambda(u/\theta) =& \lambda_\infty T + n_H \beta \int_0^T \int_{-\infty}^u \exp(-\beta(u-s)) \mathrm{d}N_s \mathrm{d}u \\
=& \lambda_\infty T + n_H \beta \int_{-\infty}^T \int_s^T \exp(-\beta(u-s)) \mathrm{d}u \mathrm{d}N_s \\
=& \lambda_\infty T + n_H \beta \int_{-\infty}^T \frac{1}{\beta}(1 - \exp(-\beta(T-s))) \mathrm{d}N_s \\
=& \lambda_\infty T + n_H \sum_{t_s < T} (1 - \exp(-\beta(T - t_s)))
\end{aligned}$$

Log likelihood for ZHawkes process

For ZHawkes processes, one needs to add the quadratic term to the intensity, which then becomes:

$$\begin{aligned}
\lambda(t/\theta) =& \lambda_\infty + \int_{-\infty}^t h_\theta(t-s) \mathrm{d}N_s + \big(\int_{-\infty}^t k_\theta(t-s) \mathrm{d}P_s \big)^2 \\
=& \lambda_\infty + \sum_{t_i < t} h(t - t_i) + \big(\sum_{t_i < t} k(t - t_i) m_{t_i} \big)^2
\end{aligned}$$

where $(t_i)_i$ are the times of the jumps and $m(t_i)_t$ the marks of the jump at time t_i. Again, in the particular case of exponential kernels, we have $\theta = (\lambda_\infty, n_H, \beta, n_Z, \omega)$, $h_\theta(t) = n_H \beta \exp(-\beta t)$, $k_\theta(t) = \sqrt{2 n_Z \omega} \exp(-\omega t)$. The log-likelihood can then entirely be defined by:

$$\ln(\mathbf{L}(t_1, t_2, ..., t_N, \theta)) = -T\lambda_\infty + \sum_{i=1}^N \big(\ln(\lambda_\infty + n_H \beta H_i + 2 n_Z \omega Z_i^2) - n_H(\exp(-\beta(T - t_i)) - 1) - 2 n_Z \omega Z_i^2 \big)$$

with

$$H_i = \sum_{k=1}^{i-1} \exp(-\beta(t_i - t_k)) \quad \text{and} \quad Z_i = \sum_{k=1}^{i-1} \exp(-\omega(t_i - t_k)) m_{t_k}$$

Log likelihood for multivariate point process (2D)

We consider a 2D process, with intensities (λ_1, λ_2) and events arrival times $((t_i^1)_{i \in [\![1,N]\!]}, (t_i^2)_{i \in [\![1,N]\!]})$, where $(t_i^1)_{i \in [\![1,N]\!]}$, respectfully $(t_i^2)_{i \in [\![1,N]\!]}$, are the times of events of process 1, respectfully process 2. The log likelihood consider the joint probability of the 2 jump processes, and thus results in:

Definition B.3.3 (Log likelihood in 2D)**.** Log likelihood of two point processes with intensities (λ_1, λ_2), the log-likelihood of N events from process 1 at times $(t_i^1)_{i \in [\![1,N]\!]}$, and N events of process 2 at times $(t_i^2)_{i \in [\![1,N]\!]}$, between 0 and time T is given by:

$$\ln(\mathbf{L}((t_i^1)_{i \in [\![1,N]\!]}, (t_i^2)_{i \in [\![1,N]\!]}, \theta)) = \sum_{i=1}^N \ln(\lambda_1(t_i^1/\theta)) + \sum_{i=1}^N \ln(\lambda_2(t_i^2/\theta)) - \int_0^T \mathrm{d}u (\lambda_1(u/\theta) + \lambda_2(u/\theta))$$

where θ is the vector of the parameters we are looking for.

B.3.2 Without exact times of events - Using an intensity estimate

Defining the (log) likelihood

It is sometimes impossible to access event times $(t_i)_i$ (like in Chapter 4 for instance). However, we can estimate λ_t for each time bin t. We can consider λ_t constant over the bin t, then the probability to observe n events in the bin t:

Definition B.3.4 (Probability of number of events in a bin)**.**

$$\mathbb{P}(\text{nb of events in bin } t = n) = \frac{\lambda_t^n}{n!}\exp(-\lambda_t)$$

The number of events we observe in bin t is our $\hat{\lambda}$ ($n = \hat{\lambda}$).
In the QHawkes framework, λ_t is defined by:

$$\lambda_t = \lambda_0 + \int_{-\infty}^{t} h(t-u)\mathrm{d}N_u + (\int_{-\infty}^{t} k(t-u)\mathrm{d}P_u)^2$$

Using the approximation $\mathrm{d}N_u \equiv \mathbb{E}(\mathrm{d}N_u) = \hat{\lambda}_u \mathrm{d}t$, we obtain:

$$\lambda_t = \lambda_0 + \sum_{u=0}^{t} h(t-u)\hat{\lambda}_u \mathrm{d}u + (\sum_{u=0}^{t} k(t-u)\mathrm{d}P_u)^2$$

where $u = 0$ represent the first bin of the day. Note that $\mathrm{d}P_t$ is the returns over the bin t, which is directly observable.

We can then define the log likelihood:

Definition B.3.5 (log likelihood of observing $\hat{\lambda}$)**.** We consider the joint density of observing the events $((\hat{\lambda}^i)_{i\in[\![1,N]\!]})$, where N represent the number of bins we have, then the log likelihood is given by:

$$\begin{aligned}
\ln(\mathbf{L}((\hat{\lambda}^i)_{i\in[\![1,N]\!]}, \theta)) &= \sum_{i=1}^{N} \ln\left(\frac{\lambda_i(\theta)^{\hat{\lambda}_i}}{\hat{\lambda}_i!}\exp(-\lambda_i(\theta))\right) = \sum_{i=1}^{N} \hat{\lambda}_i \ln(\lambda_i(\theta)) - \ln(\hat{\lambda}_i!) - \lambda_i(\theta) \\
&\sim \sum_{i=1}^{N} \hat{\lambda}_i \ln(\lambda_i(\theta)) - \left(\hat{\lambda}_i \ln(\hat{\lambda}_i) - \hat{\lambda}_i\right) - \lambda_i(\theta) \\
\ln(\mathbf{L}((\hat{\lambda}^i)_{i\in[\![1,N]\!]}, \theta)) &= \sum_{i=1}^{N} \hat{\lambda}_i \left(\ln(\lambda_i(\theta)) - \ln(\hat{\lambda}_i)\right) + \hat{\lambda}_i - \lambda_i(\theta)
\end{aligned}$$

Defining the (log) likelihood for multidimensionnal process

Definition B.3.6 (likelihood of observing $(\hat{\lambda}^1, \hat{\lambda}^2)$)**.** We consider 2 processes, with intensities (λ^1, λ^2). We observe those two processes over N bins, and the number of events from process j in bin i is $\hat{\lambda}_i^j$. Then, we can define the joint probability (likelihood), to observe $(\hat{\lambda}_i^1, \hat{\lambda}_i^2)_{i\in[\![1,N]\!]}$ over the N bins:

$$\begin{aligned}
\mathbf{L}((\hat{\lambda}_i^1)_i, (\hat{\lambda}_i^2)_i, \theta) &= \prod_{i=1}^{N} \mathbb{P}(n_1 \text{ events from process 1 in bin } i)\mathbb{P}(n_2 \text{ events from process 2 in bin } i) \\
&= \prod_{i=1}^{N} \frac{(\lambda_i^1(\theta))^{n_1}}{n_1!}\exp(-\lambda_i^1(\theta))\frac{(\lambda_i^2(\theta))^{n_2}}{n_2!}\exp(-\lambda_i^2(\theta)) \\
&= \prod_{i=1}^{N} \frac{(\lambda_i^1(\theta))^{\hat{\lambda}_i^1}}{\hat{\lambda}_i^1!}\exp(-\lambda_i^1(\theta))\frac{(\lambda_t^2(\theta))^{\hat{\lambda}_i^2}}{\hat{\lambda}_i^2!}\exp(-\lambda_i^2(\theta))
\end{aligned}$$

where θ represent the vector of parameters we are trying to estimate.

Consequently, the log likelihood writes:

$$\ln\left(\mathbf{L}((\hat{\lambda}_i^1)_i, (\hat{\lambda}_i^2)_i, \theta)\right) = \sum_{i=1}^{N} \hat{\lambda}_i^1 \ln(\lambda_i^1(\theta)) - \ln(\hat{\lambda}_i^1!) - \lambda_i^1(\theta) + \hat{\lambda}_i^2 \ln(\lambda_i^2(\theta)) - \ln(\hat{\lambda}_i^2!) - \lambda_i^2(\theta),$$

which we rewrite as:

$$\ln\big(\mathbf{L}((\hat{\lambda}_i^1)_i,(\hat{\lambda}_i^2)_i,\theta)\big)=\sum_{i=1}^{N}\hat{\lambda}_i^1\ln(\lambda_i^1(\theta))-\big(\hat{\lambda}_i^1\ln(\hat{\lambda}_i^1)-\hat{\lambda}_i^1\big)-\lambda_i^1(\theta)+\hat{\lambda}_i^2\ln(\lambda_i^2(\theta))-\big(\hat{\lambda}_i^2\ln(\hat{\lambda}_i^2)-\hat{\lambda}_i^2\big)-\lambda_i^2(\theta) \tag{B.3.2}$$

where:

- $\hat{\lambda}_i^{1,2}$: the estimated intensities of processes 1,2 in bin i
- $\lambda_i^{1,2}$: the theoretical intensities of processes 1 & 2 which depend on both the observation of $\hat{\lambda}$ and the parameters θ we are trying to estimate. In the case of a two dimensional ZHawkes, intensities are defined such that:

  $$\lambda_t^j=\lambda_\infty^j+\int_{-\infty}^{t}h_1^j(t-u)\mathrm{d}N_u^1+\int_{-\infty}^{t}h_2^j(t-u)\mathrm{d}N_u^2+(\int_{-\infty}^{t}k_1^j(t-u)\mathrm{d}P_u^1)^2+(\int_{-\infty}^{t}k_2^j(t-u)\mathrm{d}P_u^2)^2.$$

  In our discrete environment, we have ordered bins $(t_i)_{i\in[\![1,N]\!]}$, and thus we can write the intensity of process j in bin M such that

  $$\lambda_{t_M}^j=\lambda_\infty^j+\sum_{i=1}^{M-1}h_1^j(t_M-t_i)\hat{\lambda}_i^1\mathrm{d}t+\sum_{i=1}^{M-1}h_2^j(t_M-t_i)\hat{\lambda}_i^2\mathrm{d}t+(\sum_{i=1}^{M-1}k_1^j(t_M-t_i)\mathrm{d}P_i^1)^2+(\sum_{i=1}^{M-1}k_2^j(t_M-t_i)\mathrm{d}P_i^2)^2.$$

  In the case of exponential kernels, we write $h_i^j(t)=n_i^j\exp(-\kappa_i^jt)$ and $k_i^j(t)=n_{iq}^j\exp(-\kappa_{iq}^jt)$, and the intensity becomes:

  $$\begin{aligned}\lambda_{t_M}^j=&\lambda_\infty^j\\&+\sum_{i=1}^{M-1}n_1^j\exp(-\kappa_1^j(t_M-t_i))\hat{\lambda}_i^1\mathrm{d}t+\sum_{i=1}^{M-1}n_2^j\exp(-\kappa_2^j(t_M-t_i))\hat{\lambda}_i^2\mathrm{d}t\\&+(\sum_{i=1}^{M-1}n_{1q}^j\exp(-\kappa_{1q}^j(t_M-t_i))\mathrm{d}P_i^1)^2+(\sum_{i=1}^{M-1}n_{2q}^j\exp(-\kappa_{2q}^j(t_M-t_i))\mathrm{d}P_i^2)^2.\end{aligned}$$

  The parameters we want to estimate in this case are $(n_i^j)_{i,j\in[\![1,2]\!]}$, $(n_{i,q}^j)_{i,j\in[\![1,2]\!]}$, $(\kappa_i^j)_{i,j\in[\![1,2]\!]}$ and $(\kappa_{i,q}^j)_{i,j\in[\![1,2]\!]}$.

Minimize the log-likelihood

To minimise $-\ln\big(\mathbf{L}((\hat{\lambda}_i^1)_i,(\hat{\lambda}_i^2)_i,\theta)\big)$, we look at gradient descent methods. We need to calculate the gradient of $-\ln\big(\mathbf{L}((\hat{\lambda}_i^1)_i,(\hat{\lambda}_i^2)_i,\theta)\big)$ according to the parameters.

Then, with expression (B.3.2), we can write the gradient:

$$\begin{aligned}\nabla\Big(-\ln\big(\mathbf{L}((\hat{\lambda}_i^1)_i,(\hat{\lambda}_i^2)_i,\theta)\big)\Big)&=\sum_{i=1}^{N}-\nabla\Big(\hat{\lambda}_i^1\ln(\lambda_i^1(\theta))\Big)+\nabla\Big(\lambda_i^1(\theta)\Big)-\nabla\Big(\hat{\lambda}_i^2\ln(\lambda_i^2(\theta))\Big)+\nabla\Big(\lambda_i^2(\theta)\Big)\\&=\sum_{i=1}^{N}-\hat{\lambda}_i^1\nabla\Big(\ln(\lambda_i^1(\theta))\Big)+\nabla\Big(\lambda_i^1(\theta)\Big)-\hat{\lambda}_i^2\nabla\Big(\ln(\lambda_i^2(\theta))\Big)+\nabla\Big(\lambda_i^2(\theta)\Big)\\&=\sum_{i=1}^{N}-\hat{\lambda}_i^1\frac{\nabla\Big(\lambda_i^1(\theta)\Big)}{\lambda_i^1(\theta)}+\nabla\Big(\lambda_i^1(\theta)\Big)-\hat{\lambda}_i^2\frac{\nabla\Big(\lambda_i^2(\theta)\Big)}{\lambda_i^2(\theta)}+\nabla\Big(\lambda_i^2(\theta)\Big)\\&=\sum_{i=1}^{N}\nabla\lambda_i^1(\theta)\Big(-\frac{\hat{\lambda}_i^1}{\lambda_i^1(\theta)}+1\Big)+\nabla\lambda_i^2(\theta)\Big(-\frac{\hat{\lambda}_i^2}{\lambda_i^2(\theta)}+1\Big).\end{aligned}$$

In practice, for financial time series, $\hat{\lambda}\equiv\frac{\sigma^2}{\mathrm{d}t}$ where σ is the volatility, and $\mathrm{d}P$ are the returns.

B.4 "Surprise price"

We know, for Gaussian variables:

$$\mathbb{E}(X/Y) = \mathbb{E}(X) + \Sigma_{XY}\Sigma_Y^{-1}(Y - \mathbb{E}(Y))$$

So, if we consider returns as Gaussian, and we consider only the lag-1 has influence on the next return:

$$\mathbb{E}(\mathrm{d}P_t/\mathrm{d}P_{t-1}) = \mathbb{E}(\mathrm{d}P_t) + \Sigma_{t,t-1}\Sigma_{t-1}^{-1}(\mathrm{d}P_{t-1} - \mathbb{E}(\mathrm{d}P_{t-1}))$$

Since we centered/normalized the returns $\mathbb{E}(\mathrm{d}P_t) = \mathbb{E}(\mathrm{d}P_{t-1}) = 0$ and $\Sigma_{t-1} = 1$ and $\Sigma_{t,t-1} = \mathrm{cov}(\mathrm{d}P_t, \mathrm{d}P_{t-1}) = \Gamma(1)$. Now in multidimension:

$$\mathbb{E}\left(\begin{pmatrix}\mathrm{d}P_t^1\\ \mathrm{d}P_t^2\end{pmatrix} / \begin{pmatrix}\mathrm{d}P_{t-1}^1\\ \mathrm{d}P_{t-1}^2\end{pmatrix}\right) = \mathbb{E}\left(\begin{pmatrix}\mathrm{d}P_t^1\\ \mathrm{d}P_t^2\end{pmatrix}\right) + \Sigma_{t,t-1}\Sigma_{t-1}^{-1}\left[\begin{pmatrix}\mathrm{d}P_{t-1}^1\\ \mathrm{d}P_{t-1}^2\end{pmatrix} - \mathbb{E}\left(\begin{pmatrix}\mathrm{d}P_{t-1}^1\\ \mathrm{d}P_{t-1}^2\end{pmatrix}\right)\right]$$

Similarly, thanks to the normalization and centralisation we have:

$$\Sigma_{t-1} = \begin{pmatrix}1 & \mathrm{cov}(\mathrm{d}P^1, \mathrm{d}P^2)\\ \mathrm{cov}(\mathrm{d}P^1, \mathrm{d}P^2) & 1\end{pmatrix}$$

we note $\mathrm{cov}(dP^1, dP^2) = \nu$

$$\Sigma_{t,t-1} = \begin{pmatrix}\mathrm{cov}(\mathrm{d}P_t^1, \mathrm{d}P_{t-1}^1) & \mathrm{cov}(\mathrm{d}P_t^1, \mathrm{d}P_{t-1}^2)\\ \mathrm{cov}(\mathrm{d}P_{t-1}^1, \mathrm{d}P_t^2) & \mathrm{cov}(\mathrm{d}P_t^2, \mathrm{d}P_{t-1}^2)\end{pmatrix}$$

So we should have:

$$\begin{cases}\mathrm{d}\tilde{P}_t^1 = \mathrm{d}P_t^1 - \frac{1}{1-\nu^2}\Big(\big(\mathrm{cov}(\mathrm{d}P_t^1, \mathrm{d}P_{t-1}^1) - \nu\mathrm{cov}(\mathrm{d}P_t^1, \mathrm{d}P_{t-1}^2)\big)\mathrm{d}P_{t-1}^1 + \big(-\nu\mathrm{cov}(\mathrm{d}P_t^1, \mathrm{d}P_{t-1}^1) + \mathrm{cov}(\mathrm{d}P_t^1, \mathrm{d}P_{t-1}^2)\big)\mathrm{d}P_{t-1}^2\Big)\\ \mathrm{d}\tilde{P}_t^2 = \mathrm{d}P_t^2 - \frac{1}{1-\nu^2}\Big(\big(\mathrm{cov}(\mathrm{d}P_{t-1}^1, \mathrm{d}P_t^2) - \nu\mathrm{cov}(\mathrm{d}P_t^2, \mathrm{d}P_{t-1}^2)\big)\mathrm{d}P_{t-1}^1 + \big(-\nu\mathrm{cov}(\mathrm{d}P_{t-1}^1, \mathrm{d}P_t^2) + \mathrm{cov}(\mathrm{d}P_t^2, \mathrm{d}P_{t-1}^2)\big)\mathrm{d}P_{t-1}^2\Big)\end{cases}$$

B.5 Yule-Walker equations – MQGARCH framework

B.5.1 MQGARCH framework

This appendix details the calibration procedure for a Multivariate Quadratic GARCH (MQGARCH) process. Consistently with the notations therein, we consider a MQGARCH process with $N = 2$ assets, where the squared volatility of asset j, σ_j^2, is defined by

$$\sigma_{j,t}^2 = \sigma_{j,\infty}^2 + \sum_{i=1}^{N}\sum_{\tau=1}^{+\infty} L_i^j(\tau) r_{i,t-\tau} + \sum_{i,k=1}^{N}\sum_{\tau_1,\tau_2=1}^{+\infty} K_{ki}^j(\tau_1,\tau_2) r_{k,t-\tau_1} r_{i,t-\tau_2}. \tag{B.5.1}$$

To ease the development of the Yule-Walker system, we separate the linear, quadratic and cross kernels, then, σ_j^2 writes

$$\begin{aligned}
\sigma_{j,t}^2 =& \sigma_{j,\infty}^2 \\
&+ \sum_{i=1}^{N}\sum_{\tau=1}^{+\infty} L_i^j(\tau) r_{i,t-\tau} \\
&+ \sum_{i=1}^{N}\sum_{\tau=1}^{+\infty} \phi_i^j(\tau) r_{i,t-\tau}^2 \\
&+ 2\sum_{\tau_1=1}^{+\infty}\sum_{\tau_2=\tau_1+1}^{+\infty} K_i^j(\tau_1,\tau_2) r_{i,t-\tau_1} r_{i,t-\tau_2} \\
&+ \sum_{\tau=1}^{+\infty} \phi_\times^j(\tau)(r_{A,t-\tau} r_{B,t-\tau} - \overline{r_A r_B}) \\
&+ \sum_{\tau_1=1}^{+\infty}\sum_{\tau_2=\tau_1+1}^{+\infty} k_\times^j(\tau_1,\tau_2) r_{A,t-\tau_1} r_{B,t-\tau_2} \\
&+ \sum_{\tau_1=2}^{+\infty}\sum_{\tau_2=1}^{\tau_1-1} k_\times^j(\tau_1,\tau_2) r_{A,t-\tau_1} r_{B,t-\tau_2},
\end{aligned} \tag{B.5.2}$$

where A and B denote the first and second assets. In the following, we swap (A, B) and (1, 2) whichever makes the writing clearer.

Proof. The transformation of the quadratic term (for $k = i$ in Eq. (B.5.1)) goes as follows

$$\begin{aligned}
\sum_{\tau_1,\tau_2=1}^{+\infty} K_i^j(\tau_1,\tau_2) r_{i,t-\tau_1} r_{i,t-\tau_2} =& \sum_{\tau=1}^{+\infty} \phi_i^j(\tau) r_{i,t-\tau}^2 \\
&+ \sum_{\tau_1=2}^{+\infty}\sum_{\tau_2=1}^{\tau_1-1} K_i^j(\tau_1,\tau_2) r_{i,t-\tau_1} r_{i,t-\tau_2} + \sum_{\tau_1=1}^{+\infty}\sum_{\tau_2=\tau_1+1}^{+\infty} K_i^j(\tau_1,\tau_2) r_{i,t-\tau_1} r_{i,t-\tau_2} \\
=& \sum_{\tau=1}^{+\infty} \phi_i^j(\tau) r_{i,t-\tau}^2 + \sum_{\tau_1=2}^{+\infty}\sum_{\tau_2=1}^{\tau_1-1} K_i^j(\tau_1,\tau_2) r_{i,t-\tau_1} r_{i,t-\tau_2} + \sum_{\tau_2=2}^{+\infty}\sum_{\tau_1=1}^{\tau_2-1} K_i^j(\tau_1,\tau_2) r_{i,t-\tau_1} r_{i,t-\tau_2} \\
&\text{thanks to the properties of symmetry,} \\
=& \sum_{\tau=1}^{+\infty} \phi_i^j(\tau) r_{i,t-\tau}^2 + 2\sum_{\tau_1=2}^{+\infty}\sum_{\tau_2=1}^{\tau_1-1} K_i^j(\tau_1,\tau_2) r_{i,t-\tau_1} r_{i,t-\tau_2}
\end{aligned}$$

□

Additionally, if we consider the formulation of Equation (B.5.2), then the mean volatility meets the following condition

$$\overline{\boldsymbol{\sigma}^2} = \boldsymbol{\sigma}_\infty^2 + \left(\sum_{\tau=1}^{+\infty} \begin{pmatrix} \phi_1^1 & \phi_2^1 \\ \phi_1^2 & \phi_2^2 \end{pmatrix}(\tau)\right) \overline{\boldsymbol{r}^2} \iff \overline{\boldsymbol{\sigma}^2} = \boldsymbol{\sigma}_\infty^2 + \left(\sum_{\tau=1}^{+\infty} \begin{pmatrix} \phi_1^1 & \phi_2^1 \\ \phi_1^2 & \phi_2^2 \end{pmatrix}(\tau)\right) \overline{\boldsymbol{\sigma}^2}$$

B.5.2 Yule-Walker equations for a 4-step calibration

For more stability, we undertake a calibration of the model Eq. (B.5.2) in 4 steps.

During the first three steps, we consider the leverage kernel to be null and aim to determine the other kernels ($(\phi_i^j)_{i,j\in\{A,B\}}$, $(K_i^j)_{i,j\in\{A,B\}}$, $(\phi_\times^j)_{j\in\{A,B\}}$ and $(k_\times^j)_{j\in\{A,B\}}$). Therefore, to exclude the leverage effect from real data, we calibrate the linear, quadratic and cross kernels on "symmetrised" data, meaning that we use the datasets $[\sigma,\sigma]$ and $[r,-r]$ where σ and r represent our original datasets. Once step 3 is completed, we use the original datasets, in which the leverage effect is present, to determine the leverage kernels. The calibration is thus decomposed as follows:

1. Determine the QGARCH 1D of the 2 processes ($(\phi_i^i)_{i\{A,B\}}$ and $(K_i^i)_{i\{A,B\}}$)
2. Determine the QGARCH 2D cross linear and quadratic kernels ($(\phi_j^i)_{i,j\{A,B\},i\neq j}$ and $(K_j^i)_{i,j\{A,B\},i\neq j}$)
3. Determine the cross kernels of the QGARCH 2D ($(\phi_\times^i)_{i\{A,B\}}$ and $(k_\times^i)_{i\{A,B\}}$)
4. Determine the QGARCH 2D Leverage kernels ($(L_i^i)_{i\{A,B\}}$.

In order to stay agnostic on the shape of the kernels, for each step, we implement a general method of moments. We rely on a system of Yule-Walker type equations which we describe subsequently.

Covariance structure

The Yule-Walker system of equations and thus, the method of moments, rely on the following covariance structures:

$$\mathcal{C}_{ij}(\tau) = \mathbb{E}(\sigma_{i,t}^2 r_{j,t-\tau}^2) - \overline{\sigma_i^2}\,\overline{r_j^2} \tag{B.5.3}$$

$$\mathcal{C}_{ij}^r(\tau) = \mathbb{E}(r_{i,t}^2 r_{j,s}^2) - \overline{r_i^2}\,\overline{r_j^2} \tag{B.5.4}$$

$$\mathcal{D}_{ijk}(\tau_1,\tau_2) = \mathbb{E}(\sigma_{i,t}^2 r_{j,t-\tau_1} r_{k,t-\tau_2}) \tag{B.5.5}$$

$$\mathcal{D}_{p,(ij)kl}(\tau) = \mathbb{E}((r_{i,t}r_{j,t})r_{k,t-\tau_1} r_{l,t-\tau_2}) \tag{B.5.6}$$

$$\mathcal{D}_{\times,jAB}(\tau_1,\tau_2) = \mathbb{E}\left(\sigma_{j,t}^2 r_{A,t-\tau_1} r_{B,t-\tau_2}\right) \tag{B.5.7}$$

$$\mathcal{V}_{ij}(\tau) = \mathbb{E}(\sigma_{i,t}^2 r_{j,t-\tau}) \tag{B.5.8}$$

$$\mathcal{V}_{ij}^r(\tau) = \mathbb{E}(r_{i,t}^2 r_{j,t-\tau}) \tag{B.5.9}$$

Remark B.5.1. By definition of the model and in simulation, Eq. (B.5.3) and Eq. (B.5.4) should result in the same values. However in practice, there is a slight difference due to noise, which influence the calibration. Hence, in order to give reproducible guidelines we note the difference in what follows.

Yule-Walker system for QGARCH 2D without cross terms (steps 1 and 2)

For the first two steps of the calibration, we need equations linking $\mathcal{C}$ and $\mathcal{D}$ to the kernels ϕ and K. The goal is to characterise $(\phi_i^j)_{i,j\in\{A,B\}}$ and $(K_i^j)_{i,j\in\{A,B\}}$ up to a certain lag q. Hence, we want to find the values of, for $i,j\in\{A,B\}$, $\phi_i^j(\tau)_{\tau\in[\![1,q]\!]}$ and $K_i^j(\tau_1,\tau_2)_{1\leq\tau_1<\tau_2\leq q}$ (since K is symmetric, it is enough to find only its upper triangle entries). The Yule Walker equations on $(\mathcal{C}_{ij}(\tau))_{i,j\in\{A,B\},\tau\in[\![1,q]\!]}$ will give $4\times q$ equations and the Yule Walker equations on $(\mathcal{D}_{ij}(\tau_1,\tau_2))_{i,j\in\{A,B\},1\leq\tau_1<\tau_2\leq}$ will give $4\times q\frac{q-1}{2}$ equations, allowing to fully characterise $(\phi_i^j)_{i,j\in\{A,B\}}$ and $(K_i^j)_{i,j\in\{A,B\}}$ up to a certain lag q.

Without complicating things, we develop the general equations for N assets.

Lemma B.5.2 (Yule-Walker equation for $\mathcal{C}$)**.** *for $\tau > 0$, in the framework described by Eq.* (B.5.2)*, with $\phi^\times = k^\times = 0$, the covariance $\mathcal{C}$ writes*

$$\mathcal{C}_{jl}(\tau) = \sum_{i=1}^{N}\sum_{k=1}^{+\infty}\phi_i^j(k)\mathcal{C}_{il}^r(\tau-k) + 2\sum_{i=1}^{N}\sum_{k_2=\tau+1}^{+\infty}\sum_{k_1=k_2+1}^{+\infty}K_i^j(k_1,k_2)\mathcal{D}_{p,lii}(k_1-\tau,k_2-\tau)$$

Proof.

$$\begin{aligned}
\mathcal{C}_{jl}(t-s) =& \mathbb{E}(\sigma_{j,t}^2 r_{l,s}^2) - \overline{\sigma_j^2}\,\overline{r_l^2}\\
=& \sigma_{j,\infty}^2\overline{r_l^2} + \sum_{i=1}^{N}\sum_{k=1}^{+\infty}\phi_i^j(k)\mathbb{E}(r_{i,t-k}^2 r_{l,s}^2) + 2\sum_{i=1}^{N}\sum_{k_1=2}^{+\infty}\sum_{k_2=1}^{k_1-1}K_i^j(k_1,k_2)\mathbb{E}(r_{i,t-k_1}r_{i,t-k_2}r_{l,s}^2) - \overline{\sigma_j^2}\,\overline{r_l^2}\\
=& \sigma_{j,\infty}^2\overline{r_l^2} + \sum_{i=1}^{N}\phi_i^j(t-s)\overline{r_i^2 r_l^2} + \sum_{i=1}^{N}\sum_{k=1,k\neq t-s}^{+\infty}\phi_i^j(k)(\mathcal{C}_{il}^r(t-k-s) + \overline{r_i^2}\,\overline{r_l^2})\\
&+ 2\sum_{i=1}^{N}\sum_{k_1=2}^{+\infty}\sum_{k_2=1}^{k_1-1}K_i^j(k_1,k_2)\mathbb{E}(r_{i,t-k_1}r_{i,t-k_2}r_{l,s}^2) - \overline{\sigma_j^2}\,\overline{r_l^2}\\
=& \sigma_{j,\infty}^2\overline{r_l^2} + \sum_{i=1}^{N}\phi_i^j(t-s)\overline{r_i^2 r_l^2}\\
&+ \sum_{i=1}^{N}\sum_{k=1,k\neq t-s}^{+\infty}\phi_i^j(k)\mathcal{C}_{il}^r(t-k-s) + \sum_{i=1}^{N}\sum_{k=1,k\neq t-s}^{+\infty}\phi_i^j(k)\overline{r_i^2}\,\overline{r_l^2}\\
&+ 2\sum_{i=1}^{N}\sum_{k_1=2}^{+\infty}\sum_{k_2=1}^{k_1-1}K_i^j(k_1,k_2)\mathbb{E}(r_{i,t-k_1}r_{i,t-k_2}r_{l,s}^2) - \overline{\sigma_j^2}\,\overline{r_l^2}\\
=& \sigma_{j,\infty}^2\overline{r_l^2} + \sum_{i=1}^{N}\phi_i^j(t-s)\overline{r_i^2 r_l^2}\\
&+ \sum_{i=1}^{N}\sum_{k=1,k\neq t-s}^{+\infty}\phi_i^j(k)\mathcal{C}_{il}^r(t-k-s) + \sum_{i=1}^{N}\sum_{k=1}^{+\infty}\phi_i^j(k)\overline{r_i^2}\,\overline{r_l^2} - \sum_{i=1}^{N}\phi_i^j(t-s)\overline{r_i^2}\,\overline{r_l^2}\\
&+ 2\sum_{i=1}^{N}\sum_{k_1=2}^{+\infty}\sum_{k_2=1}^{k_1-1}K_i^j(k_1,k_2)\mathbb{E}(r_{i,t-k_1}r_{i,t-k_2}r_{l,s}^2) - \overline{\sigma_j^2}\,\overline{r_l^2}
\end{aligned}$$

by definition of the mean squared volatility,

$$\begin{aligned}
=& \sum_{i=1}^{N}\sum_{k=1}^{+\infty}\phi_i^j(k)\mathcal{C}_{il}^r(t-k-s) + 2\sum_{i=1}^{N}\sum_{k_2=1}^{+\infty}\sum_{k_1=k_2+1}^{+\infty}K_i^j(k_1,k_2)\mathbb{E}(r_{i,t-k_1}r_{i,t-k_2}r_{l,s}^2)\\
=& \sum_{i=1}^{N}\sum_{k=1}^{+\infty}\phi_i^j(k)\mathcal{C}_{il}^r(t-k-s) + 2\sum_{i=1}^{N}\sum_{k_2=t-s+1}^{+\infty}\sum_{k_1=k_2+1}^{+\infty}K_i^j(k_1,k_2)\mathcal{D}_{p,lii}(s-t+k_1,s-t+k_2)
\end{aligned}$$

we set $s=0$, $t=\tau$, then

$$\mathcal{C}_{jl}(\tau) = \sum_{i=1}^{N}\sum_{k=1}^{+\infty}\phi_i^j(k)\mathcal{C}_{il}^r(\tau-k) + 2\sum_{i=1}^{N}\sum_{k_2=\tau+1}^{+\infty}\sum_{k_1=k_2+1}^{+\infty}K_i^j(k_1,k_2)\mathcal{D}_{p,lii}(k_1-\tau,k_2-\tau)$$

which concludes the proof.

□

Lemma B.5.3 (Yule-Walker equation for $\mathcal{D}$)**.** *For $\tau_1 < \tau_2$, in the framework described by Eq.* (B.5.2)*, with $\phi^\times = k^\times = 0$, the covariance $\mathcal{D}$ writes*

$$\mathcal{D}_{jll}(\tau_1,\tau_2) = \sum_{i=1}^{N}\sum_{k=1}^{\tau_1-1}\phi_i^j(k)\mathcal{D}_{p,ill}(\tau_1-k,\tau_2-k) + 2\sum_{i=1}^{N}\sum_{k_2=\tau_1+1}^{+\infty}K_i^j(\tau_1,k_2)\mathcal{D}_{p,(il)il}(-\tau_1+k_2,\tau_2-\tau_1)$$

Proof. Considering $s < u < t$, then $t-u < t-s$ and we have

$$\begin{aligned}
\mathcal{D}_{jll}(t-u,t-s) =& \mathbb{E}(\sigma^2_{j,t} r_{l,u} r_{l,s}) \\
=& \sum_{i=1}^{N}\sum_{k=1}^{t-u-1} \phi_i^j(k)\mathbb{E}(r^2_{i,t-k} r_{l,u} r_{l,s}) + 2\sum_{i=1}^{N}\sum_{k_1=2}^{+\infty}\sum_{k_2=1}^{k_1-1} K_i^j(k_1,k_2)\mathbb{E}(r_{i,t-k_1} r_{i,t-k_2} r_{l,u} r_{l,s}) \\
=& \sum_{i=1}^{N}\sum_{k=1}^{t-u-1} \phi_i^j(k)\mathcal{D}_{p,ill}(t-k-u,t-k-s) + 2\sum_{i=1}^{N}\sum_{k_1=2}^{+\infty}\sum_{k_2=1}^{k_1-1} K_i^j(k_1,k_2)\mathbb{E}(r_{i,t-k_1} r_{i,t-k_2} r_{l,u} r_{l,s}) \\
=& \sum_{i=1}^{N}\sum_{k=1}^{t-u-1} \phi_i^j(k)\mathcal{D}_{p,ill}(t-k-u,t-k-s) + 2\sum_{i=1}^{N}\sum_{k_1=1}^{+\infty}\sum_{k_2=k_1+1}^{+\infty} K_i^j(k_1,k_2)\mathbb{E}(r_{i,t-k_1} r_{i,t-k_2} r_{l,u} r_{l,s}) \\
& \text{here, } t-k_2 < t-k_1 \text{ and } s<u \text{ we need } t-k_1 = u \\
=& \sum_{i=1}^{N}\sum_{k=1}^{t-u-1} \phi_i^j(k)\mathcal{D}_{p,ill}(t-k-u,t-k-s) + 2\sum_{i=1}^{N}\sum_{k_2=t-u+1}^{+\infty} K_i^j(t-u,k_2)\mathbb{E}\left((r_{i,u} r_{l,u})\, r_{i,t-k_2} r_{l,s}\right) \\
=& \sum_{i=1}^{N}\sum_{k=1}^{t-u-1} \phi_i^j(k)\mathcal{D}_{p,ill}(t-k-u,t-k-s) \\
& + 2\sum_{i=1}^{N}\sum_{k_2=t-u+1}^{+\infty} K_i^j(t-u,k_2)\mathcal{D}_{p,(il)il}(u-t+k_2,u-s)
\end{aligned}$$

we set $s=0$, $t=\tau_2$, and $\tau_1 = t-u$, then

$$\mathcal{D}_{jll}(\tau_1,\tau_2) = \sum_{i=1}^{N}\sum_{k=1}^{\tau_1-1} \phi_i^j(k)\mathcal{D}_{p,ill}(\tau_1-k,\tau_2-k) + 2\sum_{i=1}^{N}\sum_{k_2=\tau_1+1}^{+\infty} K_i^j(\tau_1,k_2)\mathcal{D}_{p,(il)il}(-\tau_1+k_2,\tau_2-\tau_1),$$

which ends the proof.

□

Yule-Walker system for QGARCH 2D with cross terms (step 3)

The goal of the third step is to characterise $(\phi^j_\times)_{j\in\{A,B\}}$ and $(K^j_\times)_{j\in\{A,B\}}$ up to a certain lag q. In this case, we need $2\times q$ equations to characterise $(\phi^j_\times)_{j\in\{A,B\}}$ and $2\times q(q-1)$ equations to characterise $(K^j_\times)_{j\in\{A,B\}}$, as $K^j_\times$ is not symmetric so we need both the upper and lower triangle entries. The equations on $(\mathcal{D}^j_\times(\tau_1,\tau_2))_{j\in\{A,B\},1\le\tau_1,\tau_2\le q}$, developed subsequently, allow to fully characterise the cross kernels $(\phi^j_\times)_{j\in\{A,B\}}$ and $(K^j_\times)_{j\in\{A,B\}}$.

Lemma B.5.4 (Yule-Walker equation for $\mathcal{D}_\times$ for $\tau_1=\tau_2=\tau$)**.** *for $\tau>0$, in the framework described by Eq.* (B.5.2)*, the covariance $\mathcal{D}_\times$ writes*

$$\begin{aligned}
\mathcal{D}_{\times,jAB}(\tau,\tau) =& \sigma^2_{j,\infty}\overline{r_A r_B} \\
& + \sum_{i=1}^{N}\sum_{k=1}^{+\infty} \phi_i^j(k)\mathcal{D}_{p,(ii)AB}(\tau-k,\tau-k) \\
& + 2\sum_{i=1}^{N}\sum_{k_1=\tau+1}^{+\infty}\sum_{k_2=k_1+1}^{+\infty} K_i^j(k_1,k_2)\mathcal{D}_{p,(AB)ii}(k_1-\tau,k_2-\tau) \\
& + \sum_{k=1}^{+\infty} \phi^j_\times(k)\left(\mathcal{D}_{p(AB)AB}(\tau-k,\tau-k) - \overline{r_A r_B}^2\right) \\
& + \sum_{k_2=\tau+1}^{+\infty}\sum_{k_1=k_2+1}^{+\infty} k^j_\times(k_1,k_2)\mathcal{D}_{p,(AB)AB}(k_1-\tau,k_2-\tau) \\
& + \sum_{k_1=\tau+1}^{+\infty}\sum_{k_2=k_1+1}^{+\infty} k^j_\times(k_1,k_2)\mathcal{D}_{p,(AB)AB}(k_1-\tau,k_2-\tau)
\end{aligned}$$

Proof. We take $s < t$, then

$$\begin{aligned}
\mathcal{D}_{\times,jAB}(t-s,t-s) =& \mathbb{E}\left(\sigma_{j,t}^2 r_{A,s} r_{B,s}\right) \\
=& \sigma_{j,\infty}^2 \mathbb{E}\left(r_{A,s} r_{B,s}\right) \\
&+ \sum_{i=1}^{N}\sum_{\tau=1}^{+\infty} \phi_i^j(\tau)\mathbb{E}\left(r_{i,t-\tau}^2 r_{A,s} r_{B,s}\right) \\
&+ 2\sum_{i=1}^{N}\sum_{\tau_1=1}^{+\infty}\sum_{\tau_2=\tau_1+1}^{+\infty} K_i^j(\tau_1,\tau_2)\mathbb{E}\left(r_{i,t-\tau_1} r_{i,t-\tau_2} r_{A,s} r_{B,s}\right) \\
&+ \sum_{\tau=1}^{+\infty} \phi_\times^j(\tau)\mathbb{E}\left((r_{A,t-\tau} r_{B,t-\tau} - \overline{r_A r_B}) r_{A,s} r_{B,s}\right) \\
&+ \sum_{\tau_1=2}^{+\infty}\sum_{\tau_2=1}^{\tau_1-1} k_\times^j(\tau_1,\tau_2)\mathbb{E}\left(r_{A,t-\tau_1} r_{B,t-\tau_2} r_{A,s} r_{B,s}\right) \\
&+ \sum_{\tau_1=1}^{+\infty}\sum_{\tau_2=\tau_1+1}^{+\infty} k_\times^j(\tau_1,\tau_2)\mathbb{E}\left(r_{A,t-\tau_1} r_{B,t-\tau_2} r_{A,s} r_{B,s}\right) \\
=& \sigma_{j,\infty}^2 \mathbb{E}\left(r_{A,s} r_{B,s}\right) \\
&+ \sum_{i=1}^{N}\sum_{k=1}^{+\infty} \phi_i^j(k)\mathcal{D}_{p,(ii)AB}(t-k-s,t-k-s) \\
&+ 2\sum_{i=1}^{N}\sum_{k_1=t-s+1}^{+\infty}\sum_{k_2=k_1+1}^{+\infty} K_i^j(k_1,k_2)\mathcal{D}_{p,(AB)ii}(s-t+k_1,s-t+k_2) \\
&+ \sum_{k=1}^{+\infty} \phi_\times^j(k)\left(\mathcal{D}_{p(AB)AB}(t-s-k,t-s-k) - \overline{r_A r_B}^2\right) \\
&+ \sum_{k_2=t-s+1}^{+\infty}\sum_{k_1=k_2+1}^{+\infty} k_\times^j(k_1,k_2)\mathcal{D}_{p,(AB)AB}(s-t+k_1,s-t+k_2) \\
&+ \sum_{k_1=t-s+1}^{+\infty}\sum_{k_2=k_1+1}^{+\infty} k_\times^j(k_1,k_2)\mathcal{D}_{p,(AB)AB}(s-t+k_1,s-t+k_2).
\end{aligned}$$

we set $t = \tau$ and $s = 0$,

$$\begin{aligned}
\mathcal{D}_{\times,jjl}(\tau,\tau) =& \sigma_{j,\infty}^2 \overline{r_A r_B} \\
&+ \sum_{i=1}^{N}\sum_{k=1}^{+\infty} \phi_i^j(k)\mathcal{D}_{p,(ii)AB}(\tau-k,\tau-k) \\
&+ 2\sum_{i=1}^{N}\sum_{k_1=\tau+1}^{+\infty}\sum_{k_2=k_1+1}^{+\infty} K_i^j(k_1,k_2)\mathcal{D}_{p,(AB)ii}(k_1-\tau,k_2-\tau) \\
&+ \sum_{k=1}^{+\infty} \phi_\times^j(k)\left(\mathcal{D}_{p(AB)AB}(\tau-k,\tau-k) - \overline{r_A r_B}^2\right) \\
&+ \sum_{k_2=\tau+1}^{+\infty}\sum_{k_1=k_2+1}^{+\infty} k_\times^j(k_1,k_2)\mathcal{D}_{p,(AB)AB}(k_1-\tau,k_2-\tau) \\
&+ \sum_{k_1=\tau+1}^{+\infty}\sum_{k_2=k_1+1}^{+\infty} k_\times^j(k_1,k_2)\mathcal{D}_{p,(AB)AB}(k_1-\tau,k_2-\tau),
\end{aligned}$$

which concludes the proof. □

One of the difficulty in including the cross terms is the non-symmetry of the covariance structure $\mathcal{D}_\times$ and of the kernels $k_\times$. Hence, we need to develop equations for $\mathcal{D}_{\times,jAB}(\tau_1,\tau_2)$ for $\tau_1 < \tau_2$ and $\tau_1 > \tau_2$.

Lemma B.5.5 (Yule-Walker equation for $\mathcal{D}_\times$ for $\tau_1 < \tau_2$)**.** *for $\tau_1 < \tau_2$, in the framework described by Eq.* (B.5.2)*, the covariance $\mathcal{D}_\times$ writes*

$$\begin{aligned}
\mathcal{D}_{\times,jAB}(\tau_1,\tau_2) =&\sigma^2_{j,\infty}\overline{r_A r_B}\\
&+\sum_{i=1}^{N}\sum_{k=1}^{\tau_1-1}\phi_i^j(k)\mathcal{D}_{p,(ii)AB}(\tau_1-k,\tau_2-k)\\
&+2\sum_{i=1}^{N}\sum_{k_2=\tau_1+1}^{+\infty}K_i^j(\tau_1,k_2)\mathcal{D}_{p,(iA)iB}(k_2-\tau_1,\tau_2-\tau_1)\\
&+\sum_{k=1}^{\tau_1-1}\phi_\times^j(k)\left(\mathcal{D}_{(AB)AB}(\tau_1-k,\tau_2-k)-\overline{r_A r_B}^2\right)\\
&+\sum_{k_1=\tau_1+1}^{+\infty}k_\times^j(k_1,\tau_1)\mathcal{D}_{p,(BA)AB}(k_1-\tau_1,\tau_2-\tau_1)\\
&+\sum_{k_2=\tau_1+1}^{+\infty}k_\times^j(\tau_1,k_2)\mathcal{D}_{p,(AA)BB}(k_2-\tau_1,\tau_2-\tau_1)
\end{aligned}$$

Proof. We consider $u < s$, then $t-s < t-u$, and we have

$$\begin{aligned}
\mathcal{D}_{\times,jAB}(t-s,t-u) =&\mathbb{E}\left(\sigma^2_{j,t}r_{A,s}r_{B,u}\right)\\
=&\sigma^2_{j,\infty}\mathbb{E}\left(r_{A,s}r_{B,u}\right)\\
&+\sum_{i=1}^{N}\sum_{k=1}^{+\infty}\phi_i^j(k)\mathbb{E}\left(r^2_{i,t-\tau}r_{A,s}r_{B,u}\right)\\
&+2\sum_{i=1}^{N}\sum_{k_1=1}^{+\infty}\sum_{k_2=k_1+1}^{+\infty}K_i^j(k_1,k_2)\mathbb{E}\left(r_{i,t-k_1}r_{i,t-k_2}r_{A,s}r_{B,u}\right)\\
&+\sum_{k=1}^{+\infty}\phi_\times^j(k)\mathbb{E}\left((r_{A,t-k}r_{B,t-k}-\overline{r_A r_B})r_{A,s}r_{B,u}\right)\\
&+\sum_{k_1=2}^{+\infty}\sum_{k_2=1}^{k_1-1}k_\times^j(k_1,k_2)\mathbb{E}\left(r_{A,t-k_1}r_{B,t-k_2}r_{A,s}r_{B,u}\right)\\
&+\sum_{k_1=1}^{+\infty}\sum_{k_2=k_1+1}^{+\infty}k_\times^j(k_1,k_2)\mathbb{E}\left(r_{A,t-k_1}r_{B,t-k_2}r_{A,s}r_{B,u}\right)\\
=&\sigma^2_{j,\infty}\overline{r_A r_B}\\
&+\sum_{i=1}^{N}\sum_{k=1}^{t-s-1}\phi_i^j(k)\mathcal{D}_{p,(ii)AB}(t-k-s,t-k-u)\\
&+2\sum_{i=1}^{N}\sum_{k_1=1}^{+\infty}\sum_{k_2=k_1+1}^{+\infty}K_i^j(k_1,k_2)\mathbb{E}\left(r_{i,t-k_1}r_{i,t-k_2}r_{A,s}r_{B,u}\right)\\
&+\sum_{k=1}^{t-s-1}\phi_\times^j(k)\left(\mathcal{D}_{(AB)AB}(t-k-s,t-k-u)-\overline{r_A r_B}^2\right)\\
&+\sum_{k_1=2}^{+\infty}\sum_{k_2=1}^{k_1-1}k_\times^j(k_1,k_2)\mathbb{E}\left(r_{A,t-k_1}r_{B,t-k_2}r_{A,s}r_{B,u}\right)\\
&+\sum_{k_1=1}^{+\infty}\sum_{k_2=k_1+1}^{+\infty}k_\times^j(k_1,k_2)\mathbb{E}\left(r_{A,t-k_1}r_{B,t-k_2}r_{A,s}r_{B,u}\right)\\
&\text{for the first quadratic term we need } t-k_1=s\\
=&\sigma^2_{j,\infty}\overline{r_A r_B}
\end{aligned}$$

$$
\begin{aligned}
&+\sum_{i=1}^{N}\sum_{k=1}^{t-s-1}\phi_i^j(k)\mathcal{D}_{p,(ii)AB}(t-k-s,t-k-u)\\
&+2\sum_{i=1}^{N}\sum_{k_2=t-s+1}^{+\infty}K_i^j(t-s,k_2)\mathcal{D}_{p,(iA)iB}(s-t+k_2,s-u)\\
&+\sum_{k=1}^{t-s-1}\phi_\times^j(k)\left(\mathcal{D}_{(AB)AB}(t-k-s,t-k-u)-\overline{r_Ar_B}^2\right)\\
&+\sum_{k_1=2}^{+\infty}\sum_{k_2=1}^{k_1-1}k_\times^j(k_1,k_2)\mathbb{E}\left(r_{A,t-k_1}r_{B,t-k_2}r_{A,s}r_{B,u}\right)\\
&+\sum_{k_1=1}^{+\infty}\sum_{k_2=k_1+1}^{+\infty}k_\times^j(k_1,k_2)\mathbb{E}\left(r_{A,t-k_1}r_{B,t-k_2}r_{A,s}r_{B,u}\right)
\end{aligned}
$$

for the last term we also have $t-k_1=s$

$$
\begin{aligned}
=&\sigma_{j,\infty}^2\overline{r_Ar_B}\\
&+\sum_{i=1}^{N}\sum_{k=1}^{t-s-1}\phi_i^j(k)\mathcal{D}_{p,(ii)AB}(t-k-s,t-k-u)\\
&+2\sum_{i=1}^{N}\sum_{k_2=t-s+1}^{+\infty}K_i^j(t-s,k_2)\mathcal{D}_{p,(iA)iB}(s-t+k_2,s-u)\\
&+\sum_{k=1}^{t-s-1}\phi_\times^j(k)\left(\mathcal{D}_{(AB)AB}(t-k-s,t-k-u)-\overline{r_Ar_B}^2\right)\\
&+\sum_{k_1=2}^{+\infty}\sum_{k_2=1}^{k_1-1}k_\times^j(k_1,k_2)\mathbb{E}\left(r_{A,t-k_1}r_{B,t-k_2}r_{A,s}r_{B,u}\right)\\
&+\sum_{k_2=t-s+1}^{+\infty}k_\times^j(t-s,k_2)\mathcal{D}_{p,(AA)BB}(s-t+k_2,s-u)
\end{aligned}
$$

we need, in the last term to work on, $t-k_2=s$

$$
\begin{aligned}
=&\sigma_{j,\infty}^2\overline{r_Ar_B}\\
&+\sum_{i=1}^{N}\sum_{k=1}^{t-s-1}\phi_i^j(k)\mathcal{D}_{p,(ii)AB}(t-k-s,t-k-u)\\
&+2\sum_{i=1}^{N}\sum_{k_2=t-s+1}^{+\infty}K_i^j(t-s,k_2)\mathcal{D}_{p,(iA)iB}(s-t+k_2,s-u)\\
&+\sum_{k=1}^{t-s-1}\phi_\times^j(k)\left(\mathcal{D}_{(AB)AB}(t-k-s,t-k-u)-\overline{r_Ar_B}^2\right)\\
&+\sum_{k_1=t-s+1}^{+\infty}k_\times^j(k_1,t-s)\mathcal{D}_{p,(BA)AB}(s-t+k_1,s-u)\\
&+\sum_{k_2=t-s+1}^{+\infty}k_\times^j(t-s,k_2)\mathcal{D}_{p,(AA)BB}(s-t+k_2,s-u).
\end{aligned}
$$

We now set $t=\tau_2$, $t-s=\tau_1$ and $u=0$, then we have

$$
\begin{aligned}
\mathcal{D}_{\times,jAB}(\tau_1,\tau_2)=&\sigma_{j,\infty}^2\overline{r_Ar_B}\\
&+\sum_{i=1}^{N}\sum_{k=1}^{\tau_1-1}\phi_i^j(k)\mathcal{D}_{p,(ii)AB}(\tau_1-k,\tau_2-k)
\end{aligned}
$$

$$
\begin{aligned}
&+2\sum_{i=1}^{N}\sum_{k_2=\tau_1+1}^{+\infty} K_i^j(\tau_1,k_2)\mathcal{D}_{p,(iA)iB}(k_2-\tau_1,\tau_2-\tau_1)\\
&+\sum_{k=1}^{\tau_1-1}\phi_\times^j(k)\left(\mathcal{D}_{(AB)AB}(\tau_1-k,\tau_2-k)-\overline{r_Ar_B}^2\right)\\
&+\sum_{k_1=\tau_1+1}^{+\infty} k_\times^j(k_1,\tau_1)\mathcal{D}_{p,(BA)AB}(k_1-\tau_1,\tau_2-\tau_1)\\
&+\sum_{k_2=\tau_1+1}^{+\infty} k_\times^j(\tau_1,k_2)\mathcal{D}_{p,(AA)BB}(k_2-\tau_1,\tau_2-\tau_1),
\end{aligned}
$$

which concludes the proof.

□

Lemma B.5.6 (Yule-Walker equation for $\mathcal{D}_\times$ for $\tau_1>\tau_2$)**.** *for $\tau_1>\tau_2$, in the framework described by Eq.* (B.5.2)*, the covariance $\mathcal{D}_\times$ writes*

$$
\begin{aligned}
\mathcal{D}_{\times,jAB}(\tau_1,\tau_2) =&\sigma_{j,\infty}^2\overline{r_Ar_B}\\
&+\sum_{i=1}^{N}\sum_{k=1}^{\tau_1-1}\phi_i^j(k)\mathcal{D}_{p,(ii)AB}(\tau_1-k,\tau_2-k)\\
&+2\sum_{i=1}^{N}\sum_{k_2=\tau_1+1}^{+\infty} K_i^j(\tau_1,k_2)\mathcal{D}_{p,(iB)iA}(k_2-\tau_1,\tau_1-\tau_2)\\
&+\sum_{k=1}^{\tau_1-1}\phi_\times^j(k)\left(\mathcal{D}_{(AB)AB}(\tau_1-k,\tau_2-k)-\overline{r_Ar_B}^2\right))\\
&+\sum_{k_1=\tau_2+1}^{+\infty} k_\times^j(k_1,\tau_2)\mathcal{D}_{p,(BB)AA}(k_1-\tau_2,\tau_1-\tau_2)\\
&+\sum_{k_2=\tau_1+1}^{+\infty} k_\times^j(\tau_1,k_2)\mathcal{D}_{p,(AB)BA}(k_2-\tau_1,\tau_1-\tau_2)
\end{aligned}
$$

Proof. We take $s<u<t$, then $t-u<t-s$ and we have

$$
\begin{aligned}
\mathcal{D}_{\times,jAB}(t-s,t-u) =&\mathbb{E}\left(\sigma_{j,t}^2 r_{A,s}r_{B,u}\right)\\
=&\sigma_{j,\infty}^2\overline{r_Ar_B}\\
&+\sum_{i=1}^{N}\sum_{k=1}^{t-u-1}\phi_i^j(k)\mathcal{D}_{p,(ii)AB}(t-k-s,t-k-u)\\
&+2\sum_{i=1}^{N}\sum_{k_1=1}^{+\infty}\sum_{k_2=k_1+1}^{+\infty} K_i^j(k_1,k_2)\mathbb{E}\left(r_{i,t-k_1}r_{i,t-k_2}r_{A,s}r_{B,u}\right)\\
&+\sum_{k=1}^{t-u-1}\phi_\times^j(k)\left(\mathcal{D}_{(AB)AB}(t-k-s,t-k-u)-\overline{r_Ar_B}^2\right))\\
&+\sum_{k_1=2}^{+\infty}\sum_{k_2=1}^{k_1-1} k_\times^j(k_1,k_2)\mathbb{E}\left(r_{A,t-k_1}r_{B,t-k_2}r_{A,s}r_{B,u}\right)\\
&+\sum_{k_1=1}^{+\infty}\sum_{k_2=k_1+1}^{+\infty} k_\times^j(k_1,k_2)\mathbb{E}\left(r_{A,t-k_1}r_{B,t-k_2}r_{A,s}r_{B,u}\right)\\
&\text{for the first quadratic term we need } t-k_1=u\\
=&\sigma_{j,\infty}^2\overline{r_Ar_B}\\
&+\sum_{i=1}^{N}\sum_{k=1}^{t-u-1}\phi_i^j(k)\mathcal{D}_{p,(ii)AB}(t-k-s,t-k-u)
\end{aligned}
$$

$$
\begin{aligned}
&+2\sum_{i=1}^{N}\sum_{k_2=t-u+1}^{+\infty}K_i^j(t-u,k_2)\mathcal{D}_{p,(iB)iA}(u-t+k_2,u-s)\\
&+\sum_{k=1}^{t-s-1}\phi_\times^j(k)\left(\mathcal{D}_{(AB)AB}(t-k-s,t-k-u)-\overline{r_Ar_B}^2\right)\Big)\\
&+\sum_{k_1=2}^{+\infty}\sum_{k_2=1}^{k_1-1}k_\times^j(k_1,k_2)\mathbb{E}\left(r_{A,t-k_1}r_{B,t-k_2}r_{A,s}r_{B,u}\right)\\
&+\sum_{k_1=1}^{+\infty}\sum_{k_2=k_1+1}^{+\infty}k_\times^j(k_1,k_2)\mathbb{E}\left(r_{A,t-k_1}r_{B,t-k_2}r_{A,s}r_{B,u}\right)\\
&\text{for the last term we also have } t-k_1=u\\
=&\sigma_{j,\infty}^2\overline{r_Ar_B}\\
&+\sum_{i=1}^{N}\sum_{k=1}^{t-s-1}\phi_i^j(k)\mathcal{D}_{p,(ii)AB}(t-k-s,t-k-u)\\
&+2\sum_{i=1}^{N}\sum_{k_2=t-s+1}^{+\infty}K_i^j(t-s,k_2)\mathcal{D}_{p,(iB)iA}(s-t+k_2,u-s)\\
&+\sum_{k=1}^{t-s-1}\phi_\times^j(k)\left(\mathcal{D}_{(AB)AB}(t-k-s,t-k-u)-\overline{r_Ar_B}^2\right)\Big)\\
&+\sum_{k_1=2}^{+\infty}\sum_{k_2=1}^{k_1-1}k_\times^j(k_1,k_2)\mathbb{E}\left(r_{A,t-k_1}r_{B,t-k_2}r_{A,s}r_{B,u}\right)\\
&+\sum_{k_2=t-u+1}^{+\infty}k_\times^j(t-u,k_2)\mathcal{D}_{p,(AB)BA}(u-t+k_2,u-s)\\
&\text{we need, in the last term to work on, } t-k_2=u\\
=&\sigma_{j,\infty}^2\overline{r_Ar_B}\\
&+\sum_{i=1}^{N}\sum_{k=1}^{t-s-1}\phi_i^j(k)\mathcal{D}_{p,(ii)AB}(t-k-s,t-k-u)\\
&+2\sum_{i=1}^{N}\sum_{k_2=t-s+1}^{+\infty}K_i^j(t-s,k_2)\mathcal{D}_{p,(iB)iA}(s-t+k_2,u-s)\\
&+\sum_{k=1}^{t-s-1}\phi_\times^j(k)\left(\mathcal{D}_{(AB)AB}(t-k-s,t-k-u)-\overline{r_Ar_B}^2\right)\Big)\\
&+\sum_{k_1=t-u+1}^{+\infty}k_\times^j(k_1,t-u)\mathcal{D}_{p,(BB)AA}(u-t+k_1,u-s)\\
&+\sum_{k_2=t-s+1}^{+\infty}k_\times^j(t-s,k_2)\mathcal{D}_{p,(AB)BA}(s-t+k_2,u-s)
\end{aligned}
$$

We now set $t=\tau_1$, $t-u=\tau_2$ and $s=0$, which results in

$$
\begin{aligned}
\mathcal{D}_{\times,jAB}(\tau_1,\tau_2)=&\sigma_{j,\infty}^2\overline{r_Ar_B}\\
&+\sum_{i=1}^{N}\sum_{k=1}^{\tau_1-1}\phi_i^j(k)\mathcal{D}_{p,(ii)AB}(\tau_1-k,\tau_2-k)\\
&+2\sum_{i=1}^{N}\sum_{k_2=\tau_1+1}^{+\infty}K_i^j(\tau_1,k_2)\mathcal{D}_{p,(iB)iA}(k_2-\tau_1,\tau_1-\tau_2)
\end{aligned}
$$

$$
\begin{aligned}
&+\sum_{k=1}^{\tau_1-1}\phi_\times^j(k)\left(\mathcal{D}_{(AB)AB}(\tau_1-k,\tau_2-k)-\overline{r_Ar_B}^2\right)\Big)\\
&+\sum_{k_1=\tau_2+1}^{+\infty}k_\times^j(k_1,\tau_2)\mathcal{D}_{p,(BB)AA}(k_1-\tau_2,\tau_1-\tau_2)\\
&+\sum_{k_2=\tau_1+1}^{+\infty}k_\times^j(\tau_1,k_2)\mathcal{D}_{p,(AA)BB}(k_2-\tau_1,\tau_1-\tau_2),
\end{aligned}
$$

which concludes the proof.

□

Appendix B.7 gives more details on who to concretely implement those systems of linear equations between kernels and covariance structures.

Yule-Walker system for QGARCH 2D for Leverage kernel (step 4)

Finally, once the linear, quadratic and cross kernels are determined, we can work on the original dataset (not symmetrised) and characterised the leverage kernels. To achieve this, additional Yule-Walker equations are required relying on $\mathcal{V}$ (defined in Equations (B.5.8) and (B.5.9)), we derive them subsequently. The goal is to characterise $(L_i^j)_{i,j\in\{A,B\}}$ up to a certain lag q. Hence, we need $4\times q$ equations which are obtained using $(\mathcal{V}_{jl}(\tau))_{j,l\in\{A,B\},1\leq\tau\leq q}$.

Lemma B.5.7 (Yule-Walker equation for $\mathcal{V}$)**.** *for $\tau>0$, in the framework described by Eq.* (B.5.2)*, the covariance $\mathcal{V}$ writes*

$$
\begin{aligned}
\mathcal{V}_{jl}(\tau)=&L_i^j(\tau)\overline{r_ir_l}\\
&+\sum_{i=1}^{N}\sum_{k=1}^{\tau}\phi_i^j(k)\mathcal{V}_{il}^r(\tau-k)\\
&+2\sum_{k_2=\tau+1}^{+\infty}K_i^j(\tau,k_2)\mathcal{V}_{(il)j}^r(k_2-\tau)\\
&+\sum_{k=1}^{\tau}\phi_\times^j(k)\mathcal{V}_{(AB)l}^r(\tau-k)\\
&+\sum_{k_2=\tau+1}^{+\infty}k_\times^j(\tau,k_2)\mathcal{V}_{(Al)B}^r(k_2-\tau)\\
&+\sum_{k_1=\tau+1}^{+\infty}k_\times^j(k_1,\tau)\mathcal{V}_{(Bl)A}^r(k_1-\tau)
\end{aligned}
$$

Proof. we consider $s<t$,

$$
\begin{aligned}
\mathcal{V}_{jl}(t-s)=&\mathbb{E}(\sigma_{i,t}^2r_{l,s})\\
=&\sigma_{j,\infty}^2\mathbb{E}(r_{l,s})\\
&+\sum_{i=1}^{N}\sum_{k=1}^{+\infty}L_i^j(k)\mathbb{E}(r_{i,t-k}r_{l,s})\\
&+\sum_{i=1}^{N}\sum_{k=1}^{+\infty}\phi_i^j(k)\mathbb{E}(r_{i,t-k}^2r_{l,s})\\
&+2\sum_{k_1=1}^{+\infty}\sum_{k_2=k_1+1}^{+\infty}K_i^j(k_1,k_2)\mathbb{E}(r_{i,t-k_1}r_{i,t-k_2}r_{l,s})\\
&+\sum_{k=1}^{+\infty}\phi_\times^j(k)\mathbb{E}\left(\left(r_{A,t-k}r_{B,t-k}-\overline{r_Ar_B}\right)r_{l,s}\right)
\end{aligned}
$$

$$
\begin{aligned}
&+\sum_{k_1=1}^{+\infty}\sum_{k_2=k_1+1}^{+\infty}k_{\times}^{j}(k_1,k_2)\mathbb{E}(r_{A,t-k_1}r_{B,t-k_2}r_{l,s})\\
&+\sum_{k_1=2}^{+\infty}\sum_{k_2=1}^{k_1-1}k_{\times}^{j}(k_1,k_2)\mathbb{E}(r_{A,t-k_1}r_{B,t-k_2}r_{l,s})\\
=&L_i^j(t-s)\overline{r_ir_l}\\
&+\sum_{i=1}^{N}\sum_{k=1}^{t-s}\phi_i^j(k)\mathcal{V}_{il}^r(t-k-s)\\
&+2\sum_{k_2=t-s+1}^{+\infty}K_i^j(t-s,k_2)\mathcal{V}_{(il)i}^r(s-t+k_2)\\
&+\sum_{k=1}^{t-s}\phi_{\times}^j(k)\mathcal{V}_{(AB)l}^r(t-k-s)\\
&+\sum_{k_2=t-s+1}^{+\infty}k_{\times}^j(t-s,k_2)\mathcal{V}_{(Al)B}^r(s-t+k_2)\\
&+\sum_{k_1=t-s+1}^{+\infty}k_{\times}^j(k_1,t-s)\mathcal{V}_{(Bl)A}^r(s-t+k_1)
\end{aligned}
$$

we set $t=\tau$ and $s=0$, and we obtain

$$
\begin{aligned}
\mathcal{V}_{jl}(\tau)=&L_i^j(\tau)\overline{r_ir_l}\\
&+\sum_{i=1}^{N}\sum_{k=1}^{\tau}\phi_i^j(k)\mathcal{V}_{il}^r(\tau-k)\\
&+2\sum_{k_2=\tau+1}^{+\infty}K_i^j(\tau,k_2)\mathcal{V}_{(il)i}^r(k_2-\tau)\\
&+\sum_{k=1}^{\tau}\phi_{\times}^j(k)\mathcal{V}_{(AB)l}^r(\tau-k)\\
&+\sum_{k_2=\tau+1}^{+\infty}k_{\times}^j(\tau,k_2)\mathcal{V}_{(Al)B}^r(k_2-\tau)\\
&+\sum_{k_1=\tau+1}^{+\infty}k_{\times}^j(k_1,\tau)\mathcal{V}_{(Bl)A}^r(k_1-\tau)
\end{aligned}
$$

□

B.6 Dealing with Multivariate Yule Walker Equations – example with the 2D linear case

This appendix is dedicated to give some insights on the calibration of a Multivariate (2D) GARCH (or Hawkes) from the Yule-Walker system. Specifically, we derive the system of Yule-Walker type equations for a 2D linear GARCH, meaning that we consider $K = \phi_\times = k_\times = 0$ in Equation (B.5.2). Thus, we only need to characterise $(\phi_i^j)_{i,j\in\{A,B\}}$. In multidimensions, for N assets, the Yule-Walker equation on $\mathcal{C}$ gives (see Lemma B.5.2), for $\tau > 0$,

$$\mathcal{C}_{jl}(\tau) = \sum_{i=1}^{N} \sum_{k=1,k\neq\tau}^{+\infty} \phi_i^j(k)\mathcal{C}_{il}^r(\tau - k).$$

We can then deduce a system on the matrices $\mathbb{C}$ and Φ, defined as

$$\mathbb{C}(\tau) = \begin{pmatrix} \mathcal{C}_{11} & \mathcal{C}_{12} \\ \mathcal{C}_{21} & \mathcal{C}_{22} \end{pmatrix}, \quad \text{and} \quad \Phi(\tau) = \begin{pmatrix} \phi_1^1 & \phi_2^1 \\ \phi_1^2 & \phi_2^2 \end{pmatrix},$$

which results in

$$\begin{pmatrix} \mathbb{C}^\top(1) \\ \mathbb{C}^\top(2) \\ \vdots \\ \mathbb{C}^\top(q) \end{pmatrix} = \begin{pmatrix} \mathbb{C}^\top(0) & \mathbb{C}^\top(-1) & \dots & \mathbb{C}^\top(-(q-1)) \\ \mathbb{C}^\top(1) & \mathbb{C}^\top(0) & \dots & \mathbb{C}^\top(-(q-2)) \\ \vdots & & & \mathbb{C}^\top(-1) \\ \mathbb{C}^\top(q-1) & \mathbb{C}^\top(q-2) & \dots & \mathbb{C}^\top(0) \end{pmatrix} \begin{pmatrix} \Phi^\top(1) \\ \Phi^\top(2) \\ \vdots \\ \Phi^\top(q) \end{pmatrix}. \tag{B.6.1}$$

Such "tensors" system is not really straight forward to implement in `python`. One solution is to flatten the equation which results in 2 systems (in 2D):

$$\begin{pmatrix} \mathcal{C}_{11}(1) \\ \mathcal{C}_{12}(1) \\ \vdots \\ \mathcal{C}_{11}(q-1) \\ \mathcal{C}_{12}(q-1) \end{pmatrix} = \mathbb{A} \begin{pmatrix} \phi_1^1(1) \\ \phi_1^1(2) \\ \vdots \\ \phi_1^1(q-1) \\ \phi_2^1(1) \\ \phi_2^1(2) \\ \vdots \\ \phi_2^1(q-1) \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \mathcal{C}_{21}(1) \\ \mathcal{C}_{22}(1) \\ \vdots \\ \mathcal{C}_{21}(q-1) \\ \mathcal{C}_{22}(q-1) \end{pmatrix} = \mathbb{A} \begin{pmatrix} \phi_1^2(1) \\ \phi_1^2(2) \\ \vdots \\ \phi_1^2(q-1) \\ \phi_2^2(1) \\ \phi_2^2(2) \\ \vdots \\ \phi_2^2(q-1) \end{pmatrix}$$

with

$$\mathbb{A} = \begin{pmatrix} \mathcal{C}_{11}(0) & \mathcal{C}_{11}(1) & \dots & \mathcal{C}_{11}(q-1) & \mathcal{C}_{21}(0) & \mathcal{C}_{21}(-1) & \dots & \mathcal{C}_{21}(-(q-1)) \\ \mathcal{C}_{12}(0) & \mathcal{C}_{12}(1) & \dots & \mathcal{C}_{12}(-(q-1)) & \mathcal{C}_{22}(0) & \mathcal{C}_{22}(-1) & \dots & \mathcal{C}_{22}(-(q-1)) \\ \vdots & & & & & & & \\ \mathcal{C}_{11}(q-1) & \mathcal{C}_{11}(q-2) & \dots & \mathcal{C}_{11}(0) & \mathcal{C}_{21}(q-1) & \mathcal{C}_{21}(q-2) & \dots & \mathcal{C}_{21}(0) \\ \mathcal{C}_{12}(q-1) & \mathcal{C}_{12}(q-2) & \dots & \mathcal{C}_{12}(0) & \mathcal{C}_{22}(q-1) & \mathcal{C}_{22}(q-2) & \dots & \mathcal{C}_{22}(0) \end{pmatrix}$$

$(\mathcal{C}_{ij}(\tau))_{i,j\in\{1,2\}}$ for $\tau \in [\![1, q]\!]$ can be computed from data and is enough to characterise the linear kernel $(\phi_i^j)_{i,j\in\{1,2\}}$ up to the lag q.

Remark B.6.1. The thought process leading to the linear system between $\mathcal{C}$ and ϕ that is presented above is my own and is by evidence not unique.

B.7 Calibration Matrices - QGARCH

The goal of the calibration up to step 3 is to deduce the kernels $(\phi_j^i)_{i,j\in\{1,2\}}$, $(K_j^i)_{i,j\in\{1,2\}}$, $(\phi_\times^i)_{i\in\{1,2\}}$ and $(k_\times^i)_{i\in\{1,2\}}$ that weight the contributions of the past on the squared volatility as defined in Eq. (B.5.2), from the observation of the covariance structures defined in Appendix B.5.2. The link between the kernels and the covariance structures is established by the Yule-Walker equations developed in Appendix B.5. To implement this Yule-Walker system of equations, we propose here a matricel environment. The calibration matrices to deduce the leverage kernels are described in a subsequent section.

B.7.1 2D QGARCH – framework

Consistently, we present a matricel system for the 2D QGARCH, considering the two assets (1 or A) and (2 or B). The goal is to link the kernels

$$\phi\!\!\!|\, = \begin{pmatrix} \phi_1^1 & \phi_2^1 \\ \phi_1^2 & \phi_2^2 \end{pmatrix}, \quad \mathbb{K} = \begin{pmatrix} K_1^1 & K_2^1 \\ K_1^2 & K_2^2 \end{pmatrix}, \quad \boldsymbol{\phi}_\times = \begin{pmatrix} \phi_\times^1 \\ \phi_\times^2 \end{pmatrix}, \quad \text{and} \quad \boldsymbol{k}_\times = \begin{pmatrix} k_\times^1 \\ k_\times^2 \end{pmatrix}$$

to the observable covariance structures

$$\mathbb{C} = \begin{pmatrix} \mathcal{C}_1^1 & \mathcal{C}_2^1 \\ \mathcal{C}_1^2 & \mathcal{C}_2^2 \end{pmatrix}, \quad \mathbb{D}_{\mathrm{d}} = \begin{pmatrix} \mathcal{D}_{111} & \mathcal{D}_{122} \\ \mathcal{D}_{211} & \mathcal{D}_{222} \end{pmatrix} \quad \text{and} \quad \mathcal{D}_\times = \begin{pmatrix} \mathcal{D}_{112} \\ \mathcal{D}_{212} \end{pmatrix}.$$

The goal of this appendix is to describe a system of tensors which describes the relationship between the kernels and the covariance structures up to a certain lag q. We construct the following system:

$$\begin{pmatrix} \mathbb{A}_{1\mathrm{d}} & \mathbb{A}_{2\mathrm{d}} & 0 & 0 \\ \mathbb{A}_{3\mathrm{d}} & \mathbb{A}_{4\mathrm{d}} & 0 & 0 \\ \mathbb{A}_{1\mathrm{d}\times} & \mathbb{A}_{2\mathrm{d}\times} & \mathbb{A}_{1\times} & \mathbb{A}_{2\times} \\ \mathbb{A}_{3\mathrm{d}\times} & \mathbb{A}_{4\mathrm{d}\times} & \mathbb{A}_{3\times} & \mathbb{A}_{4\times} \end{pmatrix} \begin{pmatrix} \phi\!\!\!|\,(\tau) \\ \mathbb{K}(\tau_1,\tau_2)_{\tau_1<\tau_2} \\ \boldsymbol{\phi}_\times(\tau) \\ \boldsymbol{k}_\times(\tau_1,\tau_2)_{\tau_1<\tau_2} \\ \boldsymbol{k}_\times(\tau_1,\tau_2)_{\tau_1>\tau_2} \end{pmatrix} = \begin{pmatrix} \mathbb{C}(\tau) \\ \mathbb{D}_{\mathrm{d}}(\tau_1,\tau_1)_{\tau_1<\tau_2} \\ \mathcal{D}_\times(\tau,\tau) \\ \mathcal{D}_\times(\tau_1,\tau_2)_{\tau_1<\tau_2} \\ \mathcal{D}_\times(\tau_1,\tau_2)_{\tau_1>\tau_2} \end{pmatrix}. \tag{B.7.1}$$

Another way to comprehend the system is to see that $\phi\!\!\!|\,$ is the time-diagonal of K and thus, one can separate the cross components from the linear and quadratic components, as follows:

$$\begin{pmatrix} \mathbb{A}_{\mathrm{d}} & 0 \\ \mathbb{A}_{\mathrm{d}\times} & \mathbb{A}_\times \end{pmatrix} \begin{pmatrix} \mathbb{K}_{\mathrm{d}} \\ \boldsymbol{K}_\times \end{pmatrix} = \begin{pmatrix} \mathbb{D}_{\mathrm{d}} \\ \mathcal{D}_\times \end{pmatrix}.$$

Let us describe the tensors system above:

- **Covariance structures**: Represented by the right hand in Eq. (B.7.1), the covariance tensors are composed of observable elements.
  - The red matrices are correlation tensors of size $q \times 2 \times 2$ (or more generally $q \times N \times N$ if N is the number of assets). They represent the covariance of the squared volatility and the covariance between the volatility and the past trends.
  - The magenta tensors are also covariance structures, capturing the influence of the past *cross* trends on future volatility. The time-diagonal $\mathcal{D}_\times(\tau,\tau)$, with $\tau \in [\![1,q]\!]$, is a tensor of 2D vectors with shape $q \times 2 \times 1$. Conversely to $\mathbb{D}_{\mathrm{d}}$, $\mathcal{D}_\times$ is non-time-symmetric, i.e, $\mathcal{D}_\times(\tau_1,\tau_2) \neq \mathcal{D}_\times(\tau_2,\tau_1)$, thus, we need to consider $\mathcal{D}_\times(\tau_1,\tau_2)$ for both $\tau_1 < \tau_2$ and $\tau_2 < \tau_1$, both of which shape $q\frac{q-1}{2} \times 2 \times 1$.
- **Kernels**: The kernels weight the feedback of past realisations on the future square volatility and characterise the process. These kernels are the unknowns we want to recover.
  - The blue tensors represent the linear and quadratic we want to calibrate (steps 1 and 2 of Appendix B.5). The shape of $\mathbb{K}_{\mathrm{d}}$ is, consistently with $\mathbb{D}_{\mathrm{d}}$, $q + \frac{q(q-1)}{2} \times 2 \times 2$ ($\phi\!\!\!|\,$ has shape $q \times 2 \times 2$ and $\mathbb{K}$ has shape $\frac{q(q-1)}{2} \times 2 \times 2$).

– The cyan tensors are the lags of the kernels weighting the feedback of the cross trend. Consistently with $\mathcal{D}_\times$, their shape is $q + q(q-1) \times 2 \times 1$.

- **Yule-Walker matrices**: These matrices are built upon observables according to the Yule-Walker system of equations developed in Appendix B.5. We explain briefly here their role, but the subsequent sections are dedicated to describe them precisely.

  – The violet calibration matrices relate the linear and quadratic kernels represented by $\mathbb{K}_\mathrm{d}$ to the correlations structure $\mathbb{D}_\mathrm{d}$.

  – The teal calibration matrices relate the kernel of the cross trends $\boldsymbol{K}_\times$ to the cross correlations $\mathcal{D}_\times$.

  – the olive components relate the linear and quadratic kernel $\mathbb{K}_\mathrm{d}$ to the cross correlations $\mathcal{D}_\times$.

The above system already considers a calibration in several steps, otherwise the upper right bloc would not be 0 as $\boldsymbol{K}_\times$ could contribute to $\mathbb{D}_\mathrm{d}$. Thus, we first solve the upper bloc for $\mathbb{K}_\mathrm{d}$ using $\mathbb{A}_\mathrm{d}\mathbb{K}_\mathrm{d} = \mathbb{D}_\mathrm{d}$, and then we solve for $\boldsymbol{K}_\times$ using $\mathbb{A}_{\mathrm{d}\times}\mathbb{K}_\mathrm{d} + \mathbb{A}_\times \boldsymbol{K}_\times = \mathcal{D}_\times$. Let us note that the first calibration step to determine $\mathbb{K}_\mathrm{d}$ is also decomposed in two, as we first characterise the self feedback (1D calibration) and then the cross-linear and quadratic feedback. The rest of this appendix is dedicated to the description of the the Yule-Walker matrices which define the system.

B.7.2 Building the Yule-Walker Matrices

The Yule-Walker matrices $\mathbb{A}_\mathrm{d}$, $\mathbb{A}_{\mathrm{d}\times}$ and $\mathbb{A}_\times$ are built according to the Yule-Walker lemmas of Appendix B.5. In fact, each matrix ($\mathbb{A}_{1\mathrm{d}}$, $\mathbb{A}_{2\mathrm{d}}$, $\mathbb{A}_{3\mathrm{d}}$, $\mathbb{A}_{4\mathrm{d}}$, $\mathbb{A}_{1\mathrm{d}\times}$, $\mathbb{A}_{2\mathrm{d}\times}$, $\mathbb{A}_{3\mathrm{d}\times}$, $\mathbb{A}_{4\mathrm{d}\times}$, $\mathbb{A}_{1\times}$, $\mathbb{A}_{2\times}$, $\mathbb{A}_{3\times}$, $\mathbb{A}_{4\times}$) replicates one term of these equations. Upon scrutiny, it appears that some terms have the same structure, and only the components, i.e. the kernels and the covariance involved in the term, change. In particular, there are 4 different structures to implement: A_1, A_2, A_3 and A_4.

A_1	A_2	A_3	A_4
$\mathbb{A}_{1\mathrm{d}}$	$\mathbb{A}_{2\mathrm{d}}$	$\mathbb{A}_{3\mathrm{d}}$	$\mathbb{A}_{4\mathrm{d}}$
$\mathbb{A}_{1\mathrm{d}\times}$	$\mathbb{A}_{2\mathrm{d}\times}$	$\mathbb{A}_{3\mathrm{d}\times}$	$\mathbb{A}_{4\mathrm{d}\times}$
$\mathbb{A}_{1\times}$	$\mathbb{A}_{2\times}$	$\mathbb{A}_{3\times}$	$\mathbb{A}_{4\times}$

Table B.1: Calibration matrices with similar structure

Specifically, restating the Equations of lemmas B.5.2, B.5.3, B.5.4, B.5.5 and B.5.6, we make the link between the matrices and the term they represent.

From lemma B.5.2, we have

$$\mathcal{C}_{jl}(\tau) = \underbrace{\sum_{i=1}^{N} \sum_{k=1, k\neq\tau}^{+\infty} \phi_i^j(k) \mathcal{C}_{il}^r(\tau - k)}_{\mathbb{A}_{1\mathrm{d}} \in \mathbb{A}_\mathrm{d}} + \underbrace{2 \sum_{i=1}^{N} \sum_{k_2=\tau+1}^{+\infty} \sum_{k_1=k_2+1}^{+\infty} K_i^j(k_1, k_2) \mathcal{D}_{p,lii}(k_1 - \tau, k_2 - \tau)}_{\mathbb{A}_{2\mathrm{d}} \in \mathbb{A}_\mathrm{d}} .$$

From lemma B.5.3, we have, for $\tau_1 < \tau_2$,

$$\mathcal{D}_{jll}(\tau_1, \tau_2) = \underbrace{\sum_{i=1}^{N} \sum_{k=1}^{\tau_1 - 1} \phi_i^j(k) \mathcal{D}_{p,ill}(\tau_1 - k, \tau_2 - k)}_{\mathbb{A}_{3\mathrm{d}} \in \mathbb{A}_\mathrm{d}} + \underbrace{2 \sum_{i=1}^{N} \sum_{k_2=\tau_1+1}^{+\infty} K_i^j(\tau_1, k_2) \mathcal{D}_{p,(il)il}(-\tau_1 + k_2, \tau_2 - \tau_1)}_{\mathbb{A}_{4\mathrm{d}} \in \mathbb{A}_\mathrm{d}}$$

From lemma B.5.4, we have, for, $\tau > 0$

$$\begin{aligned}\mathcal{D}_{\times,jAB}(\tau,\tau) - \sigma^2_{j,\infty}\overline{r_A r_B} =& \underbrace{\sum_{i=1}^{N}\sum_{k=1}^{\tau-1}\phi_i^j(k)\mathcal{D}_{p,(ii)AB}(\tau-k,\tau-k)}_{\mathbb{A}_{1\mathrm{d}\times}\in A_{\mathrm{d}\times}} \\ &+ \underbrace{2\sum_{i=1}^{N}\sum_{k_1=\tau+1}^{+\infty}\sum_{k_2=k_1+1}^{+\infty}K_i^j(k_1,k_2)\mathcal{D}_{p,(AB)ii}(k_1-\tau,k_2-\tau)}_{\mathbb{A}_{2\mathrm{d}\times}\in A_{\mathrm{d}\times}} \\ &+ \underbrace{\sum_{k=1}^{\tau-1}\phi_\times^j(k)\left(\mathcal{D}_{p(AB)AB}(\tau-k,\tau-k) - \overline{r_A r_B}^2\right)}_{\mathbb{A}_{1\times}\in\mathbb{A}_\times} \\ &+ \underbrace{\sum_{k_2=\tau+1}^{+\infty}\sum_{k_1=k_2+1}^{+\infty}k_\times^j(k_1,k_2)\mathcal{D}_{p,(AB)AB}(k_1-\tau,k_2-\tau)}_{\mathbb{A}_{2\times}\in\mathbb{A}_\times} \\ &+ \underbrace{\sum_{k_1=\tau+1}^{+\infty}\sum_{k_2=k_1+1}^{+\infty}k_\times^j(k_1,k_2)\mathcal{D}_{p,(AB)AB}(k_1-\tau,k_2-\tau)}_{\mathbb{A}_{2\times}\in\mathbb{A}_\times}\end{aligned}$$

From lemma B.5.5, we have, for, $\tau_1 < \tau_2$

$$\begin{aligned}\mathcal{D}_{\times,jAB}(\tau_1,\tau_2) - \sigma^2_{j,\infty}\overline{r_A r_B} =& \underbrace{\sum_{i=1}^{N}\sum_{k=1}^{\tau_1-1}\phi_i^j(k)\mathcal{D}_{p,(ii)AB}(\tau_1-k,\tau_2-k)}_{\mathbb{A}_{3\mathrm{d}\times}\in\mathbb{A}_{\mathrm{d}\times}} \\ &+ \underbrace{2\sum_{i=1}^{N}\sum_{k_2=\tau_1+1}^{+\infty}K_i^j(\tau_1,k_2)\mathcal{D}_{p,(iA)iB}(k_2-\tau_1,\tau_2-\tau_1)}_{\mathbb{A}_{4\mathrm{d}\times}\in\mathbb{A}_{\mathrm{d}\times}} \\ &+ \underbrace{\sum_{k=1}^{\tau_1-1}\phi_\times^j(k)\left(\mathcal{D}_{(AB)AB}(\tau_1-k,\tau_2-k) - \overline{r_A r_B}^2\right)\Big)}_{\mathbb{A}_{3\times}\in\mathbb{A}_\times} \\ &+ \underbrace{\sum_{k_1=\tau_1+1}^{+\infty}k_\times^j(k_1,\tau_1)\mathcal{D}_{p,(BA)AB}(k_1-\tau_1,\tau_2-\tau_1)}_{\mathbb{A}_{4\times}\in\mathbb{A}_\times} \\ &+ \underbrace{\sum_{k_2=\tau_1+1}^{+\infty}k_\times^j(\tau_1,k_2)\mathcal{D}_{p,(AA)BB}(k_2-\tau_1,\tau_2-\tau_1)}_{\mathbb{A}_{4\times}\in\mathbb{A}_\times}\end{aligned}$$

From lemma B.5.6, we have, for, $\tau_1 > \tau_2$

$$\begin{aligned}\mathcal{D}_{\times,jAB}(\tau_1,\tau_2) - \sigma^2_{j,\infty}\overline{r_A r_B} =& \underbrace{\sum_{i=1}^{N}\sum_{k=1}^{\tau_1-1}\phi_i^j(k)\mathcal{D}_{p,(ii)AB}(\tau_1-k,\tau_2-k)}_{\mathbb{A}_{3\mathrm{d}\times}\in\mathbb{A}_{\mathrm{d}\times}} \\ &+ \underbrace{2\sum_{i=1}^{N}\sum_{k_2=\tau_1+1}^{+\infty}K_i^j(\tau_1,k_2)\mathcal{D}_{p,(iB)iA}(k_2-\tau_1,\tau_1-\tau_2)}_{\mathbb{A}_{4\mathrm{d}\times}\in\mathbb{A}_{\mathrm{d}\times}}\end{aligned}$$

$$+\underbrace{\sum_{k=1}^{\tau_1-1}\phi_\times^j(k)\left(\mathcal{D}_{(AB)AB}(\tau_1-k,\tau_2-k)-\overline{r_A r_B}^2\right)}_{\mathbb{A}_{3\times}\in\mathbb{A}_\times}$$

$$+\underbrace{\sum_{k_1=\tau_2+1}^{+\infty}k_\times^j(k_1,\tau_2)\mathcal{D}_{p,(BB)AA}(k_1-\tau_2,\tau_1-\tau_2)}_{\mathbb{A}_{4\times}\in\mathbb{A}_\times}$$

$$+\underbrace{\sum_{k_2=\tau_1+1}^{+\infty}k_\times^j(\tau_1,k_2)\mathcal{D}_{p,(AB)BA}(k_2-\tau_1,\tau_1-\tau_2)}_{\mathbb{A}_{4\times}\in\mathbb{A}_\times}$$

Table B.1 recaps the different categories. We now proceed to the description of the four structures.

B.7.3 A_1 Structure

The matrices we call "A_1-like" are the one making the calibration between the diagonal of the observable ($\mathbb{C}$ and $\mathcal{D}_\times$) and the diagonal of a kernel (ϕ or $\phi_\times$). The characteristic equation is

$$E_{A_1}(\tau)=\sum_{u=1}^{+\infty}K(s)\mathcal{D}(\tau-s),\quad \text{where } K \text{ and } \mathcal{D} \text{ are a random kernel and covariance structure.}$$

Example – A_1 relation for $q=3$
Up to the lag $q=3$, we then have the following system

$$\begin{cases} E_{A_1}(1)=K(1)\mathcal{D}(0)+K(2)\mathcal{D}(-1)+K(3)\mathcal{D}(-2)\\ E_{A_1}(2)=K(1)\mathcal{D}(1)+K(2)\mathcal{D}(0)+K(3)\mathcal{D}(-1)\\ E_{A_1}(3)=K(1)\mathcal{D}(2)+K(2)\mathcal{D}(1)+K(3)\mathcal{D}(0)\end{cases}$$

so the Yule-Walker matrix is

$$A_1=\begin{pmatrix}\mathcal{D}(0)&\mathcal{D}(-1)&\mathcal{D}(-2)\\ \mathcal{D}(1)&\mathcal{D}(0)&\mathcal{D}(-1)\\ \mathcal{D}(2)&\mathcal{D}(1)&\mathcal{D}(0)\end{pmatrix}\text{ and when }\mathcal{D}\text{ is even, }A_1=\begin{pmatrix}\mathcal{D}(0)&\mathcal{D}(1)&\mathcal{D}(2)\\ \mathcal{D}(1)&\mathcal{D}(0)&\mathcal{D}(1)\\ \mathcal{D}(2)&\mathcal{D}(1)&\mathcal{D}(0)\end{pmatrix}$$

Generalizing to any q and for a any covariance $\mathcal{D}$, we can define

$$f_{A_1}(\mathcal{D}_{up},\mathcal{D}_{down},\Sigma)=\begin{pmatrix}\Sigma&\mathcal{D}_{up}(1)&\mathcal{D}_{up}(2)&\mathcal{D}_{up}(3)&...&\mathcal{D}_{up}((q-1))\\ \mathcal{D}_{down}(1)&\Sigma&\mathcal{D}_{up}(1)&\mathcal{D}_{up}(2)&...&\mathcal{D}_{up}((q-2))\\ \mathcal{D}_{down}(2)&\mathcal{D}_{down}(1)&\Sigma&\mathcal{D}_{up}(1)&...&\mathcal{D}_{up}((q-3))\\ &&...&&&\\ \mathcal{D}_{down}(q-1)&\mathcal{D}_{down}(q-2)&...&&\mathcal{D}_{down}(1)&\Sigma\end{pmatrix}$$

We now just need to determine $(\mathcal{D}_{up},\mathcal{D}_{down},\Sigma)$ for each of the "A_1-like" matrices.

$\mathbb{A}_{1\mathbf{d}}$

$\mathbb{A}_{1\mathrm{d}}$ is the same as for the linear GARCH model and we have

$$\mathbb{A}_{1\mathrm{d}}=f_{A_1}(\mathbb{C}^\top,\mathbb{C},\mathbb{C}(0)).$$

$\mathbb{A}_{1\mathbf{d}\times}$

$\mathbb{A}_{1\mathrm{d}\times}$ makes the relation between $\mathcal{D}_\times(\tau,\tau)$ and $\oint$. From the Yule-Walker equation in lemma B.5.4, we have

$$\mathbb{A}_{1\mathrm{d}\times}=f_{A_1}(\mathcal{D}_\times,\tau\to\mathcal{D}_\times(-\tau),\mathcal{D}_\times(0)).$$

$A_{1\times}$

Similarly, $A_{1\times}$ encodes the calibration between $\mathcal{D}_\times(\tau,\tau)$ and $\phi_\times$, from lemma B.5.4, we have

$$A_{5\times} = f_{A_1}((\mathcal{D}_{(AB)AB} - \overline{r_A r_B}^2)), \tau \to \left(\mathcal{D}_{(AB)AB}(-\tau) - \overline{r_A r_B}^2\right), \left(\mathcal{D}_{(AB)AB}(0) - \overline{r_A r_B}^2\right))$$

B.7.4 A_2 Structure

"A_2-like" matrices encode the calibration between the diagonal of the observable ($\mathbb{C}$ and $\mathcal{D}_\times$) and the non diagonal of the kernels $\mathbb{K}_{\mathrm{d}}$ and $\boldsymbol{K}_\times$. The characteristic equation is

$$E_{A_2}(\tau) = 2\sum_{k_2=\tau+1}^{+\infty}\sum_{k_1=k_2+1}^{+\infty} K(k_1,k_2)\mathcal{D}(k_1-\tau,k_2-\tau),$$

where K and $\mathcal{D}$ are a random kernel and covariance structure.

Example – A_2 **relation for** $q=3$
Up to the lag $q=3$, we then have the following system

$$\begin{cases} E_{A_2}(1) &= K(1,2)\mathcal{D}(0,1) + K(1,3)\mathcal{D}(0,2) + K(2,3)\mathcal{D}(1,2) \\ E_{A_2}(2) &= K(1,2)\mathcal{D}(-1,0) + K(1,3)\mathcal{D}(-1,1) + K(2,3)\mathcal{D}(0,1) \\ E_{A_3}(3) &= K(1,2)\mathcal{D}(-2,-1) + K(1,3)\mathcal{D}(-2,0) + K(2,3)\mathcal{D}(-1,0) \end{cases}.$$

Hence, we have the following matrix

$$\begin{pmatrix} E(1) \\ E(2) \\ E(3) \end{pmatrix} = \begin{pmatrix} \mathcal{D}(0,1) & \mathcal{D}(0,2) & \mathcal{D}(1,2) \\ \mathcal{D}(-1,0) & \mathcal{D}(-1,1) & \mathcal{D}(0,1) \\ \mathcal{D}(-2,-1) & \mathcal{D}(-2,0) & \mathcal{D}(-1,0) \end{pmatrix} \begin{pmatrix} K(1,2) \\ K(1,3) \\ K(2,3) \end{pmatrix}$$

and if $\mathcal{D}$ is causal,

$$\begin{pmatrix} E(1) \\ E(2) \\ E(3) \end{pmatrix} = \begin{pmatrix} \mathcal{D}(0,1) & \mathcal{D}(0,2) & \mathcal{D}(1,2) \\ 0 & 0 & \mathcal{D}(0,1) \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} K(1,2) \\ K(1,3) \\ K(2,3) \end{pmatrix}$$

Generalizing for any lag q and for general structure $\mathcal{D}$, we can define $f_{A_2}(\mathcal{D})$ as following: if we consider the matrix $(\mathcal{D}(i,j))_{i,j\in[\![-(q-1),(q-1)]\!]}$ then, the n^{th} row of A_2-like matrices is two times the upper triangle of $(\mathcal{D}(i,j))_{i,j\in[\![-(q-1),(q-1)]\!]}[q-n:2q-n, q-n:2q-n]$. Calling f_{A_2} the function performing the transformation $A_2 = f_{A_2}(\mathcal{D})$, we now proceed to the definition of the A_2-like matrices.

$\mathbb{A}_{2\mathbf{d}}$

For $\mathbb{A}_{2\mathrm{d}}$, we use the link between $\mathbb{C}$ and the quadratic kernel $\mathbb{K}$ so

$$\mathbb{A}_{2\mathrm{d}} = f_{A_2}(\mathbb{D}_{\mathrm{d}}^{\top}).$$

$\mathbb{A}_{2\mathbf{d}\times}$

For $\mathbb{A}_{2\mathrm{d}\times}$, we use the link between the time diagonal of $\mathcal{D}_\times$ and the off time-diagonal kernel $\mathbb{K}_{\mathrm{d}}^{\text{off-diagonal}}$ from lemma B.5.4, and we obtain

$$\mathbb{A}_{2\mathrm{d}\times} = f_{A_2}(\begin{pmatrix} \mathcal{D}_{1211} \\ \mathcal{D}_{1222} \end{pmatrix}).$$

$\mathbb{A}_{2\times}$

$\mathbb{A}_{2\times}$ encodes the calibration between the time diagonal of $\mathcal{D}_\times$ and the kernel $\boldsymbol{k}_\times$. Since $\boldsymbol{k}_\times$ is not time-symmetric (conversely to $\mathbb{K}$) we need to considered $\boldsymbol{k}_\times$ for both $\tau_1 < \tau_2$ and $\tau_2 < \tau_1$. This phenomenon intervenes in several calibration matrices, as $\mathbb{A}_{2\times}$. We set the left bloc being the one encoding the calibration of $\boldsymbol{k}_\times(\tau_1,\tau_2)_{\tau_1<\tau_2}$ and the right bloc for $\boldsymbol{k}_\times(\tau_1,\tau_2)_{\tau_1>\tau_2}$ (consistently with the framework developped above).

Then, from lemma B.5.4, and we obtain

$$\mathbb{A}_{2\times}^{\text{left bloc}} = f_{A_2}(\mathcal{D}_{p(12)12}) \quad \& \quad \mathbb{A}_{2\times}^{\text{right bloc}} = f_{A_2}(\mathcal{D}_{p(12)21}).$$

B.7.5 A_3 structure

"A_3-like" matrices are the one linking the off-time-diagonal of observable covariances ($\mathbb{D}_{\text{d}}$ and $\mathcal{D}_\times$ with $\tau_1 \neq \tau_2$) with the time-diagonal kernels (Φ and $\boldsymbol{\phi}_\times$). The characteristic equation is

$$E_{A_3}(\tau_1,\tau_2) = \sum_{u=1}^{+\infty} K(u)\mathcal{D}(\tau_1-u,\tau_2-u)$$

where K and $\mathcal{D}$ are a random kernel and covariance structure.

Example – A_3 **relation for** $q=3$
Up to the lag $q=3$, we then have the following system

$$\begin{cases} \mathcal{D}_{A_3}(1,2) = & K(3)\mathcal{D}(-2,-1) + K(2)\mathcal{D}(-1,0) + K(1)\mathcal{D}(0,1) \\ \mathcal{D}_{A_3}(1,3) = & K(3)\mathcal{D}(-2,0) + K(2)\mathcal{D}(-1,1) + K(1)\mathcal{D}(0,2) \\ \mathcal{D}_{A_3}(2,3) = & K(3)\mathcal{D}(-1,0) + K(2)\mathcal{D}(0,1) + K(1)\mathcal{D}(1,2) \end{cases}.$$

Hence, we obtain the system:

$$\begin{pmatrix} \mathcal{D}_{A_3}(1,2) \\ \mathcal{D}_{A_3}(1,3) \\ \mathcal{D}_{A_3}(2,3) \end{pmatrix} = \begin{pmatrix} \mathcal{D}(0,1) & \mathcal{D}(-1,0) & \mathcal{D}(-2,-1) \\ \mathcal{D}(0,2) & \mathcal{D}(-1,1) & \mathcal{D}(-2,0) \\ \mathcal{D}(1,2) & \mathcal{D}(0,1) & \mathcal{D}(-1,0) \end{pmatrix} \begin{pmatrix} K(1) \\ K(2) \\ K(3) \end{pmatrix}$$

and if $\mathcal{D}$ is causal, we have

$$\begin{pmatrix} \mathcal{D}_{A_3}(1,2) \\ \mathcal{D}_{A_3}(1,3) \\ \mathcal{D}_{A_3}(2,3) \end{pmatrix} = \begin{pmatrix} \mathcal{D}(0,1) & 0 & 0 \\ \mathcal{D}(0,2) & 0 & 0 \\ \mathcal{D}(1,2) & \mathcal{D}(0,1) & 0 \end{pmatrix} \begin{pmatrix} K(1,1) \\ K(2,2) \\ K(3,3) \end{pmatrix}.$$

The function to create "A_3-like" matrices f_{A_3}, has the same structure as A_2, only transposed. Hence, we define $f_{A_3}(\mathcal{D}) = f_{A_2}(\mathcal{D})^\top/2$. We now detailed the definition of each "A_3-like" matrix.

$\mathbb{A}_{3\mathbf{d}}$

$\mathbb{A}_{3\text{d}}$ encodes the feedback of Φ on the time-off-diagonal of $\mathbb{D}_{\text{d}}$. Using lemma B.5.3, we have

$$\mathbb{A}_{3\text{d}} = f_{A_3}(\mathbb{D}_{\text{d}})$$

Remark B.7.1. In python, one can use `(A3=A2.transpose(1,0,3,2)/2)`.

$\mathbb{A}_{3\mathbf{d}\times}$

$\mathbb{A}_{3\text{d}\times}$ encodes the feedback of Φ on the off-time-diagonal of $\mathcal{D}_\times$. $\mathbb{A}_{3\text{d}\times}$ is then composed of two blocs to capture the time asymmetry of $\mathcal{D}_\times$. The upper bloc characterises the feedback of $\boldsymbol{\phi}_\times$ on $\mathcal{D}_\times(\tau_1,\tau_2)$ for $\tau_1<\tau_2$ and the lower bloc for $\tau_1>\tau_2$.

Using lemmas B.5.5 and B.5.6, we can write

$$\mathbb{A}_{3\text{d}\times}^{\text{upper}} = f_{A_3}\left(\begin{pmatrix} \mathcal{D}_{p,(11)AB} \\ \mathcal{D}_{p,(22)AB} \end{pmatrix}\right) \quad \& \quad \mathbb{A}_{3\text{d}\times}^{\text{lower}} = f_{A_3}\left(\begin{pmatrix} \mathcal{D}_{p,(11)BA} \\ \mathcal{D}_{p,(22)BA} \end{pmatrix}\right).$$

The second level is done with the symmetric of the first level.

$\mathbb{A}_{3\times}$

$\mathbb{A}_{3\times}$ encodes the feedback of $\boldsymbol{\phi}_\times$ on the off-time-diagonal of $\mathcal{D}_\times$. As for $\mathbb{A}_{3\mathrm{d}\times}$, we account for the time asymmetry of $\mathcal{D}_\times$ by allocating the upper bloc of $\mathbb{A}_{3\times}$ to $\mathcal{D}_\times(\tau_1,\tau_2)$ for $\tau_1<\tau_2$ and the lower bloc for $\tau_1>\tau_2$. Using lemmas B.5.5 and B.5.6, we have

$$\mathbb{A}_{3\times}^{\text{upper}} = f_{A_3}((\mathcal{D}_{(AB)AB})) \quad \& \quad \mathbb{A}_{3\times}^{\text{lower}} = f_{A_3}((\mathcal{D}_{(AB)BA})).$$

B.7.6 A_4 structure

"A_4-like" matrices are the one linking off-time-diagonal of observable ($\mathbb{D}_\mathrm{d}$ and $\mathcal{D}_\times$ with $\tau_1\neq\tau_2$) and the off-time-diagonal of kernels ($\mathbb{K}$ and $\boldsymbol{k}_\times$). The characteristic equation is

$$E_{A_4}(\tau_1,\tau_2) = 2\sum_{u=\tau_1+1}^{+\infty} K(u,\tau_1)\mathcal{D}(u-\tau_1,\tau_2-\tau_1)$$

where K and $\mathcal{D}$ are a random kernel and covariance structure.

Example – A_4 relation for $q=3$
Up to the lag $q=3$, we then have the following system for $q=3$,

$$\begin{cases} \mathcal{D}_{A_4}(1,2) = & 2(K(2,1)\mathcal{D}(1,1)+K(3,1)\mathcal{D}(2,1)) \\ \mathcal{D}_{A_4}(1,3) = & 2(K(2,1)\mathcal{D}(1,2)+K(3,1)\mathcal{D}(2,2)) \\ \mathcal{D}_{A_4}(2,3) = & 2K(3,2)\mathcal{D}(1,1) \end{cases}.$$

Hence, we obtain the system

$$\begin{pmatrix} \mathcal{D}_{A_4}(1,2) \\ \mathcal{D}_{A_4}(1,3) \\ \mathcal{D}_{A_4}(2,3) \end{pmatrix} = \begin{pmatrix} \mathcal{D}(1,1) & \mathcal{D}(2,1) & 0 \\ \mathcal{D}(1,2) & \mathcal{D}(2,2) & 0 \\ 0 & 0 & \mathcal{D}(1,1) \end{pmatrix} \begin{pmatrix} K(1,2) \\ K(1,3) \\ K(2,3) \end{pmatrix}.$$

The function to create "A_4-like" matrices from the correlation $\mathcal{D}$ can be deduced from the example. It creates blocs of the matrix $(\mathcal{D}(i,j))_{i,j}$ along the diagonal. We now detail the structure of the A_4-like matrices.

$\mathbb{A}_{4\mathbf{d}}$

$\mathbb{A}_{4\mathrm{d}}$ encodes the link between $\mathbb{K}$ and $\mathbb{D}_\mathrm{d}$. From lemma B.5.3, we have

$$\mathbb{A}_{4\mathrm{d}} = f_{A_4}((\mathcal{D}_{(ij)ij})_{ij}).$$

$\mathbb{A}_{4\mathbf{d}\times}$

$\mathbb{A}_{4\mathrm{d}\times}$ encodes the feedback of $\mathbb{K}_\mathrm{d}$ on the off-time-diagonal of $\mathcal{D}_\times$. To account for the asymmetry of $\mathcal{D}_\times$, we build $\mathbb{A}_{4\mathrm{d}\times}$ with two blocs: the upper bloc being for $\mathcal{D}_\times(\tau_1,\tau_2)$ with $\tau_1<\tau_2$ and the lower bloc for $\tau_2<\tau_1$. From lemmas B.5.5 and B.5.6, we have

$$\mathbb{A}_{4\mathrm{d}\times}^{\text{upper}} = f_{A_4}(\begin{pmatrix}\mathcal{D}_{p(AA)AB}\\ \mathcal{D}_{p(AB)BB}\end{pmatrix}) \quad \& \quad \mathbb{A}_{4\mathrm{d}\times}^{\text{lower}} = f_{A_4}(\begin{pmatrix}\mathcal{D}_{p(BA)AA}\\ \mathcal{D}_{p(BB)BA}\end{pmatrix})$$

$\mathbb{A}_{4\times}$

$\mathbb{A}_{4\times}$ encodes the feedback of $\boldsymbol{k}_\times$ on the off-time-diagonal of $\mathcal{D}_\times$. For this particular matrix, one has to account for the asymmetry of $\mathcal{D}_\times$ and of $\boldsymbol{k}_\times$. As before, we thus build blocs for each case. From lemmas B.5.5 and B.5.6, we have

- upper left, $\mathcal{D}_\times(\tau_1 < \tau_2)$ VS $\boldsymbol{k}_\times(\tau_1 < \tau_2)$

$$\mathbb{A}_{4\times}^{\text{upper left}} = f_{A_4}(\big(\mathcal{D}_{p(AA)BB}\big))$$

- upper right, $\mathcal{D}_\times(\tau_1 < \tau_2)$ VS $\boldsymbol{k}_\times(\tau_1 > \tau_2)$

$$\mathbb{A}_{4\times}^{\text{upper right}} = f_{A_4}(\big(\mathcal{D}_{p(AB)AB}\big))$$

- lower left, $\mathcal{D}_\times(\tau_1 > \tau_2)$ VS $\boldsymbol{k}_\times(\tau_1 < \tau_2)$

$$\mathbb{A}_{4\times}^{\text{lower left}} = f_{A_4}(\big(\mathcal{D}_{p(AB)BA}\big))$$

- lower right, $\mathcal{D}_\times(\tau_1 > \tau_2)$ VS $\boldsymbol{k}_\times(\tau_1 > \tau_2)$

$$\mathbb{A}_{4\times}^{\text{lower left}} = f_{A_4}(\big(\mathcal{D}_{p(BB)AA}\big))$$

B.7.7 Calibration matrices for the leverage kernels

Leverage calibration - framework

We now explicate a method to determine the leverage kernels $(L_j^i)_{i,j\in\{1,2\}}$. For those last kernels, we aim at building a linear system such that:

$$\begin{pmatrix} \mathbb{V}(1) \\ \mathbb{V}(2) \\ \vdots \\ \mathbb{V}(q) \end{pmatrix} = \mathbb{A}_{\mathbb{L}} \begin{pmatrix} \mathbb{L}(1) \\ \mathbb{L}(2) \\ \vdots \\ \mathbb{L}(q) \end{pmatrix}$$

where

$$\mathbb{L}(\tau) = \begin{pmatrix} L_1^1(\tau) & L_2^1(\tau) \\ L_1^2(\tau) & L_2^2(\tau) \end{pmatrix} \quad \text{and} \quad \mathbb{V}(\tau) = \begin{pmatrix} \mathcal{V}_{11}(\tau) & \mathcal{V}_{12}(\tau) \\ \mathcal{V}_{21}(\tau) & \mathcal{V}_{22}(\tau) \end{pmatrix}.$$

As before, we rely on the Yule-Walker type equation to construct $\mathbb{A}_{\mathbb{L}}$. As a reminder, lemma B.5.7 gives

$$\begin{aligned} \mathcal{V}_{jl}(\tau) =& L_i^j(\tau)\overline{r_i r_l} \\ &+ \sum_{i=1}^{N}\sum_{k=1}^{\tau} \phi_i^j(k)\mathcal{V}_{il}^r(\tau - k) \\ &+ 2\sum_{k_2=\tau+1}^{+\infty} K_i^j(\tau, k_2)\mathcal{V}_{(il)i}^r(k_2 - \tau) \\ &+ \sum_{k=1}^{\tau} \phi_\times^j(k)\mathcal{V}_{(AB)l}^r(\tau - k) \\ &+ \sum_{k_2=\tau+1}^{+\infty} k_\times^j(\tau, k_2)\mathcal{V}_{(Al)B}^r(k_2 - \tau) \\ &+ \sum_{k_1=\tau+1}^{+\infty} k_\times^j(k_1, \tau)\mathcal{V}_{(Bl)A}^r(k_1 - \tau). \end{aligned}$$

The equation above demonstrates that $\mathcal{V}_{jl}$ not only depends on the leverage kernels $(L_j^i)_{i,j\in\{1,2\}}$ but also on the linear kernels $(\phi_j^i)_{i,j\in\{1,2\}}$, the quadratic kernels $(K_j^i)_{i,j\in\{1,2\}}$ and the cross kernels $(\phi_\times^i)_{i\in\{1,2\}}$ and $(k_\times^i)_{i\in\{1,2\}}$ that we determined in steps 1, 2 and 3. Hence, the equation system becomes:

$$\begin{pmatrix}\tilde{\mathbb{V}}(1)\\ \tilde{\mathbb{V}}(2)\\ \vdots\\ \tilde{\mathbb{V}}(q)\end{pmatrix} = \begin{pmatrix}\begin{pmatrix}\overline{r_1}^2 & \overline{r_1 r_2}\\ \overline{r_1 r_2} & \overline{r_2}^2\end{pmatrix} & & 0\\ & \ddots & \\ 0 & & \begin{pmatrix}\overline{r_1}^2 & \overline{r_1 r_2}\\ \overline{r_1 r_2} & \overline{r_2}^2\end{pmatrix}\end{pmatrix}\begin{pmatrix}\mathbb{L}(1)\\ \mathbb{L}(2)\\ \vdots\\ \mathbb{L}(q)\end{pmatrix}$$

where

$$\begin{pmatrix}\tilde{\mathbb{V}}(1)\\ \tilde{\mathbb{V}}(2)\\ \vdots\\ \tilde{\mathbb{V}}(q)\end{pmatrix} = \begin{pmatrix}\mathbb{V}(1)\\ \mathbb{V}(2)\\ \vdots\\ \mathbb{V}(q)\end{pmatrix} - \mathbb{A}_{1L}\begin{pmatrix}\mathbb{\phi}(1)\\ \mathbb{\phi}(2)\\ \vdots\\ \mathbb{\phi}(q)\end{pmatrix} - \mathbb{A}_{1\times L}\begin{pmatrix}\boldsymbol{\phi}_\times(1)\\ \boldsymbol{\phi}_\times(2)\\ \vdots\\ \boldsymbol{\phi}_\times(q)\end{pmatrix} - \mathbb{A}_{vL}\begin{pmatrix}\mathbb{K}_\times(1,2)\\ \mathbb{K}_\times(1,3)\\ \vdots\\ \mathbb{K}_\times(q-1,q)\end{pmatrix} - \mathbb{A}_{v\times L}\begin{pmatrix}\boldsymbol{K}_\times(1,2)\\ \boldsymbol{K}_\times(1,3)\\ \vdots\\ \boldsymbol{K}_\times(q-1,q)\end{pmatrix},$$

and $\mathbb{A}_{1L}$, $\mathbb{A}_{1\times L}$, $\mathbb{A}_{vL}$ and $\mathbb{A}_{v\times L}$ are to be determined with the Yule-Walker equation of lemma B.5.7. Notably, we have

$$\begin{aligned}\mathcal{V}_{jl}(\tau) =& L_i^j(\tau)\overline{r_i r_l}\\ &+ \underbrace{\sum_{i=1}^{N}\sum_{k=1}^{\tau}\phi_i^j(k)\mathcal{V}_{il}^r(\tau-k)}_{\mathbb{A}_{1L}}\\ &+ \underbrace{2\sum_{k_2=\tau+1}^{+\infty}K_i^j(\tau,k_2)\mathcal{V}_{(il)i}^r(k_2-\tau)}_{\mathbb{A}_{vL}}\\ &+ \underbrace{\sum_{k=1}^{\tau}\phi_\times^j(k)\mathcal{V}_{(AB)j}^r(\tau-k)}_{\mathbb{A}_{1\times L}}\\ &+ \underbrace{\sum_{k_2=\tau+1}^{+\infty}k_\times^j(\tau,k_2)\mathcal{V}_{(Al)B}^r(k_2-\tau)}_{\mathbb{A}_{v\times L}\text{ left bloc}}\\ &+ \underbrace{\sum_{k_1=\tau+1}^{+\infty}k_\times^j(k_1,\tau)\mathcal{V}_{(Bl)A}^r(k_1-\tau)}_{\mathbb{A}_{v\times L}\text{ right bloc}}.\end{aligned}$$

Using the work above, we can already determine $\mathbb{A}_{1L}$ and $\mathbb{A}_{1\times L}$ as they are "A_1-like" matrices. Hence, we have

$$\mathbb{A}_{1L} = f_{A_1}(\mathbf{0}_{2\times 2},(\mathbb{V}^r),\begin{pmatrix}\overline{\sigma_A^2 r_A} & 0\\ 0 & \overline{\sigma_B^2 r_B}\end{pmatrix}) \quad\text{and}\quad \mathbb{A}_{1\times L} = f_{A_1}(\mathbf{0}_{2\times 2},\begin{pmatrix}\mathcal{V}_{(AB)A}^r\\ \mathcal{V}_{(AB)B}^r\end{pmatrix},\begin{pmatrix}0\\0\end{pmatrix}).$$

A_5 structure

Additionally, we need to define another type of matrices for $\mathbb{A}_{vL}$ and $\mathbb{A}_{v\times L}$. Both have the same structure which we call "A_5-like". The characteristic equation is as follows

$$E_{A_5}(\tau) = \sum_{k=\tau+1}^{+\infty}K(\tau,k)\mathcal{D}(k-\tau).$$

where K and $\mathcal{D}$ are a random kernel and covariance structure.

Example – A_5 **relation for** $q=3$

Up to the lag $q = 3$, we then have the following system

$$\begin{cases} E_{A_5}(1) = K(1,2)\mathcal{D}(1) + K(1,3)\mathcal{D}(2) \\ E_{A_5}(2) = K(2,3)\mathcal{D}(1) \\ E_{A_5}(3) = 0 \end{cases}$$

so the Yule-Walker matrix is

$$A_5 = \begin{pmatrix} \mathcal{D}(1) & \mathcal{D}(2) & 0 \\ 0 & 0 & \mathcal{D}(1) \\ 0 & 0 & 0 \end{pmatrix}$$

Hence, the generalisation of "A_5-like" matrices comes down to writing each row i of A_5 as $A_5[i, q-i : 2q-2i-1] += \mathcal{D}[: q-i-1]$ and taking the other entries as zeros. Then, calling f_{A_5} the function to build "A_5-like" matrices such like $A_5 = f_{A_5}(\mathcal{D})$ we can define the last Yule-Walker matrices as

$$\mathbb{A}_{vL} = f_{A_5}(\begin{pmatrix} \mathcal{V}^r_{(AA)A} & \mathcal{V}^r_{(AB)A} \\ \mathcal{V}^r_{(AB)B} & \mathcal{V}^r_{(BB)B} \end{pmatrix}), \quad \mathbb{A}^{\text{left bloc}}_{v\times L} = f_{A_5}(\begin{pmatrix} \mathcal{V}^r_{(AA)B} \\ \mathcal{V}^r_{(AB)B} \end{pmatrix}) \quad \text{and} \quad \mathbb{A}^{\text{right bloc}}_{v\times L} = f_{A_5}(\begin{pmatrix} \mathcal{V}^r_{(BA)B}(-) \\ \mathcal{V}^r_{(BB)B}(-) \end{pmatrix}).$$

Empirically, we observe that the contribution of these kernels is not significant, and so, in practice for the sake of simplicity, we only consider $\mathbb{A}_{1L}$ and $\mathbb{A}_{1\times L}$ in the calibration.

B.8 Calibration - Proof of concept

This appendix provides a proof of concept for the proposed calibration method, demonstrating its results using synthetic data. Notably, we discuss the challenges associated with accurately retrieving the (Q)Hawkes process, an event-by-event point process, from aggregated data. These difficulties justify our decision to approximate the QHawkes process with a QGARCH model, as described by Equation (B.5.1), for the empirical study of Chapter 4.

This appendix is organised as follows. It begins by comparing the calibration of the univariate (Q)Hawkes model with that of the univariate (Q)GARCH model. This comparison shows that, in certain regimes, the calibration of the Hawkes becomes biased due to data binning whereas the QGARCH remains stable across a wider range of parameters.

The second part focuses on calibrating the multivariate (Q)GARCH using the method of moments, providing evidence that the calibration method is reliable.

To compare the calibration of the (Q)Hawkes with the calibration on the (Q)GARCH, we generate datasets in two ways:

- The thinning algorithm (see Appendix B.1) provide time and mark of events for the (Q)Hawkes, enabling the construction of price time series, according to the definition of the QHawkes in Chapter 2, i.e. $\mathrm{d}P = \pm \mathrm{d}N$. These time series are then aggregated into 1 minute bins to create the volatility time series $\sigma^{\mathscr{B}}$ and the 1-minute returns time series $\mathrm{d}P$. These synthetic time series are used to test the calibration of the (Q)Hawkes model.
- The simulation of the (Q)GARCH, as described in Equation (B.5.1), directly results in two time series at 1 minute timescale: the volatility σ^2 and the returns r. These time series are used to test the calibration of the QGARCH.

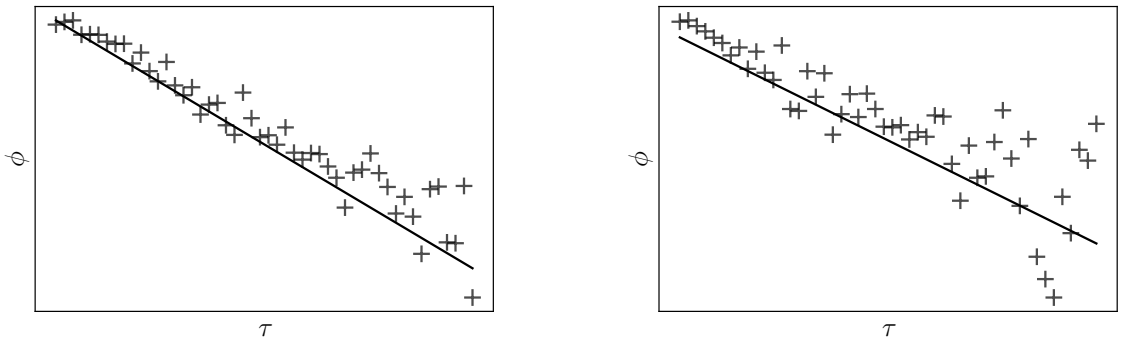
In both cases, we use exponential feedback kernels with the general form:

$$\phi(t) = n_H \beta \exp(-\beta t) \quad \& \quad k(t) = \sqrt{2 n_Z \omega} \exp(-\omega t)$$

B.8.1 Univariate cases: comparison between the (Q)Hawkes and the (Q)GARCH models

This first part is dedicated to show and discuss the calibration results of the univariate (Q)Hawkes and of the univariate (Q)GARCH. It begins with the linear case and then deals with the quadratic extension.

Univariate linear case



(a) In the case $\bar{\lambda} = 0.03\ \mathrm{min}^{-1}$ ($\lambda_\infty = 0.01\ \mathrm{min}^{-1}$).

(b) In the case $\bar{\lambda} = 1.7\ \mathrm{min}^{-1}$ ($\lambda_\infty = 0.5\ \mathrm{min}^{-1}$).

Figure B.1: Calibration of a 1D linear Hawkes with exponential kernel, for two different sets of parameters. In both cases, $n_H = 0.7$ and $\beta = 0.04\ \mathrm{min}^{-1}$ and only the baseline varies. The scatter plot is the kernel obtained by the method of moments and the plain line is the kernel used to generate the synthetic data and that we want to recover.

As a reminder, the linear Hawkes process is a point process N, whose intensity is defined as follows:

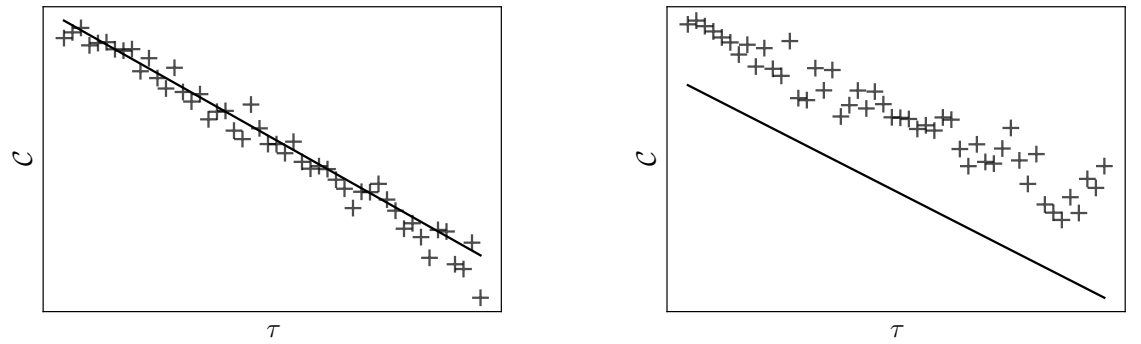
$$\lambda_t = \lambda_\infty + \int_{-\infty}^{t} \phi(t-s)\mathrm{d}N_s \quad \text{with returns defined as} \quad \mathrm{d}P_t = \pm \mathrm{d}N_t.$$

In the case of a linear GARCH process, the squared volatility σ^2 is driven by

$$\sigma_t^2 = \sigma_\infty^2 + \sum_{k=1}^{+\infty} \phi(k) r_{t-k}^2 \quad \text{with returns defined as } r_t = \sigma_t \xi_t, \quad \text{with} \quad \xi_t \sim \mathscr{N}(0,1).$$

The connection between the two models lies in the relationship between the GARCH squared volatility and the Hawkes intensity: $\frac{\sigma^2}{\Delta} \equiv \lambda_t$, where Δ is the time bin used to compute σ [70]. It is interesting to note that, by definition, the Hawkes model is a continuous process describing each event of price change whereas the GARCH model is inherently discretized in time.

The objective of the calibration is to retrieve the kernel ϕ. Specifically, in the case of a exponential kernel, $\phi = n_H \beta \exp(-\beta t)$, we aim to determine the parameters n_H, β and λ_∞ (or σ_∞^2). When the shape of the kernel is unknown, as it is often the case with empirical data, adopting a non-parametric method such as the general method of moments (GMM) allows us to remain agnostic about the kernel's shape. In this appendix, we implement the GMM on synthetic data to test its validity.



(a) In the case $\bar{\lambda} = 0.03$ min^{-1} ($\lambda_\infty = 0.01$ min^{-1}).

(b) In the case $\bar{\lambda} = 0.1.7$ min^{-1} ($\lambda_\infty = 0.5$ min^{-1}).

Figure B.2: Illustration of the estimation of the covariance structure $\mathcal{C}(\tau) = \frac{1}{\bar{\lambda}} \left(\mathbb{E}(\frac{\mathrm{d}N_t}{\mathrm{d}t} \frac{\mathrm{d}N_{t-\tau}}{\mathrm{d}t}) - \bar{\lambda}^2 \right)$ from the generation of a 1D linear Hawkes with exponential kernel, for 2 different sets of parameters. In both cases, $n_H = 0.7$ and $\beta = 0.04$ min^{-1} and only the baseline varies. The scatter plot is the obtained covariance $\mathcal{C}$, while the plain line is the theoretical covariance $\mathcal{C}$ derived in Appendix B.2.

Two key elements about the method of moments should be kept in mind. First, it relies on autocovariance structure, and in particular in the linear case, on the autocovariance of the process's activity $\mathcal{C}(\tau) = \frac{1}{\bar{\lambda}} \left(\mathbb{E}(\frac{\mathrm{d}N_t}{\mathrm{d}t} \frac{\mathrm{d}N_{t-\tau}}{\mathrm{d}t}) - \bar{\lambda}^2 \right)$ (for the GARCH, an equivalent expression is $\mathcal{C}(\tau) = \mathbb{E}(\sigma_t^2 r_{t-\tau}^2) - \overline{\sigma^2 r^2}$). These values form the linear system of equations required to recover ϕ (for more comprehensive details see [70] and Appendix B.7, in particular Section B.7.3). Secondly, the method of moments yields the values $(\phi(\tau))_{1 \leq \tau \leq q}$, where q is the lag after which the covariance and kernels are considered null.

We now proceed to the presentation and discussion of the calibration results of the Hawkes process.

Figure B.1 shows the estimations of the linear kernels ϕ calibrated on synthetic data generated by linear Hawkes processes for two sets of parameters. Specifically, we compare the results obtained from a simulation with $\bar{\lambda}\mathrm{d}t = 0.03$ to those from a second simulation with $\bar{\lambda}\mathrm{d}t = 1.7$. The observations are manifolds. Firstly, although the calibration is noisy, the method of moments successfully retrieves the exponential shape of the input kernel ϕ. Secondly, while the calibration method remains consistent across both parameter sets, it appears to yield better results when $\bar{\lambda}\mathrm{d}t \ll 1$. In this regime, the data binning to obtain $\sigma^{\mathscr{B}}$ from N and P ensures that there is at

most one event per bin, and so $\sigma^{\mathscr{B}} \approx \mathrm{d}N$. However, when $\bar{\lambda}\mathrm{d}t \geq 1$, there are more events within each bin, leading to a overestimation of the activity.

For illustration, Figure B.2 compares the autocovariance, $\mathcal{C}$, of the synthetic activity with its theoretical values (derived in Appendix B.2). As shown in Figure B.2a, when $\bar{\lambda}\mathrm{d}t \ll 1$, the estimated autocovariance matched its theoretical values. In contrast, Figure B.2b demonstrates that $\mathcal{C}$ is overestimated when $\bar{\lambda}\mathrm{d}t = 1.7$. These estimation errors in $\mathcal{C}$ result in a biased estimation of the kernel ϕ as the Yule-Walker system to find ϕ relies solely on $\mathcal{C}$.

For real data, due to the high activity of financial markets, we are situated in a regime where $\bar{\lambda}\mathrm{d}t \geq 6$. Hence, if the underlying process is a linear Hawkes, we expect the autocovariance $\mathcal{C}$ to be significantly overestimated. For the linear Hawkes process, it is possible to determine a correction to apply to $\mathcal{C}$ to retrieve the correct theoretical value and thus to correct the bias in the resulting kernel ϕ. However, extending this correction to the quadratic case is challenging.

In comparison, as shown in Figure B.3, for instance, the calibration of the univariate linear GARCH remains stable across a wide range of parameters, provided that an appropriate number of lags q is chosen (see Appendix B.7). A consistent choice is to take $q > \frac{3}{\beta}$.

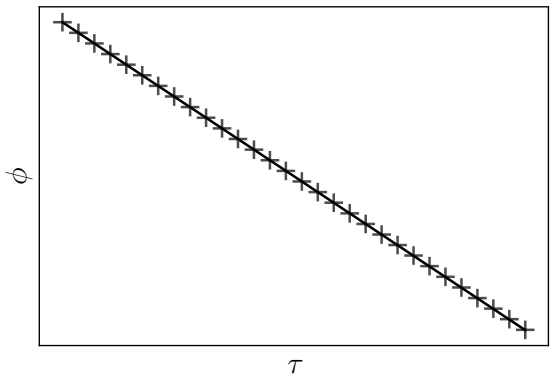


Figure B.3: Proof of concept - Calibration of a 1D linear GARCH with exponential kernel. The scatter plot is the kernel obtained by the method of moments and the plain line is the kernel used to generate the synthetic data and that we want to recover. For this simulation $n_H = 0.7$, $\beta = 0.1$ min^{-1} and $\sigma^2_\infty = 0.1$

Note also that the calibration is much less noisy for the GARCH model compared to the linear Hawkes model.

We now turn our attention to the quadratic case.

Univariate quadratic case

The quadratic case, for both the QHawkes and the QGARCH, is a lot more delicate.

Introduced in [70], the intensity of a QHawkes process is defined by

$$\lambda_t = \lambda_\infty + \int_{-\infty}^{t}\int_{-\infty}^{t} K(t-s, t-u)\mathrm{d}P_s\mathrm{d}P_s \quad \text{with returns defined as} \quad \mathrm{d}P_t = \pm\mathrm{d}N_t.$$

In the case of a QGARCH process, as defined in [12], the square volatility, σ^2, is driven by:

$$\sigma_t^2 = \sigma_\infty^2 + \sum_{i=1}^{+\infty}\sum_{j=1}^{+\infty} K(i,j) r_{t-i} r_{t-j} \quad \text{with returns defined as } r_t = \sigma_t \xi, \quad \text{with} \quad \xi \sim \mathscr{N}(0,1).$$

The connection between the two models remains the same as in the linear case, that is $\frac{\sigma^2}{\Delta} \equiv \lambda_t$.

In the quadratic cases, the kernel K can be understood as a matrix such that $\mathbb{K}_{ij} = K(i,j)$. The objective of the calibration is to estimate the entries of the matrix $\mathbb{K}$ up to a certain lag q; these entries correspond to the values $(K(i,j))_{1\leq i,j\leq q}$. Moreover, since K is symmetric, it is only necessary to determine the upper triangle of $\mathbb{K}$, specifically the values $(K(i,j))_{1\leq i\leq j\leq q}$.

Following the empirical observation of Blanc *et al.* [70], we adopt the ZHawkes framework which expresses the kernels K as the sum of a diagonal contribution and a rank-one component, meaning that $K(\tau_1,\tau_2) = \phi(\tau_1)\delta_{\tau_1-\tau_2} + k(\tau_1)k(\tau_2)$. For exponential kernels, ϕ and k take the form defined above. Note that the ZHawkes framework also greatly simplifies the simulation and limit the number of parameters to determine. Consistently with the explanation above, the method of moments, as detailed in Appendix B.7, returns the estimated diagonal of $\mathbb{K}$ up to a certain lag q, that is $K(\tau,\tau)_{\tau\in[\![1,q]\!]}$, and its upper triangle values, $K(\tau_1,\tau_2)_{1\leq\tau_1<\tau_2\leq q}$. Therefore, within the ZHawkes framework, the diagonal of $\mathbb{K}$ $(K(\tau,\tau)_{\tau\in[\![1,q]\!]})$ provides an an estimation of $\phi + k^2$, and the upper triangle values of $\mathbb{K}$ $(K(\tau_1,\tau_2)_{1\leq\tau_1<\tau_2\leq q})$ allow to rebuild the off-diagonal of $\mathbb{K}$ (since $\mathbb{K}$ is symmetric) and approximate it with a rank-one representation, so called hereafter $\tilde{k}$. Figure B.4 shows the result of such calibration on synthetic data generated under the condition $\bar{\lambda}\mathrm{d}t \ll 1$. Similar to the linear Hawkes process, this regime ensures that there are very few events per bin, thereby maintaining the relationship $\mathrm{d}P = \pm\mathrm{d}N$ and approximating $\sigma^{\mathscr{B}} \approx \mathrm{d}N$.

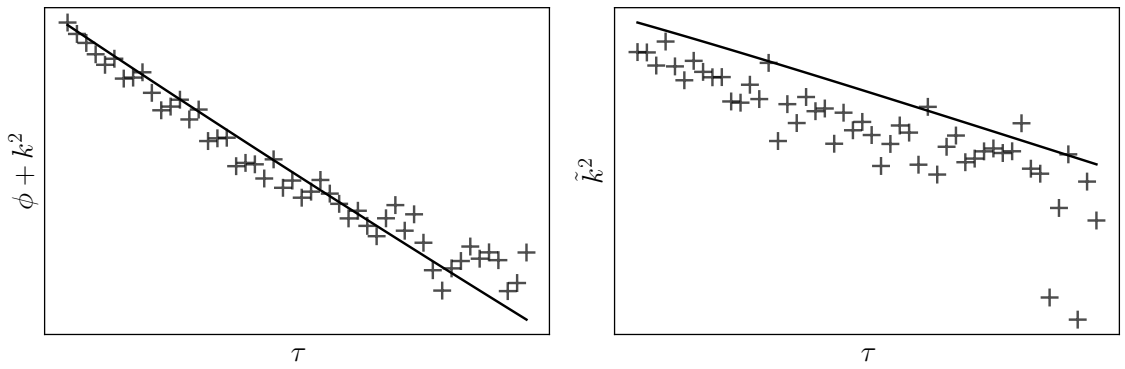


Figure B.4: Calibration of a 1D QHawkes with method of moments from a simulation with parameters $n_H = 0.6$, $\beta = 0.04$ min^{-1}, $n_Z = 0.2$, $\omega = 0.03$ min^{-1}, $\lambda_\infty = 0.003$ min^{-1}, hence $\bar{\lambda}\mathrm{d}t = 0.015$. The left panel presents the diagonal of the kernel $\mathbb{K}(\tau,\tau) = \phi(\tau) + k^2(\tau)$ up to a lag $q = 50$. The right panel presents the rank-one decomposition of the off-diagonal of $\mathbb{K}$. In both panels the scatter plot is the component from the method of moments, while the plain line is the theoretical component we want to recover.

From the method of moments estimation, one can reconstruct the entire kernel matrix $\mathbb{K}$ up to a lag q by combining the diagonal and the upper triangular entries, considering that $\mathbb{K}$ is a symmetric $q \times q$ matrix. Subsequently, the parameters n_H, n_Z, β and ω can be retrieved by minimizing $[\hat{K}(\tau_1,\tau_2) - \phi(\tau_1)\delta_{\tau_1-\tau_2} - k(\tau_1)k(\tau_2)]^2$, where $\hat{K}$ is the estimated matrix from the method of moments and ϕ and k are the parametric kernels defined above. The results of such

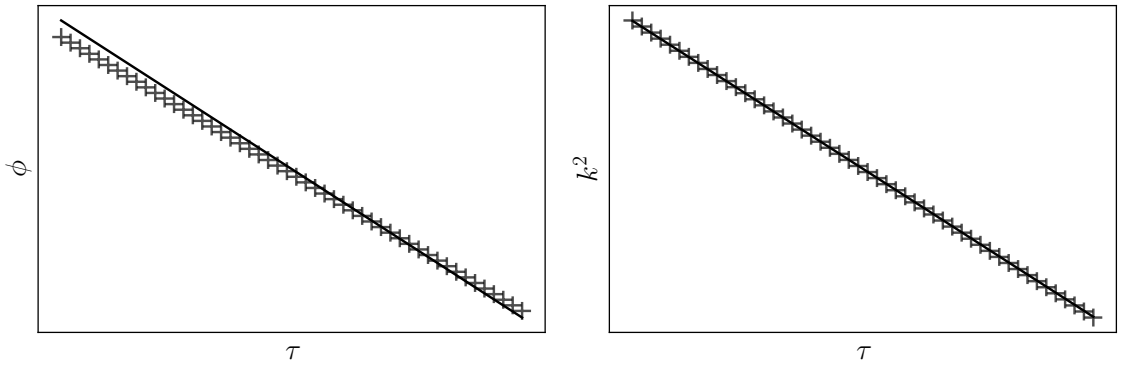
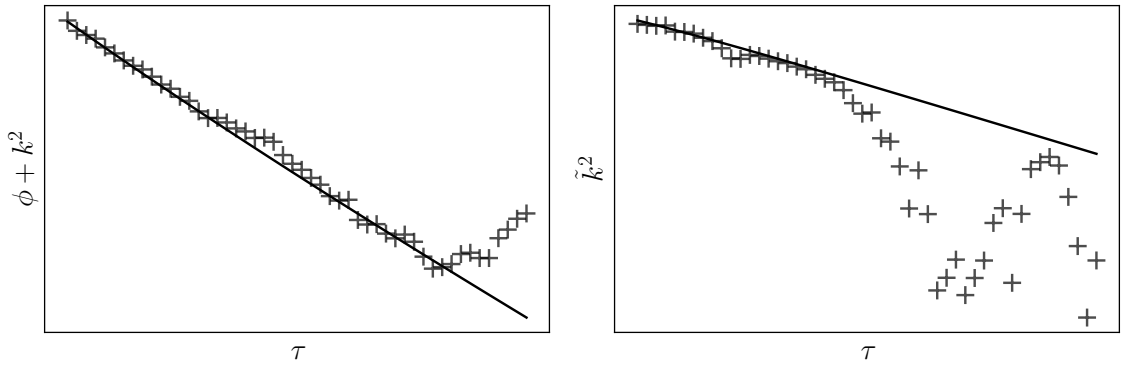


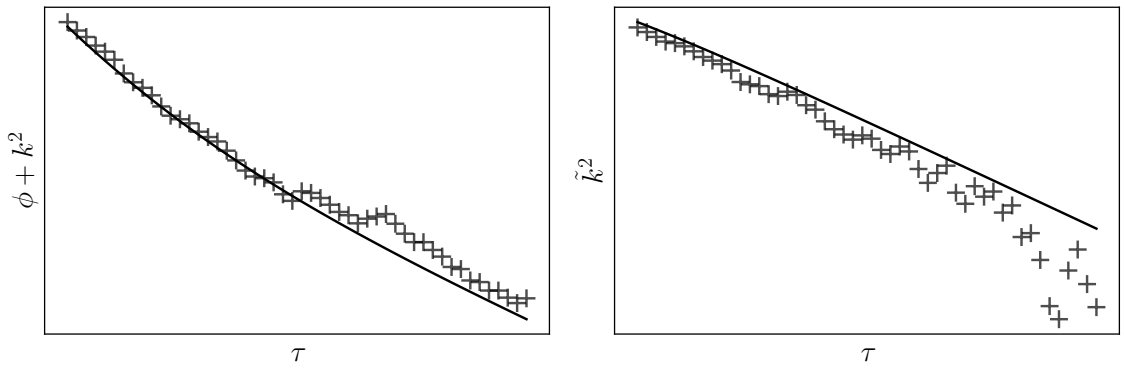
Figure B.5: Calibration of a 1D QHawkes with method of moments combined with a parametric minimisation from a simulation with parameters $n_H = 0.6$, $\beta = 0.04\ \text{min}^{-1}$, $n_Z = 0.2$, $\omega = 0.03\ \text{min}^{-1}$, $\lambda_\infty = 0.003\ \text{min}^{-1}$, hence $\bar{\lambda}\mathrm{d}t = 0.015$. The left panel presents the kernel ϕ. The right panel present the kernel k. In both panels the scatter plot is the optimised kernel obtained from the method of moments, while the plain line is the theoretical kernel with input parameters, we want to determine.

minimisation are illustrated in Figure B.5. Note that the obtained values of n_H, n_Z, β and ω can then serve as initial guesses to perform an estimation by maximum likelihood as described in Appendix B.3.

For the quadratic case, it is essential to remain in the regime where $\bar{\lambda}\mathrm{d}t \ll 1$; otherwise, the calibration results deviate significantly from the input values, and the method of moments, even when combined with maximum likelihood estimation, does not yield reliable parameters. Since, for real data at the 1-minute scale, we observe that $\bar{\lambda}\mathrm{d}t \gg 1$, properly estimating the QHawkes model as it stands is very challenging. Therefore, following the comparison and the approximation made in [70], we have turned to the QGARCH model introduced in [12]. Indeed, the method of moments to calibrate the QGARCH behaves properly for a large set of parameters as shown by Figure B.6.



(a) Calibration of a QGARCH for $\bar{\lambda} = 0.7$, obtained with $\sigma_\infty^2 = 0.1$



(b) Calibration of a QGARCH for $\bar{\lambda} = 12$, obtained with $\sigma_\infty^2 = 1.5$

Figure B.6: Proof of concept - Calibration of a 1D quadratic GARCH with exponential kernels. The scatter plot are the components obtained by the method of moments and the plain line are the theoretical components used to generate the synthetic data and that we want to recover. The left panel presents the diagonal of the kernel $K = \phi + k^2$ up to a lag $q = 50$. The right panel presents the rank one decomposition of K considering its diagonal has only zeros. For this simulation, we set $n_H = 0.7$, $n_Z = 0.2$, $\beta = 0.06$, $\omega = 0.05$ and σ_∞^2 varies between the two sub-panels.

B.8.2 Calibration of the Multivariate QGARCH

Since retrieving the univariate QHawkes from aggregated data is already challenging, for the multivariate case, we present only the calibration results on synthetic data for the MQGARCH model as a proof of concept. We first examine the 2D linear GARCH, followed by the 2D quadratic GARCH model.

Calibration of 2D linear GARCH

We now consider 2 assets, A and B, with volatility σ_A and σ_B and returns r_A and r_B. The MQGARCH volatility of the asset $i \in \{A, B\}$ is driven by

$$\sigma_{i,t}^2 = \sigma_{i,\infty}^2 + \sum_{j\in\{A,B\}} \sum_{k=1}^{+\infty} \phi_j^i(k) r_{j,t-k}^2 \quad \text{with returns defined as } r_{i,t} = \sigma_{i,t}\xi_{i,t}, \quad \text{with} \quad \xi_{i,t} \sim \mathcal{N}(0,1).$$

In the case of exponential kernels, for all $i, j \in \{A, B\}$, the kernel ϕ_j^i takes the form $\phi_j^i(k) = n_{Hij}\beta_{ij}\exp(-\beta_{ij}k)$. The goal of the calibration is then to determine the eight parameters $(n_{Hij}, \beta_{ij})_{i,j\in\{A,B\}}$. The method of moments works as before, and Appendix B.6 proposes a way to extend it intuitively for the multivariate case. Figure B.7 shows our calibration results for the set of parameters presented in Table B.2. It clearly demonstrates that the method of moments enables to recover the input kernels in the case of the 2D linear GARCH model.

	n_H	β
X_A^A	0.8	0.2
X_B^A	0.2	0.3
X_A^B	0.3	0.3
X_B^B	0.7	0.1

Table B.2: Input parameters to generate the synthetic 2D GARCH as defined above. Additionally, we set $\sigma^2_{A,\infty} = 0.05$ and $\sigma^2_{B,\infty} = 0.1$.

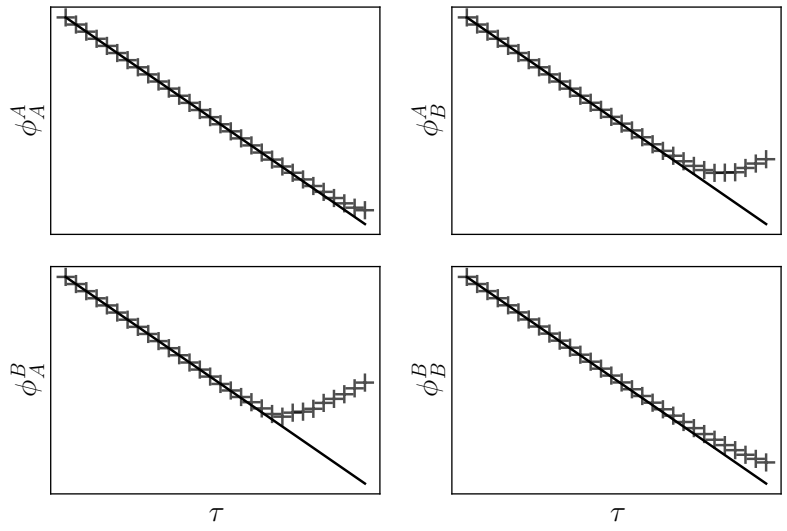


Figure B.7: Proof of concept - Calibration of a 2D linear GARCH with exponential kernels. The scatter plots are the kernels obtained by the method of moments and the plain lines are the kernels used to generate the synthetic data and that we want to recover. For this simulation, the parameters are set as defined in Table B.2.

Calibration of 2D quadratic GARCH (without cross trends feedback)

This last section focuses on our model of interest the multivariate quadratic GARCH. The goal is to provide, in a simple case, a proof of concept that the method of moments can be extended to this case and give consistent results.

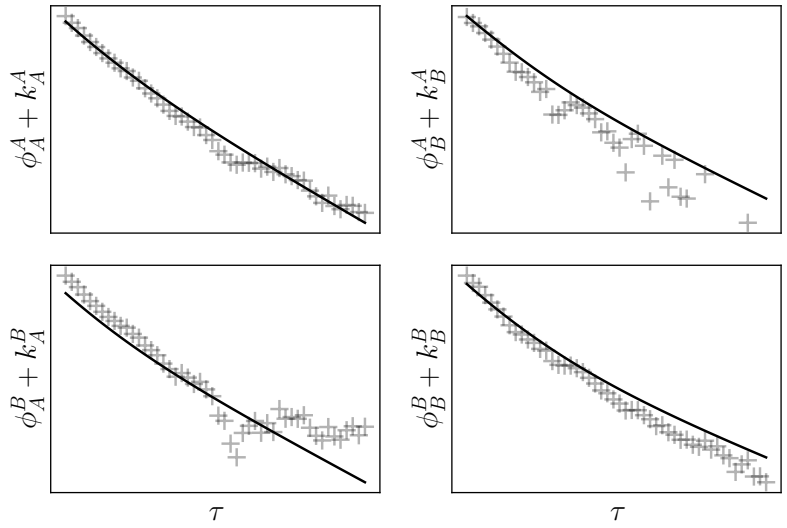
Here, we consider two assets A and B whose volatility $(\sigma_i)_{i\in\{A,B\}}$ is driven by

$$\sigma^2_{i,t} = \sigma^2_{i,\infty} + \sum_{j\in\{A,B\}} \sum_{k_1=1}^{+\infty} \sum_{k_2=1}^{+\infty} K_j^i(k_1,k_2) r_{j,t-k_1} r_{j,t-k_2},$$

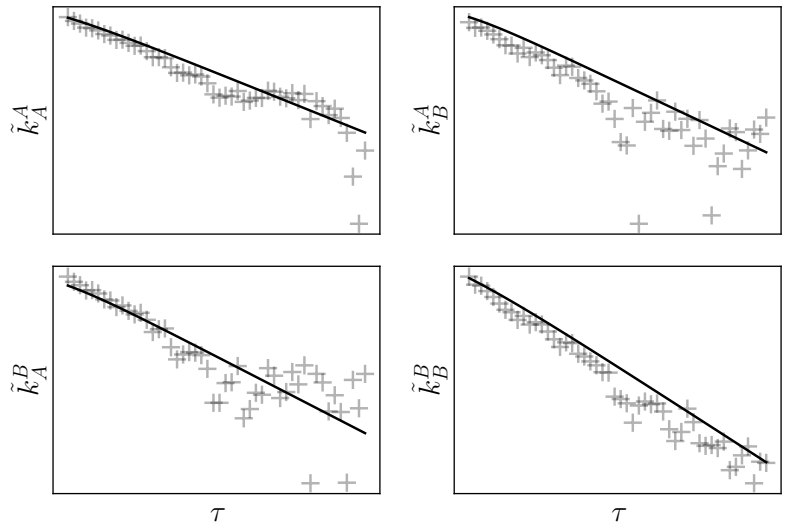
with returns defined as $r_{i,t} = \sigma_{i,t}\xi_{i,t}$, with $\xi_{i,t} \sim \mathcal{N}(0,1)$.

Note that, for the sake of simplicity, the leverage and cross-trends feedback (characterised by L and $K_\times$ in Equation (4.2.2)) are considered null in this framework. As for the univariate quadratic case above (Section B.8.1), we can adopt the ZHawkes framework [70]. Therefore, each kernel K_j^i for $i,j \in \{A,B\}$ can be decomposed as the sum of a diagonal contribution and a rank-one component, meaning that for all $i,j \in \{A,B\}$, K_j^i writes $K_j^i(\tau_1,\tau_2) = \phi_j^i(\tau_1)\delta_{\tau_1-\tau_2} + k_j^i(\tau_1)k_j^i(\tau_2)$.

For better representation, as for the univariate case, each kernel K_j^i is considered as a matrix defined by its lags, up to a certain lag q. Thus, K_j^i is represented by the symmetric matrix $(\mathbb{K}_j^i(\tau_1,\tau_2))_{1\le\tau_1,\tau_2\le q}$, where, in the ZHawkes framework, the diagonal is $(\phi_j^i(\tau) + (k_j^i)^2(\tau))_{1\le\tau\le q}$ and the upper triangle entries are $(k_j^i(\tau_1)k_j^i(\tau_2))_{1\le\tau_1<\tau_2\le q}$. As before, the method of moments (see Appendix B.7) returns, for all $i,j \in \{A,B\}$, the estimated diagonal and upper triangle values of $\mathbb{K}_j^i$ up to a certain lag q. Subsequently, the off-diagonal of each kernel $\mathbb{K}_j^i$ can be approximated with a rank-one representation, i.e. a vector $\tilde{k}_j^i$ such that $(\mathbb{K}_j^i(\tau_1,\tau_2))_{1\le\tau_1,\tau_2\le q,\tau_1\neq\tau_2} = \tilde{k}(\tau_1)\tilde{k}(\tau_2)$.



(a) Calibration of the diagonal of the kernels, for $i, j \in \{A, B\}$, $K_j^i = \phi_j^i + (k_j^i)^2$ up to a lag $q = 50$.



(b) Calibration of the rank one decomposition of off-diagonal kernels, for $i, j \in \{A, B\}$, $(K_j^i(\tau_1, \tau_2))_{1 \leq \tau_1, \tau_2 \leq q, \tau_1 \neq \tau_2} = \tilde{k}_j^i(\tau_1)\tilde{k}_j^i(\tau_2)$ up to a lag $q = 50$.

Figure B.8: Proof of concept - Calibration of a 2D quadratic GARCH with exponential kernels, with no leverage or cross trend feedback. The scatter plots are the kernels obtained by the method of moments and the plain lines are the kernels used to generate the synthetic data and that we want to recover. For this simulation, the parameters are set as defined in Table B.3.

To provide a proof of concept of the calibration of the 2D QGARCH, we generate synthetic time series of the QGARCH according to this framework, meaning that we generated two time series of volatility along with two time series of returns such that

$$\sigma_{i,t}^2 = \sigma_{i,\infty}^2 + \sum_{j \in \{A,B\}} \sum_{k=1}^{+\infty} \phi_j^i(k) r_{j,t-k}^2 + \sum_{k_1=1}^{+\infty} \sum_{k_2=1}^{+\infty} k_j^i(k_1) k_j^i(k_2) r_{j,t-k_1} r_{j,t-k_2},$$

with returns defined as $r_{i,t} = \sigma_{i,t}\xi_{i,t}$ and $\xi_{i,t} \sim \mathcal{N}(0,1)$, and where kernels take the form $\phi_j^i(\tau) = n_{H\,ij}\beta_{ij}\exp(-\beta_{ij}\tau)$ and $k_j^i(\tau) = \sqrt{2n_{Z\,ij}\omega_{ij}}\exp(-\omega_{ij}\tau)$. The value of the parameters for the simulation are described in Table B.3. Subsequently, the method of moments was applied as described in Appendix B.7 considering $L = \phi_\times = k_\times = 0$. The results of such calibration are shown Figure B.8.

Several observations can be made. First, the calibration appears to successfully replicate the

	n_H	β (min^{-1})	n_Z	ω (min^{-1})
X_A^A	0.6	0.06	0.2	0.07
X_B^A	0.1	0.1	0.15	0.1
X_A^B	0.2	0.08	0.1	0.09
X_B^B	0.4	0.04	0.21	0.06

Table B.3: Input parameters to generate the synthetic 2D QGARCH as defined above. Additionally, we set $\sigma^2_{A,\infty} = 1.2$ and $\sigma^2_{B,\infty} = 0.8$.

exponential shape and amplitude of the input kernels. However, it is evident that the calibration results do not perfectly align with the input kernels, indicating room for further improvement. Enhancements could involve increasing the number of lags q set in the method of moments, as larger values tend to yield more stable results but also require more computational time. Additionally, another approach to improve the calibration results is to utilize the method of moments' outcome as the initial point for optimization by Maximum Likelihood, as detailed in Appendix B.3, similar to the procedure used in univariate quadratic Hawkes calibration.

B.9 Kernels profiles from 2D-QGARCH Calibration on futures on indices

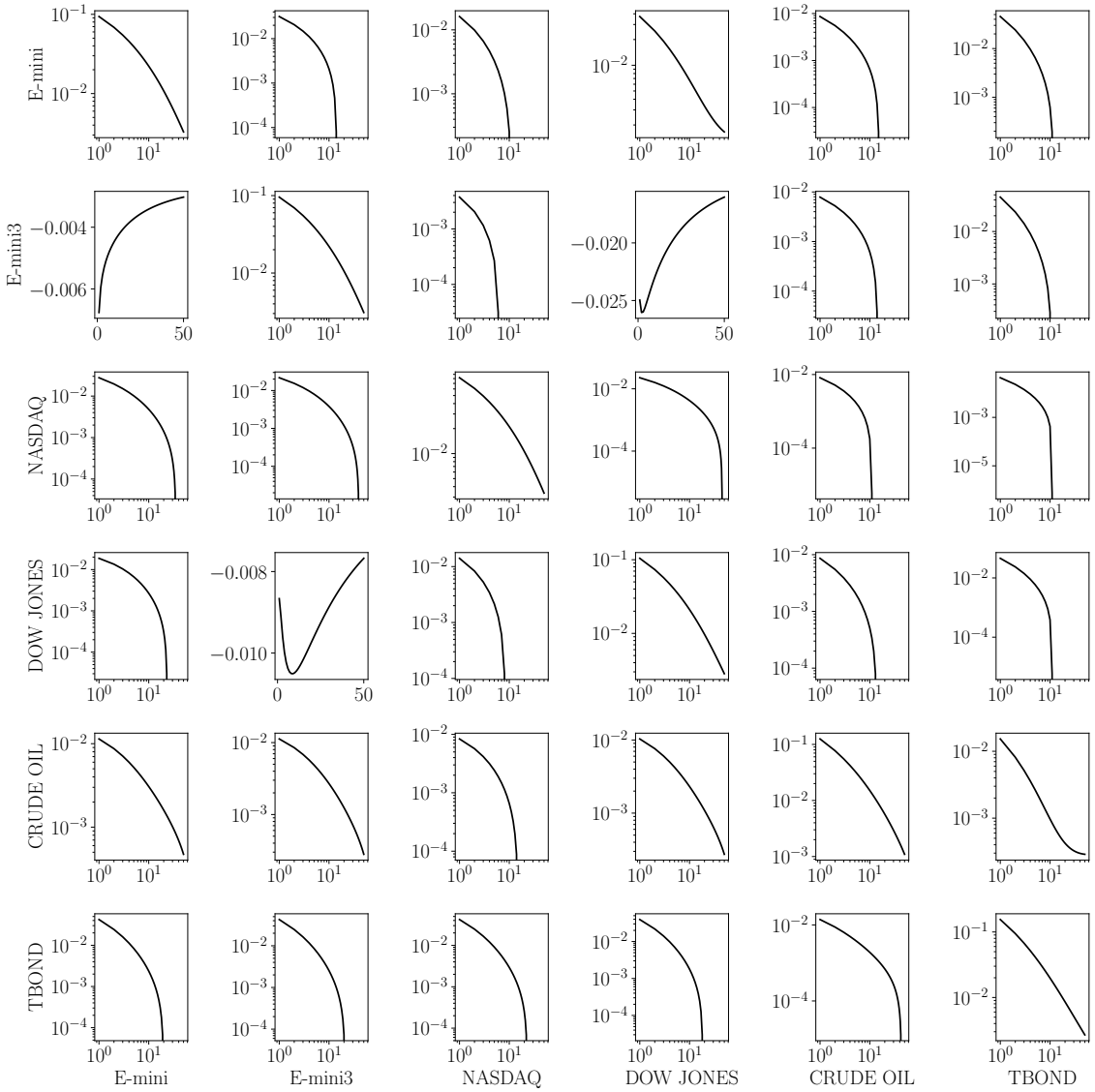


Figure B.9: (ϕ_i^j) profiles – the x-labels determine the index providing feedback and the y-labels determine the index receiving the feedback, i.e., for a kernel ϕ_i^j, i is labelled on the x-axis while j is labelled on the y-axis.

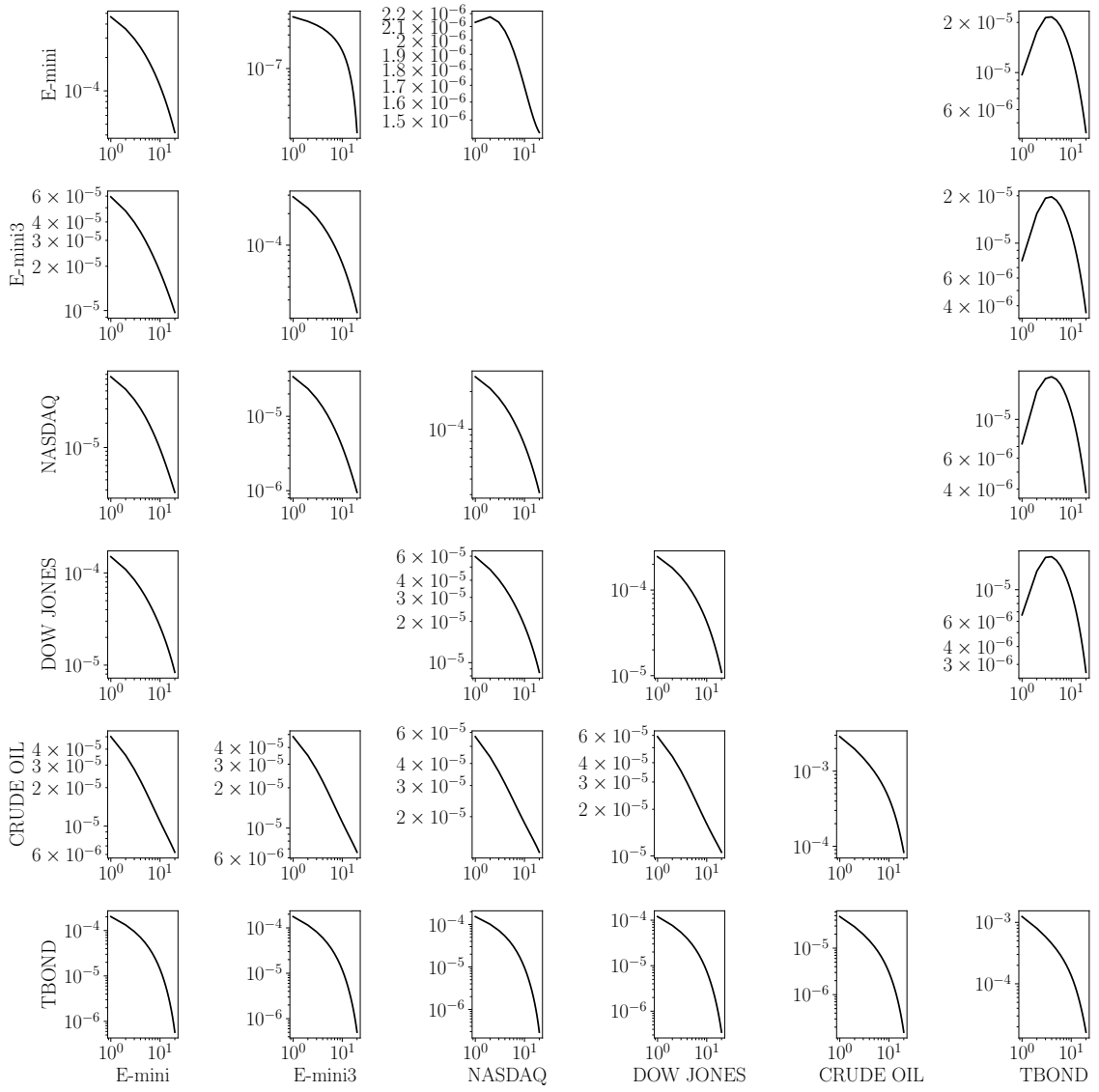


Figure B.10: $(k_i^j)^2$ profiles – the x-labels determine the index providing feedback and the y-labels determine the index receiving the feedback, i.e., for a kernel $(k_i^j)^2$, i is labelled on the x-axis while j is labelled on the y-axis.

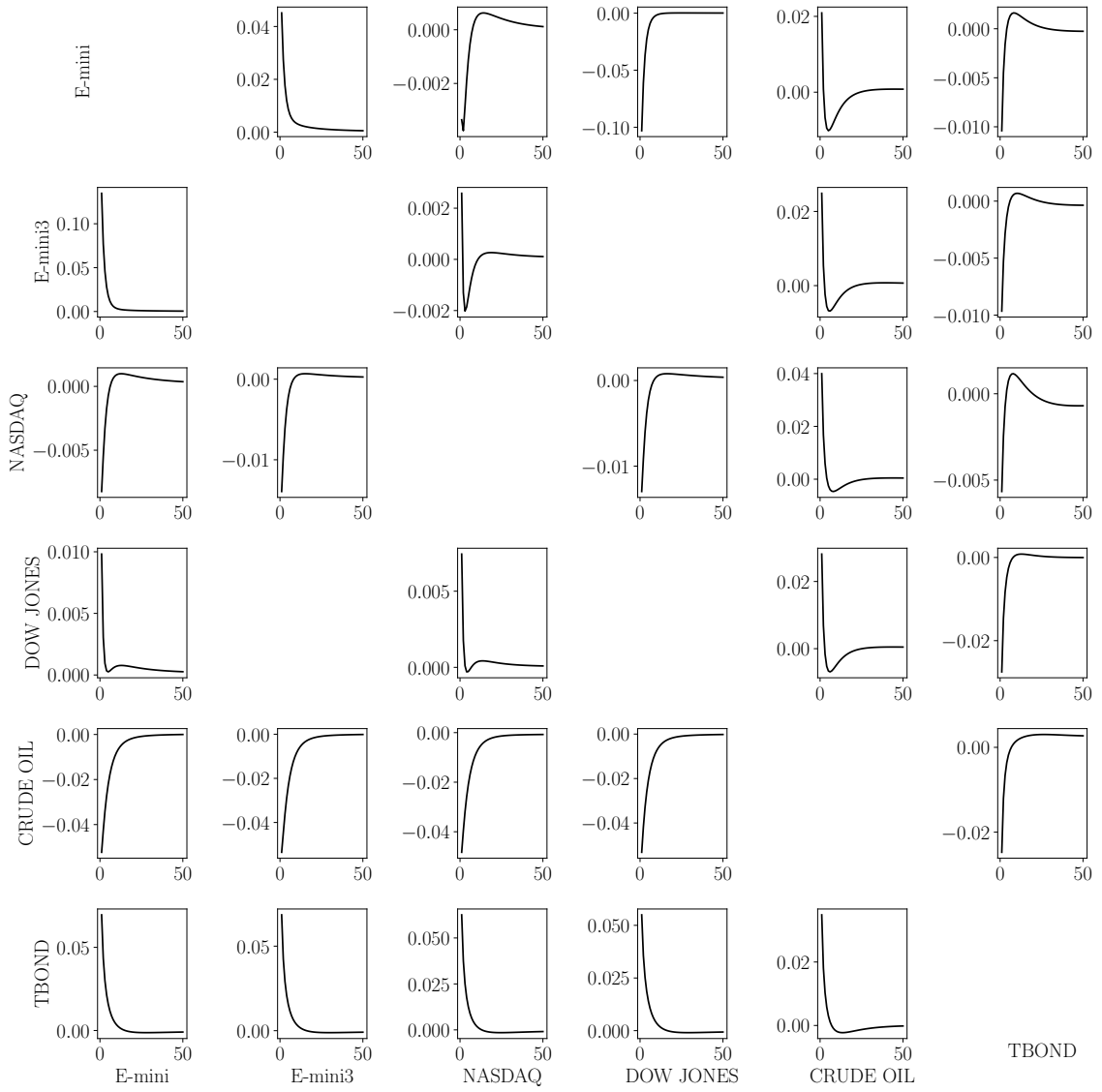


Figure B.11: $\phi^j_\times$ profiles – the y-labels determine the index receiving the feedback while the x-labels determine the index used for the correlation, i.e., the subplot (i, j) show the profile of the kernel $(\phi^i_\times)$ for the correlation between asset i and asset j.

Figure B.12: L_j^i profiles – the x-labels determine the index providing feedback and the y-labels determine the index receiving the feedback, i.e., for a kernel L_i^j, i is labelled on the x-axis while j is labelled on the y-axis.

Appendix C

Appendix of Chapter 5: Riding Wavelets: A Method to Discover New Classes of Price Jumps

C.1 Benchmark: Validation through synthetic

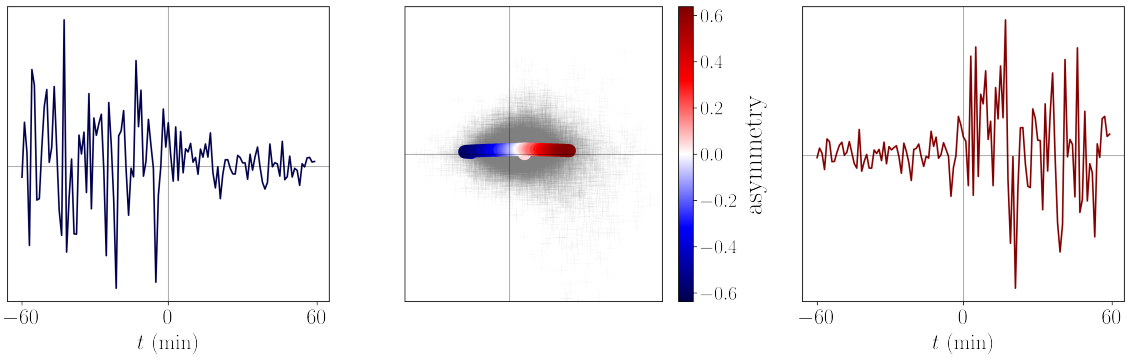


Figure C.1: Middle: Projection of our dataset on the reflexive direction D_1 (horizontal axis) and mean-reverting direction $\widetilde{D}_2$ (vertical axis). Benchmark jumps projection moves from left (anticipatory) to right (exogeneous). Left and right figures show two extreme benchmark time-series.

In order to verify that the D_1 direction indeed measures reflexivity, we create synthetic time-series with volatility profiles of varying time-asymmetry and apply our classification. Relying on [6], we construct jump time-series using the power law representation of Equation (5.1.1), with $t_c = -0.5$min, and $d = 0.5$. We adjust the parameters (N_r, N_l, p_r, p_l) to render the asymmetry of the signal. The time-series are then multiplied by a Gaussian noise (the same noise for all time-series). We then compute the features $\Phi(x)$ of each time-series x and project it on the 2D space formed by our time-asymmetry and mean-reversion directions. Fig. C.1 shows these projections, the color code corresponds to the asymmetry parameter. It clearly appears that the D_1 direction measures the time-asymmetry of the volatility.

C.2 Heterogeneous Near-Critical Branching Processes

Consider a simple branching process where a single event can trigger on average φ new events. It is well known that when $\varphi \to 1$, very large "avalanches" of events can occur. In fact, the probability that an avalanche of total size S is triggered by a single event takes, in the limit $\varepsilon = 1 - \varphi \to 0$, the following scaling form:

$$P(S|\varepsilon) \propto S^{-3/2} e^{-\varepsilon^2 S}. \tag{C.2.1}$$

Now suppose that the proximity to the critical point $\varphi = 1$ is itself random, reflecting the time varying fragility of the market and/or the intrinsic propensity of a shock to propagate across stocks. We will assume for simplicity that ε has uniform distribution between $\varepsilon_{\min}$ and 1. The observed distribution of avalanche sizes (in our case co-jump sizes) is then given by the following mixture:

$$P(S) = \frac{1}{1-\varepsilon_{\min}} \int_{\varepsilon_{\min}}^{1} \mathrm{d}\varepsilon \, P(S|\varepsilon) \propto S^{-3/2} \int_{\varepsilon_{\min}}^{1} \mathrm{d}\varepsilon \, e^{-\varepsilon^2 S}. \tag{C.2.2}$$

After a change of variable, the integral over ε can be rewritten as

$$\int_{\varepsilon_{\min}}^{1} \mathrm{d}\varepsilon \, e^{-\varepsilon^2 S} = \frac{1}{\sqrt{S}} \int_{\varepsilon_{\min}\sqrt{S}}^{\sqrt{S}} \mathrm{d}u \, e^{-u^2}$$

Now, in a intermediate regime where $S \gg 1$ but $\varepsilon_{\min}\sqrt{S} \lesssim 1$, the integral is close to $\sqrt{\pi}/2$, and one finally finds

$$P(S) \propto S^{-1-\tau} F(\varepsilon_{\min}\sqrt{S}), \qquad \tau = 1, \tag{C.2.3}$$

with $F(x)$ decreasing fast as $x \uparrow$. Hence this simple model predicts $\tau = 1$ (i.e. a Zipf law) for co-jump sizes, truncated beyond $S \sim \varepsilon_{\min}^{-2}$. From the data shown in Fig 5.18, we estimate $\varepsilon_{\min} \sim 0.1$. In other words, the market does not have to be poised extremely close to criticality to explain a broad power-tail for the co-jump size distribution.

Note finally that we could relax the hypothesis that the distribution of ε is strictly uniform. In the scaling regime, one only needs this distribution to be constant in the vicinity of $\varepsilon_{\min}$. The calculation above can be extended to cases where the distribution of ε is of a power law type close to zero, i.e. behaves as ε^γ for $\varepsilon \to 0$. In this case, one finds $P(S) \propto S^{-1-\tau}$ with $\tau = 1 + \gamma/2$.

C.3 Correlation of jump profiles in a co-jump

In this appendix, we investigate to which extent the different price profiles in a co-jump are correlated to each other. To achieve this, we consider the average correlation of the trend score Eq. (5.2.7) among the jumps in a given co-jump of size S, defined as

$$\rho = \frac{\sum_{k\neq k'=1}^{S} \widetilde{D}_3(x_k)\widetilde{D}_3(x_{k'})}{(S-1)\sum_{k=1}^{S}[\widetilde{D}_3(x_k)]^2}.$$

Fig. C.2 shows that the larger the co-jump, the more correlated are its constituents, although the effect is weak. Jumps affecting the market in its entirety are more likely to have a common external reason (exogenous) and lead to the same profile. Strongly correlated co-jumps come down to a single jump time-series which can be accessed through the average of normalized jumps:

$$\langle \overline{x} \rangle (t) = \left\langle \sigma_k^{-1} \overline{x}_k(t) \right\rangle_k$$

where $\sigma_k = \left\langle x_k^2(t) \right\rangle_t^{\frac{1}{2}}$. Fig. C.3a shows an example of co-jump, of size 83 with correlation $\rho = 0.96$. We see that the average time-series $\langle \overline{x} \rangle$ is non-zero for $t \neq 0$.

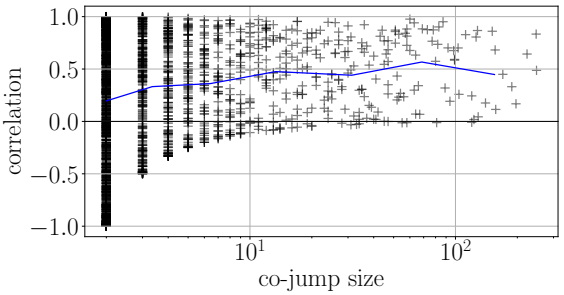
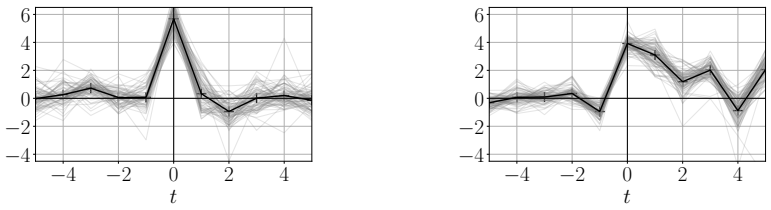


Figure C.2: Correlation measure ρ of a co-jump. Blue line represents average over bins of co-jumps of roughly the same size. As expected, the larger the co-jump the more correlated. Surprisingly, there remain weakly correlated ($\rho \approx 0$) large co-jumps.

In line with the discussion of section 5.3 about contagion-driven co-jumps, Fig. C.2 shows that there persist large co-jumps whose average correlation is close to zero, i.e. co-jumps composed of return time-series that are weakly correlated and that have no a priori reason to jump together, except through contagion. A typical example is shown on Fig. C.3b, it is of size 81 and correlation $\rho = 0.06$. As we can see, averaging its different jumps makes small sense since the average $\langle \overline{x} \rangle$ is zero for $t \neq 0$ (up to the variance). Far from being reduced to a single jump profile, such weakly large co-jumps could still be described by a small number of "hidden" profiles, depending on the "dimensionality" of the co-jump. Determining such dimensionality and hidden profiles would require applying a decomposition per co-jump, which would require more data on large co-jumps.



(a) Strongly correlated

(b) Weakly correlated

Figure C.3: Average profile of two co-jumps. The average (black curve) is taken over the profiles of the jumps involved in each co-jump (gray curves). Left: a strongly correlated co-jump that exhibits a non-zero average profile. Each jump time-series in the co-jump is a variation around this average profile. Right: a weakly correlated co-jump which has no meaningful average. The co-jumps are of size 83 and 81 respectively.

Appendix D

Appendix of Chapter 6: Why is stock volatility so much rougher than index volatility?

D.1 Sum of two log-SfBMs

D.1.1 2nd moment computation

we investigate the roughness of the following signal:

$$\mathrm{d}P_t = \exp\left(\Omega\left(t\right)\right)\mathrm{d}W_t + \exp\left(\tilde{\omega}\left(t\right)\right)\mathrm{d}B_t,$$

where Ω and $(\tilde{\omega})$ are independent SfBM defined by their Hurst exponents, respectfully, H and $\tilde{h}$, such that $\tilde{h} \ll H$, and intermittent parameters respectfully, ν and $\tilde{\nu}$. W_t and B_t are independent Brownian motions.

Based on the characterisation in Equation (6.1.2), to inquire into the roughness of the resulting signal, we want to find $\hat{h}$ such that

$$\begin{aligned} m(2,\Delta) =& \mathbb{E}\left(\left|\log\left(\exp\left(2\Omega(\Delta)\right)+\exp\left(2\tilde{\omega}(\Delta)\right)\right)-\log\left(\exp\left(2\Omega(0)\right)+\exp\left(2\tilde{\omega}(0)\right)\right)\right|^2\right) \\ =& \mathbb{E}\left(\left|\log\left(X_\Delta+Y_\Delta\right)-\log\left(X_0+Y_0\right)\right|^2\right) \\ =& K_2\Delta^{2\hat{h}} \quad \text{with } K_2 \text{ constant} \end{aligned} \tag{D.1.1}$$

where we introduce $X_t = \exp(2\Omega(t))$ and $Y_t = \exp(2\tilde{\omega}(t))$.

To develop $m(2,\Delta)$ in such form, we use the replica trick [126], which consists in writing the logarithm as follow:

$$\log(X) = \lim_{n\to 0}\frac{X^n-1}{n}.$$

Then, Equation (D.1.1) writes

$$m(2,\Delta) = \lim_{n\to 0}\frac{1}{n^2}\mathbb{E}\left(\left|(X_\Delta+Y_\Delta)^n-(X_0+Y_0)^n\right|^2\right) = \lim_{n\to 0}\frac{1}{n^2}\mathscr{E}(\Delta,n), \tag{D.1.2}$$

where we write $\mathscr{E}(\Delta,n) = \mathbb{E}\left(\left|(X_\Delta+Y_\Delta)^n-(X_0+Y_0)^n\right|^2\right)$

By stationarity, $\mathbb{E}((X_\Delta+Y_\Delta)^{2n})$ and $\mathbb{E}((X_0+Y_0)^{2n})$ behave similarly. We here treat the calculus of $\mathbb{E}((X_\Delta+Y_\Delta)^{2n})$. Using the Newton formula and the independence of X and Y, we can write:

$$\mathbb{E}((X_\Delta+Y_\Delta)^{2n}) = \sum_{i=0}^{2n}\binom{2n}{i}\mathbb{E}(X_\Delta^i Y_\Delta^{2n-i}) = \sum_{i=0}^{2n}\binom{2n}{i}\mathbb{E}(X_\Delta^i)\mathbb{E}(Y_\Delta^{2n-i})$$

Using the Appendix D.1.2 (below), we have:

$$\mathbb{E}\left((X_\Delta + Y_\Delta)^{2n}\right) = \mathbb{E}\left((X_0 + Y_0)^{2n}\right) = \sum_{i=0}^{2n} \binom{2n}{i} \mathbb{E}(X_\Delta^i)\mathbb{E}(Y_\Delta^{2n-i}) = \sum_{i=0}^{n}\sum_{j=0}^{n} \binom{n}{i}\binom{n}{i} \mathbb{E}(X_\Delta^{n+i-j})\mathbb{E}(Y_\Delta^{n-i+j}).$$

We now look at the cross product in the general case, for any integers m, n,

$$\mathbb{E}((X_\Delta + Y_\Delta)^n (X_0 + Y_0)^m) = \sum_{i=0}^{n}\sum_{j=0}^{m} \binom{n}{i}\binom{m}{j} \mathbb{E}(X_\Delta^i X_0^{m-j})\mathbb{E}(Y_\Delta^{n-i} Y_0^j). \tag{D.1.3}$$

Relying the properties of SfBM, we have:

$$\mathrm{cov}(\Omega(\Delta), \Omega(0)) = \mathbb{E}(\Omega(\Delta)\Omega(0)) - \mathbb{E}(\Omega(\Delta))\mathbb{E}(\Omega(0)) = \frac{\nu^2}{2}(T^{2H} - \Delta^{2H})$$

and so

$$\mathrm{cov}(2i\Omega(\Delta), 2j\Omega(0)) = \frac{4ij\nu^2}{2}(T^{2H} - \Delta^{2H}).$$

Given $X_t = \exp(2\Omega(t))$, we then have

$$\mathbb{E}(X_\Delta^i X_0^j) = \mathbb{E}(X_\Delta^i)\mathbb{E}(X_0^j)\exp(\mathrm{cov}(2i\Omega_\Delta, 2j\Omega_0)) = \mathbb{E}(X_\Delta^i)\mathbb{E}(X_0^j)\exp(\frac{4ij\nu^2}{2}(T^{2H} - \Delta^{2H})).$$

Similarly, for $\tilde{\omega}$, we can write

$$\mathbb{E}(Y_\Delta^i Y_0^j) = \mathbb{E}(Y_\Delta^i)\mathbb{E}(Y_0^j)\exp(\mathrm{cov}(2i\tilde{\omega}_\Delta, 2j\tilde{\omega}_0)) = \mathbb{E}(Y_\Delta^i)\mathbb{E}(Y_0^j)\exp(\frac{4ij\tilde{\nu}^2}{2}(T^{2\tilde{h}} - \Delta^{2\tilde{h}})).$$

Hence, plugging in Equation (D.1.3) for $n = m$, we obtain

$$\mathbb{E}((X_\Delta + Y_\Delta)^n (X_0 + Y_0)^n) = \sum_{i=0}^{n}\sum_{j=0}^{n} \binom{n}{i}\binom{n}{j} \mathbb{E}(X_\Delta^{n+i-j})\mathbb{E}(Y_\Delta^{n-i+j})\exp(-2(n-i)j\tilde{\nu}^2\Delta^{2\tilde{h}})\exp(-2i(n-j)\nu^2\Delta^{2H}).$$

We can now work on $\mathscr{E}(\Delta, n)$,

$$\begin{aligned}
\mathscr{E}(\Delta, n) =& 2\mathbb{E}((X_\Delta + Y_\Delta)^{2n}) - 2\mathbb{E}((X_\Delta + Y_\Delta)^n (X_0 + Y_0)^n) \\
=& 2\mathbb{E}((X_\Delta + Y_\Delta)^{2n}) - 2\mathbb{E}((X_\Delta + Y_\Delta)^n (X_0 + Y_0)^n) \\
=& 2\sum_{i=0}^{n}\sum_{j=0}^{n} \binom{n}{i}\binom{n}{j} \mathbb{E}(X_\Delta^{n+i-j})\mathbb{E}(Y_\Delta^{n-i+j})\Big(1 - \exp(-2(n-i)j\tilde{\nu}^2\Delta^{2\tilde{h}})\exp(-2i(n-j)\nu^2\Delta^{2H})\Big) \\
=& 2\sum_{i=0}^{n}\sum_{j=0}^{n} \binom{n}{i}\binom{n}{j} \mathbb{E}(X^{n+i-j})\mathbb{E}(Y^{n-i+j})\Big(1 - \exp(-2j(n-i)\tilde{\nu}^2\Delta^{2\tilde{h}} - 2i(n-j)\nu^2\Delta^{2H})\Big)
\end{aligned}$$

Note that we still have $m(2, \Delta)$ goes to 0, when Δ tends to 0.

We now work in the limit $\Delta \to 0$ with $\tilde{h} < H$, we first develop up to $o(\Delta^{2\tilde{h}})$

$$\begin{aligned}
\mathscr{E}(\Delta, n) =& 4\tilde{\nu}^2\Delta^{2\tilde{h}} \sum_{i=0}^{n}\sum_{j=0}^{n} \binom{n}{i}\binom{n}{j} \mathbb{E}(X^{n+i-j})\mathbb{E}(Y^{n-i+j}) j(n-i) + o(\Delta^{2\tilde{h}}) \\
=& 4\tilde{\nu}^2\Delta^{2\tilde{h}} \sum_{i=0}^{n-1}\sum_{j=1}^{n} \binom{n}{i}\binom{n}{j} \mathbb{E}(X^{n+i-j})\mathbb{E}(Y^{n-i+j}) j(n-i) + o(\Delta^{2\tilde{h}}) \\
=& 4\tilde{\nu}^2\Delta^{2\tilde{h}} \sum_{i=0}^{n-1}\sum_{j=1}^{n} \binom{n}{n-i}\binom{n}{j} \mathbb{E}(X^{n+i-j})\mathbb{E}(Y^{n-i+j}) j(n-i) + o(\Delta^{2\tilde{h}}) \\
=& 4\tilde{\nu}^2\Delta^{2\tilde{h}} \sum_{i=0}^{n-1}\sum_{j=1}^{n} \frac{n}{n-i}\binom{n-1}{n-1-i}\frac{n}{j}\binom{n-1}{j-1} \mathbb{E}(X^{n+i-j})\mathbb{E}(Y^{n-i+j}) j(n-i) + o(\Delta^{2\tilde{h}}) \\
=& 4\tilde{\nu}^2\Delta^{2\tilde{h}} n^2 \sum_{i=0}^{n-1}\sum_{j=1}^{n} \binom{n-1}{n-1-i}\binom{n-1}{j-1} \mathbb{E}(X^{n+i-j})\mathbb{E}(Y^{n-i+j}) + o(\Delta^{2\tilde{h}}) \\
=& 4\tilde{\nu}^2\Delta^{2\tilde{h}} n^2 \sum_{i=0}^{n-1}\sum_{j=0}^{n-1} \binom{n-1}{i}\binom{n-1}{j} \mathbb{E}(X^{n-1+i-j})\mathbb{E}(Y^{n-i+j+1}) + o(\Delta^{2\tilde{h}}) \\
=& 4\tilde{\nu}^2\Delta^{2\tilde{h}} n^2 \mathbb{E}(Y^2) \sum_{i=0}^{n-1}\sum_{j=0}^{n-1} \binom{n-1}{i}\binom{n-1}{j} \mathbb{E}(X^{n-1+i-j})\mathbb{E}(Y^{n-1-i+j}) + o(\Delta^{2\tilde{h}}) \\
\mathscr{E}(\Delta, n) =& 4\tilde{\nu}^2\Delta^{2\tilde{h}} n^2 \mathbb{E}(Y^2)\mathbb{E}\left((X+Y)^{2(n-1)}\right) + o(\Delta^{2\tilde{h}}).
\end{aligned}$$

Note that one could develop up to $o(\Delta^{2H})$ assuming $\tilde{h} < H < 2\tilde{h}$, one would obtain:

$$\begin{aligned}
\mathscr{E}(\Delta, n) =& 2\sum_{i=0}^{n}\sum_{j=0}^{n} \binom{n}{i}\binom{n}{j} \mathbb{E}(X^{n+i-j})\mathbb{E}(Y^{n-i+j})\big(2j(n-i)\tilde{\nu}^2\Delta^{2\tilde{h}} + 2i(n-j)\nu^2\Delta^{2H} + o(\Delta^{2H}) \\
=& 4\tilde{\nu}^2\Delta^{2\tilde{h}} \sum_{i=0}^{n}\sum_{j=0}^{n} \binom{n}{i}\binom{n}{j} \mathbb{E}(X^{n+i-j})\mathbb{E}(Y^{n-i+j}) j(n-i) \\
&+ 4\nu^2\Delta^{2H} \sum_{i=0}^{n}\sum_{j=0}^{n} \binom{n}{i}\binom{n}{j} \mathbb{E}(X^{n+i-j})\mathbb{E}(Y^{n-i+j}) i(n-j) + o(\Delta^{2H}) \\
\mathscr{E}(\Delta, n) =& 4\tilde{\nu}^2\Delta^{2\tilde{h}} n^2 \mathbb{E}(Y^2)\mathbb{E}((X+Y)^{2(n-1)}) + 2\nu^2\Delta^{2H} n^2 \mathbb{E}(X^2)\mathbb{E}((X+Y)^{2(n-1)}) + o(\Delta^{2H}).
\end{aligned}$$

Finally, plugging back in Equation (D.1.2), one obtains

$$\begin{aligned}
m(2, \Delta) &= \lim_{n\to 0} \frac{1}{n^2}\mathbb{E}\left(|(X_\Delta + Y_\Delta)^n - (X_0 + Y_0)^n|^2\right) \\
&= \lim_{n\to 0} 2\tilde{\nu}^2\Delta^{2\tilde{h}}\mathbb{E}(Y^2)\mathbb{E}((X+Y)^{2(n-1)}) + o(\Delta^{2\tilde{h}}) \\
m(2, \Delta) =& 4\tilde{\nu}^2\Delta^{2\tilde{h}}\mathbb{E}(Y^2)\mathbb{E}\left((X+Y)^{-2}\right) + o(\Delta^{2\tilde{h}}).
\end{aligned}$$

Or, yet again if we develop up to Δ^{2H},

$$\begin{aligned}
m(2, \Delta) &= \lim_{n\to 0} \frac{1}{n^2}\mathbb{E}\left(|(X_\Delta + Y_\Delta)^n - (X_0 + Y_0)^n|^2\right) \\
&= \lim_{n\to 0} 4\tilde{\nu}^2\Delta^{2\tilde{h}}\mathbb{E}(Y^2)\mathbb{E}\left((X+Y)^{2(n-1)}\right) + 4\nu^2\Delta^{2H}\mathbb{E}(X^2)\mathbb{E}((X+Y)^{2(n-1)}) + o(\Delta^{2H}) \\
m(2, \Delta) =& 4\mathbb{E}\left((X+Y)^{-2}\right)\left(\tilde{\nu}^2\Delta^{2\tilde{h}}\mathbb{E}(Y^2) + \nu^2\Delta^{2H}\mathbb{E}(X^2)\right) + o(\Delta^{2H})
\end{aligned}$$

All cases of $m(2k, \Delta)$ where $k \in \mathbb{N}$ can be worked out similarly. One can also develop a similar argument for $m(2k+1, \Delta)$, $k \in \mathbb{N}$ and generalized for all $m(k, \Delta)$, $k \in \mathbb{R}$ by density.

D.1.2 combinatory trick

$$
\begin{aligned}
D &= \sum_{i=0}^{n}\sum_{j=0}^{n}\binom{n}{i}\binom{n}{j} \\
&\quad \text{set } k = n+i-j \\
&= \sum_{j=0}^{n}\sum_{k=n-j}^{2n-j}\binom{n}{k+j-n}\binom{n}{j} \\
&= \sum_{k=0}^{n}\sum_{j=n-k}^{n}\binom{n}{j}\binom{n}{k-n+j} && + \sum_{k=n+1}^{2n}\sum_{j=0}^{2n-k}\binom{n}{j}\binom{n}{k-n+j} \\
&\quad \text{we set } j = j-(n-k) \\
&= \sum_{k=0}^{n}\sum_{j=0}^{k}\binom{n}{n-k+j}\binom{n}{j} && + \sum_{k=n+1}^{2n}\sum_{j=0}^{2n-k}\binom{n}{j}\binom{n}{k-n+j} \\
&\quad \text{we use: } \binom{n}{k} = \binom{n}{n-k} \\
&= \sum_{k=0}^{n}\sum_{j=0}^{k}\binom{n}{n-(n-k+j)}\binom{n}{j} && + \sum_{k=n+1}^{2n}\sum_{j=0}^{2n-k}\binom{n}{j}\binom{n}{n-(k-n+j)} \\
&= \sum_{k=0}^{n}\sum_{j=0}^{k}\binom{n}{k-j}\binom{n}{j} && + \sum_{k=n+1}^{2n}\sum_{j=0}^{2n-k}\binom{n}{j}\binom{n}{2n-k-j} \\
&\quad \text{we use: } \sum_{j=0}^{k}\binom{r}{j}\binom{s}{k-j} = \binom{r+s}{k} \\
&= \sum_{k=0}^{n}\binom{2n}{k} && + \sum_{k=n+1}^{2n}\binom{2n}{2n-k} \\
&= \sum_{k=0}^{n}\binom{2n}{k} && + \sum_{k=n+1}^{2n}\binom{2n}{k} \\
&= \sum_{k=0}^{2n}\binom{2n}{k}
\end{aligned}
$$

D.2 log-SfBM with 2 modes of volatility

We now investigate the roughness of the signal

$$\mathrm{d}P_t = \exp\left(A_0\Omega(t) + \tilde{\omega}(t)\right)\mathrm{d}W_t$$

where Ω and $\tilde{\omega}$ are two SfBM of Hurst exponent H and, respectfully, $\tilde{h}$ and $(W_t)_{t>0}$ is a Brownian motion. As before, we compute the second moment of log-volatility increments. It gives:

$$\begin{aligned}
m(2,\Delta) =& \mathbb{E}\left(\left(\left(2A_0\Omega(\Delta) + 2\tilde{\omega}(\Delta)\right) - \left(2A_0\Omega(0) + 2\tilde{\omega}(0)\right)\right)^2\right) \\
=& \mathbb{E}\left(\left(\left(2A_0\Omega(\Delta) - 2A_0\Omega(0)\right) + \left(2\tilde{\omega}(\Delta) - 2\tilde{\omega}(0)\right)\right)^2\right) \\
=& 4\mathbb{E}\big(\left(A_0\Omega(\Delta) - A_0\Omega(0)\right)^2\big) + 4\mathbb{E}\big(\left(\tilde{\omega}(\Delta) - \tilde{\omega}(0)\right)^2\big) + 8\mathbb{E}\big(A_0\Omega(\Delta) - A_0\Omega(0)\big)\mathbb{E}\big(\tilde{\omega}(\Delta) - \tilde{\omega}(0)\big) \\
=& 4A_0^2K_2\nu^2\Delta^{2H} + 4\tilde{k}_2\tilde{\nu}^2\Delta^{2\tilde{h}}
\end{aligned}$$

Thus,

$$m(2,\Delta) = 4\Big(A_0^2K_2\nu^2\Delta^{2H} + \tilde{k}_2\tilde{\nu}^2\Delta^{2\tilde{h}}\Big) \quad \text{with} \quad k_q = K_q = \frac{2^{q/2}\Gamma(\frac{q+1}{2})}{\sqrt{\pi}}$$

D.3 How to simulate a fractional Brownian motion (and derived processes)

This appendix aims at give some intuitions on fractional Brownian motions (and more generally gaussian processes defined by their auto-correlations) and how they are built. The method described here is the method introduced in Dietrich and Newsam (1997) [129]. We simply present here the main steps because we believe it helps understanding how these processes are built.

The goal is to simulate a stationary Gaussian process $Y(\tau)$ with 0 mean and variance 1. The process is defined by its correlations (translation invariant) such that

$$\mathbb{E}(Y(t)Y(t+\tau)) = r(|\tau|)$$

with some function r.

D.3.1 Create the correlation matrix R

The first step consists in building the correlation matrix R over a time grid $\{t_0, t_1, ..., t_N\}$, such that $R_{ij} = r(|t_i - t_j|)$. Note that R is a symmetric Toeplitz matrix[1], and as such, is entirely defined by its first row $R_0 = (r_0, ..., r_N)$, with $r_j = r(|t_0 - t_j|)$. t_N can be seen as the correlation range we consider.

Remark D.3.1. Note that at this stage, theoretically, generating ϵ such that $\epsilon \sim \mathcal{N}(0, I_N)$ and using the Choleski decomposition $R = AA^\top$, would allow to generate Y with $Y = A\epsilon$. However the Choleski decomposition is known to be computationally very expensive, hence, the use of the present method.

D.3.2 Create the circulant matrix S

The next step consist in creating the $(2N \times 2N)$ symmetric Toeplitz matrix S, such that:

$$\begin{cases} s_k = r_k, & \text{for} \quad k = 0, ..., N \\ s_{2N-k} = r_k, & \text{for} \quad k = 1, ..., N-1. \end{cases}$$

Hence, the first row of S is given by $S_0 = (r_0, ..., r_{N-1}, r_N, r_{N-1}, r_{N-2}, ..., r_2, r_1)$ and is enough to defined the whole matrix S (as we build it as a symmetric Toeplitz matrix). Two things are interesting about S:

- any block along the diagonal of S is a copy of R
- S is circulant[2]

.

Since S is circulant, its eigen-decomposition is known and writes $S = \frac{1}{2N} F \Lambda F^\top$, with F being the Fourier transform matrix (with entries $F_{lk} = \exp(i\frac{2\pi lk}{2})$) and $\text{diag}(\Lambda) = FS_0$. Note that, multiplying a vector v by F is a discrete Fourier transformation of v [130].

D.3.3 Generate Y

From the results above, one can then generate Y the following way:

1. Compute R
2. From R, build S
3. Compute Λ such that $\Lambda_k = (FS_0)_k$, where F is the Fast Fourier Transform matrix
4. Generate $\epsilon = \epsilon_{\text{Re}} + i\epsilon_{\text{Im}}$ such that $\epsilon_{\text{Re}}, \epsilon_{\text{Im}}$ are independent Gaussian vectors, $\epsilon_{\text{Re}}, \epsilon_{\text{Im}} \sim \mathcal{N}(0, I_{2N})$

[1] **Toeplitz matrix**: matrix in which each descending diagonal from left to right is constant.

[2] **Circulant matrix**: matrix in which all rows are composed of the same elements and each row is rotated one element to the right relative to the preceding row. It is a particular case of Toeplitz matrix.

5. Build vector $\tilde{e}$ such that $\tilde{e}_k = \epsilon_k \left(\frac{\Lambda_k}{2N}\right)^{1/2}$
6. Then, any $N+1$ subset of consecutive entries of $e = F\tilde{e}$ (from either real or imaginary part) is a realisation of Y

We give below an example on how to implement it in Python.

D.3.4 Python Algorithm

We write here a way to generate `Y`, a stationary gaussian process, defined by its correlations given in vector `cov` of size N (`cov` is the realisation of R_0).

Algorithm 7 Simulation of Y

Computation of S
```
cov = np.concatenate((cov, np.flip(cov[1:-1])), axis=0)
```
computation of Λ
```
L = np.fft.fft(cov)[None, :]
```
Generation of ϵ
```
ϵ = np.random.randn(2 * N - 2) + 1j * np.random.randn(2 * N - 2)
```
Computation of e
```
e = np.fft.fft(z * np.sqrt(L / (2 * N - 2)), axis=-1).real
```
Selection of the subset of size N
```
Y = e[:, :N]
```

ECOLE
DOCTORALE

Titre : Déchiffrer les chocs des marchés financiers : Etude des Dynamiques Endogènes de Volatilité dans des Marchés Interconnectés

Mots clés : endogéneité, multi-dimensions, Hawkes, volatilité, microstructure des marchés

Résumé : Le rôle des mécanismes de rétroaction dans la création d'instabilités sur des marchés financiers a largement été étudié dans la littérature. L'endogénéité de la volatilité et de l'activité des marchés a effectivement conduit à plusieurs krachs notoires. Le plus connu, le flash crash du 6 mai 2010, illustre comment les instabilités du marché découlent de ses caractéristiques intrinsèques. En effet, c'est une exécution excessivement rapide d'ordres de vente qui a déclenché la chute rapide suivie d'une remontée du SPmini en moins d'une heure. De plus, les instabilités des marchés sont accentuées par leur nature multidimensionnelle et leur connectivité, comme le montre la propagation de la volatilité à travers divers actifs financiers lors d'événements tels que le flash crash mentionné ci-dessus. En effet, le 6 mai 2010, le flash crash du SPmini a affecté 300 autres actifs. Cette thèse étudie empiriquement et théoriquement l'endogénéité des mouvements de prix en multi-dimension. Dans un premier temps, nous cherchons à caractériser les sauts de prix empiriques. Basés sur des recherches interdisciplinaires montrant que l'asymétrie temporelle peut être utilisée pour classer les sauts d'activité comme exogènes ou endogènes, nous développons une nouvelle méthode, non supervisée, basée sur des coefficients d'ondelettes (particulièrement adaptés pour refléter l'asymétrie temporelle) afin de mesurer la réflexivité des sauts de prix univariés. Par ailleurs, notre représentation a révélé que le retour à la moyenne et l'alignement avec la tendance sont deux caractéristiques supplémentaires, permettant d'identifier de nouvelles classes de sauts. Enfin, cette représentation permet d'étudier les propriétés réflexives des co-sauts, définis par des sauts de prix simultanés (au cours de la même minute) de plusieurs actifs. Il apparaît qu'une fraction significative des co-sauts résulte d'un mécanisme de contagion endogène. Ainsi, l'événement du 6 mai n'était pas un incident isolé, et des dynamiques endogènes couplées avec une forte connectivité contribuent aux instabilités des marchés. Parallèlement, nous étudions le modèle du Hawkes quadratique (QHawkes), utilisé pour décrire la volatilité à haute fréquence. Les processus QHawkes sont des processus de Poisson qui, par l'expression de leur intensité, décrivent l'influence du passé sur la probabilité de l'activité future. Des travaux antérieurs ont montré que le modèle QHawkes univarié reproduit plusieurs caractéristiques empiriques : les queues épaisses des distributions, le regroupement temporel de la volatilité et l'asymétrie temporelle (effets de levier et effet Zumbach). En outre, des résultats supplémentaires sur la stabilité des processus de QHawkes sont discutés, montrant que la rétroaction quadratique peut induire des événements extrêmes tout en restant stable grâce à l'équilibre les réalisations inhibitrices et excitatrices. Pour étudier le caractère multidimensionnel des marchés, nous étendons le modèle QHawkes à plusieurs dimensions, en considérant plusieurs actifs et leurs interactions croisées. Un cadre multi-actifs nécessite la prise en compte de faits stylisés supplémentaires, tels que la prévalence des co-sauts et les effets d'asymétrie temporelle croisés. En effet, ce travail met en lumière les effets de levier et de Zumbach croisés. En développant deux modèles, nous montrons que le modèle QHawkes multivarié (MQHawkes) peut reproduire les faits empiriques observés sur les marchés financiers. La calibration du modèle sur des paires d'actifs confirme que les marchés sont au bord de l'instabilité. Pour compléter, un autre modèle multivarié de volatilité dépendante de la trajectoire est étudié : le Nested Factor Model avec des processus log-SfBM comme volatilités. Ce modèle permet de réconcilier les différences de rugosité entre les indices et les actions, offrant de nouvelles perspectives sur les dynamiques de la volatilité multivariée.

Title : Unraveling Financial Market Quakes: Exploring Endogenous Volatility Dynamics in Interconnected Markets

Keywords : endogeneity, multi-dimensions, Hawkes, volatility, market microstructure

Abstract : Past research has highlighted that feedback mechanisms underlie many financial markets instabilities. Endogenous dynamics of markets volatility and activity have indeed led to various notable crashes. Case in point: the events of May 6th, 2010, commonly referred to as the 2010 flash crash, exemplify how market instabilities stem from intrinsic features of financial markets. As evidence, an excessively rapid execution of sell orders triggered the rapid decline and subsequent recovery of the SPmini within the span of an hour. Moreover, market instabilities are compounded by their multidimensional nature and interconnectedness, as demonstrated by the propagation of volatility across diverse financial assets during events like the aforementioned flashcrash. Indeed, on May 6th, 2010, the SPmini flash crash affected 300 other assets alongside the SPmini. This thesis presents both a data-driven approach and a theoretical approach to investigate the endogenous nature of price movements within a multivariate framework. Our data-driven approach aims to characterize empirical price jumps. Leveraging interdisciplinary research suggesting that the time-asymmetry of activity can be used to classify bursts of activity as exogenous or endogenous, we develop a new unsupervised method based on wavelet coefficients (particularly suitable to reflect time asymmetry) to measure reflexivity of univariate price jumps. On top of that, our wavelet-based representation revealed that mean-reversion and trend are two additional key features, permitting identification of new classes of jumps. Furthermore, this representation allows to investigate the reflexive properties of co-jumps, defined by multiple stocks experiencing price jumps within the same minute. We argue that a significant fraction of co-jumps results from an endogenous contagion mechanism. Thus, May 6th event was not an isolated incident, and the interplay of endogenous dynamics alongside high levels of interconnectedness contributes to the instabilities observed within markets. Concomitantly, our theoretical inquiry focuses on the quadratic Hawkes (QHawkes) framework, originally introduced to describe volatility dynamics at tick-by-tick level. QHawkes processes are Poisson processes, which, through the expression of their intensity, depict the influence of the past on the probability of future activity. Previous work has proved that the univariate QHawkes model replicates several empirical features of financial time series, including fat tails of the returns' distribution, volatility clustering and the time asymmetry effects (leverage and Zumbach effects). Indeed, the supplementary quadratic and leverage feedback allow to overcome the limitations of the original (linear) Hawkes framework. Besides, additional results on the stability of QHawkes processes are discussed, showing that the quadratic feedback can induce extreme events while staying stable by balancing inhibitory and excitatory realizations. To explore market interconnectedness, we extend QHawkes processes into multidimensional settings, encompassing several assets and their cross-interactions. A multi-assets framework necessitates consideration of additional stylized facts, such as the prevalence of co-jumps and cross time asymmetry effects. Indeed, this work sheds light on the cross leverage and cross Zumbach effects. Developing two frameworks, we show that the multivariate QHawkes (MQHawkes) can reproduce the empirical facts observed in financial markets. Calibrating the model on asset pairs further confirms that markets operate on the brink of instability. To be thorough, another multivariate, path-dependent volatility model is studied: the nested factor model with log-SfBM processes as volatilities. Our findings suggest that this framework reconciles differences in roughness between indices and stocks, offering further insights into the dynamics of multivariate volatility.

Institut Polytechnique de Paris
91120 Palaiseau, France