Market Microstructure and Modeling of the Trading Flow

Khalil Antoine Dayri

ÉCOLE POLYTECHNIQUE
ParisTech

CMAP

Thèse de doctorat
Mathématiques Appliquées

pour obtenir le grade de
Docteur de l'École Polytechnique

Market Microstructure and Modeling of the Trading Flow

---

présentée par
Khalil Antoine al Dayri

sous la direction de
Emmanuel Bacry
Marc Hoffmann

soutenue publiquement le 16 janvier 2011

Jury

Rapporteurs:	Fabrizio Lillo
	Nour Meddahi
Examinateurs:	Robert Almgren
	Emmanuel Bacry
	Jean-Philippe Bouchaud
	Marc Hoffmann
Directeurs de thèse:	Emmanuel Bacry et Marc Hoffmann

Centre de Mathématiques Appliquées – UMR 7641

Contents

Acknowledgements

To the gods I am indebted for having good grandfathers, good parents, a good brother and sister, good teachers, good associates, good kinsmen and friends, nearly everything good. -Marcus Aurelius

My greatest gratitude goes to my advisors, Emmanuel Bacry and Mark Hoffmann for their exceptional guidance. They have accepted me as a PhD student without hesitation even though I had been outside the academic world for a few years. Their support, both academic and personal was invaluable for the completion of this work. To Emmanuel, I am grateful for his constant availability, truly day or night, and for his vast knowledge and recommendations in all the aspects of my work, be it mathematics, finance or programming. To Marc, I am grateful for his hands off guidance and for allowing me to pursue interests that were not always to his best liking. "Our middle ground", as he calls it, was clearly biased in my direction. To both of you, thank you for your pleasantness, your constant motivations, understanding, and most of all for your patience at times I would not have been so kind with myself.

To Jean-Francois Muzy, I extend my warmest regards and acknowledgements. Calling him my third advisor would not be an exaggeration. I thank you for your time and contribution, and also for your friendship and continuous support and motivations.

Robert Almgren, my former boss at Bank of America in NY contributed significantly to my decision to pursue my PhD. I especially thank him for his guidance and advice, during my tenure with him and beyond. During my PhD, I really appreciated our numerous and lengthy conversations, and the invaluable ideas he shared with me. Thank you for being the *gold standard* of mentors and role models.

I cannot thank Mathieu H. Rosembaum enough. One of the best things I am taking from this experience is my friendship with him. Thank you for always being

there in my time of need. I am really going to miss our long and sometimes surreal conversations while we were stuck in traffic on the way to Paris.

To the Financial Risks Chair of the Risk Foundation: thank you for your support and for part funding my PhD. Also, thanks to the GIS (X-ENSAE-HEC), for part funding my PhD.

I am honored to have Fabrizio Lillo and Nour Meddahi as my thesis examiners. Many thanks to Jean-Philippe Bouchaud for being in my jury. Thank you all for your time and effort, and for traveling to Paris to be present at my defense.

I also want to mention Charles-Albert Lehalle for introducing me to Emmanuel. Without his help, my career today would have been very different indeed. Thank you also for your numerous contributions and for your passion in market microstructure. Many thanks also for Frederique Abergel, Arthur Breitman and Carl Graham for the interesting discussions we had.

These last three years at Polytechnique have been extremely enjoyable largely because working atmosphere at CMAP. I was immediately taken in as a member of the family. To Nassera, Alexandra, Sylvain, Nathalie, Wallis, Sandra: thank you for making even the most exasperating problems bearable, and for always being there when I needed you.

I send my deepest recognitions to my fellow office partners, Camille, Clement, Isabelle, Emilie, Florent, Francisco, Sylvie, Soledad, and Zixian. You have made those three years extremely pleasant. Very special regards go to Julio Alexey Cen (aka Zhihao), for, well... being Julio.

To Laurent Duvernet, my thesis brother, thank you for your motivations and help, and for painting such a pleasant picture of life in a PhD under Emmanuel and Marc. Your recommendations were certainly a major factor in my pursuit.

To my CMAP colleagues, Anna, Chao, Gerogios, Gilles, Irina, Khaled, Trung-Lap, Leticia, Jutta, Maxime, Meisam, Michael, Olivier, Xavier, Xialu: thank you and good luck!

To my friendZ and roommates: his Excellency Gaby, Grand Master Georgez and Bash Mouhandiss BAssem... Thank you for making these long study years so pleasant. To Abir, Bassel, Chadi, Daniel, Maya, Mazen, you made my return to Paris very pleasant and enjoyable.

To my American friends, Akhil, Arthur, Avishek, Chandrima, Janet, Michel, Natalia and Radhika... Your support was a light during those very dark days.

To my godfather Hanna and to Andree: you truly were my second family ever since I came to France in 2000. Many thanks to Jean-Elie, Maryline and Gerome.

Finally, I would like to thank my family for their continuous love and support. My parents who have truly devoted themselves to me, my education, and well-being. My brother George, my sister Mays, who was there for me in the most difficult times and my cousins Bassel and Ziad for their support when I was in the US and when I came back to Paris. Thank you.

Résumé

On propose une perspective originale d'analyser les différents flux hautes fréquences d'information provenant des marchés financiers et fournit des modèles simples et intuitives qui reflètent étroitement la réalité. On observe les données empiriques et note certains faits stylisés et propose des modèles pour capturer ces faits.

Dans le chapitre 1, on passe en revue les définitions et propriétés de base des marchés électroniques. En particulier, on revoit les travaux de microstructure et de modélisation du marché, et leurs relations à ce travail. On introduit la taille du "tick", qu'on utilise pour classifier les actifs et interpréter les différents résultats.

Dans le chapitre 2, on montre empiriquement que l'impact d'une seule transaction dépend de la durée inter-transactions. En effet, lorsque le taux des échanges devient plus rapide, la variance des rendements des transactions augmente fortement et que ce comportement persiste à des échelles de temps plus grossières. On montre également que la valeur du spread augmente avec l'activité et on en déduit que les carnets d'ordres sont plus vide lorsque le taux des échanges est élevé.

Dans le chapitre 3, on présente un modèle pour capturer le bruit de microstructure. Les prix des actifs sont représentés par la somme des rendements "tick" arrivant à des temps de Poisson aléatoires. Le modèle se compose d'une martingale diffusive contaminée par un bruit autocorrélé mais disparaissant aux échelles grossières. On est capable de capturer la signature de la variance et l'autocorrélation faible mais significative des rendements "tick".

Dans le chapitre 4, on utilise les processus ponctuels de Hawkes pour modéliser l'arrivée aléatoire des transactions. On modélise la transformation échelle fine - échelle grossière des prix et en particulier l'effet sur les moments des rendements des prix. On propose une technique simple d'estimation non paramétrique de la structure de dépendance des processus de Hawkes dans le cas unidimensionnel et dans quelques cas particuliers multidimensionnels. On applique la méthode à des actifs de Future et trouve des noyaux de dépendance en loi de puissance.

Abstract

We offer an original way to analyse at the various high frequency streams of information originating from financial markets and to provide simple intuitive models that closely mirror reality. We observe empirical data and report some of its stylized facts and propose models to capture these facts.

In chapter 1, we review the basic definitions and properties of electronic exchanges. In particular, we review the background work done in microstructure and trade modeling, show how they relate to our work and introduce the tick size, used to classify our assets and interpret the various results.

In chapter 2, we bring qualitative empirical evidence that the impact of a single trade depends on the intertrade time lags. We find that when the trading rate becomes faster, the return variance per trade strongly increases and that this behavior persists at coarser time scales. We also show that the spread value is an increasing function of the activity and deduce that orderbooks are more likely empty when the trading rate is high.

In chapter 3, we present a model to capture microstructure noise. Asset prices are represented as the sum of tick returns arriving at random Poisson times. The model consists of an underlying diffusive martingale which is contaminated by some vanishing autocorrelated noise. We are able to capture the signature form of the sampled realized variance and the weak but significant autocorrelation of tick returns.

In chapter 4, we use Hawkes point processes to model the random arrival of trades in the market. We model fine to coarse behavior of prices and how it affects the moments of price returns. We propose a simple non parametric estimation technique of the dependence structure of Hawkes processes in the one dimensional case and very particular multidimensional cases. We apply the method to Futures assets and find decay kernels having a power law form.

CHAPTER 1

Introduction

1.1 Motivation, Topics and Thesis Axis

This thesis was originally motivated by a desire to offer an original way to look at the various streams of information originating from financial markets and to provide models that closely mirror reality. In order to achieve that, we have to explore financial markets and their manifestation through an exchange. We do this by empirically observing data and providing original stylized facts. Then we propose to capture these facts by models that are intuitive, easy to understand and conceptually simple. We shall see that these models capture rather well some of the empirical facts, and provide a satisfying intuitive interpretation of reality.

From a **modeling perspective**, our ultimate wish is to have models that are robust across observation scales. This means that the models should capture stylized facts of the price distribution when observed at different frequencies, from the highest (also called tick frequency) to lower ones (daily or even monthly). For instance, on the coarse scale, prices are fairly well represented by a continuous time Brownian semi-martingale. On the fine scale, this structure breaks down and some original stylized facts appear. So in order to capture the statistical particularities present at the fine scale, we need to start by observing and reporting them, then provide models that replicate them. However, the model should allow for the observed disappearance of theses stylized facts at coarse scale, and the morphing of the fine scale process into a coarse scale diffusion.

In this work, we attempt to give elements that contribute to this approach. Each chapter of this thesis has a place in this enterprise. We present below the main ideas behind every of them. Please refer to the section 1.6 later in this introduction, for a more detailed presentation of each work.

In chapter 2, we observe and report empirical properties at the fine scale and see how they change as the scale increases. In particular, we look at the distribution of prices conditionally to the intensity of trading. We examine if the variance of the return changes conditionally to the time it took for these returns to arrive. So **we answer the following question: is the realized variance created by 10 trades arriving over 10 seconds similar to the realized variance created by those very same trades had they arrived during 10 minutes?** Any model that uses a transaction time clock implies that the two situations are similar. Our empirical findings show that they are not, and that trades arriving in a shorter duration have higher variance, thus showing the importance of the physical inter-trade time duration. We observe this behavior at fine scale and see how it dissipates at coarser observation scales. This chapter has appeared as an article in "Econophysics of order-driven markets", *Springer Verlag*, co-authored with Emmanuel Bacry and Jean-Francois Muzy.

In chapter 3, we turn our attention to the problem of microstructure noise. In the spirit of building parsimonious models that represent reality, and inspired by the results of chapter 2, we provide a model that captures some of the stylized facts present at the fine scale, while converging to a diffusion at coarse scale. Those facts in particular are the signature form of the sampled realized variance and the weak but significant autocorrelation of **tick returns**[1]. We model the asset price as the sum of tick returns arriving at random Poisson times. In particular, the model consists of an underlying diffusive martingale which is contaminated by some vanishing autocorrelated noise. This type model, consisting of a latent price contaminated by some microstructure noise, have been widely studied in the past with various assumptions on the underlying and on the noise processes and have been proven to successfully capture the variance signature plot while converging to a diffusion on the coarse scale. However, **we provide a model that captures, in addition to the signature plot, different microstructure properties, such as the autocovariance of tick returns, the discrete nature of tick prices, and the random arrival of market orders, and provide an original point of view to interpret these properties**. This chapter will be submitted as an article for publication soon, with Mark Hoffmann as co-authors.

In chapter 4, we further move in the direction of Statistics of Processes. By now, we are convinced that the key for a fine to coarse modeling approach lies

[1] In our work, we call **tick return** the price differential from one trade to the next. We also exclusively use the term **return** for price differentials and do not consider log returns at all.

in the modeling of the arrival times of prices. This means using point processes to model the random arrival to the market of different observable elements - the trades for instance. In particular, self exciting point processes, known as Hawkes processes, have been gaining particular traction in finance lately. These are point processes whose instantaneous intensity is linearly dependent on their past events. They have been used to model the arrival of various market events, for instance microstructure events like the arrival of trades or limit orders. Lately, they have been used by Bacry, Delattre, Hoffmann and Muzy [9] to model fine to coarse behavior of the prices and the manifestation of such behavior through the moments of price returns (variance and correlation). Inspired by their usefulness and pertinence in representing financial events, **we propose a simple non parametric estimation technique of the dependence structure of Hawkes processes and apply it to futures price data**. This chapter will be submitted as an for publication soon, with Emmanuel Bacry and Jean-Francois Muzy as co-authors.

From a **presentation perspective**, we show and interpret our results by classifying the assets according to their tick size, which we quantify in section 1.3. This is a market microstructure parameter that qualifies the traders' aversion to price movements at the highest frequency. Surprisingly we have that this parameter is not entirely explained by liquidity or coarse scale volatility. We interpret most of our empirical results according to this criterion and we see that the market microstructure of different assets changes according to their tick size.

Let us now introduce the different terms and general framework we will consider throughout this thesis. We define the continuous electronic auction in section 1.2, going over the exchange framework and the orderbook. In section 1.3, we define what we mean by tick size and how it can influence the microstructure of assets. In section 1.4, we offer a reminder of previous work that has been done in modeling the market microstructure and see in what ways they have motivated our work. In section 1.5, we present the data we use in numerical applications. Finally, we introduce the main results of the thesis in section 1.6.

1.2 Continuous Auction and Electronic Markets

In this section, we present generalities about financial markets and the regulatory framework we consider in particular. Section 1.2.1 contains generalities that are true for any type of auction based market, and section 1.2.2 expands on it to give particularities about financial exchanges.

1.2.1 Continuous Auction

We consider the framework of a **continuous** electronic market where any market participant can freely enter and exit the market at any time. Market participants are free to submit or remove their preferences to the market (also called the exchange) directly and without going through a third party (as opposed to specialist markets for example [23]). Market participants, or **traders**, like in any other auction, enter the market by submitting their ultimate preference for a particular asset. If they are buying, they inform the community of the maximum quantity of assets they are willing to buy at a particular price or differently, the price they are willing to pay for any quantity below a maximum quantity. This price is called their **bid** price. Similarly, if they are selling, they inform the community of the minimum price they are willing to receive for a particular quantity. This price is called their **ask** price. When another market participant agrees to the presented price, both participants engage in a **trade**. Naturally, if a trader has an ask price which is lower than another trader's bid a trade will occur at some price they both agree to, causing the disappearance of both preferences from the marketplace. Once all agreeing preferences have been cleared, those who are left keep their preferences known and wait for someone to engage with them. This leads to the formation in the market of a maximum buy price and a minimum sell price, called respectively called the **best bid and ask** prices. The difference between the best bid and best ask is then always strictly greater than zero and is called the bid-ask **spread** and the average between these prices is called the **midpoint** price.

In an **electronic** market, all transmission of information is handled electronically. An electronic market typically operates at very high speeds, and has the capability of publishing large amounts of information in a very short time. This means that market participants are kept extremely well informed of the state of the market. They are then able to make informed trading decisions.

The exchange keeps track of all the information currently available in the market through a record called the orderbook. We discuss the orderbook in details in the following section.

1.2.2 The Orderbook

Limit Orders and Cancel Orders

A trader communicates with the market through the use of **orders**. For example, when a market participant wants to buy the quantity q at the maximum price p, he sends what is called a **buy limit order** of q shares at the price p. When he wants to sell, he sends a **sell limit order**. The set of preferences (all the limit orders, their prices and their quantities) of all market participants is called the **orderbook**. The orderbook contains each participants' buy or sell orders for

all price levels. For example, for any given price, the orderbook contains the total volume available for trading at that price level. When a new limit order arrives at this price, it gets added to the top of the queue by an order placement engine. Similarly, a participant can decide to cancel his order through a **cancel** order and the placement engine removes his entry from that queue.

Market Orders

When a participant sends an order to buy a certain quantity q regardless of the price, called a **market** order, a matching engine fills his order with the ask orders present at the best ask. The entire volume of the market order always gets executed even if it needs several matching ask limit orders to satisfy it. A similar situation can occur in the event of a buy limit order with a price higher or equal than the best ask. These are called **marketable limit** orders and we rarely distinguish them from market orders. However, it is interesting to note that in this case, only ask orders whose prices are less or equal than the buy limit price will get executed. If any volume of the buy limit order is left, it will be placed in the orderbook as a new limit order and effectively becoming the new best bid.

Without the arrival of news orders (limit, market and cancel), the market is frozen in time. No new information is relayed, except the trivial time duration elapsed since the last order. Therefore, the process of price formation is basically the result of the dynamics of these various types of orders and their interactions. Thus a theory of prices can be achieved by modeling the dynamics of orders, in other words modeling the dynamics of the orderbook. We will discuss this modeling approach in more details in section 1.4, after we have presented other pertinent information about the orderbook.

Order Matching Algorithm

In our work, most of the markets are First In First Out markets (**FIFO**). This means that there is a time priority rule in the matching of orders. In particular, this means that in the event of two market participants having similar bids or asks, execution priority is given to whomever made his preferences known first. The orderbook keeps track of the age of the limit orders on each price level and gives the older orders priority in execution. When a market order needs more than one limit order to be entirely filled, the oldest orders get filled first, followed by the newer ones. If all the limit orders at the best ask are filled, the matching starts with the oldest order of the immediately higher price, which has effectively become the new best ask. So, a succession of trades at increasingly higher prices is triggered, one for each different filled limit order, until all the totality of the market order's volume is filled.

This version is adopted by most electronic markets with some few exceptions. A

pro rata matching, where the orders are executed proportionally to their quantity, is also common (cf [1] for a detailed look at the different matching algorithms used at the Chicago Mercantile Exchange). However, these differences are of no importance to our presentation. In fact, when required, we only use the total number of shares available at a particular price level and do not pay any attention to their decomposition by limit orders. This is also reflected in the way we handle our data in section 1.5. Indeed, when one market order hits several limit orders it results in several trades being reported, one for each limit order counterpart. In practice, we aggregate together all such transactions and consider them as one trade whose size is the size of the market order and whose price is the price of the last limit order filled.

The Tick Value

The market fixes a price grid on which traders can place their prices. The grid step is the smallest interval between two prices and it is called the **tick value**. It is measured in the currency of the asset. For a given security, it is safe to consider this grid to be evenly spaced even though the market may change it at times. In some markets the spacing of the grid can depend on the price. For example, stocks trading on the Euronext in Paris have a price dependent tick scheme. Stocks priced 0 to 9.999 € have a tick size of 0.001 € but all stocks above 10 € have a tick of 0.005 €. In our work we made sure to choose assets whose tick value does not change with the price and we were careful to keep track of exchange rules changes regarding this variable.

Published Orderbook Snapshot

In general the exchange publishes snapshots of the orderbook as often as possible. What the traders receive as information varies with the rules of the exchange, but the most important thing is that they see the total amount of shares available at a significant number of price levels (**ticks**) below the best bid and above the best ask. Depending on the market, the exchange will typically publish the total number of shares available 5 or 10 price levels (with non zero volume) deeper than the best bid or ask. A picture of what an orderbook looks like for a market participant is shown in Figure 1.1. It shows a snapshot of the March 2009 DAX futures orderbook on February 16 2009. There are 10 price levels for buy limit orders and 10 price levels for sell limit orders. The grid where the prices can be put is shown clearly and the tick value in this case is easily seen to be 0.5 €. To make the figure easier to read, we display the negative of the buy limit quantities (blue bars on the left) while keeping the sell limit quantities (red bars on the right) positive. The best bid is the pair (3 shares, 4365.5 €) and the best ask is (3 shares, 4368 €), and the spread is 2.5 € or 5 ticks.

Within each price level, there might be several orders; we represent each limit

order by its own rectangle. The abscissa of the rectangle is the price of the limit order, while its length is the quantity of the limit order. At each price level, the total length of the bar is the total number of shares available to be bought or sold at the price level.

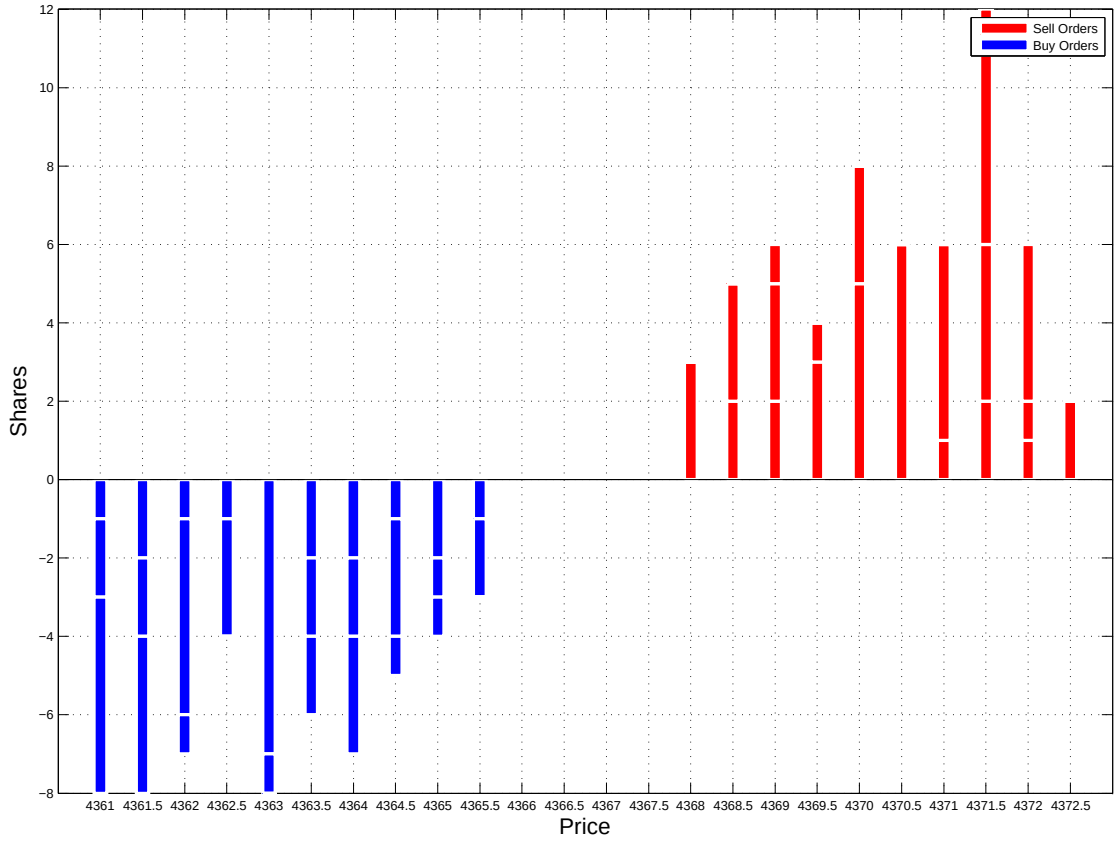


Figure 1.1: A graphical snapshot of the orderbook of the March 2009 DAX future contract on February 16 at time 13:22:48.165. The EUREX exchange only publishes the first 10 levels of each side the orderbook at any given time, and it does not publish the different orders composing each price levels. We show here a hypothetical distribution of different limit orders within each price level.

Liquidity

The **liquidity** of the market is a term used to describe the cost of trading in the market. **Direct Transaction costs** are defined as the cost differential between the actual price paid for a certain trade and the observed midpoint price at the moment the decision to trade was taken. A market where large quantities can be bought and sold with few transaction costs is called a liquid market. A small bid-ask spread, large quantities of shares available for trading, a deep orderbook, fast arrivals of new

limit orders; all reflect that the market is liquid and imply little transaction costs. More generally though, transaction costs are not limited to direct ones such as the ones implied by the spread or the depth of book, they also include the reaction of the market to any particular order. This is commonly known as **market impact** and we discuss it in more details in section 1.4.

Now that we have presented the most important market mechanisms and defined the market microstructure in which we place ourselves, we discuss the notion of tick size. It is a relative measure of the absolute tick value presented in section 1.2.2 and it is often seen as a qualitative notion that depends on the trader's aversion to a price movement of one tick.

1.3 Tick Size

We present in this section our thoughts on the tick size. We thought it pertinent to assign a whole section of the introduction to this notion and not put it in the previous section because of its importance as a general basis for the analysis of our data and results. Indeed, in chapters 2 and 3, we present and interpret the results as a function of the tick size. Furthermore, the tick size was used as an order relation to classify assets that trade on different markets and have different characteristics. In the following, we give an intuitive qualitative definition of the tick size and give our opinion as to why we think it is important to use it as a backbone for the analysis of our results. We then move to provide a quantitative definition of the tick size, which will allow us to use it as an order relation. Finally, we give descriptive statistics that show the effect of the tick size in practice.

As we explain in section 1.2.2, the tick value is a standard characteristic of any asset and is measured in its currency. It is the smallest increment by which the price can move. In all our work, all variations of transaction prices are normalized by the tick value and expressed in ticks (integers). Similarly, all variations of midpoint prices are expressed in half ticks (half integers).

As one can see in table 1.1, column **Tick Value**[2], our assets have very different tick values, corresponding to different contract sizes, and denominated in different currencies. In order to bring these assets on the same level, we qualitatively introduce the notion of **tick size**: A trader considers that an asset has a small tick when he "feels" it to be negligible, in other words, when he is not averse at all to price

[2]The Tick Value expressed in the table factors in the **lot size** i.e. the minimal number of shares that can be exchanged with one trade. That is why the tick of the DAX (lot size is 25 shares) is 12.5 € in table 1.1, while we said it was 0.5 € in section 1.2.2. This number reflects more accurately the minimal possible increment between two prices, as traders would have to exchange assets by the lot size and not by individual shares and we use it as the tick value throughout this work.

variations of the order of a single tick. In general then, the trader's perception of the tick is qualitative and empirical, and depends on many parameters such as the tick value, the price and the usual amounts traded in the asset and even his own trading strategy. All this leads to the counter-intuitive though very well known remark : the tick value is not an good absolute measure of the perceived size of the tick. It has to be viewed relatively to other market statistics. For instance, every trader "considers" that the ESX index futures has a much greater tick than the DAX index futures though the tick values are of the same orders (cf table 1.1).

We now give some quantitative aspect of the tick size. Indeed, there have been several attempts to quantify the perceived tick size. Kockelkoren, Eisler and Bouchaud in [33], write that "large tick stocks are such that the bid-ask spread is almost always equal to one tick, while small tick stocks have spreads that are typically a few ticks".

Following these lines, in chapter 2 of the thesis, we calculate the number of times (observed at times t_i) the spread is equal to 1 tick:

$$P_{=} = \frac{\#\{i, s_{t_i} = 1\}}{N} \tag{1.1}$$

where N is total number of trades in the sample, i represent the i^{th} trade, t_i its time, and s_{t_i} the value of the observed spread when the trade occurred. We show the results in table 1.1. According to this criterion, SP futures have the largest tick, with the spread equal to 1 99.8% of the time, whereas the DAX futures have the smallest tick. We use this criterion to classify our assets in part in chapter 2 of the thesis.

In a different approach, Robert and Rosenbaum in [76] introduce a model for ultra high frequency data with a parameter η quantifying the perceived tick size. They show that this parameter can easily be estimated in the following way. In a sample of non null price variations in ticks observed over the time $[0, T]$, they classify each price variation as either a k-alternation or a k-continuation ($k \in \mathbb{N}$). A k-continuation is a price variation of size k ticks whose direction is the same as the one of the preceding variation. A k-alternation is a price variation of size k ticks whose direction is opposite to the one of the preceding variation. If N_k^a is the number of k-alternations and N_k^c be the number of k-continuation, they define $\widehat{\eta}$, an estimator of η, by

$$\widehat{\eta} = \sum_{k=1}^{m} \lambda_k u_k \tag{1.2}$$

where m is the maximal absolute return variation observed in the sample and

$$u_k = \frac{1}{2}(k(\frac{N_k^c}{N_k^a} - 1) + 1)$$

are estimators of η for each k and

$$\lambda_k = \frac{N_k^a + N_k^c}{\sum_{k=1}^{m} N_k^a + N_k^c}$$

are weighting factors.

A simpler version of $\widehat{\eta}$ that uses only the variations of 1 tick can be used in general and this is the definition we use in chapter 2 and 3 of this work. In this case $\widehat{\eta}$ is reduced to:

$$\widehat{\eta} = \frac{N_1^c}{2N_1^a}. \tag{1.3}$$

The parameter η is intended to represent the aversion to price changes. If η is small, the price will be mainly doing alternations, also known as bid ask bounces. This translates into a large tick, because in this case market participants are more averse to changes in the midpoint price. If they are **market makers**[3] then they are satisfied by keeping the spread collapsed to the minimum (1 tick) and collecting the spread. If they are **market takers**[4] then a movement of only one price level is very significant for them and they are satisfied to keep trading at constant current levels. As this is put in [75], all market participants "feel more comfortable when the asset price is constant than when it is moving". So the best bid and ask price levels (and hence the midpoint price) would only move when it becomes clear that the current price level is unsustainable. Indeed, in section 2 we show that the probability of times the midpoint return is null is related to both the notion of a collapsed spread as defined in equation (1.1) and the η defined in equation (3.12).

To illustrate our thought, in table 1.1, we give for each asset during June 2009, an estimation of η and $P_=$ along with other descriptive statistics and see that the rankings of the assets using this criterion almost matches the ranking using [33]'s $P_=$ criterion.

[3]Market makers are patient traders who prefer to send limit orders and wait to be executed, thus avoiding to cross the spread but taking on volatility risk.

[4]Market takers are impatient traders who prefer to send market orders and get immediate execution, thus avoiding volatility risk but crossing the spread in the process.

Naturally, these observations have an important consequence for the asset's return distribution. Most importantly, the distribution function of a small tick asset has a much bigger support and fatter tails than that of a large tick asset (cf the difference in the probability distribution functions of the BUS5 futures against that of the SP futures for example in Figure 3.9). Other straightforward remarks that one can observe about the effect of the tick size on the distribution of returns include the increase in the probability of null midpoint returns (cf Figure 3.17) and the higher number of significant autocorrelation lags (cf Figure 3.15) as the tick size increases. This is a direct consequence of the increase in the proportion of alternations to continuations. Some of these properties were indeed noticed in [68] and require a more detailed presentation in a later work.

Finally, we point out that in the model for discrete prices studied in chapter 3, we come across a statistic that is conceptually very similar to η and behaves in a similar way as the tick size changes.

Throughout this work, some less straightforward influences of the tick size on the distribution of returns appear. Hence, we analyze and present our finding in this scope. We refer to chapter 2 and 3 for more details as it is too early to discuss these findings at this point.

1.4 Background and Models

A lot of information is contained in the orderbook and studying it opened a world of quite new research [20]. This research is intimately tied to the theory of transaction costs [19]. For example, the bid-ask spread is seen as a transaction cost, since the sender of a market order would have to pay a premium for *instant* execution. The depth of the orderbook is also a transaction cost, since the sender of a large market order would have to pay a premium for *full* execution. As we discussed in section 1.2.2-paragraph **Liquidity**, the most straightforward costs of trading constitute direct transaction costs that the trader can observe and assess at the moment of execution. However, a trader is also wary of how the market will react to his own intervention in the market, and consequently, the future value of his trades.

One question that always haunts any trader is "what will happen after I execute a trade"? Indeed, while the trader does worry about direct transaction costs (see section 1.2.2) and tries his best to reduce them, the value of the asset he holds depends on the dynamics of the market after his trade is completed. When the trader intervenes in the market, his action in itself is a new information that the market will incorporate. Hence, the market will behave differently after his intervention from how it would have behaved otherwise. So he is keen to know or even control

the future behavior after his trade, and minimize the effect on the market caused by his own trading. This overall difference in market behavior is called **market impact**. In particular, it can be quantified as a difference in prices, between the midpoint price at the time of the intervention, and the midpoint price at some time in the future. Seen in this light, the market impact is a transaction cost.

1.4.1 Transaction Impact Models

During the past decade, the explosion of the amount of available data associated with electronic markets has permitted important progress in the description of price fluctuations at the microstructure level. In particular the pioneering works of Farmer, Lillo et al. [54, 38, 43, 37] and Bouchaud et al. [22, 21, 33] relying on the analysis of orderbook data, have provided new insights in the understanding of the complex mechanism of price formation (see e.g [20] for a recent review). A central quantity in these works and in most approaches that aim at modeling prices at their microscopic level, is the market impact function, which describes the average response of prices to "events" (trade, limit and cancel orders). Indeed, the price value of some asset is obtained from its cumulated variations caused by the (random) action of sell/buy market orders. In that respect, the price dynamics is formulated as a discrete "trading time" model like:

$$p_n = \sum_{i<n} G(n-i, V_i)\varepsilon_i + diffusion \tag{1.4}$$

where n and i are transaction "times", i.e., integer indices of market orders. V_i is the quantity traded at index i, ε_i is the sign of the i^{th} market order ($\varepsilon_i = -1$ if selling and $\varepsilon_i = +1$ if buying). The function $G(k, V)$ is the bare impact corresponding to the average impact after k trades of a single trade of volume V (cf [41] for a detailed review of these models).

Among all significant results obtained within such a description, one can cite the weak dependence of impact on the volume of market orders, i.e., $G(n, V) \sim G(n) \ln V$, the long-range correlated nature of the sign of the consecutive trades ε_i and the resulting non-permanent power-law decay of impact function $G(n)$ [20]. Beyond their ability to reproduce most high frequency stylized facts, models like (2.1) or their continuous counterparts [6] have proven to be extremely interesting because of their ability to control the market impact of a given high frequency strategy and to optimize its execution cost [40].

1.4.2 Orderbook Models

Building a model that replicates empirical behavior of prices at the smallest of scales as well as the largest scales, naturally leads to modeling the dynamics of the different orders and their interactions. If we observe how the exchange works (section 1.2.2) we can see that the orders' interactions are governed by a rigid framework that guides the flow of information. This framework is the orderbook, and can be summed up by three simple interactions: limit orders arrive and fill different predetermined price levels. Cancel orders arrive and remove the limit orders from the orderbook. Market orders consume limit orders that are on one particular price level: to **top of book** (the best bid or best ask). So at every price level we have a birth-death phenomena where elements are born by the arrival of new limit orders and die according to the arrival of cancel and market orders. The whole process can be summarized as the arrival at random times of three marked point processes where the mark is size for market orders, price and size for limit orders and price and size for cancel orders.

There is a substantial number of work made in the direction of modeling these three interactions (cf [81, 20] and the references therein). Among those, we mention early attempts in [60, 27, 31, 18] which do a good job in replicating and predicting the behavior at the top of the book, such as the the spread, the depth of the best bid and ask, and probabilities of execution per unit time. However they do no capture actual price statistics, such as the variance of price returns, since the price was considered as an exogenous variable and the orderbook as a stationary process around it. On the other hand, models such in [11, 34, 82, 59, 80] addressed the price dynamics and considered the influence of the order flow on the price formation. However, these models focus mainly on explaining the variance of prices and pay secondary attention to direct consequences over the trading microstructure like the spread and the depth of book.

In [29], the authors provide a model that fills that gap. They provide a model that allows for the formation of the price as a consequence of the order flow and thus making the price a variable that is truly "formed" by the flow. This gives the advantage of recuperating immediate first order consequences of the order flow, as well as indirect second order consequences, such as the diffusive property of prices (cf [29, 81, 39]). We present in the following a quick review of their work.

Zero Intelligence

In this presentation, we describe the workings of the model for the sell limit orderbook and it is similar for the buy limit part. Hence, all market orders here are buy market orders and all limit orders are sell limit orders. For simplicity, the size of all orders has been reduced to 1, as was the tick value.

The arrival of limit orders is described as a Poisson process with constant intensity λ at each price level. The process of arrivals of all limit orders can be seen as a "Poisson" process at a rate λ share *per unit price* and per unit time and whose price is chosen "uniformly" from $[a, \infty[$ where a is the price of the best ask. Thus λ is not a straightforward Poisson intensity as it has a physical dimension of $shares/(price.time)$.

Each limit order comes with an independent Poisson process of constant intensity δ attached to it. This process serves as the cancelation process, allowing the limit order to be canceled with a probability δ per unit time. This means that each new limit order gets canceled at the time of first arrival of a Poisson process of intensity δ. δ has a physical dimension of $1/time$.

Finally, market orders hit only the best ask and they arrive according to an independent Poisson process with constant intensity μ. μ has a physical dimension of $shares/time$. Since the sizes of all orders have been reduced to 1, then each market order consumes only one limit order.

In short the quantity available at each tick price level is a birth death process, with a birth intensity described by λ, and the death intensity depends on the price level. If the price level is the best ask, then the death rate involves both the market orders rate and the cancel rate. If not, then it only involves the cancel rate.

This model can accurately predict some statistics of the orderbook that were traditionally considered as exogenous, such as the variance, the depth of book, the bid-ask spread, the price impact, and the probability and time to fill. Using dimensional analysis followed by a mean field analysis, the authors in [81] find that the spread is equivalent to $\frac{\mu}{\lambda}$, the asymptotic volume depth of the book to $\frac{\lambda}{\delta}$ and the variance to $\frac{\mu^2\delta}{\lambda^2}$. Furthermore, when confronted with real data in [39], the authors show that the model accurately predicts the empirical spread, and to a good extent the variance of financial assets. Furthermore, they show that the model produces a plausible market impact function.

The importance of this model lies in its mechanistic simplicity. Indeed, using very few parameters, and very little hypothesis about the behavior of agents or their strategies, they were able to create a bottom-up model that reproduce many empirical observations in the market. This completely transforms the modeling approach. At this level, they are modeling the most basic blocks in the process of price formation and the process of the price becomes a product of the combination of these elements. Furthermore, this modeling approach proves that some statistics are just regulated by the mechanics of the market regardless of the agents' behavior. Indeed, there are no assumptions made on the traders active in this framework, as orders arrive randomly in time, and are independent of each others and of the past.

Hence the model is called a **Zero Intelligence** market model, which reflects the lack of strategy of those sending the orders.

1.4.3 Relationship to Our Work

We find the modeling approach presented above particularly appealing. First, we relish the idea of having models that are built from the ground up as they are in good position to replicate market behavior over multiple observation scales, from fine scales to coarse scales in particular. In this framework, the modeling frontier is pushed to the most elementary level by modeling the arrival of different orders and their interactions. The price therefore becomes a byproduct of the inner workings of the model.

We also appreciate the fact that these models are conceptually simple and imitate realistically the workings of the market and do not make any assumptions about what is not observable, like the traders' strategies or behavior. Hence, their input parameters can be directly estimated from observing the market. The results of the models, when compared against real data, would be interpreted in light of assumed participants' behavior, but the behavior is never a parameter of the model.

In this work, we pretend to bring additional elements that contribute to the understanding of a fine to coarse model. For instance, if we look at the models described in section 1.4.1 equation (2.1), we notice that physical time does not play any role in the way market prices vary from trade to trade. This implies notably that the variance per trade (or per unit of volume traded) is constant and therefore that the volatility over a fixed physical time scale, is only dependent on the number of trades. In chapter 2 of the thesis, we critically examine this underlying assumption and provide empirical evidence portraying in what way this is assumption is broken.

We follow in chapter 3 with a simple model that mimics the random arrivals of trades whose prices present a mean reverting behavior at the fine scale, that vanishes at coarse scales. Tested against real data, the model was able to replicate multiscale empirical observations on the variance and the autocorrelation of trade prices. This also provides a new way to look at the tick size as defined in section 1.2.2.

Finally, chapter 4 studies a statistical method to estimate a powerful class of point processes. Inspired by previous work on modeling mutually and self exciting processes [9] and convinced by their potential in modeling the arrival of orders, we provide a non parametric moment method to estimate the decay kernel of these processes.

We introduce the works in section 1.6 of this introduction and present them in details in parts 2, 3 and 4. But now let's introduce the data we worked with.

The database we use is extensive, covering several world markets and different Futures asset classes that we describe in detail in section 1.5.

1.5 Data Presentation

The large scope of the data is necessary to show the universality of our findings, in electronic double auctions at least. Thus, we restrict our analysis to assets that trade in markets that match the framework of the electronic double auction described above (section 1.2).

In particular, we use **level 1**[5] data[6] of 10 futures contracts on assets of different classes that trade in different exchanges. The data has millisecond accuracy and was recorded over the past three years. In chapter 2 we have a date range going from 2008/08 till 2010/03, in chapter 3 we mainly use data from June 2009, and in chapter 4 we use data from October 2009.

On the EUREX exchange, localized in Zürich, Switzerland, we use futures on the DAX index (DAX) and on the EURO-STOXX 50 index (ESX), and on three interest rates based on German government debt: 10-years Euro-Bund (Bund), 5-years Euro-Bobl (Bobl) and the 2-years Euro-Schatz (Schatz). On the CBOT exchange, localized in Chicago, US, we use the futures on the Dow Jones index (DJ) and the 5-Year U.S. Treasury Note Futures (BUS5). On the CME, also in Chicago, US, we use the forex EUR/USD futures (EURO) and the the futures on the SP500 index (SP). Finally we use the Light Sweet Crude Oil Futures (CL) that trades on the NYMEX, localized in New-York, US. As for their asset classes, the DAX, ESX, DJ, and SP are equity futures, the Bobl, Schatz, Bund, and BUS5 are fixed income futures, the EURO is a foreign exchange futures and finally the CL is an energy futures. This information is summarized in table 1.1.

A futures contract is a contract allowing the purchase of an underlying physical good (be that a government bond, natural gas tankers, stocks or currency) on a future settlement date, at a price determined in the present [23]. It allows the traders of these contracts to secure the price of the good that they will physically deliver or receive at the settlement date. Therefore at any particular time, there may exist infinitely many contracts, one for every conceivable future settlement date. On an exchange however, settlement dates are standardized, one every three months (March, June, September and December) and generally three future settlement months are trading at the same time. We dealt with this issue by keeping, on each

[5]Level 1 data covers only what happens at the first level of the orderbook: every single market order is reported along with any change in the price or in the quantity at the best bid or the best ask price.

[6]Data provided by QuantHouse data providers. http://www.quanthouse.com/

day, the contract that recorded the most number of trades and, discarding the other maturities.

We also remind that when one market order hits several limit orders it results in several trades being reported (section 1.2.2). We aggregate together all such transactions and consider them as one market order whose size is the the sum of the volumes of all the reported trades and whose price is the price of the last limit order filled.

We refer the reader to table 1.1 for a summary of the information presented above and some simple statistics about the assets, such as the daily trading session, the average number of trades per day, and the proxies for the tick size η and $P_{=}$. We can see in that table that the assets are not homogenous: they have different exchanges (CBOT, CME, EUREX and NYMEX), different classes (Energy, Equity, Foreign Exchange and Interest Rates), different tick values and currency (e.g. DAX: 12.5€, DJ: 5$), different daily trading sessions, and different trading behavior: average trades per day (e.g. Schatz: 10000 trades per day, SP: 100000 trades per day) and different tick sizes (e.g. DAX: small, DJ: large). Throughout this work, we rely on this diversity in order to confirm the universality of our findings.

Futures	Exchange	Class	Tick Value	Session	# Trades/Day	1/2-η	$P_{=}$
DAX	EUREX	Equity	12.5€	8:00-17:30	56065	0.082	67.9
CL	NYMEX	Energy	10$	8:00-13:30	76173	0.188	79.8
DJ	CBOT	Equity	5$	8:30-15:15	36981	0.227	92.2
BUS5	CBOT	Interest Rate	7.8125$	7:20-14:00	22245	0.288	95.1
EURO	CME	Foreign Exchange	12.5$	7:20-14:00	42271	0.252	95.2
Bund	EUREX	Interest Rate	10€	8:00-17:15	30727	0.335	97.6
Bobl	EUREX	Interest Rate	10€	8:00-17:15	14054	0.352	99.1
ESX	EUREX	Equity	10€	8:00-17:30	55083	0.392	99.2
Schatz	EUREX	Interest Rate	5€	8:00-17:15	10521	0.385	99.4
SP	CME	Equity	12.5$	8:30-15:15	97727	0.464	99.8

Table 1.1: Data Statistics. The assets are listed from top to bottom following the increasing order of the $P_{=}$ column (see section 1.3), i.e., from the smaller (top) to the greater (bottom) "perceived" tick size. The *Exchange* column contains the exchange where the asset trades. The *Class* column contains the asset class of the underlying. The *Tick Value* column contains minimum possible price change in value (expressed in the local currency). The *Session* column indicates the considered trading hours (local time). The *# Trades/Day* is the average of the daily number of trades. η is the aversion to price change (section 1.3). $P_{=}$ is defined in equation (1.1) and reported here in percent.

1.6 Thesis Results

We introduce in this section the thesis results. We go over the main questions that we raised and present our main conclusions.

In section 1.6.1, we show that the average variance created by one trade depends on the instantaneous rate of trading. This means that the distribution of transaction time returns is dependent on the intertrade durations.

In section 1.6.2, we give a price model with random trade arrival times and integer price returns that presents a mean reverting behavior at fine scales and a diffusive behavior at coarse scales. This model is able to capture the variance signature effect and the tick returns autocorrelation.

In section 1.6.3, we give a non-parametric spectral method to estimate the decay kernels of mutually exciting point processes.

1.6.1 Intensity of Trading and the Orderbook

As we have pointed in section 1.4, we have seen in recent years the development of what is usually referred to as transaction time models. Early works on alternative time clocks [56, 26] show that an asset price can be modeled as a Brownian motion with an **alternative time clock** $X_t = B_{\tau_t}$ where B_t is the Brownian motion and τ_t a stochastic time clock independent of B_t [43]. Since then, a lot of research has been made in effort to find the best plausible time clock τ_t ([84] and the references therein) and today, there is a general consensus that the cumulated number of transactions is a good choice for the time clock τ ([7, 28]). This **transaction time clock** is rightfully seen to account for the highly intermittent nature of the volatility, a feature that manifests itself at practically all time scales, from intradaily scales (where periods of intense variations are observed, for instance around publications of important news), to monthly scales ([23, 28]).

In [43], the authors critically examine this view about transaction time. They show that while the transaction time clock is indeed an adequate Brownian subordination, it does not account for the clustered nature of the volatility nor for large price movements. Our article sheds a different light on the matter by looking at what role the physical time still plays in the series of prices, even as we look at the data in transaction time. In fact, the transaction time subordinated Brownian motion model suggests that the physical time does not play any role in the way prices vary from trade to trade. This implies notably that the variance per trade is constant as a function of the physical time and therefore that the volatility over a fixed physical time scale, is only dependent on the number of trades.

We present in our work empirical evidence showing to what extent transaction time models are able to explain volatility, and show the limit of such hypothesis. Using the extensive database presented in section 1.5 for a dozen high frequency Futures prices of varying tick sizes, we calculate the conditional first and second moments of various market statistics against the instantaneous rate of trading. We present the results graphically and show that the variance per trade, among other orderbook statistics, increases in a peculiar way as the speed of trading increases. We summarize the results below.

Realized Variance per Trade

In section 2.3, we define the Realized Variance per Trade. For a series of N trades spanning a duration of Δt, the variance per trade is the total realized variance divided by the number of trades. It is meant to represent the average variance contribution of one trade to the variance of the whole price process. Formally we go about it in the following way.

Let Δt be an intraday time scale and let N be a number of trades. We define $V(\Delta t, N)$ as the estimated price variance over the scale Δt conditioned by the fact that N trades occurred, or inversely, the estimated price variance of N trades, conditioned by the fact that they occurred during Δt seconds.

So, when $\Delta t = \Delta t_0$ is fixed and N is varying, $V(\Delta t = \Delta t_0, N)$ is seen as:

$$V(\Delta t = \Delta t_0, N) = E\left[(p_{t+\Delta t_0} - p_t)^2 \mid NT[t, t+\Delta t_0] \in [N - \delta_N, N + \delta_N]\right] \quad (1.5)$$

where E is the historical average, $NT[t, t+\Delta t_0]$ corresponds to the number of trades in the time interval $[t, t + \Delta t_0]$ and δ_N is some transactional bin size.

Along the same line, when $N = N_0$ is fixed and Δt is varying, one defines a temporal bin size $\delta_{\Delta t}$ and computes $V(\Delta t, N = N_0)$ as

$$V(\Delta t, N = N_0) = E\left[(p_{t_{i+N_0}} - p_{t_i})^2 \mid t_{i-1+N_0} - t_{i-1} \in [\Delta t - \delta_{\Delta t}, \Delta t + \delta_{\Delta t}]\right] \quad (1.6)$$

where i is the index of the i^{th} trade and t_i its time.

Finally, we define the corresponding conditional **Realized Variance per Trade** as:

$$v(\Delta t, N) = \frac{V(\Delta t, N)}{N} \quad (1.7)$$

We perform a certain number of numerical experiences summarized below.

Realized Variance is Linearly Correlated to the Number of Trades

The first thing we do is confirm the idea that the number of trades is indeed a good explanation for the volatility fluctuation. We do this by dividing the trading days into 5 minutes non overlapping intervals and looking at some measure of the realized variance during these 5 minutes. We clearly see what others have confirmed before (see for eg. [28]) that realized variance is indeed linearly correlated to the number of trades (cf Figure 2.1 and Figure 2.2).

However, these observations do not show what happens at the "second order". Indeed, if we believe this linear relationship to be true, then the realized variance *per trade* $v(\Delta t, N)$ (Eq. (1.7)) must be constant when N is held constant. As a function of Δt, any dependence that does not look like a random noise suggests that a second order contribution of the rate of trading is being ignored.

This is what we find in the early test with 5 minutes interval bin (Figure 2.2). We find that the constant variance per trade hypothesis works well for most trading regimes, but that it breaks down when the rate of incoming market orders becomes high.

Impact of a Single Trade

Our motivation of taking a bottom up approach to modeling instead of a top down one have pushed us into looking at the impact of a single trade on the following price change. We do this in section 2.4 for the many futures assets present in our database.

In this presentation, we show the results for the large tick asset SP. In Figure 1.2 below, we show how the variance of the return $r_{t_i} = p_{t_{i+1}} - p_{t_i}$ behaves when conditioned by $t_i - t_{i-1}$, the time elapsed since the previous transaction. We see that the obtained curve has a peak of 0.1 ticks^2 for very small Δt and stabilizes around an asymptotic constant value of 0.005 ticks^2 for larger Δt. This represents a 2000% increase above the asymptote. If we think of the asymptote as a regime where the variance is linearly related to the number of trade, then this figure shows that the actual behavior presents a significant break from this regime!

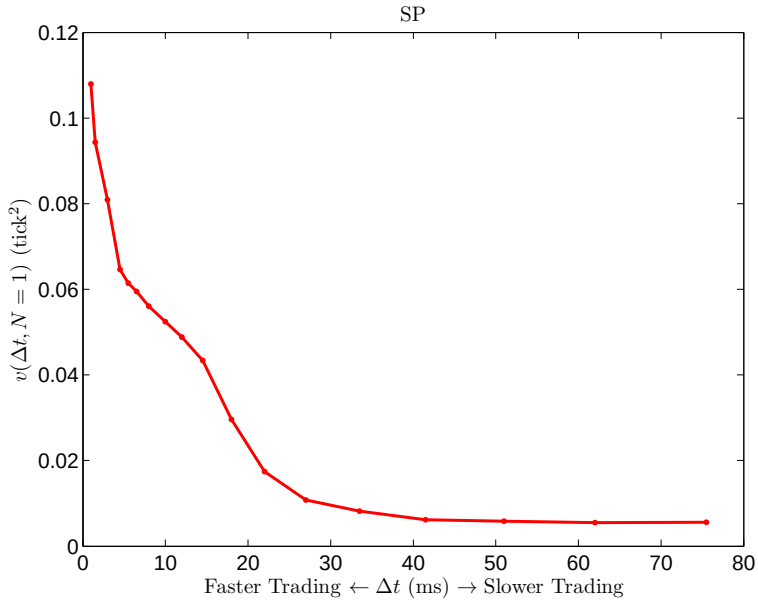
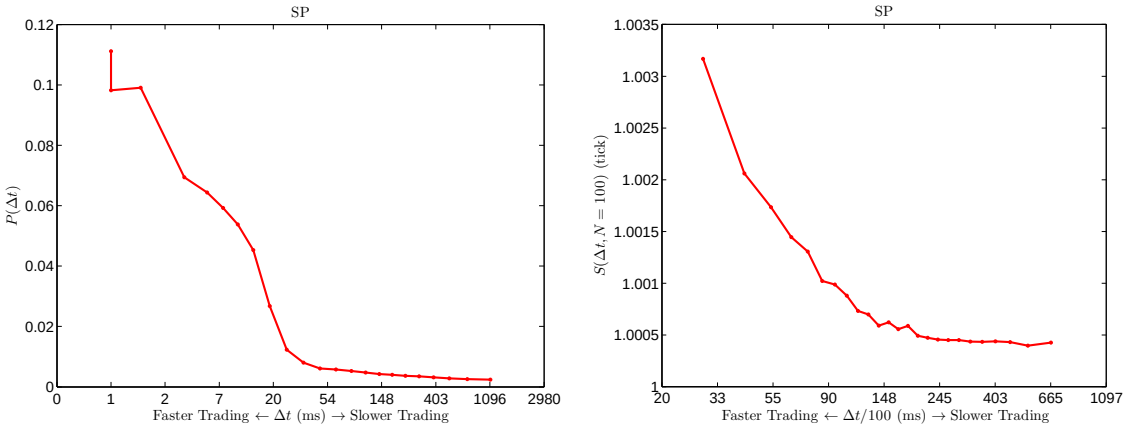


Figure 1.2: $v(\Delta t, N = 1)$ as a function of Δt over very short Δt's for SP. The variance per trade increases dramatically below $\Delta t \approx 30$ ms. The results for the other assets are shown in Figure 2.5

In order to further investigate this behavior, we try to give an idea about the distribution of the returns across different states of market activity. We do this by answering the question of whether the price moves more strongly or more often when the rate of trading varies. So we concern ourselves with the variable of tick returns conditioned to having non null returns, and find that the variance per trade of this variable is constant across all periods of market activity (Figure 2.7). This result is particularly true for large tick assets where the orderbook is always well furnished and where it is extremely rare for a trade to penetrate more than one level of it. This result mirrors [43]'s finding that the tails of the returns distribution cannot be explained by fluctuations in market activity. In fact, the distribution of returns and that of returns conditioned to them being non null have the same tails. So, while the tails cannot be explained by market activity, our result shows that on average the size of the price movement is also independent of market activity.

This observation is compatible with the general knowledge that even "impatient" traders (those who would rather send market orders and be filled immediately rather than send limit orders and wait for a market order to fill them) almost never send orders that consume more than what is available at the top of the book. Even if they have a large quantity to trade, they tend to wait for the consumed top of book to refill before trying to execute again. They do this in an effort to hide their intentions from the rest of the market, reduce their immediate execution cost, and potentially their future market impact [6]. So if the market is active enough,



(a) Probability (b) Spread

Figure 1.3: For the large tick asset SP. The plot on the left (1.3a) is the probability $P(\Delta t)$ as defined by (1.8) as a function of Δt. We see that the probability of getting a price move increases with market order rate and entirely responsible for the increase in variance seen in Figure 1.2. The plot on the right (1.3b) is the average spread over $N = 100$ trades, as a function of $\Delta t/100$. The form of the curve confirms the decrease in liquidity when the trading rate is increasing.

and the top of book is refilled fast enough, such traders are able to send many orders in a very short time that are large enough to consume the top of the book, but not large enough to consume the second level. This idea is seen in Figure 2.6, and reproduced here for the large tick asset SP in Figure 1.3. Defining $P(\Delta t)$ as the probability for the return to be non zero conditioned by the intertrade duration $t_i - t_{i-1} = \Delta t$,

$$P(\Delta t) = Prob\{r_{t_i} \neq 0 \mid t_i - t_{i-1} \in [\Delta t - \delta_{\Delta t}, \Delta t + \delta_{\Delta t}]\} \tag{1.8}$$

we show in Figure 1.3a that the frequency of price movement seen as a function of market activity increases as the rate of trading increases and is basically responsible for the behavior observed on the variance in Figure 1.2. See Figure 2.7 in subsection 2.4.2 for the plots of other assets.

As a consequence of these observations, one can say that the variance increase in high trading rate period is mostly caused by the increase in the probability that a market order moves the price by absorbing only the first level of the book (best bid or best ask). There is hardly any perforation of the book in the deeper levels. See subsection 2.4.1 for a more detailed discussion.

Before we move to the next section, we want to point out that the form of the variance curves observed is quasi universal (see figures 2.5, 2.6, 2.8). One

qualitatively observes over all assets, that the variance is almost constant over a large range of relatively slow rates of trading (or increases very slowly) and then suffers an explosion as Δt becomes smaller than a certain threshold. This behavior can be interpreted as a sudden disappearance of liquidity that we qualify by "mini liquidity crisis". We have some speculation about the origin of these crises but further research is required to confirm our thoughts. We elaborate on this point in the conclusion part of this introduction (section 1.6.1). But first, let us review our results for coarse scales.

Impact of Several Concatenated Trades

In the spirit of our bottom up approach to financial modeling, we then move to explore how this phenomenon builds up into the coarse scales. So we look at the situation when we aggregate the trades in section 2.5.1. We check whether the variance averaged locally over a large number of trades still displays a dependence with respect to the trading rate and similarly, we find that an increase in the trading rate is also reflected by an increase in the variance per trade. However, we also see that this effect is dampened, as the increase in the variance is less pronounced quantitatively. Also, the curves in figures 2.9 and 2.10 do not display the threshold-like behavior of the single trade curves and go smoothly from small to large values as the trading rate increases. This indicates that while not all activity fluctuation is transmitted to the coarse time variance, part of it is. To put it in [20]'s terms, the market slowly "digest" the fluctuations in demand and it will take an averaging over much coarser scales in order for the variance-activity dependance to disappear.

Liquidity

The final result that we present is a look at the average spread over N trades in section 2.5.2. We see in Figure 2.11 a relationship that is similar to the one observed for the variance: an overall increase of the spread value with the rate of trading for all assets suggesting that the orderbook is thinner during periods of intense trading. We reproduce this observation in Figure 1.3b.

This observation is compatible with [87], where the authors shows that the average spread is linked to the variance per trade through a two-way feedback. The first one relies on the fact that when the spread is high, the orderbook is thin and the gap between the first level and the second level of the book is equivalent to the spread. Moreover, movements of the price are the result of a change in the top of book, from its current value to the value at the second level of the book. Therefore, any price change would have to be proportional to the spread. As a consequence, the volatility per trade is also proportional to the spread. The second feedback relies on an arbitrage analysis by the market maker. In periods of large

volatility, the risk for the market maker is increased. He compensates for that risk by increasing the spread.

In all cases, an increase in the spread is synonymous with a decrease in the available liquidity.

However, our work with the single trade impact of the rate of trading show that the increase in the variance per trade comes from the number of price moves as opposed to their amplitude. This is true for all the assets except ones with very small ticks (primarily the DAX - see figures 2.6 and 2.7). So while the interpretation of a deeper book penetration is compatible with very small tick assets like the DAX, it does not hold for other ones because the book does not get penetrated by more than one level (as this is seen from the amplitude of non null price returns in Figure 2.7). So what is really going on? We reflect on this problem in our concluding remarks.

Conclusion and Further Research

Using an extensive Futures database covering many asset classes across different exchanges and displaying different tick sizes we clearly show that the distribution of price returns changes significantly and systematically with the rate of trading. Starting with the intention to shed a light on the underlying assumption behind transaction time models, namely the idea that transaction time is a good Brownian subordinator for the series of prices, we find that there is a positive correlation between the variance per trade and the rate of trading. We show this by sampling the trading day with constant trade intervals and looking at the time elapsed between the first trade and the last trade of that interval. We show results for the finest scale of $N = 1$ trade as well as the coarser scales of $N = 100, 200$.

As we have mentioned in the above paragraphs, we see that the increase in variance during periods of high activity is due to the increase in the probability of getting a price movement, and not to the amplitude of the movement. This is true except for the assets with the smallest ticks. This means that trades that penetrate the book by more than one level are extremely rare.

While this means that most of our results are not compatible with the first part of [87]'s two-way feedback effect because most of the assets we considered have ticks that are too large, they are still compatible with the second feedback effect of the market maker increasing his spread in response to the increase in total variance created by the additional arrival of trades. Hence, an increase in the total realized variance translates into an increase in the variance per trade by means of the market maker's strategy. At coarse scale, the smooth form of the spread curves in Figure 2.11, suggest the existence of a strategy dictating the average spread for each trading rate.

However, the form of the curves at $N = 1$ makes us dubious about this interpretation. Figures 2.6 and 2.4 clearly show the existence of a rate threshold above which the variance per trade has an explosive behavior. If we believe the explanation that a market maker is controlling spread, we would expect a similar behavior like with the coarse scale case: a smooth and slow increase in the variance.

This brutal and sudden change turns our attention to the rate of limit orders. Our current preliminary research has shown that the rate of limit order is strongly linearly correlated with the rate of market orders but with the market orders slightly leading (an observation also confirmed in [83]) as if the limit orders react to the arrival of market orders and not the opposite. In other words, market makers will increase their offer of liquidity only after observing the increase in the demand of liquidity. This interaction is limited by the liquidity providers' ability to see the increase in the demand of liquidity. One physical limitation naturally lies in the market framework itself. When a market order comes in, it gets matched by the exchange and a message is published about that event. The market maker receives this information and reacts to it by sending a limit order intended to replace the liquidity depleted by the market order. The market receives the limit order from the market maker and incorporates this information in the book thus refilling the liquidity consumed by the market order. We call **liquidity latency** the time elapsed from right after the execution of the market order to right after the placement of the reactionary limit order in the book. While we could not find any references that give an estimation of this latency, a discussion with high frequency traders, concluded that this latency can be very variable but no less than 5 ms, and probably averaging around 15 ms (most of it comes from the final part, the insertion of the new limit order in the orderbook by the exchange's order placement engine). Interestingly, this number is of the same order of magnitude as the threshold that we observe in Figure 1.3a (see section 2.4.1 for the results on many assets).

In fact a toy model in which market orders arrive according to a Poisson point process, and such that a limit order replaces every market order, except for those orders arriving faster than a certain threshold, would give price movement probabilities that are very similar to what we observe in Figure 1.3a.

Confirming this idea would however requires more research into the rate of limit orders and its relationship to that of market and cancel orders. It would also require access to more time accurate limit and cancel orders data, which we do not currently possess.

1.6.2 Discrete Microstructure Noise

We have established in section 1.6.1 the importance the rate of trading plays in determining the variance per trade. We now start exploring the possibilities offered

by a rate modeling approach. We begin by describing a jump process model consisting of a random walk to which we add an autocorrelated moving average process and find it particulary powerful in reproducing the signature plot effect commonly observed in financial prices [8]. In this model, the random walk is interpreted as a **latent price process**, and the autocorrelated process is interpreted as a **noise process** (cf section 1.6.2 for more details). Furthermore, our model has the particularity of having a discrete jump distribution that can replicate the autocovariance of tick returns and their unconditional distribution.

Microstructure Noise

In section 3.1, we introduce the issue of microstructure noise associated with high frequency financial data. This problem gained relevance in finance when researchers started using high frequency financial data. They were confronted with problems that were until then circumvented such as bid-ask bounces and asynchronous trading. Among these events we mention the **Signature Plot** of the Sampled Average Realized Variance and the **Epps Effect** of the correlations. In particular, the variance Signature Plot is the phenomenon which produces an increase in the realized variance as the sampling duration goes to 0, whereas the Epps Effect is the phenomenon which produces a decrease to 0 of the correlation of two correlated but asynchronous processes. We focus in chapter 3 on the Signature Plot. The classical idea for modeling the Signature Plot effect is that there exists a **latent price** process (a diffusive martingale) which gets contaminated with noise or measurement error (an idea inspired from physics were measurements are contaminated with white noise). Although we do not have a physical measurement error in financial data, the effect that appears is highly similar and the contamination process is called **microstructure noise process**. In the following paragraphs we will show the empirical stylized facts we encounter at high frequency and suggest a model to capture them.

Empirical Observations and Fine Scale Abnormalities

In this paragraph we show the empirical observations that we try to replicate in chapter 3. First we point out that we use the series of tick returns (i.e. price differentials normalized with the tick value) and not log returns. This will allow us to interpret the results by using the tick size (section 1.3) in section 3.4.3. This translates to having a series of integer price returns with often a tight compact support. This can be seen in Figure 1.4 below, where the cumulative distribution function of the Bund tick returns is displayed and where the support is practically $\{-2, -1, 0, 1, 2\}$ (cf table 3.1 in section 3.4 for a breakdown of the probabilities).

The other empirical observation is that the autocovariance of tick returns is significant but decreases very fast. The is seen on the Bund in Figure 1.4 below,

where the first lag autocovariance is clearly significant, whereas it is statistically null for bigger lags. The same phenomenon is seen in other assets, as we can see in Figure 3.7.

So on the fine scale, there are "abnormalities" in the series of tick returns which makes them weakly but significantly autocorrelated. If we are looking at the trade prices, these autocorrelations are interpreted as bid-ask bounces, whereby arriving trades "bounce" from one side of the book to the other. Nevertheless, the autocovariance effect is also present when looking at midpoint prices and last buy prices, as we can see figures 3.11 and 3.15 of chapter 3. So the origins of the autocovariance of tick returns is not merely bouncing effects, and we need a model capable of capturing it.

On the coarse scale, this autocovariance quickly disappear, and the price displays a diffusive behavior and the diffusion variance is the asymptote of the sampled realized variance seen in Figure 1.4.

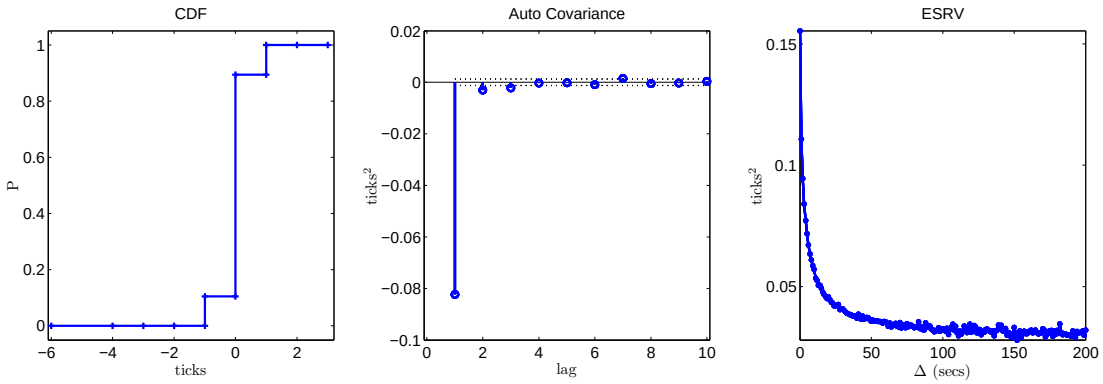


Figure 1.4: Some empirical statistics of the distribution of tick returns of the Bund futures in October 2009. The graph on the right shows the cumulative distribution function (CDF) (cf table 3.1 for details). The graph in the middle (ACF) shows the autocovariance of tick returns. The third graph on the left (ESRV) shows the empirical sampled realized variance.

We propose a discrete model built around the principle of random trade arrival time. The model displays some of the properties of high frequency prices like the autocorrelation of tick returns while being diffusive when observed at coarse scales. Consequently, the unconditional variance of sampled returns will be influenced by the sampling period, while converging to a constant (the variance of the coarse scale diffusion) when the sampling period goes to ∞. Such a model is at the heart of our view on financial modeling. First, we are faithful to the fine to coarse approach

described earlier and second, we incorporate random arrival of trades after we have seen the importance of the intensity of trading on the distribution of tick returns in section 1.6.1.

Thoughts About the General Model

The idea is to have a model of prices as a stochastic process that displays empirical behavior of prices at high frequency similar to those observed on the markets, while at the same time converging to a diffusive martingale at large scales. Other properties that we want to replicate is the arrival of trades at random times and the discrete returns of assets. So we look for a a model that can be presented with a marked point and whose marks are discrete random variables. The cumulative sum of the marks represent the price process.

In order to get a macroscopic diffusion process with some "abnormalities" at fine scales, we propose the sum of two independent point processes. The first, a compound Poisson process, contains the diffusive martingale term is little different from a random walk and we call it the **fundamental** term. The second contains all the abnormalities observed at fine scale and we call it the **noise** term [4]. In order for the sum of these two processes to diffuse at coarse scales, we make the noise term a vanishing telescopic sum which only influences the behavior of the price at the highest frequencies. This choice is motivated by two things. First, we want to be able to clearly distinguish between the fundamental term and the noise term as we do not want the noise to contribute any variance to the diffusive variance. Second, a simple vanishing telescopic noise term is enough to replicate the variance signature and the autocorrelation of tick returns. In the following, we present the model, and the closed form formulas for the sampled variance as well as the autocovariance of tick returns.

Model Description

We consider a price model process $(X_t)_{t\geq 0}$ in continuous time, consisting of the sum of two independent compound Poisson processes such that:

$$X_t = R_t + E_t \tag{1.9}$$

where R_t and E_t are independent compound Poisson processes.

The fundamental term

$$R_t = \sum_{i=1}^{N_t} \varepsilon_i$$

where N_t is a Poisson counting process of rate λ independent of ε_i, is a diffusive martingale under the assumption that ε_i are symmetrical, independent identically distributed (iid) random variables.

The noise term

$$E_t = \sum_{i=1}^{M_t} e_i$$

where M_t is a Poisson counting process of rate μ independent of e_i, is an autocorrelated vanishing process under the assumption that

$$e_i = \widetilde{\varepsilon}_i - \widetilde{\varepsilon}_{i-1}$$

where $\widetilde{\varepsilon}_i$ is a symmetrical, discrete random variable (cf section 3.4).

We have that $(e_i)_{i\geq 1}$ is a weakly stationary process such that its moments and autocovariance are independent of the index i and we define $\Gamma_e(j)$, the autocovariance function at lag j of e_i by

$$\Gamma_e(j) = E(e_i e_{i-j})$$

This model allows us to write a closed form formula for the sampled realized variance and the autocovariance of tick returns of X_t. We quickly review these results in the following.

Realized Variance

The Sampled Average Realized Variance (SRV) estimator is defined by:

$$V_T(\Delta) = E(\frac{1}{T} \sum_{k=1}^{\lfloor T/\Delta \rfloor} (X_{k\Delta} - X_{(k-1)\Delta})^2) \tag{1.10}$$

where T is the final time in the sample and Δ is the sampling period. As we have seen in 1.6.2, the plot of $V_T(\Delta)$ against Δ has a particular form and it is called the variance signature plot.

For the model (1.9), in section 3.2.1 we give an explicit formula for $V_T(\Delta)$:

$$V_T(\Delta) = E(\varepsilon_i^2)\lambda + E(e_i^2)\mu + 2\sum_{j=1}^{\infty}\Gamma_e(j)W(\mu,\Delta,j) \tag{1.11}$$

where $W(x,\Delta,j) : \mathbb{R}^2 \times \mathbb{N} \to \mathbb{R}$ is calculated by:

$$\begin{aligned} W(\mu,\Delta,0) &= \mu \\ W(\mu,\Delta,j+1) &= \mu - \frac{1}{\Delta}\sum_{k=0}^{j} P(M_\Delta > k) \end{aligned}$$

In section 3.2.2, we also give an explicit formula for the autocovariance of tick returns.

Autocovariance of Tick Returns

Let $\tau_0 = 0, \tau_1, \tau_2, \cdots, \tau_N$, be the series of jumps of either N_t or M_t (or equivalently those of the Poisson process $N_t + M_t$) put in a strict ascending order. In the vocabulary of section 1.6.1, τ_i is the transaction time clock of X_t, i.e. τ_i corresponds to the arrival of the i^{th} transaction of X.

Consider the series $J = (J_i)_{i\geq 0} = (X_{\tau_i} - X_{\tau_{i-1}})_{i\geq 0}$ of successive **tick returns** of X_t. The series J, is weakly stationary and we can define its autocovariance function at lag j, $\Gamma_J(j)$. In section 3.2.2 we exhibit a closed form formula for $\Gamma_J(j)$.

We use this model to replicate the fine to coarse behavior observed on real data (cf section 1.6.2).

Application to Financial Data

For financial asset prices, we find that taking $\widetilde{\varepsilon}_i$ as a symmetrical (centered), discrete random variable with values in $\{-1, 0, 1\}$ does a good job of replicating the empirical SRV and the autocovariance of tick returns. In this case the distribution of ε_i can be described with one parameter p:

$$p = P(\widetilde{\varepsilon}_i = 1)$$

giving e_i a distribution with a support in $\{-2, -1, 0, 1, 2\}$. We then give ε_i a distribution with a similar support that is described by two parameters $p_2^\varepsilon = P(\varepsilon_i = 2)$ and $p_1^\varepsilon = P(\varepsilon_i = 1)$ (cf section 3.4). Under these conditions, the SRV and the

autocovariance can be written as:

$$V_T(\Delta) = (8p_2^\varepsilon + 2p_1^\varepsilon)\lambda + \frac{4p}{\Delta}(1 - e^{-\Delta\mu}) \tag{1.12}$$

and the j^{th} lag ($j \geq 0$) autocovariance of J defined in equation (3.11) as:

$$\Gamma_J(0) = \frac{\lambda}{\lambda+\mu}(8p_2^\varepsilon + 2p_1^\varepsilon) + 4p\frac{\mu}{\lambda+\mu} \tag{1.13}$$

$$\Gamma_J(j) = -2p(\frac{\mu}{\lambda+\mu})^2(\frac{\lambda}{\lambda+\mu})^{j-1} \qquad \text{for } j > 0. \tag{1.14}$$

We then have that

$$\lim_{\Delta\to\infty} V_T(\Delta) = E(\varepsilon_i^2)\lambda$$

and

$$\lim_{\Delta\to 0} V_T(\Delta) = E(\varepsilon_i^2)\lambda + E(e_i^2)\mu$$

guaranteeing that $V_T(\Delta) > V_T(\infty)$ similar to what is observed on most assets (section 3.1.1).

We now explain the method we use for fitting the model and show an example of the results we get on financial data but first we note that equation (1.12) is not enough to fully determine our model. Indeed while p and μ can be determined uniquely from equation (1.12), we can only get an estimate for $E(\varepsilon_i^2)\lambda = (8p_2^\varepsilon + 2p_1^\varepsilon)\lambda$ but not p_2^ε, p_1^ε and λ separately. We see in the following paragraph how to use the distribution of J_i in order to fully define our model.

Parameter Fitting: We fit the model in three steps beginning with the SRV. We reproduce in Figure 1.5 the calibration of the model on the Bund futures. We use a moment method and do a non linear leat squares to fit the empirical SRV to equation (1.12). This uniquely determines μ, p and $\chi = E(\varepsilon_i^2)\lambda = (8p_2^\varepsilon + 2p_1^\varepsilon)\lambda$. This means that the parameters p_2^ε, p_1^ε and λ are constraint by the condition $\chi = (8p_2^\varepsilon + 2p_1^\varepsilon)\lambda$ but not uniquely determined. This first fit is here shown in Figure 1.5, upper left graph (SRV).

Once we have the values for χ, μ and p we use them to fit the autocovariance function of J, $\Gamma_J(j)$ for $j \geq 0$. We use a least squares fit of the empirical autoco-

variance to $\Gamma_J(j)$, $j \geq 0$. This means that we try to find the best model possible that fits the unconditional variance and the autocovariance of the tick returns of X. This allows us to uniquely determine λ and $E(\varepsilon_i^2) = (8p_2^\varepsilon + 2p_1^\varepsilon)$. This is shown in Figure 1.6, graph on the left (Auto Covariance).

Finally, we are left to determine p_2^ε and p_1^ε. We already know μ, λ, p and $E(\varepsilon_i^2)$ from the previous fits. The last information that we can use is the unconditional distribution of J. We use a non linear least square fit of the cumulative distribution function of J to the one suggested by the model and that gives us p_2^ε and p_1^ε uniquely. This is shown in Figure 1.6, graph on the right (PDF).

This method gives priority to reproducing the signature plot first, followed by the second moments of J_i and finally the unconditional distribution of J_i. We pay attention that the estimated parameters are consistent with the requirements of the model such that $\lambda, \mu > 0$, $\{p_2^\varepsilon, p_1^\varepsilon, p\} \subset [0, \frac{1}{2}]$ and $p_2^\varepsilon + p_1^\varepsilon \leq \frac{1}{2}$.

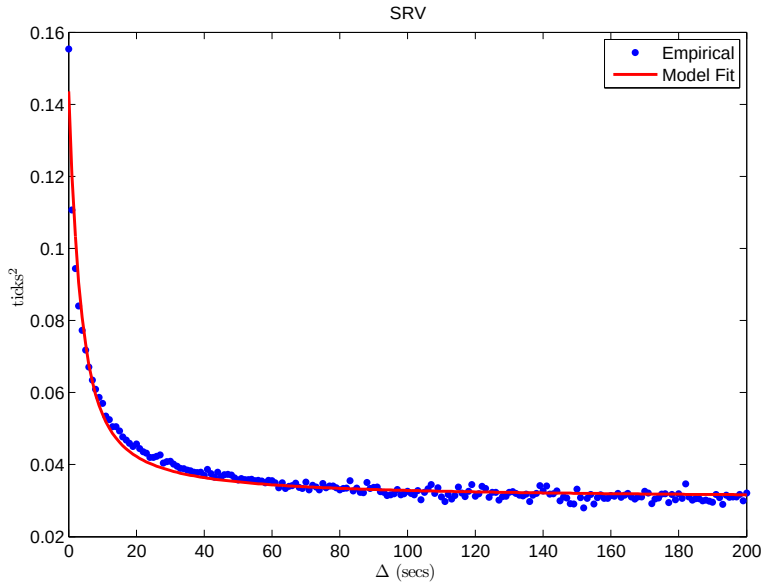


Figure 1.5: Fit result of the SRV for the process of trade prices of Bund Futures. The blue curve is for the empirical data and the red one is the one produced by the calibration of model (1.12).

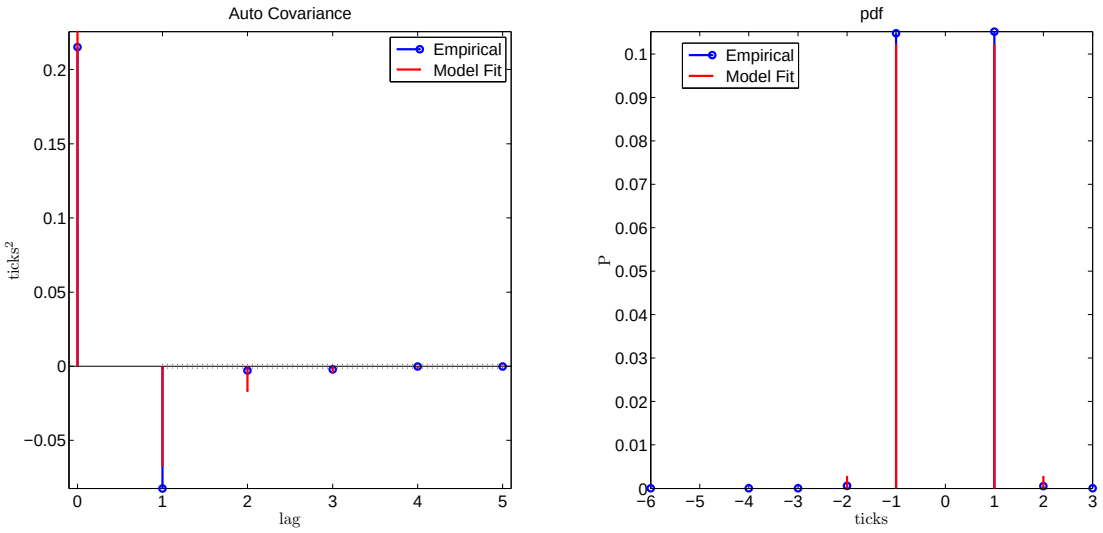


Figure 1.6: Graphic comparisons between empirical statistics for the process of trade prices of Bund Futures and the statistics produced by the calibration of model (1.12). The blue curves are for the empirical data and the red are for the calibrated model. From left to to right: in the first graph we show Γ_J for 5 lags with the 95% significance in a dotted black line, in the second we show the density of J, $f_J(x)$. $f_J(0)$ was removed to make the plot more readable.

We conclude this model by looking at how the model fairs when confronted with real data. For a concise view of the results, please see section 3.4.2 where we show and discuss the results for the Bund futures asset. We discuss the analysis done for other assets (in section 3.4.3) and the interpretation vis-a-vis of the tick size in the following paragraph.

Tick Size: In section 3.4.3 we discuss the results of the model calibration with the 10 futures presented in section 1.5. We calibrate the model with three price types, the trade price, the last buy price and the midpoint price. We mainly give here analysis and interpretations regarding the trade price and for more information please refer to section 3.4.3.

We notice that **the diffusion deviation measure**

$$d = \frac{V_T(0) - V_T(\infty)}{V_T(\infty)} \tag{1.15}$$

changes significantly with the tick size or the price type. In particular, d is larger for large tick sizes and it is the highest for the trade prices, followed by the midpoint prices and it is smallest for the bid prices. The closer d is to zero, the weaker the

signature effect. In terms of the model d translates into[7]

$$r = \frac{\mu(1 - p_0^e)}{\mu(1 - p_0^e) + \lambda(1 - p_0^\varepsilon)} \tag{1.16}$$

Now for the trade prices, we see in Figure 3.6 that the SRV displays its usual signature form for all assets and the model was able to capture it. After the fit however, we compare r to η (η is the proxy for the tick size defined in section 1.3) in Figure 3.4 and saw a linear relationship between the two: as the tick size grows, the contribution of the autocorrelated noise grows as well. For the trades prices, this represents an increase in the bid-ask bounces, a trait that is particularly present in large tick assets.

Before concluding, we point out that in section 3.5 we explore the possibility of a model that does not allow for tick returns of size zero. The model in this case becomes a true point process (i.e. meaning that the jumps of $N_t + M_t$ would correspond to a non null movement or a jump of X_t), and would have the advantage of being more easily interpreted. Indeed, if that was the case, it would suffice to use the Poisson rates μ and λ to define r in equation (1.16) instead of $\mu(1 - p_0^e)$ and $\lambda(1 - p_0^{1e})$. However, we find that our restriction made the model much less flexible with a very small range for the diffusion deviation measure d. This is seen in the fit attempt on the Bund shown in Figure 3.5.

Conclusion and Further Research We present in this paper a simple model for asset prices on the ultra high frequency scale. The model described in equation (1.9) is the sum of two independent point processes. The first one, that we call the fundamental component, consists of the cumulative sum of a series of iid random variables arriving at independent Poisson times. The second one, that we call the noise component, consists of the cumulative sum of an autocorrelated degenerate MA(1) process also arriving at independent Poisson times.

The model has the particularity of quickly converging to a diffusion as the observation scale becomes coarser, a property that is fundamental for the theory of asset pricing. However, on the microscopic scale, the model is flexible enough to reproduce one of the commonly observed statistics of prices at this scale, the signature plot effect, the phenomenon in which the Sampled Average Realized Variance (SRV) increases with the sampling period of the prices.

This model is intended almost as toy model to explore the possibilities of modeling discrete prices arriving at random times. The fact that we are working with

[7] cf subsection 3.4.3 for an explanation about this choice of r

discrete prices expressed in tick value, as opposed to logarithmic returns, puts a lot of constraints on the form of the noise component and on the estimation method. Nevertheless, the model fits in a satisfactory way to Futures data and with different definitions of the price (trade price, last buy price and midpoint price). It is capable of reproducing both the form of the observed SRV and the weak but significant autocovariance of tick returns for assets of varying tick size, which to our knowledge was not considered in previous models that portray microstructure noise.

This work has a lot of room for development. Notably, the estimation method needs improving and we need to give convergence results. Also the model itself can be improved by allowing for more flexible trade arrival times or a more complex noise error term.

1.6.3 Non Parametric Estimation of Hawkes Decay Kernels

Hawkes Processes

In chapter 4, we propose a non parametric method to estimate the decay kernel of a mutually exciting multidimensional point process, generally known as Hawkes process and apply the method to financial data. We do this in the one dimensional case (the process in this case is called self exciting) and in the special multidimensional case where the process is invariant under arbitrary permutations, i.e. when all the components of the process are identically distributed. We go through a brief review of these processes and present the main results of the paper in this introduction. We keep to the one dimensional case out of simplicity.

One dimensional Hawkes processes are a class of point processes that were first introduced by Alan Hawkes in [46, 47]. They are point processes whose instantaneous Poisson intensity is linearly dependent on the arrival past of past events. They exhibit a **self exciting** behavior by which the arrival of one event increases the probability of occurrence of new ones. In its most basic form and in the one dimensional case, the Hawkes process is a counting process defined by λ_t, the rate of arrival of events by:

$$\lambda_t = \mu + \int_{-\infty}^{t} \phi_{t-s} dN_s \tag{1.17}$$

where $\mu > 0$ is a constant **background rate**, N_t is the cumulative counting process and ϕ a positive real function called **decay kernel**.

We can clearly see in equation (1.17) that when an event occurs at time t, we have $dN_t = 1$ and hence $d\lambda_t = \phi_0$. The influence of the event is transmitted to future times through ϕ such that at time $u > t$, the increase in λ_t due to the time

t event is ϕ_{u-t}. Thus a self exciting behavior is observed (cf Figure 1.6.3 for an example of how λ_t evolves in time). It has been pointed out that this kind of self exciting behavior governs the process of order arrivals in financial markets, and Hawkes processes have been used to model it (cf. [9, 17, 24, 48]).

In section 4.1, we review the origin of these processes and their applications. Most importantly, we are concerned with the estimation of the decay kernel ϕ and the applications in finance. Our work is inspired by the previous works by Bacry, Delattre, Hoffmann and Muzy [9] and it is a continuation of [10]. In a similar way to how it was originally done, we use the autocovariance function of the cumulative counting process to give a non parametric estimation of the decay ϕ. We do that in the one dimensional case and in some special multidimensional cases, using Fourier transforms and phase recuperation techniques. Finally, we apply the method to financial data. We go over the main results in this presentation, restricting ourselves to the one dimensional case.

Definitions and Notations

Hawkes in [46] introduces the **self exciting** point process as follows. Let $(N_t)_{t\geq 0}$, the process of cumulative number of events up to time t, is defined via its random intensity process

$$\lambda_t = \mu + \int_{-\infty}^{t} \phi_{t-s} dN_s \tag{1.18}$$

where $\mu > 0$ and ϕ_t, the kernel, obeys to the following assumptions:

(H1) ϕ_t is positive and causal, i.e.,

$$\phi : \mathbb{R} \to \mathbb{R}^+, \quad \forall t < 0,\ \phi_t = 0,\ \text{and } \forall t \geq 0,\ \phi_t \geq 0, \tag{1.19}$$

(H2) the spectral radius of $\widehat{\phi}_0 = \int_{\mathbb{R}} \phi_t dt$ (i.e., its largest eigenvalue) is strictly smaller than 1.

Under these assumptions $(N_t)_{t\geq 0}$ is a point process with stationary increments and the conditional intensity λ_t is itself a stationary process with mean

$$\Lambda = E(\lambda_t) = E(dN_t)/dt.$$

that can be expressed as

$$\Lambda = \frac{\mu}{1 - \widehat{\phi}_0},$$

Before moving on, we need to introduce some more notations that will be used all along this presentation.

Notations 1.6.1. In the following

- For any function f_t, $\widehat{f}_z = \int_{\mathbb{R}} e^{-zt} f_t dt$ corresponds to its Laplace transform. The Fourier transform of f_t is then expressed $\widehat{f}_{i\omega}$, the restriction of the Laplace transform to the imaginary line.
- For any real functions a_t and b_t the convolution product of a_t and b_t is defined as $a \star b_t = \int_{\mathbb{R}} a_s b_{t-s} ds = \int_{\mathbb{R}} a_{t-s} b_s ds$.

Assumption $(H2)$ allows us to define

$$\Psi_t = \sum_{n=1}^{\infty} \phi_t^{(\star n)} \tag{1.20}$$

where $\phi_t^{(\star n)}$ is the n^{th} auto convoluted function of ϕ_t. Which translates in the Laplace domain to:

$$\widehat{\Psi}_z = \frac{\widehat{\phi}_z}{1 - \widehat{\phi}_z} \tag{1.21}$$

Covariance Operator

In section 4.3.2, we define the infinitesimal covariance operator of a stationary Hawkes process N_t and link it to the decay kernel ϕ in prop. 4.3.1. Then, we use the stationary property of the Hawkes process to define its (normalized) **covariance operator**, for any scale $h >$ and lag $\tau \in \mathbb{R}$:

$$v_\tau^{(h)} = \frac{1}{h} E\left((\int_0^h dN_s - \Lambda h)(\int_\tau^{\tau+h} dN_s - \Lambda h) \right) \tag{1.22}$$

The main result of the paper, theorem 4.3.2, will allow us to use the covariance operator in practice in order to retrieve the decay kernel. Here, we restate it in the 1D case:

Theorem 1.6.1. *Let* $g_t^{(h)} = (1 - \frac{|t|}{h})^+$. $v_\tau^{(h)}$ *can be expressed as a function of* $g_\tau^{(h)}$ *and* Ψ_τ*:*

$$v_\tau^{(h)} = \Lambda g_\tau^{(h)} \star (\delta_\tau + \widetilde{\Psi}_\tau + \Psi_\tau + \widetilde{\Psi} \star \Psi_\tau) \tag{1.23}$$

where δ *is the dirac distribution and* $\widetilde{\Psi}_\tau = \Psi_{-\tau}$

Which gives in the Laplace domain the following corollary:

Corollary. *In Laplace domain equation* (1.23) *becomes:*

$$\widehat{v}_z^{(h)} = \Lambda \widehat{g}_z^{(h)} |1 + \widehat{\Psi}_z|^2 \tag{1.24}$$

We now go over the method used to estimate ϕ from the empirical observation of $v_\tau^{(h)}$.

Estimation

Equation (1.24), shows that if we calculate $v_\tau^{(h)}$ and take its Fourier transform we get the module of $1 + \widehat{\Psi}_{i\omega}$ multiplied by $\widehat{g}_{i\omega}^{(h)} = h\text{sinc}^2(\frac{\omega h}{2\pi})$, which is null for all ω of the form $\frac{2n\pi}{h}$, $n \in \mathbb{Z}$, $n \neq 0$ (sinc is the sin cardinal function defined by $\text{sinc}(t) = \frac{\sin(\pi t)}{\pi t}$). This problem is dealt with from a practical point by using a sampling period Δ for τ in the empirical estimation of $v_\tau^{(h)}$ small enough and by setting $\Delta = h$. In this case $\widehat{g}_{i\omega}^{(h)}$ does not become 0 on $[-\pi/\Delta, \pi/\Delta]$ and we can simply divide by it (cf section 4.4.1). This allows us to observe an approximation of

$$|1 + \widehat{\Psi}_{i\omega}|^2 = \frac{\widehat{v}_{i\omega}^{(h)}}{\Lambda \widehat{g}_{i\omega}^{(h)}} \tag{1.25}$$

And now we are left with the problem of finding $1 + \widehat{\Psi}_{i\omega}$ from $|1 + \widehat{\Psi}_{i\omega}|^2$. Recuperating a frequency response function from the sole observation of its amplitude, is equivalent to finding the corresponding phase. Under certain conditions, it is possible to find a particular frequency response whose amplitude is equal to $|1 + \widehat{\Psi}_{i\omega}|^2$.

In particular, thanks to assumption $(H2)$ we have that $1+\widehat{\Psi}_z$ satisfies the **Paley-Wiener criterion** (Theorem 4.4.2):

$$\int_{\mathbb{R}} \frac{\log(|\widehat{f}_{i\omega}|)}{1+\omega^2} d\omega \quad < \quad \infty, \tag{1.26}$$

and we can find $1+\widehat{\Psi}_{i\omega}$ using the Hilbert transform. In this case $1+\widehat{\Psi}_{i\omega}$ is given by (cf Equation 4.28):

$$1+\widehat{\Psi}_{i\omega} = e^{\log(|1+\widehat{\Psi}_{i\omega}|)-iH(\log(1+\widehat{\Psi}_{i\omega}))} \tag{1.27}$$

where the operator $H(.)$ refers to the Hilbert transform.

The estimation method in the 1D case can be summarized as follows:

Main steps for kernel estimation

- Set Δ,
- Estimate the unconditional intensity Λ,
- Estimate the auto-covariance operator $v_t^{(\Delta)}$ and compute its Fourier transform $\widehat{v}_{i\omega}^{(h)}$,
- Compute $|1+\widehat{\Psi}_{i\omega}|^2$ using Eq. (1.25),
- Compute $1+\widehat{\Psi}_{i\omega}$ using Eq. (1.27),
- Go back to the initial basis and Inverse Fourier transform to get the estimation of ϕ_t.

Applications

In sections 4.5 and 4.6, we apply the estimation method to 1D and 2D point processes. In particular, we illustrate the method for simulated Hawkes processes with exponential and power law decay kernels in section 4.5, whereas we use the trade arrival times data for the Bund futures in section 4.6. In this presentation, we reproduce some of the results from the 1D case. We go through the main results below.

Application to a Simulated Hawkes Process With an Exponential Decay Kernel

In section 4.5.1 we simulate a one dimensional exponential decay Hawkes process such that

$$\phi_t \quad = \quad \alpha e^{-\beta t} \mathbb{1}_{t \geq 0}$$

An example of a simulated intensity is shown in Figure 1.6.3. The result of the fit is shown in Figure 1.8. The original ϕ is shown in red and the estimated one is in blue. We see that the estimated ϕ^e is indeed causal, and looks to be a noisy version of the real kernel. Smoothing techniques could be applied to the causal part of ϕ^e that might effectively reduce the error, for example a moving average or a low pass filter. However we have not studied the benefits of such options and do not further explore them in the course of this thesis.

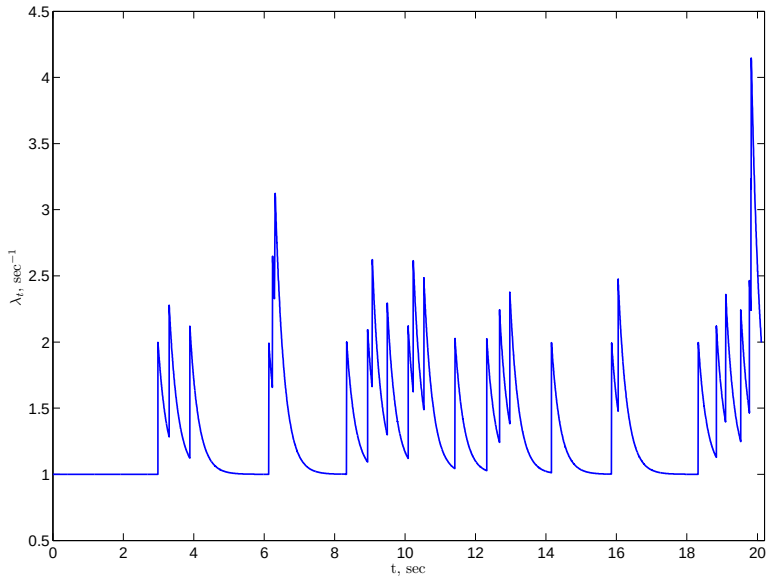


Figure 1.7: Realized Intensity λ_t of a simulated $1D$ Hawkes with $\mu = 1$, $\alpha = 1$, $\beta = 4$. We can clearly see in this case that the process is self exciting because the arrival of an event increases λ_t which triggers the arrival of more events.

Application to the Process of Trade Arrivals

When it comes to empirical data, the method is valuable because it gives an idea of the form of the decay function, which can be complemented with a parametric estimation. In the case of financial data, a power law with slow decay function is expected and our estimations confirm that. This is seen in section 4.6. We test the method for the point process of trade arrival times of Bund futures. The result is shown in Figure 1.9 for $\Delta = 0.1$. We get a causal ϕ in the form of a power law

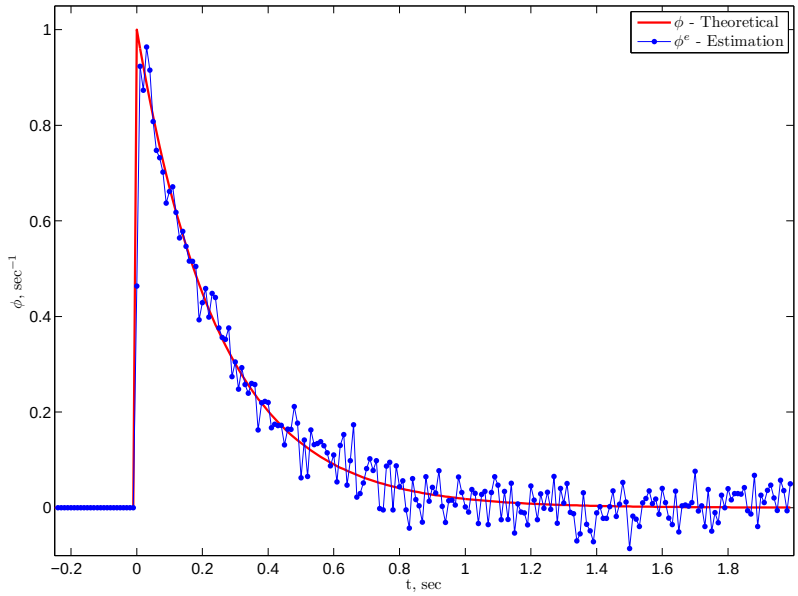


Figure 1.8: Non parametric estimation of the one dimensional Hawkes exponential kernel ϕ ((4.35)) with $\alpha = 1$, $\beta = 4$. We used $\Delta = 0.01$.

decay with an exponent close to -1.

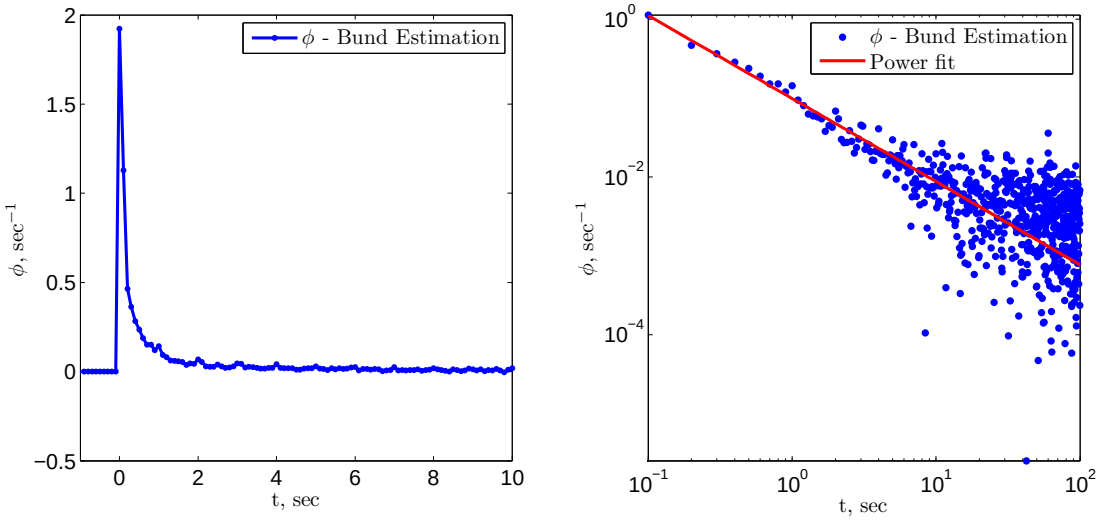


Figure 1.9: The plot on the left is the non parametric estimation of the Hawkes kernel assumed for the rate of incoming market orders of the Bund Futures. We used $\Delta = 0.1$ and $\tau_{\max} = 100$. On the right is that very same plot in log-log scale and the corresponding power law fit ax^b. We find $a = 0.09863$, $b = -1.053$.

Error Analysis

When it comes to the efficiency of the method, we point out in section 4.5.1, paragraph **Error Analysis** several properties of the error series. Most of the arguments presented here or in that section are heuristic and backed by empirical evidence. Mathematical proofs are either provided in [10] or need to be confirmed in future work.

In particular, we find that, when observed at $k\Delta$ instances, the series of errors, $(x_{k\Delta})_{k>0}$, defined by $x_{k\Delta} = \phi_{k\Delta} - \phi^e_{k\Delta}$ is a gaussian white noise, which confirms our thoughts that it can simply be removed with a white noise filter. Moreover, for a fixed Δ, we find that the L^2 estimation error

$$e^2 = \sum_{k=1}^{\frac{\tau_{\max}}{\Delta}} |\phi_{k\Delta} - \phi^e_{k\Delta}|^2 \tag{1.28}$$

decreases with the total length of the sample T, as T^{-1} (Figure 4.3).

It is also interesting to that e^2 depends on the sampling rate Δ (for T fixed). If Δ is large, then we do not have enough data points observed for $v^{(h)}_\tau$ and e^2 increases as information is lost and the Fourier transforms do not converge. If Δ is small, then the variations of the point process become uncorrelated ("Epps Effect") and the noise to signal ratio in the empirical $v^{(h)}_\tau$ becomes high, which translates into more noise in the estimation of ϕ. Indeed, Figure 4.5 shows just that, and we are able to deduce an optimal Δ that minimizes the variance.

Conclusion

In chapter 4, we adapt the estimation method to multidimensional point processes. We are able to estimate the decay kernels in the very special case of a process whose components are invariant under arbitrary permutations (cf section 4.4.2, assumption $(H3)$). In this case we are able to diagonalize the fourier transform of the covariance operator $\widehat{v}^{(h)}_{i\omega}$ using a constant orthogonal base (which does not depend on ω), and apply the 1D estimation method to its eigenvalues. Please refer to section 4.4.3 for details about the estimation process and sections 4.5.2 and 4.6.2 for an illustration with simulated data and symmetrical financial data like the series of arrivals of buy and sell orders.

CHAPTER 2

The Nature of Price Returns During Periods of High Market Activity

By studying all the trades and best bids/asks of ultra high frequency snapshots recorded from the order books of a basket of 10 futures assets, we bring qualitative empirical evidence that the impact of a single trade depends on the intertrade time lags. We find that when the trading rate becomes faster, the return variance per trade or the impact, as measured by the price variation in the direction of the trade, strongly increases. We provide evidence that these properties persist at coarser time scales. We also show that the spread value is an increasing function of the activity. This suggests that order books are more likely empty when the trading rate is high.

2.1 Introduction

During the past decade, the explosion of the amount of available data associated with electronic markets has permitted important progress in the description of price fluctuations at the microstructure level. In particular the pioneering works of Farmer's group [54, 38, 43, 37] and Bouchaud et al. [22, 21, 33] relying on the analysis of orderbook data, has provided new insights in the understanding of the complex mechanism of price formation (see e.g [20] for a recent review). A central

quantity in these works and in most approaches that aim at modeling prices at their microscopic level, is the market impact function that quantifies the average response of prices to "trades". Indeed, the price value of some asset is obtained from its cumulated variations caused by the (random) action of sell/buy market orders. In that respect, the price dynamics is formulated as a discrete "trading time" model like:

$$p_n = \sum_{i<n} G(n-i, V_i)\varepsilon_i + diffusion \tag{2.1}$$

where n and i are transaction "times", i.e., integer indices of market orders. V_i is the quantity traded at index i, ε_i is the sign of the i^{th} market order ($\varepsilon_i = -1$ if selling and $\varepsilon_i = +1$ if buying). The function $G(k, V)$ is the bare impact corresponding to the average impact after k trades of a single trade of volume V. Among all significant results obtained within such a description, one can cite the weak dependence of impact on the volume of market orders, i.e., $G(n, V) \sim G(n) \ln V$, the long-range correlated nature of the sign of the consecutive trades ε_i and the resulting non-permanent power-law decay of impact function $G(n)$ [20]. Beyond their ability to reproduce most high frequency stylized facts, models like (2.1) or their continuous counterparts [6] have proven to be extremely interesting because of their ability to control the market impact of a given high frequency strategy and to optimize its execution cost [40].

Another well known stylized fact that characterizes price fluctuations is the high intermittent nature of volatility. This feature manifests at all time scales, from intradaily scales where periods of intense variations are observed, for instance, around publications of important news to monthly scales [23]. Since early works of Mandelbrot and Taylor [57], the concept of subordination by a trading or transaction clock that maps the physical time to the number of trades (or the cumulated volume) has been widely used in empirical finance as a way to account for the volatility intermittency. The volatility fluctuations simply reflects the huge variations of the activity. The observed intradaily seasonal patterns [28] can be explained along the same line. Let us remark that according to the model (2.1), the physical time does not play any role in the way the market prices vary from trade to trade. This implies notably that the variance per trade (or per unit of volume traded) is constant and therefore that the volatility over a fixed physical time scale, is only dependent on the number of trades.

The goal of this paper is to critically examine this underlying assumption associated with the previously quoted approaches, namely the fact that the impact of a trade does not depend in any way on the physical time elapsed since previous transaction. Even if one knows that volatility is, to a good approximation, propor-

tional to the number of trades within a given time period (see Section 2.3), we aim at checking to what extent this is true. For that purpose we use a database which includes all the trades and **level 1** (i.e., best ask and best bid) ultra high-frequency snapshots recorded from the orderbooks of a basket of 10 futures assets. We study the statistics of return variations associated to one trade conditioned by the last intertrade time. We find that the variance per trade (and the impact per trade) increases as the speed of trading increases and we provide plausible interpretations to that. We check that these features are also observed on the conditional spread and impact. Knowing that the spread is a proxy to the fullness of the book and the available liquidity [87], we suspect that in high activity periods the orderbooks tend to deplete. These "liquidity crisis" states would be at the origin of considerable amounts of variance not accounted for by transaction time models.

The paper is structured as follows: In Section 2.2 we describe the futures data we used and introduce some useful notations. In Section 2.3, we study the variance of price increments and show that if it closely follows the trading activity, the variance per trade over some fixed time interval is not constant and increases for strong activity periods. Single trade variance of midpoint prices conditioned to the last intertrade duration are studied in Section 2.4. We confirm previous observations made over a fixed time interval and show that, as market orders come faster, their impact is greater. We also show that, for large tick size assets, the variations of volatility for small intertrade times translates essentially on an increase of the probability for a trade to absorb only the first level of the book (best bid or best ask). There is hardly no perforation of the book on the deeper levels.

In Section 2.5.1 we show that the single trade observations can be reproduced at coarser scales by studying the conditional variance and impact over 100 trades. We end the section by looking at the spread conditioned to the intertrade durations. This allows us to confirm that in period of high activity, the orderbook tends to empty itself and therefore the increase in the trading rate corresponds to a local liquidity crisis. Conclusions and prospects are provided in Section 2.6.

2.2 Data Description

In this paper, we study highly liquid futures data, over two years during the period ranging from 2008/08 till 2010/03. We use data of ten futures on different asset classes that trade on different exchanges. On the EUREX exchange (localized in Germany) we use the futures on the DAX index (DAX) and on the EURO STOXX 50 index (ESX), and three interest rates futures: 10-years Euro-Bund (Bund), 5-years Euro-Bobl (Bobl) and the 2-years Euro-Schatz (Schatz). On the CBOT exchange (localized in Chicago), we use the futures on the Dow Jones index (DJ) and the 5-Year U.S. Treasury Note Futures (BUS5). On the CME (also in

Chicago), we use the forex EUR/USD futures (EURO) and the the futures on the SP500 index (SP). Finally we also use the Light Sweet Crude Oil Futures (CL) that trades on the NYMEX (localized in New-York). As for their asset classes, the DAX, ESX, DJ, and SP are equity futures, the Bobl, Schatz, Bund, and BUS5 are fixed income futures, the EURO is a foreign exchange futures and finally the CL is an energy futures.

The date range of the DAX, Bund and ESX spans the whole period from 2008/08 till 2010/03, whereas, for all the rest, only the period ranging from 2009/05 till 2010/03 was available to us. For each asset, every day, we only keep the most liquid maturity (i.e., the maturity which has the maximum number of trades) if it has more than 5000 trades, if it has less, we just do not consider that day for that asset. Moreover, for each asset, we restrict the intraday session to the most liquid hours, thus for instance, most of the time, we close the session at settlement time and open at the outcry hour (or what used to be the outcry when it no longer exists). We refer the reader to Table 2.1 for the total number of days considered for each asset (column D), the corresponding intraday session and the average number of trades per day. It is interesting to note that we have a dataset with a variable number of trading days (around 350 for the DAX, Bund and ESX, and 120 for the rest) and a variable average number of orders per day, varying from 10 000 trades per day (Schatz) to 95 000 (SP).

Our data consist of **level 1** data : every single market order is reported along with any change in the price or in the quantity at the best bid or the best ask price. All the associated timestamps are the timestamps published by the exchange (reported to the millisecond).

It is important to point out that, since when one market order hits several limit orders it results in several trades being reported, we chose to aggregate together all such transactions and consider them as one market order. We use the sum of the volumes as the volume of the aggregated transaction and as for the price we use the last traded price. In our writing we freely use the terms transaction or trade for any transaction (aggregated or not). We are going to use these transactions as our "events", meaning that all relevant values are calculated at the time of, or just before such a transaction. As such, we set the following notations:

Notations 2.2.1. For every asset, let D be the total number of days of the considered period. We define:

(i) $N_k, k \in \{1 \dots D\}$ the total number of trades on the k^{th} day

(ii) t_i is the time of the i^{th} trade ($i \in [1, \sum_k N_k]$)

(iii) b_{t_i} and a_{t_i} are respectively the best bid and ask prices right before the i^{th} trade

(iv) $p_{t_i} = \frac{b_{t_i}+a_{t_i}}{2}$ is midpoint price right before the i^{th} trade

(v) $s_{t_i} = a_{t_i} - b_{t_i}$ is spread right before the i^{th} trade

(vi) $r_{t_i} = p_{t_{i+1}} - p_{t_i}$ is the return caused by the i^{th} trade, measured in ticks

(vii) $NT[s,t] = \#\{t_i, s \leq t_i < t\}$ corresponds to the number of trades in the time interval $[s,t]$

(viii) $E_t[...]$ or $E_i[...]$ indifferently refers to the historical average of the quantity in between backets, averaging on all the available days and on all the trading times $t = t_i$. The quantity is first summed up separately on each day (avoiding returns overlapping on 2 consecutive days), then the so-obtained results are summed up and finally divided by the total number of terms in the sum.

Let us note that in the whole paper, we will consider that the averaged returns are always 0, thus we do not include any mean component in the computation of the variance of the returns.

Futures	Exchange	Tick Value	D	Session	# Trades/Day	1/2-η	P_0	$P_=$
DAX	EUREX	12.5€	349	8:00-17:30	56065	0.082	49	67.9
CL	NYMEX	10$	127	8:00-13:30	76173	0.188	72.8	79.8
DJ	CBOT	5$	110	8:30-15:15	36981	0.227	72.6	92.2
BUS5	CBOT	7.8125$	126	7:20-14:00	22245	0.288	81.6	95.1
EURO	CME	12.5$	129	7:20-14:00	42271	0.252	79.5	95.2
Bund	EUREX	10€	330	8:00-17:15	30727	0.335	80.9	97.6
Bobl	EUREX	10€	175	8:00-17:15	14054	0.352	86.5	99.1
ESX	EUREX	10€	350	8:00-17:30	55083	0.392	88.3	99.2
Schatz	EUREX	5€	175	8:00-17:15	10521	0.385	89.3	99.4
SP	CME	12.5$	112	8:30-15:15	97727	0.464	96.6	99.8

Table 2.1: Data Statistics. The assets are listed from top to bottom following the increasing order of the $P_=$ column (see (2.2)), i.e., from the smaller (top) to the greater (bottom) "perceived" tick size. **D** : number of days that are considered. The **Tick Value** is the smallest variation (expressed in the local currency) by which a trading price can move. The **Session** column indicates the considered trading hours (local time). The # **Trades/Day** is the average of the daily number of trades (i.e., $\sum_{k=1}^{D} N_k/D$ using Notations 2.2.1). P_0 and $P_=$ are defined in equations (2.4) and (2.2) and reported here in percent.

"Perceived" Tick Size and Tick Value

The *tick value* is a standard characteristic of any asset and is measured in its currency. It is the smallest increment by which the price can move. In all the following, all the price variations will be normalized by the tick value to get them expressed in ticks (i.e., in integers for price variations and half-integers for midpoint-price variations).

As one can see in Table 2.1, column **Tick Value**, our assets have very different tick values. It is important to note a counter-intuitive though very well known fact : the tick value *is not* a good measure of the **perceived size** (by pratitionners) of the tick. A trader considers that an asset has a small tick when he "feels" it to be negligible, consequently, he is not averse at all to price variations of the order of a single tick. For instance, every trader "considers" that the the ESX index futures has a much greater tick than the DAX index futures though the tick values are of the same orders ! There have been several attempts to quantify the perceived tick size. Kockelkoren, Eisler and Bouchaud in [33], write that "large tick stocks are such that the bid-ask spread is almost always equal to one tick, while small tick stocks have spreads that are typically a few ticks". Following these lines, we calculate the number of times (observed at times t_i) the spread is equal to 1 tick:

$$P_{=} = \frac{\#\{i, s_{t_i} = 1\}}{N} \tag{2.2}$$

and show the results in Table 2.1. We classify our assets according to this criterion and find SP to have the largest tick, with the spread equal to 1 99.8% of the time, and the DAX to have the smallest tick.
In a more quantitative approach, in order to quantify the aversion to price changes, Rosenbaum and Robert in [76] give a proxy for the perceived tick size using last traded non null returns time-series. If N_t^a (resp. N_t^c) is the number of times a trading price makes two jumps in a row in the same (resp. different) directions, then the perceived tick size is given by $1/2 - \eta$ where η is defined by

$$\eta = \frac{N_t^c}{2N_t^a} \tag{2.3}$$

For each asset, we computed η for every single day, and average over all the days in our dataset and put the result in the $1/2 - \eta$ column in Table 2.1. We find that the rankings of the assets using this criterion almost matches the ranking using [33]'s $P_{=}$ criterion (two slight exceptions being the ESX/Schatz and BUS5/EURO ranking). One interpretation of the η based proxy is that if the tick size is large, market participant are more averse to changes in the midpoint price and market

makers are happy to keep the spread collapsed to the minimum and the midpoint would only move when it becomes clear that the current price level is unsustainable. To check that, we calculate the number of times (observed at times t_i) the return (as defined in notation 2.2.1) is null:

$$P_0 = \frac{\#\{i, r_{t_i} = 0\}}{N} \tag{2.4}$$

and show the result in Table 2.1. Again, it approximately leads to the same ranking which has nothing to do with the ranking using Tick Values.

2.3 Realized Variance Versus Number of Trades

It is widely known that, in a good approximation, the variance over some period of time is proportional to the number of trades during that time period (see e.g. [28]). Figure 2.1 illustrates this property on Bund data. On 15 minutes intraday intervals, averaging on every single day available, we look at (dashed curve) the average intraday rate of trading (i.e., the average number of trades per second) and (solid curve) the average (15 minutes) realized variance (estimated summing on 1mn-squared returns $(p_{t+1mn} - p_t)^2$). We see that the so-called intraday seasonality of the variance is highly correlated with the intraday seasonality of the trading rate [28].

In order to have more insights, we look at some daily statistics : Figure 2.2 shows a scatter plot in which each point corresponds to a given day k whose abscissa is the number of trades within this day, i.e., N_k, and the ordinate is the daily variance (estimated summing over 5-mn quadratic returns) of the same day k. It shows that, despite some dispersion, the points are mainly distributed around a mean linear trend confirming again the idea shown in Figure 2.1 that, to a good approximation, the variance is proportional to the number of trades. In that respect, trading time models (Eq. (2.1)) should capture most of the return variance fluctuations through the dynamics of the transaction rate. However, in Figure 2.2, the points with high abscissa values (i.e., days with a lot of activity) tend to be located above the linear line, whereas the ones with low abscissa (low activity) cluster below the linear line, suggesting that the variance per trade is dependant on the daily intensity of trading.

Before moving on, we need to define a few quantities. Let Δt be an intraday time scale and let N be a number of trades. We define $V(\Delta t, N)$ as the estimated price variance over the scale Δt conditioned by the fact that N trades occurred. Using notations, 2.2.1 (*vii*) and (*viii*), from a computational point of view, when

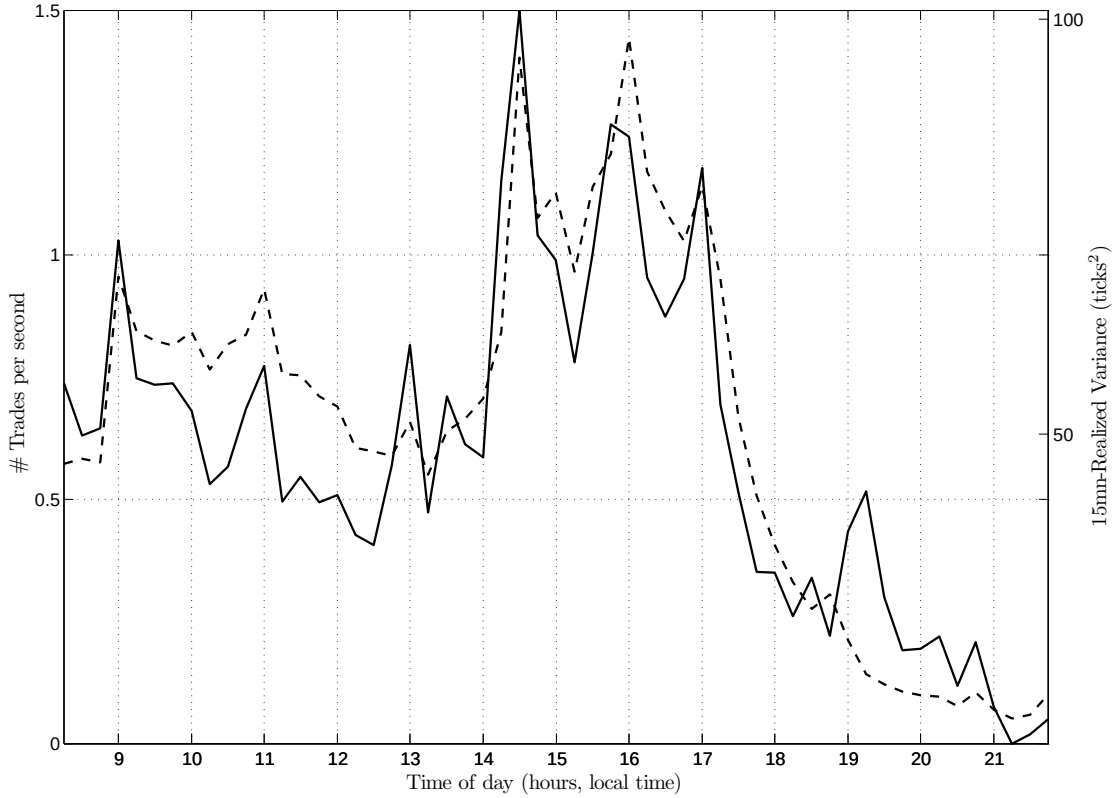


Figure 2.1: Bund intraday seasonality of both trading rate and volatility (abscissa are expressed in hours, local time). Averages are taken on all available days. Dashed line : average intraday rate of trading (average number of trades per second) using 15mn bins. Solid line : average 15-minutes-realized variance (estimated summing on 15 1mn-squared returns).

$\Delta t = \Delta t_0$ is fixed and N is varying, $V(\Delta t = \Delta t_0, N)$ is estimated as:

$$V(\Delta t = \Delta t_0, N) = E_t \left[(p_{t+\Delta t_0} - p_t)^2 \mid NT[t, t+\Delta t_0] \in [N - \delta_N, N + \delta_N] \right]. \quad (2.5)$$

where δ_N is some bin size. And, along the same line, when we study $V(\Delta t, N)$ for a fixed $N = N_0$ value over a range of different values of Δt, one defines a temporal bin size $\delta_{\Delta t}$ and computes $V(\Delta t, N = N_0)$ as[1]

$$V(\Delta t, N = N_0) = E_i \left[(p_{t_{i+N_0}} - p_{t_i})^2 \mid t_{i-1+N_0} - t_{i-1} \in [\Delta t - \delta_{\Delta t}, \Delta t + \delta_{\Delta t}] \right]. \quad (2.6)$$

Let us note that, in both cases, the bins are chosen such that each bin in-

[1]Let us point out that we used the index $i-1$ in the condition of (2.6) and not the index i since, for the particular case $N_0 = 1$ (extensively used in Section 2.4), we want to use a *causal* conditioning of the variance. For N_0 large enough, using one or the other does not really matter.

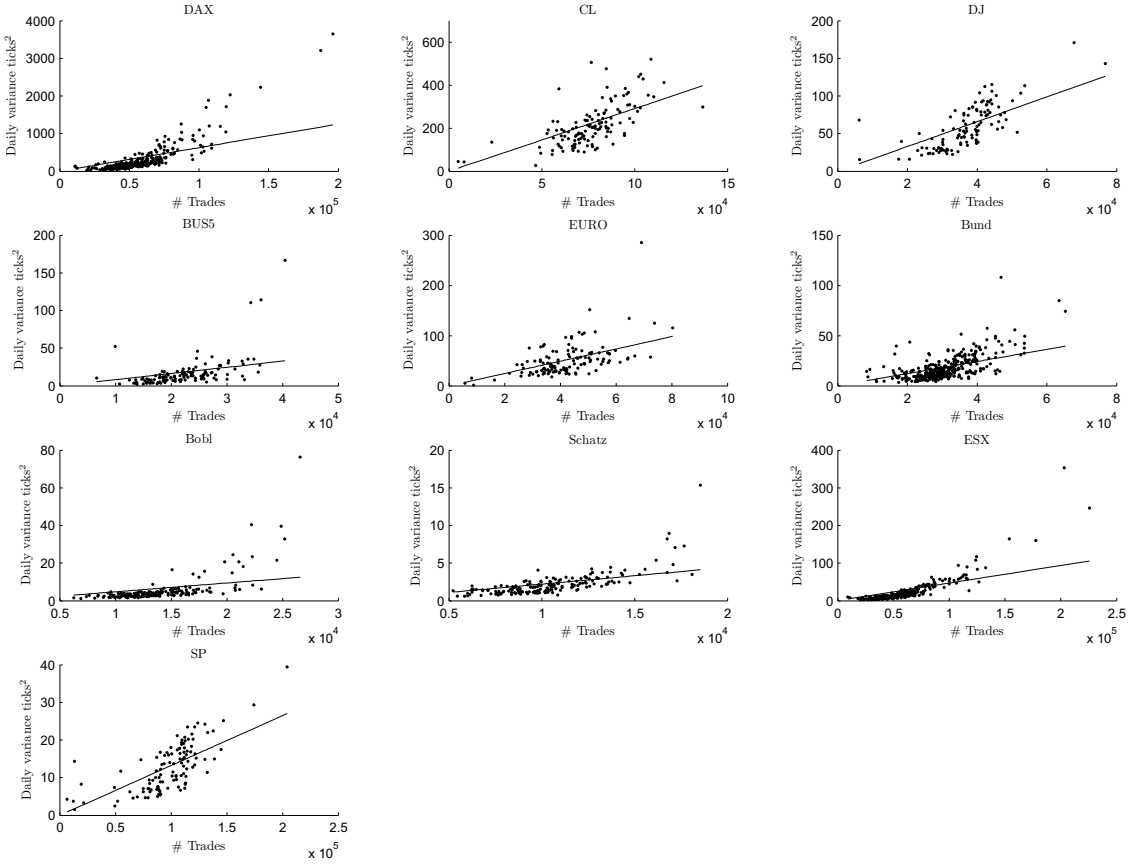


Figure 2.2: For each asset (in increasing perceived tick size $P_=$) : Daily variance (estimated summing over 5-mn quadratic returns) against daily number of trades. Each dot represents a single day. The solid line is the linear regression line with zero intercept. We see strong linearity between the variance and the number of trades but there seem to be clustering of dots above (resp. below) the solid line for days with high (resp. low) activity.

volves approximately the same number of terms. We also define the corresponding conditional **variance per trade** as:

$$v(\Delta t, N) = \frac{V(\Delta t, N)}{N}. \tag{2.7}$$

In order to test the presence of an eventual non-linear behavior in the last scatter plots (Figure 2.2), we show in Figure 2.3 the 5-minutes variance per trade $v(\Delta t = 5mn, N)$ as a function of the average intertrade duration $\frac{5mn}{N}$ as N is varying. We clearly see that the estimated curve (solid line) is below the simple average variance (dashed line) for large intertrade durations and above the average variance when the trades are less than 600 milliseconds apart. Note that we observed a similar

behavior for most of the futures suggesting a universal behavior.

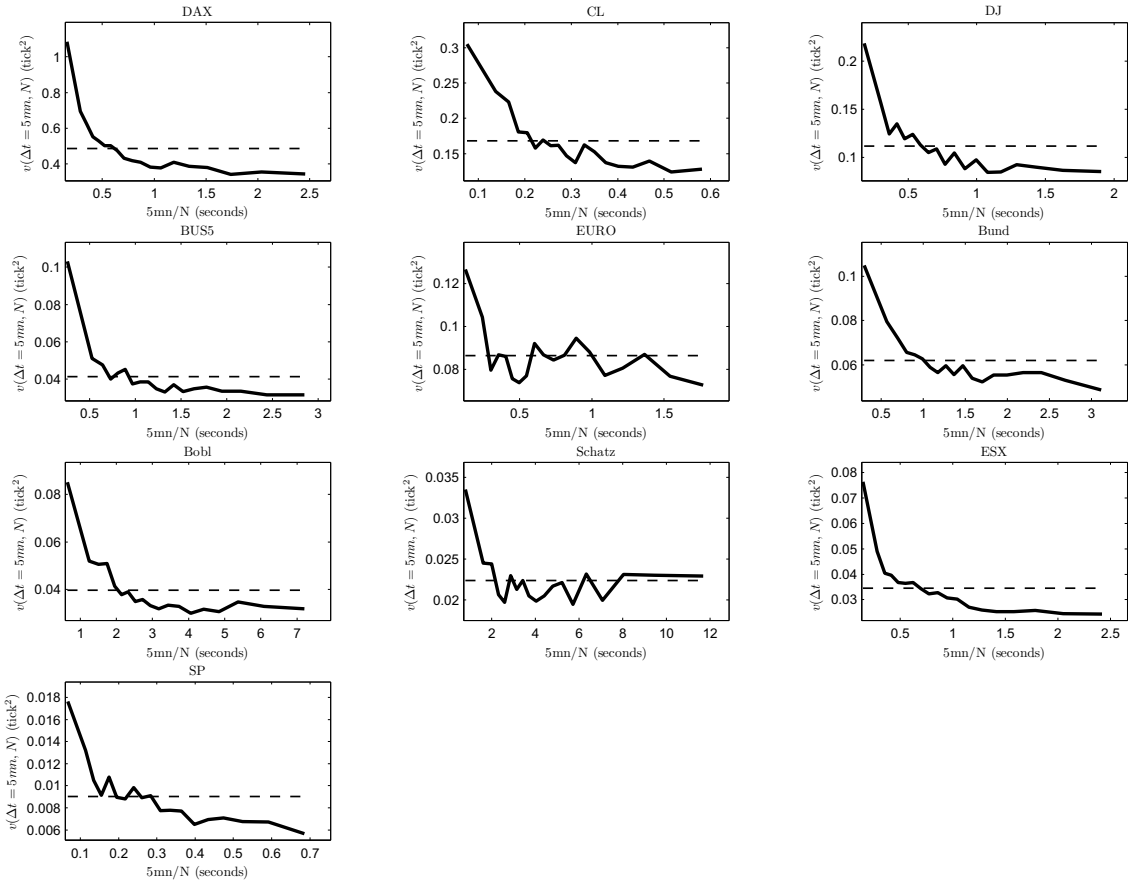


Figure 2.3: For each asset (in increasing perceived tick size $P_=$), Solid line : conditional $v(\Delta t = 5mn, N)$ variance per trade (see (2.5)) as a function of the average intertrade duration $\frac{5mn}{N}$ when varying N. Dashed line : unconditional 5mn-variance per trade. The solid line is almost constant for average times above 0.6 seconds, and it increases when the trading becomes faster.

To say that the realized variance is proportional to the number of trades is clearly a very good approximation as long as the trading activity is not too high as shown both on a daily scale in Figure 2.2 and on a 5mn-scale in Figure 2.3. However, as soon as the trading activity is high (e.g., average intertrade duration larger than $600ms$ on a 5mn-scale), the linear relationship seems to be lost. In the next section we will focus on the impact associated with a single trade.

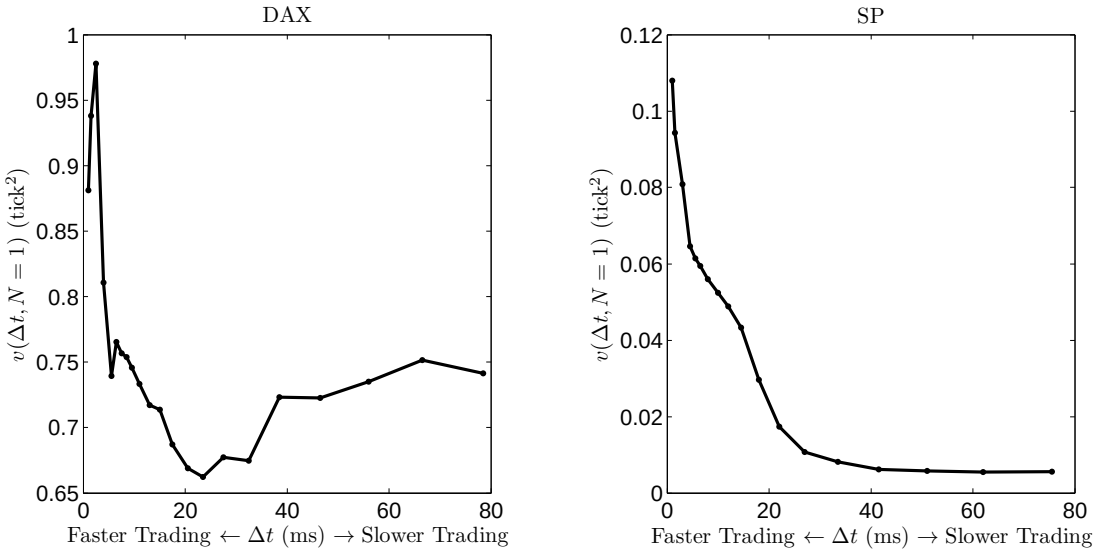


Figure 2.4: $v(\Delta t, N = 1)$ as a function of Δt over very short Δt's for DAX and SP. The variance per trade increases dramatically below a certain Δt.

2.4 Single Trade Impact on the Midpoint Price

In this section, we will mainly focus on the impact of a trade i, and more specifically on the influence of its arrival time t_i on the return $r_{t_i} = p_{t_{i+1}} - p_{t_i}$. In order to do so, it is natural to consider the return r_{t_i} conditioned by $t_i - t_{i-1}$, the time elapsed since previous transaction. We want to be able to answer questions such as : how do compare the impacts of the ith trade depending on the fact that it arrived right after or long after the previous trade ? Of course, in the framework of trading time models this question has a very simple answer : the impacts are the same ! Let us first study the conditional variance of the returns.

2.4.1 Impact on the Return Variance

In order to test the last assertion, we are naturally lead to use Eqs (2.6) and (2.7) for $N_0 = 1$, i.e,

$$v(\Delta t, N = 1) = E_i \left[r_{t_i}^2 \mid t_i - t_{i-1} \in [\Delta t - \delta_{\Delta t}, \Delta t + \delta_{\Delta t}] \right]. \tag{2.8}$$

Let us illustrate our purpose on the DAX and the SP futures. They trade on two different exchanges, (EUREX and CME) and have very different daily statistics (e.g., DAX has the smallest perceived tick and SP the largest as one can see in Table 2.1). Figure 2.4 shows for both assets $v(\Delta t, N = 1)$ (expressed in squared *tick*) as a function of Δt (in milliseconds). We notice that both curves present a

peak for very small Δt and stabilize around an asymptotic constant value for larger Δt. This value is close to 0.7 ticks^2 for the DAX and to 0.005 ticks^2 for the SP. The peak reaches 0.95 (35% above the asymptote) for the DAX, and 0.1 (2000% above the asymptote).

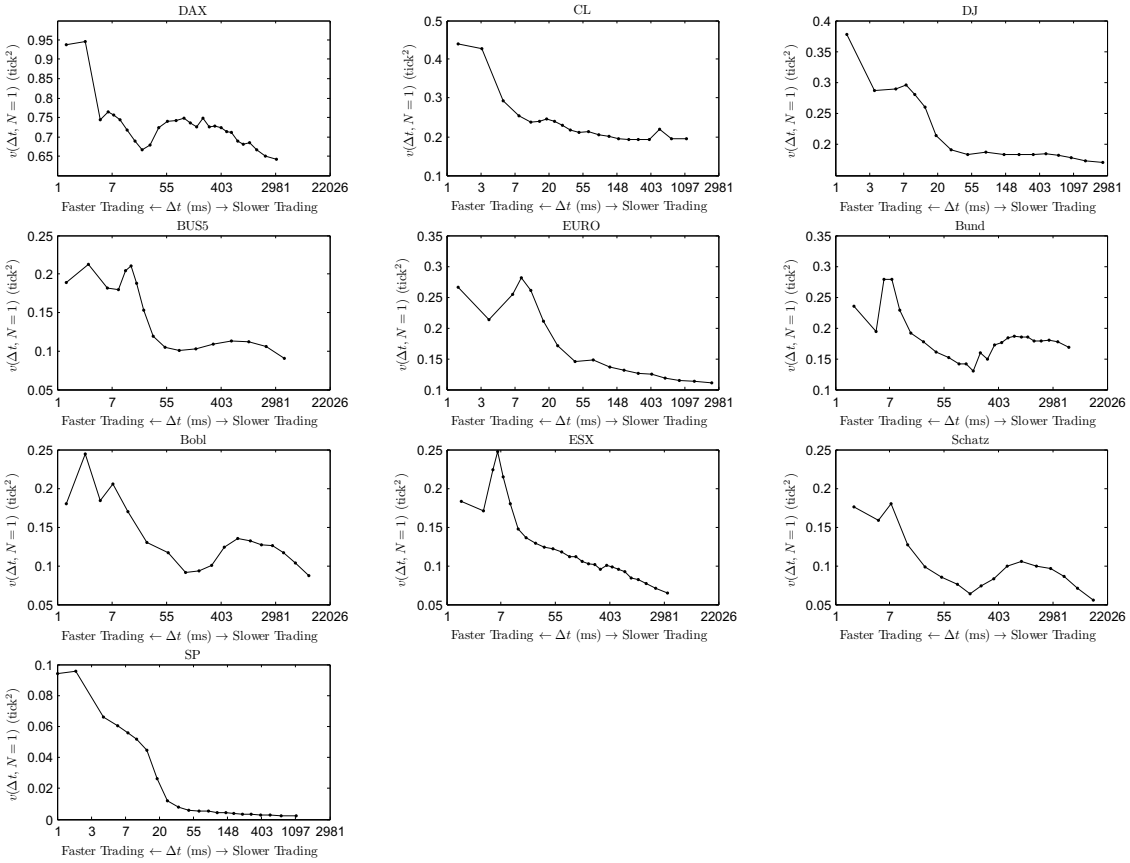


Figure 2.5: For each asset (in increasing perceived tick size $P_=$), $v(\Delta t, N = 1)$ as a function of Δt (logarithm scale). We see an "explosion" of the variance when the trading is getting faster.

Figure 2.5 switches (for all assets) to a log scale in order to be able to look at a larger time range. A quick look at all the assets show that they present a very similar behavior. One sees in particular for the ESX curve that the variance increases almost linearly with the rate of trading, and then suffers an explosion as Δt becomes smaller than 20 ms. The "same" explosion can be qualitatively observed over all assets albeit detailed behavior and in particular the minimal threshold Δt may vary for different assets.

Let us note that the variance $v(\Delta t, N = 1)$ as defined by (2.8) can be written in the following way:

$$v(\Delta t, N = 1) = P(\Delta t)A(\Delta t), \tag{2.9}$$

where $P(\Delta t)$ is the probability for the return to be non zero conditioned by the intertrade duration $t_i - t_{i-1} = \Delta t$, i.e.,

$$P(\Delta t) = Prob\{r_{t_i} \neq 0 \mid t_i - t_{i-1} \in [\Delta t - \delta_{\Delta t}, \Delta t + \delta_{\Delta t}]\} \tag{2.10}$$

and where $A(\Delta t)$ is the expectation of the squared return conditioned by the fact that it is not zero and by the intertrade duration $t_i - t_{i-1} = \Delta t$, i.e.,

$$A(\Delta t) = E_i \left[r_{t_i}^2 \mid r_{t_i} \neq 0 \text{ and } t_i - t_{i-1} \in [\Delta t - \delta_{\Delta t}, \Delta t + \delta_{\Delta t}] \right] \tag{2.11}$$

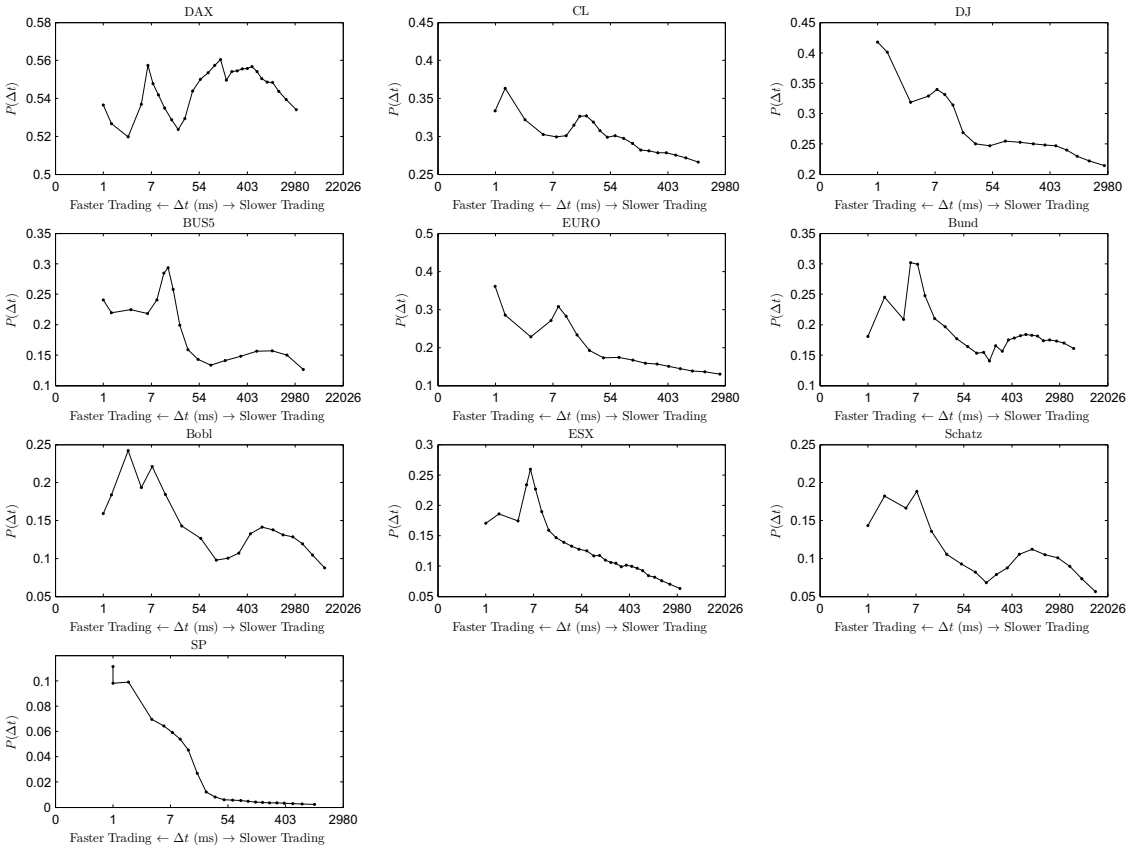


Figure 2.6: For each asset (in increasing perceived tick size $P_{=}$), Probability $P(\Delta t)$ as defined by (2.10) as a function of Δt. We see that the probability of getting a price move increases with market order rate for most assets.

In short $P(\Delta t)$ is the probability that the midpoint price moves while $A(\Delta t)$ is the squared amplitude of the move when non-zero. In Figure 2.6, we have plotted, for all assets, the function $P(\Delta t)$ for different Δt. One clearly sees that, as the trading rate becomes greater ($\Delta t \to 0$), the probability to observe a move of the

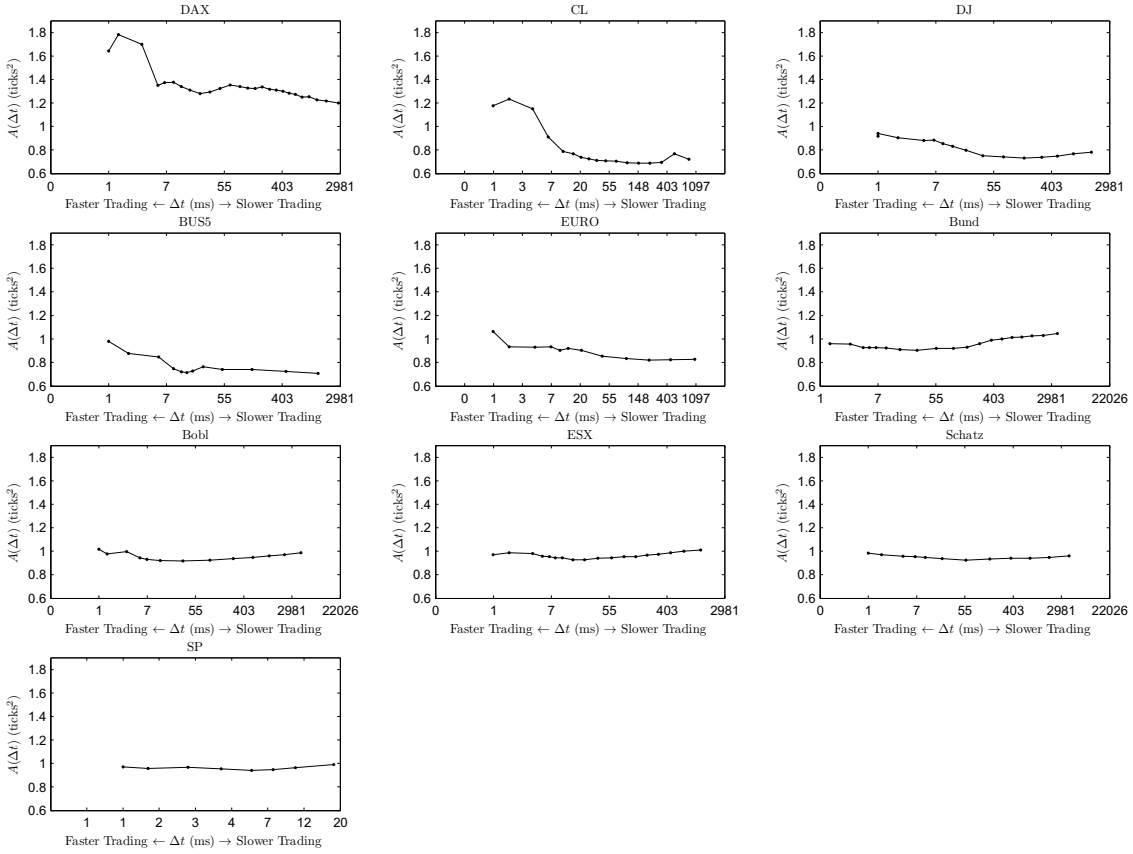


Figure 2.7: For each asset (in increasing perceived tick size $P_=$), Size of Absolute squared returns $A(\Delta t)$ as a function pf Δt. For very small tick assets, like DAX and CL we see that the absolute size of a return increases with the rate of market orders. This property quickly stops being true as the tick increases. The orderbook of a large tick asset is generally much thicker than that of a small tick asset and therefore it is extremely hard to to empty more than one level.

midpoint price increases. One mainly recovers the behavior we observed for the analog variance plots. Let us notice that (except for the DAX), the values of the moving probabilities globally decrease as the perceived ticks $P_=$ increases (for large ticks, e.g. SP, at very low activity this probability is very close to zero). The corresponding estimated moving squared amplitudes $A(\Delta t)$ are displayed in Figure 2.7. It appears clearly that, except for the smallest perceived ticks assets (DAX and CL basically), the amplitude can be considered as constant. This can be easily explained : large tick assets never make moves larger than one tick while small tick assets are often "perforated" by a market order. One can thus say that, except for very small ticks assets, the variance increase in high trading rate period is mostly caused by the increase of the probability that a market order absorb only the first level of the book (best bid or best ask). There is hardly no perforation of the book on the deeper levels.

2.4.2 Impact on the Return

Before moving to the next section, let us just look at the direct impact on the return itself, as defined for instance by [20], conditioned by the intertrade time:

$$I(\Delta t, N = 1) = E_i \left[\varepsilon_i r_{t_i} | t_i - t_{i-1} = \Delta t \right]. \tag{2.12}$$

According to [87], we expect the impact to be correlated with the variance per trade and therefore for $I(\Delta t)$ to follow a very similar shape to that of $v(\Delta t, N = 1)$ shown in Figure 2.5 . This is confirmed in Figure 2.8 where one sees that, for all assets, the impact goes from small values for large intertrade intervals to significantly higher values for small intertrade durations.

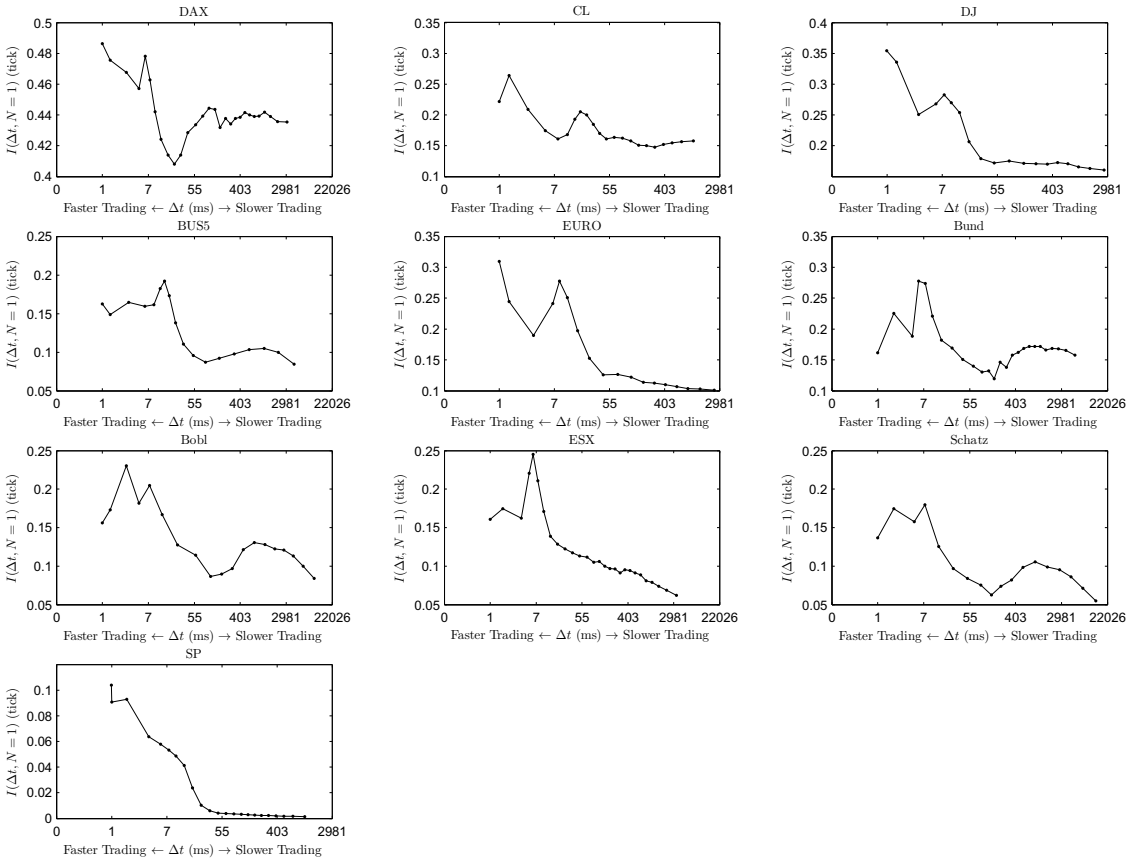


Figure 2.8: For each asset (in increasing perceived tick size $P_=$), $I(N = 1|\Delta t)$ as defined by (2.12) as a function of Δt. The shape of the curves confirms the idea that the impact is high correlated with the variance per trade.

2.5 From Fine to Coarse

2.5.1 Large Scale Conditional Variance and Impact

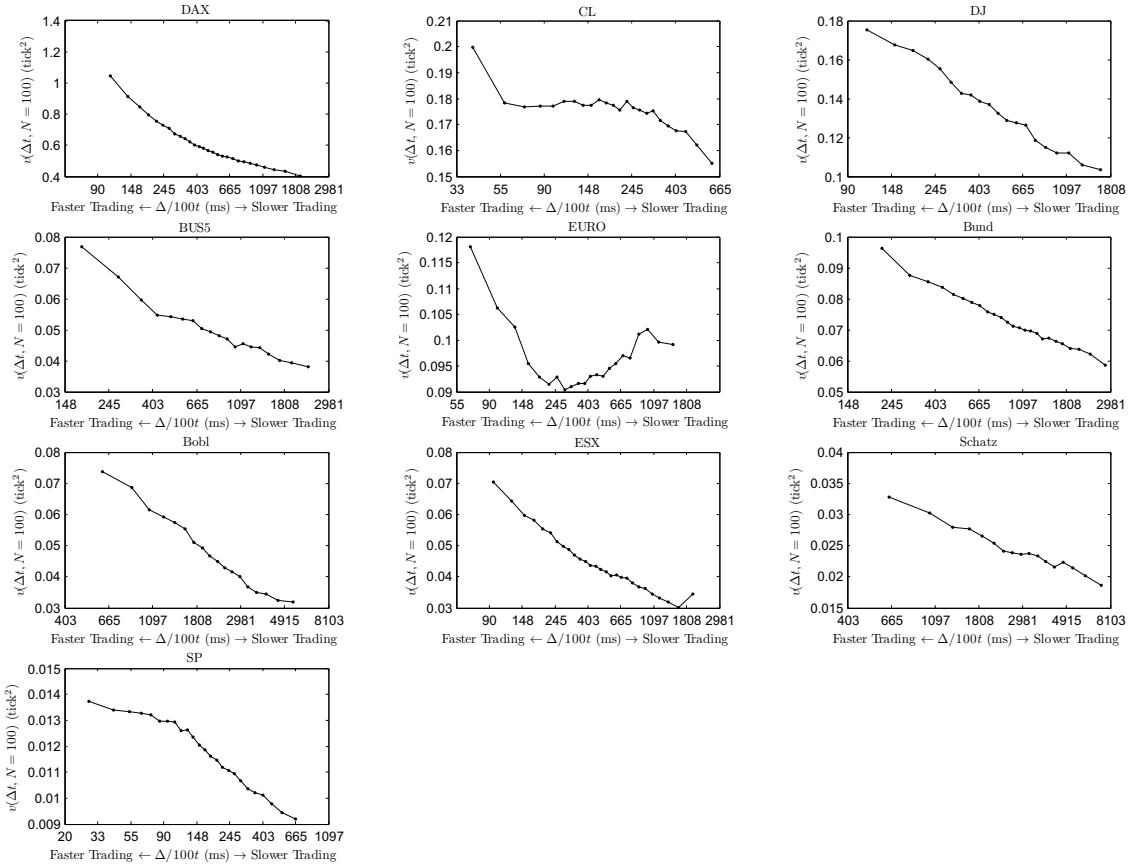


Figure 2.9: For each asset (in increasing perceived tick size $P_{=}$), $v(\Delta t, N = 100)$, as defined by (2.7), as a function of Δt. Clearly the variance of a speedy 100 trades is higher than the variance of 100 slow trades.

One of the key issue associated to our single trade study is the understanding of the consequences of our findings to large scale return behavior. This question implies the study of (conditional) correlations between successive trades, which is out of the scope of this paper and will be addressed in a forthcoming work. However one can check whether the impact or the variance averaged locally over a large number of trades still display a dependence with respect to the trading rate. Indeed, in Figure 2.3 we have already seen that this feature seems to persist when one studies returns over a fixed time (e.g., 5 min) period conditioned by the mean intertrade duration over this period. Along the same line, one can fix a large $N = N_0$ value and compute $v(\Delta t, N = N_0)$ and $I(\Delta t, N = N_0)$ as functions of Δt. Note that $v(\Delta t, N = N_0)$ is defined in Eq. (2.7) while the aggregated impact can

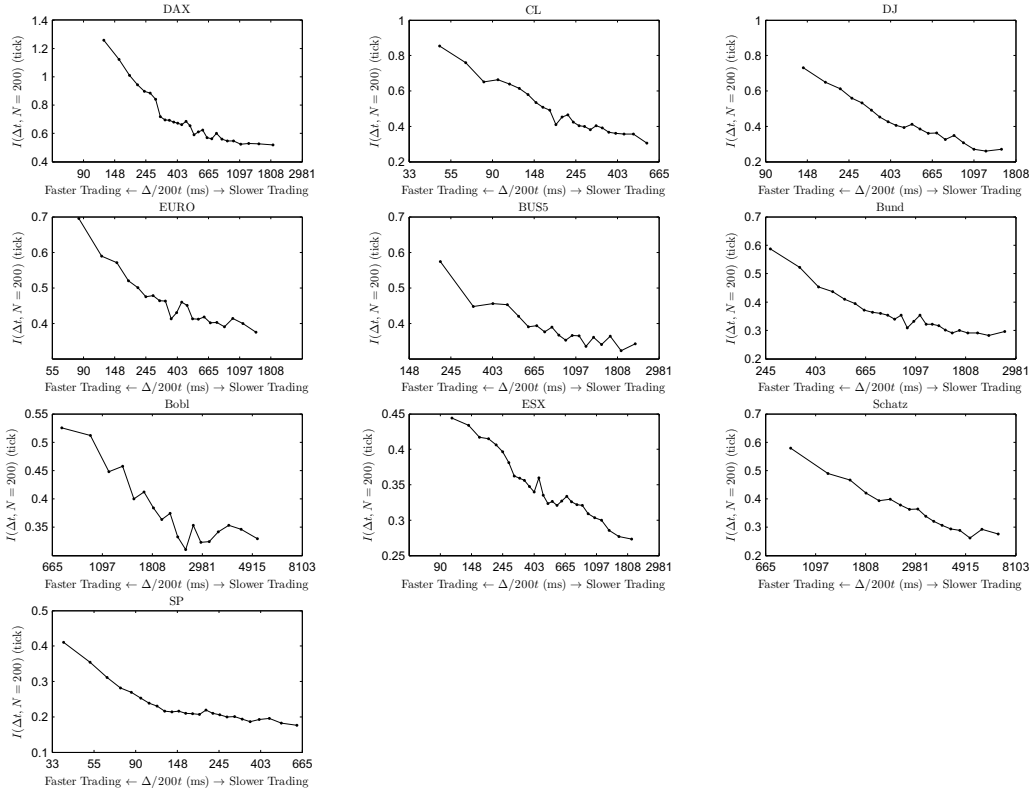


Figure 2.10: For each asset (in increasing perceived tick size $P_=$), $I(\Delta t, N = 200)$, as defined by (2.13), as a function of Δt. The impact of speedy trades propagates well into the future. Even 200 trades away, one speedy trade has caused more impact than a slower one.

be defined similarly as:

$$I(\Delta t, N = N_0) = E_i \left[\epsilon_i (p_{t_{i+N_0}} - p_{t_i}) \mid t_{i-1+N_0} - t_{i-1} \in [\Delta t - \delta_{\Delta t}, \Delta t + \delta_{\Delta t}] \right]. \tag{2.13}$$

In Figure 2.9 and 2.10 are plotted respectively the variance $v(\Delta t, N = 100)$ and the return impact $I(\Delta t, N = 200)$ as functions of Δt. One sees that at these coarse scales, the increasing of these two quantities as the activity increases is clear (except maybe for the variance of the EURO). As compared to single trade curves, the threshold-like behavior are smoothed out and both variance and return impacts go continuously from small to large values as the trading rate increases.

2.5.2 Liquidity Decreases When Trading Rate Increases

One possible interpretation of these results would be that when the trading rate gets greater and greater, the liquidity tends to decrease, i.e., the orderbook tends to empty.

In [87], the authors mention that the spread is an indicator of the thinness of the book and that the distance from the best bid or ask to the next level of the orderbook is in fact equivalent to the spread. Moreover, they bring empirical evidence and theoretical no-arbitrage arguments suggesting that the spread and the variance per trade are strongly correlated. Accordingly, we define the mean spread over N trades as

$$s_{t_i,N} = \frac{1}{N} \sum_{k=0}^{N-1} s_{t_{i+k}}, \tag{2.14}$$

and the conditional spread at the fixed scale $N = N_0$ as

$$S(\Delta t, N = N_0) = E_i \left[s_{t_i,N} \mid t_{i+N} - t_i \in [\Delta t - \delta_{\Delta t}, \Delta t + \delta_{\Delta t}] \right]. \tag{2.15}$$

Figure 2.11 displays, for each asset, $S(\Delta t, N = 100)$ as a function of $\Delta t/100$ (using log scale). One observes extremely clearly an overall increase of the spread value with the rate of trading for all assets, This clearly suggests that the orderbook is thinner during periods of intense trading. This seems to be a universal behavior not depending at all on the perceived tick size.

2.6 Concluding Remarks

In this short paper we provided empirical evidence gathered from high frequency futures data corresponding to various liquid futures assets that the impact (as measured from the return variance or using the standard definition) of a trading order on the midpoint price depends on intertrade duration. We have also shown that this property can also be observed at coarser scale. A similar study of the spread value confirmed the idea that orderbooks are less filled when trading frequency is very high. Notice that we did not interpret in any causal manner our findings, i.e., we do not assert that a high transaction rates should be responsible for the fact that books are empty. It just appears that both phenomena are highly correlated and this observation has to be studied in more details. In a future work, we also plan to study the consequences of these observations on models such those described in the introductory section (Eq. (2.1)). A better understanding of the aggregation properties (i.e., large values of N) and therefore of correlations between successive trades will also be addressed in a forthcoming study.

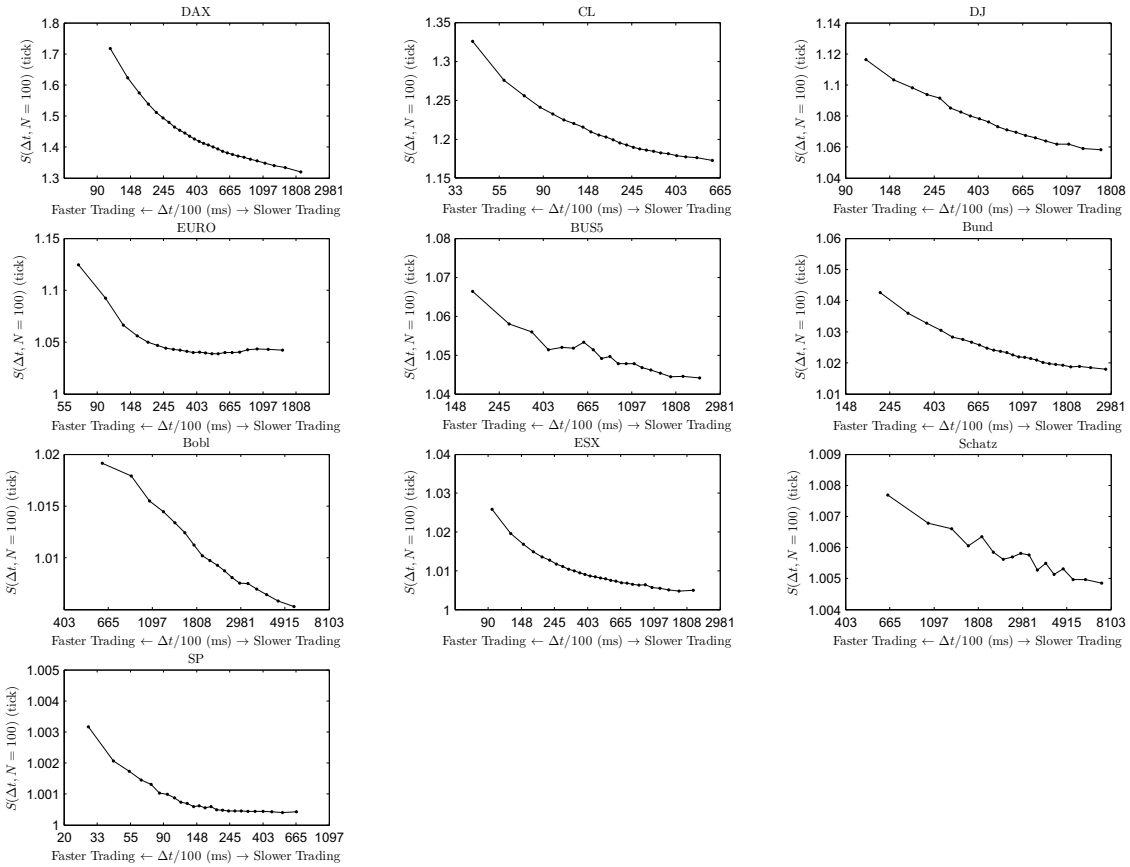


Figure 2.11: For each asset (in increasing perceived tick size $P_{=}$), $S(\Delta t, N = 100)$ as defined in (2.15) as a function of $\Delta t/100$. The form of the curves confirms that there is a strong liquidity decrease when the trading rate is increasing.

CHAPTER 3

Discrete Microstructure Noise

Now that we have established in chapter 2 the importance the rate of trading plays in determining the variance per trade, we start exploring the possibilities offered by a rate modeling approach. We begin by describing a jump process model consisting of a random walk to which we add an autocorrelated moving average process and find it particulary powerful in reproducing the signature plot effect commonly observed in financial prices. Furthermore, our model has the particularity of having a discrete jump distribution that can replicate the autocovariance of tick returns and their unconditional distribution.

3.1 Introduction

Since Bachelier and the seminal work of Black and Scholes, most popular models for price processes at a coarse time scale – for daily data, say – are Brownian diffusions, see for instance the classical textbooks of Musiela [64] or Bouchaud and Potters [23] and the references therein. In particular, diffusion models aim at describing more or less faithfully the volatility dynamics, characterized by stylized facts such as volatility clustering or leverage effect [23]. A key issue that naturally emerges when one studies high frequency data is the problem of reproducing the local characteristics of the price process such as mean-reversion and volatility behavior across scales. The discrete nature of time trade arrivals and of price variations (prices are point processes living on a tick grid), the presence of so-called

microstructure noise (described as strong mean reversion effects at small scales) make this question non trivial. At a very high frequency, price variations are also characterized by well documented stylized facts like the signature plot behavior or the Epps Effect [36].

3.1.1 The Signature Plot Effect

If $X = (X_t)_{t\geq 0}$ stands for the price of some asset at time t in continuous time (defined indifferently as the last traded price, the midpoint price between the best bid and best offer in the orderbook or the last buy or last sell price), the Sampled Realized Variance (SRV) can be defined from the quadratic variation of X over a time periods $[0, T]$ at a scale $\Delta > 0$

$$V_T(\Delta) = \frac{1}{T} \sum_{k=1}^{\lfloor T/\Delta \rfloor} \left(X_{k\Delta} - X_{(k-1)\Delta} \right)^2. \tag{3.1}$$

and whose plot as a function of Δ corresponds to the so called Signature Plot [8]. The microstructure noise effect manifests through an increase of the observed daily variance when one goes from large to small scales *i.e.* in the limit $\Delta \to 0$ (see *e.g.* Figure 3.1 for $T = 1$ month). This behavior is different from what one would expect if the data were sampled from a Brownian diffusion, in which case the function plotted in Figure 3.1 should be flat. From the perspective of statistical estimation, this leads to a simple paradox : on one side, the smaller is Δ, the larger is the dataset that can be used to estimate the volatility. However, how should one be using high-frequency data in order to obtain better estimates of the volatility is not trivial, since the realized volatility (3.1) is not stable as τ decreases.

In the literature the most popular approaches attempt to model microstructure noise with the concept of a latent price. One starts with a Brownian diffusion $\widetilde{X}_t$ defined as an efficient price, which is latent, in the sense that it cannot be observed directly. Instead the practitioner only has access to a noisy version X_t of $\widetilde{X}_t$ that accounts for microstructure noise. The most successful version – as far as mathematical development is concerned – is the additive microstructure noise model, introduced in 2001 by Gloter and Jacod [44, 45] and in the context of financial data by Ait-Sahalia, Mykland, Zhang *et al.* [3, 89, 88]. Given a sampling scale τ, one rather observes

$$X_{k\Delta} = Y_{k\Delta} + \widetilde{\varepsilon}_{k,\Delta}, \tag{3.2}$$

where the microstructure noise term $\widetilde{\varepsilon}_{n,\tau}$ satisfies $E(\widetilde{\varepsilon}_{k,\Delta}) = 0$ for obvious identifiability conditions. (Hereafter, E denotes the expectation operator.) The goal is

then to separate the noise from the true signal X_t, from which a classical volatility estimator can be performed. This has raised a vast research program over the last decade, mostly covered by econometricians and statisticians, see [4, 79, 30, 77, 78, 72, 13, 12, 14, 50, 62, 63, 73, 49] and the references therein. Whereas representation (3.2) produces an elegant pilot model to describe microstructure noise effects at the scale of a few minutes, it cannot faithfully reproduce the data as they are observed on a microscopic scale of a few seconds: for instance, the discreteness of price changes is left out and the mathematical artefact of forcing $V_T(\Delta)$ to explode when $\Delta \to 0$ becomes unavoidable.

3.1.2 Our Approach

We introduce a "fine-to-coarse" model that starts from the description of the changes of prices in continuous time compatible with the empirical observations at fine scale, and that allows us to recover a large scale diffusion behavior from the microscopic properties of the model. In essence, we invert the point of view of the latent price approach of the preceding section. Considering (3.2) again, Let us re-write the latent price model in its incremental form

$$\begin{aligned} X_{(k+1)\Delta} - X_{k\Delta} &= Y_{(k+1)\Delta} - Y_{k\Delta} + \widetilde{\varepsilon}_{k+1,\Delta} - \widetilde{\varepsilon}_{k,\Delta} \\ &= \varepsilon_{k,\Delta} + e_{k,\Delta} \end{aligned}$$

for instance. We see that our observed price has two component: a diffusive increment $\varepsilon_{k,\Delta} \sim \mathcal{N}(0, \sigma^2\Delta)$, where Δ is the diffusive volatility, and a microstructure noise term $e_{k,\Delta}$, which has the form of a moving average MA(1) process

$$e_{k,\Delta} = \widetilde{\varepsilon}_{k+1,\Delta} - \widetilde{\varepsilon}_{k,\Delta}$$

whose variance has to be set as a function of Δ. The classical approach of [3, 4, 50] assumes that $e_{k,\Delta}$ should be of order one, *i.e.* $E(e_{k\Delta}^2)$ should not depend on Δ asymptotically. This stems from the fact that Δ is related to the scheme of observation, set prior to the experiment by the statistician, and that this shall not affect in any way the nature of the microstructure noise $e_{k,\Delta}$. Let us temporarily escape from this point of view by seeing rather Δ as a time scale, irrespectively of any statistical experiment of measurement. Then, comparing the two terms $\varepsilon_{k,\Delta}$ and $e_{k,\Delta}$, we see that the diffusive component $\varepsilon_{k,\Delta}$ shall essentially be of the same order as the microstructure noise term $e_{k,\Delta}$ for instantaneous price changes, which leads us to set $E(e_{k\Delta}^2) := a^2\Delta$, up to some pre-factor $a > 0$. So alternatively, we

construct a standard microstructure noise model (in distribution) by setting

$$X_{(k+1)\Delta} - X_{k\Delta} := \sigma\Delta^{1/2}\varepsilon_k + a\Delta^{1/2}(\widetilde{\varepsilon}_{k+1} - \widetilde{\varepsilon}_k),$$

where ε_k and $\widetilde{\varepsilon}_k$ are centered independent discrete white noise with unit variance. We see that we have negative correlation in the increments by construction of the MA(1) process. Moreover, our model diffuses on coarse scales. Indeed, consider a large time scale $\Delta^\star$ in the sense that $\Delta^\star \gg \Delta$. Up to boundary effects and time shift, we have

$$\begin{aligned} X_{\Delta^\star} - X_0 &= \sum_{k\Delta=0}^{k\Delta=\Delta^\star} (X_{(k+1)\Delta} - X_{k\Delta}) \\ &= \sigma\Delta^{1/2} \sum_{k\Delta=0}^{k\Delta=\Delta^\star} \varepsilon_k + a\Delta^{1/2} \sum_{k\Delta=0}^{k\Delta=\Delta^\star} (\widetilde{\varepsilon}_{k+1} - \widetilde{\varepsilon}_k) \\ &\approx \sigma(\Delta^\star)^{1/2}\mathcal{N}(0,1) + a\Delta^{1/2}(\widetilde{\varepsilon}_{\lfloor\Delta^\star/\Delta\rfloor} - \widetilde{\varepsilon}_0) \\ &\approx \sigma(\Delta^\star)^{1/2}\mathcal{N}(0,1) \end{aligned}$$

the first term being approximated by a Gaussian random variable by the central limit theorem

$$\sigma\Delta^{1/2} \sum_{k\Delta=0}^{k\Delta=\Delta^\star} \varepsilon_k = \sigma\Delta^{1/2}\big(\tfrac{\Delta^\star}{\Delta}\big)^{1/2}\Big(\big(\tfrac{\Delta^\star}{\Delta}\big)^{-1/2} \sum_{k=0}^{k=\lfloor\Delta^\star/\Delta\rfloor} \varepsilon_k\Big) \approx \sigma\Delta^{1/2}\mathcal{N}(0,1)$$

and the second term being negligible since $\Delta \ll \Delta^\star$. In conclusion, we have a diffusive behavior toward a Brownian motion at large scales, and mean-reversion in small scales.

This is the essence of our approach. We start with ε_k and $\widetilde{\varepsilon}_k$ as centered independent discrete white noises with unit variance and we adopt the same line as before. As for the discreteness of the prices in microscopic scales, we simply have to assume that ε_k and $\widetilde{\varepsilon}_k$ live on a lattice. We still have to escape the time discretization $k\Delta$ at scale Δ. This is obtained by assuming that the discrete variables or jumps ε_k and $\widetilde{\varepsilon}_k$ are produced continuously in time following the clock of counting processes. Two independent simple Poisson processes (with different intensities) do the job, and will not essentially alter the spatial properties of our decomposition as described above. Our approach shows some similarities with Oomen [69]. However, thanks to our scale approach, we can model the jumps as integers, accounting faithfully for price discreteness. Moreover, since the empirical distribution of tick returns valued in tick values is *symmetric without skewness and not log normal with*

positive skewness so we do not model log returns, but price differentials. We will see how our use of a simple MA(1) process for the microstructure noise is powerful enough to obtain a rate modeling approach even with a trivial autoregressive noise term.

3.2 General Framework Model

We consider a price model process X_t living on the lattice $\mathbb{Z}$, each integer increment consisting of one tick. X_t can represent a trade price, a midpoint price or the last bid or last ask price. We postulate that the process X_t has a fundamental decomposition as the sum of two independent processes of the form:

$$X_t = \sum_{i=1}^{N_t} \varepsilon_i + \sum_{i=1}^{M_t} e_i \tag{3.3}$$

where N_t and M_t are independent Poisson counting processes with rates λ and μ respectively, that are also independent of their compounding jumps ε_i and e_i. We also suppose that the jumps ε_i and e_i are centered, independent (and that they have values in $\mathbb{Z}$) and satisfy

$$E(\varepsilon_i^2) < \infty, \quad E(e_i^2) < \infty.$$

If we make the additional assumption that ε_i are symmetrical (and in particular centered), independent and identically distributed random variables (later abbreviated by iid) we guarantee that

$$R_t = \sum_{i=1}^{N_t} \varepsilon_i$$

is a diffusion if observed on a coarse enough scale. This means that the process

$$R_t^{(T)} = \frac{1}{\sqrt{T}} R_{tT}$$

converges in law to a process D_t as $T \to \infty$ such that D_t is a continuous-time Markov process with continuous sample paths. We also have that R_t is a martingale which is crucial for modeling financial prices and we call it the **fundamental** component of X_t.

If we choose e_i carefully we can obtain that

$$E_t = \sum_{i=1}^{M_t} e_i$$

is asymptotically vanishing in front of R_t as $t \to \infty$. This guarantees that $X_t^{(T)} = R_t^{(T)} + E_t^{(T)}$ converges to the same asymptotic limit as $R_t^{(T)}$ as $t \to \infty$. In this context, E_t is called the **microstructure noise** component of X_t. We pay special attention to make X_t diffusive because we intend to use our process to model financial prices, which are known to display diffusive behavior over the macroscopic scale, see section 3.1 or section 1.6.

Finally, we assume that the series of random variables $(e_i)_{i\geq 0}$ is weakly stationary and that e_i are symmetrical, centered variables. This implies that the moments and autocovariance of e_i are independent of the index i, and we can define $\Gamma_e(j)$, the autocovariance function (ACF) at lag j of (e_i) by

$$\Gamma_e(j) = E(e_i e_{i-j}).$$

This assumption over (e_i) is deliberately weak because it gives us much freedom for E_t and therefore allows us to use X_t as a model for financial prices.

We now present our results regarding the Sampled average Realized Variance (SRV) defined in (3.4) and the autocovariance function of the tick returns of X_t, and conclude the first part of the paper with an application to financial data that shows that the model (3.3) can be used to estimate the variance in a high frequency setting.

3.2.1 Realized Variance

The Sampled average Realized Variance (SRV) is defined as

$$V_T(\Delta) = E(\frac{1}{T} \sum_{k=1}^{\lfloor T/\Delta \rfloor} (X_{k\Delta} - X_{(k-1)\Delta})^2) \tag{3.4}$$

where T is the final time in our sample and Δ is the sampling frequency, both measured in seconds [4]. In the financial literature [8], the plot of $V_T(\Delta)$ against Δ has a particular form, a signature if one may call it, and its distinctive plot is

called the variance signature plot. We show an example signature plot for the Bund Futures trade prices in Figure 3.1 below.

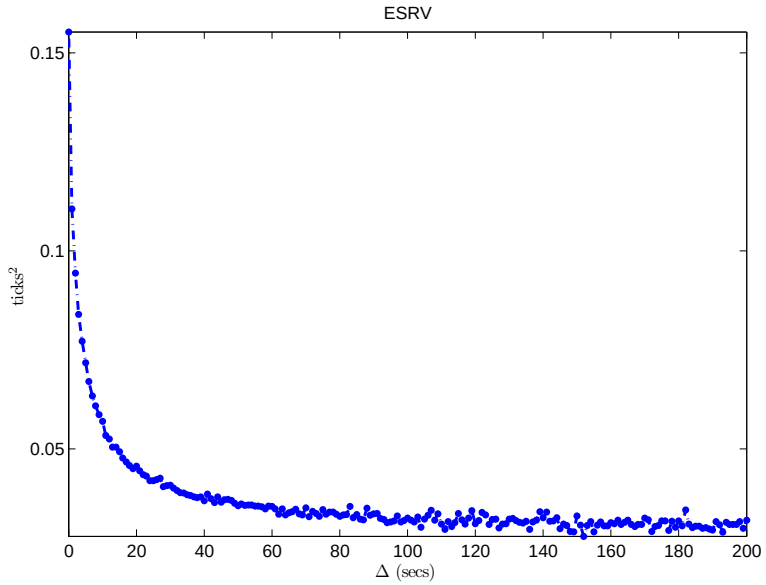


Figure 3.1: Empirical $V_T(\Delta)$ in ticks^2 against Δ in seconds for Bund futures trade price series in October 2011.

For the model (3.3), we can give an explicit formula for $V_T(\Delta)$:

Proposition 3.2.1.

$$V_T(\Delta) = E(\varepsilon_i^2)\lambda + E(e_i^2)\mu + 2\sum_{j=1}^{\infty} \Gamma_e(j) W(\mu, \Delta, j) \tag{3.5}$$

where $W(\mu, \Delta, j) : \mathbb{R}^2 \times \mathbb{N} \to \mathbb{R}$ *is given by:* $W(\mu, \Delta, 0) = \mu$ *and*

$$W(\mu, \Delta, j+1) = \mu - \frac{1}{\Delta} \sum_{k=0}^{j} P(M_\Delta > k). \tag{3.6}$$

Proof. Since ε_i and e_i are independent and centered for any $i \geq 0$ and the e_i are independent then $E(\varepsilon_i e_j) = 0$ for all $i, j \geq 0$ and $E(\varepsilon_i \varepsilon_j) = 0$ for all $i, j \geq 0$, $i \neq j$.

We then have

$$E\left((X_{k\Delta}-X_{(k-1)\Delta})^2\right) = E\left(\sum_{i\in\mathbb{N}}\varepsilon_i^2\mathbb{1}_{i\in]N_{(k-1)\Delta},N_{k\Delta}]}\right)+E\left(\sum_{i\in\mathbb{N}}e_i^2\mathbb{1}_{i\in]M_{(k-1)\Delta},M_{k\Delta}]}\right)$$
$$+E\left(2\sum_{j\geq 1}\sum_{i\in\mathbb{N}}e_ie_{i-j}\mathbb{1}_{i\in]M_{(k-1)\Delta},M_{k\Delta}]}\mathbb{1}_{i-j\in]M_{(k-1)\Delta},M_{k\Delta}]}\right)$$

Taking expectation and using that ε_i and e_i are stationary and independent of N_t and M_t, we successively have

$$E(\frac{1}{T}\sum_{k=1}^{\lfloor T/\Delta\rfloor}\sum_{i\in\mathbb{N}}\varepsilon_i^2\mathbb{1}_{i\in]N_{(k-1)\Delta},N_{k\Delta}]})=E(\varepsilon_i^2)E(\frac{1}{T}\sum_{k=1}^{\lfloor T/\Delta\rfloor}\sum_{i\in\mathbb{N}}\mathbb{1}_{i\in]N_{(k-1)\Delta},N_{k\Delta}]}),$$

$$E(\frac{1}{T}\sum_{k=1}^{\lfloor T/\Delta\rfloor}\sum_{i\in\mathbb{N}}e_i^2\mathbb{1}_{i\in]M_{(k-1)\Delta},M_{k\Delta}]}))=E(e_i^2)E(\frac{1}{T}\sum_{k=1}^{\lfloor T/\Delta\rfloor}\sum_{i\in\mathbb{N}}\mathbb{1}_{i\in]M_{(k-1)\Delta},M_{k\Delta}]})$$

and

$$E(\frac{1}{T}\sum_{k=1}^{\lfloor T/\Delta\rfloor}\sum_{j\geq 1}\sum_{i\in\mathbb{N}}e_ie_{i-j}\mathbb{1}_{i\in]M_{(k-1)\Delta},M_{k\Delta}]}\mathbb{1}_{i-j\in]M_{(k-1)\Delta})},M_{k\Delta}])=\sum_{j\geq 1}\Gamma_e(j)W(\mu,\Delta,j),$$

where

$$W(\mu,\Delta,j)=E(\frac{1}{T}\sum_{k=1}^{\lfloor T/\Delta\rfloor}\sum_{i\in\mathbb{N}}\mathbb{1}_{i\in]M_{(k-1)\Delta},M_{k\Delta}]}\mathbb{1}_{i-j\in]M_{(k-1)\Delta},M_{k\Delta}]}).$$

By inverting sums and expectations we have

$$E(\frac{1}{T}\sum_{k=1}^{\lfloor T/\Delta\rfloor}\sum_{i\in\mathbb{N}}\mathbb{1}_{i\in]N_{(k-1)\Delta},N_{k\Delta}]})=E(\frac{1}{T}\sum_{i\in\mathbb{N}}\mathbb{1}_{i\in]0,N_T]})=\lambda,$$
$$E(\frac{1}{T}\sum_{k=1}^{\lfloor T/\Delta\rfloor}\sum_{i\in\mathbb{N}}\mathbb{1}_{i\in]M_{(k-1)\Delta},M_{k\Delta}]})=E(\frac{1}{T}\sum_{i\in\mathbb{N}}\mathbb{1}_{i\in]0,M_T]})=\mu$$

which yields equation (3.5). Finally we have the explicit formula for W: the term $W(\mu, \Delta, j)$ is equal to

$$\begin{aligned} W(\mu,\Delta,j) &= E(\frac{1}{T}\sum_{k=1}^{\lfloor T/\Delta \rfloor}\sum_{i\in\mathbb{N}} \mathbb{1}_{i\in]M_{(k-1)\Delta},M_{k\Delta}]}\mathbb{1}_{i-(j-1)\in]M_{(k-1)\Delta},M_{k\Delta}]}) \\ &\quad - E(\frac{1}{T}\sum_{k=1}^{\lfloor T/\Delta \rfloor}\sum_{i\in\mathbb{N}} \mathbb{1}_{i=M_{(k-1)\Delta+j}}\mathbb{1}_{M_{k\Delta}-M_{(k-1)\Delta}>j-1}) \\ &= W(\mu,\Delta,j-1) - \frac{1}{T}\sum_{k=1}^{\lfloor T/\Delta \rfloor} P(j-1<M_\Delta) \\ &= W(\mu,\Delta,j-1) - \frac{1}{\Delta}P(M_\Delta > j-1) \end{aligned}$$

which gives (3.6).

□

A variance signature plot like the one shown in Figure 3.1 is characterized by an asymptote at ∞ and a finite limit at 0. The SRV in model (3.3) displays the same property and we discuss that in the next paragraph.

Behavior at 0 and ∞

As Δ goes to 0 or to ∞, we can calculate the limits of $W(\mu, \Delta, j)$ and therefore that of $V_T(\Delta)$.

Let $j > 1 \in \mathbb{N}$ fixed. The limits in 0 and ∞ can be calculated by:

(i) $\Delta \to \infty$:

Since $\sum_{k=0}^{j-1} P(M_\Delta > k) \leq j$, we have

$$\mu - \frac{j}{\Delta} \leq W(\mu,\Delta,j) \leq \mu \tag{3.8}$$

which yields

$$W(\mu,\Delta,j) = \mu - j\mathrm{O}(\frac{1}{\Delta})$$

So for each $j > 1$, for $\Delta \to \infty$, we have $W(\mu, \Delta, j) = \mu - \mathrm{O}(\frac{1}{\Delta})$.

(ii) $\Delta \to 0$:

We have that:

$$0 \leq W(\mu, \Delta, j) \leq \mu - \frac{1}{\Delta} P(M_\Delta > 0)$$

and

$$\begin{aligned} \mu - \frac{1}{\Delta} P(M_\Delta > 0) &= \mu - \frac{1}{\Delta}(1 - P(M_\Delta = 0)) \\ &= \mu - \frac{1}{\Delta}(1 - e^{-\mu\Delta}) \\ &= \mu - \frac{1}{\Delta}(-\mu\Delta + \mathrm{o}(\Delta)) \\ &= \mathrm{o}(1) \end{aligned}$$

So for $\Delta \to 0$, for all $j > 1$, $W(\mu, \Delta, j) = \mathrm{o}(1)$, uniformly in j.

We summarize some consequences of these computations in the following

Proposition 3.2.2. *We have for $j \in \mathbb{N}$ fixed,*

(i) $\lim_{\Delta \to 0} W(\mu, \Delta, j) = 0$

(ii) $\lim_{\Delta \to +\infty} W(\mu, \Delta, j) = \mu$.

If we furthermore assume that there exists $A > 0$ such that for all $n \geq 0$, we have $-A \leq \sum_{j=0}^{n} j\Gamma_e(j) \leq A$, we derive:

(i) $\lim_{\Delta \to 0} V_T(\Delta) = E(\varepsilon_i^2)\lambda + E(e_i^2)\mu$

(ii) $\lim_{\Delta \to +\infty} V_T(\Delta) = E(\varepsilon_i^2)\lambda + E(e_i^2)\mu + 2\mu \sum_{j=1}^{\infty} \Gamma_e(j)$

Note that the second point is seen by using the uniform bounds of $\sum_{j=0}^{n} j\Gamma_e(j)$ and $W(\mu, \Delta, j) = \mu - j\mathrm{O}(\frac{1}{\Delta})$.

We conclude this discussion about the SRV by saying that the usual form of the signature plot that we observe dictates that $\lim_{\Delta \to +\infty} V_T(\Delta)$ exists and that

$\lim_{\Delta\to 0} V_T(\Delta) > \lim_{\to\infty} V_T(\Delta)$. Therefore we must have that the noise term is more negatively than positively autocorrelated in the long run such that:

$$\sum_{j=1}^{\infty} \Gamma_e(j) < 0. \tag{3.9}$$

We discuss the nature of empirical tick returns autocovariance in more details in section 3.4 and suggest a simple expression for e_i that would match this criterion.

3.2.2 Autocovariance

We now provide a formula for the autocovariance function of the **tick returns** of X_t. Since jumps are conventionally non zero movements in the process X_t, the tick return is the increment in X_t between each arrival of either of the Poisson processes, N_t or M_t, even if the size of the movement (respectively ε_i or e_i) is null. More precisely, we know that N_t and M_t are two independent Poisson counting processes and therefore the probability of both of them jumping at the same time is null. Therefore, we can define $\tau_0 = 0, \tau_1, \tau_2, \cdots, \tau_N$ as the series of jumps of either N_t or M_t put in a strict ascending order (i.e. the series of jumps of the Poisson process $N_t + M_t$). For any $i \in \mathbb{N}$, we have that that τ_i is a jump of N_t with probability $\frac{\lambda}{\mu+\lambda}$ and similarly, it is a jump of M_t with probability $\frac{\mu}{\mu+\lambda}$.

Next we define $\mathbf{N} = \{i \in \mathbb{N} | dN_{\tau_i} = 1\}$ and $\mathbf{M} = \{i \in \mathbb{N} | dM_{\tau_i} = 1\}$ respectively as the set of jump indexes which belong to either N_t or M_t. Consider the variables $J_i = X_{\tau_i} - X_{\tau_{i-1}}$ of successive tick returns of X_t. We have that the series (J_i), $i \geq 0$ is weakly stationary (this is actually shown in the proof that will follow) and thus we can define its autocovariance function at lag j by

$$\Gamma_J(j) = E(J_i J_{i-j})$$

We have the following formula for the autocovariance at lag j of $(X_{\tau_i})_{i\geq 0}$:

$$\Gamma_J(0) = E(\varepsilon_i^2)\frac{\lambda}{\mu+\lambda} + E(e_i^2)\frac{\mu}{\mu+\lambda} \tag{3.10}$$

$$\Gamma_J(j) = \sum_{k=1}^{j} \Gamma_e(k) \binom{k-1}{j-1} (\frac{\mu}{\lambda+\mu})^{k+1} (\frac{\lambda}{\lambda+\mu})^{j-k}, \qquad \text{for } j > 0. \tag{3.11}$$

For $j = 0$, the expression of the variance of J_i, $\Gamma_J(0) = E(J_i^2)$, is seen as the

average of the variances of ε_i and e_i weighted with their respective probabilities of occurrence $\frac{\lambda}{\mu+\lambda}$ and $\frac{\mu}{\mu+\lambda}$ (since they are independent). For $j > 0$ equation (3.11) is seen by first noticing that since ε_i is independent of M_t and e_i we can write:

$$\Gamma_J(j) = E(J_i J_{i-j}) = E(e_{M_{\tau_i}} \mathbb{1}_{i\in\mathbf{M}} e_{M_{\tau_{i-j}}} \mathbb{1}_{i-j\in\mathbf{M}}).$$

Now, we suppose that τ_i and τ_{i-j} are both jumps of M_t, they are $j-1$ observations apart that could correspond to either jumps of N_t or M_t. Because N_t and M_t are independent, we can find the probability that k of these jumps are jumps of N_t. This can be written as:

$$\begin{aligned}
E(e_{M_{\tau_i}} \mathbb{1}_{i\in\mathbf{M}} e_{M_{\tau_{i-j}}} \mathbb{1}_{i-j\in\mathbf{M}}) &= E(e_{M_{\tau_i}} e_{M_{\tau_i}-1} \mathbb{1}_{i\in\mathbf{M}} \mathbb{1}_{i-j\in\mathbf{M}} \mathbb{1}_{i-s\in\mathbf{N}\forall s\in]i-j,i[} \\
&\quad + e_{M_{\tau_i}} e_{M_{\tau_i}-2} \mathbb{1}_{i\in\mathbf{M}} \mathbb{1}_{i-j\in\mathbf{M}} \mathbb{1}_{\exists s\in]i-j,i[,|i-s\in\mathbf{M}} \\
&\quad + e_{M_{\tau_i}} e_{M_{\tau_i}-3} \mathbb{1}_{i\in\mathbf{M}} \mathbb{1}_{i-j\in\mathbf{M}} \mathbb{1}_{\exists s_1,s_2\in]i-j,i[,|i-s_1,i-s_2\in\mathbf{M}} \\
&\quad + \ldots) \\
&\quad + e_{M_{\tau_i}} e_{M_{\tau_i}-j} \mathbb{1}_{i\in\mathbf{M}} \mathbb{1}_{i-j\in\mathbf{M}} \mathbb{1}_{i-s\in\mathbf{M}\forall s\in]i-j,i[}) \\
&= \sum_{k=1}^{j} E(e_{M_{\tau_i}} e_{M_{\tau_i}-k} \mathbb{1}_{i\in\mathbf{M}} \mathbb{1}_{i-j\in\mathbf{M}} \mathbb{1}_{\#\{s\in]i-j,i[,|i-s\in\mathbf{M}\}=k-1})
\end{aligned}$$

where # symbol denotes cardinality. Since e_i is stationary and independent of N_t and M_t we can write

$$\begin{aligned}
&E(e_{M_{\tau_i}} e_{M_{\tau_i}-k} \mathbb{1}_{i\in\mathbf{M}} \mathbb{1}_{i-j\in\mathbf{M}} \mathbb{1}_{\#\{s\in]i-j,i[,|i-s\in\mathbf{M}\}=k-1}) \\
=&E(e_{M_{\tau_i}} e_{M_{\tau_i}-k}) E(\mathbb{1}_{i\in\mathbf{M}} \mathbb{1}_{i-j\in\mathbf{M}} \mathbb{1}_{\#\{s\in]i-j,i[,|i-s\in\mathbf{M}\}=k-1})
\end{aligned}$$

and

$$E(e_{M_{\tau_i}} e_{M_{\tau_i}-k}) = E(e_{M_i} e_{M_{i-k}}) = \Gamma_e(k) \text{ for all } 1 \le k \le j.$$

Furthermore, we have that for any $k \in \mathbb{N}$

$$E(\mathbb{1}_{i\in\mathbf{M}} \mathbb{1}_{i-j\in\mathbf{M}} \mathbb{1}_{\#\{s\in]i-j,i[,|i-s\in\mathbf{M}\}=k}) = (\frac{\mu}{\lambda+\mu})^2 \binom{k}{j-1} (\frac{\lambda}{\lambda+\mu})^{j-1-k} (\frac{\mu}{\lambda+\mu})^k$$

therefore

$$E(e_{M_{\tau_i}} e_{M_{\tau_i}-k} \mathbb{1}_{i\in\mathbf{M}} \mathbb{1}_{i-j\in\mathbf{M}} \mathbb{1}_{\#\{s\in]i-j,i[,|i-s\in\mathbf{M}\}=k}) = \Gamma_e(k)(\tfrac{\mu}{\lambda+\mu})^2 \binom{k}{j-1} (\tfrac{\lambda}{\lambda+\mu})^{j-1-k} (\tfrac{\mu}{\lambda+\mu})^k$$

which yields (3.11) and that $(J_i)_{i\geq 0}$ is weakly stationary.

We apply the model described above to financial data. In section 3.4, we use a particular case of the general model (3.3) inspired by what we observed on the data. In the next section we review some aspects of the data and microstructure that we have detailed in the introduction (subsection 3.1.2 and also section 1.6.2, section 1.3, and section 1.5).

3.3 Data Presentation and Tick Size

Data Presentation

We use **level 1** data provided by QuantHouse Trading Solutions [1] of 10 futures contracts on assets of different classes that trade in different exchanges for the month of June 2009. The data has millisecond accuracy and it has been treated in such a way that each market order is equivalent to exactly one trade[2]. In particular we use Futures data for the DAX index (DAX), the EURO-STOXX 50 index (ESX), the 10-years Euro-Bund (Bund), the Euro-Bobl (Bobl), the Euro-Schatz (Schatz), the Dow Jones index (DJ), the 5-Year U.S. Treasury Note Futures (BUS5), the exchange rate EUR/USD (EURO) and the the futures on the SP500 index (SP) and finally the Light Sweet Crude Oil Futures (CL). For more information please refer to the introduction of the thesis (section 1.5).

Tick Size

We now give a quick reminder of the tick size and of the estimator we used for it. As we saw in section section 1.3, the tick size is a notion used to describe just how important the tick value is for the traders. On average, it describes the market's aversion to a movement in the midpoint price of half a tick.

Robert and Rosenbaum in [75] introduce a model for ultra high frequency data with a parameter η quantifying the perceived tick size. They show that this parameter can easily be estimated from the observation of the number of alternations

[1] http://www.quanthouse.com/

[2] When one market order hits several limit orders it results in several trades being reported (cf. section 1.2.2). We aggregate together all such transactions and consider them as one trade with an equivalent size and price.

and continuations of tick returns. One simple estimator of η is for example:

$$\widehat{\eta} = \frac{N_1^c}{2N_1^a}. \tag{3.12}$$

where N_1^a is the number of price variations of size 1 ticks whose direction is the same as the one of the preceding variation and N_1^c is the number of price variations whose direction is opposite to the one of the preceding variation.

The parameter η is intended to represent the aversion to price changes. If η is small, the price will be mainly doing alternations, also known as bid ask bounces. This translates into a large tick, because in this case market participants are more averse to changes in the midpoint price. As this is put in [75], all market participants "feel more comfortable when the asset price is constant than when it is moving". So the best bid and ask price levels (and hence the midpoint price) would only move when it becomes clear that the current price level is unsustainable.

We use the tick size to classify our assets and we interpret the results of our model according to this notion. Let us now describe how we used model (3.3) to capture microstructure noise and the autocovariance of tick returns.

3.4 Application to Financial Data

The variance signature plot is characterized by an increase in the calculated SRV $V_T(\Delta)$ as Δ goes to 0 (cf Figure 3.1). We provide in this section a simple discrete model, based on model (3.3), that attempts to replicate the signature plot, has a similar tick returns distribution and displays a tick autocovariance function not unlike the one we find with asset price series. In this section, the Poisson process $(N_t + M_t)_{t\geq 0}$ represents the arrival of a new price (that would be the time arrival of a trade when we are looking at trade prices or midpoint prices, or the arrival of a trade on the bid (ask) when we are looking at last sell (buy) prices). The series of price differentials between two price arrivals $(J_i)_{i\geq 0}$, represents then the series of tick returns and J_i can be equal to 0. J_i cannot be called a "jump" of X_t since it can be null. We now explain the choices we make for the particular model we use to fit financial data.

First, if we look at the distribution of tick returns, say the Bund futures for example (cf table 3.1 or Figure 3.3, graphs for the cumulative distribution function (CDF) or the probability distribution function (PDF)), we find it discrete symmetric, with a probability distribution mostly concentrated in $\{-2, -1, 0, 1, 2\}$. The autocovariance function of tick returns (Figure 3.3) is negative with significant values for lags 1 and 2. Based on these observations we suggest that the support of the

distributions of ε_i and e_i be in $\{-2,-1,0,1,2\}$ and that an autoregressive moving average is a good choice for the noise tick returns component of X_t. We remind that X_t is a continuous time discrete price process, that can represent, indifferently, a trade price, a midpoint price or a last buy or last sell price, is expressed as:

$$X_t = \sum_{i=1}^{N_t} \varepsilon_i + \sum_{i=1}^{M_t} e_i \tag{3.13}$$

In particular, based on our argument developed in section 3.1, we propose that e_i is a simple MA(1) process such that

$$e_i = \widetilde{\varepsilon}_i - \widetilde{\varepsilon}_{i-1}$$

where $\widetilde{\varepsilon}_i$ is a symmetrical, discrete random variable with values in $\{-1,0,1\}$ whose distribution can be described with one parameter:

$$p = P(\widetilde{\varepsilon}_i = 1)$$

This means e_i has a distribution that can also be described with the same parameter p such that:

$$\begin{pmatrix} P(e_i=-2) \\ P(e_i=-1) \\ P(e_i=0) \\ P(e_i=1) \\ P(e_i=2) \end{pmatrix} = \begin{pmatrix} p^2 \\ 2p-4p^2 \\ 1-4p+6p^2 \\ 2p-4p^2 \\ p^2 \end{pmatrix} \tag{3.14}$$

The support of e_i taken as $\{-2,\ldots,2\}$ is motivated by Table 3.1 for Bund futures but could presumably be adjusted to other assets with different support in a similar fashion. Furthermore, e_i is autocorrelated only at lag 1 such that:

$$\begin{aligned} \Gamma_e(1) &= -2p \\ \Gamma_e(j) &= 0 \ \forall j > 1. \end{aligned}$$

Giving ε_i the same support as e_i, in order to avoid any trivial distinctions between

tick return	nb. occurrences
-6	1
-4	1
-3	5
-2	135
-1	19005
0	137846
1	19143
2	120
3	6
4	1

Table 3.1: Empirical number of observations per tick returns the Bund Futures trade prices. We can see that returns of more than 2 ticks are very rare.

ε_i and e_i, we then need two parameters to describe the distribution of ε_i:

$$\begin{pmatrix} P(\varepsilon_i = -2) \\ P(\varepsilon_i = -1) \\ P(\varepsilon_i = 0) \\ P(\varepsilon_i = 1) \\ P(\varepsilon_i = 2) \end{pmatrix} = \begin{pmatrix} p_2^\varepsilon \\ p_1^\varepsilon \\ 1 - 2(p_1^\varepsilon + p_2^\varepsilon) \\ p_1^\varepsilon \\ p_2^\varepsilon \end{pmatrix}. \tag{3.15}$$

As a consequence we have

$$\begin{aligned} E(\varepsilon_i^2) &= 8p_2^\varepsilon + 2p_1^\varepsilon \\ E(e_i^2) &= 4p \\ \Gamma_e(1) &= -2p \\ \Gamma_e(j) &= 0 \ \forall j > 1 \\ W(\mu, \Delta, 1) &= \mu - \frac{1}{\Delta}(1 - e^{-\Delta\mu}) \end{aligned}$$

And we can then rewrite equation (3.5) as:

$$V_T(\Delta) = (8p_2^\varepsilon + 2p_1^\varepsilon)\lambda + \frac{4p}{\Delta}(1 - e^{-\Delta\mu}). \tag{3.16}$$

Also, the j^{th} lag ($j \geq 0$) autocovariance of J defined in equations (3.10) and (3.11)

as:

$$\Gamma_J(0) = \frac{\lambda}{\lambda+\mu}(8p_2^\varepsilon + 2p_1^\varepsilon) + 4p\frac{\mu}{\lambda+\mu} \tag{3.17}$$

$$\Gamma_J(j) = -2p\left(\frac{\mu}{\lambda+\mu}\right)^2\left(\frac{\lambda}{\lambda+\mu}\right)^{j-1} \qquad \text{for } j>0. \tag{3.18}$$

This model satisfies the autocovariance convergence condition of equation (3.9) guaranteeing that $\lim_{\Delta\to\infty} V_T(\Delta)$ exists and it is bigger than $\lim_{\Delta\to 0} V_T(\Delta)$. In the following, we see how this model fairs when calibrated to real data.

3.4.1 Parameter Fitting

Let us briefly explain the method we used for fitting the model and show on an example the results we get on financial data. First we note that equation (3.16) is not sufficient to fully determine our model. Indeed while p and μ can be determined uniquely, we could only get an estimate for $(8p_2^\varepsilon + 2p_1^\varepsilon)\lambda$ as a whole with p_2^ε, p_1^ε and λ undetermined but related through a linear equation. We see in the following paragraph how to use the distribution of J_i in order to fully define our model. Our main priority in this work is to replicate the signature plot effect observed in financial prices. So we start by fitting the following function:

$$V_T(\Delta) = \chi + \frac{4p}{\Delta}(1-e^{-\Delta\mu})$$

where χ replaces $(8p_2^\varepsilon + 2p_1^\varepsilon)\lambda$, to the empirical SRV defined in (3.16). We do this using a non linear least squares method under the constraints that $\chi>0$, $\mu>0$ and $0\le p\le 1/2$.

Once we have the values for χ, μ and p we use them to fit the autocovariance function of J, $\Gamma_J(j)$ for $j\ge 0$, which means we try to create the best model possible that fits the variance and the autocovariance of the jumps of X. However, even with the use of the autocovariance, the constraint $\chi = (8p_2^\varepsilon + 2p_1^\varepsilon)\lambda$ will give us the following system of equations:

$$\Gamma_J(0) = \frac{\chi}{\lambda+\mu} + 4p\frac{\mu}{\lambda+\mu}$$

$$\Gamma_J(j) = -2p(\frac{\mu}{\lambda+\mu})^2(\frac{\lambda}{\lambda+\mu})^{j-1}, \qquad \text{for } j>0.$$

with which we can only uniquely determine λ, because p_2^ε and p_1^ε are constrained by

χ. Furthermore, the model requires that $0 \leq E(\varepsilon_i^2) \leq 4$, and hence that $\lambda \geq \chi/4$. A constraint least squares fit on Γ_J gives us λ.

Finally, it remains to determine p_2^ε and p_1^ε. We already know μ, λ, p from the previous fits and have the linear relationship $(8p_2^\varepsilon + 2p_1^\varepsilon)\lambda = \chi$. The last information that we can use is the unconditional distribution of J. Setting $x = \frac{\lambda}{\mu+\lambda}$, the distribution $f_J(x)$ of J has the form

$$\begin{pmatrix} P(J=-2) \\ P(J=-1) \\ P(J=0) \\ P(J=1) \\ P(J=2) \end{pmatrix} = \begin{pmatrix} xp_2^\varepsilon + (1-x)p^2 \\ xp_1^\varepsilon + (1-x)(2p-4p^2) \\ x(1-2(p_1^\varepsilon + p_2^\varepsilon)) + (1-x)(1-4p+6p^2) \\ xp_1^\varepsilon + (1-x)(2p-4p^2) \\ xp_2^\varepsilon + (1-x)p^2 \end{pmatrix} \tag{3.19}$$

Which means that with all the other parameters estimated, the cumulative distribution function of J, is parameterizable with only p_2^ε and p_1^ε. However, using the constraint $(8p_2^\varepsilon + 2p_1^\varepsilon) = \frac{\chi}{\lambda}$ the distribution becomes only dependant on one of the two probabilities. We choose p_2^ε as the parameter and we refer to the cumulative distribution of J by $F_J(x|p_2^\varepsilon)$.

We fit $F_J(x|p_2^\varepsilon)$ to the empirical distribution of jumps again using a non linear least squares fit and we get the full distribution of ε_i and thus completely defining our model. We also put bounds on p_2^ε in a way to make it compatible with the definition of the model and the previous estimated parameters. We find the lower bound to be $\max(0, (\frac{\chi}{\lambda} - 1)/6)$ and the upper one $\min(\frac{\chi}{8\lambda}, 0.5)$.

In conclusion, this method gives priority to reproducing the signature plot first, followed by the second moments of the jumps of X_t and finally the unconditional distribution of the jumps. We were careful as to ensure that the parameters are consistent with the requirements of the model such that $\lambda, \mu > 0$, $\{p_2^\varepsilon, p_1^\varepsilon, p\} \subset [0, \frac{1}{2}]$ and $p_2^\varepsilon + p_1^\varepsilon \leq \frac{1}{2}$.

In next section, we look at the model calibration for Bund futures data, before concluding with an overview of results for other assets

3.4.2 Application to the Bund Futures

In this paragraph, we show the results of the model calibration over Bund futures. We find that the model replicates well the variance signature plot, produces jumps with similar distributions (CDF and autocovariance function) to the empirical ones.

We use 21 days covering the month of June 2009. The data is handled in the way described in section 3.3. Figure 3.2 shows that the model fits very well the signature plot. In particular, we get the following results: for the random walk component of the process $\sum_{i=1}^{N_t} \varepsilon_i$, we get $\lambda = 0.12$ for rate of arrival of jumps and $8p_2^\varepsilon + 2p_1^\varepsilon = 0.31$ for their variance; for the autocorrelated component, $\sum_{i=1}^{M_t} e_i$, we get $\mu = 0.55$ for the rate of arrival of jumps and $4p = 0.22$ for their variance. More explicitly, the statistics break down as following:

Random Walk Component		Noise Component	
λ	0.119	μ	0.549
p_2^ε	0.001	p_2^e	0.003
p_1^ε	0.152	p_1^e	0.096
p_0^ε	0.693	p_0^e	0.802
$E(\varepsilon_i^2)$	0.312	$E(e_i^2)$	0.215

Table 3.2: Calibration Result applied to the Bund Futures trade prices during November-December 2011. On the left are the statistics relating to the random walk component of the model, where as on the right are those relating for the autoregressive component or noise. We see that the noise component displays a higher arrival and a smaller variance than the martingale part.

One interpretation of these results is as following: the vanishing noise component E_t has a smaller amplitude then the fundamental martingale component R_t, but it occurs almost 5 times more often. Since these results are for trade prices, the noise component represents the bid ask bounce phenomenon where as the random walk component represents fundamental changes in the prices.

Figure 3.3 shows that the model replicates the variance of the tick returns and their unconditional distribution. The fast decaying autocorrelation function of the model corresponds well with what happens in reality where the autocovariance is significant merely at lag 1 and decays very quickly afterwards (cf also Figure 3.7 Bund graph for a more detailed look at the autocovariance, where $\Gamma_J(0)$ was ignored in order to make the figure more readable.)

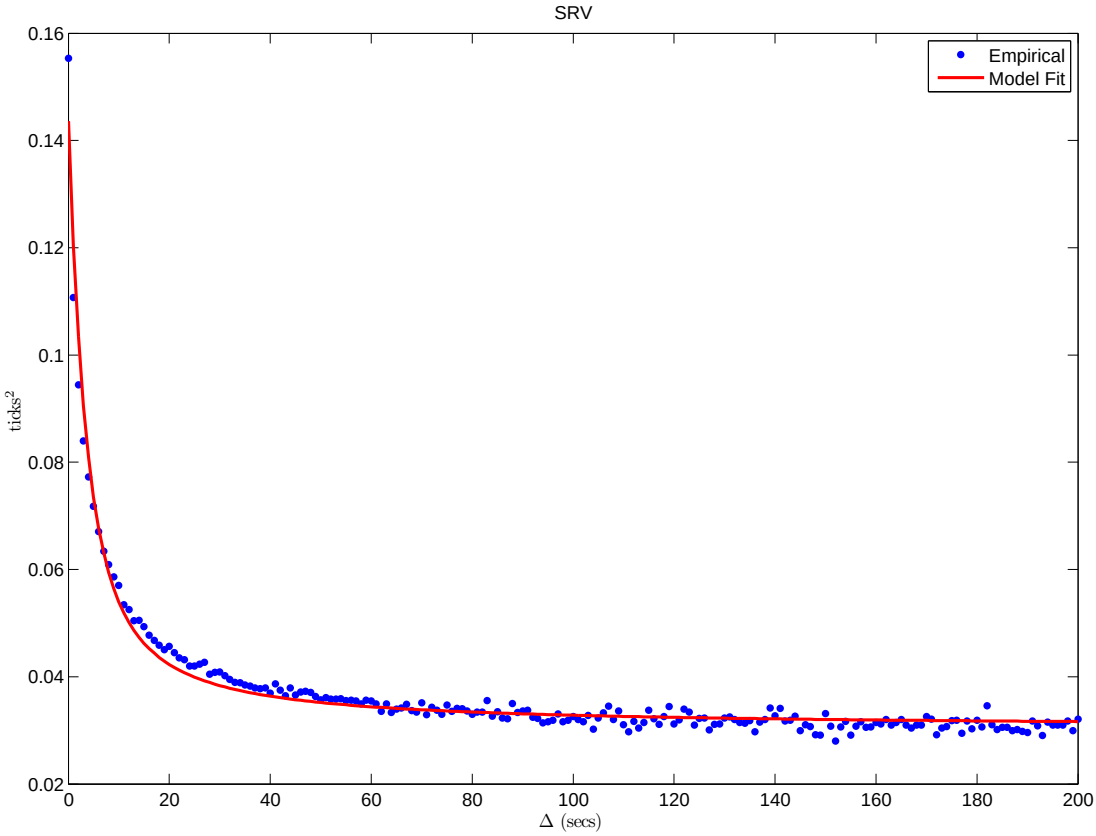


Figure 3.2: Fit result of the SRV for the process of trade prices of Bund Futures. The blue dots are for the empirical data and the red one is the one produced by the calibration of model (3.16).

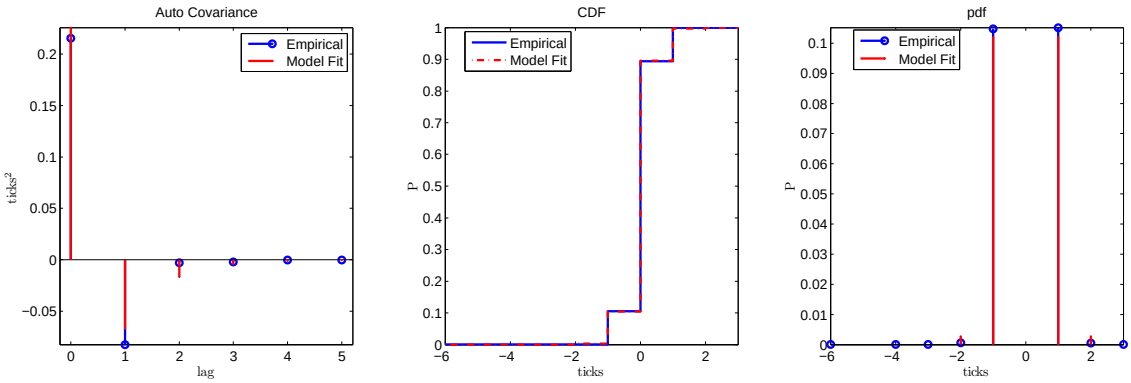


Figure 3.3: Graphic comparisons between empirical statistics for the process of trade prices of Bund Futures and the statistics produced by the calibration of model (3.16). The blue curves are for the empirical data and the red are for the calibrated model. From left to to right: in the first graph we show Γ_J for 10 lags, in the second we show $F_J(x)$ and finally we show the density of J, $f_J(x)$. $f_J(0)$ was removed to make the plot more readable

3.4.3 Results for Other Assets

We discuss in this section the results for the rest of the assets and for different types of tick returns, the traded prices, the last bid prices and the midpoint prices. The results of the fits are shown in section 3.A of the appendix, whereas the figures are shown in sections 3.B.1, 3.B.2, 3.B.3. We ordered the assets by tick size in a similar fashion to [5] and as we have discussed in section 3.3. We present in the following sections the results of the model calibration to the data and discuss them according to the type of the series and the tick size.

We focus primarily on the series of trade prices because most of the results are similar for other types of price series. However there are some interesting differences that we point us in the subsequent paragraphs.

Trade Prices

We now consider the process of trade prices. We remind that the series consists of the arrival times of market orders, coupled with the execution price expressed in ticks.

Description of the Figures: First, Figure 3.6 shows that the model is able to reproduce the signature form of the SRV for all the assets. We plot the SRV in tick^2 against Δ, the sampling duration, in seconds. The signature behavior is present across all assets regardless of the tick size. However, we notice that as the

tick increases, the **diffusion deviation measure** defined by the ratio

$$d = \frac{V_T(0) - V_T(\infty)}{V_T(\infty)}$$

increases, reflecting an increase in the noise contribution to the SRV and signifying a more pronounced departure from the efficient diffusion behavior. In terms of the model parameters, this means that $\frac{-2\mu\Gamma_e(1)}{\lambda E(\varepsilon_i^2)}$ increases. In the particular case of our model we have $-2\mu\Gamma_e(1) = 4p = E(e_i^2)$ exactly equal to the variance of the noise, and so we would expect the ratio $\frac{\mu E(e_i^2)}{\lambda E(\varepsilon_i^2)}$ to be increasing with the tick size.

Figure 3.7 shows that all the tick returns series present significant autocovariance at the first and second lag. The fitted model also has non null autocovariance of the tick returns series and it does capture some of the assets' ACF functions like the FCL, DJ and the SP. However, even though its autocovariance function has fast exponential decay (cf equation (3.18)), it is less successful in other cases where the empirical ACF decays brutally after the first lag. The DAX, for example, has significant autocovariance only at lag 1, and we do not get very adequate fits. This is due in part to the way we fit the ACF. We explained in section 3.4.1 that we fit $\Gamma_J(j)$ even when $j = 0$, which means that we also fit the variance of the series, which gives us a very good fit for the unconditional distribution as well. If we fit $\Gamma_J(j)$ for $j \geq 1$ then we would get a better fit for the ACF but we loose in the unconditional distribution of J. Nevertheless, our objective is to show that we are able to capture some of the negative autocorrelation of tick returns responsible for the signature plot, and the model is able to do that.

Figures 3.8 and 3.9 show the distribution of the tick returns compared to that of J of the fitted model. Figure 3.8 shows the cumulative distribution function with a support restricted to $[-5, 5]$ while Figure 3.9 shows the probability distribution function of all values which we put to show the distribution tail difference between small tick and large tick assets. Indeed, small tick assets show fatter tails and more extreme values, whereas large tick assets have a compact support basically restricted to $\{-2, -1, 0, 1, 2\}$. Nevertheless, even with ignoring the tails the fitted model presents a distribution that is very close to the empirical one across all assets (except maybe for the CL futures). This becomes more and more obvious with large tick assets since their distribution becomes more and more compact to the point of having a support limited to $\{-2, -1, 0, 1, 2\}$ exactly like the model in section 3.4.

Analysis of Fitting Results: The fitting results are shown in table 3.3. We present four groups of statistics: the fit statistics for R_t and E_t, some descriptive statistics about our database and finally a group that we call "tick statistics" containing results relating to the tick size of the asset. In it there is η (section 1.3) that

we have used as a proxy for the tick size and a sorting parameters for the assets and the ratios $\frac{\mu}{\lambda}$, $\frac{\mu(1-p_0^e)}{\lambda(1-p_0^\varepsilon)}$, $\frac{\mu\sqrt{E(e_i^2)}}{\lambda\sqrt{E(\varepsilon_i^2)}}$ and $r = \frac{\mu(1-p_0^e)}{\mu(1-p_0^e)+\lambda(1-p_0^\varepsilon)}$ which gives the relative importance of "noisy" ticks to "fundamental" ones.

We have seen in section 3.4.2, that the vanishing noise component has a higher occurrence rate than the fundamental one but has a smaller variance. This not true across all the other assets as we can see in assets like the EURO futures where $\mu < \lambda$ and $E(e_i) > E(\varepsilon_i)$. However, μ, λ and $E(e_i)$, $E(\varepsilon_i)$ alone, are not good indicators of the noise to martingale contribution. Good indicators of the contribution of E_t and R_t are $\mu\sqrt{E(e_i^2)}$ and $\lambda\sqrt{E(\varepsilon_i^2)}$ respectively. However, since we have some assets with fatter tails and whose variances are influenced by the values outside the support of the model's distribution, more robust indicators of the contribution of E_t and R_t are $\mu(1-p_0^e)$ and $\lambda(1-p_0^\varepsilon)$ respectively. As we were expecting, the ratios of these statistics are increasing in tick size. The larger the tick size, the bigger the contribution of the noise is. This can be interpreted as an increase in the number of bid ask bounces, which is more important in a big tick asset. Let

$$r = \frac{\mu(1-p_0^e)}{\mu(1-p_0^e) + \lambda(1-p_0^\varepsilon)}$$

be the the relative contribution of the noise term to the total movement in price. Figure 3.4 shows r plotted against η, and features a decreasing linear relationship confirming our initial thoughts. Since η is calculated as the ratio of the number of continuation to the number of alternations, this falls perfectly within the spirit of our model. As the tick increases the weight of the vanishing autocorrelated noise term increases.

We end this section by saying that we notice that the sum of the fitted rates $\mu + \lambda$ is very close to the total rate of trading displayed in the **Observation per Second** row. Although we did not set out to match this statistic, the fact the the distribution of the model matches that of the empirical observations and the fact that we were able to closely replicate the signature plot, makes it sufficient to capture the average trading rate.

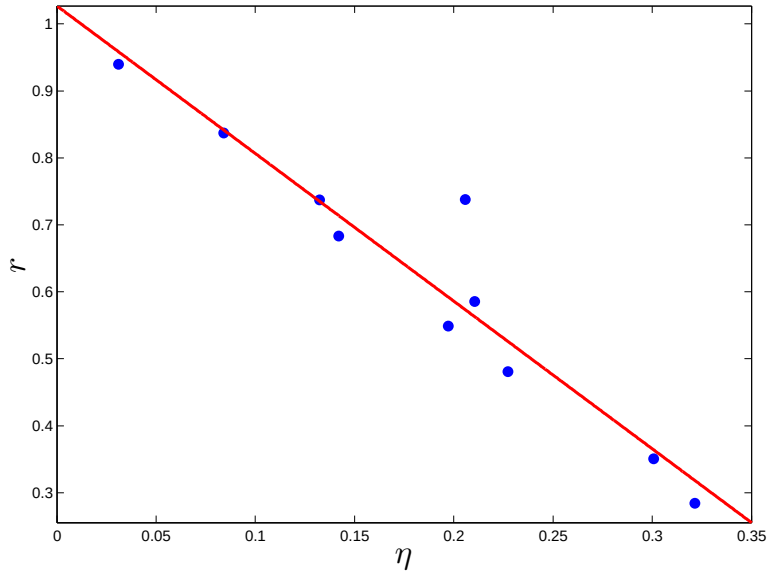


Figure 3.4: Noise to fundamental variance contribution represented by $r = \frac{\mu(1-p_0^\varepsilon)}{\mu(1-p_0^\varepsilon)+\lambda(1-p_0^\varepsilon)}$, plotted against a decreasing tick size represented by an increasing η. Each blue dot represent one asset and the red line is the linear fit.

Last Buy Prices

We now consider the process of last buy prices. We recall that the series consists of only the arrival times of market orders hit the ask price, coupled with the execution price expressed in ticks.

Description of the Figures: For the ask prices Figure 3.10 shows that the buy prices are closer to a martingale than the trade prices. This is seen by noticing that the ratio d is, on average, smaller in the case of the ask prices than it is for the trade prices. In this sense the ask prices are closer to a diffusion than the trade prices. Indeed, two of the assets (BUS5 and CL futures) no longer display the signature form of the variance they did when using trade prices. In spite of that, the other assets do present the signature variance behavior and the model fits them well.

The situation is more interesting for the autocovariance in Figure 3.11, where the empirical autocovariance of most assets is significant for more lags than in the case of trades, especially for large tick assets such as the DJ, Bobl, Schatz, Eurostoxx and SP futures. The model in this situation displays a behavior that is very similar to the empirical ACF.

Finally Figures 3.12 and 3.13 show the model comes close to the empirical distribution even though that distribution is not symmetric (the distribution of buy price returns is traditionally positively skewed).

Analysis of Fitting Results: η calculated on ask prices is not the same as η calculated on trade prices. Still r is (to some extent) increasing with the tick size (measured with η calculated with trade prices). This means that the autocorrelated noise behavior persists in a similar fashion when looking at ask prices, although this time it could not be interpreted as bid-ask bounces. There is indeed a deeper interpretation of the autocovariance of tick returns besides the bid-ask bounces originating in the orderbook, but this topic is beyond the scope of this presentation.

Midpoint Prices

We now consider the process of midpoint prices. We recall that the series consists of the arrival times of market orders, coupled with the midpoint price measured right before the arrival of the market order.

Contrary to the trade and bid prices, the series of midpoint returns has half integer support because the midpoint is the average of the bid and ask whose sum can be odd. For this reason we multiplied the series of tick returns by 2 in this case.

Description of the Figures: The mid point's efficiency (interested by the importance of d) is half way between that of the bid prices and that of the trade price. The results show that the ratio d is, on average, between that of the trade prices and that of the bid prices. Now, only one of the assets (CL futures) no longer displays the signature form of the variance. And similarly the other assets do present the signature variance behavior and the model fits them well.

Similarly to the last buy prices, the empirical autocovariance in Figure 3.15 of most assets is significant for more lags than in the case of trades, and we can see that on the EURO, DJ, Bobl, Bund, Schatz, Eurostoxx futures. The model in this situation displays a behavior that very similar to the empirical ACF.

Finally, Figures 3.12 and 3.13 show that the model **fails** to replicate the distribution of ticks. This is due to the empirical distribution of the midpoint tick returns whose series had to multiplied by 2 in order to put it on the lattice $\mathbb{Z}$. Now while this causes no problem with small tick assets whose prices often jumps by half ticks, the situation is more complicated for large tick assets whose midpoint almost never moves by less than a tick. So it is understandable that the model, whose distribution has a restricted support, can not replicate the distribution of tick returns in this case. We point out that we tested the model on large tick assets without multiplying their prices by 2 and the model performed well, similarly to the case of trade and buy prices.

Analysis of Fitting Results: Same remark as for the last buy price.

Conclusion and Further Research

We have presented in this chapter a simple model for asset prices on the high frequency scale. The model described in equation (3.3) is just the sum of two jump processes with values living on a lattice. The first one, that we call the random walk component, consists of the cumulative sum of a series of iid random variables arriving at independent Poisson times. The second one, that we can the autocorrelated noise component, consists of the cumulative sum of an MA(1) process also arriving at independent Poisson times.

The model has the particularity of converging to a diffusion as the scale becomes coarser, a property that is fundamental for the theory of asset pricing. However, on the microscopic scale, the model is flexible enough to reproduce one of the commonly observed statistics of prices at this scale, the signature plot effect, the phenomenon in which the Sampled Average Realized Variance (SRV) increases with the sampling frequency of the prices.

In practice, we use symmetric discrete compact distributions for the compounding jumps and a degenerated vanishing MA(1) process for the noise component. We wanted to see how much flexibility such a simple approach offers and we were happily surprised by the results. Indeed, we fit our model to several Futures assets and used different definitions for the price, the last traded price, the midpoint price, and the last buy price. We see that the model is robust with respect these changes and accounts for the weak but significant autocorrelation structure of tick returns and matches to a very large extent the unconditional distribution of observed tick returns. The results that we get are consistent with our expectations that the noise component becomes more important as the tick size becomes larger. The autocovariance of tick returns was not always well reproduced, however the model does provide an autocovariance structure that allows us to successfully capture the signature plot.

In the last section of this chapter below, we modify the model of section 3.4 in order to make it a true point process. This means that we discard the tick returns of size 0 and model jumps only. The following material is exploratory only, and since the results obtained only poorly fit the data – see below, Figure 3.5– we have decided not to investigate this approach further.

3.5 Supplemental material: a point process approach

What happens in the case we decide to look at the price as point process? This means that we neglect all the tick returns of size 0 and we look only at jumps. We propose in this section to construct a model in the same spirit of the model in

section 3.4, from which we artificially remove the tick returns of size 0. Let e_i be defined as in section 3.4, and define f_i as the "observed" non null part of e_i:

$$f_i = e_{\sigma(i)} \tag{3.20}$$

where σ is a random function $\mathbb{N} \to \mathbb{N}$ defined by

$$\sigma(i) = i^{th} \text{ element in } \{j \in \mathbb{N}, e_j \neq 0\}. \tag{3.21}$$

Next we define the process X_t that will represent the continuous time price process

$$X_t = \sum_{i=1}^{N_t} \varepsilon_i + \sum_{i=1}^{M_t} f_i \tag{3.22}$$

where ε_i are discrete symmetric iid random variables with support in $\{-2, -1, 1, 2\}$ and N_t, M_t are independent Poisson processes that are also independent of their respective compounding jumps ε_i and f_i. In this presentation, the arrival times of the Poisson process $(N_t + M_t)_{t\geq 0}$ corresponds to a jump of the process X_t. The jumps of X_t correspond to the observation of a new price and the series of price differentials $(J_i)_{i\geq 0}$ is no longer the series of tick returns, but the series of *non null* price returns.

In the following, we will express the unconditional distribution and the autocovariance function of f_i and provide an algorithm to calculate it. We are also going to see that f_i is weakly stationary as it can be written as a linear combination of a 2 dimensional stationary Markov Chain.

Introducing a 2 dimensional Markov Chain

First note that e_i can be written as

$$e_i = A.E_i$$

where $E_i = (\widetilde{\varepsilon}_i, \widetilde{\varepsilon}_{i-1})$ a 2 dimensional random vector, $A = (1, -1)$ and . denotes the scalar product. We easily see that E_i is a time homogenous irreducible ergodic Markov Chain of order 1 with a finite state space $\mathcal{E}$ (all the states are aperiodic and positive recurrent).

Next, define $H = \{E \in \mathcal{E}, A.E = 0\}$ and $\overline{H}$ its complementary. We can then

rewrite $\sigma(i)$ as:

$$\begin{aligned}\sigma(1) &= \min\{j > 0, E_j \in \overline{H}\} \\ \sigma(i) &= \min\{j > \sigma(i-1), E_j \in \overline{H}\};\ \text{for all } i > 1.\end{aligned}$$

This shows that the random variables $\sigma(i)$, $i \geq 1$, are strictly increasing finite stopping times (exit times from H, or entry times into $\overline{H}$, with $P(\sigma(i) < \infty) = 1$ thanks to the irreducibility of E_i). Now define the stopped Markov Chain $F_i = E_{\sigma(i)}$. We have that F_i is also an irreducible Markov Chain of order 1. This is seen by the Strong Markov Property. Indeed we have that each $i \geq 1$, the process $(E_{\sigma(i)+k})_k$, $k \geq 0$ starting at $e = E_{\sigma(i)}$ is a Markov Chain independent of the past (up to $\sigma(i)$) and conditionally to the initial condition they all have the same law. So we can write for $s_1, \ldots, s_{i-1} \in H$:

$$\begin{aligned}P(F_i = s_i | F_1 = s_1, \ldots, F_{i-1} = s_{i-1}) &= P(E_{\sigma(i)} = s_i | E_{\sigma(1)} = s_1, \ldots, E_{\sigma(i-1)} = s_{i-1}) \\ &= P(E_{\sigma(i)} = s_i | E_{\sigma(i-1)} = s_{i-1}) \\ &= P(E_{\sigma(2)} = s_i | E_{\sigma(1)} = s_{i-1})\end{aligned}$$

Now let P_E and P_F be respectively the transition matrices of E and F. The stationary probability of F is given by the vector π that satisfies $P_F'\pi = \pi$ (exists since F is irreducible). We also have that the n^{th} order transition matrices are just P_E^n and P_F^n. Furthermore, we make the assumption that F_0 is drawn according to S, which gives that $(F_i)_{i \geq 0}$ is stationary.

By construction, F can never transit into or from a state in H. Let $a \in \mathcal{E} = (H \bigcup \overline{H})$ and $b \in \overline{H}$. We then define an intermediate transition matrix $\widetilde{P}_F$. $\widetilde{P}_F(a, b)$ is the probability that the process E_i starts in a and arrives in b in "one step of F" i.e. *while completely forgetting any stops made any element of* H. This means that E_i can start in a, which may or may not be an element of H, stop many consecutive times in any intermediate element of H, and when it finally reaches b, the whole trajectory is considered as only one transition. This property is expressed with the following relationship:

$$\widetilde{P}_F(a, b) = P_E(a, b) + \sum_{n \in H} P_E(a, n)\widetilde{P}_F(n, b)$$

As a consequence, for each $b \in \overline{H}$ we have a linear system that allows us to calculate $\widetilde{P}_F(a, b)$ for all $a \in \mathcal{E}$ and in particular for all $a \in \overline{H}$. Restricting $\widetilde{P}_F$ over

elements in $\overline{H} \times \overline{H}$ gives us P_F.

First and Second Moments of f_i

f_i is now seen as

$$f_i = A.F_i$$

and since $(F_i)_{i\geq 0}$ is a stationary Markov Chain, $(f_i)_{i\geq 0}$ is also stationary. We have that its distribution lives in $\{-2,-1,1,2\}$ and since it is the distribution of e_i conditioned to $e_i \neq 0$, it can be expressed as a function of p such that

$$\begin{pmatrix} P(f_i=-2) \\ P(f_i=-1) \\ P(f_i=1) \\ P(f_i=2) \end{pmatrix} = \begin{pmatrix} \frac{p}{4-6p} \\ \frac{1-2p}{2-3p} \\ \frac{1-2p}{2-3p} \\ \frac{p}{4-6p} \end{pmatrix} \tag{3.23}$$

The autocovariance of f_i at lag j is defined by:

$$\Gamma_f(j) = E(f_i f_{i-j}) = E(A.F_i A.F_{i-j})$$

$\Gamma_f(j)$ can only be expressed as a function of P_F and we do not have simple closed form formulas for it. However it can be numerically calculated by:

$$\begin{aligned} \Gamma_f(j) &= \sum_{S,T\in\overline{H}} (A.S)(A.T)P(F_i=S)P(F_i=T|F_{i-j}=S) \\ \Gamma_f(j) &= \sum_{S,T\in\overline{H}} (A.S)(A.T)\pi(S)P_F^j(S,T) \end{aligned} \tag{3.24}$$

where $\pi(S) = P(F_i = S)$ is the stationary probability of the state S and $P_F(S,T) = P(F_i = T|F_{i-1} = S)$ is the transition matrix.

Realized Variance and Autocovariance of Tick Returns

The realized variance in this situation does not have a closed form formula and it has to be calculated using the full version of equation (3.5). Using the assumptions

of this section we can write:

$$V_T(\Delta, \mu, \lambda, p, p_1^\varepsilon) = (4 - 6p_1^\varepsilon)\lambda + \frac{2}{2-3p}\mu + 2\sum_{j=1}^{\infty} \Gamma_f(j) W(\mu, \Delta, j) \tag{3.25}$$

where $W(\mu, \Delta, j)$ is defined in equation (3.6) and $\Gamma_f(j)$ is calculated using equation (3.24). The autocovariance of the series of jumps $(J_i)\, i \geq 1$ of X_t for lag $j > 0$, is calculated as in equation (3.11) since none of the underlying assumption of the model have changed (f_i is weakly stationary and independent of ε_i, N_t and M_t) and we can write

$$\Gamma_J(j) = \sum_{k=1}^{j} \Gamma_f(k) \binom{k-1}{j-1} (\frac{\mu}{\lambda+\mu})^{k+1} (\frac{\lambda}{\lambda+\mu})^{j-k} \tag{3.26}$$

Application to Financial Data

Tested against real data, the model does not perform very well. Indeed the form of the SRV produced is extremely restricted and the ratio $d = \frac{V_T(0) - V_T(\infty)}{V_T(\infty)}$ it produces is more tightly bounded than in the case of the model in section 3.4. This is due to the fact that the autocovariance function of f has an alternating behavior, negative for odd lags and positive for even lags, which reduces the amplitude of $|\sum_{j=1}^{\infty} \Gamma_f(j) W(\mu, \Delta, j)|$ and therefore the effect of the negative reversion. So the model behaves even more poorly as the empirical ratio d of the asset increases, in other words for large tick assets. As an example we show the fit for the Bund futures in Figure 3.5.

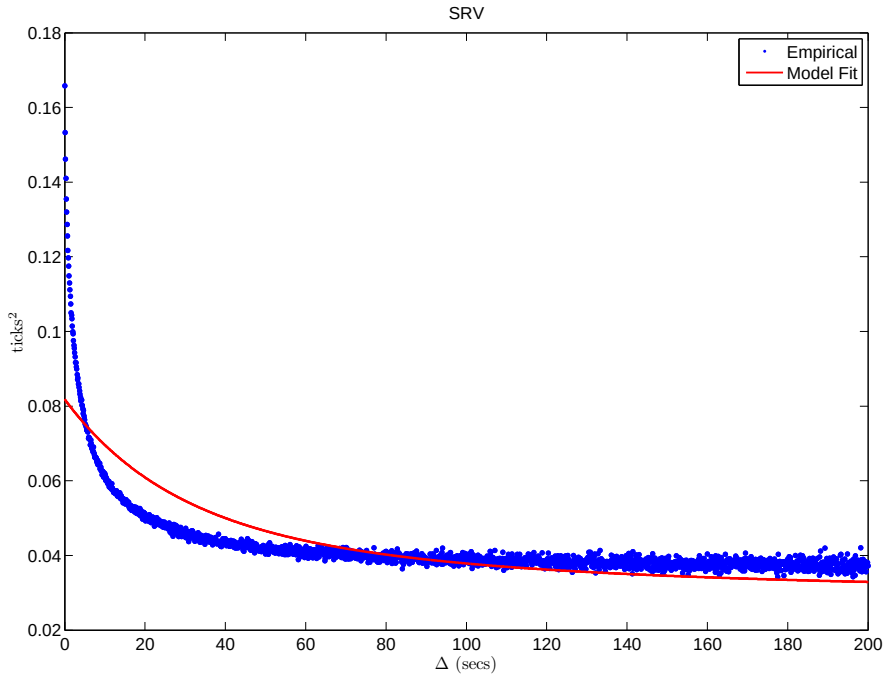


Figure 3.5: SRV plotted against Δ. The blue dots are for the empirical $V_T(\Delta)$ of the process of trade prices of Bund Futures and the red curve is the fitted curve of equation (3.25).

Appendices

3.A Tables

		BUS5	EURO	Dax	CL	DJ	Bobl	Bund	Schatz	Eurostoxx	SP
Random Walk	λ	0.227	0.459	0.406	0.142	0.696	0.159	0.162	0.128	0.176	0.982
	p_2^ε	0.018	0.013	0.067	0.497	0.007	0.002	0.001	0.000	0.000	0.000
	p_1^ε	0.109	0.125	0.163	0.003	0.098	0.102	0.134	0.041	0.064	0.016
	p_0^ε	0.746	0.724	0.540	0.000	0.790	0.792	0.728	0.919	0.871	0.967
	$E(\varepsilon_i^2)$	0.359	0.357	0.861	3.981	0.254	0.220	0.280	0.082	0.130	0.033
	$\lambda(1-p_0^\varepsilon)$	0.058	0.127	0.187	0.142	0.146	0.033	0.044	0.010	0.023	0.032
	$\lambda\sqrt{E(\varepsilon_i^2)}$	0.136	0.274	0.377	0.284	0.351	0.075	0.086	0.037	0.063	0.179
Noise Component	μ	0.320	0.216	0.343	2.392	0.622	0.205	0.435	0.118	0.459	1.780
	p_2^e	0.000	0.008	0.028	0.002	0.010	0.003	0.004	0.005	0.005	0.007
	p_1^e	0.035	0.150	0.223	0.082	0.157	0.095	0.105	0.119	0.122	0.135
	p_0^e	0.928	0.682	0.496	0.833	0.667	0.804	0.782	0.753	0.745	0.716
	$E(e_i^2)$	0.074	0.368	0.674	0.180	0.390	0.214	0.239	0.276	0.285	0.323
	$\mu(1-p_0^e)$	0.023	0.069	0.173	0.400	0.207	0.040	0.095	0.029	0.117	0.505
	$\mu\sqrt{E(e_i^2)}$	0.087	0.131	0.281	1.013	0.388	0.095	0.213	0.062	0.245	1.011
Desc. Stats	# Days	21	21	21	21	21	21	21	21	21	21
	# Trades	85687	217082	191962	585321	411036	62572	107061	46858	260616	452989
	Start Date	6/5/2009	6/5/2009	6/5/2009	6/5/2009	6/12/2009	6/5/2009	6/5/2009	6/5/2009	6/5/2009	6/12/2009
	End Date	7/7/2009	7/3/2009	7/3/2009	7/3/2009	7/10/2009	7/3/2009	7/3/2009	7/3/2009	7/3/2009	7/13/2009
	Obervations/sec	0.574	0.766	0.847	2.665	1.446	0.414	0.708	0.310	0.766	3.105
Tick Stats	η	0.335	0.324	0.286	0.239	0.229	0.201	0.146	0.134	0.084	0.031
	$\frac{\mu}{\lambda}$	1.411	0.471	0.845	16.792	0.893	1.286	2.689	0.924	2.603	1.811
	$\frac{\mu(1-p_0^e)}{\lambda(1-p_0^\varepsilon)}$	0.397	0.541	0.926	2.812	1.414	1.216	2.157	2.805	5.135	15.609
	$\frac{\mu\sqrt{E(e_i^2)}}{\lambda\sqrt{E(\varepsilon_i^2)}}$	0.638	0.478	0.747	3.566	1.107	1.267	2.485	1.698	3.861	5.666
	r	0.284	0.351	0.481	0.738	0.586	0.549	0.683	0.737	0.837	0.940

Table 3.3: Calibration results and descriptive statistics using trade prices. The assets are ordered from left right by the tick size.

		BUS5	EURO	Dax	CL	DJ	Bobl	Bund	Schatz	Eurostoxx	SP
Random Walk	λ	0.294	0.321	0.355	1.295	0.315	0.136	0.278	0.111	0.259	1.072
	p_2^ε	0.019	0.038	0.085	0.029	0.035	0.011	0.005	0.001	0.001	0.000
	p_1^ε	0.055	0.105	0.143	0.102	0.139	0.084	0.061	0.041	0.040	0.013
	p_0^ε	0.851	0.714	0.543	0.737	0.651	0.810	0.869	0.917	0.920	0.974
	$E(\varepsilon_i^2)$	0.265	0.511	0.969	0.439	0.561	0.256	0.159	0.090	0.084	0.028
	$\lambda(1-p_0^\varepsilon)$	0.044	0.092	0.162	0.341	0.110	0.026	0.036	0.009	0.021	0.028
	$\lambda\sqrt{E(\varepsilon_i^2)}$	0.151	0.229	0.349	0.858	0.236	0.069	0.111	0.033	0.075	0.178
Noise Component	μ	0.004	0.068	0.081	0.000	0.405	0.071	0.065	0.046	0.102	0.428
	p_2^e	0.138	0.012	0.047	0.000	0.002	0.001	0.002	0.001	0.001	0.000
	p_1^e	0.191	0.171	0.245	0.000	0.081	0.055	0.081	0.066	0.053	0.023
	p_0^e	0.342	0.634	0.416	1.000	0.835	0.889	0.834	0.864	0.893	0.954
	$E(e_i^2)$	1.487	0.439	0.864	0.000	0.177	0.116	0.178	0.143	0.111	0.047
	$\mu(1-p_0^e)$	0.003	0.025	0.047	0.000	0.067	0.008	0.011	0.006	0.011	0.020
	$\mu\sqrt{E(e_i^2)}$	0.005	0.045	0.075	0.000	0.171	0.024	0.027	0.017	0.034	0.092
Desc. Stats	# Days	21	21	21	21	21	21	21	21	21	21
	# Trades	45005	107874	95090	278871	192027	32216	56474	24988	128102	226837
	Start Date	6/5/2009	6/5/2009	6/5/2009	6/5/2009	6/12/2009	6/5/2009	6/5/2009	6/5/2009	6/5/2009	6/12/2009
	End Date	7/7/2009	7/3/2009	7/3/2009	7/3/2009	7/10/2009	7/3/2009	7/3/2009	7/3/2009	7/3/2009	7/13/2009
	Obervations/sec	0.302	0.381	0.420	1.270	0.676	0.213	0.374	0.166	0.377	1.555
Tick Stats	η	0.393	0.436	0.421	0.429	0.351	0.358	0.365	0.270	0.274	0.274
	$\frac{\mu}{\lambda}$	0.013	0.210	0.229	0.000	1.289	0.523	0.234	0.416	0.392	0.399
	$\frac{\mu(1-p_0^e)}{\lambda(1-p_0^\varepsilon)}$	0.059	0.270	0.293	0.000	0.611	0.304	0.297	0.677	0.520	0.712
	$\frac{\mu\sqrt{E(e_i^2)}}{\lambda\sqrt{E(\varepsilon_i^2)}}$	0.032	0.195	0.216	0.000	0.725	0.352	0.248	0.526	0.451	0.520
	r	0.056	0.212	0.226	0.000	0.379	0.233	0.229	0.404	0.342	0.416

Table 3.4: Calibration results and descriptive statistics using Last Buy Prices. The assets are ordered from left right by the tick size.

		BUS5	EURO	Dax	CL	DJ	Bobl	Bund	Schatz	Eurostoxx	SP
Random Walk	λ	0.480	0.591	0.685	0.567	0.863	0.104	0.199	0.069	0.327	1.166
	p_2^ε	0.074	0.138	0.236	0.499	0.100	0.167	0.113	0.074	0.033	0.013
	p_1^ε	0.030	0.000	0.071	0.000	0.004	0.000	0.000	0.000	0.000	0.000
	p_0^ε	0.793	0.724	0.386	0.000	0.792	0.666	0.775	0.851	0.934	0.975
	$E(\varepsilon_i^2)$	0.652	1.104	2.030	3.996	0.811	1.335	0.901	0.596	0.265	0.100
	$\lambda(1-p_0^\varepsilon)$	0.100	0.163	0.421	0.567	0.180	0.035	0.045	0.010	0.022	0.029
	$\lambda\sqrt{E(\varepsilon_i^2)}$	0.388	0.621	0.976	1.133	0.777	0.120	0.189	0.053	0.168	0.369
Noise Component	μ	0.027	0.151	0.186	35.471	0.616	0.251	0.408	0.159	0.326	1.709
	p_2^e	0.169	0.128	0.250	0.000	0.028	0.017	0.012	0.016	0.008	0.001
	p_1^e	0.147	0.203	0.000	0.001	0.223	0.191	0.169	0.190	0.148	0.052
	p_0^e	0.369	0.337	0.500	0.997	0.498	0.584	0.638	0.586	0.687	0.896
	$E(e_i^2)$	1.642	1.434	2.000	0.003	0.672	0.515	0.432	0.512	0.363	0.109
	$\mu(1-p_0^e)$	0.017	0.100	0.093	0.104	0.310	0.104	0.148	0.066	0.102	0.179
	$\mu\sqrt{E(e_i^2)}$	0.035	0.181	0.264	1.919	0.505	0.180	0.268	0.114	0.196	0.564
Desc. Stats	# Days	21	21	21	21	21	21	21	21	21	21
	# Trades	85687	217082	191962	585321	411036	62572	107061	46858	260616	452989
	Start Date	6/5/2009	6/5/2009	6/5/2009	6/5/2009	6/12/2009	6/5/2009	6/5/2009	6/5/2009	6/5/2009	6/12/2009
	End Date	7/7/2009	7/3/2009	7/3/2009	7/3/2009	7/10/2009	7/3/2009	7/3/2009	7/3/2009	7/3/2009	7/13/2009
	Obervations/sec	0.574	0.766	0.847	2.665	1.446	0.414	0.708	0.310	0.766	3.105
Tick Stats	η	0.320	0.379	0.394	0.411	0.370	0.206	0.200	0.114	0.153	0.160
	$\frac{\mu}{\lambda}$	0.057	0.256	0.272	62.576	0.714	2.416	2.053	2.324	0.995	1.466
	$\frac{\mu(1-p_0^e)}{\lambda(1-p_0^\varepsilon)}$	0.173	0.613	0.221	0.183	1.723	3.009	3.298	6.456	4.699	6.110
	$\frac{\mu\sqrt{E(e_i^2)}}{\lambda\sqrt{E(\varepsilon_i^2)}}$	0.090	0.292	0.270	1.693	0.650	1.501	1.421	2.155	1.164	1.528
	r	0.148	0.380	0.181	0.155	0.633	0.751	0.767	0.866	0.825	0.859

Table 3.5: Calibration results and descriptive statistics using Midpoint Prices. The assets are ordered from left right by the tick size.

3.B Figures

3.B.1 Trade Prices

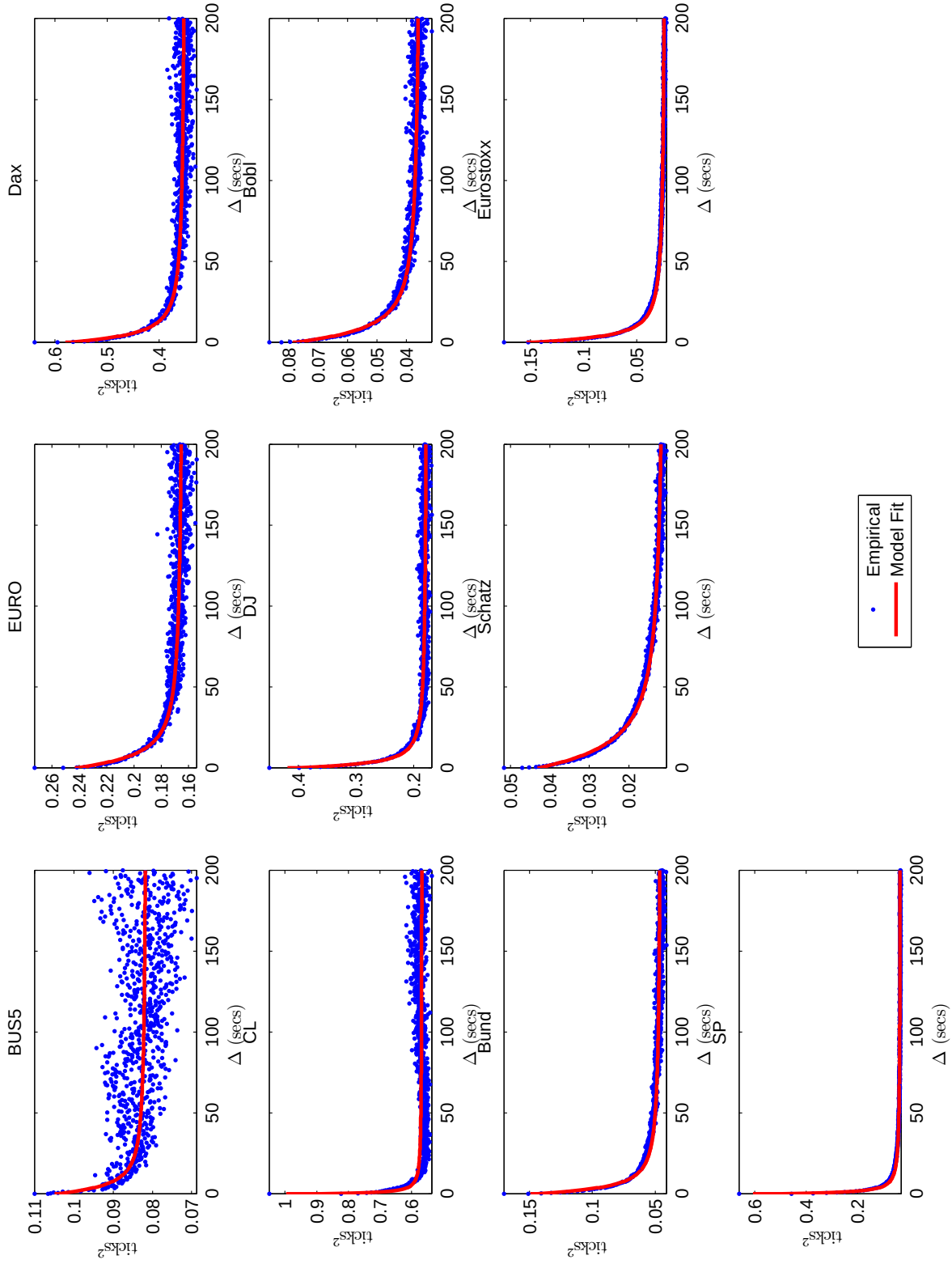


Figure 3.6: SRV fit results for the series of trade prices. The assets are ordered according to their tick size. Each blue dots represent the $V_T(\Delta)$ for the sampling duration Δ. The red curve is the function (3.16) fitted to the cloud of blue dots.

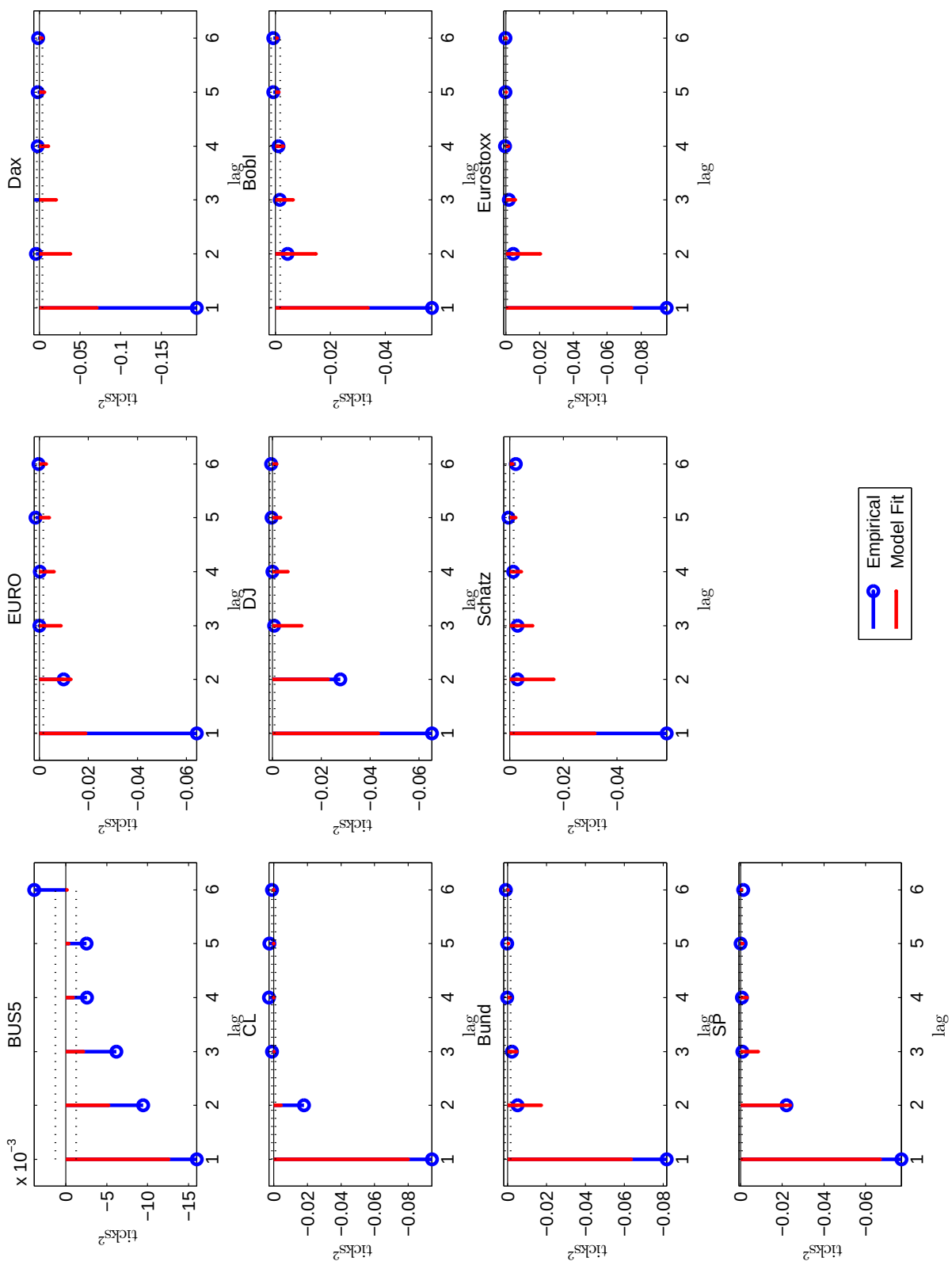


Figure 3.7: 6 lags autocovariance functions of trade tick returns. The empirical ACF is in blue and the one of the fitted model is in red. The black dotted lines are the 95% significance bounds under the assumption that the return series are iid gaussian. We can see that the empirical autocovariance for most assets is significant and that the model does capture some of it.

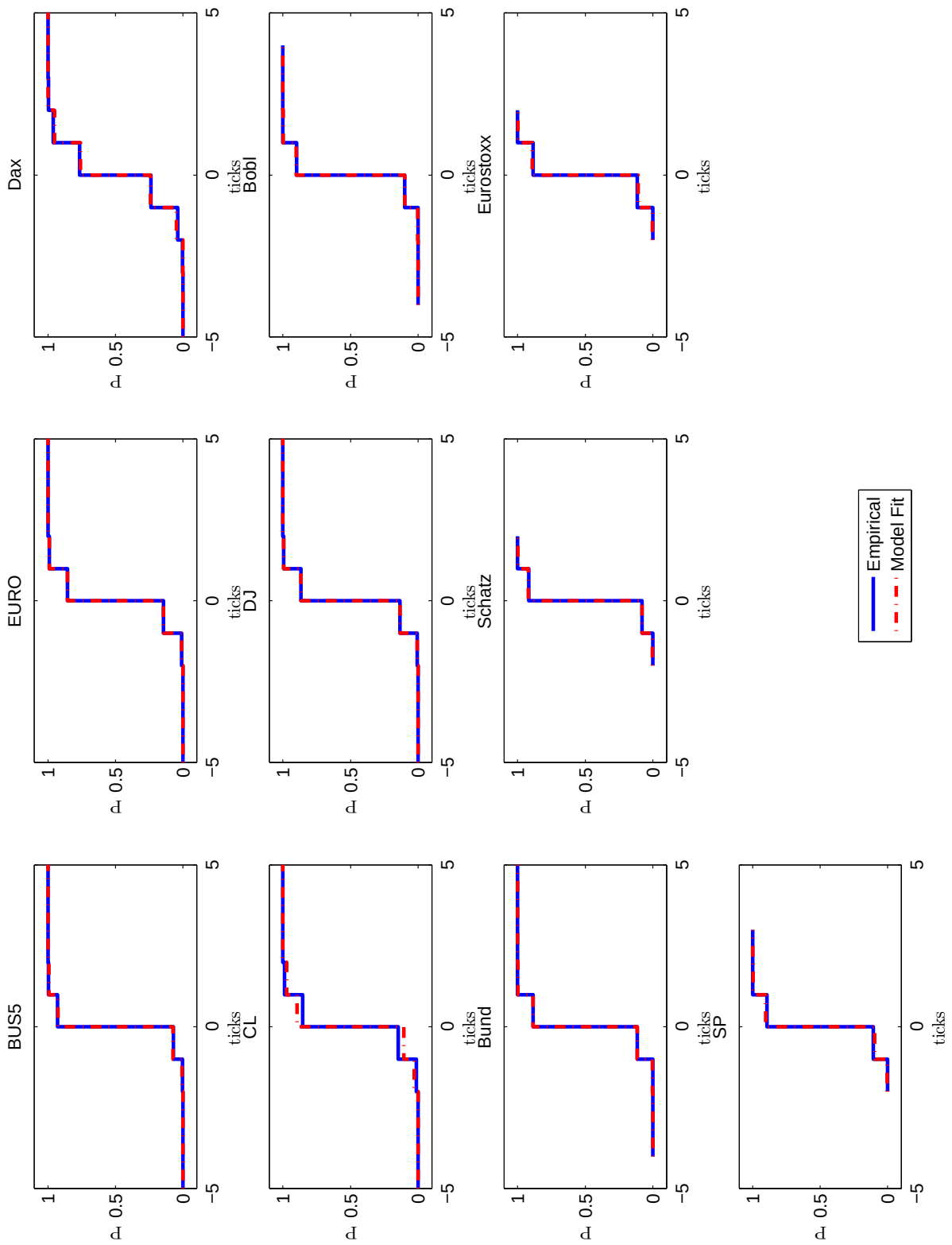


Figure 3.8: Cumulative distribution functions of trade tick returns. The blue line is the empirical cdf while the red dotted curve is the CDF of the fitted model.

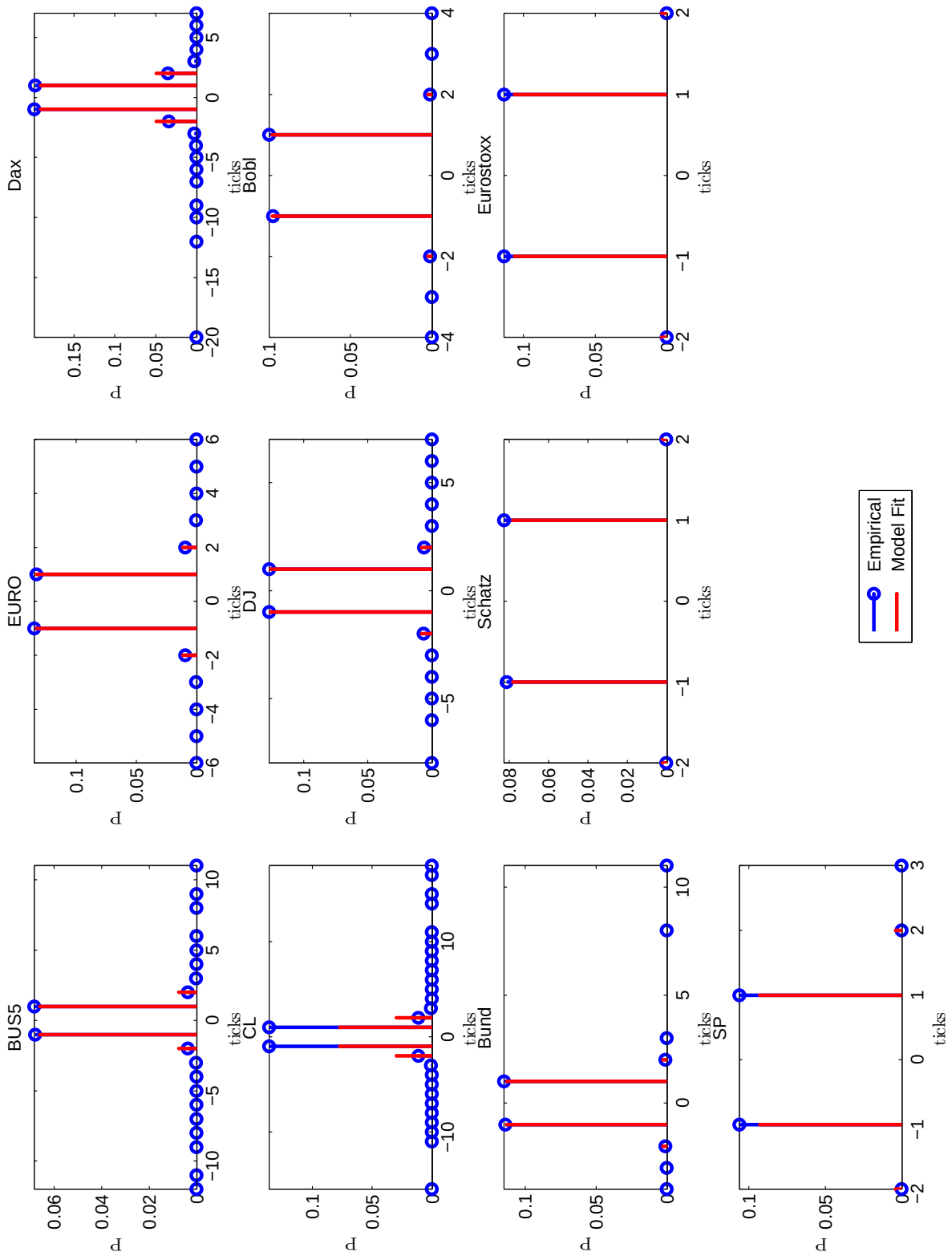


Figure 3.9: Probability distribution functions ($f_J(x)$) of trade tick returns. $f_J(0)$ was removed to improve readability. The blue line is the empirical PDF while the red line is the PDF of the fitted model. We notice that as as the tick becomes larger the distribution of tick returns becomes more compact with a support that matches that of our model. We also notice that most of the tick returns are zero valued.

3.B.2 Last Bid Prices

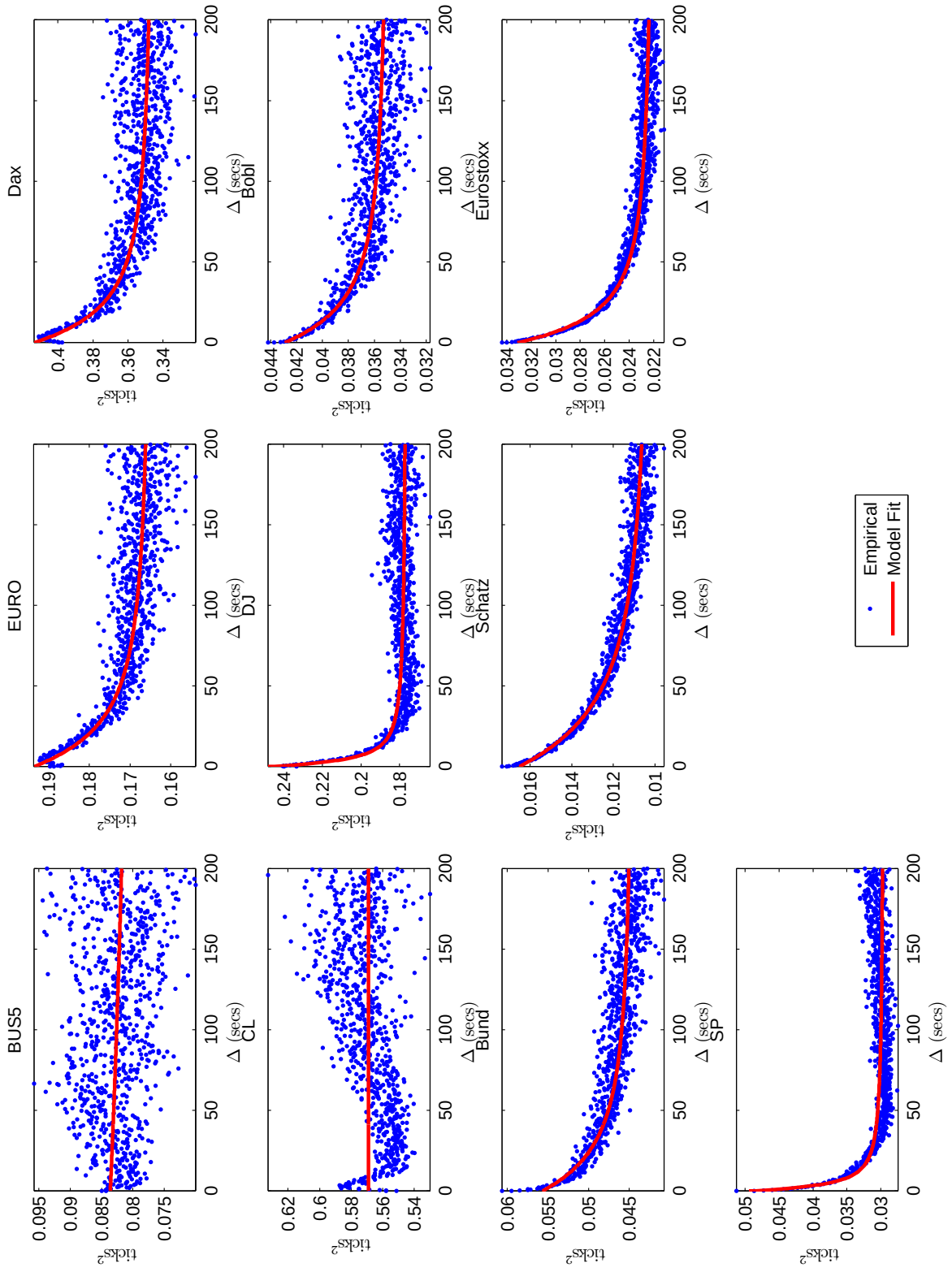


Figure 3.10: SRV fit results for the series of Last Bid prices. The assets are ordered according to their tick size. Each blue dots represent the $v_T(\Delta)$ for the sampling duration Δ. The red curve is the function (3.16) fitted to the cloud of blue dots.

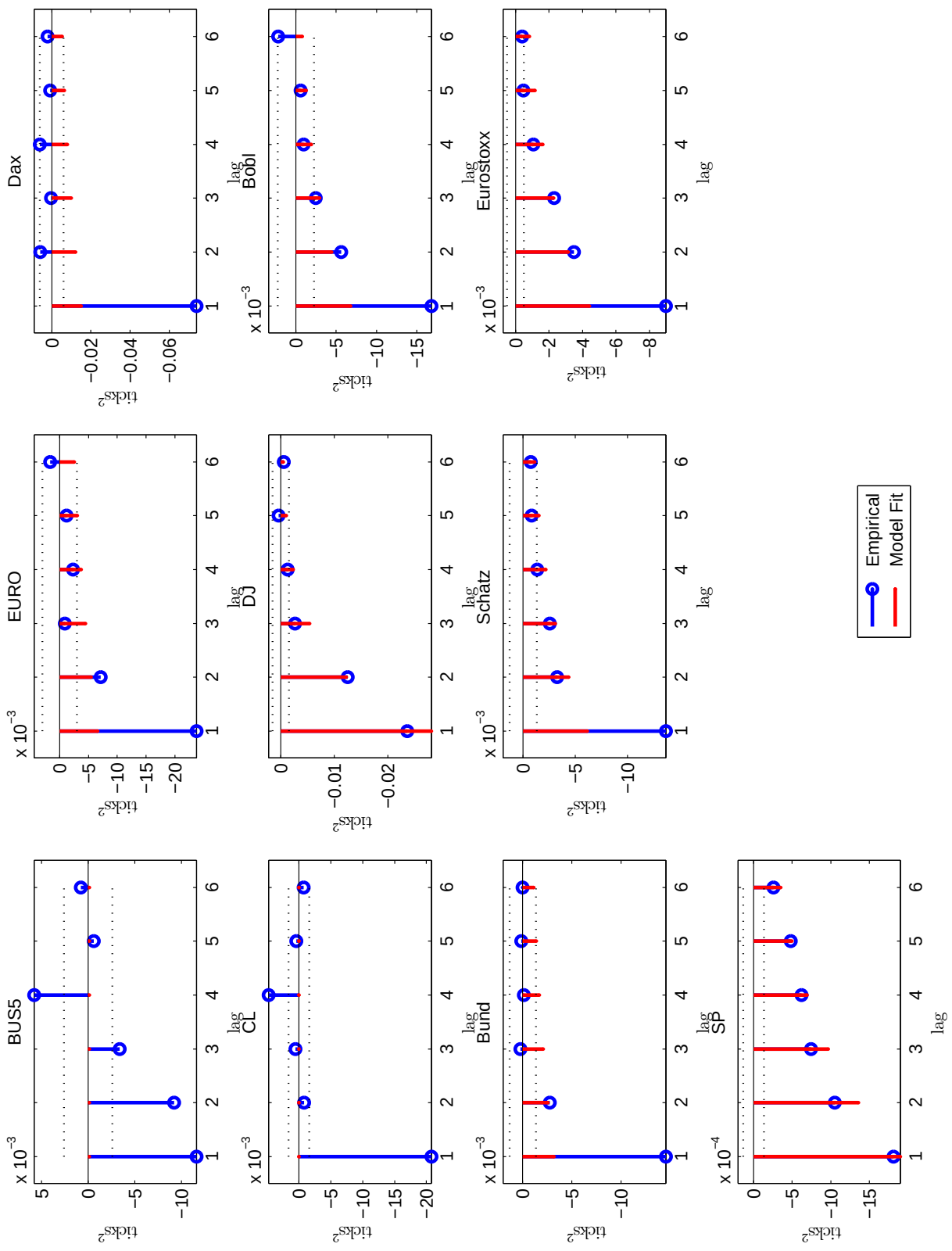


Figure 3.11: 6 lags autocovariance functions of last bid tick returns. The empirical ACF is in blue and the one of the fitted model is in red. The black dotted lines are the 95% significance bounds under the assumption that the return series are iid gaussian. We can see that the empirical autocovariance for most assets is significant and that the model does capture some of it.

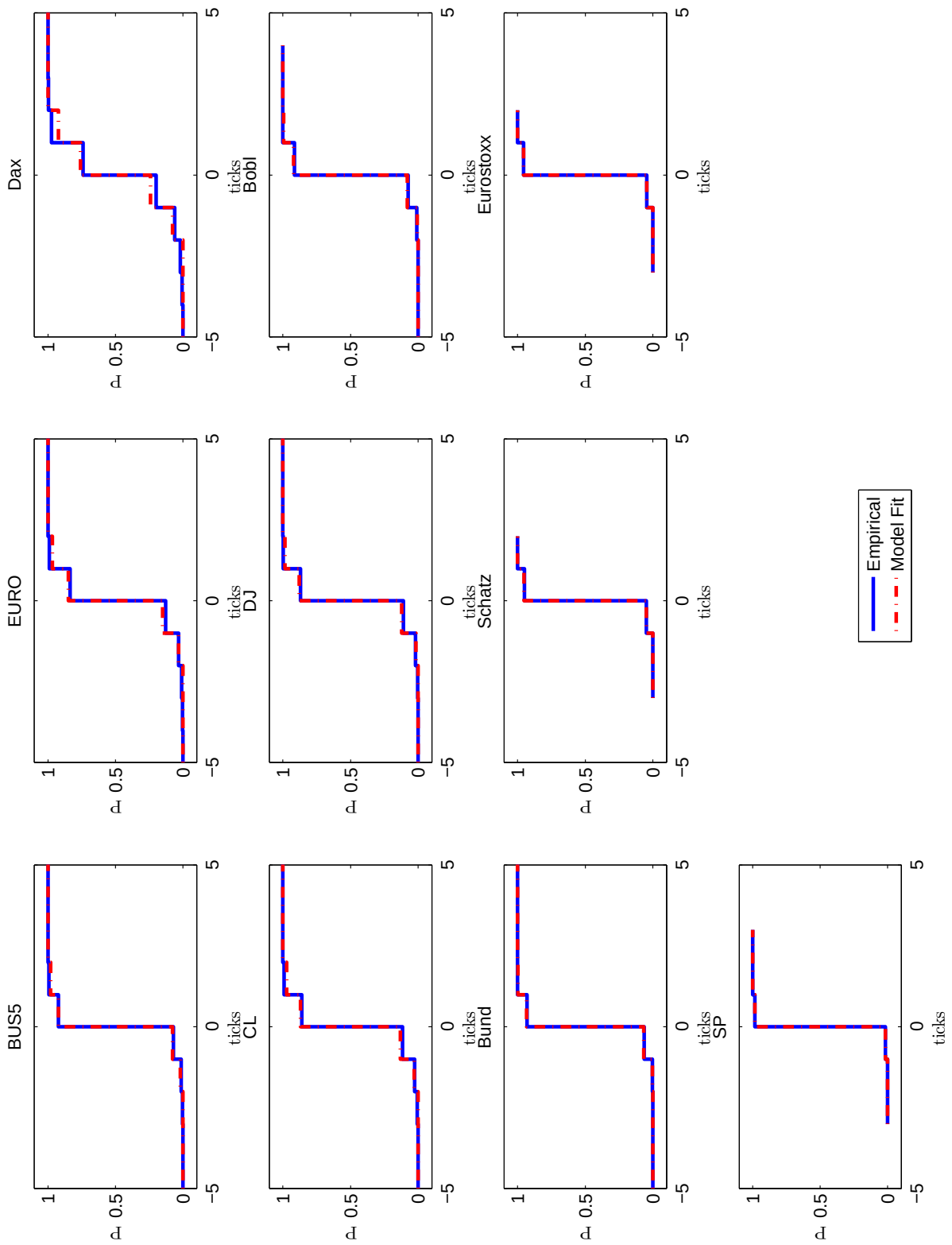


Figure 3.12: Cumulative distribution functions of last bid tick returns. The blue line is the empirical CDF while the red dotted curve is the CDF of the fitted model.

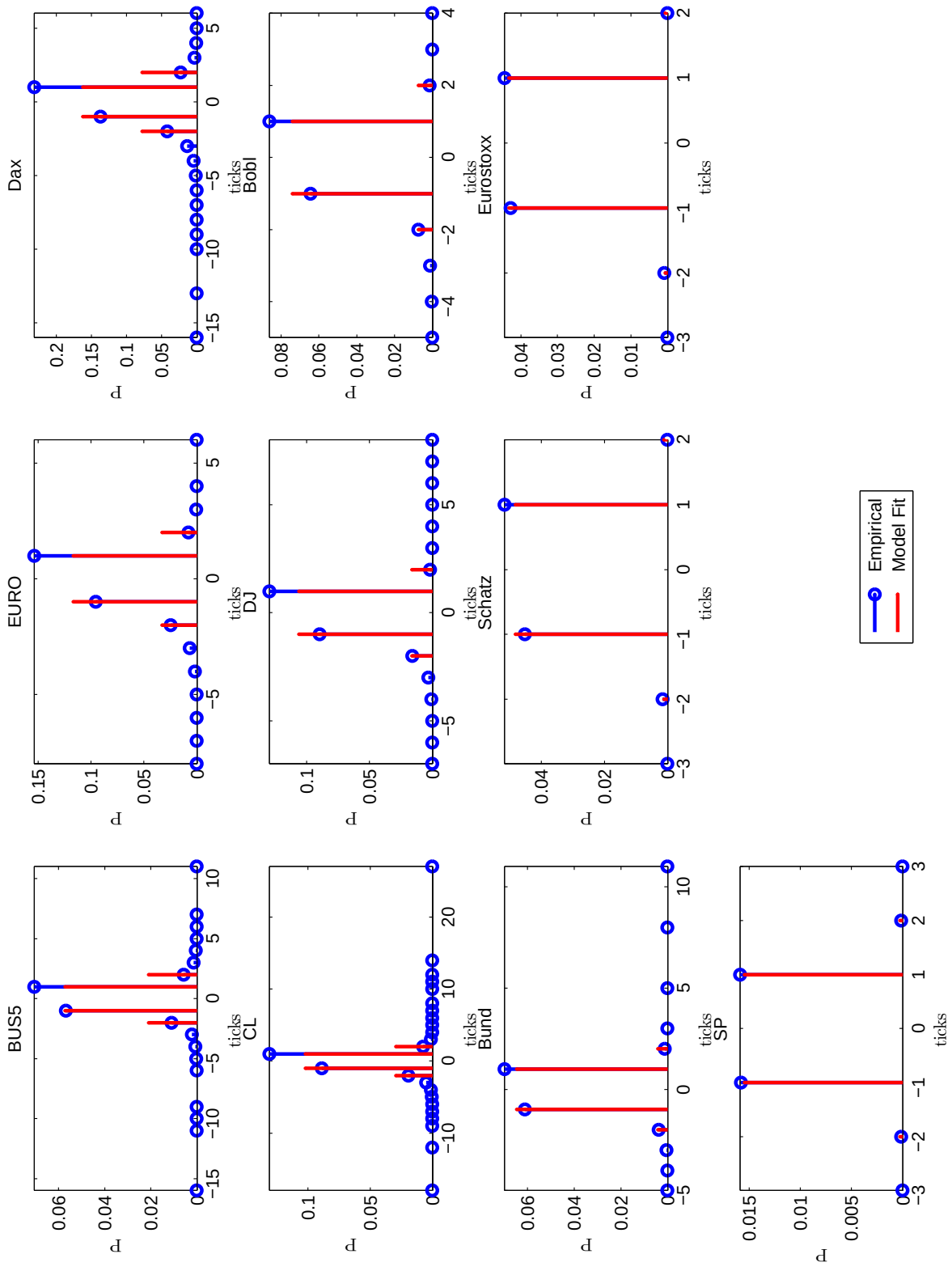


Figure 3.13: Probability distribution functions ($f_J(x)$) of last buy tick returns. $f_J(0)$ was removed to improve readability. The blue line is the empirical PDF while the red line is the PDF of the fitted model. We notice that as as the tick becomes larger the distribution of tick returns becomes more compact with a support that matches that of our model. We also notice that most of the tick returns are zero valued.

3.B.3 Midpoint Prices

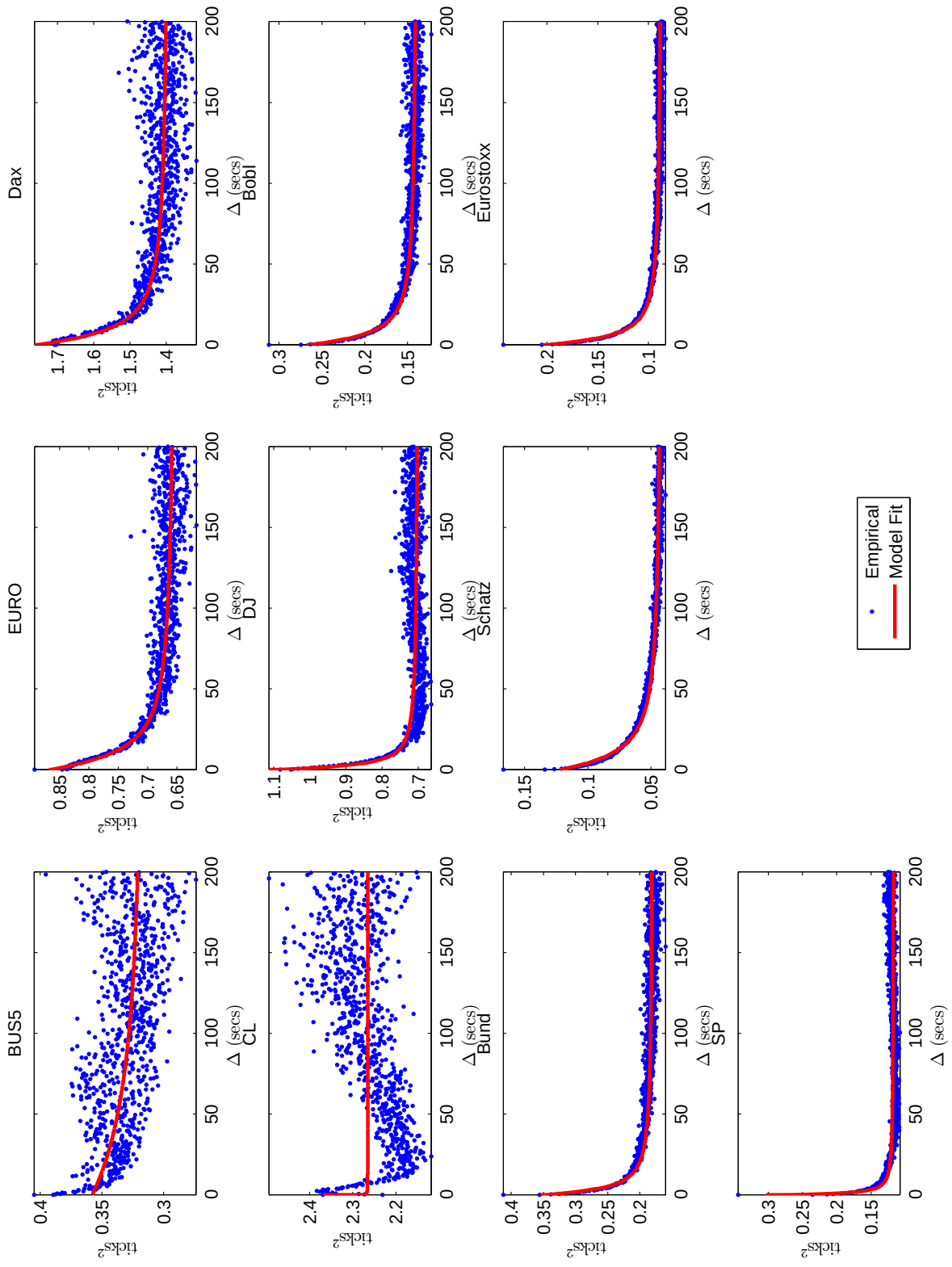


Figure 3.14: SRV fit results for the series of midpoint tick returns. The assets are ordered according to their tick size. Each blue dots represent the $V_T(\Delta)$ for the sampling duration Δ. The red curve is the function (3.16) fitted to the cloud of blue dots.

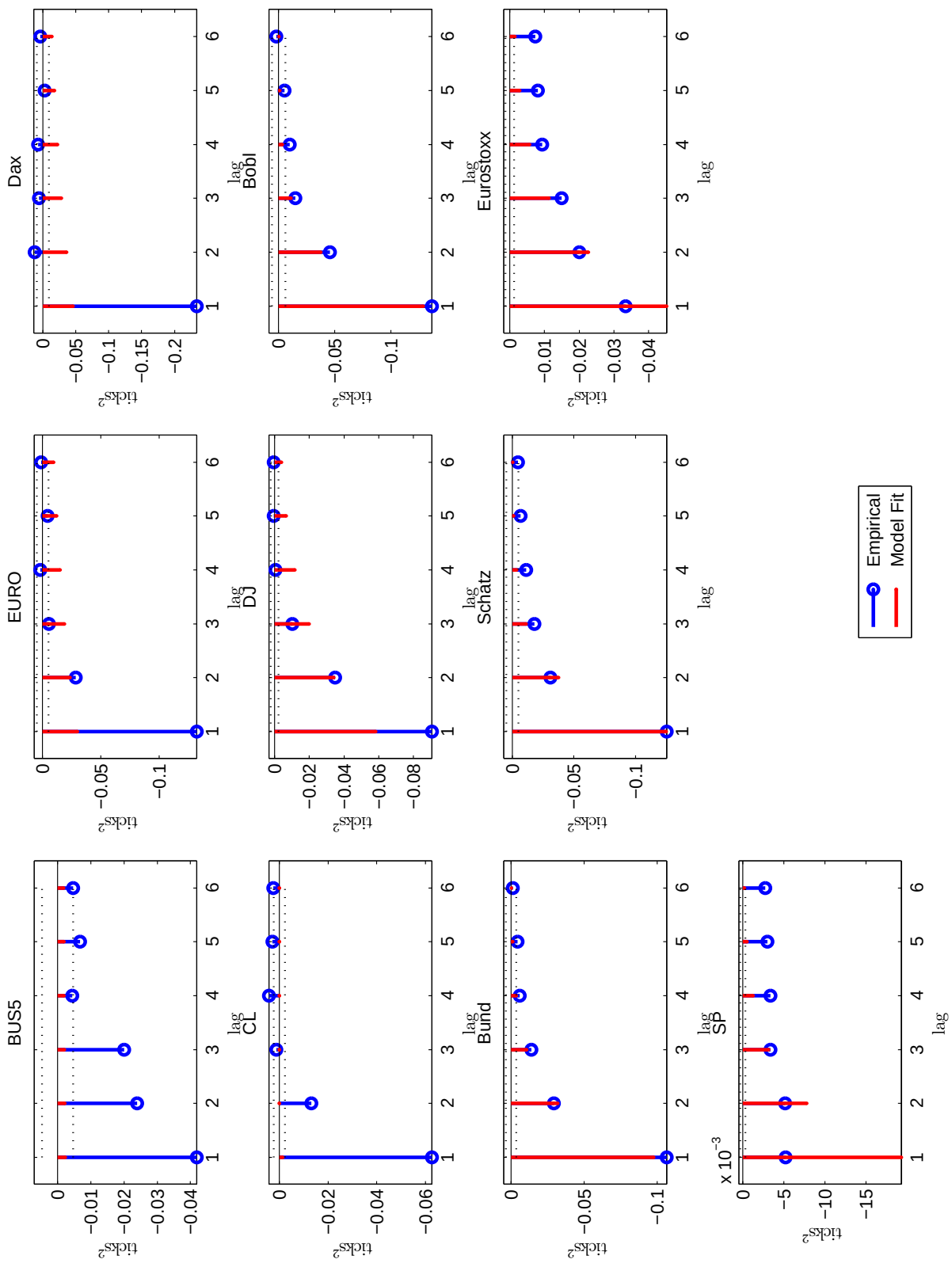


Figure 3.15: 6 lags autocovariance functions of midpoint tick returns. The empirical ACF is in blue and the one of the fitted model is in red. The black dotted lines are the 95% significance bounds under the assumption that the return series are iid gaussian. We can see that the empirical autocovariance for most assets is significant and that the model does capture some of it.

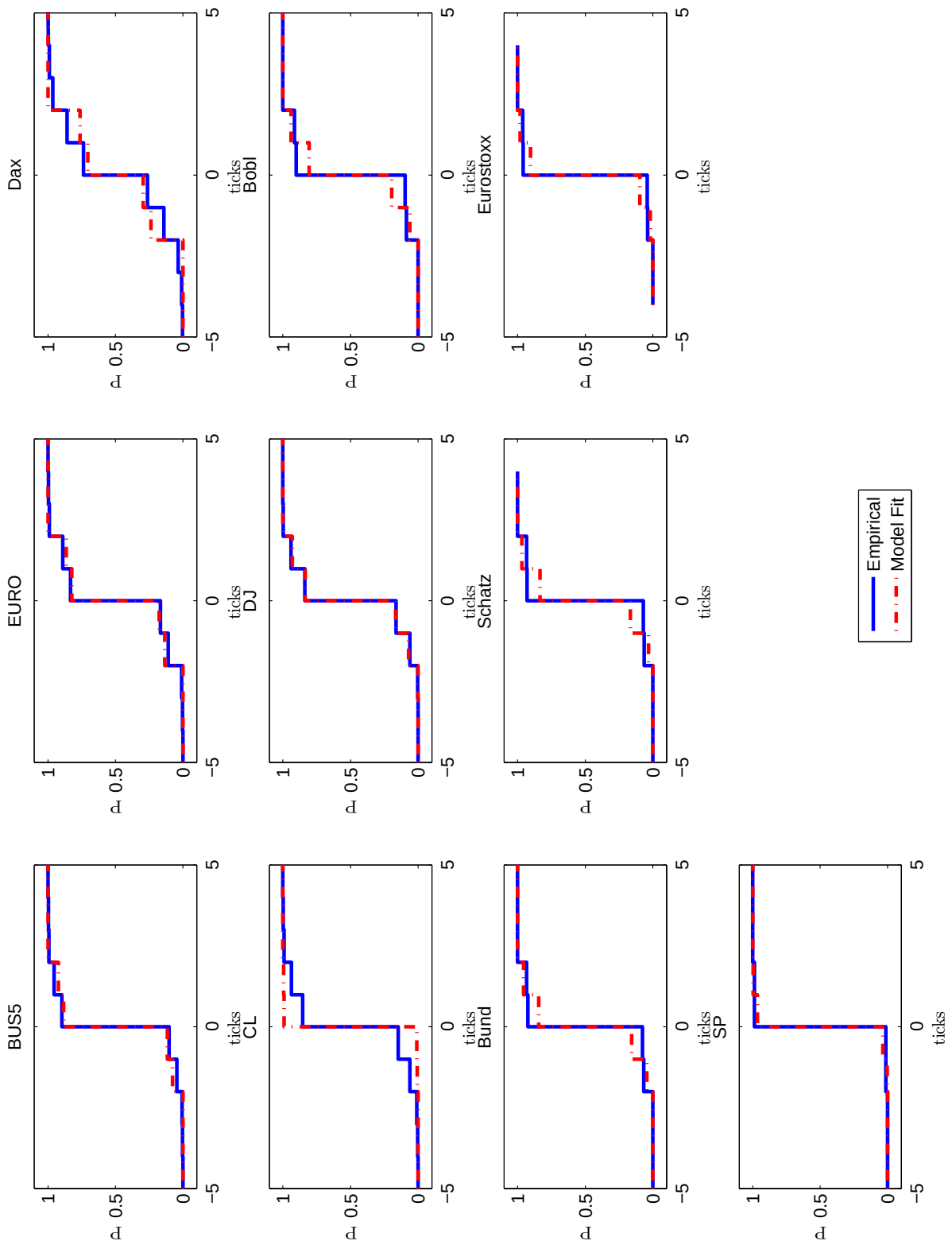


Figure 3.16: Cumulative distribution functions of midpoint tick returns. The blue line is the empirical CDF while the red dotted curve is the CDF of the fitted model.

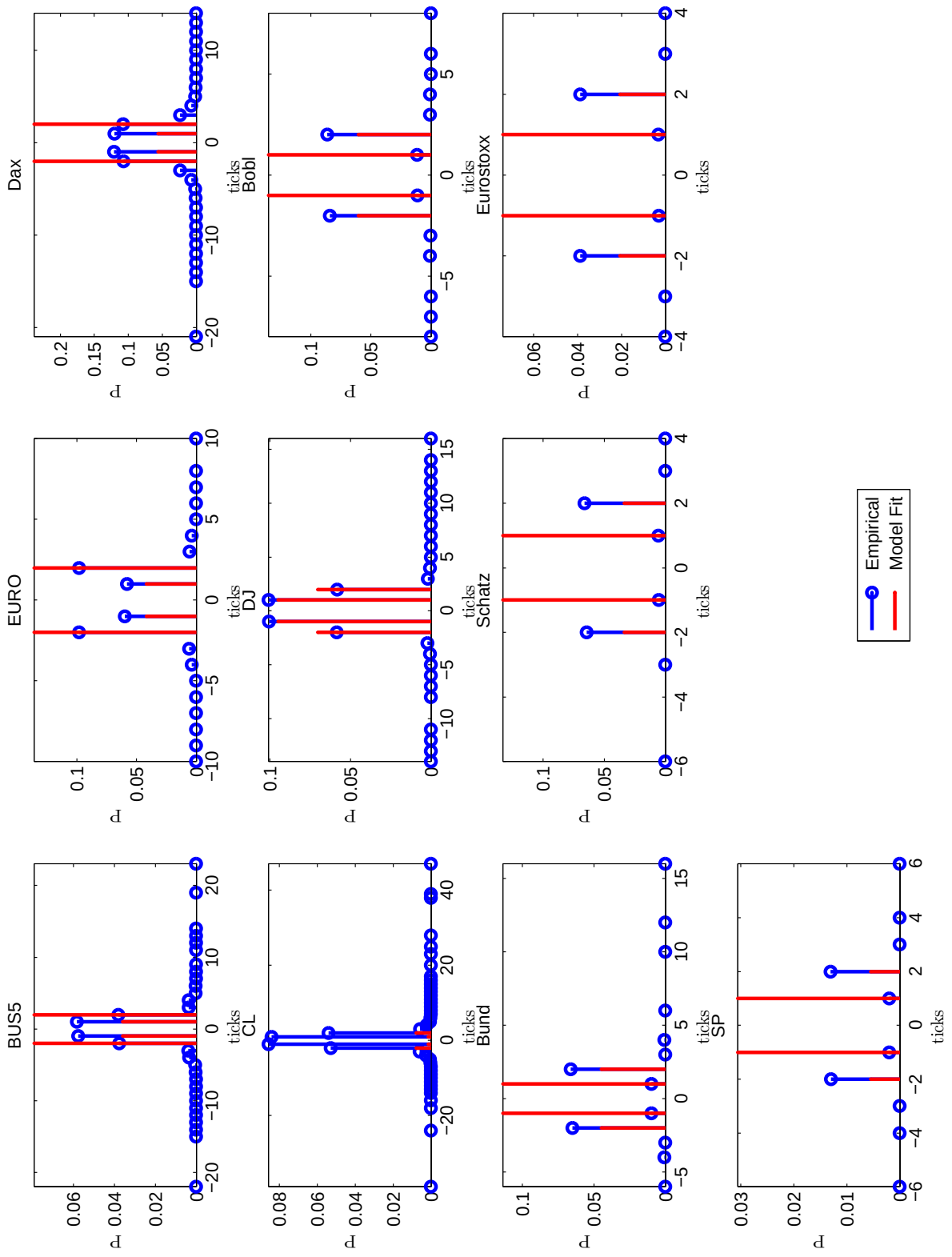


Figure 3.17: Probability distribution functions ($f_J(x)$) of midpoint tick returns. $f_J(0)$ was removed to improve readability. The blue line is the empirical PDF while the red line is the PDF of the fitted model. We notice that as as the tick becomes larger the distribution of tick returns becomes more compact with a support that matches that of our model. We also notice that most of the tick returns are zero valued.

CHAPTER 4

Hawkes Kernel Estimation

4.1 Introduction

Although the concept of self excitement was commonly used for a long time by seismologists (see [85] and the references therein), Alan Hawkes was the first to provide a well defined point process with a self exciting behavior [46, 47, 55]. This model was first introduced to reproduce the ripple effects generated after the occurrence of an earthquake [86, 2]. Since then, it has been successfully used in many areas, ranging from seismology (see e.g., [66], for a recent review), biology [74], to even criminology and terrorism [61, 32] (cf [55] and references therein for a detailed review of Hawkes processes and their applications). As far as financial applications are concerned, since transactions and price changes are discrete events, Hawkes processes have naturally been generating more and more interest. Applications can be found in the field of order arrival rate modeling [48, 24, 83], noise microstructure dynamics [9], volatility clustering [35], extreme values and VaR estimation [25] and credit modeling [42].

Hawkes models account for a self exciting behavior of events by which the arrival of one event increases the probability of occurrence of new ones. In its most basic form and in the one dimensional case, the Hawkes process is a counting process

defined by λ_t, the rate of arrival of events by:

$$\lambda_t = \mu + \int_{-\infty}^{t} \phi_{t-s} dN_s \tag{4.1}$$

where $\mu > 0$ is a constant **background** rate, N_t is the cumulative counting process and ϕ a positive real function called **decay kernel**. We can clearly see in equation (4.1) that when event occurs at time t, we have $dN_t = 1$ and hence $d\lambda_t = \phi_0$. The influence of the event is transmitted to future times through ϕ such that at time $u > t$ the increase in λ_t due to the time t event is ϕ_{u-t}. Thus a self exciting behavior is observed.

A basic issue in applications of Hawkes processes concerns their estimation. In early applications, parametric estimation used spectral analysis by means of the Bartlett spectrum (the Fourier transform of the autocovariance of the process) [15, 16] and it was indeed through that light that Hawkes presented his model, however this method was later seen to be less satisfactory.A maximum likelihood method for estimating the parameter of exponential, power law and Laguerre class of decay kernels was developed in [70, 67] and it became the standard method for estimating Hawkes processes. Furthermore, the Laguerre decay kernels were seen to be very pertinent decay kernels because they allowed to account for long term dependencies as well as offering short term flexibility. More recently, other types of estimation procedures were developed. When the form of the decay kernel is unknown non-parametric methods are desirable because they give an idea of the general form of the decay kernel. For example, by using the branching property of the Hawkes process [90], the authors in [58] and [53] were able to provide Expectation-Maximization algorithms to estimate both background rate and the decay kernel. In [74], the authors present also an algorithmic method for decay estimation by using a penalized projection method.

In this paper, we propose an alternative simple non parametric estimation method for multivariate symmetric Hawkes processes based on the Bartlett spectrum Fourier transform. We present the method and its numerical feasibility without going too much in details about convergence speeds or error optimization. By studying one dimensional and two dimensional examples, we show that our approach provides reliable estimates on both fast and slowly decaying kernels. We then apply our method to high frequency financial data and find power-law kernels. This implies that the arrival of events display long range correlations, a phenomena well known in finance, similarly to what was suggested in [17].

This paper is organized as follows. In section 4.2, we introduce a basic version of a multivariate Hawkes processes. We place ourselves in the context of an

$n-$dimentional linear Hawkes process with a constant background rate and a non-negative decay kernel. We set some notations and give out a martingale representation of the rate function λ_t in Eq. (4.1). In section 4.3, we study the autocovariance of the process. Along the same line as [47], we establish its relationship with the decay kernel in both direct and Fourier spaces. The estimation method in the case of symmetric Hawkes processes is then provided in section 4.4. This method, based on a Hilbert transform phase recuperation method, is explicitly detailed in both univariate and special symmetric bivariate cases. In section 4.5, the method is illustrated by numerical examples for both exponential and power law kernels. We also address some statistical issues concerning the estimation errors. Finally, in section 4.6, we apply our method to high frequency financial data for which we deduce a long range nature of the decay kernels.

4.2 Multivariate Hawkes Processes

4.2.1 Notations and Definitions

As introduced by Hawkes in [46] and [47], let us consider an n-dimensional point process process $(N_t)_{t\geq 0}$, where N_t^i $1 \leq i \leq n$ represents the cumulative number of events in the i^{th} component of the process N_t up to time t.

The conditional intensity vector at any time t is assumed to be a random process that depends on the past as

$$\lambda_t = \mu + \int_{-\infty}^{t} \phi_{t-s} dN_s \tag{4.2}$$

where $\mu > 0$ is a vector of size n and ϕ_t referred to as the **decay kernel**, is an $n \times n$ matrix.

Notations 4.2.1. In the following

- $\mathcal{M}_{n,p}(\mathbb{R})$ (resp. $\mathcal{M}_{n,p}(\mathbb{C})$) denotes the set of $n \times p$ matrices with values in $\mathbb{R}$ (resp. $\mathbb{C}$). For any matrix M (resp. vector v), M^{ij} (resp. v^i) denotes its elements.
- For any function f_t, $\widehat{f_z} = \int_{\mathbb{R}} e^{-zt} f_t dt$ corresponds to its Laplace transform.
- By extension if $M_t \in \mathcal{M}_{n,p}(\mathbb{R})$ then $\widehat{M_z} \in \mathcal{M}_{n,p}(\mathbb{C})$ corresponds to the matrix whose elements are the Laplace transforms of the elements of M_t.

Using these notations, one of the main results of Hawkes is that if

(H1) the kernel is positive and causal, i.e.,

$$\phi : \mathbb{R} \to \mathcal{M}_{n,n}(\mathbb{R}^+), \quad \forall t < 0,\ \phi_t = 0,\ \text{and } \forall t \geq 0,\ \phi_t^{ij} \geq 0, \tag{4.3}$$

(H2) and if the spectral radius of $\widehat{\phi_0}$ (i.e., its largest eigen value) is strictly smaller than 1,

then $(N_t)_{t\geq 0}$ is a n-dimensional point process with stationary increments. The conditional intensity λ_t is itself a stationary process with mean

$$\Lambda = E(\lambda_t) = E(dN_t)/dt.$$

Combining this last equation with Eq. (4.2), one easily gets $\Lambda = \mu + \int_{-\infty}^{t} \phi_{t-s} ds \Lambda = \mu + \int_0^{\infty} \phi_u du \Lambda$ and consequently

$$\Lambda = (\mathbb{I} - \widehat{\phi_0})^{-1} \mu,$$

where $\mathbb{I}$ refers to the $n \times n$ identity matrix.

Before moving on, we need to introduce some more notations that will be used all along the paper.

Notations 4.2.2. If $A_t \in \mathcal{M}_{m,n}(\mathbb{R})$ and $B_t \in \mathcal{M}_{n,p}(\mathbb{R})$ then the convolution product of A_t and B_t is naturally defined as $A \star B_t = \int_{\mathbb{R}} A_s B_{t-s} ds = \int_{\mathbb{R}} A_{t-s} B_s ds$. Of course it is associative and distributive however it is generally not commutative (unless A_t and B_t are commutative). The neutral element is $\delta \mathbb{I}_t$, i.e., the diagonal matrix with Dirac distribution on the diagonal. In the following we will use the notation :

$$A \star dN_t = \int_{\mathbb{R}} A_{t-s} dN_s.$$

Combining both Notations 4.2.1 and 4.2.2, it is easy to show that the convolution theorem on matrices translates in

$$\widehat{A \star B}_z = \widehat{A}_z \widehat{B}_z.$$

4.2.2 Martingale Representation of λ_t

We roughly follow a similar path to Hawkes in [47]. Let $(M_t)_{t\geq 0}$ be the martingale compensated process of $(N_t)_{t\geq 0}$ defined by:

$$dM_t = dN_t - \lambda_t dt \ . \tag{4.4}$$

Then λ_t can be represented as a stochastic integral with respect to the martingale $(M_t)_{t\geq 0}$:

Proposition 4.2.1. *One has:*

$$\lambda_t = \Lambda + \Psi \star dM_t, \tag{4.5}$$

where Ψ_t is defined as

$$\Psi_t = \sum_{n=1}^{\infty} \phi_t^{(\star n)}, \tag{4.6}$$

where $\phi_t^{(\star n)}$ refers to the n^{th} auto-convolution of ϕ_t (i.e., $\widehat{\phi^{(\star n)}}_z = (\widehat{\phi}_z)^n$)

Proof. Using equations (4.2) and (4.4) one has:

$$\lambda_t = \mu + \phi \star dN_t = \mu + \phi \star dM_t + \phi \star \lambda_t,$$

and consequently

$$(\delta \mathbb{I} - \phi) \star \lambda_t = \mu + \phi \star dM_t.$$

Let us note that the inverse of $\delta \mathbb{I}_t - \phi_t$ for the convolution product is nothing but $h_t = \delta \mathbb{I}_t + \Psi_t$. Thus convoluting on each side of the last equation by h_t, one gets

(since $h \star \phi_t = \Psi_t$)

$$\lambda_t = h \star \mu + \Psi \star dM_t.$$

Since μ is a constant one has $h \star \mu = \widehat{h}_0 \mu$. Using the convolution theorem one gets $\widehat{h}_0 \mu = (\mathbb{I} - \widehat{\phi}_0)^{-1} \mu = \Lambda$ which proves the proposition. □

4.3 The Covariance Operator of Hawkes Processes

The kernel estimator we are going to build is based on the empirical auto-covariance of $(N_t)_{t \geq 0}$. This section is devoted to the covariance operator of the n-dimensional Hawkes processes. However, we first discuss some useful results about their *infinitesimal* auto-covariance operator.

4.3.1 The Infinitesimal Covariance Operator

Let us define the infinitesimal covariance matrix:

$$\nu_{t-t'} = E(dN_t dN_{t'}^\dagger),$$

where $M^\dagger$ denotes the hermitian conjugate of a matrix M. Along the same way as Hawkes's in [47], $\nu_{t-t'}$ can be related to the kernel ϕ defining the process. We now present in proposition 4.3.1 an equation linking the infinitesimal covariance matrix to Λ and Ψ but, unlike Hawkes, we express ν explicitly as a function of Ψ and Λ instead of an integral equation in ν. This result will be at the heart of the estimation method we propose in this paper.

Proposition 4.3.1. *Let $(N_t)_{t \geq 0}$ be an n-dimensional Hawkes process with intensity λ_t as defined in Section 4.2.1 (assuming both (H1) and (H2)). Let $\widetilde{\Psi}_t = \Psi_{-t}$. We have the following result:*

$$E(dN_t dN_{t'}^\dagger) = \left(\mathbf{\Sigma} \delta_{t-t'} + \Psi_{t-t'} \mathbf{\Sigma} + \mathbf{\Sigma} \Psi_{t'-t}^\dagger + \Lambda \Lambda^\dagger + \widetilde{\Psi} \star \mathbf{\Sigma} \Psi_{t'-t}^\dagger \right) dt dt' \tag{4.7}$$

where $\mathbf{\Sigma}$ is the diagonal matrix defined by $\mathbf{\Sigma}^{ii} = \Lambda^i$ for all $1 \leq i \leq n$ and δ_t is the dirac distribution.

Proof. By using equation (4.4), we can write:

$$\begin{aligned} E(dN_t dN_{t'}^\dagger) &= E((dM_t + \lambda_t dt)(dM_{t'} + \lambda_{t'} dt')^\dagger) \\ &= E(dM_t dM_{t'}^\dagger) && (4.8a) \\ &\quad + E(\lambda_t dM_{t'}^\dagger) dt && (4.8b) \\ &\quad + E(dM_t \lambda_{t'}^\dagger) dt' && (4.8c) \\ &\quad + E(\lambda_t \lambda_{t'}^\dagger) dt dt' && (4.8d) \end{aligned}$$

We begin by noticing that for $t \neq t'$ $E(dM_t dM_{t'}^\dagger) = 0$. Indeed if $\tau > 0$, and $s < t$

$$\begin{aligned} E((M_t - M_s)(M_{t+\tau} - M_{s+\tau})^\dagger) &= E(M_t(M_{t+\tau} - M_{s+\tau})^\dagger) - E(M_s(M_{t+\tau} - M_{s+\tau})^\dagger) \\ &= E(M_t E((M_{t+\tau} - M_{s+\tau})^\dagger | M_t)) - E(M_s E((M_{t+\tau} - M_{s+\tau})^\dagger | M_s)) \end{aligned}$$

Thanks to the martingale property, we have $E(M_s E((M_{t+\tau} - M_{s+\tau})^\dagger | M_s)) = 0$ and, as $t - s$ goes to 0, we have $s + \tau > t$ and $E((M_{t+\tau} - M_{s+\tau})^\dagger | M_t) = 0$. We thus obtain:

$$E(dM_t dM_{t'}^\dagger) = \delta_{t-t'} E(dM_t dM_{t'}^\dagger)$$

As for when $t = t'$ we have for all $1 \leq i < j \leq n$

$$E(dM_t^i dM_t^j) = 0$$

since the N^i's (and hence the M^i's) for $1 \leq i \leq n$ have no jumps in common. For $i = j$ we have

$$E(dM_t^i dM_t^i) = \Lambda^i dt$$

because $E(dM_t^i dM_t^i) = E(dN_t^i dN_t^i)$ and $E(dN_t^i dN_t^i) = E(dN_t^i) = \Lambda^i dt$ since the jumps of N_t are of size 1. To sum up, the term (4.8a) becomes:

$$E(dM_t dM_{t'}^\dagger) = \mathbf{\Sigma} \delta_{t-t'} dt dt' \tag{4.9}$$

The remaining of terms can be then calculated along the same line. Replacing

λ_t's expression from equation (4.5) in the term (4.8b), gives:

$$E(\lambda_t dM_{t'}^\dagger)dt = E((\Psi \star dM_t dM_{t'}^\dagger)dt + E(\Lambda dM_{t'}^\dagger)dt$$

and using Fubini's theorem and the fact that M_t is a martingale we get:

$$E(\lambda_t dM_{t'}^\dagger)dt = \int_{\mathbb{R}} \Psi_{t-s} E(dM_s dM_{t'}^\dagger)dtds$$

and thanks to equation (4.9),

$$E(\lambda_t dM_{t'}^\dagger)dt = \int_{\mathbb{R}} \Psi_{t-s} \mathbf{\Sigma} \delta_{s-t'} ds dt' dt$$

that is simplified to:

$$E(\lambda_t dM_{t'}^\dagger)dt = \Psi_{t-t'} \mathbf{\Sigma} dt dt' \tag{4.10}$$

Similarly, the term (4.8c) becomes:

$$E(dM_t \lambda_{t'}^\dagger)dt' = \mathbf{\Sigma} \Psi_{t'-t}^\dagger dt dt' \tag{4.11}$$

and finally, using Eq. (4.9), the term (4.8d) can be written as:

$$\begin{aligned} E(\lambda_t \lambda_{t'}^\dagger)dtdt' &= \left(\Lambda E(\lambda_{t'}^\dagger) + E((\Psi \star dM_t)\lambda_{t'}^\dagger\right) dtdt' \\ &= \left(\Lambda\Lambda^\dagger + \int_{\mathbb{R}} \Psi_{t-s} E(dM_s \lambda_{t'}^\dagger ds)\right) dtdt' \\ &= \left(\Lambda\Lambda^\dagger + \int_{\mathbb{R}} \Psi_{t-s} \mathbf{\Sigma} \Psi_{t'-s}^\dagger ds)\right) dtdt' \end{aligned}$$

By setting $u = t' - s$, we get:

$$E(\lambda_t \lambda_{t'}^\dagger)dtdt' = \left(\Lambda\Lambda^\dagger + \widetilde{\Psi} \star \mathbf{\Sigma} \Psi_{t'-t}^\dagger\right) dtdt' \tag{4.12}$$

□

4.3.2 The Covariance Operator

The (normalized) covariance operator of the Hawkes process can be defined, at scale h and lag τ, by

$$v_\tau^{(h)} = h^{-1}\mathrm{Cov}\left(N_{t+h} - N_t, N_{t+h+\tau} - N_{t+\tau}\right) \ , \tag{4.13}$$

where we normalized by h in order to avoid a trivial scale dependence in expressions below. Let us note that, since the increments of N_t are stationary, the previous definition does not depend on t. Thus, it can be rewritten as

$$v_\tau^{(h)} = \frac{1}{h}E\left((\int_0^h dN_s - \Lambda h)(\int_\tau^{\tau+h} dN_s^\dagger - \Lambda^\dagger h)\right) \tag{4.14}$$

The information contained in the ACF allows one to recover the classical Signature Plot of the prices. Because it is more general, we will see that it can be used, in some case, to get a non-parametric estimation of the kernel matrix ϕ.

Indeed, let us choose a fixed value of the increment scale h and consider $v_\tau^{(h)}$ as a function of the time lag τ. This function contains information about lead-lag behavior of the process that originates in the kernel ϕ. While in reference [9], the authors used $v_0^{(h)}$ in order to estimate the parameters of an exponential kernel, we are going to use the more general $v_\tau^{(h)}$ in order to find a non parametric estimation of ϕ. Theorem 4.3.2 holds the key to that.

Theorem 4.3.2. *Let $g_t^{(h)} = (1 - \frac{|t|}{h})^+$. $v_\tau^{(h)}$ can be expressed as a function of $g_\tau^{(h)}$ and Ψ_τ:*

$$v_\tau^{(h)} = g_\tau^{(h)}\mathbf{\Sigma} + g^{(h)} \star \Psi_{-\tau}\mathbf{\Sigma} + g^{(h)} \star \mathbf{\Sigma}\Psi_\tau^\dagger + g^{(h)} \star \widetilde{\Psi} \star \mathbf{\Sigma}\Psi_\tau^\dagger \tag{4.15}$$

Proof of theorem 4.3.2. Let us begin the proof by noticing that for any function f with values in $\mathbb{R}^+$ we have:

$$h^{-1}\int_0^h \int_\tau^{\tau+h} f_{t-t'}dt'dt = f \star g_{-\tau}^{(h)} \tag{4.16}$$

It follows that

$$\begin{aligned} v_\tau^{(h)} &= \frac{1}{h} E\left(\int_0^h dN_s \int_\tau^{\tau+h} dN_s^\dagger - \int_0^h dN_s \Lambda^\dagger h - \Lambda h \int_\tau^{\tau+h} dN_s^\dagger + \Lambda\Lambda^\dagger h^2 \right) \\ &= \frac{1}{h} E\left(\int_0^h \int_\tau^{\tau+h} dN_t dN_{t'}^\dagger - \Lambda\Lambda^\dagger h^2 \right) . \end{aligned}$$

Using equation (4.7), we can split $dN_t dN_{t'}^\dagger$ into four parts and write:

$$v_\tau^{(h)} = \frac{1}{h} \int_0^h \int_\tau^{\tau+h} \left(\mathbf{\Sigma}\delta_{t-t'} + \Psi_{t-t'}\mathbf{\Sigma} + \mathbf{\Sigma}\Psi_{t'-t}^\dagger + \widetilde{\Psi} \star \mathbf{\Sigma}\Psi_{t'-t}^\dagger \right) dt dt'$$

By applying equation (4.16) to each of the terms under the double integral we get equation (4.15) and achieve the proof. □

It is more convenient to rewrite the result of theorem 4.3.2 in Laplace domain. Since $\widehat{\phi^{(n)}}_z = \widehat{\phi}_z^n$, Eq. (4.6) translates into :

$$\widehat{\Psi}_z = \sum_{n=1}^{+\infty} \widehat{\phi}_z^n = \widehat{\phi}(\mathbb{I} - \widehat{\phi})^{-1} \tag{4.17}$$

and conversely:

$$\widehat{\phi}_z = (\mathbb{I} + \widehat{\Psi}_z)^{-1} \widehat{\Psi}_z \tag{4.18}$$

Theorem 4.3.2 gives way to the following corollary.

Corollary. *In Laplace domain equation (4.15) becomes:*

$$\widehat{v}_z^{(h)} = \widehat{g}_z^{(h)} (\mathbb{I} + \widehat{\Psi}_z^*) \mathbf{\Sigma} (\mathbb{I} + \widehat{\Psi}_z^*)^\dagger \tag{4.19}$$

4.4 Non-parametric Estimation of the Kernel ϕ_t

4.4.1 The Estimation Principle

In this paper, we aim at building an estimator of ϕ_t based on empirical measurements of $v_\tau^{(h)}$. Let us note that empirical measurements of $v_\tau^{(h)}$ are naturally obtained replacing probabilistic mean by empirical mean in Eq. (4.14) (see [10] for proof of convergence of the empirical mean towards the probabilistic mean). Thus,

in order to build an estimator, we need to express ϕ_t as a function of $v_\tau^{(h)}$. In the Laplace domain, since a given $\widehat{\Psi}_z$ corresponds to a unique $\widehat{\phi}_z$ (through Eq. (4.18)), it translates in trying to express $\widehat{\Psi}_z$ as a function of $\widehat{v}_z^{(h)}$ which exactly corresponds to inverting Eq. (4.19) (i.e., computing the square root of $(\mathbb{I} + \widehat{\Psi}_z^*)\mathbf{\Sigma}(\mathbb{I} + \widehat{\Psi}_z^*)^\dagger$). Indeed, knowing $\widehat{\Psi}_z$, one easily gets $\widehat{\phi}_z$ (from Eq. (4.18)) and finally ϕ_t. Actually, let us note that, from a practical point of view, we don't need to work in the full complex domain $z \in \mathbb{C}$ of the Laplace transform. Working with the Fourier transform restriction (i.e., $z = i\omega$ with $\omega \in \mathbb{R}$) is enough to recover ϕ_t.

Discretizing and Dealing With Cancelations of $\widehat{g}_z^{(h)}$

The first problem that seems to appear for inverting this formula (4.19) (i.e., expressing $\widehat{\Psi}_z$ as a function of $\widehat{v}_z^{(h)}$ only for $z = i\omega$) is the scalar term $\widehat{g}_z^{(h)}$ that may become null. Indeed, since $g_\tau^{(h)} = (1 - \frac{|\tau|}{h})^+$, its Fourier transform, $\widehat{g}_{i\omega}^{(h)} = h\text{sinc}^2(\frac{\omega h}{2\pi}) = (4/\omega^2 h)\sin^2(\omega h/2)$, cancels for all ω of the form $\frac{2n\pi}{h}$, $n \in \mathbb{Z}$, $n \neq 0$. Actually, this is not a real problem as long as τ in the empirical estimation of $v_\tau^{(h)}$ is sampled using a sampling period Δ strictly greater than $h/2$, and choosing Δ small enough to consider that the support of $\widehat{\Psi}_{i\omega}$, $[-a, a]$ is included in $[-\pi/\Delta, \pi/\Delta]$.

Indeed, let $h < 2\Delta$ and define, to ease the notations, $\widehat{A}_z = (\mathbb{I} + \widehat{\Psi}_z^*)\mathbf{\Sigma}(\mathbb{I} + \widehat{\Psi}_z^*)^\dagger$ and let $u_\tau^{(h)} = \sum_{k=-\infty}^{\infty} v_\tau^{(h)} \delta_{\tau - n\Delta}$ be the uniform sampling of $v_\tau^{(h)}$ and $\widehat{u}_{i\omega}^{(h)} = \frac{1}{\Delta}\sum_{k=-\infty}^{\infty} \widehat{v}_{i(\omega - \frac{2k\pi}{\Delta})}^{(h)}$ its Fourier Transform. Then (4.19) becomes

$$\widehat{u}_{i\omega}^{(h)} - \frac{1}{\Delta}\sum_{k=-\infty}^{\infty} \widehat{g}_{i(\omega - \frac{2k\pi}{\Delta})}^{(h)} \mathbf{\Sigma} = \frac{1}{\Delta}\sum_{k=-\infty}^{\infty} \widehat{g}_{i(\omega - \frac{2k\pi}{\Delta})}^{(h)} (\widehat{A}_{i(\omega - \frac{2k\pi}{\Delta})} - \mathbf{\Sigma})$$

But the support of $\widehat{A}_{i\omega} - \mathbf{\Sigma}$ is also $[-a, a] \subset [-\pi/\Delta, \pi/\Delta]$. Then for all $\omega \in [-\pi/\Delta, \pi/\Delta]$ we have

$$\widehat{u}_{i\omega}^{(h)} - \frac{1}{\Delta}\sum_{k=-\infty}^{\infty} \widehat{g}_{i(\omega - \frac{2k\pi}{\Delta})}^{(h)} \mathbf{\Sigma} = \frac{1}{\Delta}\widehat{g}_{i\omega}^{(h)} (\widehat{A}_{i\omega} - \mathbf{\Sigma})$$

which gives, for all $\omega \in [-\pi/\Delta, \pi/\Delta]$

$$\widehat{A}_{i\omega} = \Delta \frac{\widehat{u}_{i\omega}^{(h)}}{\widehat{g}_{i\omega}^{(h)}} + \mathbf{\Sigma}\left(1 - \frac{\sum_{k=-\infty}^{\infty} \widehat{g}_{i(\omega - \frac{2k\pi}{\Delta})}^{(h)}}{\widehat{g}_{i\omega}^{(h)}}\right) \tag{4.20}$$

Note that h must not take a value too close to 2Δ in order to avoid the division

by values that are too close to 0. If we choose for instance $h = \Delta$, then for all $\omega \in [-\pi/\Delta, \pi/\Delta]$, $\frac{1}{\Delta}\widehat{g}_{i\omega}^{(\Delta)} > \frac{4}{\pi^2} \approx 0.4$, and $\sum_{k=-\infty}^{\infty} \widehat{g}_{i(\omega - \frac{2k\pi}{\Delta})}^{(\Delta)} = \Delta$. Hence (4.20) becomes

$$\widehat{A}_{i\omega} = \Delta \frac{\widehat{u}_{i\omega}^{(\Delta)}}{\widehat{g}_{i\omega}^{(\Delta)}} + \mathbf{\Sigma}(1 - \frac{\Delta}{\widehat{g}_{i\omega}^{(\Delta)}})$$

Now if we choose Δ such that $a \ll \frac{\pi}{\Delta}$, (i.e. the support of $\widehat{A}_{i\omega} - \mathbf{\Sigma}$ is much smaller than $[-\pi/\Delta, \pi/\Delta]$) then for all $\omega \in [-a, a]$ we have $\widehat{g}_{i\omega}^{(h)} \approx \Delta$, and $\widehat{A}_{i\omega} \approx \widehat{u}_{i\omega}^{(\Delta)}$. But for $\omega \notin [-a, a]$, $\widehat{A}_{i\omega} = \widehat{u}_{i\omega}^{(\Delta)} = \mathbf{\Sigma}$, so for all $\omega \in [-\pi/\Delta, \pi/\Delta]$

$$\widehat{A}_{i\omega} \approx \widehat{u}_{i\omega}^{(\Delta)}$$

Consequently dividing on both handsides Eq. (4.19) by $\widehat{g}_z^{(h)}$ is not, from a practical estimation point of view, a real problem. So, in the following we will write

$$(\mathbb{I} + \widehat{\Psi}_z^*)\mathbf{\Sigma}(\mathbb{I} + \widehat{\Psi}_z^*)^\dagger = \widehat{v}_z^{(h)}/\widehat{g}_z^{(h)}, \tag{4.21}$$

without bothering with eventual cancelations of $\widehat{g}_z^{(h)}$.

Computing the Square Root of $(\mathbb{I} + \widehat{\Psi}_z^*)\mathbf{\Sigma}(\mathbb{I} + \widehat{\Psi}_z^*)^\dagger$

In the estimation process, once we have estimated $v_\tau^{(h)}$ and consequently (using Discrete Fourier transform) $\widehat{v}_z^{(h)}$ (for $z = i\omega$), using this last equation, we can estimate $(\mathbb{I} + \widehat{\Psi}_z^*)\mathbf{\Sigma}(\mathbb{I} + \widehat{\Psi}_z^*)^\dagger$. We need to go from there to the estimation of $\widehat{\Psi}_z$ (then using Eq. (4.18), we get $\widehat{\phi}_z$ and then by inverse Fourier transform ϕ_t). This problem requires therefore to take the square root of the left hand side of Eq. (4.21). In dimension $n = 1$, it means being able to go from $|1 + \widehat{\Psi}_z|^2$ to $\widehat{\Psi}_z$. There is clearly a phase determination problem. We will see that this phase is uniquely determined by the hypothesis (H1) and (H2). However, in dimension $n > 1$, they are many possible solutions and determining the correct one is not necessarily possible in general. We need to make a strong additional hypothesis.

4.4.2 Further Hypothesis on the Kernel ϕ_t

In the following, we suppose that:

(H3) $\mathbf{\Sigma} = \Lambda\mathbb{I}$, $\Lambda \in \mathbb{R}^+$, and $\widehat{\phi}_z$ can be diagonalized into a matrix D with some

constant unitary matrix U (that does not depend on z):

$$\widehat{\phi}_z = U^\dagger D_z U. \tag{4.22}$$

Clearly (H3) is a strong hypothesis. If we don't consider the special non-symmetric cases when (H3) is verified, this hypothesis is obviously satisfied when all the components of the process are identically distributed, i.e. when the process is invariant under arbitrary permutations.

4.4.3 The Estimator

Using Eq. (4.17) along with (H3), one gets that $\mathbb{I} + \widehat{\Psi}_z^*$ is diagonalizable in the same basis U and that

$$\mathbb{I} + \widehat{\Psi}_z^* = U^\dagger (\mathbb{I} - D_z^*)^{-1} U, \tag{4.23}$$

and thus from (4.21) , one gets

$$E_z = U \widehat{v}_z^{(h)} U^\dagger \mathbf{\Sigma}^{-1} / \widehat{g}_z^{(h)}, \tag{4.24}$$

where E_z is the diagonal matrix with the real positive coefficients :

$$E_z^{kk} = |1 - D_z^{kk}|^{-2}. \tag{4.25}$$

So the estimation problem reduces to being able to recover the coefficients D_z^{kk} from the coefficients $E_z^{kk} = |1 - D_z^{kk}|^{-2}$. This problem is solved by the following Lemma.

Lemma 4.4.1. *Let k be fixed. Let $z = i\omega$ (with $\omega \in \mathbb{R}$). Then*

$$(1 - D_{i\omega}^{kk})^{-1} = e^{\frac{1}{2}\log|E_{i\omega}^{kk}| - iH(\frac{1}{2}\log|E_{i\omega}^{kk}|)}, \tag{4.26}$$

where the operator $H(.)$ refers to the Hilbert transform.

The proof is based on the following theorem:

Theorem 4.4.2 (Paley-Wiener [71]). *Let's suppose that we observe the amplitude* $|\widehat{f}_{i\omega}|$ *of the Fourier transform of a real filter* f_t*. If*

$$\int_{\mathbb{R}} \frac{\log(|\widehat{f}_{i\omega}|)}{1+\omega^2} d\omega \quad < \quad \infty, \tag{4.27}$$

then the filter g_t *defined by its Fourier transform*

$$\widehat{g}_{i\omega} = e^{\log(|\widehat{f}_{i\omega}|) - iH(\log(|\widehat{f}_{i\omega}|))}, \tag{4.28}$$

is the only causal filter (i.e., supported by $\mathbb{R}^+$*), which is a phase minimal filter*[1] *and which satisfies* $|\widehat{g}_{i\omega}| = |\widehat{f}_{i\omega}|$*.*

Proof of Lemma 4.4.1. It is a simple application of this theorem with $|\widehat{f}_{i\omega}| = \sqrt{E^{kk}_z} = |1 - D^{kk}_{i\omega}|^{-1}$. Indeed, let us first show that $\widehat{g}_z = (1 - D^{kk}_z)^{-1}$ is a minimal phase filter, i.e., that both the poles and zeros are such that $\Re(z) < 0$.

- Let z be a zero of $(1 - D^{kk}_z)^{-1}$, then it is a pole of D^{kk}_z and consequently of $\widehat{\phi}_z$. However, from (H1) and (H2) one concludes that $|\widehat{\phi}_z| \leq \int_0^{+\infty} |e^{-zt}| \phi_t dt$ cannot be infinite, unless $\Re(z) < 0$.
- Let z be a pole of $(1 - D^{kk}_z)^{-1}$. It thus satisfies $D^{kk}_z = 1$. Thus 1 is an eigenvalue of $\widehat{\phi}_0$ which is in contradiction with (H2). Thus $(1 - D^{kk}_z)^{-1}$ has no pole.

Consequently $\widehat{g}_z = (1 - D^{kk}_z)^{-1}$ is a phase minimal filter. Moreover, since every coefficients of ϕ_t are positive and in L^1, the Fourier transforms $D^{kk}_{i\omega}$, for any k, are continuous functions of ω and goes to 0 at infinity. Along with the fact that $(1 - D^{kk}_z)^{-1}$ has no pole and has its zeros on the half-plane $\Re(z) < 0$ we easily conclude that $|\widehat{f}_{i\omega}| = |1 - D^{kk}_z|^{-1}$ satisfies (4.28). The theorem above can be applied and the Lemma follows. □

Main steps for kernel estimation.

The different steps for the final kernel estimator, in the fully symmetric case, can be summarized as follows

- Set Δ,
- Estimate the unconditional intensity Λ

[1] a minimal phase filter is a filter whose all the zeros and the poles of its Laplace transform satisfy $\Re(z) < 0$

- Estimate the auto-covariance operator $v_t^{(\Delta)}$ and compute its Fourier transform $\widehat{v}_{i\omega}^{(h)}$
- Compute $(\mathbb{I} + \widehat{\Psi}_{i\omega}^*)(\mathbb{I} + \widehat{\Psi}_{i\omega}^*)^\dagger$ using Eq. (4.21). Diagonalize it and compute the matrix $E_{i\omega}$ defined by Eq. (4.24)
- Compute the diagonal matrix $D_{i\omega}$ using (4.26)
- Go back to the initial basis and Inverse Fourier transform to get the estimation of ϕ_t

4.4.4 Some Particular Cases

The One Dimensional Case $n = 1$

As we already pointed out, the hypothesis (H3) is always true in this case since all the functions are scalar functions (so no diagonalization is needed). So in this case the phase determination problem is solved without adding any assumption (except (H1) and (H2) of course). The kernel estimator simply consists in first computing

$$|1 + \widehat{\Psi}_{i\omega}|^2 = \frac{\widehat{v}_{i\omega}^{(h)}}{\Lambda \widehat{g}_{i\omega}^{(h)}}, \tag{4.29}$$

and then inverting the Fourier transform of

$$\widehat{\phi}_{i\omega} = 1 - e^{-\log|1+\widehat{\Psi}_{i\omega}| + iH(\log|1+\widehat{\Psi}_{i\omega}|)}. \tag{4.30}$$

The 2-dimensional Case $n = 2$

In the two-dimensional case ($n = 2$), the hypothesis (H3) is satisfied in the particular case the kernel ϕ_t is *bisymmetric*, i.e., has the form

$$\phi_t = \begin{pmatrix} \phi_t^{11} & \phi_t^{12} \\ \phi_t^{12} & \phi_t^{11} \end{pmatrix} \tag{4.31}$$

The matrix of the Laplace transform $\widehat{\phi}_z$ can be indeed decomposed as follows:

$$\widehat{\phi}_z = U^\dagger \begin{pmatrix} -\widehat{\phi}_z^{11} + \widehat{\phi}_z^{12} & 0 \\ 0 & \widehat{\phi}_z^{11} + \widehat{\phi}_z^{12} \end{pmatrix} U$$

where $U = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}$ is constant and is not a function of ω.

Diagonalizing and identifying the diagonal coefficients in both hand sides of (4.24) leads to

$$\frac{1}{\Lambda_1 \widehat{g}_{i\omega}^{(h)}}(\widehat{v}_{i\omega}^{(h)11} + \widehat{v}_{i\omega}^{(h)12}) = |1 + \widehat{\Psi}_{i\omega}^{11} + \widehat{\Psi}_{i\omega}^{12}|^2 \tag{4.32}$$

$$\frac{1}{\Lambda_1 \widehat{g}_{i\omega}^{(h)}}(\widehat{v}_{i\omega}^{(h)11} - \widehat{v}_{i\omega}^{(h)12}) = |1 + \widehat{\Psi}_{i\omega}^{11} - \widehat{\Psi}_{i\omega}^{12}|^2 \tag{4.33}$$

Applying the same method used in the case $n = 1$ to equations (4.32) and (4.33) we get an estimate of $\widehat{\Psi}$. Finally we apply

$$\widehat{\phi} = (\mathbb{I} + \widehat{\Psi})^{-1}\widehat{\Psi} \tag{4.34}$$

giving $\widehat{\phi}$. Applying the inverse transform to $\widehat{\phi}$ we finally get ϕ.

In the following section we illustrate these results and our methods on numerical simulations of 1D and 2D Hawkes processes.

4.5 Numerical Illustrations

Let us discuss some examples illustrating the estimation method as defined previously using simulated Hawkes processes. All the simulations have been performed with the thinning algorithm described in appendix 4.A.

We suppose that we have estimated the ACF from the observation of a Hawkes process $(N_t)_{t\geq 0}$ (1D or 2D), with an unknown kernel ϕ. We then strictly follow the method described in previous section where the Fourier and Hilbert transforms are replaced by their discrete versions. $v_\tau^{(h)}$ is sampled at rate Δ and, for a given fixed value of the scale h, is considered as a function of the lag $\tau = n\Delta$. As discussed previously (section 4.4.1), in order to avoid problems related to the zeros of $\widehat{g}^{(h)}$ we naturally choose $h = \Delta$. From a practical point of view, Δ has to be chosen small enough in order to work with a large sample size N but also to finely cover the support of the kernel functions ϕ, and to avoid Fourier aliasing effects. However, Δ has also to be chosen large enough in order to allow a reliable estimation of the empirical covariance that is not perturbed by the "Epps noise". Indeed, for very small lags, it is well known that the variations of point processes become

uncorrelated when studying the correlation between the returns of two assets. This is known as the Epps effect in the Finance literature (see, e.g., [9]). In fact, as illustrated below, there is an optimal choice of the sampling rate Δ. However, in order to set an order of magnitude, one can choose $\Delta \simeq 1/(10\Lambda)$ as a starting value. We denote by T the overall sample size of the process and $N = T\Delta^{-1}$ the size of the sample used to estimate to covariance function $v_{k\Delta}^{(\Delta)}$. We suppose that this latter is estimated over the interval $[-\tau_{\max}, \tau_{\max}]$ with $\tau_{\max}$ large enough so that $v_\tau^{(h)}$ is negligible when $|\tau| \geq \tau_{\max}$.

4.5.1 The Case $n = 1$

We test the estimation with two types of kernels, exponential (4.35) and power-law (4.36).

Exponential kernel

We first simulate a one dimensional Hawkes process with the exponential kernel:

$$\phi_t \quad = \quad \alpha e^{-\beta t} \mathbb{1}_{t \geq 0} \tag{4.35}$$

where we choose ($\mu = 1$, $\alpha = 1$, $\beta = 4$). This gives $\widehat{\phi}_0 = 1/4$, $\Lambda = 4/3$. Each simulated sample contains about 130000 jumps ($T \simeq 10^5$ secs). In Figure 4.1 we have reported an example of the estimated kernel function ϕ_t for a single sample. We see that ϕ that we estimated is indeed causal, and up to some noise, is very close to the real kernel (reported in a red solid line). Smoothing techniques could be applied to the causal part of ϕ that might effectively reduce the error, for example a moving average of ϕ or a low pass filter. However we have not studied the benefits (with regards to the series of error data) of such options and do not further explore them in the current presentation.

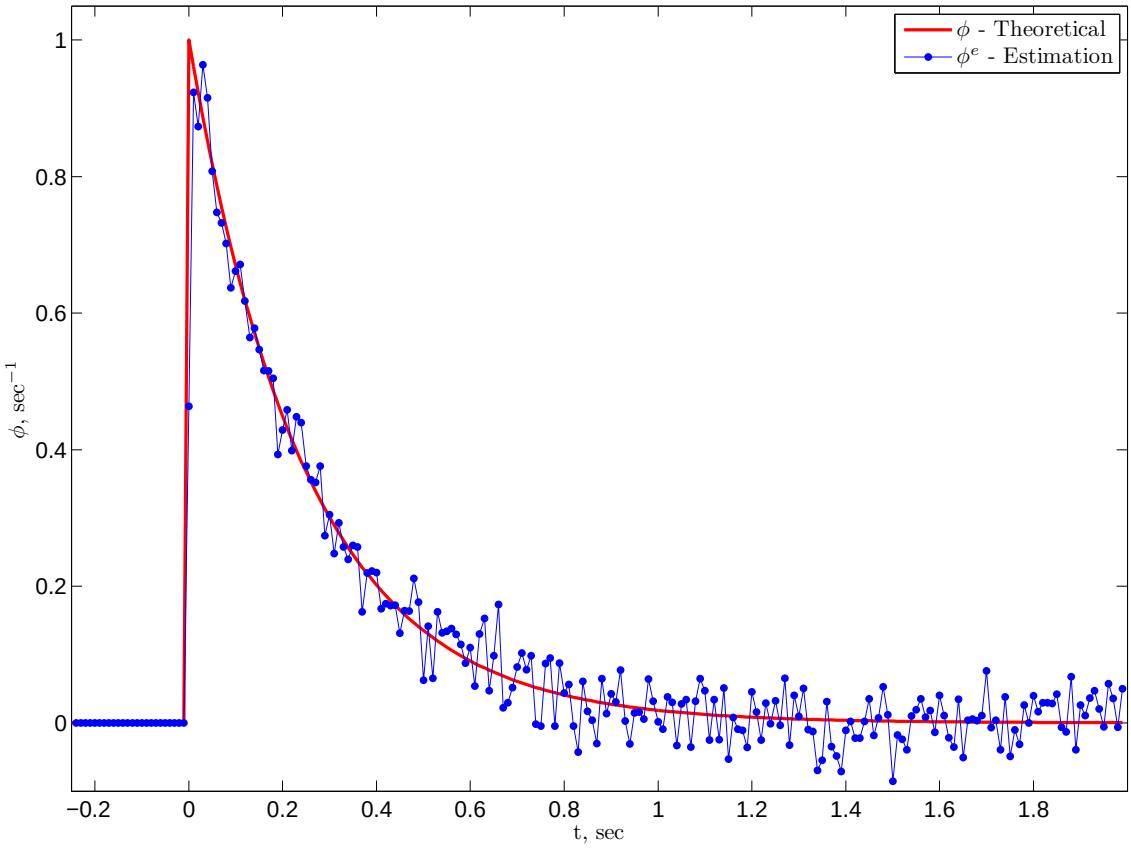


Figure 4.1: Non parametric estimation of the one dimensional Hawkes exponential kernel ϕ (4.35) with $\alpha = 1$, $\beta = 4$. We used $\Delta = 0.01$ and $\tau_{\max} = 2$.

Power Law Decay

We now consider slowly decaying kernels such as the power-law function defined as:

$$\phi_t \quad = \quad \alpha(t+\gamma)^{\beta})\mathbb{1}_{t\geq 0} \tag{4.36}$$

with $\beta < -1$. In this case we have $\widehat{\phi}_0 = -\frac{\alpha}{\beta+1}\gamma^{\beta+1}$. We choose $\alpha = 32$ and $\beta = -5$ and $\gamma = 2$ making $\widehat{\phi}_0 = 0.5 < 1$ and $\Lambda = 2$ and again With 130000 jumps ($T \simeq 65000$ secs). However, for numerical computation purposes, we used the following similar kernel with a bounded support:

$$\phi'_t \quad = \quad (\phi_t - \phi_{10})^+$$

where $x^+ = \max(x, 0)$. This would reduce the complexity of the simulation by reducing the number of calculations made when updating the intensity of Hawkes

process (cf 4.A). The kernel estimated on a single sample is reported in Figure 4.2. Once again, one can see that, up to an additive noise, the estimated kernel fits well the real one.

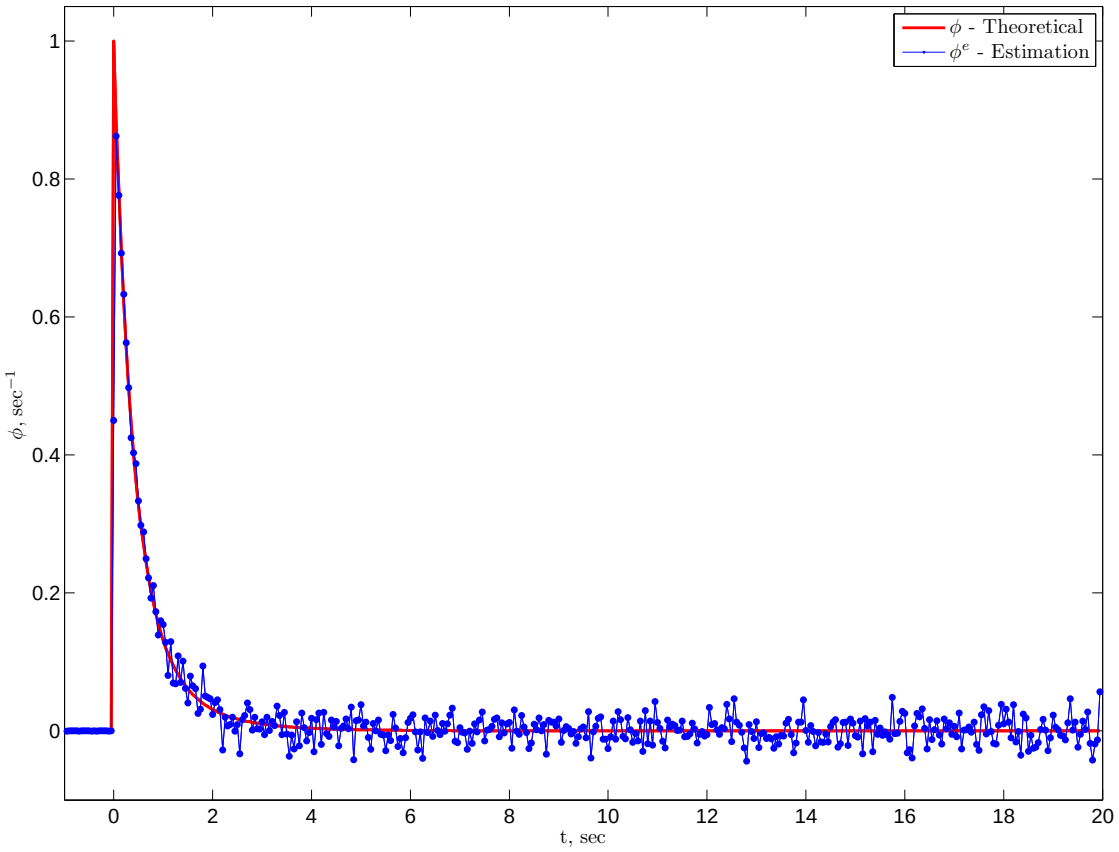


Figure 4.2: Non parametric estimation of the one dimensional Hawkes exponential kernel ϕ (4.36) with $\alpha = 32$, $\beta = -5$ and $\gamma = 2$. We used $\Delta = 0.05$ and $\tau_{\max} = 20$.

Error Analysis

Let us briefly discuss some issues related to the errors associated with our kernel estimates. In ref. [10], a central limit theorem as been proved that shows that, asymptotically, the errors of empirical covariance function estimates are normally distributed with a variance that decreases as T^{-1} (or N^{-1} for Δ fixed). One thus expects the same kind of results in the estimates of the components of ϕ.

Let us define the L^2 estimation error as:

$$e^2 = \sum_{k=1}^{\frac{\tau_{\max}}{\Delta}} |\phi_{k\Delta} - \phi^e_{k\Delta}|^2 \tag{4.37}$$

where ϕ^e is the estimated kernel. In Figure 4.3, we have reported e^2 as a function of the sample length T for an example of exponential kernel ϕ_t for $\Delta = 0.01$ fixed. As expected, one observes a behavior very close to T^{-1}.

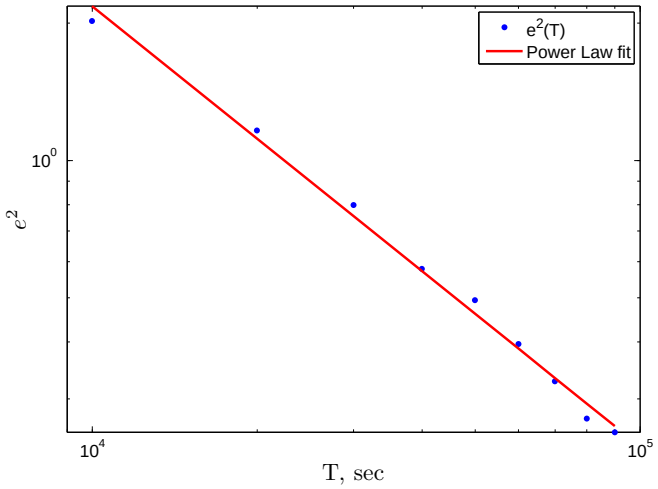


Figure 4.3: Mean square error as a function of the sample length T in log-log coordinates for 1D Hawkes processes exponential kernel ϕ (4.36) ($\alpha = 1$, $\beta = 4$). We used $\Delta = 0.01$ and $\tau_{\max} = 2$

Let us now fix $T = 10^5$ and look at the empirical squared error between the real ϕ and the estimated one for $t > 0$. We find that the series of errors are centered gaussian and mainly uncorrelated. The normality of the observed errors is illustrated in Figure 4.4 where we report, for 4 different values of t, the normal probability plots. We point out that we observe that the variance of the law slightly increases when t goes to zero (while remaining centered Gaussian) and so we have standardized the error series before plotting them in the normal probability plot of Figure 4.4. Note that we do notice a slight deviation from the Gaussian behavior in the left tail of the errors, however standard tests that we performed did not reject the null hypothesis of a Gaussian distribution.

As we discussed above, the estimation error depends, a priori, on the sampling rate Δ and should be minimal in a range between large Δ values corresponding to small N and small Δ values where the "Epps noise" becomes important. In Figure 4.5, we have plotted the measured error e^2 for different values of Δ in the exponential case with $\Delta \in [0.02\, 0.5]$, and $\tau_{\max} = 6$. We clearly observe, a minimum value around $\Delta^* = 0.15$. Notice however that, despite significative increase of the errors around the optimal value of Δ, we have observed, in both exponential and power-law cases, that the estimates of the kernel parameters remain reliable in a wide range of Δ values. Furthermore, seeing that the decay kernels go to 0 rather

quickly one may want to use a sampling rate $\Delta < \Delta^*$ in order to get a better idea of the form of the decay kernel in the neighborhood of 0.

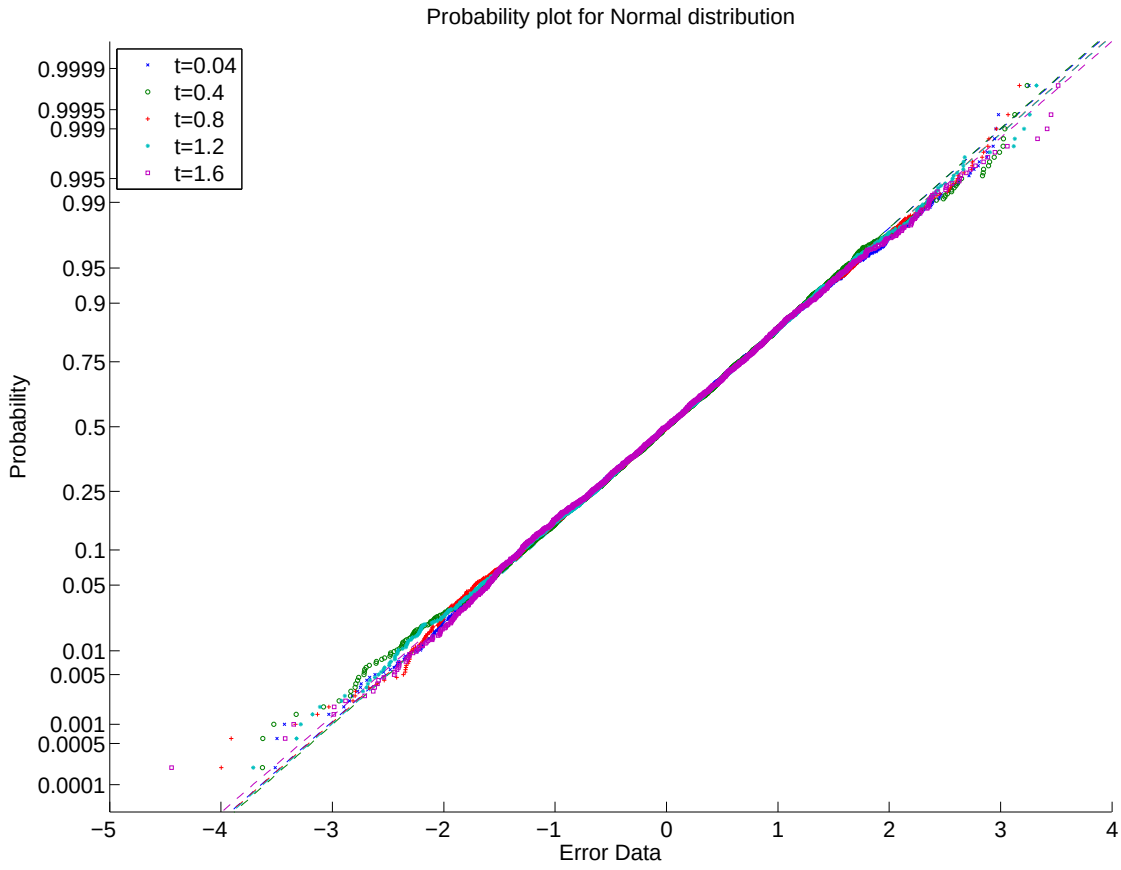


Figure 4.4: Normal probability plots obtained from the empirical error distributions at 4 different times t in the case of a 1D Hawkes processes exponential kernel ϕ ($\alpha = 1$, $\beta = 4$). We used $\Delta = 0.01$ and $\tau_{\max} = 2$. The x-axis represents the data value for each point in the data. The y-axis contains the empirical probabilities. The scale of the y-axis is not linear but proportional to the quantiles of a standard Gaussian. The dashed lines connect the 25^th and 75^th percentiles in the data. As we can see, the data points fall near their corresponding lines, indicating that the error data are reasonably Gaussian.

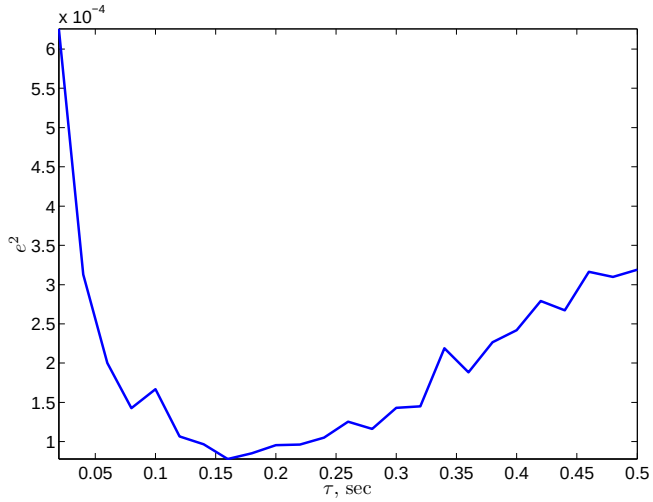


Figure 4.5: e^2 error as function of the sampling rate Δ a in the case of a 1D Hawkes processes exponential kernel ϕ ($\alpha = 1$, $\beta = 4$). We used $\tau_{\max} = 6$.

4.5.2 The Case $n = 2$

Let us now consider a 2 dimensional example of a Hawkes process with a bisymmetric kernel matrix such that:

$$\phi_t^{11} = \phi_t^{22} = \phi_{d,t} = \alpha_d e^{-\beta_d t} \mathbb{1}_{t \geq 0} \tag{4.38}$$

$$\phi_t^{12} = \phi_t^{21} = \phi_{a,t} = \alpha_a e^{-\beta_a t} \mathbb{1}_{t \geq 0} \tag{4.39}$$

where ϕ_d and ϕ_a are respectively the diagonal and anti-diagonal terms of ϕ. We use the following parameters for the simulations, $\alpha_d = 0.5$, $\beta_d = 8$, $\alpha_a = 1$, $\beta_a = 4$, $\mu = (1, 1)$. We have $T = 40.10^4$ and we simulate about 60000 jumps for each of the two components the Hawkes process. The Figure 4.6 below shows the theoretical and estimated versions of ϕ_d and ϕ_a. We see that, as in the 1D case, we get a reliable estimate of both kernels ϕ_d and ϕ_a.

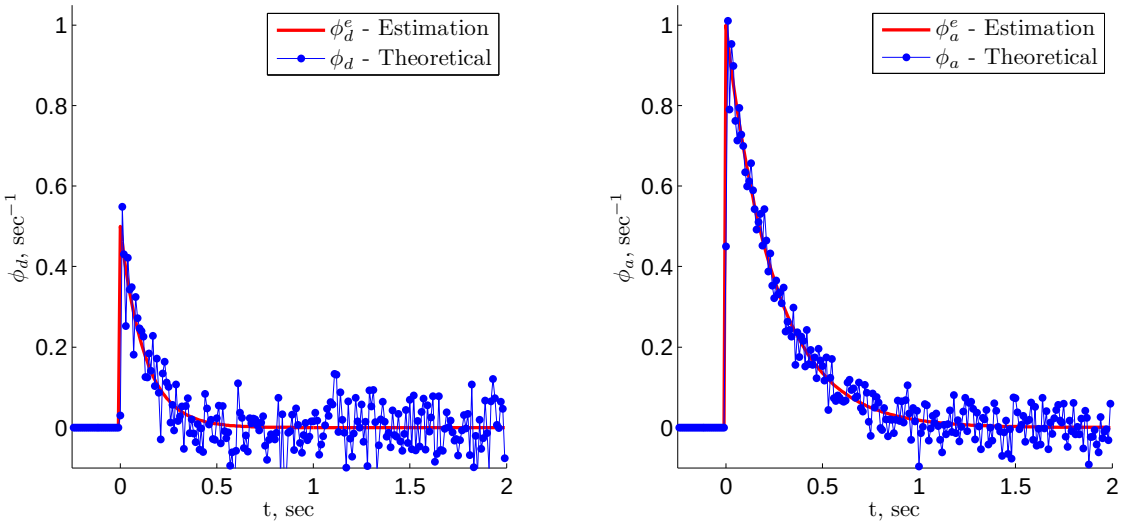


Figure 4.6: Non parametric estimation of the two dimensional Hawkes bisymmetric exponential kernel ϕ ((4.38)) with $\alpha_d = 0.5$, $\beta_d = 8$, $\alpha_a = 1$, $\beta_a = 4$, $\mu = (1, 1)$. We used $\Delta = 0.05$ and $\tau_{\max} = 2$

4.6 Application to High Frequency Market Data

As mentioned in the introduction, Hawkes point processes have found many applications, notably as models for high frequency financial data. Indeed, because of the discrete and correlated nature of trade and limit order arrival times, point processes are naturally used as statistical models of market dynamics at the microstructure level. One can mention for instance [48] where buy and sell trades arrivals are represented by a bivariate Hawkes process with exponential kernels (see also [52]). Another approach developed in [9] models high frequency price dynamics as the difference between two coupled Hawkes processes representing respectively up and down discrete price variations. The authors emphasized that such model allows one to account for the main stylized facts characterizing the observed noise microstructure (Signature plot behavior and Epps Effect).

We use **level 1** data provided by QuantHouse Trading Solutions[2] of 3 futures contracts. The data has millisecond accuracy and it has been treated in such a way that each market order is equivalent to exactly one trade[3]. In particular we mainly Futures data the 10-years Euro-Bund (Bund), and to a lesser extent for the DAX index (DAX) and the SP500 index (SP). For more information about the data

[2] http://www.quanthouse.com/

[3] When one market order hits several limit orders it results in several trades being reported (cf. section 1.2.2). We aggregate together all such transactions and consider them as one trade with an equivalent size and price.

please refer to the introduction of the thesis (section 1.5). In this paper, we use 75 days covering the period between 2009-06-01 and 2009-09-15. We also restricted our intraday dataset to the time between 9 AM and 11 AM where we think that the rate of incoming market orders is stationary and ignored the days with too little trades.

4.6.1 1D Model

In the following we test our method on the process of incoming trade times (market orders) for the Bund Future (FGBL). We then calculate the empirical ACFs for each of the days and average them out over the 75 days. We apply the estimation method to this averaged ACF.

The results are shown in the Figure 4.7 below. We get a causal ϕ that fits a power law function of the form ax^b, with $a = 0.09863$, $b = -1.053$, with 95% confidence bounds for b equal to $(-1.064, -1.043)$. So $b \approx -1$, and we assume that $\phi_t = 0$ for t larger than $\tau_{\max} = 100$ (this is not a bad approximation, seeing that the ϕ_t looks like a white noise for large t), and that it is bounded for t close to 0. We test the estimator with crude parameters $\tau_{\max} = 100$, $\Delta = 0.1$ but we check that the results are robust for different values of the parameters (see Table 4.1).

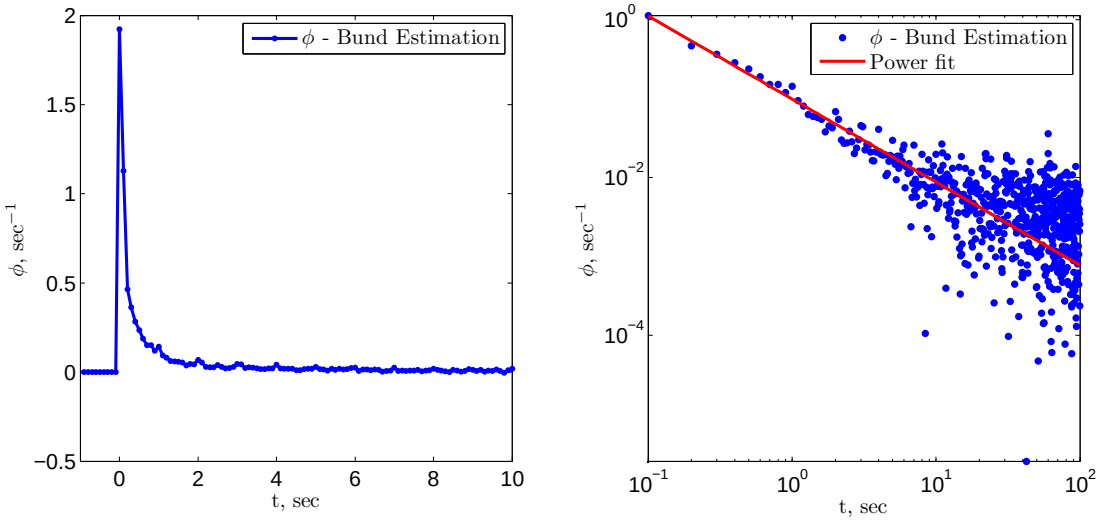


Figure 4.7: The plot on the left is the non parametric estimation of the Hawkes kernel assumed for the rate of incoming market orders of the Bund Futures. We used $\Delta = 0.1$ and $\tau_{\max} = 100$, but only show the first 10 seconds in the left plot. On the right is that very same plot in log-log scale and the corresponding power law fit ax^b. We find $a = 0.09863$, $b = -1.053$.

Furthermore, we provide in table 4.1 below results of the power law fits for $\tau_{\max} \in \{5, 10, 50, 100\}$ and $\Delta \in \{0.01, 0.05, 0.1, 0.5\}$. We see that the fit remains close to the same values ($a = 0.1$, $b = -1$) for different $\tau_{\max}$ and Δ. In particular, we see that the fit is very robust when we change $\tau_{\max}$ for a Δ fixed. Indeed, increasing $\tau_{\max}$ beyond a certain value is superfluous. This is not surprising, seeing that the estimated decay kernel is a power law that goes to 0 at ∞, so increasing $\tau_{\max}$ just produces noise in the tail. We notice also that the values of a and b stabilize as Δ becomes small. In our case, table 4.1 shows that fit becomes stable around ($a = 0.1$, $b = -1$), for $\Delta \leq 0.1$.

We conclude that our estimator is fairly robust with respect to the sampling scheme, Δ, and the range of τ.

$\tau_{\max}$	Δ	a	b
5	1	0.158909	-1.46839
5	0.5	0.12547	-1.27377
5	0.1	0.109257	-1.0123
5	0.05	0.102258	-1.00491
5	0.01	0.095562	-0.98297
10	1	0.152666	-1.40213
10	0.5	0.122812	-1.27191
10	0.1	0.104953	-1.02776
10	0.05	0.098544	-1.01665
10	0.01	0.093251	-0.98835
50	1	0.14688	-1.40274
50	0.5	0.118427	-1.29464
50	0.1	0.099606	-1.0491
50	0.05	0.093841	-1.03284
50	0.01	0.090346	-0.99546
100	1	0.146276	-1.41515
100	0.5	0.117503	-1.30292
100	0.1	0.098624	-1.05329
100	0.05	0.092975	-1.03596
100	0.01	0.089227	-0.99899

Table 4.1: Results of the Power Law fit ax^b, applied to the estimated Hawkes kernel assumed for the rate of incoming market orders of the Bund Futures. We show the results for different input parameters to the empirical ACF, $\tau_{\max}$ and Δ.

4.6.2 2D model

Trading order model

In this section we will test the estimation method with two empirical jump process that we believe are very well described by a mutually exciting Hawkes process. Similarly to the one dimensional case we test our method on the process of incoming trade times (market orders) for the Bund Future (FGBL) over the same date range and using the same intraday range. The process of prices X_t is considered as:

$$X_t = N_t^+ - N_t^-$$

where N_t^+ and N_t^- are respectively the cumulative positive and negative jumps of X_t, and we look at the two dimensional process

$$N_t = \begin{pmatrix} N_t^+ \\ N_t^- \end{pmatrix}$$

We choose in the 2-dimensional case to use the series of midpoint prices. This choice is justified by the fact that for this type of series, the matrix of ACF (shown in Figure 4.8) is indeed bisymmetric and that Λ_+ and Λ_- the respective average rates of N^+ and N^- (daily values shown in Figure 4.9) are similar. In particular, Figure 4.8 shows that hypothesis (H3), i.e. the matrix $v_\tau^{(h)}$ is bisymmetric, holds. Indeed we see that $v_\tau^{(h)11}$ (blue curve, in the plot on the left) is equal to $v_\tau^{(h)22}$ (red curve on the left) for all $t > 0$, proving that $v_\tau^{(h)11} = v_\tau^{(h)22}$ for all t, and that $v_\tau^{(h)12} = v_\tau^{(h)21}$, for all $t > 0$ also proving that $v_\tau^{(h)12} = v_\tau^{(h)21}$ for all t. Furthermore, in Figure 4.9, we show for each day Λ_1 against Λ_2. The plot shows that the $\Lambda_1 \simeq \Lambda_2$ with a very small variation and therefore that we are within our assumption that $\mathbf{\Sigma} = \Lambda\mathbb{I}$.

We point out that the choice of the midpoint series is motivated by a two factors. First, if we choose the series of traded prices, we have an important bouncing artifact that is very hard to capture by our modeling approach and we get negative decay functions on the diagonal of ϕ. Second, if we choose the series of the last traded prices of buy orders only as in [9], the bouncing artifact disappears, however the symmetry of Λ_+ and Λ_- is no longer verified. The series of midpoint prices has the advantage of having a reduced bouncing effect and of presenting symmetrical N^+ and N^- processes. Notice that in the one dimensional case these considerations are irrelevant since assumption H(3) is always trivially verified.

Still, when actually going through the estimation process, we took $\Lambda = (\Lambda_+ +$

$\Lambda_-)/2$ and we symmetrize $v_\tau^{(h)}$ with respect to τ. This reduces the fourier noise and we have that the components of $v_\tau^{(h)}$ are even which means that their Fourier Transforms are real and similarly even (this is the case for a process which is invariant under permutation, cf condition (H3)). This manipulation is only just aesthetic, as it helps reduce the noise, and we have checked that our final results are practically unchanged without it.

After verifying that $\widehat{v}_\tau^{(h)11} + \widehat{v}_\tau^{(h)12} > 0$ and $\widehat{v}_\tau^{(h)11} - \widehat{v}_\tau^{(h)12} > 0$ we can then use equations (4.32)-(4.33) in order to estimate $\widehat{\Psi}$ followed by ϕ.

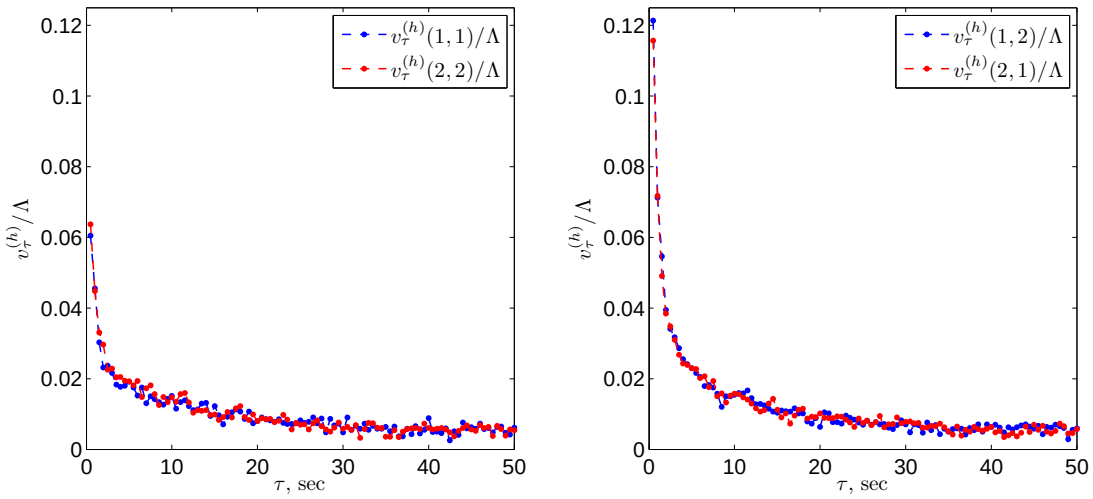


Figure 4.8: $v_\tau^{(h)}$, $\tau > 0$ of the ACF of N_t for the series of midpoints of the Bund Futures. We used $\tau_{\max} = 200$ but only show the first 50 seconds. In order to clearly see that we are in a bisymmetric world, we plotted $v_\tau^{(h)11}$ (blue curve, figure on the left) on top of $v_\tau^{(h)22}$ (red curve, figure on the left) for $h > 0$, and similarly for $v_\tau^{(h)12}$ and $v_\tau^{(h)21}$ (figure on the right). We see that the matrix is clearly bisymmetric.

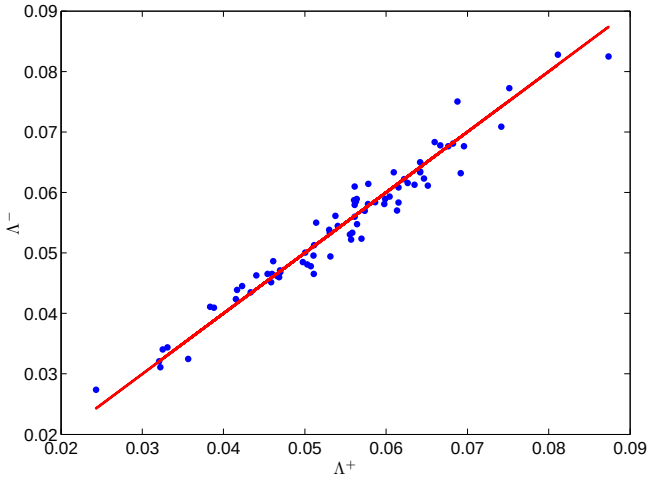


Figure 4.9: Λ_1 and Λ_2 for each of the 75 days we used (each blue dot represents one day). The red line is the $y = x$ line. We see that $\Lambda_1 \simeq \Lambda_2$.

The result of the non parametric estimation is shown in Figure 4.10. We immediately notice that ϕ^{11} is small in comparison to ϕ^{12} and can almost be considered as white noise (i.e. $\phi^{11} = 0$). This means that N^+ and N^- are not self exciting and exclusively mutually exciting. However, it is important to note that while ϕ^{11} does look flat, we notice that it becomes negative for small values of τ. This effect shows the limit of our model and we will elaborate on this point in the following paragraph.

Before doing that, we analyze the form of the ϕ^{12}. We found that ϕ^{12} fits a power law function of the form ax^b, with $a = 0.09531$, $b = -0.9931$, with 95% confidence bounds on b equal to $(-1.008, -0.9786)$. So $b \approx -1$, and we assume that $\phi_t = 0$ for t larger than $\tau_{\max} = 200$ (this is not a bad approximation, seeing that the ϕ_t looks like a white noise for large t), and that it is bounded for t close to 0. We show in Figure 4.11, the fit of ϕ^{12} to a power law in a log-log plot.

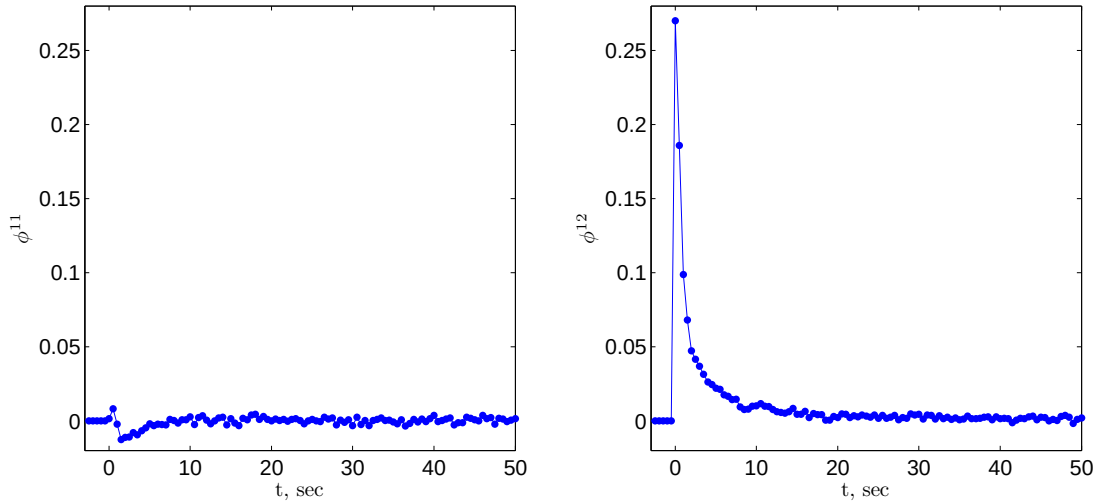


Figure 4.10: Non parametric estimation of the two dimensional Hawkes kernel assumed for the process N_t composed from the up and down jumps of the midpoint prices of the Bund Futures. We used $\tau_{\max} = 200$ but only show the first 50 seconds. We see the ϕ^{11} is negligible when compared to ϕ^{12} confirming that the N^+ and N^- are mutually but not self exciting.

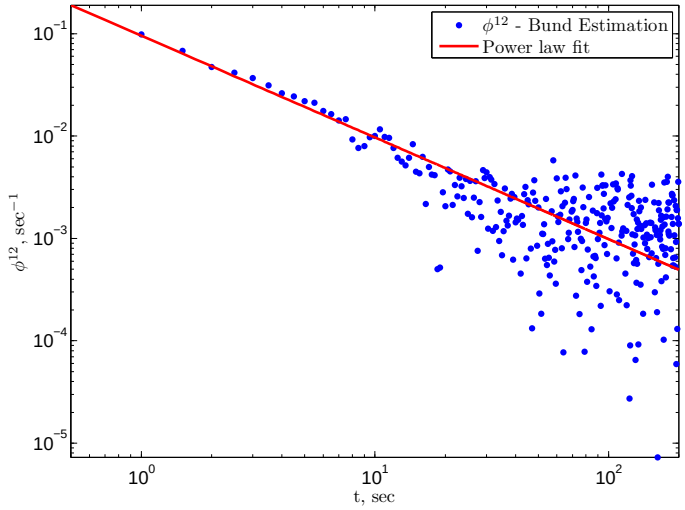


Figure 4.11: In log-log scale: Estimated ϕ^{12} (blue dots) and its corresponding power law fit (red line) ax^b. We find $a = 0.09531$, $b = -0.9931 \simeq -1$

Bouncing Effects and Inhibitory Decay Kernels

As we have seen above (Figure 4.10), ϕ^{11} can come out to be negative, which defies the assumption of our model. Indeed, if ϕ^{11} becomes negative, there is a priori no guarantees that λ_t remains positive, and a negative rate of arrival is unacceptable. As we have mentioned above, bouncing artifacts are difficult to capture with our modeling approach. The bouncing artifacts translates in practice to a strong

negative autocorrelation of the tick returns (this has been well documented in chapter 3 of this thesis). These effects are most apparent on the series of trade prices and while they are heavily dampened when we use the series of midpoint prices, they do not completely disappear. Moreover, this effect depends on the microstructure of the asset we consider, and more particularly on its **tick size**. In chapters 2 and 3, we have documented the microstructure property called the tick size. It is a measure of the traders' perceived importance of the tick value and describes the market's aversion to a movement in the price. In this paper we will limit the discussion and say that the tick size influences the strength of the autocorrelation of tick returns and increases the bouncing effects and the mean reversion of prices. This phenomenon cannot be captured by our model and we get negative decay kernels for the diagonal terms ϕ. However, when the tick is small, the effect of the tick size is minimal and the estimation we find falls within the model's assumption. We illustrate this by showing the estimation of the decay function for a large tick asset, the SP in Figure 4.12, and a small tick asset, the DAX in Figure 4.13. For the SP, we can see that ϕ^{11} is now clearly negative for small t, whereas the effect completely disappears in the DAX, where ϕ^{11} is of the same order of magnitude as ϕ^{12}, which means that the arrival of a price movement of a certain direction certainly increases the probability of arrival of a new movement, but has very little influence on its direction.

We conclude this discussion by saying that model (4.2) only assumes that λ_t needs to stay positive at all times. For that non-negative decay kernels are required. In practice however, we can conceive an a situation with a negative ϕ^{11}, provided that μ and and ϕ^{12} are chosen in such a way that λ_t has a vanishing probability of becoming negative over a large but finite period of time. This is often used to portray inhibitory effects heuristically even though the assumption of the model are not verified [51, 74]. This is exactly what we have in our situation: since $|\phi^{11}| \ll \phi^{12}$ the probability of having $\lambda_t < 0$ over the time we are conducting our experiment is infinitesimally small and the arrival of jump of N^+ (resp. N^-) **decreases** the probability of getting another jump of N^+ (resp. N^-) for a short duration afterwards, but highly increases the probability of getting a jump of N^- (resp. N^+). We also note that this effect becomes more pronounced when using the series of trade prices.

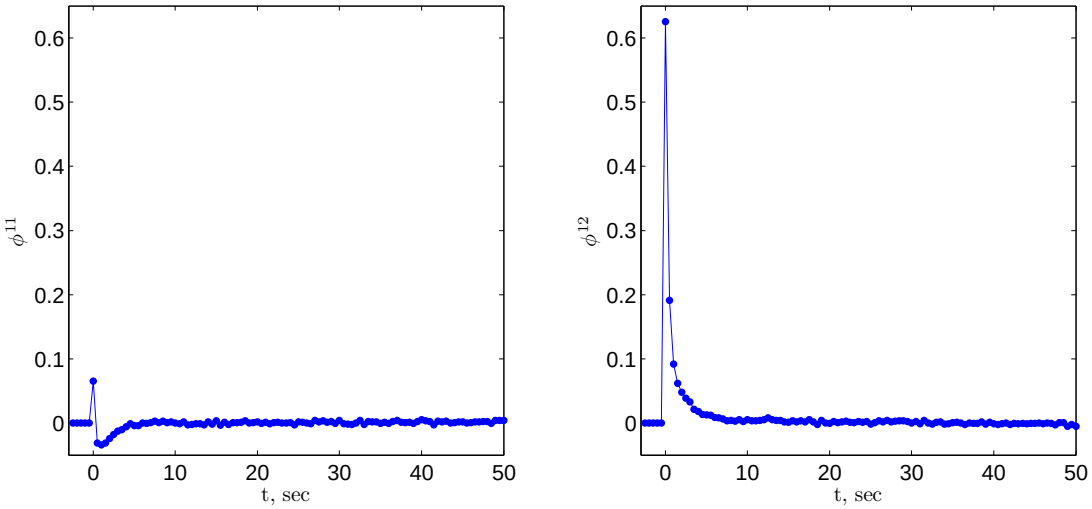


Figure 4.12: Non parametric estimation of the two dimensional Hawkes kernel assumed for the process $N_t = (N_t^+, N_t^+)$ composed from the up and down jumps of the midpoint prices of the SP Futures (a large tick asset). We used $\tau_{\max} = 200$ but only show the first 50 seconds. We see the ϕ^{11} is negative but small when compared to ϕ^{12} hinting that there are self inhibitory pressures acting on the processes N^+ and N^- .

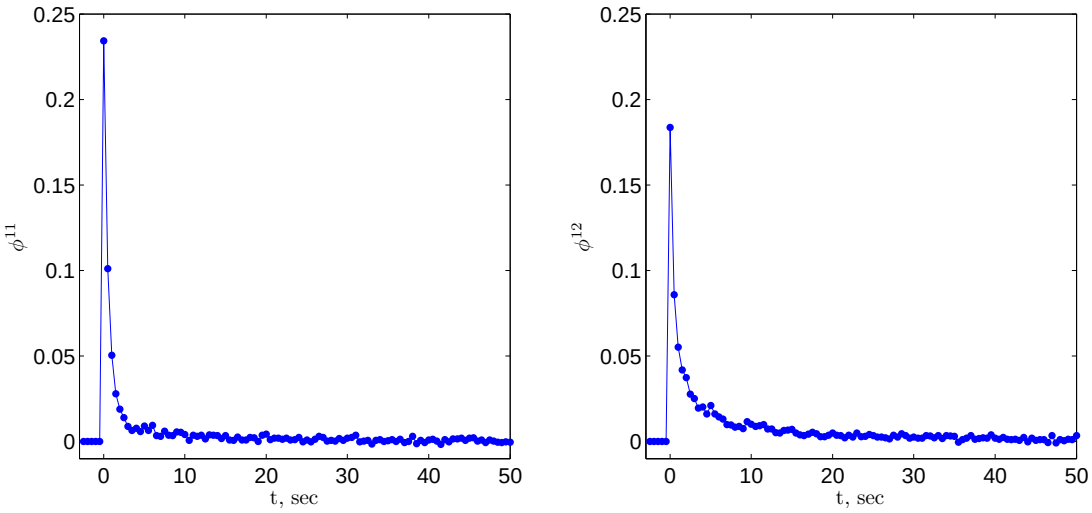


Figure 4.13: Non parametric estimation of the two dimensional Hawkes kernel assumed for the process $N_t = (N_t^+, N_t^+)$ composed from the up and down jumps of the midpoint prices of the DAX Futures (a small tick asset). We used $\tau_{\max} = 200$ but only show the first 50 seconds. We see the ϕ^{11} and ϕ^{22} are both positive and almost of the same order of magnitude. This means that the arrival of one jump of N^+ (resp. N^-) increases the similarly both probabilities of arrival of N^+ and N^-.

Appendices

4.A Example of a 2D Mutually Exciting Hawkes Process

4.A.1 A Mutually Exciting Process

In this section we turn our attention to the particular case where $n = 2$ that we use for simulation and validation of equation (4.15). We consider a symmetric, exclusively mutually exciting 2D Hawkes process with the following decay kernel:

$$\phi_t = \begin{pmatrix} 0 & \varphi_t \\ \varphi_t & 0 \end{pmatrix}$$

where

$$\varphi_t = \alpha e^{-\beta t} \mathbb{1}_{t \geq 0}$$

This type of kernel was chosen for ease of numerical considerations and also because we can easily see the mutually exciting behavior when we plot the intensity process (cf Figure 4.14).

4.A.2 Simulation

We use a thinning algorithm like the one described in [65]. The algorithm requires us to start by simulating a standard Poisson with a very high intensity Intensity M from which we remove jumps that do would not belong to the Hawkes process. We must guarantee that

$$\lambda_t^1 + \lambda_t^2 < M \text{ for all } t \in [0, T]$$

and whenever this condition is breached we stop the algorithm, increase M and start over. Assuming this condition satisfied, we apply a thinning procedure to each jump of the obtained process, rejecting it altogether with probability $1 - \frac{\lambda_t^1 + \lambda_t^2}{M}$ or marking it as a jump of N^i ($i \in \{1, 2\}$) with probability $\frac{\lambda_t^i}{M}$. Once this is done we recalculate λ_t^i, $i = 1, 2$ using Equation 4.2, when we observe the next jump and repeat the thinning procedure. Note that in the exponential case the complexity of simulating N observations is reduced to $\mathrm{O}(N)$ from $\mathrm{O}(N^2)$ for the general case.

An example of λ_t^1 and λ_t^2 over 5 seconds, with $\mu = (1, 1)$, $\alpha = 1$, $\beta = 4$ is shown in Figure 4.14. The unit of μ, α, β is seconds^{-1} because they are akin to intensities.

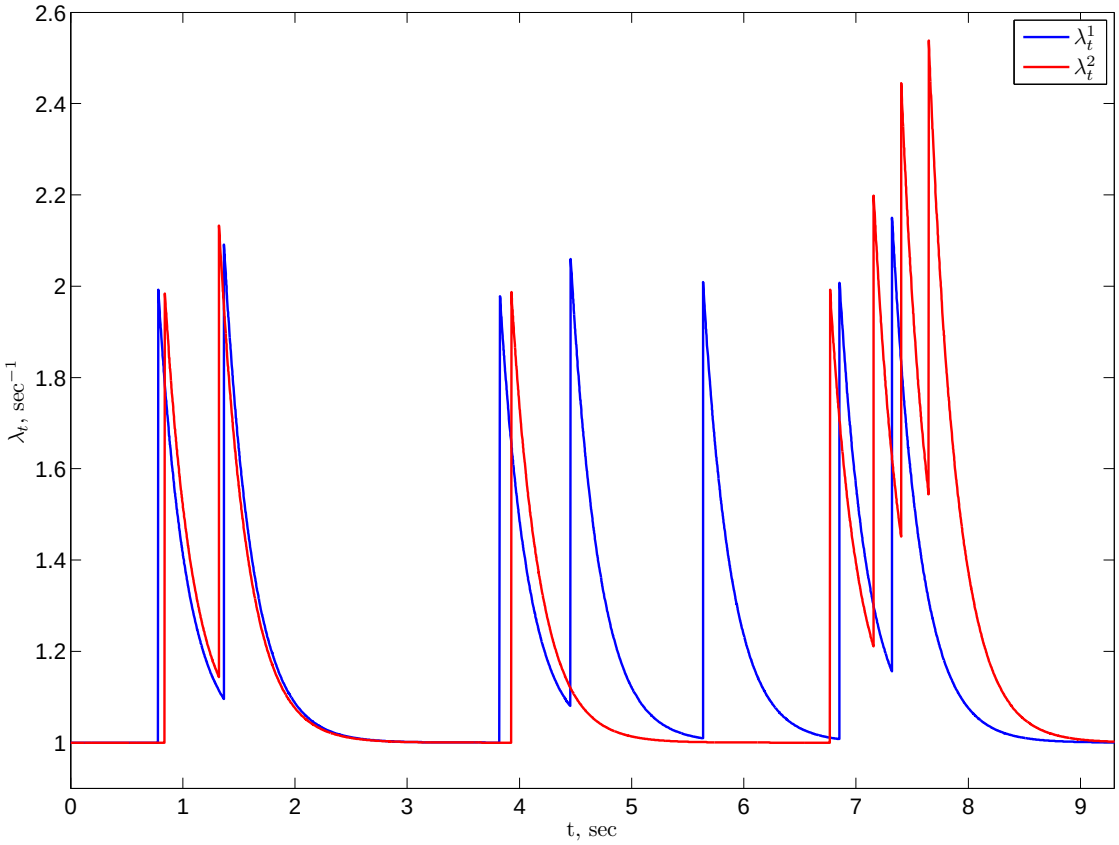


Figure 4.14: Realized Intensities λ_t^1 and λ_t^2 of a simulated $2D$ Hawkes with $\mu = (1,1)$, $\alpha = 1$, $\beta = 4$. We can clearly see in this case that the two process are mutually exciting because the arrival of one event causes the intensity of the other to increase, and increases its probability of a arrival.

4.A.3 Covariance Operator

The covariance operator defined in Equation 4.14 for a sample of length T and a fixed h and τ is calculated empirically with

$$V_\tau^{(h)}(T) = \frac{1}{T}\sum_{i=1}^{\lfloor T/h \rfloor}(N_{ih} - N_{(i-1)h} - \Lambda h)(N_{ih+\tau} - N_{(i-1)h+\tau} - \Lambda h)^\dagger \tag{4.40}$$

Using discrete Fourier transforms we can get a very good approximation of the right hand side matrix of equation (4.15). In our example, we fix $h = 0.5$ and we sample ϕ_t over $[-2, 2]$ very thinely with $\Delta = 10^{-4}$ (Δ is the sampling period for τ). Using Equation (4.18), we get the matrix $\widehat{\Psi}_{i\omega}$. As for $g_\tau^{(h)}$, its Fourier Transform is $\widehat{g}_{i\omega}^{(h)} = h\text{sinc}^2(\frac{\omega h}{2\pi})$ where $\text{sinc}(t) = \frac{\sin(\pi t)}{\pi t}$ is the sin cardinal function. Finally, we

take the inverse Fourier transforms and get the right hand side of equation (4.15). Because the matrix of $v_\tau^{(h)}$ is bisymmetric, we only need to represent its first row. This is shown in Figure 4.15.

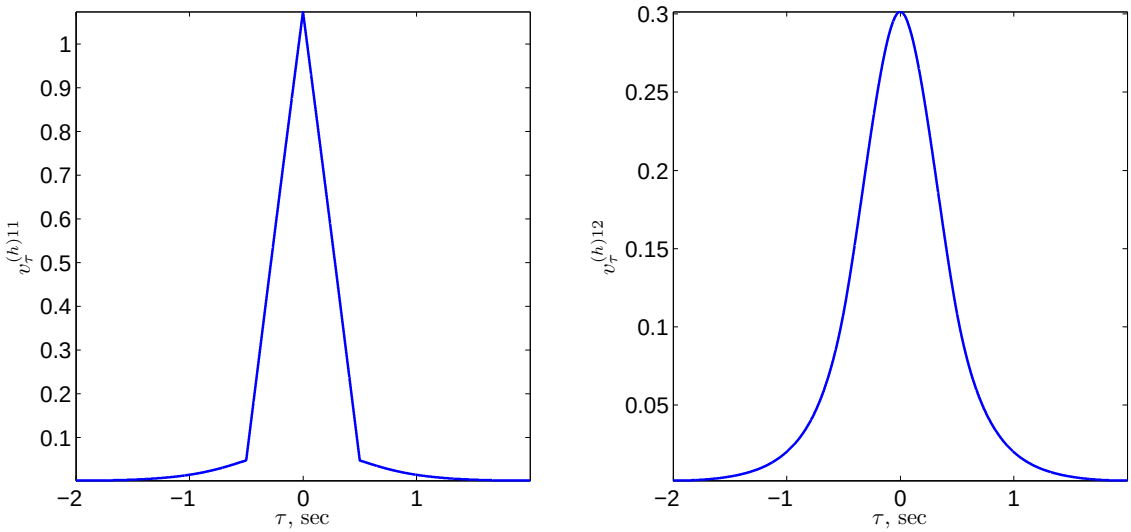


Figure 4.15: Theoretical representations of $v_\tau^{(h)}$ as a function of τ and for $h = 0.5$ computed using the right hand side of equation (4.15). The first graph is the function in the first row, first column of the matrix of $v_\tau^{(h)}$, while the second is the function in the first row, second column.

4.A.4 Convergence of the Covariance Operator

In [10], the author show that that $V_\tau^{(h)}(T)$ converges towards $v_\tau^{(h)}$ as T goes to ∞ with an L_2 error that decreases as T^{-1}. The L_2 error is defined

$$ev^2(T) = \sum_{k=1}^{\frac{\tau_{\max}}{\Delta}} |V_{k\Delta}^{(h)}(T) - v_{k\Delta}^{(h)}|^2 \tag{4.41}$$

where the sum is taken for each component of the matrix $v_{k\Delta}^{(h)}$. Looking at $ev^2(T)$ as a function of T will allow us to validate equation (4.15) and give us an idea of the convergence speed. We simulate a Hawkes process like the one described above with $\mu = (1, 1)$, $\alpha = 1$, $\beta = 4$ and $T = 4010^4$ and estimate $ev^2(kT/10)$ for $k = 1 \dots 10$. We plot the result in Figure 4.16 below (blue dots). We fit the data with a power law function of the form ax^b (red line shown in log-log plot). We find that the two error data points follow a very similar power law in T, $ev^2(T) = aT^b$ with $(a, b) = (2.0, -0.9)$ for the left graph and $(a, b) = (2.1, -0.9)$.

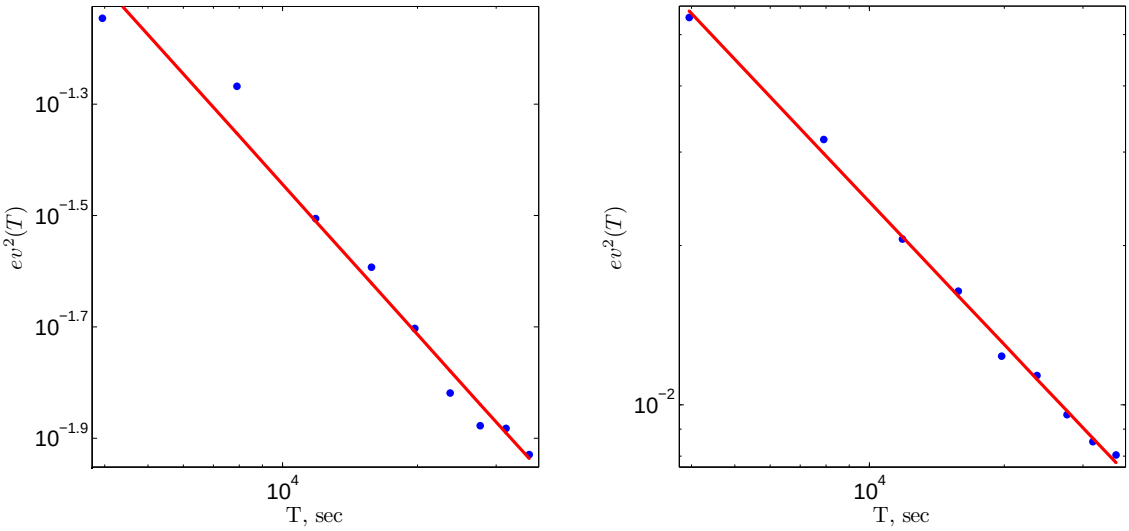


Figure 4.16: ev^2 error as a function of T in log-log scale. We find a linear relationship $bX + a$ with $b \simeq -1$ and $a \simeq 2$ for both graphs confirming the central limit theorem expressed in [10]

Bibliography

[1] CME website. http://ftp.cmegroup.com/globex/introduction/features-and-functionality/elements/matching-algorithms.html. [cited at p. 15]

[2] L. Adamopoulos. Cluster models for earthquakes: Regional comparisons. *Journal of the International Association for Mathematical Geology*, 8(4):463–475, August 1976. [cited at p. 123]

[3] Y. Ait-Sahalia, P. A. Mykland, and L. Zhang. How often to sample a continuous-time process in the presence of market microstructure noise. *The Review of Financial Studies*, 18:351–416, 2005. [cited at p. 72, 73]

[4] Y. Ait-Sahalia, P. A. Mykland, and L. Zhang. Ultra high frequency volatility estimation with dependent microstructure noise. *Journal of Econometrics*, 160(1):160–175, January 2011. [cited at p. 37, 73, 76]

[5] K. al Dayri, E. Bacry, and J.F. Muzy. Econophysics of order-driven markets. pages 155–172. Springer Verlag, 2011. [cited at p. 91]

[6] R. Almgren, C. Thum, E. Hauptmann, and H. Li. Direct estimation of equity market impact. *Risk*, July 2005. [cited at p. 21, 30, 53]

[7] T. Ané and H. Geman. Order flow, transaction clock, and normality of asset returns. *Journal of Finance*, 55(5):2259–2284, October 2000. [cited at p. 27]

[8] T. G. Andersen, T. Bollerslev, F. X. Diebold, and P. Labys. (Understanding, optimizing, using and forecasting) realized volatility and correlation. Working paper FIN-99-061, Department of Finance, Stern School of Business, New York University, 1999. [cited at p. 35, 72, 76]

[9] E. Bacry, S. Delattre, M. Hoffmann, and J. F Muzy. Modeling microstructure noise with mutually exciting point processes. January 2011. [cited at p. 12, 24, 45, 123, 131, 139, 145, 148]

[10] E. Bacry, S. Delattre, M. Hoffmann, and J. F Muzy. Scaling limits for hawkes processes. June 2011. [cited at p. 45, 51, 132, 141, 157, 158]

[11] P. Bak, M. Paczuski, and M. Shubik. Price variations in a stock market with many agents. *Physica A*, 246:430–453, 1997. [cited at p. 22]

[12] F. M. Bandi and J. R. Russell. Separating microstructure noise from volatility. *Journal of Financial Economics*, 79:655–692, 2006. [cited at p. 73]

[13] F. M. Bandi and J. R. Russell. Market microstructure noise, integrated variance estimators, and the accuracy of asymptotic approximations. *Journal of Econometrics*, 160(1):145–159, January 2011. [cited at p. 73]

[14] O. Barndorff-Nielsen, P. Hansen, A. Lunde, and N. Stephard. Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica*, 76(6):1481–1536, 2008. [cited at p. 73]

[15] M. S. Bartlett. The spectral analysis of point processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 25(2):264–296, January 1963. [cited at p. 124]

[16] M. S. Bartlett. The spectral analysis of Two-Dimensional point processes. *Biometrika*, 51(3/4):299–311, December 1964. [cited at p. 124]

[17] L. Bauwens and N. Hautsch. Modelling financial high frequency data using point processes. In T. Mikosch, J-P. Kreiss, R. A. Davis, and T. G. Andersen, editors, *Handbook of Financial Time Series*, pages 953–979. Springer Berlin Heidelberg, 2009. [cited at p. 45, 124]

[18] T. Bollerslev, I. Domowitz, and J. Wang. Order flow and the bid-ask spread: An empirical probability model of screen-based trading. *Journal of Economic Dynamics and Control*, 21(8-9):1471–1491, June 1997. [cited at p. 22]

[19] J-P. Bouchaud. An introduction to statistical finance. *Physica A: Statistical Mechanics and its Applications*, 313(1-2):238 – 251, 2002. [cited at p. 20]

[20] J-P. Bouchaud, J. D. Farmer, and F. Lillo. How markets slowly digest changes in supply and demand. In K. Schenk-Hoppe and T. Hens, editors, *Handbook of Financial Markets: Dynamics and Evolution*, pages 57–156. Elsevier: Academic Press, 2008. [cited at p. 20, 21, 22, 32, 52, 53, 66]

[21] J-P. Bouchaud, Y. Gefen, M. Potters, and M. Wyart. Fluctuations and response in financial markets: the subtle nature of 'random' price changes. *Quantitative Finance*, 10:176–190, April 2004. [cited at p. 21, 52]

[22] J-P. Bouchaud, J. Kockelkoren, and M. Potters. Random walks, liquidity molasses and critical response in financial markets. *Quantitative Finance*, 6(2):115–123, 2006. [cited at p. 21, 52]

[23] J-P. Bouchaud and M. Potters. *Theory of Financial Risk and Derivative Pricing*. Cambridge University Press, Cambridge, 2003. [cited at p. 13, 25, 27, 53, 71]

[24] C. G. Bowsher. Modelling security market events in continuous time: Intensity based, multivariate point process models. *Journal of Econometrics*, 141(2):876–912, December 2007. [cited at p. 45, 123]

[25] V. Chavez-Demoulin, A. C. Davison, and A. J. McNeil. Estimating value-at-risk: a point process approach. *Quantitative Finance*, 5(2):227–234, April 2005. [cited at p. 123]

[26] P. K. Clark. A subordinated stochastic process model with finite variance for speculative prices. *Econometrica. Journal of the Econometric Society*, 41(1):135âĂŞ155, 1973. [cited at p. 27]

[27] K. Cohen, R. Conroy, and S. F. Maier. Order flow and the quality of the market. In Y. Amihud, T. S. Y. Ho, and R. A. Schwartz, editors, *Market Making and the Changing Structure of the Securities Industry*. Beard Books, November 2002. [cited at p. 22]

[28] M. M. Dacorogna, R. Gençay, U. A. Müller, R. B. Olsen, and O. V. Pictet. *An introduction to high frequency finance*. Academic Press, 2001. [cited at p. 27, 29, 53, 58]

[29] M. G. Daniels, J. D. Farmer, L. Gillemot, G. Iori, and E. Smith. Quantitative model of price diffusion and market friction based on trading as a mechanistic random process. *Phys Rev Lett*, 90(10):108102, 2003. [cited at p. 22]

[30] F. X. Diebold and G. H. Strasser. On the correlation structure of microstructure noise in theory and practice. *Working Paper*, 2008. [cited at p. 73]

[31] I. Domowitz and J. Wang. Auctions as algorithms : Computerized trade execution and price discovery. *Journal of Economic Dynamics and Control*, 18(1):29–60, January 1994. [cited at p. 22]

[32] P. J. Brantingham E. Lewis, G. Mohler and A. Bertozzi. Self-exciting point process models of the insurgency in iraq. 6 2011. [cited at p. 123]

[33] Z. Eisler, J-P. Bouchaud, and J. Kockelkoren. The price impact of order book events: market orders, limit orders and cancellations. *Quantitative Finance*, pages 1–25, May 2011. [cited at p. 18, 19, 21, 52, 57]

[34] D. Eliezer and I. I. Kogan. Scaling laws for the market microstructure of the inter-dealer broker markets. *cond-mat/9808240*, August 1998. [cited at p. 22]

[35] P. Embrechts, J. T. Liniger, and L. Lu. Multivariate hawkes processes: an application to financial data. *Journal of Applied Probability*, 48:367–378, August 2011. [cited at p. 123]

[36] T. W. Epps. Comovements in stock prices in the very short run. *Journal of the American Statiscal Association*, 74:291–298, 1979. [cited at p. 72]

[37] J. D. Farmer, A. Gerig, F. Lillo, and S. Mike. Market efficiency and the long-memory of supply and demand: is price impact variable and permanent or fixed and temporary? *Quantitative Finance*, 6(2):107–112, April 2006. [cited at p. 21, 52]

[38] J. D. Farmer, L. Gillemot, F. Lillo, S. Mike, and S. Anindya. What really causes large price changes? *Quantitative Finance*, 4:383–397(15), August 2004. [cited at p. 21, 52]

[39] J. D. Farmer, P. Patelli, and I. I. Zovko. The predictive power of zero intelligence in financial markets. *Proceedings of the National Academy of Sciences of the United States of America*, 102(6):2254–2259, 2005. [cited at p. 22, 23]

[40] J. Gatheral. No dynamic arbitrage and market impact. *Quantitative Finance*, 10(7):749Ű759, 2010. [cited at p. 21, 53]

[41] A. N. Gerig. *A theory for market impact: How order flow affects stock price.* PhD thesis, University of Illinois at Urbana-Champaign, 2007. [cited at p. 21]

[42] K. Giesecke and L. R. Goldberg. A top down approach to Multi-Name credit. *SSRN eLibrary*, 2007. [cited at p. 123]

[43] L. Gillemot, J. D. Farmer, and F. Lillo. There's more to volatility than volume. *Quantitative Finance*, 6(5):371âĂŞ384, 2006. [cited at p. 21, 27, 30, 52]

[44] A. Gloter and J. Jacod. Diffusion with measurement errors. i. local asymptotic normality. *ESAIM : Prob. & Stat.*, 5:225–242, 2001. [cited at p. 72]

[45] A. Gloter and J. Jacod. Diffusion with measurement errors. ii. optimal estimator. *ESAIM : Prob. & Stat.*, 5:243–260, 2001. [cited at p. 72]

[46] A. Hawkes. Point spectra of some mutually exciting point processes. *Biometrika*, 58:83–90, April 1971. [cited at p. 44, 45, 123, 125]

[47] A. Hawkes. Spectra of some Self-Exciting and mutually exciting point processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 33-3:438–443, April 1971. [cited at p. 44, 123, 125, 127, 128]

[48] P. Hewlett. Clustering of order arrivals, price impact and trade path optimisation. In *Workshop on Financial Modeling with Jump processes*, 2006. [cited at p. 45, 123, 145]

[49] M. Hoffmann, A. Munk, and J. Schmidt-Hieber. Nonparamtreic estimation of the volatility under microstructure noise. 2010. arxiv:1007.4622, Math arXiv Preprint. [cited at p. 73]

[50] J. Jacod, Y. Li, P. A. Mykland, M. Podolskij, and M. Vetter. Microstructure noise in the continuous case: The pre-averaging approach. *Stochastic Process. Appl.*, 119(7):2249–2276, 2009. [cited at p. 73]

[51] M. Krumin, I. Reutsky, and S. Shoham. Correlation-Based analysis and generation of multiple spike trains using hawkes models with an exogenous input. *Frontiers in computational neuroscience*, 4(November):12, 2010. [cited at p. 152]

[52] J. Large. Measuring the resiliency of an electronic limit order book. *Journal of Financial Markets*, 10(1):1–25, 2007. [cited at p. 145]

[53] E. Lewis and G. Mohler. A nonparametric EM algorithm for a multiscale hawkes process. Joint Satistical Meetings 2011, 2011. [cited at p. 124]

[54] F. Lillo and J. D. Farmer. The long memory of the efficient market. *Studies in Nonlinear Dynamics & Econometrics*, 8(3), 2004. [cited at p. 21, 52]

[55] J. T. Liniger. *Multivariate Hawkes Processes.* PhD thesis, ETH Zurich, 2009. [cited at p. 123]

[56] B. Mandelbrot. The variation of certain speculative prices. *Journal of Business*, 36:394, 1963. [cited at p. 27]

[57] B. Mandelbrot and H. W. Taylor. On the distribution of stock price differences. *Operation Research*, 15:1057–1062, 1967. [cited at p. 53]

[58] D. Marsan and O. Lengliné. Extending earthquakes' reach through cascading. *Science*, 319(5866):1076 –1079, February 2008. [cited at p. 124]

[59] S. Maslov. Simple model of a limit order-driven market. *Physica A: Statistical Mechanics and its Applications*, 278(3-4):571–578, April 2000. [cited at p. 22]

[60] H. Mendelson. Market behavior in a clearing house. *Econometrica*, 50(6):1505–1524, November 1982. ArticleType: research-article / Full publication date: Nov., 1982 / Copyright Âl' 1982 The Econometric Society. [cited at p. 22]

[61] G. Mohler, M. Short, P. Brantingham, F. Schoenberg, and GE Tita. Self-Exciting point process modeling of crime. *Journal of the American Statistical Association*, 106(493):100–108, March 2011. [cited at p. 123]

[62] A. Munk and J. Schmidt-Hieber. Nonparametric estimation of the volatility function in a high-frequency model corrupted by noise. 2009. arXiv:0908.3163, Math arXiv Preprint. [cited at p. 73]

[63] A. Munk and J. Schmidt-Hieber. Lower bounds for volatility estimation in microstructure noise models. *A Festschrift for Larry Brown, IMS Lecture Notes Series*, 2010. To appear. [cited at p. 73]

[64] M. Musiela and M. Rutkowski. *Martingale Methods in Financial Modelling.* Springer, New York, 2004. [cited at p. 71]

[65] Y. Ogata. On lewis' simulation method for point processes. *Ieee Transactions On Information Theory*, 27:23–31, January 1981. [cited at p. 155]

[66] Y. Ogata. Seismicity analysis through point-process modeling: A review. *Pure and Applied Geophysics*, 155(2-4):471–507, August 1999. [cited at p. 123]

[67] Y. Ogata and H. Akaike. On linear intensity models for mixed doubly stochastic poisson and self- exciting point processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 44(1):102–107, January 1982. ArticleType: research-article / Full publication date: 1982 / Copyright Âl' 1982 Royal Statistical Society. [cited at p. 124]

[68] J-P. Onnela, J. TÃűyli, and K. Kaski. Tick size and stock returns. *Physica A: Statistical Mechanics and its Applications*, 388(4):441–454, February 2009. [cited at p. 20]

[69] Roel C Oomen. Statistical models for high frequency security prices. *SSRN eLibrary*, 2002. [cited at p. 74]

[70] T. Ozaki. Maximum likelihood estimation of hawkes' self-exciting point processes. *Annals of the Institute of Statistical Mathematics*, 31(1):145–155, December 1979. [cited at p. 124]

[71] R.E.A.C. Paley and N. Wiener. *Fourier transforms in the complex domain*. Colloquium Publications - American Mathematical Society. American Mathematical Society, 1934. [cited at p. 136]

[72] M. Podolskij and M. Vetter. Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Bernoulli*, 15:634–658, 2009. [cited at p. 73]

[73] M. Reiß. Asymptotic equivalence and sufficiency for volatility estimation under microstructure noise. 2010. arxiv:1001.3006, Math arXiv Preprint. [cited at p. 73]

[74] P. Reynaud-bouret and S. Schbath. Adaptive estimation for hawkes processes; application to genome analysis. *The Annals of Statistics*, 38(5):2781–2822, October 2010. Zentralblatt MATH identifier: 1200.62135; Mathematical Reviews number (MathSciNet): MR2722456. [cited at p. 123, 124, 152]

[75] C. Y. Robert and M. Rosenbaum. A new approach for the dynamics of Ultra-High-Frequency data: The model with uncertainty zones. *Forthcoming, Journal of Financial Econometrics*, 2010. [cited at p. 19, 83, 84]

[76] Christian Y. Robert and Mathieu Rosenbaum. A New Approach for the Dynamics of Ultra-High-Frequency Data: The Model with Uncertainty Zones. *Forthcoming, Journal of Financial Econometrics*, 2010. [cited at p. 18, 57]

[77] C.Y. Robert and M. Rosenbaum. A new approach for the dynamics of ultra high frequency data: the model with uncertainty zones. *Journal of Financial Econmetrics*, 2009. In press. [cited at p. 73]

[78] M. Rosenbaum. A new microstructure noise index. *Quantitative Finance*, 2007. In press. [cited at p. 73]

[79] B.-D. Seo. Realized volatility and colored market microstructure noise. Manuscript, 2005. [cited at p. 73]

[80] F Slanina. Mean-field approximation for a limit order driven market model. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, 64(5 Pt 2):056136, November 2001. PMID: 11736043. [cited at p. 22]

[81] E. Smith, J. D. Farmer, L. Gillemot, and S. Krishnamurthy. Statistical theory of the continuous double auction. *Quantitative Finance*, 3:481–514, 2003. [cited at p. 22, 23]

[82] L-H. Tang and G-S. Tian. Reaction-diffusion-branching models of stock price fluctuations. *Physica A: Statistical and Theoretical Physics*, 264(3-4):543–550, March 1999. [cited at p. 22]

[83] I. M. Toke. Econophysics of order-driven markets. chapter "Market making" behaviour in an order book model and its impact on the bid-ask spread. Springer Verlag, 2011. [cited at p. 34, 123]

[84] A. E. D. Veraart and M. Winkel. Time change. In *Encyclopedia of Quantitative Finance*. John Wiley & Sons, Ltd, 2010. [cited at p. 27]

[85] D. Vere-Jones. Stochastic models for earthquake occurrence. *Journal of the Royal Statistical Society. Series B (Methodological)*, 32(1):pp. 1–62, 1970. [cited at p. 123]

[86] D. Vere-Jones and T. Ozaki. Some examples of statistical estimation applied to earthquake data. *Annals of the Institute of Statistical Mathematics*, 34(1):189–207, December 1982. [cited at p. 123]

[87] Matthieu Wyart, Jean-Philippe Bouchaud, Julien Kockelkoren, Marc Potters, and Michele Vettorazzo. Relation between bid-ask spread, impact and volatility in order-driven markets. *Quantitative Finance*, 8(1):41–57, 2008. [cited at p. 32, 33, 54, 66, 69]

[88] L. Zhang. Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. *Bernoulli*, 12:1019–1043, 2006. [cited at p. 72]

[89] L. Zhang, P. Mykland, and Y. Ait-Sahalia. A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *J. Amer. Statist. Assoc.*, 472:1394–1411, 2005. [cited at p. 72]

[90] Jiancang Zhuang, Yosihiko Ogata, and David Vere-Jones. Stochastic declustering of space-time earthquake occurrences. *Journal of the American Statistical Association*, 97(458):pp. 369–380, 2002. [cited at p. 124]