Market activity and price impact throughout time scales

Thibault Jaisson

THÈSE

présentée pour obtenir

LE GRADE DE DOCTEUR EN SCIENCES DE L'ECOLE POLYTECHNIQUE

Spécialité : Mathématiques

par

Thibault JAISSON

Market activity and price impact throughout time scales

Soutenue le 13 Octobre 2015 devant un jury composé de :

Jean-Philippe BOUCHAUD	rapporteur
Per MYKLAND	rapporteur
Marc HOFFMANN	examinateur
Jean JACOD	examinateur
Charles-Albert LEHALLE	examinateur
Gilles PAGES	examinateur
Emmanuel BACRY	directeur de thèse
Mathieu ROSENBAUM	directeur de thèse

Remerciements

Je tiens en premier lieu à exprimer ma plus profonde gratitude envers mes directeurs de thèse Emmanuel Bacry et Mathieu Rosenbaum. Leur soutien et leur confiance durant ces trois ans m'ont permis de mener à bien ce travail. Merci pour tout ce que vous m'avez fait découvrir et pour tout ce que j'ai appris sous votre direction!

Je remercie Jean-Philippe Bouchaud et Per Mykland d'avoir accepté de rapporter ma thèse. Je suis très honoré par leur lecture attentive de ce manuscrit et leur intéret pour mon travail. Je remercie également Marc Hoffmann, Jean Jacod et Gilles Pagès d'examiner ma thèse et de participer à ma soutenance.

Je suis très reconnaissant envers Jim Gatheral pour son accueil à Baruch College l'année dernière, pour sa diponibilité et ses conseils durant mon séjour et surtout pour m'avoir introduit au sujet de la volatilité rugueuse qui est au coeur de ma thèse.

Toute ma reconnaissance va également à mes professeurs de Polytechnique et du Master, en particulier Caroline Hillairet, Stefano De Marco, Nicole El Karoui, Emmanuel Gobet et Nizar Touzi pour m'avoir fait découvrir les mathématiques financières. Je remercie aussi Charles-Albert Lehalle pour m'avoir introduit à la microstructure des marchés au Master El Karoui et surtout en stage chez Cheuvreux ainsi que pour examiner ma thèse.

Je remercie Stephane Gaïffas, Iacopo Mastromatteo et Jean-François Muzy pour leur temps et leur disponibilité ainsi que pour toute les discussions enrichissantes que j'ai pu avoir avec eux au cours de cette thèse.

Merci à tout les doctorants de Polytechnique et de Jussieu qui ont contribués à rendre mon quotidien plus agréable durant ces trois ans Alexandre, Alice, Andrea, Antoine, Aymeric, Benjamin, Camille, Candia, Charline, Eric, Etienne, Faissal, Guillaume, Gwenael, Hélène, Jean-Leopold, Jiatu, Jonathan, Laurent, Li, Manon, Massil, Nicolas, Nina, Pascal, Pauline, Pierre, Pierre-Yves, Saad, Sébastien, Thomas, Weibing et Xavier.

Je remercie égalemement tout le secrétariat du CMAP et du LPMA et plus particuliérement Nasséra Naar et Alexandra Noiret pour leur disponibilité et leur aide pendant trois ans.

Pour finir, je remercie mes parents, mes frères, ma famille, mes amis et Marion pour leurs encouragements et leur soutien durant cette thèse et bien avant.

Résumé

Cette thèse traite plusieurs problèmes posés par le caractère multi-échelle des données financières. Elle est composée de quatre parties interconnectées pouvant néanmoins être lues indépendamment.

On introduit dans la première partie le modèle RFSV à volatilité rugueuse. Dans ce modèle, le logarithme de la volatilité est modélisé par un processus d'Ornstein-Uhlenbeck fractionnaire à temps de retour à la moyenne long et exposant de Hurst petit. Nous montrons que ce modèle permet de reproduire le comportement de la surface de volatilité ainsi que de nombreux faits stylisés statistiques tels que la structure d'autocorrélation et la régularité de la volatilité. En particulier, nous semblons montrer que, contrairement à l'idée communément admise, celle-ci ne présente pas de longue mémoire en loi de puissance.

La deuxième partie porte sur le comportement en temps long des processus de Hawkes et des processus autorégressifs quasi instables, c'est-à-dire dont le noyau a une norme proche de un. Nous montrons que dans le cas où le noyau est à queue fine, les processus de Hawkes quasi instables se comportent asymptotiquement comme des processus de Cox-Ingersoll-Ross intégrés. Dans le cas où le noyau est en loi de puissance, en fonction de son exposant, la distribution limite est un processus à longue mémoire ou un processus C^1 à derivée rugueuse. En modélisant le flux d'ordres par un processus de Hawkes, ce résultat nous permet d'établir des fondements microstructurels à l'irrégularité de la volatilité évoquée dans la première partie. Nous obtenons des résultats similaires pour les processus autorégressifs.

Dans la troisième partie de cette thèse, nous proposons deux procédures statistiques permettant d'estimer les noyaux de processus de Hawkes. Nous adaptons ces algorithmes au cas où les noyaux sont à décroissance lente, situation observée sur les données. La première procédure se fonde sur l'inversion de l'équation de Wiener-Hopf tandis que la deuxiéme consiste en une maximisation de la vraisemblance par descente de gradient stochastique.

Enfin, dans la dernière partie de cette thèse, nous nous intéressons, sous des hypothèses les plus générales possible, aux conséquences de l'efficience du prix sur la structure de l'impact de marché. Nous obtenons l'existence d'un prix efficient par rapport auquel le gain ex post des ordres limites doit être nul. Finalement, nous montrons que si l'impact permanent des métaordres est linéaire en leur volume, les fluctuations de prix doivent être proportionnelles aux variations de l'anticipation par le marché du déséquilibre de volume.

Abstract

This thesis tackles several issues raised by the multi-scale properties of financial data. It consists of four connected parts which can however be read independently.

In the first part, we introduce the Rough Fractional Stochastic Volatility (RFSV) model. In this framework, the logarithm of the volatility is a fractional Ornstein-Uhlenbeck process with long mean reversion time scale and small Hurst index. We show that this model reproduces numerous stylized facts of financial data such as implied volatility properties, the autocorrelation structure and the regularity of the volatility. In particular, the RFSV sheds some new light on the supposed long memory property of volatility.

The second part deals with the long-time behavior of Hawkes and autoregressive processes which are nearly unstable, that is whose kernels have a norm close to one. We show that in the case where the kernel has a light tail, nearly unstable Hawkes processes asymptotically behaves as integrated Cox-Ingersoll-Ross (CIR) processes. In the case where it has a power law shape, depending on the power law exponent, the limiting process asymptotically exhibits either long memory or a rough behavior. This enables us to use the modeling of the order flow as a Hawkes process to obtain a microstructural foundation of the rough nature of volatility. Similar results are obtained for autoregressive processes.

In the third part of this thesis, we propose two estimation procedures for the kernels of Hawkes processes. These algorithms are in particular relevant to the case of slowly decreasing Hawkes kernels which corresponds to what is observed on financial data. The first method relies on the inversion of the Wiener-Hopf equation while the second is based on a likelihood maximization via a stochastic gradient ascent.

Finally, we are interested in deriving, under very general assumptions, consequences of market efficiency on the structure of price impact. We prove the existence of a model independent fair price with respect to which the average ex post gain of limit orders must be equal to zero. We finally show that, under a few additional assumptions, price impact must be proportional to the market anticipation of the order flow imbalance.

Keywords: Multiscale properties, Hawkes processes, autoregressive processes, nearly unstable processes, CIR, long memory, persistence, regularity, rough volatility, order flow, RFSV, estimation of Hawkes kernels, slowly decreasing kernels, Wiener-Hopf, likelihood maximization, stochastic gradient descent, market efficiency, price impact, fair markets, fair price, order flow imbalance, metaorders, anticipation of the volume, square root law, impact function.

List of papers being part of this thesis

- T. Jaisson and M. Rosenbaum, *Limit theorems for nearly unstable Hawkes processes*, Annals of Applied Probability, 25(2):600-631, 2015.
- T. Jaisson, *Market impact as anticipation of the order flow imbalance*, Quantitative Finance, 15(7):1123–1135, 2015.
- J. Gatheral, T. Jaisson and M. Rosenbaum, *Volatility is rough*, submitted, 2014.
- E. Bacry, T. Jaisson and J.-F. Muzy, *Estimation of slowly decreasing decreasing Hawkes kernels: Application to high frequency order book dynamics*, to appear in Quantitative Finance, 2015.
- T. Jaisson and M. Rosenbaum, *The different asymptotic regimes of nearly unstable autoregressive processes*, to appear in Festschrift for Ole Barndorff-Nielsen, Springer, 2015.
- T. Jaisson and M. Rosenbaum, *Rough fractional diffusions as scaling limits of nearly unstable heavy tailed Hawkes processes*, submitted, 2015.
- T. Jaisson, *Liquidity and impact in fair markets*, submitted, 2015.

Contents

Introduction

The guiding principle of this thesis is to study how various financial processes such as prices, volatility or the different order flows behave and are related at different time scales (from seconds to years). We aim at defining useful models which simultaneously reproduce most of the relevant features and at building estimation procedures in these models. Let us begin by presenting and motivating the different questions on which we want to shed some light in this thesis.

Motivations

In the world of derivatives, the most important ingredient of any model is the volatility of prices. To fit the properties of implied volatility surfaces, the most popular frameworks are the local and stochastic volatility models. However, from an econometric point of view, these models are unsatisfactory because, as opposed to what is observed on data, the volatility processes in these models are typically unable to reproduce the volatility correlation structure. To solve this issue, based on the idea that volatility presents a power law long memory, models were introduced where the volatility process is driven by a fractional Brownian motion with Hurst index larger than 1/2. However, we will see that none of these models are "rough" enough to reproduce the empirically observed regularity of the volatility. We might thus ask ourselves the following question:

Question 1. *How does the local regularity of volatility relates to its global correlation structure?*

Another important financial stylised fact is that the flow of market orders is persistent. This means that if there has been a lot of trades in the near past there will, on average, be a lot of trades in the near future. Thus, Poisson processes are not adapted to modeling the order flow. One way to reproduce this positive autocorrelation in the flow of market orders is to use Hawkes processes defined as point processes whose intensity writes as a linear regression on the past points of the process, see their definition below. However, on the one hand, we will see that the flow of market orders presents a multiscaling behavior in the sense that, whatever the time scale at which one is "looking" at this process, there are periods of high market activity (with a lot of trades) and periods of low market activity (with few trades). On the other hand, it has been shown that classical Hawkes processes cannot reproduce this feature. Indeed, they are short memory in the sense that when one considers a Hawkes process at long time scales, the periods of low and high market activity average out. Moreover, in practice,

empirical estimations of the parameters of Hawkes processes lead to processes whose stability condition is almost saturated, meaning that the L^1 norm of the regression kernel is very close to one. This leads to the following question:

Question 2. *Can nearly unstable Hawkes processes reproduce the multiscaling behavior of the order flow?*

In addition to being autocorrelated, the flow of market orders is cross-correlated with cancel and limit orders and with price changes. For example, trades tend to be followed by limit orders. One way to model this cross-correlation is to use multidimensional Hawkes processes which are self and mutually exciting point processes. However, the Hawkes influence kernels which appear in the definition of these processes and encode the causality between the different kinds of events are not directly observable on the data. We only have access to cross-correlation functions. This makes the following question important to fit Hawkes processes to the data in an appropriate way:

Question 3. *How to disentangle Hawkes influence kernels from emprirical cross-correlations?*

An important example of such cross-influence is that of trades on price moves. Indeed, market orders on average trigger price moves in the same direction. This feature called price or market impact is important for many agents such as brokers and market makers who need to take into account the effect of their or someone else's trades on the price. However, while the price is more or less a martingale because of no arbitrage arguments, the imbalance between the flow of market orders at the bid and at the ask is persistent. That is, if there has been more orders at the ask than at the bid in the past there will, on average, be more orders at the ask than at the bid in the future. This implies that the trivial impact model where the price move on any given time window is on average proportional to the imbalance between ask and bid market orders on the same period cannot be realistic. Consequently, the persistence of the order flows must be strongly linked to the structure of the influence of trades on price moves to yield a martingale price. This leads to the following question about this connection:

Question 4. *What model independent conditions on the price impact of order flows can be deduced from general no arbitrage assumptions?*

Outline

Each question presented above corresponds to a part of the thesis.

In Part I, we answer Question 1 by exhibiting a remarkable universal scaling law for the moments of the log-volatility increments. We show that modeling the log-volatility as some sort of rough fractional Brownian motion as suggested by this scaling enables us to reproduce both historical and risk neutral properties of the volatility. For example, we retrieve the smoothness and the empirical autocorrelation structure of the volatility together with the shape of the implied volatility surface. Furthermore, this model, called Rough Fractional Stochastic Volatility (RFSV for short) sheds new light on the discussion about the supposed long memory

of volatility.

Answers to Question 2 lie in Part II where we study the long time behavior of Hawkes and autoregressive processes whose stability conditions are almost saturated. For Hawkes processes, we show in Chapter II that if the kernel has a light tail, there exists a "natural" time renormalization at which the Hawkes intensity behaves as a CIR process and the Hawkes process as an integrated CIR. In Chapter III, we focus on the case where the kernel asymptotically behaves as an heavy tail power law. We show that depending on the power law exponent, the Hawkes process either asymptotically exhibits a power law long memory, or behaves as an integrated rough fractional diffusion. We also show that similar asymptotic results hold for nearly unstable autoregressive processes in Chapter IV. Therefore, modeling the order flow as a power law nearly unstable Hawkes process and taking the cumulated variance proportional to the cumulated order flow (as stated by financial economics literature) provides a microstructural foundation for the roughness of the volatility exhibited in Part I.

In Part III, we tackle Question 3 from two different perspectives. In Chapter V, we adapt the numerical procedure classically used to derive the Hawkes kernels from the empirical second order properties of the process to the case of slowly decreasing Hawkes kernels. To do this, we disentangle the cross-causality of the point process from its cross-correlation by solving numerically the Wiener-Hopf equation. We apply this procedure to fit an 8-dimensional Hawkes order book model and estimate the influence of the different events on each other. In Chapter VI, we consider the estimation of Hawkes kernels using a likelihood maximization procedure. We use the structure of the log-likelihood to propose a new parallelizable stochastic gradient ascent algorithm which outperforms usual methods such as the Majoration-Minimization (MM) algorithm. We also provide a non parametric representation of the kernels to account for slowly decreasing Hawkes kernels.

Finally, in Part IV, we present two works related to Question 4 which aim at studying the impact of market orders and metaorders under market efficiency assumptions. In Chapter VII, we revisit the notion of fair markets generalizing the approach of Madhavan-Richardson-Roomans (MRR for short). We show that, in such markets, there exists a fair price that should be considered as a reference price to compute the ex post gains of transactions. In Chapter VIII, we add the assumption that the permanent market impact of metaorders is linear in their size. We obtain a model independent relationship between the order flow and the price which generalizes the propagator model of Bouchaud et al. We apply this relationship to power law nearly unstable Hawkes processes and retrieve a power law impact function as it is observed in practice.

Let us now rapidly review the main results of this thesis.

1 Part I: How regular and persistent is really volatility?

A paramount feature of financial data is that the volatility (corresponding to the amplitude of price fluctuations) is positively autocorrelated (if prices have moved a lot yesterday, they are likely to move a lot today). Understanding this stylised fact is of course of great importance in both asset management where one needs to anticipate the amplitude of price moves and derivative pricing where the volatility process is the most important ingredient of every model.

In Chapter I, we study the regularity of the volatility and we show that, at any time scale of interest, the log-volatility essentially behaves as a fractional Brownian motion with Hurst index around 0.1. We introduce a rough version of the fractional stochastic volatility model of Comte and Renault, see [CR98], which reproduces this feature. We show that in addition to volatility surface properties, this model fits the volatility structure better than usual long memory or multifractal models and we apply it to variance prediction and derivative pricing.

1.1 Regularity of the volatility

In order to build our model from data, we look at the scaling of the increments of the (estimated) log-volatility

$$m(\Delta, q) = \mathbb{E}[|\log(\sigma_{t+\Delta}) - \log(\sigma_t)|^q].$$

Plotting $\log(m(\Delta, q))$ as a function of $\log(\Delta)$, we notice the following result which is valid for a very wide range of Δ, say from hours to months.

Result 1. *On a very wide range of time scales (at least from hours to months), the increments of the log-volatility behave as that of a fractional Brownian motion with small Hurst H index ($H \sim 0.1$):*

$$m(\Delta, q) \simeq K_q \Delta^{Hq}.$$

This leads us to model the volatility as a geometric fractional Brownian motion

$$\sigma_t = \sigma_0 e^{\eta W_t^H}.$$

However, the lack of stationarity of this model may seem unsatisfactory. In particular, the autocorrelation function of the volatility that we wish to study is not very well defined.

In order to solve this technical issue, we make our model stationary by introducing a slow mean reversion term. More precisely, we consider a rough version of the FSV model of Comte and Renault, see [CR98]

$$d\log(\sigma_t) = \eta dW_t^H + \alpha(m - \log(\sigma_t))dt$$

where the mean reversion time scale $1/\alpha$ is very large compared to the time scales of interest.

It is important to stress that in spite of the apparent similarities between the RFSV model and the classical FSV model, they are in fact very different. Indeed, on the one hand, in [CR98],

H is taken larger than $1/2$ to reproduce the supposed long memory of the volatility (periods of high volatility tend to be followed by periods of high volatility) while α is taken large to reproduce the mean reversion of the volatility (upward movements of the volatility tend to be followed by downward movements of the volatility). On the other hand, in the RFSV model, H is taken smaller than $1/2$ to reproduce the roughness of the volatility (which corresponds to some sort of mean reversion) while α is taken small to reproduce the persistence (or positive autocorrelation) of the volatility.

1.2 Volatility properties

In Chapter I, we apply the RFSV model to the description of the volatility process.

1.2.1 Persistence

A supposed stylised fact about the autocorrelation of the volatility

$$c(\Delta) = \mathrm{Cov}[\sigma_t, \sigma_{t+\Delta}]$$

is that it presents a power law long memory in the sense that c asymptotically behaves as

$$c(\Delta) \underset{\Delta \to +\infty}{\sim} \Delta^{-\gamma}$$

where $\gamma < 1$, see [ABDL03] or [DGE93]. This led some authors to use fractional Brownian motions with Hurst index $H > 1/2$ in order to reproduce this property, see [CR98] and [MO06].

In Chapter I, we will see that in our model, the autocorrelation of the volatility asymptotically satisfies the following property when the mean reverting term α becomes small.

Result 2.

$$c(\Delta) = Ae^{-B\Delta^{2H}} - \mathbb{E}[\sigma_t]^2 + o(1).$$

Furthermore, we show that this relation is consistent with empirical measures of the autocorrelation of the volatility.

1.2.2 Multiscaling

Another interesting property of the RFSV model is that as H tends to zero, its correlation structure tends towards that of the multifractal random measure of Bacry, Delour and Muzy (BDM), see [BKM08] and [BM03]. Moreover, recall that the BDM model is built to satisfy a scale invariance property in the sense that whatever the time scale at which one observes the "volatility" in this model, one "sees" the same process. We will show that although the RFSV is not strictly speaking scale invariant it apparently reproduces the empirically observed wide range scale invariance when H is small.

1.2.3 Volatility forecasting

Modeling the log-volatility as a fractional Brownian motion as suggested by the scaling laws mentioned above enables us to use properties of the conditional laws of fractional Brownian motions given in [NP00]. We obtain the following result about the prediction of the volatility.

Result 3. *The expected log-volatility and volatility conditionally on the past satisfies*

$$\mathbb{E}\left[\log\sigma_{t+\Delta}|\mathscr{F}_t\right] = \frac{\cos(H\pi)}{\pi}\Delta^{H+1/2}\int_{-\infty}^{t}\frac{\log\sigma_s}{(t-s+\Delta)(t-s)^{H+1/2}}ds$$

and

$$\mathbb{E}[\sigma^2_{t+\Delta}|\mathscr{F}_t] = \exp\left\{2\mathbb{E}\left[\log\sigma_{t+\Delta}|\mathscr{F}_t\right] + 2\frac{\Gamma(3/2-H)}{\Gamma(H+1/2)\,\Gamma(2-2H)}\nu^2\Delta^{2H}\right\}.$$

We test discrete time versions of these predictors on real data and show that they always outperform usual variance predictors such as the Autoregressive and Heterogeneous Autoregressive predictors, see [Cor09].

1.3 Option pricing

Beyond reproducing the statistical properties of the volatility process, the RFSV model can also be applied to fit volatility surfaces as it is done in [BFG15]. To achieve this, one needs to introduce some (negative) correlation between the movements of volatility and prices. One (quite ad hoc) natural way to do that is to consider the Mandelbroot-Van-Ness representation of the fractional Brownian motion

$$W_t^H = \int_{-\infty}^{t}[(t-s)_+^{H-1/2} - (-s)_+^{H-1/2}]dW_s \tag{1}$$

and to correlate the Brownian motion driving the price with the one driving the fractional Brownian motion. It is shown in [Fuk15] that, when doing so and taking H small, one retrieves a short-time at-the-money skew behaving as a power law function of the maturity, as it is observed in practice

$$\frac{\widehat{\sigma}(S_0+\sqrt{T}x,T)-\widehat{\sigma}(S_0,T)}{\sqrt{T}x} \underset{T\to 0}{\sim} \frac{c}{T^{1/2-H}}$$

where $\widehat{\sigma}$ is the implied volatility surface.

2 Part II: Nearly unstable linear processes

In Chapters II, III and IV, we study the long time behavior of linear processes (Hawkes and autoregressive processes) when their stability condition is almost saturated, that is when the L^1 norm of their kernel is close to one.

2.1 Hawkes processes

Hawkes processes were introduced in 1971 in [Haw71a] and [Haw71b] to reproduce the clustering of earthquakes. A Hawkes process N is defined as a point process whose intensity λ writes as the sum of an exogenous term μ and a linear regression on the past of the process

$$\lambda_t = \mu + \int_{-\infty}^{t} \phi(t-s) dN_s.$$

Over the last decade, these processes have been applied to several fields where it is necessary to have models which reproduce the clustering of some phenomena and thus where Poisson processes are inadapted. For example, Hawkes processes have been successfully used to reproduce the correlation of earthquakes, see [Ada76], genomic sequences, see [RBS+10], neural spikes, see [RBRGTM14], terrorist attacks, see [LMBB12], crime, see [MSB+11], price moves, see [BDHM13] or financial defaults, see [ASCDL10]. In Part II, we motivate and illustrate our results by using Hawkes processes to model the arrival of market orders.

There are many reasons for this recent popularity of Hawkes processes. First of all, Hawkes processes are tractable and very easy to simulate by thinning, see [Oga81]. Moreover, we will see that Hawkes processes can be efficiently fitted to reproduce the cross-correlations between events, see [BM14b]. Finally, let us mention the nice population dynamics interpretation of Hawkes processes that we will explicit later. This representation enables us to measure the endogeneity and causality between events in the Hawkes model, see [HO74]. In particular, we will see that the L^1 norm of the kernel corresponds to the proportion of endogenous points of the process.

2.2 Light tailed nearly unstable Hawkes processes

An important property of Hawkes processes is that under the following stability condition

$$\|\phi\|_1 = \int_0^{+\infty} \phi(s) ds < 1$$

the process admits a version with a stationary intensity. Moreover, it is shown in [BDHM12] that under this condition, Hawkes processes present weak dependence in the sense that they asymptotically behave as their expectation at long time scales:

$$\frac{N_{Tt}}{T} \underset{T \to +\infty}{\longrightarrow} \frac{\mu}{1 - |\phi|} t.$$

However, in practice, for financial order flows that we want to model as Hawkes processes, the asymptotic behavior above is not satisfied. Indeed, the cumulated order flow does not behave as a linear function of time. In fact, it presents a multiscaling behavior in the sense that whatever the time scale at which one is looking at the order flow, one observes periods of high market activity (with a lot of trades) and periods of low market activity (with few trades).

Moreover, statistical estimation results seem to show that very often, only ***nearly unstable Hawkes processes*** are able to fit the data properly, see [FS12] and [HBB13]. By nearly unstable, we mean that the L^1 norm of their kernel is close to unity. In Chapter II we formally study such processes for which the stability condition is almost violated and the kernel has a light tail.

2.2.1 Asymptotic framework

In order to have a mathematical framework in which we can study the long time behavior of Hawkes processes whose kernel's norm is close to one, we consider a sequence of Hawkes processes $(N_t^T)_{t\geq 0}$ indexed by T which corresponds to the observation scale. For a given T, (N_t^T) satisfies $N_0^T = 0$ and its intensity process (λ_t^T) is defined for $t \geq 0$ by

$$\lambda_t^T = \mu + \int_0^t \phi^T(t-s)dN_s^T,$$

where μ is a positive real number and ϕ^T a non negative measurable function on $\mathbb{R}^+$.

Let us now give more specific assumptions on the asymptotic behavior of the function ϕ^T. For $t \in \mathbb{R}^+$,

$$\phi^T(t) = a_T \phi(t),$$

where $(a_T)_{T\geq 0}$ is a sequence of positive numbers smaller than one and converging to one and ϕ is a non negative measurable function which we call the shape of the kernel and does not depend on T such that

$$\int_0^{+\infty} \phi(s)ds = 1.$$

Thus, the form of the function ϕ^T depends on T so that its shape is fixed but its L^1 norm varies with T. For a given T, this L^1 norm is equal to a_T and so is smaller than one, implying that the stability condition is in force. Since a_T tends to one, we call our sequence nearly unstable Hawkes processes.

In Chapter II, we also require the following light tail condition:

Assumption 1.

$$\int_0^{+\infty} s\phi(s)ds = m < \infty.$$

2.2.2 Observation scales

In our framework, two parameters degenerate at infinity: T and $(1-a_T)^{-1}$. The relationship between these two sequences will determine the scaling behavior of the sequence of Hawkes processes. Recall that it is shown in [BDHM12] that when $\|\phi^T\|_1$ is fixed and smaller than one, the limit of the sequence of Hawkes processes is deterministic. In our setting, if $1-a_T$ tends "slowly" to zero, we get the same kind of result.

On the contrary, if $1 - a_T$ tends too rapidly to zero, the situation is likely to be quite intricate. Indeed, for given T, the Hawkes process may already be very close to instability whereas T is not large enough to reach the asymptotic regime. The last case, which is probably the most interesting one, is the intermediate case, where $1 - a_T$ tends to zero in such a manner that a non deterministic scaling limit is obtained, while not being in the preceding degenerate settings. More precisely, in Chapter II, we make the following assumption:

Assumption 2. *There exists $\lambda > 0$ such that*

$$T(1 - a_T) \underset{T \to +\infty}{\to} \lambda.$$

2.2.3 Non degenerate scaling limit for nearly unstable Hawkes processes

We show in Chapter II the following result about the renormalized Hawkes intensity observed at a time scale T:

$$C_t^T = (1 - a_T)\lambda_{Tt}^T.$$

Result 4. *Under Assumptions 1 and 2, the sequence of renormalized Hawkes intensities (C_t^T) converges in law, for the Skorohod topology, towards the law of the unique strong solution of the following Cox Ingersoll Ross stochastic differential equation on $[0, 1]$:*

$$X_t = \int_0^t (\mu - X_s)\frac{\lambda}{m} ds + \frac{\sqrt{\lambda}}{m} \int_0^t \sqrt{X_s} dB_s.$$

Furthermore, the sequence of renormalized Hawkes process $V_t^T = (1 - a_T)N_{tT}^T / T$ converges in law, for the Skorohod topology, towards the integrated CIR process

$$\int_0^t X_s ds, \ t \in [0, 1].$$

This result implies that when the kernel's norm of a Hawkes process $\|\phi\|_1$ is close to 1, if one observes the process at the right time scale T (that is T of order $1/(1 - \|\phi\|_1)$), a non degenerate behavior (neither explosive, nor deterministic) can be obtained for a rescaled Hawkes process. More precisely, it shows that at this time scale, a nearly unstable Hawkes process asymptotically behaves as an integrated CIR process.

2.3 Heavy tailed Hawkes processes

The problem with the above result is that, in practice, the intensity of the flow of market orders does not really behave as a CIR process. In particular the auto-covariance function of the intensity is not exponentially decaying in time. Furthermore, the light tail assumption of the previous paragraph does not correspond to the data. Indeed, we observe some nearly unstable heavy Hawkes processes whose kernel's shape behaves when x tends to infinity as $1/x^{1+\alpha}$ with $\alpha \in]0, 1[$, see [BM14a] or [HBB13]. In Chapter III, we study the behavior of such heavy tailed Hawkes processes.

2.3.1 Asymptotic framework

We proceed in a framework close to that of Chapter II except that, instead of the light tail assumption, we now assume that the kernel's shape behaves as $1/x^{1+\alpha}$.

Assumption 3. *There exists $\delta > 0$ such that*

$$\lim_{x \to +\infty} \alpha x^{\alpha} \int_x^{+\infty} \phi(t) dt = \frac{\alpha \delta}{\Gamma(1-\alpha)}$$

where Γ is the Gamma function.

The link between the observation scale and the kernel's norm also needs to be modified if we want to obtain suitable limit laws as before. We thus make the following assumption:

Assumption 4. *There exists $\lambda > 0$ such that*

$$\lim_{T \to +\infty} T^{\alpha}(1 - a_T) = \lambda \delta.$$

2.3.2 The theorem

In Chapter III, we show the following result about the behavior of the renormalized heavy tailed nearly unstable Hawkes process

$$X_t^T = \frac{1 - a_T}{T \mu_T} N_{Tt}^T.$$

Result 5. *Under Assumptions 3 and 4*[1]*, X^T is tight and if X is a limiting law of X^T, there exists a Brownian motion B such that*

$$X_t = \int_0^t s f^{\alpha,\lambda}(t-s) ds + \frac{1}{\sqrt{\mu^* \lambda}} \int_0^t f^{\alpha,\lambda}(t-s) B_{X_s} ds \tag{2}$$

where $f^{\alpha,\lambda}$ is the function whose Laplace transform satisfies

$$\hat{f}^{\alpha,\lambda}(z) = \frac{\lambda}{\lambda + z^{\alpha}}.$$

We will see in Part II that $f^{\alpha,\lambda}$ can be expressed with Mittag-Leffler functions and that it asymptotically behaves as

$$f^{\alpha,\lambda}(t) \underset{t \to 0}{\sim} \frac{c}{x^{1-\alpha}} \quad \text{and} \quad f^{\alpha,\lambda}(t) \underset{t \to +\infty}{\sim} \frac{c'}{x^{1+\alpha}}.$$

The integral equation above is quite hard to interpret. When $\alpha > 1/2$, we can give a more natural formulation of the limiting behavior of nearly unstable Hawkes processes as integrated

[1] In fact we also need the following technical assumption about the sequence of exogenous intensities: There exists $\mu^* > 0$ such that $T^{1-\alpha} \mu^T \to \mu^* \delta^{-1}$.

fractional diffusions. Let (X_t) be a process satisfying (2) for $t \in [0,1]$ and assume $\alpha > 1/2$. Then X is differentiable on $[0,1]$ and the law of its derivative Y is a weak solution of the stochastic integral equation

$$Y_t = F^{\alpha,\lambda}(t) + \frac{1}{\sqrt{\mu^* \lambda}} \int_0^t f^{\alpha,\lambda}(t-s)\sqrt{Y_s} dB_s^1,$$

with B^1 a Brownian motion. Recall that the kernel $f^{\alpha,\lambda}(t-s)$ appearing in the equation for Y_t behaves in zero as that enabling to define a fractional Brownian motion with Hurst parameter $\alpha - 1/2$ from a Brownian motion as in the Mandelbroot-Van-Ness representation, see (1). This implies that Y locally behaves as such a fractional Brownian motion. In particular, for any $\varepsilon > 0$, Y has Hölder regularity $\alpha - 1/2 - \varepsilon$.

This result implies that modeling the order flow as a heavy tailed nearly unstable Hawkes process and assuming that the cumulated order flow is proportional to the cumulated variance leads to a microstructural foundation of the regularity of rough volatility models such as the RFSV model of Part I.

2.4 Nearly unstable autoregressive processes

Because of the linearity in the definition of Hawkes processes, they have very similar properties as autoregressive processes. In Chapter IV, we use the ideas of the proofs for nearly unstable Hawkes processes to extend classical results about nearly unstable autoregressive processes to the infinite order case.

Somehow as in the Hawkes case, we consider a sequence of autoregressive processes

$$y_k^n = \sum_{i=1}^{k} \phi_i^n y_{k-i}^n + \varepsilon_k$$

where the sequence of kernels has a fixed positive shape and a norm which tends to one as the observation scale n tends to infinity

$$\phi_k^n = a_n \phi_k.$$

We obtain that, in the light tail case where the shape of the kernel has an average $\sum_{k\geq 1} k\phi_k = m < +\infty$, nearly unstable autoregressive processes asymptotically behave as an Ornstein-Uhlenbeck process, as it is the case for finite order autoregressive processes, see [CS82] or [vdMPvZ99].

We now consider the heavy tail case where we make the following assumptions:

Assumption 5.

$$\phi_k \underset{k\to+\infty}{\sim} \frac{\alpha\delta}{\Gamma(1-\alpha)k^{1+\alpha}}.$$

Assumption 6.

$$\lim_{n\to+\infty} n^{\alpha}(1-a_n) = \lambda\delta.$$

We obtain the following result about the behavior of the renormalized cumulated autoregressive process

$$Z_t^n = \frac{1-a_n}{\sqrt{n}}\Big(\sum_{j=0}^{\lfloor nt\rfloor} y_j^n + (nt - \lfloor nt\rfloor)y_{\lfloor nt\rfloor+1}^n\Big).$$

Result 6. *Under Assumptions 5 and 6, Z^n converges in law for the Skorohod topology towards the process*

$$Z_t = \int_0^t F^{\alpha,\lambda}(t-s)dW_s$$

where $F^{\alpha,\lambda}(t) = \int_0^t f^{\alpha,\lambda}(s)ds$.

As for Hawkes processes, for $\alpha > 1/2$, this result can be written in terms of the derivative of Z which locally behaves as a fractional Brownian motion with Hurst index $H = \alpha - 1/2$. Indeed, if $\alpha > 1/2$, Z is differentiable and its derivative Y satisfies

$$Y_t = \int_0^t f^{\alpha,\lambda}(t-s)dW_s.$$

2.5 Beyond long memory

There are three distinct regimes for this limiting law Z depending on α:

- If $\alpha < 1/2$, then the process presents a power law long memory.
- If $\alpha > 1/2$, then the process behaves as an integrated rough diffusion.
- If $\alpha = 1/2$, then the process enjoys the same correlation structure as the log-volatility of the BDM model.

Therefore, modeling the log-volatility as a nearly unstable autoregressive process with a power law kernel allows us to retrieve the long memory of [CR98], the multifractal behavior of [BKM08] and the rough volatility of Chapter I, depending on the value of α.

3 Part III: Multiscale Hawkes estimation

Chapters V and VI focus on the estimation of multidimensional Hawkes processes whose kernels decrease slowly as it is observed on financial data. In Chapter V, we introduce a similar estimation method as in [BM14b] which consists in estimating the conditional laws of the point processes and then numerically inverting the Wiener-Hopf system. In Chapter VI, we review convex optimization techniques that are used to parametrically maximize the Hawkes likelihood and we adapt these techniques to slowly decreasing Hawkes kernels.

3.1 Hawkes processes and causality

In order to model the joint dynamics of several point processes (for example ask and bid market orders, price moves, limit and cancel orders), we will consider the multidimensional Hawkes model where there are cross influences between the different processes. By definition a family of D point processes is a multidimensional Hawkes process if the intensities of all of its components write as linear regressions over the past of the D processes:

$$\lambda_t^i = \mu^i + \sum_{k=1}^{D} \int_0^t \phi^{ij}(t-s) dN_s^j.$$

Another way to construct Hawkes processes is to consider the following population representation, see [HO74]: There are individuals 1, ..., D. Individuals of type i arrive as a Poisson process of intensity μ^i. Every individual can have children of all types and the law of the children of type i of an individual of type j who was born or migrated in t is an inhomogeneous Poisson process of intensity $\phi^{ij}(\cdot - t)$.

This construction is nice because it yields a natural way to define and measure the causality between events in the Hawkes model. For example, $\|\phi^{ij}\|_1$ corresponds to the number of events of type i directly implied by an event of type j, $\|\phi^{ij}\|_1 \Lambda^j / \Lambda^i$ corresponds to the proportion of events of type i directly implied by events of type j and μ^i / Λ^i corresponds to the proportion of events of type i which are exogenous. However in practice, the Hawkes kernels are not directly measurable from the data and these measures of causality between the different kinds of events are thus inaccessible.

In the literature, the most popular method to estimate Hawkes processes from a realization is to consider a parametrization of Hawkes kernels (often exponential, see [BGM15], or power law, see [HBB13]) and to maximise the associated Hawkes log-likelihood, see [HDB13], [LV14] or [ZZS13]. Another approach to the estimation of Hawkes processes consists in implying in a non parametric manner the Hawkes kernels from the correlation structure of the Hawkes process by numerically resolving the Wiener-Hopf system, see [BDM12] or [BM14b]. In Chapters V and VI, we will revisit these two approaches and adapt the estimation procedures to slowly decreasing Hawkes kernels.

3.2 The Wiener-Hopf equation

An easy way to empirically quantify the cross-clustering of a multidimensional point process is to consider its conditional laws which are defined as

$$g^{ij}(t)dt = \mathbb{E}[dN_t^i | dN_0^j = 1] - \mathbb{1}_{i=j}\delta(t) - \Lambda^i dt.$$

A remarkable property of Hawkes processes established in [BDM12] and [BM14b] is that their kernels and conditional laws are related by the following Wiener-Hopf equation

$$\forall\, t > 0, \quad g^{ij}(t) = \phi^{ij}(t) + \sum_{k=1}^{D} \int_0^{+\infty} g^{ik}(t-s)\phi^{kj}(s) ds.$$

Solving this equation in ϕ amounts to disentangling the underlying causality structure (kernels) from the cross-correlation structure (conditional laws) of the processes.

It is important to notice that the following result gives a model independent value to the solution of the Wiener-Hopf equation as the best linear prediction kernel. Therefore, resolving this equation makes sense even if the underlying process is not a Hawkes process.

Result 7. *Given the conditional laws and average intensity of any stationary multidimensional point process, there exists one and only one kernel $\tilde{\phi}$ which satisfies the Wiener-Hopf Equation. Moreover, this kernel minimises the mean square error of the estimation error of the intensity over the estimators which are linear on the past:*

$$\tilde{\phi} = \underset{\phi \in L^2}{\operatorname{argmin}} \mathbb{E}[(\lambda_t - \lambda_t^{\phi})^2]$$

where $\lambda_t^{\phi} = \mu + \int_{-\infty}^{t} \phi(t-s) dN_s$ *and* $\mu = (\mathbb{I} - |\phi|)\Lambda$.

3.3 Application to slowly decreasing kernels

In [BM14b], an Hawkes estimation procedure based on the Wiener-Hopf equation is presented. It consists in two steps: First estimate piecewise constant conditional laws using a uniform time grid. Second use a Gaussian quadrature scheme to obtain a linear system in the values of the kernels at the quadrature points.

However, when the kernels and conditional laws present some slowly decreasing behavior as it is observed on financial data (we observe significant kernels from 10^{-4} to 10^{3} seconds), this scheme cannot simultaneously capture the properties of the Hawkes kernels at every time scale. In order to remedy to this issue, in Chapter V we propose a modified estimation procedure of Hawkes kernels based on log-uniform time grids.

The first modification that we make to the procedure of [BM14b] is in the estimation of the conditional laws. We propose an estimation grid which takes smaller step around zero where the conditional law is high and varies rapidly. Once we have estimated the conditional laws, we need a numerical scheme to invert the Wiener-Hopf system. Again there are technical difficulties which make inapplicable many usual schemes such as renormalized Gaussian schemes. In order to overcome these difficulties, we propose the following numerical scheme.

Consider the following time grid which is uniform on $[0, T_{min}]$ and log-uniform on $[T_{min}, T_{max}]$:

$$\{t_k\}_{1 \leq k \leq K} = [0, \delta T_{min}, 2\delta T_{min}, ..., T_{min}, T_{min}e^{\delta}, T_{min}e^{2\delta}, ..., T_{max}].$$

Then assume that the kernel is piecewise linear on all the intervals $[t_k, t_{k+1}]$:

$$\phi(t) = \phi(t_k) + \frac{t - t_k}{t_{k+1} - t_k}(\phi(t_{k+1}) - \phi(t_k))$$

Then (for the sake of simplicity, we consider the dimension 1 case), the Wiener-Hopf system at the points of the grid writes as

$$\begin{aligned} g(t_n) &= \phi(t_n) + \sum_{k=0}^{N-1} \phi(t_k) \int_{t_k}^{t_{k+1}} g(t_n - s) ds \\ &+ \sum_{k=0}^{N-1} (\phi(t_{k+1}) - \phi(t_k)) \int_{t_k}^{t_{k+1}} \frac{s - t_k}{t_{k+1} - t_k} g(t_n - s) ds \end{aligned}$$

Result 8. *As a result, we get the following linear system in* $(\phi(t_k))_k$ *that we can easily solve by inverting the corresponding matrix.*

$$\begin{aligned} g(t_n) &= \phi(t_n) + \sum_{k=0}^{N-1} \phi(t_k) \int_{t_n - t_{k+1}}^{t_n - t_k} g(u) du \\ &+ \sum_{k=0}^{N-1} \frac{\phi(t_{k+1}) - \phi(t_k)}{t_{k+1} - t_k} \int_{t_n - t_{k+1}}^{t_n - t_k} (t_n - t_k - u) g(u) du. \end{aligned}$$

We test this procedure on numerical simulations of slowly decreasing Hawkes kernels and we show that it always (with "reasonable" realization sizes) allows us to retrieve the real kernels.

3.4 The Hawkes order book model

As a first application of this procedure, we consider the following 8-dimensional point process

$$N_t = (P_t^{(a)}, P_t^{(b)}, T_t^{(a)}, T_t^{(b)}, L_t^{(a)}, L_t^{(b)}, C_t^{(a)}, C_t^{(b)})$$

where:

- $P^{(a)}$ (resp. $P^{(b)}$) counts the number of upward (resp. downward) mid-price moves.
- $T^{(a)}$ (resp. $T^{(b)}$) counts the number of market orders at the ask (resp. bid) that do not move the price.
- $L^{(a)}$ (resp. $L^{(b)}$) counts the number of limit orders at the ask (resp. bid) that do not move the price.
- $C^{(a)}$ (resp. $C^{(b)}$) counts the number of cancel orders at the ask (resp. bid) that do not move the price.

We then use the causal interpretation of Hawkes processes to interpret our solution as a measure of the causality between events. The main empirical result of this study is the following.

Result 9. *Both for a large tick asset (the Bund) and a small tick asset (the Dax), the largest kernels are on the diagonal for the orders (ask or bid market, limit and cancel orders mostly imply orders of the same kind) and anti diagonal for the price moves (upward price moves mostly imply downward price moves and vice versa).*

This implies that although the order flows are persistent and impact the price, the "anti excitation" of price moves implies that the price is somehow diffusive.

3.5 The likelihood of Hawkes processes

In Chapter VI, we approach Hawkes estimation via likelihood maximization.

Given the jump times $((t_k^i)_{1\leq k\leq N^i})_{1\leq i\leq D}$ of the D processes, the log-likelihood of a Hawkes process of can we written as:

$$L = \sum_{i=1}^{D}\sum_{k=1}^{N^i} \log(\mu^i + \sum_{j=1}^{D}\int_0^{t_k^i} \phi^{ij}(s)dN_s^j) - \sum_{i=1}^{D}\int_0^T (\mu^i + \sum_{j=1}^{D}\int_0^t \phi^{ij}(s)dN_s^j)dt.$$

To solve the likelihood maximization, one then needs to chose a parametrization of the kernels. In the literature, the most popular parametrization is the fixed-scale exponential one where the kernels write as

$$\phi^{ij}(t) = a^{ij}\beta e^{-\beta t}.$$

The numerical maximization of this likelihood can be done using a Newton method, see [LV14]. However, when the Hawkes dimension becomes large, the Hessian cannot be inverted and one thus needs to use other methods. A recent class of methods which has become very popular over the last few years is the Majoration Minimization method, see [LM11] and [ZZS13], which can be applied to a very wide range of optimization frameworks including non parametric and marked ones.

3.5.1 A stochastic gradient ascent

In Chapter VI, we notice that, for the exponential parametrization, after having precomputed some weights, we can apply a stochastic gradient ascent to the likelihood maximization. Indeed the log-likelihood writes as a sum of almost independent and identically distributed, terms:

$$L(a,\mu) = \sum_{i=1}^{D}\sum_{k=1}^{N^i}\left[\log\left(\mu^i + \sum_{j=1}^{D} a^{ij}g^{ij}(t_k^i)\right) - \mu^i(t_k^i - t_{k-1}^i) - \sum_{j=1}^{D} a^{ij}\left(G^{ij}(t_k^i) - G^{ij}(t_{k-1}^i)\right)\right]$$

where

$$g^{ij}(t) = \sum_{t_{k'}^j < t} \beta e^{-\beta(t-t_{k'}^j)}$$

and

$$G^{ij}(t) = \int_0^t g^{ij}(s)ds$$

can be pre-computed efficiently at the appropriate points.

Result 10. *As a result, we can apply the following stochastic gradient ascent to likelihood maximization.*

$n \leftarrow 0$
while $Converged = False$ ***do***
  $n \leftarrow n+1$
  Choose at random (i,k) *in* $1,...,D$ *and* $1,...,N^i$

Compute $s = \mu^i + \sum_{j=1}^{D} a^{ij} g^{ij}(t_k^i)$
$\mu^i \leftarrow \mu^i + \gamma_n(\frac{1}{s} - (t_k^i - t_{k-1}^i))$
for $j \leftarrow 1$ *to* D ***do***
$a^{ij} \leftarrow a^{ij} - \gamma_n(\frac{g^{ij}(t_k^i)}{s} - (G^{ij}(t_k^i) - G^{ij}(t_{k-1}^i)))$
end for
end while

We will see that if this algorithm does not necessarily give better results than other methods, it allows us to use the literature on stochastic gradient like methods. In particular, we will see that an SVRG like method, see [JZ13], gives estimations that are consistently faster than the usual MM method.

3.5.2 Non parametric likelihood maximization

The reason why exponential kernels are so popular in the literature is that they make many computations very easy. In particular, the functions g and G can be computed in $O(N)$ while in the general case, these computations take $O(N^2)$. However, they also have drawbacks. Indeed, to use the exponential parametrization one needs to know and fix the scale of the kernel $1/\beta$. Moreover, there is no reason why the kernel should be exponential in practice. In particular, we have seen that this was not the case in finance. In order to cover more general kernel shapes as for example slowly decreasing kernels, we consider the following representation.

Result 11. *We will write the kernels as sums exponentials of different time scales*

$$\phi^{ij}(x) = \sum_{b=0}^{N_b} a_b^{ij} l^b \beta e^{-l^b \beta x}.$$

Notice that this maximization problem remains convex in (a_b^{ij}), see [LV14], and we can thus apply a similar stochastic gradient ascent algorithm as above to find the optimal (a_b^{ij}).

4 Part IV: Market impact throughout time scales

The two final chapters of the thesis focus on the impact of market orders and metaorders on the price at different time scales. We will see that general model independent assumptions enable us to get strong relationships between the structure of the order flow and the price dynamics.

4.1 The impact of trades and metaorders

We begin by reviewing some features of the impact of two kinds of orders: market orders and metaorders.

4.1.1 The impact of market orders

In order to exchange stocks, participants can use two types of orders on modern financial markets. Limit orders, which allow to buy or sell a quantity of assets at a fixed price, but with no guarantee of execution, or market orders, which immediately consume the available liquidity, irrespectively of the price level.

Understanding the dynamics of the order book which arises from these two types of orders is necessary for both practitioners, who wish to use their orders as effectively as possible, and regulators who aim at making markets as "efficient" as possible. The effects of the different order book events have thus been the subject of many empirical studies such as [EBK12] and [CKS14] which consider the impact on the price of market, limit and cancel orders.

An important property of market data is that while trades are persistent, that is, "if there has been a lot of buy trades in the past there will be a lot of buy trades in the future", prices are martingale. The impact of the order flow on the price thus cannot be "trivial". To solve this paradox, Lillo et al. [LF04] and Bouchaud et al. [BGPW04] proposed two formulations of the same model, called the propagator model, which solves this puzzle in discrete time and that we will explicit later.

Another important model, that we explain in detail in Chapter VII, for the impact of market orders is the MRR model [MRR97]. In the MRR model, the autocorrelation of trade signs is reproduced by modeling the sign dynamics as an $AR(1)$-type process and the impact of trades on a fundamental fair price is defined so that "the impact of trades is proportional to their surprise in the order flow". Moreover, somehow as in [GM85], the bid and ask prices are then set so that the average ex post gains of limit orders are equal to zero. As we will see, it is quite the case in practice.

The MRR model is thus remarkable in two ways: First, as in the propagator model, the order flow is persistent and yet the price is a martingale. Second, market makers do not make any profit on average. Our motivation in Chapters VII and VIII will be to build more realistic models which satisfy the same kind of properties as the MRR and the propagator models.

4.1.2 The impact of metaorders

Let us now introduce the concept of metaorders. In practice, when large investors want to buy or sell stocks, the volume that they wish to execute is typically much larger than the liquidity that is available in the order book. They thus need to split their orders into small transactions (market or limit orders) that they execute over a given period of time. The set of all these transactions is called a metaorder. Note that if metaorders are not directly observable, they can be loosely inferred from the flow of market orders, see [LMF05] or [TLF10].

Throughout the execution of a metaorder, prices move on average in the same direction as the order (buy metaorders imply that the price will go up on average). The importance of

describing the properties of this impact of metaorders on the price has led to many empirical studies, see for example [MVM+09] or [WG13]. These works consistently obtain a concave impact function, often described as a power law or square root function of the volume. This is sometimes referred to as the square root law.

From the theoretical point of view, let us mention two deeply different frameworks which have been developed over the last few years to explain the shape of the impact function.

- In [FGLW13], a martingale hypothesis and a fair pricing condition are introduced and explained using arguments about the perfect competition between market makers and investors. These two assumptions are used to derive relationships between the distribution of the size of metaorders and the shape of the market impact. In particular, it is shown that if the metaorder's sizes have a power law distribution of exponent $2+\gamma$ as it is observed in practice, the impact function behaves as V^γ with respect to the volume V of the metaorder. This corresponds to the empirical square root law for $\gamma = 1/2$.

- In [TLD+11], an alternative approach is taken where the square root law is reproduced by considering a "V-shaped" latent liquidity. More recently, see [DBMB14], it has been shown that such a latent liquidity could be implied by reaction-diffusion type models and that these models reproduced many stylized facts about the impact of metaorders.

4.2 Fair markets and the fair price

In Chapter VII, we define fair markets as market dynamics where the average profit of infinitesimal market making strategies is equal to zero. In such markets, we will see that if we define na_t (resp. nb_t) as the price of the next ask (resp. bid) market order after t, we have the following result.

Result 12. *The conditional expectations at time t of na_t and nb_t are equal and it is thus natural to define the fair price P of the market as this expectation:*

$$P_t = \mathbb{E}[na_t|\mathbb{F}_t] = \mathbb{E}[nb_t|\mathbb{F}_t].$$

We argue that this fair price is the natural model independent generalization of the fair price in the MRR model. Indeed, the average ex post gain of limit orders with respect to this fair price must be equal to zero in fair markets.

Result 13. *In particular, the bid and ask prices are*

$$a_t = P_t + AMI_t^{ask} \text{ and } b_t = P_t + AMI_t^{bid}$$

where AMI^{ask} and AMI^{bid} are the average market impacts of ask and bid market orders (v_t^A and v_t^B respectively correspond to the volume of ask and bid market orders in t) on the fair price:

$$AMI_t^{ask} = \mathbb{E}[P_{t^+}|v_t^A > 0, \mathscr{F}_{t^-}] - P_t \text{ and } AMI_t^{bid} = \mathbb{E}[P_{t^+}|v_t^B > 0, \mathscr{F}_{t^-}] - P_t.$$

This enables us to link the spread with the average market impact of orders without model specifications, somehow as in [WBK+08].

We show that even though we cannot really measure the fair price on data without a model, it is possible to verify that our relations about the average ex post gain of limit orders hold. We check that for all stocks, the results are within the theoretical error margin.

These results about the apparent validity of the fair market hypothesis seem to show that, as it is argued in [WBK+08], one does not need to introduce trade costs or the risk aversion of market makers to explain the size of the spread as it is sometimes done. Indeed the impact of market orders and the tick size are coherent with the bid and ask prices. This work also gives conditions on how the market impact should be related to the bid-ask dynamics in reasonable models.

4.3 A model independent impact equation

In Chapter VIII, we tackle the question of the impact of a persistent flow of market orders on a martingale price.

4.3.1 Market impact and volume anticipation

We consider the issue of what can be said about the impact of the order flow on the price by only making two general assumptions on the market dynamics, both being derived from no arbitrage arguments:

- The price is a martingale.
- The permanent impact of metaorders is a linear function of their volume.

We will see that from these two assumptions, we can deduce a strong model independent relationship between the structure of the order flow and its impact on the price. More precisely, we show the following result.

Result 14. *Price moves are proportional to the variations in the anticipation by the market of the cumulated order flow imbalance:*

$$P_t = P_0 + \kappa \lim_{s \to +\infty} \mathbb{E}[V_s^a - V_s^b | \mathscr{F}_t]$$

where V_s^a (resp. V_s^b) is the cumulated volume of ask (resp. bid) market orders from 0 to s.

This equation, which can be seen as a model independent generalization of the propagator model, allows one to closely relate the order flow and the price in a way such that the order flow can be persistent and yet the price remains a martingale.

4.3.2 Application to Hawkes processes

We then apply this equation to the case where the order flows at the ask and at the bid follow two independent Hawkes models. We use this example because the Hawkes model is a nice tractable way to fit the persistence in the order flow. In that case, we have the following result.

Result 15. *If the bid and ask order flows are two independent Hawkes processes N^b and N^a, then the price process follows some kind of continuous time propagator model where the impact of trades does not depend on the past of the market but decreases in time:*

$$P_t = P_0 + \int_0^t \zeta(t-s)(dN_s^a - dN_s^b)$$

where the kernel ζ is related to the Hawkes kernel ϕ by

$$\zeta(t) = \kappa v \frac{1}{1-\|\phi\|_1}((1-\|\phi\|_1) + \int_x^{+\infty} \phi(s)ds).$$

Moreover, when the Hawkes processes are heavy tailed and nearly unstable, we have the following power law result.

Result 16. *If the kernel's shape behaves as $1/x^{1+\alpha}$ as x tends to infinity, the impact function (defined as the average price move during the execution of a metaorder as a function of the time t) asymptotically behaves as $t^{1-\alpha}$ throughout the execution.*

From this property, we can (in an admittedly caricatural way) link most of the main results of this thesis going backward. Indeed, under the assumptions of linear permanent impact and concave impact function close to square root, this result from Part IV tells us that the right way to use Hawkes processes for order flow modelling is to take them nearly unstable with power law kernel with tail index α around $1/2$ (which somehow corresponds to the multifractal regime). From Part II, this implies a rough fractional behavior of the volatility with a Hurst exponent equal to $\alpha - 1/2$. As shown in Part I, this is precisely what is observed on the data.

Part I

How regular and persistent is really volatility?

CHAPTER I

Volatility is rough

Abstract

Estimating volatility from recent high frequency data, we revisit the question of the smoothness of the volatility process. Our main result is that log-volatility behaves essentially as a fractional Brownian motion with Hurst exponent H of order 0.1, at any reasonable time scale. This leads us to adopt the fractional stochastic volatility (FSV) model of Comte and Renault [CR98]. We call our model Rough FSV (RFSV) to underline that, in contrast to FSV, $H < 1/2$. We demonstrate that our RFSV model is remarkably consistent with financial time series data; one application is that it enables us to obtain improved forecasts of realized volatility. Furthermore, we find that although volatility is not long memory in the RFSV model, classical statistical procedures aiming at detecting volatility persistence tend to conclude the presence of long memory in data generated from it. This sheds light on why long memory of volatility has been widely accepted as a stylized fact. Finally, we provide a quantitative market microstructure-based foundation for our findings, relating the roughness of volatility to high frequency trading and order splitting.

Keywords: High frequency data, volatility smoothness, fractional Brownian motion, fractional Ornstein-Uhlenbeck, long memory, volatility persistence, volatility forecasting, option pricing, volatility surface, Hawkes processes, high frequency trading, order splitting.

1 Introduction

1.1 Volatility modeling

In the derivatives world, log-prices are often modeled as continuous semi-martingales. For a given asset with log-price Y_t, such a process takes the form

$$dY_t = \mu_t dt + \sigma_t dW_t,$$

where μ_t is a drift term and W_t is a one-dimensional Brownian motion. The term σ_t denotes the volatility process and is the most important ingredient of the model. In the Black-Scholes framework, the volatility function is either constant or a deterministic function of time. In

Dupire's local volatility model, see [Dup94], the local volatility $\sigma(Y_t, t)$ is a deterministic function of the underlying price and time, chosen to match observed European option prices exactly. Such a model is by definition time-inhomogeneous; its dynamics are highly unrealistic, typically generating future volatility surfaces (see Section 1.3 below) completely unlike those we observe. A corollary of this is that prices of exotic options under local volatility can be substantially off-market. On the other hand, in so-called stochastic volatility models, the volatility σ_t is modeled as a continuous Brownian semi-martingale. Notable amongst such stochastic volatility models are the Hull and White model [HW93], the Heston model [Hes93], and the SABR model [HKLW02]. Whilst stochastic volatility dynamics are more realistic than local volatility dynamics, generated option prices are not consistent with observed European option prices. We refer to [Gat06] and [MR06] for more detailed reviews of the different approaches to volatility modeling. More recent market practice is to use local-stochastic-volatility (LSV) models which both fit the market exactly and generate reasonable dynamics.

1.2 Fractional volatility

In terms of the smoothness of the volatility process, the preceding models offer two possibilities: very regular sample paths in the case of Black-Scholes, and volatility trajectories with regularity close to that of Brownian motion for the local and stochastic volatility models. Starting from the stylized fact that volatility is a long memory process, various authors have proposed models that allow for a wider range of regularity for the volatility. In a pioneering paper, Comte and Renault [CR98] proposed to model log-volatility using fractional Brownian motion (fBM for short), ensuring long memory by choosing the Hurst parameter $H > 1/2$. A large literature has subsequently developed around such fractional volatility models, for example [CKM03, CCR12, Ros08].

The fBM $(W_t^H)_{t\in\mathbb{R}}$ with Hurst parameter $H \in (0,1)$, introduced in [MVN68], is a centered self-similar Gaussian process with stationary increments satisfying for any $t \in \mathbb{R}$, $\Delta \geq 0$, $q > 0$:

$$\mathbb{E}[|W_{t+\Delta}^H - W_t^H|^q] = K_q \Delta^{qH}, \tag{1}$$

with K_q the moment of order q of the absolute value of a standard Gaussian variable. For $H = 1/2$, we retrieve the classical Brownian motion. The sample paths of W^H are Hölder-continuous with exponent r, for any $r < H$[1]. Finally, when $H > 1/2$, the increments of the fBM are positively correlated and exhibit long memory in the sense that

$$\sum_{k=0}^{+\infty} \mathrm{Cov}[W_1^H, W_k^H - W_{k-1}^H] = +\infty.$$

Indeed, $\mathrm{Cov}[W_1^H, W_k^H - W_{k-1}^H]$ is of order k^{2H-2} as $k \to \infty$. Note that in the case of the fBM, there is a one to one correspondence between regularity and long memory through the Hurst

[1] Actually H corresponds to the regularity of the process in a more accurate way: in terms of Besov smoothness spaces, see Section 2.1.

parameter H.

As mentioned earlier, the long memory property of the volatility process has been widely accepted as a stylized fact since the seminal analyses of Ding, Granger and Engle [DGE93], Andersen and Bollerslev [AB97] and Andersen et al. [ABDL01]. Initially, it appears that the term *long memory* referred to the slow decay of the autocorrelation function (of absolute returns for example), anything slower than exponential. Over time however, it seems that this term has acquired the more precise meaning that the autocorrelation function is not integrable, see [Ber94], and even more precisely that it decays as a power-law with exponent less than 1. Much of the more recent literature, for example [BCl1, CDP06, Chrl1], assumes long memory in volatility in this more technical sense. Indeed, meaningful results can probably only be obtained under such a specification, since it is not possible to estimate the asymptotic behavior of the covariance function without assuming a specific form. Nevertheless, analyses such as that of Andersen et al. [ABDL01] use data that predate the advent of high-frequency electronic trading, and the evidence for long memory has never been sufficient to satisfy remaining doubters such as Mikosch and Stărică in [MS00b]. To quote Rama Cont in [Con07]:

> ... the econometric debate on the short range or long range nature of dependence in volatility still goes on (and may probably never be resolved)...

One of our contributions in this paper is (we believe) to finally resolve this question, showing that the autocorrelation function of volatility does not behave as a power law, at least at usual time scales of observation. This implies that when stated in term of the asymptotic behavior of the autocorrelation function, the long memory question can simply not be answered. Nevertheless, we are able to provide explicit expressions enabling us to analyze thoroughly the dependence structure of the volatility process.

1.3 The shape of the implied volatility surface

As is well-known, the implied volatility $\sigma_{\mathrm{BS}}(k,\tau)$ of an option (with log-moneyness k and time to expiration τ) is the value of the volatility parameter in the Black-Scholes formula required to match the market price of that option. Plotting implied volatility as a function of strike price and time to expiry generates the *volatility surface*, explored in detail in, for example, [Gat06]. A typical such volatility surface generated from a "stochastic volatility inspired" (SVI) [GJ14] fit to closing SPX option prices as of June 20, 2013[2] is shown in Figure I.1. It is a stylized fact that, at least in equity markets, although the level and orientation of the volatility surface do change over time, the general overall shape of the volatility surface does not change, at least to a first approximation. This suggests that it is desirable to model volatility as a time-homogenous process, *i.e.* a process whose parameters are independent of price and time.

[2]Closing prices of SPX options for all available strikes and expirations as of June 20, 2013 were sourced from OptionMetrics (www.optionmetrics.com) via Wharton Research Data Services (WRDS).

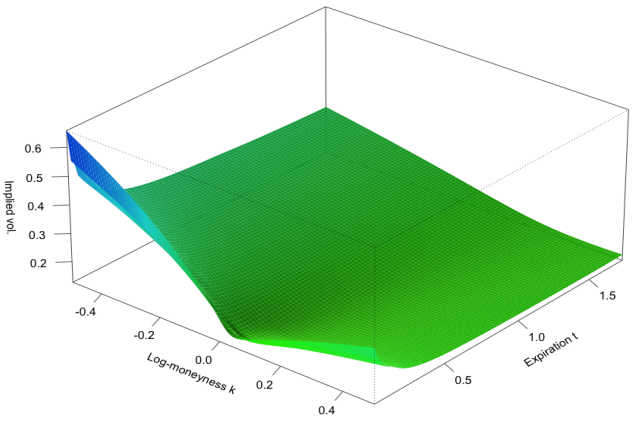


Figure I.1 – The S&P volatility surface as of June 20, 2013.

However, conventional time-homogenous models of volatility such as the Hull and White, Heston, and SABR models do not fit the volatility surface. In particular, as shown in Figure I.2, the observed term structure of at-the-money ($k = 0$) volatility skew

$$\psi(\tau) := \left| \frac{\partial}{\partial k} \sigma_{\mathrm{BS}}(k, \tau) \right|_{k=0}$$

is well-approximated by a power-law function of time to expiry τ. In contrast, conventional stochastic volatility models generate a term structure of at-the-money (ATM) skew that is *constant* for small τ and behaves as a sum of decaying exponentials for larger τ.

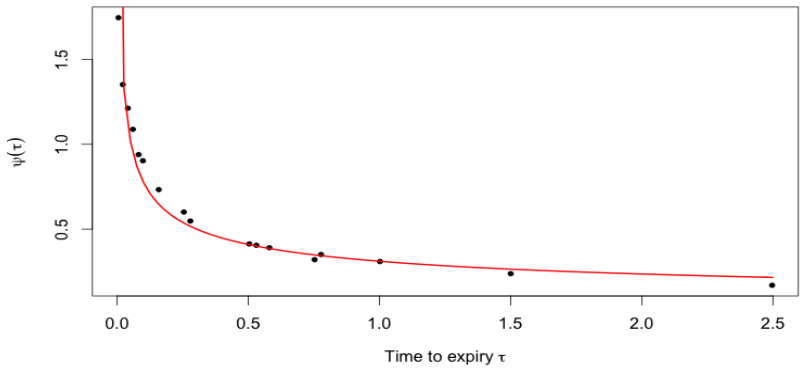


Figure I.2 – The black dots are non-parametric estimates of the S&P ATM volatility skews as of June 20, 2013; the red curve is the power-law fit $\psi(\tau) = A\tau^{-0.4}$.

In Section 3.3 of [Fuk11], as an example of the application of his martingale expansion, Fukasawa shows that a stochastic volatility model where the volatility is driven by fractional Brownian motion with Hurst exponent H generates an ATM volatility skew of the form $\psi(\tau) \sim \tau^{H-1/2}$, at least for small τ. This is interesting in and of itself in that it provides a counterexample to the widespread belief that the explosion of the volatility smile as $\tau \to 0$ (as clearly seen in Figures I.1 and I.2) implies the presence of jumps [CW03]. The main point here is that for a model of the sort analyzed by Fukasawa to generate a volatility surface with a reasonable shape, we would need to have a value of H close to zero. As we will see in Section 2, our empirical estimates of H from time series data are in fact very small.

The volatility model that we will specify in Section 3.1, driven by fBM with $H < 1/2$, therefore has the potential to be not only consistent with the empirically observed properties of the volatility time series but also consistent with the shape of the volatility surface. In this paper, we focus on the modeling of the volatility time series. A more detailed analysis of the consistency of our model with option prices is left for a future article.

1.4 Main results and organization of the paper

In Section 2, we report our estimates of the smoothness of the log-volatility for selected assets. This smoothness parameter lies systematically between 0.08 and 0.2 (in the sense of Hölder regularity for example). Furthermore, we find that increments of the log-volatility are approximately normally distributed and that their moments enjoy a remarkable monofractal scaling property. This leads us to model the log of volatility using a fBM with Hurst parameter $H < 1/2$ in Section 3. Specifically we adopt the fractional stochastic volatility (FSV) model of Comte and Renault [CR98]. We call our model Rough FSV (RSFV) to underline that, in contrast to FSV, we take $H < 1/2$. We also show in the same section that the RFSV model is remarkably consistent with volatility time series data. The issue of volatility persistence is considered through the lens of the RFSV model in Section 4. Our main finding is that although the RFSV model does not have any long memory property, classical statistical procedures aiming at detecting volatility persistence tend to conclude the presence of long memory in data generated from it. This sheds new light on the supposed long memory in the volatility of financial data. In Section 5, we apply our model to forecasting volatility. In particular, we show that RFSV volatility forecasts outperform conventional AR and HAR volatility forecasts. Finally, in Section 6, we present a market microstructure explanation for the regularities we observe in the volatility process at the macroscopic scale. We show that the empirical behavior of volatility may be explained in terms of order splitting and the high degree of endogeneity of the market ascribed to algorithmic trading. Some new results related to the RFSV model are relegated to the appendix.

2 Smoothness of the volatility: empirical results

In this section we report estimates of the smoothness of the volatility process for four assets: The DAX and Bund futures contracts, for which we estimate integrated variance directly

from high frequency data using an estimator based on the model with uncertainty zones, [RR11, RR12], and the S&P and NASDAQ indices, for which we use precomputed realized variance estimates from the Oxford-Man Institute of Quantitative Finance Realized Library[3].

2.1 Estimating the smoothness of the volatility process

Let us first pretend that we have access to discrete observations of the volatility process, on a time grid with mesh Δ on $[0,T]$: σ_0, $\sigma_\Delta,\dots,\sigma_{k\Delta},\dots$, $k\in\{0,\lfloor T/\Delta\rfloor\}$. Set $N=\lfloor T/\Delta\rfloor$, then for $q\geq 0$, we define

$$m(q,\Delta)=\frac{1}{N}\sum_{k=1}^{N}|\log(\sigma_{k\Delta})-\log(\sigma_{(k-1)\Delta})|^q.$$

In the spirit of [Ros11], our main assumption is that for some $s_q>0$ and $b_q>0$, as Δ tends to zero,

$$N^{qs_q}m(q,\Delta)\to b_q. \tag{2}$$

Under additional technical conditions, Equation (2) essentially says that the volatility process belongs to the Besov smoothness space $\mathscr{B}^{s_q}_{q,\infty}$ and does not belong to $\mathscr{B}^{s'_q}_{q,\infty}$, for $s'_q>s_q$, see [Ros09a]. Hence s_q can really be viewed as the regularity of the volatility when measured in l_q norm. In particular, functions in $\mathscr{B}^{s}_{q,\infty}$ for every $q>0$ enjoy the Hölder property with parameter h for any $h<s$. For example, if $\log(\sigma_t)$ is a fBM with Hurst parameter H, then for any $q\geq 0$, Equation (2) holds in probability with $s_q=H$ and it can be shown that the sample paths of the process indeed belong to $\mathscr{B}^{H}_{q,\infty}$ almost surely. Assuming the increments of the log-volatility process are stationary and that a law of large number can be applied, $m(q,\Delta)$ can also be seen as the empirical counterpart of

$$\mathbb{E}[|\log(\sigma_\Delta)-\log(\sigma_0)|^q].$$

Of course, the volatility process is not directly observable, and an exact computation of $m(q,\Delta)$ is not possible in practice. We must therefore proxy spot volatility values by appropriate estimated values. Since the minimal Δ will be equal to one day in the sequel, we proxy the (true) spot volatility daily at a fixed given time of the day (11 am for example). Two daily spot volatility proxies will be considered:

- For our ultra high frequency intraday data (DAX future contracts and Bund future contracts[4], 1248 days from 13/05/2010 to 01/08/2014[5]), we use the estimator of the integrated variance from 10 am to 11 am London time obtained from the model with uncertainty zones, see [RR11, RR12]. After renormalization, the resulting estimates of integrated variance over very short time intervals can be considered as good proxies for the unobservable spot variance. In particular, the one hour long window on which they are computed is small compared to the extra day time scales that will be of interest here.

[3] http://realized.oxford-man.ox.ac.uk/data/download. The Oxford-Man Institute's Realized Library contains a selection of daily non-parametric estimates of volatility of financial assets, including realized variance (rv) and realized kernel (rk) estimates. A selection of such estimators is described and their performances compared in, for example, [GO10] .

[4] For every day, we only consider the future contract corresponding to the most liquid maturity.

[5] Data kindly provided by QuantHouse EUROPE/ASIA, http://www.quanthouse.com.

- For the S&P and NASDAQ indices[6], we proxy daily spot variances by daily 5 minutes[7] realized variance estimates from the Oxford-Man Institute of Quantitative Finance Realized Library (3,540 trading days from January 3, 2000 to March 31, 2014). Since these estimates of integrated variance are for the whole trading day, we expect estimates of the smoothness of the volatility process to be biased upwards, integration being a regularizing operation. We compute the extent of this bias by simulation in Section 3.4.

In the following, we retain the notation $m(q,\Delta)$ with the understanding that we are only proxying the (true) spot volatility as explained above. We now proceed to estimate the smoothness parameter s_q for each q by computing the $m(q,\Delta)$ for different values of Δ and regressing $\log m(q,\Delta)$ against $\log\Delta$. Note that for a given Δ, several $m(q,\Delta)$ can be computed depending on the starting point. Our final measure of $m(q,\Delta)$ is the average of these values.

2.2 DAX and Bund futures contracts

DAX and Bund futures are amongst the most liquid assets in the world and moreover, the model with uncertainty zones used to estimate volatility is known to apply well to them, see [DR12]. So we can be confident in the reliability of our volatility proxy. Nevertheless, as an extra check, we will confirm the quality of our volatility proxy by Monte Carlo simulation in Section 3.4.

Plots of $\log m(q,\Delta)$ vs $\log\Delta$ for different values of q, are displayed for the DAX in Figure I.3, and for the Bund in Figure I.4.

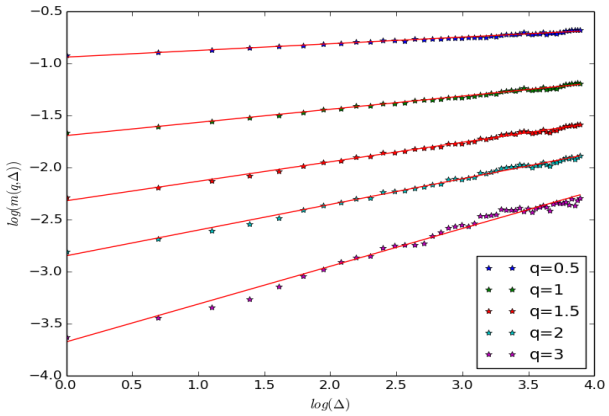


Figure I.3 – $\log m(q,\Delta)$ as a function of $\log\Delta$, DAX.

[6] And also the CAC40, Nikkei and FTSE indices in some specific parts of the paper.

[7] Taking 5 minutes time intervals to compute the realized variance because it is a good trade off to have lots of intervals every day and a small microstructure noise, see [ASMZ05].

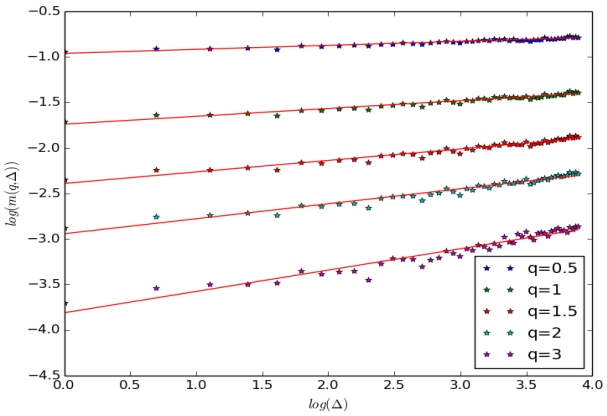


Figure I.4 – $\log m(q,\Delta)$ as a function of $\log\Delta$, Bund.

For both DAX and Bund, for a given q, the points essentially lie on a straight line. Under stationarity assumptions, this implies that the log-volatility increments enjoy the following scaling property in expectation:

$$\mathbb{E}[|\log(\sigma_\Delta) - \log(\sigma_0)|^q] = K_q \Delta^{\zeta_q},$$

where $\zeta_q > 0$ is the slope of the line associated to q. Moreover, the smoothness parameter s_q does not seem to depend on q. Indeed, plotting ζ_q against q, we obtain that $\zeta_q \sim H\,q$ with H equal to 0.125 for the DAX and to 0.082 for the Bund, see Figure I.5.

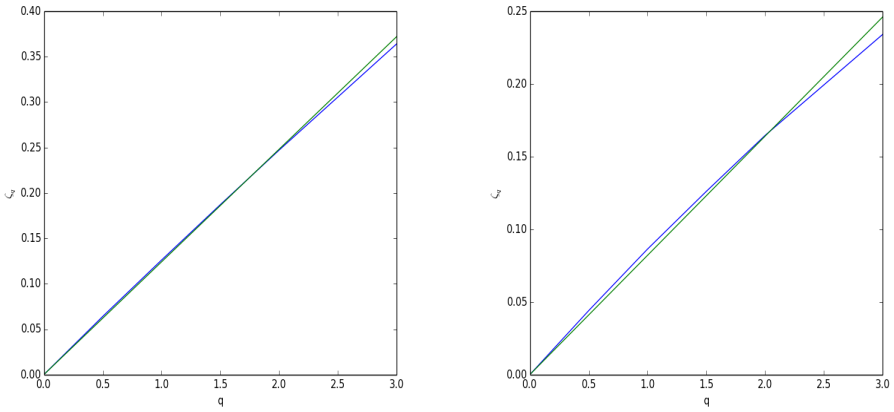


Figure I.5 – ζ_q (blue) and $0.125 \times q$ (green), DAX (left); ζ_q (blue) and $0.082 \times q$ (green), Bund (right).

We remark that the graphs for ζ_q are actually very slightly concave. However, we observe the same small concavity effect when we replace the log-volatility by simulations of a fBM with the same number of points. We conclude that this effect relates to finite sample size and is thus not significant.

2.3 S&P and NASDAQ indices

We report in Figure I.6 and Figure I.7 similar results for the S&P and NASDAQ indices. The variance proxies used here are the precomputed 5-minute realized variance estimates for the whole trading day made publicly available by the Oxford-Man Institute of Quantitative Finance.

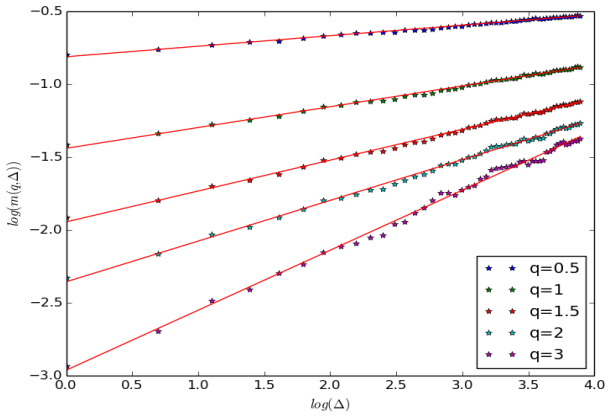


Figure I.6 – $\log m(q,\Delta)$ as a function of $\log\Delta$, S&P.

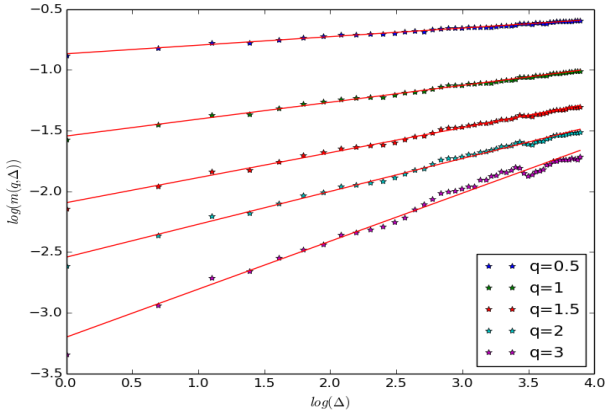


Figure I.7 – $\log m(q,\Delta)$ as a function of $\log(\Delta)$, NASDAQ.

We observe the same scaling property for the S&P and NASDAQ indices as we observed for DAX and Bund futures and again, the s_q do not depend on q. However, the estimated smoothnesses are slightly higher here: $H = 0.142$ for the S&P and $H = 0.139$ for the NASDAQ, see Figure I.8.

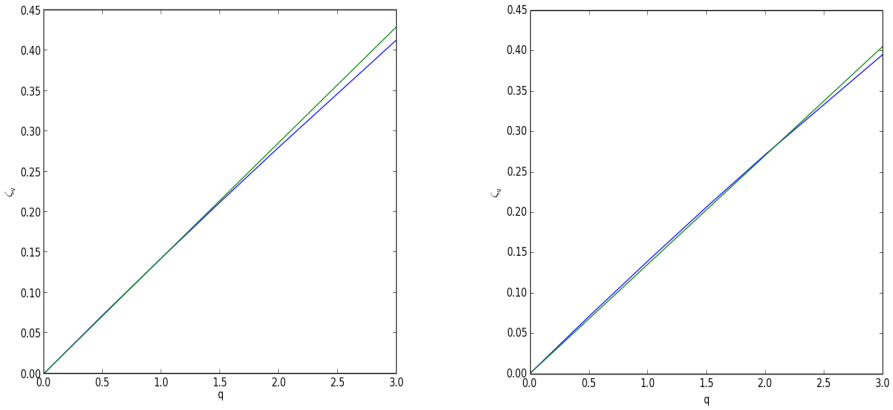


Figure I.8 – ζ_q (blue) and $0.142 \times q$ (green), S&P (left); ζ_q (blue) and $0.139 \times q$ (green), NASDAQ (right).

Once again, we do expect these smoothness estimates to be biased high because we are using whole-day realized variance estimates, as explained earlier in Section 2. Finally, we remark that as for DAX and Bund futures, the graphs for ζ_q are slightly concave.

2.4 Other indices

Repeating the analysis of Section 2.3 for each index in the Oxford-Man dataset, we find the $m(q, \Delta)$ present a universal scaling behavior. For each index and for $q = 0.5,\ 1,\ 1.5,\ 2,\ 3$, by doing a linear regression of $\log(m(q, \Delta))$ on $\log(\Delta)$ for $\Delta = 1, ..., 30$, we obtain estimates of ζ_q that we summarize in Table I.3 in the appendix.

2.5 Distribution of the increments of the log-volatility

Having established that all our underlying assets exhibit essentially the same scaling behavior[8], we focus in the rest of the paper only on the S&P index, unless specified otherwise. That the distribution of increments of log-volatility is close to Gaussian is a well-established stylized fact reported for example in the papers [ABDE01] and [ABDL01] of Andersen et al. Looking now at the histograms of the increments of the log-volatility in Figure I.9 with the fitted normal density superimposed in red, we see that, for any Δ, the empirical distributions of log-volatility

[8] We have also verified that this scaling relationship holds for Crude Oil and Gold futures with similar smoothness estimates ζ_q.

increments are verified as being close to Gaussian. More impressive still is that rescaling the 1-day fit of the normal density by Δ^H generates (blue dashed) curves that are very close to the red fits of the normal density, consistent with the observed scaling.

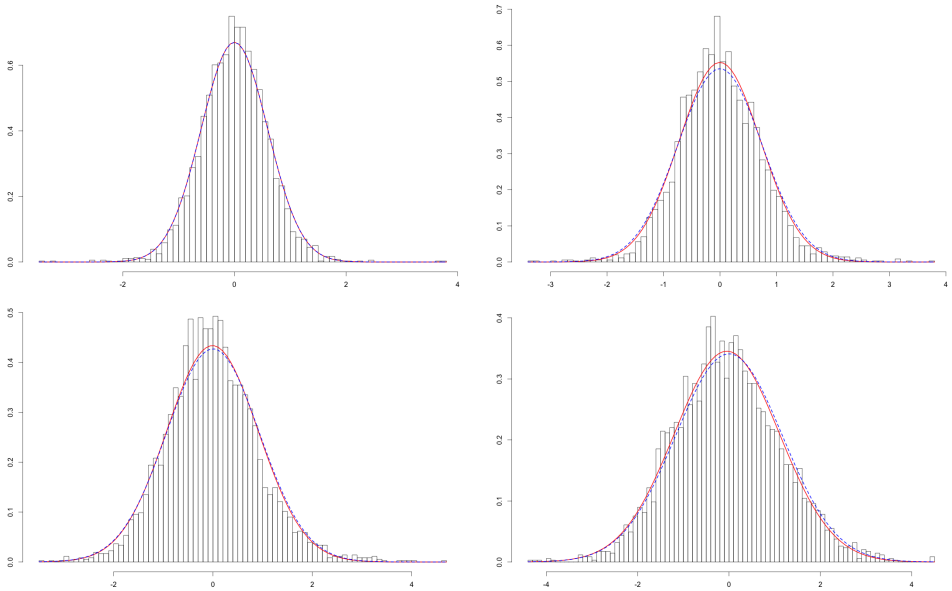


Figure I.9 – Histograms for various lags Δ (1, 5, 25 and 125 days) of the (overlapping) increments $\log\sigma_{t+\Delta} - \log\sigma_t$ of the S&P log-volatility; normal fits in red; normal fit for $\Delta = 1$ day rescaled by Δ^H in blue.

The slight deviations from the Normal distribution observed in Figure I.9 are again consistent with the computation of the empirical distribution of the increments of a fractional Brownian motion on a similar number of points.

2.6 Does H vary over time?

In order to check whether our estimations of H depends on the time interval, we split the Oxford-Man realized variance dataset into two halves and reestimate H for each half separately. The results are presented in Table I.4 in the appendix. We note that although the estimated H all lie between 0.06 and 0.20, they seem to be higher in the second period which includes the financial crisis.

3 A simple model compatible with the empirical smoothness of the volatility

In this section, we specify the Rough FSV model and demonstrate that it reproduces the empirical facts presented in Section 2.

3.1 Specification of the RFSV model

In the previous section, we showed that, empirically, the increments of the log-volatility of various assets enjoy a scaling property with constant smoothness parameter and that their distribution is close to Gaussian. This naturally suggests the simple model:

$$\log\sigma_{t+\Delta} - \log\sigma_t = \nu\left(W^H_{t+\Delta} - W^H_t\right), \tag{3}$$

where W^H is a fractional Brownian motion with Hurst parameter equal to the measured smoothness of the volatility and ν is a positive constant. We may of course write (3) under the form

$$\sigma_t = \sigma\exp\left\{\nu W^H_t\right\}, \tag{4}$$

where σ is another positive constant.

However this model is not stationary, stationarity being desirable both for mathematical tractability and also to ensure reasonableness of the model at very large times. This leads us to impose stationarity by modeling the log-volatility as a fractional Ornstein-Uhlenbeck process (fOU process for short) with a very long reversion time scale.

A stationary fOU process (X_t) is defined as the stationary solution of the stochastic differential equation

$$dX_t = \nu\, dW^H_t - \alpha\,(X_t - m)dt,$$

where $m \in \mathbb{R}$ and ν and α are positive parameters, see [CKM03]. As for usual Ornstein-Uhlenbeck processes, there is an explicit form for the solution which is given by

$$X_t = \nu\int_{-\infty}^{t} e^{-\alpha(t-s)}dW^H_t + m. \tag{5}$$

Here the stochastic integral with respect to fBM is simply a pathwise Riemann-Stieltjes integral, see again [CKM03].

We thus arrive at the final specification of our Rough Fractional Stochastic Volatility (RFSV) model for the volatility on the time interval of interest $[0, T]$:

$$\sigma_t = \exp\{X_t\},\ \ t \in [0, T], \tag{6}$$

where (X_t) satisfies Equation (5) for some $\nu > 0,\ \alpha > 0,\ m \in \mathbb{R}$ and $H < 1/2$ the measured smoothness of the volatility. Such a model is indeed stationary. However, if $\alpha \ll 1/T$, the log-volatility behaves locally (at time scales smaller than T) as a fBM. This observation is formalized in Proposition 1 below.

Proposition 1. *Let W^H be a fBM and X^α defined by (5) for a given $\alpha > 0$. As α tends to zero,*

$$\mathbb{E}\Big[\sup_{t\in[0,T]} |X^\alpha_t - X^\alpha_0 - \nu W^H_t|\Big] \to 0.$$

The proof is given in Appendix I.A.1.

Proposition 1 implies that in the RFSV model, if $\alpha \ll 1/T$, and we confine ourselves to the interval $[0, T]$ of interest, we can proceed as if the the log-volatility process were a fBM. Indeed, simply setting $\alpha = 0$ in (5) gives (at least formally) $X_t - X_s = \nu(W_t^H - W_s^H)$ and we immediately recover our simple non-stationary fBM model (3).

The following corollary implies that the (exact) scaling property of the fBM is approximately reproduced by the fOU process when α is small.

Corollary 1. *Let $q > 0,\ t > 0,\ \Delta > 0$. As α tends to zero, we have*

$$\mathbb{E}[|X_{t+\Delta}^{\alpha} - X_t^{\alpha}|^q] \to \nu^q K_q \Delta^{qH}.$$

The proof is given in Appendix I.A.2.

RFSV versus FSV

We recognize our RFSV model (6) as a particular case of the classical FSV model of Comte and Renault [CR98]. The key difference is that here we take $H < 1/2$ and $\alpha \ll 1/T$, whereas to accommodate the assumption of long memory, Comte and Renault have to choose $H > 1/2$. The analysis of Fukasawa referred to earlier in Section 1.3 implies in particular that if $H > 1/2$, the volatility skew function $\psi(\tau)$ is *increasing* in time to expiration τ (at least for small τ), which is obviously completely inconsistent with the approximately $1/\sqrt{\tau}$ skew term structure that is observed. To generate a decreasing term structure of volatility skew for longer expirations, Comte and Renault are then forced to choose $\alpha \gg 1/T$. Consequently, for very short expirations ($\tau \ll 1/\alpha$), models of the Comte and Renault type with $H > 1/2$ still generate a term structure of volatility skew that is inconsistent with the observed one, as explained for example in Section 4 of [CCR12].

In contrast, the choice $H < 1/2$ enables us to reproduce both the observed smoothness of the volatility process and generate a term structure of volatility skew in agreement with the observed one. The choice $H < 1/2$ is also consistent with what is improperly called mean reversion by practitioners, which is the fact that if volatility is unusually high, it tends to decline and if it is unusually low, it tends to increase. Finally, taking α very small implies that the dynamics of our process is close to that of a fBM, see Proposition 1. This last point is particularly important. Indeed, recall that at the time scales we are interested in, the important feature we have in mind is really this fBM like-behavior of the log-volatility.

We could no doubt have considered other stationary models satisfying Proposition 1 and Corollary 1, where log-volatility behaves as a fBM at reasonable time scales; the choice of the fOU process is probably the simplest way to accommodate this local behavior together with the stationarity property.

3.2 RFSV model autocovariance functions

From Proposition 1 and Corollary 1, we easily deduce the following corollary, where $o(1)$ tends to zero as α tends to zero.

Corollary 2. *Let $q > 0$, $t > 0$, $\Delta > 0$. As α tends to zero,*

$$\mathrm{Cov}[X_t^\alpha, X_{t+\Delta}^\alpha] = \mathrm{Var}[X_t^\alpha] - \frac{1}{2}\nu^2 \Delta^{2H} + o(1).$$

Consequently, in the RFSV model, for fixed t, the covariance between X_t and $X_{t+\Delta}$ is linear with respect to Δ^{2H}. This result is very well satisfied empirically. For example, in Figure I.10, we see that for the S&P, the empirical autocovariance function of the log-volatility is indeed linear with respect to Δ^{2H}. Note in passing that at the time scales we consider, the term $\mathrm{Var}[X_t^\alpha]$ is higher than $\frac{1}{2}\nu^2\Delta^{2H}$ in the expression for $\mathrm{Cov}[X_t^\alpha, X_{t+\Delta}^\alpha]$.

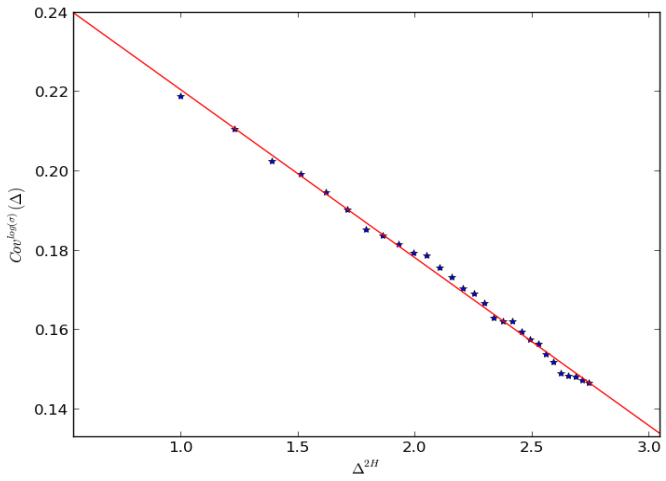


Figure I.10 – Autocovariance of the log-volatility as a function of Δ^{2H} for $H = 0.14$, S&P.

Thanks to [CKM03], we even have an exact formula for the autocovariance function of the log-volatility in the RFSV model:

$$\begin{aligned}\mathrm{Cov}[\log\sigma_t, \log\sigma_{t+\Delta}] &= \frac{H(2H-1)\nu^2}{2\alpha^{2H}}\Big\{e^{-\alpha\Delta}\Gamma(2H-1) \\ &+ e^{-\alpha\Delta}\int_0^{\alpha\Delta}\frac{e^u}{u^{2-2H}}\,du + e^{\alpha\Delta}\int_{\alpha\Delta}^{\infty}\frac{e^{-u}}{u^{2-2H}}\,du\Big\}, \qquad (7)\end{aligned}$$

and

$$\mathrm{Var}[\log\sigma_t] = \frac{H(2H-1)\nu^2}{\alpha^{2H}}\Gamma(2H-1),$$

where Γ denotes the Gamma function.

Having computed the autocovariance function of the log-volatility, we now turn our attention to the volatility itself. We have

$$\mathbb{E}[\sigma_{t+\Delta}\sigma_t] = \mathbb{E}[e^{X_t^\alpha + X_{t+\Delta}^\alpha}],$$

with X^α defined by Equation (5). Since X^α is a Gaussian process, we deduce that

$$\mathbb{E}[\sigma_{t+\Delta}\sigma_t] = e^{\mathbb{E}[X_t^\alpha]+\mathbb{E}[X_{t+\Delta}^\alpha]+\mathrm{Var}[X_t^\alpha]/2+\mathrm{Var}[X_{t+\Delta}^\alpha]/2+\mathrm{Cov}[X_t^\alpha, X_{t+\Delta}^\alpha]}.$$

Applying Corollary 2, we obtain that when α is small, $\mathbb{E}[\sigma_{t+\Delta}\sigma_t]$ is approximately equal to

$$e^{2\mathbb{E}[X_t^\alpha]+2\mathrm{Var}[X_t^\alpha]} e^{-\nu^2 \frac{\Delta^{2H}}{2}}. \tag{8}$$

It follows that in the RFSV model, $\log(\mathbb{E}[\sigma_{t+\Delta}\sigma_t])$ is also linear in Δ^{2H}. This property is again very well satisfied on data, as shown by Figure I.11, where we plot the logarithm of the empirical counterpart of $\mathbb{E}[\sigma_{t+\Delta}\sigma_t]$ against Δ^{2H}, for the S&P with $H = 0.14$.

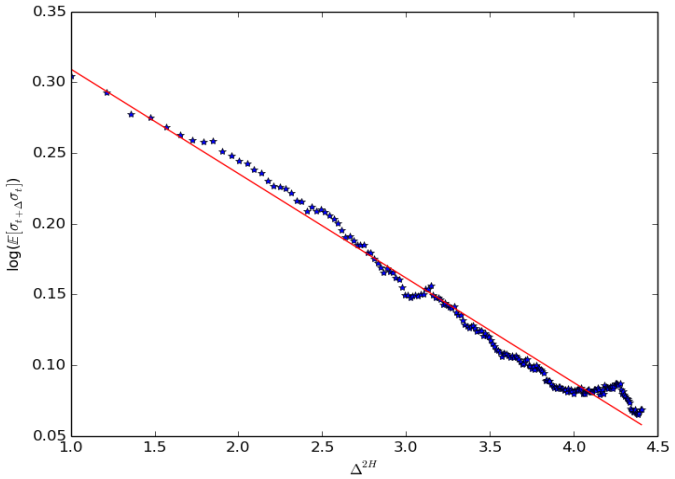


Figure I.11 – Empirical counterpart of $\log(\mathbb{E}[\sigma_{t+\Delta}\sigma_t])$ as a function of Δ^{2H}, S&P.

We note that putting Δ^{2H} on the x-axis of Figure I.11 is really crucial in order to retrieve linearity. In particular, a corollary of (8) is that the autocovariance function of the volatility does not decay as a power law as widely believed; see Figure I.12 where we show that a log-log plot of the autocovariance function does not yield a straight line.

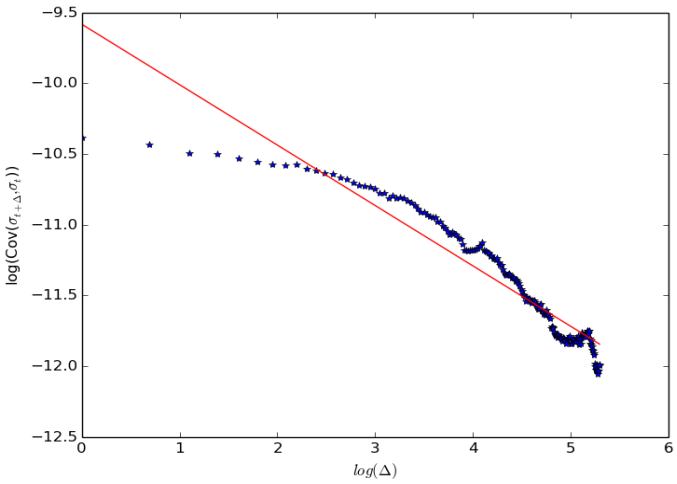


Figure I.12 – Empirical counterpart of $\log(\mathrm{Cov}[\sigma_{t+\Delta}, \sigma_t])$ as a function of $\log(\Delta)$, S&P.

3.3 RFSV versus FSV again

To further demonstrate the incompatibility of the classical long memory FSV model with volatility data, consider the quantity $m(2,\Delta)$. Recall that in the data (see Section 2) we observe the linear relationship $\log m(2,\Delta) \approx \zeta_2 \log\Delta + k$ for some constant k. Also, in both FSV and RFSV, we can consider

$$\begin{aligned} m(2,\Delta) &= \mathbb{E}\left[(\log\sigma_{t+\Delta} - \log\sigma_t)^2\right] \\ &= 2\left(\mathrm{Var}[\log\sigma_t] - \mathrm{Cov}[\log\sigma_t, \log\sigma_{t+\Delta}]\right). \end{aligned}$$

Therefore, using Equation (7), we have a closed form formula for $m(2,\Delta)$.

In Figure I.13, we plot $m(2,\Delta)$ with the parameters $H = 0.53$, corresponding to the FSV model parameter estimate of Chronopoulou and Viens in [CV12], and $\alpha = 0.5$ to ensure some visible decay of the volatility skew. The slope of $m(2,\Delta)$ in the FSV model for small lags is driven by the value of H; the lag at which $m(2,\Delta)$ begins to flatten and stationarity kicks in corresponds to a time scale of order $1/\alpha$. It is clear from the picture that to fit the data, we must have $\alpha \ll 1/T$ and the value of H must be set by the initial slope of the regression line, which as reported earlier in Section 2 is $\zeta_2 = 2 \times 0.14$.

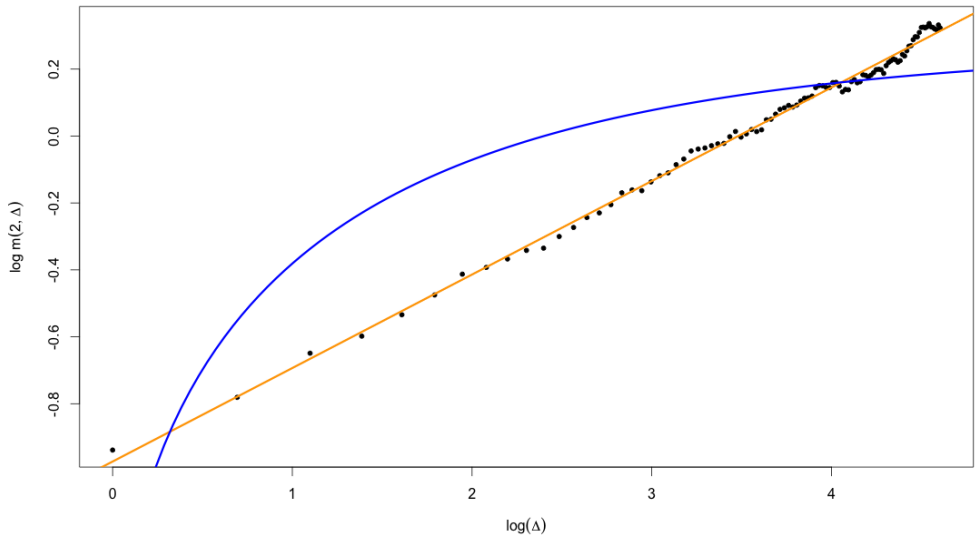


Figure I.13 – Long memory models such as the FSV model of Comte and Renault are not compatible with S&P volatility data. Black points are empirical estimates of $m(2,\Delta)$; the blue line is the FSV model with $\alpha = 0.5$ and $H = 0.53$; the orange line is the RFSV model with $\alpha = 0$ and $H = 0.14$.

3.4 Simulation-based analysis of the RFSV model

Our goal in this section is to show that in terms of smoothness measures, one obtains on simulated data from the RFSV model the same behaviors as those observed on empirical data. In particular, we would like to be able to quantify the positive bias associated with estimating H from whole-day realized variance data as in Section 2.3 relative to using data from a one-hour window as in Section 2.2.

We simulate the RFSV model for 2,000 days (chosen to be between the lengths of our two datasets). In order to account for the overnight effect, we simulate the volatility σ_t[9] and efficient price P_t[10] over the whole day. The parameters: $H = 0.14$, $\nu = 0.3$, $m = X_0 = -5$ and $\alpha = 5 \times 10^{-4}$, are chosen to be consistent with our empirical estimates from Section 2. To model microstructure effects such as the discreteness of the price grid, we consider that the observed price process is generated from P_t using the uncertainty zones model of [RR11] with tick value 5×10^{-4} and parameter $\eta = 0.25$.

[9] To simulate the fBM, we use a spectral method with 40,000,000 points (20,000 points per day). We then simulate X taking $X_{(n+1)\delta} - X_{n\delta} = \nu(W^H_{(n+1)\delta} - W^H_{n\delta}) + \alpha\delta(m - X_{n\delta})$ (with $\delta = 1/20000$).

[10] $P_{(n+1)\delta} - P_{n\delta} = P_{n\delta}\sigma_{n\delta}\sqrt{\delta}\,U_n$ where the U_n are iid standard Gaussian variables.

Exactly as in Section 2, for each of the 2,000 days, we consider two volatility proxies obtained from the observed price and based on:

- The integrated variance estimator using the model with uncertainty zones over one hour windows, from 10 am to 11 am.
- The 5 minutes realized variance estimator, over eight hours windows (the trading day).

We now repeat our analysis of Section 2, generating graphs analogous to Figures I.3, I.4, I.6 and I.7 obtained on empirical data. Figure I.14 compares smoothness measures obtained using the uncertainty zones estimator on one-hour windows with those obtained using the realized variance estimator on 8-hour windows.

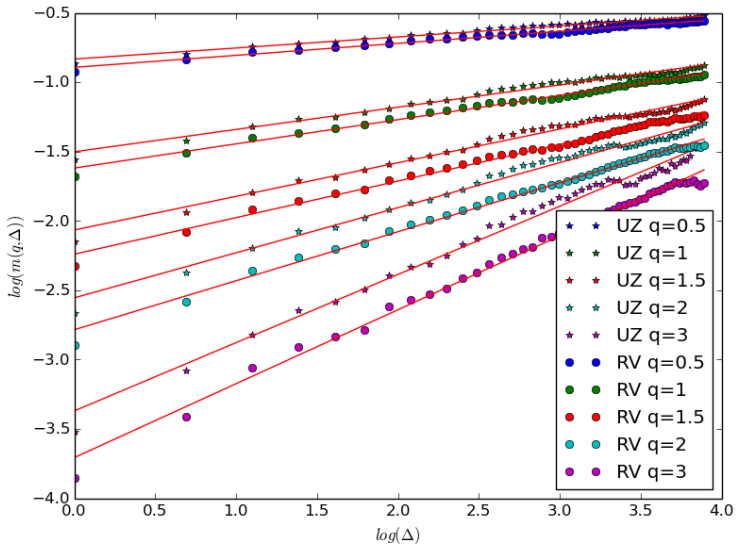


Figure I.14 – $\log(m(q,\Delta))$ as a function of $\log(\Delta)$, simulated data, with realized variance and uncertainty zones estimators.

When the uncertainty zones estimator is applied on a one-hour window (1/24 of a simulated day) as in Section 2.2, we estimate $H = 0.16$, which is close to the true value $H = 0.14$ used in the simulation. The results obtained with the realized variance estimator over daily eight-hour windows (1/3 of a simulated day) do exhibit the same scaling properties that we see in the empirical data with a smoothness parameter that does not depend on q. However, the estimated H is biased slightly higher at around 0.18. As discussed in Section 2.1, this extra positive bias is no surprise and is due to the regularizing effect of the integral operator over the longer window. We note also that the estimated values of ν ("volatility of volatility" in some sense) obtained from the intercepts of the regressions, are lower with the longer time windows, again as expected. A detailed computation of the bias in the estimated H associated

with the choice of window length in an analogous but more tractable model is presented in Appendix I.C.

We end this section by presenting in Figure I.15 a sample path of the model-generated volatility (spot volatility direct from the simulation rather than estimated from the simulated price series) together with a graph of S&P volatility over 3,500 days.

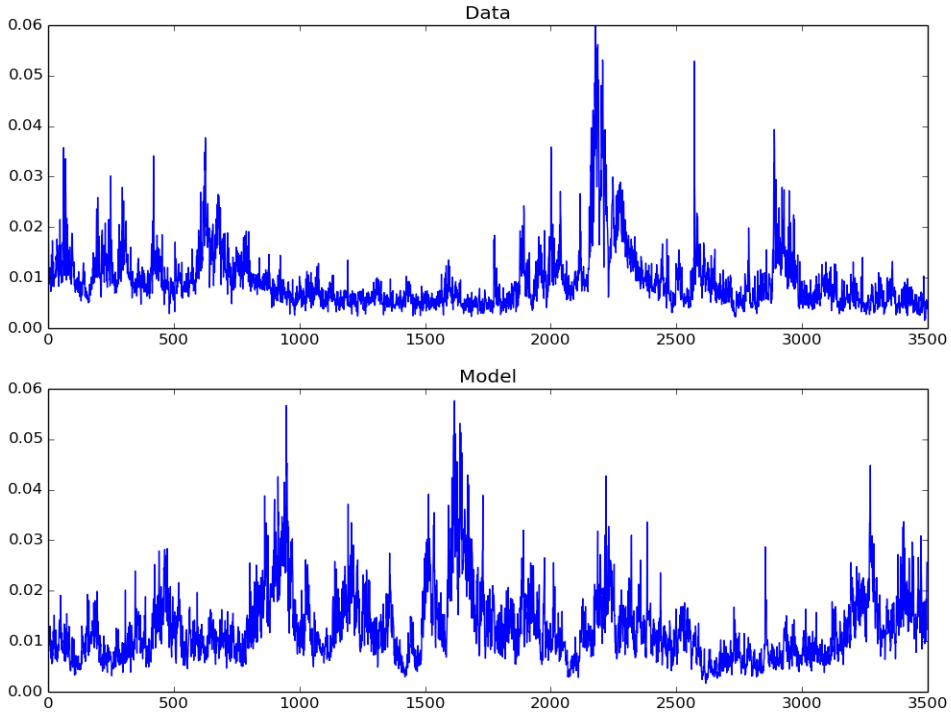


Figure I.15 – Volatility of the S&P (above) and of the model (below).

A first reaction to Figure I.15 is that the simulated and actual graphs look very alike. In particular, in both of them, persistent periods of high volatility alternate with low volatility periods. On closer inspection of the empirical volatility series, we observe that the sample path of the volatility on a restricted time window seems to exhibit the same kind of qualitative properties as those of the global sample path (for example periods of high and low activity). This fractal-type behavior of the volatility has been investigated both empirically and theoretically in, for example, [BM03, BP03, MS00a].

At the visual level, we observe that this fractal-type behavior is also reproduced in our model, as we now explain. Denote by $L^{x,H}$ the law of the geometric fractional Brownian motion with

Hurst exponent H and volatility x on $[0,1]$, that is $(e^{xW_t^H})_{t\in[0,1]}$. Then, when α is very small, the rescaled volatility process on $[0,\Delta]$: $(\sigma_{t\Delta}/\sigma_0)_{t\in[0,1]}$, has approximately the law $L^{\nu\Delta^H,H}$. Now remark that for H small, the function u^H increases very slowly. Thus, over a large range of observation scales Δ, the rescaled volatility processes on $[0,\Delta]$ have approximately the same law. For example, between an observation scale of one day and five years (1250 open days), the coefficient x characterizing the law of the volatility process is "only" multiplied by $1250^{0.14} = 2.7$. It follows that in the RFSV model, the volatility process over one day resembles the volatility process over a decade.

In order to study this multi scaling more quantitatively, let us consider as in [BM03] the quantities

$$M(q,\Delta) = \mathbb{E}[(\sum_{i=1}^{\Delta} \sigma_i^2)^q].$$

If we plot $\log(M(q,\Delta))$ as a function of $\log(\Delta)$ for different values of q, we get that both on real data and on simulations of the RFSV model, $M(\Delta,q)$ behaves as a power law function of Δ, see Figure I.16:

$$M(\Delta,q) = K_q\Delta^{\zeta_q}.$$

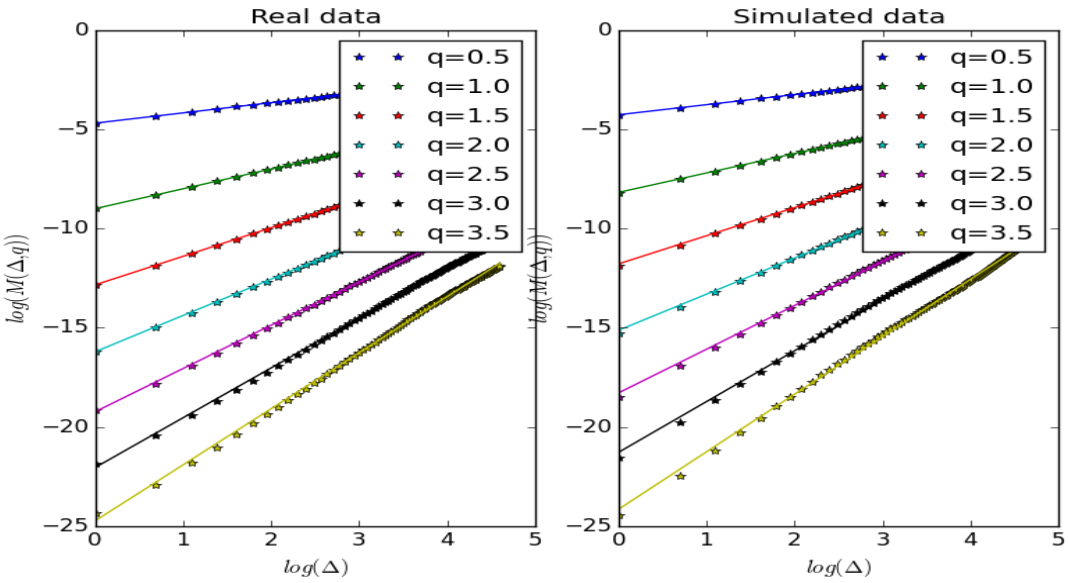


Figure I.16 – Empirical counterpart of $\log(M(\Delta,q))$ as a function of $\log(\Delta)$ on the S&P (left) and simulation (right).

Moreover, the exponent ζ_q is a concave function of q, see Figure I.17. It is explained in [BKM08] and [BM03] that this scaling and this concavity are related to the stochastic scale

invariance of the cumulated realized variance. As explained above, in the RFSV model, we retrieve this apparent stochastic scale invariance.

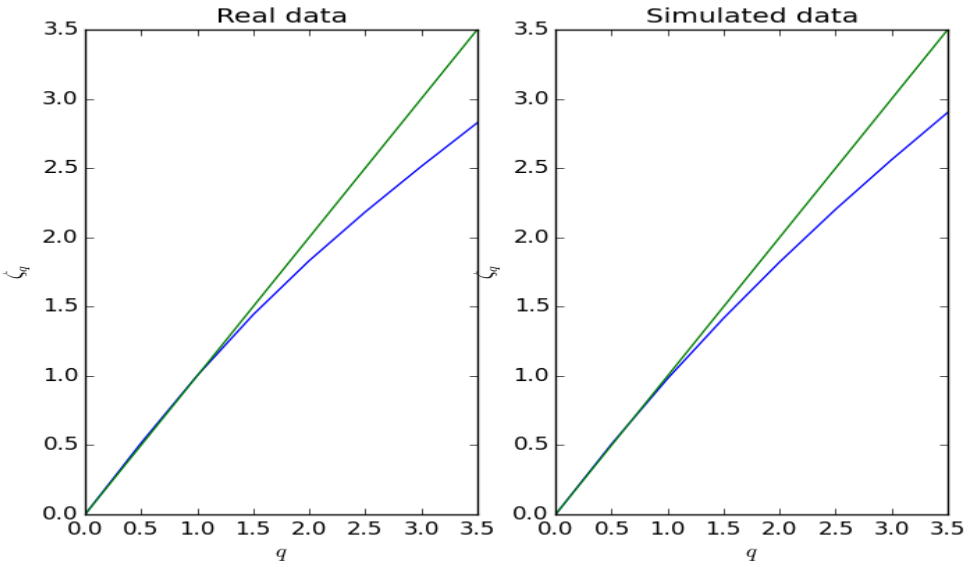


Figure I.17 – Estimates of ζ_q as a function of q on the S&P (left) and simulation (right).

4 Spurious long memory of volatility?

We revisit in this section the issue of long memory of volatility through the lens of our model. As mentioned earlier in the introduction, the long memory of volatility is widely accepted as a stylized fact. Specifically, this means that the autocovariance function $\mathrm{Cov}[\log(\sigma_t), \log(\sigma_{t+\Delta})]$ (or sometimes $\mathrm{Cov}[\sigma_t, \sigma_{t+\Delta}]$) goes slowly to zero as $\Delta \to \infty$ and often even more precisely, that it behaves as $\Delta^{-\gamma}$, with $\gamma < 1$ as $\Delta \to \infty$.

In previous sections, we showed that both in the data and in our model,

$$\mathrm{Cov}[\log(\sigma_t), \log(\sigma_{t+\Delta})] \approx A - B\Delta^{2H}$$

and

$$\mathrm{Cov}[\sigma_t, \sigma_{t+\Delta}] \approx C\, e^{-B\Delta^{2H}} - D,$$

for some constants A, B, C and D. Thus, neither in the model nor in the data does the autocovariance function decay as a power law. And neither the data nor the model exhibits long memory[11], see again Figure I.12.

[11] In fact the notion of empirical long memory does not make much sense outside the power law case. Indeed the empirical values of covariances at very large time scales are never measurable and thus one cannot conclude if

We now revisit some standard statistical procedures aimed at identifying long memory that have been used in the financial econometrics literature. In the sequel, we apply these both to the data and to sample paths of the RFSV model. Such procedures are of course designed to identify long memory under rather strict modeling assumptions; spurious results may obviously then be obtained if the model underlying the estimation procedure is misspecified .

With the same model parameters as in Section 3.4, we simulate our model over 3,500 days, which corresponds to the size of our dataset. Consider first the procedure in [ABDL01], where the authors test for long memory in the volatility by studying the scaling behavior of the quantity

$$V(t) = \mathrm{Var}\left[\int_0^t \sigma_s^2 ds\right]$$

with respect to t. In the model they consider, if $V(t)$ behaves asymptotically as $t^{2-\gamma}$ with $\gamma < 1$, then the autocorrelation function of the log-volatility should behave as $t^{-\gamma}$. Figure I.18 presents the graph of the logarithm of the empirical counterpart of $V(t)$ against the logarithm of t, on the S&P data and within our simulation framework.

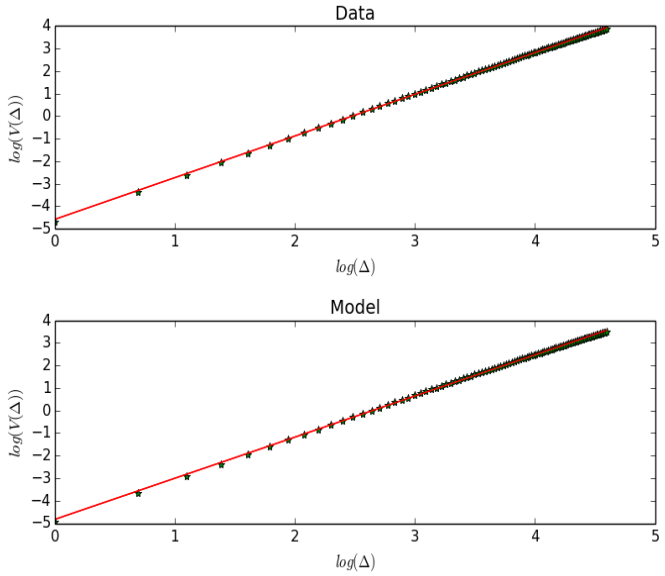


Figure I.18 – Empirical counterpart of $\log(V(t))$ as a function of $\log(t)$ on S&P (above) and simulation (below).

We note from Figure I.18 that both our simulated model and market data lead to very similar graphs, close to straight lines with slope 1.86. Accordingly, in the setting of [ABDL01], we

the series of covariances converges in general. All that we say here is that the autocovariance of the (log-)volatility does not behave as a power law.

would deduce power law behavior of the autocorrelation function with exponent 0.14 and therefore long memory. Thus, if the data are generated by a model like the RFSV model, one can easily be wrongly convinced that the volatility time series exhibits long memory.

In [ABDL03], the authors deduce long memory in the volatility by showing that the process ε_t obtained by fractional differentiation of the log-volatility $\varepsilon_t = (1-L)^d \log(\sigma_t)$, with $d = 0.4$ (which is considered as a reasonable value) and L the lag operator, behaves as a white noise. To check for this, they simply compute the autocorrelation function of ε_t. We give in Figure I.19 the autocorrelation functions of the logarithm of σ_t and ε_t, again both on the data and on the simulated path.

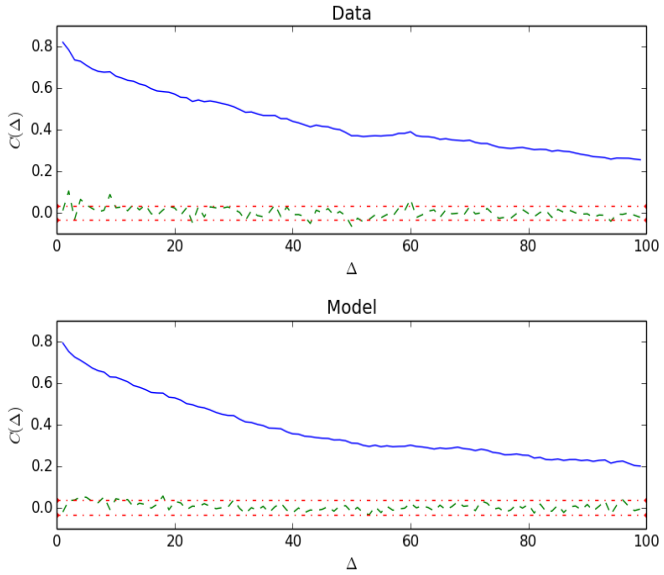


Figure I.19 – Autocorrelation functions of $\log(\sigma_t)$ (in blue) and ε_t (in green) and the Bartlett standard error bands (in red), for S&P data (above) and for simulated data (below).

Once again, the data and the simulation generate very similar plots. We conclude that this procedure for estimating long memory is just as fragile as the first, and it is easy to wrongly deduce volatility long memory when applying it.

In conclusion, it seems that classical estimation procedures identify spurious long memory of volatility in the RFSV model. Moreover, these procedures estimate the same long memory parameter from data generated from a suitably calibrated RFSV model as they estimate from empirical data. Once again, our conclusion is that although the (log-)volatility may exhibit some form of persistence, it does not present any long memory in the classical power law sense.

5 Forecasting using the RFSV model

In this section, we present an application of our model: forecasting the log-volatility and the variance.

5.1 Forecasting log-volatility

The key formula on which our prediction method is based is the following one:

$$\mathbb{E}[W^H_{t+\Delta}|\mathscr{F}_t] = \frac{\cos(H\pi)}{\pi}\Delta^{H+1/2}\int_{-\infty}^{t}\frac{W^H_s}{(t-s+\Delta)(t-s)^{H+1/2}}ds,$$

where W^H is a fBM with $H < 1/2$ and $\mathscr{F}_t$ the filtration it generates, see Theorem 4.2 of [NP00]. By construction, over any reasonable time scale of interest, as formalized in Corollary 1, we may approximate the fOU volatility process in the RFSV model as $\log\sigma^2_t \approx 2\nu W^H_t + C$ for some constants ν and C. Our prediction formula for log-variance then follows:[12]

$$\mathbb{E}\left[\log\sigma^2_{t+\Delta}|\mathscr{F}_t\right] = \frac{\cos(H\pi)}{\pi}\Delta^{H+1/2}\int_{-\infty}^{t}\frac{\log\sigma^2_s}{(t-s+\Delta)(t-s)^{H+1/2}}ds. \tag{9}$$

This formula, or rather its approximation through a Riemann sum (we assume in this section that the volatilities are perfectly observed, although they are in fact estimated), is used to forecast the log-volatility 1, 5 and 20 days ahead ($\Delta = 1,\ 5,\ 20$).

We now compare the predictive power of formula (9) with that of AR and HAR forecasts, in the spirit of [Cor09][13]. Recall that for a given integer $p > 0$, the AR(p) and HAR predictors take the following form (where the index i runs over the series of daily volatility estimates):

- AR(p):
$$\widehat{\log(\sigma^2_{t+\Delta})} = K^\Delta_0 + \sum_{i=0}^{p} C^\Delta_i \log(\sigma^2_{t-i}).$$
- HAR :
$$\widehat{\log(\sigma^2_{t+\Delta})} = K^\Delta_0 + C^\Delta_0\log(\sigma^2_t) + C^\Delta_5\frac{1}{5}\sum_{i=0}^{5}\log(\sigma^2_{t-i}) + C^\Delta_{20}\frac{1}{20}\sum_{i=0}^{20}\log(\sigma^2_{t-i}).$$

We estimate AR coefficients using the R `stats` library[14] on a rolling time window of 500 days. In the HAR case, we use standard linear regression to estimate the coefficients as explained in [Cor09]. In the sequel, we consider $p = 5$ and $p = 10$ in the AR formula. Indeed, these parameters essentially give the best results for the horizons at which we wish to forecast the

[12] The constants 2ν and C cancel when deriving the expression.

[13] Note that we do not consider GARCH models here since we have access to high frequency volatility estimates and not only to daily returns. Indeed, it is shown in [ABDL03] that forecasts based on the time series of realized variance outperform GARCH forecasts based on daily returns.

[14] More precisely, we use the default Yule-Walker method.

volatility (1, 5 and 20 days). For each day, we forecast volatility for five different indices[15].

We then assess the quality of the various forecasts by computing the ratio P between the mean squared error of our predictor and the (approximate) variance of the log-variance:

$$P = \frac{\sum_{k=500}^{N-\Delta} \left(\log(\sigma^2_{k+\Delta}) - \widehat{\log(\sigma^2_{k+\Delta})} \right)^2}{\sum_{k=500}^{N-\Delta} \left(\log(\sigma^2_{k+\Delta}) - \mathbb{E}[\log(\sigma^2_{t+\Delta})] \right)^2},$$

where $\mathbb{E}[\log(\sigma^2_{t+\Delta})]$ denotes the empirical mean of the log-variance over the whole time period.

	AR(5)	AR(10)	HAR(3)	RFSV
SPX2.rv $\Delta = 1$	0.317	0.318	0.314	**0.313**
SPX2.rv $\Delta = 5$	0.459	0.449	0.437	**0.426**
SPX2.rv $\Delta = 20$	0.764	0.694	0.656	**0.606**
FTSE2.rv $\Delta = 1$	0.230	0.229	0.225	**0.223**
FTSE2.rv $\Delta = 5$	0.357	0.344	0.337	**0.320**
FTSE2.rv $\Delta = 20$	0.651	0.571	0.541	**0.472**
N2252.rv $\Delta = 1$	0.357	0.358	0.351	**0.345**
N2252.rv $\Delta = 5$	0.553	0.533	0.513	**0.504**
N2252.rv $\Delta = 20$	0.875	0.795	0.746	**0.714**
GDAXI2.rv $\Delta = 1$	0.237	0.238	0.234	**0.231**
GDAXI2.rv $\Delta = 5$	0.372	0.362	0.350	**0.339**
GDAXI2.rv $\Delta = 20$	0.661	0.590	0.550	**0.498**
FCHI2.rv $\Delta = 1$	0.244	0.244	0.241	**0.238**
FCHI2.rv $\Delta = 5$	0.378	0.373	0.366	**0.350**
FCHI2.rv $\Delta = 20$	0.669	0.613	0.598	**0.522**

Table I.1 – Ratio P for the AR, HAR and RFSV predictors.

We note from Table I.1 that the RFSV forecast consistently outperforms the AR and HAR forecasts, especially at longer horizons. Moreover, our forecasting method is more parsimonious since it only requires the parameter H to forecast the log-variance. Compare this with the AR and HAR methods, for which coefficients depend on the forecast time horizon and must be recomputed if this horizon changes.

Remark that our predictor can be linked to that of [DRV12], where the issue of the prediction of the log-volatility in the multifractal random walk model of [BM03] is tackled. In this model,

$$\mathbb{E}[\log(\sigma^2_{t+\Delta}) | \mathscr{F}_t] = \frac{1}{\pi} \sqrt{\Delta} \int_{-\infty}^{t} \frac{\log(\sigma^2_s)}{(t-s+\Delta)\sqrt{t-s}} ds,$$

[15] In addition to S&P and NASDAQ, we also investigate CAC40, FTSE and Nikkei, over the same time period as S&P and NASDAQ. For simplicity, the parameter H used in our predictor is computed only once for each asset, using the whole time period. This yields similar results to using a moving time window adapted in time.

which is the limit of our predictor when H tends to zero.

Note also that our prediction formula may be rewritten as

$$\mathbb{E}[\log(\sigma^2_{t+\Delta})|\mathscr{F}_t] = \frac{\cos(H\pi)}{\pi}\int_0^{+\infty} \frac{\log(\sigma^2_{t-\Delta u})}{(u+1)\,u^{H+1/2}}\,du.$$

For a given small $\varepsilon > 0$, let r be the smallest real number such that

$$\int_r^{+\infty} \frac{1}{(u+1)\,u^{H+1/2}}\,du \leq \varepsilon.$$

Then we have, with an error of order ε,

$$\mathbb{E}[\log(\sigma^2_{t+\Delta})|\mathscr{F}_t] \approx \frac{\cos(H\pi)}{\pi}\int_0^{r} \frac{\log(\sigma^2_{t-\Delta u})}{(u+1)\,u^{H+1/2}}\,du.$$

Consequently, the volatility process needs to be considered (roughly) down to time $t - \Delta r$ if one wants to forecast up to time Δ in the future. The relevant regression window is thus linear in the forecasting horizon. For example, for $r = 1$, $\varepsilon = 0.35$ which is not so unreasonable. In this case, as is well-known to practitioners, to predict volatility one week ahead, one should essentially look at the volatility over the last week. If trying to predict the volatility one month ahead, one should look at the volatility over the last month.

5.2 Variance prediction

Recall that $\log\sigma^2_t \approx 2\nu W^H_t + C$ for some constant C. In [NP00], it is shown that $W^H_{t+\Delta}$ is conditionally Gaussian with conditional variance

$$\mathrm{Var}[W^H_{t+\Delta}|\mathscr{F}_t] = c\,\Delta^{2H}$$

with

$$c = \frac{\Gamma(3/2-H)}{\Gamma(H+1/2)\,\Gamma(2-2H)}.$$

Thus, we obtain the following natural form for the RFSV predictor of the variance:

$$\widehat{\sigma^2_{t+\Delta}} = \exp\left\{\widehat{\log\sigma^2_{t+\Delta}} + 2\,c\,\nu^2\Delta^{2H}\right\}$$

where $\widehat{\log(\sigma^2_{t+\Delta})}$ is the estimator from Section 5.1 and ν^2 is estimated as the exponential of the intercept in the linear regression of $\log(m(2,\Delta))$ on $\log(\Delta)$.

As in the previous paragraph, we compare in Table I.2 the performance of the RFSV forecast with those of AR and HAR forecasts (constructed on variance rather than log-variance this time).

	AR(5)	AR(10)	HAR(3)	RFSV
SPX2.rv $\Delta = 1$	0.520	0.566	0.489	**0.475**
SPX2.rv $\Delta = 5$	0.750	0.745	0.723	**0.672**
SPX2.rv $\Delta = 20$	1.070	1.010	1.036	**0.903**
FTSE2.rv $\Delta = 1$	0.612	0.621	0.582	**0.567**
FTSE2.rv $\Delta = 5$	0.797	0.770	0.756	**0.707**
FTSE2.rv $\Delta = 20$	1.046	0.984	0.935	**0.874**
N2252.rv $\Delta = 1$	0.554	0.579	**0.504**	0.505
N2252.rv $\Delta = 5$	0.857	0.807	0.761	**0.729**
N2252.rv $\Delta = 20$	1.097	1.046	1.011	**0.964**
GDAXI2.rv $\Delta = 1$	0.439	0.448	0.399	**0.386**
GDAXI2.rv $\Delta = 5$	0.675	0.650	0.616	**0.566**
GDAXI2.rv $\Delta = 20$	0.931	0.850	0.816	**0.746**
FCHI2.rv $\Delta = 1$	0.533	0.542	0.470	**0.465**
FCHI2.rv $\Delta = 5$	0.705	0.707	0.691	**0.631**
FCHI2.rv $\Delta = 20$	0.982	0.952	0.912	**0.828**

Table I.2 – Ratio P for the AR, HAR and RFSV predictors.

We find again that the RFSV forecast typically outperforms HAR and AR, although it is worth noting that the HAR forecast is already visibly superior to the AR forecast.

6 The microstructural foundations of the irregularity of the volatility

We gather in this section some ideas which may help to understand why the observed volatility appears so irregular. The starting point is the analysis of the order flow through Hawkes processes. These processes are extensions of Poisson processes where the intensity at a given time depends on the location of the past jumps. More precisely, let us consider a time period starting at 0 and denote by N_t the number of transactions between 0 and t. Assuming the point process N_t follows a Hawkes process means its intensity at time t, λ_t, takes the form:

$$\lambda_t = \mu + \sum_{0<J_i<t} \phi(t - J_i),$$

where the J_i are the past jump times, μ is a positive constant and ϕ is a non negative deterministic function called kernel.

When trying to calibrate such models on high frequency data, two main phenomena almost systematically occur:

- The L^1 norm of ϕ is close to one, see [FS12, FS13, HBB13, LC14].

- The function ϕ has a power law tail, see [BM14a, HBB13].

The first of these two facts means the degree of endogeneity of the market is very high, that is one given order endogenously generates many other orders, see [FS12, FS13, HBB13]. This recent feature of financial markets is obviously related to electronic high frequency trading, where market participants automatically react to other participants orders through their algorithms. The second observation tells us that generally, a given order influences other orders over a long time period. This is likely due to the splitting of large orders. Indeed, many orders are actually part of a metaorder whose full execution can take a large amount of time.

We believe these two phenomena together lead to a superposition effect inducing this irregular volatility. Indeed, it is explained in Chapters II and III that the macroscopic scaling limit of Hawkes processes with power law tail and kernel with L^1 norm close to one can be seen as an integrated fractional process, with Hurst parameter H smaller than $1/2$. This signifies that at large sampling scales, the dynamics of the cumulated order flow is well approximated by an integrated fractional process, with $H < 1/2$. Then, it is clearly established that there is a linear relation between cumulated order flow and integrated variance. Thus we retrieve here that because of this superposition effect, the volatility should behave as a fractional process with $H < 1/2$. The same kind of results is obtained by modeling the log-volatility as a nearly unstable heavy tail autoregressive process, see Chapter IV.

7 RFSV and pricing

In [BFG15] the authors explore the implications of the RFSV model (written under the physical measure $\mathbb{P}$), for option pricing (under the pricing measure $\mathbb{Q}$). In particular, following Mandelbrot and Van Ness, the fBM that appears in the definition (6) of the RFSV model may be represented as a fractional integral of a standard Brownian motion as follows [MVN68]:

$$W_t^H = \int_0^t \frac{dW_s}{(t-s)^\gamma} + \int_{-\infty}^0 \left[\frac{1}{(t-s)^\gamma} - \frac{1}{(-s)^\gamma} \right] dW_s, \tag{10}$$

with $\gamma = \frac{1}{2} - H$. The observed anti correlation between price moves and volatility moves may then be modeled naturally by anti correlating the Brownian motion W that drives the volatility process with the Brownian motion driving the price process. As already shown by Fukasawa [Fuk11, Fuk15], such a model with a small H reproduces the observed decay of at-the-money volatility skew with respect to time to expiry, asymptotically for short times. It is also that an appropriate extension of Fukasawa's model, consistent with the RFSV model, fits the entire implied volatility surface remarkably well, not just for short expirations. Moreover, despite that it would seem from (10) that knowledge of the entire path $\{W_s : s < t\}$ of the Brownian motion would be required, it turns out that the statistics of this path necessary for option pricing are traded and thus easily observed.

8 Conclusion

Using daily realized variance estimates as proxies for daily spot (squared) volatilities, we uncovered two startlingly simple regularities in the resulting time series. First we found that the distributions of increments of log-volatility are approximately Gaussian, consistent with many prior studies. Secondly, we established the monofractal scaling relationship

$$\mathbb{E}\left[|\log(\sigma_\Delta) - \log(\sigma_0)|^q\right] = K_q \nu^q \Delta^{q\,H}, \tag{11}$$

where H can be seen as a measure of smoothness characteristic of the underlying volatility process; typically, $0.06 < H < 0.2$. The simple scaling relationship (11) naturally suggests that log-volatility may be modeled using fractional Brownian motion.

The resulting Rough Fractional Stochastic Volatility (RFSV) model turns out to be formally almost identical to the FSV model of Comte and Renault [CR98], with one major difference: In the FSV model, $H > 1/2$ to ensure long memory whereas in the RFSV model $H < 1/2$, typically, $H \approx 0.1$. Moreover, in the FSV model, the mean reversion coefficient α has to be large compared to $1/T$ to ensure a decaying volatility skew; in the RFSV model, the volatility skew decays naturally just like the observed volatility skew, $\alpha \ll 1/T$ and indeed for time scales of practical interest, we may proceed as if α were exactly zero.

We further showed that applying standard statistical estimators to volatility time series simulated with the RFSV model would lead us to erroneously deduce the presence of long memory, with parameters similar to those found in prior studies. Despite that volatility in the RFSV model (or in the data) is not long memory, we can therefore explain why long memory of volatility is widely accepted as a stylized fact.

As an application of the RFSV model, we showed how to forecast volatility at various times cales, at least as well as Fulvio Corsi's impressive HAR estimator, but with only one parameter – H!

Finally, we explained how the RFSV model could emerge as the scaling limit of a Hawkes process description of order flow.

I.A Technical results

I.A.1 Proof of Proposition 1

Starting from Equation (5) and applying integration by parts, we get

$$X_t^\alpha = \nu W_t^H - \int_{-\infty}^t \nu\alpha e^{-\alpha(t-s)} W_s^H ds + m.$$

Therefore,

$$(X_t^\alpha - X_0^\alpha) - \nu W_t^H = -\int_0^t \nu\alpha e^{-\alpha(t-s)} W_s^H ds - \int_{-\infty}^0 \nu\alpha (e^{-\alpha(t-s)} - e^{\alpha s}) W_s^H ds.$$

Consequently,

$$\sup_{t\in[0,T]} |(X_t^\alpha - X_0^\alpha) - \nu W_t^H| \le \nu\alpha T \hat{W}_T^H + \int_{-\infty}^0 \nu\alpha(e^{\alpha s} - e^{-\alpha(T-s)})\hat{W}_s^H ds,$$

where $\hat{W}_t^H = \sup_{s\in[0,t]} |W_s^H|$. Using the maximum inequality of [NV99], we get

$$\mathbb{E}\Big[\sup_{t\in[0,T]} |(X_t^\alpha - X_0^\alpha) - \nu W_t^H|\Big] \le c\big(\nu\alpha T T^H + \int_{-\infty}^0 \nu\alpha(T\alpha e^{\alpha s})|s|^H ds\big),$$

with c some constant. The term on the right hand side is easily seen to go to zero as α tends to zero.

I.A.2 Proof of Corollary 1

We first recall Equation (2.2) in [CKM03] which writes:

$$\mathrm{Cov}[X_{t+\Delta}^\alpha, X_t^\alpha] = K\int_{\mathbb{R}} e^{i\Delta x}\frac{|x|^{1-2H}}{\alpha^2+x^2}dx,$$

with $K = \nu^2\Gamma(2H+1)\sin(\pi H)/(2\pi)$[16]. Now remark that

$$\mathbb{E}[(X_{t+\Delta}^\alpha - X_t^\alpha)^2] = 2\mathrm{Var}[X_t^\alpha] - 2\mathrm{Cov}[X_{t+\Delta}^\alpha, X_t^\alpha].$$

Therefore,

$$\mathbb{E}[(X_{t+\Delta}^\alpha - X_t^\alpha)^2] = 2K\int_{\mathbb{R}}(1-e^{i\Delta x})\frac{|x|^{1-2H}}{\alpha^2+x^2}dx.$$

This implies that for fixed Δ, $\mathbb{E}[|X_{t+\Delta}^\alpha - X_t^\alpha|^2]$ is uniformly bounded by

$$2K\int_{\mathbb{R}}(1-e^{i\Delta x})\frac{|x|^{1-2H}}{x^2}dx.$$

Moreover, $X_{t+\Delta}^\alpha - X_t^\alpha$ is a Gaussian random variable and thus for every q, its $(q+1)^{th}$ moment is uniformly bounded (in α) so that the family $|X_{t+\Delta}^\alpha - X_t^\alpha|^q$ is uniformly integrable. Therefore, since by Proposition 1,

$$|X_{t+\Delta}^\alpha - X_t^\alpha|^q \to \nu^q |W_{t+\Delta}^H - W_t^H|^q, \text{ in law},$$

we get the convergence of the sequence of expectations.

[16] This covariance is real because it is the Fourier transform of an even function.

I.B Estimations of H

I.B.1 On different indices

Index	$\zeta_{0.5}/0.5$	ζ_1	$\zeta_{1.5}/1.5$	$\zeta_2/2$	$\zeta_3/3$
SPX2.rv	0.128	0.126	0.125	0.124	0.124
FTSE2.rv	0.132	0.132	0.132	0.131	0.127
N2252.rv	0.131	0.131	0.132	0.132	0.133
GDAXI2.rv	0.141	0.139	0.138	0.136	0.132
RUT2.rv	0.117	0.115	0.113	0.111	0.108
AORD2.rv	0.072	0.073	0.074	0.075	0.077
DJI2.rv	0.117	0.116	0.115	0.114	0.113
IXIC2.rv	0.131	0.133	0.134	0.135	0.137
FCHI2.rv	0.143	0.143	0.142	0.141	0.138
HSI2.rv	0.079	0.079	0.079	0.080	0.082
KSI1.rv	0.133	0.133	0.134	0.134	0.132
AEX.rv	0.145	0.147	0.149	0.149	0.149
SSMI.rv	0.149	0.153	0.156	0.158	0.158
IBEX2.rv	0.138	0.138	0.137	0.136	0.133
NSEI.rv	0.119	0.117	0.114	0.111	0.102
MXX.rv	0.077	0.077	0.076	0.075	0.071
BVSP.rv	0.118	0.118	0.119	0.120	0.120
GSPTSE.rv	0.106	0.104	0.103	0.102	0.101
STOXX50E.rv	0.139	0.135	0.130	0.123	0.101
FTSTI.rv	0.111	0.112	0.113	0.113	0.112
FTSEMIB.rv	0.130	0.132	0.133	0.134	0.134

Table I.3 – Estimates of ζ_q for all indices in the Oxford-Man dataset.

I.B.2 On different time intervals

Index	H (first half)	H (second half)
SPX2.rk	0.115	0.158
FTSE2.rk	0.140	0.156
N2252.rk	0.083	0.134
GDAXI2.rk	0.154	0.168
RUT2.rk	0.098	0.149
AORD2.rk	0.059	0.114
DJI2.rk	0.123	0.151
IXIC2.rk	0.094	0.156
FCHI2.rk	0.140	0.146
HSI2.rk	0.072	0.129
KS11.rk	0.109	0.147
AEX.rk	0.168	0.151
SSMI.rk	0.206	0.183
IBEX2.rk	0.122	0.149
NSEI.rk	0.112	0.124
MXX.rk	0.068	0.118
BVSP.rk	0.074	0.134
GSPTSE.rk	0.075	0.147
STOXX50E.rk	0.138	0.132
FTSTI.rk	0.080	0.171
FTSEMIB.rk	0.133	0.140

Table I.4 – Estimates of H over two different time intervals for all indices in the Oxford-Man dataset

I.C The effect of smoothing

Although we are really interested in the model

$$\log\sigma_{t+\Delta} - \log\sigma_t = \nu\left(W^H_{t+\Delta} - W^H_t\right),$$

consider the more tractable (fractional Stein and Stein or fSS) model:

$$v_{t+\Delta} - v_t = \alpha\left(W^H_{t+\Delta} - W^H_t\right),$$

where $v_t = \sigma^2$. We cannot observe v_t but suppose we can proxy it by the average

$$\hat{v}^\delta_t = \frac{1}{\delta}\int_0^\delta v_u\,du.$$

We would, for example, like to estimate $m(2,\Delta) = \mathbb{E}\left[(v_{t+\Delta} - v_t)^2\right]$. However, we need to proxy spot variance with integrated variance so instead we have the estimate

$$\begin{aligned} m^\delta(2,\Delta) &= \mathbb{E}\left[(\hat{v}^\delta_{t+\Delta} - \hat{v}^\delta_t)^2\right] \\ &= \frac{1}{\delta^2}\mathbb{E}\left[\left(\int_0^\delta (v_{u+\Delta} - v_u)\,du\right)^2\right] \\ &= \frac{\alpha^2}{\delta^2}\int_0^\delta\int_0^\delta \mathbb{E}\left[(W^H_{u+\Delta} - W^H_u)(W^H_{s+\Delta} - W^H_s)\right] du\,ds \\ &= \int_0^\delta\int_0^\delta \left\{|u-s+\Delta|^{2H} - |u-s|^{2H}\right\} du\,ds, \end{aligned} \tag{12}$$

where the last step uses that:

$$\mathbb{E}\left[W^H_u W^H_s\right] = \frac{1}{2}\left\{u^{2H} + s^{2H} - |u-s|^{2H}\right\},$$

and the symmetry of the integral.

We assume that the length δ of the smoothing window is less than one day so $\Delta > \delta$. Then easy computations give

$$\begin{aligned} &\int_0^\delta\int_0^\delta |u-s+\Delta|^{2H}\,du\,ds \\ = \;& \frac{1}{2H+1}\frac{1}{2H+2}\left\{(\Delta+\delta)^{2H+2} - 2\Delta^{2H+2} + (\Delta-\delta)^{2H+2}\right\} \end{aligned}$$

and

$$\int_0^\delta\int_0^\delta |u-s|^{2H}\,du\,ds = \frac{2}{2H+1}\frac{1}{2H+2}\delta^{2H+2}.$$

Substituting back into (12) gives

$$\begin{aligned} m^\delta(2,\Delta) &= \alpha^2\Delta^{2H}\frac{1}{2H+1}\frac{1}{2H+2}\frac{1}{\theta^2}\left\{(1+\theta)^{2H+2} - 2 - 2\theta^{2H+2} + (1-\theta)^{2H+2}\right\} \\ &=: \alpha^2\Delta^{2H} f(\theta). \end{aligned}$$

where $\theta = \delta/\Delta$.

Figure I.20 shows the effect of smoothing on the estimated variance in the fSS model. Keeping δ fixed, as Δ increases, $f(\theta) = f(\delta/\Delta)$ increases towards one. Thus, in a linear regression of $\log m^\delta(2,\Delta)$ against $\log\Delta$, we will obtain a higher effective H (from the higher slope) and a lower effective ("volatility of volatility") α, exactly as we observed in the RSFV model simulations in Section 3.4.

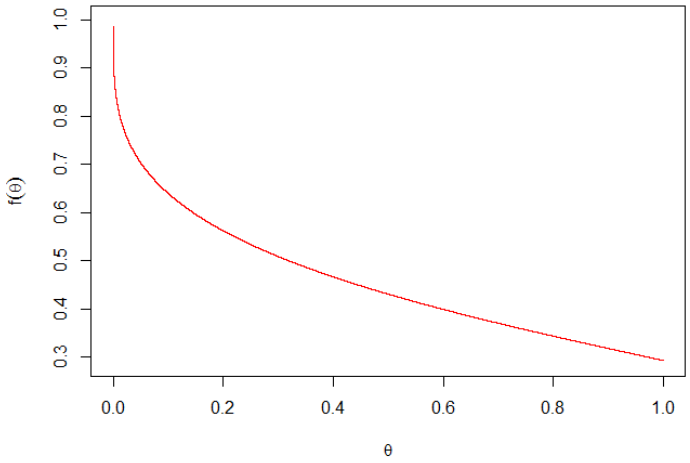


Figure I.20 – $f(\theta)$ vs $\theta = \delta/\Delta$ with $H = 0.14$.

Numerical example

In the simulation of the RSFV model in Section 3.4, we have $H = 0.14$, $\delta_1 = 1/24$ for the UZ estimate and $\delta_2 = 1/3$ for the RV estimate. We now reproduce a fSS analogue of the RFSV simulation plots of $m(2,\Delta)$ in Figure I.14. Specifically, for each $\Delta \in \{1,2,...,100\}$, with $\alpha = 0.3$ and $\delta = \delta_1$ or $\delta = \delta_2$, we compute the $m^\delta(2,\Delta)$ and regress $\log m^\delta(2,\Delta)$ against $\log\Delta$. The regressions are shown in Figure I.21 and results tabulated in Table I.5.
In Figure I.21 and Table I.5, we observe similar qualitative and quantitative biases from our fSS model simulation as we observe in our simulation of the RSFV model with equivalent parameters in Section 3.4.

Estimate	Est. α	Est. H
Exact ($\delta = 0$)	0.300	0.140
UZ ($\delta = 1/24$)	0.263	0.161
RV ($\delta = 1/3$)	0.230	0.184

Table I.5 – Estimated model parameters from the regressions shown in Figure I.21.

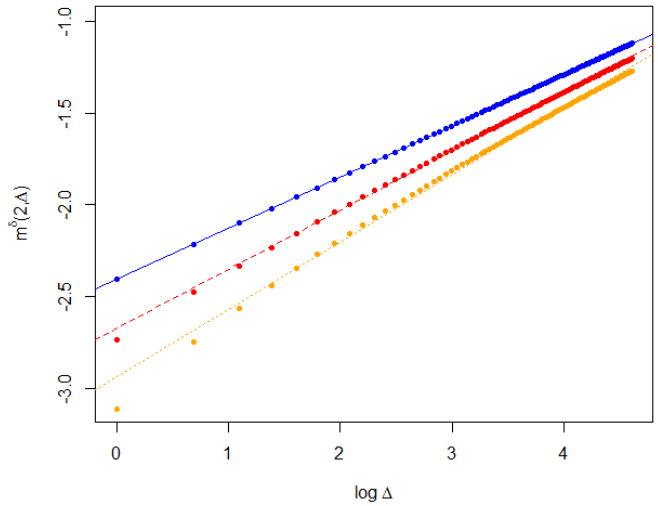


Figure I.21 – Analogue of Figure I.14 in the fSS model: The blue solid line is the true $m(2,\Delta)$; the red long-dashed line is the UZ estimate $m^{\delta_1}(2,\Delta)$; the orange short-dashed line is the RV estimate $m^{\delta_2}(2,\Delta)$.

I.D Further results

In this paragraph, we present some additional results related to the RFSV model.

I.D.1 RFSV model at intra day time scales

Until now, we have seen that the log-volatility process from one day to the other behaves as some sort of geometric fractional Brownian motion. In this paragraph, we wish to study what happens at intra day time scales. To do that, we will split every day in our high frequency database (for the DAX) into windows of one hour and estimate the realized variance on each window using the uncertainty zone estimator.

Before studying the smoothness of the obtained volatility process, one needs to deal with the intra day seasonality. Indeed, this effect is very important at intra day time scales, for example, the average variance from 4 to 5 pm is more than twice the average variance from 12 to 1 pm. To do this, we divide every hourly realized variance by the average realized variance on the corresponding time window. We then study the scaling of the increments of the log-volatility as before.

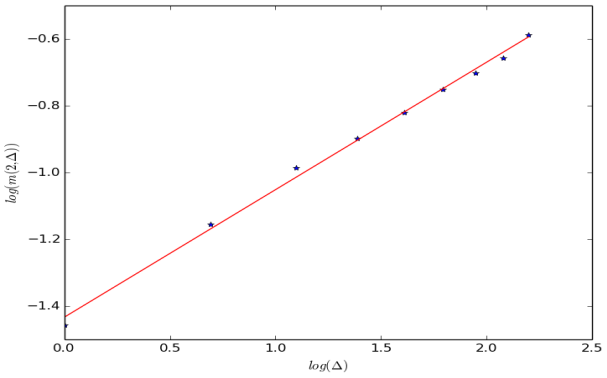


Figure I.22 – $\log(m(2,\Delta))$ as a function of $\log(\Delta)$ for the DAX on intra day time scales.

We observe the kind of scaling behavior observed at extra day time scales with $H = 0.19$. This seems to show that our model extends at intra day time scales.

I.D.2 From metaorders distribution to persistence in the order flow

In this paragraph, we consider a continuous time version of the metaorder model of [LMF05]. Representing this model as a cone of a two dimensional noise somehow as in [BM03], we obtain a simple relationship between the autocovariance function of the order flow and the metaorder size distribution. Considering the power law case, we obtain two interesting regimes: If the size distribution of metaorders has a finite expectation then the flow presents long memory. If not, the flow behaves somehow as a fractional Brownian motion of exponent smaller than one half as in the RFSV model.

I.D.2.1 The metaorder model

Let us consider a clustered order flow model where metaorders (clusters) arrive as a Poisson process and where their lengths are iid random variables (to simplify, we assume that the "intensity" of metaorders is fixed so that their volume is proportional to their length). In such a model, the (continuous time) order flow intensity V_t defined as the number of "active" metaorders at time t is thus given by:

$$V_t = \sum_{i=1}^{+\infty} \mathbf{1}_{T_i \in [t-L_i/2, t+L_i/2]}$$

where T is a Poisson process of intensity μ representing the times of the middle of metaorders and (L_i) are iid random variables of density $f_L(x)/\mu$ representing the size of metaorders.

The aim of this paragraph is, somehow as in [LMF05], to derive the correlation structure of the order flow from the distribution of L. As a tool for computations, we consider, somehow as in [BM03], the following "cone representation" of V_t:

Proposition 2. *In law, we have:*

$$V_t = N(A_t)$$

where N is a Poisson point process on $\mathbb{R} \times \mathbb{R}_+$ of intensity $f_L(x)dtdx$, $f_L(x)$ corresponds to the intensity of the arrival of metaorders of size x, A_t is the cone

$$A_t = \{(t', x); |t' - t| \leq x/2\}$$

see Figure I.23 and $N(A_t)$ is the number of points of N in A_t.

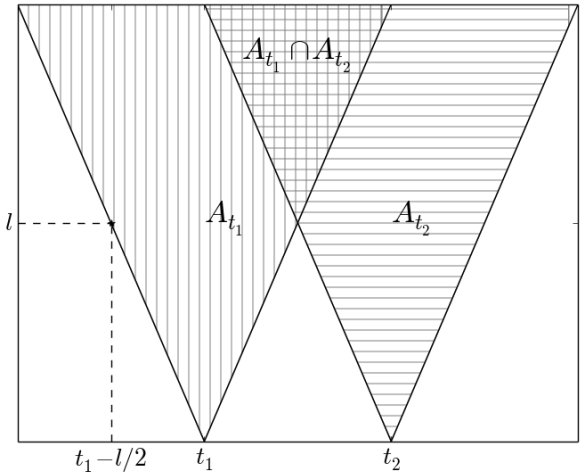


Figure I.23 – Illustration of the cones A_{t_1} and A_{t_2}.

Proof. To show this result, instead of modeling (T_i) and (L_i) "separately", we model (T_i, L_i) jointly as the Poisson point process N defined above. To end the proof, we only need to notice that $V_t = \sum_{i=1}^{+\infty} \mathbf{1}_{T_i \in [t - L_i/2, t + L_i/2]}$ is exactly equal to the number of points of (T_i, L_i) in A_t, that is $A_t(N)$. □

This representation is convenient because, the autocovariance of the order flow writes as

$$\mathrm{Cov}[V_t, V_{t+h}] = \mathrm{Var}[N(A_t \cap A_{t+h})] = \int_h^{+\infty} (t - h) f_L(t) dt. \tag{13}$$

In particular, the autocovariance function of the order flow is defined if and only if the size distribution of metaorders has a finite expectation.

I.D.2.2 Power law asymptotics

In this section, we use Equation (13) to study two interesting asymptotics where the size distribution of metaorders behaves as a power law.

The long memory case

In many works, it is mentioned that

$$f_L(x) = \frac{A}{(\delta + x)^{2+\gamma}}$$

with $\gamma \in (0,1)$, δ a cut-off scale and A a normalization constant, see for example [LMF05]. In that case, using Equation (13) we easily find as in [LMF05], that the autocorrelation of the order flow behaves as

$$\begin{aligned} \mathrm{Cov}[V_t, V_{t+h}] &= \int_h^{+\infty} (t-h) f_L(t) dt \\ &\underset{h \to +\infty}{\sim} \frac{c}{h^\gamma} \end{aligned}$$

using the dominated convergence theorem. This is consistent with the well established long memory of the order flow.

The fractional diffusion case

Let us now assume that the metaorder size distribution behaves as

$$f_L(x) = \frac{A}{(\delta + x)^{2-2H}} 1_{x \le T}$$

with $H \in (0, 1/2)$, δ a cut-off scale, T a very long stationarity time scale[17] and A a normalization constant.

Then, we easily get

$$\mathrm{Cov}[V_t, V_{t+h}] \underset{\delta \ll h \ll T}{\sim} A - Bh^{2H}$$

which is similar to the correlation structure of the log-volatility in RFSV[18].

I.D.2.3 Financial application

As for nearly unstable Hawkes and autoregressive processes, see Chapters III and IV, depending on the properties of size distribution of the metaorders, we can either get a long memory order flow or a fractionally diffusive order flow. Note also that at the interface between these two regimes, the Multifractal Random Measure model appears, see [BKM08].

[17] Somehow as for RFSV, it is necessary to introduce this time scale for the distribution to have an expectation and for V to have a correlation structure.

[18] Of course this model is not Gaussian but it is possible to obtain a Gaussian model with the same correlation structure by replacing the Poisson noise N by a Gaussian noise of same intensity.

I.D.3 RFSV and the distribution of returns

It is often argued, see [FL04] or [GGPS03] that the density of absolute price returns $|\Delta P_t|/P_t$ behaves as a power law of exponent between 3 and 4. In the RFSV model, the returns are in law equal to $N_t\sigma_t$, where $N_t \sim N(0,1)$ and are therefore light tail. However, since the observation time scale of the volatility process is not much larger than its ergodicity time scale, the empirical distribution of the volatility and thus of the returns are different from their theoretical distribution. In particular, we retrieve on simulations of the RFSV model the empirical power law behavior of the distribution of returns, see Figure I.24.

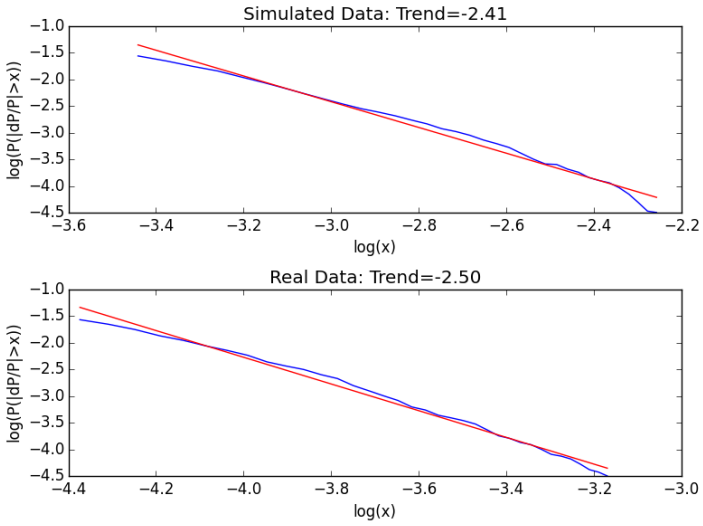


Figure I.24 – Logarithm of the probability that $|\Delta P_t|/P_t$ be higher than x as a function of $\log(x)$ on simulated data (above) and on real data (below).

I.D.4 An empirical time irreversibility

In this paragraph, we provide a simple example of the time irreversibility of the volatility process. Loosely speaking, it states that the (log-)volatility increases less often than it decreases but when it increases, it increases a lot. This property corresponds to the (inverse) skewness of the VIX, see for example [BFG15].

The first way to see this is to directly look at the log-volatility and to notice that indeed the positive increments of the log-volatility seem to be more important and less frequent than its negative increments (see the vertical black lines).

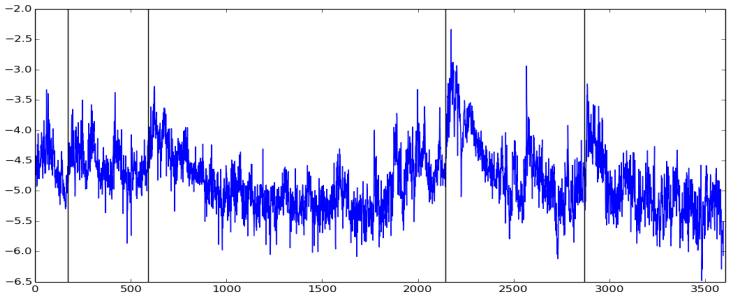


Figure I.25 – Log of the realized volatility on 3500 days.

A more formal way to study this phenomenon is to consider the empirical estimation of the quantities

$$m^3(\Delta) = \mathbb{E}[(\log(\sigma_{t+\Delta}) - \log(\sigma_t))^3]$$

and

$$M^3(\Delta) = \mathbb{E}[|\log(\sigma_{t+\Delta}) - \log(\sigma_t)|^3]$$

that we plot as a function of Δ, see Figure I.26.
We observe that although m^3 is smaller than M^3, it is significantly higher than zero especially after $\delta = 8$ days. We obtain the same kind of results for all assets.

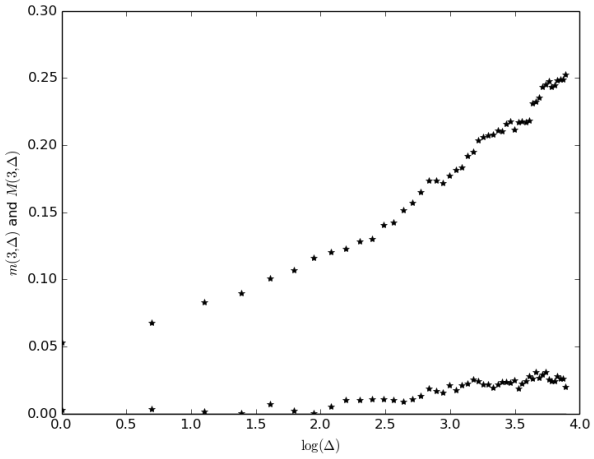


Figure I.26 – m^3 and M^3 as a function of $\log(\Delta)$ for the S&P.

Remark that this stylized fact is inconsistent with a time reversible volatility. Indeed, if the volatility were stationary[19] and time reversible[20], we would have:

$$\begin{aligned} m^3(\Delta) &= \mathbb{E}[(\log(\sigma_{t+\Delta}) - \log(\sigma_t))^3] \\ &= \mathbb{E}[(\log(\sigma_{-t-\Delta}) - \log(\sigma_{-t}))^3] \\ &= \mathbb{E}[(\log(\sigma_t) - \log(\sigma_{t+\Delta}))^3] \\ &= -m^3(\Delta) = 0. \end{aligned}$$

This property correspond to the smile of the VIX. The RFSV model does not reproduce this feature since it is time reversible. This is thus an "historical" and not a "risk neutral" property and does not corresponds to a "stochastic risk premium in the volatility" as it is said in [BFG15]. In order to solve this issue, one might wish to positively correlate the volatility of the log-volatility η with the past increments of the volatility.

I.D.5 Test of multifractality

In this paragraph, we test on our data the result of the MRM model which states that $m(2,\Delta)$ behaves as $A + B\log(\Delta)$. To do that, we plot $m(2,\Delta)$ as a function of $\log(\Delta)$, see Figure I.27.

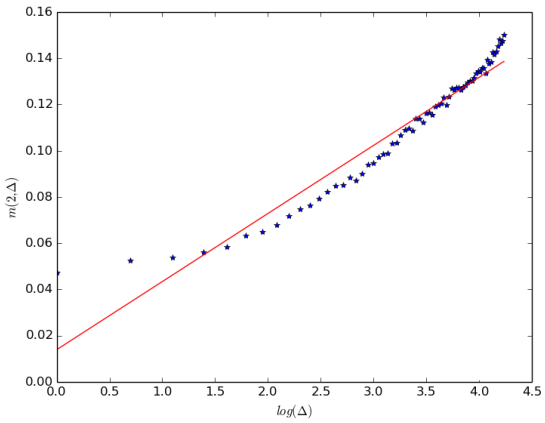


Figure I.27 – $m(2,\Delta)$ as a function of $\log(\Delta)$ for the DAX.

We get that this plot is slightly but significantly convex[21] in accordance with the RFSV model and a small Hurst index. Indeed recall that as H tends to zero, our model somehow tends towards the MRM model.

[19] A process X is stationary if the law of (X_{t+h}) is equal to the law of (X_t).

[20] A process X is time reversible if the law of (X_t) is equal to the law of (X_{-t}).

[21] This result is consistent across different kinds of data.

I.D.6 Historical skew measurement

In this paragraph, we empirically study the "skew" in the RFSV model and on historical data. To do this, we (anti-)correlate the Brownian motion driving price moves B with the fractional Brownian motion driving volatility W^H using the Mandelbrot-Van-Ness representation of the fractional Brownian motion as in Section 7:

$$\log(\sigma_t) - \log(\sigma_0) = \nu W_t^H = \nu\left(\int_0^t \frac{dW_s}{(t-s)^{1/2-H}} + \int_{-\infty}^0 \left[\frac{1}{(t-s)^{1/2-H}} - \frac{1}{(-s)^{1/2-H}}\right] dW_s\right),$$

$$dP_t = P_t \sigma_t dB_t$$

and

$$d < W_t, B_t >= -\rho dt.$$

In this model, the correlation of the increments of W^H and B writes as:

$$\mathbb{E}[(W_\Delta^H - W_0^H)(B_\Delta - B_0)] = -\rho \frac{\Delta^{1/2+H}}{1/2+H}$$

which implies that

$$\mathbb{E}[(\log(\sigma_\Delta) - \log(\sigma_0)) \int_0^\Delta \frac{dP_s}{P_s \sigma_s}] = -\rho\nu \frac{\Delta^{1/2+H}}{1/2+H}.$$

In order to check that this power law relation holds on data, we compute the empirical average $s(\Delta)$ corresponding to this expectation as a function of Δ and we plot $\log(s(\Delta))$ as a function of $\log(\Delta)$, see Figure I.28.

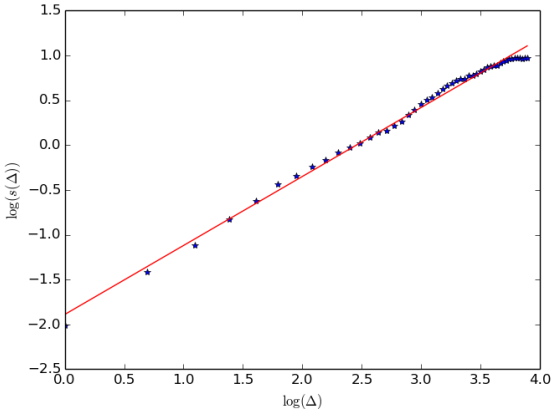


Figure I.28 – $\log(s(\Delta))$ as a function of $\log(\Delta)$ for the DAX.

We get that the power law relation holds rather well. However, we also get that the implied Hurst exponent is significantly higher[22] than the one predicted by looking at the scaling of the increments of the log-volatility. For example, for the DAX, we get a trend of 0.80 which corresponds to $H = 0.30$ instead of 0.13 in Section 2.2.

[22] This result is very preliminary and requires confirmation.

I.D.7 Rough market activity

In this chapter, we have applied our regularity measure to realized variance measurement. An alternative measure of market activity is to count the number of arriving market orders. In this paragraph, we measure the regularity of the arrival intensity of market orders. As before, we approximate the "punctual" arrival intensity of market orders as the number of market orders on a given time widow during the day: from 10 am to 11 am. We thus get a measure of this intensity λ_i for every day i. We then study of the moments of the increments of $\log(\lambda)$:

$$m^{\lambda}(q,\Delta) = \frac{1}{N}\sum_{k=1}^{N} |\log(\lambda_{k+\Delta}) - \log(\lambda_k)|^q.$$

For both the DAX and the Bund futures, we get that the obtained Hurst index is much smaller (0.05 instead of 0.13 for the DAX and 0.03 instead of 0.08 for the Bund), see Figure I.29.

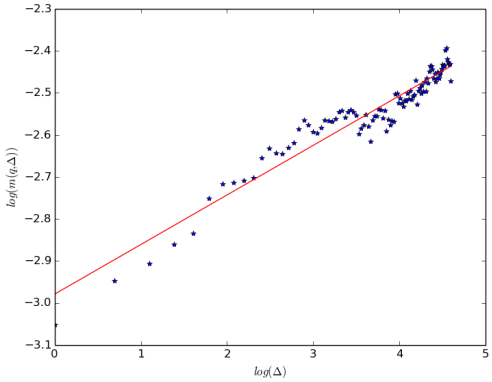


Figure I.29 – $\log(m^{\lambda}(2,\Delta))$ as a function of $\log(\Delta)$ for the DAX.

This small numerical estimation of H is coherent with simulations of a multifractal intensity model, which somehow corresponds to $H = 0$, and where $m^{\lambda}(2,\Delta) = A + B\log(\Delta)$, see Figure I.30.

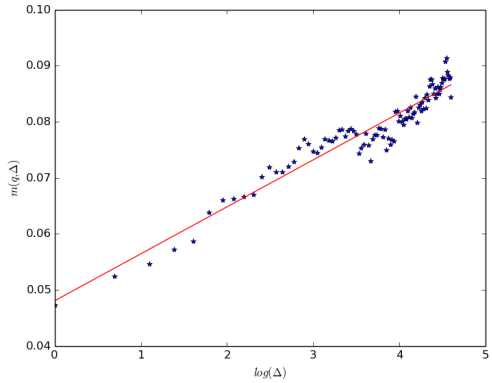


Figure I.30 – $m^{\lambda}(2,\Delta)$ as a function of $\log(\Delta)$ for the DAX.

Part II

Nearly unstable linear processes

CHAPTER II

Limit theorems for nearly unstable Hawkes processes

Abstract

Because of their tractability and their natural interpretations in term of market quantities, Hawkes processes are nowadays widely used in high frequency finance. However, in practice, the statistical estimation results seem to show that very often, only *nearly unstable Hawkes processes* are able to fit the data properly. By nearly unstable, we mean that the L^1 norm of their kernel is close to unity. We study in this work such processes for which the stability condition is almost violated. Our main result states that after suitable rescaling, they asymptotically behave like integrated Cox Ingersoll Ross models. Thus, modeling financial order flows as nearly unstable Hawkes processes may be a good way to reproduce both their high and low frequency stylized facts. We then extend this result to the Hawkes based price model introduced by Bacry *et al.* in [BDHM13]. We show that under a similar criticality condition, this process converges to a Heston model. Again, we recover well known stylized facts of prices, both at the microstructure level and at the macroscopic scale.

Keywords: Point processes, Hawkes processes, limit theorems, microstructure modeling, high frequency data, order flows, Cox Ingersoll Ross model, Heston model.

1 Introduction

A Hawkes process $(N_t)_{t\geq 0}$ is a self exciting point process, whose intensity at time t, denoted by λ_t, is of the form

$$\lambda_t = \mu + \sum_{0<J_i<t} \phi(t-J_i) = \mu + \int_{(0,t)} \phi(t-s)dN_s,$$

where μ is a positive real number, ϕ a regression kernel and the J_i are the points of the process before time t (see Section 2 for more accurate definitions). These processes have been introduced in 1971 by Hawkes, see [Haw71a, Haw71b, HO74], in the purpose of modeling earthquakes and their aftershocks, see [Ada76]. However, they are also used in various other

disciplines. In particular, in recent years, with the availability of (ultra) high frequency data, finance has become one of the main domains of application of Hawkes processes.

The introduction of Hawkes processes in finance is probably due to Bowsher, see [Bow07], who jointly studied transaction times and midquote changes, using the Hawkes framework. Then, in [BH04], Bauwens and Hautsch built so-called latent factor intensity Hawkes models and applied them to transaction data. Another pioneer in this type of approach is Hewlett. He considered in [Hew06] the particular case of the foreign exchange rates market for which he fitted a bivariate Hawkes process on buy and sell transaction data. More recently, Bacry *et al.* have developed a microstructure model for midquote prices based on the difference of two Hawkes processes, see [BDHM13]. Moreover, Bacry and Muzy have extended this approach in [BM14a] where they design a framework enabling to study market impact. Beyond midquotes and transaction prices, full limit order book data (not only market orders but also limit orders and cancellations) have also been investigated through the lenses of Hawkes processes. In particular, Large uses in [Lar07] a ten-variate multidimensional Hawkes process to this purpose. Note that besides microstruture problems, Hawkes processes have also been introduced in the study of other financial issues such as daily data analysis, see [ELL11], financial contagion, see [ASCDL10], or Credit Risk, see [EGG10].

The popularity of Hawkes processes in financial modeling is probably due to two main reasons. First, these processes represent a very natural and tractable extension of Poisson processes. In fact, comparing point processes and conventional time series, Poisson processes are often viewed as the counterpart of iid random variables whereas Hawkes processes play the role of autoregressive processes, see [DVJ02] for more details about this analogy. Another explanation for the appeal of Hawkes processes is that it is often easy to give a convincing interpretation to such modeling. To do so, the branching structure of Hawkes processes is quite helpful. Recall that under the assumption $\|\phi\|_1 < 1$, where $\|\phi\|_1$ denotes the L^1 norm of ϕ, Hawkes processes can be represented as a population process where migrants arrive according to a Poisson process with parameter μ. Then each migrant gives birth to children according to a non homogeneous Poisson process with intensity function ϕ, these children also giving birth to children according to the same non homogeneous Poisson process, see [HO74]. Now consider for example the classical case of buy (or sell) market orders, as studied in several of the papers mentioned above. Then migrants can be seen as exogenous orders whereas children are viewed as orders triggered by other orders.

Beyond enabling to build this population dynamics interpretation, the assumption $\|\phi\|_1 < 1$ is crucial in the study of Hawkes processes. To fix ideas, let us place ourselves in the classical framework where the Hawkes process (N_t) starts at $-\infty$. In that case, if one wants to get a stationary intensity with finite first moment, then the condition $\|\phi\|_1 < 1$ is required. Furthermore, even in the non stationary setting, this condition seems to be necessary in order to obtain classical ergodic properties for the process, see [BDHM12]. For these reasons, this condition is often called stability condition in the Hawkes literature.

From a practical point of view, a lot of interest has been recently devoted to the parameter $\|\phi\|_1$. For example, Hardiman, Bercot and Bouchaud, see [HBB13], and Filimonov and Sornette, see [FS12, FS13], use the branching interpretation of Hawkes processes on midquote data in order to measure the so-called degree of endogeneity of the market. This degree is simply defined by $\|\phi\|_1$, which is also called branching ratio. The intuition behind this interpretation of $\|\phi\|_1$ goes as follows: The parameter $\|\phi\|_1$ corresponds to the average number of children of an individual, $\|\phi\|_1^2$ to the average number of grandchildren of an individual,... Therefore, if we call cluster the descendants of a migrant, then the average size of a cluster is given by $\sum_{k\geq 1}\|\phi\|_1^k = \|\phi\|_1/(1-\|\phi\|_1)$. Thus, in the financial interpretation, the average proportion of endogenously triggered events is $\|\phi\|_1/(1-\|\phi\|_1)$ divided by $1+\|\phi\|_1/(1-\|\phi\|_1)$, which is equal to $\|\phi\|_1$.

This branching ratio can be measured using parametric and non parametric estimation methods for Hawkes processes, see [Oga78, Oga83] for likelihood based methods and [BDM12, RBS$^+$10] for functional estimators of the function ϕ. In [HBB13], very stable estimations of $\|\phi\|_1$ are reported for the E mini S&P futures between 1998 and 2012, the results being systematically close to one. In [FS12], values of order 0.7-0.8 are obtained on several assets. A debate on the validity of these results is currently ongoing between the two groups. In particular, it is argued in [HBB13] that the choice of exponential kernels in [FS12] may lead to spurious results, whereas various bias that could affect the study in [HBB13] are underlined in [FS13]. In any case, we can remark that both groups find values close to one for $\|\phi\|_1$, which is consistent with the results of [BDM12], where estimations are performed on Bund and Dax futures.

This seemingly persistent statistical result should definitely worry users of Hawkes processes. Indeed, it is rarely suitable to apply a statistical model where the parameters are pushed to their limits. In fact, these obtained values for $\|\phi\|_1$ on empirical data are not really surprising. Indeed, one of the most well documented stylized fact in high frequency finance is the persistence (or long memory) in flows and market activity measures, see for example [BGPW04, LF04]. Usual Hawkes processes, in the same way as autoregressive processes, can only exhibit short range dependence, failing to reproduce this classical empirical feature, see Chapter VIII for details.

In spite of their relative inadequacy with market data, Hawkes processes possess so many appealing properties that one could still try to apply them in some specific situations. In [HBB13], it is suggested to use the without ancestors version of Hawkes processes introduced by Brémaud and Massoulié in [BM01]. For such processes, $\|\phi\|_1 = 1$ but, in order to preserve stationarity and a finite expectation for the intensity, one needs to have $\mu = 0$. This is probably a relevant approach. However setting the parameter μ to 0 is not completely satisfying since this parameter has a nice interpretation (exogenous orders). Moreover it is not found to be equal to zero in practice, see [HBB13]. Finally, a time-varying μ is an easy way to reproduce seasonalities observed on the market, see [BM14a] (however, for simplicity, we work in this paper with a constant $\mu > 0$).

These empirical measures of $\|\phi\|_1$, close to one, are the starting point of this work. Indeed, our aim is to study the behavior at large time scales of nearly unstable Hawkes processes, which correspond to these estimations. More precisely, we consider a sequence of Hawkes processes observed on $[0, T]$, where T goes to infinity. In the case of a fixed kernel (not depending on T) with norm strictly smaller than one, scaling limits of Hawkes processes have been investigated in [BDHM12]. In this framework, Bacry et al. obtain a deterministic limit for the properly normalized sequence of Hawkes processes, as it is the case for suitably rescaled Poisson processes. In their price model consisting in the difference of two Hawkes processes, a Brownian motion (with some volatility) is found at the limit. These two results are in fact quite intuitive. Indeed, in the same way as Poisson processes and autoregressive models, Hawkes processes enjoy short memory properties. In this work, we show that when the Hawkes processes are nearly unstable, these weakly dependent-like behaviors are no longer observed at intermediate time scales. To do so, we consider that the kernels of the Hawkes processes depend on T. More precisely, we translate the near instability condition into the assumption that the norm of the kernels tends to one as the observation scale T goes to infinity.

Our main theorem states that when the norm of the kernel tends to one at the right speed (meaning that the observation scale and kernel's norm balance in a suitable way), the limit of our sequence of Hawkes processes is no longer a deterministic process, but an integrated Cox Ingersoll Ross process (CIR for short), as introduced in [CIR85]. In practice, it means that when observing a Hawkes process with kernel's norm close to one at appropriate time scale, it looks like an integrated CIR. Furthermore, for the price model defined in [BDHM13], in the limit, the Brownian motion obtained in [BDHM12] is replaced by a Heston model, see [Hes93] for definition. This is probably more in agreement with empirical data.

The paper is organized as follows. The assumptions and main results, notably the convergence towards an integrated CIR are given in Section 2. The case of the difference of two Hawkes processes is studied in Section 3. The proofs are relegated to Section 4 except some auxiliary results which can be found in the appendix of [JR13].

2 Scaling limits of nearly unstable Hawkes processes

We give in this section our main results about the limiting behavior of a sequence of nearly unstable Hawkes processes. We start by presenting our assumptions and defining our asymptotic setting.

2.1 Assumptions and asymptotic framework

We consider a sequence of point processes $(N_t^T)_{t\geq 0}$ indexed by T[1]. For a given T, (N_t^T) satisfies $N_0^T = 0$ and the process is observed on the time interval $[0, T]$. Furthermore, our asymptotic

[1]Of course by T we implicitly means T_n with $n \in \mathbb{N}$ tending to infinity.

setting is that the observation scale T goes to infinity. The intensity process (λ_t^T) is defined for $t \geq 0$ by

$$\lambda_t^T = \mu + \int_0^t \phi^T(t-s) dN_s^T,$$

where μ is a positive real number and ϕ^T a non negative measurable function on $\mathbb{R}^+$ which satisfies $\|\phi\|_1 < +\infty$. For a given T, the process (N_t^T) is defined on a probability space $(\Omega^T, \mathscr{F}^T, \mathbb{P}^T)$ equipped with the filtration $(\mathscr{F}_t^T)_{t\in[0,T]}$, where $\mathscr{F}_t^T$ is the σ-algebra generated by $(N_s^T)_{s\leq t}$. Moreover we assume that for any $0 \leq a < b \leq T$ and $A \in \mathscr{F}_a^T$

$$\mathbb{E}[(N_b^T - N_{a-}^T)1_A] = \mathbb{E}[\int_a^b \lambda_s^T 1_A ds],$$

which sets λ^T as the intensity of N^T. In particular, if we denote by $(J_n^T)_{n\geq 1}$ the jump times of (N_t^T), the process

$$N^T_{t\wedge J_n^T} - \int_0^{t\wedge J_n^T} \lambda_s^T ds$$

is a martingale and the law of N^T is characterized by λ^T. From Jacod [Jac75], such construction can be done. The process N^T is called a Hawkes process.

Let us now give more specific assumptions on the function ϕ^T. We denote by $\|.\|_\infty$ the L^∞ norm on $\mathbb{R}^+$.

Assumption 1. *For $t \in \mathbb{R}^+$,*

$$\phi^T(t) = a_T \phi(t),$$

where $(a_T)_{T\geq 0}$ is a sequence of positive numbers converging to one such that for all T, $a_T < 1$ and ϕ is a non negative measurable function such that

$$\int_0^{+\infty} \phi(s) ds = 1 \text{ and } \int_0^{+\infty} s\phi(s) ds = m < \infty.$$

Moreover, ϕ is differentiable with derivative ϕ' such that $\|\phi'\|_\infty < +\infty$ and $\|\phi'\|_1 < +\infty$.

Remark 1. *Note that under Assumption 1, $\|\phi\|_\infty$ is finite.*

Thus, the form of the function ϕ^T depends on T so that its shape is fixed but its L^1 norm varies with T. For a given T, this L^1 norm is equal to a_T and so is smaller than one, implying that the stability condition is in force. Note that in this framework, we have almost surely no explosion[2]:

$$\lim_{n\to+\infty} J_n^T = +\infty.$$

However, remark that we do not work in the stationary setting since our process starts at time $t = 0$ and not at $t = -\infty$.

[2]In fact, for a Hawkes process, the no explosion property can be obtained under weaker conditions, for example $\int_0^t \phi(s) ds < \infty$ for any $t > 0$, see [BDHM12].

The case where $\|\phi^T\|_1$ is larger than one corresponds to the situation where the stability condition is violated. Since $a_T = \|\phi^T\|_1 < 1$ tends to one, our framework is a way to get close to instability. Therefore we call our processes nearly unstable Hawkes processes. There are of course many other ways to make the L^1 norm of ϕ^T converge to one than the multiplicative manner used here. However, this parametrization is sufficient for applications and very convenient to illustrate the different regimes that can be obtained.

2.2 Observation scales

In our framework, two parameters degenerate at infinity: T and $(1-a_T)^{-1}$. The relationship between these two sequences will determine the scaling behavior of the sequence of Hawkes processes. Recall that it is shown in [BDHM12] that when $\|\phi\|_1$ is fixed and smaller than one, after appropriate scaling, the limit of the sequence of Hawkes processes is deterministic, as it is for example the case for Poisson processes. In our setting, if $1-a_T$ tends "slowly" to zero, we can expect the same result. Indeed, we may have T large enough so that we reach the asymptotic regime and for such T, a_T is still sufficiently far from unity. This is precisely what happens, as stated in the next theorem.

Theorem 1. *Assume $T(1-a_T) \to +\infty$. Then, under Assumption 1, the sequence of Hawkes processes is asymptotically deterministic, in the sense that the following convergence in L^2 holds:*

$$\sup_{v\in[0,1]} \frac{1-a_T}{T} |N^T_{Tv} - \mathbb{E}[N^T_{Tv}]| \to 0.$$

On the contrary, if $1-a_T$ tends too rapidly to zero, the situation is likely to be quite intricate. Indeed, for given T, the Hawkes process may already be very close to instability whereas T is not large enough to reach the asymptotic regime. The last case, which is probably the most interesting one, is the intermediate case, where $1-a_T$ tends to zero in such a manner that a non deterministic scaling limit is obtained, while not being in the preceding degenerate setting. We largely detail this situation in the next subsection.

2.3 Non degenerate scaling limit for nearly unstable Hawkes processes

We give in this section our main result, that is a non degenerate scaling limit for a sequence of properly renormalized nearly unstable Hawkes processes. Before giving this theorem, we wish to provide intuitions on how it is derived. Let M^T be the martingale process associated to N^T, that is, for $t \geq 0$,

$$M^T_t = N^T_t - \int_0^t \lambda^T_s \, ds.$$

We also set ψ^T the function defined on $\mathbb{R}^+$ by

$$\psi^T(t) = \sum_{k=1}^{\infty} (\phi^T)^{*k}(t),$$

where $(\phi^T)^{*1} = \phi^T$ and for $k \geq 2$, $(\phi^T)^{*k}$ denotes the convolution product of $(\phi^T)^{*(k-1)}$ with the function ϕ^T. Note that $\psi^T(t)$ is well defined since $\|\phi^T\|_1 < 1$. In the sequel, it will be

convenient to work with another form for the intensity. We have the following result, whose proof is given in Section 4.

Proposition 1. *For all $t \geq 0$, we have*

$$\lambda_t^T = \mu + \int_0^t \psi^T(t-s)\mu ds + \int_0^t \psi^T(t-s)dM_s^T.$$

Now recall that we observe the process (N_t^T) on $[0,T]$. In order to be able to give a proper limit theorem, where the processes live on the same time interval, we rescale our processes so that they are defined on $[0,1]$. To do that, we consider for $t \in [0,1]$

$$\lambda_{tT}^T = \mu + \int_0^{tT} \psi^T(Tt-s)\mu ds + \int_0^{tT} \psi^T(Tt-s)dM_s^T.$$

For the scaling in space, a natural multiplicative factor is $(1-a_T)$. Indeed, in the stationary case, the expectation of λ_t^T is $\mu/(1-\|\phi^T\|_1)$. Thus, the order of magnitude of the intensity is $(1-a_T)^{-1}$. This is why we define

$$C_t^T = \lambda_{tT}^T(1-a_T). \tag{1}$$

Understanding the asymptotic behavior of C_t^T will be the key to the derivation of a suitable scaling limit for our sequence of renormalized processes. We will see that this behavior is closely connected to that of the function ψ^T. About ψ^T, one can first remark that the function defined for $x \geq 0$ by

$$\rho^T(x) = T\frac{\psi^T}{\|\psi^T\|_1}(Tx) \tag{2}$$

is the density of the random variable

$$X^T = \frac{1}{T}\sum_{i=1}^{I^T} X_i,$$

where the (X_i) are iid random variables with density ϕ and I^T is a geometric random variable with parameter $1-a_T$ ($\forall k > 0, \mathbb{P}[I^T = k] = (1-a_T)(a_T)^{k-1}$). Now let $z \in \mathbb{R}$. The characteristic function of the random variable X^T, denoted by $\widehat{\rho}^T$, satisfies

$$\begin{aligned}\widehat{\rho}^T(z) = \mathbb{E}[e^{izX^T}] &= \sum_{k=1}^{\infty}(1-a_T)(a_T)^{k-1}\mathbb{E}[e^{i\frac{z}{T}\sum_{i=1}^k X_i}] \\ &= \sum_{k=1}^{\infty}(1-a_T)(a_T)^{k-1}(\hat{\phi}(\frac{z}{T}))^k = \frac{\hat{\phi}(\frac{z}{T})}{1-\frac{a_T}{1-a_T}(\hat{\phi}(\frac{z}{T})-1)},\end{aligned}$$

where $\hat{\phi}$ denotes the characteristic function of X_1. Since

$$\int_0^{+\infty} s\phi(s)ds = m < \infty,$$

the function $\hat{\phi}$ is continuously differentiable with first derivative at point zero equal to im. Therefore, using that a_T and $\hat{\phi}(\frac{z}{T})$ both tend to one as T goes to infinity, $\widehat{\rho}^T(z)$ is equivalent to

$$\frac{1}{1-\frac{izm}{T(1-a_T)}}.$$

Thus, we precisely see here that the suitable regime so that we get a non trivial limiting law for X^T is that there exists $\lambda > 0$ such that

$$T(1-a_T) \underset{T\to+\infty}{\to} \lambda. \tag{3}$$

When (3) holds, we write $d_0 = m/\lambda$. In fact we have just proved the following result.

Proposition 2. *Assume that* (3) *holds. Under Assumption 1, the sequence of random variable X^T converges in law towards an exponential random variable with parameter* $1/d_0$.

This simple result is of course not new. For example this type of geometric sums of random variable is studied in detail in [Kal97]. Note also that when X_1 is exponentially distributed, X^T is also exponentially distributed, even for a fixed T.

Assume from now on that (3) holds and set $u_T = T(1-a_T)/\lambda$ (so that u_T goes to one). Proposition 2 is particularly important since it gives us the asymptotic behavior of ψ^T in this setting. Indeed, it tells us that

$$\psi^T(Tx) = \rho^T(x)\frac{a_T}{\lambda u_T} \approx \frac{\lambda}{m}e^{-x\frac{\lambda}{m}}\frac{1}{\lambda} = \frac{1}{m}e^{-x\frac{\lambda}{m}}.$$

Let us now come back to the process C_t^T, which can be written

$$C_t^T = (1-a_T)\mu + \mu\int_0^t u_T\lambda\psi^T(Ts)ds + \int_0^t \sqrt{\lambda}\psi^T(T(t-s))\sqrt{C_s^T}dB_s^T, \tag{4}$$

with

$$B_t^T = \frac{1}{\sqrt{T}}\sqrt{u_T}\int_0^{tT}\frac{dM_s^T}{\sqrt{\lambda_s^T}}. \tag{5}$$

By studying its quadratic variation, we will show that B^T represents a sequence of martingales which converges to a Brownian motion. So, heuristically replacing B^T by a Brownian motion B and $\psi^T(Tx)$ by $\frac{1}{m}e^{-x\frac{\lambda}{m}}$ in (4), we get

$$C_t^\infty = \mu(1-e^{-t\frac{\lambda}{m}}) + \frac{\sqrt{\lambda}}{m}\int_0^t e^{-(t-s)\frac{\lambda}{m}}\sqrt{C_s^\infty}dB_s.$$

Applying Itô's formula, this gives

$$C_t^\infty = \int_0^t(\mu - C_s^\infty)\frac{\lambda}{m}ds + \frac{\sqrt{\lambda}}{m}\int_0^t\sqrt{C_s^\infty}dB_s,$$

which precisely corresponds to the stochastic differential equation (SDE) satisfied by a CIR process.

Before stating the theorem which makes the preceding heuristic derivation rigorous, we consider an additional assumption.

Assumption 2. *There exists $K_\rho > 0$ such that for all $x \geq 0$ and $T > 0$,*

$$|\rho^T(x)| \leq K_\rho.$$

Note that Assumption 2 is in fact not really restrictive. Indeed, from [Pet75] (Page 214, point 5), we get that if $\|\phi\|_\infty < \infty$ and $\int_0^{+\infty} |s|^3 \phi(s) ds < +\infty$, then Assumption 2 holds. From [Kal97] (Chapter 5, Lemma 4.1), it also holds if the random variable X_1 with density ϕ can be written (in law) under the form $X_1 = E + Y$, where E follows an exponential law with parameter $\gamma > 0$ and Y is independent of E. We now give our main theorem.

Theorem 2. *Assume that* (3) *holds. Under Assumptions 1 and 2, the sequence of renormalized Hawkes intensities (C_t^T) defined in* (1) *converges in law, for the Skorohod topology, towards the law of the unique strong solution of the following Cox Ingersoll Ross stochastic differential equation on* $[0,1]$*:*

$$X_t = \int_0^t (\mu - X_s) \frac{\lambda}{m} ds + \frac{\sqrt{\lambda}}{m} \int_0^t \sqrt{X_s} dB_s.$$

Furthermore, the sequence of renormalized Hawkes process

$$V_t^T = \frac{1 - a_T}{T} N_{tT}^T$$

converges in law, for the Skorohod topology, towards the process

$$\int_0^t X_s ds, \ \ t \in [0,1].$$

2.4 Discussion

- Theorem 2 implies that when $\|\phi\|_1$ is close to 1, if the observation time T is suitably chosen (that is of order $1/(1 - \|\phi\|_1)$), a non degenerate behavior (neither explosive, nor deterministic) can be obtained for a rescaled Hawkes process.

- This can for example be useful for the statistical estimation of the parameters of a Hawkes process. Indeed, designing an estimating procedure based on the fine scale properties of a Hawkes process is a very hard task: Non parametric methods are difficult to use and present various instabilities, see [BDM12, FS13], whereas parametric approaches are of course very sensitive to model specifications, see [FS13, HBB13]. Considering an intermediate scale, where the process behaves like a CIR model, one can use statistical methods specifically developed in order to estimate CIR parameters, see [AK12] for a survey. Of course, only the parameters λ, m and μ can be recovered this way. Therefore,

there is clearly an information loss in this approach. However, it still enables to get access to quantities which are important in practice, see Section 1. In some sense, it can be compared to the extreme value theory based method for extreme quantile estimation, where one assumes that the random variables of an iid sample belong to some max stable attraction domain. Indeed, these two methods lie between a fully parametric one, where a parametric form is assumed (for the law of the random variables or the function ϕ), and a fully non parametric one, where a functional estimator (of the repartition function or of ϕ) is used in order to reach the quantity of interest (the quantile or the L^1 norm of ϕ).

- CIR processes are a very classical way to model stochastic (squared) volatilities in finance, see the celebrated Heston model [Hes93]. Also, it is widely acknowledged that there exists a linear relationship between the cumulated order flow and the integrated squared volatility, see for example [WBK+08]. Therefore, our setting where $\|\phi\|_1$ is close to one and the limiting behavior obtained in Theorem 2 seem in good agreement with market data.

- For the stationary version of a Hawkes process, one can show that the variance of N_T^T is of order $T(1-\|\phi^T\|_1)^{-3}$, see for example [BM01]. Therefore, if $T(1-a_T)$ tends to zero, that is $\|\phi^T\|_1$ goes rapidly to one, then the variance of $\frac{(1-a_T)}{T}N_T^T$ blows up as T goes to infinity. This situation is therefore very different from the one studied here and out of the scope of this paper.

- The assumption $\int_0^{+\infty} s\phi(s)ds < +\infty$ is crucial in order to approximate ψ^T by an exponential function using Proposition 2. Let us now consider the fat tail case where the preceding integral is infinite. More precisely, let us take a function ϕ which is of order $\frac{1}{x^{1+\alpha}}$, $0<\alpha<1$, as x goes to infinity. In that case, following the proof of Proposition 2, we can show the following result, where we borrow the notations of Proposition 2.

  Proposition 3. *Let E_C^α be a random variable whose characteristic function satisfies*

  $$\mathbb{E}[e^{izE_C^\alpha}] = \frac{1}{1-C(iz)^\alpha}.$$

  Assume $\hat{\phi}(z) - 1 \sim_0 \sigma(iz)^\alpha$ for some $\sigma > 0$, $0<\alpha<1$, and $(1-a_T)T^\alpha \to \lambda > 0$. Then X^T converges in law towards the random variable $E_{\frac{\sigma}{\lambda}}^\alpha$.

  Thus, when the shape of the kernel is of order $x^{-(1+\alpha)}$, the "right" observation scale is no longer $T \sim 1/(1-\|\phi\|_1)$ but $T \sim 1/(1-\|\phi\|_1)^{\frac{1}{\alpha}}$. Remark also that if we denote by $E_{\alpha,\beta}$ the (α,β) Mittag-Leffler function, that is

  $$E_{\alpha,\beta}(z) = \sum_{n=1}^{\infty} \frac{z^n}{\Gamma(\alpha n+\beta)},$$

  see for example [SJ13], then the density ϕ_C^α of E_C^α is linked to this function since

  $$\phi_1^\alpha(x) = x^{\alpha-1}E_{\alpha,\alpha}(-x^\alpha).$$

Now let us consider the asymptotic setting where $\mu^T = \mu T^{\alpha-1}$,$\phi^T = a_T\phi$ with $a_T = 1 - \frac{\lambda}{T^\alpha}$ and ϕ as in Proposition 3. If we apply the same heuristic arguments as those used in Section 2 to the renormalized intensity

$$C_t^T = \frac{\lambda_{tT}^T(1-a_T)}{T^{\alpha-1}},$$

we get the following type of limiting law for our sequence of Hawkes intensities:

$$X_t = \mu\int_0^t \phi_{\frac{\alpha}{\lambda}}^\alpha(t-s)ds + \int_0^t \phi_{\frac{\alpha}{\lambda}}^\alpha(t-s)\frac{1}{\sqrt{\lambda}}\sqrt{X_s}dB_s.$$

These heuristic arguments are however far from a proof. Indeed, in this case, we probably have to deal with a non semi-martingale limit. Furthermore, tightness properties which are important in the proofs of this paper are much harder to show (in particular the function ϕ_C^α is not bounded). We leave this case for further research.

- In the classical time series setting let us mention the paper [BIP11] where the authors study the asymptotic behavior of unstable integer-valued autoregressive model (INAR processes). In this case, CIR processes also appear in the limit. This is in fact not so surprising since INAR processes share some similarities with Hawkes processes. In particular, they can somehow be viewed as Hawkes processes for which the kernel would be a sum of Dirac functions.

3 Extension of Theorem 2 to a price model

In the previous section, we have studied one-dimensional nearly unstable Hawkes processes. For financial applications, they can for example be used to model the arrival of orders when the number of endogenous orders is much larger than the number of exogenous orders, which seems to be the case in practice, see [FS12, HBB13]. In this section, we consider the high frequency price model introduced in [BDHM13], which is essentially defined as a difference of two Hawkes processes. Using the same approach as for Theorem 2, we investigate the limiting behavior of this model when the stability condition is close to saturation.

3.1 A Hawkes based price model

In [BDHM13], tick by tick moves of the midprice $(P_t)_{t\geq 0}$ are modeled thanks to a two dimensional Hawkes process in the following way: For $t \geq 0$,

$$P_t = N_t^+ - N_t^-,$$

where (N^+, N^-) is a two dimensional Hawkes process with intensity

$$\begin{pmatrix}\lambda_t^+\\ \lambda_t^-\end{pmatrix} = \begin{pmatrix}\mu\\ \mu\end{pmatrix} + \int_0^t \begin{pmatrix}\phi_1(t-s) & \phi_2(t-s)\\ \phi_2(t-s) & \phi_1(t-s)\end{pmatrix}\begin{pmatrix}dN_s^+\\ dN_s^-\end{pmatrix},$$

with ϕ_1 and ϕ_2 two non negative measurable functions such that the stability condition

$$\int_0^{+\infty} \phi_1(s)ds + \int_0^{+\infty} \phi_2(s)ds < 1$$

is satisfied.

This model takes into account the discreteness and the negative autocorrelation of prices at the microstructure level. Moreover, it is shown in [BDHM12] that when one considers this price at large time scales, the stability condition implies that after suitable renormalization, it converges towards a Brownian motion (with a given volatility).

3.2 Scaling limit

In the same spirit as in Section 2, we consider the scaling limit of the Hawkes based price process when the stability condition becomes almost violated. More precisely, following the construction of multivariate Hawkes processes of [BDHM12], for every observation interval $[0,T]$, we define the Hawkes process (N^{T+}, N^{T-}) with intensity

$$\begin{pmatrix} \lambda_t^{T+} \\ \lambda_t^{T-} \end{pmatrix} = \begin{pmatrix} \mu \\ \mu \end{pmatrix} + \int_0^t \begin{pmatrix} \phi_1^T(t-s) & \phi_2^T(t-s) \\ \phi_2^T(t-s) & \phi_1^T(t-s) \end{pmatrix} \begin{pmatrix} dN_s^{T+} \\ dN_s^{T-} \end{pmatrix},$$

with ϕ_1^T and ϕ_2^T two non negative measurable functions. Note that in this construction, N^{T+} and N^{T-} do not have common jumps, see [BDHM12] for details. We consider the following assumption.

Assumption 3. *For $i = 1,2$ and $t \in \mathbb{R}^+$,*

$$\phi_i^T(t) = a_T \phi_i(t),$$

where $(a_T)_{T\geq 0}$ is a sequence of positive numbers converging to one such that for all T, $a_T < 1$ and ϕ_1 and ϕ_2 are two non negative measurable functions such that

$$\int_0^{+\infty} \phi_1(s) + \phi_2(s)ds = 1 \text{ and } \int_0^{+\infty} s\big(\phi_1(s) + \phi_2(s)\big)ds = m < \infty.$$

Moreover, for $i = 1,2$, ϕ_i is differentiable with derivative ϕ_i' such that $\|\phi_i'\|_\infty < +\infty$ and $\|\phi_i'\|_1 < +\infty$.

We will also make the following technical assumption.

Assumption 4. *Let*

$$\psi_+^T = \sum_{k\geq 1} (a_T(\phi_1 + \phi_2))^{*k} \text{ and } \rho^T(x) = T\frac{\psi_+^T}{\|\psi_+^T\|_1}(Tx).$$

There exists $K_\rho > 0$ such that for all $x \geq 0$ and $T > 0$,

$$|\rho^T(x)| \leq K_\rho.$$

We work with the renormalized price process

$$P_t^T = \frac{1}{T}(N_{Tt}^{T+} - N_{Tt}^{T-}). \tag{6}$$

The following theorem states that if we consider the rescaled price process over the right time interval, that is if we take T of order $1/(1-\|\phi_1\|_1 - \|\phi_1\|_2)$, it asymptotically behaves like a Heston model, see [Hes93].

Theorem 3. *Let $\phi = \phi_1 - \phi_2$. Assume that (3) holds. Under Assumptions 3 and 4, the sequence of Hawkes based price models (P_t^T) converges in law, for the Skorohod topology, towards a Heston type process P on $[0,1]$ defined by:*

$$\begin{cases} dC_t = (\frac{2\mu}{\lambda} - C_t)\frac{\lambda}{m}dt + \frac{1}{m}\sqrt{C_t}dB_t^1 & C_0 = 0 \\ dP_t = \frac{1}{1-\|\phi\|_1}\sqrt{C_t}dB_t^2 & P_0 = 0, \end{cases}$$

with (B^1, B^2) a bidimensional Brownian motion.

4 Proofs

We gather in this section the proofs of Theorem 1, Proposition 1, Theorem 2 and Theorem 3. In the following, c denotes a constant that may vary from line to line.

4.1 Proof of Theorem 1

Let $v \in [0,1]$. From Lemma 4 in [BDHM12], we get

$$\mathbb{E}[N_{Tv}^T] = \mu T v + \mu \int_0^{Tv} \psi^T(Tv - s)s\,ds$$

and

$$N_{Tv}^T - \mathbb{E}[N_{Tv}^T] = M_{Tv}^T + \int_0^{Tv} \psi^T(Tv - s)M_s^T\,ds.$$

Thus, using that

$$\|\psi^T\|_1 = \frac{\|\phi^T\|_1}{1-\|\phi^T\|_1},$$

we deduce

$$\frac{1-\|\phi^T\|_1}{T}(N_{Tv}^T - \mathbb{E}[N_{Tv}^T]) \le \frac{1-\|\phi^T\|_1}{T}(1+\|\psi^T\|_1)\sup_{t\in[0,T]}|M_t^T| \le \frac{1}{T}\sup_{t\in[0,T]}|M_t^T|.$$

Now recall that M^T is a square integrable martingale with quadratic variation process N^T. Thus we can apply Doob's inequality which gives

$$\mathbb{E}[(\sup_{t\in[0,T]} M_t^T)^2] \le 4\sup_{t\in[0,T]}\mathbb{E}[(M_t^T)^2] \le 4\mathbb{E}[N_T^T] \le 4\mu\frac{T}{1-\|\phi^T\|_1}.$$

Therefore, we finally obtain

$$\mathbb{E}\Big[\sup_{v\in[0,1]}\Big(\frac{1-\|\phi^T\|_1}{T}(N^T_{Tv}-\mathbb{E}[N^T_{Tv}])\Big)^2\Big]\leq\frac{4\mu}{T(1-\|\phi^T\|_1)},$$

which gives the result since $T(1-\|\phi^T\|_1)$ tends to infinity.

4.2 Proof of Proposition 1

From the definition of λ^T, using the fact that ϕ is bounded on $[0,t]$, we can write

$$\lambda^T_t=\mu+\int_0^t\phi^T(t-s)dM^T_s+\int_0^t\phi^T(t-s)\lambda^T_s ds.$$

We now recall the following classical lemma, see for example [BDHM12] for a proof.

Lemma 1. *If $f(t)=h(t)+\int_0^t\phi^T(t-s)f(s)ds$ with h a measurable locally bounded function, then*

$$f(t)=h(t)+\int_0^t\psi^T(t-s)h(s)ds.$$

We apply this lemma to the function h defined by

$$h(t)=\mu+\int_0^t\phi^T(t-s)dM^T_s.$$

Thus, we obtain

$$\lambda^T_t=\mu+\int_0^t\phi^T(t-s)dM^T_s+\int_0^t\psi^T(t-s)\big(\mu+\int_0^s\phi^T(s-r)dM^T_r\big)ds. \tag{7}$$

Now remark that using Fubini theorem and the fact that

$$\psi^T*\phi^T=\psi^T-\phi^T,$$

we get

$$\begin{aligned}\int_0^t\psi^T(t-s)\int_0^s\phi^T(s-r)dM^T_r ds&=\int_0^t\int_0^t 1_{r\leq s}\psi^T(t-s)\phi^T(s-r)dsdM^T_r\\&=\int_0^t\int_0^{t-r}\psi^T(t-r-s)\phi^T(s)dsdM^T_r\\&=\int_0^t\psi^T*\phi^T(t-r)dM^T_r\\&=\int_0^t\psi^T(t-r)dM^T_r-\int_0^t\phi^T(t-r)dM^T_r.\end{aligned}$$

We conclude the proof rewriting (7) using this last equality.

4.3 Proof of Theorem 2

Before starting the proof of Theorem 2, we give some preliminary lemmas.

4.3.1 Preliminary lemmas

We start with some lemmas on ϕ and its Fourier transform $\hat{\phi}$ (the associated characteristic function).

Lemma 2. *Let $\delta > 0$. There exists $\varepsilon > 0$ such that for any real number z with $|z| \geq \delta$,*

$$|1 - \hat{\phi}(z)| \geq \varepsilon.$$

Proof. Since ϕ is bounded, $\hat{\phi}(z)$ tends to zero as z tends to infinity. Consequently, there exists $b > \delta$ such that for all z such that $|z| \geq b$,

$$|\hat{\phi}(z)| \leq \frac{1}{2}.$$

Now, let M denote the supremum of the real part of $\hat{\phi}$ on $[-b, -\delta] \cup [\delta, b]$, since $\hat{\phi}$ is continuous this supremum is attained at some point z_0. We have $M = \mathrm{Re}(\hat{\phi}(z_0)) = \mathbb{E}[\cos(z_0 X)]$, with X a random variable with density ϕ. Since ϕ is continuous, almost surely, X does not belong to $2\pi/z_0\mathbb{Z}$. Thus $M = \mathbb{E}[\cos(z_0 X)] < 1$. Therefore, taking $\varepsilon = \min(\frac{1}{2}, 1 - M)$ we have the lemma. □

Using that $\|\phi'\|_1 < +\infty$, integrating by parts, we immediately get the following lemma.

Lemma 3. *Let $z \in \mathbb{R}$. We have $|\hat{\phi}(z)| \leq c/|z|$.*

We now turn to the function ρ^T defined in (2). We have the following result.

Lemma 4. *There exist $c > 0$ such that for all real z and $T \geq 1$,*

$$|\widehat{\rho^T}(z)| \leq c(1 \wedge |\frac{1}{z}|).$$

Proof. First note that as the Fourier transform of a random variable, $|\widehat{\rho^T}| \leq 1$. Furthermore, using Lemma 2 together with the fact that

$$\int_0^{+\infty} x\phi(x)\,dx = m < +\infty,$$

we get that there exist $\delta > 0$ and $\varepsilon > 0$ such that if $|x| \leq \delta$,

$$|\mathrm{Im}(\hat{\phi})(x)| \geq \frac{m}{2}|x|$$

and if $|x| \geq \delta$,

$$|1 - \hat{\phi}(x)| \geq \varepsilon.$$

Therefore, we deduce that if $|z/T| \le \delta$,

$$|\widehat{\rho^T}(z)| = |\frac{(1-a_T)\hat{\phi}(\frac{z}{T})}{1-a_T\hat{\phi}(\frac{z}{T})}| \le \frac{(1-a_T)}{a_T|\mathrm{Im}(\hat{\phi})(\frac{z}{T})|} \le \frac{2(1-a_T)T}{a_T m|z|} \le c/|z|$$

and, thanks to Lemma 3, if $|z/T| \ge \delta$

$$|\widehat{\rho^T}(z)| \le \frac{(1-a_T)|\hat{\phi}(\frac{z}{T})|}{|1-\hat{\phi}(\frac{z}{T})|} \le \frac{c(1-a_T)T}{|z|\varepsilon} \le c/|z|.$$

□

The next lemma gives us the L^2 convergence of ρ^T.

Lemma 5. *Let $\rho(x) = \frac{\lambda}{m}e^{-\frac{x\lambda}{m}}$ be the density of the exponential random variable with parameter λ/m. We have the following convergence, where $|.|_2$ denotes the L^2 norm on $\mathbb{R}^+$:*

$$|\rho^T - \rho|_2 \to 0.$$

Proof. Using the Fourier isometry, we get

$$|\rho^T - \rho|_2 = \frac{1}{2\pi}|\widehat{\rho^T} - \widehat{\rho}|_2.$$

From Proposition 2, for given z, we have $(\widehat{\rho^T}(z) - \widehat{\rho}(z)) \to 0$. Thanks to Lemma 4, we can apply the dominated convergence theorem which gives that this convergence also takes place in L^2. □

We now give a Lipschitz type property for ρ^T.

Lemma 6. *There exists $c > 0$ such that for all $x \ge 0$, $y \ge 0$ and $T \ge 1$,*

$$|\rho^T(x) - \rho^T(y)| \le cT|x-y|.$$

Proof. We simply compute the derivative of ρ^T on $\mathbb{R}_+$, which is given by

$$(\rho^T)'(x) = T\big(\phi'(Tx)\frac{T}{\|\psi^T\|_1} + \phi' * \rho^T(Tx)\big).$$

Using that $\|\psi^T\|_1 = a_T/(1-a_T)$ together with the fact that $T(1-a_T) \to \lambda$, we get

$$|(\rho^T)'(x)| \le T(c\|\phi'\|_\infty + \|\phi'\|_1\|\rho^T\|_\infty).$$

□

We now consider the function f^T defined for $x \ge 0$ by

$$f^T(x) = \frac{m}{\lambda}\frac{a_T}{u_T}\rho^T(x) - e^{-\frac{x}{d_0}}.$$

We have the following obvious corollaries.

Corollary 1. *We have*

$$\int |f^T(x)|^2 dx \to 0.$$

Corollary 2. *There exists $c > 0$ such that for any $z \geq 0$,*

$$|f^T(z)| \leq c.$$

Corollary 3. *There exists $c > 0$ such that for any $z \geq 0$,*

$$|\widehat{f^T}(z)| \leq c(|\frac{1}{z}| \wedge 1).$$

Corollary 4. *There exists $c > 0$ such that for all $x \geq 0$, $y \geq 0$ and $T \geq 1$,*

$$|f^T(x) - f^T(y)| \leq cT|x - y|.$$

We finally give a lemma on the integrated difference associated to the function f^T.

Lemma 7. *For any $0 < \varepsilon < 1$, there exists c_ε so that for all $t, s \geq 0$,*

$$\int_{\mathbb{R}} (f^T(t-u) - f^T(s-u))^2 du \leq c_\varepsilon |t-s|^{1-\varepsilon}.$$

Proof. Defining $g^T_{t,s}(u) = f^T(t-u) - f^T(s-u)$, we easily get

$$|\widehat{g^T_{t,s}}(w)| = |e^{-iwt} - e^{-iws}||\widehat{f^T}(w)|.$$

Thus, from Corollary 3 together with the fact that

$$|\frac{e^{-iwt} - e^{-iws}}{w(t-s)}| \leq 1,$$

we get

$$\begin{aligned}
\int_{\mathbb{R}} (f^T(t-u) - f^T(s-u))^2 du &\leq c \int_{\mathbb{R}} |\widehat{g^T_{t,s}}(w)|^2 dw \\
&\leq c \int_{\mathbb{R}} |e^{-iwt} - e^{-iws}|^2 (|\frac{1}{w^2}| \wedge 1) dw \\
&\leq c \int_{\mathbb{R}} 2^{1+\varepsilon} |\frac{e^{-iwt} - e^{-iws}}{w(t-s)}|^{1-\varepsilon} (|\frac{1}{w^2}| \wedge 1) w^{1-\varepsilon} dw |t-s|^{1-\varepsilon} \\
&\leq c_\varepsilon |t-s|^{1-\varepsilon}.
\end{aligned}$$

□

4.3.2 Proof of the first part of Theorem 2

We now begin with the proof of the first assertion in Theorem 2. We split this proof into several steps.

Step 1: Convenient rewriting of C^T

In this step, our goal is to obtain a suitable expression for C_t^T. Let $d_0 = m/\lambda$. Inspired by the limiting behavior of ψ^T given in Proposition 2, we write Equation (4) under the form

$$C_t^T = R_t^T + \mu(1 - e^{-\frac{t}{d_0}}) + \frac{\sqrt{\lambda}}{m}\int_0^t e^{-\frac{t-s}{d_0}}\sqrt{C_s^T}\,dB_s^T,$$

where R_t^T is obviously defined. Using integration by parts (for finite variation processes), we get

$$C_t^T = R_t^T + \frac{\mu}{d_0}\int_0^t e^{-\frac{v}{d_0}}\,dv + \frac{\sqrt{\lambda}}{m}\int_0^t \sqrt{C_v^T}\,dB_v^T - \frac{\sqrt{\lambda}}{md_0}\int_0^t \Big(\int_0^v e^{-\frac{v-s}{d_0}}\sqrt{C_s^T}\,dB_s^T\Big)dv.$$

Then remarking that

$$\frac{\sqrt{\lambda}}{md_0}\int_0^v e^{-\frac{v-s}{d_0}}\sqrt{C_s^T}\,dB_s^T = \frac{1}{d_0}\big(C_v^T - R_v^T - \mu(1 - e^{-\frac{v}{d_0}})\big),$$

we finally derive

$$C_t^T = U_t^T + \frac{1}{d_0}\int_0^t (\mu - C_s^T)\,ds + \frac{\sqrt{\lambda}}{m}\int_0^t \sqrt{C_s^T}\,dB_s^T, \tag{8}$$

with

$$U_t^T = R_t^T + \frac{1}{d_0}\int_0^t R_s^T\,ds.$$

The form (8) will be quite convenient in order to study the asymptotic behavior of C_t^T. Indeed, we will show that U_t^T vanishes so that (8) almost represents a stochastic differential equation.

Step 2: Preliminaries for the convergence of U^T

We now want to prove that the sequence of processes $(U_t^T)_{t\in[0,1]}$ converges to zero in law, for the Skorohod topology, and therefore uniformly on compact sets on $[0,1]$ (ucp). We show here that to do so, it is enough to study a (slightly) simpler process than U^T. First, it is clear that showing the convergence of $(R_t^T)_{t\in[0,1]}$ to zero gives also the convergence of U^T. Now recall that

$$R_t^T = \mu(1-a_T) - \mu((1-e^{-\frac{t}{d_0}}) - \int_0^t a_T T\frac{\psi^T}{\|\psi^T\|_1}(Ts)ds + \sqrt{\lambda}\int_0^t \Big(\psi^T(T(t-s)) - \frac{1}{m}e^{-\frac{t-s}{d_0}}\Big)\sqrt{C_t^T}\,dB_s^T.$$

Since a_T tends to one, the first term tends to zero. For $t \in [0,1]$, Proposition 2 gives us the convergence of

$$\int_0^t a_T T\frac{\psi^T}{\|\psi^T\|_1}(Ts)\,ds$$

towards $1 - e^{-\frac{t}{d_0}}$. Using Dini's theorem, we get that this convergence is in fact uniform over $[0,1]$. Thus, using Equation (5), we see that it remains to show that $(Y_t^T)_{t\in[0,1]}$ goes to zero, with

$$Y_t^T = \int_0^t (m\psi^T(T(t-u)) - e^{-\frac{t-u}{d_0}})d\overline{M}_t^T,$$

where $\overline{M}_t^T = M_{tT}^T/T$.

Step 3: **Finite dimensional convergence of** Y^T

We now show the finite dimensional convergence of $(Y_t^T)_{t\in[0,1]}$.

Lemma 8. *For any* $(t_1,...,t_n) \in [0,1]^n$, *we have the following convergence in law:*

$$(Y_{t_1}^T,...,Y_{t_n}^T) \to 0.$$

Proof. First note that the quadratic variation of $\overline{M}^T$ at time t is given by N_{tT}^T/T^2, whose predicable compensator process at time t is simply equal to

$$\frac{1}{T^2}\int_0^{tT} \lambda_s^T ds.$$

Using this together with the fact that

$$\mathbb{E}[\lambda_t^T] = \mu + \mu\int_0^t \psi^T(t-s)ds \le \mu + \mu\frac{a_T}{1-a_T} \le cT,$$

we get

$$\mathbb{E}[(Y_t^T)^2] \le c\int_0^t (m\psi^T(T(t-s)) - e^{-\frac{t-s}{d_0}})^2 ds.$$

Now remark that

$$m\psi^T(T(t-s)) - e^{-\frac{t-s}{d_0}} = f^T(t-s),$$

where f^T is defined by $f^T(x) = 0$ for $x < 0$ and

$$f^T(x) = \frac{m}{\lambda}\frac{a_T}{u_T}\rho^T(x) - e^{-\frac{x}{d_0}}$$

for $x \ge 0$, with ρ^T the function introduced in Equation (2). From Corollary 1,

$$\mathbb{E}[(Y_t^T)^2] \to 0,$$

which gives the result. □

Step 4: A Kolmogorov type inequality for Y^T

To prove the convergence of Y^T towards 0, it remains to show its tightness. We have the following Kolmogorov type inequality on the moments of the increments of Y^T, which is a first step in order to get the tightness.

Lemma 9. *For any $\varepsilon > 0$, there exists $c_\varepsilon > 0$ such that for all $T \geq 1$, $0 \leq t, s \leq 1$,*

$$\mathbb{E}[(Y_t^T - Y_s^T)^4] \leq c_\varepsilon(|t-s|^{3/2-\varepsilon} + \frac{1}{T^2}|t-s|^{1-\varepsilon}). \tag{9}$$

Proof. Let $\mu^{\mathbb{E}[M_4^T]}$ denote the fourth moment measure of M^T, see the appendix of [JR13] for definition and properties. We have

$$\mathbb{E}[(Y_t^T - Y_s^T)^4] = \frac{1}{T^4}\int_{[0,T]^4} \Big(\prod_{i=1}^4 [f^T(t - \frac{t_i}{T}) - f^T(s - \frac{t_i}{T})]\Big)\mu^{\mathbb{E}[M_4^T]}(dt_1, dt_2, dt_3, dt_4).$$

Therefore, using Lemma A.17 of [JR13], we obtain

$$\begin{aligned}
\mathbb{E}[(Y_t^T - Y_s^T)^4] \leq{}& \frac{c}{T^3}\int_0^T |f^T(t-\frac{u}{T}) - f^T(s-\frac{u}{T})|^4 du \\
&+ \frac{c}{T^3}\int_0^T |f^T(t-\frac{u}{T}) - f^T(s-\frac{u}{T})|^3 du \int_0^T |f^T(t-\frac{u}{T}) - f^T(s-\frac{u}{T})| du \\
&+ \frac{c}{T^2}\int_0^T |f^T(t-\frac{u}{T}) - f^T(s-\frac{u}{T})|^2 du \int_0^T |f^T(t-\frac{u}{T}) - f^T(s-\frac{u}{T})|^2 du \\
&+ \frac{c}{T^3}\Big(\int_0^T |f^T(t-\frac{u}{T}) - f^T(s-\frac{u}{T})| du\Big)^2 \int_0^T |f^T(t-\frac{u}{T}) - f^T(s-\frac{u}{T})|^2 du.
\end{aligned}$$

Then, using Cauchy Schwarz inequality together with Corollary 2 and Lemma 7, we get

$$\int_0^T |f^T(t-\frac{u}{T}) - f^T(s-\frac{u}{T})| du \leq c_\varepsilon T\sqrt{|t-s|^{1-\varepsilon}}$$

and for $p = 2, 3, 4$,

$$\int_0^T |f^T(t-\frac{u}{T}) - f^T(s-\frac{u}{T})|^p du \leq c_\varepsilon T|t-s|^{1-\varepsilon},$$

which enables to conclude the proof. □

Step 5: Tightness

Let us define $\tilde{Y}^T$ the linear interpolation of Y^T with mesh $1/T^4$:

$$\tilde{Y}_t^T = Y^T_{\frac{\lfloor tT^4 \rfloor}{T^4}} + (tT^4 - \lfloor tT^4 \rfloor)(Y^T_{\frac{\lfloor tT^4 \rfloor + 1}{T^4}} - Y^T_{\frac{\lfloor tT^4 \rfloor}{T^4}}).$$

We use this interpolation since for $t - s = 1/T^4$, both terms on the right hand side of (9) have the same order of magnitude and for $t - s > 1/T^4$ the second term becomes negligible. We have the following lemma.

Lemma 10. *The sequence* $(\tilde{Y}^T)$ *is tight.*

Proof. We want to apply the classical Kolmogorov tightness criterion, see [Bil09], that states that if there exist $\gamma > 1$ and $c > 0$ such that for any $0 \le s \le t \le 1$,

$$\mathbb{E}|\tilde{Y}_t^T - \tilde{Y}_s^T|^4 \le c|t-s|^\gamma,$$

then $\tilde{Y}^T$ is tight. Remark that such inequality can of course not hold for Y^T since it is not continuous. Let $n_t^T = \lfloor tT^4 \rfloor$ and $n_s^T = \lfloor sT^4 \rfloor$. Let $0 < \varepsilon, \varepsilon' \le 1/4$ and $T \ge 1$. There are three cases:

- If $n_t^T = n_s^T$, using Lemma 9, we obtain that

$$\mathbb{E}[(\tilde{Y}_t^T - \tilde{Y}_s^T)^4]$$

is smaller than

$$|t-s|^4 T^{16}\mathbb{E}[(Y_{\frac{n_t^T+1}{T^4}} - Y_{\frac{n_t^T}{T^4}})^4] \le c_\varepsilon \frac{1}{T^{4(3/2-\varepsilon)}} T^{16}|t-s|^4 \le c_\varepsilon \frac{1}{T^{4(3/2-\varepsilon)}} T^{16}|t-s|^{1+\varepsilon'} \frac{1}{T^{4(3-\varepsilon')}}.$$

Since $0 < \varepsilon, \varepsilon' \le 1/4$, this leads to

$$\mathbb{E}[(\tilde{Y}_t^T - \tilde{Y}_s^T)^4] \le c_\varepsilon |t-s|^{1+\varepsilon'}.$$

- If $n_t^T = n_s^T + 1$,

$$\mathbb{E}[(\tilde{Y}_t^T - \tilde{Y}_s^T)^4] \le c\mathbb{E}[(\tilde{Y}_t^T - \tilde{Y}^T_{\frac{n_t^T}{T^4}})^4] + c\mathbb{E}[(\tilde{Y}^T_{\frac{n_t^T}{T^4}} - \tilde{Y}_s^T)^4] \le c_\varepsilon |t-s|^{1+\varepsilon'}.$$

- If $n_t^T \ge n_s^T + 2$, using again Lemma 9, we get

$$\begin{aligned}\mathbb{E}[(\tilde{Y}_t^T - \tilde{Y}_s^T)^4] &\le c\mathbb{E}[(\tilde{Y}_t^T - \tilde{Y}^T_{\frac{n_t^T}{T^4}})^4] + c\mathbb{E}[(\tilde{Y}^T_{\frac{n_s^T+1}{T^4}} - \tilde{Y}_s^T)^4] + c\mathbb{E}[(\tilde{Y}^T_{\frac{n_t^T}{T^4}} - \tilde{Y}^T_{\frac{n_s^T+1}{T^4}})^4] \\ &\le c_\varepsilon (\frac{1}{T^4})^{1+\varepsilon'} + c_\varepsilon |\frac{n_t^T}{T^4} - \frac{n_s^T+1}{T^4}|^{\frac{3}{2}-\varepsilon} \le c_\varepsilon |t-s|^{\min(\frac{3}{2}-\varepsilon, 1+\varepsilon')}.\end{aligned}$$

Hence the Kolmogorov criterion holds, which implies the tightness of $\tilde{Y}^T$. □

We now show that the difference between Y^T and $\tilde{Y}^T$ tends uniformly to zero.

Lemma 11. *We have the following convergence in probability:*

$$\sup_{|t-s| \le \frac{1}{T^4}} |Y_t^T - Y_s^T| \to 0.$$

Proof. Recall that for $0 \le s \le t \le 1$,

$$|Y_t^T - Y_s^T| = |\int_0^s f^T(t-u) - f^T(s-u) d\overline{M}_u^T + \int_s^t f^T(t-u) d\overline{M}_u^T|.$$

Thus, we have that $|Y_t^T - Y_s^T|$ is smaller than

$$\int_0^{sT} |f^T(t-u/T) - f^T(s-u/T)|(dN_u^T + \lambda_u du)\frac{1}{T} + \int_{sT}^{tT} |f^T(t-u/T)|(dN_u^T + \lambda_u du)\frac{1}{T}.$$

Using Corollaries 2 and 4, we obtain

$$|Y_t^T - Y_s^T| \leq c|t-s|(N_T^T + \int_0^T \lambda_u^T du) + c(N_{tT}^T - N_{sT}^T + \int_{sT}^{tT} \lambda_u^T du)\frac{1}{T}.$$

Consider now

$$\sup_{|t-s|\leq 1/T^4} |Y_t^T - Y_s^T|.$$

This is smaller than

$$c\frac{1}{T^4}(N_T^T + \int_0^T \lambda_u^T du) + 2c \max_{i=0,\dots,\lfloor T^4 \rfloor} \frac{1}{T}(N_{\frac{i+1}{T^4}T}^T - N_{\frac{i}{T^4}T}^T + \int_{\frac{i}{T^4}T}^{\frac{i+1}{T^4}T} \lambda_u^T du). \tag{10}$$

From Lemma A.5 of [JR13], we have

$$\mathbb{E}\big[N_T^T + \int_0^T \lambda_u^T du\big] \leq cT^2.$$

Thus, the first term on the right hand side of (10) tends to zero. For the second term, we use Lemma A.15 of [JR13] (with $t = \frac{i+1}{T^4}T$ and $s = \frac{i}{T^4}T$) which gives that

$$\mathbb{E}\big[\big(\frac{1}{T}(N_{\frac{i+1}{T^4}T}^T - N_{\frac{i}{T^4}T}^T + \int_{\frac{i}{T^4}T}^{\frac{i+1}{T^4}T} \lambda_u^T du)\big)^3\big] \leq \frac{c}{T^5}.$$

So, for any $\varepsilon > 0$, using Markov inequality, we get

$$\mathbb{P}\big[\frac{1}{T}(N_{\frac{i+1}{T^4}T}^T - N_{\frac{i}{T^4}T}^T + \int_{\frac{i}{T^4}T}^{\frac{i+1}{T^4}T} \lambda_u^T du) \geq \varepsilon\big] \leq \frac{c}{T^5\varepsilon^3}.$$

From this inequality, since the maximum is taken over a number of terms of order T^4, we easily deduce that the second term on the right hand side of (10) tends to zero in probability. □

We end this step by the proposition stating the convergence of Y^T.

Proposition 4. *The process Y^T converges ucp to 0 on $[0,1]$.*

Proof. We have

$$\sup_{t\in[0,1]} |Y_t^T| \leq \sup_{t\in[0,1]} |\tilde{Y}_t^T| + \sup_{t\in[0,1]} |\tilde{Y}_t^T - Y_t^T|.$$

From Lemma 8 and Lemma 10 we get that $\tilde{Y}^T$ tends to zero, in law for the Skorohod topology. This implies the ucp convergence. Applying Lemma 11 we get the result.

□

Step 6: **Limit of a sequence of SDEs**

In this last step, we show the convergence of the process $(C_t^T)_{t\in[0,1]}$ towards a CIR process. To do so, we use the fact that C^T can almost be written under the form of a stochastic differential equation. Indeed, recall that

$$C_t^T = U_t^T + \frac{1}{d_0}\int_0^t (\mu - C_s^T)ds + \frac{\sqrt{\lambda}}{m}\int_0^t \sqrt{C_s^T}dB_s^T,$$

with

$$B_t^T = \frac{1}{\sqrt{T}}\sqrt{u_T}\int_0^{tT}\frac{dM_s^T}{\sqrt{\lambda_s^T}}.$$

Then we aim at applying Theorem 5.4 in [KP91] to C^T. This result essentially says that for a sequence of SDEs where the functions and processes defining the equations satisfy some convergence properties, the laws of the solutions of the SDEs converge to the law of the solution of the limiting SDE. We now check these convergence properties.

The sequence of processes (B^T) is a sequence of martingales with jumps uniformly bounded by $c/\sqrt{\mu}$. Furthermore, for $t \in [0,1]$, the quadratic variation of (B^T) at point t is equal to

$$\frac{u_T}{T}\int_0^{tT}\frac{dN_s^T}{\lambda_s^T} = u_T\big(t + \int_0^{tT}\frac{dM_s^T}{T\lambda_s^T}\big).$$

Now, remark that

$$\mathbb{E}\big[(\int_0^{tT}\frac{dM_s^T}{T\lambda_s^T})^2\big] \le \mathbb{E}\big[\int_0^T \frac{1}{T^2\lambda_s^T}ds\big] \le c/(T\mu).$$

Therefore, we get that for any $t \in [0,1]$, the quadratic variation of (B^T) at point t converges in probability to t. Thus, we can apply Theorem VIII.3.11 in [JS87] to deduce that $(B_t^T)_{t\in[0,1]}$ converges in law for the Skorohod topology towards a Brownian motion.

Since U^T converges to a deterministic limit, we get the convergence in law, for the product topology, of the couple $(U_t^T, B_t^T)_{\in[0,1]}$ to $(0, B_t)_{\in[0,1]}$, with B a Brownian motion. The components of $(0, B_t)$ being continuous, the last convergence also takes place for the Skorohod topology on the product space.

Finally, recall that the (CIR) stochastic differential equation

$$X_t = \int_0^t (\mu - X_s)\frac{1}{d_0}ds + \frac{\sqrt{\lambda}}{m}\int_0^t \sqrt{X_s}dB_s$$

admits a unique strong solution on $[0,1]$. This, together with the preceding elements enables us to readily apply Theorem 5.4 in [KP91] to the sequence C^T, which gives the result.

4.3.3 Proof of the second part of Theorem 2

We now give the proof of the second part of Theorem 2 which deals with the sequence of Hawkes processes N^T. Let

$$V_t^T = \frac{(1-a_T)}{T} N_{tT}^T.$$

We write

$$V_t^T = \int_0^t C_s^T ds + \hat{M}_t^T,$$

where

$$\hat{M}_t^T = \frac{(1-a_T)}{T}(N_{tT}^T - \int_0^{tT} \lambda_s^T ds)$$

is a martingale. Using Doob's inequality, we obtain

$$\mathbb{E}[(\sup_{t\in[0,1]} \hat{M}_t^T)^2] \leq 4\mathbb{E}[(\hat{M}_1^T)^2] \leq 4(\frac{(1-a_T)}{T})^2 \mathbb{E}[N_T^T] \leq \frac{4\mu(1-a_T)}{T} \to 0.$$

Moreover, (C^T, t) converges in law over $[0,1]$ to (C,t) for the Skorokod topology. This last remark and Theorem 2.6 in [JMP89] on the limit of sequences of stochastic integrals give the result.

4.4 Proof of Theorem 3

We first introduce some notations. In this proof, we write

$$\phi^T = \phi_1^T - \phi_2^T \text{ and } \psi^T = \sum_{k=1}^{+\infty} (\phi^T)^{*k}.$$

Moreover, we set

$$C_t^T = \frac{\lambda_{tT}^{T+} + \lambda_{tT}^{T-}}{T}$$

and define

$$(B^1)_t^T = \int_0^{tT} \frac{dM_s^{T+} + dM_s^{T-}}{\sqrt{T(\lambda_s^{T+} + \lambda_s^{T-})}}, \quad (B^2)_t^T = \int_0^{tT} \frac{dM_s^{T+} - dM_s^{T-}}{\sqrt{T(\lambda_s^{T+} + \lambda_s^{T-})}},$$

with

$$M_s^{T+} = N_s^{T+} - \int_0^s \lambda_s^{T+} ds, \quad M_s^{T-} = N_s^{T-} - \int_0^s \lambda_s^{T-} ds.$$

Finally, we set

$$\overline{M}_t^{T+} = \frac{M_{Tt}^{T+}}{T}, \quad \overline{M}_t^{T-} = \frac{M_{Tt}^{T-}}{T}.$$

We split the proof of Theorem 3 into several steps.

Step 1: Convenient rewriting

In this first step, we rewrite the price, intensity and martingale processes under more convenient forms. We have

$$\lambda_t^{T+} - \lambda_t^{T-} = \int_0^t \phi^T(t-s)(\lambda_s^{T+} - \lambda_s^{T-})ds + \int_0^t \phi^T(t-s)(dM_s^{T+} - dM_s^{T-}).$$

Therefore, in the same way as for the proof of Proposition 1, we get

$$\lambda_t^{T+} - \lambda_t^{T-} = \int_0^t \psi^T(t-s)(dM_s^{T+} - dM_s^{T-}).$$

From this last expression, we easily obtain

$$N_t^{T+} - N_t^{T-} = \int_0^t \big(1 + \Psi^T(t-u)\big)(dM_u^{T+} - dM_u^{T-}), \tag{11}$$

with

$$\Psi^T(x) = \int_0^x \psi^T(s)ds.$$

Finally, note that

$$\overline{M}_t^{T+} - \overline{M}_t^{T-} = \frac{1}{T}(M_{Tt}^{T+} - M_{Tt}^{T-}) = \int_0^t \sqrt{C_s^T}\, d(B^2)_s^T. \tag{12}$$

Step 2: Preliminary result

For $s \in [0,1]$, we define

$$X_s^T = \frac{\lambda_{sT}^{T+} - \lambda_{sT}^{T-}}{T}.$$

We have the following important result.

Lemma 12. *The process X^T converges ucp to 0 on $[0,1]$.*

Proof. We write

$$X_t^T = \int_0^t f_1^T(t-s)d(\overline{M}_s^{T+} - \overline{M}_s^{T-}),$$

with $f_1^T(x) = \psi^T(Tx)$. Remark that Corollaries 1, 2, 3 and 4 are valid if in their statement, f^T is replaced by f_1^T. In the proof of Theorem 2, we have shown the convergence to zero of the process

$$Y_t^T = \int_0^T f^T(t-s)d\overline{M}_s^T.$$

Therefore, applying the same strategy but replacing f^T by f_1^T and $\overline{M}^T$ by $\overline{M}^{T+} - \overline{M}^{T-}$, it is clear that we get the result. □

Step 3: Convergence of (B^1, B^2)

In this step, we prove the convergence of (B^1, B^2) towards a two-dimensional Brownian motion. To do so, we study the quadratic (co-)variations of the processes. Let $i \in \{1,2\}$, $j \in \{1,2\}$. We denote by $[(B^i)^T, (B^j)^T]_t$ the quadratic co-variation of B^i and B^j at time t.

Lemma 13. *We have the following convergence in probability:*

$$[(B^i)^T, (B^j)^T]_t \to t1_{i=j}.$$

Proof. There are three cases:

• If $i = j = 1$, using that N^{T+} and N^{T-} have no common jumps, we get,

$$[(B^1)^T, (B^1)^T]_t = \int_0^{tT} \frac{dN_s^{T+} + dN_s^{T-}}{T(\lambda_s^{T+} + \lambda_s^{T-})} = t + \int_0^{tT} \frac{dM_s^{T+} + dM_s^{T-}}{T(\lambda_s^{T+} + \lambda_s^{T-})}.$$

Furthermore,

$$\mathbb{E}\Big[\Big(\int_0^{tT} \frac{dM_s^{T+} + dM_s^{T-}}{T(\lambda_s^{T+} + \lambda_s^{T-})}\Big)^2\Big] \le \frac{ct}{T\mu} \to 0.$$

Therefore we have the result for $i = j = 1$.

• If $i = j = 2$, the proof goes similarly.

• If $i = 1$ and $j = 2$,

$$[(B^1)^T, (B^2)^T]_t = \int_0^{tT} \frac{dN_s^{T+} - dN_s^{T-}}{T(\lambda_s^{T+} + \lambda_s^{T-})} = \int_0^{tT} \frac{dM_s^{T+} - dM_s^{T-} + \lambda_s^{T+}ds - \lambda_s^{T-}ds}{T(\lambda_s^{T+} + \lambda_s^{T-})}.$$

As for the case $i = j = 1$, we easily get

$$\int_0^{tT} \frac{dM_s^{T+} - dM_s^{T-}}{T(\lambda_s^{T+} + \lambda_s^{T-})} \to 0.$$

It remains to show the convergence to zero of Z_t^T defined by

$$Z_t^T = \int_0^t \frac{X_s^T}{C_s^T} ds.$$

For any $\varepsilon > 0$, we have

$$|Z_t^T| \le \int_0^t (1 \wedge |\frac{X_s^T}{\varepsilon}|) ds + \int_0^t 1_{C_s^T < \varepsilon} ds.$$

From Lemma 12, we have the convergence of the process X^T to zero. Furthermore, in Lemma 15 we will show that C^T converge in law over $[0,1]$ towards a CIR process denoted by C. Therefore, since the limiting processes are continuous, we have the joint convergence of

(X_T, C^T) to $(0, C)$. We now use Skorohod representation theorem (without changing notations). Almost surely, for T large enough, we have

$$\sup_{s\in[0,1]} |X_s^T| \le \varepsilon^2, \quad \sup_{s\in[0,1]} |C_s^T - C_s| \le \varepsilon.$$

This implies

$$\int_0^t (1 \wedge |\frac{X_s^T}{\varepsilon}|) ds + \int_0^t 1_{C_s^T<\varepsilon} ds \le \varepsilon + \int_0^1 1_{C_s<2\varepsilon} ds.$$

Recall that the set of zeros of a CIR process on a finite time interval has zero Lebesgue measure. Thus, using the dominated convergence theorem, we easily see that choosing ε conveniently, the second term in the preceding inequality can be made arbitrarily small, which ends the proof. □

Thus for any T, $(B^1)^T$ and $(B^2)^T$ are two martingales with uniformly bounded jumps and their quadratic (co-)variations satisfy Lemma 13. Consequently, Theorem VIII.3.11 of [JS87] gives us the following lemma.

Lemma 14. *We have*

$$((B^1)^T, (B^2)^T) \to (B^1, B^2),$$

in law, for the Skorohod topology, where (B^1, B^2) *is a two-dimensional Brownian motion.*

Step 4: **Convergence of** $(C^T, (B^2)^T)$

The aim of this step is to prove that the couple $(C^T, (B^2)^T)$ converges in law towards $(C, (B^2))$, with C a CIR process and B^2 a Brownian motion, independent of C. More precisely, we have the following lemma.

Lemma 15. *The couple of process* $(C^T, (B^2)^T)$ *converges in law, for the Skorohod topology, over* $[0, 1]$, *towards* (C, B^2), *where* B^2 *is a Brownian motion independent of* C *and* C *is a CIR process satisfying*

$$C_t = \int_0^t (\frac{2\mu}{\lambda} - C_s) \frac{\lambda}{m} ds + \frac{1}{m} \int_0^t \sqrt{C_s} dW_s,$$

with W *another Brownian motion, independent of* B^2.

Proof. Let us consider the process $N^T = N^{T+} + N^{T-}$. It is a point process with intensity

$$\lambda_t^T = \lambda_t^{T+} + \lambda_t^{T-} = 2\mu + a_T \int_0^t (\phi_1 + \phi_2)(t-s) dN_s^T.$$

Therefore, we are in the framework of Theorem 2: N^T is a Hawkes process whose kernel has a norm that tends to 1 at the right speed and its renormalized intensity C^T converges towards a CIR. Remark that the renormalizing factor here is $1/T$ and not $(1 - a_T)$, which is not an issue since (3) holds. Thus we get the convergence of C^T towards a CIR. To obtain the joint convergence, we just need to write the same proof as for Theorem 2 (up to obvious changes), but using this time Theorem 5.4 in [KP91] together with Lemma 14. □

Step 5: Technical results

This fifth step consists in proving two technical results. The first one is the following.

Lemma 16. *The process*

$$R_t^T = \int_0^t \int_{T(t-u)}^{+\infty} \psi^T(s) ds d(\overline{M}_u^{T+} - \overline{M}_u^{T-})$$

converges ucp to 0 *on* $[0,1]$.

Proof. We write

$$R_t^T = \int_0^t f_2^T(t-u) d(\overline{M}_u^{T+} - \overline{M}_u^{T-}),$$

with

$$f_2^T(x) = \int_{Tx}^{+\infty} \psi^T(s) ds.$$

The result follows in the same way as in the proof of Lemma 12. □

We now give the last lemma of this step.

Lemma 17. *We have*

$$\int_0^\infty \int_x^\infty \phi_i(s) ds dx < \infty.$$

Proof. Using integration by parts together with Assumption 4, we get

$$\int_0^\infty \int_x^\infty \phi_i(s) ds dx = \int_0^\infty x\phi_i(x) dx + \lim_{x\to\infty} x \int_x^\infty \phi_i(s) ds \leq 2m.$$

□

Step 6: End of the proof

We finally show Theorem 3 in this step. Using (11) we write

$$\begin{aligned} P_t^T = (1 + \frac{\|\phi\|_1}{1-\|\phi\|_1})(\overline{M}_t^{T+} - \overline{M}_t^{T-}) \\ - \int_0^t \int_{T(t-u)}^{+\infty} \psi^T(s) ds d(\overline{M}_u^{T+} - \overline{M}_u^{T-}) - (\frac{\|\phi\|_1}{1-\|\phi\|_1} - \frac{a_T\|\phi\|_1}{1-a^T\|\phi\|_1})(\overline{M}_t^{T+} - \overline{M}_t^{T-}). \end{aligned}$$

Using Theorem 2.6 in [JMP89] together with Lemma 15 and Equation (12), we get the convergence of the process $\overline{M}^{T+} - \overline{M}^{T-}$, over $[0,1]$, for the Skorohod topology, towards

$$\int_0^t \sqrt{C_s} dB_s^2.$$

Moreover, in Lemma 16, we have shown that the second term in the decomposition of P_t^T tends to zero. Finally, the third term also vanishes since $\|\phi\|_1 < 1$. This concludes the proof.

CHAPTER III

Rough fractional diffusions as scaling limits of heavy-tailed nearly unstable Hawkes processes

Abstract

We consider a sequence of Hawkes processes whose regression kernels have L^1 norms asymptotically close to one and power law decay at infinity of the form $x^{-(1+\alpha)}$, with $\alpha \in (0,1)$. Such processes are in particular used for order flow modeling in finance. We prove that after suitable rescaling, the long term limiting behavior of the sequence is that of a kind of integrated fractional Cox-Ingersoll-Ross process, with associated Hurst parameter $H = \alpha - 1/2$. This results is in contrast with the case of thin tailed regression kernels, where a classical Brownian CIR process is obtained at the limit. In particular, it shows that persistence properties in the point process can lead to an irregular behavior of the limiting process. This theoretical result enables us to give an agent-based foundation to some recent findings about the rough nature of the volatility in financial markets.

Keywords: Hawkes processes, limit theorems, stability condition, heavy tail, fractional stochastic equation, fractional Cox-Ingersoll-Ross process, volatility, long memory.

1 Introduction

A Hawkes process $(N_t)_{t\geq 0}$ is a self exciting point process, whose intensity at time t, denoted by λ_t, is of the form

$$\lambda_t = \mu + \sum_{0<J_i<t} \phi(t - J_i) = \mu + \int_{(0,t)} \phi(t-s) dN_s,$$

where μ is a positive real number, ϕ a non negative measurable function and the J_i are the points of the process before time t (see Section 2 for a more formal definition). These processes have been introduced in the early seventies by Hawkes, see [Haw71a, Haw71b, HO74], in the purpose of modeling earthquakes and their aftershocks, see [Ada76] for such application. In the last years, the probabilistic and statistical analysis of Hawkes processes

has known several interesting developments, driven by the recent use of Hawkes processes in various applied fields such as neuroscience [CSK88, PSCR11, PSCR12], sociology [BBH12, LZ13, RBRTM13, ZZS13], criminology [M$^+$13, MSB$^+$11], genome analysis [RBS$^+$10] and mostly finance [ASCDL10, BH04, Bow07, CDDM05, ELL11, EGG10].

Among the probabilistic questions raised by Hawkes processes, particular attention has been devoted to the study of their long term scaling limits. More precisely, one wishes to understand the behavior as T tends to infinity of the process

$$\alpha_T(N_{tT}),\ t \in [0,1],$$

where α_T is a suitable normalizing factor. In [BDHM12], it is shown that under the condition

$$\|\phi\|_1 = \int_0^{+\infty} \phi(s)ds < 1,$$

the asymptotic behavior of a Hawkes process is quite similar to that of a Poisson process. Indeed, as T tends to infinity,

$$\sup_{t\in[0,1]} |\frac{N_{tT}}{T} - \mathbb{E}\big[\frac{N_{tT}}{T}\big]| \to 0,$$

in probability and

$$\Big(\sqrt{T}\big(\frac{N_{tT}}{T} - \mathbb{E}\big[\frac{N_{tT}}{T}\big]\big)\Big)_{t\in[0,1]} \to \sigma(W_t)_{t\in[0,1]},$$

in law for the Skorohod topolgy, with σ an explicit constant and (W_t) a Brownian motion. This result has been extended in [Zhu13] to the case of non linear Hawkes processes.

The condition $\|\phi\|_1 < 1$ is essential in order to obtain the preceding result. It is actually very similar to the assumption $|\rho| < 1$ one makes on the autoregressive coefficient ρ when working with a discrete time stationary AR(1) process. In particular, when starting the Hawkes process at $t = -\infty$, the assumption $\|\phi\|_1 < 1$ is required in order to get a stationary intensity with finite first moment. Also, as for AR(1) processes, under this condition, Hawkes processes only exhibit weak dependence properties. Consequently, their asymptotic behavior is in that case no surprise, close to that of a Poisson process. Hence this condition is called stability condition.

In Chapter II, the authors investigate the scaling limit of Hawkes processes when the stability condition is almost violated. This means they consider a sequence of Hawkes processes satisfying the stability condition, but for which the kernel $\phi = \phi^T$ also depends on the observation scale T, such that $\|\phi^T\|_1$ tends to 1 as T goes to infinity. Such sequence is called sequence of nearly unstable Hawkes processes.

Beyond its obvious mathematical interest, considering the framework of nearly unstable Hawkes processes is motivated by empirical studies on financial data. Indeed, it has become quite standard to model the clustered nature of order flows on financial markets by means of Hawkes

processes. However, one systematically estimates L^1 norms for the regression kernels which are smaller but very close to 1, see [FS12, FS13, HBB13, LC14]. Interestingly, this empirical stylized fact has a very natural financial interpretation, namely the high degree of endogeneity of modern markets due to high frequency trading. This signifies that a large proportion of orders are just triggered by other orders, see [FS13, HBB13, JR13] for more details. Such situation is naturally modeled by nearly unstable Hawkes processes and it is proved in Chapter II that in this framework, the limiting law is that of an integrated Cox-Ingersoll-Ross process (CIR process for short). Hence, compared to the case where the stability condition is in force, the limiting behavior at first order is no longer deterministic, see also [Zhu13] for the case where $\|\phi\|_1$ is exactly equal to one and other interesting developments. Note that this CIR scaling limit seems very consistent with financial practice. Indeed, it is widely acknowledged that there exists a linear relationship between the cumulated order flow and the integrated squared volatility, see for example [WBK$^+$08], and CIR processes are a very classical model for the squared volatility.

Nevertheless, the CIR limit in law of nearly unstable Hawkes processes discussed above is obtained under the crucial assumption

$$\int_0^{+\infty} s\phi(s)ds < +\infty.$$

It is therefore quite natural to try to extend the results of Chapter II to the case of *nearly unstable heavy tailed Hawkes processes*, for which this condition is no longer satisfied. Hence we consider in this paper the situation where

$$\phi(x) \underset{x\to+\infty}{\sim} \frac{K}{x^{1+\alpha}},$$

where $\alpha \in (0,1)$ and K is a positive constant. This setting is actually much more in agreement with financial data, where one not only finds that the function ϕ has a L^1 norm close to one, but also that it has a power law tail, see [HBB13] and Chapter V. This heavy tail is also quite easy to interpret in practice: it is related to the persistence of the signed order flow (the series of $+1$, -1 where $+1$ represents a buy order and -1 a sell order). Indeed, the empirical persistent property of this process is well established and is due to the so-called order splitting phenomenon: most orders are actually parts of large orders (called metaorders), which are split in smaller orders so that prohibitive execution costs can be avoided.

Our main result is that after proper scaling, the limit laws of nearly unstable heavy tailed Hawkes processes are that of processes which can be interpreted as integrated fractional diffusions. Very loosely speaking, these processes can be viewed as the integral of fractional versions of the CIR process, where fractional Brownian motions would replace the ordinary Brownian motion. This result is quite remarkable from a probabilistic point of view. Indeed, assuming fat tail leads to a limit which is not an integrated semi-martingale. This is in strong contrast to all other scaling limits obtained for Hawkes processes. From a technical point of view, this case is of course more subtle than that investigated in Chapter II where

semi-martingale theorems are used in a quite direct manner. Moreover, Gaussian techniques are not easy to apply in our context since the limit is not a simple Gaussian functional, although it somehow involves a fractional Brownian motion.

The perhaps most surprising phenomenon obtained in our result is the value of the Hurst parameter H of the (sort of) fractional Brownian motion appearing in the limit. Indeed, fat tail meaning persistence, one would expect getting persistence in the limit and so $H > 1/2$. This is actually the contrary: an aggregation phenomenon occurs in the heavy tail case, leading to a very irregular process in the limit, its derivative behaving as a fractional Brownian motion with Hurst parameter $H < 1/2$. Coming back to financial applications, this means that in practice, the volatility process should be very irregular, which is perfectly in line with the recent empirical measures of the volatility smoothness obtained in Chapter I. Therefore, our theoretical result shows quite clearly that the rough nature of the volatility can be explained by the high degree of endogeneity of financial markets together with the order splitting phenomenon. This is to our knowledge the first agent-based explanation for the very rough behavior of the volatility.

The paper is organized as follows. We first give our assumptions together with some intuitions on the limiting behavior of our processes in Section 2. Section 3 contains our main theorems whose proofs can be found in Section 4. Finally, some technical results are relegated to an appendix.

2 Assumptions and intuitions for the results

We describe in this section our asymptotic framework together with intuitions about our main results which are given in Section 3.

We consider a sequence of point processes $(N_t^T)_{t\geq 0}$ indexed by T[1]. For a given T, (N_t^T) satisfies $N_0^T = 0$ and the process is observed on the time interval $[0, T]$. Our asymptotic setting is that the observation scale T goes to infinity. The intensity process (λ_t^T) is defined for $t \geq 0$ by

$$\lambda_t^T = \mu^T + \int_0^t \phi^T(t-s) dN_s^T,$$

where μ^T is a sequence of positive real numbers and ϕ^T a non negative measurable function on $\mathbb{R}^+$ which satisfies $\|\phi^T\|_1 < +\infty$. For a given T, the process (N_t^T) is defined on a probability space $(\Omega^T, \mathscr{F}^T, \mathbb{P}^T)$ equipped with the filtration $(\mathscr{F}_t^T)_{t\in[0,T]}$, where $\mathscr{F}_t^T$ is the σ-algebra generated by $(N_s^T)_{s\leq t}$. Moreover we assume that for any $0 \leq a < b \leq T$ and $A \in \mathscr{F}_a^T$

$$\mathbb{E}[(N_b^T - N_a^T)1_A] = \mathbb{E}[\int_a^b \lambda_s^T 1_A ds],$$

[1]Of course by T we implicitly means T_n with $n \in \mathbb{N}$ tending to infinity.

which sets λ^T as the intensity of N^T. In particular, if we denote by $(J_n^T)_{n\geq 1}$ the jump times of (N_t^T), the process

$$N^T_{t\wedge J_n^T} - \int_0^{t\wedge J_n^T} \lambda_s^T \, ds$$

is a martingale and the law of N^T is characterized by λ^T. From [Jac75], such a construction can be done and the process N^T is called a Hawkes process.

Let us now give more specific assumptions on the function ϕ^T.

Assumption 1. *For $t \in \mathbb{R}^+$,*

$$\phi^T(t) = a_T \phi(t),$$

where $(a_T)_{T\geq 0}$ is a sequence of positive numbers converging to 1 such that for all T, $a_T < 1$ and ϕ is a non negative measurable function such that $\|\phi\|_1 = 1$. Furthermore,

$$\lim_{x\to+\infty} \alpha x^\alpha \big(1 - F(x)\big) = K,$$

for some $\alpha \in (0,1)$ and some positive constant K, with

$$F(x) = \int_0^x \phi(s) ds.$$

Recall that in Chapter II, it is assumed that

$$\int_0^{+\infty} t\phi(t) dt < +\infty \tag{1}$$

and this condition leads to a CIR-type limit. Considering Assumption 1 instead of (1) will induce a completely different scaling behavior for the sequence of nearly unstable Hawkes processes. Nevertheless, in this framework, we still have almost surely no explosion[2]:

$$\lim_{n\to+\infty} J_n^T = +\infty.$$

Remark that we do not work in the stationary setting since our process starts at time $t = 0$ and not at $t = -\infty$.

Let M^T denote the martingale process associated to N^T, that is, for $t \geq 0$,

$$M_t^T = N_t^T - \int_0^t \lambda_s^T \, ds.$$

We also set ψ^T as the function defined on $\mathbb{R}^+$ by

$$\psi^T(t) = \sum_{k=1}^{\infty} (\phi^T)^{*k}(t), \tag{2}$$

[2] In fact, for a Hawkes process, the no explosion property can be obtained under weaker conditions, for example $\int_0^t \phi(s) ds < \infty$ for any $t > 0$, see [BDHM12].

where $(\phi^T)^{*1} = \phi^T$ and for $k \geq 2$, $(\phi^T)^{*k}$ denotes the convolution product of $(\phi^T)^{*(k-1)}$ with the function ϕ^T. Note that $\psi^T(t)$ is well defined since $\|\phi^T\|_1 < 1$. This function plays an important role in the study of Hawkes processes, see [BDM12]. In particular, it is proved in Chapter II that the intensity process, rescaled on $[0,1]$, can be rewritten

$$\lambda^T_{tT} = \mu^T + \int_0^{tT} \psi^T(Tt-s)\mu^T ds + \int_0^{tT} \psi^T(Tt-s) dM^T_s.$$

In term of scaling in space, a natural multiplicative factor is $(1-a_T)/\mu^T$. Indeed, in the stationary case, the expectation of λ^T_t is $\mu^T/(1-\|\phi^T\|_1)$. Thus, the order of magnitude of the intensity is $\mu^T(1-a_T)^{-1}$. This is why we define

$$C^T_t = \frac{(1-a_T)}{\mu^T}\lambda^T_{tT}.$$

Then we easily get

$$C^T_t = (1-a_T) + \int_0^t T(1-a_T)\psi^T(Ts)ds + \sqrt{\frac{T(1-a_T)}{\mu^T}}\int_0^t \psi^T(T(t-s))\sqrt{C^T_s}\, dB^T_s, \tag{3}$$

with

$$B^T_t = \frac{1}{\sqrt{T}}\int_0^{tT} \frac{dM^T_s}{\sqrt{\lambda^T_s}}.$$

From (3), we see that the asymptotic behavior of the intensity is closely related to that of $x \mapsto \psi^T(Tx)$. To analyze the limiting behavior of this function, let us remark that for $x \geq 0$,

$$\rho^T(x) = T\frac{\psi^T(Tx)}{\|\psi^T\|_1} \tag{4}$$

is the density of the random variable

$$J^T = \frac{1}{T}\sum_{i=1}^{I^T} X_i,$$

where the (X_i) are iid random variables with density ϕ and I^T is a geometric random variable with parameter $1-a_T$[3]. The Laplace transform of the random variable J^T, denoted by $\widehat{\rho}^T$, satisfies for $z \geq 0$

$$\begin{aligned}\widehat{\rho}^T(z) = \mathbb{E}[e^{-zJ^T}] &= \sum_{k=1}^{\infty}(1-a_T)(a_T)^{k-1}\mathbb{E}[e^{-\frac{z}{T}\sum_{i=1}^k X_i}] \\ &= \sum_{k=1}^{\infty}(1-a_T)(a_T)^{k-1}(\hat{\phi}(\frac{z}{T}))^k = \frac{\widehat{\phi}(\frac{z}{T})}{1-\frac{a_T}{1-a_T}(\widehat{\phi}(\frac{z}{T})-1)},\end{aligned}$$

[3] $\forall k > 0,\ \mathbb{P}[I^T = k] = (1-a_T)(a_T)^{k-1}$.

where $\widehat{\phi}$ denotes the Laplace of ϕ. We now need to compute an expansion for $\widehat{\phi}(z)$. Using integration by part, we get

$$\widehat{\phi}(z) = z\int_0^{+\infty} e^{-zt}F(t)dt = 1 - z\int_0^{+\infty} e^{-zt}\big(1-F(t)\big)dt.$$

Then using Assumption 1 together with Karamata Tauberian theorem (see for example Theorem 17.6 in [BGT89]), we get

$$\widehat{\phi}(z) = 1 - K\frac{\Gamma(1-\alpha)}{\alpha}z^{\alpha} + o(z^{\alpha}),$$

with Γ the gamma function. Set $\delta = K\frac{\Gamma(1-\alpha)}{\alpha}$ and $\nu_T = \delta^{-1}T^{\alpha}(1-a_T)$. As T goes to infinity, $\widehat{\rho}^T(z)$ is thus equivalent to

$$\frac{\nu_T}{\nu_T + z^{\alpha}}. \tag{5}$$

The function whose Laplace transform is equal to this last quantity is given by

$$\nu_T x^{\alpha-1}E_{\alpha,\alpha}(-\nu_T x^{\alpha}),$$

with $E_{\alpha,\beta}$ the (α,β) Mittag-Leffler function, that is

$$E_{\alpha,\beta}(z) = \sum_{n=0}^{\infty}\frac{z^n}{\Gamma(\alpha n+\beta)},$$

see [HMS11]. Putting this together with (3) and (4), we can expect (for $\alpha > 1/2$):

$$C_t^T \sim \nu_T\int_0^t s^{\alpha-1}E_{\alpha,\alpha}(-\nu_T s^{\alpha})ds + \gamma_T\nu_T\int_0^t (t-s)^{\alpha-1}E_{\alpha,\alpha}(-\nu_T(t-s)^{\alpha})\sqrt{C_s^T}dB_s^T,$$

with

$$\gamma_T = \frac{1}{\sqrt{\mu^T T(1-a_T)}}.$$

The process B^T can be shown to converge to a Brownian motion B. Thus, denoting by ν_∞ and γ_∞ the limits of ν_T and γ_T, passing (non rigorously) to the limit, we obtain (for $\alpha > 1/2$)

$$C_t^\infty \sim \nu_\infty\int_0^t s^{\alpha-1}E_{\alpha,\alpha}(-\nu_\infty s^{\alpha})ds + \gamma_\infty\nu_\infty\int_0^t (t-s)^{\alpha-1}E_{\alpha,\alpha}(-\nu_\infty(t-s)^{\alpha})\sqrt{C_s^\infty}dB_s. \tag{6}$$

From (6), we see that in order to get a non-deterministic asymptotic behavior for C_t^T, we need that both ν_∞ and γ_∞ are positive constants or ν_∞ is equal to zero and $\gamma_\infty\nu_\infty$ is positive. However, in the last situation, the expectation of $(C_t^\infty)^2$ would be of order $t^{2\alpha-1}$. Such cases where the variance is increasing with power law rate are incompatible with the approximate stationarity property we want to keep for our model and its limit. Therefore, only one regime seems natural and this leads us to the following assumption.

Assumption 2. *There are two positive constants λ and μ^* such that*

$$\lim_{T\to+\infty} T^\alpha(1-a_T) = \lambda\delta.$$

and

$$\lim_{T\to+\infty} T^{1-\alpha}\mu_T = \mu^*\delta^{-1}.$$

In particular, Assumption 2 implies that ν_T converges to λ and therefore the sequence of random variables (J^T) converges in law towards the random variable whose density on $\mathbb{R}^+$ is given by

$$\lambda x^{\alpha-1}E_{\alpha,\alpha}(-\lambda x^\alpha).$$

Beyond giving us the suitable asymptotic regimes for a_T and μ_T, the heuristic derivation leading to (6) provides an expression for the limiting law of the rescaled intensities of our sequence of nearly unstable heavy tailed Hawkes processes. In (6), this law appears under the form of a non-classical stochastic integral equation. Indeed, it is of Volterra-type and is therefore (a priori) neither a diffusion nor a semi-martingale. Furthermore, the main term of the Volterra kernel $x^{\alpha-1}$ exhibits a singularity at point 0, of the same kind as that of the fractional Brownian motion (B_t^H) when expressed under the form:

$$B_t^H = \frac{1}{\Gamma(H+1/2)}\Big[\int_0^t (t-s)^{H-1/2}dW_s + \int_{-\infty}^0 (t-s)^{H-1/2} - (-s)^{H-1/2}dW_s\Big], \tag{7}$$

with (W_t) a Brownian motion, see [MVN68].

The preceding computations suggest a possible approach to derive the limiting behavior of our sequence of Hawkes processes: studying the intensity of the processes. Indeed the intensities can be rewritten under the form of stochastic integral equations as (3). Consequently one can try to pass to the limit in the coefficients of the equation to obtain the limiting law, as we (non rigorously) did to get (6). This is exactly the approach used in Chapter II. However, in this more intricate case, it seems very hard to use. In particular the sequence (C_t^T) is typically not tight. Thus, instead of considering the intensities, we directly work on the Hawkes processes themselves, more in the spirit of [Zhu13].

3 Main results

We rigorously state in this section our results on the limiting behavior of nearly unstable heavy tailed Hawkes processes. We start with some technical results about the function appearing as the inverse Laplace transform of (5) in Section 2.

3.1 The function $f^{\alpha,\lambda}$

As shown by the derivations in the previous section, the function

$$f^{\alpha,\lambda}(x) = \lambda x^{\alpha-1}E_{\alpha,\alpha}(-\lambda x^\alpha)$$

plays a crucial role in our analysis. We give here elements about the regularity of this function which will be useful in the sequel. We denote by $I^\alpha f$ and $D^\alpha f$ the fractional integration and derivation operators, which are defined for a suitable measurable function f by

$$I^\alpha f(x) = \frac{1}{\Gamma(\alpha)} \int_0^x \frac{f(t)}{(x-t)^{1-\alpha}} dt$$

and

$$D^\alpha f(x) = \frac{1}{\Gamma(1-\alpha)} \frac{d}{dx} \int_0^x \frac{f(t)}{(x-t)^{\alpha}} dt.$$

The following lemma is a direct consequence of the definition of $f^{\alpha,\lambda}$ and of Section 11 of [MH08].

Proposition 1. *The function $f^{\alpha,\lambda}$ is C^∞ on $(0,1]$ and*

$$f^{\alpha,\lambda}(x) \underset{x\to 0^+}{\sim} \frac{\lambda}{\Gamma(\alpha)} x^{\alpha-1},$$

$$(f^{\alpha,\lambda})'(x) \underset{x\to 0^+}{\sim} \frac{\lambda(\alpha-1)}{\Gamma(\alpha)} x^{\alpha-2}.$$

Furthermore, $f^{\alpha,\lambda}(x)x^{1-\alpha}$ has Hölder regularity α on $(0,1]$.

For $\nu < \alpha$, $f^{\alpha,\lambda}$ is ν fractionally differentiable and

$$D^\nu f^\alpha(x) = \lambda x^{\alpha-1-\nu} E_{\alpha,\alpha-\nu}(-\lambda x^\alpha).$$

Therefore,

$$D^\nu f^\alpha(x) \underset{x\to 0^+}{\sim} \frac{\lambda}{\Gamma(\alpha-\nu)} \frac{1}{x^{1-\alpha+\nu}}$$

and

$$(D^\nu f^\alpha)'(x) \underset{x\to 0^+}{\sim} \frac{\lambda(\alpha-1-\nu)}{\Gamma(\alpha-\nu)} \frac{1}{x^{2-\alpha+\nu}}.$$

For $\nu' > 0$, f^α is ν' fractionally integrable and

$$I^{\nu'} f^\alpha(x) = \lambda \frac{1}{x^{1-\alpha-\nu'}} E_{\alpha,\alpha+\nu'}(-\lambda x^\alpha).$$

Therefore,

$$I^{\nu'} f^\alpha(x) \underset{x\to 0^+}{\sim} \frac{\lambda}{\Gamma(\alpha+\nu')} \frac{1}{x^{1-\alpha-\nu'}}$$

and for $\alpha + \nu' \neq 1$,

$$(I^{\nu'} f^\alpha)'(x) \underset{x\to 0^+}{\sim} \frac{\lambda(\alpha-1+\nu')}{\Gamma(\alpha+\nu')} \frac{1}{x^{2-\alpha-\nu'}}.$$

Proposition 1 will be a key tool in the proofs of the main results.

3.2 The limiting behavior of nearly unstable heavy tailed Hawkes processes

Let us first give some notations. We consider for $t \in [0,1]$ the renormalized Hawkes process

$$X_t^T = \frac{1-a_T}{T^\alpha \mu^* \delta^{-1}} N_{Tt}^T$$

and its associated integrated intensity

$$\Lambda_t^T = \frac{1-a_T}{T^\alpha \mu^* \delta^{-1}} \int_0^{tT} \lambda_s^T ds.$$

As explained in Section 2, the space renormalization is chosen so that the processes have an expectation of order one. We also introduce the martingale defined on $[0,1]$ by

$$Z_t^T = \sqrt{\frac{T^\alpha \mu^* \delta^{-1}}{1-a_T}} (X_t^T - \Lambda_t^T).$$

We are now ready to give our results about the limit laws of (Z^T, X^T) for the Skorohod topology.

Proposition 2. *Under Assumptions 1 and 2, the sequence (Z^T, X^T) is tight. Furthermore, if (Z,X) is a limit point of (Z^T, X^T), then Z is a continuous martingale and $[Z,Z] = X$.*

Now let (Z,X) be a couple of processes defined on some space $(\Omega, \mathscr{A}, \mathbb{P})$ with law being one of the possible limit point of the sequence of distributions associated to the sequence (Z^T, X^T). From Proposition 2, we are able to obtain the following theorem which is one of our main results.

Theorem 1. *There exists a Brownian motion B on $(\Omega, \mathscr{A}, \mathbb{P})$ (up to extension of the space) such that for $t \in [0,1]$, $Z_t = B_{X_t}$ and X is continuous and satisfies*

$$X_t = \int_0^t s f^{\alpha,\lambda}(t-s) ds + \frac{1}{\sqrt{\mu^* \lambda}} \int_0^t f^{\alpha,\lambda}(t-s) B_{X_s} ds. \tag{8}$$

Hence the limiting processes in Theorem 1 have a quite original form, which can actually be interpreted more easily by looking at their derivatives (when they exist).

3.3 The limiting volatility process

As explained in the Introduction, when it exists, the derivative of the limiting process X in Theorem 1 can be interpreted as the volatility. Actually, if the tail of the function ϕ is not too heavy, X is indeed differentiable. Let us write

$$F^{\alpha,\lambda}(t) = \int_0^t f^{\alpha,\lambda}(s) ds.$$

The following result holds.

Theorem 2. *Let (X_t) be a process satisfying (8) for $t \in [0,1]$ and assume $\alpha > 1/2$. Then X is differentiable on $[0,1]$ and the law of its derivative Y is a weak solution of the stochastic integral equation*

$$Y_t = F^{\alpha,\lambda}(t) + \frac{1}{\sqrt{\mu^* \lambda}} \int_0^t f^{\alpha,\lambda}(t-s)\sqrt{Y_s} dB_s^1,$$

with B^1 a Brownian motion. Furthermore, for any $\varepsilon > 0$, Y has Hölder regularity $\alpha - 1/2 - \varepsilon$.

3.4 Discussion

We now comment the results given in Theorem 1 and Theorem 2.

• The singularity at zero of the function $f^{\alpha,\lambda}$ appearing in our two theorems is of order $x^{\alpha-1}$. Making an analogy with the Volterra representation of the fractional Brownian motion (7), this corresponds to a Hurst parameter H equal to $\alpha - 1/2$. Thus, in the case $\alpha > 1/2$ where our volatility process is well defined, because of the square root term in front of the Brownian motion, we can somehow interpret it as a fractional CIR process with Hurst parameter equal to $\alpha - 1/2$. This corresponds to a very rough process, with Hölder regularity close to zero when α is close to $1/2$. As mentioned in the introduction, this is perfectly consistent with recent empirical measures of the volatility smoothness on financial data, see Chapter I. Note however that this model does not reproduce the geometric nature of volatility observed in Chapter I.

• A practical consequence of the preceding point is the following: When observing on a time interval of order

$$\frac{1}{(1 - \|\phi\|_1)^{1/\alpha}}$$

a Hawkes process with kernel ϕ with L^1 norm close to one and power law tail with index $1 + \alpha$, then after scaling in space, a fractional-like behavior is obtained.

• Theorem 2 relates the smoothness of the volatility process to the tail parameter α. This is particularly interesting for financial applications. Indeed, the parameter α is usually considered very hard to measure. Our theorem provides an approach where it can be obtained relying on the smoothness of the volatility, which is much easier to estimate, see Chapter I.

• The irregular volatility process appearing at the limit arises because our processes are nearly unstable with heavy tailed kernels. As explained in the introduction, in financial terms, it means that the rough behavior of the volatility can be explained by the high degree of endogeneity of modern markets combined with the persistent nature of the order flows.

• Note that Hawkes processes with L^1 norm exactly equal to one have been introduced in [BM01]. In this work, they show that in order to get a stationary intensity, the parameter μ must be equal to zero and the regression kernel has to be heavy tailed. Several additional results for the non stationary heavy tailed case ($\|\phi|_1 = 1$ and $\mu > 0$) can be found in [Zhu13].

• Compared to the approach in Chapter II, it is important to remark that our volatility process is simply the derivative of the limit of the sequence of nearly unstable heavy tailed Hawkes processes. However, contrary to what is done in Chapter II, we do not provide any result about the convergence of the sequence of intensities of the Hawkes processes. In particular, the sequence of intensities is not shown to converge towards the volatility[4]. Remark also that our assumptions are slightly weaker than those in Chapter II. In particular, we do not need the function ϕ to be bounded. Again, this is relevant for financial applications where $\phi(t)$ becomes typically very large as t tends to zero, see Chapter V.

4 Proofs

We give in this section the proofs of Proposition 2, Theorem 1 and Theorem 2. In the sequel, c denotes a positive constant which may vary from line to line (and even within the same line if no ambiguity).

4.1 Proof of Proposition 2

We show here the tightness of (Z^T, X^T). We start with the following lemma.

Lemma 1. *The sequences X^T and Λ^T are $\mathbb{C}$-tight.*

Proof. From [BDHM12], we get that the expectation of the Hawkes process N_t^T satisfies

$$\mathbb{E}[N_t^T] = \mu_T t + \mu_T \int_0^t \psi^T(t-s)s ds \leq t\mu_T(1 + \|\psi^T\|_1).$$

Therefore, since

$$\|\psi^T\|_1 \leq \frac{c}{1-a_T},$$

we get

$$\mathbb{E}[X_1^T] = \mathbb{E}[\Lambda_1^T] \leq c.$$

The tightness of X^T and Λ^T then follows, using the fact that both processes are increasing.

Moreover, since $(1-a_T)/\mu_T$ tends to zero, the maximum jump size of X^T and Λ^T (which is continuous) goes to zero as T tends to infinity. From Proposition VI-3.26 in [JS87], this implies the $\mathbb{C}$-tightness of X^T and Λ^T. □

We now give the proof of Proposition 2. It is easy to get that the angle bracket of Z^T is Λ^T. From Lemma 1, it is $\mathbb{C}$-tight. Thus, from Theorem VI-4.13 in [JS87], the sequence (Z^T) is tight. Finally, marginal tightness implies the joint tightness of (Z^T, X^T).

Let us now consider a subsequence (Z^{T_n}, X^{T_n}) converging towards a process that we denote by (Z, X). Using Proposition VI-6.26 in [JS87] together with the fact that the bracket of (Z^{T_n})

[4] Actually it can be shown that for some reasonable functions ϕ, the sequence of intensities does not converge, at least in the Skorohod topology.

is (X^{T_n}), we get that $X = [Z, Z]$.

Since $\sqrt{\frac{1-a_T}{T^\alpha}}$ goes to 0, the maximum jump size of Z^T tends to zero. Therefore, Z^T is $\mathbb{C}$-tight and so the limit Z is continuous. It remains to show that Z is a martingale. Using Corollary IX.1.19 of [JS87], Z is a local martingale. Moreover, the expectation of its bracket being finite, it is a martingale.

4.2 Proof of Theorem 1

We start with the following lemma which shows that we can work with Λ^T rather than with X^T.

Lemma 2. *The sequence of martingales $X^T - \Lambda^T$ tends to zero in probability, uniformly on* $[0, 1]$.

Proof. We have

$$X_t^T - \Lambda_t^T = \frac{1 - a_T}{T^\alpha \mu^* \delta^{-1}} M_{tT}^T.$$

Applying Doob's inequality to the martingale M^T, we get

$$\mathbb{E}[\sup_{t\in[0,1]} \{(X_t^T - \Lambda_t^T)^2\}] \leq c(\frac{1 - a_T}{T^\alpha})^2 \mathbb{E}[(M_T^T)^2].$$

Then, the bracket of M^T being N^T, we deduce

$$\mathbb{E}[\sup_{t\in[0,1]} \{(X_t^T - \Lambda_t^T)^2\}] \leq c(\frac{1 - a_T}{T^\alpha})^2 \mathbb{E}[N_T^T] \leq c\frac{\mu_T(1 - a_T)}{T^{2\alpha}} \leq c\frac{1 - a_T}{T^\alpha},$$

which ends the proof. □

We state here a lemma which will be useful in the proof of Equation (8).

Lemma 3. *The sequence of measures with density $\rho^T(x)$ defined by Equation* (4) *converges weakly towards the measure with density $\lambda x^{\alpha-1} E_{\alpha,\alpha}(-\lambda x^\alpha)$. In particular,*

$$F^T(t) = \int_0^t \rho^T(x) dx$$

converges uniformly towards

$$F^{\alpha,\lambda}(t) = \int_0^t f^{\alpha,\lambda}(x) dx.$$

Proof. The proof of this result is obtained showing that the Laplace transform of the measure with density $\rho^T(x)$ converges towards the Laplace transform of the measure with density $\lambda x^{\alpha-1} E_{\alpha,\alpha}(-\lambda x^\alpha)$. This has already been done in Section 2. □

We now give the proof of Equation (8). Let us consider a converging subsequence (Z^{T_n}, X^{T_n}) and write (Z, X) its limit. Abusing notation slightly, we write (Z^T, X^T) instead of (Z^{T_n}, X^{T_n}). Using Skorokhod's representation theorem, there exists a probability space on which one can define copies in law of the (Z^T, X^T) converging almost surely for the Skorohod topology to a random variable with the same law as (Z, X). We now work with these variables converging almost surely and their limits. The processes Z and X being continuous, we have

$$\sup_{t\in[0,1]} |X_t^T - X_t| \to 0, \quad \sup_{t\in[0,1]} |Z_t^T - Z_t| \to 0. \tag{9}$$

Let us now rewrite the cumulated intensity. For all $t \geq 0$, we have

$$\int_0^t \lambda_s^T ds = t\mu_T + \int_0^t \phi^T(t-s)(\int_0^s \lambda_u^T du)ds + \int_0^t \phi^T(t-s)M_s^T ds.$$

Now remark that

$$\begin{aligned}
\int_0^t \psi^T(t-s)\int_0^s \phi^T(s-r)M_r^T dr ds &= \int_0^t \int_0^t 1_{r\leq s}\psi^T(t-s)\phi^T(s-r)ds M_r^T dr \\
&= \int_0^t \int_0^{t-r} \psi^T(t-r-s)\phi^T(s)ds M_r^T dr \\
&= \int_0^t \psi^T * \phi^T(t-r)M_r^T dr \\
&= \int_0^t \psi^T(t-r)M_r^T dr - \int_0^t \phi^T(t-r)M_r^T dr,
\end{aligned}$$

where ψ^T is defined in Equation (2). Using this together with Lemma 3 in [BDHM12], we easily get

$$\int_0^t \lambda_s^T ds = t\mu_T + \int_0^t \psi^T(t-s)s\mu_T ds + \int_0^t \psi^T(t-s)M_s^T ds.$$

Therefore, replacing t by Tt, multiplying by $(1-a_T)/(T^\alpha \mu^* \delta^{-1})$, and writing

$$u_T = \frac{\mu_T}{\mu^* \delta^{-1} T^{\alpha-1}},$$

we get $\Lambda^T(t) = T_1 + T_2 + T_3$, with

$$\begin{aligned}
T_1 &= (1-a_T)tu_T, \\
T_2 &= T(1-a_T)u_T \int_0^t \psi^T(T(t-s))s ds, \\
T_3 &= T^{1-\alpha/2}\sqrt{\frac{(1-a_T)}{\mu^*\delta^{-1}}}\int_0^t \psi^T(T(t-s))Z_s^T ds.
\end{aligned}$$

Since u_T converges to 1, we get that T_1 goes to zero. For T_2, note that integrating by parts, we have

$$T_2 = \frac{u_T}{a_T}\int_0^t \rho^T(t-s)s ds = \frac{u_T}{a_T}\int_0^t F^T(t-s)ds.$$

Using Lemma 3 and integrating by parts again, we obtain that T_2 tends uniformly to

$$\int_0^t F^{\alpha,\lambda}(t-s)ds = \int_0^t f^{\alpha,\lambda}(t-s)sds.$$

We now turn to T_3. Remark that

$$T_3 = \frac{a_T}{\sqrt{T^\alpha(1-a_T)\mu^*\delta^{-1}}}\int_0^t \rho^T(t-s)Z_s^T ds$$

and recall that

$$Z_t^T = \sqrt{\frac{T^\alpha\mu^*\delta^{-1}}{1-a_T}}(X_t^T - \Lambda_t^T).$$

Thus, using that X^T is piecewise constant, applying integration by parts, we get (pathwise)

$$\int_0^t \rho^T(t-s)Z_s^T ds = \int_0^t F^T(t-s)dZ_s^T$$

and in the same way

$$\int_0^t f^{\alpha,\lambda}(t-s)Z_s^T ds = \int_0^t F^{\alpha,\lambda}(t-s)dZ_s^T.$$

Then,

$$\mathbb{E}\Big[\Big(\int_0^t (F^{\alpha,\lambda}(t-s) - F^T(t-s))dZ_s^T\Big)^2\Big] \le c\int_0^t \big(F^{\alpha,\lambda}(t-s) - F^T(t-s)\big)^2 ds,$$

which tends to zero thanks to Lemma 3. Furthermore, using (9), we get that

$$\int_0^t \lambda(t-s)^{\alpha-1}E_{\alpha,\alpha}(-\lambda(t-s)^\alpha)|Z_s - Z_s^T|ds$$

also tends to zero. Consequently, we finally obtain that for any t, T_3 converges to

$$\frac{1}{\sqrt{\mu^*\lambda}}\int_0^t (t-s)^{\alpha-1}E_{\alpha,\alpha}(-\lambda(t-s)^\alpha)Z_s ds.$$

Since Z is a continuous martingale, the fact that $Z_t = B_{X_t}$ is a consequence of Dambis-Dubin-Schwarz Theorem, see for example Theorem V-1.6 in [RY99].

4.3 Regularity of the solutions of Equation (8)

We start with the following lemma.

Lemma 4. *Let B be a Brownian motion and X a solution of (8) associated to B. Let H in $(0,1)$. If X has Hölder regularity H on $[0,1]$, then for any $\varepsilon > 0$, X has also Hölder regularity $((\alpha + H/2) \wedge 1) - \varepsilon$ on $[0,1]$.*

Proof. Let $\varepsilon > 0$ and $Z_t = B_{X_t}$. The function

$$t \to \int_0^t s f^{\alpha,\lambda}(t-s)ds$$

being $\mathscr{C}^1$, it is enough to show that

$$t \to \int_0^t f^{\alpha,\lambda}(t-s)Z_s ds$$

has Hölder regularity $((\alpha+H/2)\wedge 1)-\varepsilon$. Since for any $\varepsilon' > 0$, Z has Hölder regularity $(H/2-\varepsilon')$, by Lemma 6, it is $(H/2-\varepsilon)$ fractionally differentiable and $D^{H/2-\varepsilon}Z$ is continuous. Using the fact that $f^{\alpha,\lambda}$ is fractionally integrable, from Corollary 1, we get that

$$\int_0^t f^{\alpha,\lambda}(t-s)Z_s ds = \int_0^t I^{H/2-\varepsilon} f^{\alpha,\lambda}(t-s) D^{H/2-\varepsilon} Z_s ds.$$

Finally, the results about $I^{H/2-\varepsilon} f^{\alpha,\lambda}$ stated in Proposition 1 together with Lemma 8 give the result. □

Lemma 4 enables us to obtain the following result about the regularity of X and Z.

Lemma 5. *Let B be a Brownian motion, X a solution of* (8) *associated to B and $Z_t = B_{X_t}$. Then, for any $\varepsilon > 0$, almost surely,*

- *The process X has Hölder regularity $(1 \wedge 2\alpha) - \varepsilon$ on $[0,1]$.*
- *The process Z has Hölder regularity $(1/2 \wedge \alpha) - \varepsilon$ on $[0,1]$.*

Proof. Let M be the supremum of the Hölder exponents of X. From Proposition 1 together with Lemma 8, we get that $M \geq \alpha$.

Let us now assume that $M < (1 \wedge 2\alpha)$. Then we can find some $H < M$ and some $\varepsilon > 0$ such that

$$M < \big((\alpha + H/2) \wedge 1\big) - \varepsilon.$$

Thus, since X has Hölder regularity H, Lemma 4 implies that X has also Hölder regularity

$$\big((\alpha + H/2) \wedge 1\big) - \varepsilon,$$

which is a contradiction. Therefore $M \geq (1 \wedge 2\alpha)$, which ends the proof.

Since $Z_t = B_{X_t}$, the second part of the lemma is a direct consequence of the first part.

□

4.4 Proof of Theorem 2

Using Lemma 1, Lemma 5 and Corollary 2, for any $\nu \in (0, \alpha)$, we can rewrite Equation (8) as

$$X_t = \int_0^t s f^{\alpha,\lambda}(t-s)ds + \frac{1}{\sqrt{\mu^* \lambda}} \int_0^t D^\nu f^{\alpha,\lambda}(t-s) I^\nu Z_s ds.$$

Moreover, taking $\nu > 1/2$, since Z is $1-\nu$ fractionally differentiable, we get

$$I^\nu Z_s = \int_0^s D^{1-\nu} Z_u du.$$

Thus, using Fubini's theorem, we obtain

$$\begin{aligned}
\int_0^t D^\nu f^{\alpha,\lambda}(t-s) I^\nu Z_s ds &= \int_0^t \int_0^s D^\nu f^{\alpha,\lambda}(t-s) D^{1-\nu} Z_u du ds \\
&= \int_0^t \int_u^t D^\nu f^{\alpha,\lambda}(t-s) D^{1-\nu} Z_u ds du \\
&= \int_0^t \int_u^t D^\nu f^{\alpha,\lambda}(s-u) D^{1-\nu} Z_u ds du \\
&= \int_0^t \int_0^s D^\nu f^{\alpha,\lambda}(s-u) D^{1-\nu} Z_u du ds.
\end{aligned}$$

Hence, we get

$$X_t = \int_0^t Y_s ds,$$

with

$$Y_s = F^{\alpha,\lambda}(s) + \frac{1}{\sqrt{\mu^* \lambda}} \int_0^s D^\nu f^{\alpha,\lambda}(s-u) D^{1-\nu} Z_u du.$$

From Lemma 1 together with Lemma 8, we get that Y has Hölder regularity $(\alpha - \nu)$. Thus, taking ν close enough to $1/2$, we get that for any $\varepsilon > 0$, Y has Hölder regularity $(\alpha - 1/2 - \varepsilon)$. This implies that X is differentiable with derivative Y.

Now, since Z is a continuous martingale with bracket X and because $\nu > 1/2$, we can use the stochastic Fubini theorem, see for example [Ver12], to obtain

$$\begin{aligned}
D^{1-\nu} Z_s &= \frac{1}{\Gamma(\nu)} \frac{d}{ds} \int_0^s \frac{Z_v}{(s-v)^{1-\nu}} dv \\
&= \frac{1}{\Gamma(\nu)} \frac{d}{ds} \int_0^s \int_0^v \frac{1}{(s-v)^{1-\nu}} dZ_u dv \\
&= \frac{1}{\Gamma(\nu)} \frac{d}{ds} \int_0^s \int_u^s \frac{1}{(s-v)^{1-\nu}} dv dZ_u \\
&= \frac{1}{\Gamma(\nu+1)} \frac{d}{ds} \int_0^s (s-u)^\nu dZ_u.
\end{aligned}$$

Therefore,

$$Y_t = F^{\alpha,\lambda}(t) + \frac{1}{\sqrt{\mu^*\lambda}} \int_0^t D^\nu f^\alpha(t-s) \frac{1}{\Gamma(\nu+1)} \frac{d}{ds} \int_0^s (s-u)^\nu dZ_u ds.$$

Using Fubini's theorem twice and the fact that $f * (g') = (f * g)'$, we derive

$$\begin{aligned}
Y_t &= F^{\alpha,\lambda}(t) + \frac{1}{\sqrt{\mu^*\lambda}} \frac{d}{dt} \int_0^t \frac{1}{\Gamma(\nu+1)} \int_u^t D^\nu f^\alpha(t-s)(s-u)^\nu ds dZ_u \\
&= F^{\alpha,\lambda}(t) + \frac{1}{\sqrt{\mu^*\lambda}} \frac{d}{dt} \int_0^t I^{\nu+1} D^\nu f^\alpha(t-u) dZ_u \\
&= F^{\alpha,\lambda}(t) + \frac{1}{\sqrt{\mu^*\lambda}} \frac{d}{dt} \int_0^t \int_0^v I^\nu D^\nu f^\alpha(v-u) dZ_u dv \\
&= F^{\alpha,\lambda}(t) + \frac{1}{\sqrt{\mu^*\lambda}} \int_0^t f^\alpha(t-u) dZ_u.
\end{aligned}$$

Moreover, using Theorem V-3.8 of [RY99], there exists a Brownian motion B^1 such that

$$Z_t = \int_0^t \sqrt{Y_s} dB_s^1.$$

So, consider now the process $(\tilde{Y}_t)$ defined by

$$\tilde{Y}_t = F^{\alpha,\lambda}(t) + \frac{1}{\sqrt{\mu^*\lambda}} \int_0^t f^\alpha(t-u) \sqrt{Y_s} dB_s^1.$$

Going backward in the previous computations for Y_t and $D^{1-\nu} Z_s$, we remark that

$$\tilde{Y}_t = F^{\alpha,\lambda}(t) + \frac{1}{\sqrt{\mu^*\lambda}} \int_0^t D^\nu f^\alpha(t-s) \frac{1}{\Gamma(\nu+1)} \frac{d}{ds} \int_0^s (s-u)^\nu \sqrt{Y_u} dB_u^1 ds$$

and

$$\begin{aligned}
\frac{1}{\Gamma(\nu+1)} \frac{d}{ds} \int_0^s (s-u)^\nu \sqrt{Y_u} dB_u^1 &= \frac{1}{\Gamma(\nu)} \int_0^s \int_0^v \frac{1}{(s-v)^{1-\nu}} \sqrt{Y_u} dB_u^1 dv \\
&= \frac{1}{\Gamma(\nu)} \frac{d}{ds} \int_0^s \frac{1}{(s-v)^{1-\nu}} \Big(\int_0^v \sqrt{Y_u} dB_u^1 \Big) dv \\
&= \frac{1}{\Gamma(\nu)} \frac{d}{ds} \int_0^s \frac{Z_v}{(s-v)^{1-\nu}} dv \\
&= D^{1-\nu} Z_s.
\end{aligned}$$

Therefore,

$$\tilde{Y}_t = F^{\alpha,\lambda}(t) + \frac{1}{\sqrt{\mu^*\lambda}} \int_0^t D^\nu f^\alpha(t-s) D^{1-\nu} Z_s ds = Y_t.$$

Consequently,

$$Y_t = F^{\alpha,\lambda}(t) + \frac{1}{\sqrt{\mu^*\lambda}} \int_0^t f^\alpha(t-u) \sqrt{Y_s} dB_s^1.$$

III.A Technical results

In this section, we begin by gathering some useful results from [SKM93]. We denote H^λ the set of functions on $[0,1]$ with Hölder regularity λ.

Theorem 3.1 in [SKM93] gives a link between the Hölder exponent of a function and the Hölder exponent of its fractional integrals.

Proposition 3. *Let $\alpha > 0$, $\lambda > 0$ such that $\alpha + \lambda < 1$. If $\phi \in H^\lambda$, then ϕ has a fractional integral of order α and*

$$I^\alpha \phi(x) = \frac{\phi(0)}{\Gamma(1+\alpha)} x^\alpha + \psi(x)$$

with $\psi \in H^{\alpha+\lambda}$.

Lemma 13.1 of [SKM93] relates the Hölder exponent of a function and the Hölder exponent of its fractional derivatives.

Lemma 6. *If $f \in H^\lambda$ and $f(0) = 0$, then for any $\alpha < \lambda$, f admits a fractional derivative of order α and $D^\alpha f \in H^{\lambda-\alpha}$.*

Theorem 13.6 of [SKM93] gives another criterion for a function to have a fractional derivative.

Proposition 4. *If $f(x) = x^{-\mu} g(x)$ with $g \in H^\lambda$ where $\lambda > \alpha$, $-\alpha < \mu < 1$ and p such that $\mu + \alpha < 1/p$, then f is fractionally differentiable of order α and $D^\alpha f \in L^p$.*

Equation 2.20 of [SKM93] is a fractional integration by parts which can be written as follows.

Proposition 5. *If $\phi \in L^p$ and $\psi \in L^q$ with $1/p + 1/q \leq 1 + \alpha$, then ϕ and ψ have an integral of order α and*

$$\int_0^t \phi(t-s) I^\alpha \psi(s) ds = \int_0^t I^\alpha \phi(t-s) \psi(s) ds.$$

In this work, we mainly use the two following corollaries of Proposition 5.

Corollary 1. *Let $\phi \in L^r$, with $r > 1$ and $\psi \in H^\beta$. Then, for any $\alpha < \beta$, $D^\alpha \psi$ exists, belongs to $H^{\beta-\alpha}$ and*

$$\int_0^t \phi(t-s) \psi(s) ds = \int_0^t I^\alpha \phi(t-s) D^\alpha \psi(s) ds.$$

Corollary 2. *Let ϕ be continuous and ψ such that $x^\mu \psi(x) \in H^\lambda$. Then, for any $\alpha < \min(1-\mu, \lambda)$, $D^\alpha \psi$ exists, belongs to L^r for some $r > 1$ and*

$$\int_0^t \phi(t-s) \psi(s) ds = \int_0^t I^\alpha \phi(t-s) D^\alpha \psi(s) ds.$$

The following result is immediate from the definition of fractional differentiation.

Lemma 7. *If f and g are continuous fractionally differentiable of order $\nu < 1$, then for any $t \in [0,1]$,*

$$|D^\nu f(t) - D^\nu g(t)| \leq K \sup_{s\in[0,t]} |f(s) - g(s)|.$$

Finally, the next result is about the smoothness of the convolution of a power law function with a continuous function.

Lemma 8. *Let f be a differentiable function on $(0,1]$ such that for some $K > 0$ and $0 < \beta < 1$*

$$\forall x \in (0,1] \quad |f(x)| \leq \frac{K}{x^\beta} \text{ and } |f'(x)| \leq \frac{K}{x^{\beta+1}},$$

and g a continuous function on $[0,1]$. Then the convolution

$$f * g(t) = \int_0^t f(t-s)g(s)ds$$

has Hölder regularity $(1-\beta)$.

Proof. We set $x = \beta/(\beta+1)$, $G = |g|_\infty$ and we split $f*g(t+h) - f*g(t)$ into the three following terms:

$$\begin{aligned} f*g(t+h) - f*g(t) &= \int_t^{t+h} f(t+h-s)g(s)ds \\ &+ \int_{t-h}^{t} (f(t+h-s) - f(t-s))g(s)ds \\ &+ \int_0^{t-h} (f(t+h-s) - f(t-s))g(s)ds. \end{aligned}$$

The first term is bounded by $KG\frac{h^{1-\beta}}{1-\beta}$, the second by $KG(1+\frac{1}{1-\beta})h^{1-\beta}$ and the third by

$$G\int_0^{t-h}\int_{t-s}^{t+h-s} f'(u)duds \leq GK\int_0^{t-h} h\frac{1}{(t-s)^{1+\beta}}ds \leq \frac{2}{\beta}GKh^{1-\beta}.$$

□

CHAPTER IV

The different asymptotic regimes of nearly unstable autoregressive processes

Abstract

We extend the results of [CS82, Phi87, vdMPvZ99] about the convergence of nearly unstable AR(p) processes to the infinite order case. To do so, we proceed as in Chapters II and III by using limit theorems for some well chosen geometric sums. We prove that when the coefficients sequence has a light tail, nearly unstable AR(∞) processes behave as Ornstein-Uhlenbeck models. However, in the heavy tail case, we show that fractional diffusions arise as limiting laws for such processes.

Keywords: Autoregressive processes, AR(∞), nearly unstable processes, limit theorems, Ornstein-Uhlenbeck processes, fractional diffusions, volatility modeling.

1 Introduction

In the field of time series analysis, autoregressive processes (AR processes for short) probably represent the most classical class of models. In this work, a discrete time process y is said to be autoregressive if it satisfies $y_0 = \varepsilon_0$ and for $k \geq 1$,

$$y_k = \varepsilon_k + \sum_{i=1}^{k} \phi_i y_{k-i},$$

where ϕ is a sequence of non-negative coefficients and the ε_k are iid centered random variables with finite second order moment. The process is called AR(∞) if an infinite number of coefficients are non-zero and AR(p) if they are all equal to zero after rank p. A first reason for the popularity of AR processes is the fact that they are quite tractable. Furthermore, it is usually easy to give an interpretation for such dynamics in practice. That is why they are used in various fields such as population dynamics, see [CS82, Roy92], finance, see [Has91], or telecommunications, see [Ada97][1].

[1] Of course the references given here are just some examples among many others.

For classical AR processes starting at $-\infty$ and with non-negative coefficients, it is well known that a necessary condition for the existence of a stationary solution is

$$\sum_{i=1}^{+\infty} \phi_i < 1. \tag{1}$$

Thus, (1) is called stability condition. However, in many applications, notably finance, the case where this stability condition is almost saturated:

$$0 \leq 1 - \sum_{i=1}^{+\infty} \phi_i \ll 1$$

seems to be relevant, see [CS82, Mar13].

To address this near instability situation, several authors have considered the framework of sequences of AR(1) processes of the form

$$y_k^n = \rho_n y_{k-1}^n + \varepsilon_k,$$

with the autoregression parameter ρ_n tending to one as n goes to infinity, see for example [CS82, MSR12, Phi87]. In these works, it is shown that after suitable renormalization, such processes asymptotically behave as Ornstein-Uhlenbeck models. The estimation of the AR coefficient has also been extensively studied in such an asymptotic, see [Cha88, CW87] and [BC$^+$07, Cha09, CZ09] for more recent related developments. The situation where the number of coefficients in the autoregressive process is larger than one is treated in [BC13, Jeg91, vdMPvZ99]. However, to our knowledge, no result is available in the case of nearly unstable AR(∞) processes. The aim of this work is to fill this gap, using a methodology originally developed for Hawkes processes in Chapters II and III.

A Hawkes process (N_t) is a point process whose intensity at time t, denoted by λ_t, is of the form

$$\lambda_t = \mu + \int_{-\infty}^{t} \phi(t-s) dN_s,$$

with μ a constant and ϕ a non-negative measurable function, see [Haw71a]. The linearity in the previous equation implies that Hawkes processes have a common structure with autoregressive processes. In particular, under the stability condition that the L^1 norm of ϕ, denoted by $|\phi|_{L^1}$, is strictly smaller than one, Hawkes processes admit a stationary intensity. Let us also mention that the counterpart of the Yule-Walker equations linking the autocorrelations of an AR process to its coefficients is the Wiener-Hopf equation, see [BM14b, Haw71a].

In Chapters II and III, nearly unstable Hawkes processes are introduced and their asymptotic behavior is investigated. Such processes are defined as Hawkes processes for which the stability condition is almost saturated: $1 - |\phi|_{L^1} \ll 1$. The key idea to understand the limiting laws of nearly unstable Hawkes processes is to relate their dynamics with some geometric sums whose parameter tends to zero. In this work, we follow the same strategy in order to derive

asymptotic results for nearly unstable AR(∞) processes.

We show that the limiting behavior of nearly unstable AR(∞) processes strongly depends on the properties of the sequence $\phi = (\phi_1, \phi_2, \ldots)$. Two main situations have to be considered: the light tail case, where ϕ_i goes rapidly to zero (or is even equal to zero for i large enough) and the heavy tail case, where the decay of the sequence ϕ is slow. We prove that in the first situation, the cumulated process asymptotically behaves as an integrated Ornstein-Uhlenbeck model, in agreement with the results obtained in the literature for AR(p) processes. However, in the heavy tail case, a fractional limit is obtained. More precisely, the renormalized cumulated AR(∞) process behaves as a kind of integrated fractional Brownian motion with Hurst parameter smaller than 1/2, or as a process close to a fractional Brownian motion with Hurst parameter larger than 1/2, depending on the decay rate of the sequence ϕ.

The paper is organized as follows. In Section 2, we introduce the geometric sums and derive their limiting behavior as their parameter tends to zero. We describe our asymptotic framework in Section 3. Section 4 contains our main results about the limiting laws of nearly unstable AR(∞) processes and their proofs. We also interpret these results in terms of volatility modeling in Section 5. Finally, some technical lemmas are relegated to an appendix.

2 Geometric sums

In this section, we consider the asymptotic behavior of some geometric sums. Understanding these sums is actually the key point in order to be able to derive limiting laws for nearly unstable AR(∞) processes.

2.1 Geometric sums and convolution

Let $(a_n)_{n\geq 1}$ and $(\phi_i)_{i\geq 0}$ be two sequences of non-negative numbers such that $a_n < 1$, $\phi_0 = 0$ and

$$\sum_{i=1}^{+\infty} \phi_i = 1.$$

In the next sections, a_n will correspond to the sum of the autoregression coefficients and the ϕ_i to the (fixed) shape of these coefficients, see Section 3.

In this work, geometric sums, see [Kal97], are random variables of the form

$$Y^n = \sum_{i=1}^{I^n} X^i,$$

where I^n is a geometric random variable with parameter $1 - {a_n}^2$ (with the convention $\sum_1^0 = 0$) and (X^i) is a sequence of strictly positive integer-valued iid random variables such that for any $i \geq 1$,

$$\mathbb{P}[X^1 = i] = \phi_i.$$

[2] $\forall i \geq 0,\ \mathbb{P}[I^n = i] = a_n^i(1 - a_n)$.

Now we define the sequence ψ^n by

$$\psi^n = \sum_{l=0}^{+\infty} (a_n \phi)^{*l},$$

with ϕ^{*0} the sequence indexed by $\mathbb{N}$ such that $\phi_0^{*0} = 1$ and for $i > 0$, $\phi_i^{*0} = 0$ and for $n \geq 0$, $\phi^{*(n+1)} = \phi^{*n} * \phi$, where $*$ denotes the convolution operator for sequences. Note that

$$\sum_{i=0}^{+\infty} \psi_i^n = \frac{1}{1 - a_n}.$$

Actually, the random variable Y^n and the sequence ψ^n are closely related since for any $i \geq 0$,

$$\mathbb{P}[Y^n = i] = \frac{\psi_i^n}{\sum_{j \geq 0} \psi_j^n}.$$

We will use the limiting behavior of Y^n to obtain asymptotic properties for the sequence ψ^n, see Corollaries 1 and 2. This will provide us key results for our study of nearly unstable AR(∞) processes at large time scales. Indeed, in the next sections, for n tending to infinity, we consider on time intervals of the form $[0, nt]$ autoregressive processes whose coefficients sequence can be written $a_n \phi$:

$$y_k^n = \varepsilon_k + \sum_{i=1}^{k} a_n \phi_i y_{k-i}^n,$$

with a_n tending to one (nearly unstable case). Then, it is easy to obtain the following moving average representation of the process:

$$y_k^n = \sum_{i=0}^{k} \psi_{k-i}^n \varepsilon_i$$

and in particular

$$y_{\lfloor nt \rfloor}^n = \sum_{i=0}^{\lfloor nt \rfloor} \psi_{\lfloor nt \rfloor - i}^n \varepsilon_i.$$

Thanks to the previous representation, we see that the long term behavior of y^n is linked with the asymptotic properties of the $\psi_{\lfloor nt \rfloor - i}^n$ and therefore with those of Y^n / n. Indeed, for example, for nt an integer,

$$\mathbb{P}[Y^n / n = t] = \frac{\psi_{nt}^n}{\sum_{j \geq 0} \psi_j^n}.$$

We now recall some results from Chapters II and III which explain the two possible types of asymptotic behaviors for our geometric sums. Exhibiting one behavior or the other depends on the decay rate of the coefficients of ϕ, that is on the tail of the distribution of X^1.

2.2 Light tail case

The first asymptotic regime of interest is the one which occurs when the expectation of X^1 is finite. Thus we consider the following assumption.

Assumption 1. *We have*

$$\sum_{i=1}^{+\infty} i\phi_i = m < +\infty.$$

In that situation, the "appropriate" speed of convergence of a_n towards one for the nearly unstable case is given in the next assumption.

Assumption 2. *We have $a_n < 1$ and there exists $\lambda > 0$ such that*

$$n(1-a_n) \underset{n\to+\infty}{\to} \lambda.$$

Indeed, recall Proposition 2.2 of Chapter II.

Proposition 1. *Under Assumptions 1 and 2, the random variable Y^n/n converges in law towards an exponential random variable with parameter λ/m.*

Using Dini's theorem, this implies the next corollary.

Corollary 1. *Under Assumptions 1 and 2,*

$$\sup_{x\in[0,1]} \left| \sum_{i=0}^{\lfloor nx \rfloor} \psi_i^n (1-a_n) - (1-e^{-\frac{\lambda}{m}x}) \right| \underset{n\to+\infty}{\to} 0.$$

2.3 Heavy tail case

The second asymptotic regime of interest is the one which occurs when the tail of the distribution of X^1 is heavy. This can be formalized as follows.

Assumption 3. *There exists $K > 0$ and $\alpha \in (0,1)$ such that*

$$\sum_{i=N}^{+\infty} \phi_i \underset{N\to+\infty}{\sim} \frac{K}{N^\alpha}.$$

Let us now write

$$\delta = K\Gamma(1-\alpha)/\alpha,$$

with Γ the Gamma function. In that heavy tail case, the "appropriate" convergence speed of a_n towards one is given in the next assumption.

Assumption 4. *We have $a_n < 1$ and there exists $\lambda > 0$ such that*

$$n^\alpha(1-a_n) \underset{n\to+\infty}{\to} \lambda\delta.$$

Indeed, recall Proposition 2.3 of Chapter II.

Proposition 2. *Under Assumptions 3 and 4, the random variable Y^n/n converges in law towards the random variable whose Laplace transform is*

$$\frac{\lambda}{\lambda + z^\alpha}.$$

It is explained in Chapter II that this random variable has the density

$$f^{\alpha,\lambda}(x) = \lambda x^{\alpha-1} E_{\alpha,\alpha}(-\lambda x^\alpha),$$

where $E_{\alpha,\alpha}$ is the Mittag-Leffler function, see [MH08] for definition.

Writing

$$F^{\alpha,\lambda}(x) = \int_0^x f^{\alpha,\lambda}(s)ds$$

and using Dini's theorem, we have the following result.

Corollary 2. *Under Assumptions 3 and 4,*

$$\sup_{x\in[0,1]} |\sum_{i=0}^{\lfloor nx\rfloor} \psi_i^n (1-a_n) - F^{\alpha,\lambda}(x)| \underset{n\to+\infty}{\to} 0.$$

3 The asymptotic setting

As previously explained, we consider a sequence of autoregressive processes of infinite order indexed by n. More precisely, $y_0^n = \varepsilon_0$ and for any $k \geq 1$,

$$y_k^n = \varepsilon_k + \sum_{i=1}^{k} \phi_i^n y_{k-i}^n,$$

where the ϕ_i^n are non-negative coefficients. We assume that the ε_k do not depend on n, are iid centered, and satisfy (for simplicity)

$$\mathbb{E}[(\varepsilon_k)^2] = 1.$$

We are interested in the long term behavior of the cumulative sums of such autoregressive processes when the stability condition (1) is almost saturated. To study this, somehow as in Chapters II and III, we consider that the "shape" of the coefficients sequence is fixed and that their L^1 norm a_n tends to one from below. More precisely, we take

$$\phi_i^n = a_n \phi_i,$$

where

$$\sum_{i=1}^{+\infty} \phi_i = 1$$

and $a_n \to 1$ as n tends to infinity, with $a_n < 1$.
Since we want to investigate the asymptotic properties of cumulated sums of such processes, we introduce the suitably renormalized Donsker line for $t \in [0,1]$:

$$Z_t^n = \frac{1-a_n}{\sqrt{n}} \Big(\sum_{j=0}^{\lfloor nt\rfloor} y_j^n + (nt - \lfloor nt\rfloor) y_{\lfloor nt\rfloor+1}^n \Big).$$

4 Main results

4.1 The light tail case

We give here the behavior of the Donsker line when the autoregression coefficients converge rapidly to zero. Note that the next theorem covers in particular the case of AR(p) processes already studied in the literature.

Theorem 1. *Under Assumptions 1 and 2, (Z^n) converges in law towards the process Z defined by*

$$Z_t = \int_0^t (1 - e^{-\frac{\lambda}{m}(t-s)}) dW_s = \int_0^t \int_0^s \frac{\lambda}{m} e^{-\frac{\lambda}{m}(s-u)} dW_u ds,$$

where W is a Brownian motion.

This theorem means that when observed on a time scale of order $1/(1-a_n)$, a cumulated nearly unstable AR(∞) process whose coefficients sequence has a light tail behaves as an integrated Ornstein-Uhlenbeck process.

To show this result, the first step is to get the following proposition whose proof can be found in appendix.

Proposition 3. *Under Assumptions 1 and 2, the sequence (Z^n) is tight.*

Then we need to prove the finite dimensional convergence. Using the moving average representation of the process, we derive

$$Z_t^n = \sum_{i=0}^{\lfloor nt \rfloor} (1-a_n) \Big(\sum_{k=0}^{\lfloor nt \rfloor - i} \psi_k^n + (nt - \lfloor nt \rfloor) \psi_{\lfloor nt \rfloor + 1 - i}^n \Big) \frac{\varepsilon_i}{\sqrt{n}} + (nt - \lfloor nt \rfloor)(1-a_n) \frac{\varepsilon_{\lfloor nt \rfloor + 1}}{\sqrt{n}}.$$

This can be rewritten under the following integral form:

$$Z_t^n = \int_0^{\lfloor nt \rfloor / n} (1-a_n) \Big(\sum_{k=0}^{\lfloor nt \rfloor - \lfloor ns \rfloor} \psi_k^n + (nt - \lfloor nt \rfloor) \psi_{\lfloor nt \rfloor + 1 - \lfloor ns \rfloor}^n \Big) dW_s^n + (nt - \lfloor nt \rfloor)(1-a_n) \frac{\varepsilon_{\lfloor nt \rfloor + 1}}{\sqrt{n}},$$

where

$$W_s^n = \frac{1}{\sqrt{n}} \sum_{k=0}^{\lfloor ns \rfloor} \varepsilon_k.$$

From Donsker theorem in Skorohod space, we have that W_s^n converges in law towards a Brownian motion for the Skorohod topology. Furthermore, note that

$$(1-a_n) \Big(\sum_{k=0}^{\lfloor nt \rfloor - \lfloor ns \rfloor} \psi_k^n + (nt - \lfloor nt \rfloor) \psi_{\lfloor nt \rfloor + 1 - \lfloor ns \rfloor}^n \Big)$$

belongs to

$$\Big[(1-a_n) \sum_{k=0}^{\lfloor nt \rfloor - \lfloor ns \rfloor} \psi_k^n, (1-a_n) \sum_{k=0}^{\lfloor nt \rfloor + 1 - \lfloor ns \rfloor} \psi_k^n \Big].$$

Thus, from Corollary 1, as a function of s,

$$(1-a_n)\Big(\sum_{k=0}^{\lfloor nt\rfloor-\lfloor ns\rfloor}\psi_k^n+(nt-\lfloor nt\rfloor)\psi_{\lfloor nt\rfloor+1-\lfloor ns\rfloor}^n\Big)$$

tends uniformly to $1-e^{-\frac{\lambda}{m}(t-s)}$ on $[0,t]$. Also, we obviously get that

$$(nt-\lfloor nt\rfloor)(1-a_n)\frac{\varepsilon_{\lfloor nt\rfloor+1}}{\sqrt{n}}$$

tends to zero.

Then, using Theorem 2.2 of [KP91] together with the fact that Skorohod convergence implies pointwise convergence at continuity points of the limit, for given $t\in[0,1]$, we get the convergence in law of Z_t^n towards

$$\int_0^t(1-e^{-\frac{\lambda}{m}(t-s)})dW_s.$$

With the help of Cramer-Wold device, it is easy to extend this result and to show that for any $(t_1,...,t_k)\in[0,1]^k$, we have the convergence in law of $(Z_{t_1}^n,\dots,Z_{t_k}^n)$ towards

$$\Big(\int_0^{t_1}(1-e^{-\frac{\lambda}{m}(t_1-s)})dW_s,...,\int_0^{t_k}(1-e^{-\frac{\lambda}{m}(t_k-s)})dW_s\Big).$$

Together with the tightness of (Z^n), this enables us to obtain the weak converges of (Z^n) towards Z.

4.2 The heavy tail case

Let us now place ourselves under Assumption 3 which states that the coefficients sequence has a power law type behavior. Then, using the geometric sums interpretation, we see that the natural "observation scale" of the process is of order $(1-a_n)^{-1/\alpha}$. This corresponds to Assumption 4. We have the following result.

Theorem 2. *Under Assumptions 3 and 4, (Z^n) converges in law towards the process Z defined by*

$$Z_t=\int_0^t F^{\alpha,\lambda}(t-s)dW_s,$$

where W is a Brownian motion.

In Section 5, we will see that for $\alpha>1/2$, the limiting process can be viewed as an integrated rough fractional process, whereas for $\alpha<1/2$, it is close to a fractional Brownian motion with Hurst parameter larger than $1/2$.

To obtain Theorem 2, the same strategy as for the proof of Theorem 1 is used. In particular, the following proposition is proved in appendix.

Proposition 4. *Under Assumptions 3 and 4, the sequence* (Z^n) *is tight.*

The end the proof of Theorem 2 follows as previously, using the decomposition

$$Z_t^n = \int_0^{\lfloor nt\rfloor/n} (1-a_n)\Big(\sum_{k=0}^{\lfloor nt\rfloor-\lfloor ns\rfloor} \psi_k^n + (nt-\lfloor nt\rfloor)\psi_{\lfloor nt\rfloor+1-\lfloor ns\rfloor}^n\Big) dW_s^n + (nt-\lfloor nt\rfloor)(1-a_n)\frac{\varepsilon_{\lfloor nt\rfloor+1}}{\sqrt{n}},$$

the uniform convergence of

$$(1-a_n)\sum_{k=0}^{\lfloor nt\rfloor} \psi_k^n$$

towards $F^{\alpha,\lambda}(t)$ (see Corollary 2), and the convergence of W_s^n towards a Brownian motion.

5 Application to volatility modeling

Let us now interpret Theorem 2 in terms of volatility modeling on financial markets. Consider that the log-volatility process is driven at discrete times by a nearly unstable autoregressive process with heavy tailed coefficients sequence. We choose to model the log-volatility rather than the volatility itself because it is well established that it is better approximated by a linear autoregressive process than the volatility, see [ABDL03, BKM08, CR98] and Chapter I. Such model can reproduce the clustering property of the volatility at multiple time scales.

5.1 The case $\alpha > 1/2$

When $\alpha > 1/2$, applying the stochastic Fubini's theorem, see [Ver12], we get that the limiting process for the cumulated sums, Z, can be rewritten

$$Z_t = \int_0^t \int_0^s f^{\alpha,\lambda}(s-u)\, dW_u\, ds.$$

Therefore, in that case, it is differentiable and its derivative

$$Y_t = \int_0^t f^{\alpha,\lambda}(t-u)\, dW_u$$

locally behaves as a fractional diffusion with Hurst parameter $H = \alpha - 1/2$. Indeed,

$$f^{\alpha,\lambda}(x) \sim c/x^{1-\alpha}$$

when x is close to zero and recall that a fractional Brownian motion W^H can be written as

$$W_t^H = \int_{-\infty}^t \Big[\frac{1}{(t-s)^{1/2-H}} - \frac{1}{(-s)_+^{1/2-H}}\Big] dW_s.$$

In particular, proceeding as in Chapter III, we get that for any $\varepsilon > 0$, Y has Hölder regularity $\alpha - 1/2 - \varepsilon$.

Thus, in this regime, the log-volatility asymptotically behaves as a fractional Brownian motion with Hurst parameter $\alpha - 1/2$. According to Chapter I, this is consistent with empirical measures of the smoothness of the volatility process provided that $\alpha \simeq 0.6$.

5.2 The case $\alpha < 1/2$

When $\alpha < 1/2$, the behavior of Z is quite different. Proceeding as in Chapter III, we get that for any $\varepsilon > 0$, Z has Hölder regularity $1/2 + \alpha - \varepsilon$. This is not very surprising since the situation $\alpha < 1/2$ is close to that of ARFIMA processes[3], which are known to behave as a fractional Brownian motion with Hurst parameter $1/2 + \alpha$ at large time scales, see [DOT03]. In this regime, the log-volatility exhibits apparent long memory, as observed for example in [ABDL03].

5.3 The case $\alpha = 1/2$

In the critical regime $\alpha = 1/2$, we somehow asymptotically retrieve some features of the multifractal model of [BKM08]. Indeed, in [BKM08], the "log-volatility" $\omega_{l,T}(t)$ is written under the form

$$\omega_{l,T}(t) = \int_{-\infty}^{t} k_{l,T}(t-s)dW_s,$$

where W is a Brownian motion and $k_{l,T}$ is function behaving in the range $l \ll t \ll T$ as

$$k_{l,T}(t) \sim \frac{k_0}{\sqrt{t}},$$

for some model parameters l and T. Therefore, the integrated log-volatility defined as

$$\Omega_{l,T}(t) = \int_0^t \omega_{l,T}(s)ds$$

satisfies

$$\Omega_{l,T}(t) = \int_{-\infty}^{t} \big(K_{l,T}(t-s) - K(-s)\big)dW_s,$$

with

$$K_{l,T}(t) = \mathbb{I}_{t\geq 0}\int_0^t k_{l,T}(s)ds.$$

This behaves as $K_0\sqrt{t}$ in the range $l \ll t \ll T$.

Thus, this "multifractal" regime ($\alpha = 1/2$) appears as the interface between the classical long memory (log-)volatility models ($\alpha < 1/2$) and the more recent rough volatility models ($\alpha > 1/2$).

IV.A Technical results

In the sequel, c denotes a positive constant which may vary from line to line.

[3] An ARFIMA process, see [Ber94], can be written as an infinite order autoregressive process whose sum of the coefficients is equal to one and whose coefficients sequence ϕ asymptotically behaves as $\phi_i \sim c/i^{1+\alpha}$ with $0 < \alpha < 1/2$.

IV.A.1 Proof of Proposition 3

In this paragraph, we place ourselves under Assumptions 1 and 2. Before proving Proposition 3, we need a technical lemma.

Lemma 1. *There exists $c>0$ such that for any n, k, k',*

$$\mathbb{E}[y_k^n y_{k'}^n] \leq \frac{c}{1-a_n}.$$

Proof. For any k, using the moving average representation of the process, since ε is a white noise, we get

$$\mathbb{E}[(y_k^n)^2] = \sum_{j=0}^{k} (\psi_{k-j}^n)^2 \leq \sum_{j=0}^{+\infty} (\psi_j^n)^2.$$

Moreover, using Parseval's Theorem[4],

$$\begin{aligned}\sum_{j=0}^{+\infty} (\psi_j^n)^2 &= 2\int_0^{1/2} |\widehat{\psi^n}(z)|^2 dz \\ &\leq 2\int_0^{1/2} \frac{1}{|1-a_n|^2} dz,\end{aligned}$$

where we have used that

$$|\widehat{\psi^n}(z)| = |\sum_{k\geq 0} (a_n \widehat{\phi}(z))^k| = \frac{1}{|1-a_n\widehat{\phi}(z)|} \leq |\widehat{\psi^n}(0)| = \frac{1}{|1-a_n|}.$$

□

Let us now prove Proposition 3.

The result follows from an application of Kolmogorov's criterion, see [Bil09]. We show here that there exists $c>0$ such that for any n, s and t,

$$\mathbb{E}[|Z_t^n - Z_s^n|^2] \leq c|t-s|^2.$$

We begin with the case where $\lfloor nt \rfloor = \lfloor ns \rfloor$. Using Lemma 1,

$$\begin{aligned}\mathbb{E}[|Z_t^n - Z_s^n|^2] &= \frac{(1-a_n)^2}{n} n^2 |t-s|^2 \mathbb{E}[(y_{\lfloor nt \rfloor+1}^n)^2] \\ &\leq cn(1-a_n)|t-s|^2.\end{aligned}$$

Using Assumption 2, this ends the proof for $\lfloor nt \rfloor = \lfloor ns \rfloor$.

[4] In this work, the Fourier transform of a sequence (f_n) is defined for $z \in [-1/2, 1/2]$ as $\widehat{f}(z) = \sum_{k\geq 0} f_k e^{-2\pi i k z}$.

In the case $\lfloor nt \rfloor = \lfloor ns \rfloor + 1$,

$$\begin{aligned}
\mathbb{E}[|Z_t^n - Z_s^n|^2] &\leq c\big(\mathbb{E}[|Z_t^n - Z_{\lfloor nt \rfloor/n}^n|^2] + \mathbb{E}[|Z_{(\lfloor ns \rfloor+1)/n}^n - Z_s^n|^2]\big) \\
&\leq c|t-s|^2 + \frac{(1-a_n)^2}{n}(1 - ns + \lfloor ns \rfloor)^2 \mathbb{E}[(y_{\lfloor ns \rfloor+1}^n)^2] \\
&\leq c|t-s|^2 + \frac{(1-a_n)}{n}(\lfloor nt \rfloor - ns)^2 \\
&\leq c|t-s|^2.
\end{aligned}$$

We now treat the case where $t = k^t/n$ and $s = k^s/n$, where k^t and k^s are integers so that $k^t > k^s$. Using again Lemma 1 together with Assumption 2, we have

$$\begin{aligned}
\mathbb{E}[|Z_t^n - Z_s^n|^2] &= \frac{(1-a_n)^2}{n} \sum_{k_1=k^s+1}^{k^t} \sum_{k_2=k^s+1}^{k^t} \mathbb{E}[y_{k_1}^n y_{k_2}^n] \\
&\leq \frac{(1-a_n)^2}{n}(k^t - k^s)^2 \frac{c}{1-a_n} \\
&\leq c(1-a_n) n \Big(\frac{k^t - k^s}{n}\Big)^2 \leq c|t-s|^2.
\end{aligned}$$

Finally, for any $t > s$ so that $\lfloor nt \rfloor \geq \lfloor ns \rfloor + 2$, we use the decomposition

$$\mathbb{E}[|Z_t^n - Z_s^n|^2] \leq c\big(\mathbb{E}[|Z_t^n - Z_{\lfloor nt \rfloor/n}^n|^2] + \mathbb{E}[|Z_{\lfloor nt \rfloor/n}^n - Z_{(\lfloor ns \rfloor+1)/n}^n|^2] + \mathbb{E}[|Z_{(\lfloor ns \rfloor+1)/n}^n - Z_s^n|^2]\big).$$

The tightness follows.

IV.A.2 Proof of Proposition 4

In this paragraph, we place ourselves under Assumptions 3 and 4. Before proving Proposition 4, we need some technical results.

Lemma 2. *There exists $c > 0$ such that for any $|z| \leq 1/2$,*

$$|1 - \widehat{\phi}(z)| \geq c|z|^\alpha.$$

Proof. Since Assumption 3 is satisfied, there exists $c' \neq 0$ and $\delta > 0$ such that

$$1 - \widehat{\phi}(z) \underset{z \to 0}{\sim} c' z^\alpha$$

and

$$\forall |z| \leq \delta,\ |1 - \widehat{\phi}(z)| \geq |c'||z|^\alpha/2.$$

On $[\delta, 1/2]$, $z \mapsto |1 - \widehat{\phi}(z)|/|z|^\alpha$ is continuous and therefore has a minimum attained in some z_0: $|1 - \widehat{\phi}(z_0)|/|z_0|^\alpha = M$.

M is strictly positive since $\mathrm{Re}(\widehat{\phi}(z_0)) = \mathbb{E}[\cos(2\pi z_0 X^1)]$, which is strictly smaller than one because, using Assumption 3, $z_0 X^1$ does not almost surely belong to $\mathbb{N}$.

Taking $c = \min(|c'|/2, M)$ ends the proof. □

Lemma 3. *For $\alpha > 1/2$, there exists $c > 0$ such that for any n, k, k',*

$$\mathbb{E}[y_k^n y_{k'}^n] \leq c(1-a_n)^{1/\alpha-2}.$$

Proof. As before, we have

$$\mathbb{E}[(y_k^n)^2] \leq 2\int_0^{1/2} \frac{1}{|1-a_n\widehat{\phi}(z)|^2}dz.$$

Therefore, for n large enough, $a_n > 1/2$ and

$$|1-a_n\widehat{\phi}(z)|^2 = |1-a_n+a_n(1-\widehat{\phi}(z))|^2 \geq |1-a_n|^2 + \frac{|1-\widehat{\phi}(z)|^2}{4}.$$

Using Lemma 2, this implies that

$$\mathbb{E}[(y_k^n)^2] \leq \int_0^{1/2} \frac{c}{|1-a_n|^2+|z|^{2\alpha}}dz.$$

Therefore,

$$\mathbb{E}[(y_k^n)^2] \leq \int_0^{(1-a_n)^{1/\alpha}} \frac{c}{|1-a_n|^2}dz + \int_{(1-a_n)^{1/\alpha}}^{1/2} \frac{c}{|z|^{2\alpha}}dz,$$

which ends the proof. □

Lemma 4. *For $\alpha < 1/2$, there exists $c > 0$ such that for any n, k, k',*

$$\mathbb{E}[y_k^n y_{k'}^n] \leq \begin{cases} c(1-a_n)^{2\alpha-1} & \text{if } k = k' \\ c|k-k'|^{2\alpha-1} & \text{if } k \neq k'. \end{cases}$$

Proof. The fact that

$$\mathbb{E}[(y_k^n)^2] \leq c(1-a_n)^{2\alpha-1}$$

is a direct consequence of the proof of Lemma 3 together with the inequality $1-2\alpha \geq 2-1/\alpha$ for $\alpha < 1/2$.

To prove the other inequality, remark that using the moving average representation of the process, for $k \geq k'$, we get

$$\mathbb{E}[y_k^n y_{k'}^n] = \sum_{j=0}^{k'} \psi_{k-j}^n \psi_{k'-j}^n \leq \sum_{j=0}^{+\infty} \psi_j^n \psi_{j+k-k'}^n.$$

Therefore, using Parseval's theorem together with Lemma 2 and the fact that the ψ_j are non-negative, we obtain

$$\mathbb{E}[y_k^n y_{k'}^n] \leq c\mathrm{Re}\Big(\int_0^{1/2} \frac{e^{2\pi i|k-k'|z}}{|1-a_n\widehat{\phi}(z)|^2}dz\Big) \leq c\mathrm{Re}\Big(\int_0^{1/2} \frac{e^{2\pi i|k-k'|z}}{|z|^{2\alpha}}dz\Big).$$

Thus, using Abel's theorem, we finally obtain

$$\mathbb{E}[y_k^n y_{k'}^n] \leq \frac{c}{|k-k'|^{1-2\alpha}}.$$

□

Lemma 5. *For $\alpha = 1/2$, for any $\varepsilon \in (0,1)$, there exists $c > 0$ such that for any n, k, k',*

$$\mathbb{E}[y_k^n y_{k'}^n] \le \begin{cases} c(1-a_n)^{-\varepsilon} & \text{if } k = k' \\ c + c\big((1-a_n)^2|k-k'|\big)^{-\varepsilon} & \text{if } k \neq k'. \end{cases}$$

Proof. Using the same proof as for Lemma 3, we get

$$\mathbb{E}[(y_k^n)^2] \le c|\log(1-a_n)|.$$

The first inequality follows. We obtain the second inequality using that for $k > k'$,

$$\begin{aligned} \mathbb{E}[y_k^n y_{k'}^n] &\le c\mathrm{Re}\Big(\int_0^{1/2} \frac{e^{2\pi i|k-k'|z}}{(1-a_n)^2 + |z|}\,dz\Big) \\ &\le c\int_0^{(1-a_n)^2} \frac{1}{(1-a_n)^2}\,dz + c\mathrm{Re}\Big(\int_{(1-a_n)^2}^{1/2} \frac{e^{2\pi i|k-k'|z}}{|z|}\,dz\Big) \\ &\le c + c|\log\big((1-a_n)^2|k-k'|\big)|. \end{aligned}$$

□

Let us now prove Proposition 4.

In the case where $\alpha > 1/2$, the proof that

$$\mathbb{E}[|Z_t^n - Z_s^n|^2] \le c|t-s|^2$$

is almost the same as in the light tail case replacing the use of Lemma 1 by that of Lemma 3.

The case $\alpha < 1/2$ is slightly more complicated. We now show that

$$\mathbb{E}[|Z_t^n - Z_s^n|^2] \le c|t-s|^{1+\eta},$$

for some $\eta > 0$. As before, we begin with the case where $\lfloor nt \rfloor = \lfloor ns \rfloor$. Using Lemma 4 together with Assumption 4 and the fact that $1 - \alpha(2\alpha+1) \ge 0$, we get

$$\begin{aligned} \mathbb{E}[|Z_t^n - Z_s^n|^2] &= \frac{(1-a_n)^2}{n} n^2|t-s|^2 \mathbb{E}[(y_{\lfloor nt \rfloor + 1}^n)^2] \\ &\le c\frac{(1-a_n)^2}{n} n^2|t-s|^{1+\alpha(2\alpha+1)} \frac{1}{n^{1-\alpha(2\alpha+1)}} (1-a_n)^{2\alpha-1} \\ &\le c|t-s|^{1+\alpha(2\alpha+1)}. \end{aligned}$$

Using the same arguments as for the light tail case, we get a similar bound for $\lfloor nt \rfloor = \lfloor ns \rfloor + 1$.

We now treat the case where $t = k^t/n$ and $s = k^s/n$, where k^t and k^s are integers so that $k^t > k^s$. Using again Lemma 4, we have

$$\mathbb{E}[|Z_t^n - Z_s^n|^2] = \frac{(1-a_n)^2}{n} \sum_{k_1=k^s+1}^{k^t} \sum_{k_2=k^s+1}^{k^t} \mathbb{E}[y_{k_1}^n y_{k_2}^n].$$

This is smaller than (with obvious notation)

$$\frac{(1-a_n)^2}{n}\Big(\frac{c(k^t-k^s)}{(1-a_n)^{1-2\alpha}}+c\sum_{\Delta k=1}^{k^t-k^s-1}\frac{1}{(\Delta k)^{1-2\alpha}}\#\{(k_1,k_2)\in[k^s+1,k^t];|k_1-k_2|=\Delta k\}\Big).$$

Therefore, using Assumption 4, we get

$$\begin{aligned}\mathbb{E}[|Z_t^n-Z_s^n|^2] &\leq \frac{(1-a_n)^2}{n}\Big(\frac{c(k^t-k^s)}{(1-a_n)^{1-2\alpha}}+c\sum_{\Delta k=1}^{k^t-k^s-1}\frac{1}{(\Delta k)^{1-2\alpha}}(k^t-k^s)\Big)\\ &\leq \frac{(1-a_n)^2}{n}\Big(\frac{c(k^t-k^s)}{(1-a_n)^{1-2\alpha}}+c(k^t-k^s)^{1+2\alpha}\Big)\\ &\leq c((t-s)^{1+\alpha+2\alpha^2}+(t-s)^{1+2\alpha}).\end{aligned}$$

The result for any s and t in $[0,1]$ is obtained as in the light tail case.

Finally, the proof for $\alpha = 1/2$ is obtained the same way, using Lemma 5 instead of Lemma 4.

Part III

Multiscale Hawkes estimation and application to finance

CHAPTER V

Estimation of slowly decreasing Hawkes kernels: Application to high frequency order book dynamics

Abstract

We present a modified version of the non parametric Hawkes kernel estimation procedure studied in [BM14b] that is adapted to slowly decreasing kernels. We show on numerical simulations involving a reasonable number of events that this method allows us to estimate faithfully a power-law decreasing kernel over at least 6 decades. We then propose a 8-dimensional Hawkes model for all events associated with the first level of some asset order book. Applying our estimation procedure to this model, allows us to uncover the main properties of the coupled dynamics of trade, limit and cancel orders in relationship with the mid-price variations.

Keywords: Hawkes processes, kernel estimations, power law kernels, high frequency, order book dynamics, trades, limit orders, cancel orders, market impact.

1 Introduction

The understanding of the price formation mechanism remains one of the most challenging problem in quantitative finance. In most modern financial exchanges, assets are traded via a continuous double auction: agents can choose to buy or sell a stock at certain prices posting *limit orders* in the order book or to execute the available limit orders using *market orders*. As long as their their limit orders are not executed, agents can also use *cancel orders* to take them off[1]. The intrinsic complexity of the formation of order book (it is the result of market, limit and cancel orders of a large number of anonymous traders) implies that in spite of its great practical and theoretical importance, there are relatively few full order book models (see [Roş09b, SFGK03] for "zero-intelligence", i.e. purely random model of book dynamics and

[1] See e.g., [CDL13, SFGK03] for more details on the functioning of continuous double auctions.

[CDL13, GDKB13, HLR13] for Markovian models of the order book).

A class of simpler models does not take into account the amount of liquidity and focuses exclusively on asset prices which are described as the result of the *impact* of various type of orders: Former models considered exclusively the impact of market orders [BFL09, BGPW04] while some recent models account for the influence of all types of order book events [CDL13, EBK12]. In a recent work, some of us [BM14a] introduced a price impact model using the framework of Hawkes processes. This model shares features with former impact models but allows one to describe the joint dynamics of both trades and price moves. In particular, unlike previous impact models, this Hawkes model takes into account the influence of price changes on themselves and on trades.

Hawkes processes were defined in [Haw71a] as point processes whose intensity function is a linear regression on the past of the process

$$\lambda_t^N = \mu + \int_{(-\infty,t)} \phi(t-s)\, dN_s.$$

The self and mutually exciting nature of Hawkes processes makes them naturally adapted to the modeling of multivariate counting processes. They have for example been successfully applied in many domains such as the study of earthquakes [Oga99], neurobiology [RBRGTM14] or sociology [MSB+11]. In Finance, they have been recently used to model the arrival of trades or limit orders [Hew06, Bow07, Lar07], price variations at microstructure level [BDHM13, BM14a, HBB13], financial contagion [ASCDL10] or credit risk [EGG10].

In this work, we will propose a generalization of the impact model of [BM14a], in order to account for the interaction between price variations and any event (market, limit or cancel order) at the level I of the order-book. This model turns out to be an 8-dimensional Hawkes process that will allow us to have a direct measure of the endogeneity and the causality between these events, see [FS12] or [HBB13] for related discussions about the use of Hawkes processes to measure the market endogeneity.

As far as estimation problems are concerned, the few former non parametric estimates of Hawkes kernels in the context of high frequency finance lead to power law decreasing kernels whose exponent is slightly higher than one [BDM12, BM14a, HBB13]. As we will see, this implies that one needs to estimate the kernels on a wide range of time scales (typically from 10 microseconds to 100 seconds) to capture the dynamics of the order book. For that purpose we propose an improvement of the numerical scheme introduced in [BM14a, BM14b] in order to handle slowly decreasing kernels. It will be illustrated on some numerical examples, that this method allows us to estimate the kernel on a large range of time scales which cannot be done within the Gaussian scheme of [BM14b].

The paper is organized as follows. After recalling the main definitions and properties of multidimensional Hawkes processes, we present the main non-parametric estimation principles

in Section 2. Section 3 explains how these principles can be adapted to the non-parametric estimation of slowly decaying kernels. The so-obtained algorithm is presented and tested on numerical simulations at the end of this Section. The level I book model is then introduced in Section 4. It is calibrated on high frequency data corresponding to DAX and BUND front future contracts during one year (from June 2013 to June 2014). We then comment our results as far as endogeneity and mutual influence of all types of events are concerned. Discussion in relationship with former works and prospects for future research are provided in the concluding Section 5.

2 Multidimensional Hawkes processes and estimation principles

In this section we recall the definition of a D-dimensional Hawkes processes and briefly review its main properties. For more details we refer the reader to [BM14b].

2.1 Basic definitions

Definition 1. *A D-dimensional point process $N_t = (N_t^i)_{1\leq i\leq D}$ is a Hawkes process if for every i, the i^{th} component of its intensity function λ_t is a linear regression of the past jumps of N_t, i.e.,*

$$\lambda_t^i = \mu^i + \sum_{j=1}^{D} \int_{(-\infty,t)} \phi^{ij}(t-s) dN_s^j, \quad \forall i \in [1, D], \tag{1}$$

where $\mu = \{\mu^i\}_{1\leq i\leq D}$ is the so-called exogenous intensity *(with $\mu^i \geq 0$), and $\phi = \{\phi^{ij}\}_{1\leq i,j\leq D}$ the so-called* Hawkes kernel *matrix, where each $\phi^{ij}(t)$ is a real positive and causal function[2]. Using matrix convolution notation, the system of Equation* (1) *simply rewrites*

$$\lambda_t = \mu + \phi * dN_s. \tag{2}$$

Let us recall the following well known stability condition of the Hawkes processes, see for example [Haw71b].

Proposition 1. *If the matrix of the norms L^1 of the kernels defined as*

$$||\phi|| = (||\phi^{ij}||_1)_{1\leq i,j\leq D}$$

has a spectral radius strictly smaller than one then N admits a version with stationary increments.

Let us remark that Brémaud and Massoulié [BM96] have shown that this stability criterion remains valid when, in Equation (1), λ_t is a nonlinear positive Lipschitz function of $\{dN_s\}_{s<t}$ and where the elements of the kernel ϕ are not restricted to be positive. A particularly interesting generalization of Hawkes processes considered in [HRBR12] is

$$\lambda_t = \left(\mu + \phi * dN_s\right)^+ \tag{3}$$

[2] By definition, a function is causal if its support is included in $\mathbb{R}^+$.

where $(x)^+ = x$ if $x > 0$ and $(x)_+ = 0$ otherwise. This extension allows one to account for inhibitory effects when $\phi^{ij}(t) < 0$.

We will now assume that the condition of Proposition 1 is satisfied and consider a stationary Hawkes process. The next simple result describes the behavior of the first order statistics of N, see for example [BM14b].

Proposition 2. *The average of the intensity vector of N:*

$$\Lambda = (\mathbb{E}[\lambda_t^1], ..., \mathbb{E}[\lambda_t^D])^T$$

satisfies

$$\Lambda = (\mathbb{I} - ||\phi||)^{-1}\mu. \tag{4}$$

2.2 Second-order properties and Wiener-Hopf system

The second order statistics of Hawkes processes are naturally characterized by their infinitesimal covariance functions.

Definition 2. *The infinitesimal covariance ν of N is the $D \times D$ matrix whose elements are*

$$\nu^{ij}(t - t')dtdt' = \mathbb{E}[dN_t^i dN_{t'}^j] - \Lambda^i \Lambda^j dtdt',$$

where Σ is the $D \times D$ matrix whose diagonal is the average intensities Λ^i.

The following result is shown in [BDM12].

Proposition 3. *ν can be expressed as a functional of ϕ:*

$$\nu = (\delta\mathbb{I} + \psi) * \Sigma(\delta\mathbb{I} + \psi^T) \tag{5}$$

where $\delta(t)$ stands for the Dirac distribution, Σ is the $D \times D$ matrix whose diagonal is the average intensities Λ^i, and where the matrix $\psi(t)$ is is defined as:

$$\psi(t) = \sum_{k \geq 1} \phi^{(*k)}(t), \tag{6}$$

in which we used the notation $\phi^{(*k)}(t) = \phi * \phi * \ldots * \phi$ (where ϕ is repeated k times).

Let us mention that there is an alternative "population representation" of Hawkes processes (see e.g., [HO74]) according to which they are built as the times of arrival and birth of the following population process:

- There are individuals of type 1, ..., D.
- Migrants of type i arrive at a Poisson rate of μ^i.
- Every individual can have children of (a priori) all types and the children of type i of an individual of type j who was born or migrated in t are distributed as an inhomogeneous Poisson process of intensity $\phi^{ij}(\cdot - t)$.

In this model, $||\phi^{ij}||$ appears as the average number of children of type i of an individual of type j. We thus use $||\phi^{ij}||$ as a measure of "causality" between j and i. Similarly, $\Lambda^j/\Lambda^i||\phi^{ij}||$ appears as the proportion of individuals of type i whose parent is an individual of type j.

In this construction, the elements ψ^{ij} of the matrix ψ also have a natural interpretation. The descendants of type i of each exogenous event of type j occurring at time t are a point process of average intensity $\psi^{ij}(\cdot - t)$. Therefore, $||\phi^{ij}||$ appears as the average number of individuals of type i which descend from an individual of type j. Similarly, $\Lambda^j/\Lambda^i||\psi^{ij}||$ appears as the proportion of individuals of type i whose "ancestor" is an individual of type j.

As explained in [BM14b], since the jumps are always of size one, the second order statistics of the Hawkes process can also be summed up by its conditional laws.

Definition 3. *We define the conditional laws denoted* $(g^{ij})_{1\leq i,j\leq D}$ *as the non singular part of the measure* $\mathbb{E}[dN_t^i|dN_0^j = 1]$*:*

$$g^{ij}(t)dt = \mathbb{E}[dN_t^i|dN_0^j = 1] - \mathbb{1}_{i=j}\delta(t) - \Lambda^i dt.$$

They are linked with the infinitesimal covariance by the equation

$$g(t) = \nu(t)\Sigma^{-1} - \delta(t)\mathbb{I}. \tag{7}$$

The following proposition, see [BM14b], links the conditional laws and the kernels of Hawkes processes through a Wiener-Hopf system of equations.

Proposition 4. *Given the conditional laws of a Hawkes process* g*, its kernel* ϕ *is the only*[3] *solution of the Wiener-Hopf system:*

$$g(t) = \phi(t) + g * \phi(t), \quad \forall\, t > 0. \tag{8}$$

Remark 1. *The linear structure of Hawkes processes implies that they have properties similar to that of autoregressive processes. In particular, the previous proposition can be seen as the counter part of the Yule-Walker Equation.*

2.3 Existence and uniqueness of solutions of the Wiener-Hopf system

The Wiener-Hopf system (8) is at the core of our estimation procedure. In [BM14b], it is shown that if g is the conditional law of a Hawkes process as defined in Definition 3 then the system (8) admits the unique causal solution which is the Hawkes kernel ϕ.

Let us now leave the Hawkes model and assume that Σ, $\nu \in L^1$ and g are respectively the diagonal average matrix, the infinitesimal covariance function and the conditional law of any stationary D-dimensional point process. Using the Wiener-Khintchine Theorem, we get that

[3] The uniqueness of the solution of Equation (8), considered as an equation in ϕ, is shown in [BM14b].

the matrix $\widehat{\nu}(z)$ is positive definite for any $z \in \mathbb{R}$. We can thus apply Theorem 8.2 of [GK58] to decompose ν as

$$\nu(t) = Y(t) * Y^T(-t)$$

where Y is causal. We then set

$$\psi = (Y - \delta \mathbb{I})\sqrt{\Sigma^{-1}}$$

and ϕ such that

$$\hat{\phi} = \hat{\psi}(\mathbb{I} + \hat{\psi})^{-1}$$

to get that ϕ satisfies Equation (5) and thus Equation (8) (these two equations are equivalent for $g(t) = \nu(t)\Sigma^{-1} - \delta(t)\mathbb{I}$).

We have thus shown the following result.

Theorem 1. *If $g \in L^1$ is the conditional law of a point process then Equation* (8) *has one and only one solution (in L^1).*

The previous theorem shows that it makes sense to try to find the solution of the Wiener-Hopf system for empirical data. Of course, the so-obtained solution is no longer necessarily positive. We will comment on the interpretation of a negative ϕ later.

2.4 Model independent origin of the Wiener-Hopf equation

We now show that, from a model independent point of view, solving the Wiener-Hopf system amounts to finding the kernel matrix ϕ (and the intensity vector μ) that provide the best linear predictor (in the sense of mean-square error) of the intensity. This remark leads the authors of [HRBR12, RBRGTM14] to consider an alternative estimation algorithm relying on the minimization of a contrast function, a proxy for the mean square error.

Let us consider a 1-dimensional (for the sake of simplicity) stationary point process (not necessarily a Hawkes process) which admits a mean intensity Λ and a conditional law continuous and in L^2

$$\begin{aligned} g(t-t')dtdt' = g(t'-t)dtdt' &= \mathbb{E}[dN_t > 0 | dN_{t'} > 0]dt' - \Lambda\delta_t(dt')dt - \Lambda dtdt' \\ &= \mathbb{E}[dN_t dN_{t'}]/\Lambda - \Lambda\delta_t(dt')dt - \Lambda dtdt'. \end{aligned}$$

We are looking for the optimal kernel $\tilde{\phi}$ in L^2 and intensity $\tilde{\mu}$ such that the linearly estimated intensity defined as

$$\lambda_t^{\phi} = \mu + \int_{-\infty}^{t} \phi(t-s)dN_s$$

minimises the mean square error

$$\varepsilon(\phi) = \mathbb{E}[(\lambda_t - \lambda_t^{\phi})^2].$$

The first thing to notice is that

$$\mathbb{E}[\lambda_t - \lambda_t^{\phi}] = \Lambda - \mu - \int \tilde{\phi}\Lambda$$

so that the optimal μ is related to the optimal ϕ by

$$\tilde{\mu} = \Lambda - \int \tilde{\phi}\Lambda.$$

Let us now denote $\bar{\lambda} = \lambda - \Lambda$ and $\bar{N}_t = N_t - \Lambda t$. $\varepsilon(\phi)$ can be rewritten:

$$\begin{aligned}
\varepsilon(\phi) &= \mathbb{E}[(\bar{\lambda}_t - \bar{\lambda}_t^{\phi})^2] \\
&= \mathbb{E}[(\bar{\lambda}_t)^2] - 2\int_{-\infty}^{t} \phi(t-s)\mathbb{E}[\bar{\lambda}_t d\bar{N}_s] + \int_{-\infty}^{t} \phi(t-s)\int_{-\infty}^{t} \phi(t-s')\mathbb{E}[d\bar{N}_{s'} d\bar{N}_s] \\
&= \mathbb{E}[(\bar{\lambda}_t)^2] - 2\Lambda\int_{-\infty}^{t} \phi(t-s)g(t-s)ds \\
&+ \Lambda\int_{-\infty}^{t} \phi(t-s)\int_{-\infty}^{t} \phi(t-s')(g((t-s)-\phi(t-s'))ds' + \delta_s(ds'))ds \\
&= \mathbb{E}[(\bar{\lambda}_t)^2] - 2\Lambda\int_{0}^{+\infty} \phi(x)g(x)dx + \Lambda\int_{0}^{+\infty} \phi(x)(g*\phi(x) + \phi(x))dx.
\end{aligned}$$

Finding this $\tilde{\phi}$ is thus equivalent to minimizing

$$F(\phi) = \int_{0}^{+\infty} \phi(x)[g*\phi(x) + \phi(x) - 2g(x)]dx$$

If $\tilde{\phi} \in L^2$ is the optimal kernel then for every test function η in L^2 and ε, let us differentiate F in the direction η:

$$F(\tilde{\phi} + \varepsilon\eta) = F(\tilde{\phi}) + \varepsilon\int_{0}^{+\infty} [\eta(x)(g*\tilde{\phi}(x) + \tilde{\phi}(x) - 2g(x)) + \tilde{\phi}(x)(g*\eta(x) + \eta(x))]dx + o(\varepsilon).$$

Therefore, since for all $\frac{\partial F(\phi^{\eta,\varepsilon})}{\partial\varepsilon}(\varepsilon = 0) = 0$ and using the symmetry of g

$$\int_{0}^{+\infty} \tilde{\phi}(x)(g*\eta(x))dx = \int_{0}^{+\infty} \eta(x)(g*\tilde{\phi}(x))dx$$

and thus

$$0 = \frac{\partial F(\phi^{\eta,\varepsilon})}{\partial\varepsilon}(\varepsilon = 0) = \int_{0}^{+\infty} \eta(x)(2g*\tilde{\phi}(x) + 2\tilde{\phi}(x) - 2g(x))dx.$$

Since the above equation is true for any test function, $\tilde{\phi}$ must satisfy the Wiener-Hopf equation:

$$g(y) = \tilde{\phi}(y) + g*\tilde{\phi}(y).$$

Adding the fact that $\phi \mapsto \varepsilon(\phi)$ is convex, continuous and coercive in L^2 (which is reflexive), we have shown the following result.

Theorem 2. *If g is continuous and in L^2, $\phi \mapsto \varepsilon(\phi)$ has a unique minimum in L^2 and this minimum satisfies the Wiener-Hopf equation.*

2.5 Estimation principles

As explained in [BM14b], the procedure for solving numerically the Wiener-Hopf system (8) follows Nyström method [Nys30]. More precisely, given a realization if the process N on an interval $[0, T]$, the procedure for non parametric estimation is the following

1. Compute an estimation $\tilde{g}$ of the matrix function g using a "fine enough" time grid (values of g outside of this time grid will be computed using an interpolation scheme).
2. Use a quadrature method to discretize the Wiener-Hopf system (8) on an interval $[T_{min}, T_{max}]$. If we use the quadrature points $\{t_k\}_{1 \leq k \leq K}$ along with the quadrature weights $\{w_k\}_{1 \leq k \leq K}$, one gets, the system

$$\tilde{g}^{ij}(t_n) = \tilde{\phi}^{ij}(t_n) + \sum_{l=1}^{D} \sum_{k=1}^{K} w_k \tilde{g}^{il}(t_n - t_k) \tilde{\phi}^{lj}(t_k), \quad \forall n \in [0, K], \tag{9}$$

   for all $i, j \in [1, D]$.
3. Inverse this so-obtained KD^2 linear system. This leads to the estimation of the matrix kernel at the quadrature points $\{\tilde{\phi}^{ij}(t_n)\}_{i,k \in [1,D], n \in [1,K]}$ as well as an estimation of $||\phi||_1$ (using quadrature).
4. Estimate empirically the average intensity Λ (just counting the number of points on each component of the process) and estimate μ using (4).

A thorough study of this algorithm has been done in [BM14b] (in the framework on a Gaussian quadrature) and showed that it performs a fast and efficient non parametric estimation of a Hawkes process. It proved particularly competitive in the case the realization process has a large number of jumps and the matrix kernel is not well localized in time. Let us point out that [BM14b] showed that this algorithm leads to precise result even if some elements of the kernel matrix have moderate negative values.

However, as we will show in the next section, when some elements of the kernel matrix are power-law, the quality of the estimation can be improved significantly. In this case, the use of Gaussian quadrature is not adapted and leads to inaccurate estimations because the "mass" of the power-law kernel exponents are spread over a large range of scales. We shall study this phenomenon in the framework of high frequency financial data.

3 Non parametric estimation of slowly decaying kernels

One of the main issue when dealing with high frequency financial data is that the range of scales involved in the dynamics is rather large. This translates into kernels with large support, typically from $T_{min} = 100$ microseconds to $T_{max} = 100$ seconds, slowly decaying (power-law like), varying very quickly at small lags (around T_{min}) and much slower at larger lags (around T_{max}). Let us follow step by step the algorithm of Section 2.5 and explain how to adapt it to this framework.

3.1 Using an adapted sample-grid for g

Step 1. of the algorithm (estimation of g) is clearly the first sensitive step. Since the kernels are varying very quickly at small lags, more care should be taken for estimating the behavior of g in the neighbourhood of 0 than anywhere else. In Appendix V.A, we describe precisely the algorithm we shall use in this paper for the estimation of g. It basically consists in two step procedure: (i) empirical computation of g on a time-grid which is more dense around 0 than anywhere else, (ii) linear interpolation for computing the values of g at any time (in all our numerical experiments, we have checked than higher-order approximation does not bring relevant precision to the overall algorithm).

3.2 Towards an adapted quadrature scheme

The choice of the quadrature (step 2. in the algorithm) is the next (and last) sensitive step. The Gaussian quadrature used in [BM14b] for solving the Wiener-Hopf system is clearly not adapted to this situation. Indeed, the Gaussian quadrature points are roughly uniform over $[0, T]$ whereas the kernels are varying very quickly around 0 and slowly anywhere else. In order to get precise estimation of kernels on the whole range $[T_{min}, T_{max}]$, the Gaussian quadrature is constrained by their behavior at small lags ($t \simeq T_{min}$) and would imply using an number of quadrature points of the order of $K = T_{max}/T_{min} = 10^6$ which, even for $D = 1$, is far to many to be able to solve numerically the system (9).

One way to circumvent this difficulty could consist in adding points close to zero by using the change of variable proposed in [Slo81]. Let $\{t_k, w_k\}_{1 \le k \le K}$ correspond to the Gaussian quadrature scheme on $[\log(T_{min}), \log(T_{max})]$, we then set

$$(t'_k, w'_k) = (e^{t_k}, e^{t_k} w_k). \tag{10}$$

We then replace the system (9) by the new system:

$$\tilde{g}^{ij}(t'_n) = \tilde{\phi}^{ij}(t'_n) + \sum_{l=1}^{D} \sum_{k=1}^{K} w'_k \tilde{g}^{il}(t'_n - t'_k) \tilde{\phi}^{lj}(t'_k). \tag{11}$$

In order to test this new estimation scheme, we simulate a one-dimensional Hawkes process with parameters

$$\phi(t) = \frac{0.06}{(0.005 + t)^{1.3}} \quad \text{and} \quad \mu = 0.05. \tag{12}$$

The simulation is performed on a time-period long enough (10^7 seconds) so that the Monte Carlo error is small enough compared to the "quadrature error".

Remark 2. *This kernel checks the same kind of "multiscale behavior" as kernels involved in high frequency financial dynamics. Indeed, 90% of its L^1 norm is spread over 5 decades: from 10^{-3} to 10^2 seconds, in the sense that $\int_0^{+\infty} \phi(t)dt = 0.98$, $\int_0^{10^{-3}} \phi(t)dt = 0.05$ and $\int_{10^2}^{+\infty} \phi(t)dt = 0.05$.*

We applied the algorithm described in Section 2.5, in which we use the procedure described in Appendix V.A to estimate g using $h_{min} = 1ms$, $h_{max} = 1000s$ and $h_\delta = 0.05$, and in which

we replaced (9) by (11). We chose $K = 200$, $T_{min} = 1ms$ and $T_{max} = 2000s$. Figure V.1, displays the so-obtained results. On the left, the log-log plot of both the estimated kernel and the theoretical kernel (as well as the conditional law g) are displayed. On the right, the integral on $[0, t]$ of both theoretical and estimated kernels are displayed.

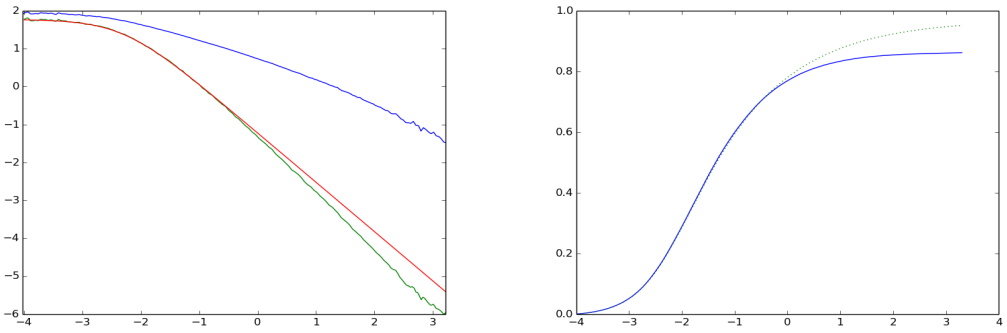


Figure V.1 – Left: Estimation of the conditional law g and kernel ϕ in the case of a 1-dimensional Hawkes of size $T = 10^7$ seconds, with parameters (12). The estimation of g is performed using the procedure described in Appendix V.A using $h_{min} = 1ms$, $h_{max} = 1000s$ and $h_\delta = 0.05$, and kernel estimation is performed using algorithm of Section 9 with Gaussian quadrature followed by the change of variable (10) ($K = 200$, $T_{min} = 1ms$, $T_{max} = 2000s$ and thus $\delta = 0.077$). The three curves correspond to $\log_{10} - \log_{10}$ empirical conditional law (blue), theoretical kernel (red) and estimated kernel (green) for 200 quadrature points.
Right: Cumulated theoretical kernel $\int_0^t \phi(s)ds$ (green dots) and cumulated estimated kernel $\int_0^t \tilde{\phi}(s)ds$ (blue) as a function of $\log_{10}(t)$.

We observe that, for large times ($t \geq 1s$), the estimated kernel does not approximate well the theoretical kernel. Moreover, this procedure significantly underestimates the cumulated kernel. Indeed, we get $||\tilde{\phi}||_1 = 0.85$ instead of $||\phi||_1 = 0.98$. This error on the estimation of the norms will be very important in our study of the causality between market events. Indeed the norms of the kernels $||\phi^{ij}||_1$ will have a direct interpretation as the average number of events of type i caused by an event of type j.

Let us try to give an intuitive explanation of the bad performance of this method: Recall that the quadrature approximates

$$\int_0^{+\infty} g(t'_n - s)\phi(s)ds \simeq \sum_{k=1}^{K} w'_k g(t'_n - t'_k)\phi(t'_k), \quad \forall n \in [1, K]$$

Thanks to the change of variable (10) there are more quadrature points around 0 than around any other time. That ensures that the quadrature approximation captures the fast variation of $\phi(s)$ around $s = 0$. However, g is also varying quickly around 0, the estimation error illustrated in Figure V.1 comes from the bad approximation of $g(t'_n - s)$ around $s = t'_n$ (i.e., $k = n$ in the

previous quadrature formula).

In the next section, we propose a new quadrature scheme that will capture all behavior of the conditional law g next to t_n for any n.

3.3 The adapted quadrature scheme for slow-decaying kernels

Let us consider the time grid which is uniform from 0 to T_{min} and log-uniform between T_{min} and T_{max} using the same grid-size $\delta < 1$

$$\{t_k\}_{1\leq k\leq K} = [0, \delta T_{min}, 2\delta T_{min}, ..., T_{min}, T_{min}e^{\delta}, T_{min}e^{2\delta}, ..., T_{max}]. \tag{13}$$

The adapted quadrature scheme consists in considering that the kernels are piecewise affine on $[t_k, t_{k+1}]$:

$$\phi(t) = \phi(t_k) + \frac{t - t_k}{t_{k+1} - t_k}(\phi(t_{k+1}) - \phi(t_k)).$$

Under this linear hypothesis, Equation (8), in the $D = 1$ dimensional case, becomes (the generalization to dimension $D > 1$ is straightforward):

$$\begin{aligned} g(t_n) &= \phi(t_n) + \sum_{k=0}^{N-1} \phi(t_k) \int_{t_k}^{t_{k+1}} g(t_n - s)ds \\ &+ \sum_{k=0}^{N-1} (\phi(t_{k+1}) - \phi(t_k)) \int_{t_k}^{t_{k+1}} \frac{s - t_k}{t_{k+1} - t_k} g(t_n - s)ds \\ &= \phi(t_n) + \sum_{k=0}^{N-1} \phi(t_k) \int_{t_n - t_{k+1}}^{t_n - t_k} g(u)du \\ &+ \sum_{k=0}^{N-1} \frac{\phi(t_{k+1}) - \phi(t_k)}{t_{k+1} - t_k} \int_{t_n - t_{k+1}}^{t_n - t_k} (t_n - t_k - u)g(u)du \end{aligned}$$

Thus, using the adapted quadrature, the system (9) becomes (in the $D = 1$ case)

$$\begin{aligned} \tilde{g}(t_n) &= \tilde{\phi}(t_n) + \sum_{k=0}^{N-1} \tilde{\phi}(t_k) \int_{t_n - t_{k+1}}^{t_n - t_k} \tilde{g}(u)du \\ &+ \sum_{k=0}^{N-1} \frac{(\tilde{\phi}(t_{k+1}) - \tilde{\phi}(t_k))(t_n - t_k)}{t_{k+1} - t_k} \int_{t_n - t_{k+1}}^{t_n - t_k} \tilde{g}(u)du \\ &- \sum_{k=0}^{N-1} \frac{\tilde{\phi}(t_{k+1}) - \tilde{\phi}(t_k)}{t_{k+1} - t_k} \int_{t_n - t_{k+1}}^{t_n - t_k} u\tilde{g}(u)du. \end{aligned} \tag{14}$$

The empirical values of the integrals $\int_0^x \tilde{g}(u)du$ and $\int_0^x u\tilde{g}(u)du$ can easily be computed beforehand at all abscissa of the form $x = t_n - t_k$ using the linear interpolation of the conditional laws described in Section 3.1 (see also Appendix V.A).

As before, this is a linear equation in the values of ϕ at the quadrature points. Solving it, one gets an estimation of ϕ at these points. A simple linear interpolation can be performed to get

estimation at other points.

Applying the same testing procedure as before, we get, see Figure V.2 that this scheme works perfectly with a reasonable number of quadrature points.

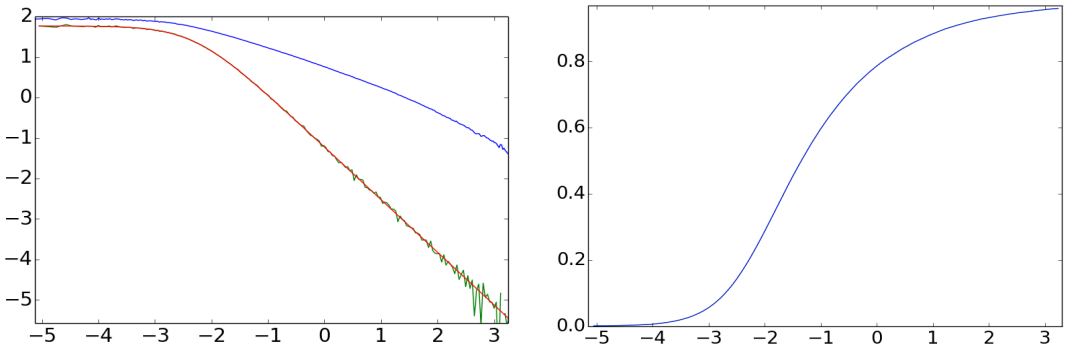


Figure V.2 – Left: Estimation of the conditional law g and kernel ϕ for the same Hawkes process realization as in Figure V.1. The kernel estimation is performed using the algorithm of Section 9 where the quadrature of step 3. has been replaced by the adapted quadrature given by (14) ($K = 200$, $T_{min} = 1ms$, $T_{max} = 2000s$ and thus $\delta = 0.077$). The three curves correspond to $\log_{10} - \log_{10}$ empirical conditional law (blue), theoretical kernel (red) and estimated kernel (green) for 200 quadrature points. The estimation and the theoretical kernels perfectly match. Right: Cumulated theoretical kernel $\int_0^t \phi(s)ds$ (green dots) and cumulated estimated kernel $\int_0^t \tilde{\phi}(s)ds$ (blue) as a function of $\log_{10}(t)$.

3.4 The adapted estimation procedure

Here is the step by step estimation procedure that will be used all along this paper:

1. Compute an estimation $\tilde{g}$ of the matrix function g using the procedure described in Section 3.1.

2. Use the adapted quadrature method to discretize the Wiener-Hopf system (8) on an interval $[T_{min}, T_{max}]$. The quadrature points are given by (13) and the scheme by (14) (this last equation corresponds to the one dimensional scheme, but generalization to any dimension is obvious).

3. Inverse this so-obtained KD^2 linear system. This leads to the estimation of the matrix kernel at the quadrature points $\{\tilde{\phi}^{ij}(t_n)\}_{i,k\in[1,D],n\in[1,K]}$ as well as an estimation of $||\phi||_1$ (using quadrature).

4. Estimate empirically the average intensity Λ and estimate μ using (4).

4 Hawkes model for Level I Order book events

4.1 Definition of the model

Multivariate Hawkes models provide a natural framework to account for the impact of past events of various type on the rate of arrival of future events. In [BM14a], this framework has been considered to extent classical impact price models (as in e.g. [BGPW04]) by accounting in real time for the "impact" of market orders on price changes but also for the auto-regressive dynamics of price changes and their retro-action on the rate of market order arrivals. Along the same line, one can extend the model of [BM14a] by accounting, within a multidimensional Hawkes model, for the cross and self influencing dynamics of all event types in the order book. For the sake of parsimony and simplicity, we consider only events occurring at the Level I of the order book, i.e., events that change the state of the order book at the best bid or best ask levels. More precisely we consider the following 8 dimensional counting process:

$$N_t = (P_t^{(a)}, P_t^{(b)}, T_t^{(a)}, T_t^{(b)}, L_t^{(a)}, L_t^{(b)}, C_t^{(a)}, C_t^{(b)})$$

where:

- $P^{(a)}$ (resp. $P^{(b)}$) counts the number of upward (resp. downward) mid-price moves.
- $T^{(a)}$ (resp. $T^{(b)}$) counts the number of market orders at the ask (resp. bid) that do not move the price.
- $L^{(a)}$ (resp. $L^{(b)}$) counts the number of limit orders at the ask (resp. bid) that do not move the price.
- $C^{(a)}$ (resp. $C^{(b)}$) counts the number of cancel orders at the ask (resp. bid) that do not move the price.

Remark 3. *Let us stress that mid-price moves can correspond to the occurrence of a market order or a cancel order that eats all the available liquidity at best bid or best ask or to a limit order placed in the spread between best bid and best ask. We choose to not distinguish these events in order to handle a process with relatively low dimension.*

The previous counting process can be considered as an 8-dimensional Hawkes process characterized by 8 exogenous intensities and 64 kernels. We denote by $\phi^{N^j \to N^i}$ the kernel ϕ^{ij} coding the influence of the j^{th} process on the i^{th} intensity. For example $\phi^{T^{(b)} \to P^{(a)}}$ corresponds to the influence of the trades at the bid that left the mid-price unchanged on the upward price moves. These intensities and kernels can be estimated from empirical book data using the procedure described in Section 3.4.

Notice that, in the following, we never impose "by hand" any (bid,downward)-(ask,upward) symmetry. However, it can be seen in Appendix V.B that, as expected, this symmetry is rather well satisfied (for example, $\phi^{T^{(b)} \to P^{(a)}} \simeq \phi^{T^{(a)} \to P^{(b)}}$) in our estimations. This indicates that, as suggested on simulations, while we perform our kernel estimations on processes with many

dimensions and on many time scales, our method and the amount of data that we use, allows us to retrieve significant results about the kernels. Indeed, an unstable method would hardly allows one to retrieve this symmetry.

4.2 Description of the database

The financial data used in this paper have been provided by the company QuantHouse EUROPE/ASIA (http://www.quanthouse.com). It consists in all level-I order book data[4] of BUND and DAX future contracts. For every day, we only keep the most liquid maturity and we use data over one year from June 2013 to June 2014. Each file lists all the changes in the first limit (best ask or best bid) of the order book at a micro second precision. We can thus easily precisely compute from this data the times of the different market events that will be of interest here (ask or bid market, limit or cancel orders at the first limit and upward and downward mid price moves).

At the time scales that we shall study (from 10^{-5} to 10^{2} seconds), the order book dynamics strongly depend on the tick size of the asset (or more precisely by the quotient between the average spread and the tick size, see [BFL09] or [DR12] for characterization and differences between small and large tick assets). In that respect, the DAX corresponds to a "small-tick" asset while the Bund is a "large tick" asset.

The number of events in our sample is summarized in Table V.1.

	$P^{(a)}$	$P^{(b)}$	$T^{(a)}$	$T^{(b)}$	$L^{(a)}$	$L^{(b)}$	$C^{(a)}$	$C^{(b)}$
xFDAX	9.36	9.35	2.66	2.67	19.7	19.6	23.3	23.1
xFGBL	1.72	1.72	3.40	3.48	29.8	29.8	26.6	26.5

Table V.1 – Total number of events (in millions).

We have checked using numerical simulations of multidimensional Hawkes processes with power-law kernels, that the size of the sample is sufficient to provide reliable estimations of the shapes of the kernels. We have also checked that our main results do not depend on intraday seasonal effects: if one selects a 1 hour intraday time slice, we obtain the same results (up to some statistical noise).

4.3 Conditional law estimations

To estimate the 64 conditional laws we proceed as explained in Section 3.1 (see also Appendix V.A) taking the parameters $h_\delta = 0.05$, $h_{min} = 0.1$ milliseconds and $h_{max} \simeq 1059$ seconds (see (17)) on the period from 1 June 2013 to 1 June 2014 (252 open days).

[4] That is the times of market orders and limit and cancel orders at the first limit and the state of the first bid-ask queues at these times.

Surprisingly most of the conditional laws between financial events have very similar properties. As an illustration, Figure V.3 displays, for the Bund, $g^{P^{(a)}\to P^{(b)}}$, $g^{T^{(a)}\to T^{(a)}}$, $g^{L^{(a)}\to L^{(a)}}$, $g^{C^{(a)}\to C^{(a)}}$, $g^{L^{(a)}\to L^{(b)}}$ and $g^{L^{(a)}\to C^{(a)}}$ in $\log_{10}-\log_{10}$.

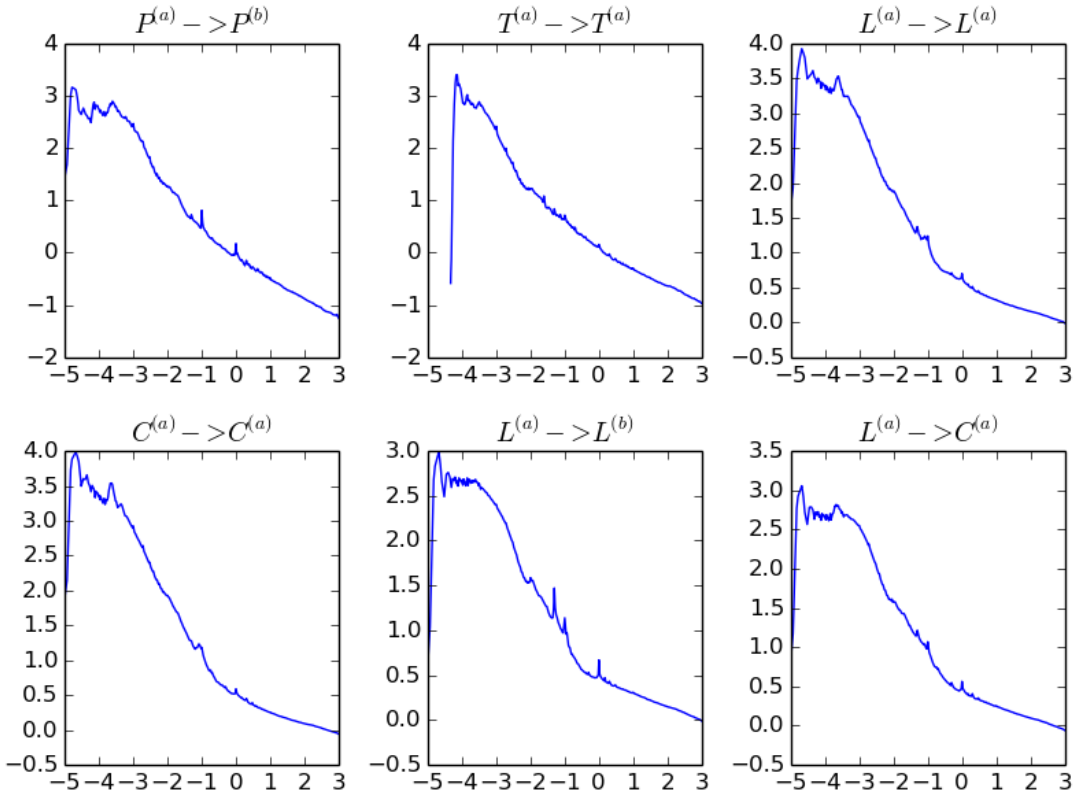


Figure V.3 – $\log_{10}-\log_{10}$ empirical conditional laws $g^{P^{(a)}\to P^{(b)}}$, $g^{T^{(a)}\to T^{(a)}}$, $g^{L^{(a)}\to L^{(a)}}$ and $g^{C^{(a)}\to C^{(a)}}$ for the Bund.

The first thing that we notice is that above a time scale $\tau_1 \sim 0.1$ seconds, these conditional laws behave as $t^{-\gamma}$ with $\gamma < 1$. For $g^{T^{(a)}\to T^{(a)}}$, this is the well known long-range memory of the order flows. Between τ_1 and $\tau_2 \sim 0.3ms$, the conditional laws roughly behaves as a power law of exponent of order 1. Below, τ_2, the conditional laws "saturate".

We also observe a few "bumps" on the conditional law which do not correspond to noise in the estimation. We believe that there are two kinds of bumps. The first kind of bumps appear at "round" times (0.01, 0.1, 0.5, 1 and 2 seconds), we believe that they are due to the automatization of trading. For example, if an algorithm posts a limit order every second, this implies a bump in the conditional law at 1 second. The second kind of bumps appears around $\tau_2 \sim 0.3$ milliseconds. We believe that it corresponds to the average reaction time (i.e., the average "latency") of the agents to an event. The decreasing of the conditional law below 0.03

milliseconds is an artefact of our data.

Once the conditional laws have been estimated, we solve the Wiener-Hopf system, following steps 2. and 3. in Section 3.4 with the parameters $K = 100$ (so that $\delta \simeq 0.15$), $T_{min} = 0.1$ milliseconds and $T_{max} \simeq 140$ seconds and compute the exogenous intensity estimates following step 4. We shall first comment on the values of the exogenous intensities.

4.4 Exogenous intensities

Let us recall that, within the Hawkes model, the exogenous intensity μ^i can be interpreted as the rate of (Poisson) events of type i that are "coming" from an exogenous source of information, i.e., that are not "caused" by any other past event in the model. In that respect, the ratio:

$$R^i = \frac{\mu^i}{\Lambda^i} \tag{15}$$

represents an exogeneity ratio, namely the ratio between the number of exogenous events and the total number of events of type i. Notice that, in the one dimensional case, according to Proposition 2, one has simply: $R = 1 - ||\phi||_1$.

	$P^{(a)}$	$P^{(b)}$	$T^{(a)}$	$T^{(b)}$	$L^{(a)}$	$L^{(b)}$	$C^{(a)}$	$C^{(b)}$
μ	2.37e-2	2.39e-2	1.06e-2	1.14e-2	2.07e-2	2.27e-2	1.53e-2	7.51e-3
R	2.7%	2.7%	4.3%	4.5%	1.1%	1.2%	0.7%	0.4%

Table V.2 – Estimated exogenous intensities (in s^{-1}) and the corresponding exogenous ratio (15) of the DAX futures.

	$P^{(a)}$	$P^{(b)}$	$T^{(a)}$	$T^{(b)}$	$L^{(a)}$	$L^{(b)}$	$C^{(a)}$	$C^{(b)}$
μ	7.13e-3	7.10e-3	1.41e-2	1.45e-2	3.83e-2	3.83e-2	4.00e-2	4.39e-2
R	4.4%	4.4%	4.5%	4.5%	1.4%	1.4%	1.6%	1.8%

Table V.3 – Estimated exogenous intensities (in s^{-1}) and the corresponding exogenous ratio (15) of the BUND futures.

The estimated intensities and the corresponding exogenous ratio for the DAX and the BUND futures are reported respectively in Tables V.2 and V.3. These tables reveal that, for both assets, the level of exogeneity is very low: R is only a few percent, meaning that most of the events can be considered to be directly triggered by past events within this model. See Chapters II and III for theoretical studies of the diffusive long term behavior of such almost purely endogenous Hawkes processes. For comparison, a simple one dimensional Hawkes model accounting for mid-price jump events as in refs [FS12] provides an exogenous ratio of respectively $R = 0.059$ for the DAX and $R = 0.10$ for the BUND future. This means that the

8-dimensional Hawkes model provides a better description of mid-price changes than a simple one dimensional model that has to involve a larger amount of "external" sources of information.

Let us also remark that, for both assets, market orders are more exogenous than limit and cancel orders. This is not surprising since several studies tend to show that market orders are "leading" limit and cancel. Thus, degree of endogeneity of the limit and cancel orders should be greater than the one of the market orders.

Finally, one can see that though limit and market orders have strikingly similar exogenous ratio for small and large tick assets, the mid-price changes and cancel occurrences are more endogenous in the case of small tick asset (DAX).

4.5 Matrix of kernel norms $||\phi||_1$

Before discussing the precise shape of the estimated kernels, we use the values of the norms of the kernels ϕ^{ij} to comment the main mutual and self excitations that occur in the order book[5]. The matrices of the kernel norms $||\phi||_1$ we obtained for the DAX and the BUND are reported in Figure V.4. For the sake of simplicity, we have represented the norm values using a colormap from blue to red. Notice that blue values correspond to negative kernel norms[6]. This feature has a natural interpretation as an inhibitory effect within the non-linear version of the Hawkes model described by Equation (3). We have checked, on numerical simulations, that as long as the realized quantity (2) remains most of the time positive, the non parametric estimation procedure introduced in Section 3.4, when applied to the non linear Hawkes process (3), leads to reliable estimations even for negative kernels.

As expected, the overall symmetry (ask/bid) is fairly well recovered empirically on the matrix shape. Any 2×2 sub-matrix with homogeneous inputs (i.e., same type of inputs to be chosen among Price changes, Trades, Cancels or Limits) and homogeneous outputs is symmetric. Thus for instance the kernel $\phi^{T^{(a)} \to C^{(b)}}$ seems to have the same norm as the kernel $\phi^{T^{(b)} \to C^{(a)}}$. Let us point out that, since the Dax is a stock index we could have expected some discrepancy between the ask side and the bid side (statistics of downward jumps are slightly different than statistics of upward jumps on equity markets) however, this discrepancy is negligible on intraday data.

One striking feature that clearly appears on both matrices is the anti-diagonal shape of the $(P^{(a)}, P^{(b)})$ sub-matrix and the diagonal shape of the $(T^{(a)}, T^{(b)})$, $(L^{(a)}, L^{(b)})$, $(C^{(a)}, C^{(b)})$ sub-matrices, see also Paragraphs V.B.1, V.B.2, V.B.3 and V.B.4 in appendix where the corresponding normalized cumulated kernels are plotted. This means that, on the one hand, market, limit and cancel orders mainly cause events of the same type and sign. This property

[5] Let us recall that the norm $||\phi^{ij}||_1$ represents the mean number of events of type i triggered by an event of type j.

[6] Let us remark that we allow an abuse of language in the sense that $||\phi||_1$ stands for $\int_0^\infty \phi(s)ds$ which is not really a norm unless $\phi(t) > 0$.

can be mainly attributed to the splitting of metaorders, see [LMF05], and to a less extent to some herding behavior of agents. We will see in the next section that these diagonal kernels are well described by roughly power-law decreasing functions. As far as mid-price jumps are concerned, the cross-exciting structure between $P^{(a)}$ and $P^{(b)}$ implies a strongly mean-reverting behavior, which is of the main characteristic of the price microstructure and which somehow guarantees the absence of long range correlation in prices, see [BM14a], in agreement with market efficiency. Note that this cross price kernel can be linked to the kernel of the propagator model, see [BGPW04] or Chapter I.

Finally, let us point out that that price changes appear to be the components that influences the most the other components, i.e., a price change "drives" the dynamics more than anything else. We will come back on this feature and on other features of this matrix (impact of orders on the price, impact of the price on events, impact of trades on the liquidity,...) in the following sections.

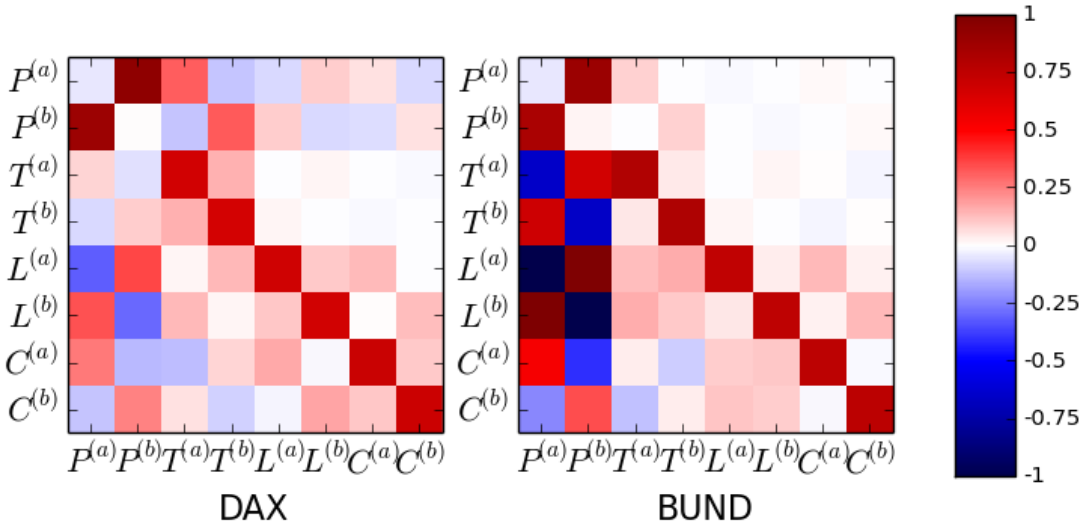


Figure V.4 – The matrix of estimated $||\phi^{ij}||$.

4.6 Matrix of norms $||\psi||_1$

Let us recall that the kernel ϕ^{ij} can be considered as the "bare" impact of an event of type j on an event of type i. If one wants to account for the "dressed" impact, i.e. the impact associated with all the cascade triggered by some event, one has to estimate the function ψ^{ij} as defined by (6) whose matrix of norms can be computed as:

$$||\psi||_1 = ||\phi||_1 (\mathbb{I} - ||\phi||_1)^{-1}.$$

More precisely, if one introduces:

$$||\bar{\psi}^{ij}||_1 = \frac{\mu^j}{\Lambda^i} ||\psi^{ij}||_1$$

then, according to the population interpretation mentionned in section 2.2 and Proposition 2, $||\bar{\psi}^{ij}||$ corresponds to the fraction of events of type i triggered by *exogenous* events of type j (while $||\psi||_1$ corresponds to the average number of events of type i indirectly generated by an event of type j). The matrix norm $||\bar{\psi}||$ is displayed in Figure V.5.

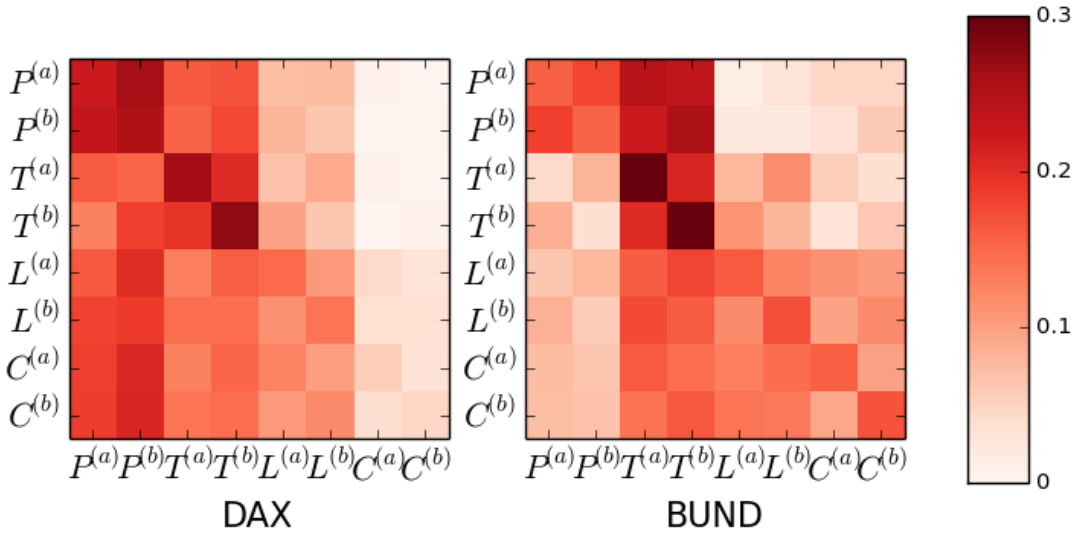


Figure V.5 – The matrix of estimated $||\bar{\psi}^{ij}||$.

One can see that exogenous limit and cancel orders have a poor influence on mid-price variations that are mainly caused by exogenous trades and price jumps. This indicates that exogenous information is mainly incorporated into prices through trades or orders that directly shift the mid-price. One can also remark that in the case of large tick asset (BUND), an exogenous price move mainly impacts the price itself but not the order flow that does not move the price while for the small tick asset (DAX), an exogenous price variation impacts all type of events.

4.7 Shape of the kernels

In all former studies where non parametric estimations of Hawkes kernels involved in the dynamics of order flows were performed [BDM12, BM14a, HBB13], power-law kernels with exponents close to 1 have been observed. This property can be directly linked to the strong persistence of the order flow dynamics, see Chapter VIII, mainly caused by the splitting of large orders and, to a lesser extent, to the herding behavior of agents. This feature remains true in the 8-dimensional description adopted in this paper when one focuses on the kernels involved in the self-excitation of market, limit and cancel orders and in the cross excitation of mid-price moves (the kernels with highest norms, see Figure V.4).

Figure V.6 displays $log_{10} - log_{10}$ plots of the kernels $\phi^{T^{(a)} \to T^{(a)}}$, $\phi^{L^{(a)} \to L^{(a)}}$, $\phi^{C^{(a)} \to C^{(a)}}$ and $\phi^{P^{(a)} \to P^{(b)}}$. It clearly appears that these kernels loosely behave as a power law of exponent slightly higher than one.

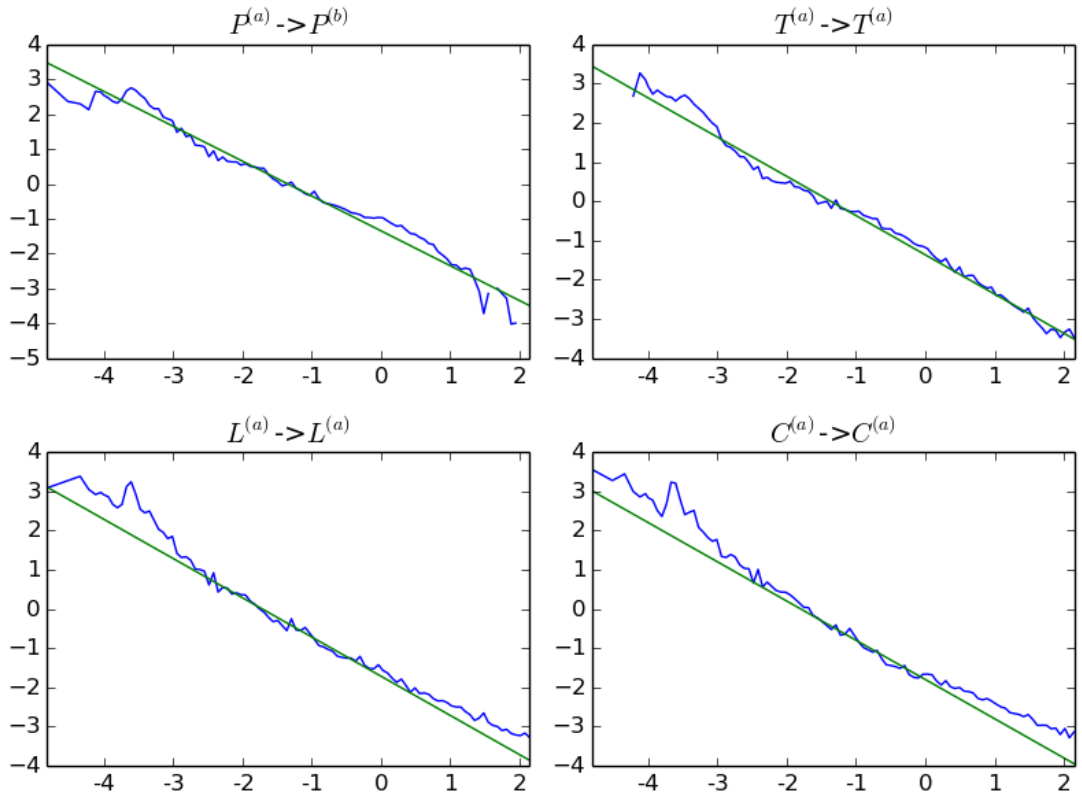


Figure V.6 – $\log_{10} - \log_{10}$ plots of $\phi^{P^{(a)} \to P^{(b)}}$, $\phi^{T^{(a)} \to T^{(a)}}$, $\phi^{L^{(a)} \to L^{(a)}}$ and $\phi^{C^{(a)} \to C^{(a)}}$ (in blue) and a straight line of trend -1 in green. These kernels display loose power-law behaviors with an exponent close to 1. More precisely, the exponent seems to be higher than 1 below 0.1 second and lower than one above 0.1 second.

Let us stress that this "roughly power law behavior" is not true for all kernels. For example, the kernels $\phi^{T^{(a)} \to P^{(a)}}$ and $\phi^{P^{(a)} \to L^{(a)}}$ are "localised" below the millisecond, see Figure V.7 and the cumulated kernels in Appendix V.B. Empirical kernels have a much "richer" behavior than conditional laws which as stated earlier all have the same stylized facts.

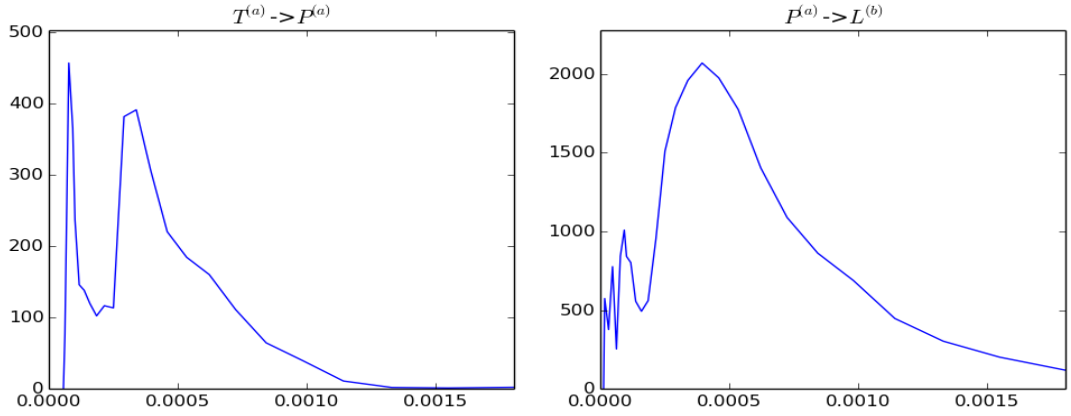


Figure V.7 – $\phi^{T^{(a)} \to P^{(a)}}$ and $\phi^{P^{(a)} \to L^{(a)}}$ as a function of time below the millisecond.

4.8 The impact of the order flows on the price

For the sake of clarity, we split the two 8×8 matrices of kernels into 2×2 sub-matrices. This representation of the kernels is provided in Appendix V.B for both DAX and BUND. Each sub-plot represents the normalized cumulated kernels $\frac{\Lambda^j}{\Lambda^i} \int_0^t \phi^{ij}(u)du$ as a function of $\log_{10}(t)$ (so there are 2×64 of them). This normalization for the kernel is natural in the population dynamics interpretation of Hawkes processes, see [HO74]. Indeed, $\frac{\Lambda^j}{\Lambda^i}||\phi^{ij}||$ corresponds to the proportion of i whose parent is a j (while $||\phi^{ij}||$ corresponds to the average number of children of type i for an individual of type j).

In this section, we focus on the impact of the flow of orders on the price, i.e., on the events which do not instantaneously move the price and that have an impact on mid-price variations. It corresponds to Paragraphs V.B.5, V.B.6 and V.B.7 in appendix.

One can see that the trades have a very localized delayed *self* impact on price variations (by *self* we mean *in the same direction*, e.g., an ask market order implies an eventual upward price move). Indeed, it seems that it takes 0.3 milliseconds for the market to incorporate the information of the trade. This is the averaged time for other agents to react to this order (this corresponds to the "latency" delay we have observed on the conditional laws in Section 4.3 and the kernels in Section 4.7). After this short time scale, the kernel is negligible, that is the cumulated kernel remains constant, and the trade does not imply any more price changes.

As far as limit and cancel orders are concerned, their impacts appear to be less localized and to be different when one compares the DAX (small tick) and the BUND (large tick). For the DAX, limit orders impact the price by exiting the price moves in the opposite direction (limit orders at the bid trigger upward price moves and conversely) and by slightly inhibiting the price moves in the same direction. For the BUND, limit orders mainly have an inhibitory effect one price changes (limit orders at the bid inhibit downward price moves). As opposed to the impact of market orders, the impact of limit orders is not very fast. While the information of

market orders is almost immediate (0.3 milliseconds), a limit order needs time to be significant.

Cancel orders impact the price in a very similar way as limit orders but in the opposite direction.

4.9 The impact of price jumps on the order flows

The kernels that correspond to the impact of price jumps on the order flows correspond to Paragraphs V.B.8, V.B.9 and V.B.10 where the normalized cumulated kernels[7] are plotted.

The most striking feature is that the effect of a mid-price change on market orders strongly depends on the tick size (as it is also observed on Figure V.4). For the DAX (small tick), this effect is rather small and *self-exciting* (upward price moves imply ask trades). On the contrary, for the BUND (large tick), this effect is far more important and is *cross-exciting* (upward price moves trigger bid trades and inhibit ask trades). This can be interpreted in terms of adverse selection: On the one hand, for large tick assets, after an upward price move, agents will not want to execute market orders at the new ask price because this price will be too high. On the other hand , the new bid price will be more interesting and agent will thus execute bid market orders.

On both assets, the influence of price moves on the liquidity is mostly *self* in the sense that upward price moves excite bid limit orders and ask cancel orders (and inhibit ask limit orders and bid cancel orders). The excitations are mostly localized (around 0.3 milliseconds) and correspond to the averaged time reactions of market makers to the new information while the inhibition kernels are long term. One can also notice that for the DAX, the short term exciting effect of say $\phi^{P^{(a)} \to C^{(a)}}$ is balanced by a longer term inhibitory effect (around 10 seconds). We believe that this effect is linked to the reversion of the liquidity to its stationary state a "long" time after the price change has occurred.

4.10 The impact of market orders on liquidity and vice-versa

The kernels that correspond to the impact of the market order flow on limit and cancel order flows are in Paragraphs V.B.11 and V.B.12 where the normalized cumulated kernels are plotted.

For both assets, trades have very fast positive influence on the opposite limit orders. This is due to the fact that the underlying "efficient" price has moved and the liquidity must adapt itself. At longer time scales, the effect is reversed (the cross kernels become negative, i.e., trades inhibit cross limit orders). This is due to the reversion of the liquidity towards its stationary state. As far as the influence of market orders on the rate of cancel orders is concerned, for the BUND, trades imply cancel orders in the same direction at very short time scale and then, at longer time scales, this influence becomes negative. This effect is very small which means that the proportion of limit and cancel orders directly implied by a trade

[7] See beginning of the previous section or of Appendix V.B

is small. However, we see on Figure V.5 that when we "dress" this impact with intermediate events, exogenous trades indirectly imply a significant proportion of limit and cancel orders.

Concerning the reverse impact of limit and cancel order flows on market order flow, the corresponding kernels are displayed in Paragraphs V.B.13 and V.B.14. One can see that ask (resp. bid) cancel and bid (resp. ask) limit orders excite the trading rate at the ask (resp. the bid).

4.11 The impact of the limit order flow on the cancel order flow and vice-versa

The kernels that correspond to the dynamics between limit and cancel order flows correspond to Paragraphs V.B.3, V.B.4, V.B.15 and V.B.16 which display the corresponding normalized cumulated kernels.

As for market orders, the limit/limit kernels and the cancel/cancel kernels are mostly self-exciting (e.g., $\phi^{L^{(a)}\to L^{(a)}}$ is dominant and $\phi^{L^{(a)}\to L^{(b)}}$ is negligible). Remark also that cancel orders have a short term influence on cross limit orders ($\phi^{C^{(a)}\to L^{(b)}}$ is localized around 0.3 milliseconds) which corresponds to the information of the cancel order and a longer term impact on self limit orders ($\phi^{C^{(a)}\to L^{(a)}}$ is a power law) which corresponds to the return of the liquidity.

In a similar way, limit orders have a short term influence on cross cancel orders and a longer term impact on self cancel orders.

5 Discussion and concluding remarks

We have presented a numerical solution of the Wiener-Hopf Equation (8) when some kernels behave as power laws of exponents slightly higher than one and are thus significant over a very wide range of time scales (from 100 microseconds to 100 seconds!) and in a rather large number of dimensions. This lead us to an efficient algorithm for non-parametric estimation of these kernels.

Using the natural causal interpretation of Hawkes processes, we applied this algorithm to the study of high frequency financial data (timestamped with a precision of a micro-second). It allowed us to disentangle the cross-influences between eight types of first limit events occurring in order books.

Our approach allowed us to retrieve some well known stylized facts about market dynamics:

- The order flows are strongly self excited: The main influence of market, limit and cancel orders is on themselves. This is linked to the well known persistence of order flows and to the splitting of meta-orders into sequences of orders. Moreover, orders also have an

influence of orders of the same "direction". For example, ask trades excite bid limit orders and ask cancel orders. The same kind of behavior had already been observed in [EBK12] by directly looking at the correlation functions between these events.

- The prices are efficient: To balance the strong persistence of the order flow imbalance, the influence of price changes on themselves is mostly cross. That is upward price moves cause downward price moves. This is consistent the measure of the "bare propagators" of [EBK12]. Indeed, in [EBK12], the bare impact of events on the price is shown to decrease in time. Not however that this bare impact is not exactly equivalent to our kernels since it does not fully "disentangle" the influences of all the events on all the events but only the influence of orders on price changes.

- The orders impact the price even if they do not move it mechanically. Again, this is consistent with the measures of the propagators, of [EBK12], of the events that do not move the price.

The generality of our Hawkes model allowed us to account for much richer dynamics than previous works and to describe and quantify the influences between all types of events. We have thus found some new and more subtle results that to our knowledge had never been observed.

- Price moves have a retro influence on orders flows: These are the kernels that the tick size influences the most. For large tick assets, adverse selection prevails and this effect is cross: upward price moves excite bid trades and vice versa. For small tick assets, this effect is self because of the persistence of order flows.

- The Hawkes framework and our estimation of the kernels allows us not only to have a measure of "causality" between the different events but also of the time scales at which this causality appears. We have found that there are loosely two kinds of influences: Fast and localized influences, for which the volume of the kernel is localized around the reaction time of the market (on our data 0.3 milliseconds) and influences which are spread over a wide range of time scales and whose corresponding kernels thus behave somehow as power law functions of exponents slightly higher than one.

- Note also that as opposed to [EBK12], our model is a real time model and not a discrete time model where each period corresponds to an event. This allows us for example, to see that market and limit orders have different impacts. Market orders have localized impacts while limit orders take time to have an impact.

- Finally, we can measure the endogeneity of all market events. We find in particular that market orders are more exogenous than limit and cancel orders. Therefore, when looking at the "dressed kernels" ψ, we find that the trades are leading in the sense that, exogenous trades have more influence than exogenous limit and cancel orders.

One thing that probably lacks in our purely Hawkes order book model is that the arrival of events does not directly depend on the state of the order book but only on the past events. For

example, it is not clear that in our model, when the spread is important, liquidity will arrive. On the other side of the spectrum of order book models, we can mention [GDKB13, HLR13] where the order book dynamics are purely Markovian and thus do not depend on the history of the order flow. We believe that "reality" is between these two approaches: the intensity of market events depends both on their history and on the state of the order book.

The stability and tractability of our numerical estimation method over extremely wide ranges of time scales and as the dimension of the model increases imply that we have many leads to improve our order book model. For example, we can add marks to account for the size of the different events, see [BM14b] or exogenous terms to account for external news, see [RPL14].

We conclude by underlying that of course, the relevance of Hawkes models is not restricted to finance. Therefore, our procedure can be applied to the study of any field where the measure of complex relations between point processes is necessary.

6 An alternative approach: Dirac Hawkes processes

In this section, we want to apply this estimation procedure to model the joint dynamics of trades and price changes. A problem that appears when doing so is that some trades imply price changes instantaneously while "classical" Hawkes processes almost surely do not present simultaneous jumps.

6.1 Dealing with Kernels involving Diracs

To account for this fact, we introduce "Dirac" Hawkes processes and we present our new associated kernel estimation procedure.

6.1.1 Construction of Dirac Hawkes process

An alternative way to build Hawkes processes proposed in [HO74] is to consider the following population dynamics model:

- There are individuals of type 1, ..., D.
- For every type i, individuals of type i migrate as a Poisson process of intensity μ^i.
- Individuals of type j can have children of all types.
- The children of type i of an individual of type j who was born or migrated in t are born according to a inhomogeneous Poisson process of intensity $\phi^{ij}(\cdot - t)$.

Considering this population representation of Hawkes processes, it is possible (by considering that individuals have a non null probability of having children at age 0) to define Hawkes processes of kernel

$$\phi = d_\phi \delta_0 + \phi_c$$

where ϕ_c is continuous and d_ϕ is a matrix. The stability condition thus becomes $\rho(d_\phi + |\phi_c|) < 1$.

6.1.2 Infinitesimal correlation

Let us begin by recalling the definition of the infinitesimal covariance, ν[8]:

$$\nu^{ij}(t-t')dtdt' = \mathbb{E}[dN_t^i dN_{t'}^j] - \Sigma^i \Sigma^j dtdt'$$

where Σ^i is the average intensity of the i^{th} process ($\mathbb{E}[dN_t^i]/dt$). We show in Appendix V.C that the following formula, proven in [BDM12] remains valid for Dirac Hawkes processes

$$\nu = (\delta\mathbb{I} + \psi) * \Sigma(\delta\mathbb{I} + \tilde{\psi}^T)$$

where $\psi = \sum_{k\geq 1} \phi^{*k}$.

Therefore, if we define the conditional laws as

$$g^{ij}(t)dt = \frac{\mathbb{E}[dN_t^i dN_0^j | dN_0^j > 0]}{\mathbb{E}[dN_0^j | dN_0^j > 0]} - \Sigma^i dt$$

which when the Dirac part of the kernel is null is the usual conditional law of the process[9]. We have the following relation between g and μ.

Proposition 5.

$$g = \nu^T \Sigma^{-1} = (\delta\mathbb{I} - \psi) * \Sigma(\delta\mathbb{I} - \tilde{\psi}^T)\Sigma^{-1}. \tag{16}$$

Proof.

$$\begin{aligned}
\nu^{ij}(t'-t)dtdt' &= \mathbb{E}[dN_t^i dN_{t'}^j] - \Sigma^i \Sigma^j dtdt' \\
&= \frac{\mathbb{E}[dN_{t'-t}^j dN_0^i | dN_0^i > 0]}{\mathbb{E}[dN_0^i | dN_0^i > 0]} \mathbb{E}[dN_0^i | dN_0^i > 0] P(dN_0^i > 0) - \Sigma^i \Sigma^j dtdt' \\
&= \Sigma^i g^{ji}(t'-t)dtdt'.
\end{aligned}$$

□

Let us now decompose g between its Dirac part and a its regular part:

$$g = d_g \delta_0 + g_c$$

where d_g is a matrix and g_c is a function.

[8] Note that in this paragraph, we define the infinitesimal covariance and the conditional laws in a slightly different manner than in the rest of the thesis because we do not take off their Dirac part.

[9] Except that we do not take off the Dirac part.

Remark 4. *The Dirac part of the conditional law can be estimated using the following empirical average*

$$\hat{d}_{g\,ij} = \frac{\sum_{k=1}^{n_j} dN^i_{T_k^j} dN^j_{T_k^j}}{\sum_{k=1}^{n_j} dN^j_{T_k^j}}$$

where the (T_k^j) *are the jump times of* j.

The estimation of the continuous part of the conditional law is the same as in the non Dirac case.

6.1.3 Dirac estimation

The Dirac part of ϕ, d_ϕ can be estimated by keeping the Dirac parts in (5):

$$d_g\Sigma = (\mathbb{I} - d_\phi)^{-1}\Sigma((\mathbb{I} - d_\phi)^{-1})^T = XX^T$$

which writes

$$Y = XX^T$$

with $Y = d_g\Sigma$ and $X = (\mathbb{I} - d_\phi)^{-1}\sqrt{\Sigma}$.

For a given symmetric Y, there is a priori an infinite number of X which satisfy this equation. However, if we make the additional assumption that we have ordered the indexes so that d_ϕ is lower triangular then X must also be lower triangular (because $(\mathbb{I} - d_\phi)^{-1} = \sum d_\phi^k$) and $Y = XX^T$ has a unique solution which can be solved using the Cholesky algorithm. This lower triangular assumption can be understood as an assumption about the causality between simultaneous events. For example, when there is a trade and a price move at the same time, we (might) want to say that it is the trade which "caused" the price move and not the price move which "caused" the trade. We therefore want to set to zero the corresponding element of d_ϕ.

Once we have X, we easily get

$$d_\phi = \mathbb{I} - \sqrt{\Sigma}X^{-1}.$$

6.1.4 Regular part estimation

Covoluting Equation (5) by $(I\delta - \phi)$ on the left as in [BM14b], we get that for $t > 0$:

$$\begin{aligned} (I\delta - \phi) * g(t) &= \Sigma((\mathbb{I} - d_\phi)^T)^{-1}\Sigma^{-1}\delta \\ d_g\delta + g_c - d_\phi d_g\delta - d_\phi g_c - \phi_c d_g - \phi_c * g_c &= \Sigma((\mathbb{I} - d_\phi)^T)^{-1}\Sigma^{-1}\delta \end{aligned}$$

We remark that d_ϕ has been set so that the terms in δ disappear and there remains to solve:

$$g_c = d_\phi g_c + \phi_c d_g + \phi_c * g_c$$

which is not far from a Wiener-Hopf system. We can apply a slightly modified version of the scheme presented in the previous paragraph to solve it.

$$\begin{aligned} g_c(t_n) &= d_\phi g_c(t_n) + \phi(t_n) d_g + \sum_{k=0}^{N-1} \phi(t_k) \int_{t_n - t_{k+1}}^{t_n - t_k} g_c(u) du \\ &+ \sum_{k=0}^{N-1} \frac{(\phi(t_{k+1}) - \phi(t_k))(t_n - t_k)}{t_{k+1} - t_k} \int_{t_n - t_{k+1}}^{t_n - t_k} g_c(u) du \\ &- \sum_{k=0}^{N-1} \frac{\phi(t_{k+1}) - \phi(t_k)}{t_{k+1} - t_k} \int_{t_n - t_{k+1}}^{t_n - t_k} u g_c(u) du \end{aligned}$$

6.2 A Hawkes impact model

We begin our empirical study by applying our estimation procedure to the model presented in [BM14a]: The 4-dimensional counting process

$$(PA, PB, TA, TB)$$

where

- PA (resp. PB) counts the number of upward (resp. downward) mid-price moves
- TA (resp. TB) counts the number of trades at the ask (resp. bid)

is modeled as a 4-dimensional Hawkes process.

In the next paragraph we present our estimation results. To take into account the Dirac part of the kernel in our representation of the results, instead of plotting $\int_0^t \phi^{ij}(s) ds$, we plot $d_\phi^{ij} + \int_0^t \phi_c^{ij}(s) ds$

We get similar results as in the 8-dimensional model above. The main difference is that since we include the trades which move the price in the processes TA and TB, we observe a "Dirac impact" in zero in the kernels $\phi^{TA->\rightarrow PA}$ and $\phi^{TB->\rightarrow PB}$.

6.3 Estimation results

6.3.1 DAX

Let us begin by presenting the average intensities Λ and the exogenous intensities μ:

	PA	PB	TA	TB
μ	1.04e-2	1.02e-2	1.85e-2	1.85e-2
Λ	7.80e-1	7.80e-1	4.44e-1	4.45e-1

Table V.4 – μ and Λ for the DAX.

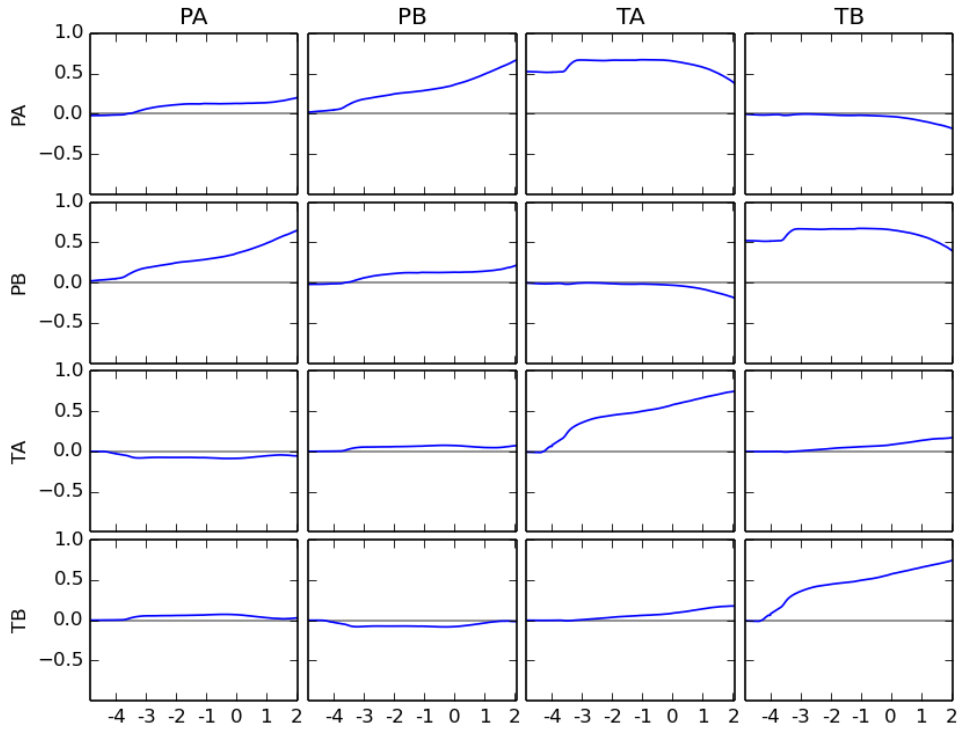


Figure V.8 – $d_\phi + \int_0^t \phi_c(s)ds$ as a function of $\log(t)$ for the DAX.

6.3.2 Bund

We do the same for the Bund.

	PA	PB	TA	TB
μ	1.99e-4	3.47e-4	1.69e-2	1.67e-2
Λ	1.48e-1	1.48e-1	3.62e-1	3.69e-1

Table V.5 – μ and Λ for the Bund.

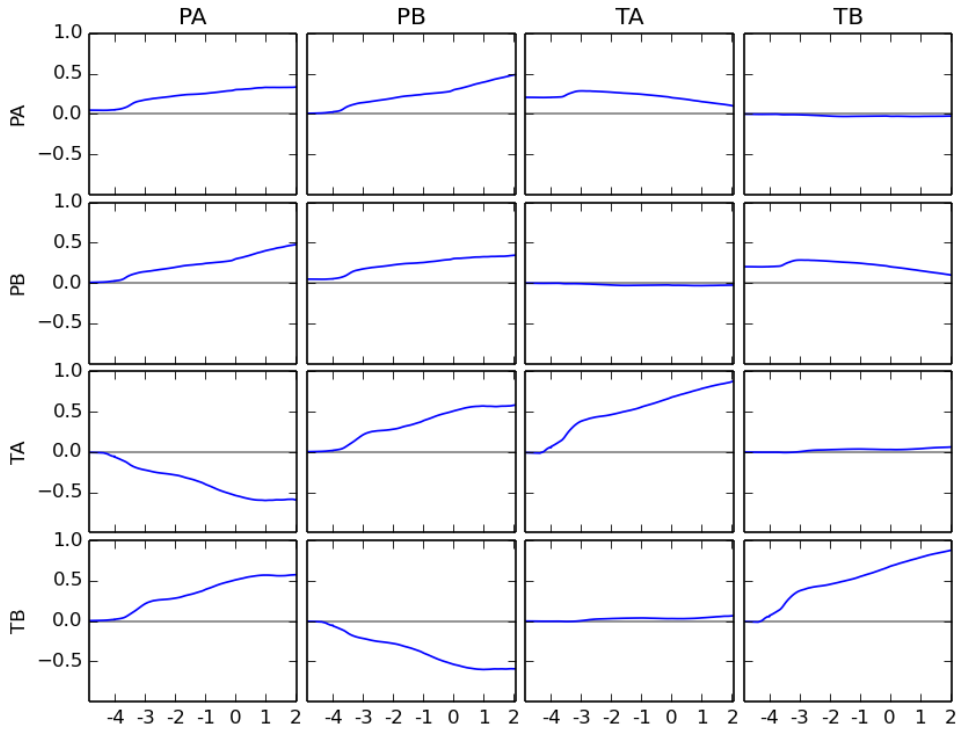


Figure V.9 – $d_\phi + \int_0^t \phi_c(s)ds$ as a function of $\log(t)$ for the Bund.

V.A Multiscale estimation of the conditional laws

The aim of this appendix is to present our multi scale estimation procedure of the conditional law g defined by (3).

V.A.1 Estimation procedure of g

Let us assume that, for all (i,j), we have the times of the events i $(T_k^i)_{k\leq n_i}$ and j $(T_k^j)_{k\leq n_j}$ and we want an empirical estimation of the conditional laws g^{ij}.

To do that, we will choose a time grid $(t_l)_{l\leq n}$ small enough so that we can approximate $g^{ij}(\frac{t_l+t_{l+1}}{2})$ by $\frac{1}{t_{l+1}-t_l}\int_{t_l}^{t_{l+1}} g^{ij}(s)ds$ but large enough so that there is enough points (k,k') such that $T_k^i - T_{k'}^j \in [t_l, t_{l+1}]$ to have a good approximation of $\int_{t_l}^{t_{l+1}} g^{ij}(s)ds$ by $\frac{1}{n_j}\sum_{k=1}^{n_i}\sum_{k'=1}^{n_j} \mathbb{1}_{T_k^i - T_{k'}^j \in [t_l,t_{l+1}]}$. We thus approximate $g^{ij}(\frac{t_l+t_{l+1}}{2})$ by

$$\hat{g}^{ij}(\frac{t_l+t_{l+1}}{2}) = \frac{1}{t_{l+1}-t_l}\frac{1}{n_j}\sum_{k=1}^{n_i}\sum_{k'=1}^{n_j} \mathbb{1}_{T_k^i - T_{k'}^j \in [t_l,t_{l+1}]}$$

and we take a piecewise affine interpolation of $\hat{g}^{i,j}$ between the points $\frac{t_l+t_{l+1}}{2}$.

V.A.2 Choice of the grid

The first natural grid that one can take is a uniform grid: $[0, h, 2h, ..., h_{max}]$. However, if one does that, one cannot get a good estimation of the conditional law on a large range of time scales. For example, if one takes $h = 0.0001s$ to have a good estimation of the conditional law at low time scales, one cannot estimate get a good estimation around one second. Indeed, there will be very few points of i between 1 and 1.0001 seconds after a point of j.

To solve this problem, we will consider a grid that is uniform between 0 and h_{min} and log-uniform between h_{min} and h_{max}:

$$[0, h_{min}h_\delta, h_{min}2h_\delta, ..., h_{min}, h_{min}e^{h_\delta}, h_{min}e^{2h_\delta}, ..., h_{max}]. \tag{17}$$

Doing this, we will have enough points between two points of the grid but the claw will not vary too much between two points of the grid.

V.B Estimation results for the full order book model: the cumulated kernel matrix $\int_0^t \phi(s)ds$

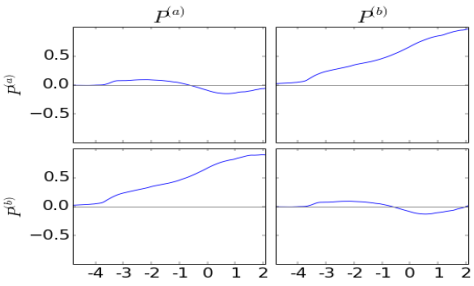
In this Appendix, we represented the resulting kernels estimated on the high frequency database (see Section 4). For each i and j in

$$\{P^{(a)}, P^{(b)}, T^{(a)}, T^{(b)}, L^{(a)}, L^{(b)}, C^{(a)}, C^{(b)}\}$$

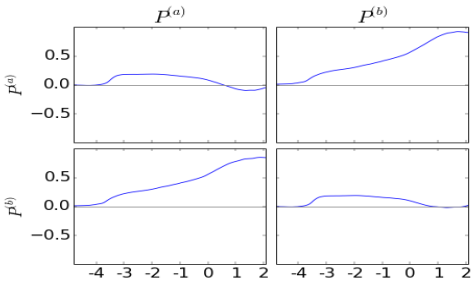
we have represented the estimated normalized cumulated kernels $\frac{\Lambda^j}{\Lambda^i}\int_0^t \phi^{i\to j}(u)du$ as a function of $\log_{10}(t)$. This normalization for the kernel is natural because in the population dynamics interpretation of Hawkes processes $\frac{\Lambda^j}{\Lambda^i}||\phi^{i\to j}||_1$ corresponds to the proportion of i whose parent is a j (while $||\phi^{i\to j}||_1$ corresponds to the average number of children of type i for an individual of type j).

There are 2×64 plots that are organized by assets and by types of events.

V.B.1 Influence of price moves on price moves

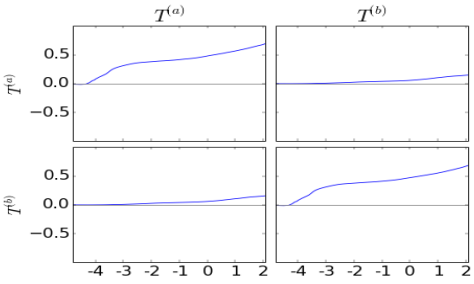


(a) xFDAX

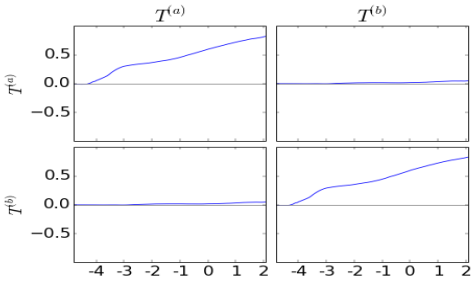


(b) xFGBL

V.B.2 Influence of trades on trades

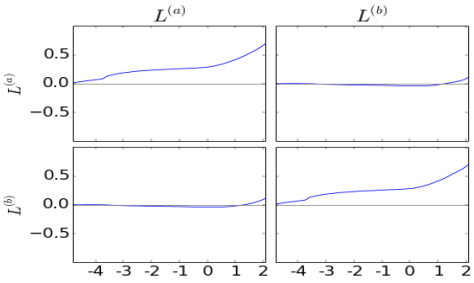


(a) xFDAX

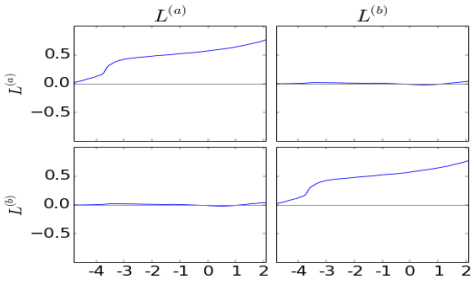


(b) xFGBL

V.B.3 Influence of limit orders on limit orders

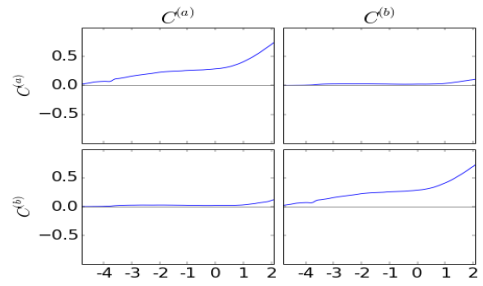


(a) xFDAX

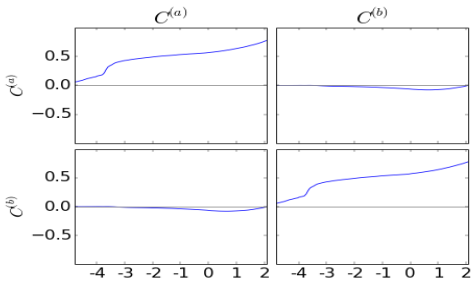


(b) xFGBL

V.B.4 Influence of cancel orders on cancel orders

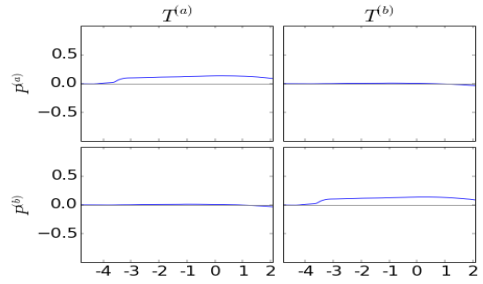


(a) xFDAX

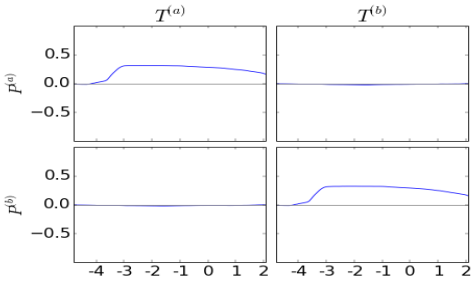


(b) xFGBL

V.B.5 Influence of trades on price changes

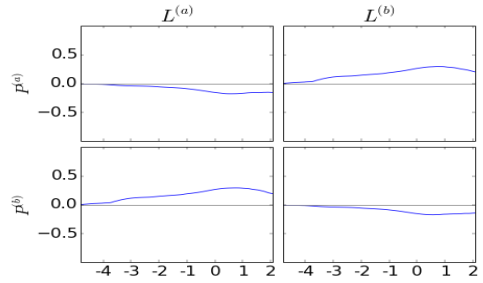


(a) xFDAX

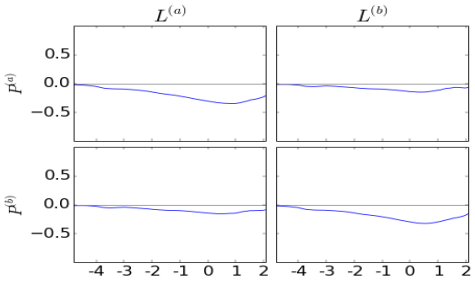


(b) xFGBL

V.B.6 Influence of limit orders on price changes

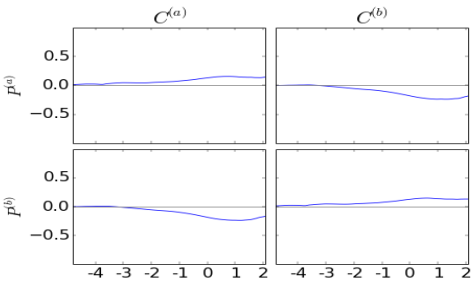


(a) xFDAX

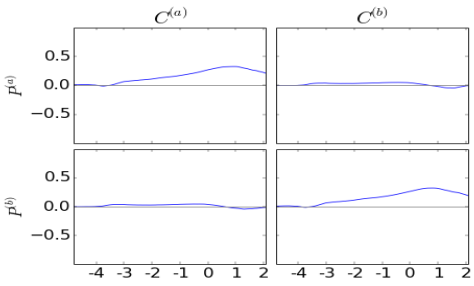


(b) xFGBL

V.B.7 Influence of cancel orders on price changes

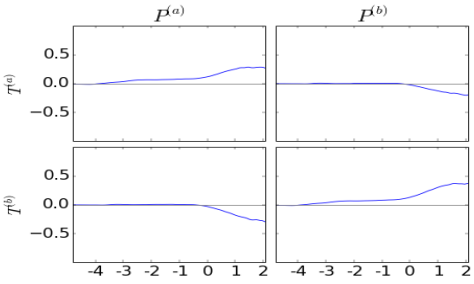


(a) xFDAX

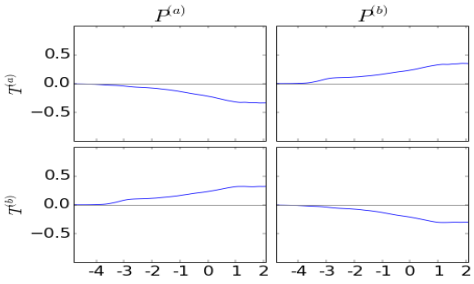


(b) xFGBL

V.B.8 Influence of price changes on trades

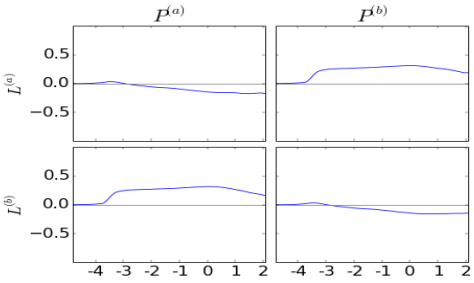


(a) xFDAX

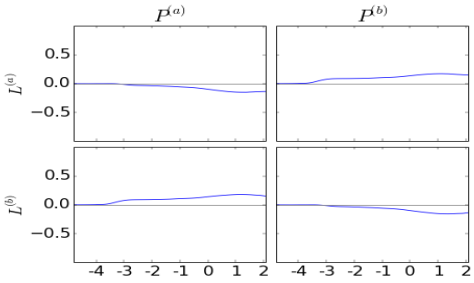


(b) xFGBL

V.B.9 Influence of price changes on limit orders

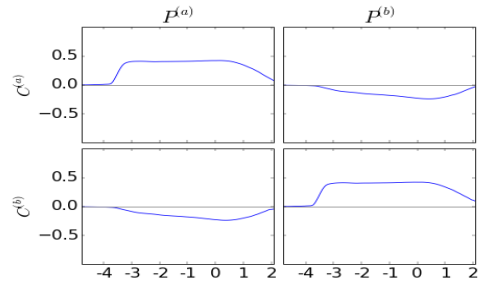


(a) xFDAX

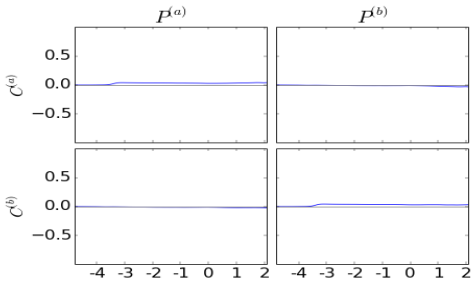


(b) xFGBL

V.B.10 Influence of price changes on cancel orders

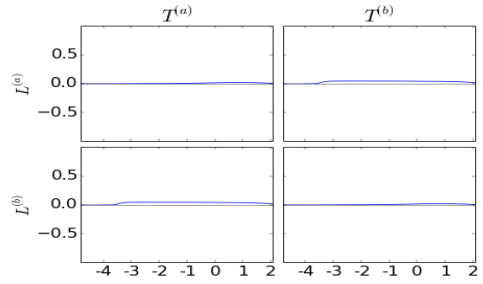


(a) xFDAX

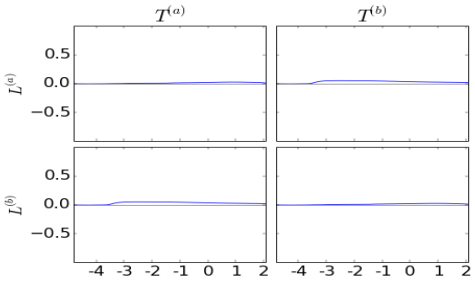


(b) xFGBL

V.B.11 Influence of trades on limit orders

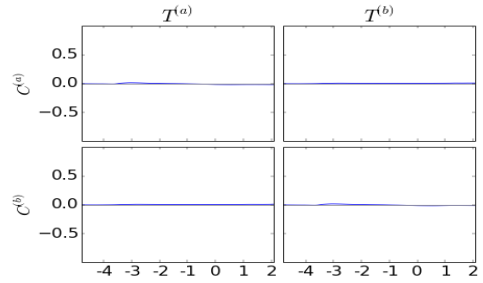


(a) xFDAX

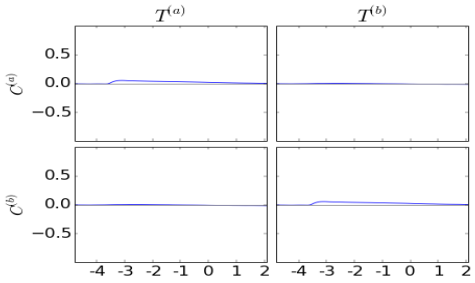


(b) xFGBL

V.B.12 Influence of trades on cancel orders

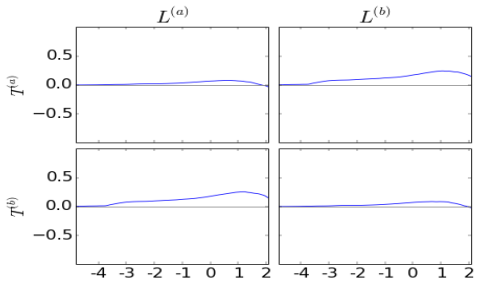


(a) xFDAX

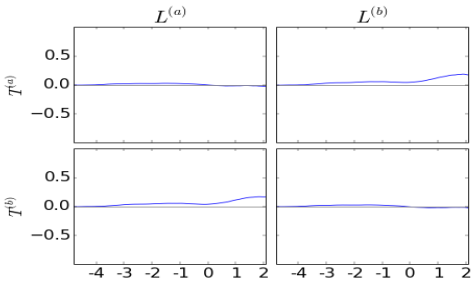


(b) xFGBL

V.B.13 Influence of limit orders on trades

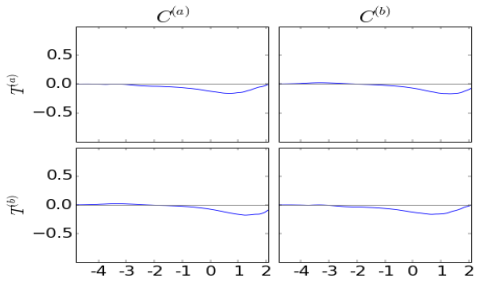


(a) xFDAX

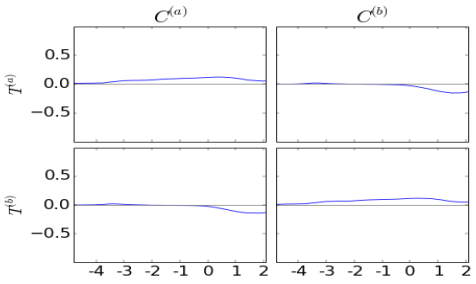


(b) xFGBL

V.B.14 Influence of cancel orders on trades

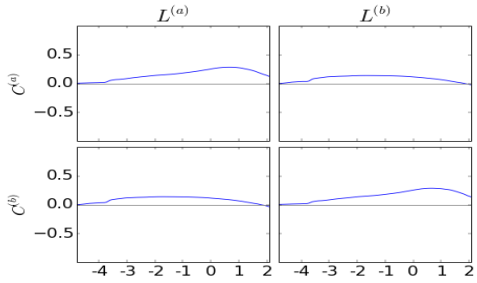


(a) xFDAX

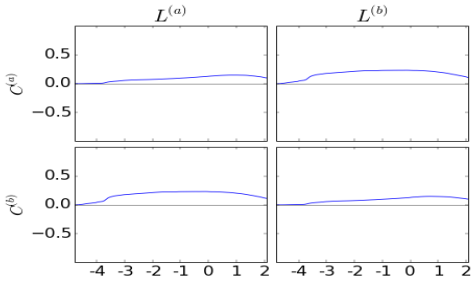


(b) xFGBL

V.B.15 Influence of limit orders on cancel orders

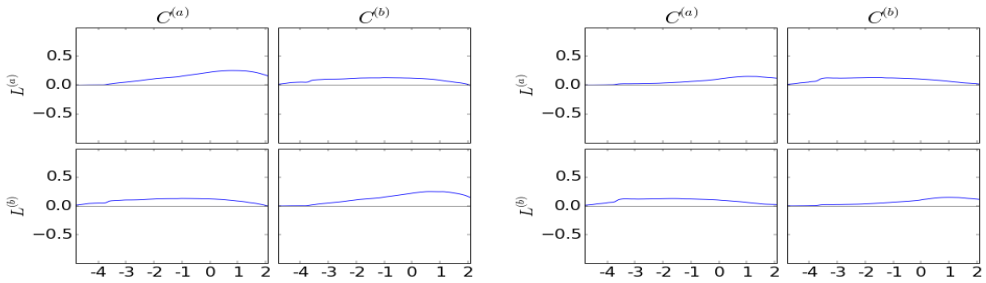


(a) xFDAX



(b) xFGBL

V.B.16 Influence of cancel orders on limit orders



(a) xFDAX (b) xFGBL

V.C Infinitesimal covariance for Dirac Hawkes processes

The aim of this appendix is to show the generalization of the link between the infinitesimal covariance and the Hawkes kernel

$$\nu = (\delta\mathbb{I} - \tilde{\psi}) * \Sigma(\delta\mathbb{I} - \psi^T)$$

when the kernel has a Dirac part. To do that, we consider a Hawkes process of kernel

$$\phi = d_\phi \delta_0 + \phi_c$$

as constructed in Section 6.1.1.

Remark 5. *The points now arrive as "clusters" of simultaneous events and for every cluster, we can define the "initial" point of the cluster as the "ancestor" of the other points of the cluster.*

V.C.1 A multidimensional branching process

Let us begin by considering the following population model[10]: We start with an individual of type i. This individual has children of type 1, ..., D. The number of children of type j of this individual being a Poisson random variable of parameter a_{ij}. The children then have children of type 1, ..., D. The number of children of type j of an individual of type k being a Poisson random variable of parameter a_{kj}. All the Poisson variables are independent.

Denoting the matrix $A = (a_{ij})_{ij}$, under the stability condition

$$\rho(A) < 1,$$

where ρ denotes the spectral radius, the total number of individual is almost surely finite.

We now define the generating function of the numbers of individuals.

[10] This model will reproduce the laws of the number of individuals in the clusters whose initial individual is of type i.

Definition 4. *Let us denote N_k^i, the total number of individuals of type k (the index i indicating that the original individual is of type i). Then if we call f_i the function defined as*

$$f_i(z_1,...,z_D) = \mathbb{E}[z_1^{N_1^i} \times ... \times z_D^{N_D^i}],$$

f_i is linked to the moments of (N_k^i) by the following proposition.

Proposition 6.

$$\frac{\partial f_i}{\partial z_k}(1,...,1) = \mathbb{E}[N_k^i]$$

and

$$\frac{\partial^2 f_i}{\partial z_k \partial z_{k'}}(1,...,1) = \begin{cases} \mathbb{E}[N_k^i N_{k'}^i] & \text{if } k \neq k' \\ \mathbb{E}[(N_k^i)^2] - \mathbb{E}[N_k^i] & \text{if } k = k'. \end{cases}$$

We will use the generating function of Poisson processes that we denote as

$$P_a(z) = \mathbb{E}[z^{N_a}],$$

where N_a is a Poisson process of parameter a, and the basic properties:

$$P_a(1) = 1,\ P_a'(1) = a \text{ and } P_a''(1) = a^2.$$

We will make extensive use of the following branching property.

Proposition 7. *f_i checks*

$$f_i(z_1,...,z_D) = z_i P_{a_{i1}}(f_1(z_1,...,z_D)) \times ... \times P_{a_{iD}}(f_D(z_1,...,z_D)) \tag{18}$$

Proof. The proof is similar to that of the classical Galton-Watson branching property.

We begin by the following representation of the N_j^i:

$$N_j^i = \begin{cases} \sum_{k=1}^D \sum_{l=1}^{Poiss_{ik}} (N_j^k)^l & \text{if } j \neq i \\ 1 + \sum_{k=1}^D \sum_{l=1}^{Poiss_{ik}} (N_i^k)^l & \text{if } j = i. \end{cases}$$

where the $Poiss_{ik}$ are independent Poisson variables of parameter a_{ik} and the $(N_j^k)^l$ are independent random variables whose laws are that of N_j^k. This implies that:

$$\begin{aligned} f_i(z_1,...,z_D) &= \mathbb{E}[z_1^{N_1^i} \times ... \times z_D^{N_D^i}] \\ &= z_i \Pi_{k=1}^D \mathbb{E}[z_1^{\sum_{l=1}^{Poiss_{ik}} (N_1^k)^l} \times ... \times z_D^{\sum_{l=1}^{Poiss_{ik}} (N_D^k)^l}] \\ &= z_i \Pi_{k=1}^D \mathbb{E}[f_k(z_1,...,z_D)^{Poiss_{ik}}] \\ &= z_i P_{a_{i1}}(f_1(z_1,...,z_D)) \times ... \times P_{a_{iD}}(f_D(z_1,...,z_D)). \end{aligned}$$

□

V.C.1.1 First order properties

Differentiating with respect to z_k Equation (18), we get:

$$\frac{\partial f_i}{\partial z_k}(z) = \mathbb{1}_{i=k}\frac{f_i(z)}{z_i} + z_i \sum_{j=1}^{D} \frac{\partial f_j}{\partial z_k}(z) P'_{a_{ij}}(f_j(z)) \frac{f_i(z)}{z_i P_{a_{ij}}(f_j(z))}. \tag{19}$$

Taking $z = (1, ..., 1)$ and using $f_i(1, ..., 1) = 1$, this yields

$$\mathbb{E}[N_k^i] = \mathbb{1}_{i=k} + \sum_{j=1}^{D} \mathbb{E}[N_k^j] a_{ij}.$$

Therefore, if we define the matrix E as $E_{ik} = \mathbb{E}[N_k^i]$, we get the following first order properties.

Proposition 8.

$$E = (\mathbb{1} - A)^{-1}.$$

V.C.1.2 Second order properties

Differentiating with respect to $z_{k'}$ Equation (19), we get:

$$\begin{aligned}
\frac{\partial^2 f_i}{\partial z_k \partial z_{k'}}(z) &= \frac{\mathbb{1}_{i=k}}{z_i}\frac{\partial f_i}{\partial z'_k}(z) + \frac{\mathbb{1}_{i=k'}}{z_i}\frac{\partial f_i}{\partial z_k}(z) - 2\frac{\mathbb{1}_{i=k=k'}}{z_i^2} f_i(z) \\
&+ \sum_{j=1}^{D} \frac{\partial^2 f_j}{\partial z_k \partial z_{k'}}(z) P'_{a_{ij}}(f_j(z)) \frac{f_i(z)}{z_i P_{a_{ij}}(f_j(z))} \\
&+ \sum_{j=1}^{D} \frac{\partial f_j}{\partial z_k}(z) \frac{\partial f_j}{\partial z'_k}(z) P''_{a_{ij}}(f_j(z)) \frac{f_i(z)}{z_i P_{a_{ij}}(f_j(z))} \\
&+ \sum_{j=1}^{D} \frac{\partial f_j}{\partial z_k}(z) P'_{a_{ij}}(f_j(z)) \Big[\frac{\frac{\partial f_i}{\partial z'_k}(z)}{z_i P_{a_{ij}}(f_j(z))} \\
&- \frac{f_i(z)}{(z_i P_{a_{ij}}(f_j(z)))^2} \big(P_{a_{ij}}(f_j(z)) \mathbb{1}_{k'=i} + z_i \frac{\partial f_j}{\partial z_k}(z) P'_{a_{ij}}(f_j(z)) \big) \Big]
\end{aligned}$$

which, using Proposition 8, for $z = (1, ..., 1)$ implies:

$$\frac{\partial^2 f_i}{\partial z_k \partial z_{k'}}(1, ..., 1) = \sum_{j=1}^{D} a_{ij} \frac{\partial^2 f_j}{\partial z_k \partial z_{k'}}(1, ..., 1) + E_{ik} E_{ik'} - \mathbb{1}_{i=k=k'}. \tag{20}$$

This yields that

$$\frac{\partial^2 f_i}{\partial z_k \partial z_{k'}}(1, ..., 1) = \sum_{j=1}^{D} E_{ij} E_{jk} E_{jk'} - E_{ik} \mathbb{1}_{k=k'}.$$

V.C.2 Back to Hawkes estimation

Let us consider a Hawkes process whose kernel has a Dirac part

$$\phi = \delta d_\phi + \phi_c.$$

In that case, the orders arrive as simultaneous independent "clusters" and we denote:

- Σ_i, the average intensity of N^i.
- Λ_i, the average intensity of arrival of clusters whose first point is i.

Remark 6. *The number of points of type k in a cluster whose first point is of type i has the same law as N_k^i in the previous paragraph replacing A by d_ϕ^t.*

Summing over all the possible first points of the clusters, this implies the following result.

Proposition 9. *Σ and Λ are linked by*

$$\Sigma_i = \sum_{j=1}^{D} E_{ij}\Lambda_j$$

where $E = (\mathbb{I} - d_\phi^t)^{-1}$.

V.C.2.1 Dirac infinitesimal autocorrelation

Let us now study the Dirac infinitesimal autocorrelation defined bellow.

Definition 5.

$$C_{kk'} = \frac{\mathbb{E}[dN_t^k dN_t^{k'}]}{dt}.$$

Summing over all the possible first points of the clusters, it checks

$$C_{kk'} = \sum_{i=1}^{D} \Lambda_i \left(\frac{\partial^2 f_i}{\partial z_k \partial z_{k'}}(1,...,1) + \frac{\partial f_i}{\partial z_k}(1,...,1)\mathbb{1}_{k=k'}\right)$$

which yields after calculations

$$C_{kk'} = \sum_{j=1}^{D} \Sigma_j E_{jk} E_{jk'}.$$

Proposition 10. *In matrix notations, this writes*

$$C = E^t \Sigma E = (\mathbb{I} - d_\phi)^{-1}\Sigma((\mathbb{I} - d_\phi)^{-1})^t$$

since according to the previous paragraph $E = ((\mathbb{I} - d_\phi)^{-1})^t$.

V.C.2.2 Infinitesimal autocorrelation

Let us now define:

Definition 6. *The renormalized intensity l*

$$l_t = (\mathbb{I} - |\phi|)^{-1}\lambda_t$$

and the compensated point process M

$$dM_t = dN_t - l_t dt.$$

Using Proposition 10 implies the following result:

Proposition 11. *M is a martingale and*

$$\mathbb{E}[dM_t dM_t^t] = \mathbb{E}[dN_t dN_t^t] = (\mathbb{I} - d_\phi)^{-1}\Sigma((\mathbb{I} - d_\phi)^{-1})^t dt.$$

Setting the vector $S = (\mathbb{I} - |\phi|)^{-1}\mu$[11] and the matrix function $\bar{\psi} = (\sum_{k\geq 0}\phi^{*k}) * \phi_c$, we have the following result.

Proposition 12. *l checks:*

$$l_t = S + \bar{\psi} * dM_t.$$

Proof. Decomposing dN in the definition of λ yields

$$l_t = (\mathbb{I} - d_\phi)^{-1}\mu + (\mathbb{I} - d_\phi)^{-1}\phi_c * dM_t + (\mathbb{I} - d_\phi)^{-1}\phi_c * l_t$$

which, multiplying both sides by $(\mathbb{I} - d)$ writes

$$(\mathbb{I}\delta - d_\phi\delta - \phi_c) * l_t = \mu + \phi_c * dM_t.$$

Convoluting both sides by $\sum_{k\geq 0}(d_\phi\delta + \phi_c)^{*k}$ implies the proposition. □

Using the decomposition:

$$\mathbb{E}[dN_t dN_{t'}^t] = \mathbb{E}[l_t l_{t'}^t]dt dt' + \mathbb{E}[dM_t l_{t'}^t]dt' + \mathbb{E}[l_t dM_{t'}^t]dt + \mathbb{E}[dM_t dM_{t'}^t],$$

the following expressions (implied by Propositions 11 and 12 and proceeding as in [BDM12]):

- $\mathbb{E}[dM_t dM_{t'}^t] = \mathbb{1}_{t=t'}(\mathbb{I} - d_\phi)^{-1}\Sigma((\mathbb{I} - d_\phi)^{-1})^t dt$
- $\mathbb{E}[l_t dM_{t'}^t] = \bar{\psi}^t(t - t')(\mathbb{I} - d_\phi)^{-1}\Sigma((\mathbb{I} - d_\phi)^{-1})^t dt'$
- $\mathbb{E}[dM_t l_{t'}^t] = ((\mathbb{I} - d_\phi)^{-1})^t\Sigma(\mathbb{I} - d_\phi)^{-1}\tilde{\bar{\psi}}(t - t')dt$[12]
- $\mathbb{E}[l_t l_{t'}^t] = \bar{\psi}^t * \tilde{\bar{\psi}}(t - t')(\mathbb{I} - d_\phi)^{-1}\Sigma((\mathbb{I} - d_\phi)^{-1})^t + SS^t$

[11] S is the diagonal of Σ.

[12] For every function f, $\tilde{f}$ is defined as $\tilde{f}(t) = f(-t)$.

and the following decomposing of ψ:

$$\psi = (\sum_{k\geq 1} d_\phi^k)\delta + \bar{\psi}(\mathbb{I} - d_\phi)^{-1}$$

we finally get our result:

Theorem 3. *The infinitesimal autocorrelation function ν checks:*

$$\nu(t-t')dtdt' = \mathbb{E}[dN_t dN_{t'}^t] - SS^t dtdt' = (\mathbb{I}\delta + \psi)^{-1} * \Sigma((\mathbb{I}\delta + \psi)^{-1})^t (t-t')dtdt'.$$

V.D Error analysis

In this section, we study the error that we make when we apply our estimation procedure. Note that there are many other elements on this topic in [BM14b].

V.D.1 Bootstrap error analysis

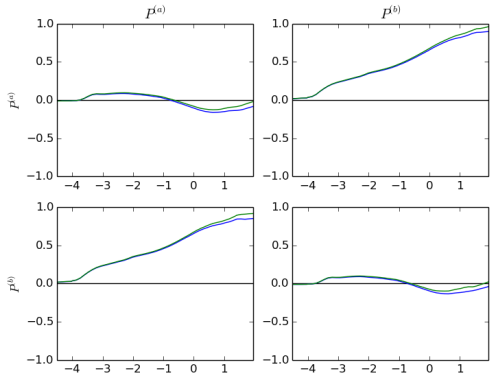
In this paragraph, we use a bootstrap method to obtain empirical error bounds on the kernel estimations and to show that our comments are robust. We apply it to the DAX but of course we obtain similar results for the Bund.

V.D.1.1 Description of the procedure

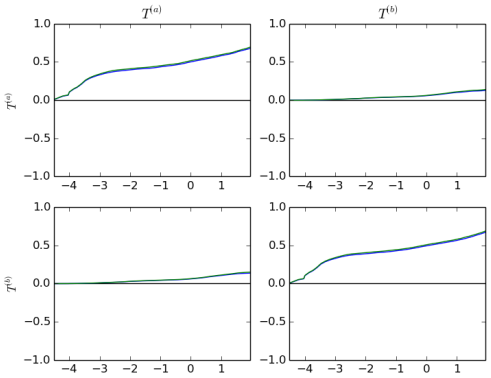
Recall that our data consists of 250 trading days. In order to study the estimation error, we choose at random 125 days among these 250 days and we compute the kernels on these days (choosing the same estimation parameters as before). We do this 100 times. We thus have 100 kernel estimations each performed on a set of 125 days. We then compute the 10% and 90% percentiles of the cumulated kernels and we plot these cumulated kernels.

V.D.1.2 Results

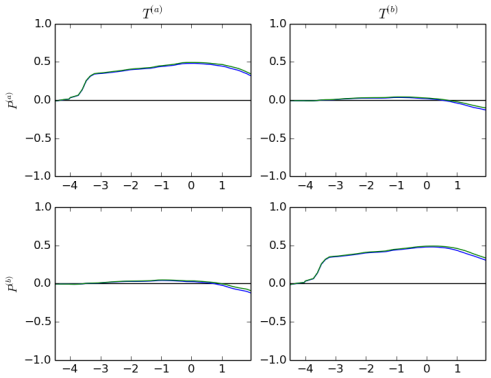
Below we present 10% and 90% cumulated kernels defined above corresponding to the influence of price moves on price moves, trades on trades, trades on price moves and limit orders on cancel orders.



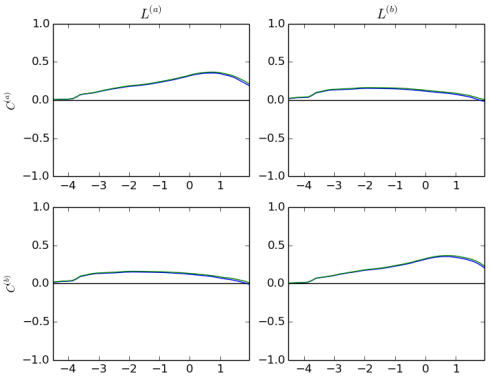
(a) Price moves on price moves.



(b) Trades on trades.



(c) Trades on price moves.



(d) Limit orders on cancel orders.

Figure V.26 – 10% and 90% cumulated kernels as a function of $\log 10(t)$.

We obtain that the error is always very small compared to the cumulated kernels themselves. It is the largest for the kernels encoding the influence of price moves on price moves on price moves and is always below 7%. Our comments about the causality between market events thus seem to be robust to this error.

V.D.2 Other simulations

In this section, we test our estimation procedure on different simulations of Hawkes processes. See [BM14b] for many other examples.

V.D.2.1 Multidimensional power law

As a first example, we consider slowly decreasing power law Hawkes kernels in dimension 4

$$\phi^{ij}(t) = \frac{a^{ij}}{(\delta + t)^{\beta}}$$

where $\beta = 1.3$, $\delta = 0.0005$ and the a^{ij} are chosen uniformly random between 0 and 0.006. We also take $\mu = 0.1$ and a simulation time $T = 100000$. We then apply the estimation procedure

to the simulated data with estimation parameters $h_\delta = 0.1$, $\delta = 0.05$, $T_{min} = h_{min} = 0.0001$, $T_{max} = h_{max} = 220$. As before, we plot the integral of the kernels as a function of the log of the time (the error is below 1%).

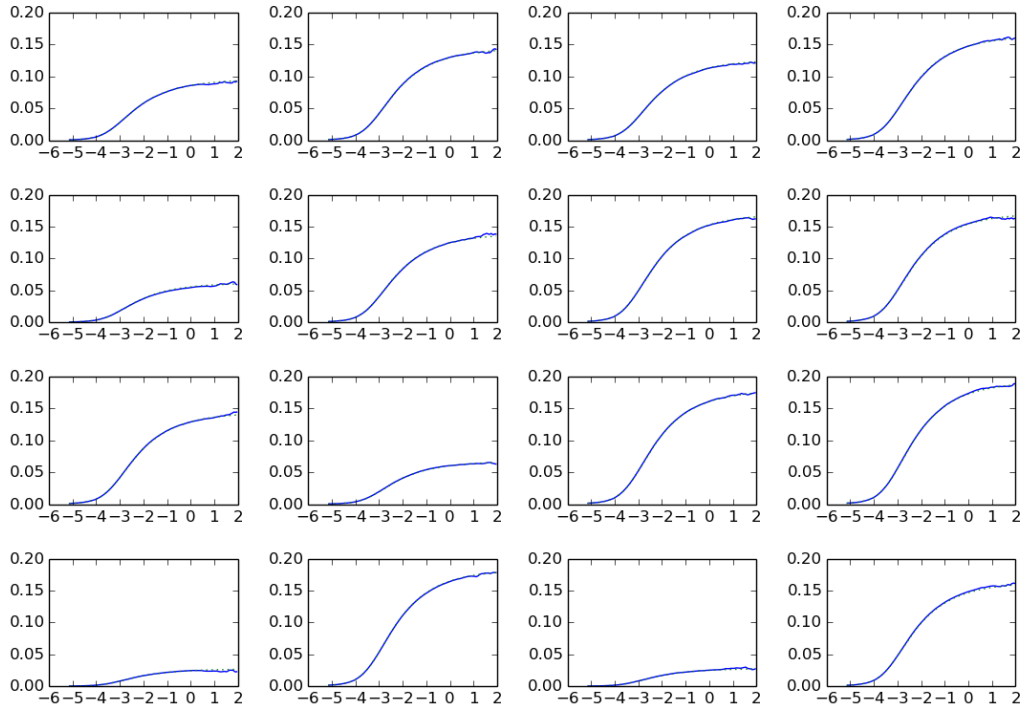


Figure V.27 – Cumulated theoretical kernels (in green) and estimated kernels (in blue) as a function of $\log 10(t)$ for power law kernels.

V.D.2.2 Exponential kernels

We now consider exponential kernels[13] in dimension 4

$$\phi^{ij}(t) = a^{ij}\beta^{ij}e^{-\beta^{ij}t}$$

where the a^{ij} are chosen uniformly random between 0 and 0.5 and the β^{ij} are chosen uniformly random between 0 and 10. We also take $\mu = 0.1$ and a simulation time $T = 100000$. We then apply the estimation procedure to the simulated data with estimation parameters $h_\delta = 0.1$, $\delta = 0.05$, $T_{min} = h_{min} = 0.2$, $T_{max} = h_{max} = 20$. As before, we see that our procedure retrieves the theoretical kernel (the error is below 1%).

[13] This estimation procedure is not optimal for such kernels.

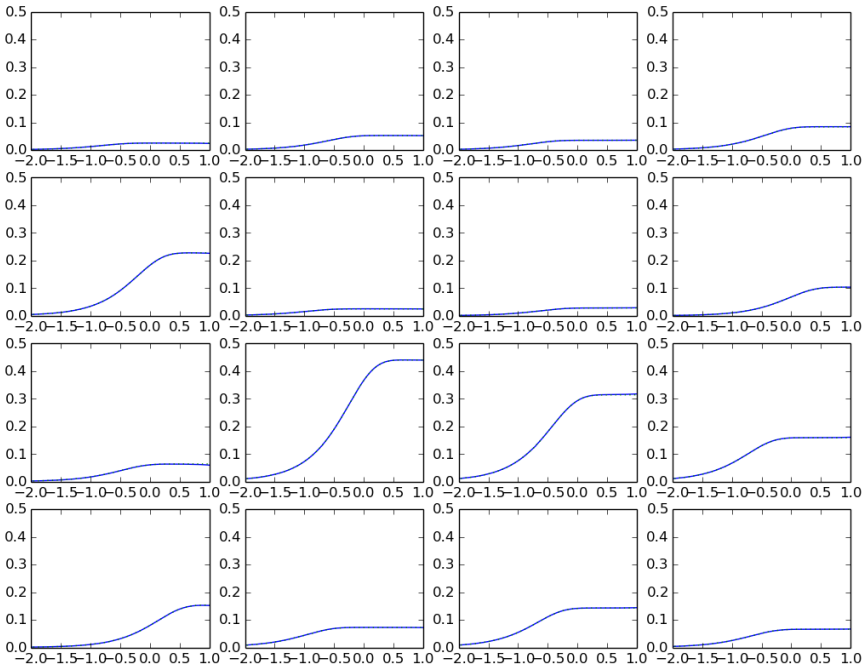


Figure V.28 – Cumulated theoretical kernels (in green) and estimated kernels (in blue) as a function of $\log 10(t)$ for exponential kernels.

V.D.2.3 Zero and negative kernels

To show that this procedure also works when the kernels are null or even negative, see [BM14b], we simulate a one dimensional exponential Hawkes process with kernel $\phi(t) = -0.03e^{-t}$ and $\mu = 1$ (so that the Hawkes intensity is (almost) never equal to zero). Again, plotting the estimated and the theoretical kernel, we get that our method retrieves the kernel.

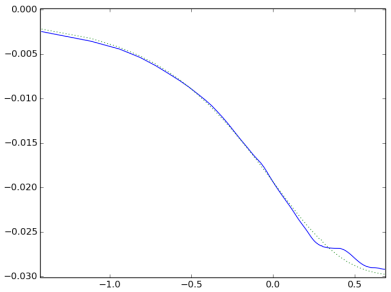


Figure V.29 – Cumulated theoretical kernel (in green) and estimated kernel (in blue) as a function of $\log 10(t)$ for a negative exponential kernel.

V.D.3 Dependence with respects to the parameters

In order to have a better understanding of where the estimation error that we make when applying our procedure comes from, let us vary the different estimation parameters and study how the error varies. Note that a heuristic to choose the optimal estimation parameters is presented in [BM14b].

To do this, we place ourselves in dimension one and we consider the Hawkes kernel

$$\phi^{theo}(t) = \frac{0.015}{(0.005+t)^{1.3}}.$$

We define the estimation error as

$$\varepsilon = \max_t \left| \int_0^t \phi^{theo}(s) - \phi^{esti}(s) ds \right| / \int \phi^{theo}$$

where ϕ^{esti} is the estimated kernel. If not specified otherwise, we chose the parameters: $T = 1000000$, $\mu = 0.1$, $h_\delta = 0.1$, $\delta = 0.03$, $T_{min} = 0.0001$, $T_{max} = 300$.

V.D.3.1 Dependence with respect to h_δ

Let us first study the influence of h_δ on ε. To do this, we plot ε as a function of h_δ for different values of δ.

We get that when h_δ is small, ε decreases with h_δ because of the noise in the conditional law and that when h_δ is large, ε increases with h_δ because the piecewise linear approximation becomes bad when the steps are too large.

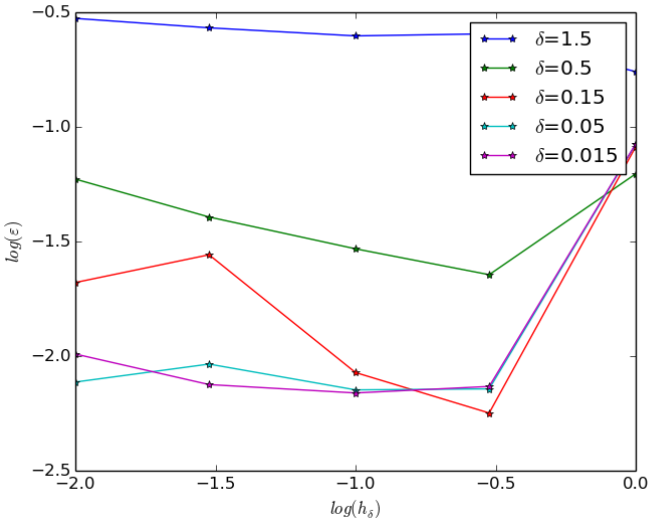


Figure V.30 – $\log 10(\varepsilon)$ as a function of $\log 10(h_\delta)$ for different numbers of quadrature points.

V.D.3.2 Dependence with respect to δ

In a similar way, we now plot ε as a function of δ for different h_δ.

We get that when h_δ is not too large, the error increases with δ.

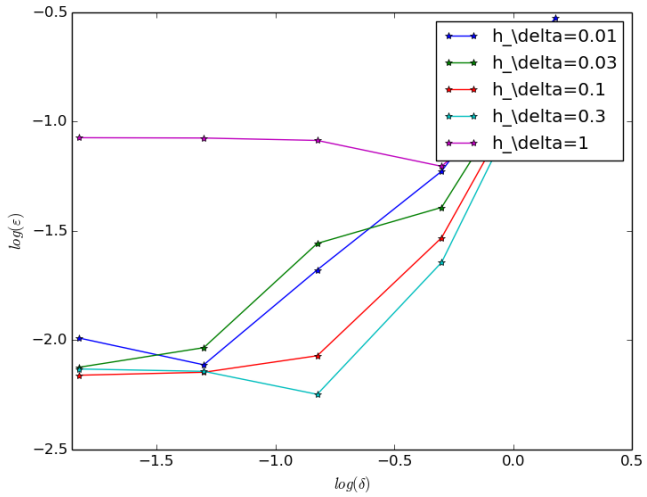


Figure V.31 – $\log 10(\varepsilon)$ as a function of $\log 10(\delta)$ for different values of h_δ.

Note that around our choice of parameters, the estimation error does not depend much on the parameters and our results are thus robust to (reasonable) changes of estimation parameters.

CHAPTER VI

Parametric and non parametric estimation of Hawkes processes with a stochastic gradient descent

Abstract

In this paper, we show that the log-likelihood of the realization of Hawkes processes can be written as a sum on all the points of the process of terms that can be rapidly computed. This leads us to propose a stochastic gradient descent to estimate the Hawkes parameters by likelihood maximization. We show that a SVRG version of this stochastic gradient descent always outperforms usual likelihood maximization methods. We also propose a multiscale representation of Hawkes kernels which yields a fast non parametric estimation procedure of Hawkes processes.

Keywords: Hawkes processes, machine learning, convex optimization, stochastic gradient descent, likelihood maximization, slowly decreasing.

1 Introduction

Hawkes processes were introduced in [Haw71b] and are defined as point processes $(N^i)_{1\leq i\leq D}$ whose ith component's intensity writes as the sum of an exogenous constant term μ^i and a linear regression on the past of the processes

$$\lambda_t^i = \mu^i + \sum_{j=1}^{D} \phi^{ij}(t-s)dN_s^j$$

where $(\phi^{ij})_{1\leq i,j\leq D}$ are positive, causal[1] measurable functions.

The self and mutually exciting nature of these processes as well as their remarkable tractability makes them natural candidates to reproduce the cross-clusterization of different types of events

[1]Causal means that its support is included in $\mathbb{R}_+$.

that is present in many fields. For example, over the last few years Hawkes processes have been successfully applied to seismology, see [Oga99], neurobiology, see [RBRGTM14], or sociology [LM11]. Moreover, a nice feature of Hawkes processes is that under their cluster representation, see [HO74], the L^1 norm of ϕ^{ij}, $|\phi^{ij}|$ corresponds to the average number of points of i directly "generated" by a point of j, see Chapter V for a discussion on this topic. This property implies that it is tempting to use the Hawkes framework to estimate the kernels in order to measure the causality between different kinds of events.

The estimation of Hawkes kernels and intensities from empirical data has been the subject of many studies. There are two main classes of estimation methods: The first one, see Chapter V or [BM14b] relies on the resolution of the Wiener-Hopf Equation which links the kernels to the conditional laws of Hawkes processes. The second class of methods which is the most frequently used relies on some likelihood maximization. The most usual parametrization of Hawkes processes that is used to apply likelihood maximization is exponential one:

$$\phi^{ij}(x) = a^{ij}\beta e^{-\beta x}.$$

We will see that, in many cases, this parametrization is very convenient to obtain efficient algorithms. Moreover the intensity of exponential Hawkes processes is a Markov process. Notice that sometimes, a power law parametrization is used, see [HBB13].

When the dimension D of the problem increases (say above 100) and Newton methods become impractical to minimise the likelihood[2], the most popular method to minimise the likelihood is the Majoration Minimization (MM) method, see [LM11] or [ZZS13]. Let us also mention that in many cases, one does not simply wish to minimise the likelihood but introduces penalizations to take advantage of prior knowledge on the sparse or low rank structure of the process, see [BGM15] or [ZZS13]. This paper is focused on accelerating the speed of the maximization algorithm so such penalizations are outside the scope of this work. However, there are no reasons why one should not be able to combine the algorithm presented below with penalizations.

In this paper, we use the particular structure of the Hawkes exponential log-likelihood to introduce a stochastic gradient descent in Section 2. We show on simulated data in Section 3 that with a slight modification, this algorithm is consistently faster than the usual MM algorithms. We also propose in Section 4 a non parametric representation of Hawkes kernels to show this algorithm can be applied to non exponential Hawkes processes and even to slowly decreasing Hawkes kernels. We conclude in Section 5.

[2]Because the Hessian is a $D(D+1) \times D(D+1)$ matrix.

2 The stochastic gradient descent

2.1 Rewriting the Hawkes likelihood

Given a realization of D point processes $((t_k^i)_{k=1,\dots,N^i})_{i=1,\dots,D}$, the log-likelihood of a point process on this realization is, see [DVJ02]

$$L = \sum_{i=1}^{D}\sum_{k=1}^{N^i} \log(\lambda^i_{t_k^i}) - \sum_{i=1}^{D}\int_0^T \lambda_s^i ds.$$

In the case of a Hawkes model given by the coefficients a^{ij} and μ^i, this rewrites,

$$L(a,\mu) = \sum_{i=1}^{D}\sum_{k=1}^{N^i} \log(\mu^i + \sum_{j=1}^{D} a^{ij}\beta \int_0^{t_k^i} e^{-\beta(t_k^i - s)} dN_s^j) - \sum_{i=1}^{D}\int_0^T \Big(\mu^i + \sum_{j=1}^{D} a^{ij}\beta\int_0^s e^{-\beta(s-u)} dN_u^j\Big) ds.$$

Notice that if β is fixed, $L(a,\mu)$ is convex, see [LV14], but if β is varying this is no longer the case and the optimization becomes much more complicated. However, a priori, this does not write as a sum of independent terms (or at least terms which can be computed rapidly) and thus one cannot apply a stochastic gradient ascent to maximise the likelihood.

The basis of our algorithm is to notice that $L(a,\mu)$ rewrites as

$$L(a,\mu) = \sum_{i=1}^{D}\sum_{k=1}^{N^i} \Big[\log\Big(\mu^i + \sum_{j=1}^{D} a^{ij} g^{ij}(t_k^i)\Big) - \mu^i(t_k^i - t_{k-1}^i) - \sum_{j=1}^{D} a^{ij}\big(G^{ij}(t_k^i) - G^{ij}(t_{k-1}^i)\big)\Big] \quad (1)$$

where

$$g^{ij}(t) = \sum_{t_{k'}^j \le t} \beta e^{-\beta^(t - t_{k'}^j)}$$

and

$$G^{ij}(t) = \int_0^t g^{ij}(s) ds.$$

The first step of our algorithm is to precompute $(g^{ij}(t_k^i))$ and $(G^{ij}(t_k^i))$ which can be done in a time $O(N)$, where $N = \sum_{i=1}^{D} N^i$ is the total number of points, see [LV14] using the relations:

$$g^{ij}(t_k^i) = g^{ij}(t_{k-1}^i) e^{-\beta(t_k^i - t_{k-1}^i)} + \sum_{t_{k-1}^i < t_{k'}^j < t_k^i} \beta e^{-\beta(t_k^i - t_{k'}^j)}$$

and

$$G^{ij}(t_k^i) = G^{ij}(t_{k-1}^i) + g^{ij}(t_{k-1}^i)\frac{1 - e^{-\beta(t_k - t_{k-1})}}{\beta} + \sum_{t_{k-1}^i < t_{k'}^j < t_k^i} (1 - e^{-\beta(t_k^i - t_{k'}^j)}).$$

2.2 The algorithm

The likelihood thus appears as a sum over the total number of points of weakly dependent terms

$$L(a,\mu) = \sum_{i=1}^{D}\sum_{k=1}^{N^i}\Big[F(a,\mu,i,k)\Big]$$

where

$$F(a,\mu,i,k) = \log\Big(\mu^i + \sum_{j=1}^{D} a^{ij} g^{ij}(t_k^i)\Big) - \mu^i(t_k^i - t_{k-1}^i) - \sum_{j=1}^{D} a^{ij}\big(G^{ij}(t_k^i) - G^{ij}(t_{k-1}^i)$$

whose derivatives are

$$\frac{\partial F(a,\mu,i,k)}{\partial a^{i'j}} = \mathbb{I}_{i=i'}\Big(\frac{g^{ij}(t_k^i)}{\mu^i + \sum_{j'=1}^{D} a^{ij'} g^{ij'}(t_k^i)} - (G^{ij}(t_k^i) - G^{ij}(t_{k-1}^i))\Big)$$

and

$$\frac{\partial F(a,\mu,i,k)}{\partial \mu^{i'}} = \mathbb{I}_{i=i'}\Big(\frac{1}{\mu^i + \sum_{j'=1}^{D} a^{ij'} g^{ij'}(t_k^i)} - (t_k^i - t_{k-1}^i)\Big).$$

This decomposition of the likelihood as a sum of almost independent and identically distributed random variables suggests the following SGD algorithm to find the optimal a and μ.

Algorithm 1 SGD

$n \leftarrow 0$
while $Converged = False$ **do**
  $n \leftarrow n+1$
  Choose at random (i,k) in $1,...,D$ and $1,...,N^i$
  Compute $s = \mu^i + \sum_{j'=1}^{D} a^{ij'} g^{ij'}(t_k^i)$
  $\mu^i \leftarrow \mu^i + \gamma_n(\frac{1}{s} - (t_k^i - t_{k-1}^i))$
  for $j \leftarrow 1$ to D **do**
    $a^{ij} \leftarrow a^{ij} + \gamma_n(\frac{g^{ij}(t_k^i)}{s} - (G^{ij}(t_k^i) - G^{ij}(t_{k-1}^i)))$
  end for
end while

2.3 SVRG modification

We will see in the next section on simulated data that, in its current form, this algorithm does not systematically outperforms state-of-the-art optimization algorithm such as the MM algorithm. However, this representation enables us to apply all the methods in the SGD literature such as the SVRG algorithm, see [JZ13], that we describe bellow. We will see that this algorithm will almost always outperforms the MM algorithm.

Algorithm 2 SVRG

$n \leftarrow 0$
while $Converged = False$ **do**
  $n \leftarrow n+1$
  if n% N==0 **then**
    $\tilde{a} \leftarrow a$
    $\tilde{\mu} \leftarrow \mu$
    for $i, j \leftarrow 1$ to D **do**
      Compute $\frac{\partial L}{\partial a^{ij}}(\tilde{a}, \tilde{\mu})$ and $\frac{\partial L}{\partial \mu^i}(\tilde{a}, \tilde{\mu})$
    end for
  end if
  Choose at random (i, k) in $1, ..., D$ and $1, ..., N^i$
  Compute $s = \mu^i + \sum_{j'=1}^{D} a^{ij'} g^{ij'}(t_k^i)$
  Compute $\tilde{s} = \tilde{\mu}^i + \sum_{j'=1}^{D} \tilde{a}^{ij'} g^{ij'}(t_k^i)$
  $\mu^i \leftarrow \mu^i + \gamma_n(\frac{1}{s} - (t_k^i - t_{k-1}^i) - \frac{1}{\tilde{s}} - (t_k^i - t_{k-1}^i) + \frac{1}{N^i}\frac{\partial L}{\partial \mu^i}(\tilde{a}, \tilde{\mu}))$
  for $j \leftarrow 1$ to D **do**
    $a^{ij} \leftarrow a^{ij} + \gamma_n(\frac{g^{ij}(t_k^i)}{s} - (G^{ij}(t_k^i) - G^{ij}(t_{k-1}^i)))$
  end for
end while

Remark 1. *An important property of the SGD and SVRG algorithms presented above is that the estimations of* $((a^{ij})_j, \mu^i)$ *and* $((a^{i'j})_j, \mu^{i'})$ *are independent for* $i \neq i'$. *Therefore, these algorithms can be entirely parallelized into* D *independent calculations. In practice, this is very important for speed and memory concerns.*

3 Performance on simulated data

3.1 The Majoration-Minimization estimation method

In this paragraph, we rapidly describe the Majoration-Minimization algorithm presented in [ZZS13] which is one of the most popular estimation procedure for high dimensional Hawkes processes[3]. We will then compare the convergence speed of this method with our SGD and SVRG gradient descents.

The MM algorithm consists in two steps.

- In the majoration step, the log-likelihood is bounded by Q a surrogate function of a, μ

[3]Note that the MM algorithm is in fact equivalent to the Expectation-Maximization (EM) algorithm of [HDB13].

and the previous estimates of $\tilde{a}$ and $\tilde{\mu}$

$$\begin{aligned} L(a,\mu) \leq Q(a,\mu,\tilde{a},\tilde{\mu}) &= \sum_{i=1}^{D}\sum_{k=1}^{N^i}\Big[p_k^{i0}\log(\frac{\mu^i}{p_k^{i0}}) + \sum_{j=1}^{D} p_k^{ij}\log\Big(\frac{a^{ij}g^{ij}(t_k^i)}{p_k^{ij}}\Big) \\ &- \mu^i(t_k^i - t_{k-1}^i) - \sum_{j=1}^{D} a^{ij}\big(G^{ij}(t_k^i) - G^{ij}(t_{k-1}^i)\big)\Big] \end{aligned}$$

where

$$p_k^{ij} = \frac{\tilde{a}^{ij}g^{ij}(t_k^i)}{\tilde{\mu}^i + \sum_{j=1}^{D}\tilde{a}^{ij}g^{ij}(t_k^i)}$$

and

$$p_k^{i0} = \frac{\tilde{\mu}^i}{\tilde{\mu}^i + \sum_{j=1}^{D}\tilde{a}^{ij}g^{ij}(t_k^i)}.$$

- In the minimization step, the surrogate function is minimised by taking the closed form results

$$\mu^i = \frac{\sum_{k=1}^{N^i} p_k^{i0}}{T}$$

and

$$a^{ij} = \frac{\sum_{k=1}^{N^i} p_k^{ij}}{G^{ij}(T)}.$$

Finally, the estimates are modified $\tilde{a} = a$ and $\tilde{\mu} = \mu$ and the algorithm loops.

3.2 Performance comparison

Let us now simulate Hawkes processes with different parameters using the Ogata thinning algorithm, see [Oga81], and apply the three different estimation procedure described above (SGD, SVRG and MM). We will compare the performance of these algorithms using the L^1 error function

$$\varepsilon = \sum_{i,j}|a^{ij} - \tilde{a}^{ij}| / \sum_{i,j} a^{ij}$$

where a is the "real Hawkes matrix" (used for the simulation) and $\tilde{a}$ is the estimated Hawkes matrix.

As an example of Hawkes coefficients, we consider matrices of different "communities" where each individual interacts with all the other individuals of his community (including himself) but not with individuals of other communities. In other words, the matrix a is equal to zero everywhere except on squares on the diagonal. The sizes of the communities are: 1, 2, ..., c where c is the number of communities (therefore the dimension is $D = c(c+1)/2$). All the intensities of the interactions are equal and the matrix is renormalized so that its maximum eigenvalue is equal to 0.8. For example, on Figure VI.1, we plot a for different values of c (and thus of D). We also chose $\beta = 1$ and $\mu = 1$.

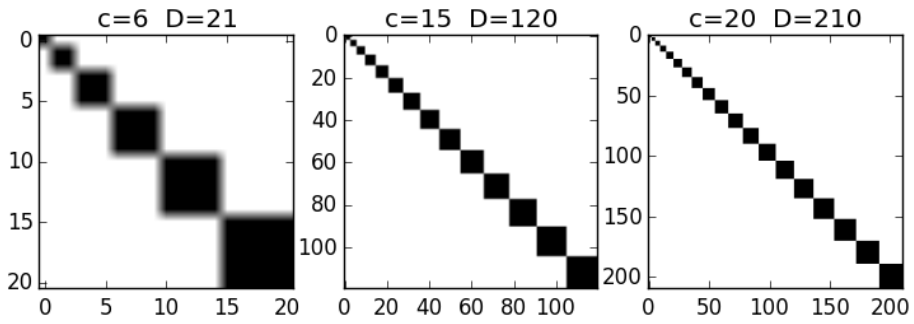


Figure VI.1 – The matrix a for different values of c and D: Left: $c = 6$ and $D = 21$. Center: $c = 15$ and $D = 120$. Right: $c = 20$ and $D = 210$. The white corresponds to 0 and the black corresponds to a number which is fixed for every matrix.

In the following paragraphs, we compare the performance of our algorithms for different values of c (and thus D) and T. To do this, we plot $\log(\varepsilon)$ as a function of $\log(t)$ where t is the execution time.

We get that the SVRG algorithm always outperforms the state-of-the-art MM algorithm. This is especially true when the number of points and the dimension increase.

3.2.1 $D = 21$ $(c = 6)$

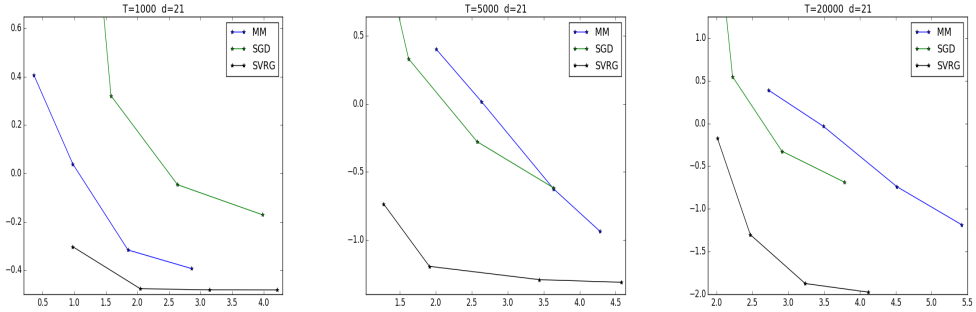


Figure VI.2 – $\log(\varepsilon)$ as a function of the execution $\log(t)$ for $D = 21$ and $T = 1000$ $(N/D = 2980)$ (left), $T = 5000$ $(N/D = 15000)$ (middle) and $T = 20000$ $(N/D = 5900)$ (right).

3.2.2 $D = 120$ $(c = 15)$

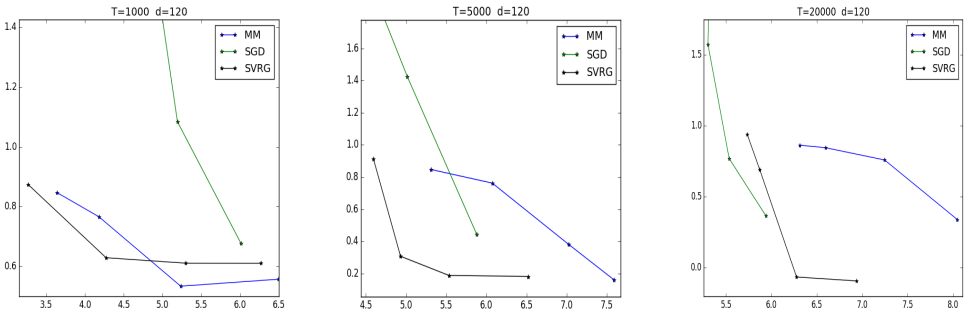


Figure VI.3 – $\log(\varepsilon)$ as a function of the execution $\log(t)$ for $D = 120$ and $T = 1000$ ($N/D = 2700$) (left), $T = 5000$ ($N/D = 13500$) (middle) and $T = 20000$ ($N/D = 54000$) (right).

3.2.3 $D = 210$ $(c = 20)$

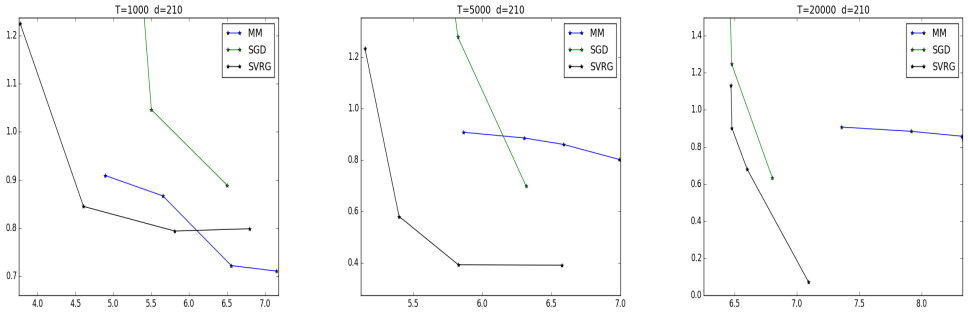


Figure VI.4 – $\log(\varepsilon)$ as a function of the execution $\log(t)$ for $D = 210$ and $T = 1000$ ($N/D = 2670$) (left), $T = 5000$ ($N/D = 13300$) (middle) and $T = 20000$ ($N/D = 53000$) (right).

4 Multiscale Hawkes processes

An important practical issue with the above exponential parametrization of Hawkes kernels is that in practice, there is no reason why the kernels should be exponential and there are even less reasons why the time scales $1/\beta$ should be known. To solve this issue, an algorithm is presented in [LV14] where the Hawkes kernels are represented as sums of exponentials:

$$\phi^{ij}(x) = \sum_{b=1}^{N_b} a_b^{ij} b\beta e^{-b\beta x}$$

where the coefficients (a_b^{ij}) must be estimated. It is shown on examples that this method has the advantage of fitting much more general kernels than the exponential representation. However, in many applications such as seismology, see [BM14b], and finance, see Chapter V, the kernels which appear asymptotically behave as power laws of exponent slightly higher than one and thus present weight at many time scales. For example, as explained in Chapter V, to reproduce the behavior of kernels at the millisecond time scale and at the second time scale, one needs to take $\beta = 1$ and $N_b = 1000$. This introduces far too many parameters to have a precise estimation with a reasonable computation time and data size.

4.1 The algorithm

To solve this issue, we propose a similar representation as that of [LV14] where the kernels appear as sums of exponentials. However, to account for the multiscaling behavior of the kernels, we take the different time scales of the exponential functions log-uniform:

$$\phi^{ij}(x) = \sum_{b=0}^{N_b} a_b^{ij} l^b \beta e^{-l^b \beta x}.$$

Notice that the SVRG and the SGD algorithms can easily be adapted to this framework. For example, the SGD algorithm now writes.

Algorithm 3 SGD Multibeta

$n \leftarrow 0$
while $Converged = False$ **do**
  $n \leftarrow n+1$
  Choose at random (i,k) in $1,...,D$ and $1,...,N^i$
  Compute $s = \mu^i + \sum_{b=1}^{N_b} \sum_{j=1}^{D} a_b^{ij} g_b^{ij}(t_k^i)$
  $\mu^i \leftarrow \mu^i + \gamma_n(\frac{1}{s} - (t_k^i - t_{k-1}^i))$
  for $j \leftarrow 1$ to D **do**
    for $b \leftarrow 0$ to N_b **do**
      $a_b^{ij} \leftarrow a_b^{ij} - \gamma_n(\frac{g_b^{ij}(t_k^i)}{s} - (G_b^{ij}(t_k^i) - G_b^{ij}(t_{k-1}^i)))$
    end for
  end for
end while

Where

$$g_b^{ij}(t) = \sum_{t_{k'}^j \le t} l^b \beta e^{-l^b \beta (t - t_{k'}^j)}$$

and

$$G_b^{ij}(t) = \int_0^t g_b^{ij}(s) ds.$$

can be computed rapidly as before.

4.2 Numerical test

To test this procedure, we apply it to a simulation of a power law Hawkes process whose shape does not depend on (i, j) and is a power law function of time

$$\phi_b^{ij}(x) = \frac{C_{ij}}{(\delta + x)^{1+\alpha}}$$

and were the C_{ij} are set so that the matrix of kernels'norms is equal to that presented at the beginning of Section 3.2. We take the parameters $\mu = 1$, $\delta = 0.02$, $\alpha = 0.65$, $c = 6$ (so that $D = 21$) and $T = 20000$ ($N/D = 59300$).

To estimate the parameters, we take $\beta = 0.05$, $l = 3$, $N_b = 9$, so that $\beta l^{N_b} = 984$ and we use the SVRG algorithm with a number of iteration large enough so that the L^1 error "saturates". We present the theoretical and estimated cumulated kernels $\int_0^t \phi^{ij}(s)ds$ as a function of $\log 10(t)$ in Figure VI.5. We see that in spite of the relatively low number of exponentials in the kernel representation (10 in total and only 2 per decades), we get a very good fit of the cumulated kernels[4] over multiple decades.

5 Conclusion and extensions

We have shown that by pre computing the right terms the log-likelihood of Hawkes processes could be written as a sum of almost independent terms. Therefore, one can apply methods in the literature on stochastic gradient descents. In particular, the SVRG algorithm systematically outperforms the classical MM algorithm used to estimate Hawkes processes in high dimensions. We have also shown that this algorithm could be parallelized and applied to general kernels shape and even to multiscale kernels. Another important advantage of this procedure is that it can easily be applied to more general Hawkes processes. In particular, it is possible to take into accounts marks, a varying μ and penalizations.

[4] It is explained in Chapter V that for multiscale Hawkes kernels, plotting the integral of the cumulated kernel $\int_0^t \phi^{ij}(s)ds$ as a function of $\log(t)$ is a natural representation of the kernels.

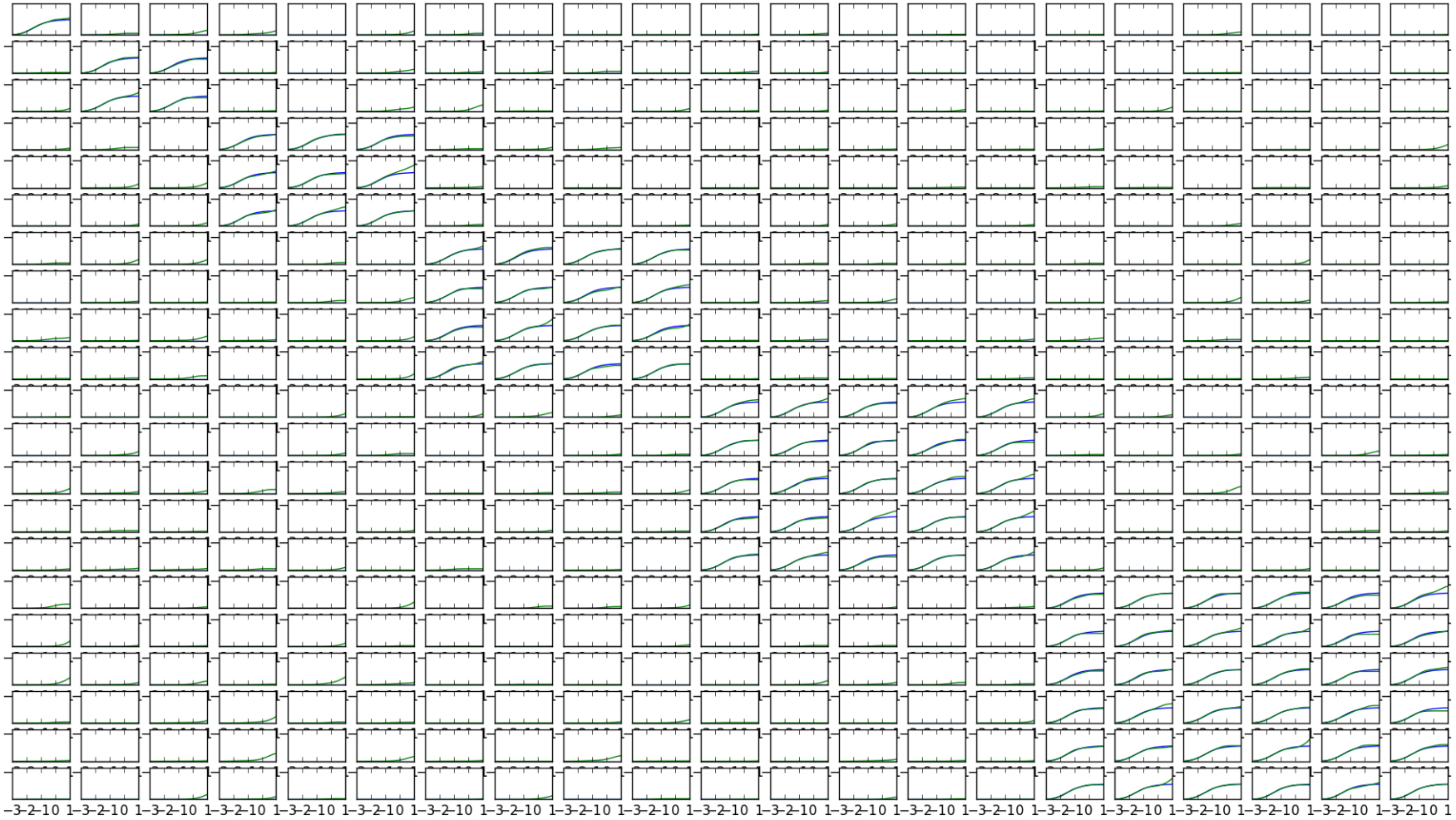


Figure VI.5 – $\int_0^t \phi^{ij}(s)ds$ as a function of $\log_{10}(t)$ for the theoretical kernel (blue) and the estimated kernel (green) for t from 10^{-3} to 10.

Part IV

Market impact throughout time scales

CHAPTER VII

Liquidity and impact in fair markets

Abstract

We develop a theory which applies to any market dynamics that satisfy a fair market assumption on the nullity of the average profit of simple market making strategies. We show that for any such fair market, there exists a martingale fair price which corresponds to the average liquidation value (at the ask or the bid) of an infinitesimal quantity of stock. We show that this fair price is a natural reference price to compute the ex post gain of limit orders. Using only the fair market assumption, we link the spread to the impact of market orders on the fair price. We use our definition of the fair price to build empirical tests of the relevance of this notion whose results are consistent with our theoretical predictions.

Keywords: Fair price, market orders, limit orders, bid-ask spread, liquidity, price impact, market efficiency.

1 Introduction

One can argue that the main aim of financial markets is to give investors the possibility to balance their portfolio as cheaply and as quickly as possible. To do that, most modern financial exchanges allow their participants to buy or sell assets on an electronic order book via a continuous double auction: Throughout the day, participants can either use limit orders which enable them to buy or sell a given number of shares at a given price but with no guarantee of execution, thus adding liquidity to the order book, or they can use market orders which consume available limit orders thus taking liquidity from the order book. Understanding the nature and dynamics of this liquidity is of great importance for practitioners who wish to improve the efficiency of their strategies. For example, brokers who execute large trades need to split them[1] into many transactions and to wait for new liquidity to arrive between their orders. It is also crucial for regulators and exchanges who aim at improving the global

[1]Such large trades are typically much larger than the liquidity available in the order book.

efficiency of the market, for example by tuning the tick value[2] or transaction costs.

Several statistical models of order books have been proposed to understand the dynamics of liquidity. Among them, we can cite "zero intelligence" models, see [CDL13], [SFGK03] or [Roş09b], which assume that market and limit orders arrive at Poisson rates. More recently, this topic has been tackled using mean field games, see [LLL+13] or general Markovian dynamics, see [HLR13]. In Chapter V, the arrival of these orders is modeled as Hawkes processes whose kernels are empirically fitted to reproduce the causality between the different events. However, in most of these models, simple market making strategies can yield significant profits on average which we will see is not compatible with empirical computations, especially for small tick assets[3]. In this work, we do not give a complete modeling of order book dynamics but rather conditions that reasonable models should satisfy. We show that given the impact of market orders on a fair price that we will define, we can derive the bid-ask spread dynamics.

The most common measure of liquidity is the spread[4]. The links between the spread and other market variables (such as the volatility, the daily volume, the risk aversion of market makers or the transaction costs) have already been the subject of much literature. The classical view is that limit orders are mainly posted by market makers who sell at the ask and buy at the bid and thus "earning the spread", while market orders are used by other agents which can be seen as brokers or investors who want to execute a number of shares for various reasons (investment strategies, hedging, regulatory constraints,...).

In the literature, three main reasons are proposed to explain the size of the spread: The first one is order processing costs, see [HS97]. However, in modern electronic financial markets, these costs are negligible compared to the spread[5]. The second one is the market makers' risk aversion (see for example [AS08] and [GLFT12] where the average return of market makers is balanced by the inventory risk they take on). The third reason, which we will develop in the next paragraphs, is adverse selection. Throughout this work, we perform empirical measures which seem to show that this reason is sufficient to explain most of the spread behavior.

Let us now explain where this adverse selection comes from. When an institutional investor executes a buy metaorder[6], the price on average goes up. This phenomenon, called the market impact of metaorders, has been subject to many empirical and theoretical studies, see [ATHL05], [FGLW13], [Kyl85] or [MVM+09]. The impact of metaorders can be understood in two different ways: It can either be viewed as informational in the sense that investors have a better information on the future of the price and trade because of this information, thus revealing the price, see [FGLW13] or [Kyl85]. Market impact can also be seen as mechanical in

[2] On electronic markets, assets cannot be exchanged at any price but only at certain discrete prices. The tick value of an asset is the discretization step, see [DR12].

[3] A small tick asset is an asset whose average spread is significantly larger than the tick.

[4] The spread is the difference between the price of the lowest sell limit order (ask price) and the price of the highest buy limit order (bid price) in the order book.

[5] At least ten times smaller according to [WBK+08].

[6] A metaorder is a large transaction incrementally executed by an investor, see [FGLW13].

the sense that even uninformed transactions move prices[7] through the imbalance in supply and demand they create.

Whatever the reason for this impact, during the execution of their metaorders, investors mostly use market orders and therefore, the impact of metaorders is translated onto the order flow into an impact of market orders. This impact has also been the subject of many empirical and theoretical studies, see [BHS95], [BGPW04], [EBK12] or [MRR97]. Because of this impact that market orders have on the price, the expected ex post gain of a limit order with respect to the mid price is not equal to the half spread: It is on average equal to the half spread minus the average market impact of a market order on the mid price. This phenomenon is called adverse selection.

Theoretical studies using adverse selection to explain the bid-ask spread are often rooted to [GM85], where the bid and ask prices are set by market makers so that the expected ex post gain of their limit orders with respect to the fundamental value of the stock is equal to zero. In [MRR97], the authors propose a simple $AR(1)$-type order flow model where the impact on a "fair price" is proportional to the surprise in the order flow[8] so that the fair price is a martingale. In this model, bid and ask prices are set so that the expected ex post gain of a limit order with respect to the fair price is equal to zero. In [WBK$^+$08], Bouchaud et al. apply these ideas to the more realistic FARIMA volume model, see [Ber94]. Assuming that most of price moves are due to volume impact and not exogenous information, they obtain nice relationships between the spread, market impact and the volatility per trade which are consistent with empirical measures. This theory has been extended to large tick assets in [DR12].

In most of the models cited above and in many other studies of the behavior of liquidity, spread and impact, there exists an unobservable reference price (called for example fundamental value [MRR97], efficient price [DRR13], [RR11] or fair price [GLFT12]). This price often represents some martingale underlying value estimated by market participants based on the available information of the stock based on the available information and is used as a reference price to compute ex post gains. However, it is often not clear what this price really means and why it is compared to this price that market making strategies should not have positive ex post gains. In this paper, we assume that market dynamics are given (that is a model for limit and market orders) where limit orders can be posted at any price (in practice this corresponds to small tick assets). We will see that in this framework, we in fact do not need to postulate the existence of a reference price. Instead, we only assume that our market dynamics satisfy a fair market assumption which states that *infinitesimal market making strategies are not profitable on average.*

Our first aim is to show that for any such model, there exists a martingale fair price which generalises that of [MRR97]. This fair price can be thought of as the expected price for buying or selling an infinitesimal quantity of stock using limit orders. We will see that this fair price

[7] See Chapter VIII for a more detailed discussion on this topic.

[8] That is the sign of the next order minus its conditional expectation, see below.

has nice properties which make it a natural reference price to compute ex post gains. Our second aim is to show that with no other assumptions, we can compute relations between the impact of market orders on this fair price and the liquidity in the order book. We revisit important theoretical results on the spread in light of our fair price. We show that even though on data we cannot compute independently of a model the impact function of market orders on the fair price or even the value of the fair price, we can estimate average ex post gains compared to our fair price. We obtain that our theoretical predictions hold empirically.

This paper is organized as follows. In Section 2, we specify our no arbitrage hypothesis and show the existence of a model independent fair price. We propose a way to empirically check the existence of our fair price that we apply to real data. In Section 3, we give a general definition of the market impact of market orders. Using our fair price as a reference price instead of the mid price, we show that we can easily compute the bid and ask prices from the fair price dynamics. Furthermore, we empirically check that adverse selection and tick value explain most the bid-ask spread dynamics. We conclude in Section 4.

2 Fair market and fair price

In this section, we present our framework and our fair market assumption. We notably show the first of its implications: The existence of a model independent fair price that can be used to compute the ex post gains of limit orders.

2.1 General framework

Long term investors rebalance their portfolios using metaorders. At a given time, there is a priori no reason why the flows of buy and sell metaorders should match. Therefore, for the market to be fluid, that is for investors to always find counterparts to execute their trades, there must be intermediaries, called market makers, who provide liquidity at the bid and at the ask in the order book. Since there are many market makers, competition between them should imply that the spread and the liquidity are set so that simple market making strategies are at most marginally "profitable" (the profitability of a strategy remaining to be defined). Indeed, if a simple market making strategy were profitable, other market makers would place limit orders just in front of the limit orders of the strategy thus making it less profitable until it is no longer worth doing so, see [WBK+08].

In what follows, we assume that market makers are risk neutral (in practice, this can be justified by the small time scales of market making strategies) and order processing costs are null (in practice negligible). Therefore, the profitability of a market making strategy is defined as its average profit.

The concept of "average" seen as expectation being undefined outside of a model, we assume that we are given market dynamics. This means we have a probabilistic model of buy and sell market orders and liquidity functions that we define below.

Definition 1. *We set v_t^A (resp. v_t^B) as the volume of buy (resp. sell) market orders at time t: If there is a buy (resp. sell) market order of volume v at time t then $v_t^A = v$ (resp. $v_t^B = v$) else $v_t^A = 0$ (resp. $v_t^B = 0$). The processes v_t^A and v_t^B are assumed to be equal to zero except over a finite number points over any finite time interval.*

Definition 2. *We set Liq_t^{ask} (resp. Liq_t^{bid}) as the ask (resp. bid) cumulated liquidity function at time t: For any price P, $Liq_t^{ask}(P)$ (resp. $Liq_t^{bid}(P)$) is the sum of the volumes of all ask (resp. bid) limit orders whose prices are smaller (resp. higher) than P at time t. These two processes are assumed to be right continuous with left limits.*

Definition 3. *We set*

$$a_t = argmin_x\{Liq_t^{ask}(x) > 0\} \text{ (resp. } b_t = argmax_x\{Liq_t^{bid}(x) > 0\})$$

as the ask (resp. bid) price at time t.

Definition 4. *The market M is defined as the process:*

$$M_t = (\sum_{s\leq t} v_s^A, \sum_{s\leq t} v_s^B, Liq_t^{ask}, Liq_t^{bid})$$

which is right continuous with left limits in time. We write $\mathcal{F} = (\mathcal{F}_t)_t$ for the natural filtration of M and $\mathcal{F}^*$ for the natural filtration of the left continuous (in t) version[9] of M.

Note that $\mathcal{F}$ and $\mathcal{F}^*$ are defined so that $\mathcal{F}_t^* \subset \mathcal{F}_t$ and if there is a trade (or any order book event) at time t, it is known from $\mathcal{F}_t$ but not from $\mathcal{F}_t^*$.

We also choose the convention that if a market maker sets a limit order at time t and if a market order arrives at time t, then the limit order is executed (provided price priority is respected).

In this work, we take the tick value equal to zero (that is limit orders can be posted at any price). In practice, this assumption corresponds to small tick assets.

Definition 5. *An infinitesimal market making strategy is a strategy that uses only very small*[10] *limit orders and that ends with no inventory.*

Instead of making an assumption on the nullity of the ex post gain of limit orders compared to an arbitrary price, we directly make an assumption on the profit of market making strategies. This fair market hypothesis on the market dynamics can be understood as a perfect competition hypothesis between market makers:

Assumption 1. *An infinitesimal market making strategy does not affect the market dynamics and the expected profit of an infinitesimal market making strategy is zero.*

[9] The necessity to introduce such a filtration will appear in Section 3.

[10] Say of one share, so that this strategy does not impact the rest of the market.

We use this formal assumption on the market dynamics to define a *fair price* which will be the expectation of the price at which an agent can buy or sell an infinitesimal quantity of stocks using limit orders. Our fair price can be viewed as a generalization of the fair price of the model of [MRR97] that we will present in the next paragraph as an illustration of our fair market framework.

Remark 1. *In [Don12], the same kind of arguments about perfect competition between market makers is used to compute properties of the impact of metaorders.*

2.2 The example of the MRR model

In this subsection, we present the Madhavan Richardson Roomans (MRR for short) market model introduced in [MRR97], which is a fair market model by construction. We compute the expectation of what we call the next bid and ask prices after time n. This simple computation illustrates how a fair price appears in any fair market.

2.2.1 The MRR model

The MRR model is a simple impact model in which orders are autocorrelated but prices are martingale. Moreover, the ask (resp. bid) price is defined as the expected future fair price if there is a buy (resp. sell) order. Therefore, by construction, the ex post gain of limit orders with respect to the martingale fair price is null and Assumption 1 is satisfied.

In this discrete time model, the n^{th} market order is either a unit volume buy ($\varepsilon_n = +1$) or sell ($\varepsilon_n = -1$) order and the sign process is an $AR(1)$-type process:

$$\mathbb{E}[\varepsilon_n|\mathscr{F}_{n-1}] = \rho\varepsilon_{n-1}$$ [11]

with $0 < \rho < 1$.

The market impact of an order on the fair price p is proportional to the surprise of the order flow in the sense that:

$$p_{n+1} - p_n = \theta(\epsilon_n - \mathbb{E}[\epsilon_n|\mathscr{F}_{n-1}]) + \zeta_n = \theta(\epsilon_n - \rho\epsilon_{n-1}) + \zeta_n$$

with $\theta > 0$ and $\mathscr{F}_{n-1}$ is the sigma algebra generated by $(\varepsilon_k)_{k<n}$ and $(p_k)_{k\leq n}$ and where ζ is an independent white noise corresponding to the effect of outside information on the price that we set here to zero to simplify.

In this model, imposing an ex post gain equal to zero for bid and ask limit orders gives the ask price

$$a_n = p_n + \theta(1 - \rho\epsilon_{n-1}),$$

[11] Therefore, $\mathbb{P}[\varepsilon_n = +1|\mathscr{F}_{n-1}] = (\rho\varepsilon_{n-1} + 1)/2$.

the bid price

$$b_n = p_n + \theta(-1 - \rho\epsilon_{n-1})^{12}$$

and thus the spread

$$\Phi_n = a_n - b_n = 2\theta.$$

2.2.2 Next bid and ask prices

In the MRR model, bid and ask prices are taken so that the (expected) ex post gain of limit orders with respect to the fair price is equal to zero. Let us now show that conversely, given the dynamics of the ask (or bid) price, we can redefine the fair price. To do this, we need to introduce the following quantities:

Definition 6. *The next ask (resp. bid) price after time n denoted by na_n (resp. nb_n) is the price of the first ask market order after n (n included). More formally:*

$$na_n = a_{\inf\{k\geq n;\epsilon_k=+1\}} \text{ and } nb_n = b_{\inf\{k\geq n;\epsilon_k=-1\}}.$$

Given the dynamics of (a, b, ϵ), we can derive the fair price as the conditional expectation of the next ask or the next bid price (since they are equal).

Proposition 1. *We have $\mathbb{E}[na_n|\mathscr{F}_{n-1}] = \mathbb{E}[nb_n|\mathscr{F}_{n-1}] = p_n$.*

Proof. Because of the Markov property of the sign process and the invariance of the model with respect to translation of the fair price, the conditional expectation of the next bid price writes as

$$\mathbb{E}[nb_n|\mathscr{F}_{n-1}] = p_n + f^b(\varepsilon_{n-1})$$

for some deterministic function f^b. Let us decompose this expectation:

$$\mathbb{E}[nb_n|\mathscr{F}_{n-1}] = \mathbb{E}[\mathbb{I}_{\varepsilon_n=+1} nb_{n+1}|\mathscr{F}_{n-1}] + \mathbb{E}[\mathbb{I}_{\varepsilon_n=-1} b_n|\mathscr{F}_{n-1}]$$

and use that:

- b_n is in $\mathscr{F}_{n-1}$ so that

$$\mathbb{E}[\mathbb{I}_{\varepsilon_n=-1} b_n|\mathscr{F}_{n-1}] = b_n\mathbb{P}[\varepsilon_n = -1|\mathscr{F}_{n-1}].$$

- By the tower property of conditional expectations

$$\begin{aligned}\mathbb{E}[\mathbb{I}_{\varepsilon_n=+1} nb_{n+1}|\mathscr{F}_{n-1}] &= \mathbb{E}[\mathbb{I}_{\varepsilon_n=+1}\mathbb{E}[nb_{n+1}|\mathscr{F}_{n-1}, \varepsilon_n = +1]|\mathscr{F}_{n-1}] \\ &= \mathbb{P}[\varepsilon_n = +1|\mathscr{F}_{n-1}]\mathbb{E}[nb_{n+1}|\mathscr{F}_{n-1}, \varepsilon_n = +1] \\ &= \mathbb{P}[\varepsilon_n = +1|\mathscr{F}_{n-1}](p_n + \theta(1 - \rho\varepsilon_{n-1}) + f^b(+1)).\end{aligned}$$

[12] Note than a_n and b_n are in $\mathscr{F}_{n-1}$.

We thus get that for $\varepsilon_{n-1} = \pm 1$,

$$\begin{aligned}\mathbb{E}[nb_n|\mathscr{F}_{n-1}] = p_n + f^b(\varepsilon_{n-1}) &= (p_n + \theta(-1-\rho\varepsilon_{n-1})) \times (1-\rho\varepsilon_{n-1})/2 \\ &+ (p_n + \theta(1-\rho\varepsilon_{n-1}) + f^b(+1)) \times (1+\rho\varepsilon_{n-1})/2.\end{aligned}$$

Taking $\varepsilon_{n-1} = +1$, this yields that $f^b(+1) = 0$ and then taking $\varepsilon_{n-1} = -1$, this yields that $f^b(-1) = 0$ and therefore $\mathbb{E}[nb_n|\mathscr{F}_{n-1}] = p_n$.

In the same way, we can show that $\mathbb{E}[na_n|\mathscr{F}_{n-1}] = p_n$. □

Let us now show that Proposition 1 does not come from the particular structure of the MRR model but only from the nullity of the average profit of market making strategies in the MRR model. This implies that the fair price can be generalized to any market model satisfying Assumption 1.

2.3 General definition of the fair price

In a more general way, let us assume that we have known continuous time dynamics for the ask price, the bid price and market orders and that they satisfy Assumption 1. We begin by defining in continuous time the next bid and ask prices.

Definition 7. *The next ask (resp. bid) price after time t denoted na_t (resp. nb_t) is the price of the first ask market order after (including) t. More formally:*

$$na_t = a_{\inf\{\tau \geq t; v^A_\tau > 0\}}$$

and

$$nb_t = b_{\inf\{\tau \geq t; v^B_\tau > 0\}}.$$

We have the following important result.

Proposition 2. *Under Assumption 1, we have:*

$$\mathbb{E}[na_t|\mathscr{F}^*_t] = \mathbb{E}[nb_t|\mathscr{F}^*_t].$$

Proof. If $\mathbb{E}[na_t|\mathscr{F}^*_t] > \mathbb{E}[nb_t|\mathscr{F}^*_t]$, maintaining two limit orders just in front of the best ask and the best bid prices from t until they are executed yields a positive profit on average. In the same way, if $\mathbb{E}[na_t|\mathscr{F}^*_t] < \mathbb{E}[nb_t|\mathscr{F}^*_t]$, this strategy yields a negative profit on average. □

Remark 2. *This proposition might at first sight seem unnatural because the ask price is always above the bid price but we have shown that it is true in the MRR model and we will see that it is well satisfied on data.*

We can now define the fair price.

Definition 8. *The fair price P_t is defined by:*

$$P_t = \mathbb{E}[na_t|\mathscr{F}^*_t] = \mathbb{E}[nb_t|\mathscr{F}^*_t]. \tag{1}$$

We have the first following result about the fair price.

Proposition 3. *The fair price is a $\mathscr{F}^*$-martingale:*

$$\mathbb{E}[P_{t+s}|\mathscr{F}_t^*] = \mathbb{E}[na_{t+s}|\mathscr{F}_t^*] = \mathbb{E}[na_t|\mathscr{F}_t^*] = P_t.$$

Proof. If $\mathbb{E}[na_{t+s}|\mathscr{F}_t^*] > \mathbb{E}[na_t|\mathscr{F}_t^*] = \mathbb{E}[nb_t|\mathscr{F}_t^*]$, posting a bid order in t just in front of the best bid and an ask order in $t+s$ just in front of the best ask gives a positive profit on average. □

It is thus possible to obtain the fair price from any model in which there is an ask price, a bid price and an order flow process which satisfy Assumption 1. We will see in the next section that it is possible to proceed the other way around and to get the bid and ask prices from any impact model on a fair price (as it is done in the construction of the MRR framework).

Moreover, since the fair price is a martingale, the response function associated to it defined almost as in [BGPW04] as

$$R(\delta) = \mathbb{E}[P_{t+\delta} - P_t | v_t^A > 0]$$

does not depend on δ. Indeed, for $\delta_2 > \delta_1$,

$$R(\delta_2) - R(\delta_1) = \mathbb{E}[\mathbb{E}[P_{t+\delta_2} - P_{t+\delta_1}|\mathscr{F}_{t+\delta_1}^*]|\, v_t^A > 0] = 0.$$

We will later see that this makes the fair price a natural reference price to compute impact functions.

Remark 3. *In the previous proofs, we have assumed that we could always place ourselves just in front of the ask or bid prices. In fact there is the tick value/priority issue which implies that one cannot place limit orders at any position in the order book and thus Proposition 2 should become:*

$$|\mathbb{E}[na_t|\mathscr{F}_t^*] - \mathbb{E}[nb_t|\mathscr{F}_t^*]| \leq 2\alpha$$

where α is the tick value. For many assets, we cannot neglect the tick value but in this work, we assume that it is null.

2.4 Verification of Proposition 2 on data

The aim of this paragraph is to propose a way to empirically check Proposition 2 on real data. By definition of the conditional expectation, Proposition 2

$$\mathbb{E}[na_t - nb_t|\mathscr{F}_t^*] = 0$$

implies:

$$\forall E_t \in \mathscr{F}_t^* \quad \frac{\mathbb{E}[na_t 1_{E_t}]}{\mathbb{P}(E_t)} = \frac{\mathbb{E}[nb_t 1_{E_t}]}{\mathbb{P}(E_t)}.$$

By definition of $(\mathscr{F}_t^*)$, we get that for any measurable set E in the image space where $(M_{t-y})_{y>0}$[13] takes value, denoted $\tilde{\sigma}$, we have

$$\frac{\mathbb{E}[na_t 1_E((M_{t-y})_{y>0})]}{\mathbb{P}[(M_{t-y})_{y>0} \in E]} = \frac{\mathbb{E}[nb_t 1_E((M_{t-y})_{y>0})]}{\mathbb{P}[(M_{t-y})_{y>0} \in E]}.$$

Let us now assume that the market is stationary in the sense that for every E,

$$\frac{\mathbb{E}[(na_t - nb_t) 1_E((M_{t-y})_{y>0})]}{\mathbb{P}[(M_{t-y})_{y>0} \in E]} \tag{2}$$

does not depend on t.

To empirically check that Proposition 2 is valid, we need to consider a family of past events $(E^i)_i$ (that are frequent enough) and for every i, we use a law of large numbers to show that (2) is equal to zero. Here is the list of events that we consider:

- $E^0 = \tilde{\sigma}$.
- $E^1((M_{t-y})_{y>0})$: The last market order before t is a buy order.
- $E^2((M_{t-y})_{y>0})$: The last market order before t is a sell order.
- $E^3((M_{t-y})_{y>0})$: The last mid price move before t is upward.
- $E^4((M_{t-y})_{y>0})$: The last mid price move before t is downward.

Remark 4. *We could have taken any event on the past that happens often enough.*

Description of our database

Our database is a list of 25 futures[14]. Standard notations are used to name the assets. For each of these futures, we consider all the trading days of January 2012. For each of these days, we only take the most liquid maturity. We have a list of all the first limit order book events (market orders, first limit limit orders and first limit cancellations) of the day. We cut off the first and last hours of the trading day.

We then consider a regular time grid (t_n) of 10 seconds interval[15] and for every t_n, we compute the $E^i((M_{t-y})_{y>0})$ and the next bid and ask prices. Averaging over the N points of the month,

[13] Where M is the market process, see Definition 4.

[14] Data kindly provided by QuantHouse EUROPE/ASIA, http://www.quanthouse.com.

[15] We could have taken any reasonable time grid. We choose this one to have enough points to approximate the expectation by its empirical average on this time grid. It is useless to take points which are too close together since then they would be too dependent.

we can compute the empirical approximations Δ^i of $\mathbb{E}[(na_t - nb_t)1_{E^i}((M_{t-y})_{y>0})]/\mathbb{P}(E^i((M_{t-y})_{y>0}))$:

$$\Delta^i = \frac{\frac{1}{N}\sum_{k=1}^{N} 1_{E^i_{t_n}}(na_{t_n} - nb_{t_n})}{\frac{1}{N}\sum_{k=1}^{N} 1_{E^i_{t_n}}}. \tag{3}$$

We present our results in Table VII.1.

Stock	Δ^0	Δ^1	Δ^2	Δ^3	Δ^4	Tick	Spread/Tick[16]
xFGBL	3.31e-3	3.33e-3	3.29e-3	3.33e-3	3.29e-3	1.e-2	1.09
xFGBS	3.27e-3	3.27e-3	3.26e-3	3.21e-3	3.28e-3	5.e-3	1.03
xFGBM	4.05e-3	4.10e-3	4.01e-3	4.05e-3	4.06e-3	1.e-2	1.04
xFDAX	6.26e-2	6.30e-2	6.20e-2	5.76e-2	6.75e-2	5.e-1	1.70
xFSXE	6.34e-1	6.42e-1	6.23e-1	6.33e-1	6.34e-1	1.	1.03
xFEURO	3.42e-5	3.41e-5	3.43e-5	3.40e-5	3.44e-5	1.e-4	1.16
xFAD	3.27e-5	3.96e-5	2.38e-5	3.42e-5	3.11e-5	1.e-4	1.26
xFJY	7.67e-7	5.9e-7	2.92e-6	8.36e-7	7.02e-7	1.e-6	1.21
xFBP	2.74e-5	2.11e-5	3.15e-5	2.94e-5	2.53e-5	1.e-4	1.40
xFCAD	2.94e-5	2.68e-5	3.10e-5	3.01e-5	2.87e-5	1.e-4	1.21
xFCH	3.25e-5	3.14e-5	3.04e-5	2.52e-5	3.83e-5	1.e-4	1.56
xFSP	1.66e-1	1.69e-1	1.63e-1	1.68e-1	1.65e-1	2.5e-1	1.01
xFND	1.01e-1	9.77e-2	1.04e-1	1.00e-1	1.01e-1	2.5e-1	1.21
xFDJ	4.43e-1	4.47e-1	4.36e-1	4.40e-1	4.46e-1	1.	1.39
xFGOLD	3.18e-2	3.08e-2	3.31e-2	3.10e-2	3.26e-2	1.e-1	1.93
xFCOPP	2.31e-4	2.28e-4	2.34e-4	2.43e-4	2.19e-4	5.e-4	2.13

Table VII.1 – Expectations of the difference between the next bid and the next ask for 5 different events during January 2012.

As we have theoretically shown under the fair market assumption, we observe that for all events, the difference is smaller than one tick. Therefore for small tick assets, it is marginal and our prediction is empirically quite satisfied.

2.5 The fair price as a reference price

In a rather large class of market dynamics, we have shown that it is always possible to define a martingale "fair price" which is the expectation of the price at which it is possible for agents to infinitesimally buy or sell assets using limit orders.

[16] Average of the spread divided by the tick value.

Definition 9. *We define the expected ex post gain of an ask (resp. bid) limit order compared to a price process as the expectation of the difference between the price of the limit order and the price process just after the limit order (resp. the difference between the price process after the limit order and the price of the limit order).*

In this work, the most important property of the fair price is the following result. It explains why it should be considered as a rigorous reference price to compute ex post gains.

Proposition 4. *Under Assumption 1, the expected ex post gain of a limit order compared to the fair price is equal to zero.*

Proof. Let us assume that at time t, the expected ex post gain of an ask limit order is strictly positive. This means that the expectation $\mathscr{E}$ of the fair price just after the execution of this order is smaller than the execution price of the order A. Consider the following infinitesimal market making strategy:

- Set an ask limit order just in front of this ask limit order.
- Right after the eventual execution of this ask limit order, set a bid limit order which follows the best bid.

Then, by definition of the fair price (expectation of the price of the next trade at the bid), the expected execution price B of the bid limit order is the value of the fair price right after the execution of the ask limit order $\mathscr{E}$. Therefore, the expected profit of this strategy is $A - B = A - \mathscr{E} > 0$. □

In the next section, we apply Proposition 4 to limit orders in various contexts.

3 Bid-ask spread, fair price and impact

3.1 Necessary and sufficient conditions on the bid and ask prices

In the previous section, we have given a way to define a fair price and we have shown that the ex post gain of any limit order compared to this price has to be null under the fair market hypothesis. Here, we proceed the other way around and show that given a model of the impact of the flow of market orders on the martingale fair price, we can derive no arbitrage bid and ask prices (as it is done in the construction of the MRR model). To do so, we proceed in the same vein as in [MRR97] and [WBK+08] applying Proposition 4 to limit orders placed at the best bid and ask prices.

Proposition 5. *Given dynamics for the fair price and the order flow (P_t, v_t^A, v_t^B), the bid and ask prices at time t, a_t and b_t must satisfy:*

$$a_t = \mathbb{E}[P_{t^+} | v_t^A > 0, \mathscr{F}_t^*]^{17} \tag{4a}$$

[17] The notation $\mathbb{E}[P_{t^+} | v_t^A > 0, \mathscr{F}_t^*]$ means $\lim_{\varepsilon \to 0+} \mathbb{E}[P_{t+\varepsilon} | \tau_t^A \in [t, t+\varepsilon], \mathscr{F}_t^*]$ that we assume to always exist, where τ_t^A is the time of the next ask order after (including) t.

and

$$b_t = \mathbb{E}[P_{t^+} | v_t^B > 0, \mathscr{F}_t^*]. \tag{4b}$$

Proof. If $a_t > \mathbb{E}[P_t^+ | v_t^A > 0, \mathscr{F}_t^*]$, placing an infinitesimal ask limit order just in front of a_t between t and $\min(t+\varepsilon, \tau_t^A)$ gives a positive average ex post gain for ε small enough. □

In spite of the recurrence of the concept of market impact of market orders in the microstructure literature, there is no consensus on how it should be defined or even on what price it should be computed. In our framework, we can give a natural and model independent definition of the local average market impact of a trade.

Definition 10. *For any given market dynamics, we define the local average market impact at time t of a bid or ask market order as:*

$$AMI_t^{ask} = \mathbb{E}[P_{t^+} | v_t^A > 0, \mathscr{F}_t^*] - P_t$$
$$AMI_t^{bid} = \mathbb{E}[P_{t^+} | v_t^B > 0, \mathscr{F}_t^*] - P_t.$$

By calling it local, we want to stress that this value can depend on the market state at time t and is thus difficult to compute empirically.

Remark 5. *For any $h > 0$, we could have taken the definitions:*

$$AMI_t^{ask} = \mathbb{E}[P_{t+h} | v_t^A > 0, \mathscr{F}_t^*] - P_t$$

and

$$AMI_t^{bid} = \mathbb{E}[P_{t+h} | v_t^B > 0, \mathscr{F}_t^*] - P_t.$$

Indeed, since P is a martingale, the definition of the impact does not depend on the horizon. This independence is another good reason to consider the impact of market orders on the fair price and not on an arbitrary price as it is often done.

The following result is immediate.

Proposition 6. *With this definition of market impact, Proposition 5 can be rewritten:*

$$a_t = P_t + AMI_t^{ask}$$

and

$$b_t = P_t + AMI_t^{bid}.$$

Remark 6. *This implies the dynamics of (a_t, b_t, v_t^A, v_t^B) can be deduced from the dynamics of (P_t, v_t^A, v_t^B). There is thus an equivalence between these two dynamics under the fair market hypothesis.*

Proposition 6 implies that the spread

$$\Phi_t = a_t - b_t = AMI_t^{ask} - AMI_t^{bid}$$

is completely explained by adverse selection. This is in fact not surprising since Assumption 1 essentially says that adverse selection "compensates" the position of liquidity that market makers set in the order book. In the next paragraph, we show that we can test Proposition 5 on data and we will see that it is indeed well satisfied (especially for small tick assets).

3.2 Verification of Proposition 5 on data

As it is not possible to compute model independently the fair price, neither is it possible to do it for the average market impact of trades AMI. However let us show that we can empirically compute adequate ex post gains and thus check Propositions 5.

By definition of P_t, Equations (4a) and (4b) can be rewritten:

$$a_t = \mathbb{E}[P_{t^+} | v_t^A > 0, \mathscr{F}_t^*] = \mathbb{E}[na_{t^+} | v_t^A > 0, \mathscr{F}_t^*]^{18}$$

and

$$b_t = \mathbb{E}[P_{t^+} | v_t^B > 0, \mathscr{F}_t^*] = \mathbb{E}[nb_{t^+} | v_t^B > 0, \mathscr{F}_t^*].$$

Let us now proceed as in Section 2.4. We need the following lemma:

Lemma 1. *Set*

$$a_t^\varepsilon = \mathbb{E}[na_{t+\varepsilon} | \tau_t^A \in [t, t+\varepsilon], \mathscr{F}_t^*]$$

and assume for all $E_t \in \mathscr{F}_t^*$,

$$\mathbb{E}[(a_t^\varepsilon - a_t) 1_{E_t} 1_{\tau_t^A \in [t,t+\varepsilon]}] / \mathbb{P}[\tau_t^A \in [t, t+\varepsilon]] \to 0$$

which is true in any reasonable model since $a_t^\varepsilon \to a_t$. *Then, we have*

$$\forall E \in \tilde{\sigma},\ \frac{\mathbb{E}[a_t 1_E((M_{t-}y)_{y>0}) | v_t^A > 0]}{\mathbb{P}[(M_{t-}y)_{y>0} \in E | v_t^A > 0]} = \frac{\mathbb{E}[na_{t^+} 1_E((M_{t-}y)_{y>0}) | v_t^A > 0]}{\mathbb{P}[(M_{t-}y)_{y>0} \in E | v_t^A > 0]}$$

and

$$\forall E \in \tilde{\sigma},\ \frac{\mathbb{E}[b_t 1_E((M_{t-}y)_{y>0}) | v_t^B > 0]}{\mathbb{P}[(M_{t-}y)_{y>0} \in E | v_t^B > 0]} = \frac{\mathbb{E}[nb_{t^+} 1_E((M_{t-}y)_{y>0}) | v_t^B > 0]}{\mathbb{P}[(M_{t-}y)_{y>0} \in E | v_t^B > 0]}.$$

Proof. By definition of the conditional expectation for all $E_t \in \mathscr{F}_t^*$,

$$\frac{\mathbb{E}[a_t^\varepsilon 1_{E_t} 1_{\tau_t^A \in [t,t+\varepsilon]}] / \mathbb{P}[\tau_t^A \in [t, t+\varepsilon]]}{\mathbb{P}[E_t, \tau_t^A \in [t, t+\varepsilon]] / \mathbb{P}[\tau_t^A \in [t, t+\varepsilon]]} = \frac{\mathbb{E}[na_{t+\varepsilon} 1_{E_t} 1_{\tau_t^A \in [t,t+\varepsilon]}] / \mathbb{P}[\tau_t^A \in [t, t+\varepsilon]]}{\mathbb{P}[E_t, \tau_t^A \in [t, t+\varepsilon]] / \mathbb{P}[\tau_t^A \in [t, t+\varepsilon]]}.$$

Going to the limit as ε tends to zero and replacing the event E_t by $(M_{t-}y)_{y>0} \in E$ as in Section 2.4 ends the proof. □

Let us now assume that

$$\frac{\mathbb{E}[(a_t - na_{t^+}) 1_E((M_{t-}y)_{y>0}) | v_t^A > 0]}{\mathbb{P}[(M_{t-}y)_{y>0} \in E | v_t^A > 0]}$$

and

$$\frac{\mathbb{E}[(b_t - nb_{t^+}) 1_E((M_{t-}y)_{y>0}) | v_t^B > 0]}{\mathbb{P}[(M_{t-}y)_{y>0} \in E | v_t^B > 0]}$$

[18] Which in fact means $\lim_{\varepsilon \to 0+} \mathbb{E}[na_{t+\varepsilon} | \tau_t^A \in [t, t+\varepsilon], \mathscr{F}_t^*]$.

do not depend on t.

Considering the same E^i as before and summing over all the N^{ask} and N^{bid} ask and bid market order times (t_n^{ask}) and (t_n^{bid}) of the assets of our database during January 2012, we need to show that

$$\Delta A^i = \frac{\frac{1}{N^{ask}} \sum_{k=1}^{N^{ask}} 1_{E^i_{t_n^{ask}}} (a_{t_n^{ask}} - n a_{(t_n^{ask})^+})}{\frac{1}{N^{ask}} \sum_{k=1}^{N^{ask}} 1_{E^i_{t_n^{ask}}}} \simeq 0 \tag{5a}$$

and

$$\Delta B^i = \frac{\frac{1}{N^{bid}} \sum_{k=1}^{N^{bid}} 1_{E^i_{t_n^{bid}}} (b_{t_n^{bid}} - n b_{(t_n^{bid})^+})}{\frac{1}{N^{bid}} \sum_{k=1}^{N^{bid}} 1_{E^i_{t_n^{bid}}}} \simeq 0. \tag{5b}$$

In Table VII.2, we list the empirical ΔA^i and ΔB^i for our list of futures after the events (E^i) defined by Equations (5a) and (5b).

Stock	ΔA^0	ΔA^1	ΔB^0	ΔB^1	Tick	NOrders
xFGBL	-1.00e-6	1.91e-4	-1.09e-6	-2.08e-4	1.e-2	8.00e5
xFGBS	-1.16e-6	6.00e-5	4.72e-7	-7.95e-5	5.e-3	1.98e5
xFGBM	-1.08e-6	1.62e-4	7.41e-7	-1.84e-4	1.e-2	3.11e5
xFDAX	1.35e-3	8.10e-3	1.48e-3	-5.77e-3	5.e-1	1.23e6
xFSXE	3.74e-4	2.05e-2	4.36e-4	-2.66e-2	1.	8.27e5
xFEURO	2.44e-8	1.15e-6	2.52e-8	-1.02e-6	1.e-4	1.63e6
xFAD	1.64e-7	7.66e-7	1.59e-7	-3.74e-7	1.e-4	6.37e5
xFJY	4.01e-8	4.41e-8	6.39e-10	4.14e-10	1.e-6	1.27e4
xFBP	1.49e-7	2.93e-7	1.29e-7	-2.55e-7	1.e-4	4.66e5
xFCAD	1.43e-7	5.90e-7	1.45e-7	-1.68e-7	1.e-4	3.98e5
xFCH	2.47e-7	-9.47e-7	1.66e-7	1.34e-6	1.e-4	1.90e5
xFSP	2.73e-5	2.56e-3	2.63e-5	-2.64e-3	2.5e-1	2.98e6
xFND	3.73e-4	2.51e-3	3.61e-4	-1.32e-3	2.5e-1	1.12e6
xFDJ	7.97e-4	2.53e-3	9.24e-4	1.22e-3	1.	7.33e5
xFGOLD	1.88e-3	6.89e-3	6.11e-4	1.53e-4	1.e-1	9.02e5
xFCOPP	2.08e-6	6.10e-6	2.15e-6	-1.79e-6	5.e-4	3.66e5

Table VII.2 – Average next ask (resp. bid) after an ask (resp.bid) market order compared to the ask (resp. bid) price just before the order after some events.

As expected, we see that the ΔA^i and ΔB^i are always significantly smaller than a tick.

The nullity of average ex post gains of limit orders placed at the best prices for small tick assets shows that simple market making strategies can only be marginally profitable on average. Therefore, we have empirically shown that, as postulated in [MRR97] and [WBK+08], adverse selection and the tick value explain the bid and ask prices and thus the spread. The risk aversion of market makers and their order processing costs seem to have negligible influences on the spread.

3.3 Response functions, time horizon and market making

In [WBK+08], similar conditions on the profit of market making strategies are studied. The difference is that the ex post gain of limit orders is computed compared to the mid price a time δt after the order: $E[a_t - m_{t+\delta t} | v_t^A > 0]$ for ask limit orders and $E[m_{t+\delta t} - b_t | v_t^B > 0]$ for bid limit orders.

However, the mid price is not a martingale in general. In particular, the empirical response function is not constant but significantly (compared to the tick value) increases over time, see [BGPW04] and they thus need to consider more complex market making strategies taking into account the dehedging of market makers. They show that, the ex post gain of market makers depends on the horizon of their strategies. In particular, it is stated that fast market making should be more profitable than slow market making.

This is inconsistent with our framework where the average profit of a market making strategy does not depend on the "horizon" of the market maker. Let us show that in our framework we can also measure the increment of market making profit over time and that as measured in [WBK+08] it increases over time. However, we also show that it is always smaller than the tick and thus coherent with our framework for small tick assets.

In order to empirically show that the average profit of market making strategies does not seem to depend much on the horizon of market makers, we plot the ask (resp. bid) response function of the fair price defined as:

$$RA^P(\delta) = \mathbb{E}[P_{t+\delta} - P_t | v_t^A > 0]$$

and

$$RB^P(\delta) = \mathbb{E}[P_{t+\delta} - P_t | v_t^B > 0]$$

which also write

$$RA^P(\delta) = \mathbb{E}[na_{t+\delta} - P_t | v_t^A > 0]$$

and

$$RB^P(\delta) = \mathbb{E}[nb_{t+\delta} - P_t | v_t^B > 0]$$

by definition of the fair price.

In our framework, this function should not depend on δ since P is a martingale. Let us notice that since we cannot empirically compute model independently the fair price before the orders, we cannot measure RA^P and RB^P. However, we can compute the empirical variations of the

ask (resp. bid) response function on the fair price defined as the empirical average over the ask and bid order times of January 2012:

$$RA^P(\delta) - RA^P(0^+) = \frac{1}{N^{ask}} \sum_{i=1}^{N^{ask}} (na_{t_n^{ask}+\delta} - na_{(t_n^{ask})^+})$$

and

$$RB^P(\delta) - RB^P(0^+) = \frac{1}{N^{bid}} \sum_{i=1}^{N^{bid}} (nb_{t_n^{bid}+\delta} - nb_{(t_n^{bid})^+})$$

for three liquid assets, see Figures VII.1, VII.2 and VII.3.

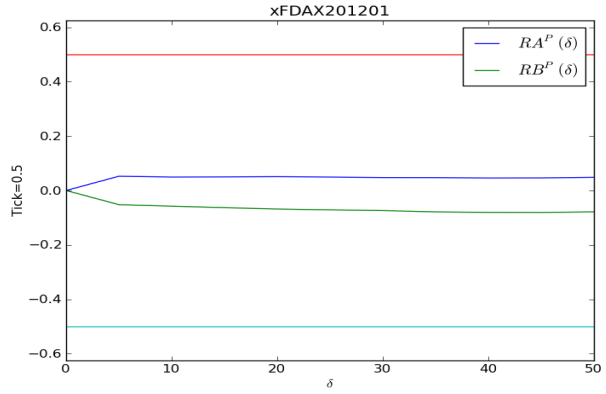


Figure VII.1 – Ask (in blue) and bid (in green) response functions for the DAX future. The two horizontal lines correspond to plus (in red) and minus (in cyan) the tick value.

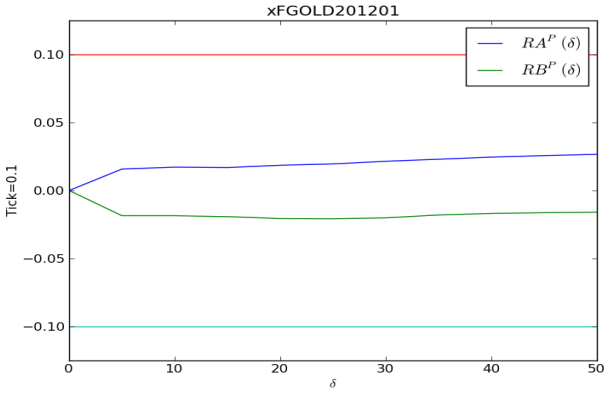


Figure VII.2 – Ask (in blue) and bid (in green) response functions for the GOLD future. The two horizontal lines correspond to plus (in red) and minus (in cyan) the tick value.

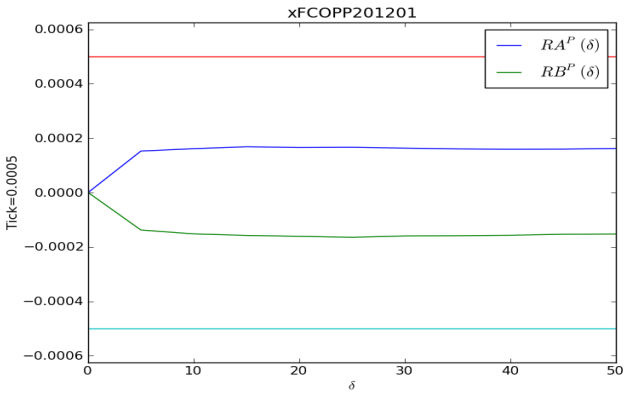


Figure VII.3 – Ask (in blue) and bid (in green) response functions for the COPP future. The two horizontal lines correspond to plus (in red) and minus (in cyan) the tick value.

We observe on Figures VII.1, VII.2 and VII.3 that in accordance with our predictions, the variations of the response functions computed on the fair price are small compared to the tick value.

Remark 7. *Notice that there is a slight increment of the ask response function at very small time scales (below the second) which implies that fast market making might be marginally more profitable than slow market making. However, this increment is always small compared to the tick value.*

3.4 On tick value, market making and liquidity

We have shown evidence that for small tick assets, the distance of some liquidity from the best price was perfectly compensated by the impact of market orders which consumed this liquidity. This gave ex post gains close to zero for market makers. For large tick assets, it is no longer the case. Indeed, market makers cannot any more perfectly compete by putting liquidity as close to the mid price as they might wish therefore, the competition is in the speed at which they take the priority.

This seems to make the markets less efficient. On the other hand, a large tick market might attract market makers and thus liquidity. Moreover, a large tick simplifies the information flow that practitioners get. The optimal tick value, should be computed with these arguments in mind. See [DR12] for an example of computation of an optimal tick value.

4 Conclusion

We have studied continuous time market dynamics which only satisfy a fair market hypothesis. Theoretically, this hypothesis can be justified by arguments of perfect competition between market makers. In this general framework, we have given precise and model independent definitions of concepts such as fair price, market impact and liquidity that are often present

in the literature on order book dynamics. Taking our fair price as a reference price, we have derived a relationship between, market impact and liquidity which is supported by real data analysis.

From a practical point of view, the strategies of market participants such as brokers depend a lot on liquidity and impact. Such agents should thus calibrate their strategies using models which accurately reproduce stylized relationships between these quantities. In particular, in such models, the ex post gains of limit orders compared to a martingale reference price should be equal to zero.

Our framework implies that market making strategies are not profitable. Obviously, this is not rigorously true because it is the competition between market makers that makes our hypothesis reasonable. The first reason for the coherence between our framework and the existence of market makers is the tick value which can lead to profits for market makers. The second reason is that we have only considered very simple strategies and small statistical arbitrages may remain for more complex strategies. Finally some market participants are encouraged to provide liquidity in order to obtain advantages for other market activities.

CHAPTER VIII

Market impact as anticipation of the order flow imbalance

Abstract

In this paper, we assume that the permanent market impact of metaorders is linear and that the price is a martingale. Those two hypotheses enable us to derive the evolution of the price from the dynamics of the flow of market orders. For example, if the market order flow is assumed to follow a nearly unstable Hawkes process, we retrieve the apparent long memory of the flow together with a power law impact function which is consistent with the celebrated square root law. We also link the long memory exponent of the sign of market orders with the impact function exponent. One of the originalities of our approach is that our results are derived without assuming that market participants are able to detect the beginning of metaorders.

Keywords: Market impact, metaorder, market order flow, Hawkes processes, long memory, square root law, market efficiency.

1 Introduction

Loosely speaking, the *market impact* is the link between the volume of an order (either market order or metaorder[1]) and the price moves during and after the execution of this order, see [Bou10]. Accurate modeling of market impact is of paramount importance. Practically, market impact greatly affects brokers, market makers but also large investors in the design of their optimal strategies. For example, market impact tends to decrease the profits of investment strategies by creating an execution cost that increases with the volume of the transaction. From a theoretical point of view, it is argued in [BFL09] and [Shi80] that endogenous (that is independent of information outside of the market) changes in supply and demand have more influence on price fluctuations than exogenous information. Therefore, understanding the price formation process necessarily goes through understanding market impact.

[1] A metaorder is a large transaction executed incrementally, see [FGLW13].

Here we focus on metaorders. More precisely, we study the impact function of metaorders, which is the expectation of the price move with respect to time during and after the execution of the metaorder. We call permanent market impact of a metaorder the limit in time of the impact function (that is the average price move between the start of the metaorder and a long time after its execution). The literature on this topic is often rooted to the seminal work of Kyle [Kyl85]. In the latter, it is shown that an informed trader who has private information on the future of the price should incrementally execute his metaorder, thereby slowly revealing the price to the rest of the market. However, it is also deduced that the market impact should be linear throughout the execution of the metaorder, which does not agree with empirical studies that consistently give a strictly concave market impact. We refer for example to [BR13] and [MVM$^+$09] where the estimated impact function is close to a power law with respect to time throughout the execution, with exponent being between 0.5, see [BR13], and 0.7, see [MVM$^+$09]. These empirical results are often referred to as the square root law.

The first main theoretical explanations of the concavity of the impact function connect it to the behavior of large investors, see [GGPS06], the persistence in the order flow, see [BGPW04] and [LF04], or the size distribution of metaorders, see [FGLW13]. In [TLD$^+$11], Toth *et al.* show that the diffusivity of the price can be linked to the V-shape of the latent order book. They propose a model in which they numerically show that the concavity of the impact function depends on the long memory of the order flow. Recently, Farmer *et al.* [FGLW13] have managed to relate this concavity to the distribution of the sizes of metaorders. More precisely, in their model, they get a power law impact function whose exponent ν is linked to the exponent of the distribution tail of the size of metaorders β by $\nu = \beta - 1$, which quite fits empirical data. To obtain this, they use two conditions: A martingale hypothesis on the price, and a fair pricing condition, which is derived from a Nash equilibrium between investors and states that the average ex post gain of a metaorder of any size should be zero. Let us mention that in [Don12], similar hypothesis are used and the fair pricing condition is viewed as a “second order condition" in the sense that it minimizes the market makers’ P&L volatility.

In the works mentioned above, it is assumed that market makers can precisely detect when a metaorder is being executed. Although reasonable for large metaorders, it is a strong postulate, especially at the beginning of the execution and for small metaorders. Here, we try to build a theory which does not require such an assumption. Indeed, market makers only see a flow of market orders and should thus set their prices according to this flow, and not to the underlying flow of metaorders.

It is well known, see [LF04], that the sign (buy or sell) of market orders presents persistence (more precisely, the correlation function of the sign process behaves as a power law with exponent lower than one). This persistence can be explained by the fact that the market order flow reflects the partition of metaorders into sequences of same-sided market orders. Indeed, it is shown in [LMF05] that an order flow obtained from the fragmentation of metaorders exhibits long memory. In [TPLF11], it is empirically proved that the persistence in the sign process mostly comes from splitting (and not herding).

In [BGPW04], Bouchaud *et al.* propose a way to obtain a diffusive price process from a long memory order flow modeled as a discrete time FARIMA process (see [Ber94] for definition). To do so, they consider that the impact of a market order decreases in time as a power law (where the exponent of the power law is linked to the exponent of the long memory). However, in this work, the impact is not related to the notion of metaorder. In particular, if one executes a metaorder in this model, its permanent impact is null, see [BFL09]. Moreover, while the permanent impact of metaorders is not easy to compute, it does not seem to be zero in practice, see [MVM$^+$09]. Our aim is to somehow place ourselves between [BGPW04] and [FGLW13]. We assume that the permanent impact of metaorders does exist with a specific (linear) form derived from no price manipulation arguments. Furthermore, we consider that market makers only see the flow of market orders that they essentially understand as a superposition of fragmented metaorders.

There are two manners to understand why metaorders impact the price. In most papers, see for example [FGLW13] and [Kyl85], market impact is viewed as a way to pass on private information to the price. In these models, large investors react to information signals on the future expectation of the price using metaorders. In such approaches, metaorders reveal “fundamental" price moves but do not really cause them. In particular, if a metaorder is executed for no reason, it does not have any long term impact on the price. Here, we use the other vision of market impact. We assume that it is mechanical in the sense that a metaorder moves the price through its volume, by creating a long term imbalance in supply and demand, independently of the informativeness of the metaorder[2]. Choosing between these two paradigms requires brokerage data where the client accepts to say whether he is trading because he has directional views on the price or for another reason (risk management, hedging, regulatory constraints,...). This has been done in [CS07], [Obi07] and more more recently in [WG13] where it is shown that after a few days, the impact of “informed trades" is larger than that of “cash flow trades" (which even tends to be null), see also [ACG07] and [MPS04]. However, at intraday time scales, which are of interest here, the two previous impact functions are very close.

In some papers, see for example [Gat10] and [HS04], the mechanical vision of market impact is implicitly assumed to derive important results. In these two articles, the authors derive conditions in their models so that a round trip[3] is not profitable on average. In [HS04], it is applied to the Almgren-Chriss optimal liquidation framework, see [AC01], and the authors obtain that the permanent impact of a trade must be linear in its size. In [Gat10], this hypothesis is considered in a more general model and once again, it is obtained that if there is permanent impact, it has to be linear. However, these reasonings only hold if the impact of a trade does not depend on the reason investors are executing their trades, therefore, under the hypothesis that the impact is mechanical.

[2] In Section 2.2, we illustrate this vision of market impact with a simple investor model.

[3] A round trip is defined in [HS04] as a trading strategy which starts and ends with a null inventory.

In this work, we make two natural hypotheses: The linearity of the permanent market impact of a metaorder (which arises when impact is mechanical and there is no price manipulation but can also hold if impact is informational, see Section 2), and a martingale assumption on the price. These two hypotheses give us a simple and general relationship between the dynamics of the order flow and the evolution of the price (Equation (1)). Let us emphasize that this relationship is not between the price and the flow of specific metaorders which are not observable by market makers. We then apply this formula to the example of an order flow modeled by a nearly unstable Hawkes process.

Hawkes processes, introduced in [Haw71a], have been successfully applied to seismology, neurophysiology, epidemiology, and reliability and are nowadays very popular in finance. Among their recent applications in this field, let us cite the studies of tick by tick prices in [BDHM13], trades in [Bow07], order books in [Lar07], financial contagion in [ASCDL10] and credit risk in [EGG10]. Using Hawkes processes to model the flow of market orders in continuous time is natural. Indeed, it is a simple way to introduce clustering in the order flow and to reproduce the empirical correlation functions of buy and sell orders, see [BDM12].

We will see that we recover many well known stylized facts of market impact under the Hawkes assumption. In particular, at the scale of individual market orders, we will compute a continuous time propagator model which generalizes the discrete time propagator proposed in [BGPW04]. In such models, the impact of a market order on the price does not depend on the past of the market order flow but is decreasing in time. At the scale of metaorders, we will compute a power law impact function, where the exponent of the power law ν is linked to the exponent of the long memory of the order flow γ by the relation: $\gamma = 2\nu - 1$. We thus somehow recover the square root law.

The paper is organized as follows. In Section 2, we explain and justify the use of the linear permanent impact assumption and we compute our general impact equation. In Section 3, we apply this equation to the example of an order flow modeled by nearly unstable Hawkes processes. We derive the price process implied by such an order flow and show that the impact function of a metaorder is asymptotically a power law with exponent lower than one. We conclude in Section 4.

2 Computing price moves from the order flow

The aim of this section is to present our assumptions and to see how they imply our impact equation that links the dynamics of the price to that of the order flow.

2.1 Linear permanent impact under the mechanical impact assumption

In this section, we explain why it is reasonable to assume that the permanent market impact (PMI for short) of a metaorder is linear in its volume under the mechanical vision of market

impact.

During and after the execution of a metaorder of volume V, the price process is impacted. Indeed, if one buys a large quantity of stock, the price moves upward on average. We assume that a long time after the execution of the metaorder (compared to the time it takes to execute it), the market somehow returns to some kind of stationary state except that the price has on average moved by a value that only depends on V. The permanent market impact, is by definition this average price move between the start of the execution of the order and a long time after the end of this execution:

$$PMI(V) = \lim_{s \to +\infty} \mathbb{E}[P_s - P_0 | V],$$

where $s = 0$ corresponds to the beginning of the metaorder and P_s is the mid price process. We assume that the impact of metaorders is large enough so that we can neglect the spread and thus the mid, ask and bid prices are close.

In order to measure this impact, one thus has to wait for the market to absorb the volume of the order and to get to its new equilibrium.

Remark 1. *The notion of expectation is in fact not defined outside of a model. Here, we make the assumption (that is implicit in most works on market impact) that the dynamics of the market satisfy stationarity conditions which allow us to compute expectations as empirical averages.*

Remark 2. *The need to wait a long time (metaorders can last days), the variations of the market state (market impact depends on the stock, volatility, market activity,...) as well as the relative small number of recorded metaorders (in [MVM+ 09], they detect around 100 000 metaorders per year on 74 stocks traded on the LSE, in [BR13] they use 12 500 metaorders) make it very difficult to have robust empirical laws on the permanent impact. For example, in [BR13], the empirical permanent impact of a metaorder with respect to its duration is derived by averaging over metaorders on many stocks and market states (time periods) which might seem arguable. We can however say that the permanent market impact seems different from zero.*

We place ourselves in the mechanical impact paradigm explained in the introduction and give a simple reason, based on “no price manipulation" arguments, why this impact should be linear in the volume of the metaorder:

$$PMI(V) = kV.$$

In [HS04], it is shown that in the framework of Almgren and Chriss [AC01], where the impact is decomposed into execution costs and permanent market impact, the absence of price manipulations implies that the permanent market impact must be linear.

In [Gat10], these arguments are generalized to a more realistic time dependent impact model. The model states that if a trader executes an order at a trading rate $\dot{x}_s$ at time s (assuming that the volume x_t executed up to time t is differentiable), then he impacts the price S in t

by: $\int_0^t f(\dot{x}_s)G(t-s)ds$. Therefore, if without the trader, the price is a Brownian motion with volatility σ, the price satisfies:

$$S_t = S_0 + \int_0^t f(\dot{x}_s)G(t-s)ds + \int_0^t \sigma dB_s,$$

where f is the impact function and G is the decay kernel, which represents the decrease of the impact of a trade through time.

In this model, using the same kind of arguments as in [Gat10], we can easily show that if there is permanent impact, then this impact is linear in volume. More precisely, we have the following proposition:

Proposition 1. *If there is no price manipulation in the sense of [Gat10][4] and there is permanent impact, that is* $G(s) \underset{s\to+\infty}{\to} G_\infty > 0$, *then* f *is linear.*

The proof is given in appendix.

Remark 3. *In the previous model, the cumulated executed volume of the trader* $x_t = \int_0^t \dot{x}_s\, ds$ *is assumed to be differentiable which is not the case in practice. Indeed, the volume is executed by individual market orders. However, if* $f(v) = kv$ *as above, the model can be rewritten:*

$$S_t = S_0 + k\int_0^t G(t-s)dx_s + \int_0^t \sigma dB_s,$$

and generalized to real order flows. The permanent impact of a metaorder of volume V *is thus linear in its volume:*

$$PMI(V) = G_\infty kV.$$

Therefore, in a rather general model, the permanent impact of a metaorder has to be linear. However, let us point out that the preceding arguments assume that even traders who execute metaorders without information have a permanent impact on the price. Therefore, the impact has to be mechanical as explained in the introduction. This could seem unrealistic but in fact can be understood in terms of the aversion that asset managers have to holding large positions of an asset. In the next paragraph, we present a simple investor model which explains how even uninformed metaorders can impact an indifference price.

2.2 A toy investor model

A classical view is to consider that investors who use metaorders have a better information about the future of the price than other market participants, see for example [Kyl85]. However, this is based on the economical concept of an “informational price" whereas empirical studies seem to show that the main drive of price fluctuations is not information, see [Shi80], but volume, see [BFL09]. Indeed, in [Shi80], Shiller argues that: “The movements in stock price indexes could not realistically be attributed to any objective new information, since

[4] In [Gat10], a price manipulation is a round trip whose average cost is negative.

movements in the indexes are “too big" relative to actual subsequent-movements in dividends"[5].

In this section, we consider a simple example of investor model, which is in fact a particular case of the CAPM model (see for example [Sha64]). This model enables us to illustrate how an uninformed metaorder can have a long term mechanical impact on the price. This impact is due to the risk aversion of asset managers (that is their unwillingness to hold large positions of one asset) and not to information asymmetry.

We assume that there are n investors in our market. There is a fixed number N of shares that are spread between the agents and let us call P the price. Every investor $(I_i)_{i=1...n}$ estimates that the distribution of the yield of a share has an average E_i and a variance Σ_i and chooses the number of shares N_i in his portfolio optimizing a mean variance criterion with risk aversion λ_i:

$$\begin{aligned} N_i &= \operatorname{argmax}_x\{x(E_i - P) - \lambda_i x^2 \Sigma_i\} \\ &= \frac{E_i - P}{2\lambda_i \Sigma_i}. \end{aligned}$$

Moreover,

$$\sum_{i=1}^{n} N_i = N,$$

which gives the indifference price P of investors:

$$P = \frac{\sum_{i=1}^{n} \frac{E_i}{2\lambda_i \Sigma_i} - N}{\sum_{i=1}^{n} \frac{1}{2\lambda_i \Sigma_i}}.$$

Let us now assume that the total number of shares becomes $N - N_0$ due to the action of some non optimizing agent needing to buy some shares (for cash flow reasons for example). The new indifference price is

$$P^+ = P + \frac{N_0}{\sum_{i=1}^{n} \frac{1}{2\lambda_i \Sigma_i}} = P + kN_0.$$

This means that in this framework, a metaorder has an impact on the indifference price that is linear in its size. We thus have illustrated (in a quite naive model admittedly!) how the impact of a trade might not depend on its informativeness. In Appendix VIII.B, in order to have an order of magnitude of the impact of typical metaorders in this framework, we input reasonable values for the parameters of our model. We find that our theoretical uninformed permanent impact of metaorders is much smaller than the empirical impact of [WG13]. This result goes in the direction of empirical studies such as [WG13] which measure a small permanent impact for uninformed metaorders.

[5] I thank a referee for pointing out that, in fact, this reasoning only holds if one does not take into account the stochasticity of discount rates which might be a strong approximation.

Remark 4. *Let us stress again that in contrast to [FGLW13] and [Kyl85], in our framework, the impact of an order is purely due to its volume and not to any exogenous information. Therefore, all traders are equivalent and there are no noise traders. In particular, individual trades should be considered as small metaorders.*

2.3 Impact dynamics

We have shown that it is reasonable to consider that every metaorder executed by a market participant has a long term impact on the price that is linear in its size. Here, we moreover assume a martingale property for the price. We use these assumptions to get a general impact equation linking the order flow dynamics and the price process.

Let us denote by $[0, S]$ the time during which metaorders are being executed (which can be thought of as the trading day) and by V_t^a and V_t^b the cumulated volumes of market orders at the ask and at the bid up to a time $t \in [0, S]$. We consider the following assumption that is also presen in [KO11].

Assumption 1. *The daily cumulated volume of buy (resp. sell) market orders V_S^a (resp. V_S^b) is proportional to the sum of the volume of the N_S^a buy (resp. N_S^b sell) metaorders of the day*

$$V_S^a = \frac{\chi}{2} \sum_{i=1}^{N_S^a} v_i^a,$$

where v_i^a is the volume of the i^{th} buy metaorder of the day and χ is a constant.

Remark 5. *χ can be understood as "the level of intermediation in the market". For example, if one assumes that metaorders are executed only via market orders and market orders are only used to execute metaorders then $\chi = 2$. In practise, χ is smaller than 2 because investors also use limit orders.*

To fix ideas, we assume that at the beginning and at the end of the day, there are auctions whose prices are denoted by P_0^- and P_S^+. Then our assumption on the linearity of the permanent market impact goes as follows.

Assumption 2. *P_S^+ is on average equal to the price of the opening auction P_0^- plus the impact of the metaorders of the day which is taken linear in their volume:*

$$\begin{aligned} P_S^+ &= P_0^- + k\Big(\sum_{i=1}^{N_S^a} v_i^a - \sum_{i=1}^{N_S^b} v_i^b\Big) + Z_S \\ &= P_0^- + \kappa(V_S^a - V_S^b) + Z_S, \end{aligned}$$

where $\kappa = k\frac{2}{\chi}$ is a positive constant and Z is a martingale random walk.

Finally, we consider the following assumption.

Assumption 3. *The price P is a martingale.*

Assumption 3 implies that

$$\begin{aligned} P_t &= \mathbb{E}[P_S^+|\mathscr{F}_t] \\ &= P_0^- + \mathbb{E}[\kappa(V_S^a - V_S^b)|\mathscr{F}_t] + \mathbb{E}[Z_S|\mathscr{F}_t], \end{aligned}$$

where $\mathscr{F}_t$ is the information set of market participants at time t.

Remark 6. *Until now, we have mostly mentioned the impact of metaorders. The previous equation describes how this impact as well as the link between market orders and metaorders imply that the order flow has an impact on the price. Indeed, the flow of market orders changes the expectation of the cumulated order flow imbalance which corresponds to the expectation of the cumulated imbalance of metarders whose permanent impact is known. From now on, we will mostly directly study the impact of the order flow.*

The noise term $\mathbb{E}[Z_S|\mathscr{F}_t]$ corresponds to price moves that are caused by other factors than volume (for example public announcements). For the sake of simplicity, we consider that $Z_S = 0$ and thus

$$P_t = P_0^- + \mathbb{E}[\kappa(V_S^a - V_S^b)|\mathscr{F}_t].$$

Let us now extend the buy and sell market order flow processes to $\mathbb{R}_+$ (until now they were only defined on a trading day $[0, S]$) and assume that these processes satisfy the following technical assumption.

Assumption 4. *The quantity $\mathbb{E}[V_s^a - V_s^b|\mathscr{F}_t]$ converges to some finite limit when s tends to infinity.*

This assumption can be understood in this way: $\mathscr{F}_t$ does not provide any information about the order flow imbalance which will occur at time $t' \gg t$. Indeed, under this assumption, for all $h > 0$,

$$\mathbb{E}[(V_{s+h}^a - V_{s+h}^b) - (V_s^a - V_s^b)|\mathscr{F}_t] \underset{s\to+\infty}{\to} 0.$$

Therefore, the order flow imbalance between s and $s+h$ is asymptotically (in s) not predictable at time t. For example, in the next section, we will assume that the market order flows are Hawkes processes for which this assumption is satisfied. Under this hypothesis, we have that if S is large enough compared to the characteristic convergence scale of the function $s \mapsto \mathbb{E}[V_s^a - V_s^b|\mathscr{F}_t]$, when computing the price P_t, S can be seen as infinity. This implies the main result of this paper.

Theorem 1. *Under Assumptions 1, 2, 3 and 4*

$$P_t = P_0 + \kappa \lim_{s\to+\infty} \mathbb{E}[V_s^a - V_s^b|\mathscr{F}_t]. \tag{1}$$

Proof. Starting from Assumptions 1 and 2 and taking off Z to simplify we get:

$$P_S^+ = P_0^- + \kappa(V_S^a - V_S^b).$$

Using Assumption 3, this implies that

$$P_t = P_0^- + \mathbb{E}[\kappa(V_S^a - V_S^b)|\mathscr{F}_t].$$

Therefore, if S is large enough, Assumption 4 yields the result. □

Remark 7. *To fix ideas, we have assumed that metaorders stop at the end of the day S. However, some metaorders last days (or even weeks). Therefore, S should thus not be considered as the end of the day but the end of a period that is very large compared to the execution of metaorders.*

From Equation (1), and any example of order flow dynamics, we can thus derive a price process which satisfies the following properties:

- The price is a martingale even if the order flow exhibits persistence.
- The permanent impact of a metaorder is linear in its size, independently of the execution.
- The price process only depends on the global market order flow and not on the individual executions of metaorders. We thus do not need to assume that the market "sees" the execution of metaorders as it is usually done (see for example [FGLW13]).

Equation (1) can be seen as a rigorous and model independent generalization of the postulate "the impact is proportional to the innovation in the order flow" of [GM85] and [MRR97]. Indeed, in our model, market orders move the price because they change the anticipation that market makers have about the future of the order flow. Furthermore, considering, as in Equation (1), that price moves are due to the "surprise in the order flow" (that is the variation of the expected cumulated order flow imbalance) is a general way to solve the following apparent "paradox" (which is one of the challenges of impact models): Order flows are persistent and yet prices are martingales.

From the empirical point of view, our framework is supported by [Hir11] where it is shown that the trades of some arbitragers (namely high frequency traders) anticipate the order flows of other investors. Therefore, arbitragers move prices through their trades until it is no longer worth doing so; that is until the price correctly anticipates the future order flow imbalance.

In the next section, we give an example of realistic order flow dynamics and compute the associated price process. We will see that the linearity of the permanent impact of a metaorder is not incompatible with the observed concave power law impact functions.

3 Application to the Hawkes order flow example

We now want to apply Equation (1) to a reasonable example of market order flow dynamics. More precisely, we model the buy and sell market order flows as two independent (nearly unstable) stationary Hawkes processes which somehow reproduce the persistence of the sign of market orders. Applying Equation (1) to these processes, we derive a simple price model that is very similar to the propagator model of [BGPW04].

3.1 Impact of individual market orders

Before deriving the impact of individual market orders under the Hawkes order flow assumption, we need to introduce Hawkes processes and the propagator model.

3.1.1 Hawkes processes

Hawkes processes have been introduced in [Haw71a]. They have recently been applied to many fields of finance, see the references in introduction. For example, they were used to study tick by tick prices in [BDHM13] and order flows in [BM14a]. We model the arrival times of buy and sell market orders as the jump times of two independent Hawkes processes.

Remark 8. *We consider that all market orders have the same volume v.*

By definition, a Hawkes process N is a self exciting point process whose intensity λ at time t depends linearly on its past, see [Haw71a]:

$$\lambda_t = \mu + \int_0^t \phi(t-s) dN_s,$$

where μ is a positive constant and ϕ is a positive function supported on $\mathbb{R}_+$ which satisfies the stability condition:

$$\int_0^{+\infty} \phi(s) ds < 1.$$

The main reason for the popularity of Hawkes processes in the modeling of market activity is that they are a simple way to reproduce the persistence of this point process, see among many other works [FS13] or [HBB13]. Indeed, the self exciting nature of Hawkes processes implies that if there has recently been a lot of points, there will be (on average) a lot of points in the coming period, see [BDM12] and [BM14b] for links between the auto covariance of the point process and the Hawkes kernel. For other statistical studies showing the relevance of Hawkes processes to model the arrival of trades see [Bow07] or [HBB13] for Kolmogorov-Smirnov testing of whether price changes occur as Hawkes processes.

A nice property of Hawkes processes that we will use in order to compute the price dynamics is the following, see [BDHM12]:

Proposition 2. *$\forall t \geq 0$:*

$$\lambda_t = \mu + \mu \int_0^t \psi(t-s)ds + \int_0^t \psi(t-s)dM_s,$$

where $\psi = \sum_{k=1}^{+\infty} \phi^{*k}$, ϕ^{*k} being the k^{th} convolution product of ϕ and $dM_t = dN_t - \lambda_t dt$.

Therefore, since M is a martingale and ψ is supported on $\mathbb{R}_+$, for $0 \leq s, r \leq t$,

$$\mathbb{E}[N_t - N_s | F_r] = \int_s^t \mathbb{E}[\lambda_u | F_r] du$$

and

$$\mathbb{E}[\lambda_u | F_r] = \mu + \mu \int_0^u \psi(u-x)dx + \int_0^r \psi(u-x)dM_x$$

This formula allows us to compute the anticipation of the order flow and will be very useful to apply Equation (1) in the Hawkes context.

Hawkes processes are a natural extension of Poisson processes and can be used to model the empirical clustering of same-sided trades. More precisely, we will later show that a Hawkes order flow can almost reproduce the empirical long memory of the signed flow of market orders.

3.1.2 Propagator model

In [BGPW04], the order flow is modeled as a FARIMA process that enables to obtain the long memory property of this flow observed in practice. In order to have a diffusive price, the authors propose a price impact model called propagator model where the impact of a market order does not depend on the past of the order flow but is transient (its permanent market impact is null). Proceeding essentially as in [GSS12], this model can be extended to continuous time the following way:

Definition 1. *A continuous time propagator model is an impact model where the price process P is linked to the cumulated signed volume of market orders $V = V^a - V^b$ by:*

$$P_t = P_0 + \int_0^t \zeta(t-s)dV_s,$$

where ζ is a decay kernel.

Remark 9. *There is a small difference between this model and the model of [GSS12]: In [GSS12], V is the trading rate of a given trader while here it is the order flow of all the market.*

3.1.3 Computation of the price dynamics

From now on, we model the arrivals of buy and sell market orders as the points of two independent Hawkes processes N^a and N^b with identical kernel ϕ and exogenous intensity μ. We also assume that the volume of each market order is constant equal to v.

Proposition 3. *In the Hawkes order flow model, assuming that the price and volume satisfy the dynamics Equation* (1)*, then the price process follows a propagator model:*

$$P_t = P_0 + \int_0^t \zeta(t-s)(dN_s^a - dN_s^b), \tag{2}$$

with $\zeta(t) = \kappa v(1 + \int_t^{+\infty} \psi(u) - \int_0^t \psi(u-s)\phi(s)dsdu)$.

The proof is given in appendix.

Let us notice that the kernel does not tend to 0 since there is permanent impact.

Our propagator kernel ζ compensates the correlation of the order flow implied by the Hawkes kernel ϕ to recover a martingale price.

3.1.4 Properties of our propagator model

In this paragraph, we place ourselves in a general framework which includes the result of Proposition 3 and generalizes it to critical Hawkes order flows (with $\int \phi = 1$, see [BM01] for existence). We show that a continuous time propagator price model combined with a Hawkes order flow can lead to a martingale price under a simple condition on ζ and ϕ. We then use this condition to obtain a nicer expression of ζ. Indeed, if:

- The buy and sell market order flows are two independent Hawkes processes with kernel ϕ:

$$\lambda^{a/b} = \mu + \int_0^t \phi(t-s)dN_s^{a/b}.$$

- The price process follows a propagator model:

$$P_t = P_0 + \int_0^t \zeta(t-s)(dN_s^a - dN_s^b).$$

Then,

- If there is no trade in t, the price is differentiable in t with derivative

$$P_t' = \int_0^t \zeta'(t-s)(dN_s^a - dN_s^b).$$

- If there is a buy trade in t (which happens with intensity $\lambda_t^a = \mu + \int_0^t \phi(t-s)dN_s^a$), the price jumps upward of $\zeta(0)$.

- If there is a sell trade in t (which happens with intensity $\lambda_t^b = \mu + \int_0^t \phi(t-s)dN_s^b$), the price jumps downward of $\zeta(0)$.

Therefore,

$$\lim_{h\to 0} \frac{E[P_{t+h}|\mathscr{F}_t] - P_t}{h} = \zeta(0)\int_0^t \phi(t-s)(dN_s^a - dN_s^b) + \int_0^t \zeta'(t-s)(dN_s^a - dN_s^b).$$

Therefore, we have the following result:

Proposition 4. *In the framework defined above, if the price is a martingale then:*

$$\zeta'(x) = -\zeta(0)\phi(x). \tag{3}$$

Given a function ζ, we can compute a unique ϕ_ζ which satisfies (3)*. Conversely, given a ϕ and a $\zeta(0)$ (or ζ anywhere else), we can compute a unique ζ_ϕ which satisfies* (3)*.*

Remark 10. *Let us stress that in the general framework of this paragraph, the Hawkes order flows can be critical ($\int \phi = 1$). Moreover, the criticality of the Hawkes process is equivalent to the transience of the impact of a market order:*

$$\lim_{x\to+\infty} \zeta(x) = 0 \Leftrightarrow \int \phi = 1.$$

In this critical case, Assumption 4 is not satisfied. We are thus outside of the framework of Equation (1)*.*

Going back to the framework of Proposition 3 (in which $\int \phi < 1$), since our price P_t is an expectation conditionally on $\mathscr{F}_t$, it is by construction a martingale and therefore, the implied ζ must satisfy (3). Indeed, we have:

$$\begin{aligned}
\zeta'(t) &= \kappa\nu[-\psi(t) + \int_0^t \psi(t-s)\phi(s)ds - \int_t^{+\infty} \psi(u-t)\phi(t)dy] \\
&= \kappa\nu[-\psi(t) + \psi * \phi(t) - \phi(t)|\psi|] \\
&= -\kappa\nu(|\psi| + 1)\phi(t)
\end{aligned}$$

since

$$\psi * \phi = \psi - \phi$$

and thus

$$\zeta'(t) = -\zeta(0)\phi(t).$$

Theorem 2. *In the setting of Proposition 3, we can get another expression for ζ:*

$$\zeta(x) = \zeta(0)(1 - \int_0^x \phi(s)ds) = \kappa\nu\frac{1}{1-\int\phi}((1 - \int \phi) + \int_x^{+\infty} \phi(s)ds). \tag{4}$$

Proof. $\zeta'(x) = -\zeta(0)\phi(x)$ therefore

$$\zeta(x) = C - \zeta(0)\int_0^x \phi(s)ds = \zeta(0) - \zeta(0)\int_0^x \phi(s)ds.$$

Moreover, using Proposition 3,

$$\zeta(0) = \kappa v(1 + \int_0^{+\infty} \psi(u)du) = \kappa v/(1 - \int \phi).$$

□

The ratio between the temporary and the permanent impact of an order satisfies

$$\lim \zeta/\zeta(0) = 1 - \int \phi.$$

Therefore, when $\int \phi$ is close to one which, we will see, corresponds to the apparent long memory asymptotic, the temporary impact is much larger than the permanent impact. This is due to the fact that in this asymptotic a market order on average "generates" a lot of market orders, see next paragraph.

3.1.5 Cluster representation of Hawkes processes

In this paragraph, we recall the branching construction of Hawkes processes which will allow us to interpret $\int \phi$ in terms of market impact.

Let us define a population model (see the Introduction of [BM01]): At time zero, there are no individuals. Some individuals (migrants) arrive as a uniform Poisson process with intensity μ. If a migrant arrives at time s, the birth dates of its children form a Poisson process of intensity $\phi(. - s)$, with $\int \phi < 1$. In the same way, if a child is born in s', the birth dates of its children form a Poisson process of intensity $\phi(. - s')$. We call N_t the number of individuals who were born or migrated until t. We have that N is a point process of intensity

$$\lambda_t = \mu + \int_0^t \phi(t - s)dN_s,$$

and therefore, N is a Hawkes process.

If we now name a cluster the set of all individuals who descend from the same migrant, a Hawkes process is the superposition of independent clusters which arrive at a rate μ. This is the Poisson cluster representation of Hawkes processes, see [HO74]. Replacing "individual" by "market order" and "cluster" by "metaorder", a Hawkes process can be seen as the superposition of independent executions of metaorders. In this interpretation, the average size of a metaorder is $1/(1 - \int \phi)$. Therefore, when $\int \phi$ is close to one, a metaorder is on average made of a large number of market orders. The parameter $\int \phi$ can thus be seen as the degree of endogeneity of the market and has recently been the subject of many investigations, see [FS13], [HBB13] and Chapter II.

3.2 Impact of metaorders

In this paragraph, we study the impact of a large group of trades that we see as a metaorder in an asymptotic framework. To do so, we assume that the Hawkes order flows are nearly unstable (their kernel's norm is close to one) which enables to retrieve the persistence property which is observed on the data.

3.2.1 Apparent long memory

It is well known that the sign of market orders presents long memory: Its autocorrelation function asymptotically behaves as a power law with exponent lower than one, see [LF04]. Recall that the Fourier transforms of the kernel ϕ and of the autocorrelation $Cov(.,h)$ of the increments of length h[6] of a stationary Hawkes process (denoted respectively by $\hat{\phi}$ and $\widehat{Cov}(.,h)$) are linked by:

$$\widehat{Cov}(z,h) = h\frac{\frac{\mu}{1-\int\phi}\hat{g}_z^h}{|1-\hat{\phi}(z)|^2},$$

where $\hat{g}_z^h$ is the Fourier transform of $g_\tau^h = (1-|\tau|/h)^+$, see [BDM12].

Therefore, since

$$\widehat{Cov}(0,h) = \int_0^\infty Cov(\tau,h)d\tau$$

and

$$\int\phi = \hat{\phi}(0),$$

the infinity of the integral of the correlation function, which defines long memory, implies that the norm of the kernel must be equal to one. Such "long memory" Hawkes processes have been studied in [BM01].

However, estimations performed in [BM14a], [FS13] and [HBB13] seem to show that $\int\phi$ is close to one but strictly lower than one ($\int\phi \sim 0.9$). Such nearly unstable Hawkes processes have been studied in Chapter II, where it is shown that their diffusive limit is an integrated Cox-Ingersoll-Ross process. However, the latter work assumes that the shape of the kernel has an average in the sense that $\int x\phi(x)dx < +\infty$, which is not really in agreement with the power law persistence of the sign process. Qualitative arguments and numerical simulations presented in [BM14a] show that taking a nearly unstable Hawkes process with ϕ behaving asymptotically as $1/x^{1+\alpha}$ should lead to apparent long memory up to the time scale $1/(1-\int\phi)^{1/\alpha}$ (that is when the gap τ between the considered h-increments is much lower than $1/(1-\int\phi)^{1/\alpha}$). This is consistent with the empirical kernel estimations performed in Chapter V and in [BM14b] where estimations of the kernel associated to the trade process yield power law kernels whose norm is close to one.

In order to give a formal version of this statement, we proceed as in Chapter II and consider a sequence of renormalized Hawkes processes whose kernels'norm tends to one "faster" than the observation scale T tends to infinity.

[6] $Cov(\tau,h) = \mathbb{E}[(N_{t+\tau+h}-N_{t+\tau})(N_{t+h}-N_t)] - (\mathbb{E}[N_{t+h}-N_t])^2$.

Asymptotic framework

For every observation scale $T > 0$, we define the stationary Hawkes process N^T of exogenous intensity $\mu^T = C_\mu(1 - a_T)T^{2\alpha-1}$ and of kernel $\phi^T = a_T\Phi$, where $\int \Phi = 1$ and (a_T) is a sequence of positive real numbers smaller than one which tends to one as T tends to infinity. Furthermore, we assume that

$$\Phi(x) \underset{x\to+\infty}{\sim} \frac{\alpha c^\alpha}{x^{1+\alpha}},$$

where $c > 0$ and $\alpha \in]0, \frac{1}{2}[$.

Since T is the observation scale, we consider the process N^T_{Tt} that we renormalise by considering $X^T_t := A_T N^T_{Tt}$, with $A_T = (1 - a_T)/(T\mu^T) = 1/(T^{2\alpha})$. Of course, A_T and μ^T are chosen so that the expectation and the covariance of the sequence of processes converge.

Considering the process at time scales below $1/(1 - \int \phi)^{1/\alpha}$ means that the observation scale T is much smaller than $1/(1 - a_T)^{1/\alpha}$. It can be formally written the following way:

Assumption 5. $T(1 - a_T)^{1/\alpha} \underset{T\to+\infty}{\to} 0$.

Result

Let $h > 0$ be fixed, and the covariance of the h-increments of X^T be denoted by:

$$C^{X,T}_{\tau,h} := Cov[(X^T_{t+\tau+h} - X^T_{t+\tau}), (X^T_{t+h} - X^T_t)].$$

The following result holds.

Theorem 3. *Under Assumption 5:*

$$C^{X,T}_{\tau,h} \to C^X_{\tau,h} = K(|\tau + h|^{2H} + |\tau - h|^{2H} - 2|\tau|^{2H}), \tag{5}$$

with $H = 1/2 + \alpha$ and K is a constant explicitly given in the proof in appendix.

This limiting function satisfies the long memory property: for any $h > 0$,

$$C^X_{\tau,h} \underset{\tau\to+\infty}{\sim} \frac{C_h}{\tau^{1-2\alpha}}.$$

The proof is given in appendix.

Remark 11. *Equation* (5)*, corresponds exactly to the correlation function of a fractional Brownian motion of Hurst index H.*

Remark 12. *A more precise behavior of such "heavy tailed nearly unstable Hawkes processes" is heuristically discussed in Chapter II.*

In [BM14a], it is shown that the time scale $1/(1 - \int \phi)^{1/\alpha}$ up to which a Hawkes process can present apparent long memory can be rather large (10^3 seconds). This implies that modeling financial order flows as nearly unstable Hawkes processes is consistent with their empirical apparent persistence.

3.2.2 Modeling metaorders

In the above framework, N^a and N^b are the flows of anonymous market orders. This corresponds to modeling the market from the point of view of a passive agent who does not use orders and does not know who uses the different orders. In order to compute the impact function of metaorders, it is convenient to look at the market from the point of view of someone who is executing a (buy) metaorder, see [BM14a]. To do so, we consider that the total order flow is the sum of anonymous orders (which are modeled as two independent Hawkes processes as before) and labelled orders which correspond to "our" order flow. More precisely, during the execution of a metaorder, we model our order flow as a Poisson process $P^{F,\tau}$ of intensity F on $[0,\tau]$. The total buy and sell order flows are thus $N^a + P^{F,\tau}$ and N^b.

Let us compute the impact function of a metaorder in our model. From the point of view of the rest of the market, there is no difference between our orders and anonymous orders. It thus seems natural to assume that the price impact of our orders is the same as the price impact of anonymous orders. Therefore, according to the propagator model:

$$P_t = P_0 + \int_0^t \zeta(t-s)(dN_s^a - dN_s^b) + \int_0^t \zeta(t-s) dP_s^{F,\tau}.$$

The processes N^a and N^b having the same average intensities, taking expectations, we get the following result:

Proposition 5. *The impact function MI of a metaorder executed between 0 and τ is:*

$$MI(t) := \mathbb{E}[P_t - P_0] = F \int_0^{t\wedge\tau} \zeta(t-s) ds. \tag{6}$$

In particular, the permanent impact of this metaorder is

$$PMI = F\tau \lim_{t\to+\infty} \zeta(t).$$

Proof. Using the expression of the price, we write the market impact as

$$MI(t) = \mathbb{E}[\int_0^t \zeta(t-s)(dN_s^a - dN_s^b)] + \mathbb{E}[\int_0^t \zeta(t-s) dP_s^{F,\tau}].$$

Since the propagator kernel perfectly compensates the Hawkes kernel, the first expectation is equal to zero. Using the expectation of the Poisson process, the second expectation is equal to the result. □

3.2.3 Apparent power law impact function

Here, we consider that the Hawkes processes used to model the order flows are close to criticality (which corresponds to empirical measures, see Section 3.2.1). We show that if the length of the metaorder τ is small compared to the correlation scale $1/(1-\int\phi)^{1/\alpha}$ that appears in Section 3.2.1: $1 \ll \tau \ll 1/(1-\int\phi)^{1/\alpha}$, then the renormalized impact function is close to a

power law with exponent $1-\alpha$.

In order to formally express and show this statement, we consider the same sequence of Hawkes processes as in Section 3.2.1. Let us also denote $(\tau^T)_T$ the sequence of sizes of metaorders. In the same vein as before, we formally write that the length of metaorders is small compared to the correlation scale as:

Assumption 6. $\tau^T \underset{T\to+\infty}{\to} +\infty$ *and* $\tau^T(1-a_T)^{1/\alpha} \underset{T\to+\infty}{\to} 0$.

Let us define the renormalized impact for $t \in [0,1]$:

$$RMI^T(t) = \frac{1-a_T}{(\tau^T)^{1-\alpha}} MI^T(t\tau^T).$$

We have the following result:

Theorem 4. *Under Assumption 6, the market impact is asymptotically a power law:*

$$RMI^T(t) \underset{T\to+\infty}{\to} K' t^{1-\alpha},$$

with $K' = \frac{c^\alpha \kappa \nu \Phi}{1-\alpha}$.

The proof is given in appendix.

Of course, the renormalizing constant $(1-a_T)(\tau^T)^{\alpha-1}$ is chosen again so that the renormalized impact function converges. This power law impact function is close to the “square root law" which is empirically observed, see [MVM$^+$09]. Let us now see how this impact exponent $\nu = 1-\alpha$ can be linked to the order flow persistence exponent γ.

3.2.4 Link between the exponents

For a price process implied by Equation (1) and for a Hawkes order flow with kernel whose norm is close to one and whose distribution tail is of order $1/x^{1+\alpha}$, we have shown that:

- The exponent of the long memory of the order flow is $\gamma = 1-2\alpha$, see Theorem 3.
- The exponent of the impact power law is $\nu = 1-\alpha$, see Theorem 4.

Therefore, we have as in [BGPW04]:

$$\nu = (1+\gamma)/2.$$

On the data, γ is stock dependent and seems to vary from 0.2 to 0.7, see [BGPW04], which in our framework implies that ν must vary between 0.6 and 0.85, which quite fits empirical data, see [MVM$^+$09]. Note however that on the data, these two exponents do not seem to be linked, see [MTB14].

Remark 13. *Theoretically retrieving the shape of the impact function from other market variables is an important problem in market microstructure. Three main possibilities have been suggested to solve it. The first one is to consider the risk aversion of investors and market makers as it is done in [GGPS06]. Another way is to consider statistical models for the flow of market orders and price changes as in [BM14a]. Here we proceed somehow as in [BFL09], [FGLW13] or [TLD+11] by introducing a martingale condition on the price which allows us to imply the impact function only from the order flow. Let us notice that the link between the exponent of the impact function and the exponent of the long memory of the order flow that we get is the same as in [BM14a] and [BFL09].*

Reversion after the metaorders

In the same way, we can study the reversion of the price after the completion of the metaorder. For time scales that are large but small compared to the correlation scale $1/(1-\int\phi)^{1/\alpha}$, we get the following theorem.

Theorem 5. *Under Assumption 6, the market impact asymptotically behaves as a power law after the completion of the metaorder:*

$$RMI^T(1+t) \underset{T\to+\infty}{\to} K'((t+1)^{1-\alpha} - t^{1-\alpha}),$$

with $K' = \frac{c^\alpha \kappa \nu \Phi}{1-\alpha}$ *(since we have renormalized the time by* τ^T*,* $RMI^T(1)$ *corresponds to the renormalized market impact at the end of the metaorder).*

The proof is essentially the same as that of Theorem 4.

Defining the renormalized permanent market impact $RPMI$ of a metaorder as the limit of the renormalized impact function

$$RPMI^T = \lim_{t\to+\infty} RMI^T(t),$$

we get

$$RPMI^T = \kappa \nu F (\tau^T)^\alpha (1 - a_T).$$

This quantity tends to zero as T tends to infinity because of Assumption 6. In practice, this means that the permanent impact of metaorders is much smaller than the "peak impact[7]".

Discussion and comparison with measures of permanent market impact

In [BR13], [MVM+09] and [WG13], empirical measures of the permanent market impact of metaorders are performed. As predicted by the theory of [FGLW13], it seems[8] that this permanent impact is a square root function of the volume and is equal to two thirds of the peak impact.

[7] That is the impact at the time of completion of the metaorder.

[8] Any measure of the permanent market impact is very noisy and requires a huge number of metaorders.

At first sight, these results thus appear to be inconsistent with our small[9] and linear permanent impact. However, as recently explained in [BKB$^+$14], these empirical measurements do not correspond to the permanent impact of an isolated metaorder, as computed above, but of realized metaorders. Therefore, since metaorders are themselves clustered[10], these measurements can probably not be seen as estimates of the permanent impact of an isolated metaorder. Actually, in [BKB$^+$14], when "deconvolving" the market impact from past and future metaorders, a much smaller permanent impact is empirically obtained, in agreement with our model.

4 Conclusion

We have recalled why the permanent market impact of a metaorder should be linear. Using this linearity as well as a martingale hypothesis on the price, we have computed an impact equation (Equation (1)) that allows us to retrieve price dynamics from any order flow model. We applied it to the example of order flows modeled by nearly unstable Hawkes processes. In this model, the sign of market orders presents apparent long memory and we have shown that we can recover many stylized facts of market impact. In particular, we computed a power law impact function whose exponent is linked to the long memory of the sign of market orders.

The example of the Hawkes order flow allowed us to derive simple formulas on the future expectation of the volume. However, this model might be too simplistic. It would be interesting to consider more realistic frameworks such as the one presented in [LMF05] where the order flow is really built as a superposition of independent metaorders. In addition to their natural interpretation, these order flow models manage to link the size distribution of metaorders to the long memory property of the order flow. However, such models are in a way much more complex and getting closed form prediction formulas for the order flow imbalance is probably very intricate.

VIII.A Proofs

VIII.A.1 Proof of Proposition 1

We proceed as in [Gat10]:

For all $(v_1, v_2, T) > 0$, we consider that on $[0, \theta T]$, we buy stock at a rate v_1 and on $[\theta T, T]$ sell stock at a rate v_2. We take $\theta = v_2/(v_1 + v_2)$ so that this strategy is a round trip. It is shown in [Gat10] (Equation (4)) that the expectation of the cost of this strategy is:

$$E = v_1 f(v_1) \int_0^{\theta T} \int_0^t G(t-s) ds dt + v_2 f(v_2) \int_{\theta T}^T \int_{\theta T}^t G(t-s) ds dt - v_2 f(v_1) \int_{\theta T}^T \int_0^{\theta T} G(t-s) ds dt$$

[9] Compared to the peak impact.

[10] More precisely, the signs of the metaorders are in fact autocorrelated, see [BKB$^+$14].

Writing $G = G_\infty + \tilde{G}$ with $\tilde{G} = G - G_\infty$:

$$\begin{aligned}\frac{E(v_1+v_2)^2}{T^2 v_1 v_2} &= \frac{1}{2}G_\infty(f(v_1)v_2 - f(v_2)v_1) \\ &+ \frac{(v_1+v_2)^2}{v_1 v_2}\Big(v_1 f(v_1)\int_0^\theta \frac{1}{T}\int_0^{tT}\tilde{G}(s)dsdt + v_2 f(v_2)\int_\theta^1 \frac{1}{T}\int_0^{T(t-\theta)}\tilde{G}(s)dsdt \\ &- v_2 f(v_1)\int_\theta^1 \frac{1}{T}\int_{T(t-\theta)}^{Tt}\tilde{G}(s)dsdt\Big).\end{aligned}$$

Therefore, using Cesàro's lemma (applied to $\tilde{G} \to 0$), we show that only the first term does not tend to zero when T tends to infinity and thus for E not to be strictly negative when T becomes large enough, we must have: $v_1 f(v_2) \le v_2 f(v_1)$.

Symmetrically, we get: $v_1 f(v_2) = v_2 f(v_1)$ therefore f must be linear.

VIII.A.2 Proof of Theorem 3

Let us write:

$$\theta(x) = \int_0^{+\infty} \frac{e^{iu}}{u^x}du \text{ for } x \in]0,1[\text{ and } \theta(x) = \int_0^{+\infty} \frac{e^{iu}-1}{u^x}du \text{ for } x \in]1,2[.$$

By definition of X^T, the covariance of its h-increments is linked to the covariance of the Th-increments of N^T by[11]

$$C^{X,T}_{\tau,h} = A_T^2 C^{N,T}_{T\tau,Th}.$$

Let us now compute the Fourier transform of the correlation function of the h-increments of X^T. For a fixed $z \in \mathbb{R}^*$, we have

$$\hat{C}^{X,T}_{z,h} = \int_{-\infty}^{+\infty} C^{X,T}_{\tau,h} e^{iz\tau}d\tau = \frac{A_T^2}{T}\hat{C}^{N,T}_{\frac{z}{T},Th}.$$

Theorem 1 of [BDM12] yields

$$\hat{C}^{N,T}_{\frac{z}{T},Th} = Th\frac{\frac{\mu^T}{1-a_T}\hat{g}^{Th}_{\frac{z}{T}}}{|1-\hat{\phi}^T_{\frac{z}{T}}|^2} = Th\frac{\frac{\mu^T}{1-a_T}T\hat{g}^h_z}{|1-\hat{\phi}^T_{\frac{z}{T}}|^2}$$

with $g^u_t = (1-\frac{|t|}{u})^+$.

We now state the following technical lemma.

Lemma 1. *For $z \in \mathbb{R}^*$*

$$\hat{\phi}^T_{\frac{z}{T}} = a_T\Big[1 + \theta(1+\alpha)\alpha\big(\frac{cz}{T}\big)^\alpha + o\big(\frac{1}{T^\alpha}\big)\Big].$$

[11] $C^{X,T}_{\tau,h} = \mathbb{E}[(X^T_{t+\tau+h} - X^T_{t+\tau})(X^T_{t+h} - X^T_t)] - (\mathbb{E}[X^T_{t+h} - X^T_t])^2$.

Proof. The Fourier transform writes:

$$\begin{aligned}\hat{\phi}^T_{\frac{z}{T}} &= \int_0^{+\infty} \phi^T(t) e^{it\frac{z}{T}} dt \\ &= \int_0^{+\infty} \phi^T(t) dt + \int_0^{+\infty} \phi^T(t)(e^{it\frac{z}{T}} - 1) dt \\ &= a_T\big(1 + T^{-\alpha}\int_0^{+\infty} \phi(\frac{Tu}{z})(e^{iu} - 1)\frac{T^{1+\alpha}}{z} du\big).\end{aligned}$$

However, for a fixed $u > 0$, our hypothesis on the asymptotic behavior of ϕ implies:

$$\phi(\frac{Tu}{z})(e^{iu} - 1)\frac{T^{1+\alpha}}{z} \to (e^{iu} - 1)\frac{\alpha c^\alpha z^\alpha}{u^{1+\alpha}}.$$

Moreover, since ϕ is bounded, there exists $C > 0$ such that $|\phi(x)| \leq \frac{C}{x^{1+\alpha}}$ and thus

$$|\phi(\frac{Tu}{z})(e^{iu} - 1)\frac{T^{1+\alpha}}{z}| \leq z^\alpha \frac{C}{u^{1+\alpha}}|e^{iu} - 1|.$$

We can therefore apply the dominated convergence theorem to obtain the lemma.

□

Using Lemma 1, we get that

$$\hat{C}^{N,T}_{\frac{z}{T},Th} = hT^2 \frac{\frac{\mu^T}{1-a_T}}{|\theta(1+\alpha)|^2(\frac{c}{T})^{2\alpha} a_T^2 \alpha^2} \frac{\hat{g}^h_z}{|z|^{2\alpha}}\Big[1 + o(1) + O\big(T^\alpha(1 - a_T)\big)\Big].$$

Replacing μ^T by its expression, leads to

$$\lim_{T\to+\infty} \hat{C}^{X,T}_{z,h} = \frac{hC_\mu}{|\theta(1+\alpha)|^2 c^{2\alpha}\alpha^2} \frac{\hat{g}^h_z}{|z|^{2\alpha}}.$$

Let us now state a domination lemma for $\hat{C}^{X,T}_{z,h}$.

Lemma 2. *There is a constant $C > 0$ such that for all $z \in \mathbb{R}^*$*

$$|\hat{C}^{X,T}_{z,h}| \leq C \min(\frac{1}{|z|^2}, 1)(1 + \frac{1}{z^{2\alpha}}).$$

Proof. We begin the proof by noticing that there exist $c_1 > 0$ and $c_2 > 0$ such that

$$|\hat{g}^h_z| \leq c_1 \min(\frac{1}{|z|^2}, 1)$$

and

$$|1 - \hat{\phi}^T_{\frac{z}{T}}| \geq c_2 \min(|\frac{z}{T}|^\alpha, 1).$$

If we now use these inequalities together with Theorem 1 of [BDM12] and the definition of X^T, we obtain

$$|\hat{C}^{X,T}_{z,h}| \leq \frac{hC_\mu T^{-2\alpha}}{c_2^2 \min(|\frac{z}{T}|^\alpha, 1)^2} c_1 \min(\frac{1}{|z|^2}, 1)$$

which ends the proof. □

The Fourier transform inversion formula writes

$$C_{t,h}^{X,T} = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \hat{C}_{z,h}^{X,T} e^{-itz} dz.$$

Thanks to the previous lemma, we can use the dominated convergence theorem to get that for a fixed t, $C_{t,h}^{X,T}$ has a limit when T tends to infinity and

$$C_{t,h}^{X} := \lim_{T\to+\infty} C_{t,h}^{X,T} = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \frac{hC_\mu}{|\theta(1+\alpha)|^2 c^{2\alpha}\alpha^2} \frac{\hat{g}_z^h}{|z|^{2\alpha}} e^{-itz} dz.$$

We now state the following lemma which is given by a simple change of variables.

Lemma 3. *The function whose Fourier transform is* $|z|^{-2\alpha}$ *is* $\frac{1}{2Re\left(\theta(1-2\alpha)\right)|t|^{1-2\alpha}}$.

Therefore, if we set $K = \frac{hC_\mu}{2Re\left(\theta(1-2\alpha)\right)|\theta(1+\alpha)|^2 c^{2\alpha}\alpha^2}$, then the function $t \mapsto Kg_t^h * \frac{1}{|t|^{1-2\alpha}}$ has the same Fourier transform as $C_{t,h}^X$ and thus

$$\lim_{T\to+\infty} C_{t,h}^{X,T} = Kg_t^h * \frac{1}{|t|^{1-2\alpha}}. \tag{7}$$

Computing the convolution ends the proof:

$$\lim_{T\to+\infty} C_{t,h}^{X,T} = K(|t+h|^{2H} + |t-h|^{2H} - 2|t|^{2H}) \text{ with } H = 1/2+\alpha.$$

VIII.A.3 Proof of Proposition 3

Starting from Equation (1):

$$P_t = P_0 + \kappa\nu \lim_{s\to+\infty} \mathbb{E}[N_s^a - N_s^b | \mathscr{F}_t].$$

We rewrite $N_s^{a/b} = M_s^{a/b} + \int_0^s \lambda^{a/b} du$ which implies:

$$P_t = P_0 + \kappa\nu \lim_{s\to+\infty} \mathbb{E}[M_s^a - M_s^b + \int_0^t \lambda_u^a - \lambda_u^b du + \int_t^s \lambda_u^a - \lambda_u^b du | \mathscr{F}_t].$$

We use proposition 2 to replace $\lambda^{a/b}$ and we get:

$$P_t = P_0 + \kappa\nu \lim_{s\to+\infty} \mathbb{E}[M_s^a - M_s^b + \int_0^t \lambda_u^a - \lambda_u^b du + \int_t^s \int_0^u \psi(u-x)(dM_x^a - dM_x^b) du | \mathscr{F}_t].$$

Using that $M^{a/b}$ is a martingale and that:

$$dM_x = dN_x - \lambda_x dx = dN_x - (\mu + \int_0^t \phi(x-r) dN_r) dx,$$

we get:

$$\begin{aligned} P_t &= P_0 + \kappa\nu[N_t^a - N_t^b + \int_t^{+\infty}\int_0^t \psi(u-x)(dN_x^a - dN_x^b)du \\ &+ \int_t^{+\infty}\int_0^t \psi(u-x)\int_0^x \phi(x-r)(dN_r^a - dN_r^b)dxdu]. \end{aligned}$$

Inverting the integrals in the previous equation, we get:

$$N_t^a - N_t^b = \int_0^t (dN_s^a - dN_s^b),$$

and

$$\begin{aligned} \int_t^{\infty}\int_0^t \psi(u-x)(dN_x^a - dN_x^b)du &= \int_0^t\int_t^{\infty} \psi(u-x)du(dN_x^a - dN_x^b) \\ &= \int_0^t\int_{t-s}^{\infty} \psi(u')du'(dN_s^a - dN_s^b). \end{aligned}$$

Moreover

$$\begin{aligned} &\int_t^{\infty}\int_0^t \psi(u-x)\int_0^x \phi(x-r)(dN_r^a - dN_r^b)dxdu \\ =\ &\int_0^t\int_t^{\infty}\int_r^t \psi(u-x)\phi(x-r)dxdu(dN_r^a - dN_r^b) \\ =\ &\int_0^t\int_{t-s}^{\infty}\int_0^{t-s} \psi(u'-x')\phi(x')dx'du'(dN_s^a - dN_s^b), \end{aligned}$$

where we have made the changes of variables: $u' = u - r$ and $x' = x - r$.
Replacing those three terms in the previous equation ends the proof:

$$P_t = P_0 + \kappa\nu\Big[\int_0^t (1 + \int_{t-s}^{+\infty}\psi(u)du - \int_{t-s}^{+\infty}\int_0^{t-s}\psi(u-x)\phi(x)dxdu)(dN_s^a - dN_s^b)\Big].$$

VIII.A.4 Proof of Theorem 4

We start from the impact equation (6), that we apply to our asymptotic:

$$\begin{aligned} RMI^T(t) &= F\frac{1-a_T}{(\tau^T)^{1-\alpha}}\int_0^{t\tau^T}\zeta^T(s)ds \\ &= F\frac{1-a_T}{(\tau^T)^{1-\alpha}}\int_0^t \frac{\kappa\nu\tau^T}{(1-a_T)(\tau^T)^{\alpha}}\Big[(1-a_T)(\tau^T)^{\alpha} + \int_{s\tau^T}^{+\infty}\phi^T(x)dx(\tau^T)^{\alpha}\Big]ds \\ &\to \kappa\nu F\int_0^t \frac{c^{\alpha}}{s^{1-\alpha}}ds = \frac{c^{\alpha}\kappa\nu\Phi}{1-\alpha}t^{1-\alpha}. \end{aligned}$$

Using the dominated convergence theorem with $(1-a_T)(\tau^T)^{\alpha} \to 0$ and

$$\int_{s\tau^T}^{+\infty}\phi^T(x)dx \sim c^{\alpha}/(\tau^T s)^{\alpha}$$

ends the proof.

VIII.B Numerical example for the toy model of Section 2.2

In this appendix, we try to build a concrete numerical example of the toy investor model of Section 2.2. In this model, we recall that the price is given by

$$P = \frac{\sum_{i=1}^{n} \frac{E_i}{2\lambda_i \Sigma_i} - N}{\sum_{i=1}^{n} \frac{1}{2\lambda_i \Sigma_i}}.$$

Let us consider a stock whose value is $P = 20\$$, whose volatility is $\sigma = 25\%$, such that there is a total of $N = 10^9 shares$ and that the annual turnover is 100% (so that the average daily volume is $ADV = N/250 = 4.10^6 shares/day$, taking 250 days a year). Then if the investors are identical (that is have the same E_i, Σ_i and λ_i) and expect to sell their stocks in one year with an average yearly yield of 7% (so that $E_i = 21.4\$$), the trend of the permanent impact k checks:

$$k = \frac{1}{\sum_{i=1}^{n} \frac{1}{2\lambda_i \Sigma_i}} = \frac{E_i - P}{N} = 1.4.10^{-9}\$/share$$

and the permanent impact of ΔN shares is

$$\Delta P = k \Delta N.$$

Let us compare it to the impact function of [WG13]:

$$impact(\Delta N) = 2.8.\sigma \sqrt{\frac{\Delta N}{ADV}}.$$

We find that for a metaorder of $\Delta N = 4.10^6 shares$ (one day's worth of volume), the impact function of [WG13] is equal to 0.7% while our permanent impact is equal to 0.028%.

This result is coherent with a result [WG13] which states that the permanent market impact of uninformed trades is much smaller than their impact at the end of the trade. It also allows us to get an order of magnitude of the permanent impact of such trades.

Bibliography

[AB97] T. G. Andersen and T. Bollerslev. Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance*, 4(2):115–158, 1997.

[ABDE01] T. G. Andersen, T. Bollerslev, F. X. Diebold, and H. Ebens. The distribution of realized stock return volatility. *Journal of Financial Economics*, 61(1):43–76, 2001.

[ABDL01] T. G. Andersen, T. Bollerslev, F. X. Diebold, and P. Labys. The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, 96(453):42–55, 2001.

[ABDL03] T. G. Andersen, T. Bollerslev, F. X. Diebold, and P. Labys. Modeling and forecasting realized volatility. *Econometrica*, 71(2):579–625, 2003.

[AC01] R. Almgren and N. Chriss. Optimal execution of portfolio transactions. *Journal of Risk*, 3:5–40, 2001.

[ACG07] G. J. Alexander, G. Cici, and S. Gibson. Does motivation matter when assessing trade performance? An analysis of mutual funds. *Review of Financial Studies*, 20(1):125–150, 2007.

[Ada76] L. Adamopoulos. Cluster models for earthquakes: Regional comparisons. *Journal of the International Association for Mathematical Geology*, 8(4):463–475, 1976.

[Ada97] A. Adas. Traffic models in broadband networks. *Communications Magazine, IEEE*, 35(7):82–89, 1997.

[AK12] M. B. Alaya and A. Kebaier. Parameter estimation for the square-root diffusions: Ergodic and nonergodic cases. *Stochastic Models*, 28(4):609–634, 2012.

[AS08] M. Avellaneda and S. Stoikov. High-frequency trading in a limit order book. *Quantitative Finance*, 8(3):217–224, 2008.

[ASCDL10] Y. Aït-Sahalia, J. Cacho-Diaz, and R. J. Laeven. Modeling financial contagion using mutually exciting jump processes. Technical report, National Bureau of Economic Research, 2010.

[ASMZ05] Y. Aït-Sahalia, P. A. Mykland, and L. Zhang. How often to sample a continuous-time process in the presence of market microstructure noise. *Review of Financial studies*, 18(2):351–416, 2005.

[ATHL05] R. Almgren, C. Thum, E. Hauptmann, and H. Li. Direct estimation of equity market impact. *Risk*, 57, 2005.

[BBH12] C. Blundell, J. Beck, and K. A. Heller. Modeling reciprocating relationships with Hawkes processes. In *Advances in Neural Information Processing Systems*, pages 2600–2608, 2012.

[BC+07] B. Buchmann, N. H. Chan, et al. Asymptotic theory of least squares estimators for nearly unstable processes under strong dependence. *The Annals of statistics*, 35(5):2001–2017, 2007.

[BC11] S. R. Bentes and M. M. Cruz. Is stock market volatility persistent? A fractionally integrated approach. 2011.

[BC13] B. Buchmann and N. H. Chan. Unified asymptotic theory for nearly unstable ar (p) processes. *Stochastic Processes and their Applications*, 123(3):952–985, 2013.

[BDHM12] E. Bacry, S. Delattre, M. Hoffmann, and J. F. Muzy. Scaling limits for Hawkes processes and application to financial statistics. *arXiv preprint arXiv:1202.0842*, 2012.

[BDHM13] E. Bacry, S. Delattre, M. Hoffmann, and J.-F. Muzy. Modeling microstructure noise with mutually exciting point processes. *Quantitative Finance*, 13(1):65–77, 2013.

[BDM12] E. Bacry, K. Dayri, and J.-F. Muzy. Non-parametric kernel estimation for symmetric Hawkes processes. application to high frequency financial data. *The European Physical Journal B*, 85(5):1–12, 2012.

[Ber94] J. Beran. *Statistics for long-memory processes*, volume 61. CRC Press, 1994.

[BFG15] C. Bayer, P. K. Friz, and J. Gatheral. Pricing under rough volatility. *Available at SSRN*, 2015.

[BFL09] J.-P. Bouchaud, J. D. Farmer, and F. Lillo. How markets slowly digest changes in supply and demand. *Handbook of financial markets: dynamics and evolution*, 1:57, 2009.

[BGM15] E. Bacry, S. Gaïffas, and J.-F. Muzy. A generalization error bound for sparse and low-rank multivariate Hawkes processes. *arXiv preprint arXiv:1501.00725*, 2015.

[BGPW04] 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*, 4(2):176–190, 2004.

[BGT89] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular variation*, volume 27. Cambridge university press, 1989.

[BH04] L. Bauwens and N. Hautsch. Dynamic latent factor models for intensity processes. 2004.

[BHS95] B. Biais, P. Hillion, and C. Spatt. An empirical analysis of the limit order book and the order flow in the Paris bourse. *The Journal of Finance*, 50(5):1655–1689, 1995.

[Bil09] P. Billingsley. *Convergence of probability measures*, volume 493. Wiley. com, 2009.

[BIP11] M. Barczy, M. Ispany, and G. Pap. Asymptotic behavior of unstable INAR (p) processes. *Stochastic Processes and their Applications*, 121(3):583–608, 2011.

[BKB+14] X. Brokmann, J. Kockelkoren, J.-P. Bouchaud, et al. Slow decay of impact in equity markets. *Available at SSRN 2471528*, 2014.

[BKM08] E. Bacry, A. Kozhemyak, and J.-F. Muzy. Log-normal continuous cascades: Aggregation properties and estimation. Application to financial time-series. *arXiv preprint arXiv:0804.0185*, 2008.

[BM96] P. Brémaud and L. Massoulié. Stability of nonlinear Hawkes processes. *Annals of Probability*, 24(3):1563–1588, 1996.

[BM01] P. Brémaud and L. Massoulié. Hawkes branching point processes without ancestors. *Journal of applied probability*, 38(1):122–135, 2001.

[BM03] E. Bacry and J. F. Muzy. Log-infinitely divisible multifractal processes. *Communications in Mathematical Physics*, 236(3):449–475, 2003.

[BM14a] E. Bacry and J.-F. Muzy. Hawkes model for price and trades high-frequency dynamics. *Quantitative Finance*, 14(7):1147–1166, 2014.

[BM14b] E. Bacry and J.-F. Muzy. Second order statistics characterization of Hawkes processes and non-parametric estimation. *arXiv preprint arXiv:1401.0903*, 2014.

[Bou10] J.-P. Bouchaud. Price impact. *Encyclopedia of quantitative finance*, 2010.

[Bow07] C. G. Bowsher. Modeling security market events in continuous time: Intensity based, multivariate point process models. *Journal of Econometrics*, 141(2):876–912, 2007.

[BP03] J.-P. Bouchaud and M. Potters. *Theory of financial risk and derivative pricing: From statistical physics to risk management*. Cambridge University Press, 2003.

[BR13] N. Bershova and D. Rakhlin. The non-linear market impact of large trades: Evidence from buy-side order flow. *Available at SSRN 2197534*, 2013.

[CCR12] F. Comte, L. Coutin, and E. Renault. Affine fractional stochastic volatility models. *Annals of Finance*, 8(2-3):337–378, 2012.

[CDDM05] V. Chavez-Demoulin, A. C. Davison, and A. J. McNeil. A point process approach to value-at-risk estimation. *Quantitative Finance*, 5(2):227–234, 2005.

[CDL13] R. Cont and A. De Larrard. Price dynamics in a Markovian limit order market. *SIAM Journal on Financial Mathematics*, 4(1):1–25, 2013.

[CDP06] Z. Chen, R. T. Daigler, and A. M. Parhizgari. Persistence of volatility in futures markets. *Journal of Futures Markets*, 26(6):571–594, 2006.

[Cha88] N. H. Chan. The parameter inference for nearly nonstationary time series. *Journal of the American Statistical Association*, 83(403):857–862, 1988.

[Cha09] N. H. Chan. Time series with roots on or near the unit circle. In *Handbook of Financial Time Series*, pages 695–707. Springer, 2009.

[Chr11] A. Chronopoulou. Parameter estimation and calibration for long-memory stochastic volatility models. In F. G. Viens, M. C. Mariani, and I. Florescu, editors, *Handbook of Modeling High-Frequency Data in Finance*, pages 219–231. John Wiley & Sons, 2011.

[CIR85] J. C. Cox, J. E. Ingersoll, and S. A. Ross. A theory of the term structure of interest rates. *Econometrica: Journal of the Econometric Society*, pages 385–407, 1985.

[CKM03] P. Cheridito, H. Kawaguchi, and M. Maejima. Fractional Ornstein-Uhlenbeck processes. *Electron. J. Probab*, 8(3):14, 2003.

[CKS14] R. Cont, A. Kukanov, and S. Stoikov. The price impact of order book events. *Journal of financial econometrics*, 12(1):47–88, 2014.

[Con07] R. Cont. Volatility clustering in financial markets: Empirical facts and agent-based models. In G. Teyssière and A. P. Kirman, editors, *Long Memory in Economics*, pages 289–309. Springer Berlin Heidelberg, 2007.

[Cor09] F. Corsi. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196, 2009.

[CR98] F. Comte and E. Renault. Long memory in continuous-time stochastic volatility models. *Mathematical Finance*, 8(4):291–323, 1998.

[CS82] W. Cumberland and Z. Sykes. Weak convergence of an autoregressive process used in modeling population growth. *Journal of Applied Probability*, pages 450–455, 1982.

[CS07] J. Coval and E. Stafford. Asset fire sales (and purchases) in equity markets. *Journal of Financial Economics*, 86(2):479–512, 2007.

[CSK88] E. Chornoboy, L. Schramm, and A. Karr. Maximum likelihood identification of neural point process systems. *Biological cybernetics*, 59(4-5):265–275, 1988.

[CV12] A. Chronopoulou and F. G. Viens. Estimation and pricing under long-memory stochastic volatility. *Annals of Finance*, 8(2-3):379–403, 2012.

[CW87] N. H. Chan and C.-Z. Wei. Asymptotic inference for nearly nonstationary AR(1) processes. *The Annals of Statistics*, pages 1050–1063, 1987.

[CW03] P. Carr and L. Wu. What type of process underlies options? A simple robust test. *Journal of Finance*, 58(6):2581–2610, 2003.

[CZ09] N. H. Chan and R.-M. Zhang. Inference for nearly nonstationary processes under strong dependence with infinite variance. *Statistica Sinica*, 19(3):925, 2009.

[DBMB14] J. Donier, J. F. Bonart, I. Mastromatteo, and J.-P. Bouchaud. A fully consistent, minimal model for non-linear market impact. *Minimal Model for Non-Linear Market Impact (November 29, 2014)*, 2014.

[DGE93] Z. Ding, C. W. Granger, and R. F. Engle. A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1(1):83–106, 1993.

[Don12] J. Donier. Market impact with autocorrelated order flow under perfect competition. *Available at SSRN 2191660*, 2012.

[DOT03] P. Doukhan, G. Oppenheim, and M. S. Taqqu. *Theory and applications of long-range dependence*. Springer, 2003.

[DR12] K. Dayri and M. Rosenbaum. Large tick assets: implicit spread and optimal tick size. *arXiv preprint arXiv:1207.6325*, 2012.

[DRR13] S. Delattre, C. Y. Robert, and M. Rosenbaum. Estimating the efficient price from the order flow: a Brownian Cox process approach. *Stochastic Processes and their Applications*, 2013.

[DRV12] J. Duchon, R. Robert, and V. Vargas. Forecasting volatility with the multifractal random walk model. *Mathematical Finance*, 22(1):83–108, 2012.

[Dup94] B. Dupire. Pricing with a smile. *Risk Magazine*, 7(1):18–20, 1994.

[DVJ02] D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes*. Springer, 2002.

[EBK12] Z. Eisler, J.-P. Bouchaud, and J. Kockelkoren. The price impact of order book events: market orders, limit orders and cancellations. *Quantitative Finance*, 12(9):1395–1419, 2012.

[EGG10] E. Errais, K. Giesecke, and L. R. Goldberg. Affine point processes and portfolio credit risk. *SIAM Journal on Financial Mathematics*, 1(1):642–665, 2010.

[ELL11] P. Embrechts, T. Liniger, and L. Lin. Multivariate Hawkes processes: An application to financial data. *Journal of Applied Probability*, 48:367–378, 2011.

[FGLW13] J. D. Farmer, A. Gerig, F. Lillo, and H. Waelbroeck. How efficiency shapes market impact. *Quantitative Finance*, 13(11):1743–1758, 2013.

[FL04] J. D. Farmer and F. Lillo. On the origin of power-law tails in price fluctuations. *Quantitative Finance*, 4(1):7–11, 2004.

[FS12] V. Filimonov and D. Sornette. Quantifying reflexivity in financial markets: Toward a prediction of flash crashes. *Physical Review E*, 85(5):056108, 2012.

[FS13] V. Filimonov and D. Sornette. Apparent criticality and calibration issues in the Hawkes self-excited point process model: application to high-frequency financial data. *arXiv preprint arXiv:1308.6756*, 2013.

[Fuk11] M. Fukasawa. Asymptotic analysis for stochastic volatility: Martingale expansion. *Finance and Stochastics*, 15(4):635–654, 2011.

[Fuk15] M. Fukasawa. Short-time at-the-money skew and rough fractional volatility. *arXiv preprint arXiv:1501.06980*, 2015.

[Gat06] J. Gatheral. *The volatility surface: A practitioner's guide*, volume 357. John Wiley & Sons, 2006.

[Gat10] J. Gatheral. No-dynamic-arbitrage and market impact. *Quantitative Finance*, 10(7):749–759, 2010.

[GDKB13] A. Gareche, G. Disdier, J. Kockelkoren, and J.-P. Bouchaud. Fokker-Planck description for the queue dynamics of large tick stocks. *Physical Review E*, 88(3):032809, 2013.

[GGPS03] X. Gabaix, P. Gopikrishnan, V. Plerou, and H. E. Stanley. A theory of power-law distributions in financial market fluctuations. *Nature*, 423(6937):267–270, 2003.

[GGPS06] X. Gabaix, P. Gopikrishnan, V. Plerou, and H. E. Stanley. Institutional investors and stock market volatility. *The Quarterly Journal of Economics*, 121(2):461–504, 2006.

[GJ14] J. Gatheral and A. Jacquier. Arbitrage-free SVI volatility surfaces. *Quantitative Finance*, 14(1):59–71, 2014.

[GK58] I. Gokhberg and M. G. Krein. Systems of integral equations on the half-line with kernels depending on the difference of the arguments. *Uspekhi matematicheskikh nauk*, 13(2):3–72, 1958.

[GLFT12] O. Guéant, C.-A. Lehalle, and J. Fernandez-Tapia. Dealing with the inventory risk: A solution to the market making problem. *Mathematics and Financial Economics*, pages 1–31, 2012.

[GM85] L. R. Glosten and P. R. Milgrom. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of financial economics*, 14(1):71–100, 1985.

[GO10] J. Gatheral and R. C. Oomen. Zero-intelligence realized variance estimation. *Finance and Stochastics*, 14(2):249–283, 2010.

[GSS12] J. Gatheral, A. Schied, and A. Slynko. Transient linear price impact and Fredholm integral equations. *Mathematical Finance*, 22(3):445–474, 2012.

[Has91] J. Hasbrouck. Measuring the information content of stock trades. *The Journal of Finance*, 46(1):179–207, 1991.

[Haw71a] A. G. Hawkes. Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 438–443, 1971.

[Haw71b] A. G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971.

[HBB13] S. J. Hardiman, N. Bercot, and J.-P. Bouchaud. Critical reflexivity in financial markets: A Hawkes process analysis. *The European Physical Journal B*, 86(10):1–9, 2013.

[HDB13] P. F. Halpin and P. De Boeck. Modeling dyadic interaction with Hawkes processes. *Psychometrika*, 78(4):793–814, 2013.

[Hes93] S. L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of financial studies*, 6(2):327–343, 1993.

[Hew06] P. Hewlett. Clustering of order arrivals, price impact and trade path optimisation. In *Workshop on Financial Modeling with Jump processes, Ecole Polytechnique*, pages 6–8, 2006.

[Hir11] N. Hirschey. Do high frequency traders anticipate buying and selling pressure. *Work. Pap., McCombs Sch. Bus., Univ. Texas Austin*, 2011.

[HKLW02] P. S. Hagan, D. Kumar, A. S. Lesniewski, and D. E. Woodward. Managing smile risk. *Wilmott Magazine*, pages 84–108, 2002.

[HLR13] W. Huang, C.-A. Lehalle, and M. Rosenbaum. Simulating and analyzing order book data: The queue-reactive model. *arXiv preprint arXiv:1312.0563*, 2013.

[HMS11] H. J. Haubold, A. M. Mathai, and R. K. Saxena. Mittag-Leffler functions and their applications. *Journal of Applied Mathematics*, 2011.

[HO74] A. G. Hawkes and D. Oakes. A cluster process representation of a self-exciting process. *Journal of Applied Probability*, pages 493–503, 1974.

[HRBR12] N. R. Hansen, P. Reynaud-Bouret, and V. Rivoirard. Lasso and probabilistic inequalities for multivariate point processes. *arXiv preprint arXiv:1208.0570*, 2012.

[HS97] R. D. Huang and H. R. Stoll. The components of the bid-ask spread: A general approach. *Review of Financial Studies*, 10(4):995–1034, 1997.

[HS04] G. Huberman and W. Stanzl. Price manipulation and quasi-arbitrage. *Econometrica*, 72(4):1247–1275, 2004.

[HW93] J. Hull and A. White. One-factor interest-rate models and the valuation of interest-rate derivative securities. *Journal of Financial and Quantitative Analysis*, 28(02):235–254, 1993.

[Jac75] J. Jacod. Multivariate point processes: predictable projection, Radon-Nikodym derivatives, representation of martingales. *Probability Theory and Related Fields*, 31(3):235–253, 1975.

[Jeg91] P. Jeganathan. On the asymptotic behavior of least-squares estimators in AR time series with roots near the unit circle. *Econometric Theory*, 7(03):269–306, 1991.

[JMP89] A. Jakubowski, J. Mémin, and G. Pagès. Convergence en loi des suites d'intégrales stochastiques sur l'espace 1 de Skorokhod. *Probability Theory and Related Fields*, 81(1):111–137, 1989.

[JR13] T. Jaisson and M. Rosenbaum. Limit theorems for nearly unstable Hawkes processes. Version with technical appendix. *Available at https://hal.archives-ouvertes.fr/ccsd-00935038/*, 2013.

[JS87] J. Jacod and A. N. Shiryaev. *Limit theorems for stochastic processes*, volume 288. Springer-Verlag Berlin, 1987.

[JZ13] R. Johnson and T. Zhang. Accelerating stochastic gradient descent using predictive variance reduction. In *Advances in Neural Information Processing Systems*, pages 315–323, 2013.

[Kal97] V. V. Kalashnikov. *Geometric sums: bounds for rare events with applications: risk analysis, reliability, queueing*, volume 413. Springer, 1997.

[KO11] A. P. Kyle and A. Obizhaeva. Market microstructure invariants: Theory and implications of calibration. *Available at SSRN 1978932*, 2011.

[KP91] T. G. Kurtz and P. Protter. Weak limit theorems for stochastic integrals and stochastic differential equations. *The Annals of Probability*, 19(3):1035–1070, 1991.

[Kyl85] A. S. Kyle. Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, pages 1315–1335, 1985.

[Lar07] J. Large. Measuring the resiliency of an electronic limit order book. *Journal of Financial Markets*, 10(1):1–25, 2007.

[LC14] M. Lallouache and D. Challet. Statistically significant fits of Hawkes processes to financial data. *Available at SSRN 2450101*, 2014.

[LF04] F. Lillo and J. D. Farmer. The long memory of the efficient market. *Studies in Nonlinear Dynamics & Econometrics*, 8(3), 2004.

[LLL+13] J.-M. Lasry, C.-A. Lehalle, P.-L. Lions, et al. Efficiency of the price formation process in presence of high frequency participants: a mean field game analysis. Technical report, 2013.

[LM11] E. Lewis and G. Mohler. A nonparametric EM algorithm for multiscale Hawkes processes. *preprint*, 2011.

[LMBB12] E. Lewis, G. Mohler, P. J. Brantingham, and A. L. Bertozzi. Self-exciting point process models of civilian deaths in Iraq. *Security Journal*, 25(3):244–264, 2012.

[LMF05] F. Lillo, S. Mike, and J. D. Farmer. Theory for long memory in supply and demand. *Physical Review E*, 71(6):066122, 2005.

[LV14] R. Lemonnier and N. Vayatis. Nonparametric markovian learning of triggering kernels for mutually exciting and mutually inhibiting multivariate Hawkes processes. In *Machine Learning and Knowledge Discovery in Databases*, pages 161–176. Springer, 2014.

[LZ13] L. Li and H. Zha. Dyadic event attribution in social networks with mixtures of Hawkes processes. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, pages 1667–1672. ACM, 2013.

[M+13] G. Mohler et al. Modeling and estimation of multi-source clustering in crime and security data. *The Annals of Applied Statistics*, 7(3):1525–1539, 2013.

[Mar13] J. Markevičiūtė. *Asymptotic results on nearly nonstationary processes.* PhD thesis, Vilniaus universitetas, 2013.

[MH08] A. Mathai and H. J. Haubold. Mittag-Leffler functions and fractional calculus. *Special Functions for Applied Scientists*, pages 79–134, 2008.

[MO06] R. V. Mendes and M. J. Oliveira. A data-reconstructed fractional volatility model. *arXiv preprint math/0602013*, 2006.

[MPS04] M. Mitchell, T. Pulvino, and E. Stafford. Price pressure around mergers. *The Journal of Finance*, 59(1):31–63, 2004.

[MR06] M. Musiela and M. Rutkowski. *Martingale methods in financial modeling*, volume 36. Springer, 2006.

[MRR97] A. Madhavan, M. Richardson, and M. Roomans. Why do security prices change? A transaction-level analysis of NYSE stocks. *Review of Financial Studies*, 10(4):1035–1064, 1997.

[MS00a] R. N. Mantegna and H. E. Stanley. *Introduction to econophysics: Correlations and complexity in finance.* Cambridge University Press, 2000.

[MS00b] T. Mikosch and C. Stărică. Is it really long memory we see in financial returns. In P. Embrechts, editor, *Extremes and integrated risk management*, pages 149–168. Risk Books, 2000.

[MSB+11] G. O. Mohler, M. B. Short, P. J. Brantingham, F. P. Schoenberg, and G. E. Tita. Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, 106(493), 2011.

[MSR12] J. Markevičiūtė, C. Suquet, and A. Račkauskas. Functional central limit theorems for sums of nearly nonstationary processes. *Lithuanian mathematical journal*, 52(3):282–296, 2012.

[MTB14] I. Mastromatteo, B. Toth, and J.-P. Bouchaud. Agent-based models for latent liquidity and concave price impact. *Physical Review E*, 89(4):042805, 2014.

[MVM+09] E. Moro, J. Vicente, L. G. Moyano, A. Gerig, J. D. Farmer, G. Vaglica, F. Lillo, and R. N. Mantegna. Market impact and trading profile of hidden orders in stock markets. *Physical Review E*, 80(6):066102, 2009.

[MVN68] B. B. Mandelbrot and J. W. Van Ness. Fractional Brownian motions, fractional noises and applications. *SIAM review*, 10(4):422–437, 1968.

[NP00] C. J. Nuzman and V. H. Poor. Linear estimation of self-similar processes via Lamperti's transformation. *Journal of Applied Probability*, 37(2):429–452, 2000.

[NV99] A. Novikov and E. Valkeila. On some maximal inequalities for fractional Brownian motions. *Statistics & Probability Letters*, 44(1):47–54, 1999.

[Nys30] E. J. Nyström. Über die praktische auflösung von integralgleichungen mit anwendungen auf randwertaufgaben. *Acta Mathematica*, 54(1):185–204, 1930.

[Obi07] A. Obizhaeva. Information vs. liquidity: Evidence from portfolio transition trades. *Manuscript, University of Maryland*, 2007.

[Oga78] Y. Ogata. The asymptotic behavior of maximum likelihood estimators for stationary point processes. *Annals of the Institute of Statistical Mathematics*, 30(1):243–261, 1978.

[Oga81] Y. Ogata. On Lewis' simulation method for point processes. *Information Theory, IEEE Transactions on*, 27(1):23–31, 1981.

[Oga83] Y. Ogata. Likelihood analysis of point processes and its application to seismological data. *Bull. Int. Statist. Inst*, 50:943–961, 1983.

[Oga99] Y. Ogata. Seismicity analysis through point-process modeling: A review. *Pure and applied geophysics*, 155(2-4):471–507, 1999.

[Pet75] V. V. Petrov. *Sums of independent random variables*. Berlin, 1975.

[Phi87] P. C. Phillips. Towards a unified asymptotic theory for autoregression. *Biometrika*, 74(3):535–547, 1987.

[PSCR11] V. Pernice, B. Staude, S. Cardanobile, and S. Rotter. How structure determines correlations in neuronal networks. *PLoS computational biology*, 7(5):e1002059, 2011.

[PSCR12] V. Pernice, B. Staude, S. Cardanobile, and S. Rotter. Recurrent interactions in spiking networks with arbitrary topology. *Physical review E*, 85(3):031916, 2012.

[RBRGTM14] P. Reynaud-Bouret, V. Rivoirard, F. Grammont, and C. Tuleau-Malot. Goodness-of-fit tests and nonparametric adaptive estimation for spike train analysis. *The Journal of Mathematical Neuroscience (JMN)*, 4(1):1–41, 2014.

[RBRTM13] P. Reynaud-Bouret, V. Rivoirard, and C. Tuleau-Malot. Inference of functional connectivity in neurosciences via Hawkes processes. In *1st IEEE Global Conference on Signal and Information Processing*, 2013.

[RBS+10] P. Reynaud-Bouret, S. Schbath, et al. Adaptive estimation for Hawkes processes; application to genome analysis. *The Annals of Statistics*, 38(5):2781–2822, 2010.

[Ros08] M. Rosenbaum. Estimation of the volatility persistence in a discretely observed diffusion model. *Stochastic Processes and their Applications*, 118(8):1434–1462, 2008.

[Ros09a] M. Rosenbaum. First order p-variations and Besov spaces. *Statistics & Probability Letters*, 79(1):55–62, 2009.

[Roş09b] I. Roşu. A dynamic model of the limit order book. *Review of Financial Studies*, 22(11):4601–4641, 2009.

[Ros11] M. Rosenbaum. A new microstructure noise index. *Quantitative Finance*, 11(6):883–899, 2011.

[Roy92] T. Royama. *Analytical population dynamics*, volume 10. Springer, 1992.

[RPL14] M. Rambaldi, P. Pennesi, and F. Lillo. Modeling FX market activity around macroeconomic news: A Hawkes process approach. *arXiv preprint arXiv:1405.6047*, 2014.

[RR11] 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 Econometrics*, 9(2):344–366, 2011.

[RR12] C. Y. Robert and M. Rosenbaum. Volatility and covariation estimation when microstructure noise and trading times are endogenous. *Mathematical Finance*, 22(1):133–164, 2012.

[RY99] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293. Springer, 1999.

[SFGK03] E. Smith, J. D. Farmer, L. Gillemot, and S. Krishnamurthy. Statistical theory of continuous double auction. *Quantitative Finance*, 3:481–514, 2003.

[Sha64] W. F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, 19(3):425–442, 1964.

[Shi80] R. J. Shiller. Do stock prices move too much to be justified by subsequent changes in dividends? *National Bureau of Economic Research Cambridge, Mass., USA*, 1980.

[SJ13] P. Shah and R. Jana. Results on generalized Mittag-Leffler function via Laplace transform. *Applied Mathematical Sciences*, 7(12):567–570, 2013.

[SKM93] S. G. Samko, A. A. Kilbas, and O. I. Marichev. Fractional integrals and derivatives. *Theory and Applications, Gordon and Breach, Yverdon*, 1993.

[Slo81] I. H. Sloan. Quadrature methods for integral equations of the second kind over infinite intervals. *Mathematics of Computation*, 36(154):511–523, 1981.

[TLD+11] B. Toth, Y. Lemperiere, C. Deremble, J. De Lataillade, J. Kockelkoren, and J.-P. Bouchaud. Anomalous price impact and the critical nature of liquidity in financial markets. *Physical Review X*, 1(2):021006, 2011.

[TLF10] B. Toth, F. Lillo, and J. D. Farmer. Segmentation algorithm for non-stationary compound Poisson processes. *The European Physical Journal B*, 78(2):235–243, 2010.

[TPLF11] B. Toth, I. Palit, F. Lillo, and J. D. Farmer. Why is order flow so persistent? *arXiv preprint arXiv:1108.1632*, 2011.

[vdMPvZ99] T. van der Meer, G. Pap, and M. C. van Zuijlen. Asymptotic inference for nearly unstable AR(p) processes. *Econometric Theory*, 15(02):184–217, 1999.

[Ver12] M. Veraar. The stochastic Fubini theorem revisited. *Stochastics An International Journal of Probability and Stochastic Processes*, 84(4):543–551, 2012.

[WBK+08] M. Wyart, J.-P. Bouchaud, J. Kockelkoren, M. Potters, and M. Vettorazzo. Relation between bid–ask spread, impact and volatility in order-driven markets. *Quantitative Finance*, 8(1):41–57, 2008.

[WG13] H. Waelbroeck and C. Gomes. Is market impact a measure of the information value of trades? Market response to liquidity vs. informed trades. *Market Response to liquidity vs. Informed Trades (July 9 2013)*, 2013.

[Zhu13] L. Zhu. Nonlinear Hawkes processes. *arXiv preprint arXiv:1304.7531*, 2013.

[ZZS13] K. Zhou, H. Zha, and L. Song. Learning social infectivity in sparse low-rank networks using multi-dimensional Hawkes processes. In *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics*, pages 641–649, 2013.