



# Jump-Diffusion Models for Financial Bubbles Modelling: A Multi-scale Type-II Bubble Model With Self-Excited Crashes

*Author:* [Niels Cariou-Kotlarek](#)

Professor : [Patrick Cheridito](#), [Didier Sornette](#), [Alexander Wehrli](#)<sup>1</sup>

Paper's type : Master's Thesis

EIDGENÖSSISCHE TECHNISCHE HOCHSCHULE ZÜRICH

DEPARTMENT OF MATHEMATICS  
DEPARTMENT OF MANAGEMENT, TECHNOLOGY AND ECONOMICS

Date: September 12, 2022

---

<sup>1</sup> Alphabetically ordered.



## **Abstract**

The thesis contributes to the ongoing efforts to develop more accurate and effective models of financial bubbles. We present a novel approach to modelling financial bubbles by developing a type-II bubble model aiming at presenting both positive and negative bubble episodes. We achieve this by considering a process with features such as positive feedback, self-excitation effects, non-stationary dynamics as well as multiple regimes, integral to the complex interactions in modern, computerized markets. The first part of the thesis outlines a novel extension of the integer-valued autoregressive process from its univariate form to the multidimensional setting. This extension serves as the underpinning for the proposed bubble model. Specifically, we introduce the INVAR process as a generalisation of the integer-valued autoregressive process (INAR) to the multivariate domain to discretise the Hawkes processes used in the bubble model. In the second part, known stylized facts of financial markets are discussed, and we explain the use of temporal point processes and rough models in finance. The third part presents a first step in constructing type-II bubble model and compares it to current literature, highlighting its unique features, such as the use of a bivariate Hawkes process that has both upward and downward jumps, a multi-scale mispricing index and a regime process. The analysis reveals that adding opposite sign Hawkes processes is not a promising approach to bubble modelling, however some of the tools we developed could be useful to improve existing bubble models.



## Acknowledgements

Studying financial bubbles under the guidance of my supervisors has been an honour.

I am deeply thankful to Professor Sornette, Professor Cheridito and Dr Wehrli for the great opportunity to be part of their research project. They have been supportive and were always available to give me great feedback. I would also like to highlight how grateful I am to Professor Sornette for his patience, warmth and support throughout the thesis; we had the opportunity to get into deep conversations about academics, research, and personal projects.

I would also like to thank ETH for offering the opportunity to write my Master's thesis on a subject that connects two of the topics I am the most interested in, namely mathematics and finance.

Finally, I am grateful for the pieces of advice and thoughtful discussions I had with my friends and the support from my family throughout the thesis. In particular, I am thinking of Émir, Bianca, Robin, Heike and Erol.

I am now looking forward to my next opportunities, which would have not happened without all the help I have received so far.

Niels D. C. Cariou-Kotlarek



# CONTENTS

<b>Introduction</b>	<b>1</b>
<b>1 Invar Processes</b>	<b>3</b>
1.1 Introduction . . . . .	3
1.2 Motivation: A Simpler Existence Condition . . . . .	4
1.3 INVAR Theory . . . . .	6
1.3.1 Setting . . . . .	6
1.3.2 The INVAR( $p$ ) Process . . . . .	7
1.3.3 The INVAR( $\infty$ ) Process, Existence and Properties . . . . .	8
1.3.4 Convergence of INVAR Processes to Hawkes Processes . . . . .	10
1.4 Numerical Study of INVAR( $\infty$ ) Processes . . . . .	11
1.4.1 Synthetic Time Series Experiments . . . . .	11
1.4.2 Observation . . . . .	12
1.4.3 Edge Effect . . . . .	12
<b>2 Properties of Financial Markets</b>	<b>15</b>
Introduction . . . . .	15
2.1 Properties of Financial Data . . . . .	15
2.1.1 Known Properties of Financial Data . . . . .	15
2.1.2 Modelling is a Sequence of Dilemmas . . . . .	18
2.1.3 Correlation Inside the Data and (Spurious) Long Memory . . . . .	20
2.1.4 Market Impact, Memory and Roughness . . . . .	20
2.1.5 Volatility, Memory and Roughness . . . . .	22
2.1.6 Being Stationary and the Lack Thereof . . . . .	24
2.1.7 Jump and Rough Models for Asset Pricing . . . . .	26
2.2 Point Processes Used for Financial Data . . . . .	27
2.2.1 Endo-Exo Impact, a Long Lasting Spurious Relationship . . . . .	27
2.2.2 Criticality of the Markets . . . . .	29
2.2.3 Non-Stationarity for TPP . . . . .	30
<b>3 Modelling Financial Bubbles</b>	<b>31</b>
Introduction . . . . .	31
3.1 What are Financial Bubbles? . . . . .	32

3.1.1	The Universality of Bubbles . . . . .	32
3.1.2	Dragon Kings, the Hope of Predicting Future Catastrophes . . . . .	32
3.1.3	Diagnosing Bubbles . . . . .	33
3.1.4	Financial Bubbles . . . . .	33
3.1.5	Are There No Universality to Bubbles? . . . . .	34
3.1.6	Rational Expectation Bubbles . . . . .	35
3.1.7	Type-I and II Financial Bubbles . . . . .	35
3.1.8	Financial Bubbles Stylized Facts . . . . .	36
3.1.9	Non-Local Mispricing Index Used in Type-I Bubble Models . . . . .	37
3.2	Our New Type-II Bubble . . . . .	38
3.2.1	Idea of the Model . . . . .	38
3.2.2	Model's Equations . . . . .	39
3.2.3	A Bubbly Market of Type-II . . . . .	40
3.2.4	Mathematical Framework . . . . .	41
3.2.5	Discrete Time Grid Model . . . . .	41
3.2.6	Synthetic Bubbles and Crashes . . . . .	41
3.2.7	Analysis of the Type-I Bubble Model . . . . .	42
3.2.8	Upward Jumps Added to the Model . . . . .	43
3.2.9	Drawdown and Bubble Regimes . . . . .	44
3.2.10	Do Negative Bubbles Appear? . . . . .	45
3.2.11	Further Improvements to the Model . . . . .	46
3.3	Real Experiment . . . . .	48
3.3.1	Mimicking Real Data . . . . .	48
3.3.2	Outcome of the Comparison of the Models . . . . .	50
	<b>Outline of the Work of this Thesis</b>	<b>53</b>
<b>A</b>	<b>Temporal Point Processes (TPP)</b>	<b>55</b>
A.1	Point Process Theory . . . . .	55
A.2	Lévy Process and Jump-Diffusion Model . . . . .	57
A.3	Matrices, Eigenvalues, Matrix Valued Polynomials and Norms . . . . .	58
A.4	Matrix Analysis . . . . .	58
A.5	Polynomial Eigenvalue Problem . . . . .	59
A.6	Lemmas for Spectral Radius Orderings . . . . .	60
<b>B</b>	<b>Proofs for INVAR processes</b>	<b>63</b>
B.1	Definition . . . . .	63
B.2	Proof Existence INVAR . . . . .	63
B.3	Proof VAR Representation of INVAR . . . . .	67
B.4	Proof Covariance Function of an INVAR . . . . .	69
B.5	Proof Weak Convergence from INVAR to Hawkes . . . . .	70
B.6	Lemmas and Proofs . . . . .	74
<b>C</b>	<b>EM Algorithm</b>	<b>79</b>
C.1	Multidimensional Equations . . . . .	79
C.2	Truncating the Kernel . . . . .	80
<b>D</b>	<b>Additional Discussions for the Type-II Model</b>	<b>81</b>
D.1	Stopping Time and Regime Process . . . . .	81
D.1.1	Mathematical Framework of the Regime Process . . . . .	81
D.1.2	Phony Successive Changes of Regimes . . . . .	82
D.1.3	Bubbly Market Regimes . . . . .	82
D.1.4	Multi-Scale Regime Switching . . . . .	83
D.2	Hindsight Bias . . . . .	85
D.2.1	What is the Realism of the Investors? . . . . .	85

D.2.2	Delayed Compensation . . . . .	85
D.2.3	Transient Overcompensation . . . . .	87
D.3	Mispricing Index . . . . .	88
D.3.1	A Difference Equation for the Mispricing Index . . . . .	88
D.3.2	Multi-Scale Mispricing Index . . . . .	89
D.3.3	Numerical Experiments . . . . .	91
Bibliography		<b>104</b>



# INTRODUCTION

The study of bubbles and crises has been the subject of extensive research in natural sciences before being explored in finance. The efficient market hypothesis, which postulates that asset prices reflect all available information, has long been a cornerstone of finance theory and has served as a barrier to the study of bubbles in financial markets. However, recent evidence has shown that markets exhibit phases of inefficiency, and bubbles are now widely recognized as a natural phenomenon in financial markets [188]. The possible origins, underlying mechanisms and dynamics of bubbles have been intensively debated by researchers, practitioners and policymakers [113, 172, 192] due to their impact, which are commonly referred to as "Black Swans" [195]. Despite that, functional models capturing accurately the growth and development of bubbles are still lacking, and the mathematical and financial literature has yet to understand all the inherent properties of bubbles. This thesis aims to improve the current state-of-the-art models of financial bubbles by developing novel tools and ideas. While many mathematical models have been developed to describe and predict the behaviour of bubbles, the search for a model that accurately reproduces the variety of stylized facts of real-world financial bubbles is still ongoing. In this thesis, we focus on multivariate binned counting processes and use them to create a novel type-II bubble. Most, if not all successful bubble models in the last decade have been of type-I, and our research is the first of its kind on the topic of type-II bubbles. Type-II bubbles are more flexible than their type-I counterparts but are also more difficult to model and calibrate. They are also considered are more natural as they do not require the assumption of omniscience from investors or that rational market participants are aware of the risk of a crash contrary to type-I bubbles. We generalised the type-I bubble model proposed by Malevergne and al. [131], by incorporating both positive and negative jumps. Our hope is that this would enable us to find a useful example for the class of type-II bubbles and reproduce more stylized facts of real-world financial bubbles, including negative bubbles. However, our analysis reveals that our model was not able to conclusively reproduce stylized facts not already covered by simpler models. Specifically, we note that we are missing in the model a mechanism to produce faster-than-exponential losses which would create negative bubbles. We tried variants of that model, but it seems that this path is not conclusive. Therefore, our results suggest that adding opposite signs Hawkes processes is not a promising approach to bubble modelling, at least not in our setting. The question of how to reproduce the stylized facts of bubbles remains an exciting open and challenging research question. Nevertheless, this thesis contributes to the ongoing efforts to develop more accurate and effective models of financial bubbles, and we also propose other mechanisms that could be added to type-I bubbles to improve their expressiveness.

Note that traditionally, bubbles have been modelled using the closing day to day mid-price, which limits the use of continuous models. In order to discretise Hawkes processes we use their INVAR approximation, which is a generalization of the integer-valued autoregressive process (INAR) to the multivariate case. While there has been growing interest in integer-valued processes, including INAR processes, little research has been conducted on their multivariate counterparts. We show that this generalization still holds when considering

"infinite memory", meaning the current value of the process depends on the whole past history. Drawing on important lessons from other fields, such as time series analysis and econometrics, we demonstrate that the INVAR process is of utmost importance, as the INAR( $\infty$ ) process is the discrete analogue of the Hawkes process, which is the prototypical self-excited point process.

While the properties of integer-valued AR and VAR processes with limited memory (for  $p < \infty$ ) have been extensively studied in the literature, little research has focused on infinite memory (for  $p = \infty$ ). We prove an important theorem concerning the convergence in law of INVAR processes to multivariate Hawkes processes when the mesh decreases to 0. In that sense, the INVAR process has application in any field where Hawkes processes are used: online communities and network connections [164, 179, 180], healthcare criminology, seismology [33, 163], finance [92, 162, 205], LOB modelling [1, 9, 11] but also for general binned time series [4, 79, 111, 169, 176]. As we see in Chapter 2, Hawkes processes are commonly used in finance for modelling both micro and macroscopic behaviour, which makes the INVAR process a useful models having applications in a variety of scenarios.

In Chapter 1, we prove that the properties true for univariate INAR( $\infty$ ) processes naturally generalise to multiple dimensions. For example, both INAR and multivariate INAR processes benefit from an ARMA representation and from the convergence in law to univariate or multivariate Hawkes processes. Our intuition stems from the theory of INAR processes [117, 122] and also from the theory of VAR processes [31]. Our research leads us to propose a novel way to characterise stable nonnegative (in the sense that the coefficients are all positive) VAR processes, a sufficient condition on the spectral radius of the sum of the matrix coefficient, which was the key to connect the theory between INVAR( $p$ ) and INVAR( $\infty$ ) processes.

Chapter 2 of this thesis focuses on financial markets and summarizes a decade-long work of our research group. We highlight the ways in which temporal point processes have reshaped the way we understand markets [69], particularly at the level of their microstructure and agent interaction, as well as their mesostructure properties [172, 174, 205, 206, 211, 212]. We discuss the current opposition between jump and rough models [17, 49, 80, 81, 128] and the interesting prospect of using quadratic and exponential Hawkes process [21, 51]. These questions provide valuable insights into what an appropriate model is, what purposes it must serve, and what theoretical considerations are needed to show its correctness. Particularly, the endo-exo problem, discussed in Subsection 2.2.1, provides a valuable perspective on bubble modelling.

Building on the tools developed in Chapter 1 and the intuition built in Chapter 2, the third and final chapter details our research on modelling bubbles using a bivariate jump process. This model is the first of its kind by being a type-II bubble and is developed upon other type-I bubble models such as [131, 209]. We also explore and propose a new regime model in Subsection 3.2.9, and improve on the mispricing index used in the literature of rational-expectation bubbles in Appendix D.3.

We have gathered the code for the experiments in a [public GitHub repository](#) where all experiments can be reproduced.

# CHAPTER

## 1

# INVAR PROCESSES

*This chapter is based on [34], our current work in progress with Alexander Wehrli, one of the supervisors of this thesis. The experiments related to this chapter are unavailable at the time of writing, but will be soon with the preprint released.*

## 1.1 Introduction

(Vector) AR processes are one the most successful, flexible and easy to use models for the analysis of (multivariate) time series [219]. Introduced by Christopher Sims a few decades ago, VAR models have gained a permanent place in applied macroeconomics as well as other fields [181]. One of the rationale for their resilience regardless of the field is the Wold theorem which states that any vector of time series has a VAR representation under mild regularity conditions. The development of integer-valued (counting) time series models was founded on that basis to model counting phenomena and they also take the form of general system of difference equations:  $X_n - \sum \alpha_k X_{n-k} = \varepsilon_n + \sum \beta_k \varepsilon_{n-k}$ , for  $n \in \mathbb{Z}$ , constrained such that the difference equations yield integer-valued processes.

The first contributions to integer-valued time series models were the discrete time series generated by mixtures [101, 102, 103]. Then, McKenzie [138] and Al-Osh with Alzaid [157] defined the INAR process. The modern definition of the INAR( $p$ ),  $p < \infty$  process is attributed to [109] where existence was proven, later extended to  $p = \infty$  in [117]. It is now an accepted model proposed in textbooks [77]. The connexion between processes and integer-valued processes has been a growing field [15, 29, 30, 37, 76, 174, 204]. A recent review presenting the latest advances in this field for the interested reader lies in [76]. It is particularly noteworthy that in [117], Kirchner established the fundamental relationship between INAR processes and Hawkes processes, proving that infinite memory INAR processes are the discrete analogue of Hawkes processes, converging in law as the bin size approaches zero. This highlights the connection between the Hawkes process and the temporal point process AR equivalent. INMA processes are their moving average counterpart and were explored recently in [37]. Their definition can also be found in [76] and would be the discrete analogue of shot noise processes. Multivariate INAR have been proposed in [122] and for completeness, the multivariate INARMA can also be found in the literature (as well as MINMA), but it has been studied less than its autoregressive only counterpart [114].

Fitting Hawkes processes can be challenging due to the computational complexity of the likelihood function, see [33]. The least squares method is also as computationally expansive [35]. An interesting alternative is using neural networks to approximate the intensity function, first proposed by Du et al. in [61]. Finally, EM methods to perform MLE are also very expensive [123, 212], and all methods scale in the number of

events of the time series. One solution is the use of binned temporal point processes, which may be estimated via the same aforementioned methods but with a complexity cost reduced from  $n$  to the number of bins  $B$  (or to the length of the kernel). This alleviates the computational burden when fitting Hawkes processes to large counts of events [208].

Furthermore, data does not always respect the hypothesis of simple point processes that events never happen simultaneously. Data may have imprecisions or sign of bundling effects, be rounded or aggregated or the original phenomenon might be naturally discrete... Binned data is prevalent in fields such as finance and economics [79, 169], healthcare [111], social sciences [3, 158], sports [176], and natural phenomena [4]. Financial data suffers from binning where microstructural noise effects such as the bid-ask bounce effect [193] and that bundling effect [69] can be observed in studies of the limit order book. Using traditional continuous estimation methods in such cases can lead to errors, as the fundamental assumption of the simpleness of the process is violated and the data time stamps are only approximations of the true time stamps.

[115, 116] are the first papers to share some thoughts about how to estimate the parameters of an INAR( $p$ ) and INVAR( $p$ ) process. Using naively the maximum likelihood estimator (MLE) is now accepted as being particularly flawed and inefficient for INAR processes [115, 208]. The conditional least square (CLS) estimation reaches similar performances but for a significantly reduced complexity and implementation cost [115]. Sornette's research group noticed that the CLS-estimator is not powerful enough to untangle endogenous from exogenous dependencies, as it relies on the first-order moments [69, 208, 212], unable to discriminate the origins of slowly varying trends and persistent stochastic fluctuations. On the contrary, the EM algorithm [57, 139] succeeds in this task at the cost of a greater computational cost, see [174, 212]. This is most apparent in the case of non-stationary dynamics [206, 208, 212]. However, this EM algorithm has not yet been used in the case of INVAR processes, but shows promising results for multidimensional Hawkes processes [212].

For completeness, two recent papers have used the EM algorithm and compared it to the CLS-estimator, see [179, 180]. The latent variable space for their EM algorithm is the space of the time stamps where the weights are the probability that a time stamp is in a certain time area. Conversely, our approach employs the branching structure as the missing information in the EM algorithm. This strategy results in a less computationally intensive approach and allows us to choose to parameterize the memory kernel. In contrast, Shlomovich et al.'s method relies entirely on a parametric selection of the underlying kernel, which makes incorporating nonstationarities considerably more challenging than our approach. Additionally, while [180] compares the two EM techniques in the univariate setting, our methodology has yet to be presented for the multivariate case [179]. Given the considerably more intricate branching structure involved in the multivariate case, we anticipate that our EM algorithm will exhibit some hidden strengths against theirs.

In this paper, we extend the theoretical results from [117] to the multivariate case. We feed our intuition on VAR processes, for which we have proven an easier criterion of existence for (IN)VAR( $p$ ) processes when  $p < \infty$  in Theorem 1.2.1. By taking the limit on  $p$  in Theorem 1.2.1 on the infinite sum, we get an existence condition for (IN)VAR( $\infty$ ), which will be the condition we prove to be sufficient in Theorem 1.3.8. This extension could not be derived by the previous method where we construct a matrix. We extend the uni-dimensional result of convergence of INAR to Hawkes processes [117] to multiple dimensions in Theorem 1.3.13). We also propose a VAR representation of the INVAR process in Theorem 1.3.10, an MA representation in Theorem 1.3.11 and provide the autocovariance structure of the INVAR( $\infty$ ) process in Theorem 1.3.12.

Finally, we conclude the paper in Section 1.4 with numerical experiments where we compare the results of the EM algorithm to the CLS algorithm, on the roadmap proposed from [208]. We expect our proposed methodologies to carry over to multivariate time series allowing to disentangle endo and exo influences. Furthermore, the EM algorithm uses a parametric estimation which can be easily adapted to include further improvements: spline estimation, Bayesian optimisation... Whereas improvements in the CLS method would be more difficult.

## 1.2 Motivation: A Simpler Existence Condition

[109] proves that for  $p \in \mathbb{N}_{>0}$ , a stationary INAR( $p$ ) process with coefficients  $\{\alpha_k\}_{k \in \{1, \dots, p\}}$  exists when its characteristic equation (a polynomial with  $\lambda$  as the unknown:  $\lambda^p = \sum_{k=1}^p \alpha_k \lambda^{p-k}$ ) has no solution outside

the unit circle. Mimicking the theory of AR( $p$ ) processes, they construct a matrix with the coefficients  $\alpha_k$  on the first row, a diagonal of ones starting from the second line [109]. This representation projects an INAR( $p$ ) process onto a bigger dimensional space, where it behaves as an INAR(1) process, and the latter benefits from a simple induction formula. This proves that an INAR( $p$ ) process has to satisfy the same characteristic equation and condition for existence as an AR( $p$ ) process.

It is later extended in [121] where Latour uses Rouché's Theorem to prove a simpler equivalent condition to verify. The author proves that when  $\sum_{k=1}^p |\alpha_k| < 1$ , all the roots of the characteristic polynomial are inside the unit circle<sup>1</sup>. He also proves that INAR and AR processes have a very similar structure (for example from a second-order point of view). Much later, some work has been done on infinite integer-valued univariate autoregressive processes (INAR( $\infty$ )) [117]. The author proved in this paper that the condition  $p \in \mathbb{N}_{\geq 0}$ ,  $\sum_{k=1}^p \alpha_k < 1$  extends to the case  $p = \infty$ , and that INAR( $\infty$ ) also admit an AR( $\infty$ ) representation (thus an MA( $\infty$ ) representation as well).

Latour in [122] proposes a *multivariate* extension to the INAR process (that we call INVAR). Using a matrix representation to project an INAR( $p$ ) to an INAR(1) in a higher-dimensional space, he observes that if we assume stationarity for the INVAR process, a simple expression for the mean is found. The discussion reads:

Let us assume we have a (multivariate) IN(V)AR( $p$ ),  $p < \infty$ , process called  $\{X_n\}_{n \in \mathbb{Z}}$ , and we use the same notation as in the Definition 1.3.5. Then, we define for all  $n \in \mathbb{Z}$ :  $Y_n = (X_n, X_{n-1}, \dots, X_{n-p+1})$ ,  $\tilde{\varepsilon}_n = (\varepsilon_n, 0, \dots, 0)$ . We call their means  $\mu_Y$ ,  $\mu_{\tilde{\varepsilon}}$  (and the latter is nonzero because  $\varepsilon$  is Poisson with a strictly positive coefficient). Finally, we define the matrix  $A$  (notice that  $A$  is the transpose of the companion matrix of the reciprocal characteristic polynomial of the difference equations of an INVAR( $p$ ) process, see Appendix A.5):

$$A = \begin{pmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ \text{Id} & 0 & \cdots & 0 & 0 \\ 0 & \text{Id} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \text{Id} & 0 \end{pmatrix}. \quad (1.2.1)$$

The original stochastic difference equations are translated into the following:

$$Y_n = A \circ Y_{n-1} + \tilde{\varepsilon}_n. \quad (1.2.2)$$

For a constant mean process  $Y_n$  (we assumed stationarity of the solution process), using Proposition 1.3.4:

$$\mu_Y = A \cdot \mu_Y + \mu_{\tilde{\varepsilon}}. \quad (1.2.3)$$

Eq. (1.2.3) has a solution if and only if the spectral radius of  $A$  is strictly smaller than 1, Section 2.1 of [175], which is then:

$$\mu_Y = (\text{Id} - A)^{-1} \mu_{\tilde{\varepsilon}}. \quad (1.2.4)$$

Latour in [122] concludes that an INVAR( $p$ ) process exists if and only if the previously introduced matrix  $A$  respects that the eigenvalues of  $\det(\text{Id} - Az)$  are outside the unit circle, or equivalently, that the eigenvalues of  $\det(\text{Id} - \sum_{k=1}^p A_k z^k)$  are outside the unit circle (the equivalence is from [130] problem 2.1). Existence is hence related to the classical polynomial eigenvalue problem, see Definition A.5.1.

This is interesting because the condition "the eigenvalues of  $\det(\text{Id} - Az)$  are outside the unit circle" is equivalent to the condition " $A$  has eigenvalues smaller than 1", matching the condition for the existence of a solution to the problem of inversion (1.2.3).

Proceeding from (1.2.3), it is clear that all subsequent lines are trivially satisfied as  $\mu_{\tilde{\varepsilon}}$  is zero except the first. By using the assumption that the process  $X$  is stationary, (1.2.3) is equivalent to:

$$\mu_X = \sum_{k=1}^p A_k \mu_X + \mu_{\varepsilon}. \quad (1.2.5)$$

---

<sup>1</sup>Equivalently, the connected equation  $1 - \sum \alpha_i z^i$  has roots only outside the unit circle [121]. This is due to the correspondence between the characteristic equation and reciprocal characteristic equation, see our discussion related to (A.5.2).

By the same reasoning as before, this equation admits a solution if and only if  $\sum_{k=1}^p A_k$  has all its eigenvalues smaller than 1.

This heuristic hints that if we have "the eigenvalues of  $\det(\text{Id} - Az)$  are outside the unit circle", then we have " $\sum_{k=1}^p A_k$  has eigenvalues smaller than 1". We indeed prove:

$$\rho\left(\sum_{k=1}^p A_k\right) < 1 \iff \rho(A) < 1, \quad (1.2.6)$$

see notation  $\rho$  in Definition 1.3.1. Using the notation from the following section in this theorem:

**Theorem 1.2.1** (Alternative Condition for the Existence of the VAR( $p$ )). *Let  $p \in \mathbb{N}_{>0}$ . Assuming a sequence of matrices  $\forall k \in \mathbb{N}_{>0}, A_k \in \mathbb{M}_d, d \in \mathbb{N}_{>0}$ .  $\{X_n\}_{n \in \mathbb{Z}}$ , a stationary process, solution to the VAR( $p$ ) difference equations with coefficients  $A_k$ , exists almost surely when all eigenvalues of the sum of the term-wise absolute value of the reproduction matrices are less than 1:  $\rho(\sum_{k=1}^p |A_k|) < 1$ . For general VAR processes, this condition is only necessary.*

Naturally, the same statement is true for INVAR( $p$ ) processes.

This is a direct consequence of Lemma A.6.2 for nonnegative matrices. Our discussion outlines a condition on the sum of the matrices  $\sum_{k=1}^p A_k$  in order to have existence of the VAR (and by extension INVAR) process. This condition is useful because it is easily computable (instead of computing the companion matrix's eigenvalues) and makes the verification much simpler<sup>2</sup>. It was also expected as the natural generalisation of what we have in one dimension but has not yet appeared in the literature to the best of our knowledge.

## 1.3 INVAR Theory

### 1.3.1 Setting

We say that when we sum from  $n$  to  $m \in \mathbb{Z}$  such that  $m < n$ , the sum is equal to 0:  $\forall m < n \in \mathbb{Z}, \forall \{x_i\}_{i \in \mathbb{Z}} \subset \mathbb{R}^{\mathbb{Z}} : \sum_{i=n}^m x_i := 0$ . We use the convention that when we write a sum up to a random vector of vectors, it means a "component-wise" sum: for each coordinate  $k$  of the vectors:  $(\sum_{i=1}^{N_t} J_i)_k = \sum_{i=1}^{(N_t)_k} (J_i)_k$ .

We will speak about  $d$ -dimensional vector, where  $d \in \mathbb{N}_{>0}$ , fixed throughout this paper. We define  $\mathcal{B}$  to be the Borel-sets in  $\mathbb{R}^d$ . The convention we will use is that subscripts represent the sequence's index, whereas superscripts represent the index in the  $d$ -dimensional vector. For example,  $N_n^i$  represents the  $i$ -th coordinate of the  $n$ -th element of the sequence  $N$ .

The set  $\mathcal{B}_b$  denotes the Borel sets that are also bounded. The set of point measures on  $\mathbb{R}^d$  is denoted by  $M_p$ . As in [117], we define the space of nonnegative continuous functions on  $\mathbb{R}^d$  with compact support by  $C_K^+$  and we say that a family of point process  $\{m_n\}_{n \in \mathbb{N}_{\geq 0}}$  converges vaguely to a point process  $m^*$  when:  $\forall f \in C_K^+, \lim_{n \rightarrow \infty} \int f dm_n \rightarrow \int f dm^*$ . This convergence defines the vague topology on  $M_p$  and we call the Borel  $\sigma$ -algebra generated by this topology  $\mathcal{M}_p$ . We will refer to functions  $\Psi$  as being continuous with respect to the vague topology, in the sense that  $\lim_{n \rightarrow \infty} \Psi(m_n) = \Psi(m^*)$ .

We refer to the algebra of all real  $p \in \mathbb{N}_{>0}$  dimensional square matrices by  $\mathbb{M}_d$ . We call  $\text{Id} \in \mathbb{M}_d$  the identity matrix (diagonal matrix with only one on the diagonal) of the matching size with the rest of the equation. We also use the notation  $\mathbf{1}$  to denote a vector of ones,  $\mathbf{0}$  to denote a vector of zeros, and finally  $\cdot^T$  to denote the transposition operator.

**Definition 1.3.1** (Spectral Radius). Let  $A \in \mathbb{M}_d$ , the **spectral radius** of a matrix is defined as the maximum of all the modulus of the complex eigenvalues:

$$\rho(A) := \max(|\lambda_1|, |\lambda_2|, \dots, |\lambda_d|). \quad (1.3.1)$$

We work within a stochastic basis  $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \in \mathbb{R}_+}, \mathbb{P})$  and consider random vectors from  $\Omega$  to  $\mathbb{R}^d$ . The operator  $\mathbb{E}$  in this space is classically defined as  $\int_{\Omega} X d\mathbb{P} \in \mathbb{R}^d$  and similarly we define the covariance operator. We will be using  $L^q$  spaces whose norm is defined for  $X$  a random vector:  $\|X\|_{L^q}^q = \int_{\Omega} |X|^q d\mathbb{P} = \mathbb{E}[|X|^q] \in \mathbb{R}$ .

---

<sup>2</sup>Computing the eigenvalues of the companion matrix scales in the memory size  $p$  times the dimension of the process  $d$ . Computing the eigenvalues of the sum of matrix coefficient scales only in the dimension of the process  $d$ .

We can use any norm on  $\mathbb{R}^d$  (as they are all equivalent) and for convenience, we use the  $l^1$  norm, coined  $|\cdot|$ , in other words:  $|N| = \sum_{i=1}^d |N^i|$ . This will help avoid technical details. We call respectively for all  $q > 1$ ,  $L^q$  bounded and finite when a random vector's  $L^q$  norm is respectively bounded or finite.

Multivariate point processes are the direct extension of the previous definitions. We write a  $d$ -variate point process  $N$  as  $N_t = (N_t^1, \dots, N_t^d)$ .

### 1.3.2 The INVAR( $p$ ) Process

**Definition 1.3.2** (Univariate Thinning Operator). Suppose that we have an integer-valued random variable  $X$ ,  $\alpha \geq 0$ ,  $\xi_l^{(\alpha)} \sim \text{Pois}(\alpha)$  independent over  $l \in \mathbb{N}_{\geq 0}$  and independent of  $X$ .

We coin  $\circ$  the **thinning operator** as defined in [121, 122], that reads:

$$\alpha \circ X = \sum_{k=1}^X \xi_k^{(\alpha)}. \quad (1.3.2)$$

We say that two thinning operations are independent when their respective Poisson realisations are independent.

Since for all  $d \in \mathbb{N}_{>0}$ ,  $\mathbb{R}^d$  is a complete separable metric space, most results from  $\mathbb{R}$  extend naturally to the multivariate case. Similarly to the 1-D case scenario, we define the multivariate thinning operator by:

**Definition 1.3.3** (Multivariate Thinning Operator). Let  $X$  be a positive, integer-valued, random,  $d$ -vector. We coin  $A$  a matrix with entries  $(\alpha_{i,j})_{i,j \in \{1, \dots, d\}}$  and the **multivariate thinning operator** is defined as a matrix of thinning operators, as in Definition 1.3.2, as also defined in [121, 122]:

$$A \circ X = \begin{pmatrix} \sum_{j=1}^d \alpha_{1,j} \circ X^j \\ \vdots \\ \sum_{j=1}^d \alpha_{d,j} \circ X^j \end{pmatrix}. \quad (1.3.3)$$

**Proposition 1.3.4** (Properties of the Thinning Operator). *As long as  $A$  is a matrix and  $X$  is a positive random integer vector with matching size:*

- $\mathbb{E}[A \circ X] = A \mathbb{E}[X]$ ,
- $\mathbb{E}[(A \circ X)(A \circ X)^T] = \text{Diag}(B \mathbb{E}[X]) + A \mathbb{E}[XX^T]A^T$ ,

where  $B$  is the covariance matrix of the operator  $A \circ$ .

Another property that we will use (the derivation is straightforward):

- $\text{Cov}[A \circ X] = \text{Diag}(B \mathbb{E}[X]) + \text{Cov}(A \cdot X) = \text{Diag}(B \mathbb{E}[X]) + A \mathbb{E}[XX^T]A^T - A \mathbb{E}[X]\mathbb{E}[X^T]A^T$ .

*Proof.* The two first points are from [122], Lemma 2.1. In particular, it goes directly from the definition that the mean of the operator  $A \circ$  is  $A$ .

For the third bullet point, it suffices to use the second property and develop the expression of the variance in terms of moments of order 1 and 2.  $\square$

**Definition 1.3.5** (INVAR( $p$ )). Let  $p \in \mathbb{N}_{>0} \cup \{\infty\}$ , a  $d$ -dimensional vector of positive numbers  $A_0$ , a matrix sequence that we call reproduction matrices  $\{A_n\}_{n \in \mathbb{N}_{>0}}$  nonnegative, let  $\varepsilon_n \sim \text{Pois}(A_0)$ ,  $n \in \mathbb{Z}$  i.i.d. (each  $\varepsilon_n$  is  $d$ -dimensional, and all components are independent) and  $\xi_l^{(n,k),(i,j)} \sim \text{Pois}((A_k)_{i,j})$  independent over  $n \in \mathbb{Z}, k \in \mathbb{N}_{>0}, l \in \mathbb{N}_{\geq 0}, i, j \in \{1, \dots, d\}$  and independent of  $\{\varepsilon_n\}_{n \in \mathbb{N}_{\geq 0}}$ .

A  $d$ -dimensional time series  $\{X_n^{(\Delta)}\}_{n \in \mathbb{Z}}$  is called a **multivariate integer-valued autoregressive process (INVAR( $p$ ))**, if it satisfies:

$$\varepsilon_n = X_n - \sum_{k=1}^p A_k \circ X_{n-k},$$

or equivalently with more details

$$\begin{aligned} \begin{pmatrix} \varepsilon_n^1 \\ \vdots \\ \varepsilon_n^d \end{pmatrix} &= X_n - \begin{pmatrix} \sum_{k=1}^p \sum_{j=1}^d \alpha_{1,j} \circ X_{n-k}^j \\ \vdots \\ \sum_{k=1}^p \sum_{j=1}^d \alpha_{d,j} \circ X_{n-k}^j \end{pmatrix} \\ &= X_n - \begin{pmatrix} \sum_{k=1}^p \sum_{j=1}^d \sum_{l=1}^{X_{n-k}^j} \xi_l^{(n,k),(1,j)} \\ \vdots \\ \sum_{k=1}^p \sum_{j=1}^d \sum_{l=1}^{X_{n-k}^j} \xi_l^{(n,k),(d,j)} \end{pmatrix}. \end{aligned} \quad (1.3.4)$$

In the sequence of matrix,  $n \in \mathbb{N}_{\geq 0}$ ,  $i, j \in \{1, \dots, d\}$ ,  $(A_n)_{i,j}$  would represent the influence from dimension  $j$  onto dimension  $i$ .

**Remark 1.3.6.** In one dimension, the INAR process can be represented alternatively as a branching process, which endows it with the interpretation of the size of a population, where each individual is alive for one-time step but can have offspring in any of the next  $p$  time steps. The  $d$ -dimensional INVAR process can be interpreted as  $d$  populations that interact with each other. When the sequence of matrices  $A_k$  is always diagonal, every population is independent.

**Remark 1.3.7. Generalisation:** it is possible to consider the parameter  $A_0$  to be zero. This case is however degenerate (the process eventually fades away) unless some notion of "criticality" comes into play. As mentioned in the univariate case, there should be no problem to prove the existence of a stationary, critical INVAR( $\infty$ ) similarly to how it was done for Hawkes processes. Critical Hawkes processes exist and were proven to exist in [27] and some additional properties were found in [170]. This is however not the goal of this paper, hence we assume that at least a coordinate of  $A_0$  is nonzero. Also, it is hidden behind our definition that the innovation of our series (the vector  $\varepsilon_n$ ) has independent coordinates. For the purpose of this paper where we want to prove convergence to Hawkes processes, for which the exogenous rates are usually considered independent, we do not need to consider more generality. It would however not be too complicated to weaken this assumption.

### 1.3.3 The INVAR( $\infty$ ) Process, Existence and Properties

**Theorem 1.3.8** (Existence of the INVAR( $\infty$ )).  $\{X_n\}_{n \in \mathbb{Z}}$ , a stationary process, solution to the INVAR( $\infty$ ) difference equations, exists almost surely (see Definition 1.3.5, in (1.3.4)), when the maximal modulus of the eigenvalues of the sum of the reproduction matrices is less than 1:  $\rho(\sum_{k=1}^{\infty} A_k) < 1$ . Using  $\nu := A_0$ , we have:

$$\mathbb{E}[X_n] = \nu \odot (\text{Id} - K)^{-1} \mathbf{1}, \quad n \in \mathbb{Z}, \quad (1.3.5)$$

where  $\sum_{k=1}^{\infty} A_k =: K$ .

We reserve the name stationary INVAR( $\infty$ ) to the unique (almost-surely) solution to the difference equation, such that  $\rho(K) < 1$ .

**Remark 1.3.9.** We do not require  $\sum_{k=1}^{\infty} A_k$  to be absolutely summable (in other words  $\|\sum_{k=1}^{\infty} A_k\| < \infty$ ). The summands are all positive so the sum diverges only monotonically.

*Proof.* See in appendix Proof B.2. □

We have previously discussed how INVAR processes and VAR processes share similar properties due to their identical starting equations, and how the theories of the two have mutually influenced each other. As noted in [130], under mild conditions, a VAR(1) process  $Y_n$  has a MA representation (vector moving average)

$Y_n = (\text{Id} - A_1)^{-1} \cdot A_0 + \sum_{i=0}^{\infty} A_1^i \varepsilon_{n-i}$ , where  $\varepsilon$  is some white noise. The mean of such a process would then be  $(\text{Id} - A_1)^{-1} A_0$ .

In the case of integer-valued processes, we use a thinning operator instead of vector multiplication. In 1 dimension, the two operators behave similarly, both are symmetric (recall Definition A.4.3):  $x * y = y * x$  and  $x \odot y = y \odot x$ . However, in multi-dimension, only the thinning operator is symmetric:  $x * y$  has a different meaning from  $y * x$ . Eq. (1.3.5) could have been rewritten as  $\mathbb{E}[X_n] = ((\text{Id} - K)^{-1} \mathbf{1}) \odot \nu$ , which resembles the expectation of a VAR process (written in the previous paragraph).

This similarity is reassuring, and we also observe a similarity between INVAR and multivariate Hawkes processes. Stationary multivariate Hawkes processes require that the reproduction matrix has a spectral radius strictly less than 1, equal to the integral of the self-excitation function. The condition for the existence of INVAR( $\infty$ ) is that the infinite sum  $\sum_{k=1}^{\infty} A_k$  has a spectral radius smaller than 1. In that sense,  $\sum_{k=1}^{\infty} A_k \approx \int \Phi$ , for well-chosen matrices  $A_k$ : the discrete analogue of the integral of the kernel function.

We now present alternative representations which will assist in deriving the second-order moments.

**Theorem 1.3.10** (VAR( $\infty$ ) Representation of an INVAR( $\infty$ )). *Given a stationary INVAR( $\infty$ )  $\{X_n\}_{n \in \mathbb{Z}}$ , with notation from the definition, we define the multivariate process:*

$$u_n := X_n - \sum_{k=1}^{\infty} A_k X_{n-k} - \nu, \quad n \in \mathbb{Z}. \quad (1.3.6)$$

*Then, the sequence  $\{u_n\}_{n \in \mathbb{Z}}$  is stationary and is a white noise with covariance matrix  $\text{Diag}(\nu \odot (\text{Id} - K)^{-1} \mathbf{1})$ .*

*By rewriting the definition of  $\{u_n\}_{n \in \mathbb{Z}}$ , we get the VAR representation of an INVAR:*

$$(X_n - \nu \odot (\text{Id} - K)^{-1} \mathbf{1}) - \sum_{k=1}^{\infty} A_k (X_{n-k} - \nu \odot (\text{Id} - K)^{-1} \mathbf{1}) = u_n, \quad n \in \mathbb{Z}.$$

*If we coin  $\mu_X := \nu \odot (\text{Id} - K)^{-1} \mathbf{1}$  (see (1.3.5)), we get:*

$$(X_n - \mu_X) - \sum_{k=1}^{\infty} A_k (X_{n-k} - \mu_X) = u_n, \quad n \in \mathbb{Z}. \quad (1.3.7)$$

*Proof.* See in appendix Proof B.3. □

**Theorem 1.3.11** (MA( $\infty$ ) Representation of an INVAR( $\infty$ )). *Given a stationary INVAR( $\infty$ )  $\{X_n\}_{n \in \mathbb{Z}}$  with mean  $\mu_X$ , with notation from the definition, and  $\{u_n\}_{n \in \mathbb{Z}}$  its associated white noise sequence (see Theorem 1.3.10), then we get that the INVAR process satisfies:*

$$X_n - \mu_X = \sum_{k=0}^{\infty} B_k u_{n-k}, \quad n \in \mathbb{Z}. \quad (1.3.8)$$

*The sequence of coefficients  $\{B_k\}_{k \in \mathbb{N}_{\geq 0}}$  is defined  $B_0 = \mathbf{1}$ , and for  $k > 0$ :  $B_k = \sum_{i=1}^k A_i B_{k-i}$ . Remark that each matrix of the sequence is nonnegative, and the sequence is summable with  $|\sum_{k=1}^{\infty} B_k| = |(\text{Id} - K)^{-1} \mathbf{1}| < \infty$  for a vector norm  $|\cdot|$ .*

*Proof.* The proof simply relies on Wiener-Khinchin's and Herglotz's Theorem [31] and the possibility to represent an AR process as an MA one. We leave it to the reader, as it is technical and does not add to the general understanding of the theorems. □

**Theorem 1.3.12** (Covariance Function of an INVAR( $\infty$ )). *Thanks to the MA( $\infty$ ) representation of an INVAR( $\infty$ ), the autocovariance function is directly obtained:*

$$R(j) = \sum_{k=0}^{\infty} B_k \text{Diag}(\nu \odot (\text{Id} - K)^{-1} \mathbf{1}) B_{k+|j|}^T, \quad j \in \mathbb{Z}, k \in \mathbb{N}_{\geq 0}. \quad (1.3.9)$$

*In the following, the inequality refers to a component-wise inequality:*

$$\sum_{j=0}^{\infty} R(j) \leq (\text{Id} - K)^{-1} \mathbf{1} \cdot \text{Diag}(\nu \odot (\text{Id} - K)^{-1} \mathbf{1}) \cdot ((\text{Id} - K)^{-1} \mathbf{1})^T. \quad (1.3.10)$$

We assume the matrix norm  $\|\cdot\|$  is consistent with the vector norm (any norm, not necessarily  $l_1$ )  $|\cdot|$  (see Theorem A.4.2)<sup>3</sup>. By the assumption on  $K$  and using that all terms in the autocovariance matrices are positive, we have:

$$\left\| \sum_{j=0}^{\infty} R(j) \right\| \leq \|(\text{Id} - K)^{-1} \mathbf{1} \cdot \text{Diag}(\nu \odot (\text{Id} - K)^{-1} \mathbf{1}) \cdot ((\text{Id} - K)^{-1} \mathbf{1})^T\| \leq |(\text{Id} - K)^{-1} \mathbf{1}|^3 \|\text{Diag}(\nu)\| < \infty. \quad (1.3.11)$$

*Proof.* See in appendix Proof B.4.  $\square$

### 1.3.4 Convergence of INVAR Processes to Hawkes Processes

**Theorem 1.3.13** (Weak Convergence of INVAR( $\infty$ ) to Hawkes). Let  $\Delta > 0$  small, and  $\{X_n^{(\Delta)}\}_{n \in \mathbb{N}_{\geq 0}}$  a stationary INVAR( $\infty$ ) process with immigration vector parameter  $\Delta \cdot \nu$  (the exogenous component corresponding to the coefficients of the  $\varepsilon$  of the Definition 1.3.5) and reproduction matrix coefficient  $\Delta \cdot \Phi(k \cdot \Delta)$  (the endogenous component corresponding to the coefficients of the  $\xi$ ) with  $k \in \mathbb{N}_{\geq 0}$ . Stationarity imposes  $\rho(\sum_k \Delta \Phi(k \Delta)) < 1$ . We define the corresponding family of point processes:

$$N^{(\Delta)}(A) = \sum_{k \in \mathbb{N}_{\geq 0} : k \Delta \in A} X_k^{(\Delta)}, \text{ for } A \in \mathcal{B}_b. \quad (1.3.12)$$

By calling  $N$  a multivariate Hawkes process with exogenous rate  $\nu > 0$  and piecewise continuous positive kernel  $\phi$ . Then we have the weak convergence:

$$N^{(\Delta)} \xrightarrow{w} N, \text{ when } \Delta \rightarrow 0. \quad (1.3.13)$$

This means that a multivariate Hawkes process is nothing more than the superposition of infinite memory INVAR processes. It is possible to extend this theorem to the case  $\nu = 0$ , in the same way, it was done for Hawkes processes, see discussion above Theorem A.1.4.

*Proof.* See in appendix Proof B.5.  $\square$

**Remark 1.3.14. Non-homogeneity:** It is possible to assume a dependence of the exogenous rate  $\nu$  and the kernel  $\Phi$  on time to account for non-homogeneity. If  $\Phi$  varies with time, the time at which the event is triggered will impact the shape of the kernel, as described in Definition A.1.3. The proof for the convergence of an INVAR process with time-dependent parameters would require replacing terms involving  $\Phi$  and  $\nu$  in most inequalities with the supremum of the function over the interval. A sufficient condition for this supremum to be bounded is that the two functions are at least piecewise continuous, which covers most cases encountered in practice.

**Positivity:** It is generally assumed in the literature that the self-excitation matrix function is positive, and this assumption is maintained in this paper. While an extension to a more general setting with self-inhibition mechanisms is possible, it would require ensuring the kernel is absolutely integrable and making some other minor technical considerations.

Directly, by taking the conditional expectations, conditioned on  $\sigma(X_k^{(\Delta)} : k \leq n-1)$  upon an INVAR( $\infty$ ) process (for the same parameters as above in the theorem), we get an analogue expression of the conditional intensity to a Hawkes process:

$$\frac{\mathbb{E}[X_n^{(\Delta)} | \sigma(X_k^{(\Delta)} : k \leq n-1)]}{\Delta} = \nu(n) + \sum_{k=1}^{\infty} \Phi(k \cdot \Delta) X_{n-k}^{(\Delta)}, \forall n \in \mathbb{Z}, \quad (1.3.14)$$

<sup>3</sup>Actually, the choice of the norms is not important because norms in  $\mathbb{R}^d$  as well as in  $\mathbb{R}^d \times \mathbb{R}^d$  are all equivalent. Hence, once the inequality is proven for a consistent norm, up to a multiplicative constant, the inequality holds for any norms.

where the conditional intensity of a Hawkes process would be from (A.1.4):

$$\lambda(t) = \nu(t) + \int_0^t \Phi(t-s) dN_s, \quad \forall t \in \mathbb{R}.$$

This analogue to the conditional intensity is precious to simulate step-by-step INVAR processes by sampling Poisson processes.

We would have the same equivalence for the moments of the processes and other characterising functions (relate the (1.3.5) to the mean of a multivariate Hawkes process, which has the same expression).

Intuitively, the counting variable  $X^{(\Delta)}$  represents the number of events in one interval of length  $\Delta$ :  $\forall k \in \mathbb{N}_{\geq 0} : X_k^{(\Delta)} = N(\Delta l, \Delta(k+1))$ . When the time step  $\Delta$  gets smaller, the process allows for more interaction between the different events (as two events being in the same bin do not interact with each other) and when we increase the number of interactions between the bins (the parameter  $p$  of an INAR( $p$ ) process), we allow the influence of a longer period of the history onto the current value of the process.

It is also reassuring to see the correspondence between the discrete and continuous versions about the stability conditions: for INVAR( $\infty$ ), a sufficient and necessary condition is  $\rho(\sum_k \Delta \Phi(\Delta k)) < 1$ ; for Hawkes processes, we require  $\rho(\int \Phi) < 1$ .

The proposed proof follows the canonical approach from Brémaud and Massoulié [27]. In a nutshell, the tightness of the approximating family is first proved, which by Prokhorov's Theorem yields weak subsequential limits for all subsequences. To conclude, we need to show that all those weak subsequential limits have the same distribution, the one of a Hawkes process.

**Proposition 1.3.15** (Kernel Uniform Bound). *We define for a piecewise continuous kernel (normalised and causal) function from  $\mathbb{R}$  to  $\mathbb{R}^d$ ,  $h$ , the matrix  $K^{(\Delta)}$ :*

$$K^{(\Delta)} := \Delta \sum_{k=1}^{\infty} h(k \cdot \Delta). \quad (1.3.15)$$

We state that there exists a constant  $\delta > 0$  and a matrix  $\tilde{K}$  such that uniformly on  $\Delta$ :

$$\rho(K^{(\Delta)}) \leq \rho(\tilde{K}) < 1, \quad \forall \Delta \in ]0, \delta[. \quad (1.3.16)$$

## 1.4 Numerical Study of INVAR( $\infty$ ) Processes

### 1.4.1 Synthetic Time Series Experiments

Estimating branching process can be easily seen as an incomplete data estimation problem, for which the solution is the expectation-maximization (EM) algorithm in order to perform MLE. In such context, the branching structure is the missing information which we describe probabilistically. Using the EM algorithm for the estimation of INVAR processes is utterly valuable: the difficulty to estimate branching processes' parameters is increased due to the increase of hidden dependencies, one process triggering events in other dimensions, and the EM may render tractable such highly problematic estimation. We provide a description of the EM algorithm for the multivariate estimation of the parameters of an INVAR process in Appendix C.

In this study, we compare the CLS estimator [115, 116] with the EM algorithm [206, 212] for estimating the parameters of INVAR processes. Although the CLS estimator is not meant to stand alone, it is a valuable preliminary methodology for choosing a parametric model, especially in the continuous setting of Hawkes processes. However, this comparison is still of interest as the CLS estimator is one of the few available methods for modeling binned data.

We estimate the parameters of INVAR processes for various synthetic time series with chosen arbitrary parameters, and aggregate the results through Monte-Carlo simulation. This allows us to obtain an estimate of the true parameters and deduce certain properties about the estimators. The sampled time series include a burn-in period, and initial estimates for starting the EM algorithm are chosen through likelihood comparison using a grid search across a starting domain.

We present a comparison of estimating an INVAR's constant through time exogenous rate  $\nu$  and branching ratio  $\eta$ , while the kernel is fixed as geometric and known in the case of the EM algorithm. We are currently

investigating more sophisticated experiments to be published in the future, including allowing the parameters to evolve through time and testing the methodologies over real-life datasets and applications.

### 1.4.2 Observation

We present the results of three experiments conducted on synthetic time series. The first experiment involved a low branching ratio of 0.4, and the results are shown on the top panel of Fig. 1.1. The second experiment was conducted for a branching ratio close to criticality, i.e., 0.99, and the results are presented in the middle of Fig. 1.1. The third experiment was also performed for an almost critical branching ratio of 0.99 but with exogenous rates ten times larger than those in the second experiment. The results are shown at the bottom of Fig. 1.1.

Our analysis reveals that the Expectation-Maximization (EM) algorithm outperforms the Conditional Least Squares (CLS) estimator by a significant margin, especially when the edge effect is negligible, see Subsection 1.4.3 for a definition. The EM yields smaller variance<sup>4</sup> for the estimates and effectively disentangles the endo-exo dynamics, as evidenced by the highly accurate cross-excitation estimates. Additionally, the EM algorithm produces only plausible results, in contrast to the CLS estimator, which produces negative estimates. This can be attributed to the non-parametric nature of the CLS estimation. Moreover, the estimates obtained using the CLS algorithm appear to be slightly more biased than those obtained using the EM algorithm.

Nevertheless, it is important to note that the edge effect has a strong impact on the accuracy of the estimation and must be accounted for to ensure sound results. The CLS algorithm is not affected by this effect.

Overall as it can be expected, the multivariate estimation is more challenging than in the univariate and the continuous case. However, the EM algorithm successfully decompose the activity into its endogenous and exogenous contributions.

### 1.4.3 Edge Effect

We document how estimates of the branching ratio in the INAR setup are affected by finite fitting windows. The unobserved offspring after the end of the fitting window are actually already accounted for in the previous estimators [208] as it is visible in (C.1.9). However, the unobserved events prior to the fitting window are not accounted for and can create a bias in the estimate, which has been named the edge effect bias. It can be exacerbated under the following conditions:

1. High immigration intensity which yields an increased number of unobserved points prior to the fitting window, illustrated by the observed increase in variance between the estimates shown as the second and third panels of Fig. 1.1,
2. Longer memory kernel time scales, which increase the likelihood that offspring from unobserved points fall within the fitting window,
3. Higher branching ratios, which raise on average the offspring count of unobserved points before the fitting window, potentially materialised within the fitting window, highlighted by the increased variance between the estimates shown in the first and second panels of Fig. 1.1,

Several potential solutions to this issue have been proposed, but their efficacy is yet to be documented and experimented. Possible solutions include conducting the estimation for a time  $t_0 > 0$  and utilizing information on  $[0, t_0]$  as a source of events in the E-step, sampling unobserved points in the past from the estimates at the present times, or use some kind of parametric assumption to add dummy origin points before time 0.

---

<sup>4</sup>We expect the variance of the estimator to increase once we also estimate the kernel, but the variance should nonetheless be lower.

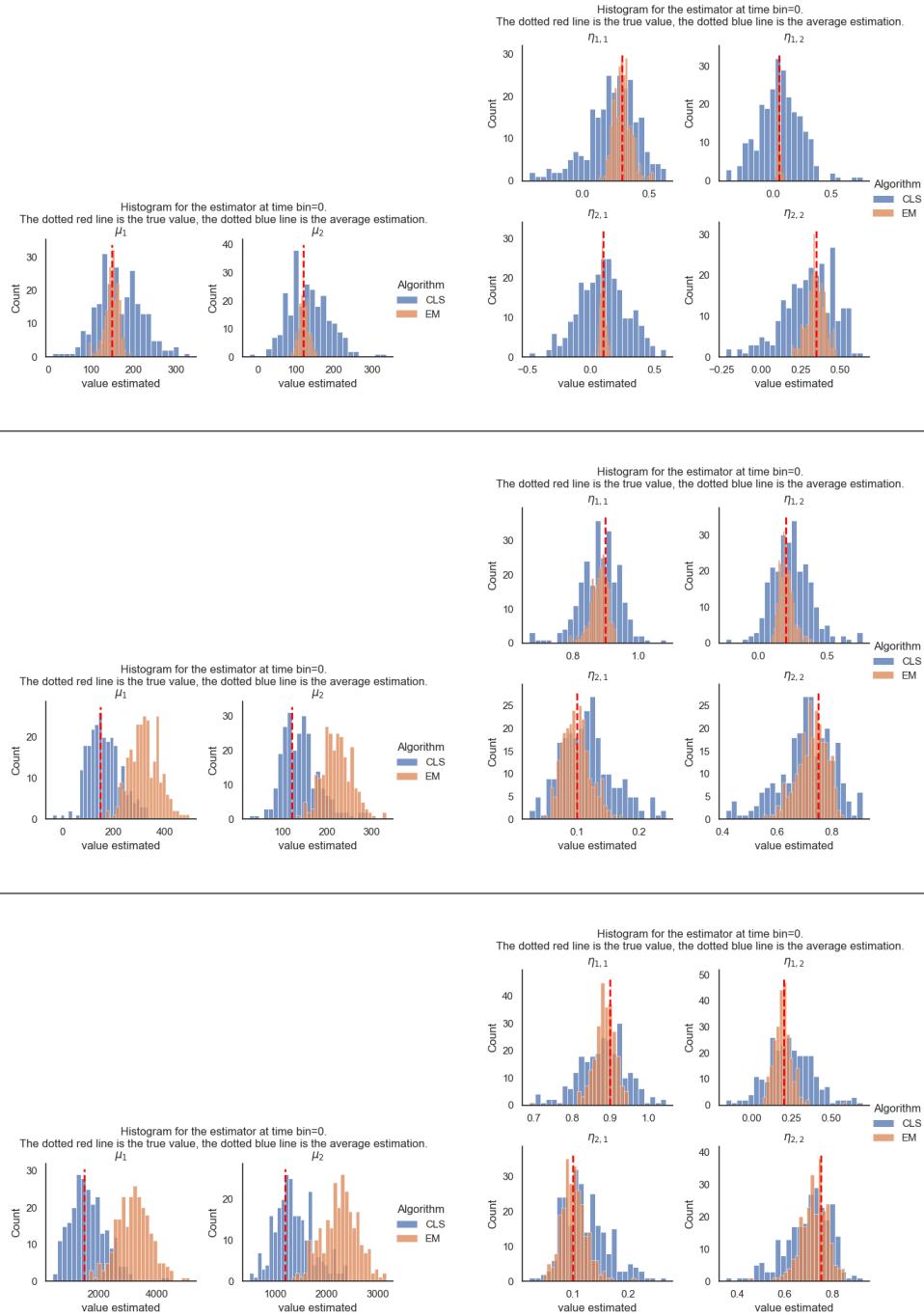


Figure 1.1: Histogram for the estimator's value when a process for constant time parameters are estimated. For the first line of images, the time series have true exogenous rate [150, 120] and branching ratio [[0.3, 0.05], [0.1, 0.35]], for a spectral radius of 0.4. For the second line of images, the true exogenous rate is [150, 120] and branching ratio [[0.9, 0.2], [0.1, 0.75]], for a spectral radius of 0.99. Finally, the third line of images shows the estimation for a true exogenous rate of [1500, 1200] and branching ratio [[0.9, 0.2], [0.1, 0.75]], for a spectral radius of 0.99. The fitting window is for all estimation of length 2000. These estimations clearly present the superiority of the EM algorithm over the CLS estimation as well as the edge effect described in the text.



## CHAPTER

# 2

# PROPERTIES OF FINANCIAL MARKETS

## Introduction

*The hunt for a perfect statistical model of financial markets is still going on.*

In [21], a quote from 2017 that remains relevant to this day.

The field of financial mathematics encompasses various sub-disciplines, with one such area focusing on the pricing and prediction of financial instruments. While the efficient market hypothesis posits that markets are semimartingale and free from arbitrage opportunities, there exist in practice higher-order dependencies that can be examined to generate profit.

This chapter provides an overview of the literature on two primary subjects in financial mathematics: macroscopic stylized facts and microstructural properties of markets. We investigate the relationship between these two topics and compare the approaches of quantitative researchers using jump processes versus rough models. Specifically, we focus our discussion on the concept of memory and volatility.

Through our examination, we aim to develop a deeper understanding of financial markets and the expected observable phenomena, particularly those related to roughness, volatility, and memory. Although our model in the subsequent chapter is limited to daily price observations and cannot replicate lower frequency data, a thorough comprehension of market features will be a strength for the choice of the model and parameters as well as for other calibration issues that we will face in the last chapter, in particular because Hawkes processes are used for modelling both micro and macroscopic behaviours. This makes the INVAR process introduced in the previous chapter particularly attractive.

We conclude this chapter by proposing a research idea that would solve the current paradox of criticality and unify the microscopic observations that markets are not critical with macroscopic rough models.

## 2.1 Properties of Financial Data

### 2.1.1 Known Properties of Financial Data

The traditional approach to market valuation is based on an event-driven approach, where the return is justified by events occurring around a specific time. This approach aims to rationalize market behaviour. Since various events have diverse impacts, it is expected that different markets exhibit distinct characteristics. However, empirical evidence shows that financial time series share certain statistical similarities, known as stylized facts [44]. Below, we enumerate some of them.

Regime	Time scale (sec)	Behaviour
Long-term	$\geq 10^4$	Behavioural mechanisms, long-term herding, hedging strategies, optimal execution of very large portfolio transactions
Short-term	$\sim 10^3 - 10^4$	Complex order types, margin/leveraged trading, herding in algorithmic strategies, optimal portfolio execution strategies, order splitting
High frequency	$\sim 10^0 - 10^2$	High-frequency / automated, limit order book based trading strategies, complex order types
Ultrahigh frequency	$\leq 10^{-1}$	Latency, microstructure, HFT

Table 2.1: Self-referential mechanisms at different time scales. More about it in [82, 205, 212].

**Properties of the Distribution of the Returns.**

- **Average gain:** Not all markets provide returns, and on average, gains for one agent mean losses for another. Returns depends on the period and the market.
- **Gain/loss asymmetry:** Markets exhibit more downward movements than upward, resulting in a negative skewness of the returns' distribution.
- **Fat-tailed return distributions:** Empirical evidence shows that market returns follow a distribution with fatter tails than the normal distribution. This distribution is well approximated by power laws with low exponents [44, 66, 108, 153]. In [44, 153], it is argued that returns should have finite variance (it excludes power laws with  $\alpha < 2$ ), which indicates a strong kurtosis. Moreover, the distribution changes from a power law to a Gaussian as the time scale increases [83, 108].
- **Non-stationarity and stability of the markets:** While the concept of criticality was initially believed to be a stylized fact, it was later refuted by the FS group (as discussed in Subsection 2.2.1). It is now established that markets are stable, yet they remain highly non-homogeneous due to the presence of strong exogenous non-stationarities and endogenous heterogeneities.
- **Momentum:** also known as trend following or "follow the winner", is characterized by positive returns being followed by further positive returns, similar to how investors "ride" a bubble. The momentum effect has spurred a proliferation of literature on time series momentum strategies (TSMOM), as documented in studies such as [145, 213].

**Memory of the Returns.**

- **Martingale property** (absence of linear autocorrelations, strong market efficiency hypothesis): The martingale property, which reflects the strong market efficiency hypothesis, states that linear autocorrelations of asset returns are mostly insignificant, as any existing linear correlation is quickly exploited by arbitrageurs, thereby eliminating any arbitrage opportunities. This suggests that market agents are rational, and the market prices respond accurately and instantaneously to any new information [58]. While this holds true for most time scales, microstructure effects may cause some autocorrelation in short time scales due to microstructure effect. For example, the returns of SLM (NYSE) stock daily returns exhibit negligible linear autocorrelation, as seen in [44].
- **Memory in the absolute returns:** The absolute returns' autocorrelation function decays slowly [58], as a power law with an exponent of around 1/4 as shown in [44].
- **Market Impact:** Metaorders' market impact behaves non-linearly and approximately like a square root [200]. However, recent studies such as [12, 112, 216] have challenged this view. Market impact and metaorders are defined in [12], and higher participation rates generally result in a more linear market impact profile.

Stylized facts are not required to be entirely accurate, and there may be some inconsistencies. For instance, the first lag in the autocorrelation sequence is usually statistically-significantly negative, indicating a negative autocorrelation and contradicting the martingale property. These inconsistencies can be attributed to market frictions such as transaction costs and latency.

### Property of the Volatility of the Returns.

- **Volatility clustering:** We observe positive autocorrelation of volatility measures, indicating that high-volatility events tend to cluster over time. This indicates long-term memory of volatility, as the tail of the autocorrelation function behaves as a power law with a coefficient smaller than 0.5. Mandelbrot [132] was perhaps the first to reference this phenomenon, and it has since been studied extensively [5, 26, 45].
- **Leverage Effect:** future volatility negatively correlates with the past asset's returns. It is a well-established relationship between stock returns and both implied and realised volatility: volatility increases when the stock price falls [20, 171]. This effect has been widely studied [32, 68], both on the macroscopic and microscopic scales: [64], and is crucial for practitioners in derivative pricing frameworks for reproducing implied volatility surfaces, which gave rise to a range of models, such as stochastic volatility models where the correlation coefficient between the two Brownian motions is negative. A negative stock return innovation increases the financial leverage of a company since the value of equity decreases for a fixed level of debt, hence increases the future volatility. This effect breaks up-down symmetry when looking at the transformation of returns to the opposite of returns ( $r \rightarrow -r$ ) [20, 171].
- **Trading volume is correlated with volatility:** Trading volume is positively correlated with volatility, indicating a long memory pattern. This phenomenon has been observed [129] and later mentioned in literature [44, 73].
- **Asymmetry in time scales of the volatility:** Coarse-grained volatility measures predict fine-scale volatility better than the other way around. This phenomenon, also called acausality, may have roots in the roughness of the process presented in the next point.
- **Roughness:** The roughness of financial markets is still a topic of debate. Some researchers suggest that the log volatility behaves as a fractional Brownian motion [17, 49, 80, 81, 128], which we define in Definition 2.1.1. The fractality of financial markets has been extensively reviewed in [108], where the authors thoroughly expose all the different evidence and occurrences of fractality in markets [133]. We discuss this property further in Subsection 2.1.5.
- **Zumbach Effect:** The correlation between past squared returns ( $r$ ) and future volatility ( $\sigma$ ) is greater than the correlation between future squared returns and past volatility, loosely written  $\text{Cov}(r_{t+\tau}^2, \sigma_t^2) < \text{Cov}(r_t^2, \sigma_{t+\tau}^2)$ . We also distinguish the weak and strong Zumbach effect, where we have just described the weak version of it and the strong one is the conditional law of future volatility depends not only on past volatility but also on past returns [21, 51]. Another way to interpret this effect is to say that price trends drive volatility [41]. Notably, the Zumbach effect is invariant under the opposite transformation of returns ( $r \rightarrow -r$ ), unlike the leverage effect.
- **Time reversal asymmetry:** A stationary time series is said to be time reversible if its finite-dimensional distributions are all invariant to the reversal of time indices ( $t \rightarrow -t$ ). A general definition is given in [36, 220, 221]. Otherwise, we say it is time irreversible. The stylized fact is called the time-reversal (a)symmetry (TRS/TRA) property. TRS is a central concept in science, but financial time series are TRA, as emphasized by Zumbach in [220] and [222]. Two main effects are accounting for TRA in financial time series: the leverage effect and the Zumbach effect. Interestingly, some models exhibit a weak Zumbach effect, such as rough volatility models [65], while all continuous-time stochastic volatility models are TRS [51].
- **Excess Volatility Puzzle:** defined by Shiller in [178], refers to the volatility of the equity market that cannot be justified by variation in subsequent dividends. See also [48, 205].

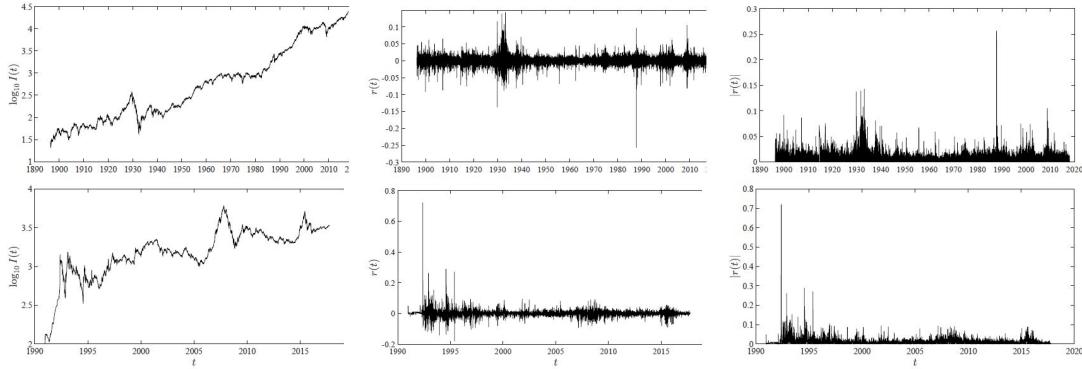


Figure 2.1: Figure from [108]. Each line represents a stock market index, with upper panels presenting the Dow Jones index and the lower panel the Shanghai Stock Exchange index. The left plots presents the time series of daily closing prices, in the middle we observe the daily returns and finally on the right the daily volatilities.

We present in Fig. 2.1 the time series of daily closing prices, returns and volatilities of two stock market indices. Similarly for the same underlyings, we show the distribution of the daily returns, the autocorrelation of returns as well as the correlation of the volatility in Fig. 2.2. These plots provide empirical evidence for the aforementioned stylized facts.

## 2.1.2 Modelling is a Sequence of Dilemmas

The choice of a model is highly involved because a stylised fact may be a partial property of a more fundamental rationale. Common oppositions about time series are deterministic trends against stochastic fluctuations, inhomogeneity against clustering, heterogeneity against coupling/contagion, independent/spontaneous activity against dependent/triggered activity [212]. Recently, a debate on the roughness of markets has emerged, and in [212], the discussion on stability and criticality concludes that markets are stable (refer to Definition A.1.3 for a definition of the stability).

Criticality is deeply connected with memory and proving that markets are stable suggests that markets possess a shorter memory than previously believed. The apparent long memory that was identified was attributed to unidentified exogenous effects. In the reference book [89], the author mentions how persistent "semi-systematic"<sup>1</sup> errors can lead to spurious memory effects (serial correlation), this despite measurements being independent by design. In our case, this translates into not considering certain sources of exogenous impacts.

Models with short memory but long observable impact, such as the threshold autoregressive process from [199], can produce long autocorrelations and spurious estimates of fractional integration. Similarly, [85] demonstrates that gradual changes in the autoregressive structure can have a similar effect. We discuss this more in detail in the following subsection.

These spurious relationships are pitfalls that researchers and practitioners must be aware of to perform sound analysis. Studies indicate that these issues are widespread in financial data, emphasizing the need for more sophisticated tools to capture various exogenous dynamics together with time-varying endogenous components.

In the following subsection, we focus on three critical aspects of financial modeling: memory, power laws, and stationarity. Our aim is to provide a comprehensive overview of these issues and their relevance to financial modeling.

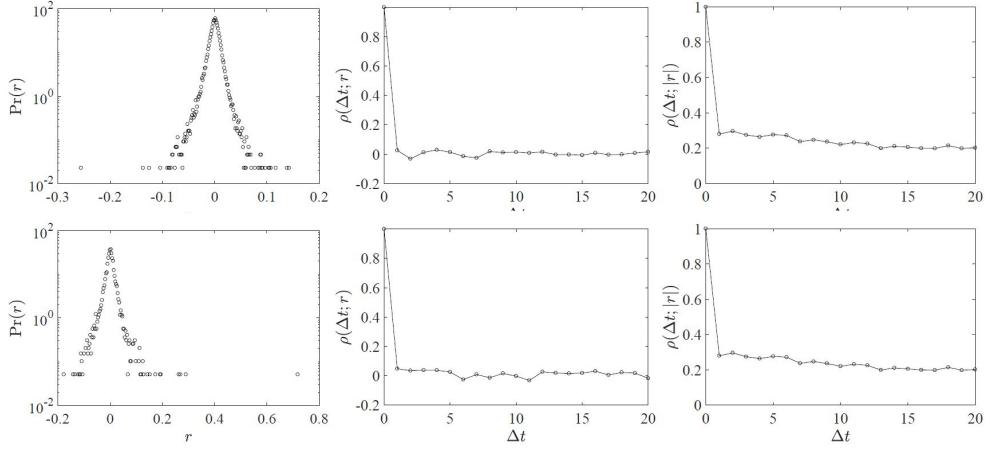


Figure 2.2: This figure, taken from [108], shows the data for two stock market indices: the Dow Jones index in the upper panels and the Shanghai Stock Exchange index in the lower panel. The left plots show the fat-tailed distribution of returns, which is not consistent with the Gaussian hypothesis that would result in an inverted parabola shape. In the middle plots, we observe the lack of linear autocorrelation of returns. Finally, the right plots depict volatility's memory.

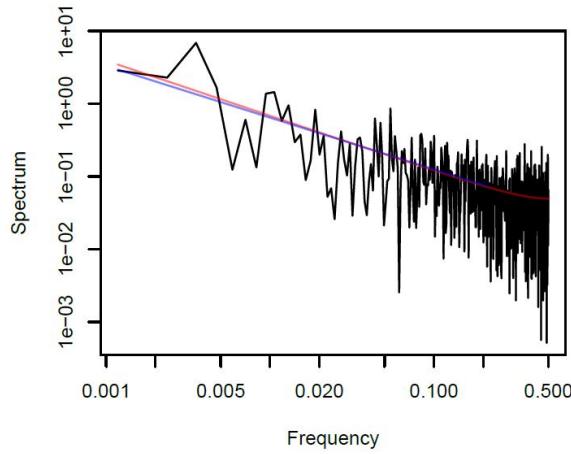


Figure 2.3: This figure from [84] presents the periodogram of the time series of Nile River minima in log-log coordinates. We observe the characteristic slope of power law increases.

### 2.1.3 Correlation Inside the Data and (Spurious) Long Memory

It is surprising that to this day, certain financial and risk-related models still assume that data is independently and identically distributed over time, without any correlation or clustering of values [84]. However, this assumption is invalidated by the long-range memory phenomenon that frequently occurs in economic data [197].

We call the autocorrelation function the function  $\rho(\tau)$  that describes the linear interdependence of a stochastic process at different points in time  $X_t$  and  $X_{t+\tau}$  where  $\tau$  refers to some lag  $\tau \in \mathbb{Z}$ .

The autocorrelation function or sequence (ACF / ACS) is loosely referred to as memory: processes with a "slowly" vanishing ACS are called long memory, in contrast to the fast decaying ones called short memory. We call slow, ACS decaying as a power law ( $\rho(\tau) \sim \tau^{-\alpha}$ ,  $0 < \alpha < 1$ ,  $\tau \gg 1$ ), whereas fast would be for example an exponential decay ( $\rho(\tau) \sim \exp(-\tau)$ ,  $\tau \gg 1$ ).

On one hand, when the decay is fast, the sum of the decays is finite  $\sum_{\tau_k=-\infty}^{\infty} \rho(\tau_k) < \infty$ , i.e. finite memory or short range dependence (SRD); on the other hand, we observe for slow decay a diverging sum  $\sum_{\tau_k=-\infty}^{\infty} \rho(\tau_k) = \infty$ , i.e. infinite memory or long range dependence (LRD)<sup>2</sup>. For example, ARMA( $p, q$ ),  $p, q \in \mathbb{N}_{\geq 0}$  are SRD processes.

The slow decay is often referred to as the Hurst phenomenon (going back to [98], and find a modern presentation in [84, 152]). A well-known example is the Nile River's annual minimum water levels, for which data has been compiled for over a thousand years, which exhibits long memory.

Under summability conditions, the autocorrelation sequence and spectral density are Fourier transforms of each other, known as Wiener-Khinchin's and Herglotz's theorem [31]. The LRD property is observed throughout the spectral density when it diverges at zero and decays slowly. Specifically, the spectral density exhibits the following asymptotic equivalence:  $S(f) \sim |f|^{-\beta}$ ,  $f \ll 1$ , with exponents related as  $1 = \alpha + \beta$ , where  $\alpha$  is the exponent of the asymptotic equivalence of the ACS (see the proof of the equivalence in textbooks [60, 168, 215]).

Even though the Hurst coefficient has never been written or used by Professor Hurst himself, it is now a common measure of memory length of the process and relates to the previously introduced coefficient  $\alpha$  and  $\beta$  by  $H = 1 - \alpha/2 = \frac{1}{2}(1 - \beta)$ , hence  $H \in ]0, 1[$ . We call rough a process with Hurst coefficient different from  $1/2$ <sup>3</sup>. Because of the previous equalities, we know that a rough process is LRD for  $H > 1/2$  and SRD for  $H < 1/2$ .

However, the whole scientific community does not root for the idea of infinite memory processes. As recalled in [152], some argued that infinite memory is not physically realistic.

A more problematic issue is that processes without memory, such as Markovian processes, can exhibit Hurst characteristics: a shifting mean model based on Gaussian independent synthetic series, for which the Hurst phenomenon is replicated even though the process is completely Markovian (hence without memory) and stationary [118]; in the same fashion, Hurst proposed a card game with similar features [99], and both experiments are synthesised in [25]. In [18], the authors found that a deterministic trend may also induce the Hurst phenomenon and in [85], we observe that a slow change in the autoregressive structure can also have this effect. This misleading relationship is now widely accepted and observed over a wide range of phenomenon [69, 155, 208, 212] and originated from non-stationarities due to a deterministic trend or from unit root processes.

### 2.1.4 Market Impact, Memory and Roughness

Market impact refers to the idea that an order has the potential to drive prices with force equal to the amount spent. Academics have questioned whether this impact is transient, with prices eventually reverting to their fundamental value, or whether it has a lasting effect, a discussion that evokes the unit-rooters versus detrenders debate. The latter effect is commonly known as permanent market impact. Research has been conducted to explore the relationship between endogeneity, market impact, and roughness.

<sup>1</sup>Coined by Newcomb, and many famous names of statistics acknowledged this problem, such that Pearson, Jeffreys...

<sup>2</sup>In the case of a continuous autocorrelation function, the sum is replaced by an integral, and the definitions remain largely the same.

<sup>3</sup>From the previous equalities, the autocorrelation of a rough process with Hurst coefficient  $H$  is asymptotically following a power law with exponent  $2H - 2$  and the spectral density follows asymptotically a power law with exponent  $2H - 1$ . See more details in the papers [84, 152].

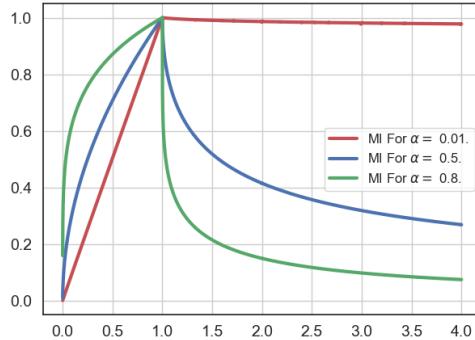


Figure 2.4: In [112], the authors show that the transient market impact is proportional to  $\int_0^t \mathbf{1}_{[0,s]}(t-u)u^{-\alpha} du$ , which we plot for various parameter values of  $\alpha$ . In particular, the blue line represents the square root law, as described in the text.

Modelling market impact is a complex issue, as it is not directly observable. It is defined as the difference between the price after a metaorder has been executed and the price in the absence of such an order. This difference, however, is not independent of the decision-making process that gave rise to the order. In other words, the price trajectory depends on the original transaction, and we can only observe what occurs after the order is executed, whereas we desire knowledge of the outcome had the root transaction not taken place. For example, a trend-following agent that anticipates a price increase due to a metaorder would affect the market by adding a non-independent noise term to the measurements, thereby rendering the price movement dependent on the decision. The market impact curve is typically observed to have two components (we describe the buy case, but it symmetrically works for sell orders), first, an increasing phase when the metaorders are executed and a decreasing part when the market recovers and the transient price increases dissipates.

Over the last decade, Hawkes processes have been effectively used to analyse market impact. In [112], the authors modelled the market using a price process that was the difference between the two components of a bivariate Hawkes process. Market impact was measured as the difference in price arising from metaorders. Assuming no-arbitrage and that the market impact exists with a transient and permanent part as well as assuming that the order flow is a linear Hawkes process, they find that the market impact  $\psi$  is a power law of exponent  $1 - \varepsilon$ :  $\psi(t) \propto t^{1-\varepsilon}$ , where  $\varepsilon \in ]0, 1[$ , where  $\phi$  is the Hawkes' kernel which follows a power law of exponent  $-(1 + \varepsilon)$ :  $\phi(x) \sim x^{-(1+\varepsilon)}$ ,  $x \rightarrow \infty$ . They also discovered that the permanent market impact was a linear function of quantity. The shape of the market impact, as described in the paper, is illustrated in Fig. 2.4.

The present research findings are supported by prior empirical studies. The shape of the kernel has been confirmed to approximate a power law, as documented in a number of publications, including [9, 90, 123, 212, 208]. The power law kernel is typically approximated using the Pareto tail kernel from [24], sometimes miscredited to [90]. However, recent observations suggest that a truncated power law might provide an even better fit due to practical computational and numerical limitations, as noted in [208, 212].

Secondly, the parameters of the kernel, particularly the exponent, align with theoretical considerations. Researchers have reported a range of values for the exponent, including approximately 0.45 in [90], close to 0.6 in [206, 208, 212], and 0.5 in [12]. The results in [112] indicates that market impact might follow the square-root law proposed in [200] and empirically confirmed in [12]. The exponent coefficient  $\varepsilon$  may differ across markets, especially concerning market liquidity, as well as the aggressiveness of trades during market crises. However, the authors of [212] observe that the characteristic time scale used for the kernels and the power law exponent are mostly independent of the degrees of freedom given to the exogenous rate, which significantly impacts the branching ratio. Consequently, the kernel parameters are estimated robustly, and the values obtained reflect an accurate representation. In particular, [90] establishes a relationship between the time evolution of the characteristic time scales used for the kernel of the Hawkes process and Moore's law.

The empirical results suggest that the actual market impact is more significant than the widely accepted square-root law and may have an exponent  $\varepsilon > 0.5$ . The higher the exponent, the faster the market recovers from the impact, with the extreme case  $\varepsilon = 1$  implying no transient price movement, and at the end of the metaorder, the market instantaneously drops to the terminal value. This situation seems pathological, as it raises the question of how the market is aware of the completion of the metaorder. In a perfectly liquid market with high participation rates (defined as the ratio between the volume traded by the metaorder and the volume traded by the entire market during the execution interval, as described in [216]), a linear market impact is expected instead of a square root, indicating no memory of the orders [12]. This has the interpretation that when the metaorder has a short duration and high impact on the market, the latter cannot withstand the shock, resulting in a strong linear impact, with relaxation only after the metaorder is complete.

Some other papers suggest that the law is exponential [216]. They also propose the concept of an impact surface to characterize more broadly the market impact of orders and metaorders.

The present discussion on market impact is linked to the concepts of roughness and memory. The ongoing investigation aims to establish connections between micro and macro dynamics is consistent, assuming that markets are critical.

Deriving connections between micro and macroscopic models (like the Hawkes process and Heston) is an open field of research and a very exciting topic, being the dream of theorists to start from an agent level perspective and upon aggregation recover a certain class of stochastic models. The first paper in this direction proved that rescaled Hawkes processes tend to converge to a diffusion process driven by standard Brownian motion, under the assumption of thin-tailed memory of the Hawkes process and non-critical markets [13]. This result has been extended to nearly unstable Hawkes processes [106], Hawkes processes with thick tails [107], and finally generalized to multidimensional Hawkes processes [167].

In [112], it was discovered that the appropriately rescaled Hawkes process model (assuming a nearly unstable Hawkes process sequence, as defined in [106]) yields macroscopic behavior of the price that follows a rough Heston model with Hurst parameter  $H = \varepsilon - 1/2$ , where  $\varepsilon$  denotes the exponent of the power law kernel. The authors require the exponent to correspond to a thick-tailed distribution, which means that  $0.5 < \varepsilon < 1$ . This establishes a one-to-one correspondence between the roughness coefficient  $H$  and the long memory parameter of the order flow  $\varepsilon$ . Therefore, criticality can be linked to the presence of roughness, and the heaviness of the tail can be related to the degree of roughness. Furthermore, research in the direction of generalized Hawkes processes, specifically quadratic Hawkes processes, has been conducted [51].

When we model a phenomenon related to human participation, and we coin the participation  $A(t)$ ,  $\forall t \in \mathbb{R}$ , an autoregressive model would be given with a kernel  $\phi$ , and exogenous influence  $\mu$ :

$$A(t) = \mu(t) + \int_{-\infty}^t A(\tau)\phi(t-\tau) d\tau. \quad (2.1.1)$$

This formulation can be rewritten without an autoregressive component, as demonstrated in [182] and discussed in [183, 185]:

$$A(t) = \int_{-\infty}^t \mu(\tau)K(t-\tau) d\tau. \quad (2.1.2)$$

Assuming that the kernel  $\phi$  follows a power law, the relationship between  $\phi$  and  $K$  is given by:  $\phi(t) \propto 1/(t - t_c)^{1+\varepsilon} \implies K(t) \propto 1/(t - t_c)^{1-\varepsilon}$ . One interesting observation is the consistency between these heuristics and the relationship between the Hurst coefficient and the exponent of the kernel:  $H = \varepsilon - 1/2$ . A rough model is essentially based on the fractional kernel  $K$ , which can be described as a power law with a coefficient of  $H - 1/2$  (as presented later in (2.1.4)). Combining  $H = \varepsilon - 1/2$  with the idea that the fractional kernel  $K$  is a power law with coefficient  $H - 1/2$ , we recover a power law kernel with exponent  $\varepsilon - 1$ .

## 2.1.5 Volatility, Memory and Roughness

We define the fractional Brownian motion and then we explain how it is related to volatility processes.

**Definition 2.1.1** (Fractional Brownian Motion). A fractional Brownian motion (fBm)  $\{W_t^H\}_{t \in \mathbb{R}_+}$  is defined as the continuous stochastic process, centred Gaussian process with covariance function (from [134, 152, 201]):

$$R_h(t, \tau) := \mathbb{E}[W_t^H W_\tau^H] = \frac{1}{2}(t^{2H} + \tau^{2H} - |t - \tau|^{2H}), \quad t, \tau \geq 0. \quad (2.1.3)$$

Further information on fractional Brownian motion can be obtained through the textbook [201] or the article [14]. Specifically, it is evident that fractional Brownian motion cannot be categorized as a semimartingale<sup>4</sup> or a Markov process, which holds significant implications for the practical and theoretical utilization of fractional Brownian motion models. Notably, a fractional Brownian motion with a Hurst coefficient of  $1/2$  exhibits independent increments, which is not the case when  $H \neq 1/2$ .

The Mandelbrot-van Ness representation [134] of an fBm with Hurst coefficient  $H \in (0, 1)$ , with  $B$  a Brownian motion, is:

$$W_t^H = \int_{-\infty}^0 \left( (t-s)^{H-\frac{1}{2}} - (-s)^{\frac{1}{2}} \right) dB_s + \int_0^t (t-s)^{H-\frac{1}{2}} dB_s. \quad (2.1.4)$$

Directly we get that:

$$\text{Var}(W_H(T) - W_H(0)) = V_H T^{2H} \text{ where } V_H \text{ is a constant.} \quad (2.1.5)$$

**Remark 2.1.2.** Note that the trajectories of the sample paths are almost surely  $H - \varepsilon$  Hölder, which is rough for small  $H$ . This is given by the kernel  $(t-s)^{H-1/2}$ , making the process  $\int_0^t (t-s)^{H-\frac{1}{2}} dB_s$   $H - \varepsilon$  Hölder.

**Remark 2.1.3.** Fractional Brownian motion is a self-similar (or self-affine) process, meaning its distribution scales according to:

$$Y_t \stackrel{(d)}{=} c^{-H} Y_{ct},$$

where  $H$  is the Hurst exponent previously introduced. These processes have a time-dependent variance but with (strongly) stationary increments.

As said in [45], self-similarity does not imply long-range dependence (some  $\alpha$ -stable Lévy processes are counter example), and long range dependence does not imply self-similarity see [39]. Finally, long range memory, self-similarity and arbitrage-free models are also totally independent properties, as demonstrated in [39]. The confusion between these notions can be attributed to the fBm which is a self-similar model with long range memory that is not a semimartingale.

One of the most common knowledge in financial mathematics is that volatility is serially correlated with long-term dependence [6, 7, 185]. The long memory of the volatility is philosophically associated with a large hierarchy of time scales of investors from HFT to large funds who trade at the time scale of years, referred to as the order splitting phenomenon: most orders are part of metaorders split into batches and executed over larger periods. A fascinating and related fact is the roughness of volatility. It is now almost accepted that volatility is rough, emphasized and pushed by Gatheral's and Rosenbaum's group: "Volatility is rough" [81]; further evidence in [17, 49, 64, 112, 128]. One rough volatility model is the infamous rough Heston model [42], for which an algorithm for simulation lies in [105].

Rough volatility refers to models where the log volatility process behaves as a fractional Brownian motion (fBm) with a Hurst coefficient smaller than 0.5, and empirically found to be around 0.15, depending on what market we are interested in.

Fractional Brownian motion has two remarkable properties which makes it suitable to be included in market' models:

1. Long-range dependence with slow decay  $\sim T^{2H-2}$ ,
2. Ability to produce trajectories with varying levels of Hölder regularity, i.e., roughness.

---

<sup>4</sup>It is worth noting that fractional Brownian motion cannot be classified as a semimartingale due to its infinite p-variation for every interval of positive length for  $H < 1/2$  (where  $p > 2$ ), or p-variation equal to zero for  $H > 1/2$ , leading to a constant process, which is incompatible with the continuous semimartingale definition.

The significance of rough models lies in their ability to link properties of the macroscopic and microstructural scales: Point 1 is a macroscopic property whereas Point 2 presents a microstructure property.

Interestingly, the initial financial models based on fractional Brownian motion (for example [43]) employed a Hurst coefficient of  $1/2 < H$ , implying a long-range dependence. However, recent recommendations suggest using a Hurst coefficient less than  $1/2$ .

The reason to model the log volatility as a fractional Brownian motion is twofold. First, the log volatility is known to be distributed as a Gaussian, as supported by numerous studies (e.g., [6, 7, 81]). Secondly, there seems to be a monofractal scaling relationship between the log volatility and the Hurst exponent  $H$ , i.e. the log volatility is self-affine with Hurst exponent  $H$ .

In [81], the authors identified a linear relationship between the empirical moment of order  $q > 0$  of the log volatility increment and the increment size:  $\mathbb{E}[|\log(\sigma_\Delta) - \log(\sigma_0)|^q] \approx c\Delta^{Hq}$  where  $p \approx 0.1$  by linear regression, which hints that the logarithm of volatility behaves like a fBm. They mention in the paper that a natural and simple model for the volatility would be  $\sigma_t = \sigma \exp(\nu W_t^H)$ ,  $\sigma > 0$ ,  $\nu > 0$ . It is however not stationary and instead, one would consider the exponential of a stationary fractional Ornstein-Uhlenbeck process defined by the equation  $dX_t = \nu dW_t^H - \alpha(X_t - m) dt$ ,  $m \in \mathbb{R}$ ,  $\nu > 0$ ,  $\alpha > 0$  where the differential operator is a simple Riemann-Stieltjes integral and the Rough Fractional Stochastic Volatility (RFSV) would be the process  $\sigma_t = \exp(X_t)$  (see [40] for the rationale behind the existence of a solution).

The interplay between memory and roughness lies in how we define memory. Some definitions (the one we used with the autocorrelation function) yield that for  $H < 1/2$ , the fBm is SRD. However, when we consider the power spectrum density as an indicator of memory, we find that for any Hurst coefficient  $0 < H < 1$ ,  $H \neq 1/2$ , the fBm is LRD [75, 124] (and (SRD) only for  $H = 1/2$ ). Also in [81], Gatheral, Jaisson and Rosenbaum show that classical statistical procedure aiming to detect long memory relationship (applied to their rough model, the RFSV) identifies long memory, despite the model being SRD. The connections between roughness and memory are intricate and it is possible that accepting that volatility is rough would also mean accepting that we observe (spurious) LRD, hereby contradicting decades of time-series analyses at showing that volatility is long dependent as well as the intuition people have of volatility. Given these challenges, [81] recommends to focus on the quantifiable concept of roughness rather than memory, as the former is more robust and less susceptible to varied definitions and estimation difficulties. In particular, multifractality has been identified as a key property of financial markets, and further work is needed to elucidate its mechanisms and implications [108]. However, it is worth noting that roughness itself is computed from estimated volatility, which can be subject to microstructure noise error, complicating the relationship between memory, roughness, and market impact [46].

## 2.1.6 Being Stationary and the Lack Thereof

The notion of stationarity constitutes a fundamental assumption in numerous statistical models. Stationarity can be defined in various ways, one is referring to the  $p$ -th order stationary process, where all moments up to order  $p$  remain invariant over time. In layman terms, a stationary process allows for the same model to be used throughout the time series, while a non-stationary process requires the use of a non-stationary model, harder to calibrate.

Trends in the mean represent the most common sources of non-stationarity, which can occur due to either the presence of a unit root or a deterministic trend. To illustrate these violations, we offer two prototypical examples:

We give one stereotypical example for each of the two types of violation discussed:

$$u_t = a + b \cdot t + \varepsilon_t \quad (2.1.6)$$

$$v_t = a + v_{t-1} + \varepsilon_t \quad (2.1.7)$$

$u_t$  in (2.1.6) is a typical trend-stationary model whereas  $v_t$  in (2.1.7) is the classical AR(1) model, with coefficient equal to 1, or random walk, which is not stationary but difference-stationary.

Unit root models are characterized by exogenous shocks that have permanent effects on the system, and they are commonly referred to as difference-stationary models. On the other hand, mean-reverting models, also known as trend-stationary models, exhibit stationarity once the deterministic trend is accounted for.

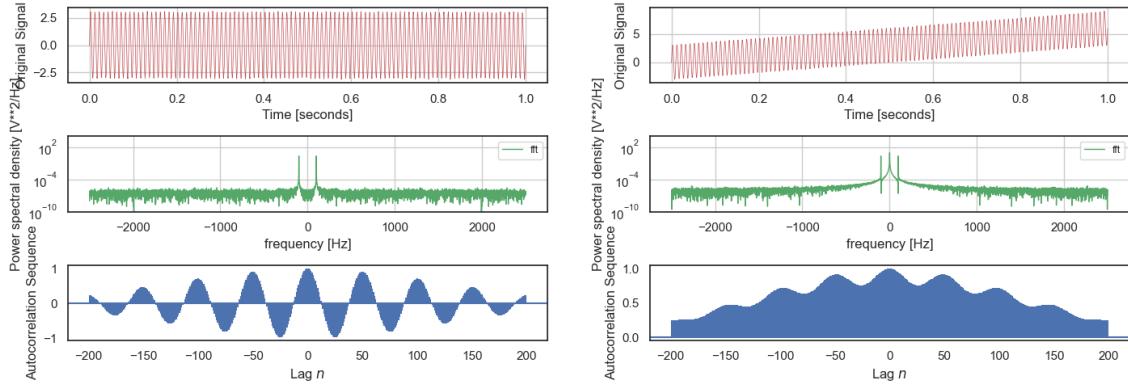


Figure 2.5: In this illustration, we depict the signal, periodogram, and autocorrelation function (ACF) for two different signals , on the left we have a sinus wave with little white background noise and on the right the same signal to which we add a drift. This drift induces a spurious long-memory effect in the ACS. The periodic pattern evident in both ACS, manifested as sinusoidal waves, denotes the presence of seasonality within the data.

The identification of whether an economic or financial time series possesses a deterministic trend or a unit root is critical in economics due to the different modelling approaches. This issue has been debated extensively in the financial data literature [125, 203]. In the former, researchers study the increments of the time series, whereas the recommended approach for the latter involves regression methods that depend on the trend. The incorrect identification and detrending of unit-root models may lead to spurious relationships, such as the Slutsky effect [159]<sup>5</sup>. We demonstrate this issue with a simple example in Figure 2.5. Further discussion of spurious relationships is provided in Subsection 2.2.1.

Initially, trend-stationary models were believed to be applicable to most (if not all) time series [54]. However, in the 1980s, researchers began considering unit autoregressive roots, as evidenced by the well-known Nelson-Plosser test [150], which provided statistical evidence supporting the hypothesis of a unit root for macroeconomic time series such as GNP, employment, wages, and interest rates to name only a few.

Unit root tests are notoriously challenging, and it remains difficult for researchers to reject the unit root hypothesis for many economic time series [54, 119, 125, 160, 203]. In many cases, the question of identification remains unresolved. More recently, additional spurious relationships have been discovered, making the identification problem even more challenging. For example, in [161], Phillips showed that when two time series are statistically independent but have a trend, they admit a regression representation, thereby exhibiting a spurious relationship. Remarkably, trends and unit roots may be two sides of the same coin: Phillips also showed that a Brownian motion (a continuous AR(1) process) can be represented by deterministic functions of time with random coefficients. Furthermore, it is known that standardised difference-stationary discrete time series converge weakly to a Brownian motion; hence, they are representable with deterministic time functions. A remarkable conclusion from Phillips on this topic is, quoting him [161]: *Our results show that such specifications (trend stationary processes) are not, in fact, real alternatives to a UR model at all. Since the UR processes have limiting representations entirely in terms of these functions (deterministic), it is apparent that we can mistakenly "reject" a UR model in favour of a trend "alternative" when, in fact, that alternative model is nothing other than an alternate representation of the UR process itself.*

In finance there are multiple causes behind non-stationarity. One is regime shifts; the Honk-Kong Stock Index (HSI) showed one behaviour from the years 70 until before the year 2000 but radically changed (in terms of market returns and volatility) ever since. This shift can be attributed to exogenous factors, such as political intervention, wherein Hong Kong split from the United Kingdom and became a Chinese province in 1987. Fig. 2.6 displays the log price of the HSI, highlighting the two trends that characterize the pre-2000

<sup>5</sup>In time-series analysis, it is well-known that if there is seasonality in the data, the ACF should exhibit periodicity, as shown in Fig. 2.5, where the amplitude decreases as the lag increases. The presence of a trend in the data can also lead to a slow decay of the entire ACF, and these effects can compound one another.

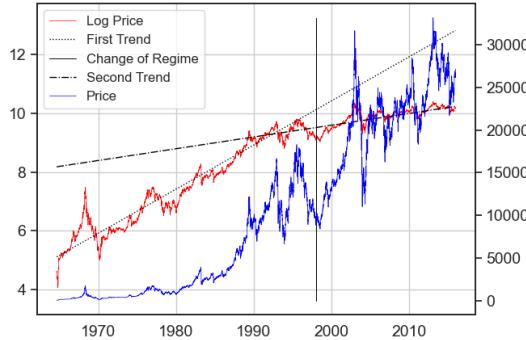


Figure 2.6: Honk-Kong Stock Index from 1960 to 2020. We notice two trends in the log price: until 2000 and then until today. This exhibits the non-stationarity of financial time series. The trends and regimes are visually identified, but they are usually identified as such using robust statistical methods.

and post-2000 periods.

Financial time series are highly non-stationary. Another one is how bubbles are the product of unsustainable growth, and a change of regime/some kind of non-stationarity has to happen before the singularity. Time-dependency was also observed for the parameters of the Hawkes process in numerous studies and was, in the end, one of the missing tools of the endo-exo discussion (see Subsection 2.2.1). Time-dependency is a way to resolve the difficulty of non-stationarity, classically done in many studies [69, 71, 174, 206, 208, 212].

### 2.1.7 Jump and Rough Models for Asset Pricing

There is an opposition between researchers standing for jump processes and for rough models. By design, they represent two visions of the markets: one considers that the market is discontinuous and that we observe shocks. The other one considers the price as a continuous process whose volatility is "rougher" and hence can vary in value suddenly, creating unexpected variations in price.

We introduced log Lévy models in Section A.2 as a mean of modelling equity [137]. While log Lévy models have demonstrated strong performance in approximating the implied volatility smile in terms of both floating smile and volatility smile, as established in previous research [16, 104, 143], they are limited in their ability to reproduce more complex market phenomena and are subject to several drawbacks. These include imprecise calibration for multiple maturities simultaneously [196], belonging to the class of "sticky moneyness" models, which results in the implied volatility of options in the model being independent of time [196], and a smile that flattens too quickly with maturity.

Hawkes processes have emerged as prominent candidates for financial modelling, primarily due to their ability to exhibit clustering. It has also been noted that the autocorrelation function of the generated log volatility asymptotically follows the function of the offspring density i.e. the kernel, suggesting that careful selection of the kernel function suffices for good calibration [10, 64]. Additionally, Hawkes processes are easily adaptable to incorporate desired features such as roughness, achieved via nearly unstable kernels.

Recently, bivariate Hawkes processes have received attention in their ability to model the joint buy-sell orders, where the price can be modelled as the difference between the two coordinates of the process [64, 167] with increments of the size of the tick. [11] proposes a microstructure model based on the difference of two Hawkes processes, enabling the study of the market impact within the context of mid-quote price. This model accounts for high frequency stylized traits such as endogeneity, mechanisms preventing statistical arbitrages, asymmetry in the liquidity of the bid and ask of the order book, and the presence of metaorders split over an extensive period of time [64] creating a long lasting memory effect.

In other words, Hawkes processes are in essence high frequency/microstructural models. They have found widespread applications, such as modelling equity-indexed annuities [88], limit order books [1, 92, 144], and financial bubbles. Despite common belief that peaks of volatility introduce crashes, empirical studies have shown that only 30% of crashes are preceded by a burst of volatility, while the remaining part presents a

sudden burst of volatility only during and after the crash, making volatility an unreliable indicator of the maturation of a bubble or an incoming crash, as demonstrated in [187], an extensive empirical study of 40 historical bubbles. Hawkes processes reproduce this pattern, where current negative jumps may generate more jumps in the future, creating a chain reaction of downward jumps when the market starts dropping without showing any forewarning signs before the crash. Hawkes processes can also be discretised to account for coarser time grids or used with bigger jump size, embodying the aggregation of batches of jumps creating variation of a few percent.

Stochastic volatility models, driven by a Brownian motion, occupy a position on the opposite end of the financial modelling spectrum. They are praised for their simplicity, typically necessitating only a few parameters. Furthermore, they incorporate some notable macroscopic characteristics and are easily understood. However, they are frequently criticized for their inability to account for the long memory of observed return volatility. Moreover, they are known to lack certain essential properties, which hinder their reliability when calibrated to actual data, such as accurately fitting the short-end of the volatility smile.

Fractional Brownian motion-driven price processes were considered a potential solution, but were found to contain arbitrage opportunities and related issues, as highlighted in [38, 165]. As mentioned in [165], the NFLVR condition, which is equivalent to the existence of an equivalent martingale measure [55, 56], does not necessarily ensure the absence of arbitrage in asset price processes in the strict sense: there are examples of asset price processes where the NFLVR condition does not hold and there are no arbitrage. On the other hand, rough stochastic volatility models offer an arbitrage-free alternative [141] that exhibits appealing properties, see the introduction of [64, 87]. Additionally, there exist alternatives to rough models that are able to reproduce "rough" behaviour: an energetic OU process (the OU-OU process), with strong mean-reversion and high volatility [166]. Lastly, some academics have attempted to combine stochastic volatility models with jump-diffusion models, such as the Bates model [16].

## 2.2 Point Processes Used for Financial Data

### 2.2.1 Endo-Exo Impact, a Long Lasting Spurious Relationship

We start this subsection with a quote from [185]: *Accordingly, large shocks should result from really bad surprises. It is a fact that exogenous shocks exist, as epitomized by the recent events of Sept. 11, 2001, and there is no doubt about the existence of utterly exogenous bad news that move stock market prices and create strong bursts of volatility. A case that cannot be refuted is the market turmoil in Japan following the Kobe earthquake of Jan. 17, 1995 that led to a total cost estimated around 200 billion dollars. Indeed, as long as the science of earthquake prediction is still in its infancy, destructive earthquakes are not endogenized in advance in stock market prices by rational agents ignorant of seismological processes. One may also argue that the invasion of Kuwait by Iraq on Aug. 2, 1990 and the coup against Gorbachev on Aug., 19, 1991 were strong exogenous shocks. However, some could argue that precursory fingerprints of these events were known to some insiders, suggesting the possibility that the action of these informed agents may have been reflected in part in stock markets prices. Even more difficult is the classification (endogenous versus exogenous) of the hierarchy of volatility bursts that continuously shake stock markets.*

Trying to summarise a decade-long reflection on untangling endogenous and exogenous impact in a short subsection would be presumptuous. Theses have been written on this topic [205], and we will limit ourselves to giving a feeling about what was at stake.

There has been a debate (spoiler alert, won by the supervisor of this thesis, Prof Didier Sornette) between the two research groups: Filimonov / Sornette (FS) and Hardiman / Bouchaud (HB). Both groups have fit Hawkes processes (endo-exo processes) to financial markets to capture its properties.

First, Sornette's group claimed that the market has increased from being somewhat self-exciting with a branching ratio of  $0.3^6$  to moderately self-exciting with a branching ratio of 0.7 in [71]. This shift in the value of the reflexivity of the markets is attributed to high-frequency trading and other computational methods [206]. Later, the HB group affirmed that the markets are critical and with a long memory [9, 90] (with a branching ratio close to 1) when they fit wide observation windows to the data. It was also claimed that the

---

<sup>6</sup>As explained in Definition A.1.3.

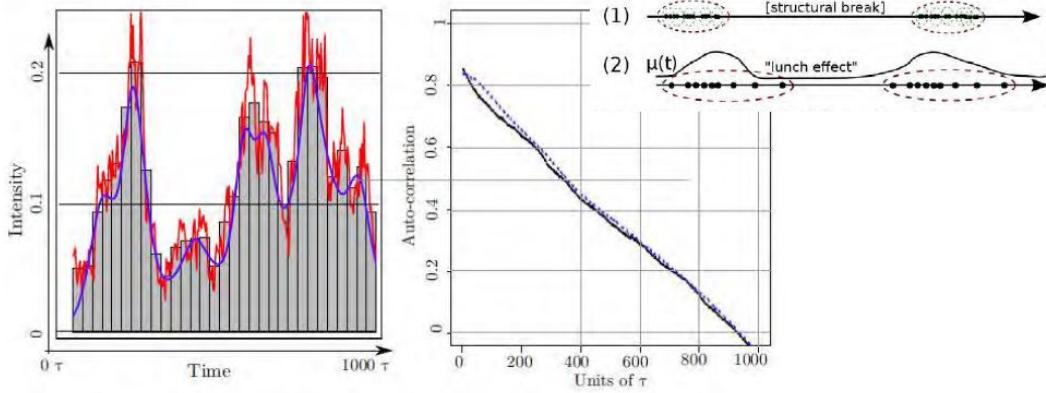


Figure 2.7: This experiment has been made and presented by Professor Sornette's. After sampling a time series generated by a near-critical Hawkes process, we collect and represent in red on the left plot the true exo-endo intensity. The histogram of the sampled time series is also represented. Then, we estimate the parameters of a purely exogenous process with intensity drawn in blue. It is clear that despite the inaccuracy of the model, we have been able to recover most of the dynamics of the intensity. Then, we sample a time series from that estimated model and compare its autocorrelation sequence to the original's time series, both drawn on the right graph. It is clear that both time series are identical from the ACF point of view, which indicates that higher order moments are required to perform a sound estimation. On the right plot,  $\tau$  represents the characteristic time of the Hawkes process' kernel.

memory with the branching ratio was growing with the window size [120]. However, the FS group proved it to be a spurious apparent criticality in [69] and later strongly rejected the hypothesis of criticality within the univariate micro-structure [208, 212]. They estimated the branching ratio at around 0.7 - 0.8, up to 0.85. We give more details below.

As exposed in [212], time series and temporal point process models are similar in approach. In both cases, we have two main classes of models, the autoregressive AR and moving average MA ones (which in the domain of TPP, each correspond respectively to the Hawkes and the shot noise processes, as exposed in [117, 174] or Section 1.3). Similarly, we match the time series problem of unit-root/trends to the endo-exo problem faced in TPP modelling. In both fields, it is a difficult problem to untangle and can lead to spurious relationships; for example, untangling endogenous phenomena from exogenous is not easy: misspecifying the exogenous rate  $\nu$  yields a bias in the estimation of the branching ratio: when underestimated, exogenous events are attributed to the self-excitation component of the intensity, and the branching ratio is overestimated. We show an extreme example of untangling in Fig. 2.7 where the non-homogeneous exogenous rate (Poisson) matches almost exactly the stochastic and realised, conditional intensity of a Hawkes process. Then, in such a case the autocorrelation functions are identical despite the process being fundamentally different: one is a pure exogenous process whereas the other one is endogenous. It also indicates that the first-order moment does not carry enough information to fully distinguish exogeneity from endogeneity. Another example in [69] lies in Fig. 16 where they estimated the branching ratio of Poisson processes (hence with no self-excitation mechanism) but recovered criticality.

Despite the complexity of the task, it seems to be a solved problem thanks to Sornette's group's extensive work on this topic in the case of TPP, and the brilliant idea of using the EM algorithm to estimate the parameters and whose formulation naturally untangles endogenous from exogenous events. The mistake was underestimating external shocks of exogeneity and confusing them for endogeneity<sup>7</sup>. A remarkable final achievement of solving the endo-exo problem is the prowess of providing a possible explanation to the excess volatility puzzle [48, 178] with the intrinsic endogeneity of the markets, providing a conclusive answer to a half-a-century-discussion [207].

We conclude this subsection with a small digression on the EM algorithm, which is the main tool used to solve the endo-exo problem. As already noted by [69], and further explained [212], second-order moment-based

<sup>7</sup>So-called spurious effect of inference endo-exo

estimations are performing poorly at discriminating between endo and exo trends. All the nonstationarities we have mentioned produce a complex layering where different causes create the same moments pattern, so we can expect difficulties in estimating the parameters with a moment-based estimator. Alternatively, likelihood estimators have a perspective on the data (exploiting higher-order features) that enable them to perform an efficient model identification. A maximum likelihood estimator is much more powerful in such a scenario, and the EM algorithm is perfectly suited where the dependency on the immigration naturally unfolds. The EM algorithm shows even stronger features in the multivariate setting where it decomposes and allows tracking of the origin of the different events, whereas a naive MLE would fail to converge easily. It is however not yet so popular in financial applications, perhaps due to its iterative nature and computationally intensive algorithm.

### 2.2.2 Criticality of the Markets

In the article [69], the authors make numerous precise statements about what could yield biases in the estimation of the parameters of the model. In particular, this paper shows one of the first occurrences of using the EM algorithm in the field of finance.

Some biases are caused by:

#### Flaws in the Data:

- Hawkes processes' likelihood might present several local minima, in particular in the context of real, with missing features, noisy and sometimes erroneous data, as shown in Fig. 7 of [69].
- It is now common knowledge that overnight trading, or more generally trading from outside the regular trading hours, represents a growing proportion of the total trades of the day. In 2010 for example, the number of changes in the mid-price happening outside of the regular trading hours amounted to over 65% of the total of changes over a day for the E-mini S&P 500 Futures Contract.
- The data might be biased by the way it is registered in the books. The FAST/FIX protocol (one of the most common protocols for communicating financial data) bundles different trades in the same message. This indicates that the timestamps of real data from the markets suffer from overhead due to processing the data on exchange platforms, the latency of the communicated message and also does not respect the "simple point process" assumption of the Hawkes processes and assumption of the uniqueness of time stamp per message because of the bundling effect.

#### Flaws from the Calibration of the Model:

- Time series may exhibit non-stationarity. We mention the possibility of seeing intraday trading patterns (the infamous U-shaped pattern of trading activity during the day, [53, 63], indicating that the hour following the open is much more active than the hour around noon) and this suggests that time bars are a poor way to process data<sup>8</sup>. But more importantly in this context, when one tries to fit a simple Hawkes process, the larger the window, the bigger the differences between piecewise stationary regimes are, affecting tremendously the calibration of the model. See different regimes in Fig. 14 and 15 of [69]. They show that concatenating different regimes gives criticality to the calibrated model. This is reminiscent of our previous discussion from Subsection 2.1.3.
- They also stressed that outliers in the data are interpreted as long waiting times. The model endogenizes the event, whereas it could have been due to a succinct burst in the exogenous rate. This bias is directly linked to the edge effect (see Fig. 4 and 5 from [69]).
- The choice of the kernel plays a crucial role as well, and the misspecification of it induces a noticeable bias for power law kernels.

---

<sup>8</sup>We refer to the notion of bars and time bars vs tick or event bars from the insightful book [53]. In particular, CPU cycles' time would perhaps be a better way of stamping the data rather than arbitrary time intervals [62].

All of the mentioned points are potential reasons why the criticality of markets is falsely identified. When the shortcomings are accounted for and the underlying model is not changed, a sub-critical Hawkes process is fit (and criticality is strongly rejected) as seen in [212]. In this paper, the authors have also compared the estimations of the parameters when the window of observations becomes wider. They postulate that the reason why most experiments observe a branching ratio increasing towards 1 as the window size  $T$  grows towards infinity is an artefact and a spurious conclusion of the over-simplistic choice of the model. The authors observe in [212] that the optimal number of degrees of freedom required for an appropriate fit of the exogenous rate scales approximately linearly with the fitting time, and the scaling depends on the market: one degree of freedom per hour is expected for the EUR/USD market, whereas the E-mini mid-price process would require three degrees of freedom per hour in average.

Furthermore, it is striking that in the experiments, high spurious branching ratios are achieved upon completely exogenous data because the exogeneity parameter is forced to be constant. They also notice some indication that spuriously high endogenous excitation and long-memory correlate, and that the spectral radius of the branching matrix and memory of the kernels drop when one increases the flexibility of the exogenous rate.

A potential reason why the parameters of the Hawkes process require to be non-stationary is related to the heterogeneity of the profiles of traders, with different reaction times, sensitivity... This argument of heterogenous reflexive agent behaviour is supported by the recent paper [162] where they have shown that algorithmic traders (high-frequency market makers) have higher branching ratios than other agents. All the different actors are not active at the same time and might act upon the new information with different delays.

### 2.2.3 Non-Stationarity for TPP

In conclusion, the discussion presented in Subsection 2.1.6 presents non-stationarity in datasets when modelling with temporal point processes and proposed solutions. To accommodate for regime shifts, [212] suggests the approach of dividing the data into segments, and performing parameter estimation on each segment. This is also implemented in [71, 74, 90, 136, 206, 212], among other studies. By rolling out the window of fixed length, a time-variant estimation of the parameters is obtained. Log splines, with varying knot positions, are also employed in the literature to represent the general shape of the time-dependent parameter. Although log splines are more difficult to implement and do not allow for perfect parallelization, they efficiently reduce the number of degrees of freedom needed to represent the parameter's value over time. The degree of freedom is selected through optimizing the BIC criterion [91], as suggested in [206, 208]. Logsplines are typically used only on the exogenous parameter. However, the Hawkes( $p, q$ ) model incorporates a spline of order  $q$  for the branching ratio, where the parameters  $p, q$  denote the number of knots of the spline. It is noteworthy that there is a significantly greater degree of heterogeneity in the exogeneity than in the endogeneity, thus  $p \gg q$ . Further on, [206] proposes to expand its Hawkes( $p, q$ ) model to the MA-Hawkes( $p, q$ ). However, this model did not show a particular improvement of the fit, relatively to the difficulty of implementing the estimation and the increase of technical issues.

## CHAPTER

# 3

## MODELLING FINANCIAL BUBBLES

### Introduction

The modelling of financial bubbles has benefited from the work of Prof Sornette's research group performed over the last few years. The goal of bubble-hunters is to understand the mechanisms and characteristics of bubbles in order to prevent them from occurring. A framework for mathematically characterizing financial bubbles has been proposed in [173] where the authors summarise the existence of two specific characterisations of bubbles, the type-I and the type-II. In both cases, the crash is central to recovering market efficiency despite exuberant price deviations.

Some type-I bubbles have been developed, such as those built on the logic of rational expectation and nonstationarity, as seen in [131, 209]. In this work, we challenge the current research by analysing potential extensions to these models, by considering a bivariate jump process to create a first of its kind type-II bubble. We leverage the capacities of type-II bubble models to capture some stylized facts not captured by previous models and, consequently, may inspire a fundamental shift in the way bubbles are modelled. Our aim is to develop a class of models that captures a wider range of crash/bubble phenomena, occurring in various asset classes, including equity, currencies, and commodities. However, current models still struggle to reproduce certain stylized facts, such as fast rebounds, negative bubbles, and convincing after-crash dynamics. Moreover, current models do not provide a good framework for long multi-scaled crashes. The challenge lies in defining the efficient phase and its connection to the inefficient phase. Current model transition from the efficient phase to the inefficient phase in a convincing manner, but the converse remains an open debate. We hope that adding a second dimension to the Hawkes process will yield improvements in that direction.

Our methodology involves reviewing the literature on bubbles and analysing the similarities and differences between real-life bubble instances in Section 3.1. We note that bubbles are rampant in a variety of fields, and the connections between them are strong, as seen in [108]. This includes positive and negative bubbles, general observable patterns, as well as the related LPPLS framework, dragon kings, and mechanisms of self-organizing structures during crashes.

Then, in Section 3.2, we present and investigate a novel type-II bubble and the usual stylized facts considered for bubble models in Subsection 3.1.8. Our model is composed of a bivariate Hawkes process and a multi-scale non-local mispricing index, with phases corresponding to the amount of compensation for positive jumps. All of these different mechanisms are detailed and rationalised in the text. We conclude this section by looking at real prices and searching for the parameters that best mimic the patterns in empirical data.

We emphasize that all the experiments of this chapter can be found in the [public GitHub repository](#).

## 3.1 What are Financial Bubbles?

### 3.1.1 The Universality of Bubbles

The question of the origin of crises touches every field: *"Are large biological extinctions such as the Cretaceous/Tertiary boundary due to a meteorite, extreme volcanic activity or self-organised critical extinction cascades? Are commercial successes due to a progressive reputation cascade or the result of a well-orchestrated advertisement? Determining the chain of causality for extreme events in complex systems requires disentangling interwoven exogenous and endogenous contributions with either no clear or too many signatures."* Quote from [185], which is the first sentence of the abstract.

Crises are rampant in all systems:

- Natural disasters, such as earthquakes, volcanic eruptions, hydrodynamic turbulences etc,
- Natural phenomena such as lightning strikes, diseases and epidemics, epileptic seizures, magnetic storms etc,
- Human-made projects, such as failures of engineering structures, material failure symptoms, crashes in the stock markets, protests and revolutions, traffic jams etc,

as well as in many other phenomena. The impact of crises on our life lead to flourishing literature on this topic [110, 184, 185, 195] Crises are intrinsically associated with bubbles. In complex systems, crises can be understood as the burst of an accumulation of slight deviations from a natural state. When taken separately, they are no harm to the complex global system but they lead the system to an unsustainable path.

The related research questions are:

- How can we model these phenomena?
- Can we forecast crises?
- How can we mitigate or prevent bubbles?

They could be better answered if we understood the origin of crises. One class we understand is earthquakes, particularly with the ETAS model: [95, 146, 147, 148, 149, 154]. Conversely, financial bubbles are also getting a growing interest. Both financial and seismology questions are deeply rooted in the issue of identifying whether the shock is due to external or internal shocks (in earthquakes study, we talk about mainshocks and aftershocks [95]). This discussion could be extended to many other fields (natural, artificial or social), and we leave the interested reader to see the references [183, 185].

Understanding bubbles and crises is critical in finance as they relate to the decades-long controversy of excess volatility puzzle [48, 178]. If markets are efficient, crises should be attributed to unpredictable exogenous shocks. Conversely, if we accept the reflexivity of the market, crises could be due to an aggregation of small perturbations that would end in excess volatility and excessive valuation of assets.

### 3.1.2 Dragon Kings, the Hope of Predicting Future Catastrophes

The power law distributions epitomize the remarkable property of scale invariance: assuming a function  $f$  follows a power law implies that scaling the input  $x$  by a constant  $c$  gives an output proportional to the output  $f(x)$ :  $f(c \cdot x) \propto f(x)$ <sup>1</sup>. In other words, the ratio of the probabilities of getting an event  $c$  times stronger is proportional to the ratio of the sizes. Extreme events are scaled versions of their smaller counterparts, which can be interpreted as the principle that the underlying mechanism is the same at any scale. Extreme phenomena are unpredictable because of the common mechanism and the independent nature of the events.

In such context, extensive literature and zoological ideas tried to embody these extreme, unpredictable and fearful concepts: [8, 184, 195, 214]. For example, the "*Black Swan*" is the idea that an event beyond the realm of normal expectations happens making it difficult to comprehend its existence, and which is inappropriately rationalised in hindsight. Nevertheless, Pr Sornette, with his "*Dragon Kings*" highlights that, despite the scalable nature of the distribution, some self-organising mechanisms might still exist. Some examples are phase transition, bifurcation, tipping point etc. More can be found in [184, 190]. The idea of "*Dragon Kings*" opens the path to answering the motivating questions of this section. We will focus next on financial bubbles.

---

<sup>1</sup>Take for example the function  $f(x) = x^p, p \in \mathbb{R}$ .

### 3.1.3 Diagnosing Bubbles

We describe the aforementioned self-organising mechanisms, as presented in [186]. Stated shortly, bubbles are processes with unsustainable growth, showing faster-than-exponential growth. Exponential growth is the expected growth of markets (due to the compounding of interest rates). Bubbles are however growing faster than the markets which makes them unsustainable. The growth is attributed to a positive feedback mechanism which fuels the bubble to the point where it cannot grow more because it reached a singularity point. Hence, the shape of bubbles is closer to a hyperbolic power law than to the exponential function, see Figure 3 in [186].

In the article, the authors detail some mechanisms that can produce a positive feedback mechanism. They mention:

- **Technical Arguments:** when prices go up, option hedging and portfolio insurance technics encourage investors to buy more of the underlying stock. Under the classical Black-Scholes option theory, a call option's risk is eliminated by buying more of the underlying. This is one of the many of the technical foundations of positive feedback.
- **Behavioural Mechanisms:** herding behaviour, imitation, momentum... classical trading biases.

Bubbles grow by waves and not regularly accelerating, which can be explained by many logical reasonings [186]. We summarise one thought process here: assume an asset is priced below its expected fundamental value. Due to the momentum effect, the price exceeds its target. When re-evaluated later, it creates momentum in the opposite direction, which is the empirical rationale behind the oscillations. The nonlinear relationship between the amplitude of the oscillation and the frequency accelerates the bubble growth. During bubble formation, we observe increasingly larger oscillations with smaller amplitude, a phenomenon known as log-periodicity. This is natural and has been observed in a variety of dynamic systems, such as hydrodynamic turbulence, chemical and biological growth processes. These oscillations can be seen as smaller nested bubbles compared to the overall larger bubble. These observations led to the development of the log-periodic power law singularity model (LPPLS), which can be found in [186] or [217].

The crash follows a similar pattern to the bubble, with a negative bubble pattern and the same oscillation with faster-than-exponential decay. In Fig. 8 of [186], we can see a positive and negative bubble pattern during the crises of 2007 in the S&P price. Positive and negative bubbles refer to irrational growth and decline, respectively, where the mechanism behind is a buying frenzy coupled with a fear of missing out or, conversely, a fear of losing, creating a panic to sell and exit the market [78].

We finally clarify that our discussion is not exactly related to the determination of a fundamental value. Transient non-sustainable exuberant market patterns are distinct from the existence of a profound mispricing. Instead, we identify non-sustainable dynamics that are bound to crash (positive bubble) or rally (negative bubble) and seek to determine when it will occur and how much percentage change will result.

### 3.1.4 Financial Bubbles

Financial crises in financial markets are typically characterized by drops in value that are greater than 10%, while market corrections range from 5% to 10%. These drops may take several years to reach their minimum, and may be followed by a bearish market phase during which the market continues to decline steadily for extended periods of time. In [217], the authors define a crash as a drop of at least 15% in less than three weeks. Thus, over the course of 100 years, one can expect to see three major crises or crashes (with drops greater than 30%), eight crises (with drops greater than 20%), and 100 corrections. However, it is important to note that these quantities depend on the type of market and there are significant differences between Asian, European, and American markets. Also, contrary to a naive belief, markets are relatively stable because changes of over one percent happen every week or two on average (even though they cluster in time).

The paper [173] summarises the global idea of how to mathematically define a financial bubble. At a high level, we call a financial bubble a situation where the market price of an asset exhibits an irrational increase and exceeds the asset's true value or fundamental value. Two schools of thought have emerged regarding the origins of bubbles. The first group states that rational expectation bubbles are possible, meaning that bubbles can occur in efficient markets (with the first occurrence documented in [22, 23]). The second group

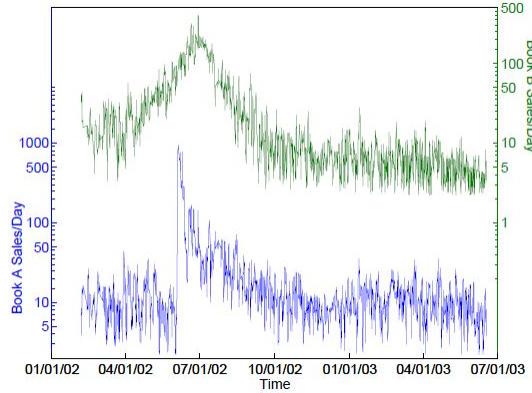


Figure 3.1: This figure taken from [185], depicts the sales frenzies for two books where the trigger for them has been clearly identified. We present the time evolution of sales per day for Book A (bottom, blue, left scale), which is "Strong Women Stay Young" by Dr M. Nelson, and Book B (top, green, right scale), which is "Heaven and Earth (Three Sisters Island Trilogy)" by N. Roberts. The pattern difference between the two is striking, with Book A experiencing an exogenous peak on June 5, 2002, while Book B reaches a maximum endogenously on June 29, 2002.

suggests that a violation of market efficiency is necessary for a bubble to occur, and that a subsequent crash is required to restore efficiency globally.

### 3.1.5 Are There No Universality to Bubbles?

We argued that bubbles and crises are universal.

However, there exist diverse mechanisms that underlie different types of bubbles, which exhibit a range of characteristics and origins. Owing to the complex nature of the system, bubbles are triggered by distinct factors, and consequently, different bubbles present diverse patterns. This has been observed at varying time scales, ranging from high-frequency flash crashes [206] to large market crashes [70, 110].

We focus on one critical differentiation, namely, endogenous crises from exogenous crises. Certain events can be attributed to an internal self-excitation among agents, while others are a direct response to external shocks. We expect these two types of crises to exhibit different patterns, some of which are discussed in [189]. This concept has been previously examined in [183] and further developed in [185]. If we coin  $A_t$  the response to a phenomenon,  $t_c$  the time when the event occurred, and we set  $t$  close to  $t_c$ , then empirical evidence indicates that classical human responses follow a power law response [156]:

$$A_t \sim 1/(t - t_c)^{1-\theta}. \quad (3.1.1)$$

In the case of an exogenous critical response, we observe:

$$A_t \sim 1/(t - t_c)^{1+\theta}, \quad (3.1.2)$$

and finally, in the case of a critical endogenous contagion, we would observe:

$$A_t \sim 1/(t - t_c)^{1-2\theta}. \quad (3.1.3)$$

The difference between the two types of bubbles can be observed in Fig. 3.1, which displays the sales of two books exhibiting different patterns. These patterns have been recently observed in financial data [135], where a careful study links news, price jumps, and volatility.

Flash crashes represent a particular type of market crash, distinct from other forms of bubbles and crashes. They are characterized by large, extreme, and transient drops in prices and are the result of both endogenous hyper-activity and external factors. The first recorded flash crash occurred on May 6, 2010, in the U.S. equity market, where a sudden and almost instantaneous drop of 10% was followed by a swift rebound. Ever since,

many other events of the same type were observed on the markets, as proof that markets' nature has changed since the crisis of 2007 and we are entering a digital world of finance. Flash crashes are a feature of modern electronic financial markets. High-frequency trading is commonly attributed as the primary cause of flash crashes, as noted in the report by BlackRock [59] or the paper [206]. Other reason include the increasing prevalence of algorithmic trading on Electronic Broking Services (EBS), as mentioned in the report [194], around 70% of trades were due to algorithmic trading on EBS in 2016, as well as market fragmentation and shortage of liquidity [151] among others. These different concepts have been discussed in [206] as well as in [62], where they emphasise the role of the concentration of liquidity provision (fragmentation of the market), as well as the so-called toxicity of the flow due to the reduced participation of retail investors.

We detail the literature on the topic of financial bubbles in the next subsection.

### 3.1.6 Rational Expectation Bubbles

The rational expectations market can be described as follows. Given an asset  $S$  with present market price  $S_0$  that entitles the holder to a series of diverse payoff streams during its lifetime. If the market is rational in a basic sense, the fundamental price of this asset should be the market-implied present value of exactly those payoff streams:  $S_0$ . Some authors proposed to define bubbles in such markets as called rational expectation bubbles.

However, the postulate of efficient market implies that financial bubbles should not be able to develop. A celebrated result by Tirole [198] confirms this intuition in a discrete setting, later extended to the continuous setting in [173]. According to this result, rational expectation bubbles cannot emerge in a market where all investors are rational and utility-maximizing. Instead, at least one agent must be irrational for bubbles to develop. The only remaining possibility for rational bubbles would be that crashes are entirely exogenous, which is unrealistic and proven to be false empirically because we observe bubbles in practice. If they are not rational, what irrational mechanisms crystallise bubbles?

Researchers have found that agents tend to be overconfident and undervalue risks. Bubbles tend to burst when agents realize the mispricing, even if only partially, and the price adjusts. Uninformed investors who ignore the possibility of a crash are likely to believe that the deal is an excellent opportunity, further fueling the bubble. Conversely, rational and informed investors who correctly assess the possibility of a crash may also ride the wave because they have no reason to burst it prematurely [2, 113] and benefit from the momentum. Suppose an investor were entirely certain about the mispricing. In that case, they would either try to leverage the arbitrage opportunity or stay out of the market, but not trading would reveal to their clients that they do not have superior private information that can lead to profit. Therefore, investors have no incentive to go against bubbles.

Some authors state that bubbles are caused by a break in market efficiency (in opposition to the rational expectation market). Asymmetric information, noise trading (referring to irrational traders), herding, and mimicking effects are some of the underlying mechanisms that have been identified to create and feed bubbles, as reviewed in [113]. We refer the interested reader to [173] for further discussion.

### 3.1.7 Type-I and II Financial Bubbles

We have seen that instantaneous rational bubbles do not exist: some kind of market failure must appear in the markets for bubbles to develop (both in discrete and continuous cases). In a recent publication, the authors of [173] offer two frameworks to model markets with failures: type-I and type-II bubbles. Both rely on explosive semimartingales that encompass all discrete and continuous-time bubble models in the literature featuring market inefficiencies. In line with them, we use their two categories and focus on the type-II group.

Type-I bubbles assume the entire market is efficient, whereas type-II bubbles emerge in an efficient market and transform into a non-efficient one and the market crash is the means through which the market regains efficiency. Theories supporting the concept of type-II bubbles include heterogeneous beliefs, which propose that bubbles arise from the prior heterogeneity of beliefs in the asset's fundamental value, as well as behavioural finance theory and complex system theory, which propose that bubbles are the result of noise investors and positive feedback mechanisms due to imitation and collective herding behaviour.

Several models for these two types are presented in the paper [173]. In Subsection 5.3, the authors compare an efficient market and a type-I bubble. A type-I bubble allows for bubbles because it includes a

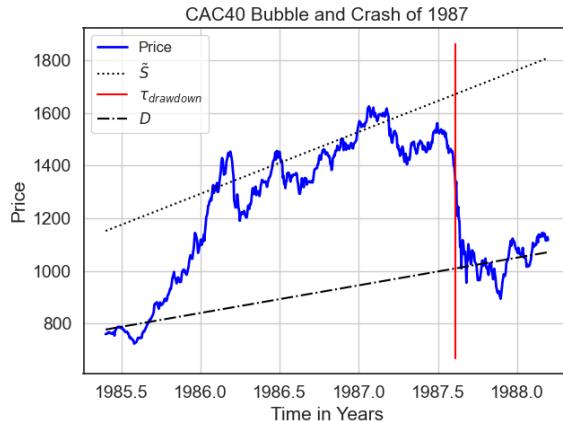


Figure 3.2: The figure shows a time series representing the crash of the CAC40 stock price bubble, where the characteristic triplet  $(\tilde{S}, \tau_{\text{drawdown}}, D)$  is indicated. The estimated returns over each period are provided as rough approximations computed by hand for illustrative purposes.

positive feedback loop for returns. The higher the probability of a crash, the larger the desired returns for investors matching the risk of a crash.

However, type-II bubbles seem to have more realistic foundations. They do not necessitate the assumption of omniscience from investors or that rational market participants are aware of the risk of a crash. For this reason, in this thesis, we focus on proposing a novel type-II bubble model that can fit real data. It is essential to note that both approaches are equivalent when investigating why market efficiency can fail. In general, it is impossible to distinguish type-I from type-II bubbles from a single bubble episode.

Following the idea of [173], we characterise bubbles by the triplet:

1. A pre-drawdown process  $(\tilde{S}_t)_{t \in \mathbb{R}_+}$ ,
2. A stopping time representing when the drawdown starts  $\tau_{\text{drawdown}}$ ,
3. A drawdown process  $(D_t)_{t \in \mathbb{R}_+}$ .

We think that instead of calling the first process a pre-drawdown process, a more explicit name would be a drawup process.

Then, the asset price is given by:

$$S_t(:=\tilde{S}_t^{\tau_{\text{drawdown}}, D}) = \tilde{S}_t \mathbf{1}_{t < \tau_{\text{drawdown}}} + D_t \mathbf{1}_{\tau_{\text{drawdown}} \leq t}. \quad (3.1.4)$$

Under the model of type-II bubbles, we assume that  $D$  is an efficient market, whereas the whole process  $S$  is inefficient. At the stopping time  $\tau_{\text{drawdown}}$ , the inefficiency is resolved, and the price collapses to a fairer level.

By efficient, we refer to the efficiency defined in [173]. They distinguish between efficient phases of the market and inefficient ones. Efficiency can have different meanings (see Subsection 4.1.3. of [173]), and in our case, we talk about one of the weakest forms: No Free Lunch with Vanishing Risk (NFLVR). A sufficient condition for NFLVR is that the process is a martingale.

We can create a market that allows for bubbles by iterating the phases of drawups and drawdowns. Then, we would index each pair of regimes by an index and connect the different phases continuously. This is detailed in Subsection 3.2.3.

### 3.1.8 Financial Bubbles Stylized Facts

Good bubble models are first characterised by the shape of their sampled trajectories, which encourages us to list the stylized facts that bubble modellers are looking after. The first notable fact is that investors have

different sensitivity to market changes and investment horizons which will create different shapes and sizes of bubbles. Investors are aware of the fact that markets are not efficient and use historical value as an anchor of what could be the true price. Directly following from our discussion about the criticality of markets, bubbles have a strong endogenous part and we observe volatility clustering around the time of the crash. The endogeneity is understood either as a fear of bankruptcy or a fear of missed opportunity, leading to faster-than-exponential growths and declines, which gives bubbles their characteristic hyperbolic shape [131, 186] with a finite time singularity. At the peak of the bubble, the stock price may plateau and/or crash in a sequence of smaller drops [217]. Crashes usually happen in less than three years, and we see an asymmetry in the shape of the bubble and crash, where the crash happens quicker. Crashes happen on average every 8 years and the drop can be from 20 to 50 percent, corresponding to a change in log price of 0.2 up to 0.7. These numbers vary depending on the market and its volatility. Another characteristic of bubbles which appeared over the last decades is the fast rebound after a crash, that most older bubble models do not account for.

### 3.1.9 Non-Local Mispricing Index Used in Type-I Bubble Models

We describe an approach that combines a type-I bubble with a rational expectation type of market, as mentioned in Subsection 3.1.6. The instantaneous risk-return relationship is not compatible with the development of bubbles, which makes instantaneous rational expectation bubbles impossible. If an instantaneous risk-return relationship is assumed, bubbles should crash instantaneously in a single significant negative jump, rather than over weeks and months, which is typically the case [110]. Additionally, bubbles often plateau for some time at their peak before crashing<sup>2</sup>. In such a case, the crash hazard rate vanishes inside the model, whereas in reality, it does not.

In other words, the opinion of the agents regarding the possibility of a crash should not be modelled locally. In [131], the authors propose a non-local relationship. They offer to account for the behavioural anchoring on past price levels, which translates mathematically to considering a non-local mispricing index based on real trading methods: anchoring on price or return is a common practice by traders. Some of the many patterns used are “support” and “resistance” levels which are anchors. Some other models have tried to incorporate anchoring in their definition [100] and it helps getting faster-than-exponential growth, a feature that we expect from a bubble model as detailed in Subsection 3.1.8.

The model proposed in [131] assumes that investors estimate the mispricing by an exponential moving average and associates the mispricing index with a crash hazard rate. Both are interconnected because we want to model a positive feedback mechanism between the two: the higher the risk of a crash, the higher the required rate of returns to keep investors in the market..

Hereinafter, we summarise the outcome model. More details can be found in the original paper. We work in a complete probability space endowed with a (discrete) stochastic basis  $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \in \mathbb{R}_+}, \mathbb{P})$ , which reflects natural views on the market, like investors never forget:  $\forall s < t, \mathcal{F}_s \subset \mathcal{F}_t$ . The market is composed of two assets: one risk-free asset with return  $r_f$  and a risky asset  $\{S_t\}_{t \in \mathbb{R}_+}$  with return  $\{r_t\}_{t \in \mathbb{R}_+}$ . The model is discrete with an interval's length equal to 1.

**Risky Asset Dynamics.** We define  $r_t$  to be the log returns:

$$r_t = \ln\left(\frac{S_t}{S_{t-1}}\right), \quad (3.1.5)$$

$$r_t = \mu_t + \sigma_t \cdot \varepsilon_t - \kappa J_t \cdot I_t, \quad (3.1.6)$$

where  $\{\mu_t\}_{t \in \mathbb{R}_+}, \{\sigma_t\}_{t \in \mathbb{R}_+}, \kappa$  are  $\mathcal{F}_t$  - predictable ( $\kappa$  could be time-dependent but it has been chosen to be constant),  $\varepsilon_t \sim \mathcal{N}_{0,1}$  normal variables i.i.d.. Finally,  $J_t$  is a positive, independent random variable of  $\mathcal{F}_{t-1}$ ,  $I_t$  is a i.i.d. Bernoulli random variable where  $\mathbb{P}(I_t | \mathcal{F}_{t-1}) = \lambda_t$ . For all times  $t$ , the random variables  $J_t, I_t, \varepsilon_t$  are mutually independent conditionally on  $\mathcal{F}_{t-1}$ .

---

<sup>2</sup>For example the CAC40 was at 6000 points at the end of the year 2000, 5800 at the end of the year 2001 (with a peak at 6800 for one day) and crashes aggregated over the three following years before the CAC40 reached its lowest point of the period with 2600 points

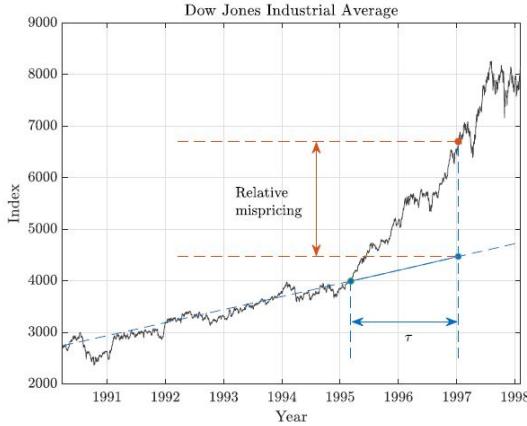


Figure 3.3: Illustration of the ratio of mispricing where  $\tau$  is the comparison period, plot from [131].

**Efficiency of the Overall Market.** We also assume average constant return rates, denoted here by  $\bar{r}$ . This last condition accounts for the no-arbitrage condition:  $r_t - \bar{r}$  is a martingale and  $\bar{r}$  can be interpreted as the required rate for risky assets (without including the bubbly part of the price):

$$\mathbb{E}[r_t | \mathcal{F}_{t-1}] (=: \bar{r}) = \mu_t - \kappa \lambda_t. \quad (3.1.7)$$

**Mispricing Index.** As explained previously, we are searching for an index representing the mispricing between the current (inflated) price and what it should have been. The details are given in Appendix D.3. We summarise the mispricing index as the discrete process  $X_t$ :

$$X_t = \frac{1}{s} \ln\left(\frac{\delta_{t-1,a_\tau}}{\delta_{ref}}\right), \quad (3.1.8)$$

$$\lambda_t = L(X_t). \quad (3.1.9)$$

where  $L$  denotes the logistic function  $L(x) := \frac{1}{1+e^{-x}}$ .

Note that  $\{X_t\}_{t \in \mathbb{R}_+}$  satisfies the equation:

$$X_t = a_\tau \cdot X_{t-1} + \frac{1-a_\tau}{s} (r_{t-1} - \bar{r}) + (1-a_\tau) \frac{-\ln \delta_{ref}}{s}. \quad (3.1.10)$$

**Volatility Model.** Finally, the model is completed by imposing the volatility to follow a GARCH(1,1) process, with coefficients  $\alpha, \beta \in [0, 1]$ ,  $\alpha + \beta \leq 1$ , and basis volatility  $\bar{\sigma}^2$ .

$$\sigma_t^2 = \bar{\sigma}^2 (1 - \alpha - \beta) + \alpha (r_{t-1} - \bar{r})^2 + \beta \sigma_{t-1}^2. \quad (3.1.11)$$

Note that when  $\beta = 1, \alpha = 0$ , we recover a constant volatility.

Directly, we see that this model is a semimartingale and is hence efficient. This model is then indeed a type-I bubble model.

## 3.2 Our New Type-II Bubble

### 3.2.1 Idea of the Model

Our work comes into the continuation of Sornette's group of modelling financial bubbles. Their prior and ongoing studies have successfully captured many of the characteristics of financial bubbles. However, they

failed to capture negative bubbles patterns, the crashes are still too fast and we wish plateau phases before a crash and generally the phase before a crash to be even less volatile. In this work, we propose to extend certain aspects of their model to gain further insights into the emergence and development of bubbles. Specifically, we model the daily closing day mid-price, building upon the work of Malevergne et al. [131], who presented a type-I bubble. We generalize their work by introducing a conscious bias in the market during drawups that is accounted for during drawdowns, leading to even more significant crashes. We accomplish this by adding an additional coordinate to the counting process, resulting in a total of two coordinates. The first coordinate represents the downward jumps, as before, while the second coordinate represents the upward jumps.

In summary, we present a novel multi-scale regime rational expectation bubble of type-II with non-local mispricing, incorporating a positive feedback mechanism that results in faster-than-exponential growth in asset prices. The novelty comes from the fact that compared to the current literature, our bubble is of type-II, which allows for even more exuberant dynamics during the drawup phase. This is a departure from the type-I counterparts, as presented in [131, 209]. Additionally, our work is in line with the current trend in Sornette's research group, as we incorporate self-exciting mechanisms in the jumps that have shown promise for type-I bubbles, despite not yet being published [209].

Our work opens the path to further research on the topic of type-II bubbles, which despite being challenging and difficult to work with, have the potential to change the way we model bubbles in many areas and in particular in finance. We expect that the additional coordinate will provide more flexibility in the possible dynamics of the diffusion and it also allows the fundamental returns of a stock to be negative while still observing an overall positive trend in the stock price. These mechanisms are underpinned by known behavioural and psychological mechanisms of traders.

### 3.2.2 Model's Equations

We work in a complete probability space endowed with a stochastic basis  $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \in \mathbb{R}_+}, \mathbb{P})$ . The market is composed of two assets: one risk-free asset with return  $r_f$  and a risky asset  $\{S_t\}_{t \in \mathbb{R}_+}$  with return  $\{r_t\}_{t \in \mathbb{R}_+}$ . We propose to model the whole market's price  $S$  through the equation, for all  $t \in \mathbb{R}_+$ :

$$\frac{dS_t}{S_t} = \mu(t) dt + \sigma(t) dW_t + (dZ_t^+ - \Omega_t d\Lambda_t^+) - (dZ_t^- - d\Lambda_t^-), \quad (3.2.1)$$

where  $\{W_t\}_{t \in \mathbb{R}_+}$  is a Brownian motion,  $\{Z_t^\pm\}_{t \in \mathbb{R}_+}$  the components of a bivariate Hawkes process and  $\{\Lambda_t^\pm\}_{t \in \mathbb{R}_+}$  their respective compensators.  $\{\Omega_t\}_{t \in \mathbb{R}_+}$  is a regime process that represents how much compensation of the upward jumps is included in the equation and which can take any value in  $\mathbb{R}_+$ . When  $\Omega_t = 1$ , the process  $\{S_t\}_{t \in \mathbb{R}_+}$  is efficient, or mathematically is a semimartingale. In summary, this model is a classical geometric Brownian motion with time dependent parameters which includes a bivariate jump part with regimes.

We define the bivariate Hawkes process' intensity as for all  $t \in \mathbb{R}_+$ :

$$\lambda(t | \mathcal{F}_{t-}) = \nu(t) + \int_0^t \Phi(t-s) dN_s = \begin{pmatrix} \nu_+ \\ \nu_- \end{pmatrix}(t) + \int_0^t \begin{pmatrix} \eta_{+,+} & \eta_{+,-} \\ \eta_{-,+} & \eta_{-,-} \end{pmatrix}(t-s) \odot h(t) \cdot d \begin{pmatrix} N_s^+ \\ N_s^- \end{pmatrix}, \quad (3.2.2)$$

for an exogenous rate  $\nu > 0$  and endogenous kernel  $\Phi$  and we untangle the branching ratio  $\eta$  and the shape of the kernel represented by normalised kernel  $h$ . The jump process is  $\{Z_t^\pm\}_{t \in \mathbb{R}_+}$  with jumps' size  $\{J_i^\pm\}_{i \in \mathbb{N}_{\geq 0}}$ :

$$Z_t^\pm := \sum_{i=1}^{N_t^\pm} J_i^\pm, \quad \forall t \in \mathbb{R}_+,$$

with compensator:

$$\Lambda_t^\pm := \mathbb{E}[J^\pm] \int_0^t \lambda_u^\pm du, \quad \forall t \in \mathbb{R}_+.$$

The first dimension denoted by a "+" embodies upward jumps and the second dimension coined with "−" embodies the downward jumps. We observe from empirical evidence that both Hawkes' parameters are linearly related to the mispricing factor  $\{X_t\}_{t \in \mathbb{R}_+}$ , in the sense  $\nu = \nu^1 + \nu^2 L(X_t)$ , where  $L$  represents as

before the logistic function. The dependency of the value of the parameters on the mispricing allows the following mechanism: when prices increase, the probabilistic number of downward jumps increase for the next step, which produces an acceleration of the returns: the higher the prices, the faster the returns increase. The underlying psychological mechanisms is that investors are conscious of the possibility of a comeback to reality through a crash/correction, hence they need to be remunerated proportionally to accept the risk of negative jumps and crashes. This builds on the standard and fundamental return-risk relationship, which naturally produce an "acceleration" of the drawup where the returns are the "speed" of the price, increasing as long as there is no downward jump. In this scenario, this rational expectation relationship between returns and risk of a crash generate the empirically observed super-exponential growth, and the mispricing index allows to correlate the probability of a crash with the prospect of higher future returns.

Then, it makes sense to write (3.2.2) for all  $t \in \mathbb{R}_+$  as:

$$\begin{aligned} \lambda(t \mid \mathcal{F}_{t-}) &= \nu^1(t) + \nu^2(t) \cdot L(X_t) + \int_0^t ((\eta^1(t-s) + \eta^2(t-s) \cdot L(X_t)) \odot h(t)) \cdot N_s \\ &= \begin{pmatrix} d_+^1 \\ d_-^1 \end{pmatrix}(t) + \begin{pmatrix} d_+^2 \\ d_-^2 \end{pmatrix}(t) L(X_t) \\ &\quad + \int_0^t \left( \begin{pmatrix} d_{+,+}^1 & d_{+,-}^1 \\ d_{-,+}^1 & d_{-,-}^1 \end{pmatrix}(t-s) + \begin{pmatrix} d_{+,+}^2 & d_{+,-}^2 \\ d_{-,+}^2 & d_{-,-}^2 \end{pmatrix}(t-s) L(X_t) \right) \odot h(t) \cdot d \begin{pmatrix} N_s^+ \\ N_s^- \end{pmatrix}. \end{aligned}$$

The mispricing index is an idea from [131] and we quickly summarise its properties in Appendix D.3.

### 3.2.3 A Bubbly Market of Type-II

Recall the type-II bubble formalism (recall (3.1.4)), for all  $t \in \mathbb{R}_+$ :

$$S_t = \tilde{S}_t \mathbf{1}_{t < \tau_{\text{drawdown}}} + D_t \mathbf{1}_{\tau_{\text{drawdown}} \leq t}. \quad (3.2.3)$$

By looking at the jump-diffusion equation (3.2.1), there is no doubt that this process is a type-II bubble. The stopping time to distinguish the drawup and drawdown phase is  $\tilde{\tau}$  as the first time  $\Omega_t$  hits 1<sup>3</sup>. It is a stopping time because  $\Omega_t$  is adapted. Then, the drawup process is simply the solution to (3.2.1), and the drawdown process would be the solution to (3.2.1) again but with  $\Omega_t \equiv 1$ , using the appropriate initial condition so the overall process is continuous.

The process before the stopping time is inefficient, in the sense that it does not satisfy the NFLVR condition since it is not a semimartingale, whereas after the stopping time and as long as the process remains in the drawdown regime, it is efficient by the fact that it is a semi-martingale.

When we have an observed time series of length  $T$ :  $[0, T]$ , the two regimes are iterated, in the sense that over a time observation of  $[0, T]$ , we might observe an unknown number of bubbles, composed of pairs of regimes described as above (drawups and drawdowns). The equations for drawup and drawdown regimes are the same for each pair. Still, we compute solutions starting with a different initial condition by connecting the different paths of pairs of regimes continuously.

In a similar fashion as we have described bubbles by a triplet in Subsection 3.1.7, we can characterise bubbly markets by a quadruplet:

1. A drawup process  $(\tilde{S}_t)_{t \in \mathbb{R}_+}$ ,
2. A family of stopping times representing when the following drawdown starts  $\{\tau_{\text{drawdown}}^i\}_{i \in \mathbb{N}_{\geq 0}}$ ,
3. A drawdown process  $(D_t)_{t \in \mathbb{R}_+}$ ,
4. A family of stopping times representing when the following bubble starts  $\{\tau_{\text{drawup}}^i\}_{i \in \mathbb{N}_{\geq 0}}$ .

---

<sup>3</sup>And we can define the drawup regime when the regime process is not equal to 1.

### 3.2.4 Mathematical Framework

Stability and existence of the process defined via (3.2.2) is ensured whenever the branching ratio of the endogenous part of the jump component is smaller than 1, namely  $\rho(\eta^1(t) + \eta^2(t)) < 1$ . However, some research suggests the presence of a transient super-critical phase [209]. We integrated that in our model by authorising the model to be critical whenever  $L(X_t)$  is larger than a certain threshold, close to 1. The model would still be overall stable despite transient critical phases. For this study, it did not seem relevant to include the results.

The model accounts for good and bad news through each respective dimension of the Hawkes process. The self-excitation mechanisms also mimic what is actually happening in reality. It has been presented in [131, 209] that both the self-excitation mechanism and non-local mispricing index were great tools to get explosive bubble with faster-than-exponential returns, while at the same time, getting crashes that present nested patterns of smaller crashes instead of being instantaneous.

Another positive point of the model is the returns of a drifted Brownian motion are uncorrelated by definition, as well as the returns of the jump process when the jumps are Hawkes. This means that the clustering would not be visible from the ACF and that our Lévy processes respect the stylized facts of no autocorrelation for returns we have described in Section 2.1. The self-excitation mechanism from Hawkes processes is natural as it embodies behavioural biases in markets (the herding phenomenon, a classic bias for investors) and it quantitatively helps recover the clustering effect, visible in the returns as well as in the realised volatility process. From the point of view that the jump process embodies the news process, it makes sense to model good or bad news as a self-exciting point process because news may be related (for example, in the electric car industry: good news about batteries is highly correlated to innovation in electric cars). The jump process has the natural interpretation of representing a flow of news. News can drive price up: for example a technological breakthrough in the company field, or down: for example unexpected difficulties in getting the raw materials or an increase in the price of commodities.

Finally, the regime process is a tool that recently increases in popularity in the asset modelling literature, for example in [144].

### 3.2.5 Discrete Time Grid Model

In this subsection, we assume that we are working on a discrete-time grid with an interval equal to  $\Delta_{\text{Haw}}$ . We approximate the multivariate Hawkes process  $N$  from (3.2.1) by its associated INVAR process  $N^{\Delta_{\text{Haw}}}$ , as explained in the beginning of this thesis in Theorem 1.3.13. We will refer to this chapter or interchangeably directly to the working paper [34]. This approximation is sound: we are working with a stationary Hawkes process (with a spectral radius smaller than one) and our assumptions on the parameters  $\nu, \eta$  match the ones of the theorem, namely being piecewise continuous and positive.  $\Delta_{\text{Haw}}$  here equals one day, the frequency of our data. We assume that the observations of the time series before the windows are zero.

[34] tells us that for every time  $t \in \mathbb{R}$  and matching time step  $n \in \mathbb{Z}$  (where  $t \approx n\Delta_{\text{Haw}}$ ):

$$\nu(t) + \int_0^t \Phi(t-s) dN_s \approx \nu(n) + \sum_{j=1}^{n-1} \Phi(n-j) \cdot N_j^{\Delta_{\text{Haw}}}. \quad (3.2.4)$$

This yields the discrete (multivariate) analogue to the conditional intensity for all  $n \in \mathbb{Z}$  of the conditional intensity (3.2.2):

$$\lambda(n | \mathcal{F}_{n-1}) \approx \nu(n) + \sum_{j=1}^{n-1} (\eta_{n-j} \odot h(j\Delta_{\text{Haw}})) \cdot N_j^{\Delta_{\text{Haw}}}. \quad (3.2.5)$$

For simplicity, we assume in the rest of this study that the  $J_i^\pm$  are constant with magnitude  $\kappa_\pm$ , the Hawkes kernel  $\Phi$  is geometric in all dimensions, with constant parameter, and the volatility  $\sigma^2$  is constant for simplicity.

### 3.2.6 Synthetic Bubbles and Crashes

Here we present the characteristics of synthetic realizations of our model. By sampling a synthetic time series for a model, we express visually some characteristic properties of the model in order to evaluate its initial

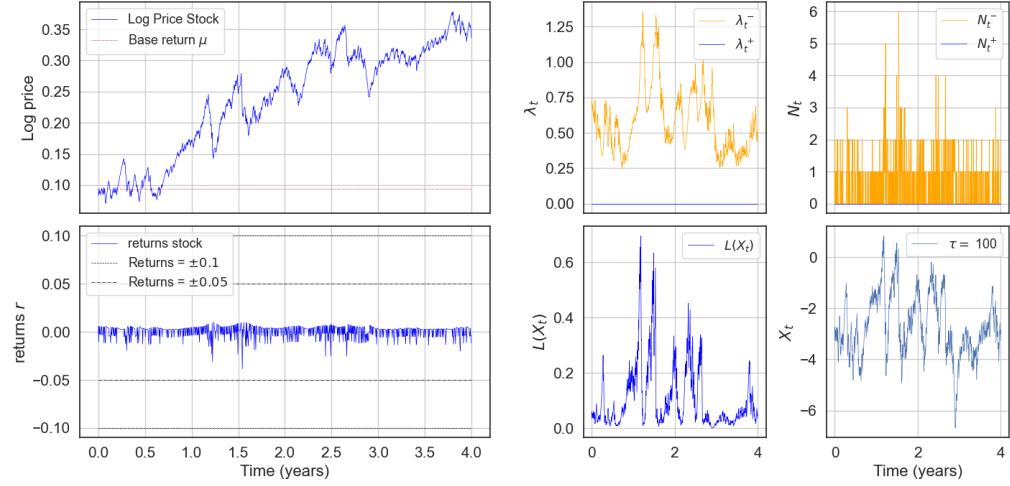


Figure 3.4: We sampled a path for the type-I bubble model detailed in (3.2.6) where the drift and volatility are null, which allows us to focus on the jump process. We plot the log price, the returns, the underlying conditional intensity as well as the number of jumps and the mispricing index for every time  $t \in \mathbb{N}_{\geq 0}$ , between 0 and 4 years. We see that the increase of  $\lambda_t$  yields clustering of the jumps and provokes a crash after a build-up phase. The blatant clustering of returns is expected from Hawkes processes. The shape of these synthetic bubble and crash episodes are very realistic and we observe many of the stylized facts of bubbles presented in Subsection 3.1.8. One remarkable fact is the absence of realised volatility before a crash, as we can observe at time 14 just before the crash happening at 14.5 when the volatility increases and remains high afterwards. This effect is emphasized when the parameters are dependent on the mispricing index. Without the non-local mispricing index, we could not get such an important localised clustering effect. Note that excessively low returns are associated with excessively high number of jumps. The parameters for this trajectory are:  $d_1^1 = [0, 0.1]$ ,  $d_2^1 = [0, 0.8]$ ,  $d_{1..}^1 = [[0.0, 0.0], [0.0, 0.7]]$ ,  $d_{2..}^1 = [[0.0, 0.0], [0.0, 0.2]]$ ,  $\kappa = [0, -0.008]$ ,  $\tau = 100$ ,  $s = 0.0002$ ,  $\bar{X} = -4$ .

correctness.

The general procedure to evaluate a bubble model is that once a model seems to have appropriate dynamics and shapes of trajectories, we perform an estimation procedure on synthetic data as well as on real data and see how confident we are in the estimation of the parameters. If we are able to recover the original parameters used to sample the synthetic data, or that the estimation of the parameters for real data is confident, then we consider the model as being plausible. Another important point to consider is that we want the model to be as simple as possible, following the Occam's razor principle.

In order to present the results clearly, we separate the different components of the model: Brownian motion, upward and downward jumps. This allows us to see the contribution of each part of the equation into the trajectory.

### 3.2.7 Analysis of the Type-I Bubble Model

We first study the fully compensated part of the model by setting the parameters of the upward jumps to be null. This amounts to simplify (3.2.1) into the following model:

$$\frac{dS_t}{S_t} = \mu(t) dt + \sigma(t) dW_t - (dZ_t^- - d\Lambda_t^-). \quad (3.2.6)$$

We represent a characteristic trajectory in Fig. 3.4, and see that the rises of the mispricing index  $X_t$  leads to clustering of negative jumps associated with subsequent crashes or drawdowns, producing shapes in the asset price of usual speculative bubbles. As it has been documented in the literature, the dynamics are convincing and we clearly notice localised clustering patterns associated to faster-than-exponential returns

and visible outliers in the negative returns. We also observe an asymmetry in the shape of the bubble pre and post crash. Rather than crashing in a single jump, our model displays multiple jumps around time 1.5 and the synthetic time series exhibit realistically looking drawdowns. This is visible around time 1.25 where the crash is proceeded by a highly volatile phase, which we also see in the values of  $L(X_t)$  on the right plot. Remark that this model displays the counter-intuitive behaviour that bubbles are often associated with smaller relative volatility as mentioned in [187].

Furthermore, the jump size  $\kappa$  is highly (inversely) entangled to the ratio of endogeneity  $\eta$ , and it is known that estimating jointly the jumps size and number of jumps per bin is a difficult problem to solve, with currently no clear solution.

We do not perform any estimation for this model as it is currently worked on by another member of Sornette's research group, and the estimations reveal that this model is convincing and captures most dynamics already [209].

### 3.2.8 Upward Jumps Added to the Model

Compared to the model (3.2.6) presented in the previous subsection, our contribution lies in integrating an opposite sign jump process  $Z_t^+$  into the model (3.2.6), as written in (3.2.1). Due to the symmetric nature of the equations we are working with, we expect the dynamics of  $Z_t^+$  to be the symmetric of  $Z_t^-$ . We start by studying the two extreme scenarios when (1):  $\Omega_t \equiv 1$  and (2):  $\Omega_t \equiv 0$ .

**(1):  $\Omega_t \equiv 1$**  When we compensate the positive jumps' contribution at all time, our model still respects the type-I bubble framework, meaning the overall model is a semimartingale. We sampled a characteristic price trajectory in Fig. 3.5 and as expected, we note the symmetric behaviour of both components of the Hawkes process. The added dimension often negatively interferes with the first dimension, which makes the returns spiky and jumpy with extreme returns both positive and negative, and the curve smooth over many periods of time. Also, we do not observe negative bubble patterns. In order to obtain a negative bubble, we would need the parameters of the upward Hawkes process to grow when the price drops, creating an acceleration in the compensator's value. Here, we have the opposite effect where the dependency on  $L(X_t)$  makes the parameter shrink when the price decreases.

**(2):  $\Omega_t \equiv 0$**  We now present the opposite stand where there are no compensation at all times. We sampled and plot a characteristic trajectory in Fig. 3.6. Since the positive jump part is not compensated, we expect the trajectory to diverge. To quantify this divergence mathematically, we compute the expected return from the positive jumps, called here for simplicity  $r_i$  (daily returns at time  $t_i$ ), and we use the notations from our model (3.2.2).

The Hawkes process has the exogenous rate  $\nu_+$ , branching ratio for positive jumps  $\eta_{+,+}$  and we set the cross excitation to zero:  $\eta_{+,-} = 0$ . The expected returns read for all  $i \in \mathbb{N}_{\geq 0}$ :

$$\mathbb{E}[r_i] = \kappa_+ \mathbb{E}[\lambda_i^+] = \kappa_+ \frac{\nu_+}{1 - \eta_{+,+}}. \quad (3.2.7)$$

Then, by bounding  $L(X_t)$  by 0 on one side and 1 on the other, we get for all  $i \in \mathbb{N}_{\geq 0}$ :

$$\kappa_+ \frac{d_+^1}{1 - d_{+,+}^1} \leq \mathbb{E}[r_i] \leq \kappa_+ \frac{d_+^1 + d_+^2}{1 - d_{+,+}^1 - d_{+,+}^2}. \quad (3.2.8)$$

By excluding the dependency on mispricing in the parameters  $d_+^2 = d_{+,+}^2 = 0$ , the expected return for all  $i \in \mathbb{N}_{\geq 0}$  can be described as  $\mathbb{E}[r_i] = \kappa_+ \frac{d_+^1}{1 - d_{+,+}^1}$ .

The theorised divergence is observed in Fig. 3.6. So, we conclude that phases when  $\Omega_t = 0$  require careful monitoring as prices quickly diverge. The addition of positive jumps to the model pushes prices higher, reducing the importance of negative jumps in the overall drawup phase and strengthening the bubble during drawups. It would be interesting in further research to estimate the coefficient  $\bar{r}$  while allowing for positive jumps, to see if perhaps the returns are negative in certain markets.

The underlying philosophical justification of drifting positive jumps could be an irrational behaviour of markets participants: the investors' unrealistic and unsustainable expectations. It accounts for a collection of different psychological biases, such as overconfidence, unrealistic prospects, over-evaluation of returns, and under-evaluation of risk. In short, there is a rampant psychological bias toward positive news considered to

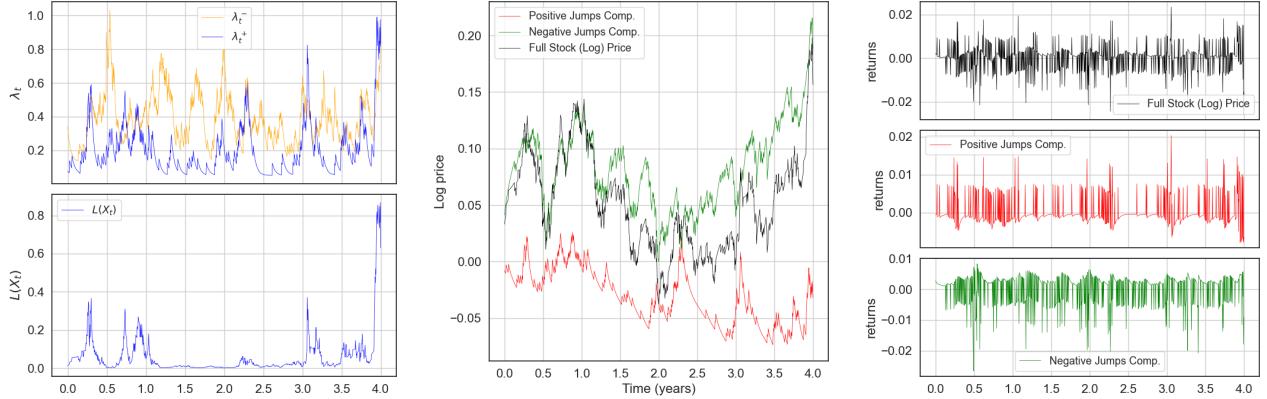


Figure 3.5: Sampled a path of the model (3.2.1) where  $\Omega_t \equiv 1$  and the drift and volatility functions are null. The left plot presents the dynamics of the jump component of our model. We note the correlation between  $\lambda_t^+$  and  $\lambda_t^-$  due to the nature of our model, which makes both components of the model negatively interfering with each other. The middle plot shows the evolution of the log price and we see phases when both dimensions are going in opposite directions: around time 2.25 for example, where the negative jumps drop sharply whereas the positive component increased. This makes the price approximatively constant. Both dimension may also not interfere, just like at time 0.5 when the negative jumps make the price drop without much variation in the positive jump part. On the contrary, at the end of the observation period, both dimensions are increasing at the same time and the price sky-rockets. Finally, the right plots decompose the different components of the returns and the middle and bottom plot suggest the symmetrical nature of the model. The parameters for this trajectory are:  $d^1 = [0.05, 0.1]$ ,  $d^2 = [0.2, 0.5]$ ,  $d_{\cdot}^1 = [[0.7, 0.0], [0.0, 0.7]]$ ,  $d_{\cdot}^2 = [[0.2, 0.0], [0.0, 0.2]]$ ,  $\kappa = [0.008, -0.008]$ ,  $\tau = 100$ ,  $s = 0.0002$ ,  $\bar{X} = -4$ .

be deserved and expected, which makes investors irrational regarding positive news. Hence, it makes sense to add an excessive return due to positive jumps on top of the two other mechanisms: the pure martingale process ( $B_m$ ) and the faster-than-exponential bubble part (the downward jumps). However, we expect the irrationality to be corrected during drawdowns, which is what we discuss in Appendix D.2.

It is clear that neither all-time compensation or no compensation are good enough mechanisms alone, and for this reason we propose the regime process to be a predictable process that changes over time, such that the dynamics of the bubble can be more complex. We first detail the regime process and then show the dynamic of the multi-regime bubble model.

### 3.2.9 Drawdown and Bubble Regimes

So far, our analysis has focused on a constant regime process  $\Omega_t$  for model (3.2.1), resulting in a single regime model. Now, we would like to offer a generalisation including multiple regimes (a regime-switching model). Compared to before, the plots are color-coded with shades from green to red to indicate drawup and drawdown regimes, respectively. We provide here a brief overview of our novel bi-regime model, explain the mathematical framework in Appendix D.1.1, and extend it to a multi-regime framework in Appendix D.1.4.

We define the drawdown (resp. bubble) phase as occurring when the time of the current local maximum is smaller (resp. larger) than the time of the current local minimum. Here, local refers to the segment of the path in some neighbourhood of the current time, specifically the left-side window of length  $t\Delta$ , where  $t$

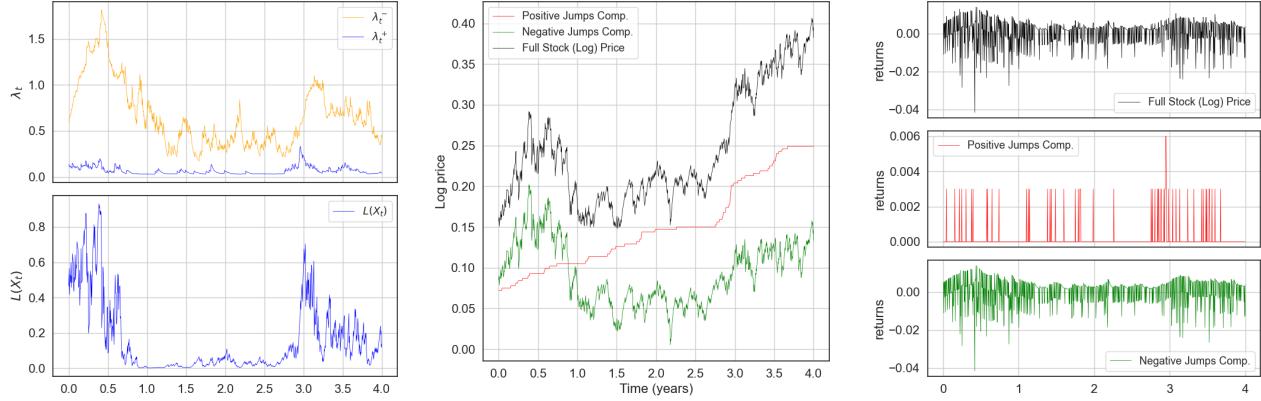


Figure 3.6: Sampled a path of the model (3.2.1) where  $\Omega_t \equiv 0$  and the drift and volatility functions are null. The left plot presents the dynamics of the jump component of our model. We remark as in Fig. 3.5 the correlation between  $\lambda_t^+$  and  $\lambda_t^-$ . The middle plot informs us about the divergence of the positive jumps when we remove the compensation mechanism. We clearly observe strong sporadic increases which drive prices further up, especially during bubbly phases (like right before time 3). The parameters for this trajectory are:  $d^1 = [0.03, 0.1]$ ,  $d^2 = [0.1, 0.5]$ ,  $d_{\cdot, \cdot}^1 = [[0.3, 0.0], [0.0, 0.7]]$ ,  $d_{\cdot, \cdot}^2 = [[0.2, 0.0], [0.0, 0.2]]$ ,  $\kappa = [0.003, -0.008]$ ,  $\tau = 100$ ,  $s = 0.0002$ ,  $\bar{X} = -4$ .

indicates the time scale, with the right edge at the current time. In other words, for all  $t \in \mathbb{R}$ , we have:

$$M(t, t^\Delta) := \max_{u \in [t-t^\Delta, t]} \{S_u\}, \quad (3.2.9)$$

$$\tau_M(t, t^\Delta) := \operatorname{argmax}_{u \in [t-t^\Delta, t]} \{S_u\}, \quad (3.2.10)$$

$$m(t, t^\Delta) := \min_{u \in [t-t^\Delta, t]} \{S_u\}, \quad (3.2.11)$$

$$\tau_m(t, t^\Delta) := \operatorname{argmin}_{u \in [t-t^\Delta, t]} \{S_u\}, \quad (3.2.12)$$

where  $\tau_m$  and  $\tau_M$  are stopping times. Regime modelling based on extrema has been used in previous works, such as [209, 210].

We also require that the market has not undergone a change in the opposite direction greater than a "space" threshold  $s\Delta$  since the last extrema. Stated clearly, if a pair of stopping times are ordered such that we are in a drawdown phase, but the market increases by a value greater than  $s\Delta$  since the last minimum, then the current regime is a drawup. Conversely, if a pair of stopping times are ordered such that we are in a drawup phase but the market decreases by a value greater than  $s\Delta$  since the last maximum, then the current regime is a drawdown. In other words, if we are in a certain regime and the market changes in the opposite direction where the opposite direction is down for drawup, up for drawdown, then we invert the regime.

The mathematical expressions for the two conditions that trigger a forced change of regime are as follows, where the first expression is for the drawup regime and the second is for the drawdown regime:

$$\ln(M(t, t^\Delta)) - \ln(S_t) > s\Delta, \quad (3.2.13)$$

$$\ln(S_t) - \ln(m(t, t^\Delta)) > s\Delta. \quad (3.2.14)$$

In Fig. D.1, we illustrate the mechanism behind the regime-switching.

### 3.2.10 Do Negative Bubbles Appear?

Now that the regime process has been properly introduced in Subsection 3.2.9, we follow these guidelines and add it to the model. We sample a characteristic sampled trajectory in Fig. 3.7. The plots are color-coded with shades of green and red to indicate drawup and drawdown regimes, respectively. The regime

process embodies that investors expect high returns, and the absence of positive news triggers pessimism. This behaviour is logical but overly optimistic during drawup phases making it irrational. The irrationality is expected to be corrected during drawdowns, and we would hope to observe compared to the type-I bubble model bigger crashes, negative bubbles, and fast recovery after the crash.

The Fig. 3.7 illustrates that despite a thorough study of the parameters' space and the possible alternatives to the model, we have not been able to recreate the stylized facts we desired, like negative bubble patterns, a greater variety of crash types, nested bubbles, among others. The added value of a second dimension is the even faster acceleration during the drawup regime, but which does not justify the added complexity to the model. When there is a change of regime, we would have expected to see a drop in the value of the upward jumps process. In spite of the minor decrease, it is of minimal effect compared to what we observe for negative jumps. It is not possible to increase the size of the drop for positive jumps because there is a gap between the price drop, when the mispricing index is at its maximum and so is the intensity  $\lambda_t^+$ , and the moment when there is a regime shift meaning when we start compensating the positive jumps. Furthermore, the decrease is not just due to a perfect timing of the regime but is also a random event, as illustrated in Fig. 3.5. Finally, negative bubbles require a acceleration of the losses (just like we incorporated an acceleration of the returns to create positive bubbles) and no such mechanism is present in our model. Despite the nice philosophical intuition and hope built up on a single dimensional Hawkes process, this model is hence not conclusive.

The idea of multi-regime switching presented in Appendix D.1.4 is also not contributing towards our goal, for the same rationale.

### 3.2.11 Further Improvements to the Model

We have presented the main ideas of the model previously. Unfortunately, it has not been conclusive. In an attempt to address the unresolved stylized facts, we have explored several variations of the bivariate jump diffusion process. Herein, we describe the three most promising ideas that we have considered, and which could potentially improve other models:

- Multi-scale mispricing. In order to create larger build-ups for bubbles that build-up over years, and the bubble crashing in multiple shocks. This involves combining the indices or the logarithm of indices, in order to create more progressive acceleration of prices that better mimics the shape of a hyperbolic law. This approach allows for the use of larger time scales where smaller time scales build momentum. We provide a detailed explanation of this idea in Appendix D.3.2. This idea helped towards our goal of having more nested bubble patterns, and fast rebound after a crash.
- Delayed compensation. We have explored the use of delayed compensation to counteract the divergence of positive jumps described in Subsection 3.2.8. By reintroducing compensation in a delayed manner, we create a drop in price visible in returns that mimics empirical results. This approach also transiently makes the price trajectory a supermartingale, which could be an expected feature of markets. We explain this idea in Appendix D.2 and in particular in Appendix D.2.2. However, this approach is limited by the fact that it makes prices partially predictable, given correct identification of phases.
- Transient overcompensation. Lastly, we propose the use of transient overcompensation by authorizing the regime process to take a value greater than 1. This approach is expected to induce a phase of overcompensation that would help drive prices down. The bubble model alternates between phases of undercompensation (submartingale), efficiency (martingale), and overcompensation (supermartingale), which feels natural. However, this mechanism only generates downward trends and not negative bubbles. We provide further explanation of this idea in Appendix D.2.

Furthermore, we have explored a modification to the model's equations (3.2.1) by substituting  $L(-X_t)$  for  $L(X_t)$ . This approach has yielded insightful observations about the role of  $X_t$ , which, to the best of our knowledge, have not been previously documented in the literature. Specifically, as depicted in Figure 3.8, the substitution of  $L(-X_t)$  for  $L(X_t)$  results in a reversal of the entire dynamic, such that instead of observing faster-than-exponential price increases, we now observe faster-than-exponential crashes and exponential bubbles. This finding is significant because it suggests that rather than attempting to generalize

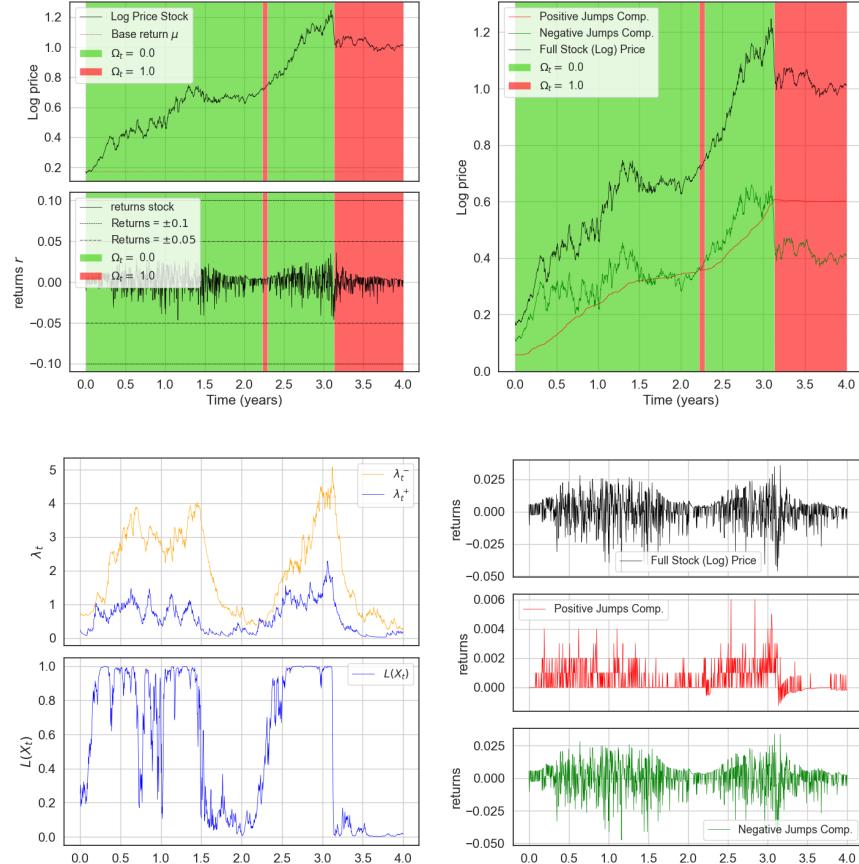


Figure 3.7: Sampled a path of the model (3.2.1) where  $\Omega_t$  follows the dynamics guided by the discussion from Subsection 3.2.9. The dynamics for this figure assume the drift and volatility functions to be null. The top left plot shows the evolution of the log price as well as the returns throughout the study period. Specifically looking at the bubble that builds up after time 2, we see that the compensation phase (starting briefly after 3 years) misses a large part of the crash, preventing the positive jumps' trajectory from shrinking back after the crash and which keeps the price on a diverging trajectory just like when  $\Omega_t \equiv 0$ . This is obvious from the returns where most of the positive jumps returns plotted in red are positive. The parameters for this trajectory are:  $d^1 = [0.03, 0.1], d^2 = [0.1, 0.5], d_{\cdot,\cdot}^1 = [[0.7, 0.0], [0.0, 0.7]], d_{\cdot,\cdot}^2 = [[0.2, 0.0], [0.0, 0.2]], \kappa = [0.001, -0.008], \tau = 100, s = 0.0002, \bar{X} = -4, \Delta = 250$ .

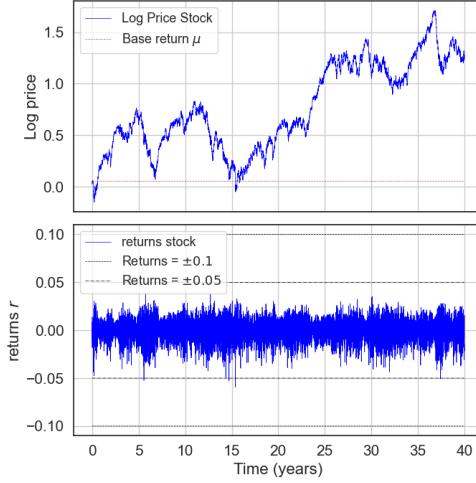


Figure 3.8: Price dynamics and returns for the self-excited model with only one dimension to the Hawkes process, where  $X_t$  has been replaced with  $-X_t$  in (3.2.1) which changes the dynamics since  $L(-X_t)$  increases as the price decreases. The drift and volatility functions are null for this figure. The parameters for the trajectory are:  $d^1 = [0.0, 0.1]$ ,  $d^2 = [0.0, 0.5]$ ,  $d_{\cdot,\cdot}^1 = [[0.0, 0.0], [0.0, 0.7]]$ ,  $d_{\cdot,\cdot}^2 = [[0.0, 0.0], [0.0, 0.2]]$ ,  $\kappa = [0.0, -0.005]$ ,  $\tau = 100$ ,  $s = 0.0004$ ,  $\bar{X} = -4$ . We observe exponential growth phases and faster-than-exponential crashes, in contrast to the original model where exponential crashes and faster-than-exponential growth phases are observed. Here the model is a pure jump process, and there is neither drift or Brownian motion.

all bubbles using the same model, a more fruitful approach may involve developing a collection of distinct bubble models, each tailored to the specific underlying mechanisms driving the bubble in question.

It should be noted, however, that the use of  $L(-X_t)$  in the model (3.2.1) is not without its limitations. While this modification successfully captures certain stylized facts, such as the rapid rebound following the crash, faster-than-exponential decrease, and the characteristic calm period preceding the crash, it fails to account for others, such as the peak of returns occurring at the local minimum of the crash. Furthermore, the effectiveness of this model must be verified through further research and engineering, such as fitting the model to real-world data, which is typically the next step in the modeling of bubbles.

Also, quickly summarised, it is difficult to include in the same model both  $L(X_t)$  and  $L(-X_t)$  because it makes the model highly volatile, oscillating rapidly between the two types of bubbles.

### 3.3 Real Experiment

#### 3.3.1 Mimicking Real Data

The natural next step in our analysis is fitting our model on realised time series. We have access to some traditional markets (CAC40, DAX, DJIA, EuroStoxx600, HSI, Nasdaq, Nikkei, SP500), and we would like to see what sets of parameters can reproduce the observed bubbly behaviour. In the upper plot of Fig. 3.9, we present the realised price for specific markets and their log price. In the bottom plot, we show the corresponding returns.

Nonetheless, our analysis revealed that our model has not fulfilled the promises we were hopping for it. Despite our efforts, it seems that simpler models (and in particular [209]'s in development model) are as powerful with less parameters. An alternative to their model would be to substitute in their model  $L(-X_t)$  for  $L(X_t)$ , which could give a different class of bubbles, as it has been discussed in the previous section. For this reason, we do not estimate the parameters of our model off real time series and we leave for further research the study of the model including  $L(-X_t)$ .

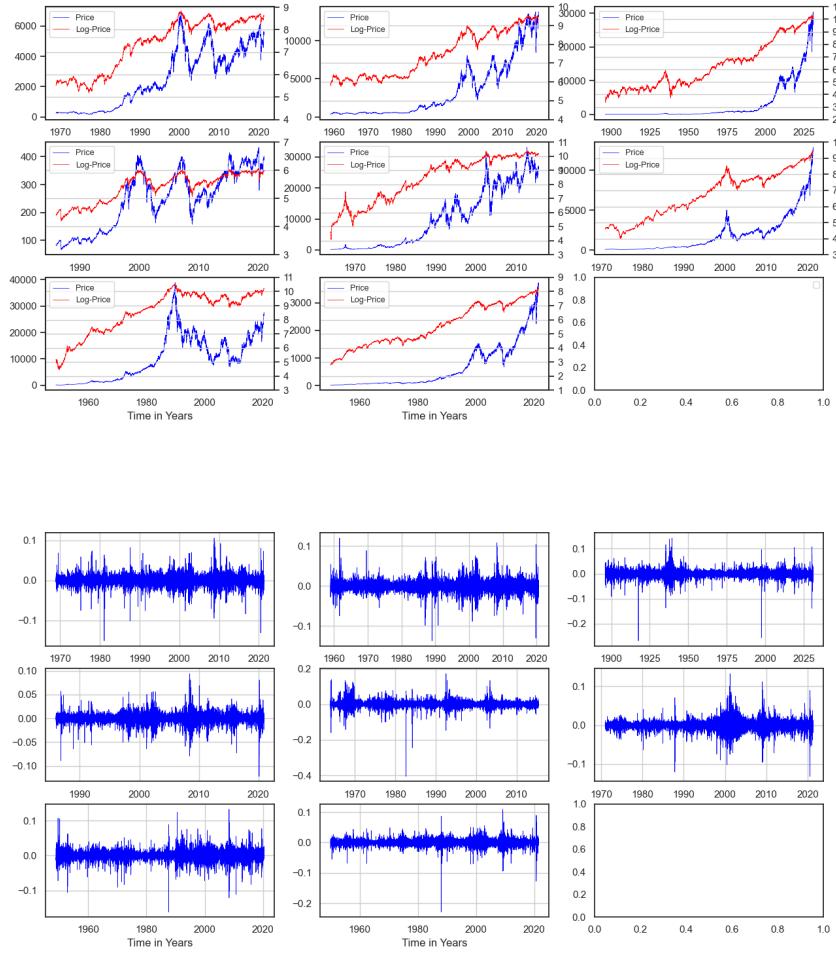


Figure 3.9: We present the log price evolution of the closing mid price of some markets and the increments of the log prices. In order, left to right, top to bottom, the markets are the CAC40, DAX, DJIA, EuroStoxx600, HSI, Nasdaq, Nikkei, SP500, where the last point of the time series corresponds to the last trading day of December 2020. We notice two groups of stocks: CAC40, DAX, EuroStoxx600 and HSI, Nasdaq, Nikkei, and SP500. The data associated with the DJIA go from 1896 to 2020 but because there are more than 250 data entries per year, the plot shows a final entry in 2027. By observing the log price, we see the characteristic bubbly shapes, with the stylized facts we have documents. The returns show strong outliers both for positive and negative returns that are related to peak of bubbles and crashes.

### 3.3.2 Outcome of the Comparison of the Models

By combining the different components of the model that we discussed, we end up with our new model: a Hawkes Jump-Diffusion Process with Non-Local Multi-Scale Mispicing Index and Multi-Scale Regime Switching. However, we observed that adding a dimension to the Hawkes process did not fulfil its promise of capturing a broader range of stylized facts, despite strong behavioural underpinnings. Although the model's positive jump component allows for negative fundamental returns, the market still experiences excessively endogenous periods that are too frequent and long, resulting in specific funnel-shaped returns. While the regime process has limited capabilities to contribute to negative bubbles, we believe it holds potential and could be used in other models.

The multi-scale regime-switching mechanism proposed in this study offers a promising alternative for financial markets. This approach captures the diverse perspectives and sensitivities of market participants by incorporating a hierarchy of time and space scales, leading to more accurate and robust identification of drawdown phases. There is still a challenge related to phase identification in regimes. In our regime definition, we use a space threshold  ${}^s\Delta$  based on the assumption that drift is negligible. For instance, for longer time scales such as six years, we could expect 42% returns (7% per year) that would exceed the threshold set at 38% with our definition. The solution would be to detrend the time series before estimating any parameters. Generally, this raises financial engineering questions, such as whether to use a local or global estimator for regimes. Regime identification adds complexity to the model, making parameter estimation difficult.

We are confident that the multi-scale mispricing index is a powerful alternative to the mispricing index. As a conclusion, we present a sample trajectory of our model in Fig. 3.10, using three regimes ( $\Omega_t \in \{0, 0.5, 1\}$ ) and comparing the mispricing index, including three sub-indices, to the mispricing computed with only one index. We also add a Brownian motion to the dynamics so that the curve can be interpreted as a realistic trajectory for a price process. The figure illustrates two stylized facts that this mechanism captures better, the fast rebound and the nested bubbles patterns.

Perhaps an alternative lies in the model where we use  $L(-X_t)$  instead of  $L(X_t)$  as the acceleration factor for the parameters, creating positive feedback the further the price goes lower. The roles of upward and negative jumps are inverted, meaning that we always compensate for the upward jumps, and negative jumps are only compensated during drawups. The model shows slower bubble growth but faster crashes. Exactly like before, the downward jump process seems to have a limited impact on the dynamics of the price and does not give rise to a negative bubble, however we work out the parameters and characterisations of the model. More generally, as we said in our introduction, despite the universality of bubbles, they come in different sizes and shapes and it would make sense to use different models depending on the characteristics of the bubble and market. By changing the model depending on the type of bubble, we would be able to account for more bubble types.

We saw that real data presented in Figure 3.9, display a greater level of explosiveness for the returns than what our model is currently able to produce, even when the model parameters are set closer to criticality. This inability to accurately capture such explosiveness with our linear Hawkes process may potentially be addressed through the adoption of a nonlinear Hawkes process [28, 218], as previously postulated by Wehrli [205] as a possible solution to the modelling of fast-accelerating intensities. At shorter time scales, the exogenous intensity spikes in a sharper way than the branching ratio during crises, which the Hawkes( $p, q$ ) model (a univariate, linear Hawkes process) is unable to capture. This lack of fit also extends to the inability of linear Hawkes processes to adequately account for certain empirical stylized facts, including fat-tails and endogenous instabilities such as the Zumbach effect. As such, an extension to nonlinear processes such as those described in [72] may offer a potential solution to these limitations. It is worth noting that while these nonlinear processes may be less relevant in coarser grid models, this nevertheless suggests that the jump part of the model may have been an incorrect choice, regardless of the scale. Furthermore, it is not necessarily clear why positive and negative news should be treated as two separate processes given that information is often entangled and does not follow a linear relationship. Nonlinear Hawkes processes offer a different representation than linear Hawkes processes, which can lead to increased complexity in simulation and estimation, rendering the use of the EM method impractical. Given these limitations, we recommend the use of the published model from [131], or the ongoing work presented in the working paper [209], which represents a univariate simplification of our model and can be tested with the proposed improvements, such as the multi-index mispricing.

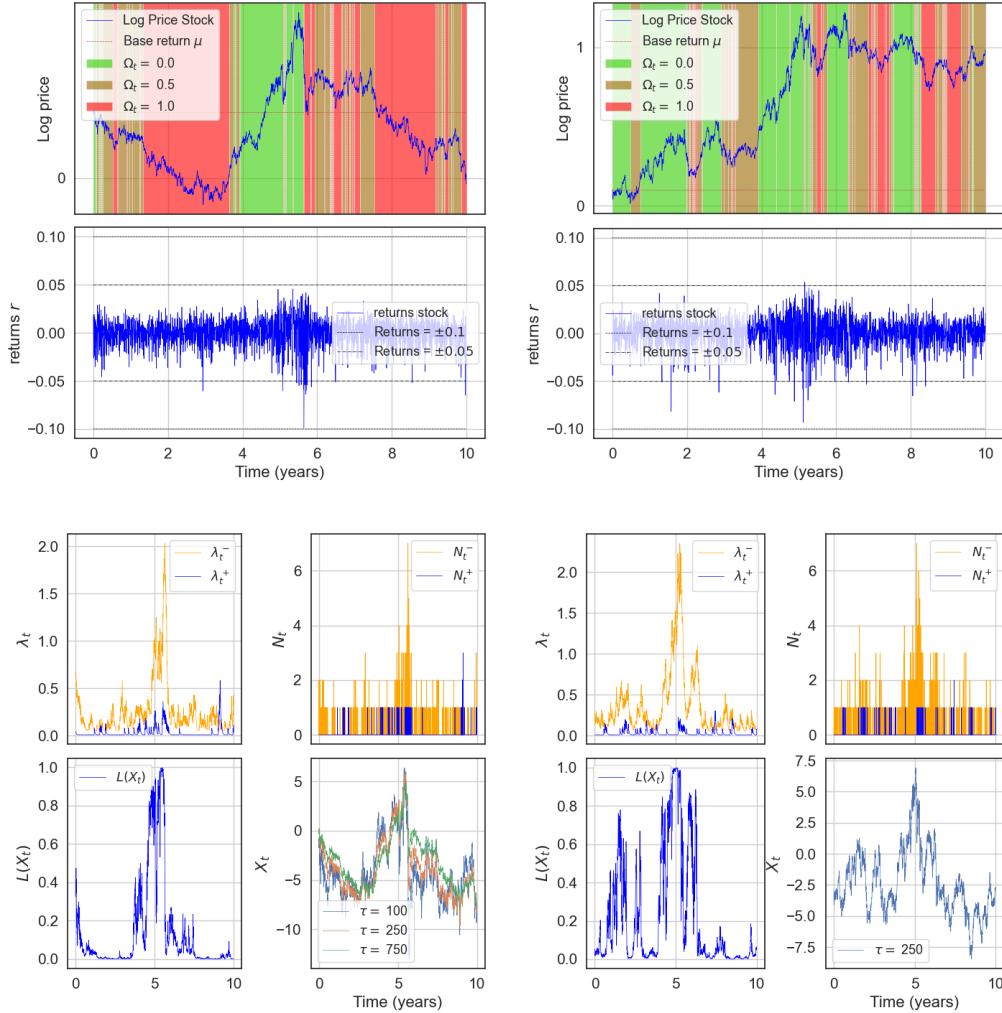


Figure 3.10: We sample a trajectory from the model (3.2.1) where we compare the model with and without a multi-scale mispricing index. On the left, the specific parameters are  $\tau = [100, 250, 750]$ ,  $s = [0.0003, 0.0002, 0.0001]$ , whereas, on the right, they are  $\tau = [250]$ ,  $s = [0.0002]$ . The common parameters are  $d^1 = [0.01, 0.05]$ ,  $d^2 = [0.001, 0.2]$ ,  $d_{\cdot,\cdot}^1 = [[0.7, 0.0], [0.0, 0.6]]$ ,  $d_{\cdot,\cdot}^2 = [[0.1, 0.0], [0.0, 0.35]]$ ,  $\kappa = [0.005, -0.02]$ ,  $\bar{X} = -4$ ,  $\Delta = [250, 750]$ ,  $\sigma = 0.15$  (per annum). We observe a hierarchy of bubbles on the left figures that we do not observe on the right.  $X_t$  is less volatile on the left and take more time to reach high values, which prevents bubbles from appearing and disappearing constantly. It also prevents from  $L(X_t)$  to reach 1 which is desirable, as  $L(X_t)$  represents an extreme case. Instead, bubbles take longer to develop (compare the period 3 - 5 years on both plots) and longer to crash (compare the period 5-8 years). In short, the stylized facts of bubbles seems more accurately captured by the multi-index mispricing model, including nested bubbles as well as fast rebound after a crash.



# OUTLINE OF THE WORK OF THIS THESIS

In this thesis, we have studied the INVAR process and we have proven that the properties true for univariate INAR( $\infty$ ) processes naturally generalise to multiple dimensions. For example, both INAR and multivariate INAR processes benefit from an ARMA representation and from the convergence in law to univariate or multivariate Hawkes processes. INVAR processes have a plethora of application aside from financial bubbles, and on which we are currently working already as follow-up works. The approximation is useful in many scenarios for numerical and practical reasons, and we use it to discretise our new type-II bubble model: a Hawkes Jump-Diffusion Process with Non-Local Multi-Scale Mispricing Index and Multi-Scale Regime Switching. Our study represents a pioneering contribution to the field of type-II bubbles. These models exhibit greater flexibility in their dynamics than type-I bubble but are also more sensitive and prone to rapid explosion due to the lack of constraints imposed on their behaviour during the drawup regime. Paradoxically, this challenge also presents an opportunity for success, as it could help capture some stylized facts not displayed by previous models and, consequently, may inspire a fundamental shift in the way bubbles are modelled.

Our model shows strong features but we doubt that the added layers of complexity are worth the improvements. In particular, the regime process has been specifically one of the challenges we faced. The research about interesting type-II bubbles is still ongoing and it seems that a regime process should be at the center of it, balancing phases of sub and supermartingality. Our research also suggests that the bivariate Hawkes process is not the way to proceed. Instead, we recommend using our novel ideas to improve already existing models, such as the in progress self-exciting type-I bubble that is soon to be published.

Further research could be conducted to explore these new ideas in conjunction with financial engineering to test their robustness. Some of the ideas include the regime process, the multi-scale mispricing index, the overcompensation phase, or the proposed reversed model using  $L(-X_t)$ . We are particularly hopeful that the multi-scale mispricing index will be successful in increasing the expressiveness of bubble models. Mathematical characterization of inherent patterns in bubbles such as volatility, conditional laws, and returns could be a promising path for bubble research, aligned with [189]. The paper suggests that endogenous crashes present a different volatility pattern from exogenous crashes, and identifying the appropriate model that captures these patterns could be of significant value.

A final disruptive view about bubble modelling is that linear Hawkes processes are unable to account for some empirical stylized facts that exponential Hawkes process could be able to reproduce.

Factually, quadratic Hawkes processes capture some interesting microstructural stylized facts and present more explosiveness in their dynamics. Note that quadratic Hawkes processes are a second-order approxima-

tion of the exponential Hawkes process, with conditional intensity:

$$\lambda(t) = \phi_0(t) + \int_0^t \phi_1(t-s) dN_s + \int_0^t \int_0^t \phi_2(t-s, t-u) dN_s dN_u.$$

For this reason, this suggests that a promising candidate for modelling public markets at the microstructural level would be the exponential Hawkes process whose conditional intensity would have this form:

$$\lambda(t) = \mu(t) \exp \left( \int_0^t \phi(t-s) dN_s \right).$$

Such models may also be applicable to financial bubble modelling, but pose several challenges that require further investigation. One significant difficulty is that exponential Hawkes processes lose the branching structure of linear Hawkes processes, making their estimation even more challenging.

## APPENDIX

# A TEMPORAL POINT PROCESSES (TPP)

### A.1 Point Process Theory

For a more complete description of point process theory, we recommend the book from Daley and Vere-Jones [50]. Hereinafter, we give a small insight on some definitions that are at the core of the idea of Hawkes processes.

**Definition A.1.1** (Conditional Intensity Function). When  $N$  is a point process with natural filtration  $\mathcal{F}$ , we call the left-continuous and adapted process, "stochastic **conditional intensity function** of the point process", defined as:

$$\lambda(t \mid \mathcal{F}_{t^-}) = \lim_{h \rightarrow 0^+} \frac{\mathbb{P}(N_{t+h} - N_t > 0 \mid \mathcal{F}_{t^-})}{h} = \frac{\mathbb{P}(dN_t > 0 \mid \mathcal{F}_{t^-})}{dt}, \quad (\text{A.1.1})$$

which for simple point processes can be rewritten in terms of a conditional expectation:

$$\lambda(t \mid \mathcal{F}_{t^-}) = \lim_{h \rightarrow 0^+} \frac{\mathbb{E}(N_{t+h} - N_t \mid \mathcal{F}_{t^-})}{h} = \frac{\mathbb{E}(dN_t \mid \mathcal{F}_{t^-})}{dt}. \quad (\text{A.1.2})$$

The latter expression is more intuitive because one might think from the first expression that the intensity is a probability. Hereinafter, the conditioning is implied in the function  $\lambda$  for conciseness.

**Definition A.1.2** (Inhomogeneous Hawkes process). Let  $N(t) = (N_t^1, \dots, N_t^d)$  be a simple multivariate counting process, with natural associated filtration  $\{\mathcal{F}_t\}_{t \in \mathbb{R}_+}$ , whose conditional intensity function satisfies for all  $m \in \{1, \dots, d\}$  as a left-continuous adapted stochastic process  $\lambda^m$  given by the Stieltjes integral:

$$\lambda^m(t) = \nu^m(t) + \sum_{n=1}^d \int_0^t \phi_s^{m,n}(t-s) dN_s^n, \quad t \in \mathbb{R}_+. \quad (\text{A.1.3})$$

Then, we call the point process  $N$  a **Hawkes process**.

Note that  $\forall t \in \mathbb{R}$ ,  $\nu(t) \in \mathbb{R}^d$ ,  $\Phi(t) \in \mathbb{R}^{d \times d}$ , and we impose that  $\phi$  is a kernel function called Hawkes' excitation function, which we require to be causal ( $\phi(t) = 0$ ,  $\forall t < 0$ ), and  $\nu \geq 0$ . It is common to assume that  $\Phi$  is piecewise continuous, and additionally, to this, we require that the time dependency of  $\nu$  and  $\Phi$  to be piecewise continuous as well. When we write  $\phi_s$ , we mean that the kernel may change shape depending on when an event takes place.

We represent alternatively the conditional intensity in vectorial representation, where we write vectors with the same notation as for scalars but there should be no confusion from the context:

$$\lambda(t) = \nu(t) + \int_0^t \Phi(t-s) dN_s, \quad t \in \mathbb{R}_+, \quad (\text{A.1.4})$$

where  $\Phi$  is the matrix composed of the kernel functions and the product  $\Phi(t-s) dN_s$  is a matrix-vector product. It will be convenient to write the above equation with the integration operator  $*$ :

$$\lambda(t) = \nu(t) + \Phi(t-\cdot) * N., \quad t \in \mathbb{R}_+. \quad (\text{A.1.5})$$

**Definition A.1.3** (Branching Ratio). In the univariate case, the integral of a homogeneous kernel function is called the **branching ratio**, and plays a crucial role for the dynamics of the process:

$$\eta := \int_{\mathbb{R}} \phi(t) dt. \quad (\text{A.1.6})$$

In the multivariate case, we integrate term by term the kernel  $\Phi$  and this gives a branching ratio matrix  $\eta \in \mathbb{R}^{d \times d}$ . We rewrite from (A.1.4) the term  $\Phi$  as an untangled product of a memory shape  $h$  (normalised kernel), a function from  $\mathbb{R}$  to  $\mathbb{R}^d$  and of a branching ratio  $\eta$ , also a function from  $\mathbb{R}$  to  $\mathbb{R}^d$ :

$$\Phi(t-\cdot) * N. = (\eta \odot h)(t-\cdot) * N., \quad t \in \mathbb{R}_+, \quad (\text{A.1.7})$$

where  $\odot$  represents the Hadamard product / element-wise product (see Definition A.4.3). While  $h^{m,n}$  represents the density of the offspring arrival times generated by the process  $n$  for the process  $m$ ,  $\eta^{m,n}$  represents the average number of children generated by the process  $n$  for the process  $m$ . Finally, it is well known that the spectral radius of the matrix  $\eta$  plays the role of the univariate branching ratio [93]. See the definition of spectral radius in Definition 1.3.1.

As above, using the branching ratios, we can untangle the shape of the kernel with its branching ratio with a normalised kernel  $h$  for the case of Hawkes processes. From (A.1.3):

$$\lambda^m(t | \mathcal{F}_{t-}) = \nu^m(t) + \sum_{n=1}^d \sum_{\{k: t_k^n < t\}} \eta^{m,n}(t_k^n) \cdot h^{m,n}(t - t_k^n).$$

An important theorem about the Hawkes process is related to its existence and stationarity. Due to its structure, non-stationary Hawkes processes explode (in the sense that they diverge) extremely quickly. The usual sufficient condition for existence, well-definedness and stability (stationarity) is that the branching ratio is strictly smaller than one. In the multivariate case, it is also well known that a sufficient condition for stability is the spectral radius (Definition 1.3.1) of the branching ratio is strictly smaller than 1 (see [28, 52, 126, 127, 202]).

We summarise some important properties in the next statement:

**Theorem A.1.4** (Properties of the Hawkes Process). *Let us assume we have a multivariate Hawkes process, as defined previously with exogenous rates  $\nu \geq 0$  and positive and causal kernel  $\Phi$ . Also we define the spectral radius as in Definition 1.3.1.*

#### Stationarity of a Hawkes Process.

*A sufficient condition for existence and stationarity of a Hawkes process is that the branching ratio matrix has a spectral radius strictly smaller than 1:*

$$\rho \left( \int_{\mathbb{R}^+} \Phi \right) < 1.$$

#### Characterisation of a Hawkes Process.

*A Hawkes process can also be uniquely characterised by the following equation as long as it is stationary:*

$$\mathbb{E} [N_{[a,b]} | \mathcal{F}_a] = \mathbb{E} \left[ \int_a^b \lambda(s) ds | \mathcal{F}_a \right], \quad \forall a < b \in \mathbb{R}. \quad (\text{A.1.8})$$

*There exists a unique stationary process that satisfies (A.1.8).*

*Proof.* **Stationarity of a Hawkes Process.**

This is a well known fact. A proof can be found in [28, 52, 126, 127, 202].

**Characterisation of a Hawkes Process.**

The interested reader can read about this proposition in [94, 117].  $\square$

**Definition A.1.5** (Compensator). Assuming an  $\mathcal{F}_t$ -adapted counting process  $\{N_t\}_{t \in \mathbb{R}_+}$  with left-continuous conditional intensity  $\lambda_t$ , we define the **compensator** as:

$$\Lambda(t) = \int_0^t \lambda(u) du, \quad (\text{A.1.9})$$

whereby this definition guarantees the martingale property,  $\forall t > s > 0$ :

$$\mathbb{E}[N_t - \Lambda(t) | \mathcal{F}_s] = N_s - \Lambda(s). \quad (\text{A.1.10})$$

Very naturally, we extend this definition to jump-diffusion processes (Definition A.2.2), where the jumps' distribution has to be taken into account as well to preserve the martingale property.

## A.2 Lévy Process and Jump-Diffusion Model

We provide a formal definition for Lévy processes of jump-diffusion. For a more in-depth introduction to the topic, we refer the interested reader to [96, 196].

**Definition A.2.1** (Compound Poisson Process). Let  $\{Z_i\}_{i \in \mathbb{N}_{\geq 0}}$  be a sequence of independent and identically distributed random variables with a square integrable distribution, and let  $\{N_t\}_{t \in \mathbb{R}_+}$  be a Poisson process. Then, the compound Poisson process is defined by

$$Y_t = \sum_{k=1}^{N_t} Z_k. \quad (\text{A.2.1})$$

More generally, we define Lévy processes of jump-diffusion as follows.

**Definition A.2.2** (Lévy Process Jump-Diffusion). Let  $\{W_t\}_{t \in \mathbb{R}_+}$  be a Brownian motion on a given probability space, let  $\{N_t\}_{t \in \mathbb{R}_+}$  be a counting process, and let  $\{Z_i\}_{i \in \mathbb{N}_{\geq 0}}$  be a sequence of independent and identically distributed random variables with a square-integrable distribution. Finally, let  $\mu$  and  $\sigma$  be two adaptable processes.

Then, the following process is a **jump-diffusion process**:

$$Y_t = \mu t + \sigma W_t + \sum_{k=1}^{N_t} Z_k. \quad (\text{A.2.2})$$

Equivalently, we will write in differential form

$$dY_t = \mu dt + \sigma dW_t + Z_{N_t} dN_t. \quad (\text{A.2.3})$$

The part with the counting process is called the jump process.

One of the most famous jump-diffusion models is the Merton model from [142], wherein the jump-diffusion model  $\{Z_i\}_{i \in \mathbb{N}_{\geq 0}}$  are i.i.d. and taken as Gaussian. We can also consider a model where  $\mu, \sigma, Z_i$  are all multiplied by  $\tilde{Y}_t$ , in which case we would have characterised the logarithm of a jump-diffusion process. This class of model is called log Lévy models. See [47] for a list of almost 400 references on the subject.

## A.3 Matrices, Eigenvalues, Matrix Valued Polynomials and Norms

### A.4 Matrix Analysis

We will say for  $A, B \in \mathbb{M}_d$ , that  $A \leq B$  if the inequality holds component-wise (and we similarly define the symbols  $<$ , and  $\geq, >$ ).

We will also talk about nonnegative matrices (in the sense that each component is greater or equal to 0) and positive matrices (each component is strictly greater than 0).

**Theorem A.4.1** (Perron-Frobenius Theorem, Theorem 8.3.1 [97]). *Let a nonnegative matrix  $A \geq 0 \in \mathbb{M}_d$ , then:*

$$\rho(A) \text{ is an eigenvalue of } A. \quad (\text{A.4.1})$$

We now continue by detailing some properties of the norms of matrices:

**Theorem A.4.2** (Norms of Matrices). *Let  $A \in \mathbb{M}_d$ . We define a class of matrix norms induced by the  $q$ -norms for vectors, for  $q \in [1; \infty]$ . They are defined:*

$$\|A\|_q = \sup_{x \neq 0} \frac{\|Ax\|_q}{\|x\|_q}, \quad (\text{A.4.2})$$

and reduce to simpler estimates in the following cases:

- $\|A\|_1 = \max_{j \leq n} (\sum_{i=1}^n |a_{i,j}|),$
- $\|A\|_\infty = \max_{i \leq n} (\sum_{j=1}^n |a_{i,j}|).$

Finally, we talk about a matrix norm  $\|\cdot\|$  consistent with a vector norm  $|\cdot|$  if for all matrix  $A$  and vector  $x$  with matching size, it holds:

$$|Ax| \leq \|A\| |x|. \quad (\text{A.4.3})$$

*Proof.* [86] proved these facts in 5.6.3. and 5.6.7.  $\square$

**Definition A.4.3** (Hadamard Product). We define the binary operator **Hadamard Product**  $\odot$  as the term-wise multiplication operator.

We define it between two vectors, for any coordinate  $i$ :

$$(x \odot y)_i := (x^i \cdot y^i)_i, \quad (\text{A.4.4})$$

or between a vector and a matrix, for any pair of coordinates  $i, j$ :

$$(x \odot A)_{i,j} := (x^i A_{i,j})_{i,j}, \quad (\text{A.4.5})$$

$$(A \odot x^T)_{i,j} := (A_{i,j} x^j)_{i,j}. \quad (\text{A.4.6})$$

We will also need the notation Diag and some of its properties:

**Definition A.4.4** (Diagonal). We call the **diagonal** operator Diag which takes a vector and returns a matrix, where the entries on the diagonal correspond to the entries of the vectors.

Directly, we have the following properties:

- Diag is a linear operator:  $\text{Diag}(\sum \lambda x_i) = \sum \lambda \text{Diag}(x_i)$ ,
- For a matrix  $A$  and a random vector  $X$ ,  $\mathbb{E}[\text{Diag}(AX)] = \text{Diag}(A\mathbb{E}[X])$ ,
- We have for any matrix norm  $\|\cdot\|$  consistent with a vector norm  $|\cdot|$ , for two vectors  $x, y \in \mathbb{R}^n$ :  $\|\text{Diag}(x \odot y)\| = \|\text{Diag}(x) \odot y\| \leq \|\text{Diag}(x)\| |y|$ .

**Definition A.4.5** (Convergence of Matrices). We say that a **sequence of matrices converges** if and only if every of its components converge: for a sequence of matrices  $\{A_n\}_{n \in \mathbb{N}_{\geq 0}}$ :  $\lim_{n \rightarrow \infty} A_n = A \iff \forall i, j \in \{1, \dots, d\}$ ,  $\lim_{n \rightarrow \infty} (A_n)_{i,j} = A_{i,j}$ . The same applies to an infinite sum of matrices.

## A.5 Polynomial Eigenvalue Problem

Let  $p \in \mathbb{N}_{>0}$ . The characteristic polynomial of a matrix  $A \in \mathbb{M}_p$  is defined as  $p_A(z) = \det(A - \text{Id } z)$ ,  $z \in \mathbb{C}$  and the eigenvalues of  $A$  are the roots of  $p_A$ . Finding eigenvalues comes down to searching for the roots, for some coefficients  $\alpha_k \in \mathbb{R}, \forall k \in \mathbb{N}_{>0}$  that depend on the matrix  $A$ :

$$z^p = \sum_{k=1}^p \alpha_k z^{p-k}, z \in \mathbb{C}.$$

The reciprocal monic polynomial of  $p_A$  (reciprocal characteristic equation) is defined as  $p_{A,r}(z) = \alpha_1^{-1} z^p \cdot p(1/z), z \in \mathbb{C}$ :

$$\alpha_1^{-1} = \alpha_1^{-1} \sum_{k=1}^p \alpha_k z^k, z \in \mathbb{C}.$$

Its roots are the inverse of the roots of  $p_A$ .

The matrix:

$$C(p_A) = \begin{pmatrix} 0 & 0 & \cdots & 0 & \alpha_1 \\ 1 & 0 & \cdots & 0 & \alpha_2 \\ 0 & 1 & \cdots & 0 & \alpha_3 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & \alpha_p \end{pmatrix}, \quad (\text{A.5.1})$$

is referred to as the companion matrix (see Definition 3.3.13 in [97]). The zeros of the polynomial  $p_A$  are also the eigenvalues of the companion matrix  $C(p_A)$  [140].

The following matrix is the inverse of the companion matrix of  $p_A$ :

$$C(p_A)^{-1} = \alpha_1^{-1} \begin{pmatrix} \alpha_2 & 1 & 0 & \cdots & 0 \\ \alpha_3 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha_p & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \end{pmatrix}. \quad (\text{A.5.2})$$

Its characteristic polynomial is the reciprocal monic polynomial of  $p_A$  (see [97] Problem 3.3.P25).

Likewise, for monic matrix polynomials  $P(z) = \text{Id } z^p - \sum_{k=1}^p A_k z^{p-k}$ ,  $A_k \in \mathbb{M}_d, \forall k \in \mathbb{N}_{>0}$ , we have its associated (block) companion matrix:

$$C(P) = \begin{pmatrix} 0 & 0 & \cdots & 0 & A_1 \\ \text{Id} & 0 & \cdots & 0 & A_2 \\ 0 & \text{Id} & \cdots & 0 & A_3 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \text{Id} & A_p \end{pmatrix}. \quad (\text{A.5.3})$$

A scalar  $z$  is called an eigenvalue of  $P$  if and only if there exists a vector  $v$  such that  $P(z)v = 0$ . One can check for the existence of such a vector by computing the determinant:  $\exists z \in \mathbb{R}, v \in \mathbb{R}^d, P(z)v = 0 \iff \exists z \in \mathbb{R} : \det(P(z)) = 0$ . Eigenvalues of  $P$  are the eigenvalues of  $C(P)$ . Finally, as before we can define the reciprocal monic matrix polynomials and inverse in a similar fashion.

The polynomial eigenvalue problem, i.e. computing and locating the eigenvalues of matrix polynomials, is central in scientific computing.

**Definition A.5.1** (Polynomial Eigenvalue Problem). The polynomial eigenvalue problem relates to finding a nonzero eigenvector  $v$ , corresponding to an eigenvalue  $z$  satisfying for  $P(z) = \text{Id} - \sum_{k=1}^p A_k z^k$ ,  $A_k \in \mathbb{M}_d, k \in \mathbb{N}_{>0}$  the equation  $P(z)v = 0$ .

We assume that  $\det(P(z))$  is not identically zero (which would happen if  $A_0$  was singular for example).

## A.6 Lemmas for Spectral Radius Orderings

**Lemma A.6.1** (Lemma 1). *Let us define the two sequences of nonnegative matrices  $\forall k \in \mathbb{N}_{>0}, A_k, B_k \geq 0 \in \mathbb{M}_d$ . We fix the condition that there exist two indices  $i < j$  in that sequence such that  $B_i \leq A_i$  and  $A_j \leq B_j$ , such that  $A_i + A_j \leq B_i + B_j$ , whereas for all other indices  $k \neq i, j \in \mathbb{N}_{>0}$ ,  $A_k = B_k$ .*

*Naturally, we define the companion matrices (written here in total generality for some sequence of matrices  $C$ )*

$$A(C_1, C_2, \dots, C_p) = \begin{pmatrix} C_1 & C_2 & \cdots & C_{p-1} & C_p \\ \text{Id} & 0 & \cdots & 0 & 0 \\ 0 & \text{Id} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \text{Id} & 0 \end{pmatrix}. \quad (\text{A.6.1})$$

By assuming that  $\rho(\sum_{k=1}^p B_k) < 1$ , we have the following inequality:

$$\rho(A(A_1, A_2, \dots, A_p)) \leq \rho(A(B_1, B_2, \dots, B_p)). \quad (\text{A.6.2})$$

The equality case happens only when  $B_i = A_i, B_j = A_j$ . Similar results can be found when we assume instead  $\det(\sum_{k=1}^p B_k) > 1$ .

*Proof.* In order to prove this theorem, we work with the associated (reciprocal) monic matrix characteristic polynomials of each companion matrix. Keeping the notation from the statement, we write  $P_A(z) = \text{Id} - \sum_{k=1}^p A_k z^k, P_B(z) = \text{Id} - \sum_{k=1}^p B_k z^k$ .

We also coin respectively  $z_A^*, z_B^*$  the smallest eigenvalue of  $P_A, P_B$  (which equal to the inverse of the largest eigenvalue of their respective companion matrix). Our goal is to prove  $|z_B^*| \leq |z_A^*|$ . By Perron-Frobenius Theorem A.4.1 for nonnegative matrices, we know  $z_A^*, z_B^* \in \mathbb{R}$ .

Now, note that because  $\rho(\sum_{k=1}^p B_k) < 1$  by assumption, we have  $\det(P_A(z)), \det(P_B(z)) > 0$  for  $z \in [0, 1]^1$ . This indicates that the polynomial  $P_A$  (resp.  $P_B$ ) is strictly positive on the half-line  $[0, z_A^*]$  (resp.  $[0, z_B^*]$ ). In order to show that  $z_B^* < z_A^*$ , we show that  $\det(P_A(z))$  is strictly positive on  $[0, z_B^*]$ .

We state this known lemma: a matrix  $M \in \mathbb{M}_d, \forall x \in \mathbb{R}^+ : x^T M x > 0 \implies \det(M) > 0$ . We know that over  $[0, z_B^*]$ ,  $P_B$  is strictly positive. This shows directly  $\forall z \in [0, z_B^*]$ :

$$0 < x^T (\text{Id} - \sum_{k=1}^p B_k z^k) x < x^T (\text{Id} - \sum_{k=1}^p A_k z^k) x, \quad (\text{A.6.3})$$

using the assumption that  $\sum_{k=1}^p A_k \leq \sum_{k=1}^p B_k$ , which concludes the proof.  $\square$

**Lemma A.6.2** (Lemma 2). *Keeping the notation from Lemma A.6.1, if we have  $\rho(\sum_{k=1}^p A_k) < 1$ :*

$$\rho(\sum_{k=1}^p A_k) = \rho(A(\sum_{k=1}^p A_k, 0, \dots, 0)) < 1 \iff \rho(A(A_1, \dots, A_p)) < 1. \quad (\text{A.6.4})$$

The inequality A.6.4 would be in the other direction when  $\rho(\sum_{k=1}^p A_k) > 1$ .

*Proof.*

$$\rho(\sum_{k=1}^p A_k) = \rho(A(\sum_{k=1}^p A_k, 0, \dots, 0)),$$

since  $A(\sum_{k=1}^p A_k, 0, \dots, 0)$  is a lower triangular matrix whose eigenvalues are the ones from  $\sum_{k=1}^p A_k$  and all the others are 0.

The only thing to show is that

$$\rho(A(\sum_{k=1}^p A_k, 0, \dots, 0)) < 1 \iff \rho(A(0, \dots, 0, \sum_{k=1}^p A_k)) < 1. \quad (\text{A.6.5})$$

---

<sup>1</sup> And using monotonicity of  $\rho$  on the ordering of matrices, as well as the complex Jordan's decomposition.

This equivalence is direct by looking at the characteristic polynomials  $\xi_1(z) = \text{Id} - \sum_{k=1}^p A_k z$ ,  $\xi_2(z) = \text{Id} - \sum_{k=1}^p A_k z^p$ . If  $z_*$  is a solution to the first eigenvalue problem, then  $z_*^{1/p}$  is a solution to the second problem (and inversely). Since all eigenvalues of  $\xi_2$  have the same modulus, this means that the largest eigenvalue of  $\xi_2$  is  $z_*^{1/p}$ , where we can take its value to belong to  $\mathbb{R}$ . It follows that  $z_* < 1$  if and only if  $z_*^{1/p} < 1$ .

In the end, the statement follows directly from the previous statement, Lemma A.6.1 and the equivalence we just derived in (A.6.5):

$$\rho(A(\sum_{k=1}^p A_k, 0, \dots, 0)) < \rho(A(A_1, \dots, A_p)) < \rho(A(0, \dots, 0, \sum_{k=1}^p A_k)).$$

□



## APPENDIX

# B

## PROOFS FOR INVAR PROCESSES

### B.1 Definition

We say, where the definition from  $\mathbb{R}$  to  $\mathbb{R}^d$  naturally extends:

**Definition B.1.1** (Uniformly Integrable). We call **uniformly integrable** a family of random counting processes  $\{N_n\}_{n \in \mathbb{N}_{\geq 0}}$  if:

$$\forall A \in \mathcal{B}_b \lim_{K \rightarrow \infty} \sup_{n \in \mathbb{N}_{\geq 0}} \mathbb{E}[\mathbf{1}_{|N_n(A)| > K} |N_n(A)|] = 0. \quad (\text{B.1.1})$$

With this definition, it is direct to see that for all  $q > 1$ :  $L^q$  bounded variables are uniformly integrable (proven for the univariate case in [19]).

**Definition B.1.2** (Uniformly Tight). We call **uniformly tight** a family probability measure  $\{\mathbb{Q}^{(i)}\}_{i \in \mathbb{N}_{\geq 0}}$  on  $(M_p, \mathcal{M}_p)$  if:

$$\forall \varepsilon > 0, \exists K \subset M_p : \sup_{i \in \mathbb{N}_{\geq 0}} \mathbb{Q}^{(i)}[K^c] < \varepsilon. \quad (\text{B.1.2})$$

### B.2 Proof Existence INVAR

*Proof.* Proof of Theorem 1.3.8. Our proof is constructive. In a branching structure fashion, we create different objects that represent generations and families of events. We call generation process  $G_n^{(g,r,s)} \in \mathbb{N}_{\geq 0}^d$  the  $g$ -th generation from an immigrant  $s$ , immigrated at time  $r$ , and  $n$  represents the lag since the first immigrant (immigrated at time  $r$ ) when the generation lives. In the first generation, there is a 1 in every coordinate because every family needs to start with at least one member.  $G$  represents a possible genealogy tree, where at the root there is an ancestor represented by a 1. We define  $G$  through this recursion:

$$G_n^{(0,r,s)} = \mathbf{1}_{n=0}, \quad \forall n, r \in \mathbb{Z}, s \in \mathbb{N}_{>0}, \quad (\text{B.2.1})$$

$$G_n^{(g,r,s)} = \sum_{k=1}^n A_k \circ G_{n-k}^{(g-1,r,s)}, \quad \forall g \in \mathbb{N}_{>0}, n, r \in \mathbb{Z}, s \in \mathbb{N}_{>0}, \quad (\text{B.2.2})$$

where we mean by  $\mathbf{1}$  a  $d$ -dimensional vector with ones if  $n = 0$  and 0 otherwise.

Then, we aggregate the whole offspring tree of that immigrant  $s$  immigrated at time  $r$  by summing over  $g$  and we call it  $F_n^{(r,s)} \in \mathbb{N}_{\geq 0}^d$ :

$$F_n^{(r,s)} = \sum_{g=0}^{\infty} G_n^{(g,r,s)}, \quad \forall n \in \mathbb{Z}. \quad (\text{B.2.3})$$

Let  $\varepsilon_n \sim Pois(A_0) \in \mathbb{N}_{\geq 0}^d, n \in \mathbb{Z}$ , i.i.d.. The candidate series, solution to the stochastic difference equations, called  $\{\tilde{X}_n\}_{n \in \mathbb{Z}}$ , is the superposition of all the families composed of different immigrants  $s$  immigrated at times  $r$ :

$$\tilde{X}_n = \sum_{r=-\infty}^n \sum_{s=1}^{\varepsilon_r} F_{n-r}^{(r,s)}, \quad \forall n \in \mathbb{Z}. \quad (\text{B.2.4})$$

The sum from 1 to a random vector  $\varepsilon_r$  represents a component-wise sum: the parenthesis index corresponds to the indexing within the family, whereas the second index represents the coordinate of the vector:

$$\sum_{s=1}^{\varepsilon_r} F_{n-r}^{(r,s)} := \begin{pmatrix} \sum_{s=1}^{\varepsilon_r^1} F_{n-r}^{(r,s),1} \\ \vdots \\ \sum_{s=1}^{\varepsilon_r^d} F_{n-r}^{(r,s),d} \end{pmatrix}.$$

Now we see that if for an  $r \in \mathbb{Z}, \varepsilon_r$  is 0, the family  $F_{0-r}$  does not appear and the genealogy  $G_0^{(0,r,s)}$  does not even start. However, if it did start, it is through the definition of  $G$  that some cross-influence (the different coordinates) would appear!

If we suppose that our time series starts at time 0, then we could sum over  $r \in \{0, 1, \dots\}$ . However, in total generality, we need to account for a different starting point before 0. This sum is well defined as being nested increasing sums<sup>1</sup>.

The expectation of  $\sum_{s=1}^{\varepsilon_r} F_{n-r}^{(r,s)}$  results in a Hadamard product (introduced in Definition A.4.3) and using the expectation property in the univariate case from Proposition 1.3.4 gives:

$$\mathbb{E} \left[ \sum_{s=1}^{\varepsilon_r} F_{n-r}^{(r,s)} \right] = \begin{pmatrix} \nu^1 \mathbb{E}[F_{n-r}^{(r,1),1}] \\ \vdots \\ \nu^d \mathbb{E}[F_{n-r}^{(r,1),d}] \end{pmatrix} = \nu \odot \mathbb{E}[F_{n-r}^{(r,1)}].$$

We wish to prove by induction on the index  $g \in \mathbb{N}_{\geq 0}$ , by calling  $\sum_{k=1}^{\infty} A_k = K$  and using  $\mathbf{1}$  as the  $d$ -dimensional vector of ones:

$$\mathbb{E} \left[ \sum_{n=0}^{\infty} G_n^{(g,r,s)} \right] = K^g \cdot \mathbf{1}. \quad (\text{B.2.5})$$

The induction step goes like this:

$$\begin{aligned} \mathbb{E} \left[ \sum_{n=0}^{\infty} G_n^{(g,r,s)} \right] &= \mathbb{E} \left[ \sum_{n=0}^{\infty} \sum_{k=1}^n A_k \circ G_{n-k}^{(g-1,r,s)} \right] \\ &= \sum_{n=0}^{\infty} \sum_{k=1}^n A_k \mathbb{E} \left[ G_{n-k}^{(g-1,r,s)} \right] \\ &= \sum_{k=0}^{\infty} A_k \sum_{n=0}^{\infty} \mathbb{E} \left[ G_n^{(g-1,r,s)} \right] \\ &= K \cdot \mathbb{E} \left[ \sum_{n=0}^{\infty} G_n^{(g-1,r,s)} \right], \end{aligned}$$

---

<sup>1</sup>For any infinite sum in this proof, we do not repeat the arguments of well-definedness of the sums, or why we can exchange the sum and expectation operators in the future, which always boils down to the fact that we are working sums of positive numbers.

where we have used the expectation property from Proposition 1.3.4, and to go from line 2 to 3, we used that we are summing positive numbers, which also gives that we can interchange the order of the sums. We highlight that the product on the last line is a matrix-vector product. To conclude, we need to verify that the property is true at  $g = 0$ :  $\mathbb{E}[\sum_{n=0}^{\infty} G_n^{(0,r,s)}] = \sum_{n=0}^{\infty} \mathbf{1}_{n=0} = \mathbf{1} = K^0 \cdot \mathbf{1}$ .

Then, it is well known that  $\sum_{g=0}^M K^g \xrightarrow[M \rightarrow \infty]{} (\text{Id} - K)^{-1}$ , when the spectral radius of  $K$  is strictly less than 1. This means that  $\mathbb{E}[\sum_{n=0}^{\infty} F_n^{(r,s)}] = \mathbb{E}[\sum_{n=0}^{\infty} \sum_{g=0}^{\infty} G_n^{(g,r,s)}] = (\text{Id} - K)^{-1} \cdot \mathbf{1}$ .

In the following, we use again the Hadamard product notation introduced in Definition A.4.3. We compute the expectation of the candidate process  $\tilde{X}$ :

$$\begin{aligned}\mathbb{E}[\tilde{X}_n] &= \mathbb{E}\left[\sum_{r=-\infty}^n \sum_{s=1}^{\varepsilon_r} F_{n-r}^{(r,s)}\right] \\ &= \sum_{r=-\infty}^n \mathbb{E}\left[\sum_{s=1}^{\varepsilon_r} F_{n-r}^{(r,s)}\right] \\ &= \sum_{r=-\infty}^n \nu \odot \mathbb{E}[F_{n-r}^{(r,s)}] \\ &= \nu \odot \mathbb{E}\left[\sum_{r=-\infty}^n F_{n-r}^{(r,s)}\right].\end{aligned}$$

Then, we have thanks to the monotone convergence Theorem:

$$\begin{aligned}\mathbb{E}[\tilde{X}_n] &= \nu \odot \left(\sum_{r=-\infty}^n \mathbb{E}[F_{n-r}^{(r,1)}]\right) = \nu \odot \left(\sum_{r=-\infty}^n \sum_{g=0}^{\infty} \mathbb{E}[G_{n-r}^{(g,r,1)}]\right) \\ &= \nu \odot \left(\sum_{g=0}^{\infty} \mathbb{E}\left[\sum_{r=-\infty}^n G_{n-r}^{(g,r,1)}\right]\right) = \nu \odot \left(\sum_{g=0}^{\infty} K^g\right) \cdot \mathbf{1} = \nu \odot (\text{Id} - K)^{-1} \mathbf{1}. \quad (\text{B.2.6})\end{aligned}$$

Now, it remains to prove that the candidate  $\tilde{X}_n$  is indeed an INVAR process, or in other words, it solves the stochastic difference equations (1.3.4)) and that it is unique. We need to prove that  $\tilde{X}_n$  verifies:

$$\begin{aligned}\varepsilon_n &= \tilde{X}_n - \sum_{k=1}^{\infty} A_k \circ \tilde{X}_{n-k} \\ &= \tilde{X}_n - \begin{pmatrix} \sum_{k=1}^{\infty} \sum_{j=1}^d \sum_{l=1}^{\tilde{X}_{n-k}^j} \xi_l^{(n,k),(1,j)} \\ \vdots \\ \sum_{k=1}^{\infty} \sum_{j=1}^d \sum_{l=1}^{\tilde{X}_{n-k}^j} \xi_l^{(n,k),(d,j)} \end{pmatrix}. \quad (\text{B.2.7})\end{aligned}$$

In the definition of an INVAR process, we had defined the  $\xi_l^{(n,k),(i,j)}$  as being a sequence of random variables following a Poisson distribution with parameter  $A_k$ , independent over  $n \in \mathbb{Z}, k \in \mathbb{N}_{>0}, l \in \mathbb{N}_{\geq 0}, i, j \in \{1, \dots, d\}$ .  $i, j$  represents the origin of the events (to dimension  $i$  from dimension  $j$ ), and  $n, k$  represents together when the event happened. First, we explicitly write  $G_n^{(g,i,j)}$  from (B.2.2) using a sequence of Poisson random variable  $\xi_{g,r,s,m}^{(n,k),(i,j)}$ . One can interpret each term as being the number of offspring at time  $n$ , with parents at time  $n - k$ , from dimension  $j$  onto dimension  $i$  and belonging to the  $g$ -th generation of the family immigrant  $s$  at time  $r$ :

$$G_n^{(0,r,s)} = \mathbb{1}_{n=0}, \quad \forall n, r \in \mathbb{Z}, s \in \mathbb{N}_{>0},$$

$$G_n^{(g,r,s)} = \begin{pmatrix} \sum_{k=1}^n \sum_{j=1}^d \sum_{m=1}^{G_{n-k}^{(g-1,r,s),j}} \xi_{g,r,s,m}^{(r+n,k),(1,j)} \\ \vdots \\ \sum_{k=1}^n \sum_{j=1}^d \sum_{m=1}^{G_{n-k}^{(g-1,r,s),j}} \xi_{g,r,s,m}^{(r+n,k),(d,j)} \end{pmatrix}, \quad \forall g \in \mathbb{N}_{>0}, n, r \in \mathbb{Z}, s \in \mathbb{N}_{>0}.$$

We may write  $\forall n \in \mathbb{Z}$ :

$$\begin{aligned} \tilde{X}_n &= \sum_{r=-\infty}^n \sum_{s=1}^{\varepsilon_r} F_{n-r}^{(r,s)} \\ &= \sum_{r=-\infty}^n \sum_{s=1}^{\varepsilon_r} \sum_{g=0}^{\infty} G_{n-r}^{(g,r,s)} \\ &= \sum_{r=-\infty}^{n-1} \sum_{s=1}^{\varepsilon_r} \sum_{g=0}^{\infty} G_{n-r}^{(g,r,s)} + \varepsilon_n. \end{aligned}$$

And actually, on the last line, we can replace the most inner sum from  $g = 0$  to  $\infty$  by a sum starting at 1, because  $\forall r \leq n-1, G_{n-r}^{(0,r,s)} = 0$ .

We now replace  $G_{n-r}^{(g,r,s)}$  by its expression in terms of  $\xi$  from above:

$$\begin{aligned} \tilde{X}_n &= \sum_{r=-\infty}^{n-1} \sum_{s=1}^{\varepsilon_r} \sum_{g=1}^{\infty} G_{n-r}^{(g,r,s)} = \sum_{r=-\infty}^{n-1} \sum_{s=1}^{\varepsilon_r} \sum_{g=1}^{\infty} \begin{pmatrix} \sum_{k=1}^{n-r} \sum_{j=1}^d \sum_{m=1}^{G_{n-r-k}^{(g-1,r,s),j}} \xi_{g,r,s,m}^{(r+n-r,k),(1,j)} \\ \vdots \\ \sum_{k=1}^{n-r} \sum_{j=1}^d \sum_{m=1}^{G_{n-r-k}^{(g-1,r,s),j}} \xi_{g,r,s,m}^{(r+n-r,k),(d,j)} \end{pmatrix} \\ &= \begin{pmatrix} \sum_{r=-\infty}^{n-1} \sum_{s=1}^{\varepsilon_r} \sum_{g=1}^{\infty} \sum_{k=1}^{n-r} \sum_{j=1}^d \sum_{m=1}^{G_{n-r-k}^{(g-1,r,s),j}} \xi_{g,r,s,m}^{(n,k),(1,j)} \\ \vdots \\ \sum_{r=-\infty}^{n-1} \sum_{s=1}^{\varepsilon_r} \sum_{g=1}^{\infty} \sum_{k=1}^{n-r} \sum_{j=1}^d \sum_{m=1}^{G_{n-r-k}^{(g-1,r,s),j}} \xi_{g,r,s,m}^{(n,k),(d,j)} \end{pmatrix}. \end{aligned}$$

Remark that we can permute the two sums with variables  $r$  and  $k$ . Then, we may extend  $\sum_{k=1}^{n-r}$  to go until  $\infty$  because for any  $i$ ,  $G_{n-r-k}^{(g-1,r,s),i}$  is 0 when  $k > n - r$ . We can also reduce the sum  $\sum_{r=-\infty}^{n-1}$  to go only up to  $n - k$  since the terms for  $r = n - k + 1, \dots, n - 1$  are zero. The previous sums read:

$$\tilde{X}_n = \sum_{r=-\infty}^{n-1} \sum_{s=1}^{\varepsilon_r} \sum_{g=1}^{\infty} G_{n-r}^{(g,r,s)} = \sum_{k=1}^{\infty} \sum_{j=1}^d \begin{pmatrix} \sum_{r=-\infty}^{n-k} \sum_{s=1}^{\varepsilon_r} \sum_{g=1}^{\infty} \sum_{m=1}^{G_{n-r-k}^{(g-1,r,s),j}} \xi_{g,r,s,m}^{(n,k),(1,j)} \\ \vdots \\ \sum_{r=-\infty}^{n-k} \sum_{s=1}^{\varepsilon_r} \sum_{g=1}^{\infty} \sum_{m=1}^{G_{n-k-r}^{(g-1,r,s),j}} \xi_{g,r,s,m}^{(n,k),(d,j)} \end{pmatrix}. \quad (\text{B.2.8})$$

The concluding argument to prove that the stochastic difference equations are satisfied by  $\tilde{X}_n$  is to notice that in every component of the vector, the number of elements summed over through the five sums corresponds exactly to the corresponding components of  $\tilde{X}_{n-k}$ , by (B.2.4):

$$\sum_{r=-\infty}^{n-k} \sum_{s=1}^{\varepsilon_r} \sum_{g=1}^{\infty} \sum_{m=1}^{G_{n-r-k}^{(g-1,r,s),j}} \xi_{g,r,s,m}^{(n,k),(d,j)} = \tilde{X}_{n-k}^j, \quad \forall j \in \{1, \dots, d\}.$$

The argument is the same as in the uni-dimensional case and more details can be found in [117].

In conclusion, we have proven  $\forall n \in \mathbb{Z}$ :

$$\begin{aligned}
\tilde{X}_n &= \sum_{k=1}^{\infty} \sum_{j=1}^d \begin{pmatrix} \sum_{r=-\infty}^{n-k} \sum_{s=1}^{\varepsilon_r} \sum_{g=1}^{\infty} \sum_{m=1}^{G_{n-r-k}} \xi_{g,r,s,m}^{(n,k),(1,j)} \\ \vdots \\ \sum_{r=-\infty}^{n-k} \sum_{s=1}^{\varepsilon_r} \sum_{g=1}^{\infty} \sum_{m=1}^{G_{n-r-k}} \xi_{g,r,s,m}^{(n,k),(d,j)} \end{pmatrix} + \varepsilon_n \\
&= \sum_{k=1}^{\infty} \sum_{j=1}^d \begin{pmatrix} \sum_{l=1}^{\tilde{X}_{n-k}^j} \xi_l^{(n,k),(1,j)} \\ \vdots \\ \sum_{l=1}^{\tilde{X}_{n-k}^j} \xi_l^{(n,k),(d,j)} \end{pmatrix} + \varepsilon_n.
\end{aligned}$$

This shows that the stochastic difference equations (1.3.4) is satisfied almost surely.

To prove uniqueness, we consider two processes satisfying the same equations. We call them  $X_n$  and  $Y_n$ , defined on the same probability space and having the same offspring. For all  $n \in \mathbb{Z}$ :

$$\mathbb{E}[|X_n - Y_n|] \leq \sum_{k=1}^{\infty} \mathbb{E}[|A_k \circ X_{n-k} - A_k \circ Y_{n-k}|],$$

by triangular inequality. Also, the difference is well defined because by assumption, the  $L^1$  norm of each process is finite, see (B.2.6).

We know that for each component of the vector of the differences  $A_k \circ X_{n-k} - A_k \circ Y_{n-k}$ , we can prove equality in the  $L^1$  sense which gives almost sure equality as done in [117]. This proves that both processes are equal almost surely.

We proved the process is unique and satisfies the original conditions. Hence, we have constructed an INVAR( $\infty$ ) process respecting the property (1.3.5) and it is uniquely defined almost surely.  $\square$

### B.3 Proof VAR Representation of INVAR

*Proof.* Proof of Theorem 1.3.10. First, the equation is well-defined because the INVAR process exists and is well-defined. Furthermore, the sum aggregates positive terms so the limit of the sum is well-defined. We fix  $n, n' \in \mathbb{Z}$  for the rest of this proof.

We get that  $\mathbb{E}[u_n] = 0$  by tower property; we define the natural filtration  $\{\mathcal{F}_n\}_{n \in \mathbb{N}_{\geq 0}} = \{\sigma(X_k : k \leq n)\}_{n \in \mathbb{N}_{\geq 0}}$ , and we use that  $\mathbb{E}[\sum_{k=1}^{\infty} A_k X_{n-k}] = \sum_{k=1}^{\infty} A_k \mathbb{E}[X_n] = K\mu_X$  (by stationarity):

$$\begin{aligned}
\mathbb{E}[u_n] &= \mathbb{E}[X_n - \sum_{k=1}^{\infty} A_k X_{n-k} - \nu] \\
&= \mathbb{E}\left[\mathbb{E}[X_n | \mathcal{F}_{n-1}] - \sum_{k=1}^{\infty} A_k X_{n-k} - \nu\right] \\
&= \mu_X - K\mu_X - \nu \\
&= (\text{Id} - K)\mu_X - \nu \\
&= \mathbf{0}.
\end{aligned}$$

When  $n < n'$ , the autocovariance matrix sequence looks like:

$$\begin{aligned}
\mathbb{E}[u_n u_{n'}^T] &= \mathbb{E}[\mathbb{E}[u_n u_{n'}^T | \mathcal{F}_{n-1}]] \\
&= \mathbb{E}[\mathbb{E}[u_n | \mathcal{F}_{n-1}] u_{n'}^T] \\
&= 0.
\end{aligned}$$

Finally, in order to compute the covariance matrix of  $u_n$ , we develop the expression  $\mathbb{E}[u_n u_n^T]$ . We use the Diag operator from Definition A.4.4:

$$\begin{aligned}
\mathbb{E}[u_n u_n^T] &= \mathbb{E} \left[ \left( X_n - \sum_{k=1}^{\infty} A_k X_{n-k} - \nu \right) \cdot (u_n)^T \right] \\
&= \mathbb{E}[X_n \cdot X_n^T] + \mathbb{E} \left[ X_n \cdot \left( - \sum_{k=1}^{\infty} A_k X_{n-k} - \nu \right)^T \right] \\
&\quad + \mathbb{E} \left[ \left( - \sum_{k=1}^{\infty} A_k X_{n-k} - \nu \right) \cdot (u_n)^T \right].
\end{aligned}$$

The last term vanishes because  $\mathbb{E}[u_n^T] = \mathbf{0}$ ,  $\mathbb{E}[X_{n-k} u_n^T] = \mathbf{0}, k > 0$ .

Then, we replace  $X_n$  by its value from the difference equations (1.3.4):  $\varepsilon_n + \sum_{k=1}^{\infty} A_k \circ X_{n-k}$ :

$$\begin{aligned}
\mathbb{E}[X_n \cdot X_n^T] &= \mathbb{E}[(\varepsilon_n + \sum_{k=1}^{\infty} A_k \circ X_{n-k}) \cdot (\varepsilon_n + \sum_{k=1}^{\infty} A_k \circ X_{n-k})^T] \\
&= \mathbb{E}[\varepsilon_n \varepsilon_n^T + \varepsilon_n (\sum_{k=1}^{\infty} A_k \circ X_{n-k})^T + (\sum_{k=1}^{\infty} A_k \circ X_{n-k}) \varepsilon_n^T + (\sum_{k=1}^{\infty} A_k \circ X_{n-k})(\sum_{k=1}^{\infty} A_k \circ X_{n-k})^T] \\
&= \text{Diag}(\nu) + \nu \nu^T + \nu \mathbb{E}[(\sum_{k=1}^{\infty} A_k \circ X_{n-k})^T] + \mathbb{E}[(\sum_{k=1}^{\infty} A_k \circ X_{n-k})] \nu^T + \\
&\quad \mathbb{E}[(\sum_{k=1}^{\infty} A_k \circ X_{n-k})(\sum_{k=1}^{\infty} A_k \circ X_{n-k})^T],
\end{aligned}$$

where we have used  $\mathbb{E}[\varepsilon_n] = \nu$ ,  $\mathbb{E}[\varepsilon_n \varepsilon_n^T] = \text{Diag}(\nu^2) + \nu \nu^T$  because the random vector is Poisson, and the assumption that the  $\varepsilon_n$  is independent of history.

The second term:

$$\begin{aligned}
\mathbb{E} \left[ X_n \cdot \left( - \sum_{k=1}^{\infty} A_k X_{n-k} - \nu \right)^T \right] &= \mathbb{E} \left[ \left( \varepsilon_n + \sum_{k=1}^{\infty} A_k \circ X_{n-k} \right) \cdot \left( - \sum_{k=1}^{\infty} A_k X_{n-k} - \nu \right)^T \right] \\
&= -\nu \mathbb{E} \left[ \left( \sum_{k=1}^{\infty} A_k X_{n-k} + \nu \right)^T \right] - \mathbb{E} \left[ \left( \sum_{k=1}^{\infty} A_k \circ X_{n-k} \right) \cdot \left( \sum_{k=1}^{\infty} A_k X_{n-k} + \nu \right)^T \right] \\
&= -\nu \nu^T - \nu \mathbb{E} \left[ \left( \sum_{k=1}^{\infty} A_k X_{n-k} \right)^T \right] - \\
&\quad \mathbb{E} \left[ \left( \sum_{k=1}^{\infty} A_k \circ X_{n-k} \right) \cdot \left( \sum_{k=1}^{\infty} A_k X_{n-k} \right)^T \right] - \mathbb{E} \left[ \left( \sum_{k=1}^{\infty} A_k \circ X_{n-k} \right) \right] \cdot \nu^T.
\end{aligned}$$

By recombining these two quantities, some terms cancel out (in particular,  $\mathbb{E}[(\sum_{k=1}^{\infty} A_k X_{n-k})^T] = \mathbb{E}[(\sum_{k=1}^{\infty} A_k \circ X_{n-k})^T]$ ). This is visible by a conditioning argument on the sequence of the  $X_n$ .

It remains:

$$\mathbb{E}[u_n u_n^T] = \text{Diag}(\nu) + \mathbb{E}[(\sum_{k=1}^{\infty} A_k \circ X_{n-k})(\sum_{k=1}^{\infty} A_k \circ X_{n-k})^T] - \mathbb{E} \left[ \left( \sum_{k=1}^{\infty} A_k \circ X_{n-k} \right) \cdot \left( \sum_{k=1}^{\infty} A_k X_{n-k} \right)^T \right]. \quad (\text{B.3.1})$$

Directly by conditioning on  $\mathcal{F}_{\infty}$  and factorising the elements:

$$\begin{aligned}
& \mathbb{E}\left[\left(\sum_{k=1}^{\infty} A_k \circ X_{n-k}\right)\left(\sum_{k=1}^{\infty} A_k \circ X_{n-k}\right)^T\right] - \mathbb{E}\left[\left(\sum_{k=1}^{\infty} A_k \circ X_{n-k}\right) \cdot \left(\sum_{k=1}^{\infty} A_k X_{n-k}\right)^T\right] \\
& = \mathbb{E}\left[\text{Cov}\left(\sum_{k=1}^{\infty} A_k \circ X_{n-k} \mid \mathcal{F}_{\infty}\right)\right], \tag{B.3.2}
\end{aligned}$$

where by Proposition 1.3.4:

$$\mathbb{E}\left[\text{Cov}\left(\sum_{k=1}^{\infty} A_k \circ X_{n-k} \mid \sigma(X_{n-k}, k > 0)\right)\right] = \mathbb{E}[\text{Diag}(A_k X_{n-k})].$$

Then, this gives the following equalities, where we used what we computed to go from first to second line, then linearity of the expectation, by stationarity (see the properties after Definition A.4.4) we get the third line and finally, we use the linearity of the operator Diag, and we conclude on the last equality using  $\mu_X = \nu + K\mu_X$ :

$$\begin{aligned}
\mathbb{E}[u_n u_n^T] &= \text{Diag}(\nu) + \mathbb{E}\left[\text{Cov}\left(\sum_{k=1}^{\infty} A_k \circ X_{n-k} \mid \sigma(X_{n-k}, k > 0)\right)\right] \\
&= \text{Diag}(\nu) + \sum_{k=1}^{\infty} \mathbb{E}[\text{Diag}(A_k X_{n-k})] \\
&= \text{Diag}(\nu) + \sum_{k=1}^{\infty} \text{Diag}(A_k \mathbb{E}[X_0]) \\
&= \text{Diag}(\nu) + \text{Diag}(K \cdot \mu_X) \\
&= \text{Diag}(\mu_X).
\end{aligned}$$

□

## B.4 Proof Covariance Function of an INVAR

*Proof.* Proof of Theorem 1.3.12. The autocovariance matrix function is obtained by identification with the statements from [31], in particular, it follows from (11.1.12) and (11.1.13). The statement reads that if for  $t \in \mathbb{N}_{\geq 0}$ ,  $X_t = \sum_{j \in \mathbb{Z}} C_j Z_{t-j}$  where the coefficients  $C_j$  are summable, with multivariate white noise  $Z_t$  with covariance matrix  $\Sigma$ , then the causal process  $X_t$  is stationary, with mean 0 and covariance function  $\Gamma(h) = \sum_{j \in \mathbb{Z}} C_{j+h} \cdot \Sigma \cdot C_j^T$ ,  $h \in \mathbb{Z}$ .

Since  $\text{Cov}(u_n, u_n) = \text{Diag}(\nu \odot (\text{Id} - K)^{-1} \mathbf{1})$ , and since the coefficients in the MA representation (cf. Theorem 1.3.11) are summable, we can use the proposition to conclude that the autocovariance sequence is:

$$R(j) = \sum_{k=0}^{\infty} B_k \text{Diag}(\nu \odot (\text{Id} - K)^{-1} \mathbf{1}) B_{k+|j|}^T, \quad j \in \mathbb{Z}, k \in \mathbb{N}_{\geq 0}.$$

We have the sum of the autocovariance matrices (we mean by inequality the component-wise inequality):

$$\begin{aligned}
\sum_{j=0}^{\infty} R(j) &= \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} B_k \text{Diag}(\nu \odot (\text{Id} - K)^{-1} \mathbf{1}) B_{k+j}^T \\
&\leq \sum_{k=0}^{\infty} \sum_{j=-k}^{\infty} B_k \text{Diag}(\nu \odot (\text{Id} - K)^{-1} \mathbf{1}) B_{k+j}^T \\
&= \left( \sum_{k=0}^{\infty} B_k \right) \text{Diag}(\nu \odot (\text{Id} - K)^{-1} \mathbf{1}) \left( \sum_{j=0}^{\infty} B_j^T \right).
\end{aligned}$$

We were able to add terms to the sum because the covariance matrix is by definition positive and the coefficients  $B$  are positive. The bound is given by replacing  $\sum_{k=0}^{\infty} B_k = (\text{Id} - K)^{-1} \mathbf{1}$ .  $\square$

## B.5 Proof Weak Convergence from INVAR to Hawkes

*Proof.* Proof of Theorem 1.3.13. Let  $\delta > 0$  and we start with a family of point processes  $(N^{(\Delta)})_{\Delta \in (0, \delta)}$  as in the statement of the theorem, to which corresponds a family of probability measures  $(\mathbb{P}^{(\Delta)})_{\Delta \in (0, \delta)}$ . The family of probability measures is uniformly tight as defined in Definition B.1.2, see Lemma B.6.4, hence sequentially compact<sup>2</sup> by Prokhorov's Theorem [50] or in [19]. Now, we take any sequence  $\tilde{\Delta}$  converging to 0. There exists a subsequence of  $\tilde{\Delta}$  that converges weakly to a point process called  $N^*$ , and the subsequence  $\{\Delta_n\}_{n \in \mathbb{N}_{\geq 0}}$ . If the limit does not depend on the choice of subsequence, then the original sequence converges weakly to  $N^*$  (see Theorem 2.3 in [19]). Thus, we only need to show that any subsequential limit candidate  $N^*$  solves (A.1.8), which means that all limits are equal (in distributions) and are Hawkes processes, as explained in Theorem A.1.4. This would show that the sequence  $(N^{(\Delta)})_{\Delta \in (0, \delta)}$  converges weakly to a Hawkes process. We call  $\{N_n\}_{n \in \mathbb{N}_{\geq 0}}$  the multivariate point processes associated to the sequence  $\{\Delta_n\}_{n \in \mathbb{N}_{\geq 0}}$ . Because of the subscript, we replace the time evaluation of  $N_n$  to be written with parentheses  $(N_n)_t =: N_n(t)$ .

We take cylindrical sets that generate the filtration  $\mathcal{F}_t$ . For a chosen multivariate point process  $N$ :

$$\begin{aligned} \mathfrak{B}_a^N := \{ \{ \omega \in \Omega : \bigcup_{i=1}^k \{N([s_i, t_i])(w) \in D_i\} \} : \\ -\infty < s_i < t_i \leq a, D_i \subset \mathbb{N}_{\geq 0}^d, k \in \mathbb{N}_{\geq 0} \}. \end{aligned} \quad (\text{B.5.1})$$

What we would like to show in the following is that for all  $a < b$ ,  $A^* \in \mathfrak{B}_a^{N^*}$ :

$$\mathbb{E} \left[ \mathbf{1}_{A^*} N_{[a,b]}^* \right] = \mathbb{E} \left[ \mathbf{1}_{A^*} \int_a^b \lambda(s) \, ds \right], \quad (\text{B.5.2})$$

where the conditional intensity is the one from a Hawkes process.

This would show that the limiting process is a Hawkes process. As stated, the Hawkes' unicity shows that the subsequence's choice was arbitrary. We have:

$$\begin{aligned} \mathbb{E} [\mathbf{1}_{A_n} N_n((a, b))] &= \mathbb{E} \left[ \mathbf{1}_{A_n} \sum_{k \in \mathbb{Z}: k\Delta_n \in (a, b]} X_k^{(\Delta_n)} \right] \\ &= \mathbb{E} \left[ \mathbf{1}_{A_n} \sum_{k \in \mathbb{Z}: k\Delta_n \in (a, b]} \left( \varepsilon_k^{(\Delta_n)} + \sum_{l=1}^{\infty} (\Delta_n \cdot \Phi(l \cdot \Delta_n)) \circ X_{k-l}^{(\Delta_n)} \right) \right] \\ &= \mathbb{E} \left[ \mathbf{1}_{A_n} \sum_{k \in \mathbb{Z}: k\Delta_n \in (a, b]} \left( \Delta_n \cdot \nu + \sum_{l=1}^{\infty} \Delta_n \cdot \Phi(l \cdot \Delta_n) X_{k-l}^{(\Delta_n)} \right) \right] \\ &= \mathbb{E} \left[ \mathbf{1}_{A_n} \sum_{k \in \mathbb{Z}: k\Delta_n \in (a, b]} \Delta_n \left( \nu + \int_{-\infty}^{k\Delta_n} \Phi(k \cdot \Delta_n - s) \, dN_n(s) \right) \right], \end{aligned} \quad (\text{B.5.3})$$

where the second to last line follows from the fact that whenever  $k$  is such that  $k\Delta_n > a$ , the immigrants  $\varepsilon_k^{(\Delta_n)}$  and the innovation in  $X_k^{(\Delta_n)}$  are independent of an event up to time  $a$ , and hence we can replace the thinning by its mean; in the last line, we rewrote the sum as an integral with respect to the random measure  $N_n$ .

---

<sup>2</sup>Then, for any sequence  $(N^{(\Delta)})_{\Delta \in (0, \delta)}$ , we can extract a converging subsequence  $(N^{(\Delta_n)})_{\Delta_n \in (0, \delta)}$  and we call the limit  $N^*$ .

We will show that the left-hand sides (LHS) of (B.5.3) converges to the LHS of (B.5.2), as well as the right-hand sides (RHS).

First, we study the **LHS**:

We define the mapping, for arbitrary  $(D_i)_{i < k}$  sets as in (B.5.1):

$$\begin{aligned}\Psi : (M_p, \mathcal{M}_p) &\rightarrow (\mathbb{R}, \mathcal{B}) \\ m &\mapsto \mathbb{1}_{\bigcup_{i=1}^k \{N_{[s_i, t_i]}(w) \in D_i\}} m([a, b]).\end{aligned}$$

We would like to use the continuous mapping Theorem, and for that we need to prove that this mapping is almost surely continuous. This mapping is trivially (vaguely) continuous (with respect to the vague topology) on  $M_p \setminus \{m : m(D_\Psi)\}$  where  $D_\Psi$  is defined as the set of discontinuities  $D_\Psi := \{a, b\} \cup \bigcup_{i=1}^k \{s_i, t_i\}$ . By Theorem B.6.1, if we have that  $\mathbb{P}[N(D_\Psi) > 0] = 0$ , then this would prove that  $\Psi(N_n) \xrightarrow{w} \Psi(N^*)$ . Since  $D_\Psi$  is finite, we just need to show that for all the mentioned points, the probability is 0. Sufficiently, we can show that  $\forall t \in \mathbb{R}$ ,  $\mathbb{P}[N_t^* > 0] = 0$ . We show this in the following, using  $\mathbf{0}$  as the vector of zeros. Also, we recall we are using the  $l^1$  norm. This allows us to say  $\mathbb{1}_{N > 0} \leq |N|$  for  $N$  a multivariate point process and that  $|N| = \mathbf{1}^T N$ , because every component of a point process is positive.

Note that:

$$\begin{aligned}\mathbb{P}[N_t^* > \mathbf{0}] &= \mathbb{E}[\mathbb{1}_{N_t^* > \mathbf{0}}] \\ &\leq \mathbb{E}[|N_t^*|] \\ &= \mathbb{E}[\mathbf{1}^T N_t^*]\end{aligned}\tag{B.5.4}$$

$$= \lim_{n \rightarrow \infty} \mathbf{1}^T \mathbb{E}[N_n(\{t\})]\tag{B.5.5}$$

$$= \lim_{n \rightarrow \infty} \Delta_n \mathbf{1}^T \nu \odot (\text{Id} - K^{(\Delta)})^{-1} \mathbf{1} = 0,\tag{B.5.6}$$

where we have used the Theorem B.6.1 to go from (B.5.4) to (B.5.5) and that the norm is a continuous function, and on the last line the equality is due to Lemma B.6.2. In other words, the probability must be equal to 0.

This shows

$$\mathbb{1}_{A_n} N_n([a, b]) \xrightarrow[n \rightarrow \infty]{w} \mathbb{1}_{A^*} N_{[a, b]}^*,$$

and in conclusion, with the uniform integrability of  $N_n([a, b])$  implying uniform integrability of  $\mathbb{1}_{A_n} N_n([a, b])$ , we get the convergence of the means (and not in means) by the Lemma B.6.7:

$$\lim_{n \rightarrow \infty} \mathbb{E}[\mathbb{1}_{A_n} N_n([a, b])] = \mathbb{E}[\mathbb{1}_{A^*} N_{[a, b]}^*],\tag{B.5.7}$$

which is what we desired.

Second, we prove the convergence of the **RHS**.

The convergence of the RHS is equivalent to proving the convergence:

$$\begin{aligned}\lim_{n \rightarrow \infty} \sum_{k \in \mathbb{Z}: k \Delta_n \in (a, b]} \mathbb{E} \left[ \Delta_n \mathbb{1}_{A_n} \int_{-\infty}^{k \Delta_n} \Phi(k \cdot \Delta_n - s) dN_n(s) \right] \\ = \int_a^b \mathbb{E} \left[ \mathbb{1}_{A^*} \int_{-\infty}^t \Phi(t - s) dN_s^* \right] dt.\end{aligned}$$

We organise the rest of the proof as follows:

- We cut the improper integral into a tail and a body, and we analyse both pieces,

- Then, by continuity argument and uniform integrability (proved by showing that the random vector is  $L^2$  bounded), we get the convergence of the means for the body,
- We conclude by showing the tails converge as well.

We chose  $-M < a$ , a lower bound of the integral's domain. We need to show that for all  $t \in [a, b]$ :

$$\mathbb{1}_{A_n} \int_{-M}^t \Phi(k \cdot \Delta_n - s) dN_n(s) \xrightarrow[n \rightarrow \infty]{w} \mathbb{1}_{A^*} \int_{-M}^t \Phi(t-s) dN_s^*. \quad (\text{B.5.8})$$

This is direct by proceeding in the same fashion as before with the mapping  $\Psi$ . Indeed, the integral is nothing more than a sum over the events of the point process, and for this reason, we can define the almost identical mapping and prove the same properties by using the Theorem B.6.1 and Lemma B.6.2.

Next, we argue that the random vector  $\int_{-M}^t \Phi(t-s) dN_n(s)$  has a bounded  $L^2$  norm (Lemma B.6.5). This shows that the random vector  $\int_{-M}^t \Phi(k \cdot \Delta_n - s) dN_n(s)$  is uniformly integrable (the implication follows directly from the Definition B.1.1 and Proposition B.6.3), hence is  $\mathbb{1}_{A_n} \int_{-M}^t \Phi(k \cdot \Delta_n - s) dN_n(s)$  uniformly integrable. As before, this proves convergence of the means thanks to Lemma B.6.7:

$$\mathbb{E} \left[ \mathbb{1}_{A_n} \int_{-M}^t \Phi(k \cdot \Delta_n - s) dN_n(s) \right] \xrightarrow{n \rightarrow \infty} \mathbb{E} \left[ \mathbb{1}_{A^*} \int_{-M}^t \Phi(t-s) dN_s^* \right]. \quad (\text{B.5.9})$$

We introduce two quantities for conciseness:

$$\begin{aligned} E_{n,M_\varepsilon}(t) &:= \mathbb{E} \left[ \mathbb{1}_{A_n} \int_{-M_\varepsilon}^t \Phi(k \cdot \Delta_n - s) dN_n(s) \right], \\ E_{M_\varepsilon}^*(t) &:= \mathbb{E} \left[ \mathbb{1}_{A^*} \int_{-M_\varepsilon}^t \Phi(t-s) dN_s^* \right]. \end{aligned}$$

Using this notation, we have just proven that

$$\forall M_\varepsilon > 0, t \in ]a, b], \quad \lim_{n \rightarrow \infty} |E_{n,M_\varepsilon}(t) - E_{M_\varepsilon}^*(t)| = 0.$$

We are now ready to look at the difference between the two terms of the RHS. We will prove that the first line converges to 0:

$$\begin{aligned} &\left| \sum_{k \in \mathbb{Z}: k\Delta_n \in (a,b]} \mathbb{E} \left[ \Delta_n \mathbb{1}_{A_n} \int_{-\infty}^{k\Delta_n} \Phi(k \cdot \Delta_n - s) dN_n(s) \right] - \int_a^b \mathbb{E} \left[ \mathbb{1}_{A^*} \int_{-\infty}^t \Phi(t-s) dN_s^* \right] dt \right| \\ &\leq \left| \sum_{k \in \mathbb{Z}: k\Delta_n \in (a,b]} \mathbb{E} \left[ \Delta_n \mathbb{1}_{A_n} \int_{-M}^{k\Delta_n} \Phi(k \cdot \Delta_n - s) dN_n(s) \right] - \int_a^b \mathbb{E} \left[ \mathbb{1}_{A^*} \int_{-M}^t \Phi(t-s) dN_s^* \right] dt \right| \\ &+ \left| \sum_{k \in \mathbb{Z}: k\Delta_n \in (a,b]} \mathbb{E} \left[ \Delta_n \mathbb{1}_{A_n} \int_{-\infty}^{-M} \Phi(k \cdot \Delta_n - s) dN_n(s) \right] \right| \\ &+ \left| \int_a^b \mathbb{E} \left[ \mathbb{1}_{A^*} \int_{-\infty}^{-M} \Phi(t-s) dN_s^* \right] dt \right|, \end{aligned}$$

and where we have cut the difference into three pieces and used the triangular inequality. We call the terms respectively first, second and third summand.

We show that there exists a constant  $M_\varepsilon$  and an  $n_\varepsilon$  such that each term is bounded by  $\varepsilon > 0$ , arbitrary.

#### First summand.

We add and subtract the term  $\int_a^b E_n(t) dt$  in order to get the correct approximation, and identify the notation  $E, E^*$ :

$$\begin{aligned} & \left| \sum_{k \in \mathbb{Z}: k\Delta_n \in (a,b]} \Delta_n E_{n,M_\varepsilon}(k \cdot \Delta_n) - \int_a^b E_{M_\varepsilon}^*(t) dt \right| \\ & \leq \left| \sum_{k \in \mathbb{Z}: k\Delta_n \in (a,b]} \Delta_n E_{n,M_\varepsilon}(k \cdot \Delta_n) - \int_a^b E_{n,M_\varepsilon}(t) dt \right| + \left| \int_a^b E_{n,M_\varepsilon}(t) dt - \int_a^b E_{M_\varepsilon}^*(t) dt \right|. \end{aligned}$$

The second term converges to zero by dominated convergence. Indeed,

- $\lim_{n \rightarrow \infty} E_n = E^*$ ,
- $E_{n,M_\varepsilon}$  is bounded uniformly with respect to  $n$ :  $\forall t \in [a,b] :$

$$\begin{aligned} |E_{n,M_\varepsilon}(t)| &= \left| \mathbb{E} \left[ \mathbf{1}_{A_n} \int_{-M_\varepsilon}^t \Phi(t-s) dN_n(s) \right] \right| \\ &\leq \left| \mathbb{E} \left[ \int_{-M_\varepsilon}^b \Phi(t-s) dN_n(s) \right] \right| \\ &\leq \sup \|\Phi\| \cdot \sum_{k \in \mathbb{Z}: k\Delta_n \in [-M_\varepsilon, b]} \mathbb{E} \left[ |X_k^{(\Delta_n)}| \right] \\ &\leq \sup \|\Phi\| (\lceil M_\varepsilon + b \rceil) \mathbf{1}^T \nu \odot (\text{Id} - \tilde{K})^{-1} \mathbf{1} < \infty, \end{aligned}$$

where  $\sup \|\Phi\| < \infty$  because the kernel is assumed to be piecewise continuous, and we used the bound from Lemma B.6.2 by bounding the sum of expectations with the maximal number of elements in the sum times the expectation of each term.

By this, we can take  $N_\varepsilon^{(1)} \in \mathbb{N}_{\geq 0}$  large enough to get a bound by  $\varepsilon/2$ .

Also, the first term converges by noticing that we may use the bound from Lemma B.6.6:

$$\begin{aligned} & \left| \sum_{k \in \mathbb{Z}: k\Delta_n \in (a,b]} \Delta_n E_{n,M_\varepsilon}(k \cdot \Delta_n) - \int_a^b E_{n,M_\varepsilon}(t) dt \right| \\ & \leq \sum_{k \in \mathbb{Z}: k\Delta_n \in (a,b]} \int_{k\Delta_n}^{(k+1)\Delta_n} |E_{n,M_\varepsilon}(k \cdot \Delta_n) - E_{n,M_\varepsilon}(t)| dt \\ & \leq \sum_{k \in \mathbb{Z}: k\Delta_n \in (a,b]} \int_{k\Delta_n}^{(k+1)\Delta_n} \frac{\varepsilon}{2(b-a+\delta)} dt \\ & \leq \frac{b-a+\Delta_n}{\Delta_n} \Delta_n \frac{\varepsilon}{2(b-a+\delta)} \\ & \leq \frac{\varepsilon}{2}. \end{aligned}$$

We have used the bound for the number of elements in the sum in the first to the last line, as detailed in Lemma B.6.2.

In other words, if we set  $N_\varepsilon = \max(N_\varepsilon^{(1)}, N_\varepsilon^{(2)})$ , we get that for  $n > N_\varepsilon$ :

$$\left| \sum_{k \in \mathbb{Z}: k\Delta_n \in (a,b]} \Delta_n E_{n,M_\varepsilon}(k \cdot \Delta_n) - \int_a^b E_{M_\varepsilon}^*(t) dt \right| < \varepsilon. \quad (\text{B.5.10})$$

This is the bound we were looking for the first term.

### Second summand and third summand.

We consider the integrand of the second summand first. Using Lemma B.6.2 and taking the limit (see B.6.3), we have that  $\mathbb{E}[N_{dt}^*]/dt \leq \nu \odot (\text{Id} - K)^{-1} \mathbf{1}$  (component-wise). By taking  $M_\varepsilon^{(1)}$  large enough such that the norm of this vector is smaller than  $\varepsilon$ :  $\left| \int_{M_\varepsilon^{(1)}+a}^{\infty} \Phi(s) ds \cdot \nu \odot (\text{Id} - K)^{-1} \mathbf{1} \right| < \varepsilon/(b-a)$  ( $M_\varepsilon^{(1)}$  exists by continuity and integrability of the kernel, the fact that  $\text{Id} - K$  is invertible and  $\nu$  is finite), we get:

$$\begin{aligned} \left| \mathbb{E} \left[ \mathbb{1}_{A^*} \int_{-\infty}^{-M_\varepsilon^{(1)}} \Phi(t-s) dN_s^* \right] \right| &\leq \left| \mathbb{E} \left[ \int_{-\infty}^{-M_\varepsilon^{(1)}} \Phi(t-s) dN_s^* \right] \right| \\ &\leq \left| \int_{M_\varepsilon^{(1)}+a}^{\infty} \Phi(s) ds \cdot \nu \odot (\text{Id} - K)^{-1} \mathbf{1} \right| \\ &< \frac{\varepsilon}{b-a}. \end{aligned} \quad (\text{B.5.11})$$

Furthermore, we can also bound this term uniformly with respect to  $n$  by taking  $M_\varepsilon^{(2)}$  large enough

$$\mathbb{E} \left[ \Delta_n \mathbb{1}_{A_n} \int_{-\infty}^{-M_\varepsilon^{(2)}} \Phi(k \cdot \Delta_n - s) dN_n(s) \right] \leq \frac{\varepsilon}{\lceil b-a \rceil}. \quad (\text{B.5.12})$$

Integrating (B.5.11) and summing (B.5.12) over the right interval gives the desired  $\varepsilon$  bound of the second and third summand. By fixing  $M_\varepsilon = \max(M_\varepsilon^{(1)}, M_\varepsilon^{(2)})$ , we have a minimal value for the truncation of the tail such that the bound of the summands hold.

### Conclusion.

With these three summands bounded, whenever one takes an  $M > M_\varepsilon$  and an  $N > N_\varepsilon$ , one obtains that the difference between the RHS is less than  $3\varepsilon$ . In other words, we have proven that the RHS converge.

In conclusion, when  $\Delta \rightarrow 0$ , the point process  $N^{(\Delta)}$  converges weakly to a Hawkes process  $N^*$  with the parameter from the theorem. By unicity,  $N^*$  is the desired process  $N$ .

□

## B.6 Lemmas and Proofs

We now proceed with the lemmas useful for the main proofs.

**Theorem B.6.1** (Continuous Mapping Theorem). *Let us assume we have a family of point process and a point process such that  $N_n \xrightarrow[n \rightarrow \infty]{} N$ . We also assume that a bounded, measurable function taking values in  $\mathbb{R}^+$  with compact support and which discontinuity set  $D_f \in \mathcal{B}$ :  $\mathbb{P}[N(D_f) > 0] = 0$ . Then, we have that:*

$$\int f dN_n \xrightarrow[n \rightarrow \infty]{} \int f dN. \quad (\text{B.6.1})$$

*Proof.* The continuous mapping Theorem is rampant in weak convergence theory, and a good reference could be [177] Theorem 1.10 or in the book [19]. □

**Lemma B.6.2** (Bounds for the Expectation). *Let us assume we constructed the approximating sequence from (1.3.12),  $N^{(\Delta)}$ .*

*Then, we have these (in)equalities, using the approximating kernel  $K^{(\Delta)}$  from (1.3.15):*

$$\mathbb{E}[N^{(\Delta)}(\{k\Delta\})] = \Delta \cdot \nu \odot (\text{Id} - K^{(\Delta)})^{-1} \mathbf{1}, \quad (\text{B.6.2})$$

*and when  $a < b$ , there is a maximum of  $\lceil \frac{b-a}{\Delta} \rceil$  segments of length  $\Delta$  contained in a segment  $[a, b]$ , such that:*

$$\mathbb{E}[N^{(\Delta)}([a, b])] < (b-a+2\Delta) \cdot \nu \odot (\text{Id} - K^{(\Delta)})^{-1} \mathbf{1}. \quad (\text{B.6.3})$$

*In the case that the norm  $|\cdot|$  is the  $l^1$  norm, we directly have from the above:*

$$\mathbb{E} \left[ |N^{(\Delta)}(\{k\Delta\})| \right] = \Delta \mathbf{1}^T \nu \odot (\text{Id} - K^{(\Delta)})^{-1} \mathbf{1}. \quad (\text{B.6.4})$$

*Proof.* We use (1.3.12):  $N^{(\Delta)}(A) := \sum_{k \in \mathbb{N}_{\geq 0} : k\Delta \in A} X_k^{(\Delta)}$ , for  $A \in \mathcal{B}_b$ .

Remark that  $N^{(\Delta)}$  is nonzero almost surely only on sets including multiples of  $\Delta$ , the ticks when the process grows.

Then, the result follows by definition of the process and saying that the building block of  $N^{(\Delta)}$  is  $X^{(\Delta)}$  and its expectation by (1.3.5) is:  $\mathbb{E}[X^{(\Delta)}] = \Delta \cdot \nu \odot (\text{Id} - K^{(\Delta)})^{-1}\mathbf{1}$ .

For the second inequation, notice that the number of multiples of  $\Delta$  in  $[a, b]$  is less or equal to  $\lceil \frac{b-a}{\Delta} \rceil + 1 \leq (b-a)/\Delta + 2$ .  $\square$

**Proposition B.6.3** ( $(N^{(\Delta)}(A))_{\Delta \in (0, \delta)}$  is Uniformly Integrable). *Let us assume we constructed the approximating sequence of (1.3.12),  $N^{(\Delta)}$ .*

*Then for all  $A \in \mathcal{B}_b$ ,  $(N^{(\Delta)}(A))_{\Delta \in (0, \delta)}$  is uniformly integrable.*

*Proof.* In [67], we see a generalisation of Markov inequality to  $\mathbb{R}^d$ .

The sequence is uniformly integrable by the following argument. We can take  $A$  closed without loss of generality, since  $|N^{(\Delta)}(A)| < |N^{(\Delta)}(\bar{A})|$ . Considering a compact interval  $A$ , we may write it  $[a, b]$ . We set the constant  $M_\varepsilon = \varepsilon^{-1}(b-a+2\delta)\mathbf{1}^T\nu \odot (\text{Id} - \tilde{K})^{-1}\mathbf{1}$ . Then:

$$\begin{aligned} \mathbb{P}\left(\left|N^{(\Delta)}([a, b])\right| > M_\varepsilon\right) &\leq M_\varepsilon^{-1}\mathbb{E}\left[\left|N^{(\Delta)}([a, b])\right|\right] \\ &< M_\varepsilon^{-1}(b-a+2\Delta)\mathbf{1}^T\nu \odot (\text{Id} - K(\Delta))^{-1}\mathbf{1} \\ &< M_\varepsilon^{-1}(b-a+2\delta)\mathbf{1}^T\nu \odot (\text{Id} - \tilde{K})^{-1}\mathbf{1} \\ &= \varepsilon. \end{aligned}$$

From line 1 to 2, we used the Lemma B.6.2 to bound the expectation.

An alternative proof would be using the  $L^2$  bound we found for the sequence in Lemma B.6.5.  $\square$

**Lemma B.6.4** (The Family  $N^{(\Delta)}$  is Tight). *Let us assume we constructed the approximating sequence of (1.3.12),  $N^{(\Delta)}$ .*

*Then the corresponding family of the probability measures  $(\mathbb{P}^{(\Delta)})_{0 < \Delta < \delta}$  on  $(M_p, \sigma(\mathcal{M}_p))$  is uniformly tight.*

*Proof.* The claim follows from Proposition 11.1.VI in [50] and by proving that we have:

$$\forall A \in \mathcal{B}_b, \forall \varepsilon > 0, \exists M < \infty : \sup_{0 < \Delta < \delta} \mathbb{P}(N^{(\Delta)}(A) > M) < \varepsilon, \quad (\text{B.6.5})$$

which is direct since the interval  $A$  is compact.  $\square$

**Lemma B.6.5** (Bounded  $L^2$  Norm). *Let  $M > 0, t \in \mathbb{R}^+$ , a sequence of INVAR( $\infty$ ) processes  $N_n$  and  $\Phi$  their kernel (as in the proof of Theorem 1.3.13).*

*Then the random variable  $\int_{-M}^t \Phi(t-s) dN_n(s)$  has bounded  $L^2$  norm:*

$$\exists K_{constant} > 0 : \mathbb{E}\left[\left|\int_{-M}^t \Phi(t-s) dN_n(s)\right|^2\right]^{1/2} < K_{constant}. \quad (\text{B.6.6})$$

*The bound  $K_{constant}$  is independent of  $n$  and  $\Delta_n$ .*

*Proof.* We want to bound  $\mathbb{E}\left[\left|\int_{-M}^t \Phi(t-s) dN_n(s)\right|^2\right]$ . We prove it for the  $l_2$  norm  $|\cdot|$ .

First, we use that  $N_n([a, b])$  can be rewritten:  $\sum_{k \in \mathbb{Z} : k\Delta_n \in (a, b]} X_k^{(\Delta_n)}$ .

$$\begin{aligned} \mathbb{E}\left[\left|\int_{-M}^t \Phi(t-s) dN_n(s)\right|^2\right] &= \mathbb{E}\left[\left|\sum_{k \in \mathbb{Z} : k\Delta_n \in (-M, t]} \Phi(k\Delta_n) X_k^{(\Delta_n)}\right|^2\right] \\ &\leq \mathbb{E}\left[\left|\sum_{k=1}^{\lceil (M+t)/\Delta_n \rceil} \Phi(k\Delta_n) X_{-k}^{(\Delta_n)}\right|^2\right]. \end{aligned}$$

Using the fact that the norm is the  $l_2$  norm:

$$\begin{aligned}\mathbb{E} \left[ \left| \int_{-M}^t \Phi(t-s) dN_n(s) \right|^2 \right] &\leq \mathbb{E} \left[ \left| \sum_{k=1}^{\lceil (M+t)/\Delta_n \rceil} \Phi(k\Delta_n) X_{-k}^{(\Delta_n)} \right|^2 \right] \\ &= \sum_{i=1}^d \mathbb{E} \left[ \sum_{k=1}^{\lceil (M+t)/\Delta_n \rceil} ((\Phi(k\Delta_n) X_{-k}^{(\Delta_n)})^i)^2 \right].\end{aligned}$$

Now, we bound the second moment by the variance, and then distribute the variance (the random variable inside the expectation is one dimensional) on each term of the inside sum:

$$\begin{aligned}\mathbb{E} \left[ \left| \int_{-M}^t \Phi(t-s) dN_n(s) \right|^2 \right] &\leq \sum_{i=1}^d \text{Var} \left[ \sum_{k=1}^{\lceil (M+t)/\Delta_n \rceil} (\Phi(k\Delta_n) X_{-k}^{(\Delta_n)})^i \right] \\ &\leq \sum_{i=1}^d \sum_{k=1}^{\lceil (M+t)/\Delta_n \rceil} \sum_{l=1}^{\lceil (M+t)/\Delta_n \rceil} \sup \|\Phi\|^2 \text{Cov}((X_{-k}^{(\Delta_n)})^i, (X_{-l}^{(\Delta_n)})^i).\end{aligned}$$

On the last line, we identify the autocovariance sequence coined here  $R^{(\Delta_n)}$  of an INVAR( $\infty$ ) process that we have previously derived in Theorem 1.3.12:

$$\mathbb{E} \left[ \left| \int_{-M}^t \Phi(t-s) dN_n(s) \right|^2 \right] \leq \sup \|\Phi\|^2 \sum_{i=1}^d \sum_{k=1}^{\lceil (M+t)/\Delta_n \rceil} \sum_{l=1}^{\lceil (M+t)/\Delta_n \rceil} (R^{(\Delta_n)}(|l-k|))_{i,i}.$$

Now, we add to the sum  $\sum_{l=1}^{\lceil (M+t)/\Delta_n \rceil} (R^{(\Delta_n)}(|l-m|))_{i,i}$  terms so the sum index on  $l$  covers the whole integer axis  $\mathbb{Z}$ :  $\sum_{l=1}^{\lceil (M+t)/\Delta_n \rceil} (R^{(\Delta_n)}(|l-m|))_{i,i} \leq \sum_{l \in \mathbb{Z}} (R^{(\Delta_n)}(l))_{i,i}$ , which is justified since the autocovariance matrices having only positive values, independently of time. Finally, we may bound roughly  $\sum_{l \in \mathbb{Z}} (R^{(\Delta_n)}(l))_{i,i}$  by  $\|\sum_{l \in \mathbb{Z}} (R^{(\Delta_n)}(l))\|$  for any matrix norm using Theorem 1.3.12 (the norm is always bigger or equal to one of the coordinates, by triangular inequality).

In conclusion, by using the bound on the norm of the sum of the autocovariance sequence from Theorem 1.3.12 (coined  $C_{R^{(\Delta_n)}} = \|(\text{Id} - K)^{-1} \mathbf{1}\|^3 \|\text{Diag}(\nu)\|$ ):

$$\begin{aligned}\mathbb{E} \left[ \left| \int_{-M}^t \Phi(t-s) dN_n(s) \right|^2 \right] &\leq \sup \|\Phi\|^2 \sum_{i=1}^d \sum_{k=1}^{\lceil (M+t)/\Delta_n \rceil} \left\| \sum_{l \in \mathbb{Z}} (R^{(\Delta_n)}(l)) \right\| \\ &\leq d \cdot \sup \|\Phi\|^2 \Delta_n \cdot C_{R^{(\Delta_n)}} \left( \frac{M+t}{\Delta_n} + 1 \right) \\ &\leq d \cdot \sup \|\Phi\|^2 C_{R^{(\Delta_n)}} (M+t+\delta) < \infty,\end{aligned}$$

and this bound is independent of  $n$  and  $\Delta_n$ . □

**Lemma B.6.6** (Uniform Continuity Bound for the Theorem). *In the proof of Theorem 1.3.13, we assume that the kernel is uniformly continuous. Then, we get the following bound:*

$$\forall a \leq s < t \leq b, t-s < \Delta_n : |E_{n,M_\varepsilon}(t) - E_{n,M_\varepsilon}(s)| \leq \frac{\varepsilon}{2(b-a+\delta)}. \quad (\text{B.6.7})$$

$\delta$  here is the constant set at the beginning of the proof of Theorem 1.3.13.

*Proof.* Uniform continuity grants us a constant  $\delta_{\text{u.c.}} > 0$ , which gives for any  $t_0 > 0$ ,

$$|t-t_0| < \delta_{\text{u.c.}} \wedge t \Rightarrow \|\Phi(t) - \Phi(t_0)\| < \frac{\varepsilon}{4(b-a+\delta)(M_\varepsilon+b+\delta)} \frac{1}{\mathbf{1}^T \nu \odot (\text{Id} - \tilde{K})^{-1} \mathbf{1}}.$$

Now, choose  $N_\varepsilon^{(2)}$  large such that,  $\forall n \geq N_\varepsilon^{(2)}$  :

$$\Delta_n < \min \left\{ \delta_{\text{u.c.}}, \frac{\varepsilon}{4(b-a+\delta) \cdot \sup \|\Phi\|} \frac{1}{\mathbf{1}^T \nu \odot (\text{Id} - \tilde{K})^{-1} \mathbf{1}} \right\}.$$

Bounding  $\Delta_n$  by  $\delta_{\text{u.c.}}$  is purely cosmetic and is not required in the rest of the proof. Let  $a \leq s < t \leq b$  with  $t-s < \Delta_n < \delta_{\text{u.c.}}$ , then

$$\begin{aligned} |E_{n,M_\varepsilon}(t) - E_{n,M_\varepsilon}(s)| &= \left| \sum_{k\Delta_n \in [-M_\varepsilon, t[} \mathbb{E} \left[ 1_{A_n} \Phi(t - k\Delta_n) X_k^{(\Delta_n)} \right] - \sum_{k\Delta_n \in [-M_\varepsilon, s[} \mathbb{E} \left[ 1_{A_n} \Phi(s - k\Delta_n) X_k^{(\Delta_n)} \right] \right| \\ &\leq \left( \sum_{k\Delta_n \in [-M_\varepsilon, s]} \|\Phi(t - k\Delta_n) - \Phi(s - k\Delta_n)\| + \sum_{k\Delta_n \in ]s, t[} \|\Phi(t - k\Delta_n)\| \right) \mathbb{E} \left[ |X_0^{(\Delta_n)}| \right] \\ &\leq \left( \frac{M_\varepsilon + s + \Delta_n}{\Delta_n} \frac{\varepsilon}{4(b-a+\delta)(M_\varepsilon + b + \delta)} \frac{1}{\nu \odot (\text{Id} - \tilde{K})^{-1} \mathbf{1}} + \sup \|\Phi\| \right) \\ &\quad \cdot \Delta_n \cdot \mathbf{1}^T \nu \odot (\text{Id} - K^{(\Delta)})^{-1} \mathbf{1} \\ &= \frac{\varepsilon}{4(b-a+\delta)} \frac{M_\varepsilon + s + \Delta_n}{M_\varepsilon + b + \delta} \frac{\mathbf{1}^T \nu \odot (\text{Id} - K^{(\Delta)})^{-1} \mathbf{1}}{\mathbf{1}^T \nu \odot (\text{Id} - \tilde{K})^{-1} \mathbf{1}} \\ &\quad + \Delta_n \cdot \sup \|\Phi\| \mathbf{1}^T \nu \odot (\text{Id} - K^{(\Delta)})^{-1} \mathbf{1} \\ &\leq \frac{\varepsilon}{4(b-a+\delta)} + \Delta_n \cdot \sup \|\Phi\| \mathbf{1}^T \nu \odot (\text{Id} - \tilde{K})^{-1} \mathbf{1} \\ &\leq \frac{\varepsilon}{4(b-a+\delta)} + \frac{\varepsilon}{4(b-a+\delta)} \\ &= \frac{\varepsilon}{2(b-a+\delta)}. \end{aligned}$$

The sup of the norm of  $\Phi$  is bounded because it is a continuous function.

We have used, from the first line to the second the linearity of the expectation,  $\forall k, \tilde{k} \in \mathbb{N}_{\geq 0} : \mathbb{E}[X_k^{(\Delta_n)}] = \mathbb{E}[X_{\tilde{k}}^{(\Delta_n)}] = \mathbb{E}[X_0^{(\Delta_n)}]$ , we factorised by this common expectation and finally we bounded the norm of the sum of the expectations by the product of the matrix norm (of the kernel  $\Phi$ ) and the expectation of the norm of  $X_0^{(\Delta_n)}$  (by consistency of the norm  $|\cdot|$ ). Then, to pass from line 2 to line 3 we have used that: since  $t-s < \Delta_n$ , the second sum over elements in  $]s, t[$  is a simple term. Furthermore, there can only be  $\frac{M_\varepsilon + s + \Delta_n}{\Delta_n}$  elements in the first sum (Lemma B.6.2) and we multiply this quantity by the uniform continuity bound.

To be able to write the inverse of  $\mathbf{1}^T \nu \odot (\text{Id} - \tilde{K})^{-1} \mathbf{1}$ , we need the matrix  $1 - \tilde{K}$  to be invertible, but also that at least one component of  $\nu$  is nonzero, which is an assumption of the original model.

We bounded both  $\frac{M_\varepsilon + s + \Delta_n}{M_\varepsilon + b + \delta}$  and  $\frac{\mathbf{1}^T \nu \odot (\text{Id} - \tilde{K})^{-1} \mathbf{1}}{\mathbf{1}^T \nu \odot (\text{Id} - K^{(\Delta)})^{-1} \mathbf{1}}$  by 1, thanks to the assumptions on the constants, namely:

- $\Delta_n < \delta$ ,
- $\rho(\tilde{K}) \leq \rho(K^{(\Delta)})$ .

□

**Lemma B.6.7** (Weak Convergence Gives Convergence of Means). *If the sequence of  $d$ -dimensional random vectors  $\{X_n\}_{n \in \mathbb{Z}}$  are uniformly integrable and we have the weak convergence of this sequence to some random vector  $X^*$ , then we have convergence of the mean:  $\lim_{n \rightarrow \infty} E[X_n] = E[X^*]$ .*

*Proof.* We prove this by relying on the existing proof for the one-dimensional case from [19] (Theorem 3.5). We assume it is true in the one-dimensional case.

Then, we can say that for each of the  $d$  components of the vector  $X_n$ , separately, the original assumptions hold (the uniform integrability is direct and weak convergence holds by the continuous mapping Theorem for example). In other words, we get that:

$$\forall i \in \{1, \dots, d\} : \lim_{n \rightarrow \infty} \mathbb{E}[X_n^i] = \mathbb{E}[X^{*,i}],$$

where  $i$  represents the index of the vector. This is equivalent to the convergence of the whole vector, as desired:

$$\lim_{n \rightarrow \infty} \mathbb{E}[X_n] = \mathbb{E}[X].$$

□

## APPENDIX

### C

### EM ALGORITHM

#### C.1 Multidimensional Equations

The EM algorithm exists in many variants, and one can find the equations for estimating the parameters of the Hawkes process in [206] and of the univariate INAR process in [208].

To simplify the notation and improve clarity, we assume that the number of bins is equal across all dimensions. We denote the number of bins as  $B$ , where  $B$  is a positive integer. Suppose we have access to a multidimensional time series of length  $T$  that is divisible by a time interval  $\Delta$ . We can then define the number of bins as  $B \triangleq T/\Delta$ . Using the branching structure of the INVAR, as defined in Definition 1.3.5, we can decompose the bin counts as follows for  $l \in \{1, \dots, B\}$ :

$$\begin{aligned} X_l^m &= I_l^m + \sum_{n=1}^d O_l^{m,n}, \quad m \in \{1, \dots, d\}, \\ I_l^m &\sim \text{Pois}(\nu_l^m), \quad m \in \{1, \dots, d\}, \\ O_l^{m,n} &\sim \text{Pois}\left(\sum_{k=1}^{l-1} \Phi_{l-k}^{m,n} X_k^n\right), \quad m, n \in \{1, \dots, d\}. \end{aligned} \tag{C.1.1}$$

The two processes  $I^m, O^{m,n}$  are unobserved and will be considered as our latent variables completing the knowledge of the original model.

The E-step requires us to compute the weights, for  $l \in \{1, \dots, B\}, m, n \in \{1, \dots, d\}$ :

$$\pi_l^{I,m} := \mathbb{E}[I_l^m | X_l, \theta] = X_l^m \frac{\nu_l^m}{\lambda_l^m}, \tag{C.1.2}$$

$$\pi_l^{O,m,n} := \mathbb{E}[O_l^{m,n} | X_l, \theta] = X_l^m \frac{\sum_{k=1}^{l-1} \Phi_{l-k}^{m,n} X_k^n}{\lambda_l^m}, \tag{C.1.3}$$

granted by the fact that when conditioned on the realised bin count, each of the random variable can be viewed as a binomial random variable with appropriate parameter.

Notice that they can be assimilated to probabilities when normalised by the individual bin count.

All of the processes are independent and hence, the complete data likelihood called  $L_c$  is by definition:

$$L_c(\theta | X, I, O) = \prod_{m=1}^d \left[ \prod_{l=1}^B p_{I^m}(I_l^m) \prod_{n=1}^d p_{O^{m,n}}(O_l^{m,n}) \right], \tag{C.1.4}$$

where  $p_{I^m}, p_{O^{m,n}}$  are respectively the probability mass functions of  $I^m, O^{m,n}$  for all  $m, n \in \{1, \dots, d\}$ . Then, the complete data likelihood reads:

$$\log L_c(\theta | X, I, O) = l_{c,\nu} + l_{c,\Phi},$$

$$l_{c,\nu} = \sum_{m=1}^d \sum_{l=1}^B I_l^m \log(\nu_l^m) - \nu_l^m - \log(I_l^m!), \quad (\text{C.1.5})$$

$$l_{c,\Phi} = \sum_{m=1}^d \sum_{n=1}^d \sum_{l=1}^B \left[ O_l^{m,n} \log \left( \sum_{k=1}^{l-1} \Phi_{l-k}^{m,n} X_k^n \right) - \sum_{k=1}^{l-1} \Phi_{l-k}^{m,n} X_k^n - \log(O_l^{m,n}!) \right]$$

$$= \sum_{m=1}^d \sum_{n=1}^d \sum_{l=1}^B \left[ O_l^{m,n} \log \left( \sum_{k=1}^{l-1} \Phi_{l-k}^{m,n} X_k^n \right) - \log(O_l^{m,n}!) - \Phi_{B-l}^{m,n} X_l^n \right]. \quad (\text{C.1.6})$$

We can maximise the two parts of the likelihood independently, since they rely on different sets of parameters.

This gives us the objective function to be maximised, where we omitted the constant terms:

$$\mathbb{E}_{I,O|\hat{\theta},X}[\log L_c] = \sum_{m=1}^d \sum_{l=1}^B \pi_l^{I,m} \log(\nu_l^m) - \nu_l^m + \sum_{m=1}^d \sum_{n=1}^d \sum_{l=1}^B \left[ \pi_l^{O,m,n} \log \left( \sum_{k=1}^{l-1} \Phi_{l-k}^{m,n} X_k^n \right) - \Phi_{B-l}^{m,n} X_l^n \right]. \quad (\text{C.1.7})$$

We use the notation  $\Phi = \eta \odot h$  for  $h$  the corresponding normalised kernel, and  $\Phi = \eta \odot H$  the cumulative kernel. Then, the M-step yields the estimates for the parameters. The exogenous rate and endogenous rate (branching ratio) can be solved in closed form and yield, assuming constant value through time:

$$\hat{\nu}_k^m \equiv \sum_{l=1}^B \pi_l^{I,m} / B, \quad (\text{C.1.8})$$

$$\hat{\eta}_k^{m,n} \equiv \frac{\sum_{l=1}^B \pi_l^{O,m,n}}{\sum_{l=1}^B H_{B-l}^{m,n} X_l^n}. \quad (\text{C.1.9})$$

Each iteration is composed of the E-step:

- Compute this iteration's estimated intensity function,
- Compute the weights  $\pi_l^{I,O}$ ,

and of the M-step:

- Estimate the kernel by maximising the likelihood function for either a parametric or non-parametric choice,
- Calculate the endogenous and exogenous rates.

A choice for the discrete offspring kernel in analogy to the point process setup is thus for example based on geometric offspring:

$$h_j = (1-p)^j p, \quad 0 < j \in \mathbb{N}_{\geq 0}. \quad (\text{C.1.10})$$

## C.2 Truncating the Kernel

By truncating the kernel, we can reduce the memory length of the kernel used in the equations. We do that by assuming that  $O_l^{m,n}$  is impacted only by a maximum of the previous  $p$  steps:

$$O_l^{m,n} \sim \text{Pois} \left( \sum_{k=1}^{l-1 \wedge p} \Phi_k^{m,n} X_{l-k}^n \right), \quad m, n \in \{1, \dots, d\}.$$

The algorithm's performance are improved by truncating the kernel, because the sum of the kernel values is computationally intensive and the sum is the bottleneck in the computations. By truncating the kernel, we can reduce the scaling factor from  $B$  to  $p$ , which enhances the algorithm's efficiency.

## APPENDIX

# D

## ADDITIONAL DISCUSSIONS FOR THE TYPE-II MODEL

### D.1 Stopping Time and Regime Process

#### D.1.1 Mathematical Framework of the Regime Process

We continue the discussion related to the choice of stopping times that define the drawdown and drawup phases, as introduced in Subsection 3.2.9. If we define the drawdown regime through the stopping time as  $\tau_{\text{drawdown}} = \max(u \in [0, T] : S_u < (1 - \varepsilon)S_{u-\Delta_\varepsilon})$ , where  $\Delta_\varepsilon$  is the length for retrospective comparison,  $\varepsilon$  is a chosen lag parameter, we argue that this definition may miss some smoother crashes that do not occur in a single big jump.

If we denote respectively by 'Drawup Phase' and by 'Drawdown Phase' the times to be respectively in a bubble and in a drawdown phase, we may rewrite them as the sets of the times  $t$  respecting:

$$\text{Drawup Phase} := \{\tau_m(t, {}^t\Delta) < \tau_M(t, {}^t\Delta)\}_{t \in \mathbb{R}}, \quad (\text{D.1.1})$$

$$\text{Drawdown Phase} := \{\tau_M(t, {}^t\Delta) < \tau_m(t, {}^t\Delta)\}_{t \in \mathbb{R}}. \quad (\text{D.1.2})$$

The definition of stopping time using a symmetric window around time  $t$ , is shown in [73] to impose restrictive conditions on the price increments surrounding the extrema. Specifically, the distribution of returns before and after the extrema is positively and negatively skewed, respectively, independent of the underlying generating process. Consequently, this definition sets critical model characteristics at the point where the two regimes switch. Our study employs a one-sided window, which is expected to introduce less bias.

In addition to the concept of local extrema, we define regimes using the notion of abnormal price difference, a highly intuitive idea with a natural interpretation for the space threshold  ${}^s\Delta$ . When the price process follows a geometric Brownian motion, time and space (or price) are related. Assuming our model is  $\frac{dS_t}{S_t} = \mu dt + \sigma dW_t$ ,  $\{W_t\}_{t \in \mathbb{R}_+}$  a Brownian motion. Solving the GBM equation yields  $\ln(S_t) = (\mu - \frac{\sigma^2}{2})T + \sigma\sqrt{T}\mathcal{N}(0, 1)$ , where  $\mathcal{N}(0, 1)$  is a standard normal random variable. We ignore the first term<sup>1</sup> and observe that a reasonable threshold would be  $\sigma\sqrt{T}$  (one standard deviation away)<sup>2</sup>. We define  $\sigma\sqrt{{}^t\Delta^j} =: {}^s\Delta^j, j \in \{1, \dots, \delta_{\text{scales}}\}$  as

<sup>1</sup>Why can we ignore it? If the horizon is small, then the first term is dominated by the second. For longer horizons, however, it might be necessary to discount the price so the result is not corrupted.

<sup>2</sup>Here we are hiding some assumptions. For example,  $\sigma$  should not be too big, nor  $\mu$ . Furthermore,  $T$  should also not be too big, or the first term  $(\mu - \frac{\sigma^2}{2})T$  would dominate  $\sigma\sqrt{T}\mathcal{N}(0, 1)$ . See more details in [191].

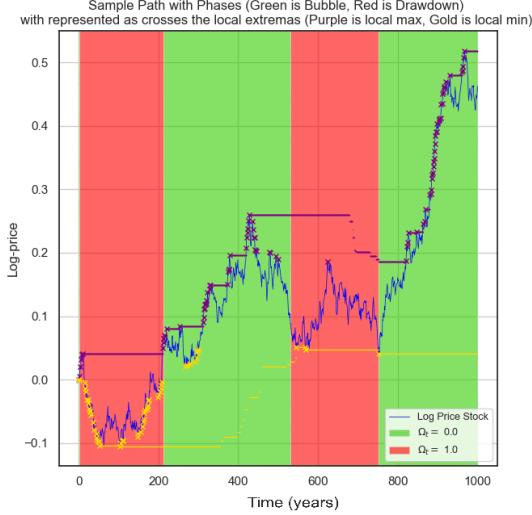


Figure D.1: We consider a simple geometric Brownian motion trajectory in log price, where the regime does not affect the trajectory, to present an analysis of how the stopping times  $\tau_m$  and  $\tau_M$  evolve over time. We set  $t\Delta = 250$  and plot the stock price as a blue line with the background colored green and red to represent the drawup and drawdown regimes. We also mark the extrema with gold and purple crosses for  $\tau_m$  and  $\tau_M$ , respectively. To indicate the time interval over which the stopping time represents a local extremum, we draw a straight line at the same height as the marker and of the same color as the type of extremum. At around time 50, we observe that the maximum is taken from the beginning of the time series (from time 5), but the minima are from later times during the drawdown regime. When the market has increased sufficiently (at time 210),  $\tau_M$  comes back in front of the last  $\tau_m$  from time 51, triggering a regime switch on the following day.

a "space" threshold, in contrast to our previous time threshold.

The standard deviation over a period of time  $t\Delta^j$ -long is  $\sigma\sqrt{t\Delta^j}$  for a daily standard deviation  $\sigma$ . For  $\sigma = 0.01$  and  $t\Delta^1 = 100$  and  $t\Delta^2 = 250$ , we obtain two space scales,  $s\Delta^1 = 0.1$  and  $s\Delta^2 = 0.15$ . If we are in a drawup, the threshold for the decrease would be a drop of approximately 10% and 15%, respectively. These values appear to be reasonable for establishing a threshold at which the market changes regime.

### D.1.2 Phony Successive Changes of Regimes

In some specific scenarios, we observe successive changes in regimes. This is considered as a feature of our regime-switching model since it is very natural that there are periods of incertitude when investors do not know in what regime the market lies. Successive regime changes occur when the price remains stagnant for the duration  $t\Delta$ , and minor price movements can alter the ordering of the minimum and maximum values which trigger regime changes. The threshold induced by  $s\Delta$  accentuates this phenomenon. It is known that for a Brownian motion  $B_t$ , starting at the value  $x$  at time 0, and for all  $\varepsilon > 0$ , there exists infinitely many times  $t$  in the ball  $(0, \varepsilon)$  at which  $B_t = 0$ . This implies that when the stock price reaches the limit (the threshold  $s\Delta$ ), and because of the Brownian motion component of our price, we expect the price to change regimes multiple times. We illustrate this mechanism in Fig. D.2.

### D.1.3 Bubbly Market Regimes

Then, suppose we want to work with a market involving multiple bubbles, in the spirit of our definition of bubbly market, see Subsection 3.2.3. In that case, we could define the sequence of stopping times representing the beginning of each drawdown/drawup regime respectively as:

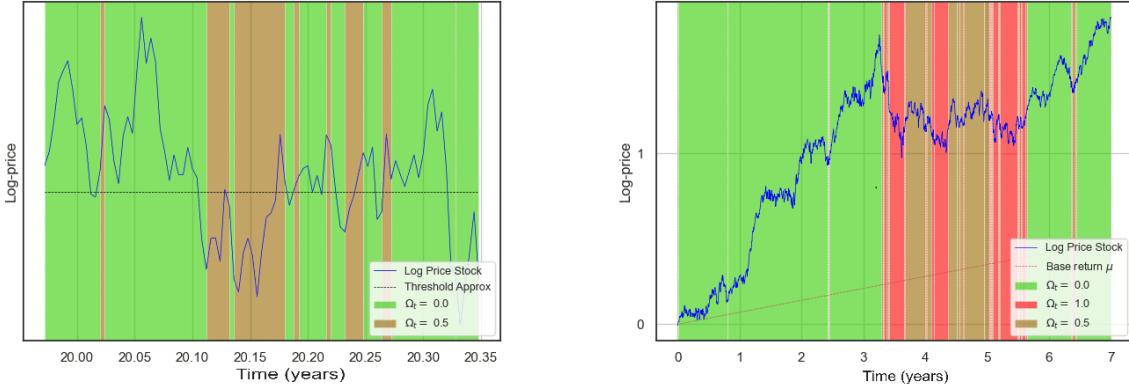


Figure D.2: The impact of the threshold  ${}^s\Delta$  on the regime-switching model is illustrated in the two plots of the figure. The plots are color-coded with green and red to indicate drawup and drawdown regimes, respectively. The left plot exhibits the "regime hesitation" phenomenon that occurs when the stock price approaches the threshold value, resulting in numerous regime changes quickly arbitrarily. This is clearly observed at time 20.12 when a slight increase in price triggers a regime switch. The right plot showcases the efficacy of the multi-scale mechanism in identifying the lows of the price, leading to regime switches depending on the scale of the price drop. Notably, the regime switching dynamics accurately identifies micro-bubbles (small-scale drawup phases within a larger-scale drawdown), as evidenced by regime switches at times 3.6 and 4.4. There are also brief instances of regime hesitation observed at times 3.3 and 5.5 when the asset's price decreases.

$$\tau_{\text{drawdown}}^i({}^t\Delta) := \inf\{t > \tau_{\text{drawdown}}^{i-1} : \tau_M(t, {}^t\Delta) < \tau_m(t, {}^t\Delta)\}, i \in \mathbb{N}_{>0}, \quad \tau_{\text{drawdown}}^0 := -1, \quad (\text{D.1.3})$$

$$\tau_{\text{drawup}}^i({}^t\Delta) := \inf\{t > \tau_{\text{drawup}}^{i-1} : \tau_m(t, {}^t\Delta) < \tau_M(t, {}^t\Delta)\}, i \in \mathbb{N}_{>0}, \quad \tau_{\text{drawup}}^0 := 0. \quad (\text{D.1.4})$$

We highlight that  $\tau_{\text{bubble}}^i$  and  $\tau_{\text{drawdown}}^i$  are stopping times for all  $i \in \mathbb{N}_{\geq 0}$ . The rationale behind setting the first  $\tau_{\text{drawdown}}^0$  at a negative time is purely practical for the dynamics to start in a drawup regime (the process starts at 0).

#### D.1.4 Multi-Scale Regime Switching

It is widely recognized that market practitioners rely on historical data to identify the level of mispricing, and therefore, a delayed stopping time approach seems appropriate. However, the precise identification of the transition point from a bubble to a drawdown regime seems artificial, as having a single stopping time suggests that all market participants synchronously re-evaluate the price of assets. Moreover, different institutions and organisms exhibit varying degrees of sensitivity to market movements. High-frequency traders, for example, react swiftly to changes in the market, while larger institutional asset managers might take longer to react. Furthermore, institutions often employ a hierarchy of different anchors to evaluate trends and patterns.

To address these concerns, we propose a multi-scale regime-switching mechanism that incorporates a hierarchy of time and space scales to account for varying degrees of sensitivity to changes in the price. Specifically, we fix a number of time scales  $\delta_{\text{scales}}$  and associate them with time scales  ${}^t\Delta^j_{j \in \{1, \dots, \delta_{\text{scales}}\}}$  and space scales  ${}^s\Delta^j_{j \in \{1, \dots, \delta_{\text{scales}}\}}$ . The joint determination of these scales provides a more nuanced approach to identifying drawdown phases, taking into account the diverse perspectives and sensitivities of market participants.

More rigorously, we define the following objects:

- A list of windows  $\{{}^t\Delta^j\}_{j \in \{1, \dots, \delta_{\text{scales}}\}}$ , that for convenience we assume ordered with  ${}^t\Delta^1$  being the smallest window,

- A list of space windows  $\{{}^s\Delta^j\}_{j \in \{1, \dots, \delta_{\text{scales}}\}}$ , each related to one of the above windows,
- Two series of sequences of stopping time  $\{\tau_m(t, {}^t\Delta^j)\}_{j \in \{1, \dots, \delta_{\text{scales}}\}}$ ,  $\{\tau_M(t, {}^t\Delta^j)\}_{j \in \{1, \dots, \delta_{\text{scales}}\}}$  associated to each window, see the original definition in (3.2.10) and (3.2.12). For the sake of clarity and notation conciseness, we will coin  $\tau_m(t, {}^t\Delta^j) =: \tau_{m,j}(t)$  and  $\tau_M(t, {}^t\Delta^j) =: \tau_{M,j}(t)$ .

Then, we define the regimes in the same way we defined the two phases for one time scale, in the sense that we leverage the ordering of extrema and count how many of these pairs are ordered, while considering the space thresholds as well. We weight each pair's contribution to the total regime value. We define the stochastic processes  $\{\{A_t^j\}_{j \in \{1, \dots, \delta_{\text{scales}}\}}\}_{t \in \mathbb{R}_+}$  taking the value 0 or 1 depending on the previously introduced conditions, and jump between the two values when the conditions are met. We call them flag processes. Then, we propose the following adapted regime process  $\{\Omega_t\}_{t \in \mathbb{R}_+}$  defined using the sum of flag processes:

$$\Omega_t = \sum_{j=1}^{\delta_{\text{scales}}} \omega^j A_t^j. \quad (\text{D.1.5})$$

Here, the  $\{\omega^j\}_{j \in \{1, \dots, \delta_{\text{scales}}\}}$ ,  $\sum_j \omega^j = 1$  are the weights associated to each time scale, that we can choose for simplicity to be  $\omega^j = 1/\delta_{\text{scales}}$ , for all  $j \in \{1, \dots, \delta_{\text{scales}}\}$ . To conclude, we summarise the algorithm for computing the values of the  $\{\{A_t^j\}_{j \in \{1, \dots, \delta_{\text{scales}}\}}\}_{t \in \mathbb{R}_+}$  in Algorithm 1.

We represent in Fig. D.3 what is visually the mechanism behind the regime-switching in a multi-scale framework. The higher the value of  $\Omega_t$ , the more "in-the-drawdown" we are. The two special cases are  $\Omega_t = 0$  when the process is completely in a drawup regime and  $\Omega_t = 1$  when the process is in a drawdown regime.

We believe that using a hierarchy of time scales to define the regime process is the correct one. We propose interpreting the parameter  $\Omega$  as a percentage or probability of being in a drawdown phase, with higher values indicating greater rationality among market participants. Additionally, the weights  $\omega^j$  represent the impact of market participants who trade at a specific time scale on the global pessimism in the markets.

Our experiments suggest that it did not improve the performance of the studied model due to its original flaws. Nonetheless, we believe that the proposed approach offers a novel perspective on defining regime processes in financial markets and may be worthy of further investigation in future studies.

---

**Algorithm 1:** Multi-Scale Regime Process  $\Omega_t$  Calculation

---

```

1 We want to compute  $\Omega_t$ , for a fixed  $t$ .
2 for  $j \in \{1, \dots, \delta_{\text{scales}}\}$  do
3   Compute  $\tau_{m,j}(t)$  and  $\tau_{M,j}(t)$  as defined in (3.2.12) and defined in (3.2.10), and the extremum
    $m^j(t), M^j(t)$ .
4   if  $\tau_{M,j}(t) < \tau_{m,j}(t)$  then
5     if  $\ln(S_t) - \ln(m^j(t)) \leq {}^s\Delta^j$  then
6       |  $A_t^j = 1$ 
7     else
8       |  $A_t^j = 0$ 
9     end
10   else
11     if  $\ln(M^j(t)) - \ln(S_t) \geq {}^s\Delta^j$  then
12       |  $A_t^j = 1$ 
13     else
14       |  $A_t^j = 0$ 
15     end
16   end
17 end
18 Compute  $\Omega_t$  by using (D.1.5).

```

---

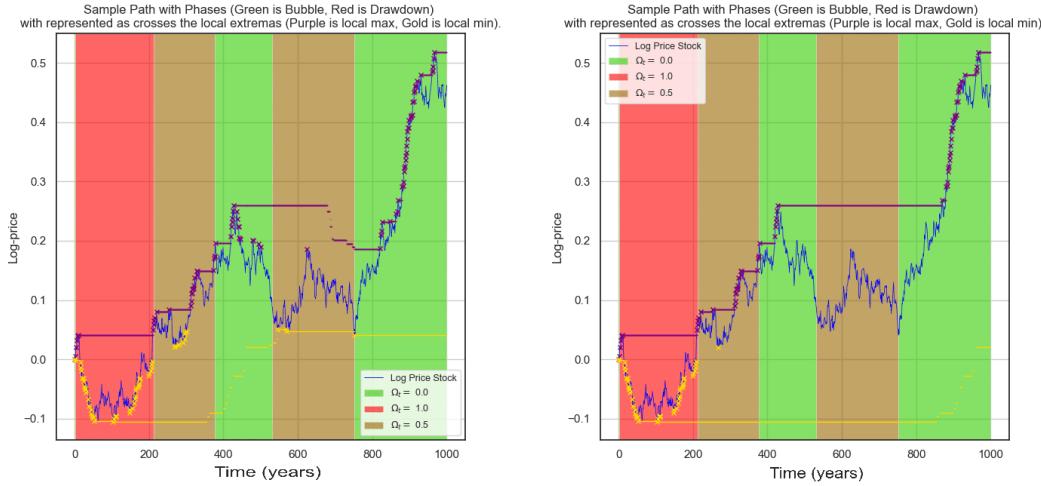


Figure D.3: Following the format of Figure D.1, we present a visual explanation of the multi-scale regime. We consider a simple geometric Brownian motion trajectory in log price, where the regime does not affect the trajectory. The plots serve to demonstrate the regime-switching mechanism of a bubble model with two time scales. The left plot displays the stopping times associated with  $\Delta^1 = 250$ , while the right plot displays those associated with  $\Delta^2 = 750$ .

## D.2 Hindsight Bias

### D.2.1 What is the Realism of the Investors?

In Subsection 3.2.8, we observed that drawup phases make the price deviate from its trend due to the additional upward jumps. The identified problem is the lack of crash in the upward jump component to bring the price process back to a fair level. The growth during drawups does not seem to be a problem and could be an interesting feature of our model, but for the lack of drop during drawdowns. We believe that the uncontrollable increase during drawups models the overly optimistic expectations of investors. When the bubble bursts, many investors are often prone to the "hindsight bias", where they change their perspective of the market brutally and realise how over-confident they have been. At the moment of the crash, investors stop their exuberance, because of fear and structural constraints: they cannot afford to be exposed to risk, and re-evaluate their previous decisions/mistakes. In particular, investors account for the positive bias they had about the market, making the crash even stronger.

We hoped that a drop would naturally appear in our model during drawdowns. Alternatively, we propose two potential add-ons that would balance out the unrealistic increases: delayed compensation and transient overcompensation. The former is an explosive way of reintroducing not included compensation, whereas the latter introduces phases when there is an overcompensation. Both ideas propose to include in the model an oscillation between explosive sub and supermartingale regimes. This framework feels natural and closer to reality [186].

### D.2.2 Delayed Compensation

The first mechanism we study so to introduce a drop in the price during a crash is the delayed compensation. Delayed compensation aims at modelling investors' realism, a comeback to rationality: we propose to reincorporate the compensation of upward jumps. In mathematical terms, the delayed part refers to the compensator's integral part that covers the bubble phase, which has not been included for previously explained reasons. If we say that we observe one pair (drawup, drawdown) on the time interval  $[0, T], T > 0$  where the crash occurs at time  $T_J : 0 < \tau_{\text{drawdown}} < T$ , the missed compensation would be the quantity  $\int_0^{\tau_{\text{drawdown}}} \lambda_u^+ du =: \Lambda_{\tau_{\text{drawdown}}}^+$ . Then, at each time step following the start of the drawdown  $T_J$ , we deduce

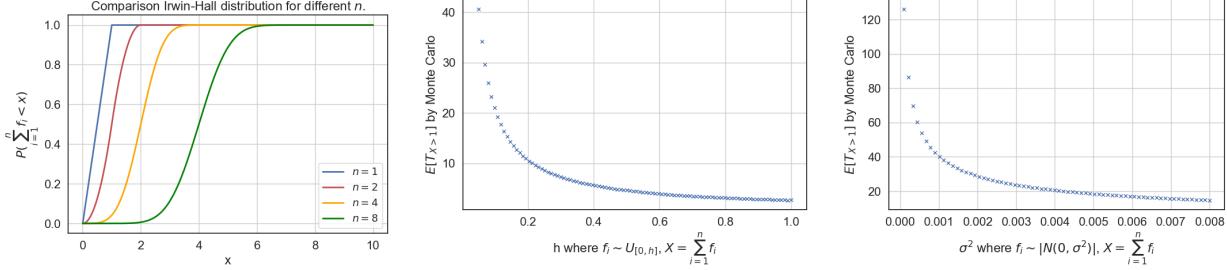


Figure D.4: The first image shows the distribution of a sum of  $n$  uniformly sampled random variable on  $[0,1]$  for  $n$  varying with the colors. In the middle, we show the evolution of the average duration of the delayed compensation  $\mathbb{E}[T_{d.c.}]$  (as defined in the text (D.2.2)) when the  $r_i$  are taken uniform, whereas on the final image they are taken Gaussian. For both images, we have computed the expectation with Monte Carlo methods.

from the price process a fraction of the missed compensation:

$$\frac{dS_t}{S_t} = \mu(t) dt + \sigma(t) dW_t + (dZ_t^+ - \Omega_t d\Lambda_t^+) - (dZ_t^- - d\Lambda_t^-) - g(\Lambda_{\tau_{\text{drawdown}}}^+, t).$$

The quantity  $\Lambda_{\tau_{\text{drawdown}}}^+$  represents the integral of the drawup that has not been compensated.  $g$  is the link between the total missed compensation and the impact on the stock market at time  $t$ . We chose  $g$  guided by the following intuition. Everyday investors re-evaluate the price of their assets: the reassessment of past positive information is done by a fraction of the investors who change the value of their investment in hindsight, despite the absence of new information. We decide to take a random fraction following a distribution. A simple choice would be the family of fractions to be i.i.d., coined  $\{f_i\}_{i \in \mathbb{N}_{\geq 0}}, f_i \sim \mathcal{L}$ , for some distribution  $\mathcal{L}$ , and at each time a fraction  $f_i$  of the investors have reassessed the price to fairer levels. On a grid  $t_i, i \in \mathbb{N}_{\geq 0}$  with fractions of the total  $f_i$ , the function  $g(x, t_i)$  would look like:

$$g(x, t_i) = x \cdot f_i. \quad (\text{D.2.1})$$

We call hereinafter  $T_{\text{d.c.}}$  the first time when the sum of fractions of released compensation is greater than 1, where d.c. stands for delayed compensation:

$$T_{\text{d.c.}} := \inf\{n \in \mathbb{N}_{\geq 0} : \sum_{i=1}^n f_i \geq 1\}. \quad (\text{D.2.2})$$

A first intuitive choice would be to set  $\mathcal{L}$  as the uniform law. The advantage of taking the  $\{f_i\}_{i \in \mathbb{N}_{\geq 0}}$  uniform is that we know the distribution of a sum of uniform random variables as well as the tractable horizon over which the missed compensation is incorporated. We call this time the duration of the delayed compensation. The sum of uniform random variables on  $[0, 1]$  follows a Irwin-Hall distribution, and one can prove directly that  $f_i \sim \mathcal{U}_{[0,1]}$  implies that  $\mathbb{P}(\sum_{i=1}^n f_i \leq 1) = \frac{1}{n!}$ . In turns, it shows that the expectation  $\mathbb{E}[T_{\text{d.c.}}] = e$ .

The same reasoning can be done when we take  $f_i \sim \mathcal{U}_{[0, h]}, h > 0$ , even though  $\mathbb{E}[T_{\text{d.c.}}]$  has a less straightforward form. We plot on Figure D.4 how  $\mathbb{E}[T_{\text{d.c.}}]$  varies as a function of  $h$ .

Another option for the distribution of the fractions  $\{f_i\}_{i \in \mathbb{N}_{\geq 0}}$  would be to sample from the absolute value of a normal distribution  $\mathcal{N}(0, \sigma^2)$  where  $\sigma^2$  can be chosen, and we plot as before how the average duration  $\mathbb{E}[T_{\text{d.c.}}]$  varies depending on  $\sigma^2$  on the same Fig. D.4.

Including the delayed compensation mechanism helps the model stay in line with a trend (mean-reversion phenomenon) since every drawdown will push the price back to the trend, which increases the stability of the model. Furthermore, bubbles of type-II have an efficient drawdown, and the abrupt change of regime from inefficient to efficient would probably require a phase of adaptation. This is because all investors cannot change their positions overnight. Just as in any system suffering a change of regime, there is a transient

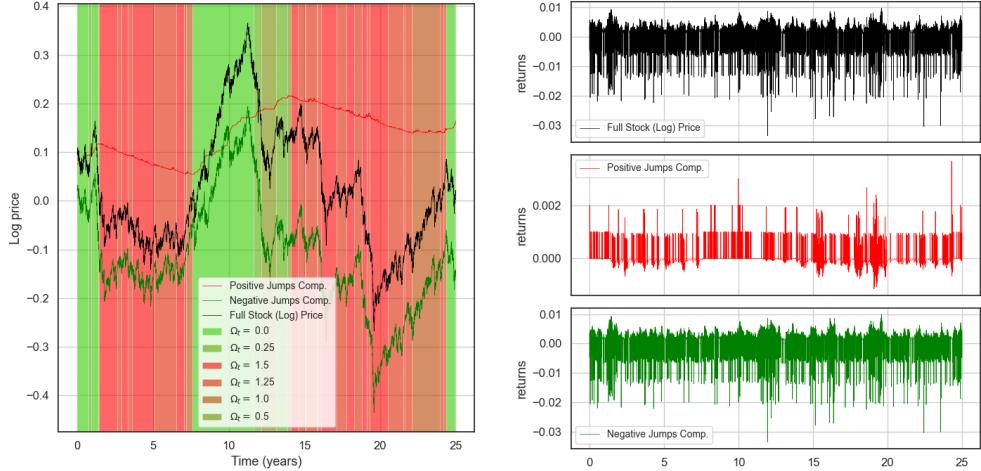


Figure D.5: We simulate a sample path with three regimes  $t\Delta^1 = 250$ ,  $t\Delta^1 = 750$ ,  $t\Delta^1 = 1500$  and associated weights  $\omega^1 = 0.25$ ,  $\omega^2 = 0.25$  and  $\omega^3 = 1$ . The sample path indicates that the transient phase of overcompensation prevents the divergence of prices that has been observed. Specifically, we see two distinct phases of overcompensation, the first ranging from time 2 to 8, and the second from time 14 to 24. Additionally, with a bubble phase in the middle from time 8 to 14. The model is still considered a type-II bubble, as there are periods during which  $\Omega_t$  was equal to 1, as between times 22.5 to 23.8. The positive jumps curve reflect the opinions of investors, oscillating between phases of over-optimism (which are mathematically translated into submartingale phases) and phases of over-pessimism (which are mathematically translated into supermartingale phases).

phase of over-compensation and instability. The drawdown is, however, eventually efficient. The fact that we can choose the distribution of the fractions  $f_i$  allows us to select the length of the transition phase and decide how abrupt the change of mindset can be.

In order to combine this idea with the multi-regime switching, we could release the delayed compensation when the regime's value exceeds a threshold ( $\Omega_t$  larger than a constant), or fix a probability that a certain amount of overdue compensation is released each day, for example with probability  $\Omega_t$ .

### D.2.3 Transient Overcompensation

Another approach to stabilise the model leverages the multi-scale regime-switching paradigm from Subsection D.1.4 and overcompensate the price periodically, making the drawdown stronger and taking the price back to lower values. In order to do that, we set the weights of the regimes such that their sum is greater than 1. This idea is interesting because it would help create a basic negative trend in the price, while respecting the type-II bubble paradigm if we ensure that there is a combination of weights such that the sum equals 1. When  $\Omega_t < 1$  the market is in a submartingale regime when  $\Omega_t > 1$ , the market is in a supermartingale regime and when  $\Omega_t = 1$  the market is in a martingale regime.

Overcompensation seems more accurate than delayed compensation because it is difficult to understand how investors would be able to re-evaluate the prices accurately in hindsight. However, the flaw of overcompensation is that there is no acceleration phenomenon and the decrease in price is linear in log price: we simulated a sample path with three regimes  $t\Delta^1 = 250$ ,  $t\Delta^1 = 750$ ,  $t\Delta^1 = 1500$  and associated weights  $\omega^1 = 0.25$ ,  $\omega^2 = 0.25$  and  $\omega^3 = 1$ . The sample path lies in Fig. D.5. We see that this mechanism effectively reduces the upward bias.

## D.3 Mispricing Index

### D.3.1 A Difference Equation for the Mispricing Index

Malevergne and al. [131] proposes to use an anchoring price defined for a discrete price process  $\{S_n\}_{n \in \mathbb{Z}}$ .

In short, the mispricing index embodies how overvalued the price is. Although defining a fundamental price level is challenging, it is possible to develop a relative metric that measures the price change over time. This is achieved by anchoring the metric to past price levels, which is a common practice in trading that employs various patterns such as "support" or "resistance" levels. It is only natural to model the price using the abstract concept that practitioners also use. We include this metric inside our model and increase the intensity of the jump processes when the price diverges from its reference value, thus making the model more prone to instability and crashes when it is "overvalued".

This mechanism is intuitive because, contrary to popular belief, sharp price increases often precede low returns [217]. We could define an asset as mispriced when it deviates significantly from its historical exponential moving average. The rational expectation mechanism is established such that the greater the divergence from the expectation, the higher the expected returns, owing to a greater probability of a crash. This puts the acceleration of the returns at the center of how we model bubbles.

We call  $\bar{r}$  the mean returns and  $\tau$  a time scale. The anchoring price is defined as  $S_{n-\tau} \exp(\bar{r}\tau)$ , and the ratio  $\frac{S_n}{S_{n-\tau} \exp(\bar{r}\tau)}$  gives how much of the current price is a bubble. In this appendix, we summarise what they proposed.

We call it the ratio of mispricing, for all  $n \in \mathbb{Z}$ :

$$\delta_{n,\tau} := \frac{S_n}{S_{n-\tau} \exp(\bar{r}\tau)}. \quad (\text{D.3.1})$$

The remarkable property of  $\frac{1}{\tau} \ln(\delta_{n,\tau})$  is that it can be rewritten as, for all  $n \in \mathbb{Z}$ :

$$\frac{1}{\tau} \ln(\delta_{n,\tau}) = \frac{1}{\tau} \sum_{k=1}^{\tau} \ln \left( \frac{S_{n-\tau+k}}{S_{n+k-1}} \right) - \bar{r} = \frac{1}{\tau} \sum_{k=1}^{\tau} r_{n-\tau+k} - \bar{r}. \quad (\text{D.3.2})$$

In other words, it represents the average daily excess return over the long-term growth rate  $\bar{r}$ .

The expression does not behave smoothly, and we approximate it by an exponentially smoothed version by the following. The approximation relies on the observation that the solution to the equation (with function variable  $\ln(\delta_{n,a_\tau})$ ):

$$\ln(\delta_{n,a_\tau}) = (1 - a_\tau)(r_n - \bar{r}) + a_\tau \ln(\delta_{n-1,a_\tau}), \quad (\text{D.3.3})$$

is the function:

$$\ln(\delta_{n,a_\tau}) = (1 - a_\tau) \sum_{k=0}^{\infty} a^k (r_{n-k} - \bar{r}), \quad (\text{D.3.4})$$

which corresponds to (D.3.2), and by identification we can define for all  $n \in \mathbb{Z}$ :

$$\ln(\delta_{n,a_\tau}) \approx \frac{1}{\tau} \ln(\delta_{n,\tau}), \quad \text{where } a_\tau \approx 1 - \frac{1}{\tau}, \quad (\text{D.3.5})$$

such that our original objects follow approximately (D.3.3).

We now define the mispricing index as the discrete process  $\{X_n\}_{n \in \mathbb{N}_{\geq 0}}$ :

$$X_n := \frac{1}{s} \ln \left( \frac{\delta_{n-1,a_\tau}}{\delta_{ref}} \right), \quad \forall n \in \mathbb{Z}, \quad (\text{D.3.6})$$

where we have that  $s$  is the steepness of the logistic function (reactivity of the market) and  $\delta_{ref}$  would be the average daily mispricing ratio.  $X_0$  has to be chosen.

We chose the risk of a crash to be:

$$\lambda_n = L(X_n), \quad (\text{D.3.7})$$

where  $L$  denotes the logistic function  $L(x) := \frac{1}{1+e^{-x}}$ . Through  $L$ ,  $\delta_{ref}$  represents the mispricing beyond which the probability for a jump to occur is greater than 1/2.

Thanks to the recurrence relationship of  $\delta_{n,a_\tau}$ , a similar equation applies to  $\{X_n\}_{t \in \mathbb{R}_+}$ , and reads for all  $n \in \mathbb{Z}$ :

$$X_n = a_\tau \cdot X_{n-1} + \frac{1-a_\tau}{s} (r_{n-1} - \bar{r}) + (1-a_\tau) \frac{-\ln \delta_{ref}}{s}. \quad (\text{D.3.8})$$

Associated to the mispricing factor are the parameters:

$$\{X_0, \tau, s, \delta_{ref}\}. \quad (\text{D.3.9})$$

Interestingly, this mechanism does not provide any indication of the fundamental value (real price) of the underlying asset but rather provides insight into recent price variations. This aligns with the arguments presented in [186], as summarized in Subsection 3.1.3, where the LPPLS algorithm, a tool for detecting bubbles and forecasting their outcomes, focuses on the dynamics of how the current price was attained rather than computing a fundamental mispricing. In that regard, the previously introduced mispricing index is not as much a mispricing index as more of a price-variation indicator.

### D.3.2 Multi-Scale Mispricing Index

We offer to rework the mispricing index because we notice that compared to real stock, the original model produces increases in price that are of a shape that is sometimes off. The returns are faster-than-exponential and indeed we observe a peak, however, the base of the peak looks too sharp and the change in regime is too sudden, as it can be seen in [131] Fig. 2 time 1700 and time 3000, the increase of the price at the beginning of the bubble. On the contrary, we observe empirically a smoother transition from drawdown to drawup (recall the log-periodic power law singularity model LPPLS, [217] as well as the presentation of this idea in [186] in Fig. 7 and 8).

We suggest using multiple mispricing indices to create a novel index as a convex combination of the previous indices.

Let us assume we have a sequence of mispricing indices  $\{X_n^i\}_{i \in \mathbb{N}_{\geq 0}}$  and weights  $\{w_X^i\}_{i \in \mathbb{N}_{\geq 0}}$ , summing to 1. Each index (numbered  $i$ ) comes with its own parameters  $\{X_0^i, \tau^i, s^i, \delta_{ref}^i\}$ . Then, we can define the multi-scale mispricing index process as, for all  $n \in \mathbb{Z}$ :

$$X_n := \sum_{i \in \mathbb{N}_{\geq 0}} w_X^i X_n^i. \quad (\text{D.3.10})$$

This linear combination can either be viewed as:

- A convex combination of the previously defined indices: we do a weighted arithmetic mean between the indices,
- The logarithm of a geometric weighted average of the mispricing ratios:  $\sum_{i \in \mathbb{N}_{\geq 0}} w_X^i \frac{1}{s^i} \ln\left(\frac{\delta_{n-1,a_\tau}^i}{\delta_{ref}^i}\right)$ .

Because of the way the indices are defined, indices with longer time scales are a smooth-out version of the indices with shorter time scales, but when we take combinations of them, we are not simply averaging the time scales but really creating a novel indicator that is more complex in its dynamics. Note that the weights  $w_X^i$  are multiplied by the parameter  $1/s^i$ . Hence one can fix the former and change the latter to get the desired result.

Another possibility to construct more sophisticated mispricing indices is a logistic combination of the indices:

$$L = \sum_{i \in \mathbb{N}_{\geq 0}} w_X^i L(X_n^i), \quad (L(x) = \frac{1}{1 + e^{-x}}). \quad (\text{D.3.11})$$

In such case, the weights  $w_X^i$  and the parameters  $s^i$  cannot be replaced by a unique parameter. However, this combination has not proven to show more interesting features than the linear combination. Further experimenting would need to be done to identify the strength of each method.

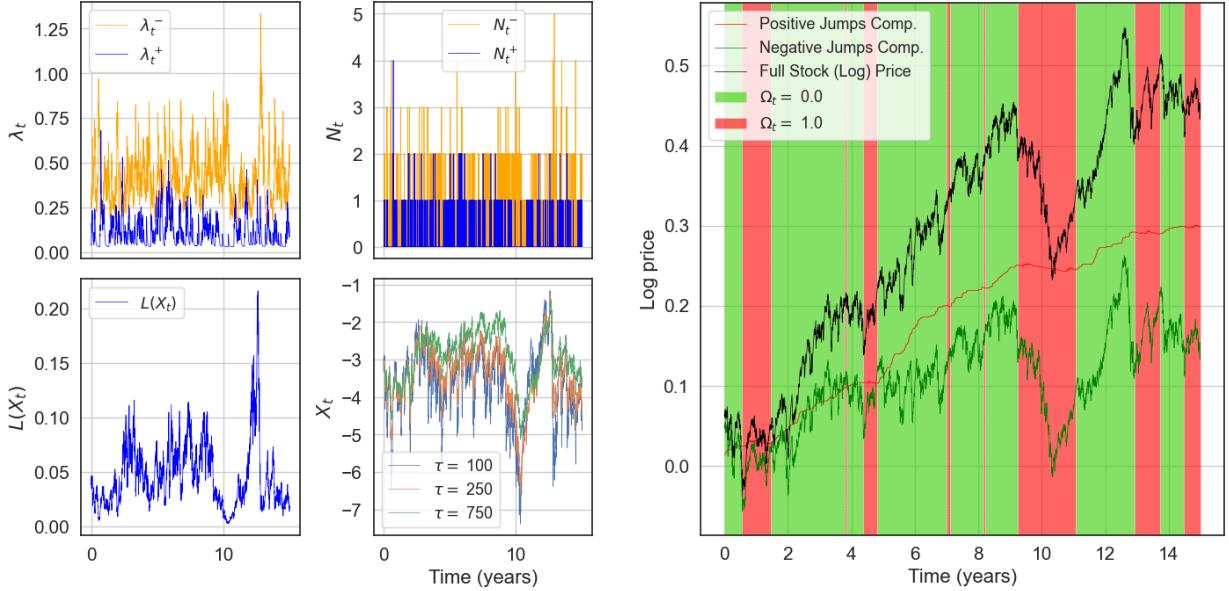


Figure D.6: We sample a trajectory of the model including both positive and negative jumps (but no Brownian motion) when we use a multi-scale mispricing index, with a bi-regime process. This approach enables the mispricing associated with larger time scales to grow, leading to the development of nested bubbles with accelerating returns that may last for years. This is visible between time 2 and 8, where the longer time scale  $\tau = 750$  kept the mispricing high despite the crash around year 4.5 and allowed the bubbly behaviour to continue afterwards, while the smaller time scale  $\tau = 100$  drove the mispricing index down generating smaller scales bubbles. We also observe a stronger crash at year 9, when the large time scale remained high and intensified the crash by keeping the intensity higher. The parameters for this trajectory are:  $d^1 = [0.03, 0.1]$ ,  $d^2 = [0.1, 0.5]$ ,  $d_{\cdot}^1 = [[0.7, 0.0], [0.0, 0.7]]$ ,  $d_{\cdot}^2 = [[0.2, 0.0], [0.0, 0.2]]$ ,  $\kappa = [0.1, -0.8]$ ,  $\tau = [100, 250, 750]$ ,  $s = [0.0003, 0.0002, 0.0001]$ ,  $\bar{X} = -4$ ,  $\Delta = 250$ .

### D.3.3 Numerical Experiments

We aim to mitigate the problems associated with short and long time scales by combining different indices. By utilizing a multi-scale mechanism, smaller variations on shorter time scales trigger a momentum which is then tempered by the larger time scales, preventing an explosion of the value of  $X_t$  that would occur if each time scale were considered separately. This approach enables the mispricing associated with larger time scales to grow, leading to the development of nested bubbles with accelerating returns that may last for years. We suggest that leveraging a wide range of time scales, from 100 to 750 days or more. To illustrate the superiority of a multi-scale mispricing index, we present a sample path in Fig. D.6. Our model shows that a multi-scale mispricing index permits  $L(X_t)$  to vary less quickly, which slows down the growth of bubbles, especially at their inception. The price increases still demonstrates faster-than-exponential growth, and the large time scales result in stronger crashes because the parameter's value remains high after the crash. Additionally, we observe different lengths of bubbles, nested bubbles, plateaus, and spikes. Our proposed model is interpretable and allows for the consideration of different sensitivities. A strong opinion impacts the market's overall mispricing, and longer time scales slow down the explosiveness of shorter ones.

We also note that other combinations are possible, such as combining the logistic function of the mispricing indices as proposed in the previous subsection. But it was not clear from our experimentation which combination was more realistic.



# BIBLIOGRAPHY

- [1] Frédéric Abergel and Aymen Jedidi. “Long-time behavior of a hawkes process–based limit order book”. In: *SIAM Journal on Financial Mathematics* 6.1 (2015), pp. 1026–1043.
- [2] Dilip Abreu and Markus K Brunnermeier. “Bubbles and crashes”. In: *Econometrica* 71.1 (2003), pp. 173–204.
- [3] Eben Afrifa-Yamoah and Ute A Mueller. “Modeling digital camera monitoring count data with intermittent zeros for short-term prediction”. In: *Heliyon* 8.1 (2022), e08774.
- [4] Emrah Altun, Deepesh Bhati, and Naushad Mamode Khan. “A new approach to model the counts of earthquakes: INARPQX (1) process”. In: *SN Applied Sciences* 3.2 (2021), pp. 1–17.
- [5] Torben G Andersen et al. “Modeling and forecasting realized volatility”. In: *Econometrica* 71.2 (2003), pp. 579–625.
- [6] Torben G Andersen et al. “The distribution of realized exchange rate volatility”. In: *Journal of the American statistical association* 96.453 (2001), pp. 42–55.
- [7] Torben G Andersen et al. “The distribution of realized stock return volatility”. In: *Journal of financial economics* 61.1 (2001), pp. 43–76.
- [8] Sasho Andonov. *Safety Accidents in Risky Industries: Black Swans, Gray Rhinos and Other Adverse Events*. CRC Press, 2021.
- [9] E. Bacry, K. Dayri, and J. F. Muzy. “Non-parametric kernel estimation for symmetric Hawkes processes. Application to high frequency financial data”. In: *The European Physical Journal B* 85.5 (May 2012). ISSN: 1434-6036. DOI: [10.1140/epjb/e2012-21005-8](https://doi.org/10.1140/epjb/e2012-21005-8). URL: <http://dx.doi.org/10.1140/epjb/e2012-21005-8>.
- [10] Emmanuel Bacry and Jean-François Muzy. “First- and Second-Order Statistics Characterization of Hawkes Processes and Non-Parametric Estimation”. In: *IEEE Transactions on Information Theory* 62 (2016), pp. 2184–2202.
- [11] Emmanuel Bacry and Jean-François Muzy. *Hawkes model for price and trades high-frequency dynamics*. 2014. DOI: [10.48550/ARXIV.1301.1135](https://arxiv.org/abs/1301.1135). URL: <https://arxiv.org/abs/1301.1135>.
- [12] Emmanuel Bacry et al. “Market impacts and the life cycle of investors orders”. In: *Market Microstructure and Liquidity* 1.02 (2015).
- [13] Emmanuel Bacry et al. *Scaling limits for Hawkes processes and application to financial statistics*. 2012. arXiv: [1202.0842 \[math.PR\]](https://arxiv.org/abs/1202.0842).

- [14] Michał Balcerak and Krzysztof Burnecki. “Testing of fractional Brownian motion in a noisy environment”. In: *Chaos, Solitons & Fractals* 140 (2020).
- [15] Maurice Stevenson Bartlett. *An introduction to stochastic processes: with special reference to methods and applications*. CUP Archive, 1978.
- [16] David S Bates. “Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options”. In: *The Review of Financial Studies* 9.1 (1996), pp. 69–107.
- [17] Mikkel Bennedsen, Asger Lunde, and Mikko S Pakkanen. “Decoupling the short-term and long-term behavior of stochastic volatility”. In: *Journal of Financial Economics*, forthcoming (2022).
- [18] R. N. Bhattacharya, Vijay K. Gupta, and Ed Waymire. “The Hurst effect under trends”. In: *Journal of Applied Probability* 20.3 (1983), pp. 649–662. DOI: [10.2307/3213900](https://doi.org/10.2307/3213900).
- [19] Patrick Billingsley. *Convergence of probability measures*. John Wiley & Sons, 2013.
- [20] Fischer Black. “Studies of stock market volatility changes”. In: *1976 Proceedings of the American statistical association business and economic statistics section* (1976).
- [21] Pierre Blanc, Jonathan Donier, and J-P Bouchaud. “Quadratic Hawkes processes for financial prices”. In: *Quantitative Finance* 17.2 (2017), pp. 171–188.
- [22] Olivier Blanchard and Mark Watson. *Bubbles, Rational Expectations and Financial Markets*. NBER Working Papers 0945. National Bureau of Economic Research, Inc, 1982. URL: <https://EconPapers.repec.org/RePEc:nbr:nberwo:0945>.
- [23] Olivier Jean Blanchard. “Speculative bubbles, crashes and rational expectations”. In: *Economics Letters* 3.4 (1979), pp. 387–389. ISSN: 0165-1765. DOI: [https://doi.org/10.1016/0165-1765\(79\)90017-X](https://doi.org/10.1016/0165-1765(79)90017-X). URL: <https://www.sciencedirect.com/science/article/pii/016517657990017X>.
- [24] Thierry Bochud and Damien Challet. “Optimal approximations of power laws with exponentials: application to volatility models with long memory”. In: *Quantitative Finance* 7.6 (2007), pp. 585–589.
- [25] Duane C. Boes and Jose D. Salas. “Nonstationarity of the mean and the hurst Phenomenon”. In: *Water Resources Research* 14.1 (1978), pp. 135–143. DOI: <https://doi.org/10.1029/WR014i001p00135>. eprint: <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/WR014i001p00135>. URL: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/WR014i001p00135>.
- [26] Tim Bollerslev and Jonathan H Wright. “Semiparametric estimation of long-memory volatility dependencies: The role of high-frequency data”. In: *Journal of econometrics* 98.1 (2000), pp. 81–106.
- [27] Pierre Brémaud and Laurent Massoulié. “Hawkes Branching Point Processes without Ancestors”. In: *Journal of Applied Probability* 38.1 (2001), pp. 122–135. ISSN: 00219002. URL: <http://www.jstor.org/stable/3215746>.
- [28] Pierre Brémaud and Laurent Massoulié. “Stability of Nonlinear Hawkes Processes”. In: *The Annals of Probability* 24.3 (1996), pp. 1563–1588. ISSN: 00911798. URL: <http://www.jstor.org/stable/2244985>.
- [29] David R Brillinger. “Comparative aspects of the study of ordinary time series and of point processes”. In: *Developments in statistics*. Vol. 1. Elsevier, 1978, pp. 33–133.
- [30] David R Brillinger. “Time series, point processes, and hybrids”. In: *Canadian Journal of Statistics* 22.2 (1994), pp. 177–206.
- [31] Peter J Brockwell and Richard A Davis. *Time series: theory and methods*. Springer Science & Business Media, 2009.
- [32] John Y Campbell and Ludger Hentschel. “No news is good news: An asymmetric model of changing volatility in stock returns”. In: *Journal of financial Economics* 31.3 (1992), pp. 281–318.
- [33] Niels Cariou-Kotlarek. “Hawkes Iterative Time Dependent Estimation of Parameters”. In: *ETH Zurich Working Paper Preprint* (Apr. 2022). URL: <https://github.com/Code-Cornelius/ITiDeEP>.
- [34] Niels Cariou-Kotlarek and Alexander Wehrli. “INVAR( $\infty$ ), Existence, Relation to Multivariate Hawkes Processes and Estimation”. In: *Working paper* (Feb. 2023).

- [35] Álvaro Cartea, Samuel N Cohen, and Saad Labyad. “Gradient-based estimation of linear Hawkes processes with general kernels”. In: *arXiv preprint arXiv:2111.10637* (2021).
- [36] Yi-Ting Chen, Ray Y Chou, and Chung-Ming Kuan. “Testing time reversibility without moment restrictions”. In: *Journal of Econometrics* 95.1 (2000), pp. 199–218.
- [37] Zezhun Chen and Angelos Dassios. “Cluster point processes and Poisson thinning INARMA”. In: *Stochastic Processes and their Applications* 147 (2022), pp. 456–480.
- [38] Patrick Cheridito. “Arbitrage in fractional Brownian motion models”. In: *Finance and Stochastics* 7.4 (2003), pp. 533–553.
- [39] Patrick Cheridito. “Gaussian moving averages, semimartingales and option pricing”. In: *Stochastic Processes and their Applications* 109.1 (2004), pp. 47–68. ISSN: 0304-4149. DOI: <https://doi.org/10.1016/j.spa.2003.08.002>. URL: <https://www.sciencedirect.com/science/article/pii/S0304414903001315>.
- [40] Patrick Cheridito, Hideyuki Kawaguchi, and Makoto Maejima. “Fractional Ornstein-Uhlenbeck processes”. In: *Electronic Journal of Probability* 8.none (2003), pp. 1–14. DOI: <10.1214/EJP.v8-125>. URL: <https://doi.org/10.1214/EJP.v8-125>.
- [41] Rémy Chicheportiche and Jean-Philippe Bouchaud. “The fine-structure of volatility feedback I: Multi-scale self-reflexivity”. In: *Physica A: Statistical Mechanics and its Applications* 410 (2014), pp. 174–195.
- [42] Fabienne Comte, Laure Coutin, and Eric Renault. “Affine fractional stochastic volatility models”. In: *Annals of Finance* 8.2 (2012), pp. 337–378.
- [43] Fabienne Comte and Eric Renault. “Long memory in continuous-time stochastic volatility models”. In: *Mathematical finance* 8.4 (1998), pp. 291–323.
- [44] Rama Cont. “Empirical properties of asset returns: stylized facts and statistical issues”. In: *Quantitative finance* 1.2 (2001), p. 223.
- [45] Rama Cont. “Volatility clustering in financial markets: empirical facts and agent-based models”. In: *Long memory in economics*. Springer, 2007, pp. 289–309.
- [46] Rama Cont and Purba Das. “Rough volatility: fact or artefact?” In: *arXiv preprint arXiv:2203.13820* (2022).
- [47] Rama Cont and Peter Tankov. “Nonparametric calibration of jump-diffusion option pricing models.” In: *The Journal of Computational Finance* 7 (2004), pp. 1–49.
- [48] David M Cutler, James M Poterba, and Lawrence H Summers. *What moves stock prices?* 1988.
- [49] José Da Fonseca and Wenjun Zhang. “Volatility of volatility is (also) rough”. In: *Journal of Futures Markets* 39.5 (2019), pp. 600–611.
- [50] D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes: Volume I, Elementary theory and methods*. Springer, 2002.
- [51] Aditi Dandapani, Paul Jusselin, and Mathieu Rosenbaum. “From quadratic Hawkes processes to super-Heston rough volatility models with Zumbach effect”. In: *Quantitative Finance* 21.8 (2021), pp. 1235–1247.
- [52] Angelos Dassios and Hongbiao Zhao. “Exact simulation of Hawkes process with exponentially decaying intensity”. In: *Electronic Communications in Probability* 18.none (2013), pp. 1–13. DOI: <10.1214/ECP.v18-2717>. URL: <https://doi.org/10.1214/ECP.v18-2717>.
- [53] Marcos Lopez De Prado. *Advances in financial machine learning*. John Wiley & Sons, 2018.
- [54] David N. DeJong and Charles H. Whiteman. “Reconsidering ‘trends and random walks in macroeconomic time series’”. In: *Journal of Monetary Economics* 28.2 (1991), pp. 221–254. ISSN: 0304-3932. DOI: [https://doi.org/10.1016/0304-3932\(91\)90051-0](https://doi.org/10.1016/0304-3932(91)90051-0). URL: <https://www.sciencedirect.com/science/article/pii/0304393291900510>.
- [55] Freddy Delbaen and Walter Schachermayer. “The fundamental theorem of asset pricing for unbounded stochastic processes”. In: *Mathematische Annalen* 312.2 (1998), pp. 215–250.

- [56] Freddy Delbaen and Walter Schachermayer. *The mathematics of arbitrage*. Springer Science & Business Media, 2006.
- [57] Arthur P Dempster, Nan M Laird, and Donald B Rubin. “Maximum likelihood from incomplete data via the EM algorithm”. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 39.1 (1977), pp. 1–22.
- [58] Zhuanxin Ding, Clive W.J. Granger, and Robert F. Engle. “A long memory property of stock market returns and a new model”. In: *Journal of Empirical Finance* 1.1 (1993), pp. 83–106. ISSN: 0927-5398. DOI: [https://doi.org/10.1016/0927-5398\(93\)90006-D](https://doi.org/10.1016/0927-5398(93)90006-D). URL: <https://www.sciencedirect.com/science/article/pii/092753989390006D>.
- [59] Noel Archard Managing Directory. “*Flash Crash*” perceptions study. Tech. rep. BlackRock, 2010.
- [60] Paul Doukhan, George Oppenheim, and Murad Taqqu. *Theory and applications of long-range dependence*. Springer Science & Business Media, 2002.
- [61] Nan Du et al. “Recurrent marked temporal point processes: Embedding event history to vector”. In: *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 2016, pp. 1555–1564.
- [62] David Easley, Marcos M Lopez De Prado, and Maureen O’Hara. “The microstructure of the “flash crash”: flow toxicity, liquidity crashes, and the probability of informed trading”. In: *The Journal of Portfolio Management* 37.2 (2011), pp. 118–128.
- [63] David Easley, Marcos M López De Prado, and Maureen O’Hara. “The volume clock: Insights into the high-frequency paradigm”. In: *The Journal of Portfolio Management* 39.1 (2012), pp. 19–29.
- [64] Omar El Euch, Masaaki Fukasawa, and Mathieu Rosenbaum. “The microstructural foundations of leverage effect and rough volatility”. In: *Finance and Stochastics* 22 (2018), pp. 241–280.
- [65] Omar El Euch et al. “The Zumbach effect under rough Heston”. In: *Quantitative Finance* 20.2 (2020), pp. 235–241.
- [66] Cheoljun Eom, Taisei Kaizoji, and Enrico Scalas. “Fat tails in financial return distributions revisited: Evidence from the Korean stock market”. In: *Physica A: Statistical Mechanics and its Applications* 526 (2019), p. 121055.
- [67] Konstantinos Ferentinos. “On Tchebycheff’s type inequalities”. In: *Trabajos de Estadistica y de Investigacion Operativa* 33.1 (1982), pp. 125–132.
- [68] Stephen Figlewski and Xiaozu Wang. “Is the ‘Leverage Effect’ a leverage effect?” In: *Working Paper, Leonard N. Stern School of Business, New York University*. (2000).
- [69] Vladimir Filimonov and Didier Sornette. “Apparent criticality and calibration issues in the Hawkes self-excited point process model: application to high-frequency financial data”. In: *Quantitative Finance* 15.8 (2015), pp. 1293–1314.
- [70] Vladimir Filimonov and Didier Sornette. “Power law scaling and “Dragon-Kings” in distributions of intraday financial drawdowns”. In: *Chaos, Solitons & Fractals* 74 (2015). Extreme Events and its Applications, pp. 27–45. ISSN: 0960-0779. DOI: <https://doi.org/10.1016/j.chaos.2014.12.002>. URL: <https://www.sciencedirect.com/science/article/pii/S0960077914002124>.
- [71] Vladimir Filimonov and Didier Sornette. “Quantifying reflexivity in financial markets: Toward a prediction of flash crashes”. In: *Physical Review E* 85.5 (May 2012). ISSN: 1550-2376. DOI: [10.1103/physreve.85.056108](https://doi.org/10.1103/physreve.85.056108). URL: <http://dx.doi.org/10.1103/PhysRevE.85.056108>.
- [72] Vladimir Filimonov and Didier Sornette. “Self-excited multifractal dynamics”. In: *EPL (Europhysics Letters)* 94.4 (May 2011), p. 46003. DOI: [10.1209/0295-5075/94/46003](https://doi.org/10.1209/0295-5075/94/46003). URL: <https://doi.org/10.1209%2F0295-5075%2F94%2F46003>.
- [73] Vladimir Filimonov and Didier Sornette. “Spurious trend switching phenomena in financial markets”. In: *The European Physical Journal B* 85.5 (2012), pp. 1–5.
- [74] Vladimir Filimonov et al. “Quantification of the high level of endogeneity and of structural regime shifts in commodity markets”. In: *Journal of international Money and finance* 42 (2014), pp. 174–192.

- [75] Patrick Flandrin. “On the spectrum of fractional Brownian motions”. In: *IEEE Transactions on information theory* 35.1 (1989), pp. 197–199.
- [76] Konstantinos Fokianos. “Multivariate Count Time Series Modelling”. In: *Econometrics and Statistics* (2021). ISSN: 2452-3062. DOI: <https://doi.org/10.1016/j.ecosta.2021.11.006>. URL: <https://www.sciencedirect.com/science/article/pii/S2452306221001374>.
- [77] Konstantinos Fokianos. *Regression models for time series analysis*. Wiley-Interscience, 2002.
- [78] Antoine Fosset, Jean-Philippe Bouchaud, and Michael Benzaquen. “Non-parametric estimation of quadratic Hawkes processes for order book events”. In: *The European Journal of Finance* 28.7 (2022), pp. 663–678.
- [79] R Keith Freeland and Brendan PM McCabe. “Analysis of low count time series data by Poisson autoregression”. In: *Journal of time series analysis* 25.5 (2004), pp. 701–722.
- [80] Masaaki Fukasawa, Tetsuya Takabatake, and Rebecca Westphal. “Consistent estimation for fractional stochastic volatility model under high-frequency asymptotics”. In: *Mathematical Finance* n/a.n/a (). DOI: <https://doi.org/10.1111/mafi.12354>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/mafi.12354>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/mafi.12354>.
- [81] Jim Gatheral, Thibault Jaisson, and Mathieu Rosenbaum. “Volatility is rough”. In: *Quantitative Finance* 18.6 (2018), pp. 933–949. DOI: [10.1080/14697688.2017.1393551](https://doi.org/10.1080/14697688.2017.1393551). eprint: <https://doi.org/10.1080/14697688.2017.1393551>. URL: <https://doi.org/10.1080/14697688.2017.1393551>.
- [82] Ramazan Gençay et al. *An introduction to high-frequency finance*. Elsevier, 2001.
- [83] Shoaleh Ghashghaie et al. “Turbulent cascades in foreign exchange markets”. In: *Nature* 381.6585 (1996), pp. 767–770.
- [84] M. Ghil et al. “Extreme events: dynamics, statistics and prediction”. In: *Nonlinear Processes in Geophysics* 18.3 (2011), pp. 295–350. DOI: [10.5194/npg-18-295-2011](https://doi.org/10.5194/npg-18-295-2011). URL: <https://npg.copernicus.org/articles/18/295/2011/>.
- [85] Clive WJ Granger and Zhuanxin Ding. “Varieties of long memory models”. In: *Journal of econometrics* 73.1 (1996), pp. 61–77.
- [86] Franklin A Graybill. *Matrices with applications in statistics*. 512.896 G7 1983. 1983.
- [87] Hamza Guennoun et al. “Asymptotic behavior of the fractional Heston model”. In: *SIAM Journal on Financial Mathematics* 9.3 (2018), pp. 1017–1045.
- [88] Donatiene Hainaut. “Impact of volatility clustering on equity indexed annuities”. In: *Insurance: Mathematics and Economics* 71.C (2016), pp. 367–381. DOI: [10.1016/j.insmatheco.2016.07.003](https://doi.org/10.1016/j.insmatheco.2016.07.003). URL: <https://ideas.repec.org/a/eee/insuma/v71y2016icp367-381.html>.
- [89] Frank R Hampel et al. *Robust statistics: the approach based on influence functions*. Vol. 196. John Wiley & Sons, 2011.
- [90] Stephen J. Hardiman, Nicolas Bercot, and Jean-Philippe Bouchaud. “Critical reflexivity in financial markets: a Hawkes process analysis”. In: *The European Physical Journal B* 86.10 (Oct. 2013). ISSN: 1434-6036. DOI: [10.1140/epjb/e2013-40107-3](https://doi.org/10.1140/epjb/e2013-40107-3). URL: <http://dx.doi.org/10.1140/epjb/e2013-40107-3>.
- [91] Trevor Hastie et al. *The elements of statistical learning: data mining, inference, and prediction*. Vol. 2. Springer, 2009.
- [92] Alan G. Hawkes. “Hawkes processes and their applications to finance: a review”. In: *Quantitative Finance* 18.2 (2018), pp. 193–198. DOI: [10.1080/14697688.2017.1403131](https://doi.org/10.1080/14697688.2017.1403131). eprint: <https://doi.org/10.1080/14697688.2017.1403131>. URL: <https://doi.org/10.1080/14697688.2017.1403131>.
- [93] Alan G. Hawkes. “Spectra of Some Self-Exciting and Mutually Exciting Point Processes”. In: *Biometrika* 58.1 (Apr. 1971), pp. 83–90.
- [94] Alan G. Hawkes and David Oakes. “A Cluster Process Representation of a Self-Exciting Process”. In: *Journal of Applied Probability* 11.3 (1974), pp. 493–503. ISSN: 00219002. URL: <http://www.jstor.org/stable/3212693>.

- [95] Agnès Helmstetter and Didier Sornette. “Subcritical and supercritical regimes in epidemic models of earthquake aftershocks”. In: *Journal of Geophysical Research: Solid Earth* 107.B10 (2002), ESE 10-1-ESE 10–21. DOI: <https://doi.org/10.1029/2001JB001580>. eprint: <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2001JB001580>. URL: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2001JB001580>.
- [96] Norbert Hilber et al. *Computational methods for quantitative finance: Finite element methods for derivative pricing*. Springer Science & Business Media, 2013.
- [97] Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.
- [98] H. E. Hurst. “Long-Term Storage Capacity of Reservoirs”. In: *Transactions of the American Society of Civil Engineers* 116.1 (1951), pp. 770–799. DOI: [10.1061/TACEAT.0006518](https://doi.org/10.1061/TACEAT.0006518). eprint: <https://ascelibrary.org/doi/pdf/10.1061/TACEAT.0006518>. URL: <https://ascelibrary.org/doi/abs/10.1061/TACEAT.0006518>.
- [99] Harold Edwin Hurst. “A suggested statistical model of some time series which occur in nature”. In: *Nature* 180.4584 (1957), pp. 494–494.
- [100] Andreas Hüslér, Didier Sornette, and Cars H Hommes. “Super-exponential bubbles in lab experiments: evidence for anchoring over-optimistic expectations on price”. In: *Journal of Economic Behavior & Organization* 92 (2013), pp. 304–316.
- [101] Patricia A Jacobs and Peter AW Lewis. “Discrete time series generated by mixtures II: asymptotic properties”. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 40.2 (1978), pp. 222–228.
- [102] Patricia A Jacobs and Peter AW Lewis. “Discrete time series generated by mixtures. I: Correlational and runs properties”. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 40.1 (1978), pp. 94–105.
- [103] Patricia A Jacobs and Peter AW Lewis. “Stationary discrete autoregressive-moving average time series generated by mixtures”. In: *Journal of Time Series Analysis* 4.1 (1983), pp. 19–36.
- [104] Antoine Jacquier, Martin Keller-Ressel, and Aleksandar Mijatović. “Large deviations and stochastic volatility with jumps: asymptotic implied volatility for affine models”. In: *Stochastics An International Journal of Probability and Stochastic Processes* 85.2 (2013), pp. 321–345.
- [105] Antoine Jacquier and Mugad Oumgari. “Deep curve-dependent PDEs for affine rough volatility”. In: *arXiv preprint arXiv:1906.02551* (2019).
- [106] Thibault Jaisson and Mathieu Rosenbaum. “Limit theorems for nearly unstable Hawkes processes”. In: *The Annals of Applied Probability* 25.2 (Apr. 2015). ISSN: 1050-5164. DOI: [10.1214/14-aap1005](https://doi.org/10.1214/14-aap1005). URL: <http://dx.doi.org/10.1214/14-AAP1005>.
- [107] Thibault Jaisson and Mathieu Rosenbaum. “Rough fractional diffusions as scaling limits of nearly unstable heavy tailed Hawkes processes”. In: *The Annals of Applied Probability* 26.5 (2016), pp. 2860–2882.
- [108] Zhi-Qiang Jiang et al. “Multifractal analysis of financial markets: a review”. In: *Reports on Progress in Physics* 82.12 (2019), p. 125901.
- [109] Du Jin-Guan and Li Yuan. “THE INTEGER-VALUED AUTOREGRESSIVE (INAR(p)) MODEL”. In: *Journal of Time Series Analysis* 12.2 (1991), pp. 129–142. DOI: <https://doi.org/10.1111/j.1467-9892.1991.tb00073.x>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9892.1991.tb00073.x>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9892.1991.tb00073.x>.
- [110] Anders Johansen and Didier Sornette. “Shocks, Crashes and Bubbles in Financial Markets”. In: *Brussels Economic Review* 53.2 (2010), pp. 201–253. URL: <https://EconPapers.repec.org/RePEc:bxr:bxrceb:2013/80942>.
- [111] Robert C Jung, Martin Kukuk, and Roman Liesenfeld. “Time series of count data: modeling, estimation and diagnostics”. In: *Computational Statistics & Data Analysis* 51.4 (2006), pp. 2350–2364.

- [112] Paul Jusselin and Mathieu Rosenbaum. “No-arbitrage implies power-law market impact and rough volatility”. In: *Mathematical Finance* 30.4 (2020), pp. 1309–1336.
- [113] Taisei Kaizoji and Didier Sornette. “Market Bubbles and Crashes”. In: *Encyclopedia of Quantitative Finance* (2010).
- [114] D Karlis. “Modelling multivariate times series for counts”. In: *Handbook of Discrete-Valued Time Series, Handbooks of Modern Statistical Methods* (2016), pp. 407–424.
- [115] M. Kirchner and A. Bercher. “A nonparametric estimation procedure for the Hawkes process: comparison with maximum likelihood estimation”. In: *Journal of Statistical Computation and Simulation* 88.6 (2018), pp. 1106–1116. DOI: [10.1080/00949655.2017.1422126](https://doi.org/10.1080/00949655.2017.1422126). eprint: <https://doi.org/10.1080/00949655.2017.1422126>. URL: <https://doi.org/10.1080/00949655.2017.1422126>.
- [116] Matthias Kirchner. “An estimation procedure for the Hawkes process”. In: *Quantitative Finance* 17.4 (2017), pp. 571–595. DOI: [10.1080/14697688.2016.1211312](https://doi.org/10.1080/14697688.2016.1211312). eprint: <https://doi.org/10.1080/14697688.2016.1211312>. URL: <https://doi.org/10.1080/14697688.2016.1211312>.
- [117] Matthias Kirchner. “Hawkes and INAR( $\infty$ ) processes”. In: *Stochastic Processes and their Applications* 126.8 (2016), pp. 2494–2525. ISSN: 0304-4149. DOI: <https://doi.org/10.1016/j.spa.2016.02.008>. URL: <https://www.sciencedirect.com/science/article/pii/S0304414916000399>.
- [118] Vit Klemeš. “The Hurst Phenomenon: A puzzle?” In: *Water Resources Research* 10.4 (1974), pp. 675–688. DOI: <https://doi.org/10.1029/WR010i004p00675>. eprint: <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/WR010i004p00675>. URL: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/WR010i004p00675>.
- [119] Denis Kwiatkowski et al. “Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?” In: *Journal of Econometrics* 54.1 (1992), pp. 159–178. ISSN: 0304-4076. DOI: [https://doi.org/10.1016/0304-4076\(92\)90104-Y](https://doi.org/10.1016/0304-4076(92)90104-Y). URL: <https://www.sciencedirect.com/science/article/pii/030440769290104Y>.
- [120] Mehdi Lallouache and Damien Challet. “The limits of statistical significance of Hawkes processes fitted to financial data”. In: *Quantitative Finance* 16.1 (2016), pp. 1–11.
- [121] Alain Latour. “Existence and Stochastic Structure of a Non-negative Integer-valued Autoregressive Process”. In: *Journal of Time Series Analysis* 19.4 (1998), pp. 439–455. DOI: <https://doi.org/10.1111/1467-9892.00102>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-9892.00102>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-9892.00102>.
- [122] Alain Latour. “The Multivariate Ginar(p) Process”. In: *Advances in Applied Probability* 29.1 (1997), pp. 228–248. DOI: [10.2307/1427868](https://doi.org/10.2307/1427868).
- [123] Erik Lewis and George Mohler. “A nonparametric EM algorithm for multiscale Hawkes processes”. In: *Journal of Nonparametric Statistics* 1.1 (2011), pp. 1–20.
- [124] Ming Li. “On the long-range dependence of fractional Brownian motion”. In: *Mathematical Problems in Engineering* 2013 (2013).
- [125] Gilberto Libanio. “Unit roots in macroeconomic time series: theory, implications, and evidence”. In: *Nova Economia* 15.3 (2005), pp. 145–176. URL: <https://EconPapers.repec.org/RePEc:nov:artigo:v:15:y:2005:i:3:p:145-176>.
- [126] Kar Wai Lim et al. “Simulation and calibration of a fully Bayesian marked multidimensional Hawkes process with dissimilar decays”. In: *Asian Conference on Machine Learning*. PMLR. 2016, pp. 238–253. DOI: [10.48550/ARXIV.1803.04654](https://doi.org/10.48550/ARXIV.1803.04654).
- [127] Thomas Josef Liniger. “Multivariate hawkes processes”. PhD thesis. ETH Zurich, 2009.
- [128] Giulia Livieri et al. “Rough volatility: evidence from option prices”. In: *IIE transactions* 50.9 (2018), pp. 767–776.
- [129] Ignacio N Lobato and Carlos Velasco. “Long memory in stock-market trading volume”. In: *Journal of Business & Economic Statistics* 18.4 (2000), pp. 410–427.
- [130] Helmut Lütkepohl. *Introduction to multiple time series analysis*. Springer Science & Business Media, 2013.

- [131] Yannick Malevergne, Didier Sornette, and Ran Wei. “A model of financial bubbles and drawdowns with non-local behavioral self-referencing”. In: *Swiss Finance Institute Research Paper* 21-96 (2021).
- [132] Benoît Mandelbrot. “The Variation of Certain Speculative Prices”. In: *The Journal of Business* 36 (1963). URL: <https://EconPapers.repec.org/RePEc:ucp:jnlbus:v:36:y:1963:p:394>.
- [133] Benoit B Mandelbrot. *Fractals and scaling in finance: Discontinuity, concentration, risk. Selecta volume E*. Springer Science & Business Media, 2013.
- [134] Benoit B. Mandelbrot and John W. Van Ness. “Fractional Brownian Motions, Fractional Noises and Applications”. In: *SIAM Review* 10.4 (1968), pp. 422–437. DOI: [10.1137/1010093](https://doi.org/10.1137/1010093). eprint: <https://doi.org/10.1137/1010093>. URL: <https://doi.org/10.1137/1010093>.
- [135] Riccardo Marcaccioli, Jean-Philippe Bouchaud, and Michael Benzaquen. “Exogenous and endogenous price jumps belong to different dynamical classes”. In: *Journal of Statistical Mechanics: Theory and Experiment* 2022.2 (Feb. 2022), p. 023403. DOI: [10.1088/1742-5468/ac498c](https://doi.org/10.1088/1742-5468/ac498c). URL: <https://doi.org/10.1088/1742-5468/ac498c>.
- [136] Michael Mark, Jan Sila, and Thomas A Weber. “Quantifying endogeneity of cryptocurrency markets”. In: *The European Journal of Finance* (2020), pp. 1–16.
- [137] Till Massing. “What is the best Lévy model for stock indices? A comparative study with a view to time consistency”. In: *Financial Markets and Portfolio Management* 33.3 (2019), pp. 277–344.
- [138] Ed McKenzie. “Some simple models for discrete variate time series 1”. In: *JAWRA Journal of the American Water Resources Association* 21.4 (1985), pp. 645–650.
- [139] Geoffrey J McLachlan and Thriyambakam Krishnan. *The EM algorithm and extensions*. John Wiley & Sons, 2007.
- [140] Aaron Melman. “Generalization and variations of Pellet’s theorem for matrix polynomials”. In: *Linear Algebra and its Applications* 439.5 (2013), pp. 1550–1567. ISSN: 0024-3795. DOI: <https://doi.org/10.1016/j.laa.2013.05.003>. URL: <https://www.sciencedirect.com/science/article/pii/S0024379513003145>.
- [141] R Vilela Mendes, Maria João Oliveira, and AM Rodrigues. “The fractional volatility model: no-arbitrage, leverage and completeness”. In: *arXiv preprint arXiv:1205.2866* (2012).
- [142] Robert C. Merton. “Option pricing when underlying stock returns are discontinuous”. In: *Journal of Financial Economics* 3.1 (1976), pp. 125–144. ISSN: 0304-405X. DOI: [https://doi.org/10.1016/0304-405X\(76\)90022-2](https://doi.org/10.1016/0304-405X(76)90022-2). URL: <https://www.sciencedirect.com/science/article/pii/0304405X76900222>.
- [143] Aleksandar Mijatović and Peter Tankov. “A new look at short-term implied volatility in asset price models with jumps”. In: *Mathematical Finance* 26.1 (2016), pp. 149–183.
- [144] Maxime Morariu-Patrichi and Mikko S Pakkanen. “State-dependent Hawkes processes and their application to limit order book modelling”. In: *Quantitative Finance* (2021), pp. 1–21.
- [145] Tobias J Moskowitz, Yao Hua Ooi, and Lasse Heje Pedersen. “Time series momentum”. In: *Journal of financial economics* 104.2 (2012), pp. 228–250.
- [146] Shyam Nandan, Guy Ouillon, and Didier Sornette. “Are large earthquakes preferentially triggered by other large events?” In: *Earth and Space Science Open Archive* (2022), p. 31. DOI: [10.1002/essoar.10510831.1](https://doi.org/10.1002/essoar.10510831.1). URL: <https://doi.org/10.1002/essoar.10510831.1>.
- [147] Shyam Nandan, Guy Ouillon, and Didier Sornette. *Triggering of large earthquakes is driven by their twins*. 2021. DOI: [10.48550/ARXIV.2104.04592](https://arxiv.org/abs/2104.04592). URL: <https://arxiv.org/abs/2104.04592>.
- [148] Shyam Nandan et al. “Forecasting the rates of future aftershocks of all generations is essential to develop better earthquake forecast models”. In: *Journal of Geophysical Research: Solid Earth* 124.8 (2019), pp. 8404–8425.
- [149] Shyam Nandan et al. “Global models for short-term earthquake forecasting and predictive skill assessment”. In: *The European Physical Journal Special Topics* 230.1 (2021), pp. 425–449.

- [150] Charles R. Nelson and Charles R. Plosser. “Trends and random walks in macroeconomic time series: Some evidence and implications”. In: *Journal of Monetary Economics* 10.2 (1982), pp. 139–162. ISSN: 0304-3932. DOI: [https://doi.org/10.1016/0304-3932\(82\)90012-5](https://doi.org/10.1016/0304-3932(82)90012-5). URL: <https://www.sciencedirect.com/science/article/pii/0304393282900125>.
- [151] Maureen O’Hara and Mao Ye. “Is market fragmentation harming market quality?” In: *Journal of Financial Economics* 100.3 (2011), pp. 459–474. ISSN: 0304-405X. DOI: <https://doi.org/10.1016/j.jfineco.2011.02.006>. URL: <https://www.sciencedirect.com/science/article/pii/S0304405X11000390>.
- [152] PE O’Connell et al. “The scientific legacy of Harold Edwin Hurst (1880–1978)”. In: *Hydrological Sciences Journal* 61.9 (2016), pp. 1571–1590. DOI: [10.1080/02626667.2015.1125998](https://doi.org/10.1080/02626667.2015.1125998). eprint: <https://doi.org/10.1080/02626667.2015.1125998>. URL: <https://doi.org/10.1080/02626667.2015.1125998>.
- [153] Robert Rupert Officer. “The distribution of stock returns”. In: *Journal of the american statistical association* 67.340 (1972), pp. 807–812.
- [154] Yosihiko Ogata. “Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes”. In: *Journal of the American Statistical Association* 83.401 (1988), pp. 9–27. DOI: [10.1080/01621459.1988.10478560](https://doi.org/10.1080/01621459.1988.10478560). eprint: <https://www.tandfonline.com/doi/pdf/10.1080/01621459.1988.10478560>. URL: <https://www.tandfonline.com/doi/abs/10.1080/01621459.1988.10478560>.
- [155] Arek Ohanessian, Jeffrey R. Russell, and Ruey S. Tsay. “True or Spurious Long Memory? A New Test”. In: *Journal of Business & Economic Statistics* 26.2 (2008), pp. 161–175. ISSN: 07350015. URL: <http://www.jstor.org/stable/27638972>.
- [156] Joao Gama Oliveira and Albert-László Barabási. “Darwin and Einstein correspondence patterns”. In: *Nature* 437.7063 (2005), pp. 1251–1251.
- [157] Mohamed A Al-Osh and Aus A Alzaid. “First-order integer-valued autoregressive (INAR (1)) process”. In: *Journal of Time Series Analysis* 8.3 (1987), pp. 261–275.
- [158] Xanthi Pedeli and Dimitris Karlis. “A bivariate INAR (1) process with application”. In: *Statistical modelling* 11.4 (2011), pp. 325–349.
- [159] Torben Mark Pedersen. “The Hodrick–Prescott filter, the Slutsky effect, and the distortionary effect of filters”. In: *Journal of economic dynamics and control* 25.8 (2001), pp. 1081–1101.
- [160] Pierre Perron, Mototsugu Shintani, and Tomoyoshi Yabu. “Testing for Flexible Nonlinear Trends with an Integrated or Stationary Noise Component”. In: *Oxford Bulletin of Economics and Statistics* 79.5 (2017), pp. 822–850. DOI: <https://doi.org/10.1111/obes.12169>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/obes.12169>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/obes.12169>.
- [161] Peter C. B. Phillips. “New Tools for Understanding Spurious Regressions”. In: *Econometrica* 66.6 (1998), pp. 1299–1325. ISSN: 00129682, 14680262. URL: <http://www.jstor.org/stable/2999618>.
- [162] Marcello Rambaldi, Emmanuel Bacry, and Jean-François Muzy. “Disentangling and quantifying market participant volatility contributions”. In: *Quantitative Finance* 19.10 (2019), pp. 1613–1625.
- [163] Alex Reinhart. “A review of self-exciting spatio-temporal point processes and their applications”. In: *Statistical Science* 33.3 (2018), pp. 299–318.
- [164] Marian-Andrei Rizoiu et al. “Hawkes processes for events in social media”. In: *Frontiers of multimedia research*. 2017, pp. 191–218.
- [165] Leonard Christopher Gordon Rogers. “Arbitrage with fractional Brownian motion”. In: *Mathematical Finance* 7.1 (1997), pp. 95–105.
- [166] Leonard Christopher Gordon Rogers. “Things we think we know”. In: *Preprint, available at https://www.skokholm.co.uk/lcgr/downloadable-papers* (2019).
- [167] Mathieu Rosenbaum and Mehdi Tomas. “From microscopic price dynamics to multidimensional rough volatility models”. In: *Advances in Applied Probability* 53.2 (2021), pp. 425–462.

- [168] Walter Rudin. *Real and complex analysis (mcgraw-hill international editions: Mathematics series)*. McGraw-Hill Publishing Co. New York, 1987, p. 483.
- [169] Tina Hviid Rydberg and Neil Shephard. “Dynamics of trade-by-trade price movements: decomposition and models”. In: *Journal of Financial Econometrics* 1.1 (2003), pp. 2–25.
- [170] A. Saichev and D. Sornette. “Superlinear scaling of offspring at criticality in branching processes”. In: *Physical Review E* 89.1 (Jan. 2014). DOI: [10.1103/physreve.89.012104](https://doi.org/10.1103/physreve.89.012104). URL: <https://doi.org/10.1103%2Fphysreve.89.012104>.
- [171] Muhammad Surajo Sanusi. “Investigating the sources of Black’s leverage effect in oil and gas stocks”. In: *Cogent Economics & Finance* 5.1 (2017), p. 1.
- [172] Michael Schatz. “Financial Modeling of Bubbles and Crashes”. PhD thesis. ETH Zurich, 2020.
- [173] Michael Schatz and Didier Sornette. “Inefficient Bubbles and Efficient Drawdowns in Financial Markets”. In: *International Journal of Theoretical and Applied Finance* 23.07 (2020), p. 2050047. DOI: [10.1142/S0219024920500478](https://doi.org/10.1142/S0219024920500478). eprint: <https://doi.org/10.1142/S0219024920500478>. URL: <https://doi.org/10.1142/S0219024920500478>.
- [174] Michael Schatz, Spencer Wheatley, and Didier Sornette. “The ARMA Point Process and its Estimation”. In: *Econometrics and Statistics* (2021). ISSN: 2452-3062. DOI: <https://doi.org/10.1016/j.ecosta.2021.11.002>. URL: <https://www.sciencedirect.com/science/article/pii/S2452306221001349>.
- [175] Eugene Seneta. *Non-negative matrices and Markov chains*. Springer Science & Business Media, 2006.
- [176] Golnaz Shahtahmassebi and Rana Moyeed. “An application of the generalized Poisson difference distribution to the Bayesian modelling of football scores”. In: *Statistica Neerlandica* 70.3 (2016), pp. 260–273.
- [177] Jun Shao. *Mathematical statistics*. Springer Science & Business Media, 2003.
- [178] Robert J. Shiller. “Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?” In: *The American Economic Review* 71.3 (1981), pp. 421–436. ISSN: 00028282. URL: <http://www.jstor.org/stable/1802789>.
- [179] Leigh Shlomovich, Edward AK Cohen, and Niall Adams. “A parameter estimation method for multivariate binned Hawkes processes”. In: *Statistics and Computing* 32.6 (2022), p. 98.
- [180] Leigh Shlomovich et al. “Parameter Estimation of Binned Hawkes Processes”. In: *Journal of Computational and Graphical Statistics* (2022), pp. 1–11.
- [181] Christopher A Sims. “Macroeconomics and reality”. In: *Econometrica: journal of the Econometric Society* (1980), pp. 1–48.
- [182] Anne Sornette and Didier Sornette. “Renormalization of earthquake aftershocks”. In: *Geophysical Research Letters* 26.13 (July 1999), pp. 1981–1984. DOI: [10.1029/1999gl900394](https://doi.org/10.1029/1999gl900394). URL: <https://doi.org/10.1029%2F1999gl900394>.
- [183] D Sornette and A Helmstetter. “Endogenous versus exogenous shocks in systems with memory”. In: *Physica A: Statistical Mechanics and its Applications* 318.3 (2003), pp. 577–591.
- [184] Didier Sornette. “Dragon-Kings, Black Swans and the Prediction of Crises”. In: *Dragon-Kings, Black Swans and the Prediction of Crises* 2 (July 2009), pp. 1–18. DOI: [10.2139/ssrn.1596032](https://doi.org/10.2139/ssrn.1596032).
- [185] Didier Sornette. “Endogenous versus exogenous origins of crises”. In: *Extreme events in nature and society*. Springer, 2006, pp. 95–119.
- [186] Didier Sornette and Peter Cauwels. “Financial bubbles: mechanisms and diagnostics”. In: *Review of Behavioral Economics* 2.3 (2015), pp. 279–305.
- [187] Didier Sornette, Peter Cauwels, and Georgi Smiljanov. “Can we use volatility to diagnose financial bubbles? Lessons from 40 historical bubbles”. In: *Quantitative Finance and Economics* 2.1 (2018), pp. 486–594.
- [188] Didier Sornette and Frank Cuypers. “Why stock markets crash: Critical events in complex financial systems”. In: *Physics Today* 57.3 (2004), pp. 78–79.

- [189] Didier Sornette, Yannick Malevergne, and J.F. Muzy. “What causes crashes?” In: *Risk* 16 (Feb. 2003), pp. 67–71.
- [190] Didier Sornette and Guy Ouillon. “Dragon-kings: mechanisms, statistical methods and empirical evidence”. In: *The European Physical Journal Special Topics* 205.1 (2012), pp. 1–26.
- [191] Didier Sornette, Spencer Wheatley, and Peter Cauwels. “The fair reward problem: The illusion of success and how to solve it”. In: *Advances in Complex Systems* 22.03 (2019), p. 1950005.
- [192] Joseph E Stiglitz. “Symposium on bubbles”. In: *Journal of economic perspectives* 4.2 (1990), pp. 13–18.
- [193] Sasha Stoikov. “The micro-price: a high-frequency estimator of future prices”. In: *Quantitative Finance* 18.12 (2018), pp. 1959–1966. DOI: [10.1080/14697688.2018.1489139](https://doi.org/10.1080/14697688.2018.1489139). eprint: <https://doi.org/10.1080/14697688.2018.1489139>. URL: <https://doi.org/10.1080/14697688.2018.1489139>.
- [194] Report submitted by a Study Group established by the Markets Committee. *Monitoring of fast-paced electronic markets*. Tech. rep. Bank for International Settlement, Sept. 2018.
- [195] Nassim Nicholas Taleb. *The black swan: The impact of the highly improbable*. Vol. 2. Random house, 2007.
- [196] Peter Tankov. “Financial Modeling with Lévy Processes”. École thématique. Lecture. Banach Center, Institute of Mathematics of the Polish Academy of Sciences (IMPAN), Warsaw, Oct. 2010. URL: <https://cel.archives-ouvertes.fr/cel-00665021>.
- [197] Gilles Teyssière and Alan P Kirman. *Long memory in economics*. Springer Science & Business Media, 2006.
- [198] Jean Tirole. “On the Possibility of Speculation under Rational Expectations”. In: *Econometrica* 50.5 (1982), pp. 1163–1181. ISSN: 00129682, 14680262. URL: <http://www.jstor.org/stable/1911868>.
- [199] Howell Tong. “Threshold models in time series analysis—30 years on”. In: *Statistics and its Interface* 4.2 (2011), pp. 107–118.
- [200] Bence Tóth et al. “Anomalous Price Impact and the Critical Nature of Liquidity in Financial Markets”. In: *Phys. Rev. X* 1 (2 Oct. 2011), p. 021006. DOI: [10.1103/PhysRevX.1.021006](https://doi.org/10.1103/PhysRevX.1.021006). URL: <https://link.aps.org/doi/10.1103/PhysRevX.1.021006>.
- [201] Ciprian A Tudor. “Fractional Brownian Motion and Related Processes”. In: *Analysis of Variations for Self-similar Processes*. Springer, 2013, pp. 3–25.
- [202] Pierre Brémaud and Laurent Massoulié. “Power spectra of general shot noises and Hawkes point processes with a random excitation”. In: *Advances in Applied Probability* 34.1 (2002), pp. 205–222. DOI: [10.1239/aap/1019160957](https://doi.org/10.1239/aap/1019160957).
- [203] Daniel Ventosa-Santularia. “Spurious regression”. In: *Journal of Probability and Statistics* 2009 (2009).
- [204] Almut E.D. Veraart. “Modeling, simulation and inference for multivariate time series of counts using trawl processes”. In: *Journal of Multivariate Analysis* 169 (2019), pp. 110–129. ISSN: 0047-259X. DOI: <https://doi.org/10.1016/j.jmva.2018.08.012>. URL: <https://www.sciencedirect.com/science/article/pii/S0047259X16302068>.
- [205] Alexander Wehrli. “Financial Point Process modelling, the endo-exo problem, flash crashes and the excess volatility puzzle explained”. PhD thesis. ETHZurich, 2021.
- [206] Alexander Wehrli and Didier Sornette. “Classification of flash crashes using the Hawkes(p,q) framework”. In: *Quantitative Finance* 22.2 (2022), pp. 213–240. DOI: [10.1080/14697688.2021.1941212](https://doi.org/10.1080/14697688.2021.1941212). eprint: <https://doi.org/10.1080/14697688.2021.1941212>. URL: <https://doi.org/10.1080/14697688.2021.1941212>.
- [207] Alexander Wehrli and Didier Sornette. “Excess financial volatility explained by endogenous excitations revealed by EM calibrations of a generalized Hawkes point process”. In: *Swiss Finance Institute Research Paper, Preprint* 21-35 (2021).

- [208] Alexander Wehrli, Spencer Wheatley, and Didier Sornette. “Scale-, time- and asset-dependence of Hawkes process estimates on high frequency price changes”. In: *Quantitative Finance* 21.5 (2021), pp. 729–752. DOI: [10.1080/14697688.2020.1838602](https://doi.org/10.1080/14697688.2020.1838602). eprint: <https://doi.org/10.1080/14697688.2020.1838602>. URL: <https://doi.org/10.1080/14697688.2020.1838602>.
- [209] Ran Wei, Alexander Wehrli, and Didier Sornette. “A financial bubble model with non-local behavioral self-referencing and self-excited crashes”. In: *ETH Zurich Working Paper* (2022).
- [210] Rebecca Westphal and Didier Sornette. “Market impact and performance of arbitrageurs of financial bubbles in an agent-based model”. In: *Journal of Economic Behavior & Organization* 171 (2020), pp. 1–23.
- [211] Spencer Wheatley, Vladimir Filimonov, and Didier Sornette. “The Hawkes process with renewal immigration & its estimation with an EM algorithm”. In: *Computational Statistics & Data Analysis* 94 (2016), pp. 120–135.
- [212] Spencer Wheatley, Alexander Wehrli, and Didier Sornette. “The endo-exo problem in high frequency financial price fluctuations and rejecting criticality”. In: *Quantitative Finance* 19.7 (2019), pp. 1165–1178.
- [213] Kieran Wood et al. “Trading with the Momentum Transformer: An Intelligent and Interpretable Architecture”. In: *arXiv preprint arXiv:2112.08534* (2021).
- [214] Michele Wucker and Tavia Gilbert. *You Are What You Risk*. Pegasus Books, 2021.
- [215] Kōsaku Yosida. *Functional analysis*. Springer Science & Business Media, 2012, p. 513.
- [216] Elia Zarinelli et al. “Beyond the square root: Evidence for logarithmic dependence of market impact on size and participation rate”. In: *Market Microstructure and Liquidity* 1.02 (2015), p. 1550004.
- [217] Dongshuai Zhao and Didier Sornette. “Bubbles for Fama from Sornette”. In: *Swiss Finance Institute Research* (2021), pp. 21–94.
- [218] Lingjiong Zhu. “Central Limit Theorem for Nonlinear Hawkes Processes”. In: *Journal of Applied Probability* 50.3 (2013), pp. 760–771. DOI: [10.1239/jap/1378401234](https://doi.org/10.1239/jap/1378401234).
- [219] Eric Zivot and Jiahui Wang. “Vector autoregressive models for multivariate time series”. In: *Modeling financial time series with S-PLUS®* (2006), pp. 385–429.
- [220] Gilles Zumbach. “Time reversal invariance in finance”. In: *Quantitative Finance* 9.5 (2009), pp. 505–515.
- [221] Gilles Zumbach. “Volatility conditional on price trends”. In: *Quantitative Finance* 10.4 (2010), pp. 431–442.
- [222] Gilles Zumbach and Paul Lynch. “Heterogeneous volatility cascade in financial markets”. In: *Physica A: Statistical Mechanics and its Applications* 298.3-4 (2001), pp. 521–529.