DISS. ETH NO. ………………

FINANCIAL MODELING OF BUBBLES AND CRASHES

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

MICHAEL GEORG SCHATZ

MSc Mathematics, ETH Zurich

born on 22.05.1990

citizen of Austria

accepted on the recommendation of

Prof. Dr. Didier Sornette

Prof. Dr. Martin Schweizer

2020

Abstract

The aim of this thesis is to develop and study mathematical models of financial markets experiencing bubbles and crashes. Our study is based on a model structure that combines increasingly positive returns (the build-up of a bubble) with a sudden reversal and an accumulation of large negative returns (the crash). The occurrence of financial crashes, fueled by bubbles, are widely recognized as disruptive events that have significant adverse effects on the performance of mathematical tools in many financial problem settings. However, models that explicitly incorporate the stylized features of bubble-and-crash dynamics are yet to permeate most areas of (mathematical) finance.

To address this shortcoming, in the first part of the thesis we develop a general stochastic model framework based on the dialectic view that build-up and implosion of financial bubbles are inherently linked. We show that the framework encompasses a wide range of models from the bubble literature in (financial) economics that focus on market failure as a driver of financial distress. As such, it functions both as a tool to compare the various assumptions underlying such models and as a good starting point to robustify financial problem settings that are affected by bubbles and crashes. Based on this general framework, we pick several specific models combining (possibly explosive) diffusions and single jump processes, study their stochastic and statistical properties and show how they can be applied to a common problem in mathematical finance, the pricing and hedging of European options. Where indicated, we apply our methods to financial data sets.

The second part of the thesis introduces a class of point processes that combines two well-established cluster mechanisms, a self-exciting structure commonly referred to as Hawkes process, and a shot-noise structure known as Neyman-Scott process. As such, these processes are useful in modeling the accumulation of points – for example, the clustering of negative returns during financial crashes within the framework mentioned above. However, given that such point processes transcend the mostly diffusive setting of the first part and show a similarity to ARMA time series of counts, which confirms their wide potential applicability beyond finance, this line of work deserves a distinct part in this thesis. In many cases, missing data complicates statistical estimation for this type of process, and explicit mathematical expressions exist only for (Markovian) special cases. To address this, we derive an Expectation-Maximization (EM) algorithm that allows us to estimate general inhomogeneous and non-parametric specifications and marks. We test the algorithm in a range of simulation and case studies.

Kurzfassung

Das Studienobjekt dieser Doktorarbeit sind mathematische Modelle von Spekulationsblasen und Crashs in Finanzmärkten. Die Struktur der Modelle, die wir untersuchen, basiert auf einer Kombination von stark ansteigenden Renditen (der Aufbau einer Blase) und einer überraschenden Trendumkehr mit grossen Verlusten (Crash). Es wird oft betont, dass das relativ häufige Auftreten solcher Kursverluste signifikante Auswirkungen auf das reibungslose Funktionieren mathematischer Modelle in finanztheoretischen Problemstellungen hat. Dennoch sind Lösungsansätze, die solche Blasen-Crash-Dynamiken explizit miteinbeziehen, in vielen Bereichen der Finanzmathematik unterrepräsentiert.

Um dieses Problem zu adressieren, entwickeln wir im ersten Teil dieser Arbeit ein allgemeines stochastisches Framework, basierend auf dem dialektischen Ansatz, dass der Aufbau einer Finanzblase und ihr Crash als zwei Seiten einer Medaille betrachtet werden müssen. Aus diesem Framework lassen sich viele in der Literatur gängige Modelle für Finanzblasen ableiten, die Marktineffizienz als zentrales Merkmal einer Finanzblase sehen. Als solches erlaubt unser Framework einerseits, zugrunde liegende Hypothesen verschiedener Modelle zu vergleichen, und andererseits, robuste Lösungsansätze für Problemstellungen in der Finanzökonomie zu gestalten. Ausgehend von diesem generellen Ansatz untersuchen wir stochastische und statistische Merkmale mehrerer Modelle, die explosive Diffusionsprozesse mit Sprungprozessen verknüpfen, und wenden sie auf ein klassisches Problem in der Finanzmathematik an, der Preisermittlung und Replikation von europäischen Optionen. Die entwickelten Methoden testen wir an Daten.

Im zweiten Teil der Arbeit untersuchen wir eine Art von Punktprozess, die zwei bekannte Häufungsmechanismen kombiniert: einerseits den Hawkes-Prozess (mit Selbstauslösung) und andererseits den Neyman-Scott-Prozess (mit Clustering). Als solche können diese Prozesse zwar nützlich sein, um beispielsweise die Akkumulierung von Kursverlusten während eines Crashs zu modellieren, gehen allerdings über das Setting mit Diffusionsprozessen im ersten Teil hinaus und werden daher in einem separaten Teil der Arbeit behandelt. Die weite Anwendbarkeit wird durch eine Analogie zu ARMA Zeitreihen angedeutet. Statistische Schlussfolgerungen für diese Prozesse werden oft durch fehlende Daten erschwert und explizite mathematische Lösungsansätze gibt es bislang nur für Markov'sche Spezialfälle. Wir entwickeln daher einen EM-(engl.: expectation maximization) Algorithmus, der statistische Rückschlüsse für eine Vielzahl von zeitinhomogenen und nicht-parametrischen Spezifikationen und Modelltypen erlaubt. Der Algorithmus wird in mehreren Simulations- und Fallstudien untersucht.

Acknowledgments

First and foremost I would like to thank Didier Sornette, my doctoral advisor, who has been an endless source of ideas and a constant challenger of my thoughts and methods, as well as an impressive role model of an interdisciplinary researcher with diverse interests, not lacking the ability to focus and deep-dive. I would like to thank Spencer Wheatley – coauthor, colleague, friend and mentor – from whom I learned a lot about statistics, discipline, quality food and the fact that being an academic is much more than writing scientific papers. Many thanks to Li Lin for a fruitful collaboration and Rick Paik Schoenberg for kindly hosting me at UCLA statistics department and the ability to learn from his immense knowledge on point processes. I am very grateful for many discussions with Martin Herdegen, Martin Schweizer, Sebastian Herrmann, Josef Teichmann and the participants of the German Statistics and Stochastic Days and the Bachelier Finance Society World Congress, during which I started to digest the mathematical-finance-perspective of things and learned that there is no substitute for mathematical rigor in stochastic and statistical modeling. I want to especially thank Martin Herdegen for his patience with my questions and excellent comments on my research.

Shout-out to my office colleagues Ke, Ali, Dongshuai, Richard, Guilherme, Herman, Tobias, Rebecca, Chahat, Ran, Giuseppe, JC, Sumit and Alexander, for excellent discussions on and off topic and the productive, casual atmosphere in the office. I am especially grateful for numerous discussions on (macro-)economics with Richard Senner, leading to the conclusion that applying mathematical models – no matter how rigorous – to financial data is always political in its conclusions and implications, just like all of (financial) economics. And finally, for their permanent, unconditional support during all these years and the constant reminder that math is not everything, I am indebted to Verena, my parents, and friends throughout Switzerland, Austria and the world.

Financial support by the Swiss National Science Foundation (grant No. 2-77156-16 entitled *Financial mathematics of positive-feedback bubbles and crashes*) for parts of this thesis is gratefully acknowledged.

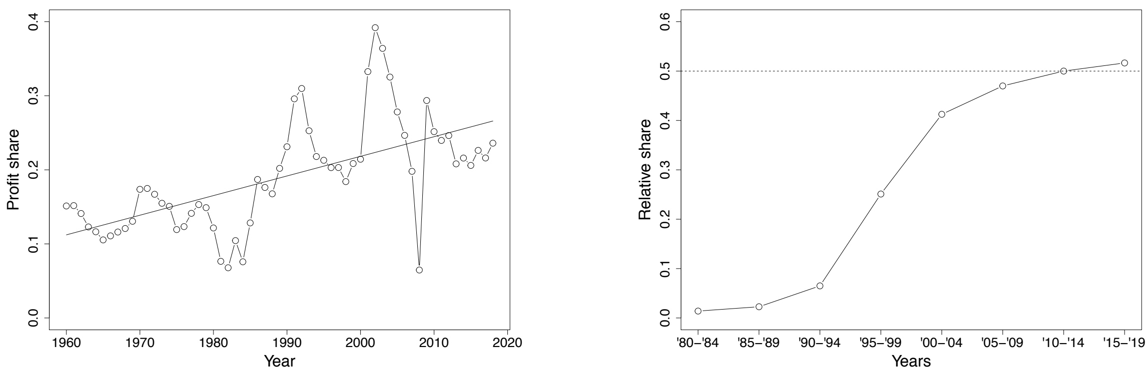
Contents

Part I

Introduction

Barter economies, financial crises and mathematical models

During the past several decades, the financial sector has become an increasingly substantial part of most capitalist economies. A common illustration of this phenomenon, which has been termed *financialization*, is the profit share of financial firms in the U.S., increasing both in mean and volatility, as depicted in figure 1a. This development, nevertheless, is dwarfed by the shift in attention and interest of academic theorists and mathematicians working with economic models. *Mathematical finance*, effectively a sub-discipline of *mathematical economics*, has become a dominant object of research, see figure 1b. At first, a possible explanation that comes to mind is that finance, as a growing and "innovative" industry, simply offers a larger playing field for ambitious modelers.



(a) Profit share of financial corporations in the U.S. economy, excluding Federal reserve banks. *Source: bea.gov, U.S. Department of Commerce.*

(b) Relative share of search results for "Mathematical finance" as opposed to "Mathematical economics". *Source: Google scholar.*

Upon a closer look, it seems that purely financial problems are a more adequate fit for the mathematical tools employed, naturally attracting more attention. Indeed, one of the most celebrated results of mathematical economics, the theory of general equilibrium by Arrow-Debreu [11] with an extension to a stochastic setting by Radner [206], seems to work better in finance than it does in economics. This (theoretically) very deep and intriguing result studies what is called a *pure exchange*, or *barter*, economy, which treats money as neutral and has been criticized for lacking the means to model the labour market or real economic variables like consumption, aggregate demand and the availability of money and credit. Trading in a financial market, however, can be more adequately modeled as a pure exchange economy. In fact, the most central concepts in (mathematical) finance, the efficient market hypothesis (the hypothesis that asset prices multiplied by appropriate state price densities follow a martingale – [89, 210]) and the fundamental theorem of asset pricing (proving that this martingale property is in fact equivalent to the absence of riskless gains by trading – [70, 119]) are essentially equivalent to the existence of an Arrow-Radner equilibrium (see, e.g., [81] or [149]). These concepts and results proved fruitful in all kinds of financial problem settings – modeling of assets prices, pricing derivatives, managing risks, utility maximization, (mark-to-market) accounting – and are utilized in areas as diverse as risk-adjusted compensation of company executives and funded retirement plans. Finance – especially the mathematics-infused part of the discipline – does a great job if it can detach itself from issues that go beyond a pure exchange economy, that is, if real economic variables are sufficiently well-behaved and market participants are rational.

Unfortunately, over the course of history, several disruptive events – *financial crises* – have shown that this detachment can utterly fail. Some economic theorists have studied this possibly toxic co-dependence of financial markets and their underlying economy. Hyman Minsky [185], extending ideas from John Maynard Keynes, elaborates at length how economic prosperity (of all things!) can lead to a substantial change in the credit structure and financial instability. Going a step further, Charles Kindleberger [160] expounds how private money can add to the (bank-)credit-fueled boom-bust dynamics of Minsky, causing what he calls a *speculative mania*, also known as *overtrading* or simply *financial bubble*. At a micro-level, this idea of speculative trading has been made explicit in models of financial market participants, using asymmetric information, heterogeneous beliefs and positive feedback activity to show that asset prices can rise beyond reasonable levels, creating the possibility of a sudden crash. A detailed account of many of these mechanisms, with a focus on positive feedback activity, can be found in [225]. In effect, these departures from rational pure exchange behavior lead to inhomogeneity, unstable "runaway" dynamics, disruptive regime changes and singularities in asset prices and financial key variables, which we will collectively refer to as bubble-and-crash dynamics.

This thesis contributes to the literature in (mathematical) finance by studying properties of processes that explicitly incorporate such bubble-and-crash dynamics. The main reason for using these processes is to robustify the use of mathematical models in financial applications. At the very best, these models can provide a predictive edge and allow one to model the probabilistic structure of a financial bubble-and-crash episode. At least, they extend the financial economist's tool-kit with models that are not oblivious to the possibility of dramatic, disruptive events. Our main object of interest is a stochastic model for a price time series $(S_t)_{t\in[0,\infty)}$ of any stock or financial asset. Chapter 1 introduces a simple framework based on three *bubble-characteristics*.

- A pre-drawdown process, $\tilde{S}$, which serves as a model for bubble dynamics and can incorporate superexponential growth and explosive behavior.
- The random time of the drawdown, τ_J, being essentially what has become known as the *Minsky moment*, the time when instabilities built up in the background materialize. To call on Keynes' analogy of musical chairs, it is the time when the music stops.
- The shape of the drawdown, X, which serves as a model for a financial crash. The simplest version of a drawdown is an instant drop in asset price values at τ_J.

These three bubble characteristics are linked to describe the full stock price S as

$$S_t = \tilde{S}_t \mathbb{1}_{\{t<\tau_J\}} + X_t \mathbb{1}_{\{\tau_J \leqslant t\}}, \quad t \in [0,\infty). \tag{1}$$

Based on this decomposition, subsequent chapters study modeling aspects related to special cases of (1). Chapters 2 and 3 are concerned with jump-diffusion models of S, where the crash is represented by a single jump, while chapter 4 studies a model for a pre-drawdown process $\tilde{S}$ based on momentum and trend-following behavior.

The last chapter 5 concerns point processes, a model class that deals with the accumulation and clustering of point occurrences in time. They are a useful modeling tool in situations where the occurrence of certain events is believed to trigger further events. As such, they seem to be a promising tool to model the drawdown X in (1) as the accumulation of large negative price changes. We introduce a point process analogue of well-established ARMA (autoregressive moving-average) time series, discuss some of its properties and derive an algorithm for (non-)parametric statistical inference. Due to the generality of this model class with many applications beyond finance and the statistical focus of this chapter, it deserves a distinct part in this thesis, only loosely connected to chapters 1–4.

Before we review in detail the contributions of this thesis, a word of caution is in order. The models studied here are concerned with behavior of a single asset price. While these dynamics are often informed by more complex models operating in the background, it should be clear that, in order to fully understand financial crises, none of these models is sufficiently wholesome to even remotely substitute for a thorough qualitative analysis as put forward by economists like Hyman Minsky or Charles Kindleberger. Key drivers of financial systems such as liquidity structures, credit and debt, leverage, yields, interest rates or the level of interconnectedness of financial institutions contribute towards a better understanding of financial distress, with signs of excessive growth in asset prices being only one possible indicator. That said, many applications in financial engineering start with postulating a certain stochastic model for asset prices. We believe they can benefit from incorporating bubble-and-crash price dynamics into their model setting.

A detailed overview of the thesis

This is a cumulative thesis, which is based on a series of published, submitted and unpublished papers, all of which are joint work with Didier Sornette. Chapters 1,2 and 3 are first-authored and based on the preprint [213], the publication [212] and unpublished work, respectively. Chapter 4 is published in [175] and joint work with Lin Li. My contributions include a major revision of an earlier model setup, joint writing of the preliminary setting and derivations in section 4.2, writing of the setting and derivation of the model properties in section 4.3, as well as formulating and proving the results in section 4.6. Chapter 5 (jointly first-authored with Spencer Wheatley) is an extension of the preprint [249]. My contributions are substantial in all parts of the paper that go beyond an early draft in [248].

1 A mathematical framework for inefficient market bubbles. In this chapter we propose a general framework that encompasses a wide range of financial bubble models, sharing the view that *bubbles are driven by market failure*. Possible mechanisms that have been noted in the literature include asymmetric information, heterogeneous beliefs, noise trading and positive feedback mechanisms. The object of interest is a stochastic model for the price time series $(S_t)_{t\in[0,\infty)}$ of any stock or financial asset. After a discussion of rational expectation bubbles in section 1.2, in section 1.3 we present a simple framework based on three *bubble-characteristics* $(\tilde{S}, \tau_J, X)$ as introduced above (see bullet points preceding equation (1)). We combine this decomposition with the abstract notion of market efficiency – which will take familiar forms in many examples – and require that

$$\tilde{S} \text{ violates market efficiency.}$$

This condition makes explicit the understanding that bubbles are driven by market failures.

This leads to two possible interpretations of inefficient market bubbles. *Type-I bubbles* require the full stock price S to be an efficient market. This view posits that inefficiencies lead to the presence of a drawdown at τ_J as well as exuberant behavior in $\tilde{S}$, which can, however, be interpreted as two sides of the same coin and do in fact balance each other. *Type-II bubbles* require the drawdown process X to be an efficient market. This view posits that inefficiencies lead to irrational behavior and overvaluation that is eventually corrected at τ_J.

Sections 1.4 and 1.5 present several examples in both continuous and discrete time and compare our approach to rational expectation bubbles. Section 1.6 discusses implications for bubble detection and bubble modeling.

2 Uniform integrability of single jump processes. This chapter deals with a special case of a Type-I inefficient market bubble that can be described by a stochastic differential equation

$$dS_t = b(S_t)\mathbb{1}_{\{t<\tau_J\}}dt + \sigma(S_t)1_{\{t<\tau_J\}}dW_t - \frac{b(S_{t-})}{h(S_{t-})}dJ_t, \tag{2}$$

with initial value $S_0 \in (0, \infty)$ and coefficient functions $b, \sigma, h : \mathbb{R} \to [0, \infty)$, where $(W_t)_{t\in[0,\infty)}$ is a Brownian motion, $(J_t)_{t\in[0,\infty)}$ is a $\{0, 1\}$-valued single jump process with

$$\mathbb{P}\left[dJ_t = 1 | \mathcal{F}_{t-}, J_{t-} = 0\right] = h(S_{t-}), \tag{3}$$

such that h is the *hazard rate* of J, and τ_J denotes the time of the jump. By construction, this process is a so-called supermartingale with the property that $\mathbb{E}\left[S_\infty\right] = \mathbb{E}\left[S_{\tau_J}\right] \leqslant S_0$, and the essential question of interest is whether or not it is a uniformly integrable martingale, that is, whether

$$\mathbb{E}\left[S_\infty\right] = S_0 \quad \text{or} \quad \mathbb{E}\left[S_\infty\right] < S_0. \tag{4}$$

To answer this question, we derive a deterministic condition on the functions b, σ and h in section 2.4. Section 2.5 covers examples from the literature and discusses our result from the viewpoint of rational expectation bubbles.

3 Quadratic hedging during financial bubbles. This chapter starts from the understanding of a bubble expounded in chapter 1 to tackle a standard problem in mathematical finance using a well-established bubble model. In particular, we study replication of derivative payoffs through dynamic trading in a well-known example of a Type-I bubble, the Johansen-Ledoit-Sornette (JLS) model. It describes a bubbly price process as the solution $(S_t)_{t\in[0,T]}$ of a stochastic differential equation

$$\frac{dS_t}{S_t} = \kappa h(t)\mathbb{1}_{\{t<\tau_J\}}dt + \sigma dW_t - \kappa dJ_t, \quad t \in [0,T], \tag{5}$$

where W is a real-valued Brownian motion and J is a single jump process that jumps from 0 to 1 at time τ_J with deterministic hazard rate h. The essential building block of the JLS model is an explosive[1] structure of the hazard rate, which is derived based on the analysis of herding behavior and positive feedback in a network of traders. As such, the model is a typical example of a Type-I bubble in the framework of chapter 1. For a European contingent claim $f(S_T)$ with deterministic payoff function $f : [0,\infty) \to [0,\infty)$, we study the quadratic hedging problem

$$\min_{(V_0,\theta)} \mathbb{E}\left[\left(f(S_T) - (V_0 + \int_0^T \theta_t dS_t)\right)^2\right], \tag{6}$$

over all possible portfolios with initial value $V_0 \in (0,\infty)$ and stock holdings $(\theta_t)_{t\in[0,T]}$.

We derive explicit results for the process in (5) in section 3.3, relying heavily on the existing literature for the quadratic hedging problem (6). In section 3.4, we present a simulation study and a case study based on S&P500 stock market data of the 1987 *Black Monday* crash.

4 A simple bubble mechanism: time-varying momentum horizon. In this chapter we develop a model that makes explicit one of the bubble drivers mentioned above: positive feedback caused by momentum trading strategies and behavioral characteristics such as over-reaction and herding. We describe an asset price process S and its momentum X as a two-dimensional stochastic process $(S_t, X_t)_{t\in[0,\infty)}$ where S is a exponential diffusion with linear drift and the price momentum X is calculated by geometric averaging with time horizon $1/\theta_t$, that is,

$$dX_t = -\theta_t X_t dt + \theta_t \frac{dS_t}{S_t}. \tag{7}$$

An essential feature of the model is that we use a varying time horizon $(\theta_t)_{t\in[0,\infty)}$ that depends on the momentum in a linear specification

$$\theta_t = \theta^* + \eta X_t, \tag{8}$$

for $\eta \in [0,1)$ and $\theta^* \in (0,\infty)$. Using the idea that investors focus on the momentum as a leading indicator of pricing we restrict to the special case $S_t = \lambda(X_t)$ for a deterministic function λ. Section 4.2 shows that this is the case only if drift and diffusion coefficients of S depend solely on X and S can be written as

$$S_t = \text{const} \times \left(1 + \frac{\eta}{\theta^*}X_t\right)^{1/\eta}, \quad t \in [0,\infty). \tag{9}$$

In section 4.3 we discuss the behavior of resulting diffusion processes, which foremost depends on whether $\eta = 0$ (mild behavior) or $\eta > 0$ (wild behavior and possible explosion). Section 4.4 discusses a calibration procedure and implements statistical tests for bubble detection.

[1] By *explosive*, here we mean finite time divergence in the sense that $\lim_{t\to T} h(t) = \infty$.

5 The ARMA point process and its estimation. This chapter introduces a point process that extends both the well-known Hawkes process and the Neyman-Scott process. In particular, we define the ARMA point process (ARMApp) as a measure-valued counting process $N : \Omega \times \mathcal{B}(\mathbb{R}) \to \mathbb{N}_0$ that combines

- a self-exciting structure, where each point triggers new points with intensity $\theta : [0, \infty) \to [0, \infty)$, and
- a standard cluster structure, where each point of some immigrant process N^μ triggers new points with intensity $\phi : [0, \infty) \to [0, \infty)$. N^μ is usually a Poisson process with possibly time-dependent intensity μ.

A useful tool in point process modeling is the intensity function, which for the ARMApp is given by

$$\lambda(t) = \mu(t) + \int_{-\infty}^{t} \theta(t-s)dN_s^\mu + \int_{-\infty}^{t} \phi(t-s)dN_s. \tag{10}$$

Section 5.3 discusses similar processes in the literature and provides integral equations for second order statistics, with explicit solutions in case of exponential intensities θ and ϕ. We justify the name ARMA point process by comparison with *integer-valued autoregressive moving-average time series* in section 5.4. A set of heuristic arguments leads us to a conjecture (without proof) on their equivalence in terms of limiting distributions. The main part in section 5.5 derives an MCEM (Monte-Carlo Expectation Maximization) algorithm for the ARMApp and various extensions. We test the performance of the algorithm in a simulation study in section 5.6 and a case study using S&P500 E-mini futures data in section 5.7.

Part II

Modeling financial bubbles and crashes

Chapter 1

A mathematical framework for inefficient market bubbles

1.1 Introduction

The term *asset price bubble* typically describes a situation where the market price of an asset exceeds its fair price, referred to as the asset's *fundamental value*. Asset price bubble models can essentially be grouped into two main paradigms – those where bubbles appear in an efficient market and those that are based on a violation of market efficiency.

The first group has become known as *rational expectation bubbles*, which tries to square the idea of asset over-valuation (and thus an inefficiency) with efficient markets. This inherent discrepancy prohibits the existence of bubbles in rational market equilibria and leads to joint hypothesis issues and internal inconsistencies in bubble detection; see section 1.2.1 below. While these problems have been discussed in the literature for discrete time interpretation of the rational expectations paradigm, we argue, using results from the literature, that they also permeate the more recent continuous time modeling approach. Despite these issues, presumably because such models can readily be described in a simple, unified framework, rational expectation bubbles are still used extensively in both discrete and continuous time models, ranging from economics and econometrics to mathematical finance.

Meanwhile, we have seen a flurry of models that are based on the understanding that bubbles are caused by a break of market efficiency in the sense of *perfect information rational expectations*. Various mechanisms have been proposed – such as asymmetric information [5, 6], heterogeneous beliefs [118, 214] and noise trading such as positive feedback activity [69, 221, 225] in combination with limits to arbitrage [1, 68, 222, 223]. In one way or another, such mechanisms create unsustainable behavior in asset prices and the risk of a crash. To capture the essence of such models, we propose to dissect the stock price into three so-called *bubble characteristics* –

1. a pre-drawdown process $\tilde{S}$,
2. the random time of the drawdown τ_J and (1.1)

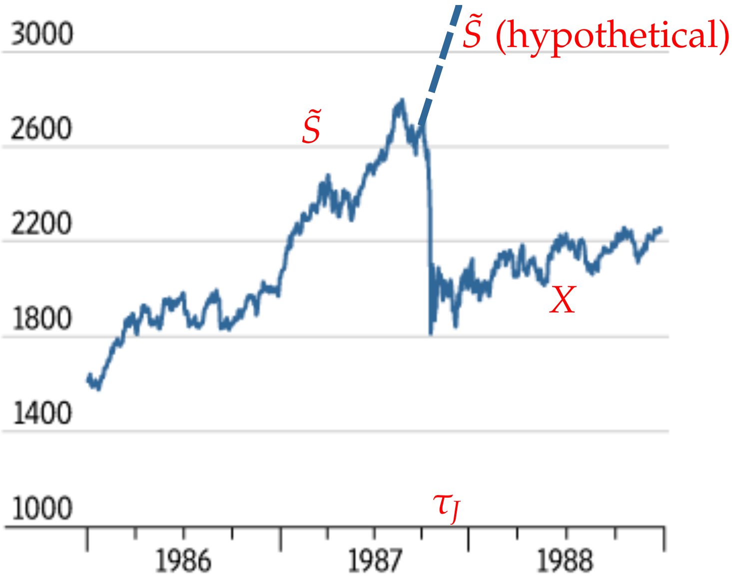


Figure 1.1: The Dow Jones stock price bubble and its bubble characteristics $(\tilde{S}, \tau_J, X)$. *Source: WSJ Markets Data Group.*

3. the shape of the drawdown X.

These three bubble characteristics allow one to describe the full stock price $(S_t)_{t\in[0,\infty)}$ as

$$S_t = \tilde{S}_t \mathbb{1}_{\{t<\tau_J\}} + X_t \mathbb{1}_{\{\tau_J \leqslant t\}}, \tag{1.2}$$

following $\tilde{S}$ up to the crash at τ_J and X afterwards. Figure 1.1 shows such a decomposition for the Dow Jones stock price bubble and crash of 1987. To complement our framework, we need a notion of *market efficiency*. We include such a notion through an abstract condition, which distinguishes efficient from inefficient markets and allows us to classify the bubble characteristics (1.1). The pre-drawdown process $\tilde{S}$ is understood to show exuberant, excessive behavior, thus we require that

$$\tilde{S} \text{ violates market efficiency,} \tag{1.3}$$

in a suitable way made precise below. The drawdown (τ_J, X) is understood as a mechanism to restore market efficiency, where two main bubble types can be extracted from the literature, gradually departing from rational expectation bubbles.

1.1.1 Type-I bubbles.

We can model the risk and shape of a drawdown (τ_J, X) being caused by an inefficiency, while extraordinary (pre-drawdown) returns ensure that this risk is "appropriately" priced it in the market: in this case,

$$\text{the full price process } S \text{ is an efficient market,} \tag{1.4}$$

while both the pre-drawdown process $\tilde{S}$ and the drawdown process X – seen in isolation – are inefficient. We call this a type-I bubble.

This understanding of a bubble has been put forward in section (B) of [27] and continued in [153, 155]. Note that, while originally this type of bubble has been discussed within the rational expectations framework, the exact valuation of fundamentals plays a minor role here. The essence of this understanding of a bubble can be captured by a quote from [27], that "during the duration of the bubble, [the bubble process] is growing faster (...) because asset holders have to be compensated for the probability of a crash." What drives a type-I bubble is the mere existence of the risk of a crash (captured in our framework by τ_J and X), compensated by a pre-crash process that earns abnormal returns ($\tilde{S}$ violating market efficiency) such that the full price process S is an efficient market.

This conception of a bubble resolves an important methodological issue of (classic) rational expectation bubbles: a type-I inefficient bubble is verified by the occurrence of a crash and does not rely on an infinite payoff structure. However, in such models the precise mechanism of how *rational market participants are aware of the crash risk and price it in the market* remains unclear. While a rational, omniscient investor would stay invested in such a market, the underlying stock price S needs to be postulated a priori.

1.1.2 Type-II bubbles.

In order to avoid this issue of postulating market efficiency for S to accommodate rational investors, we can either look at markets that are not populated by rational investors with symmetric information or add limits to arbitrage. In the absence of rational investors going against exuberant development, $\tilde{S}$ is driven by some failure of market efficiency. To arrive at a full-fledged bubble model, we assume that the very cause for this inefficiency disappears/gets resolved at τ_J and price levels collapse to "appropriate" levels: in this case, within the decomposition (1.2),

$$\text{the drawdown process } X \text{ is an efficient market,} \tag{1.5}$$

while the full process S is inefficient. We call this a type-II bubble.

This representation captures the view for instance espoused by [90] and [196] *(inefficiency: (ex-post) overestimation of future payoffs)*, [214] *(inefficiency: heterogeneous beliefs and overconfidence)* or [1] *(inefficiency: noise trading and synchronization risk)*. While the underlying mechanisms causing a break of market efficiency are quite diverse,[1] the nature of the crash in each case is captured by decomposition (1.2) and the time of the crash represents the time where rationality is restored. Such models avoid the above problem of assuming an efficient market process S, as noted above for type-I bubbles. However, for type-II bubbles the efficient process X (and thus a model for S) can in general only be determined under additional assumptions on the assets payoff structure. Note that in case one is interested only in why market efficiency can fail (that is, only in the

[1]The reader may be surprised to see [90] in the list of references to underpin inefficient market bubbles, as the author is essentially arguing that markets are efficient before, during and after the crash. However, a (hypothetical) omniscient observer with a proper valuation of future payoffs correctly sees a failure of market efficiency, which is restored only after the crash. For details see section 1.4.3.3.

process $\tilde{S}$) and not in the specific role of a drawdown, then the distinction between type-I and type-II bubbles is not relevant. This includes many of the early papers, e.g., [6, 68, 69].

1.1.3 Bubbles as overvaluation.

One may like to think of bubbles as overvaluation, and for both bubble types irrationality in the pre-drawdown process $\tilde{S}$ leads to a valuation that is *too high* in the following sense.

1. In a type-I bubble, the asset is overvalued for those who mistake $\tilde{S}$ for the true dynamics of the price process. Projecting the pre-drawdown stock dynamics to final payoffs (ignoring the risk of a drawdown)[2], one overvalues the true payoff of the asset. An omniscient observer, assessing the correct underlying S, does not see an overvaluation, only the risk of a drawdown compensated by high returns. A rational, omniscient investor stays invested.
2. In a type-II bubble, the asset is overvalued, even if the true dynamics S are known – X represents the correct valuation of the final payoff. The omniscient observer sees a bubble and, consequently, the risk of a drawdown as possible revaluation. A rational, omniscient investor would exit the market eventually, depending on the likelihood of possible crash times.

Below we provide a simple framework for such bubble characteristics $(\tilde{S}, \tau_J, X)$ and market efficiency conditions that allows one to accommodate, compare and ultimately justify several approaches in the literature. Many of these models have not yet been integrated in a wider framework and have been perceived as quite diverse attempts to bubble modeling, thus preventing efficient dialogue between research disciplines. Our framework is proposed as a way to strengthen the understanding of bubbles as a failure of market efficiency and serve as a basis for simple models that – contrary to rational expectation models – are based on *the* essential property of a bubble: the possibility of a drawdown fueled by market failure.

In section 1.2 we discuss alternative approaches in the literature – with a focus on rational expectation bubbles and their shortcomings.[3] In particular, we collect results from the literature that characterize rational expectation bubbles in both discrete and continuous time. In section 1.3 we provide a general framework for the bubble characteristics (1.1) and the two types of inefficient market bubbles discussed above. Sections 1.4 and 1.5 discuss applications and examples of inefficient market bubbles in continuous and discrete time, respectively. We end each section with a comparison to the rational expectations framework, highlighting the main differences and possible similarities. Exploiting the simple structure of our framework, we close with a discussion of issues in bubble detection and bubble modeling and some further research questions in section 1.6. The appendix in section 1.7 summarizes some auxiliary definitions and compiles our technical results.

[2] In other words, this means operating under the belief that $\tau_J \equiv \infty$ or $X \equiv \tilde{S}$, almost surely.

[3] As the main purpose of the present paper is to introduce a novel framework, the discussion of rational expectation bubbles is not devised as a comprehensive literature review and limited to general ideas and seminal publications.

1.2 Review of bubble literature

Besides the literature mentioned in the introduction (modeling a break of market efficiency), the lion's share of bubble models emerges from the rational expectations framework. Thus, our discussion in this section will mainly focuses on this strand of the literature.

1.2.1 Rational expectation bubbles

This approach is based on the theoretical definition of an asset's fundamental value in rational expectation markets.

1.2.1.1 Definitions

Setting (Rational expectation bubbles). For simplicity, we assume appropriately discounted prices and dividend streams and thus 0 interest rate.[4] Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,\infty]}, \mathbb{P})$ be a filtered probability space satisfying the usual conditions, $\tau : \Omega \to [0, \infty]$ the lifetime of a non-negative asset $(S_t)_{t\in[0,\tau]}$ with cumulative dividend stream $(D_t)_{t\in[0,\tau]}$ and a final payoff S_τ. Then, following [147], one can define the wealth process $(V_t)_{t\in[0,\tau]}$ of the asset by

$$V_t = S_t + D_t, \quad t \in [0, \tau]. \tag{1.6}$$

Finally, we assume that both S and D are RCLL semimartingales and that *rational expectations* holds in the form of *absence of arbitrage opportunities*. To describe this in our general continuous setting, we use the condition of *No Free Lunch with Vanishing Risk (NFLVR)* and the notion of *self-financing, admissible* trading strategies; see [71] for details.[5] A general version of the first fundamental theorem of asset pricing in [71] then implies that there exists a nonempty set $\mathcal{Q}$ of $\mathbb{P}$-equivalent measures such that V is a $\mathbb{Q}$-local martingale for every $\mathbb{Q} \in \mathcal{Q}$.

All definitions above include discrete time models by using $\tau : \Omega \to \mathbb{N} \cup \{\infty\}$ and the index set $[0, \tau] \cap \mathbb{N}$.

Fundamental values and bubble definition. If we make the additional assumption that the market generated by S is complete (that is, $\mathcal{Q}$ is a singleton), then the fundamental value of an asset can simply be defined as

$$S_t^* = E_{\mathbb{Q}}\left[(D_\tau - D_t) + S_\tau \mathbb{1}_{\{\tau<\infty\}} | \mathcal{F}_t\right], \quad t \in [0, \tau], \tag{1.7}$$

the expectation of future payoff streams under the unique market implied probability $\mathbb{Q}$. In general, however, the market generated by S is incomplete and the straightforward definition (1.7) does not extend immediately. We can distinguish two notions of fundamental value and use the naming from [128].

[4] This assumption requires the a priori choice of a bank account, which is bounded from below and can be used as a numeraire. Thus, it is inflicted with a certain loss of generality; see, e.g., the discussion in [123].

[5] Below we always refer to self-financing, admissible strategies when using the terms *trading* and *replication*.

Definition 1.2.1 (Strong rational expectation bubble). Assume the setting above. Then the fundamental value of an asset is defined as

$$S_t^* = \operatorname*{ess\,sup}_{\mathbb{Q}\in\mathcal{Q}} E_{\mathbb{Q}}\left[(D_\tau - D_t) + S_\tau \mathbb{1}_{\{\tau<\infty\}}|\mathcal{F}_t\right], \quad t \in [0,\tau]. \tag{1.8}$$

As above, we say that S has a rational expectation bubble at t if $S_t^* < S_t$ and call $B = S - S^*$ the bubble component of S.

Definition 1.2.2 (Rational expectation $\mathbb{Q}$-bubble). Assume the setting above. For some $\mathbb{Q} \in \mathcal{Q}$, we define a $\mathbb{Q}$-fundamental value as

$$S_t^{\mathbb{Q}} = E_{\mathbb{Q}}\left[(D_\tau - D_t) + S_\tau \mathbb{1}_{\{\tau<\infty\}}|\mathcal{F}_t\right], \quad t \in [0,\tau]. \tag{1.9}$$

We say that S has a rational expectation $\mathbb{Q}$-bubble at t if $S_t^{\mathbb{Q}} < S_t^* = S_t$ and call $B^{\mathbb{Q}} = S - S^{\mathbb{Q}}$ the $\mathbb{Q}$-bubble component of S.

Remark 1.2.1.

(a) In the case of a complete market, the definitions collapse to the fundamental value in (1.7). It is clear that, if S has a strong bubble, then S has a $\mathbb{Q}$-bubble for every $\mathbb{Q} \in \mathcal{Q}$.

(b) The fundamental value in (1.8) has the the *superreplication property* S. By [165]'s optional decomposition, it is equal to the cheapest starting capital that allows one to replicate all payoffs of S on $[t,\tau]$ and, as such, has a solid economic interpretation as *fundamental* value. For a general discussion of fundamental values; see, e.g., section 6.1 in [128]

(c) [211] were among the first to recognize that the fundamental value does not immediately extend from complete markets and discuss both definitions in incomplete markets; in fact, many of the early papers on rational expectation bubbles ignore this issue. Subsequently [179], [130], [180] and [128] use the superreplication price (1.8) and the corresponding bubble definition,[6] while [148] and [21] use a dynamic version of (1.9).

Example 1.2.1.

(a) **Special case: discrete time bubbles.** The classical framework is based on an infinite time horizon $\tau \equiv \infty$ and uses a discrete set of trading dates; see, e.g., [98], [27] and [78]. [40] provides an excellent overview of the early literature, [107] a recent overview. To arrive at the classical framework in the setting above, we add the structural assumption that for some $R \in [0,\infty)$ there exists a "discounting" measure $\mathbb{Q}_R \in \mathcal{Q}$ with the property that, for all $n, \tau \in \mathbb{N}$,

$$E_{\mathbb{Q}_R}[S_{n+\tau}|\mathcal{F}_n] = \mathbb{E}\left[\frac{S_{n+\tau}}{(1+R)^\tau}\Big|\mathcal{F}_n\right] \quad \text{and} \quad E_{\mathbb{Q}_R}[D_{n+\tau} - D_{n+\tau-1}|\mathcal{F}_n] = \mathbb{E}\left[\frac{D_{n+\tau} - D_{n+\tau-1}}{(1+R)^\tau}\Big|\mathcal{F}_n\right]. \tag{1.10}$$

[6] Let us note that [128] use a dynamic version of (1.8), which allows for the concept of *bubble birth*.

The measure $\mathbb{Q}_R \in \mathcal{Q}$ discounts future prices and payoffs and ensures a constant required rate of return R. Using equation (1.10) and the $\mathbb{Q}_R$-martingale property of the value process $V = S + D$ yields the standard one-step No-Arbitrage condition, cf. chapter 7 of [41],

$$(1+R)S_n = \mathbb{E}\left[S_{n+1} + D_{n+1} - D_n | \mathcal{F}_n\right], \quad n \in \mathbb{N}. \tag{1.11}$$

Assuming the market is complete, R defines a unique martingale measure $\mathbb{Q}_R$ and the fundamental value (1.7) takes the well-known form of future discounted dividend payments, that is,

$$S_n^* = E_{\mathbb{Q}_R}\left[(D_\infty - D_n)|\mathcal{F}_n\right] = \mathbb{E}\left[\sum_{k=1}^{\infty} \frac{D_{n+k} - D_{n+k-1}}{(1+R)^k} \middle| \mathcal{F}_n\right], \quad n \in \mathbb{N} \cup \{\infty\}. \tag{1.12}$$

The assumption of completeness, however, is often implicit and unjustified; see example 4.4 in [211].

(b) **Special case: strict local martingale price process.** In the seminal paper of [179], the authors have realized that the definition of a fundamental value in equation (1.8) allows for bubbles even when restricting to a finite time horizon $\tau \equiv T \in (0, \infty)$. This is possible only for continuous time models and thus a priori excluded in the classical bubble literature. Assume the setting above, zero dividend payments and, for simplicity, that the market is complete. Then we have a single payoff S_T and the fundamental value is given by $S_t^* = E_{\mathbb{Q}}\left[S_T|\mathcal{F}_t\right]$. The essential notion is that of a strict local martingale.

Definition 1.2.3 (Strict local martingale). A local martingale S is a strict local martingale if it fails to be a martingale.

As such, strict local martingales are fair games locally but fail to satisfy the martingale property. Thus, if S is a strict local martingale, we have

$$E_{\mathbb{Q}}\left[S_T|\mathcal{F}_0\right] < S_0 \tag{1.13}$$

and a strong bubble according to definition 1.2.1.

To understand the underlying mechanism, we can invoke [165]'s optional decomposition theorem, which tells us that there exists a trading strategy with initial capital $E_{\mathbb{Q}}\left[S_T|\mathcal{F}_0\right]$, replicating S_T almost surely. In addition, a trivial buy-and-hold strategy on $[0, T]$ with initial capital S_0 is always a replicating strategy; see figure 1.2 for a visualization. Note that the existence of two such strategies does not interfere with No-Arbitrage (NFLVR), as exploiting the difference would require a net short position in the asset, which, as a strict local martingale, is necessarily unbounded. Unbounded portfolio losses, however, are excluded by the admissibility condition.

1.2.1.2 Modeling rational expectation bubbles

One may be tempted to formulate the following for *rational expectation markets*.

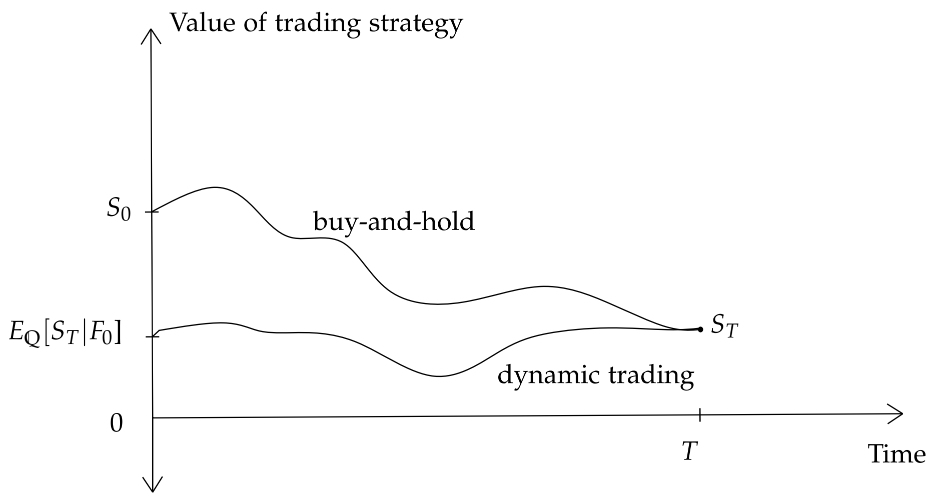


Figure 1.2: Sample paths of two self-financing, admissible strategies with final payoff $H = S_T$ in a strict local martingale model.

> *Given an asset S with present market price S_0 that entitles the holder to a series of diverse payoff streams during its lifetime, then, 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.*

A deviation of the fundamental value from an asset's market price in a rational expectation market seems thus inconsistent with the notion of a market equilibrium populated by rational traders. Making precise this inconsistency in a rational equilibrium model of traders, a celebrated result by [240] for strong rational expectation bubbles is often used in the literature to (theoretically) dismiss the existence of discrete time bubbles. This reasoning can be extended to bubbles in continuous time, as we argue below. These arguments do not apply to the case of $\mathbb{Q}$-bubbles that fail to be strong bubbles, which we discuss in a separate paragraph.

Strong rational expectation bubbles. The fundamental value (1.8) in definition 1.2.1 is the superreplication price associated to S. By [165]'s optional decomposition, S_t^* is equal to the cheapest starting capital that allows one to replicate all payoffs of S on $[t, \tau] \cap [0, \infty)$ and, as such, has a solid economic interpretation as *fundamental* value (see, e.g., the discussion in section 6.1 in [128]). This defining property of a strong rational expectation bubble implies that, given $S_t^* < S_t$, no investor preferring more to less would accept a price S_t for payoffs that may be replicated with strictly smaller initial capital S_t^*. In discrete time complete markets, this rationale was made precise in [240] (see the discussion therein and especially propositions 3 and 6).

Proposition 1.2.1. *Assume the setting above with a discrete set of trading dates, $\tau \equiv \infty$ and market completeness. Then in the following situations a strong rational expectation bubble (definition 1.2.1) cannot emerge as the equilibrium[7] price process of an economy populated by rational, utility-maximizing investors.*

1. *We have a finite number of short-run maximizing investors.*

[7] For definitions of equilibria with short-run maximizing investors (myopic REE) and long-run maximizing investors (fully dynamic REE) we refer to [240].

2. *We have any (finite or infinite) number of long-run maximizing investors.*

In theorem 3.1 of [211], this was extended to a slightly more general equilibrium framework and incomplete markets. For continuous time models in finite time $\tau \equiv T \in (0, \infty)$ (which includes strict local martingales as in example 1.2.1(b)), we can utilize and reformulate theorem 3.2 of [149].

Proposition 1.2.2. *Assume the setting above and let $\tau \equiv T \in (0, \infty)$. Then a strong rational expectation bubble (definition 1.2.1) cannot emerge as the equilibrium*[8] *price process of an economy populated by rational, utility-maximizing investors.*

While these negative results depend on the specific notion of *rational expectations equilibrium* employed and some leave room for bubbles in special types of models,[9] they show that strong rational expectation bubbles are necessarily driven by market inefficiencies that transcend equilibria in markets populated by rational, utility-maximizing investors.

In this sense, propositions (1.2.1) and (1.2.2) confirm the intuitive reasoning above that there is no such thing as a rational (expectation) bubble. To arrive at a price process that follows a strong rational expectations bubble, the behavior of agents need to depart *in one way or another* from the standard assumption – that is, there needs to be at least one agent that is not rational and/or not utility-maximizing. However, in this case we agree with [240] that "(...) research should be devoted to the explanation of actual price bubbles by non-rational behavior" and one should strive to explicitly model such behavior. Eventually, this insight leads to the models discussed in section 1.1 and, thus, is the cornerstone of our general framework introduced in section 1.3.

Rational expectation $\mathbb{Q}$-bubbles. In the above setting, assume for some $t \in [0, \tau)$ and $\mathbb{Q} \in \mathcal{Q}$ we have that

$$S_t^{\mathbb{Q}} < S_t^* = S_t. \tag{1.14}$$

Then S has a $\mathbb{Q}$-bubble but not a strong bubble. As the superreplication price S^* equals the asset price, in this situation the buy-and-hold strategy is the cheapest way of replicating the payoffs of S and the arguments for strong bubbles above do not apply.

However, there is a caveat: strict inequality in $S_t^{\mathbb{Q}} < S_t^*$ and the superreplication property imply that any investor seeking to replicate the payoffs of S on $[t, \tau] \cap [0, \infty)$ with initial capital $S_t^{\mathbb{Q}}$ will fail to do so with positive probability. In other words, no market participant is able to generate the payoff stream of S with initial capital $S_t^{\mathbb{Q}}$. The argument that the asset is *overpriced* at S_t does not hold in this situation and the relation of processes satisfying (1.14) to asset price bubbles (understood as mispricing) remains questionable.

1.2.1.3 Detecting rational expectation bubbles

If we choose to sideline these theoretical concerns, failing to explicitly model the generating mechanism of a bubble leads to challenges in the empirical detection of rational expectation bubbles.

[8] A market equilibrium in this setting is defined in section 2.7 of [149].

[9] See, for example, the overlapping generations model of [241] or examples 4.1-4.4 in [211].

Discrete infinite time. As mentioned in example 1.2.1(a), classical rational expectation bubbles rely on an infinite time horizon $\tau \equiv \infty$. This leads to several requirements on rational expectation bubbles that do not appear to be essential characteristics of asset price bubbles, but rather structural requirements of the infinite horizon framework.

1. As noted, assets with bounded lifetime cannot have a bubble. This implies that a test for bubbles is necessarily inconclusive before the asset has reached the end of its lifetime.
2. If one puts aside the latter point by *assuming* an infinite payoff structure, a bubble process B may be any martingale.

These points imply that the infinite horizon framework does not produce testable implications and additional hypothesis on the price and dividend structure are necessary for bubble detection. For a more detailed and extensive discussion of such *joint hypothesis* issues see, e.g., [40], [114] or [144].

As an example, a popular test for rational expectation bubbles is based on equation (1.11); see, e.g., [77], [198] or [134]. One assumes that one-period dividend payments $(d_t)_{t\in\mathbb{N}}$ follow a linear autoregressive process of the form

$$d_{t+1} = c + \delta d_t + \epsilon_t, \quad t \in \mathbb{N}, \tag{1.15}$$

for $\delta < 1$, $c \in \mathbb{R}$ and a sequence of normally distributed random variables $(\epsilon_t)_{t\in\mathbb{N}}$. Then equation (1.12) implies that the stock price follows a linear autoregressive process as well. Consequently, only if there is a bubble component B one will be able to detect non-stationary behavior $\delta > 1$ in a linear model of the form (1.15) for the stock price. Fortunately, such tests for non-stationarity, based on *unit-root tests* have been applied and developed for several decades and are very well understood (Dickey-Fuller, Phillips-Perron, etc.). Unfortunately, however reliably one can detect $\delta > 1$ in a linear specification (1.15), the existence of a bubble is based on a *joint hypothesis* on required returns, dividend streams and the stock price. In other words, non-stationary behavior can be attributed to a misspecified model of future dividends, time-varying/stochastic expected returns *or* a bubble component. In section 1.5.3.1 we provide a detailed comparison of this approach with our framework and review some of the issues raised above in greater detail.

Continuous finite time. Example 1.2.1(b) discusses the fact that, for a finite deterministic lifetime $\tau \equiv T \in (0, \infty)$, S has a bubble if and only if it is a strict local martingale. There have not been many attempts to apply the strict local martingale approach for the purpose of bubble detection. One that is worth mentioning is based on the strict local martingale property of homogeneous diffusions with explosive volatility; see [145], [146] and chapter 4 of [205]. In this specific framework, bubble detection is possible if and only if the volatility increases sufficiently *before* a crash happens. However, [232] analyzed over 40 bubbles in a model free approach and conclude that more often than not volatility is low before the crash. Thus, empirically, explosive volatility does not seem to be a reliable bubble indicator.

A class of diffusion processes with a crash (see section 1.4.4 for details) has been shown to be a strict local martingale if and only if the process is explosive and its relative jump size increases to one as a function of pre-drawdown price levels. While the first condition may well be tested, the second one seems unobservable even ex-post.

1.2.2 Other approaches

Below we describe two approaches that use a notion of bubbles based solely on stylized (empirical) facts of the resulting price process. A break of market efficiency is not explicitly modeled and thus an essential feature of an asset price bubble left out, whence these approaches are not ideal or fully satisfying – however, they avoid using the fundamental value in a bubble definition.

1.2.2.1 Growth and decline

[219] and [220] consider a geometric Brownian motion with a drift that changes at some random time τ. The asset price in their model follows the stochastic differential equation

$$\frac{dS_t}{S_t} = \left(\mu_1 \mathbb{1}_{\{t<\tau\}} + \mu_2 \mathbb{1}_{\{\tau \leqslant t\}}\right) dt + \sigma dW_t, \quad t \in [0, T]. \tag{1.16}$$

The authors define a bubble as a process that satisfies (1.16) with $\mu_1 > 0 > \mu_2$. This translates into the simple statement that an asset experiences a bubble if initially it has positive expected returns followed by negative expected returns. In section 1.4.2.3 below we see that this approach can be included in our framework.

1.2.2.2 Drift burst hypothesis

In a recent article, [51] postulate the so-called *drift burst hypothesis* to model short-lived *flash crashes* in high-frequency tick data. In particular, on a finite time horizon $[0, T]$, in a diffusion model

$$\frac{dS_t}{S_t} = \mu(t)dt + \sigma(t)dW_t, \quad t \in [0, T], \tag{1.17}$$

they look at explosive drift $\mu(t) = (T-t)^{-\alpha}$ accompanied by explosive diffusion $\sigma(t) = (T-t)^{-\beta}$ for $\alpha \in (1/2, 1)$ and $\beta \in (0, 1/2)$. Now, if $\alpha - \beta < \frac{1}{2}$, a Girsanov measure change, shows that the process S satisfies No-Arbitrage; see lemma 1.7.3. Explosive drift has already been considered in [155] and [153] to model asset price bubbles, where the authors use a crash with accelerating hazard rate to allow for the explosive drift $\mu(t) = (T-t)^{-\alpha}$. While there are interesting parallels between the two approaches, [51] try to adhere to the No-Arbitrage condition (by imposing $\alpha - \beta < \frac{1}{2}$), whereas adding a crash explicitly allows for arbitrage in the process given by (1.17). We will see that this difference is an essential feature of our definition of a bubble in the following section.

1.3 General definition of a bubble

As we have seen in propositions 1.2.1 and 1.2.2 above, rational expectation bubbles are essentially situations where *absence of arbitrage* is fulfilled, while the existence of a rational equilibrium in an economy with utility-maximizing investors is violated. This inconsistency is not ruling out rational bubbles, if one accepts the notion that bubbles represent *market failure*. However, not explicitly modeling the underlying mechanism of such market failure seems not satisfactory and leads to issues in detecting such bubbles, as discussed in section 1.2.1.3. To accommodate models from the literature that explicitly model market failures to describe bubbles, cf. the literature cited in section 1.1, we employ a nonlinear decomposition that complies with basic stylized facts of a bubble and allows one to transcend the rational expectations framework. In particular, we introduce a general definition of the bubble characteristics (1.1) to decompose the stock price as

$$S_t = \tilde{S}_t \mathbb{1}_{\{t<\tau_J\}} + X_t \mathbb{1}_{\{\tau_J \leqslant t\}}. \tag{1.18}$$

The use of the pre-crash process $\tilde{S}$ allows (or forces) one to model market failure as the driving mechanism of a bubble.

1.3.1 Notation

We use the following notation throughout the paper. For two real numbers $a, b \in \mathbb{R}$ let $a \wedge b = \min\{a, b\}$. For a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and random times $\tau_1, \tau_2 \colon \Omega \to [0, \infty]$ we use the *stochastic interval* notation

$$[[\tau_1, \tau_2)) = \{(\omega, t) \in \Omega \times [0, \infty) \colon \tau_1(\omega) \leqslant t < \tau_2(\omega)\}. \tag{1.19}$$

Similarly we use $[[\tau_1, \tau_2]]$ and $((\tau_1, \tau_2]]$. For a stochastic process $(X_t)_{t\in[0,\infty)}$ we denote its left-continuous version by $(X_{t-})_{t\in[0,\infty)}$, that is, the process with the property that $X_{t-} = \lim_{s\nearrow t} X_s$ for all $t \in [0, \infty)$. For a stochastic processes $(X_t)_{t\in[0,\infty)}, (Y_t)_{t\in[0,\infty)}$ and a random time $\tau \colon \Omega \to [0, \infty]$ we denote by $(X_t^{\tau,Y})_{t\in[0,\infty)}$ the stopped process with the property that

$$X_t^{\tau,Y} = X_t \mathbb{1}_{\{t<\tau\}} + Y_t \mathbb{1}_{\{\tau \leqslant t\}}. \tag{1.20}$$

In the special case of $Y = X_\tau$ we write $X_t^\tau = X^{\tau,X_\tau}$. For a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ with a right-continuous filtration $\mathbb{F} = (\mathcal{F}_t)_{t\in[0,\infty)}$ and an $\mathbb{F}$-stopping time $\tau : \Omega \to [0, \infty)$ we define the (itself right-continuous) filtration $(\mathcal{F}_{t\wedge\tau-})_{t\in[0,\infty)}$ consisting of the σ-algebras $\mathcal{F}_{t\wedge\tau-}$ given by

$$\mathcal{F}_{t\wedge\tau-} = \sigma\left(\{A \cap \{s < \tau\} \colon 0 \leqslant s \leqslant t, A \in \mathcal{F}_s\} \cup \mathcal{F}_0\right). \tag{1.21}$$

For a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we say that a filtration $\mathbb{F} = (\mathcal{F}_t)_{t\in[0,\infty)}$ satisfies the *usual hypothesis* if $\mathcal{F}_0$ is trivial and $\mathbb{F}$ is right-continuous and $\mathbb{P}$-complete. A filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ with the property that $\mathbb{F}$ satisfies the usual hypothesis is called *stochastic basis*. For two filtrations $\mathbb{F} = (\mathcal{F}_t)_{t\in[0,\infty)}$ and $\mathbb{G} = (\mathcal{G}_t)_{t\in[0,\infty)}$ we write $\mathbb{F} \subseteq \mathbb{G}$ if $\mathcal{F}_t \subseteq \mathcal{G}_t$ for all $t \in [0, \infty)$. We assume familiarity with the concept of semimartingales and (local) martingales; see, e.g., chapter I of [143].[10] We will call X a $(\mathbb{P}, \mathbb{F})$-(semi)martingale or simply $\mathbb{F}$-(semi)martingale if X is a

[10]While the concepts are introduced to cover rather general processes, the essential mechanism can be understood entirely without previous knowledge of continuous time stochastic processes. The reason we choose semimartingales is twofold. For one, they allow for the notion of trading in the sense of an Itô integral. Of course, more general concepts of

(semi)martingale on a stochastic basis $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$.

1.3.2 Market setting

Let $(\Omega, \mathcal{G}, \mathbb{G}, \mathbb{P})$ be a stochastic basis, let $(D_t)_{t\in[0,\infty)}$ be an RCLL $\mathbb{G}$-semimartingale representing dividend payments and let $(B_t)_{t\in[0,\infty)}$ be an RCLL $\mathbb{G}$-semimartingale representing a money market account with the property that

$$\mathbb{P}\left[\inf_{t\in[0,\infty)} B_t > 0\right] = 1. \tag{1.22}$$

We assume that $D \equiv 0$ and units are denominated in B such that $B \equiv 1$, which corresponds to no dividends and zero interest rate.[11] For simplicity, throughout the rest of the chapter we consider only nonnegative stochastic processes, although most definitions and results extend to general stochastic processes (that is, negative prices) with little adaption.

The definitions in the following section can be adapted to discrete time (using piecewise constant processes and filtrations) and a finite time horizon $T \in (0,\infty)$ (replacing the intervals $[0,\infty)$ and $[0,\infty]$ with $[0,T]$ and $[0,T] \cup \{\infty\}$, respectively.)

1.3.3 Definitions

To describe an asset's market price, we use semimartingales as the natural class for which the Itô integral is well-defined and thus provide a concept of *trading*. To account for possible explosive behavior in bubble prices, we allow a priori for explosive semimartingales. For the following (somewhat non-standard) definition, cf. section 2 of [48].

Definition 1.3.1 (Semimartingales with possible explosion ($\mathcal{S}_E$)). Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a stochastic basis, let $(S_t)_{t\in[0,\infty)}$ be an RCLL, $\mathbb{F}$-adapted stochastic process taking values in $[0,\infty]$, let $(\tau_n)_{n\in\mathbb{N}}$ be the $\mathbb{F}$-stopping times[12] with the property that

$$\tau_n = \inf\{t \in [0,\infty) | S_t \geqslant n\}, \tag{1.23}$$

and assume that S^{τ_n} is an $\mathbb{F}$-semimartingale for all $n \in \mathbb{N}$. Then S is called an $\mathbb{F}$-*semimartingale with possible explosion* and the predictable $\mathbb{F}$-stopping time[13] τ given by

$$\tau = \lim_{n\to\infty} \tau_n = \inf\{t \in [0,\infty) | S_{t-} = \infty\} \tag{1.24}$$

is called *explosion time* of S. We call S *explosive* $\mathbb{F}$-*semimartingale* if $\mathbb{P}[\tau < \infty] > 0$ and denote by $\mathcal{S}_E(\mathbb{F})$ the set of all such possibly explosive $\mathbb{F}$-semimartingales for a given stochastic basis.

stochastic integration (and thus trading) exist. Secondly, semimartingales cover most asset price models in the literature, especially those in bubble modeling.

[11] By applying notions of market efficiency to the value process $V = S + D$ instead of a stock price S, the assumption of zero dividends is made without loss of generality. The a priori choice of a numeraire B, however, is afflicted with a certain loss of generality as represented by (1.22); see the discussion in [123].

[12] See, e.g., theorem 1.27 on page 7 of [143].

[13] See, e.g., proposition 1.2(d) on page 51 of [87].

Definition 1.3.2 (Efficient market condition). Let the setting in section 1.3.2 be fulfilled and let $\mathcal{S}_E(\mathbb{G})$ be the set of possibly explosive $\mathbb{G}$-semimartingales as introduced in definition 1.3.1. Then a binary operator $\mathcal{E} : \mathcal{S}_E(\mathbb{G}) \to \{0,1\}$ is called *efficient market condition* with the following terminology

$$\mathcal{E}(S) = 1 \iff S \text{ is an } \mathcal{E}\text{-efficient market (given the information in } \mathbb{G}). \tag{1.25}$$

Definition 1.3.3 (Asset price drawdown). Let the setting in section 1.3.2 be fulfilled, let $\tilde{S} \in \mathcal{S}_E(\mathbb{G})$ with (possibly infinite) explosion time τ, let $\tau_J : \Omega \to [0,\infty]$ be a random time with $\tau_J < \tau$ on $\{\tau < \infty\}$, let $(X_t)_{t\in[0,\infty)}$ be an RCLL $\mathbb{G}$-semimartingale and let $(S_t)_{t\in[0,\infty)}$ be the stopped process

$$S_t = \tilde{S}_t^{\tau_J, X} = \tilde{S}_t \mathbb{1}_{\{t<\tau_J\}} + X_t \mathbb{1}_{\{\tau_J \leqslant t\}}, \quad t \in [0,\infty). \tag{1.26}$$

Then (τ_J, X) is called *asset price drawdown for* $\tilde{S}$ if for $t \in [0,\infty)$ it holds that

$$\mathbb{P}\left[\tau_J \leqslant t\right] > 0 \Longrightarrow \mathbb{E}\left[S_t\right] < \mathbb{E}\left[\tilde{S}_t\right] \in (0,\infty]. \tag{1.27}$$

Definition 1.3.4 (Bubble). Let the setting in section 1.3.2 be fulfilled, S be a $\mathbb{G}$-semimartingale and $\mathcal{E}$ be an efficient market condition. Then S has an $\mathcal{E}$*-bubble* if and only if there exists a filtration $\mathbb{F} \subseteq \mathbb{G}$ satisfying the usual hypothesis, a possibly explosive semimartingale $\tilde{S} \in \mathcal{S}_E(\mathbb{F}) \cap \mathcal{S}_E(\mathbb{G})$ and an asset price drawdown (τ_J, X) for $\tilde{S}$ such that

1. $S = \tilde{S}^{\tau_J, X}$.
2. $\forall Y \in \mathcal{S}_E(\mathbb{F}) \cap \mathcal{S}_E(\mathbb{G})\colon Y\mathbb{1}_{\{\cdot<\tau_J\}} = \tilde{S}\mathbb{1}_{\{\cdot<\tau_J\}} \Longrightarrow \mathcal{E}(Y) = 0$.

3a. $\mathcal{E}(S) = 1$ (type-I bubble).

3b. $\mathcal{E}(S) = 0$ and $\mathcal{E}(X) = 1$ (type-II bubble).

If 1,2,3a hold, we call S an inefficient bubble of type-I; if 1,2,3b hold, we call S an inefficient bubble of type-II. The triplet $(\tilde{S}, \tau_J, X)$, defining the structure of the bubble, is referred to as *bubble characteristics*.

1.3.4 Comments on technical conditions

(a) The definition 1.3.3 of an asset price drawdown is very general and allows for, e.g., upward jumps in some sample paths, as long as the overall expectation of the stopped process is declining. The drawdown condition (1.27) could be replaced by a stronger pathwise condition,

$$\mathbb{P}\left[X\mathbb{1}_{\{\tau_J \leqslant \cdot\}} \leqslant S\mathbb{1}_{\{\tau_J \leqslant \cdot\}}\right] = 1 \text{ and } \mathbb{P}\left[X\mathbb{1}_{\{\tau_J \leqslant \cdot\}} < S\mathbb{1}_{\{\tau_J \leqslant \cdot\}}\right] > 0. \tag{1.28}$$

We found this condition to be too restrictive to include, e.g., diffusive drawdowns as described in section 1.4.2.3. On the other hand, the assumption that $\mathbb{E}\left[S_t^{\tau_J,X}\right] < \infty$ for the stopped process, implicit in condition (1.27), may be relaxed using the pathwise condition (1.28).

(b) Instead of the notion of a fundamental price – the cornerstone of rational expectation bubbles – above definitions merely depend on an understanding of market efficiency. Such an efficient market condition $\mathcal{E}$ can take various forms; see section 1.4.1.3 below for examples. To arrive at a meaningful definition of a bubble, it is essential to use a meaningful notion of market efficiency, for example

$$\mathcal{E}(S) = 1 \iff \text{trading in } (B, S) \text{ using } \mathbb{G}\text{-predictable trading strategies}$$
$$\text{does not allow for } \textit{riskless profits},$$

or the more restrictive

$$\mathcal{E}(S) = 1 \iff \text{trading in } (B, S) \text{ using } \mathbb{G}\text{-predictable trading strategies}$$
$$\text{does not allow for } \textit{profits on average.}$$

(c) Condition 2. of definition 1.3.4 ensures that there is no efficient, $\mathbb{F}$-adapted market that is indistinguishable from $\tilde{S}$ on $[[0, \tau_J))$, in which case we can confidently label $\tilde{S}$ – or, more precisely, the part of $\tilde{S}$ that is relevant in the decomposition of S – an inefficient market. Note that in the case of a type-I bubble this requires $\mathbb{F} \subsetneq \mathbb{G}$, otherwise S itself would be such an $\mathbb{F}$-adapted, efficient market. In most examples below it holds that

$$Y \in \mathcal{S}_E(\mathbb{F}) \cap \mathcal{S}_E(\mathbb{G}) \text{ and } Y\mathbb{1}_{\{\cdot<\tau_J\}} = \tilde{S}\mathbb{1}_{\{\cdot<\tau_J\}} \implies Y = \tilde{S}, \tag{1.29}$$

and condition 2. can be succinctly stated as $\mathcal{E}(\tilde{S}) = 0$.

1.3.5 Discussion

(a) We understand bubbles as times of irrational exuberance and unsustainable growth – this part is captured by the break in market efficiency of the pre-drawdown process $\tilde{S}$, the apparent price development for an *observer ahead of the crash*. It seems that extraordinary profits are present in the market. However, the consequence of such price developments is the existence of a drawdown – this is captured by the stopped process $S = \tilde{S}^{\tau,X}$, which is the true dynamics perceived by an *omniscient observer*. As such, we formalize the minimal criteria a bubble process should satisfy, based on the following observations.

  (i) [185], who bases his *Financial Instability Hypothesis* on the observation that the "fundamental instability of a capitalist economy is a tendency to explode – to enter into a boom or euphoric state" and that these "sustained economic growth and business cycle booms (...) generate conditions conducive to disaster for the entire economic system",

  (ii) [160], who, within a more general *Anatomy of a Typical Crisis*, calls a bubble "an upward movement of prices that then implodes" as opposed to "(...) the technical language of some economists, [where] a bubble is any deviation from fundamentals",

  (iii) [38], who describe a bubble through "(i) a *run-up phase*, in which bubble imbalances form, and (ii) a *crisis phase*, during which the risk that has built up in the background materializes (...)" and "stress that the run-up and crisis phases cannot be seen in isolation - they are two sides of the same coin."

As already discussed in the introduction, there are two types of inefficient bubbles, which differ in their understanding of how the drawdown counteracts exuberance.

(b) **Type-I bubbles:** The motivation behind a type-I bubble is as follows: we view a bubble as a stock price development so exceptional that, in an efficient market, it can only be explained by the possibility of a crash. Or, looking at the other side of the coin: a bubble is equivalent to the risk premium of a crash, which leads to exorbitant (inefficient) stock price development. The essential assumption 3a in the definition of a type-I bubble is efficiency of the full process

$$S = \tilde{S}^{\tau_J, X} = \tilde{S}\mathbb{1}_{\{\cdot < \tau_J\}} + X\mathbb{1}_{\{\tau_J \leqslant \cdot\}}. \tag{1.30}$$

This type of bubble rests on the view that, while instabilities and the risk of a drawdown are inherently present in the market, as a whole it does a good job in assessing and pricing this risk.

(c) **Type-II bubbles:** Alternatively, one may argue that, while a bubble materializes as a temporary departure from market efficiency, once this inefficiency gets resolved, prices return to efficient levels. In the full price process,

$$S = \tilde{S}^{\tau_J, X} = \tilde{S}\mathbb{1}_{\{\cdot < \tau_J\}} + X\mathbb{1}_{\{\tau_J \leqslant \cdot\}}, \tag{1.31}$$

X represents those efficient levels. During a type-I bubble, market failure creates an inherent instability that is (efficiently) priced in the market, whereas for type-II bubbles market failure leads to exuberant prices and a drawdown happens when this inefficiency gets resolved and prices return to efficient levels. The difference $\tilde{S} - X$ may be called the *bubble component*, while X (as an efficient process) may be called the *fundamental component*. While this yields a similar additive decomposition as for rational bubbles, note that for inefficient bubbles it is not the case that all components $\tilde{S}$, X *and* $\tilde{S} - X$ are efficient markets.

(d) Processes with explosive dynamics are generally not considered in asset price theory as obviously no frenzy would ever cause a stock to be valued at ∞ at any point in time. However, adding (the possibility of) a downturn enlarges the picture and allows for more excessive processes accompanied by a looming crash. This theoretical reasoning has a straightforward analogy in practice, where ignoring[14] the possibility of a drawdown seems to create large riskless returns during a bubble that pass most measures of risk.[15] We believe that acknowledging (or forcing to acknowledge) the risk of a drawdown in times of large, accelerating growth can lead to more robust models for financial applications and risk management.

1.4 Continuous time modelling

While the definitions above can be applied to both continuous and discrete time models, to increase readability we want to treat them separately as there is hardly any overlap of these model types in the literature.

[14] Carelessly or deliberately - exposure to bubbles can take various forms.

[15] See, e.g., [28] for an account on this issue in the setting of financial stability and monetary policy.

To present a classification of bubbles in sections 1.4.2 and 1.4.3 below, at various points we make use of propositions. Their proofs are largely trivialities, merely checking the conditions of our main definition 1.3.4 above, but included for the convenience of the reader.

1.4.1 Examples of bubble characteristics

In this section we provide examples of the essential building blocks of continuous time inefficient market bubbles, that is, the bubble characteristics $(\tilde{S}, \tau_J, X)$ and a market efficiency condition $\mathcal{E}$. These building blocks will be used in sections 1.4.2 and 1.4.3 to present a variety of bubble models from the literature.

1.4.1.1 Examples of semimartingales with explosion $\tilde{S}$

The following examples are intended to serve as illustrations of (possibly) explosive pre-drawdown processes $\tilde{S}$ in the characteristic description of a bubble $S \longleftrightarrow (\tilde{S}, \tau_J, X)$. In continuous time modeling, Itô processes with explosion play an important role.

Definition 1.4.1. Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a stochastic basis, let $(W_t)_{t\in[0,\infty)}$ be a real valued $\mathbb{F}$-adapted[16] Brownian motion, let $(S_t)_{t\in[0,\infty)}$ be a nonnegative, continuous, $\mathbb{F}$-adapted process and $\tau_n, \tau : \Omega \to [0, \infty]$ be $\mathbb{F}$-predictable stopping times given by

$$\tau_n = \inf\{t \in (0, \infty] | S_t \geqslant n\} \text{ and } \tau = \lim_{n\to\infty} \tau_n \tag{1.32}$$

with the property that there exists a constant $S_0 \in (0, \infty)$ and $(\mathcal{F}_{t\wedge\tau-})_{t\in[0,\infty)}$-predictable processes $\alpha\colon [[0, \tau)) \to \mathbb{R}$ and $\beta\colon [[0, \tau)) \to (0, \infty)$ such that, for all $n \in \mathbb{N}$, S is a unique[17] strong solution of the stochastic integral equation

$$S_{t\wedge\tau_n} = S_0 + \int_0^{t\wedge\tau_n} \alpha_s ds + \int_0^{t\wedge\tau_n} \beta_s dW_s, \quad t \in [0, \infty). \tag{1.33}$$

Then S is called a possibly explosive *Itô-process*. If $\tau \equiv \infty$, S is called non-explosive Itô-process or simply Itô-process.

It is clear that an explosive Itô-process S is an explosive semimartingale with explosion time τ.

Geometric Brownian motion. Let $\tau \equiv \infty$ and for $\mu \in \mathbb{R}$, $\sigma \in (0, \infty)$ let $\alpha\colon [[0, \infty)) \times (0, \infty) \to \mathbb{R}$ and $\beta\colon [[0, \infty)) \times (0, \infty) \to (0, \infty)$ be given by

$$\alpha_t = \mu S_t, \quad \beta_t = \sigma S_t. \tag{1.34}$$

Then S is a geometric Brownian motion given by $S_t = S_0 \exp\left(\left(\mu - \frac{\sigma}{2}\right) t + \sigma W_t\right)$ for $t \in [0, \infty)$ and thus a non-explosive Itô-process.

[16] Although this will be the case most of the time, we do not a priori assume that $\mathbb{F}$ is generated by W.

[17] Uniqueness here is considered as *pathwise uniqueness*, in the sense that if there exist two solutions X^1 and X^2, then $\mathbb{P}\left[\{\omega | \forall t\colon (w, t) \in [[0, \tau)) \Rightarrow X^1_t(\omega) = X^2_t(\omega)\}\right] = 1$

Lipschitz continuous coefficients. Let $\tilde{\alpha} : [0, \infty) \times \Omega \times [0, \infty) \to \mathbb{R}$ and $\tilde{\beta} : [0, \infty) \times \Omega \times [0, \infty) \to [0, \infty)$ be real-valued functions that satisfy $\tilde{\alpha}(\cdot, \cdot, 0) \geqslant 0$ and $\tilde{\beta}(\cdot, \cdot, 0) = 0$, and are *locally random Lipschitz continuous* uniformly for $t \in [0, \infty)$, that is, for $f \in \{\tilde{\alpha}, \tilde{\beta}\}$ and for every $n \in \mathbb{N}$ there exists a finite random variable $K_n : \Omega \to (0, \infty)$ such that for all $x, y \in [0, n]$ and all $\omega \in \Omega$ it holds that

$$\sup_{t \in [0,\infty)} |f(t, \omega, x) - f(t, \omega, y)| \leqslant K_n(\omega)|x - y|. \tag{1.35}$$

This ensures (see, e.g., chapter 5 in [203]) that there exists a strong solution of equation (1.33) with

$$\alpha_t(\omega) = \tilde{\alpha}(t, \omega, S_t(\omega)), \quad \beta_t(\omega) = \tilde{\beta}(t, \omega, S_t(\omega)) \tag{1.36}$$

up to a predictable stopping time τ, the exit time of the domain $[0, \infty)$ at its upper boundary ∞. The boundary conditions on $\tilde{\alpha}$ and $\tilde{\beta}$ at the lower boundary 0 ensure that the process is non-negative. For homogeneous functions $\tilde{\alpha}(t, \omega, S_t(\omega)) = \tilde{\alpha}(S_t(\omega))$ and $\tilde{\beta}(t, \omega, S_t(\omega)) = \tilde{\beta}(S_t(\omega))$ one can proceed using *Feller's test for explosion* (see, e.g., corollary 4.4 in [50]) to determine whether the process is explosive. For a specific example see section 1.4.2.2 below.

Stochastic exponential. Assume there exists an $\mathbb{F}$-predictable stopping time $\tau : \Omega \to [0, \infty)$ and predictable processes $\mu : [[0, \tau)) \to \mathbb{R}$, $\sigma : [[0, \tau)) \to [0, \infty)$ with $\mathbb{P}\left[\int_0^\tau \sigma_t^2 dt < \infty\right] = 1$. Let $\alpha\colon [[0, \tau)) \times (0, \infty) \to \mathbb{R}$ and $\beta\colon [[0, \tau)) \times (0, \infty) \to [0, \infty)$ be given by

$$\alpha_t(\omega) = \mu_t(\omega) S_t(\omega), \quad \beta_t(\omega) = \sigma_t(\omega) S_t(\omega) \quad \text{for } (\omega, t) \in [[0, \tau)). \tag{1.37}$$

Then theorem 5.3 in [202] ensures that equation (1.33) has a unique strong solution on $[[0, \tau))$. Moreover, for this special form of α and β, one can write S as stochastic exponential of the process $\int_0^\cdot \mu_s ds + \int_0^\cdot \sigma_s dW_s$, that is,

$$S_t = \exp\left(\int_0^t \left(\mu_s - \frac{\sigma_s^2}{2}\right) ds + \int_0^t \sigma_s dW_s\right) \quad \text{on } [[0, \tau)). \tag{1.38}$$

The process S is explosive if and only if $\mathbb{P}\left[\int_0^\tau \mu_t dt = \infty\right] > 0$ and strictly positive if and only if $\mathbb{P}\left[\int_0^\tau \mu_t dt = -\infty\right] = 0$. For a specific example see section 1.4.2.1 below.

1.4.1.2 Examples of drawdowns (τ_J, X)

Continuous time crash. In many continuous models the drawdown X takes the form of a single jump ("crash") happening at a totally inaccessible[18] random time τ_J with a crash size depending on the information available at that time. We can adapt a standard approach in the literature to construct such a random time; see e.g. section 6.5 in [23]. Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a stochastic basis, let $\tilde{S}$ be a $\mathbb{F}$-semimartingale with explosion time τ, let $(\Lambda_t)_{t \in [0,\infty)}$ be an $\mathbb{R} \cup \{\infty\}$-valued, continuous, increasing, $(\mathcal{F}_{t \wedge \tau -})_{t \in [0,\infty)}$-adapted process with the property that $\Lambda_\tau = \infty$. Then, on a larger probability space $(\Omega, \mathcal{G}, \mathbb{P})$, we can construct a random time $\tau_J : \Omega \to [0, \infty)$ and a filtration

[18] A totally inaccessible random time almost surely avoids predictable stopping times; see, e.g., definition 2.20 in [143]. A predictable stopping time can be approximated by a sequence $(\tau_n)_{n \in \mathbb{N}}$ of stopping times $\tau_n < \tau$, i.e., it can be "foreseen".

$\mathbb{G} \supseteq \mathbb{F}$ with the property that $\tilde{S}$ is a $\mathbb{G}$-semimartingale and

$$\mathbb{1}_{\{\tau_J \leqslant t\}} - \Lambda_{t \wedge \tau_J} \text{ is a } \mathbb{G}\text{-local martingale;} \tag{1.39}$$

see lemma 1.7.1 for details. To model the crash size, let $\kappa : \Omega \to [0,1]$ be a $\mathcal{G}_{\tau_J-}$-measurable random variable and the drawdown X be given by

$$X_t = \mathbb{1}_{\{\tau_J \leqslant t\}} \tilde{S}_{\tau_J -} (1-\kappa) \mathbb{1}_{\{\tau_J \leqslant t\}}, \quad t \in [0,\infty). \tag{1.40}$$

Then the stopped process is the single jump process

$$\tilde{S}_t^{\tau_J, X} = \tilde{S}_t \mathbb{1}_{\{t<\tau_J\}} + \tilde{S}_{\tau_J -} (1-\kappa) \mathbb{1}_{\{\tau_J \leqslant t\}}, \quad t \in [0,\infty), \tag{1.41}$$

and is, by construction, a $\mathbb{G}$-semimartingale. Moreover, (τ_J, X) is a drawdown as in definition 1.3.3.

Note that the semimartingale property follows from this construction only due to the continuity of the hazard process Λ. See example 3.12 in [125] for a single jump process with a discontinuous hazard process that fails to be a semimartingale.

Bifurcation. Depending on the dynamics of the underlying markets and its fundamentals, it can be reasonable to model the asset price drawdown as a crash that only happens with a certain probability and include an alternative scenario, where the potential positive influence of exuberant behavior turns out to be sustainable.

Assume the setting of the last paragraph and in addition assume that $(\Omega, \mathcal{G}, \mathbb{G}, \mathbb{P})$ is rich enough to support a Bernoulli variable Y on $\mathcal{G}_{\tau_J-}$ with probability of success $\alpha \in [0,1]$. Define the drawdown X to be

$$X_t = \begin{cases} \tilde{S}_{\tau_J -} \mathbb{1}_{\{\tau_J \leqslant t\}}, & \text{if } Y = 1 \\ \tilde{S}_{\tau_J -} (1-\kappa) \mathbb{1}_{\{\tau_J \leqslant t\}}, & \text{if } Y = 0 \end{cases} \quad \text{for } t \in [0,\infty). \tag{1.42}$$

Again, the $\mathbb{G}$-semimartingale property of the stopped process $\tilde{S}^{\tau_J, X}$ follows from construction.

Diffusive drawdown. While a single jump (crash) at τ_J can be a reasonable model assumption if there are liquidity problems during times of market distress, in general a drawdown materializes in some time interval $[\tau_J, \cdot]$, which may be modeled by a diffusive drawdown. For this, let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a stochastic basis, W be a real-valued $\mathbb{F}$-Brownian motion and $\tilde{S}$ be a possibly explosive Itô process as in definition 1.4.1 with explosion time τ and $\alpha \geqslant 0$. Let τ_J be a random time constructed by a $(\mathcal{F}_{t \wedge \tau -})_{t \in [0,\infty)}$-adapted hazard process as given by equation (1.39). For two constants $\mu \in (-\infty, 0)$ and $\sigma \in (0,\infty)$ define the drawdown X to be

$$X_t = \tilde{S}_t \mathbb{1}_{\{t<\tau_J\}} + \tilde{S}_{\tau_J} \exp\left(\left(\mu - \frac{\sigma^2}{2} \right)(t-\tau_J) + \sigma \left(W_t - W_{\tau_J} \right) \right) \mathbb{1}_{\{\tau_J \leqslant t\}}, \quad t \in [0,\infty). \tag{1.43}$$

Let $\mathbb{G}$ be the filtration generated by $\mathbb{F}$ and the stopped process $S = \tilde{S}^{\tau_J,X}$ and satisfying the usual hypothesis. Observe that W remains a Brownian motion with respect to $\mathbb{G}$ and $W_t - W_{\tau_J}$ is a restarted Brownian motion independent of $\mathcal{G}_{\tau_J}$. In particular, $\tilde{S}$ and S are a $\mathbb{G}$-semimartingale with continuous sample paths.

1.4.1.3 Examples of efficient market conditions $\mathcal{E}$

Below we present measures of market efficiency in increasing strength that are common in the literature. Before stating the definitions, let us give a descriptive introduction.

1. **Exclude markets that allow for riskless gains:** The weakest condition is classical No-Arbitrage, which is known as *No Free Lunch with Vanishing Risk* developed for possibly unbounded processes in continuous time. For a full characterization in a general continuous-time equilibrium model set-up see [149]. It is similar to the (slightly weaker) *law of one price*, which basically means that the same good cannot trade for different prices. [218] calls this *riskless arbitrage*.
2. **Exclude markets that allow for large gains with little risk:** Starting with [54] in an intention to get reasonable bounds for relative pricing approaches without resorting to strong assumptions on collective risk preferences, there has been an effort to rule out not only riskless arbitrage opportunities but also arbitrage opportunities that are afflicted with little risk, so called *Good Deals*. A common measure of riskiness employed is the Sharpe-ratio; see [54], [244] or [25]. *Riskiness*, however, can be captured by a variety of utility/performance measures different from the Sharpe ratio; see [162]. We employ the notion of good-deals not for pricing purposes, but to classify markets with a payoff structure that is "too good to be true". [218] calls this type of arbitrage *risky arbitrage*, although he does not specify a particular measure to contrast risk and return.
3. **Exclude markets with stochastic or time-dependent required returns:** In the economics literature an additional, arguably strong assumption on risk preferences is sometimes employed to analyze markets and simplify the description of asset prices. It is assumed that investors have, on average, constant risk-aversion such that the required return on the asset is constant over time; see, e.g., [27], [155], [228], [199] or chapter 7 of [41] for a textbook treatment. Note that, while this condition is derived from *rational expectations* or *No-Arbitrage* by a priori fixing a risk-neutral measure, it is considerably stronger than the No-Arbitrage condition in point 1. above and one has to be careful when using it, especially in negations.[19] A more general version of this condition would be to a priori fix some level of risk premia. In fact, most absolute asset pricing models essentially use this condition, and the pricing measure is informed by certain (macro-)economic variables.

Definition 1.4.2 (No Free Lunch with Vanishing Risk (NFLVR)). Let the setting in section 1.3.2 be fulfilled and let $S \in \mathcal{S}_E(\mathbb{G})$ be a non-negative semimartingale with possible explosion. Then

[19] A market with stochastically varying risk premia can hardly be called inefficient, although it violates constant required rate of return. This lies at the heart of [91]'s joint hypothesis problem and [172]'s (incomplete) critique of the efficient market hypothesis.

S is called a (NFLVR)-efficient market if it does not explode and no riskless profits can be made through trading with self-financing, $\mathbb{G}$-predictable strategies with bounded losses. In continuous time, this definition involves a limiting description for *riskless*; see [71] for details.

Remark 1.4.1. The fundamental theorem of asset pricing in [71] implies that a (non-negative) $\mathbb{G}$-semimartingale S satisfies (NFLVR) if and only if there exists at least one measure $\mathbb{Q} \approx \mathbb{P}$ such that S is a $(\mathbb{Q}, \mathbb{G})$-local martingale.

For the next definition, we restrict to a finite time horizon to simplify presentation and avoid additional assumptions.

Definition 1.4.3 (No-Good-Deal (NGD) #1). Let the setting in section 1.3.2 be fulfilled, let $S \in \mathcal{S}_E(\mathbb{G})$ be a non-negative semimartingale with possible explosion, let $T \in [0, \infty)$, $K \in (0, \infty]$, let $\mathcal{Q}_T$ be the set of local martingale measures for S equivalent to $\mathbb{P}$ on $\mathcal{G}_T$ and let $f : (0, \infty) \to \mathbb{R}$ be a convex function. Then S is called a *static* $(\mathrm{NGD})_{K,f}$-efficient market if there exists a measure $\mathbb{Q} \in \mathcal{Q}_T$ such that for the density process $Z_t^{\mathbb{Q}} = \mathbb{E}\left[\frac{d\mathbb{Q}}{d\mathbb{P}} | \mathcal{G}_t\right]$, $t \in [0, T]$, it holds that

$$\sup_{t \in [0,T]} \mathbb{E}\left[f\left(\frac{Z_T^{\mathbb{Q}}}{Z_t^{\mathbb{Q}}} \right) \middle| \mathcal{G}_t \right] < K. \tag{1.44}$$

Remark 1.4.2. Depending on the choice of the function f, the bound in equation (1.44) allows one to derive a bound on the expected utilities of $\mathbb{Q}$-valued payoffs, for details and specific choices for f see [162] and [3]. In the setting of a jump diffusion in the filtration generated by a Brownian motion and a pure jump process, an instantaneous (or *dynamic*) version of *No-Good-Deal* with a quadratic f has been introduced in [25]. The explicit description of the density of an equivalent martingale measure in such models allows one to derive instantaneous conditions; see appendix 1.7.2.

Definition 1.4.4 (Constant required return (CR)). Let the setting in section 1.3.2 be fulfilled, let $S \in \mathcal{S}_E(\mathbb{G})$ be a non-negative semimartingale with possible explosion and let $r \in [0, \infty)$. Then S is called a $(\mathrm{CR})_r$-efficient market if S satisfies (NFLVR) and $(e^{-rt} S_t)_{t \in [0, \infty)}$ is a $(\mathbb{P}, \mathbb{G})$-local martingale.

Example 1.4.1. We close this section viewing the above definitions from two different angles, a general and a specific one. Let $f(x) = x^2 - 1$, such that (1.44) is a bound on the *variance* of the densities.

1. In terms of the set of $\mathbb{P}$-equivalent local martingale measures $\mathcal{Q}$ for a strictly positive $\mathbb{G}$-semimartingale S on $[0, T]$, $T \in [0, \infty)$.
   (a) (NFLVR) $\iff \mathcal{Q} \neq \varnothing$ (a pricing measure exists).
   (b) $(\mathrm{NGD})_{K,f} \iff$ there exists a measure $\mathbb{Q} \in \mathcal{Q}$ such that its density process satisfies (1.44) (and thus gives rise to a market without "K-Good Deals").

(c) $(\mathrm{CR})_r \iff$ there exists a measure $\mathbb{Q} \in \mathcal{Q}$ such that its density process satisfies[20]

$$\frac{d[S, Z^{\mathbb{Q}}]_t}{S_t Z_t^{\mathbb{Q}}} = -r\,dt, \quad t \in [0, T]. \tag{1.45}$$

If $e^{-r\cdot}S$ and S are true $\mathbb{P}$- and $\mathbb{Q}$-martingales, respectively, this condition simplifies to $\mathbb{E}_{\mathbb{Q}}[S_T|\mathcal{F}_t] = e^{-r(T-t)}\mathbb{E}[S_T|\mathcal{F}_t]$, for $t \in [0, T]$.

Note that if $\mathbb{P} \in \mathcal{Q}$, then S fulfills $(\mathrm{NGD})_{\epsilon,f}$ for any $\epsilon > 0$ and $(\mathrm{CR})_0$.

2. In terms of a one-period model starting at S_0 with two states of the world $\Omega = \{\omega_1, \omega_2\}$ with equal probability $\mathbb{P}(\omega_1) = \mathbb{P}(\omega_2) = \frac{1}{2}$. At time $T = 1$ the price either moves to $S_1(\omega_1) = S_0(1+u)$ or to $S_1(\omega_2) = S_0(1+d)$ for $d < u \in [-1, \infty)$. Then it holds that

   (a) (NFLVR) $\iff$ $(d < 0 < u)$ or $(d = u = 0)$.

   (b) $(\mathrm{NGD})_{K,f} \iff$ (NFLVR) and $\left(2 + \frac{8ud}{(u-d)^2} < K\right)$.

   (c) $(\mathrm{CR})_r \iff$ (NFLVR) and $r = \log\left(1 + \frac{u+d}{2}\right)$.

   Condition (2a) describes that riskless profit is possible only if the market moves up or down almost surely, (2b) is derived from explicitly calculating the variance of the random variable $Z = \frac{d\mathbb{Q}}{d\mathbb{P}}$ for this simple model. Finally, condition (2c) follows from calculating the continuously discounted growth rate for the discrete process.

1.4.2 Examples of type-I bubbles

Following definition 1.3.4, a type-I bubble is given by its characteristic triplet $(\tilde{S}, \tau_J, X)$, where $\tilde{S}$ fails market efficiency, while the full stock price S is an efficient market. We do not focus on the generating mechanism of inefficiency (for which we refer to the original papers), but classify resulting processes as inefficient market bubbles for various market efficiency conditions. A stronger market efficiency condition $\mathcal{E}$ leads to a weaker notion of a bubble.

1.4.2.1 JLS model

The setting considered here is a simple version the classical model of [155],[153], which is based on an analysis of a hierarchical system of traders. Let the setting in section 1.3.2 be fulfilled with a finite time horizon $[0, T]$ for some $T \in [0, \infty)$, let $\alpha \in [0, \infty)$, $\mu_0, \sigma_0 \in (0, \infty)$ be constants, let $\mu : [0, T) \to [0, \infty)$ be given by

$$\mu(t) = \frac{\mu_0}{(T-t)^\alpha}. \tag{1.46}$$

For a constant $\kappa \in (0, 1)$, consider a random time τ_J given by a (deterministic) hazard process

$$\Lambda_t = \int_0^t \frac{\mu(s)}{\kappa}\,ds, \quad t \in [0, T] \tag{1.47}$$

[20] Here we use the quadratic variation $[X, Y]$, defined for any two semimartingales X, Y; see, e.g., section II.6 in [203].

as in (1.39). The probability that there is no crash on $[0, T]$ is $1 - e^{-\Lambda_T}$ and nonzero if and only if $\alpha \in [0, 1)$. Moreover, let $(W_t)_{t\in[0,T]}$ be an $\mathbb{F}$-Brownian motion and let the pre-drawdown process $\tilde{S}$ be an Itô process as in section 1.4.1.1 with (possible) explosion time T that satisfies the stochastic differential equation

$$\frac{d\tilde{S}_t}{\tilde{S}_t} = \mu(t)dt + \sigma_0 dW_t. \tag{1.48}$$

It is easy to see that $\tilde{S}$ explodes to ∞ at T if and only if $\alpha \in [1, \infty)$. Finally, to complement the drawdown (τ_J, X), let X be a crash as defined in (1.40) with a relative jump size κ.

Then the stopped process $S = \tilde{S}^{X,\tau_J}$ satisfies a stochastic differential equation of the form

$$\frac{dS_t}{S_{t-}} = \mu(t)\mathbb{1}_{\{t<\tau_J\}}dt + \sigma_0\mathbb{1}_{\{t<\tau_J\}}dW_t - \kappa dJ_t, \quad t \in [0, T] \tag{1.49}$$

where $(J_t)_{t\in[0,T]}$ is the single jump process $J = \mathbb{1}_{\{\tau_J \leqslant \cdot\}}$. Our simple formulation implies in particular that the process is constant after τ_J.

Proposition 1.4.1. *Let the setting above be fulfilled, let $K \in (\mu_0/\sigma_0, \infty]$, S be given as in (1.49) with bubble characteristics $(\tilde{S}, \tau_J, X)$ and (NGD) refer to the dynamic No-Good-Deal condition from section 1.7.2.3. Then it holds that*

(a) $\alpha \geqslant \frac{1}{2} \iff S$ has a (NFLVR)-bubble.

(b) $\alpha > 0 \iff S$ has a $(\text{NGD})_K$-bubble.

(c) $\alpha = 0 \implies S$ has a $(\text{NGD})_{\frac{\mu_0}{\sigma_0}}$-bubble.

Proof of proposition 1.4.1. Lemma 1.7.3 implies that $\tilde{S}$ violates (NFLVR) if and only if $\alpha \in [\frac{1}{2}, \infty)$. By definition 1.7.1, $\tilde{S}$ violates $(\text{NGD})_K$ if for some $t \in [0, T]$ we have

$$\frac{\mu_0}{\sigma_0(T-t)^\alpha} \geqslant K. \tag{1.50}$$

Thus, $\tilde{S}$ violates $(\text{NGD})_K$ if and only if $\alpha > 0$ and $(\text{NGD})_{\frac{\mu_0}{\sigma_0}}$ if $\alpha = 0$. On the other hand, the full price process S in (1.49) is a $\mathbb{P}$-martingale by definition and satisfies (NFLVR) and $(\text{NGD})_K$ for all $K > 0$. □

Remark 1.4.3.

(a) While the behavior of the random time τ_J is derived from analyzing a hierarchical network of traders, the pre-drawdown process $\tilde{S}$ in equation (1.48) is chosen to ensure market efficiency. It is not a priori clear why $\tilde{S}$ should follow this behavior with instantaneous return μ. As noted in the introduction, it needs to be *postulated* to ensure rational investors do not exit the market.

(b) In particular, $\tilde{S}$ is chosen such that S is a $\mathbb{P}$-martingale, which is a considerably stronger assumption than dictated by market efficiency (NFLVR) or (NGD). One possibility is to extend the JLS model as in [126], where the authors assume that the crash only accounts for some part of the drift in the pre-drawdown process. Another possibility is to focus on the behavior close to explosion time T, in which case mild additional assumptions suffice; see the discussion in section 1.7.3 in the appendix.

1.4.2.2 Andersen-Sornette model

The following model has been introduced in [228] and [8]. Let the setting in section 1.3.2 be fulfilled and W be an $\mathbb{F}$-Brownian motion, let $m \in (1, \infty)$, $\mu_0, \sigma_0 \in (0, \infty)$ be constants, let $\sigma \colon [0, \infty) \to [0, \infty)$ and $\mu \colon [0, \infty) \to [0, \infty)$ be functions given by $\mu(x) = \frac{m\sigma_0^2}{2} x^{2m-2} + \mu_0 x^{m-1}$ and $\sigma(x) = \sigma_0 x^{m-1}$ for $x \in [0, \infty)$ and let the pre-drawdown process $\tilde{S}$ be the Itô-process given by the stochastic differential equation

$$\frac{d\tilde{S}_t}{\tilde{S}_t} = \mu(\tilde{S}_t)dt + \sigma(\tilde{S}_t)dW_t, \quad t \in [0, \infty). \tag{1.51}$$

Feller's test for explosion shows immediately that $\tilde{S}$ is an explosive Itô-process with a predictable explosion time $\tau : \Omega \to [0, \infty)$. Alternatively it is rather straightforward (see [228]) to derive an explicit solution for $\tilde{S}$ with $\tau(\omega) = \inf\{t \in (0, \infty) : \mu_0 t + \sigma_0 W_t(\omega) = \mu_0 T_c\}$ for some $T_c \in (0, \infty)$ depending on the starting value $\tilde{S}_0 \in (0, \infty)$. For a constant relative jump size $\kappa \in (0, 1)$, consider a drawdown (τ_J, X) with random time τ_J given by a hazard process

$$\Lambda_t = \int_0^t \frac{\mu(S_s)}{\kappa} ds, \quad t \in [0, \infty). \tag{1.52}$$

as in (1.39) and a crash X as in (1.40). Then we get that the stopped process $S = \tilde{S}^{\tau_J, \kappa}$ satisfies the stochastic differential equation

$$\frac{dS_t}{S_{t-}} = \mu(S_t)\mathbb{1}_{\{t<\tau_J\}}dt + \sigma(S_t)\mathbb{1}_{\{t<\tau_J\}}dW_t - \kappa dJ_t, \quad t \in [0, \infty), \tag{1.53}$$

where $(J_t)_{t\in[0,\infty)}$ is the single jump process $J = \mathbb{1}_{\{\tau_J \leqslant \cdot\}}$.

Proposition 1.4.2. *Let the setting above be fulfilled and S be given as in (1.53). Then S has a (NFLVR)-bubble.*

Proof of proposition 1.4.2. As argued above, $\tilde{S}$ has an explosion time τ with $\mathbb{P}[\tau < T] > 0$ for all $T > 0$ and thus, in particular, $\tilde{S}$ violates (NFLVR). Lemma 1.7.2 confirms that S is a $\mathbb{P}$-local martingale[21] and satisfies (NFLVR). The proof of proposition 1.4.2 is thus completed. □

Remark 1.4.4. Mirroring the situation discussed in remark 1.4.3 above, here μ and σ are chosen to model a simple form of inefficiency: nonlinear growth of the pre-drawdown process $\tilde{S}$. The subsequent definition of Λ as in equation (1.52), however, needs to be postulated to ensure that S is an efficient market. Moreover, its specific form is not necessary for S to satisfy (NFLVR) and other choices are possible; see the discussion in section 3 of [228] and the ideas in section 1.7.3 below.

[21] In fact, S is even a true martingale, as shown in chapter 2.

1.4.2.3 Growth and decline

In section 1.2.2.1 we have discussed an approach by [219], who model a bubble by a process S given by an Itô process of the form

$$\frac{dS_t}{S_t} = \left(\mu_1 \mathbb{1}_{\{t<\tau_J\}} + \mu_2 \mathbb{1}_{\{\tau_J \leqslant t\}}\right) dt + \sigma dW_t, \quad t \in [0, T], \tag{1.54}$$

for a random time of $\tau_J : \Omega \to (0, T)$, uniformly distributed on $(0, T)$ and independent of the $\mathbb{F}$-Brownian motion W. They say there is a bubble if $\mu_1 > 0 > \mu_2$. To capture this definition in our framework, we need to define its bubble characteristics $(\tilde{S}, \tau_J, X)$. Let $\tilde{S}$ be an Itô process as in definition 1.4.1 given by $\alpha_t = \mu_1 \tilde{S}_t$ and $\beta_t = \sigma \tilde{S}_t$. Then $\tilde{S}$ is a geometric Brownian motion. Moreover, let (τ_J, X) be a diffusive drawdown as defined in section 1.4.1.2 with X associated with the parameters μ_2 and σ and a filtration $\mathbb{G}$, being the smallest filtration satisfying the usual hypothesis that S is adapted to. The process S in (1.54) is then equal to the stopped process $S = \tilde{S}^{\tau_J, X}$. While the decomposition of a bubble can be readily be applied to this understanding of a bubble, its derivation is based merely on stylized facts and does not a priori allow for an interpretation using market efficiency. A (rather ad-hoc) possibility would be to choose a dynamic $(\mathrm{NGD})_K$-condition as in definition 1.7.2, with a constant $K = \frac{\mu_1}{\sigma}$.[22] Then $\tilde{S}$ violates $(\mathrm{NGD})_K$ while S fulfills $(\mathrm{NGD})_K$ and S has a $(\mathrm{NGD})_K$-bubble in the sense of definition 1.3.4.

Remark 1.4.5. As for the examples of type-I bubbles in sections 1.4.2.1 and 1.4.2.2 above, the behavior of S needs to be postulated. If S is given as above, the return μ_1 may be seen as remuneration for negative returns to follow. However, while even an informed investor cannot observe τ_J and thus cannot pinpoint the exact time when the instantaneous returns turn negative ($\mu_2 < 0$), the unconditional expected return eventually turns negative on $[0, T]$, and it needs to be argued why market participants stay invested in such an environment, an argument that seems to be missing in [219].

1.4.3 Examples of type-II bubbles

Following definition 1.3.4, a type-II bubble is given by its characteristic triplet $(\tilde{S}, \tau_J, X)$, where $\tilde{S}$ fails market efficiency, while the drawdown X is an efficient market. Below we discuss examples of type-II bubbles featuring overconfidence, limits to arbitrage or overestimation of future payoffs to derive $\tilde{S}$ as the (partial) equilibrium price process emerging under these inefficiencies. As in section 1.4.2.1, it should be clear that extracting the pre-crash $\tilde{S}$ is a simplification of the underlying model and may omit important additional features, but allows us to concisely describe the resulting bubble process.

1.4.3.1 Heterogeneous beliefs on dividend payments

Following the seminal paper by [118], there has been an effort to describe a deviation from rational prices in a dynamic partial equilibrium framework using heterogeneous beliefs. This includes

[22] To apply the (NGD) condition in this setting, we need to invoke theorem 7.13 in [177], which clarifies the structure of $\mathbb{P}$-equivalent measures on $\mathcal{G}_T^S$.

the models in [192, 214]. As elaborated in chapter 1 of [118], instead of a homogeneous set of investors (the *representative investor*) we look at several investor classes with subjective beliefs about the distribution of future dividends. All investors are assumed rational and risk-neutral. To describe this model in our framework (and ensure that the correction is downward), we need to assume that at least one investor class is overconfident, that is, they overestimate future dividend payoffs. Then the bubble characteristics can be described as follows (here we omit, for simplicity, a description of the filtrations involved).

1. The pre-crash process $\tilde{S}$ is derived as (partial) equilibrium price if investors' beliefs are distinct and short-selling is prohibited (cf. proposition 2 in [118]).
2. One way to represent the fundamental value X in these models is to take an objective position and define X as the (risk-neutral) expectation of future dividends (cf. proposition 3 in [118]).[23] As described in section 5 of [214], a crash may be represented as a random time where investor beliefs collapse or fundamentals become observable, whence the new equilibrium price should settle at X.
3. The models work with risk neutral agents, such that the market efficiency condition is constant required returns as given by definition 1.4.4.

To sum up, the price process is given by

$$S = \tilde{S}^{\tau_J, X} = \tilde{S}\mathbb{1}_{\{\cdot < \tau_J\}} + X\mathbb{1}_{\{\tau_J \leqslant \cdot\}} \tag{1.55}$$

and has a type-II bubble in the sense of definition 1.3.4: the process $\tilde{S}$ exceeds risk-neutral valuation, while at time τ_J the bubble collapses and S follows the efficient process X.

1.4.3.2 Noise trading and synchronisation risk

The following model is from [1] and seeks to describe bubbles that persist due to a synchronization problem among informed, rational traders, which is introduced through an unknown starting date of the mispricing. In particular, in the model it is assumed that the price grows deterministically with a growth rate $g \in (0, \infty)$. The fundamental uncertainty in the model is introduced by the random time $t_0 : \Omega \to [0, \infty)$, which is used to specify risk premia[24] given by $g\mathbb{1}_{[0,t_0)} + 0\mathbb{1}_{[t_0,\infty)}$. However, t_0 is unknown to the market participants and the mispricing can persist. Information about t_0 is sequentially revealed to informed investors over some time window $[t_0, t_0 + \eta]$, who then, maximizing their payoff, try to ride the bubble for a while but sell-out not too late. After a fraction $\kappa \in [0, 1]$ of informed traders wants to sell out, a crash happens. The model is completed with an exogenous bursting of the bubble at a time $t_0 + \overline{\tau}$ if a crash does not happen before. The bubble characteristics can be described as follows.

1. Let the pre-crash $\tilde{S}_t$ be an Itô process on a stochastic basis $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ as in definition 1.4.1 given by $\alpha_t = g\tilde{S}_t$ and $\beta_t \equiv 0$. Then $\tilde{S}$ is the deterministically growing process $\tilde{S}_t = e^{gt}$.

[23] Alternatively, one can use subjective fundamental values, that is, the risk-neutral expectation of future dividends under subjective beliefs of one of the investor classes, cf. footnote 7 on page 1186 of [214]. The (overconfident) belief in being able to resell the asset to other market participants leads investor to pay more than their subjective fundamental value.

[24] Recall that, for simplicity, we assume discounted prices and thus $r = 0$.

2. Let the drawdown (τ_J, X) be given by a process X with $X_t = e^{gt}\mathbb{1}_{[0,t_0)} + e^{gt_0}\mathbb{1}_{[t_0,\infty)}$ and a random time $\tau_J^{\text{ex/end}}$ with

   (a) $\tau_J^{\text{ex}} = t_0 + \overline{\tau}$ if we have an exogenous crash as in section 5.1 of [1] or

   (b) $\tau_J^{\text{end}} = t_0 + \eta\kappa + \tau^*$ for some uniquely specified $\tau^* \in (0, \overline{\tau} - \eta\kappa)$ if we have an endogenous crash as in section 5.2. of [1].

3. The market efficiency condition $\mathcal{E}$ is a version of constant required return (definition 1.4.4), where the required return, by definition, is given by $g\mathbb{1}_{[0,t_0)} + 0\mathbb{1}_{[t_0,\infty)}$.

Assume that $\mathbb{G}$ is generated by $\mathbb{F}$ and t_0. Then the stopped process $S = \tilde{S}^{\tau_J, X}$ has a type-II bubble in the sense of definition 1.3.4: the $\mathbb{G}$-semimartingales $\tilde{S}$ and S violate efficiency and grow at the exuberant return g beyond t_0, while at time τ_J the bubble collapses and S follows the efficient $\mathbb{G}$-semimartingale X.

1.4.3.3 An efficient inefficient bubble

Here we deal with a mechanism that was initially discussed by [90] and appears in a similar form in [196]. Both argue that agents are rational during boom and crash and thus dismiss the claim that the *Dow Jones 1987 drop* and the *Nasdaq 1997-2003 epsiode*, respectively, were bubbles characterized by a deviation from fundamentals. Let us describe the essence of such models in a simplified example to show how this can be described as an inefficient market bubble.

Setting. Let $T \in [0, \infty)$, let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ be a stochastic basis, let $(W_t)_{t\in[0,T]}$ be an $\mathbb{F}$-Brownian motion and $\mu \in \mathbb{R}$, $\sigma \in (0, \infty)$ be constants and assume dividend payments $(D_t)_{t\in[0,T]}$ given by $D_t = 0$ for $t < T$ and a final payoff

$$D_T = \exp\left(\mu T - \frac{\sigma^2}{2} T + \sigma W_T\right). \tag{1.56}$$

A standard way to model misperceptions about future payoffs would be to introduce a measure $\mathbb{Q}$ representing perceived distribution of the final payoff (or, as in [196], of the excess profitability). For simplicity, we will instead just assume that rational agents belief the final payoff to be

$$\tilde{D}_T = d \exp\left(\mu T - \frac{\sigma^2}{2} T + \sigma W_T\right) \tag{1.57}$$

for some $d > 1$. Within this setting, we can now define the characteristic triplet of our bubble.

Characteristic triplet.

1. A drawdown process $(X_t)_{t\in[0,T]}$, given by the conditional expectation of the true (discounted) final payoff,

$$X_t = \mathbb{E}_{\mathbb{P}}\left[e^{-\mu(T-t)} D_T | \mathcal{F}_t\right], \quad t \in [0, T), \tag{1.58}$$

   and $X_T = 0$.

2. A random time $\tau_J : \Omega \to (0, T)$, independent of W. This represents the time when agents realize that their beliefs about the final payoff were incorrect. We assume that the filtration $\mathbb{G}$ is generated by W and τ_J.
3. A pre-crash process $(\tilde{S}_t)_{t\in[0,T]}$ derived under belief that the final payoff equals $\tilde{D}_T$, that is,

$$\tilde{S}_t = \mathbb{E}_{\mathbb{P}}\left[e^{-\mu(T-t)}\tilde{D}_T|\mathcal{F}_t\right], \quad t \in [0, T), \tag{1.59}$$

and $\tilde{S}_T = 0$.

Finally, we assume investors (on average) require a return μ and let $\mathcal{E}$ be the market efficiency condition given by

$$\mathcal{E}(S) = 1 \iff S_t + D_t \text{ satisfies (CR)}_\mu,$$

for a $\mathbb{G}$-semimartingale S. Here we have to adjust definition 1.4.4 for nonzero dividends.

Proposition 1.4.3. *Assume the setting above and let S be the stopped process*

$$S_t = \tilde{S}_t^{\tau_J, X} = \tilde{S}_t \mathbb{1}_{\{t<\tau_J\}} + X_t \mathbb{1}_{\{\tau_J \leqslant t\}}, \quad t \in [0, T]. \tag{1.60}$$

Then S has a type-II inefficient market bubble.

Proof of proposition 1.4.3. From 1.58 we can infer that X is given by

$$X_t = \exp\left(\mu t - \frac{\sigma^2}{2} t + \sigma W_t\right), \quad t \in [0, T). \tag{1.61}$$

Similarly, from 1.59,

$$\tilde{S}_t = d \exp\left(\mu t - \frac{\sigma^2}{2} t + \sigma W_t\right), \quad t \in [0, T). \tag{1.62}$$

Then, clearly, $e^{-\mu t}(X_t + D_t)_{t\in[0,T]}$ is a $(\mathbb{P}, \mathbb{G})$-martingale and X satisfies $(CR)_\mu$, while $e^{-\mu t}(\tilde{S}_t + D_t)_{t\in[0,T]}$ and $e^{-\mu t}(S_t + D_t)_{t\in[0,T]}$ fail to be $(\mathbb{P}, \mathbb{G})$-martingales, respectively. The conditions of definition 1.3.4 are fulfilled and S has an inefficient market bubble (of type-II). □

Remark 1.4.6.

(a) As argued in [90], this effect of overvaluing future payoffs may be amplified by a change in risk premia. A sudden increase in risk aversion lowers today's *discounted* expected value even further.

(b) Given their beliefs, agents act rational all the time. Upon realizing overconfidence, the market efficiently adjusts to the value X of true expected payoffs. Thus, objectively, agents behave irrational prior to τ_J, which allows for the interpretation of bubbly behavior (overvaluation), verified by the crash. In this sense, there is an analogy between the "efficient" model here and models based on market failures: this analogy is covered by the decomposition $(\tilde{S}, \tau_J, X)$. The existence of a bubble is a question of perspective.

1.4.4 Comparison to rational expectation bubbles

We present a model class that allows for partly overlapping occurrence of rational expectation and type-I inefficient market bubbles. A range of bubble models in the literature (see, e.g., sections 1.4.2.1 to 1.4.2.2) are based on single jump processes of the form

$$dS_t = \alpha(t, S_t)\mathbb{1}_{\{t<\tau_J\}}dt + \beta(t, S_t)\mathbb{1}_{\{t<\tau_J\}}dW_t - \frac{\alpha(t, S_{t-})}{h(t, S_{t-})}dJ_t, \quad t \in [0, \infty), \tag{1.63}$$

where $(W_t)_{t\in[0,\infty)}$ is a Brownian motion, α and β are suitably regular coefficient functions, $(J_t)_{t\in[0,\infty)}$ is a $\{0, 1\}$-valued single jump process, whose jump time $\tau_J = \inf\{t > 0 \colon J_t = 1\}$ is associated with a hazard process

$$\Lambda_{t\wedge\tau_J} = \int_0^{t\wedge\tau_J} h(s, S_s)ds, \quad \text{for } t \in [0, \infty). \tag{1.64}$$

Within our framework, this amounts to using a possibly explosive Itô process as in definition 1.4.1 given by α and β and a crash (τ_J, X) as in section 1.4.1.2 where X as in (1.40) is given by

$$X_t \equiv \tilde{S}_{\tau_J-}\,(1-\kappa) \qquad \text{on } [[\tau_J, \infty)). \tag{1.65}$$

with a relative jump size

$$\kappa = \frac{\alpha(\tau_J, S_{\tau_J-})}{S_{\tau_J-}h(\tau_J, S_{\tau_J-})}. \tag{1.66}$$

Based on their examination of single jump processes with a deterministic hazard rate in [125], [126] consider (within a more general setting) a solution to the SDE (1.63) assuming a finite time horizon $T \in [0, \infty)$ and coefficients

$$\alpha(t, S_t) = \phi'(t)S_t,\ \beta(t, S_t) = \sigma_0 S_t \text{ and } h(t, S_t) = h(t) \tag{1.67}$$

for $\sigma_0 \in (0, \infty)$ and continuously differentiable functions $\phi, h : [0, T) \to (0, \infty)$. They show, in particular, that the process $(S_t)_{t\in[0,T]}$ is a strict local martingale if and only if

$$\int_0^T h(t)dt = \infty \quad \text{and} \quad \int_0^T \left(h(t) - \phi'(t)\right) dt < \infty. \tag{1.68}$$

Similarly, in chapter 2 we consider the setting of a homogeneous diffusion with

$$\alpha(t, S_t) = \alpha(S_t),\ \beta(t, S_t) = \beta(S_t) \text{ and } h(t, S_t) = h(S_t) \tag{1.69}$$

for suitably regular functions α, β and h and show that S fails to be a uniformly integrable martingale if and only if

$$\lim_{x\to\infty} \frac{\alpha(x)}{h(x)x} = 1 \text{ or } \lim_{x\to\infty} \frac{h(x)x^2}{\beta^2(x)} = 0. \tag{1.70}$$

Comparison and discussion. Note that single jump processes as above generate an incomplete market, and the choice of fundamental value in the rational expectations literature is not unique, cf. the discussion in section 1.2.1. For simplicity, in the proposition below we content ourselves with diagnosing a bubble using either of the two fundamental values.

Proposition 1.4.4. *Let the setting in section 1.3.2 be fulfilled and assume that the asset follows (1.63) with one of the following specifications under some $\mathbb{P}$-equivalent measure $\mathbb{Q}$.*

1. *The specification in section 1.4.2.1 with $\alpha \in [1, \infty)$ and asset lifetime $\tau = T$.*
2. *The specification in section 1.4.2.2 with asset lifetime $\tau = \tau_J$.*

Then S has both a rational expectation bubble[25] *and an inefficient market bubble (definition 1.3.4).*

Proof of proposition 1.4.4. For 1, we can check that (1.68) is fulfilled and thus S a $\mathbb{Q}$-strict local martingale. As such, it holds that $E_{\mathbb{Q}}[S_T|\mathcal{F}_0] < S_0$ and we have two (and only two) possibilities:

(a) the superreplication price (1.8) is equal to S_0, in which case we have a rational expectation $\mathbb{Q}$-bubble as in definition 1.2.2, or

(b) the superreplication price (1.8) is strictly smaller than S_0, in which case we have a strong bubble as in definition 1.2.1.

In any case, S has a rational expectations bubble. Being a (strict) local martingale, we know that S satisfies (NFLVR). At the same time, the pre-drawdown process $\tilde{S}$, given by

$$d\tilde{S}_t = \frac{\mu_0}{(T-t)^\alpha}dt + \sigma_0 dW_t, \quad t \in [0, T], \tag{1.71}$$

explodes $\mathbb{P}$-a.s. on $[0, T]$ and thus violates (NFLVR), whence S has a (NFLVR)-inefficient market bubble of type-I. This proves 1. Using the setting of section 1.4.2.2 and the results implied by (1.70), the second part follows analogously. □

Remark 1.4.7. Although there is some overlap, following our discussion in section 1.2.1 we suggest this is coincidental: if one accepts that single jump processes of the form (1.63) are reasonable models for financial bubbles, then differentiating between *bubble* or *no bubble* solely based on the behavior of the relative size of the crash as suggested by (1.68) and (1.70) seems far-fetched. The analysis in section 4 in [126], specifically table 4.4, supports this statement, as for this type of single jump processes we see no fundamental differences in optimal investment strategies for strict local martingales and martingales.

1.5 Discrete time modelling

In this section we assume the setting in section 1.3.2, but restrict to discrete time, which simplifies the definitions in section 1.3.3 and allows us to state the following succinct definition of a bubble.

[25] Either a strong rational expectation bubble (definition 1.2.1) or a rational expectation $\mathbb{Q}$-bubble (definition 1.2.2).

1.5.1 Discrete time inefficient market bubble

Definition 1.5.1. Let the setting in section 1.3 be fulfilled, $(S_n)_{n\in\mathbb{N}}$ be a $\mathbb{G}$-adapted discrete time stochastic process and $\mathcal{E}$ be an efficient market condition. Then S has a *$\mathcal{E}$-bubble* if and only if there exists a process $(\tilde{S}_n)_{n\in\mathbb{N}}$ adapted to a filtration $\mathbb{F}\subseteq\mathbb{G}$, a random time $\tau_J:\Omega\to\mathbb{N}\cup\infty$ and a $\mathbb{G}$-adapted process $(X_n)_{n\in\mathbb{N}}$ such that

1. $S=\tilde{S}\mathbb{1}_{\{\cdot<\tau_J\}}+X\mathbb{1}_{\{\tau_J\leqslant\cdot\}}$.
2. $\mathbb{P}\left[\tau_J\leqslant n\right]>0\Longrightarrow\mathbb{E}\left[X_n\mathbb{1}_{\{\tau_J\leqslant n\}}\right]<\mathbb{E}\left[\tilde{S}_n\mathbb{1}_{\{\tau_J\leqslant n\}}\right]$ for $n\in\mathbb{N}$.
3. For every $\mathbb{F}$-adapted processes Y it holds that $Y\mathbb{1}_{\{\cdot<\tau_J\}}=\tilde{S}\mathbb{1}_{\{\cdot<\tau_J\}}\Longrightarrow\mathcal{E}(Y)=0$.

4a. $\mathcal{E}(S)=1$ (type-I bubble).

4b. $\mathcal{E}(S)=0$ and $\mathcal{E}(X)=1$ (type-II bubble).

The pair (τ_J,X) is called *drawdown*. The bubbly stock price is fully described by its *bubble characteristics* $(\tilde{S},\tau_J,X)$.

To see that this is equivalent to the general definition 1.3.4 when restricted to discrete time, we need to note the following.

1. Any discrete semimartingale with explosion as in definition 1.3.1 has explosion time $\tau=\infty$. A discrete process can reach infinity in finite time if and only if it jumps to ∞ at some random time τ while being finite at $\tau-1$. Then S^{τ_n}, the sequence of stopped processes in definition 1.3.1, cannot be a sequence of non-explosive processes.
2. Any adapted discrete process is of finite variation and a (discrete time) semimartingale.

1.5.2 Examples

Below we provide examples of the essential building blocks of discrete time inefficient market bubbles (the bubble characteristics $(\tilde{S},\tau_J,X)$ and a market efficiency condition $\mathcal{E}$) and an example of a bubble model from the literature.

In discrete time, as any adapted process $\tilde{S}$ is a valid pre-drawdown process and explosion cannot occur, we omit stating examples for pre-drawdown processes.

1.5.2.1 Drawdown as a discrete time crash

Let $(\Omega,\mathcal{F},\mathbb{F}=(\mathcal{F}_n)_{n\in\mathbb{N}},\mathbb{P})$ be a stochastic basis and $(p_n)_{n\in\mathbb{N}}$ be a sequence of $[0,1]$-valued $\mathbb{F}$-predictable random variables. Then we can define a random time $\tau_J:\Omega\to\mathbb{N}$ with the property that

$$\mathbb{P}\left[\tau_J=n|\mathcal{F}_{n-1},\tau_J\geqslant n\right]=p_n,\quad n\in\mathbb{N}. \tag{1.72}$$

For an $\mathbb{F}$-adapted pre-drawdown process $(\tilde{S}_n)_{n\in\mathbb{N}}$ we may choose a process $(X_n)_{n\in\mathbb{N}}$ with $X_0=0$ and the property that, for $n\in\mathbb{N}$,

$$\begin{aligned}X_n&\in[0,\tilde{S}_{n-1}),\quad\text{if }\tau_J=n,\\X_n&\equiv X_{n-1},\qquad\text{else.}\end{aligned} \tag{1.73}$$

If we let $\mathbb{G}$ be the filtration generated by $\mathbb{F}$, τ_J and X, the crashed process $S = \tilde{S}\mathbb{1}_{\{\cdot<\tau_J\}} + X\mathbb{1}_{\{\tau_J \leqslant \cdot\}}$ is a $\mathbb{G}$-semimartingale and (τ_J, X) is a drawdown as in definition 1.3.3.

1.5.2.2 Efficient market condition

For examples of market efficiency conditions we refer to section 1.4.1.3, which covers discrete time models without adaption. Very common in the literature is the condition of constant required returns introduced in definition 1.4.4. In discrete time, it can be stated using a one-step condition.

Definition 1.5.2 (Constant required return (CR)). Let the setting in section 1.3.2 be fulfilled in discrete time, let $(S_n)_{n\in\mathbb{N}}$ be an $\mathbb{G}$-adapted stochastic process that satisifies (NFLVR), and let $r \in \mathbb{R}$. Then S is a (CR)$_r$-efficient market if for $R = e^r - 1$ we have

$$\mathbb{E}\left[S_{n+1}|\mathcal{G}_n\right] = (1+R)S_n, \quad n \in \mathbb{N}. \tag{1.74}$$

With a slight abuse of notation, we use (CR)$_R$ to refer to constant required return with simple return R as given by the one-step condition (1.74). Note that we have assumed zero dividends and the wealth process is equal to the stock price. In the presence of a cumulative dividend process $(D_n)_{n\in\mathbb{N}}$, equation 1.74 has to be replaced with

$$\mathbb{E}\left[S_{n+1} + D_{n+1} - D_n|\mathcal{G}_n\right] = (1+R)S_n, \quad n \in \mathbb{N}. \tag{1.75}$$

1.5.2.3 The periodically collapsing bubble of Evans

As a classical example of a bubble model in the literature that has originally been used to show the shortcomings of early cointegration tests, the periodically collapsing bubble of [88] can be described in our framework and represents arguably the simplest model of a type-I bubble as in definition 1.5.1. Given parameters $R, \delta \in \mathbb{R}$, $\alpha \in (\delta/(1+R), \infty)$, $\pi \in (0,1)$ and a sequence $(u_n)_{n\in\mathbb{N}}$ of positive, iid random variables with $\mathbb{E}\left[u_1\right] = 1$, the bubble process $(S_n)_{n\in\mathbb{N}}$ is defined as

$$\begin{aligned}
S_{n+1} &= (1+R)S_n u_{n+1} && \text{for } S_n \leqslant \alpha \\
S_{n+1} &= \begin{cases} \left(\delta + \frac{1+R}{\pi}\left(S_n - \frac{\delta}{1+R}\right)\right) u_{n+1} & \text{with probability } \pi \\ \delta u_{n+1} & \text{else} \end{cases} && \text{for } S_n > \alpha
\end{aligned}$$

To identify the bubble characteristics, we define a pre-drawdown process $\tilde{S}$ given by

$$\begin{aligned}
\tilde{S}_{n+1} &= (1+R)\tilde{S}_n u_{n+1} && \text{for } \tilde{S}_n \leqslant \alpha \\
\tilde{S}_{n+1} &= \left(\delta + \frac{1+R}{\pi}\left(\tilde{S}_n - \frac{\delta}{1+R}\right)\right) u_{n+1} && \text{for } \tilde{S}_n > \alpha
\end{aligned}$$

and a drawdown (τ_J, X) as in definition 1.3.3 given by

$$\mathbb{P}\left[\tau_J = n | \tilde{S}_n, \tau_J \geqslant n\right] = \begin{cases} 1-\pi & \text{for } \tilde{S}_n > \alpha \\ 0 & \text{else} \end{cases} \tag{1.76}$$
$$X_n = \delta u_n \quad \text{for } n \in \mathbb{N},$$

to get $S = \tilde{S}^{\tau_J, X}$. With a market efficiency condition of constant required returns given by (1.74) above,

$$\mathcal{E}(S) = 1 \leftrightarrow \text{S satisfies (CR)}_R, \tag{1.77}$$

then S has a $(CR)_R$-bubble according to definition 1.5.1: the pre-drawdown process $\tilde{S}$ grows at a rate $\frac{1+R}{\pi}$ while the full process S grows at rate $1+R$. To arrive at the periodically collapsing model with a series of crashes $\left(\tau_J^{(n)}\right)_{n\in\mathbb{N}}$, simply restart the process after a crash $\tau_J^{(i)}$ until the next crash at $\tau_J^{(i+1)}$.

Remark 1.5.1. Note that the model was originally described in the rational expectations framework, discussed in section 1.2.1, with an additive decomposition where a fundamental component F was added to the bubble S. This can be mimicked without changes in our framework. To see the difference of the two approaches, consider the case $\pi \equiv 1$: in the rational expectations framework, this is a valid bubble model. In our approach, however, the existence of a crash (and higher growth countering it) is an essential feature of a bubble and $\pi \equiv 1$ would not be an inefficient market bubble. Moreover, a bubble in our approach can naturally end after a crash, whereas within the rational expectations framework the bubble continues ad infinitum.

1.5.3 Comparison to rational expectation bubbles

As we have discussed in the introduction, (discrete time) rational expectation bubbles satisfy a one-step no-arbitrage condition, thus they cannot meaningfully be described in our framework, whose main characteristics is a break of market efficiency in the pre-drawdown process.

1. In section 1.5.3.1 we compare the two approaches and discuss the joint hypothesis issue mentioned in section 1.2.1.3 in a simplified setting.
2. In section 1.5.3.2 we discuss, using examples from the literature, how researchers seem to implicitly add natural elements of our framework to rational expectations bubbles.

1.5.3.1 Efficient vs. inefficient market bubbles

Consider the standard discrete time set-up from section 1.2.1 given by a stock price $(S_n)_{n\in\mathbb{N}}$, a cumulative dividend stream $(D_n)_{n\in\mathbb{N}}$ and the constant-return No-Arbitrage condition (1.74). Our goal is to illustrate the joint hypothesis issue discussed in section 1.2.1.3 in a simplified setting. For this, assume we want to model a technology stock that does not pay dividends in the near future $[0, T]$ for some $T \in \mathbb{N}$ and possibly experiences a bubble.[26]

[26]Take, for instance, the case study of Zynga in [103] or the recent development of the Bitcoin price.

Rational expectation bubble. Then, as in discussed in example 1.2.1(a), the standard description of a stock price $S = S^* + B$ with fundamental value S^* given by

$$S_n^* = \begin{cases} \sum_{i=T-n+1}^{\infty} \left(\frac{1}{1+R}\right)^i \mathbb{E}\left[D_{n+i} - D_{n+i-1} | \mathcal{F}_n\right], & n \in \{0, \ldots, T\}, \\ \sum_{i=1}^{\infty} \left(\frac{1}{1+R}\right)^i \mathbb{E}\left[D_{n+i} - D_{n+i-1} | \mathcal{F}_n\right], & n \in \{T+1, \ldots\}, \end{cases} \tag{1.78}$$

it holds that

$$\mathbb{E}\left[S_{n+1}^* | \mathcal{F}_n\right] = (1+R) S_n^*, \quad n \in \{0, \ldots, T\}. \tag{1.79}$$

Using the standard argument in the rational bubble literature, from equations (1.75) and (1.78) we can derive that the bubble component $B = S - S^*$ necessarily satisfies the martingale property on $[0, \infty)$,

$$\mathbb{E}\left[B_{n+1} | \mathcal{F}_n\right] = (1+R) B_n, \quad n \in \mathbb{N}. \tag{1.80}$$

We can conclude that the process with or without a bubble grows at the same rate and fails to cover the intuitive notion that *the expected return is larger if the stock price experiences a bubble.* Even if the true required return R is known (e.g., inferred from a consumption model for the pricing kernel), a bubble cannot be detected. This reasoning extends to the situation of multiple dividends, in the sense that the only way to derive testable hypothesis on bubble processes are assumptions on future dividends.

Inefficient market bubble (type-I). Assume instead there is the possibility of disclosure of bad news leading to a lower expected valuation of future dividends and a drop by a multiplicative fraction of $\kappa \in (0,1)$. At a date $n \in [0, T]$ this occurrence can be described by a random time τ_J with $\mathbb{P}\left[\tau_J = n+1 | \tau_J > n\right] = p$. There is a probability $(1-p)^T$ that there is no downward correction on the whole time interval and high payoffs are realized. Then the full price process up to and including the crash can be written as

$$S_n^{\tau_J, \kappa} = S_n \mathbb{1}_{\{n < \tau_J\}} + (1-\kappa) S_{n-1} \mathbb{1}_{\{\tau_J = n\}}, \quad n \in \{0, \ldots, \tau_J\}, \tag{1.81}$$

where S represents the high payoffs in case of no downward correction and is assumed to grow at some constant (conditional) return R',

$$\mathbb{E}\left[S_{n+1} | \mathcal{F}_n, \text{no drop at } n+1\right] = (1+R') S_n, \quad n \in \{0, \ldots, T\}. \tag{1.82}$$

The unconditional expected growth (that is, the expected return that a rational trader requires to hold the asset) is thus equal to

$$\mathbb{E}\left[S_{n+1}^{\tau_J, \kappa} \middle| \mathcal{F}_n, \tau_J > n\right] = [(1+R')(1-p) + p(1-\kappa)] S_n^{\tau_J, \kappa}, \quad n \in \{0, \ldots, \tau_J\}, \tag{1.83}$$

with a growth rate $1 + R = (1+R')(1-p) + p(1-\kappa) < 1 + R'$. The objective market efficiency condition is therefore No-Arbitrage with expected return R. The full model consists of

1. a *pre-drawdown process* S growing with expected return R',
2. *a crash* described by the random time τ_J and a fraction κ, and
3. an *objective price process* $S^{\tau_J,\kappa}$ growing with (true) expected return $R < R'$.

In other words, the pre-drawdown process may grow at a significantly higher rate R', remunerating for the possibility of a crash (that is, revaluation) and violating market efficiency, while the de facto price process $S^{\tau_J,\kappa}$ grows at R. Contrary to the rational expectations framework, if we know the true required return is equal to R', we can detect a bubble.

Remark 1.5.2.

(a) For the uninformed investor that ignores the possibility of a crash (i.e., trades in the belief that pre-drawdown process S is the true underlying), the high returns given by R' seem like a marvelous deal. However, even for rational, informed traders (think of hedge funds, banks or other large players) that correctly assess the probability of a crash, p and require a return R', it is fully rational to stay invested. In this conception of a bubble, the question in the rational bubble literature of why informed traders are not deflating a bubble is naturally answered: they don't have a reason to do so.[27]

(b) To conclude this section, note that, while the above simplified examples show the theoretical difference of the two frameworks, neither of the above descriptions allows one to detect a bubble before the crash if the required rate of return is unknown – all processes (including the bubbly pre-drawdown process) are assumed to have constant expected return and measuring that return (in the absence of dividends) will be inconclusive.[28] One possible remedy is provided by the observation that pre-drawdown trading behavior of noise traders is often characterized by positive feedback mechanisms, frenzy, overoptimism and thus increasing returns; see, e.g., [69, 137, 186, 221, 230] and section 1.4.2.1 to 1.4.2.2 for continuous time models of increasing (accelerating) returns.

1.5.3.2 Sidelining rational expectations in the literature

Below we show that some of the additional assumption and slight departures from rational expectations applied in the literature are, essentially, not too different from the basic ideas of our framework. We hope that this strengthens the case for our inefficient markets approach to bubbles introduced in section 1.3, which explicitly incorporates these ideas. For this, recall the bubble characteristics $(\tilde{S}, \tau_J, X)$ as introduced in definition 1.5.1, where $\tilde{S}$ is *a pre-drawdown process that violates market efficiency* and (τ_J, X) represents *the possibility of a drawdown*. These departures from the rational expectations paradigm seem to be further indications for its inadequacy in constituting a reasonable theoretical framework – explicitly including those stylized features can improve the modeling process significantly.

[27] This may be important to keep in mind in view of the growing amount of empirical bubble literature that present short sale bans/constraints as a driving mechanism of a bubble build-up. While short sale constraints will naturally have an influence on the shape of the bubble and are a driving force in many type-II bubble models (see, e.g., section 1.4.3.1), they seem not essential to generate the *exuberant returns & crash*-pattern that is characteristic for bubbles.

[28] Apart, of course, from a fundamental analysis, that can be highly uncertain if expected payoffs are in the distant future.

A pre-drawdown process that violates market efficiency. Some authors study log-prices instead of prices, e.g., among others, [199] or [198]. The main idea is to take a log-linear approximation of the constant-return No-Arbitrage condition

$$(1 + R)S_n = \mathbb{E}\left[S_{n+1} + D_{n+1} - D_n(1 + R)|\mathcal{F}_n\right], \quad n \in \mathbb{N}, \tag{1.84}$$

for a stock price $(S_n)_{n\in\mathbb{N}}$ and cumulative dividend payments $(D_n)_{n\in\mathbb{N}}$, as elaborated in [42] or section 7 of [41]. Then it is argued that non-stationary behavior in an autoregressive specification for log-prices implies the existence of a bubble. However, this leads to behavior transcending the *rational expectations* framework, which can be seen as follows. Assume that the bubble component is large compared to single-period dividends $\tilde{D}_n = D_n - D_{n-1}(1 + R)$, that is,

$$S_n \approx B_n \gg \tilde{D}_n = D_n - D_{n-1}(1 + R), \quad n \in \mathbb{N} \tag{1.85}$$

and assume that, as in [199], the log stock price $s = \log(S)$ follows a linear autoregressive process

$$s_{n+1} = c + \delta s_n + \epsilon_n, \quad n \in \mathbb{N} \tag{1.86}$$

with $c, \delta \in (0, \infty)$. For simplicity of the argument, assume that $\epsilon_n \equiv 0$ for all n. Then, by recursion, we get for the price S that

$$S_{n+1} = e^{s_{n+1}} \geqslant S_0^{(\delta^{n+1})}, \quad n \in \mathbb{N}. \tag{1.87}$$

Now the standard approach in such bubble tests is to look for an explosive root ($\iff \delta > 1$) in equation (1.86). However, for $\delta > 1$, S cannot fulfill (1.84), violating the original model assumptions and one would detect a bubble precisely in situations where the No-Arbitrage condition (1.84) is *violated*. This test is thus close in spirit to point 2 in definition 1.3.4: the pre-drawdown process shows exuberant behavior and violates market efficiency.

Introducing a drawdown through the backdoor. Therefore, to ensure that the one-step No-Arbitrage condition (1.84) can still be satisfied in case of so-called mildly explosive log-prices, [199], [169] and [198] specify a stock market crash (or, rather, an *end date* of the bubble) by a random time of regime change. We refer to section 2.3 in [169] for a detailed discussion. The rational expectation framework requires that the crash is completely exogenous – imposed by the modeler – and rational market participants are unaware of its possibility, otherwise it would reflect earlier in the price process. Imposing such a regime change with a crash to a *more reasonable* price level can be viewed as the acknowledgement of the fact that a bubble will not last forever (that is, until infinity), violating the prediction of the rational expectations framework. Thus, this model assumption is close in spirit to point 3 in our definition 1.3.4: there exists the risk of a crash and the price process including the crash constitutes a *reasonable* price process. However, in the present case, the (necessarily) exogenous description of a crash does not allow for the possibility that the pre-drawdown behavior of the stock price be related to the probabilistic structure of the crash.

1.6 Discussion

Let us close with a few lessons learned for the two essential tasks of **bubble detection** and **bubble modeling** within our framework. Bubble detection essentially refers to the binary decision *bubble* or *no bubble*, informed by pre-drawdown data; bubble modeling refers to an attempt at giving a full (probabilistic) description of an asset's price evolution in a bubbly episode, either pre- or post-drawdown.

Rational expectation models try to square the idea of overvaluation with market efficiency, which results in bubble models that are consistent with utility-maximizing investors only in a rather special situation: propositions 1.2.1 and 1.2.2 imply that a payoff at infinity and infinitely many market participants are necessary to allow for rational expectation bubbles.[29] As discussed in section 1.2.1.3, this type of models entirely relies on additional hypotheses, as, by construction, a payoff at infinity cannot be verified. One may well, of course, reverse the logic of proposition 1.2.1 and 1.2.2 and define a bubble as a process that **does not** emerge as the equilibrium price process of an economy populated by utility-maximizing investors. This means, however, that some kind of market failure drives the evolution of the stock price. To arrive at sensible models of bubbles, this implies that one should strive to explicitly model this inefficiency.

In an attempt to do so, we define inefficient market bubbles in section 1.3 based on a violation of efficiency, in accord with the possibility (risk) of a drawdown. Based on the bubble characteristics $(\tilde{S}, \tau_J, X)$, we introduce two types of bubbles (definition 1.3.4), which both are based on an inefficient pre-drawdown process $\tilde{S}$ and differ in their interpretation of the drawdown (τ_J, X). A type-I bubble can be understood as a first (partial) departure from rational expectations: while we acknowledge and model the existence of an instability (τ_J, X), the full stock price process $S = \tilde{S}^{\tau_J, X}$ is still an efficient market. More severely, in a type-II bubble the stock price S departs from efficient levels (represented by X) and returns to efficiency with a crash at τ_J. That said, in general, without additional information on the probabilistic structure of τ_J, we will not be able to distinguish type-I and type-II inefficient bubbles by looking at a single bubble episode.

Bubble detection, however, should take place on $[0, \tau_J)$ and thus merely concerns detecting inefficiency of $\tilde{S}$, irrespective of the type. This simpler task still exposes us to well-known joint hypothesis problems. Recall the discussion in section 1.4.3.3, suggesting that efficiency in *price levels* – and thus also the existence of a bubble – is a matter of perspective, as high valuation may a priori always be explained by high expected future payoffs (or profitability) and a drawdown as a result of changes in expectations and risk aversion. Instead, for bubble detection, we have to focus our attention on models that generate testable signals of inefficiency. Notable examples are situations characterised by heterogeneous beliefs and overconfidence (e.g., [214]), leading to increased trading volume, or momentum trading (e.g., [175]), imitation and herding (e.g., [155]), leading to accelerating returns. These testable signals in *price evolution* and variables such as *trading volume* allow for bubble detection, verified by the occurrence of a crash. Such models have the potential to avoid *joint hypothesis problems* in testing for market efficiency, first comprehensively described by [91] and still open for discussion; see, e.g., [109]. As a straightforward example, if we find the dynamics of price or return to be explosive in finite time (e.g., the models in

[29] As demonstrated in the overlapping generations model of [241].

sections 1.4.2.1 and 1.4.2.2), then at some point we can expect a regime change. We know the exuberant/explosive/accelerating behavior necessarily has to end and can constitute a bubble without further assumptions (see example 1.6.2 below).

Having said that, the second task of **bubble modeling** is even more ambitious. If we aim for an accurate prediction of a drawdown or, more generally, increase robustness of financial problem formulations vis-à-vis asset price bubbles, the full structure of a bubble will become important. The decomposition $(\tilde{S}, \tau_J, X)$ is a good starting point, but will likely require choosing between type-I and type-II bubble as well as additional assumptions on the probabilistic structure of the drawdown.[30]

Further research. Apart from an immediate extension of the framework in section 1.3 to

1. a higher dimensional setting including several price processes and market instruments,
2. other types of market efficiency such as statistical arbitrage or macroeconomic specifications of risk premia, and
3. a more detailed description of the correction mechanism in type-II bubbles (in the present setting merely a jump from $\tilde{S}_{\tau_J-}$ to X_{τ_J}), which is not always as immediate as during the 1987 Dow Jones market crash (cf. figure 1.1),

let us give two examples of research questions induced by the decomposition of inefficient market bubbles that may be promising to investigate.

Example 1.6.1 (Explosive processes with a jump.). The breakdown of an asset price with a bubble within the framework of section 1.3 provides a solid theoretical justification for using explosive processes (in combination with a random time of a crash) in asset price modeling; see the processes in sections 1.4.2.1-1.4.2.2 above for applications in the literature. Such processes have received only little attention in the stochastic processes literature, with the notable exception of [124] and [126], who study single jump processes with deterministic, explosive jump intensity in combination with an independent Brownian motion. In particular, absolutely continuous changes of measure, which are essential to arbitrage theory and option pricing, have not been investigated yet in situations such as

1. section 1.7.3, which is similar to [126] but allows for a more general relation of drift and jump intensity, and
2. section 1.4.2.2, where the jump intensity depends on the Brownian evolution.

Example 1.6.2 (Bubble detection.). Detecting a bubble with bubble characteristics $(\tilde{S}, \tau_J, X)$ in a market obeying a market efficiency condition $\mathcal{E}$ (cf. the definitions in section 1.3) in its build-up phase (that is, prior to τ_J) amounts to asserting whether the pre-drawdown process $\tilde{S}$ violates market efficiency. One may use the weak market efficiency condition $\mathcal{E} =$ (NFLVR) and a homogeneous diffusion of the form

$$d\tilde{S}_t = \mu_0 \tilde{S}^\alpha dt + \sigma_0 \tilde{S}^\beta dW_t, \quad t \in [0, \infty), \tag{1.88}$$

[30] E.g., an assumption on $\tilde{S}$ in section 1.4.2.1, cf. remark 1.4.3, an assumption on τ_J in section 1.4.2.2, cf. remark 1.4.4, a distributional assumption on τ_J in section 1.4.3.1 or an assumption on t_0 influencing τ_J in section 1.4.3.2.

for parameters $\mu_0, \sigma_0, \alpha, \beta \in (0, \infty)$ and a Brownian motion $(W_t)_{t \in [0,\infty)}$. Then a version of Feller's test for explosion (see, e.g., corollary 4.4 in [50]) can be used in combination with a sufficient condition for $\tilde{S}$ to violate (NFLVR),

$$\exists t \in [0, \infty) \colon \mathbb{P}\left[\tilde{S}_t = \infty\right] > 0 \Longrightarrow \tilde{S} \text{ violates (NFLVR).} \tag{1.89}$$

Thus, the problem is reduced to estimating $(\hat{\mu}_0, \hat{\sigma}_0, \hat{\alpha}, \hat{\beta})$ from data – for which an adaption of classic results for non-explosive diffusions, cf. [2] and references therein, may be warranted. While it seems rather extreme to use what is essentially a relative pricing approach on a single asset, this is very robust and avoids joint hypothesis problems.

1.7 Appendix

1.7.1 Technical results

Below we collect results from the literature that are used above; applied/extended to our setting of explosive processes.

Lemma 1.7.1. *Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a stochastic basis, let S be an $\mathbb{F}$-semimartingale with explosion as in definition 1.3.1 with explosion time τ and let $(\Lambda_t)_{t \in [0,\infty)}$ be a continuous, $\mathbb{F}$-adapted, increasing process with*

$$\begin{aligned} \Lambda &< \infty \quad \text{on } [[0, \tau)), \\ \Lambda &\equiv \infty \quad \text{on } [[\tau, \infty)). \end{aligned} \tag{1.90}$$

Then there exists a filtered probability space $(\hat{\Omega}, \mathcal{G}, \mathbb{G}, \hat{\mathbb{P}})$ and a random time τ_J with hazard process Λ such that the compensated jump process

$$\mathbb{1}_{\{\tau_J \leqslant t\}} - \Lambda_{t \wedge \tau_J}, \quad t \in [0, \infty), \tag{1.91}$$

is a $\mathbb{G}$-martingale and S is a $\mathbb{G}$-semimartingale with explosion.

Proof of lemma 1.7.1. We follow a standard approach in the literature (see, e.g., section 6.5 in [23]) and take $(\hat{\Omega}, \mathcal{G}, \hat{\mathbb{P}})$ to be the product extension of $(\Omega, \mathcal{F}, \mathbb{P})$ that supports a uniformly $[0,1]$-distributed random variable ξ, define

$$\tau_J = \inf\{t \in [0, \infty) | e^{-\Lambda_t} \leqslant \xi\} \tag{1.92}$$

and let $\mathbb{G}$ be the smallest right-continuous filtration generated by $\mathbb{F}$ and $\mathbb{1}_{\{\tau_J \leqslant \cdot\}}$. Then, by construction, τ_J is a $\mathbb{G}$-stopping time with $\mathbb{P}\left[\tau_J < \tau\right] = 1$ and the compensated jump process

$$\mathbb{1}_{\{\tau_J \leqslant t\}} - \Lambda_{t \wedge \tau_J}, \quad t \in [0, \infty), \tag{1.93}$$

is a $\mathbb{G}$-martingale. Moreover, by construction and continuity of Λ, any $\mathbb{F}$-local martingale is a $\mathbb{G}$-local martingale (this is the so-called *martingale invariance property*) and hence any $\mathbb{F}$-semimartingale (with explosion) is a $\mathbb{G}$-semimartingale (with explosion). This completes the proof of lemma 1.7.1. □

Next we show that a class of strictly positive semimartingales with explosion, whose stochastic logarithm is a quasi-left-continuous local submartingale with explosion, allows for a straightforward construction of a crash such that the crashed process is a local martingale. The result can in principle be extended to processes that may hit zero (and stay there), however then more care is needed in describing the stochastic exponential and logarithm; see [167] for a note on stochastic exponentials and logarithms on stochastic intervals.

Lemma 1.7.2. *Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a stochastic basis, let $(S_t)_{t\in[0,\infty)}$ be an $\mathbb{F}$-semimartingale with explosion as in definition 1.3.1 with explosion time τ such that $S > 0$ and $S_- > 0$, let $(Y_t)_{t\in[0,\infty)}$ be the stochastic logarithm[31] of S, assume that there exists a continuous, $\mathbb{F}$-adapted, increasing process $(A_t)_{t\in[0,\infty)}$ with $A \equiv \infty$ on $[[\tau, \infty))$ such that, for every $n \in \mathbb{N}$, there exists some decomposition $Y^{\tau_n} = Y_0 + M^n + A^{\tau_n}$ for a local martingale M^n and an $\mathbb{F}$-stopping times τ_n, and let $\kappa \in (0, 1)$ be a constant. Then A/κ is the hazard process of a random time τ_J with the property that the stopped process*

$$S_t^{\tau_J,\kappa} = S_t \mathbb{1}_{\{t<\tau_J\}} + S_{\tau_J -} (1-\kappa)\, \mathbb{1}_{\{\tau_J \leqslant t\}}, \quad t \in [0, \infty) \tag{1.94}$$

is a $(\mathbb{P}, \mathbb{G})$-local martingale, where $\mathbb{G}$ is the filtration generated by $\mathbb{F}$ and $\mathbb{1}_{\{\tau_J \leqslant \cdot\}}$.

Proof of Lemma 1.7.2. Continuity of A and the assumption on Y imply that, for every $n \in \mathbb{N}$, there exists a unique (up to indistinguishability) local martingale M^n and continuous finite variation process A^n such that $Y^{\tau_n} = Y_0 + M^n + A^n$. By definition, S fulfills the stochastic integral equation

$$S_t = \int_0^t S_{s-} dY_s = \int_0^t S_{s-} \left(dM_s^n + dA_s^n\right), \quad \text{on } [[0, \tau_n]]. \tag{1.95}$$

This implies that the crashed process $S^{\tau_J,\kappa}$ solves the integral equation

$$S_t^{\tau_J,\kappa} = \int_0^t \left(S_{s-}^{\tau_J,\kappa}\right) \mathbb{1}_{\{s<\tau_J\}} \left(d\left(M_s^n\right) + d\left(A_s^n\right) - \kappa d\mathbb{1}_{\{\tau_J \leqslant s\}}\right), \quad \text{on } [[0, \tau_n]]. \tag{1.96}$$

The construction in lemma 1.7.1 ensures that for the filtration $\mathbb{G}$, generated by $\mathbb{F}$ and $\mathbb{1}_{\{\tau_J \leqslant \cdot\}}$, it holds that $\frac{1}{\kappa} A^{\tau_J} - \mathbb{1}_{\{\tau_J \leqslant \cdot\}}$ and M^n are $\mathbb{G}$-local martingales. Then, using $A = A^n$ on $[[0, \tau_n]]$ and equation (1.96), we know that, for every n, $(S^{\tau_J,\kappa})^{\tau_n}$ is a $\mathbb{G}$-local martingale. We can conclude, using the result of theorem 44(e) in [203], that $S^{\tau_J,\kappa}$ is a $\mathbb{G}$-local martingale. The proof of lemma 1.7.2 is thus completed. □

With a straightforward application of Girsanov's theorem, we get a characterization of strictly positive Itô-processes that satisfy (NFLVR).

Lemma 1.7.3. *Let the setting in section 1.3.2 be fulfilled on a time horizon $[0, T]$ for some $T \in [0, \infty)$, let S be a strictly positive Itô-process with explosion time $\tau : \Omega \to [0, \infty]$ as defined in definition 1.4.1, assume that the filtration $\mathbb{F} \subseteq \mathbb{G}$ is generated W and assume that every $\mathbb{F}$-local martingale is a $\mathbb{G}$-local martingale. Then S is a (NFLVR)-efficient market on $[0, T]$ if and only if the following properties are fulfilled.*

[31] The stochastic logarithm L of a strictly positive process X with explosion is a process that, for any $n \in \mathbb{N}$, is given as the solution of the integral equation $dL_t = 1/X_t dX_t$ on $[[0, \tau_n]]$. It is thus well-defined on $\cup_{n\in\mathbb{N}}[[0, \tau_n]] = [[0, \tau))$.

1. $\mathbb{P}[\tau > T] = 1$,
2. $\mathbb{P}\left[\int_0^T \left(\frac{\alpha_s}{\beta_s}\right)^2 ds < \infty\right] = 1$ *and*
3. $\mathbb{E}\left[\exp\left(-\int_0^T \frac{\alpha_s}{\beta_s} dW_s - \frac{1}{2}\int_0^T \left(\frac{\alpha_s}{\beta_s}\right)^2 ds\right)\right] = 1.$

Proof of Lemma 1.7.3. Below we write g_s for a process g that may depend on (s, S_s). First assume that S satisfies *No Free Lunch with Vanishing Risk* on $[0, T]$. By the fundamental theorem of asset pricing in [71], there exists a measure $\mathbb{Q}$ equivalent to $\mathbb{P}$ such that S is a $(\mathbb{Q}, \mathcal{G})$-local martingale on $[0, T]$. This implies that S does not explode $\mathbb{Q}$-a.s., hence $\mathbb{P}$-a.s. on $[0, T]$. Moreover, a result in [237][32] shows that S is a $(\mathbb{Q}, \mathcal{F})$-local martingale. Then, by the martingale representation theorem[33] there exists a predictable, square-integrable process $f : [0, T] \times \Omega \to \mathbb{R}$ such that the martingale density process $(Z_t)_{t\in[0,T]}$ given by $Z_t = \mathbb{E}\left[\frac{d\mathbb{Q}}{d\mathbb{P}}\middle|\mathcal{F}_t\right]$ has the form

$$Z_t = \exp\left(-\int_0^t f_s dW_s - \frac{1}{2}\int_0^t f_s^2 ds\right), \quad t \in [0, T]. \tag{1.97}$$

By the Girsanov-Meyer theorem[34] we know that Z defines a $\mathbb{Q}$-local martingale $(N_t)_{t\in[0,T]}$ through

$$N_t = \int_0^t \beta_s dW_s + \int_0^t \beta_s f_s ds, \quad t \in [0, T]. \tag{1.98}$$

As S is a $\mathbb{Q}$-local martingale, so is $S - N = \int_0^{\cdot} (\alpha_s - \beta_s f_s)\, ds$, which is the case if and only if we have $\frac{\alpha}{\beta} = f$. This implies that $\frac{\alpha}{\beta}$ is a square integrable process with probability 1 and that

$$\exp\left(-\int_0^T \frac{\alpha_s}{\beta_s} dW_s - \frac{1}{2}\int_0^T \left(\frac{\alpha_s}{\beta_s}\right)^2 ds\right) \tag{1.99}$$

is a well defined random variable with expected value 1.

Conversely, assume that conditions 1–3 hold. Then we can define a measure $\mathbb{Q}$ equivalent to $\mathbb{P}$ by its Radon-Nikodym derivative on $\mathcal{F}_T$,

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \exp\left(-\int_0^T \frac{\alpha_s}{\beta_s} dW_s - \frac{1}{2}\int_0^T \left(\frac{\alpha_s}{\beta_s}\right)^2 ds\right). \tag{1.100}$$

Again we apply the Girsanov-Meyer theorem to get that the process

$$\int_0^t \beta_s dW_s + \int_0^t \frac{\alpha_s}{\beta_s}\beta_s ds = S_t, \quad t \in [0, T] \tag{1.101}$$

is a $(\mathbb{Q}, \mathbb{F})$-local martingale and hence, by assumption, a $(\mathbb{Q}, \mathbb{G})$-local martingale, completing the proof of lemma 1.7.3. □

[32] See theorem 3.6 in [101].
[33] See, e.g., corollary 4 in chapter IV of [203].
[34] See, e.g., theorem 20 in chapter III of [203]

1.7.2 Measure change and No-Good-Deal (NGD) for jump diffusions

Below we extend No-Good-Deal bounds introduced in section 1.4.1.3 above to a jump diffusion setting, where a useful instantaneous formulation can be derived. For simplicity we restrict ourselves to a finite time horizon $[0, T]$ for some $T \in (0, \infty)$.

1.7.2.1 Jump diffusion with integrable intensity

The following is a standard construction, details of which can found in, e.g., chapter III.5 of [143] or chapter 6.6 of [23]. Let the setting in section 1.3.2 be fulfilled, let $\mathbb{F} \subseteq \mathbb{G}$ be a filtration satisfying the usual hypothesis, let W be an $\mathbb{F}$-Brownian motion, let $(\lambda_t)_{t\in[0,T]}$ be an $\mathbb{F}$-progressively measurable, locally integrable process, $(N_t)_{t\in[0,T]}$ be a conditional Poisson counting process with intensity λ, further assume that the filtration $\mathbb{G}$ is generated by W and N, let μ, σ, κ be $\mathbb{G}$-predictable processes with $\kappa \geqslant -1$, suitably integrable, such that the equation

$$S_t = S_0 + \int_0^t S_{s-} \left(\mu_s ds + \sigma_s dW_s + \kappa_s dN_s \right), \tag{1.102}$$

has a unique strong solution $(S_t)_{t\in[0,T]}$ and let $(\gamma_t)_{t\in[0,T]}$ be an $\mathbb{G}$-predictable process with values in $(-1, \infty)$ satisfying

$$\int_0^T \frac{1}{\sigma_s^2} (\kappa_s(1+\gamma_s)\lambda_s)^2 ds < \infty \text{ and } \int_0^T (1+\gamma_s)\lambda_s < \infty, \quad \mathbb{P}\text{-a.s.} \tag{1.103}$$

Then the process $(Z_t)_{t\in[0,T]}$ given by

$$Z_t = 1 + \int_0^t Z_{s-} \left(-\frac{1}{\sigma_s}(\mu_s + \kappa_s(1+\gamma_s)\lambda_s) dW_s + \gamma_s \left(dN_s - \lambda_s ds \right) \right) \tag{1.104}$$

defines a measure $\mathbb{Q} \approx \mathbb{P}$ if $\mathbb{E}[Z_T] = 1$, in which case S is a $\mathbb{Q}$-local martingale. In fact, all measures $\mathbb{Q} \approx \mathbb{P}$ such that S is a local $\mathbb{Q}$-martingale admit a density of the form (1.104).

1.7.2.2 A single jump diffusion with possibly explosive intensity

The local integrability of the process λ above is essential to have a well-defined Poisson process N. If we restrict ourselves to the first jump of N, we can allow for a possibly non-integrable intensity and explosive behavior up to the jump. This has been rigorously elaborated in [126] in the setting of deterministic hazard rate. In particular, let the setting in section 1.3.2 be fulfilled, let $h : [0, T) \to [0, \infty)$ and $\kappa : [0, T) \to (0, 1]$ be differentiable functions, let $\mu_0 \in \mathbb{R}$, $\sigma_0 \in (0, \infty)$ be constants, let J be a single jump process with hazard rate h, let $\tau_J = \inf\{t \in [0, \infty) | J_t = 1\}$ be the random time of the jump, let W be a Brownian motion and assume that $\mathbb{G}$ is the filtration generated by W and J, let $(S_t)_{t\in[0,T]}$ be the solution of the integral equation

$$S_t = S_0 + \int_0^t S_{s-} \left((\mu_0 + \kappa(s)h(s)\mathbb{1}_{\{s \leqslant \tau_J\}}) ds + \sigma_0 dW_s - \kappa(s) dJ_s \right), \tag{1.105}$$

and let $\gamma : [0,T) \to (-1,\infty)$ be a differentiable function that is uniformly bounded away from -1 such that

$$\int_0^T (\kappa(t)h(t)\gamma(t))^2 dt < \infty \text{ and } \int_0^T \mathbb{1}_{\{\mathbb{P}[T \leqslant J]>0\}} h(t)(1+\gamma(t))dt < \infty. \tag{1.106}$$

Then the process $(Z_t)_{t\in[0,T]}$ given by

$$Z_t = 1 + \int_0^t Z_{s-}\left(-\frac{1}{\sigma_0}(\mu_0 - \kappa(s)h(s)\gamma(s)\mathbb{1}_{\{s\leqslant \tau_J\}})dW_s + \gamma_s\left(dJ_s - h(s)\mathbb{1}_{\{s\leqslant \tau_J\}}ds\right)\right) \tag{1.107}$$

satisfies $\mathbb{E}[Z_T] = 1$ and defines a measure $\mathbb{Q} \approx \mathbb{P}$; with S being a $\mathbb{Q}$-local martingale. Note that μ and h might not be defined at T, which is negligible in the integral. This is a straightforward (though non-trivial) extension of equation (1.104) to the single jump setting. The first condition of (1.106) ensures that a Girsanov change of measure is well-defined, whereas the second condition ensures that, in case there is non-negligible probability that the jump does not happen on $[0,T]$, this persists under the new measure.

1.7.2.3 No-Good-Deal

From equations (1.104) and (1.107) one can derive straightforward bounds on the instantaneous Sharpe ratios of S and Sharpe ratios of derivatives and portfolios derived from S. See appendix A in [25], where the seminal work in [115] is extended to the jump diffusion setting. In a slight abuse of notation (recall that we used an absolute bound on the price in equation (1.44), such a bound is called instantaneous or dynamic *No-Good-Deal bound*. [162] have generalized this to Lévy models.

Definition 1.7.1 (No-Good-Deal (NGD)). Let the setting in 1.7.2.1 be fulfilled and $K \in (0,\infty]$. Then S satisfies *dynamic* $(\text{NGD})_K$ if there exists a measure $\mathbb{Q} \approx \mathbb{P}$ with a density of the form (1.104) for a $\mathbb{G}$-predictable process γ such that for almost every $t \in [0,T]$ it holds that

$$\frac{1}{\sigma_t^2}(\mu_t + \kappa_t(1+\gamma_t)\lambda_t)^2 + \gamma_t^2\lambda_t < K. \tag{1.108}$$

For a single jump diffusion with possibly explosive intensity, as a direct extension of definition 1.7.1, the No-Good-Deal condition accounts for the fact that the process follows a Geometric Brownian motion after the first jump.

Definition 1.7.2 (No-Good-Deal (NGD) for a single jump diffusion). Let the setting in 1.7.2.2 be fulfilled and and $K \in (0,\infty]$. Then S satisfies *dynamic* $(\text{NGD})_K$ if there exists a measure $\mathbb{Q} \approx \mathbb{P}$ with a density of the form (1.104) for an $\mathbb{G}$-predictable process γ such that for almost every $t \in [0,T]$ it holds that

$$\frac{1}{\sigma_0^2}(\mu_0 - \kappa(t)h(t)\gamma_t\mathbb{1}_{\{t\leqslant \tau_J\}})^2 + \gamma_t^2 h(t)\mathbb{1}_{\{t\leqslant \tau_J\}} < K. \tag{1.109}$$

As quantities given by the market implied measure $\mathbb{Q}$, the two quantities in the sum of (1.108) and (1.109) represent the diffusion risk and $\gamma^2 h$ represents the jump risk of the *instantaneous market price of risk*.

1.7.3 Beyond the bubble framework – further assumptions

As noted above in sections 1.4.2.1-1.4.2.2, the framework presented in this paper allows – necessarily – for a great deal of slack in the modeling process. We have criticized the rational expectations framework for its additional assumptions necessary in bubble detection (constant expected return and stationary dividend payments; see section 1.2.1.3) – here we want to discuss (considerably weaker) additional assumptions within our framework that implicitly have been used in the literature. For our discussion we need to use straightforward extensions of section 1.7.2 that have NOT been rigorously presented in the literature yet.

General setting. For this purpose, we generalize the setting of the section above. Assume the setting in section 1.3.2 and let $T \in [0, \infty)$ be a finite time horizon, let $\sigma_0 \in (0, \infty)$ be a constant, let $\mu : [0, T) \to [0, \infty)$ be a continuous function with $\lim_{t \to T} \mu(t) = \infty$ and let $\tilde{S}$ be an Itô process as in section 1.4.1.1 with (possible) explosion time T that satisfies the stochastic differential equation

$$\frac{d\tilde{S}_t}{\tilde{S}_t} = \mu(t)dt + \sigma_0 dW_t. \tag{1.110}$$

Assume we want to model a (NGD)$_\infty$-bubble with a crash as in section 1.4.1.2 given by a constant relative crash size κ and a random time τ_J with hazard process Λ given by

$$\Lambda_t = \int_0^t h(s)ds, \quad t \in [0, T], \tag{1.111}$$

for a positive continuous function $h : [0, T) \to \mathbb{R}$. Then the crashed process S is given by

$$\frac{dS_t}{S_{t-}} = \mu(t)\mathbb{1}_{\{t<\tau_J\}}dt + \sigma_0 \mathbb{1}_{\{t<\tau_J\}}dW_t - \kappa dJ_t, \quad t \in [0, T], \tag{1.112}$$

where $(J_t)_{t\in[0,T]}$ is the single jump process $J = \mathbb{1}_{\{\tau_J \leqslant \cdot\}}$. In our setting, the pre-drawdown process $\tilde{S}$ violates (NGD)$_\infty$, as μ is unbounded. To get a bubble as in definition 1.3.4, we need to specify μ and h such that the crashed process S satisfies (NGD)$_\infty$.

No-Good-Deal bounds and additional assumptions. In section 1.4.2.1 we made the choice $\mu = \kappa h$, which implies that S is a martingale (under the probability measure $\mathbb{P}$). However, for the crashed process to satisfy (NGD)$_\infty$ we have a much weaker sufficient condition: that there exists a function $\gamma : [0, T) \to (-1, \infty)$, as in section 1.7.2.1, bounded away from -1 and suitably integrable, such that the No-Good-Deal bound (1.108) up to the first jump is fulfilled. Let us look at this a little more closely. If we define $\phi : [0, T) \to \mathbb{R}$ by

$$\phi(t) = \frac{1}{\sigma_0}\left(\mu(t) - \kappa(1 + \gamma(t))h(t)\right), \quad t \in [0, T), \tag{1.113}$$

then S fulfills (NGD)$_\infty$ if

1. ϕ is square integrable
2. $(1 + \gamma)h$ is integrable if h is and

3. the No-Good-Deal bound

$$\left(\phi(t)^2 + \gamma(t)^2 h(t)\right) \mathbb{1}_{\{t \leqslant \tau_J\}} < \infty, \quad t \in [0, T]. \tag{1.114}$$

is fulfilled. We can write the expected instantaneous returns process $(R_t)_{t\in[0,T]}$ as

$$R(t) = \mu(t) - \kappa h(t) = \sigma_0 \phi(t) + \kappa h(t)\gamma(t). \tag{1.115}$$

This formulation allows for the interpretation of ϕ as being the *market price of diffusion risk* and γ the *market price of jump risk*, cf., e.g., [25]. The conditions imposed by $(\mathrm{NGD})_\infty$ are rather weak, so one might want to add additional assumptions on risk-preferences of investors in the market (via assumptions on R, ϕ, and γ). For example, in increasing generality,

(a) Arguably quite restrictive, one can assume that investors are risk-neutral on average during the bubble and thus the required return satisfies $R \equiv 0$. This is implies immediately that $\mu = \kappa h$.

(b) One may impose a No-Good-Deal bound $K \in (0, \infty)$ on S, such that equation (1.114) is fulfilled with bound K.

(c) No additional assumption, merely that S has a $(\mathrm{NGD})_\infty$-bubble.

A specific model – JLS model. To analyze the implications of above assumptions (a)–(c) for a specific model choice, we consider the JLS model; see [225] and references therein. In this model, a network of traders is analyzed to find that herding, positive feedback and imitative behavior in the market leads to systemic instability and thus a possible financial crash given by a hazard rate

$$h(t) = B_1(T-t)^{-\alpha} + B_2(T-t)^{-\alpha} \cos\left(\omega \ln(T-t) - \psi\right) \tag{1.116}$$

for parameters $\alpha, B_1, B_2, \omega, \psi \in (0, \infty)$ chosen such that h is non-negative. For a detailed derivation we refer to [225] and [217]. This is a so-called risk-driven bubble, as the possibility of a crash (as opposed to the price process) is specified; see section 5 in [225] for details on this terminology.

Using the assumptions above, we seek to derive conditions on the price S (or, equivalently, its return μ) to yield reasonable bubble models given by equation (1.112) based on a crash given by the hazard rate h.

(a) If we assume risk-neutral investors we get that $\mu = \kappa h$ and thus μ necessarily behaves like h in equation (1.116).

(b) If we assume a finite No-Good-Deal bound K, equation (1.114) implies that

$$\lim_{t\to T} |\mu(t)(T-t)^{\alpha} - \kappa(B_1 + B_2 \cos\left(\omega \ln(T-t) - \psi\right))| = 0 \tag{1.117}$$

Thus, with speed of convergence depending on α and K, μ eventually shows explosive behavior and log-periodic oscillations around the critical time T.

(c) If we only assume that S satisfies $(\text{NGD})_\infty$ then little can be said about the exact behavior of μ. As an example, we may have market prices of risk given by $\gamma(t) = 1/(B_1 + B_2 \cos(\omega \ln(T-t) - \psi))$ and $\phi \equiv 0$ and, from equation (1.115), $\mu(t) = (T-t)^{-\alpha}$. Thus, although market prices of risk ϕ and γ are both bounded, μ may show no log-periodic oscillations (while S still satisfies No-Arbitrage!). As another example, in the case $\alpha \in (0, \frac{1}{2})$ we could have $\gamma \equiv 0$ and $\phi = (c - \kappa h)/\sigma_0$ for some constant $c \in (0, \infty)$. Then ϕ is square integrable and an admissible market price of risk while $\mu = c - \kappa h + \kappa h \equiv \text{const}$ on $[0, T]$.

Discussion. In the papers that introduced the JLS model (see [153], [154] or [225] for an overview) it is assumed that the No-Arbitrage condition holds *and* investors are risk-neutral, thus $\mu = \kappa h$ and the price mimics the explosive behavior and log-periodic oscillations of h. With this assumption, an analysis of the price process leads to a detailed knowledge about the hazard of a crash.

Responding to criticism from [142] regarding the assumption of *risk-neutrality*, in [155] it is claimed that changing risk preferences within a bounded expected return R or some bounded market price of crash risk γ[35] cannot prevent that explosive behavior and log-periodic signatures are apparent in the price. [96] extends the criticism presented in [142], pointing out that the expected required return need not be constant. In response to this, [233] and [255] introduce the general case of assuming only No-Arbitrage[36], however the discussion is reduced to the case $\gamma \equiv 0$; in this case it is claimed that log-periodic signatures of h persist. Long story short, the issue that additional assumptions are necessary to derive a relationship $\mu \leftrightarrow h$ is well known and has been discussed in the literature, but a comprehensive study has not been presented yet.

The above can be seen as a first step[37] towards such a comprehensive study, with a full description of possible discount factors and a clear framework that allows one to accurately determine the assumptions that are necessary to reach the desired conclusions. We have seen that, in general, No-Arbitrage (or the stronger $(\text{NGD})_\infty$) alone does not ensure that explosive behavior and log-periodic signature of h is observable in the price (see the examples in point (c) above) – in contradiction to some claims in the literature. In particular, it is possible that an extreme change in risk premia could lead to a situation where the increasing risk of a crash is not reflected in the price (resp. μ) until the crash happens. However, with reasonably weak additional assumptions (see point (b) – $(\text{NGD})_K$ for some $K \in (0, \infty)$ or, e.g., imposing a bound on required return R) one can infer that μ behaves like h around the critical time T and thus show explosive *and* log-periodic behavior.

To close the section, we want to note that similar conclusions, within the respective model framework, can be drawn for *return-driven* bubbles as defined in [225], where an analysis of (or assumptions on) market structures leads to a model for μ and one wants to infer information about the hazard of a crash h.

[35] Called ν and K, respectively, in [155].

[36] In the sense of (NFLVR) – the existence of an equivalent measure $\mathbb{Q}$ with associated density process $Z^{\mathbb{Q}}$. In [233] the equivalent terminology of a stochastic discount factor $M = Z^{\mathbb{Q}}$ is used.

[37] Note that we make the restrictive assumption that γ (and thus ϕ) are deterministic functions and thus, although covering a large variety, we can only apply our statements to a subclass of pricing kernels.

Chapter 2

Uniform integrability of single jump processes

2.1 Introduction

Local martingales that are not uniformly integrable martingales have recently gained increased attention in the stochastic processes and mathematical finance literature, being linked to special cases in arbitrage pricing theory ([83], [72], [113]) and to the occurence of bubbles ([179], [57], [130], [147], [148], [128], [22]). Based on the seminal paper of Loewenstein and Willard [179], theorem 4.1 in [148] characterizes three types of local martingales that can be used to model bubbles[1], depending on the model horizon:

(a) General local martingales on an infinite time horizon.

(b) Local martingales that are not uniformly integrable martingales on a stochastically unbounded, but finite time horizon.

(c) Strict local martingales on a bounded time horizon.

The result is based on the fact that local martingales, while being *instantaneous fair games*, may show a drop in expectation in the long term. Table 2.1 summarizes such a classification of local martingales. To date most of the literature is limited to a finite time horizon, thereby immediately excluding processes in (a) and (b). While processes in (a) rely on an infinite time horizon and seem somewhat ill-suited for financial modeling, processes in (b) are readily conceivable.[2] In the present paper we introduce a natural class of candidates for bubble processes on a finite but stochastically unbounded time horizon. We combine a homogeneous diffusion with a single jump (characterized by state dependent hazard rate and jump size) and provide a necessary and sufficient deterministic criterion to decide whether they are uniformly integrable martingales.

[1] Below, we refer to such processes as *mathematical (finance) bubbles*.

[2] One may argue, for example, that the finite model horizon for a stock price of a large corporation is more realistically described by an unbounded random rather than a bounded deterministic lifetime.

Table 2.1: Characterization of RCLL non-negative local martingales by martingale property and uniform integrability (UI). The uniform integrability property has explanatory power on the loss of mass only for true martingales. Strict local martingales are true supermartingales even if they are uniformly integrable.

	Non-UI	*UI*
Strict local martingale	$\exists t \in [0,\infty)\colon \mathbb{E}[M_t] < M_0$	
Martingale	$\mathbb{E}[M_\infty] < M_0$	$\mathbb{E}[M_\infty] = M_0$

While many models of mathematical bubbles lack a well-defined empirical basis, a single jump has a straightforward interpretation as a financial drawdown.

In particular, we look at processes $(S_t)_{t\in[0,\infty)}$ defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,\infty)}, \mathbb{P})$ that satisfy a homogeneous version of the stochastic differential equation

$$dS_t = b(t, S_t)\mathbb{1}_{\{t<\tau_J\}}dt + \sigma(t, S_t)1_{\{t<\tau_J\}}dW_t - \frac{b(t, S_{t-})}{h(t, S_{t-})}dJ_t, \tag{2.1}$$

with coefficient functions $b, \sigma, h : [0, \infty) \times \mathbb{R} \to [0, \infty)$, where $(W_t)_{t\in[0,\infty)}$ is an $(\mathcal{F}_t)_{t\in[0,\infty)}$-Brownian motion, $(J_t)_{t\in[0,\infty)}$ is a $\{0, 1\}$-valued single jump process with

$$\mathbb{P}\left[dJ_t = 1|\mathcal{F}_{t-}, J_{t-} = 0\right] = h(t, S_{t-}), \tag{2.2}$$

such that h is the *hazard rate* (also known as *intensity process*) of J, and τ_J denotes the time of the jump. For most reasonable choices of b, σ and h such that $b(t, x) \leqslant h(t, x)x$ for all $(t, x) \in [0, \infty) \times [0, \infty)$ (see specific examples in section 2.5 and the detailed setting in section 2.3 below) the resulting process S is non-negative and a local martingale. Intuitively, the local martingale property can be seen from the fact that S is the sum of a Brownian integral and a single jump process that grows instantaneously by $b(t, S_{t-})$ and has an expected instantaneous decline of $-(b(t, S_{t-})/h(t, S_{t-}))\mathbb{P}\left[dJ_t = 1|\mathcal{F}_{t-}, J_{t-} = 0\right] = -b(t, S_{t-})$. For $\tau_J < t$, it remains constant at S_{τ_J}. Above questions on uniform integrability and strict local martingality have been answered in various special cases of (2.1). As a simple example, assume that for constants $\lambda \in [0, \infty)$ and $\kappa \in (0, 1]$ we have

$$b(t, S_t) = \kappa\lambda S_t,\ \sigma(t, S_t) \equiv 0 \text{ and } h(t, S_t) \equiv \lambda. \tag{2.3}$$

One can directly calculate the expected value of $S_\infty = S_{\tau_J}$ as

$$\mathbb{E}[S_\infty] = \mathbb{E}[S_{\tau_J}] = (1-\kappa)S_0\mathbb{E}\left[e^{\kappa\lambda\tau_J}\right] = (1-\kappa)S_0\int_0^\infty \lambda e^{-\lambda t}e^{\kappa\lambda t}dt = S_0 \tag{2.4}$$

for $\kappa < 1$ and $\mathbb{E}[S_\infty] = \mathbb{E}[0] = 0$ for $\kappa = 1$. In this simple case S is a uniformly integrable martingale if and only if $\kappa < 1$. Similarly, it is easy to check that S is indeed a martingale. Let us present three variants of equation (2.1) in somewhat increasing generality, the last illustrating the setting in this article, where the question of interest is whether a process can be used as a model of mathematical bubbles in (stochastic) finite time.

Linear characteristics and time-dependent hazard rate. Based on their examination of single jump processes with a deterministic hazard rate in [125], the authors in [127] consider (within a more general setting) a solution to the SDE (2.1) assuming a finite time horizon $T \in [0, \infty)$ and coefficients

$$b(t, S_t) = \phi'(t) S_t,\ \sigma(t, S_t) = \sigma_0 S_t \text{ and } h(t, S_t) = h(t) \tag{2.5}$$

for $\sigma_0 \in (0, \infty)$ and continuously differentiable functions $\phi, h : [0, T) \to (0, \infty)$. They show, in particular, that the process $(S_t)_{t \in [0,T]}$ is a strict local martingale if and only if

$$\int_0^T h(t) dt = \infty \quad \text{and} \quad \int_0^T \left(h(t) - \phi'(t)\right) dt < \infty. \tag{2.6}$$

Due to the finite time window ($S_\infty = S_T$), any true martingale in this setting is immediately uniformly integrable.

Driftless homogeneous diffusion. There has been a lot of interest in the strict local martingale property of stochastic exponentials based on diffusions, see, e.g., [73], [164], [135], [184] and references therein. One can apply those results to a special case of (2.1) with a homogeneous diffusion function and zero drift,

$$b(t, S_t) \equiv 0,\ \sigma(t, S_t) = \sigma(S_t) \text{ and } h(t, S_t) \equiv 0. \tag{2.7}$$

In particular, for diffusion coefficients σ with $\sigma(\cdot) \neq 0$ and $\sigma^{-2}(\cdot)$ locally integrable on $(0, \infty)$, one can show that the process $(S_t)_{t \in [0,\infty)}$ is

1. a strict local martingale on any interval $[0, T]$ or $[0, \infty)$ if $\int_c^\infty x/\sigma^2(x) dx < \infty$ for some $c \in (0, \infty)$ and
2. a martingale that is not uniformly integrable if $\int_c^\infty x/\sigma^2(x) dx = \infty$ for all $c \in (0, \infty)$,

see, e.g., corollary 4.3 in [184]. Note that, as in the last example, for such pure diffusion processes the question whether $(S_t)_{t \in [0,\infty)}$ is a uniformly integrable martingale is trivial, as almost surely we have $S_\infty = 0$.[3]

Objectives in this chapter - state-dependent drift and diffusion coefficients. Below we consider homogeneous, state-dependent coefficient functions

$$b(t, S_t) = b(S_t),\ \sigma(t, S_t) = \sigma(S_t) \text{ and } h(t, S_t) = h(S_t) \tag{2.8}$$

for locally Lipschitz continuous b, σ and locally Hölder continuous h. As such we (partly) extend the homogeneous, state-dependent setting of a pure diffusion as in [184] and others to a single jump framework as in [126]. Our main result in section 2.4 below is concerned with a deterministic necessary and sufficient criterion on b, σ and h to decide whether $(S_t)_{t \in [0,\infty)}$ is a uniformly integrable martingale.

If one accepts that local martingales that are not uniformly integrable martingales are suitable processes to model bubbles, our result contributes to the financial literature on bubbles in several dimensions:

[3]One can see this by applying the classification of chapters 2 and 4 in [50] to the case of a driftless diffusion.

1. Single jump processes as in (2.1) (where the single jump J represents a financial crash of relative size b/h) are a simple and tractable alternative to include crash risk in financial models and serve as simple tool to integrate empirical features of bubbly markets into mathematical models. The main result below allows us to bridge one of the gaps between

   (a) the literature on bubbles based on explosive processes and a crash as in [228] with

   (b) the mathematical finance notion of bubbles as non-uniformly integrable martingales or strict local martingales discussed in section 2.5.2.

   See section 2.5.1.2 below for a specific example.

2. The classification of mathematical bubbles can be extended from single jump processes with deterministic intensity as in [126] to jumps whose hazard rate is random (state-dependent), allowing for more realistic description of crash risk. Moreover, we (partly) extend the setting of a pure diffusion as in [184], covering various models in the literature,[4] to include the financially relevant case of a crash, see sections 2.5.1.1 and 2.5.1.3 below for examples.

3. Equation (2.6) implies that single jump models with a deterministic hazard rate as in [126] can be mathematical bubble models only if there is an almost sure jump on a finite time interval $[0, T]$. Models based on a homogeneous diffusion as considered below feature a crash distributed on $[0, \infty)$. For an investor with deterministic finite investment horizon (as is standard in the literature) there is a nonzero probability that the crash does not happen within his investment horizon, a reasonable assumption in financial problem settings.

We close with a discussion of assumptions and open questions in section 2.5.3.

2.2 Notation

The following notation is used throughout the chapter. Unless stated otherwise, we consider stochastic processes *unique up to indistinguishability* and require stochastic integral equations on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to hold $\mathbb{P}$-a.s. We assume familiarity with the notions of a *martingale*, *supermartingale* and *local martingale*.[5] A stochastic process $(X_t)_{t\in[0,\infty)}$ is *uniformly integrable* if

$$\lim_{n\to\infty} \sup_{t\in[0,\infty)} \mathbb{E}\left[|X_t| \mathbb{1}_{\{n<|X_t|\}}\right] = 0. \tag{2.9}$$

For a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and random times $\sigma, \tau \colon \Omega \to [0, \infty]$ we use the *stochastic interval* notation

$$\begin{aligned} [[\sigma, \tau]] &= \{(\omega, t) \in \Omega \times [0, \infty] \colon \sigma(\omega) \leqslant t \leqslant \tau(\omega)\} \\ [[\sigma, \tau)) &= \{(\omega, t) \in \Omega \times [0, \infty] \colon \sigma(\omega) \leqslant t < \tau(\omega)\}. \end{aligned} \tag{2.10}$$

[4] For example, the CEV model or geometric Brownian motion.
[5] For an introduction, see, e.g., chapter 1 in [204].

For a stochastic process $(X_t)_{t\in[0,\infty)}$ we denote its left-continuous version by $(X_{t-})_{t\in[0,\infty)}$, that is, the process with the property that $X_{t-} = \lim_{s\nearrow t} X_s$ for all $t \in [0,\infty]$. For a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,\infty)}, \mathbb{P})$ with a right-continuous filtration and an $(\mathcal{F}_t)_{t\in[0,\infty)}$-stopping time $\tau : \Omega \to [0,\infty)$ we define the (itself right-continuous) filtration $(\mathcal{F}_{t\wedge\tau-})_{t\in[0,\infty)}$ consisting of the σ-algebras $\mathcal{F}_{t\wedge\tau-}$ given by

$$\mathcal{F}_{t\wedge\tau-} = \sigma\left(\{A \cap \{s < \tau\} : 0 \leqslant s \leqslant t, A \in \mathcal{F}_s\} \cup \mathcal{F}_0\right). \tag{2.11}$$

2.3 Setting

2.3.1 Definitions

Let $b : [0,\infty) \to [0,\infty)$ and $\sigma : [0,\infty) \to [0,\infty)$ be locally Lipschitz continuous functions with $\sigma^{-1}(0) = \{0\}$, let $B_0 \in (0,\infty)$, $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,\infty)}, \mathbb{P})$ be a filtered probability space with a right-continuous and $\mathbb{P}$-complete filtration, let $(W_t)_{t\in[0,\infty)}$ be a real valued $(\mathcal{F}_t)_{t\in[0,\infty)}$-Brownian motion, let $B\colon [0,\infty) \times \Omega \to (0,\infty]$ be the unique strictly positive process with the property that

1. for $(\mathcal{F}_t)_{t\in[0,\infty)}$-stopping times $(\tau_n)_{n\in\mathbb{N}}\colon \Omega \to [0,\infty)$ given by $\tau_n = \inf\{t \geqslant 0\colon B_t \geqslant n\}$, for all $n \in \mathbb{N}$ on $[[0,\tau_n]]$ we have

$$\begin{aligned} &\int_0^t \left(b(B_s)ds + \sigma^2(B_s)\right) ds < \infty \text{ and} \\ &B_t = B_0 + \int_0^t b(B_s)ds + \int_0^t \sigma(B_s)dW_s, \end{aligned} \tag{2.12}$$

   and

2. for the predictable $(\mathcal{F}_t)_{t\in[0,\infty)}$-stopping time $\tau\colon \Omega \to [0,\infty]$ given by $\tau = \sup_{n\in\mathbb{N}} \tau_n$ it holds that B is $(\mathcal{F}_{t\wedge\tau-})_{t\in[0,\infty)}$ -adapted,

let $h : [0,\infty) \to [0,\infty)$ be a locally Hölder continuous function with the property that[6]

$$\begin{aligned} &\frac{b(x)}{h(x)x} \in [0,1] \text{ for all } x \in (0,\infty) \text{ and} \\ &\lim_{x\to\infty} \frac{b(x)}{h(x)x} \text{ exists,} \end{aligned} \tag{2.13}$$

let $J : [0,\infty] \times \Omega \to \{0,1\}$ be an RCLL stochastic ("single jump") process with the property that for all $t \in [0,\infty)$ we have

$$\mathbb{P}[J_t = 1 | \mathcal{F}_{t\wedge\tau-}] = 1 - e^{-\int_0^{t\wedge\tau} h(B_s)ds}, \tag{2.14}$$

let $\tau_J : \Omega \to \mathbb{R}$ be the $\mathbb{P}$-a.s. unique random time with the property that for all $t \in [0,\infty)$ it holds that $\mathbb{P}[J_{t\wedge\tau} = 1] = \mathbb{P}[\tau_J \leqslant t \wedge \tau]$, let $(\mathcal{G}_t)_{t\in[0,\infty)}$ be the filtration generated by $(\mathcal{F}_{t\wedge\tau-})_{t\in[0,\infty)}$ and

[6]In the following, we employ the convention that $0/0 = 0$ to allow for $b(x) = h(x) = 0$, for some $x \in [0,\infty)$, while retaining notational convenience. The quantity $b(x)/h(x)x$ is the relative jump size and thus not relevant in cases where $h(x) = 0$. Let us also note here that assumption (A) below excludes the case $b \equiv h \equiv 0$.

J and let $S\colon [0,\infty) \times \Omega \to [0,\infty)$ be the $(\mathcal{G}_t)_{t\in[0,\infty)}$-adapted RCLL process with the property that for all $t \in [0,\infty)$ we have

$$S_t = B_0 + \int_0^t b(S_s)\mathbb{1}_{\{s<\tau_J\}}ds + \int_0^t \sigma(S_s)\mathbb{1}_{\{s<\tau_J\}}dW_s - \int_0^t \frac{b(S_{s-})}{h(S_{s-})}dJ_s. \tag{2.15}$$

2.3.2 Assumptions

Moreover, we assume that

(A) $\forall n \in \mathbb{N} \cap [B_0,\infty)\colon \mathbb{P}[\tau_n < \infty] = 1$, and

(B) $\lim_{x\to\infty} h(x)x^2/\sigma^2(x)$ exists and is finite.

2.3.3 Comments to the setting

Uniqueness of solutions of SDEs is understood as pathwise uniqueness. The local Lipschitz assumption on b and σ ensure that the integral equation (2.12) has a unique strong solution up to the random time τ, see, e.g., theorem 4.3 in [202]. The time τ is called explosion time of B. Local Lipschitz conditions on b and σ, $b \geqslant 0$ and $\sigma(0) = 0$ ensure strict positivity of B, see, e.g., theorem 4.1 in chapter 9 of [104]. The σ-algebra $(\mathcal{F}_{t\wedge\tau-})_{t\in[0,\infty)}$ includes precisely the information of the trajectories of B up to its explosion time τ. Uniqueness holds in law and pathwise, see, e.g., chapter 1 in [50].

The process J is a single jump process that jumps from 0 to 1 at time τ_J, for construction (and thus existence) see section 6.5 in [23]. For a measurable function $h : [0,\infty) \to [0,\infty)$ the process $(h(B_t)\mathbb{1}_{\{t<\tau\}})_{t\in[0,\infty)}$ can be understood as the $(\mathcal{F}_{t\wedge\tau-})_{t\in[0,\infty)}$-martingale intensity process of J (cf., e.g., chapter 6 of [23]). The additional requirement of local Hölder continuity is used in the application of the Feynman-Kac formula. We show below (equation (2.34) in the proof of theorem 2.4.1) that strict positivity of B and assumption (A) imply that $\mathbb{P}\left[\tau_J < \tau\right] = 1$ and thus in particular that $J_\infty = 1$. Similarly, assumption (B) is needed in the proof and clearly restricts the choice of σ and h, whereas equation (2.13) merely implies the natural condition that the relative jump size is in $[0,1]$ and excludes the special case of periodic behavior at infinity (and is therefore not listed as a distinct assumption in section 2.3.2).

Existence and uniqueness of the solution to equation (2.15) is guaranteed by the semimartingale property of the integrator (cf. Section 3 of [48] for a discussion of the semimartingale property of a stopped, time-inhomogeneous jump diffusion). Proposition 3.2 in [48] (for $X = S$ and $T_\Delta = \tau_J$) implies that $(S_t)_{t\in[0,\infty)}$ is a non-negative local martingale and thus, by Fatou's lemma, a non-negative supermartingale with the property that for all $(\mathcal{G}_t)_{t\in[0,\infty)}$-stopping times $\rho\colon \Omega \to [0,\infty)$ it holds that $\mathbb{E}\left[S_\rho\right] \leqslant S_0$. Alternatively, one can

1. check that W is still a Brownian motion with respect to $(\mathcal{G}_t)_{t\in[0,\infty)}$ and
2. use the local martingale property of the compensated jump process $\left(J_t - \int_0^{t\wedge\tau} h(B_s)ds\right)_{t\in[0,\infty)}$ (cf. section 6.5 of [23]).

Then S can be expressed as integrals with respect to the local martingales $J - \int_0^{\cdot\wedge\tau} h(B_s)ds$ and W and is itself a local martingale. The process S can be called a *single jump local martingale* as it follows the diffusion B and has a single jump at τ_J; thus obeying the equation

$$S = B\mathbb{1}_{\{\cdot<\tau_J\}} + \left(1 - \frac{b(B_{\tau_J})}{h(B_{\tau_J})B_{\tau_J}}\right) B_{\tau_J}\mathbb{1}_{\{\tau_J \leqslant \cdot\}} . \tag{2.16}$$

2.4 Main result

Proposition 2.4.1. *Assume the setting in section 2.3,* $\mathbb{P}[\tau_J < \tau] = 1$ *and let* $f : [0,\infty) \to [0,1]$ *be a measurable function. Then it holds that*

$$\mathbb{E}\left[\left(1 - f(B_{\tau_J})\right) B_{\tau_J}\right] = \mathbb{E}\left[\int_0^{\tau} h(B_t)\left(1 - f(B_t)\right) B_t e^{-\int_0^t h(B_s)ds} dt\right]. \tag{2.17}$$

Proof. We observe that $(\tau_n)_{n\in\mathbb{N}}$ and τ are $(\mathcal{F}_{t\wedge\tau-})_{t\in[0,\infty)}$-stopping times and thus τ an $(\mathcal{F}_{t\wedge\tau-})_{t\in[0,\infty)}$-predictable stopping time. Moreover, B has continuous trajectories on $[[0,\tau))$ and is $(\mathcal{F}_{t\wedge\tau-})_{t\in[0,\infty)}$-adapted. We can conclude that

$$(1 - f(B))\, B\, \mathbb{1}_{\{\cdot<\tau\}} \colon [0,\infty) \times \Omega \to [0,\infty) \tag{2.18}$$

is an $(\mathcal{F}_{t\wedge\tau-})_{t\in[0,\infty)}$-predictable process. Then the claim follows from

$$\mathbb{E}\left[\left(1 - f(B_{\tau_J})\right) B_{\tau_J}\, \mathbb{1}_{\{\tau_J<\tau\}}\right] = \mathbb{E}\left[\left(1 - f(B_{\tau_J})\right) B_{\tau_J}\right] \tag{2.19}$$

and part (ii) of Corollary 6.3. in [150]. The proof of proposition 2.4.1 is thus completed. □

Lemma 2.4.1. *Assume the setting and the assumptions in section 2.3, let* $v \colon (0,\infty) \to [0,\infty)$ *be a twice differentiable function with the property that it satisfies the ordinary differential equation*

$$\frac{1}{2}\sigma(x)^2\frac{\partial^2 v}{\partial x^2}(x) + b(x)\frac{\partial v}{\partial x}(x) = h(x)v(x), \quad x \in (0,\infty) \tag{2.20}$$

with boundary condition $\lim_{n\to\infty} v(n) = \infty, n \in \mathbb{N}$. *Then it holds that*

$$\lim_{n\to\infty} \frac{n}{v(n)} = 0 \iff \left(\lim_{x\to\infty} \frac{b(x)}{h(x)x} < 1\right) \text{ and } \left(\lim_{x\to\infty} \frac{h(x)x^2}{\sigma^2(x)} > 0\right). \tag{2.21}$$

Proof. First we note that

$$\frac{b(x)x}{\sigma^2(x)} = \frac{b(x)}{h(x)x}\frac{h(x)x^2}{\sigma^2(x)}. \tag{2.22}$$

Using assumption (B) and the existence of $\lim_{x\to\infty} b(x)/(h(x)x) \in [0,1]$, there are constants $p_0, q_0 \in [0,\infty)$ with the property that

$$\lim_{x\to\infty}\left|\frac{b(x)x}{\sigma^2(x)} - p_0\right| = 0, \quad \lim_{x\to\infty}\left|\frac{h(x)x^2}{\sigma^2(x)} - q_0\right| = 0. \tag{2.23}$$

Equation (2.22) implies that $p_0 \leqslant q_0$ and

$$p_0 < q_0 \iff \left(\lim_{x\to\infty} \frac{b(x)}{h(x)x} < 1\right) \text{ and } \left(\lim_{x\to\infty} \frac{h(x)x^2}{\sigma^2(x)} > 0\right). \tag{2.24}$$

As elaborated in Chapter 9.12 of [24], a substitution $y = 1/x$ transforms the solution of the ODE (2.20) to a solution $w\colon (0,\infty) \to [0,\infty)$ of the ODE

$$\frac{\partial^2 w}{\partial y^2}(y) + \left(\frac{2}{y} - \frac{2}{y^2}\frac{b\left(\frac{1}{y}\right)}{\sigma^2\left(\frac{1}{y}\right)}\right)\frac{\partial w}{\partial y}(y) - \frac{2}{y^4}\frac{h\left(\frac{1}{y}\right)}{\sigma^2\left(\frac{1}{y}\right)}w(y) = 0, \quad y \in (0,\infty), \tag{2.25}$$

with boundary condition

$$\lim_{n\to\infty} w\left(\frac{1}{n}\right) = \lim_{n\to\infty} v(n) = \infty. \tag{2.26}$$

The continuity theorem for solutions of ODEs, see theorem 3 on page 177 in [24], and equation 2.23 ensures that the behavior of the solution w of (2.25) at 0 does not change if we substitute the coefficient functions using p_0 and q_0 to arrive at the ODE

$$\frac{\partial^2 w}{\partial y^2}(y) + \left(\frac{2}{y} - \frac{2}{y}p_0\right)\frac{\partial w}{\partial y}(y) - \frac{2}{y^2}q_0 w(y) = 0, \quad y \in (0,\infty), \tag{2.27}$$

which has a *regular singular point* at 0. To analyze the behavior of w around 0, we have to look at the solutions $r_2 < r_1 \in \mathbb{R}$ of the *indicial equation* of the ODE (2.27), $r(r-1) + 2(1-p_0)r - 2q_0 = 0$, with solutions

$$r_{2,1} = p_0 - \frac{1}{2} \pm \frac{1}{2}\sqrt{4p_0^2 - 4p_0 + 8q_0 + 1}. \tag{2.28}$$

Applying the corollary after theorem 7 and theorem 8 of chapter 9 in [24], we get that in a small enough neighborhood of 0, the function w can be written as

$$\begin{aligned} w(y) &= \alpha w_1(y) + \beta w_2(y), \text{ with} \\ w_1(y) &= y^{r_1}\left(1 + \sum_{k=1}^{\infty} a_k y^k\right) \text{ and} \\ w_2(y) &= y^{r_2}\left(1 + \sum_{k=1}^{\infty} b_k y^k\right) + C w_1(y)\ln(y)\mathbb{1}_{\{r_1 \in r_2 + \mathbb{N}\}}. \end{aligned} \tag{2.29}$$

Now we look at two separate cases.

- For $p_0 = q_0 = 0$, equation (2.28) shows that $r_1 = 0, r_2 = -1$. Then (2.29) implies that $\lim_{n\to\infty} n/w(1/n) > 0$.
- For $q_0 > 0$, equation (2.28) shows that $r_1 > 0$, $r_2 = -1$ for $p_0 = q_0$ and $r_2 < -1$ for $p_0 < q_0$. As $r_1 > 0$, using equation (2.26) and $w_1(0) = 0$ we get that $\beta > 0$. Now (2.29) shows that $\lim_{n\to\infty} n/w(1/n) = 0$ if and only if $p_0 < q_0$.

In both cases, we have that $\lim_{n\to\infty} n/v(n) = \lim_{n\to\infty} n/w(1/n) = 0$ if and only if $p_0 < q_0$.

Now we can conclude with referring to equation (2.24) above. The proof of lemma 2.4.1 is thus completed. □

Theorem 2.4.1. *Assume the setting and the assumptions in section 2.3, then it holds that*

$$\mathbb{E}\left[B_{\tau_J} - \frac{b(B_{\tau_J})}{h(B_{\tau_J})}\right] = B_0 \iff \left(\lim_{x\to\infty} \frac{b(x)}{h(x)x} < 1\right) \text{ and } \left(\lim_{x\to\infty} \frac{h(x)x^2}{\sigma^2(x)} > 0\right). \tag{2.30}$$

Proof. Recall the $(\mathcal{F}_t)_{t\in[0,\infty)}$-stopping times $(\tau_n)_{n\in\mathbb{N}}$ given by $\tau_n = \inf\{t \geqslant 0 \colon B_t \geqslant n\}$ for $n \in \mathbb{N}$ and $\tau = \lim_{n\to\infty} \tau_n$. Let $t \in [0,\infty)$, fix $n \in \mathbb{N} \cap [B_0,\infty)$, and let $T = t \wedge \tau_n$. Then integration by parts and Itô's formula show that

$$\begin{aligned} B_T e^{-\int_0^T h(B_s)ds} = B_0 - \int_0^T B_s h(B_s) e^{-\int_0^s h(B_u)du} ds \\ + \int_0^T e^{-\int_0^s h(B_u)du} b(B_s) ds + \int_0^T e^{-\int_0^s h(B_u)du} \sigma(B_s) dW_s. \end{aligned} \tag{2.31}$$

For fixed n, the integrand in the stochastic integral is bounded. Taking expectations, we get

$$\mathbb{E}\left[B_T e^{-\int_0^T h(B_s)ds}\right] = B_0 - \mathbb{E}\left[\int_0^T \left(1 - \frac{b(B_s)}{h(B_s)B_s}\right) h(B_s) B_s e^{-\int_0^s h(B_u)du} ds\right]. \tag{2.32}$$

Using $\mathbb{P}[\tau_n < \infty] = 1$, bounded convergence and monotone convergence, respectively, we get with $t \to \infty$ that

$$\mathbb{E}\left[B_{\tau_n} e^{-\int_0^{\tau_n} h(B_s)ds}\right] = B_0 - \mathbb{E}\left[\int_0^{\tau_n} \left(1 - \frac{b(B_s)}{h(B_s)B_s}\right) h(B_s) B_s e^{-\int_0^s h(B_u)du} ds\right]. \tag{2.33}$$

Equation (2.33), $B_{\tau_n} = n$ and monotone convergence imply that

$$\begin{aligned} \mathbb{P}\left[\tau_J < \tau\right] &= 1 - \mathbb{E}\left[e^{-\int_0^{\tau-} h(B_s)ds}\right] \\ &= 1 - \mathbb{E}\left[e^{-\int_0^{\tau} h(B_s)ds}\right] \\ &= 1 - \lim_{n\to\infty} \mathbb{E}\left[e^{-\int_0^{\tau_n} h(B_s)ds}\right] \\ &\geqslant 1 - \lim_{n\to\infty} \frac{B_0}{n} \\ &= 1. \end{aligned} \tag{2.34}$$

In particular, equation (2.34) shows that the jump happens with probability 1 on $[[0,\tau))$. Using again monotone convergence and proposition 2.4.1 we let $n \to \infty$ in equation (2.33) to arrive at

$$\lim_{n\to\infty} n\mathbb{E}\left[e^{-\int_0^{\tau_n} h(B_s)ds}\right] = B_0 - \mathbb{E}\left[\left(1 - \frac{b(B_{\tau_J})}{h(B_{\tau_J})B_{\tau_J}}\right) B_{\tau_J}\right]. \tag{2.35}$$

Let $v_n : (0,n) \to [0,1]$ be the function with the property that for $x \in (0,n)$ it holds that

$$v_n(x) = \mathbb{E}\left[e^{-\int_0^{\tau_n} h(B_s)ds} \,\Big|\, B_0 = x \right], \tag{2.36}$$

and let $D_n = (0,n)$ an open domain with boundary $\delta D_n = \{0\} \cup \{n\}$. Note that the setting in section 2.3 ensures that b, σ are continuous and h is Hölder continuous on $[0,n]$. Using the Feynman-Kac-formula for a degenerate second order differential operator on D_n and attainable boundary n,[7] we get that v_n is twice differentiable, satisfies the ordinary differential equation

$$\frac{1}{2}\sigma(x)^2 \frac{\partial^2 v_n}{\partial x^2}(x) + b(x)\frac{\partial v_n}{\partial x}(x) = h(x)v_n(x), \quad x \in (0,n), \tag{2.37}$$

and is uniquely determined by the boundary condition

$$v_n(n) = \mathbb{E}\left[e^{-\int_0^{\tau_n} h(B_s)ds} \,\Big|\, B_0 = n \right] = 1. \tag{2.38}$$

For all $n \in \mathbb{N}$, $x \in (0,n)$, equation (2.33) implies that

$$v_n(x) = \mathbb{E}\left[e^{-\int_0^{\tau_n} h(B_s)ds} \,\Big|\, B_0 = x \right] \leqslant \frac{x}{n} \tag{2.39}$$

and thus $\lim_{n\to\infty} v_n(x) = 0$ for all $x \in (0,\infty)$, which is used in equation (2.44) below. Now, let $v : (0,\infty) \to [0,\infty)$ be the function given by

$$v(x) = \frac{v_n(x)}{v_2(1)\cdots v_n(n-1)} \quad \text{for} \quad x \in (0,n]. \tag{2.40}$$

The boundary condition $v_n(n) = 1$ implies the equality

$$\frac{v_{n+1}(n)}{v_2(1)\cdots v_n(n-1)v_{n+1}(n)} = \frac{v_n(n)}{v_2(1)\cdots v_n(n-1)}, \quad n \in \mathbb{N}, \tag{2.41}$$

and thus, together with uniqueness of v_n, we can conclude that v is well-defined, independent of n. Moreover, v is a twice differentiable function with the property that

$$v(n) > 0 \text{ and } \frac{v(x)}{v(n)} = v_n(x) \quad \text{for all } n \in \mathbb{N}, x \in (0,n]. \tag{2.42}$$

By definition, v satisfies the ordinary differential equation

$$\frac{1}{2}\sigma(x)^2 \frac{\partial^2 v}{\partial x^2}(x) + b(x)\frac{\partial v}{\partial x}(x) = h(x)v(x), \quad x \in (0,\infty), \tag{2.43}$$

and, using equations (2.39) and (2.42) for arbitrary x, the boundary condition

$$\lim_{n\to\infty} v(n) = \lim_{n\to\infty} \frac{v(x)}{v_n(x)} = \infty. \tag{2.44}$$

[7] See, e.g., theorem 1.1 in chapter 13 of [105].

Using equations (2.35), (2.36) and (2.42) we get that

$$B_0 - \mathbb{E}\left[\left(1 - \frac{b(B_{\tau_J})}{h(B_{\tau_J})B_{\tau_J}}\right) B_{\tau_J}\right] = \lim_{n\to\infty} n\mathbb{E}\left[e^{-\int_0^{\tau_n} h(B_s)ds}\right] \\ = v(B_0) \lim_{n\to\infty} \frac{n}{v(n)}. \tag{2.45}$$

Finally, lemma 2.4.1 implies that

$$B_0 - \mathbb{E}\left[\left(1 - \frac{b(B_{\tau_J})}{h(B_{\tau_J})B_{\tau_J}}\right) B_{\tau_J}\right] = 0 \tag{2.46}$$

if and only if $\lim_{x\to\infty} b(x)/(h(x)x) < 1$ and $\lim_{x\to\infty} h(x)x^2/\sigma^2(x) > 0$. The proof of theorem 2.4.1 is thus completed. □

Corollary 2.4.1. *Assume the setting and the assumptions in section 2.3. Then the process S is a uniformly integrable martingale if only if* $\lim_{x\to\infty} b(x)/(h(x)x) < 1$ *and* $\lim_{x\to\infty} h(x)x^2/\sigma^2(x) > 0$.

Proof. The single jump process S is a positive local martingale and, thus, by Doob's martingale convergence theorem, a supermartingale with

$$\mathbb{E}\left[S_\infty\right] \leqslant B_0. \tag{2.47}$$

Thus S is a uniformly integrable martingale if and only if

$$\mathbb{E}\left[\left(1 - \frac{b(B_{\tau_J})}{h(B_{\tau_J})B_{\tau_J}}\right) B_{\tau_J}\right] = B_0.$$

Theorem 2.4.1 completes the proof of corollary 2.4.1. □

The following corollary covers a special case that is very common in the literature, see, e.g., the examples in sections 2.5.1.1 and 2.5.1.2 below.

Corollary 2.4.2. *Assume the setting and the assumptions in section 2.3 and let* $\lim_{x\to\infty} b(x)x/\sigma^2(x) > 0$. *Then the process S is a uniformly integrable martingale if and only if*

$$\lim_{x\to\infty} \frac{b(x)}{h(x)x} < 1. \tag{2.48}$$

Proof. From $\lim_{x\to\infty} b(x)x/\sigma^2(x) > 0$ and $\forall x\colon b(x)/(h(x)x) \in [0,1]$ we know that $\lim_{x\to\infty} h(x)x^2/\sigma^2(x) > 0$. Then the claim follows from corollary 2.4.1. □

2.5 Applications and Discussion

2.5.1 Examples

2.5.1.1 Geometric Brownian motion

First we discuss the situation where the underlying process B is a geometric Brownian motion. In contrast to processes with deterministic jump intensity given by equation (2.5) and the example in section 2.5.1.2 below, this allows to construct a process that is not a uniformly integrable martingale (and thus a mathematical bubble), while the underlying diffusion is not explosive.

To see this, let $\sigma_0, c, \epsilon \in (0,\infty)$, $\mu_0 \in [\sigma_0^2/2, \infty)$ and let $b(x) = \mu_0 x, \sigma(x) = \sigma_0 x$ and

$$h(x) = \left(\mu_0 \left(1 + \frac{c}{\epsilon}\right)\right) \mathbb{1}_{\{x \leqslant \epsilon\}} + \left(\mu_0 \left(1 + \frac{c}{x}\right)\right) \mathbb{1}_{\{\epsilon < x\}} \tag{2.49}$$

for $x \in (0,\infty)$. Then the process B given by equation (2.12) is transient geometric Brownian motion with explosion time $\tau \equiv \infty$. Let us check the assumptions of section 2.3.2. From the discussion on page 197 in [158] we know that $\mathbb{P}[\tau_n < \infty] = 1$ for all $n \geqslant B_0$. Moreover, it holds that

$$\frac{h(x)x^2}{\sigma^2(x)} = \frac{\mu_0}{\sigma_0^2}\left(1 + \frac{c}{x}\right), \quad x \in (\epsilon, \infty)\,. \tag{2.50}$$

Thus assumptions (A) and (B) are satisfied and corollary 2.4.2 shows that the resulting single jump process S is *not* a uniformly integrable martingale. S follows a geometric Brownian motion until the time of the jump that is distributed according to the hazard rate $(h(B_t))_{t\in[0,\infty)}$. Note that for $\mu_0 < \sigma_0^2/2$ it holds that $\mathbb{P}[\tau_n = \infty] > 0$ for $n > B_0$ and assumption (A) is not satisfied.

2.5.1.2 Andersen-Sornette model

In [228] and [8] a model of bubbles has been introduced that is based on super-exponential diffusive growth and a crash represented by a single jump. The process satisfies the above assumptions and can thus be shown to be a mathematical bubble for a suitable jump intensity, highlighting a possible link between the two approaches of

1. bubbles driven by positive feedback mechanisms, superexponential growth and a failure of market efficiency as in [228] and
2. mathematical bubbles as discussed in the introduction and section 2.5.2 below.

A similar link for processes based on deterministic jump intensity has been discussed in [126]. To replicate the setting from [228], assume the setting in section 2.3, let $m \in (1,\infty)$, $\mu_0, \sigma_0 \in (0,\infty)$, let $b\colon [0,\infty) \to [0,\infty)$ and $\sigma\colon [0,\infty) \to [0,\infty)$ be given by $b(x) = (m\sigma_0^2/2)x^{2m-1} + \mu_0 x^m$ and $\sigma(x) = \sigma_0 x^m$ for $x \in [0,\infty)$, let $T_c \in (0,\infty)$. In [228] it has been shown that for $\alpha = \frac{1}{m-1}$ the process B of (2.12) is given by

$$B_t = \alpha^\alpha \frac{1}{(\mu_0(T_c - t) - \sigma_0 W_t)^\alpha}, \quad \text{for } (\omega, t) \in [[0,\tau)), \tag{2.51}$$

with explosion time $\tau : \Omega \to [0, \infty)$ given by $\tau(\omega) := \inf\{t \in (0, \infty) : \mu_0 t + \sigma_0 W_t(\omega) = \mu_0 T_c\}$. As described in [228], the very form of μ and σ can be deduced from the Stratonovich formulation of a non-linear SDE

$$dB_t = \mu_0 B_t^m dt + \sigma_0 B_t^m \circ dW_t, \tag{2.52}$$

where $\circ$ denotes the Stratonovich integral.[8] This has been introduced as a straightforward extension of a non-linear differential equation $dx = x^m dt$ for $m > 1$ as a means of describing self-reinforcing behavior, leading to super-exponential growth. From equation (2.51) we can deduce that assumption (A) is fulfilled. Let $\kappa : [0, \infty) \to [0, 1]$, the *relative jump size*, be any measurable function such that

1. $\lim_{x\to\infty} \kappa(x) = 1$ and
2. $h(x) = \frac{b(x)}{x\kappa(x)}$ is locally Hölder continuous and fulfills assumption (B).

Then corollary 2.4.2 implies that S is *not* a uniformly integrable martingale. In [228] it has been assumed that the relative jump size $\kappa(\cdot) \equiv \kappa \in (0, 1)$ is constant, and corollary 2.4.2 shows that in this case S is a uniformly integrable martingale.

2.5.1.3 Jump (to default) extended CEV model

Extending diffusive models with jumps can enhance their ability to capture defaults, crashes or market anomalies. An example of this is the jump to default extended CEV model, introduced in [44]. Let us discuss a similar jump extended model and its classification as a mathematical bubble based on the results above. For this, assume the setting in section 2.3, let $\mu_0 \in [0, \infty)$, $\sigma_0 \in (0, \infty)$, $\alpha, \beta \in \mathbb{R}$, $b(x) = \mu_0 x^\alpha$ and $\sigma(x) = \sigma_0 x^\beta$. For $\alpha = 1$ the process B (usually defined with absorption at 0) given by equation (2.12) is called the *constant elasticity of variance (CEV)* model, introduced in [60] and [84]. In [44] they allow for the parameter range $\beta \in (-\infty, 1)$ and $\alpha = 2\beta - 1$ and include a possible default (a jump to 0) to arrive at a local martingale in the form of (2.15), which they called *jump to default extended CEV model*. Due to the fixed relative jump size of 1, the resulting process is a non-uniformly integrable martingale. Note that the latter parameter range fails to be included in section 2.3.1, as the resulting diffusion coefficient is not Lipschitz continuous at 0.

Instead consider the parameter range $\beta \in (1, \infty)$, $\alpha \in (1, 2\beta - 1]$. Then example 4.4 in [184] and its preceding remark show that the discounted process

$$B_t e^{-\int_0^{\tau\wedge t} \frac{b(B_s)}{B_s} ds}, \quad t \in [0, \infty) \tag{2.53}$$

is a strict local martingale and thus not a uniformly integrable martingale. Similar to [44], instead of discounting with the drift $\mu_0 x^{\alpha-1}$, let us add a jump to the model with $h(x) = \frac{\mu_0}{\kappa} x^{\alpha-1}$ for some $\kappa \in (0, 1)$. Then corollary 2.4.1 shows that the local martingale S, given by (2.16), is

1. not a uniformly integrable martingale for $\alpha \in (1, 2\beta - 1)$ and
2. a uniformly integrable martingale for $\alpha = 2\beta - 1$ and $\mu_0 \geqslant \sigma_0^2/2$.

[8] Introduced in [236].

Note that assumption (A) for the CEV model with $\alpha = 2\beta - 1 \geqslant 1$ is fulfilled if and only if $\mu_0 \geqslant \sigma_0^2/2$, see theorem 5.1 in [50]. In particular, corollary 2.4.1 does not apply to the case $\mu_0 = 0$ and $\beta = 2$, in which the underlying diffusion $(B_t)_{t\in[0,\infty)}$ is given by the inverse Bessel process

$$B_t = B_0 + \int_0^t B_s^2 dW_s, \quad t \in [0, \infty), \tag{2.54}$$

a classical example of a strict local martingale.

2.5.2 Comments on mathematical bubble models

Based on the seminal paper [179], several authors (see, e.g., [57], [130], [147], [148] and [128]) have contributed towards the attempt to describe financial bubbles through a deviation of a continuous stock price from its fundamental value.

To describe the essential idea of these approaches (see, e.g., [205] for a comprehensive introduction to mathematical bubble modeling), assume that some process S on a probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,\infty)}, \mathbb{P})$ describes a stock market (accompanied by the usual bank account as numéraire) on some time interval $[0, \tau]$, where $\tau : \Omega \to [0, \infty)$ is a $\mathbb{P}$-a.s. finite, but possibly unbounded stopping time. Requiring the absence of arbitrage opportunities,[9] we know that in a complete market there exist a unique equivalent measure $\mathbb{Q} \approx \mathbb{P}$ such that S is a local $\mathbb{Q}$-martingale. Further assuming a complete market and the absence of dividends, one can define the fundamental price of an asset at time 0 as the expected value of the final payoff S_τ, that is, $\mathbb{E}_{\mathbb{Q}}[S_\tau|\mathcal{F}_0]$. Then it becomes clear that corollary 2.4.1 above gives necessary and sufficient conditions to classify single jump processes of the form (2.15) (with the choice $\tau = \tau_J$) as mathematical bubbles.

This reasoning is complicated by the fact that a market generated by single jump processes is, in general, incomplete. Therefore, it is not immediately clear how to define the fundamental value of an asset and there exist several competing approaches in the literature, see, e.g., [148] and [128] or the discussion in section 1.2. For a definite classification as either of the two approaches it will be necessary to examine how the processes studied in this chapter behave under an equivalent change of measure.

Models of mathematical bubbles to date have mostly been considered on a finite time horizon $[0, T]$ for some deterministic $T \in (0, \infty)$. While this is generally rationalized in many financial problems by a finite time investment horizon of the agent, this immediately excludes non-uniformly integrable martingales studied in corollary 2.4.1. If we distinguish, however, *investment horizon* – constraining the trading activity of a market participant – and *model horizon* – the lifetime of a financial asset –, then non-uniformly integrable martingales are readily conceivable as mathematical bubble models.

[9] In the form of *No free lunch with vanishing risk*, developed by Delbaen and Schachermayer in [70] and [71].

2.5.3 Discussion

2.5.3.1 Relaxing assumptions on characteristics

Processes as defined by (2.15) in the setting in section 2.3 are well defined for very general coefficient functions b, σ, h. The local Lipschitz condition on b and σ can be significantly relaxed, see theorem 4.53 in [86] or proposition 2.2 in [50], yielding a so-called *weak solution* B for (2.12). Moreover, the Hölder condition on h can be relaxed to mere measurability, see, e.g., section 6.5 in [23]. The additional assumptions we make are in order to apply the Feynman-Kac formula and analyse the expectation of equation (2.34) in theorem 2.4.1. Very recently there has been an effort by [94, 95] to extend the stochastic representation of Dirichlet boundary problems for a degenerate differential operator to more general[10] diffusion processes. Using these results, one may be able to extend the analysis in this chapter to such processes.

2.5.3.2 Relaxing the assumptions in section 2.3.2

Assumption (B) in section 2.3.2 is necessary to guarantee that the ordinary differential equation appearing in lemma 2.4.1 has at most regular singular points and can be analyzed using Frobenius series. There exist some results on the (asymptotic) behavior around *irregular singular points*, see, e.g., sections 3.4 and 3.5 in [20] for a textbook treatment. However, there is no unified treatment of such equations and an extension may only be fruitful for particular examples, if at all.

Assumption (A) in 2.3.2 ensures that

$$\mathbb{P}\left[\lim_{t\to\infty} B_t = 0\right] = 0 \tag{2.55}$$

for the process B given by equation (2.12). If we drop this assumption, we cannot conclude that the jump happens with probability 1 on $[0, \infty)$ and additional terms appear in the probabilistic representation of equation (2.37),[11] thus leading to a non-trivial extension of the approach in this chapter.

2.5.3.3 Analysis of the martingale property

Let us close with a revisit of the discussion in the introduction. Assume we know that a process S is not a uniformly integrable martingale. Then the obvious next question to ask is whether the process is a strict local martingale or a martingale that is not uniformly integrable. To make this point clear, consider the following simple example. Assume the setting in section (2.3), let $\sigma_0, \mu_0 \in (0, \infty)$ and let $b(x) = \mu_0 x, \sigma(x) = \sigma_0 x$ and $h(x) = \mu_0$ for $x \in (0, \infty)$. Then the process B given by equation (2.12) is a geometric Brownian motion and the process S follows B up to the jump, where it jumps to 0 (we have a *relative jump size* $b(x)/(h(x)x) \equiv 1$). It is clear that S is not

[10] They allowed for σ that is not Lipschitz continuous to cover, e.g., the *Heston stochastic volatility model*, the *SABR model* and the *CEV* model.

[11] For details, see equation (1.17) and theorem 1.1 on page 311 of [105].

a uniformly integrable martingale. However, as h is not state-dependent it holds that

$$\mathbb{E}\left[S_t\right] = \mathbb{E}\left[B_t \mathbb{1}_{\{t<\tau_J\}}\right] = \mathbb{E}\left[B_t\right]\mathbb{E}\left[\mathbb{1}_{\{t<\tau_J\}}\right] = B_0, \tag{2.56}$$

which implies that the supermartingale S is a true martingale. It is crucial to have this simple, state-independent form of h to evaluate the expectation, in general one is confronted with a parabolic problem (cf. chapter 15 in [105]) as opposed to the elliptic problem we encountered in theorem 2.4.1. Thus it is not immediately clear in general whether the process is a true martingale.

Chapter 3

Quadratic hedging during financial bubbles

3.1 Introduction

We study possibilities to price and hedge European options, below referred to as the *option problem*, in a market that experiences a financial bubble. A financial market will be described by a two-dimensional stochastic process $(B_t, S_t)_{t\in[0,T]}$, where $B \equiv 1$ is a bank account and S represents the price process of a financial asset. Trading in this market can be represented by a pair $\left(V_0, (\theta_t)_{t\in[0,T]}\right)$ with initial value $V_0 \in \mathbb{R}$ and predictable asset holdings θ, such that the total value $(V_t)_{t\in[0,T]}$ of the trading strategy is given by

$$V_t = V_0 + \int_0^t \theta_s dS_s, \quad t \in [0,T]. \tag{3.1}$$

Such a strategy is called self-financing, as holdings in the bank account are implied by condition (3.1). For simplicity, we are interested in path-independent European options with a payoff $f(S_T)$ for some suitable payoff function f. As an example, in the well-known Black-Scholes model, where S follows a Brownian motion and generates a complete market, any payoff is perfectly replicable by some self-financing trading strategy (V_0, θ) with

$$f(S_T) = V_0 + \int_0^T \theta_s dS_s. \tag{3.2}$$

By no-arbitrage arguments, this solves the hedging problem with a *fair* option price V_0 and a hedging strategy θ. Instead of a geometric Brownian motion, below we use a single jump diffusion, where the single jump ought to represent a financial crash. This will create an incomplete market, where we can only hope to approximate the final payoff with a strategy (V_0, θ), that is,

$$f(S_T) \approx V_0 + \int_0^T \theta_s dS_s. \tag{3.3}$$

A solution to the option problem is a pair (V_0, θ) that best approximates the final payoff under a suitable error norm.

The study below is based on the Johansen-Ledoit-Sornette (JLS) model of [153, 155], which describes a bubbly price process as the solution $(S_t)_{t\in[0,T]}$ of a stochastic differential equation

$$\frac{dS_t}{S_{t-}} = \kappa h(t)\mathbb{1}_{\{t<\tau_J\}}dt + \sigma dW_t - \kappa dJ_t, \tag{3.4}$$

where W is a real-valued Brownian motion, J is a single jump process that jumps from 0 to 1 at time τ_J with deterministic hazard rate h, and $\kappa \in (0,1)$ is the constant relative jump size. The essential building block of the JLS model is an explosive[1] yet integrable version of the hazard rate h, which is derived based on the analysis of herding behavior and positive feedback in a network of traders. Both herding and positive feedback mechanisms have been argued to contribute to the formation of financial bubbles. We are aware that describing a financial crash with a single jump is rather simplistic, however, can be seen as a good first order approximation, especially when trading halts, short sale restrictions or funding restrictions impede trading during financial drawdowns, cf. the discussion in [53].

It should be stressed that our approach to the option problem is inextricably linked to the understanding that an asset price during a financial bubble can be described (or at least reasonably approximated) by the JLS model in (3.4). There exist other approaches to model financial bubbles, leading to a different formulation of the option problem and applicable solution techniques. In mathematical finance, a popular approach is the so-called *strict local martingale* approach, which models bubbles as processes that show a deviation from their theoretical fundamental value.[2] For such models, there exist a range of results concerned with option pricing, see, e.g., [57, 58, 129, 130, 208]. These results are, besides the *bubble*-terminology, unrelated to the approach taken here.

Much closer in their modeling approach is a series of papers by [17, 18, 19] concerned with a stochastic volatility jump diffusion, where the jump intensity is proportional to the stochastic volatility of the process. Theoretical option prices are determined by using a given investor utility function. Inferring parameter values from market prices of options, they find that jumps have a negative expected value and interpret this possibility of downward jumps as *crash risk*. While [178] discuss an optimal investment problem in a combined portfolio of stock and options, hedging-by-replication of option payoffs has not been dealt with in this model class.

To solve the option problem in our setting, we utilize the framework of quadratic hedging and choose a strategy (V_0, θ) that minimizes the quadratic loss function

$$\mathbb{E}\left[\left(f(S_T) - (V_0 + \int_0^T \theta_t dS_t)\right)^2\right], \tag{3.5}$$

first introduced and studied by [102] and [30]. The price process (3.4) is a martingale by definition, in which case the optimal hedging strategy is equivalent to the so-called risk-minimizing strategy of [102], see [216] or [197] for an introduction to quadratic hedging approaches. While for our

[1] By *explosive*, we mean finite time divergence in the sense that $\lim_{t\to T} h(t) = \infty$.

[2] See [205] for an introduction, [128] for an overview and chapter 1 for a discussion and comparison to bubbles as understood in this thesis.

setting it suffices to start from the ideas in [102], extensions to (3.4) may benefit from [157], who provide semi-explicit results on quadratic hedging ideas for time-inhomogeneous Lévy processes. Unrelated to the notion of financial bubbles, a result that is similar to ours is covered in chapter 7 of [166],[3] where the option problem is solved for a jump diffusion with dynamics

$$\frac{dS_t}{S_{t-}} = \kappa\lambda dt + \sigma dW_t - \kappa dN_t, \tag{3.6}$$

for drift and diffusion coefficients μ, σ and N being a Poisson process with constant intensity λ. Our result is not covered by the solution in [166] due to two important differences of (3.4) as compared to (3.6), which lie at the heart of our understanding of a financial bubble, see chapter 1 for an extended discussion.

1. The jump intensity h in (3.4) is time-dependent and explosive, which leads to (highly) inhomogeneous behavior of pre-crash returns and crash risk.

2. We stop the compensated Poisson process $\kappa\lambda dt - \kappa dN_t$ at its first jump, whence return increments cease to be independent (large, increasing returns – the bubble build up – stopped abruptly at the time of the crash featuring a large negative return).

This chapter is organized as follows. Sections 3.2.1 and 3.2 introduce some relevant notation and the setting used in this chapter, respectively. Section 3.3 studies the application of the quadratic hedging framework to the model in 3.4 and gives explicit formulas for European call options. Section 3.4 applies these results to a specific form of the hazard rate h used in the literature and combines the theoretical hedging results with a fitting method for financial bubbles in a (preliminary) case study on simulated and real data. Finally, section 3.5 concludes and gives an overview of possible extensions and further research.

3.2 Setting

3.2.1 Notation

We use the following notation throughout this chapter. On a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ we consider stochastic processes *unique up to indistinguishability* and require stochastic equations to hold $\mathbb{P}$-a.s. In a slight misuse of notation, for random times $\tau_1, \tau_2 : \Omega \to [0, \infty)$ we say that a stochastic equation holds on $[\tau_1, \tau_2]$ if it holds for all $(\omega, t) \in \Omega \times [0, \infty]$ such that $\tau_1(\omega) \leqslant t \leqslant \tau_2(\omega)$. We denote by $\mathcal{L}^2(\mathcal{F})$ the space of all $\mathcal{F}$-measurable random variables X with $\mathbb{E}\left[X^2\right] < \infty$. For a stochastic process $(X_t)_{t\in[0,T]}$ we denote its left-continuous version by $(X_{t-})_{t\in[0,T]}$, that is, the process with the property that $X_{t-} = \lim_{s\nearrow t} X_s$ for all $t \in [0, T]$. For a semimartingale $(X_t)_{t\in[0,T]}$ we denote by $(\mathcal{E}(X)_t)_{t\in[0,T]}$ the *stochastic exponential* of X, satisfying

$$d\mathcal{E}(X)_t = \mathcal{E}(X)_t dX_t, \quad \mathcal{E}(X)_0 = 1. \tag{3.7}$$

[3]And which, in turn, follows from a special case of the material in section 10.4 of [55].

We denote the distribution function of standard normally distributed random variable with $\Phi(\cdot)$. At several occasions below we use the signifier *Black-Scholes* or simply *BS* to refer to various processes and properties related to the well-known results of [26], who solved the hedging problem for a geometric Brownian motion.

The following definitions and notations are related to the formulation of the hedging problem and are taken from sections 2 and 3 of [197], which we refer to for details. On a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ let $\mathcal{M}^2(\mathbb{P})$ be the space of square integrable $(\mathcal{F}_t)_{t\in[0,T]}$-martingales. For two square integrable martingales $X, Y \in \mathcal{M}^2(\mathbb{P})$ let $(\langle X, Y\rangle_t)_{t\in[0,T]}$, the *predictable covariation*, be the unique predictable process associated to X and Y such that $XY - \langle X, Y\rangle$ is a martingale on $[0, T]$. We use $\langle X\rangle = \langle X, X\rangle$. For a price process $S \in \mathcal{M}^2(\mathbb{P})$, let $\mathcal{L}^2(S)$ be the space of predictable processes $(\theta_t)_{t\in[0,T]}$ such that the stochastic integral

$$\int_0^t \theta_s dS_s, \quad t \in [0, T] \tag{3.8}$$

is in $\mathcal{M}^2(\mathbb{P})$. For a market $(1, S)$, where 1 represents a deterministic bank account, a so-called admissible self-financing trading strategy can be described by a pair $(V_0, \theta) \in \mathbb{R} \times \mathcal{L}^2(M)$. Due to the self-financing property (no cash in- or outflow apart from initial endowment V_0), the associated wealth process $(V_t)_{t\in[0,T]}$ is given by

$$V_t = V_0 + \int_0^t \theta_s dS_s, \quad t \in [0, T]. \tag{3.9}$$

3.2.2 Definitions

Price process. Let $T \in [0, \infty)$, let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ be a filtered probability space with a right-continuous and $\mathbb{P}$-complete filtration, let $(W_t)_{t\in[0,\infty)}$ be a real-valued $(\mathcal{F}_t)_{t\in[0,T]}$-Brownian motion, let $h : [0, T] \to [0, \infty)$ be an integrable function with primitive $\Gamma_\cdot = \int_0^\cdot h(s)ds$, let $J : [0, T] \times \Omega \to \{0, 1\}$ be an RCLL $(\mathcal{F}_t)_{t\in[0,T]}$-adapted stochastic ("single jump") process with the property that for all $t \in [0, T]$ we have

$$\mathbb{P}\left[J_t = 1|\mathcal{F}_{t-}\right] = 1 - e^{-\int_0^t h(s)ds}, \tag{3.10}$$

let $\tau_J : \Omega \to [0, T]$ be the $\mathbb{P}$-a.s. unique $(\mathcal{F}_t)_{t\in[0,T]}$-random time with the property that for all $t \in [0, T]$ it holds that $\mathbb{P}[J_t = 1] = \mathbb{P}\left[\tau_J \leqslant t\right]$, let $S_0 \in (0, \infty)$, $\kappa \in (0, 1)$ and $\sigma \in (0, \infty)$ be constants, let $S : [0, T] \times \Omega \to (0, \infty)$ be the $(\mathcal{F}_t)_{t\in[0,T]}$-adapted RCLL process with the property that

$$S_t = S_0 + \int_0^t S_s\kappa h(s)\mathbb{1}_{\{s<\tau_J\}}ds + \int_0^t S_s\sigma dW_s - \kappa\int_0^t S_{s-}dJ_s, \quad t \in [0, T], \tag{3.11}$$

and let $\tilde{S} : [0, T] \times \Omega \to (0, \infty)$ be the $(\mathcal{F}_t)_{t\in[0,T]}$-adapted RCLL process with the property that

$$\tilde{S}_t = S_0 + \int_0^t \tilde{S}_s\kappa h(s)ds + \int_0^t \tilde{S}_s\sigma dW_s, \quad t \in [0, T]. \tag{3.12}$$

European payoff. Moreover, let $f : [0, \infty] \to [0, \infty]$ be a payoff function of a European option on S with the property that $f(S_T) \in \mathcal{L}^2(\mathcal{F}_T)$, and let $V^f : [0, T] \times \Omega \to [0, \infty)$ be the *value process* associated with f, that is, for all $t \in [0,T]$ it holds that

$$V_t^f = E\left[f(S_T)|\mathcal{F}_t\right]. \tag{3.13}$$

3.2.3 Comments to the setting

An example of a filtration $(\mathcal{F}_t)_{t\in[0,T]}$ in the above setting would be the filtration generated by a Brownian motion W and the single jump process J. Note that in this case, as J jumps at a totally inaccessible stopping time, the Brownian motion remains a Brownian motion under the larger filtration. The process S is a single jump diffusion and has an explicit solution given by

$$S_t = S_0 \exp\left(\int_0^{t\wedge\tau_J} \kappa h(s)ds - \frac{\sigma^2}{2}t + \sigma W_t\right)\left(1 - \kappa \mathbb{1}_{\{\tau_J \leqslant t\}}\right), \quad t \in [0,T], \tag{3.14}$$

while $\tilde{S}$ denotes its *crash-free* analogue

$$S_t = S_0 \exp\left(\int_0^{t} \kappa h(s)ds - \frac{\sigma^2}{2}t + \sigma W_t\right), \quad t \in [0,T], \tag{3.15}$$

which is well-defined due to integrability of h and used below to simplify derivations and calculations.

3.3 Hedging in a single jump framework

3.3.1 General results

The following lemma examines the structure of the value process in a our single jump setting and follows immediately from the properties of a random time with deterministic hazard rate, see, e.g., corollary 5.1 in [23].

Lemma 3.3.1. *Assume the setting in section 3.2. Then it holds that*

$$\begin{aligned} V_t^f = {} & \mathbb{1}_{\{\tau_J \leqslant t\}} E\left[f\left((1-\kappa)\tilde{S}_{\tau_J}\mathcal{E}\left(\sigma(W_T - W_{\tau_J})\right)\right)|\mathcal{F}_t\right] \\ & + \mathbb{1}_{\{t<\tau_J\}} \int_t^T \mathbb{E}\left[f\left((1-\kappa)\tilde{S}_u\mathcal{E}\left(\sigma(W_T - W_u)\right)\right)|\mathcal{F}_t\right] e^{\Gamma_t - \Gamma_u} h(u)du \\ & + \mathbb{1}_{\{t<\tau_J\}} \mathbb{E}\left[f(\tilde{S}_T)|\mathcal{F}_t\right] e^{\Gamma_t - \Gamma_T}. \end{aligned} \tag{3.16}$$

The next result concerns an optimal hedging strategy in a market $(1, S)$. The problem formulation (3.17) has been introduced in [102].

Proposition 3.3.1. *Assume the setting in section 3.2. Then the process S is a square integrable martingale and the quadratic hedging problem*

$$\min_{(V_0,\theta)\in\mathbb{R}\times\mathcal{L}^2(S)} \mathbb{E}\left[\left(f(S_T) - (V_0 + \int_0^T \theta_t dS_t)\right)^2\right] \tag{3.17}$$

has a minimum at (V_0^*, θ^*) *given by* $V_0^* = V_0^f$ *and a strategy* θ^* *which solves*

$$\langle V^f, S\rangle_t = \int_0^t \theta_s^* d\langle S, S\rangle_s, \quad t \in [0, T] \tag{3.18}$$

A proof of proposition 3.3.1 is given in section 3.6. Due to linearity of (3.18), we can expect the optimal hedge to be

- a version of the optimal hedging strategy for a jump diffusion (3.6) up to the crash time τ_J and
- equal to the Black-Scholes delta after the crash.

The following result confirms this intuition and will be the basis for more explicit formulas below.

Theorem 3.3.1. *Assume the setting in section 3.2 and assume there exists a real-valued function* $F \in \mathcal{C}^{1,2}([0, T] \times [0, \infty))$ *such that the pre-crash value process can be written as*

$$V_t^f \mathbb{1}_{\{t<\tau_J\}} = F(t, S_t)\mathbb{1}_{\{t<\tau_J\}} \tag{3.19}$$

Then the optimal hedging strategy is given by

$$\theta_t^* = \left(\frac{\partial F}{\partial S}(t, S_t) + \frac{\kappa\kappa_t^f h(t)}{S_t(\sigma^2 + \kappa^2 h(t))}\right)\mathbb{1}_{\{t<\tau_J\}} + \frac{\partial V^f}{\partial S}(t, S_t)\mathbb{1}_{\{\tau_J \leqslant t\}}, \tag{3.20}$$

where the process $\kappa^f : [0, T] \times \Omega \to \mathbb{R}$ *is given by*

$$\kappa_t^f = -\kappa S_t \frac{\partial F}{\partial S}(t, S_t) - E\left[f\left((1-\kappa)S_t\mathcal{E}\left(\sigma(W_T - W_t)\right)\right)|\mathcal{F}_t\right] + F(t, S_t), \quad t \in [0, T]. \tag{3.21}$$

A proof of theorem 3.3.1 is given in section 3.6.

3.3.2 European Call option

Corollary 3.3.1. *Assume the setting in section 3.2 and assume that* $f(x) = (x - K)^+$ *is the payoff function of a European call with strike* $K \in (0, \infty)$. *Then, with*

$$\begin{aligned} d_{1,2}(u) &= \frac{\kappa(\Gamma_u - \Gamma_t) \pm \frac{\sigma^2}{2}(T-t) + \ln\frac{(1-\kappa)S_t}{K}}{\sigma\sqrt{T-t}}, \\ \hat{d}_{1,2} &= \frac{\kappa(\Gamma_T - \Gamma_t) \pm \frac{\sigma^2}{2}(T-t) + \ln\frac{S_t}{K}}{\sigma\sqrt{T-t}}, \\ \bar{d}_{1,2} &= \frac{\pm\frac{\sigma^2}{2}(T-t) + \ln\frac{(1-\kappa)\tilde{S}_t}{K}}{\sigma\sqrt{T-t}}, \end{aligned} \tag{3.22}$$

and

$$\begin{aligned}\phi_t &= (1-\kappa)\int_t^T e^{(1-\kappa)(\Gamma_t-\Gamma_u)}\Phi(d_1(u))h(u)du + e^{(1-\kappa)(\Gamma_t-\Gamma_T)}\Phi(\hat{d}_1),\\ \psi_t &= \int_t^T e^{\Gamma_t-\Gamma_u}\Phi(d_2(u))h(u)du + e^{\Gamma_t-\Gamma_T}\Phi(\hat{d}_2),\end{aligned} \tag{3.23}$$

the value process of the call is given by

$$V_t^f = \mathbb{1}_{\{t<\tau_J\}}\left[S_t\phi_t - K\psi_t\right] + \mathbb{1}_{\{\tau_J \leqslant t\}}\left[S_t\Phi(\bar{d}_1) - K\Phi(\bar{d}_2)\right] \tag{3.24}$$

A proof of corollary 3.3.1 is given in section 3.6.

Corollary 3.3.2. *Assume the setting in section 3.2 and assume that $f(x) = (x-K)^+$ is the payoff function of a European call with strike $K \in (0,\infty)$. Then, using the notation in (3.22), (3.23) and*

$$\kappa_t^f = (1-\kappa)S_t\left(\phi_t - \Phi\left(\bar{d}_1\right)\right) - K\left(\psi_t - \Phi\left(\bar{d}_2\right)\right), \tag{3.25}$$

the optimal hedging strategy can be expressed as

$$\theta^* = \left(\phi_t + \frac{\kappa\kappa_t^f h(t)}{S_t(\sigma^2+\kappa^2 h(t))}\right)\mathbb{1}_{\{t<\tau_J\}} + \Phi(\bar{d}_1)\mathbb{1}_{\{\tau_J \leqslant t\}}. \tag{3.26}$$

A proof of corollary 3.3.2 is given in section 3.6.

3.4 Hedging in the Johansen-Ledoit-Sornette model

Below we apply the results from section 3.3 to a model introduced in [153, 154, 155], which is known as the JLS (Johansen-Ledoit-Sornette) or LPPLS (Log-Periodic-Power-Law-Singularity) model. In this model, price dynamics are specified up to a critical point T_c. To embed the JLS model in the hedging problem of section 3.2 with given maturity T, in the case $T_c < T$ we assume that the price follows a geometric Brownian motion on $(T_c, T]$, cf. the model setting in [127]. The next section reviews the existing methodology of describing the JLS model.

3.4.1 Setting

Assume the setting in section 3.2, let $T_c \in (0,\infty)$, $B' \in (0,\infty)$, $C' \in (-B', B')$, $\alpha \in (-1,0)$, $\omega \in (0,\infty)$, $\rho' \in \mathbb{R}$ and assume that

$$h(t) = \left[B'(T_c-t)^\alpha + C'(T_c-t)^\alpha \cos\left(\omega\ln(T_c-t)-\rho'\right)\right]\mathbb{1}_{\{t<T_c\}}, \quad t\in[0,T]. \tag{3.27}$$

3.4.2 Comments to the setting

The choice of parameters ensures that h is non-negative, see the discussion in [243]. The associated hazard process $\Gamma = \int_0^{\cdot} h(s)ds$ is given by

$$\Gamma_t = [B(T_c - t)^m + C(T_c - t)^m \cos(\omega \ln(T_c - t) - \rho)] \mathbb{1}_{\{t \leqslant T_c\}}, \quad t \in [0, T]. \tag{3.28}$$

for parameters $m = 1 + \alpha \in (0, 1)$, $\rho \in \mathbb{R}$,

$$B = \frac{-\kappa B'}{m} \in (-\infty, 0) \text{ and } \quad C = \frac{-\kappa C'}{\sqrt{m^2 + \omega^2}} \in \left(\frac{Bm}{\sqrt{m^2 + \omega^2}}, \frac{-Bm}{\sqrt{m^2 + \omega^2}}\right). \tag{3.29}$$

In the literature, equation (3.28) is the most common way to describe the JLS model. The resulting stock price follows equation (3.11), which, in differential notation, reads

$$\frac{dS_t}{S_{t-}} = \kappa h(t) \mathbb{1}_{\{t < \tau_J\}} dt + \sigma dW_t - \kappa dJ_t, \quad t \in [0, T]. \tag{3.30}$$

In particular, before the crash, the log-price follows

$$\ln(S_t) = A + \Gamma_t - \frac{\sigma^2}{2} t + W_t, \quad t \in [0, \tau_J), \tag{3.31}$$

for a constant $A = \ln(S_0) - \Gamma_0$. Equation (3.31) is the cornerstone of most JLS fitting procedures, where the Itô term $-(\sigma^2/2)t$ is generally ignored.

3.4.3 Performance on generic data

To confirm the feasibility of the quadratic hedging strategy in (3.26), we test its performance on simulated data. To this end, we choose some arbitrary time horizon $T = 200$ and look at a European call option with maturity T and payoff function $f(x) = (x - K)^+$ for varying strike levels $K \in (0, \infty)$.

Data generation. We simulate a set of daily price time series according to (3.31) with parameters given by

$$\begin{aligned} &\kappa = 0.25, \quad \sigma = 0.0086, \quad A = 6.0763 \quad B = -0.0253, \\ &C = 0.0015, \quad m = 0.5099154 \quad \omega = 8.4173 \quad \rho = 3.01937, \end{aligned} \tag{3.32}$$

and time of explosion $T_c = T = 200$, which we set for simplicity to be equal to option maturity. For each price time series, the time of the crash is simulated in $[0, T_c]$ with hazard rate (3.27). Note that there is a non-zero probability that the crash does not happen. Figure 3.1 shows sample paths of S along with

$$\exp\left(A + [B(T_c - t)^m + C(T_c - t)^m \cos(\omega \ln(T_c - t) - \rho)]\right), \quad t \in [0, T_c], \tag{3.33}$$

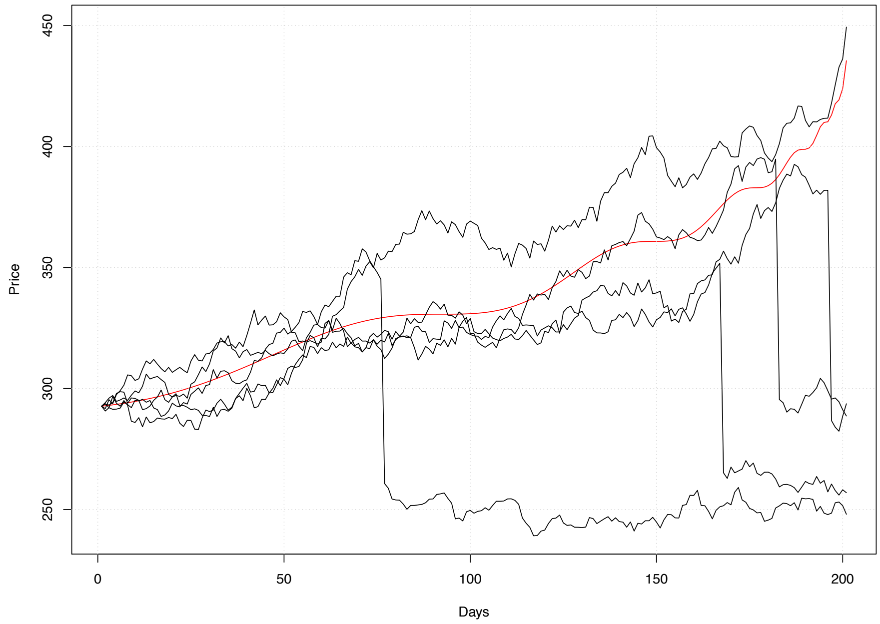


Figure 3.1: Simulated sample paths based on the specification (3.31) with parameter values in (3.32). The red (smooth) line shows deterministic pre-crash returns.

the deterministic compensator of the crash. For simplicity, we assume that all parameters are known. As we see below in section 3.4.4, all parameters can be inferred from the pre-crash stock price data, except for κ.

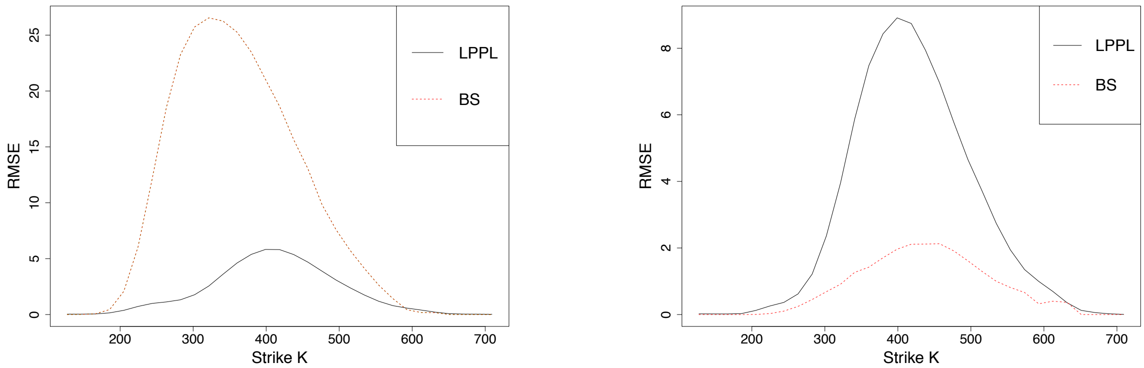
Mean squared hedging error. For simulated sample paths $(S^{(i)})_{i\in\{0,\dots,M\}}$ and a given (discretized) daily trading rule $(V_0,(\theta_i)_{i\in\{0,\dots,T-1\}})$, we approximate the expected squared hedging error

$$\mathbb{E}\left[\left(f(S_T)-(V_0+\int_0^T\theta_t dS_t)\right)^2\right]\approx\frac{1}{M}\sum_{i=1}^{M}\left(f\left(S_T^{(i)}\right)-\left(V_0+\sum_{t=0}^{T-1}\theta_t\left(S_{t+1}^{(i)}-S_t^{(i)}\right)\right)\right)^2 \tag{3.34}$$

for both the JLS strategy (V_0^*,θ^*) given by Corollaries 3.3.1 and 3.3.2 and the Black-Scholes hedging strategy $(V_0^{\text{BS}},\theta^{\text{BS}})$. Figure 3.2a shows these errors for a wide range of possible strike values $K\in[50,500]$ and $M=15000$ sample paths.

Note that the right-hand side of (3.34)

- includes a discretization error along with the replication error on the left-hand side of (3.34), and
- an additional model error for the Black-Scholes strategy, which is based on the assumption that S follows a geometric Brownian motion (with square integrable drift).



(a) Generated with $M = 15000$ simulated price time series.

(b) Generated with a subset of $M = 3033$ crash-free price time series. Black-Scholes shows no model error in this case and consists of discretization error only.

Figure 3.2: The square root of the mean square hedging error (3.34) for JLS and Black-Scholes hedging specifications, respectively.

As an (approximate) upper bound on the discretization error for θ^{BS}, figure 3.2b shows the hedging error (3.34) conditional on no crash happening, in which case there should be no replication error for the Black-Scholes strategy.

Some comments on hedging behavior. Let us take a crude qualitative look at the behavior of the optimal strategies (V_0^*, θ^*) given by corollaries 3.3.1 and 3.3.2 and the Black-Scholes hedging strategy $(V_0^{\mathrm{BS}}, \theta^{\mathrm{BS}})$ for a single price time series. Below we differentiate between the (theoretical) option value process given by (3.24) for the JLS model and

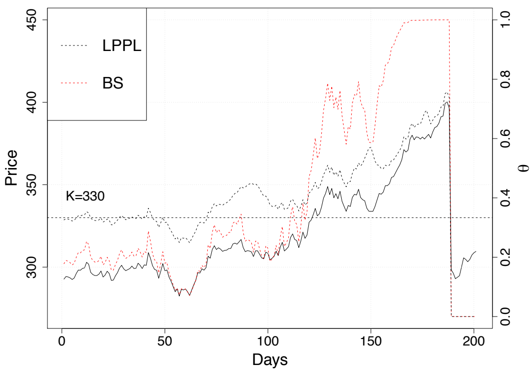
$$S_t \Phi\left(\frac{\frac{\sigma^2}{2}(T-t) + \ln\frac{S_t}{K}}{\sigma\sqrt{T-t}}\right) - K\Phi\left(\frac{-\frac{\sigma^2}{2}(T-t) + \ln\frac{S_t}{K}}{\sigma\sqrt{T-t}}\right), \quad t \in [0,T], \tag{3.35}$$

for the Black-Scholes model, and the value process of a replicating strategy given by

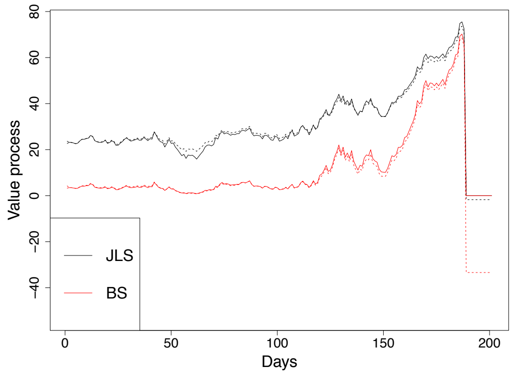
$$V_0 + \sum_{s=0}^{\lfloor t \rfloor - 1} \theta_s \left(S_{s+1} - S_s\right), \quad t \in [0,T]. \tag{3.36}$$

Referring to figures 3.3a and 3.3b, we can note the following qualitative behavior, consistent with known properties of the so-called Black-Scholes Greeks.

- For out-of-the-money options, the increased variability in the stock price leads in general to $\theta^* > \theta^{\mathrm{BS}}$. Similarly, for in-the-money options, in general we see $\theta^* < \theta^{\mathrm{BS}}$.
- Increased variability in the stock price and convexity of the payoff function implies that the JLS value process (3.24) is an upper bound of the BS value process (3.35), in particular we have $V_0 > V_0^{\mathrm{BS}}$ and options are priced higher in the JLS model.



(a) A single price time series (solid line - left axis) simulated from (3.31) along with Black-Scholes-Delta θ^{BS} and JLS optimal hedging strategy θ^* (dashed line- right axis) as in corollary 3.3.2.



(b) Theoretical option value (solid lines - see (3.24) and (3.35)) and value process of the hedging strategies (dashed lines - see (3.36)).

Figure 3.3: Hedging a call option on a simulated price time series for JLS and Black-Scholes hedging specifications, respectively.

- Figure 3.3a shows that, before the crash, the Black-Scholes strategy closely follows its theoretical value process. On the other hand, the JLS strategy "insures" against the crash by holding more cash relative to Black-Scholes, which results in a smaller loss when the crash finally happens.

These effects depends on the values of κ and h, which introduce non-diffusive randomness into the model through adding the risk of a single crash. As the bullet points above explain, the JLS hedging strategy accounts for this increased randomness. A statistical analysis using a jump-diffusion setting while avoiding the model assumptions in section 3.4.1 will not allow for that, as the crash is the first and only jump in our methodology.

3.4.4 Performance during historic bubble episodes

Here we employ a fitting method from [75] to dynamically infer the parameters in (3.27) from a given price time series and derive a quadratic hedging strategy. In particular, we look at the S&P500 index price process during the 1987 crash episode, see figure 3.4. We study the performance of hedging a European call, choose a maturity date T =1987/12/01 and set the origin (start of hedging) at t_0 =1986/07/01. For every $t \in [t_0, T]$, we fit the price process $(S_s)_{s<t}$ as in [75] based on equation (3.31), and calculate the optimal hedge θ_t^* using (3.26). To this end, additional to the parameters $(B', C', \alpha, \omega, \rho')$, we need to set an (expected) crash size $\kappa^* \in (0, 1)$ a-priori and decide dynamically when the crash has happened, that is, for which t it holds that $t \geqslant \tau_J$. This leads to two main complications that are dealt with as follows.

1. A financial crash does in general not show as a single jump, but unfolds over several days or even months. Two immediate (naive) approaches to deal with this come to mind.[4]

[4]As noted in the introduction, short sale restrictions, trading halts or borrowing issues may in some cases exogenously

Figure 3.4: Price time series of the S&P500 Index before and during the 1987 *Black Monday* crash episode.

(a) We halt rehedging after a critical relative loss $\overline{\kappa} \in (0,1)$ has materialized. To be precise, we set

$$\tau_J = \inf\left\{ t\colon 1 - \frac{S_t}{\sup_{s \leqslant t} S_t} \geqslant \overline{\kappa} \right\} \tag{3.37}$$

and do not hedge on $[\tau_J, T]$. Below we use $\overline{\kappa} = 0.1$, which yields an a-posteriori crash size

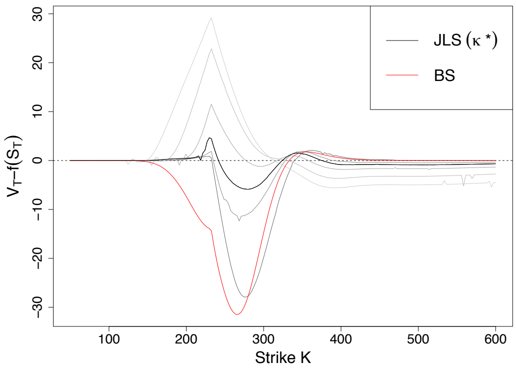
$$\kappa^* = 1 - \frac{S_T}{S_{\tau_J - 1}} = 24\%. \tag{3.38}$$

(b) We ignore a materializing drawdown and assume $t < \tau_J$ in the optimal strategy (3.26), for all $t \in [0, T]$. In a sense, we only hedge against daily drawdowns, and assume a decent hedging performance for

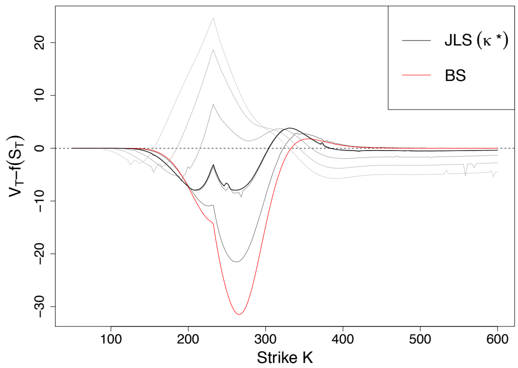
$$\kappa^* = \sup_{t \in [0, T-1]} \left(1 - \frac{S_t}{S_{t+1}} \right) = 20.4\%. \tag{3.39}$$

Both approaches violate the original model assumption, do not incorporate all available information and can likely be improved upon.

determine a crash period that is, in effect, a single jump.



(a) Hedging error (3.40) of final value of the hedging strategy in 1a above, with $\bar{\kappa} = 0.1$.



(b) Hedging error (3.40) of final value of the hedging strategy in 1b above.

Figure 3.5: Hedging a call option during the 1987 S&P500 bubble episode for JLS and Black-Scholes hedging specifications. Grey lines show results of hedging strategy for $\kappa \in (0.1, 0.2, 0.3, 0.4, 0.5)$ in decreasing intensity (0.1 in dark gray – 0.5 in light grey).

2. The size of the drawdown, κ^* as given in the two latter approaches, is not known a-priori. We do not deal with this issue directly but report performance of hedging strategies for varying values of κ to show robustness against mis-specification.

Figures 3.5a and 3.5b show the quantity

$$V_0 + \int_0^T \theta_s dS_s - f(S_T) \tag{3.40}$$

for a range of possible strike values $K \in [50, 600]$. For both specifications mentioned in 1 above, we see that the quadratic hedge does a good job in replicating the option value across strike ranges, provided the a-priori chosen crash size is not too far from κ^*.

3.5 Discussion and further research

Above we have shown that the quadratic hedging approach for the martingale process

$$\frac{dS_t}{S_{t-}} = \kappa h(t) \mathbb{1}_{\{t<\tau_J\}} dt + \sigma dW_t - \kappa dJ_t, \quad t \in [0, T], \tag{3.41}$$

leads to explicit results in corollaries 3.3.1 and 3.3.2, defining a strategy (V_0^*, θ^*) that optimally hedges the payoff of a European call option. Using simulated data and perfect parameter knowledge, we quantify the edge over a simple Black-Scholes strategy in section 3.4.3. Applying the result to a historic bubble episode (the 1987 S&P500 bubble and crash), in section 3.4.4 we discussed two of our simplified model assumptions.

Besides an obvious extension to European puts and other option types, there are several other assumptions that can be dropped and interesting research questions that can be dealt with, loosely collected below.

1. Proposing equation (3.41) together with a bank account $B \equiv 1$, we have implicitly assumed a risk-free rate $r \equiv 0$. To account for, e.g., some constant risk free rate $r \in \mathbb{R}$, we can start with a price process $\tilde{S}$ and a bank account $\tilde{B}_t = e^{rt}$, and assume that equation (3.41) describes the discounted price process $\tilde{S}/\tilde{B}$, cf. the elaborations in chapter 7 of [166], which can be adapted to our setting.

2. More generally, we can drop the martingale assumption altogether and assume the (discounted) asset price follows

$$\frac{dS_t}{S_{t-}} = (\mu(t) + \kappa h(t)) \mathbb{1}_{\{t<\tau_J\}} dt + \sigma dW_t - \kappa dJ_t, \quad t \in [0,T], \tag{3.42}$$

for a suitable function $\mu : [0,T] \to \mathbb{R}$. As elaborated in [216], quadratic hedging in this case becomes more involved and there exist different approaches. If, instead of (3.17), we look for a so-called risk-minimizing strategy, then in case we have $\kappa\mu(t) > -(\sigma^2 + \kappa^2 h(t))$ for all $t \in [0,T]$, this can be solved using the *minimal martingale measure*, see [215]. In particular, for $\gamma(t) = (\kappa\mu(t))/(\sigma^2 + \kappa^2 h(t)) > -1$ the minimal martingale measure $\mathbb{Q} \sim \mathbb{P}$ is given by a density $(Z_t^{\mathbb{Q}})_{t\in[0,T]}$ with

$$Z_t^{\mathbb{Q}} = \mathcal{E}\left(\int_0^{\cdot} \kappa h(s)\gamma(s)\sigma dW_s + \int_0^{\cdot} \gamma_s \left(dJ_s - h(s)ds\right) \right)_t, \quad t \in [0,T], \tag{3.43}$$

and the stock price can be written as

$$\frac{dS_t}{S_{t-}} = \kappa h^{\mathbb{Q}}(t) \mathbb{1}_{\{t<\tau_J\}} dt + \sigma dW_t - \kappa dJ_t, \quad t \in [0,T], \tag{3.44}$$

for $h^{\mathbb{Q}}(t) = (1+\gamma(t))h(t)$, the $\mathbb{Q}$-hazard rate of J. Then the results in [215] imply that the risk-minimizing strategy (V_0^*, θ^*) is given by

$$\begin{aligned} V_0^* &= \mathbb{E}_{\mathbb{Q}}\left[f(S_T)|\mathcal{F}_t\right] \\ \theta_t^* &= \frac{d\langle V^f, S\rangle_t^{\mathbb{P}}}{d\langle S, S\rangle_t^{\mathbb{P}}}, \quad t \in [0,T]. \end{aligned} \tag{3.45}$$

It is rather straightforward to check that corollary 3.3.1 extends to this case with $h^{\mathbb{Q}}$ substituted for h. Based on this updated result, corollary 3.3.2 extends unchanged.

While this flexible model description may allow for a better fit to option market data, for (reasonable) values of constant excess drift $\mu(t) \equiv \mu \in (0,\infty)$ we did not see a significant change in the optimal strategy (V_0^*, θ^*).

3. Another possible step towards more flexible model dynamics is to simply substitute your favourite stochastic volatility process $(\sigma_t)_{t\in[0,T]}$ for the constant σ in (3.41). An analogy of proposition 3.3.1 for this extension can be found in [157]. However, this will in general not yield an explicit solution as in corollaries 3.3.1 and 3.3.2.

A much more straightforward extension is to allow for a (suitably well-behaved) stochastic process $(\kappa_t)_{t\in[0,T]}$ as model for the relative crash size. This can both improve fit to empirical option data as well as incorporate statistical knowledge on the crash size (which is, a-priori, inherently unquantifiable in our model, see the discussion in section 3.4.4 above).

4. As already suggested above, relative pricing approaches are essentially a two-way street, and can be utilized as in section 3.4.4, deriving parameter values from historical stock prices to calculate option prices; or by inferring parameter values from option price data. In our setting, a combined approach may be utilized as follows. As a first step, using common JLS fitting methods to infer parameter values based on (3.31), followed by a fit to option market data to infer $(\kappa_t)_{t\in[0,T]}$ and a possible excess drift μ as in (3.42). This can be used to derive interesting tests on joint hypotheses regarding the martingale assumption of the JLS model and the crash awareness of the market.

3.6 Proofs

Proof of proposition 3.3.1. Theorem 3.6 (a) in [127] shows that S is $\mathbb{P}$-martingale. Moreover, it holds that

$$\mathbb{E}\left[S_t^2\right] \leqslant \mathbb{E}\left[\tilde{S}_t^2\right] = S_0^2 e^{2\left(\int_0^t \kappa h(s)ds\right)+\sigma^2 t}, \tag{3.46}$$

and square integrability of S follows from integrability of h. Finally, the Galtchouk-Kunita-Watanabe decomposition of the payoff $f(S_T) \in \mathcal{L}^2(\mathcal{F}_T)$ implies that the self-financing strategy (V_0^*, θ^*) with $V_0^* = \mathbb{E}[f(S_T)]$ and

$$\theta_t^* = \frac{d\langle V^f, S\rangle_t}{d\langle S, S\rangle_t} \tag{3.47}$$

is the unique minimizer of (3.17), which completes the proof of proposition 3.3.1. □

Proof of theorem 3.3.1. In view of proposition 3.3.1, we need to calculate

$$\theta_t^* = \frac{d\langle V^f, S\rangle_t}{d\langle S\rangle_t}, \tag{3.48}$$

where $V^f = \mathbb{E}[f(S_T)|\mathcal{F}_\cdot]$ is the value process associated to the payoff function f. On $[\tau_J, T]$, the stock price S follows a geometric Brownian motion, and equation (3.48) shows, by standard Black-Scholes arguments, that

$$\theta_t^* = \frac{\partial V^f}{\partial S}(t, S_t), \quad \text{on } [\tau_J, T]. \tag{3.49}$$

To derive the optimal hedging strategy on $[0, \tau_J)$, we apply Itô's formula to $F(t, \tilde{S}_t)$,[5] which gives

$$dF = \left(\frac{\partial F}{\partial t} + \frac{\sigma^2}{2}\frac{\partial^2 F}{\partial^2 \tilde{S}}\right) dt + \frac{\partial F}{\partial \tilde{S}} d\tilde{S} \tag{3.50}$$

[5] Alternatively, we could apply Itô's formula to $F(t, S_t)$ for the discontinuous process S. We chose the present derivation, which utilizes the single jump nature of S more explicitly.

For ΔV^f given by

$$\Delta V_t^f = \mathbb{E}\left[f\left((1-\kappa)\tilde{S}_{\tau_J\wedge t}\mathcal{E}\left(\sigma(W_T - W_{\tau_J\wedge t})\right)\right)|\mathcal{F}_t\right] - F(t,\tilde{S}_t), \quad t\in[0,T], \tag{3.51}$$

the (predictable) jump size of the value process, let us define

$$\kappa_t^f = -\kappa\tilde{S}_t\frac{\partial F}{\partial\tilde{S}}(t,\tilde{S}_t) - \Delta V_t^f, \quad t\in[0,T]. \tag{3.52}$$

By assumption on the function F, on $[0,\tau_J)$ it holds that $V_t^f = F(t,\tilde{S}_t)$. Thus, omitting below the argument $(t,\tilde{S}_t)$ of F, we get the dynamics

$$\begin{aligned}
\mathbb{1}_{\{t<\tau_J\}}dV_t^f &= \mathbb{1}_{\{t<\tau_J\}}dF + \Delta V_t^f dJ_t \\
&= \left(\frac{\partial F}{\partial t} + \frac{\sigma^2}{2}\frac{\partial^2 F}{\partial^2\tilde{S}}\right)\mathbb{1}_{\{t<\tau_J\}}dt + \frac{\partial F}{\partial\tilde{S}}\mathbb{1}_{\{t<\tau_J\}}d\tilde{S} + \Delta V_t^f dJ_t \\
&= \left(\frac{\partial F}{\partial t} + \frac{\sigma^2}{2}\frac{\partial^2 F}{\partial^2\tilde{S}}\right)\mathbb{1}_{\{t<\tau_J\}}dt + \frac{\partial F}{\partial\tilde{S}}\mathbb{1}_{\{t<\tau_J\}}dS + (\kappa\tilde{S}_t\frac{\partial F}{\partial\tilde{S}} + \Delta V_t^f)dJ_t \\
&= \left(\frac{\partial F}{\partial t} + \frac{\sigma^2}{2}\frac{\partial^2 F}{\partial^2\tilde{S}}\right)\mathbb{1}_{\{t<\tau_J\}}dt + \frac{\partial F}{\partial\tilde{S}}\mathbb{1}_{\{t<\tau_J\}}dS - \kappa_t^f dJ_t.
\end{aligned} \tag{3.53}$$

As V^f and S (stopped at τ_J) are $\mathbb{P}$-martingales, so is the process $(M_t)_{t\in[0,T]}$ given by

$$M_t = \int_0^t\left[\left(\frac{\partial F}{\partial t} + \frac{\sigma^2}{2}\frac{\partial^2 F}{\partial^2\tilde{S}}\right)\mathbb{1}_{\{t<\tau_J\}}ds - \kappa_s^f dJ_s\right], \quad t\in[0,T], \tag{3.54}$$

which is the case only if for all $t\in[0,\tau_J)$ we have

$$\left(\frac{\partial F}{\partial t} + \frac{\sigma^2}{2}\frac{\partial^2 F}{\partial^2\tilde{S}}\right) = \kappa_t^f h(t). \tag{3.55}$$

Combining (3.53) and (3.55) implies for all $t\in[0,\tau_J]$ that

$$V_t^f = \int_0^t\frac{\partial F}{\partial\tilde{S}}dS_s + \int_0^t\kappa_s^f\left(h(s)ds - dJ_s\right). \tag{3.56}$$

Using equations (3.11) and (3.56) we get for all $t\in[0,\tau_J]$ that

$$\begin{aligned}
\langle V^H, S\rangle_t &= \langle\int_0^\cdot\frac{\partial F}{\partial\tilde{S}}dS_s, S\rangle_t + \langle\int_0^\cdot\kappa_s^f\left(h(s)ds - dJ_s\right), S\rangle_t \\
&= \int_0^t\frac{\partial F}{\partial\tilde{S}}d\langle S\rangle_s + \int_0^t\kappa\kappa_s^f S_s d\langle\int_0^\cdot h(u)du - dJ_u\rangle_s \\
&= \int_0^t\frac{\partial F}{\partial\tilde{S}}d\langle S\rangle_s + \int_0^t\kappa\kappa_s^f S_s h(s)ds.
\end{aligned} \tag{3.57}$$

and equation (3.48) implies

$$\begin{aligned}\theta_t^* = \frac{d\langle V^H, S\rangle_t}{d\langle S\rangle_t} &= \frac{\frac{\partial F}{\partial \tilde{S}} d\langle S\rangle_t + \kappa\kappa_t^f S_t h(t)dt}{d\langle S\rangle_t} \\ &= \frac{\partial F}{\partial \tilde{S}} + \frac{\kappa\kappa_t^f h(t)}{S_t\left(\sigma^2 + \kappa^2 h(t)\right)}, \quad t \in [0, \tau_J).\end{aligned} \tag{3.58}$$

Together with (3.49) this shows

$$\theta_t^* = \left(\frac{\partial F}{\partial S}(t, S_t) + \frac{\kappa\kappa_t^f h(t)}{S_t(\sigma^2 + \kappa^2 h(t))}\right)\mathbb{1}_{\{t<\tau_J\}} + \frac{\partial V^f}{\partial S}(t, S_t)\mathbb{1}_{\{\tau_J \leqslant t\}}, \quad t \in [0, T], \tag{3.59}$$

and we can conclude with noting that, on $[0, \tau_J)$, we have

$$\begin{aligned}\kappa_t^f &= -\kappa\tilde{S}_t\frac{\partial F}{\partial \tilde{S}}(t, \tilde{S}_t) - \mathbb{E}\left[f\left((1-\kappa)\tilde{S}_{\tau_J \wedge t}\mathcal{E}\left(\sigma(W_T - W_{\tau_J\wedge t})\right)\right)|\mathcal{F}_t\right] + F(t, \tilde{S}_t) \\ &= -\kappa S_t\frac{\partial F}{\partial S}(t, S_t) - \mathbb{E}\left[f\left((1-\kappa)S_t\mathcal{E}\left(\sigma(W_T - W_t)\right)\right)|\mathcal{F}_t\right] + F(t, S_t).\end{aligned} \tag{3.60}$$

The proof of theorem 3.3.1 is thus completed. □

Proof of corollary 3.3.1. We use the following fact that for a lognormally distributed random variable $X \sim \log\mathcal{N}(\mu, \sigma^2)$ it holds that

$$\mathbb{E}\left[(X-K)^+\right] = e^{\mu+\frac{\sigma^2}{2}}\Phi\left(\frac{\mu+\sigma^2-\ln K}{\sigma}\right) - K\Phi\left(\frac{\mu-\ln K}{\sigma}\right), \tag{3.61}$$

where Φ denotes the cumulative distribution function of the standard normal distribution. Recall from lemma 3.3.1 that the value process of a European call with payoff function $f(x) = (x-K)^+$ is given by

$$\begin{aligned}V_t^f = {}& \mathbb{1}_{\{\tau_J \leqslant t\}}\mathbb{E}\left[\left((1-\kappa)\tilde{S}_{\tau_J}\mathcal{E}\left(\sigma(W_T - W_{\tau_J})\right) - K\right)^+|\mathcal{F}_t\right] \\ &+ \mathbb{1}_{\{t<\tau_J\}}\int_t^T \mathbb{E}\left[\left((1-\kappa)\tilde{S}_u\mathcal{E}\left(\sigma(W_T - W_u)\right) - K\right)^+|\mathcal{F}_t\right]e^{\Gamma_t - \Gamma_u}h(u)du \\ &+ \mathbb{1}_{\{t<\tau_J\}}\mathbb{E}\left[(\tilde{S}_T - K)^+|\mathcal{F}_t\right]e^{\Gamma_t - \Gamma_T}.\end{aligned} \tag{3.62}$$

Below we use equation (3.61) to calculate the conditional expectations in (3.62). For $t \in [0, T]$ and an $\mathcal{F}$-measurable random variable Y we use the notation $\mathbb{E}_t[Y] = \mathbb{E}[Y|\mathcal{F}_t]$. For the second summand in (3.62) it holds that

$$\begin{aligned}
&\mathbb{E}_t\left[\int_t^T \left((1-\kappa)\tilde{S}_u\mathcal{E}\left(\sigma(W_T-W_u)\right)-K\right)^+ e^{\Gamma_t-\Gamma_u}h(u)du\right] \\
&=\int_t^T \mathbb{E}_t\left[\left((1-\kappa)\tilde{S}_u\mathcal{E}\left(\sigma(W_T-W_u)\right)-K\right)^+\right] e^{\Gamma_t-\Gamma_u}h(u)du \\
&=\int_t^T (1-\kappa)\tilde{S}_t\mathbb{E}_t\left[\left(\frac{\tilde{S}_u\mathcal{E}\left(\sigma(W_T-W_u)\right)}{\tilde{S}_t}-\frac{K}{\tilde{S}_t(1-\kappa)}\right)^+\right] e^{\Gamma_t-\Gamma_u}h(u)du.
\end{aligned} \tag{3.63}$$

From (3.12) we know that

$$\frac{\tilde{S}_u\mathcal{E}\left(\sigma(W_T-W_u)\right)}{\tilde{S}_t} \approx \log\mathcal{N}\left(\int_t^u \kappa h(s)ds-\frac{\sigma^2}{2}(T-t),\sigma^2(T-t)\right) \tag{3.64}$$

and we can apply equation 3.61 to see that, for $t\in[0,T]$, equation (3.63) can be written as

$$(1-\kappa)\tilde{S}_t\int_t^T e^{-(1-\kappa)(\Gamma_u-\Gamma_t)}\Phi(d_1(u))h(u)du-K\int_t^T e^{-(\Gamma_u-\Gamma_t)}\Phi(d_2(u))h(u)du \tag{3.65}$$

with

$$d_{1,2}(u)=\frac{\kappa(\Gamma_u-\Gamma_t)\pm\frac{\sigma^2}{2}(T-t)+\ln\frac{(1-\kappa)\tilde{S}_t}{K}}{\sigma\sqrt{T-t}},\quad u\in[t,T]. \tag{3.66}$$

Moreover, for every $t\in[0,T]$,

$$\frac{\tilde{S}_T}{\tilde{S}_t}\sim\log\mathcal{N}\left(\int_t^T \kappa h(s)ds-\frac{\sigma^2}{2}(T-t),\sigma^2(T-t)\right) \tag{3.67}$$

implies that

$$\begin{aligned}
\mathbb{E}_t\left[(\tilde{S}_T-K)^+e^{\Gamma_t-\Gamma_T}\right] &= \tilde{S}_t\mathbb{E}_t\left[\left(\frac{\tilde{S}_T}{\tilde{S}_t}-\frac{K}{\tilde{S}_t}\right)^+\right]e^{\Gamma_t-\Gamma_T} \\
&= e^{(1-\kappa)(\Gamma_t-\Gamma_T)}\Phi\left(\widehat{d}_1\right)-Ke^{\Gamma_t-\Gamma_T}\Phi\left(\widehat{d}_2\right),
\end{aligned} \tag{3.68}$$

with

$$\widehat{d}_{1,2}=\frac{\kappa(\Gamma_T-\Gamma_t)\pm\frac{\sigma^2}{2}(T-t)+\ln\frac{\tilde{S}_t}{K}}{\sigma\sqrt{T-t}}. \tag{3.69}$$

Adding equations (3.65) and (3.68) confirms that ϕ and ψ in (3.24) are given by (3.22). Finally, on $[\tau_J,T]$ the price process S follows a geometric Brownian motion started as S_{τ_J}, whence the value process $(V_t^f)_{t\in[\tau_J,T]}$ is equal to the Black-Scholes value process $S_t\Phi(\bar{d}_1)-K\Phi(\bar{d}_2)$ with

$$\bar{d}_{1,2}=\frac{\pm\frac{\sigma^2}{2}(T-t)+\ln\frac{(1-\kappa)\tilde{S}_t}{K}}{\sigma\sqrt{T-t}}. \tag{3.70}$$

The proof of corollary 3.3.1 is thus completed. □

Proof of corollary 3.3.2. Let $F : [0,T] \times [0,\infty) \to [0,\infty)$ be a function given by

$$F(t, S_t) = S_t \phi_t - K\psi_t \tag{3.71}$$

with ϕ and ψ as in equation (3.23). A short direct calculation using (3.22) and (3.23) shows that

$$\frac{\partial F}{\partial S}(t, S_t) = \phi_t, \quad t \in [0,T]. \tag{3.72}$$

In order to use theorem 3.3.1 to derive the optimal hedging strategy, we first observe that, for all $t \in [0,T]$,

$$\mathbb{E}\left[((1-\kappa)S_t\mathcal{E}\left(\sigma(W_T - W_t)\right) - K)^+ \,|\mathcal{F}_t\right] = (1-\kappa)S_t\Phi(\tilde{d}_1) - K\Phi(\tilde{d}_2), \tag{3.73}$$

with $\bar{d}_{1,2}$ given by (3.70). Combining this with (3.72) implies that

$$\begin{aligned} \kappa_t^f &= -\kappa S_t \frac{\partial F}{\partial S}(t, S_t) - \mathbb{E}\left[f\left((1-\kappa)S_t\mathcal{E}\left(\sigma(W_T - W_t)\right)\right)|\mathcal{F}_t\right] + F(t, S_t) \\ &= -\kappa S_t\phi_t - (1-\kappa)S_t\Phi(\bar{d}_1) + K\Phi(\bar{d}_2) + S_t\phi_t - K\psi_t \\ &= (1-\kappa)S_t\left(\phi_t - \Phi(\bar{d}_1)\right) - K\left(\psi_t - \Phi(\bar{d}_2)\right), \quad t \in [0,T]. \end{aligned} \tag{3.74}$$

Recall from (3.24) that

$$V_t^f \mathbb{1}_{\{\tau_J \leqslant t\}} = \left[S_t\left(\phi_t - \Phi(\bar{d}_1)\right) - K\left(\psi_t - \Phi(\bar{d}_2)\right)\right]\mathbb{1}_{\{\tau_J \leqslant t\}}, \quad t \in [0,T], \tag{3.75}$$

which implies by standard Black-Scholes arguments that

$$\frac{\partial V_t^f}{\partial S}\mathbb{1}_{\{\tau_J \leqslant t\}} = \Phi(\bar{d}_1)\mathbb{1}_{\{\tau_J \leqslant t\}}, \quad t \in [0,T]. \tag{3.76}$$

Finally, plugging equations (3.72), (3.74) and (3.76) into (3.20) shows that

$$\begin{aligned} \theta_t^* &= \left(\frac{\partial F}{\partial S}(t, S_t) + \frac{\kappa\kappa_t^f h(t)}{S_t(\sigma^2 + \kappa^2 h(t))}\right)\mathbb{1}_{\{t<\tau_J\}} + \frac{\partial V^f}{\partial S}(t, S_t)\mathbb{1}_{\{\tau_J \leqslant t\}} \\ &= \left(\phi_t + \frac{\kappa\kappa_t^f h(t)}{S(\sigma^2 + \kappa^2 h(t))}\right)\mathbb{1}_{\{t<\tau_J\}} + \Phi(\bar{d}_1)\mathbb{1}_{\{\tau_J \leqslant t\}}, \quad t \in [0,T], \end{aligned} \tag{3.77}$$

for

$$\kappa_t^f = (1-\kappa)S_t\left(\phi_t - \Phi\left(\bar{d}_1\right)\right) - K\left(\psi_t - \Phi\left(\bar{d}_2\right)\right). \tag{3.78}$$

The proof of corollary 3.3.2 is thus completed. □

Chapter 4

A simple bubble mechanism: time-varying momentum horizon

4.1 Introduction

We provide a simple model that builds on the fundamental assumption that financial markets continuously exhibit transitions between phases of growth, exuberance (conventionally called bubbles) and crises [226]. In this framework, most crises are actually endogenous and the consequence of procyclical (also called positive) feedbacks [229, 230, 231]. In our stylized framework, the excess return and volatility of the price generating process are assumed to be functions of the momentum only, which is estimated over a time horizon that may vary as a function of the momentum itself. The intermittent positive feedbacks between momentum and the time horizon over which momentum is estimated leads to regime shifts from non-bubble to transient bubble phases. As a result, the obtained price process exhibits the major known stylized facts of empirical financial time series. In particular, the model provides a straight-forward mechanism for the formation and evolution of bubbles, which combines the concept that momentum trading is ubiquitous amongst investors and the perversive occurrence of bubbles.

Indeed, momentum or trend following strategies constitute one of the most popular technical investment techniques used by hedge-funds, mutual funds and individual investors in general [9, 10]. Momentum, the tendency for rising asset prices to rise further, and falling prices to keep falling, enjoys strong empirical support [111, 112, 151, 152] and provides an improved explanatory power of risk premia in factor model regressions [43, 92]. Momentum effects have been documented in the US, Europe and Asia Pacific markets [92], and across different asset classes [12]. In addition, momentum is not only present in cross-sectional investigations for individual firms but also detected from the aggregated time series like stock indices and futures contracts [121, 189].

Momentum is often attributed to investors' behavioral characteristics represented by their beliefs, such as over-confidence, self-attribution and confirmation biases [64], and under-reaction/over-reaction [15], or to preferences, such as the disposition effect derived from prospect theory and mental accounting operations [110]. Yet antoher explanation for momentum effects is based on herding [74, 133]. [195] promoted the concept of convention to emphasize that prices reflect investor's beliefs. In self-fulfilling prophecies, investors make these beliefs come true through. Building on this understanding, [253] formulated a simple model for self-referential behavior in financial markets where agents build strategies based on their belief about the existence of correlation between some flow of information and price changes. Through their market impact, their strategies ensure those beliefs materialize, leading to excess volatility and regime shifts. [93] argued that trend-following strategies cause both short-term trends in prices, as well as longer-term oscillations and boom-bust cycles.

Previous models developed by our group implementing the concept of positive feedbacks to characterize bubbles have emphasized mostly the influence of past prices level on future returns [56, 138, 174, 176, 228, 230] and the reinforcing role of social influence of the price formation process [156]. The role of a nonlinear response to momentum [140] has been suggested to be at the origin of the super-exponential price growth characterising bubbles [170, 230] and also as a model of hyperinflation [234]. Here, we take a different route by constructing the simplest continuous price process whose expected returns and volatility are functions of momentum only. The momentum is measured as a simple continuous moving average of past prices over a given time horizon. This momentum is then used to surmise future returns by the investors. We formulate a simple self-consistent framework to embody the market impact of momentum strategies. The key idea is that investors use momentum strategies because they believe that these strategies have value. Due to their market impact, the momentum strategies in turn modify the price structure, leading to a reinforcement of the price momentum. In addition, taking into account the fact that competition between momentum traders make them decrease their time horizons as the price momentum develops, this leads to transient self-fulfilling bubbles via the positive feedback mechanism between the triplet of price, momentum and the time horizon at which momentum is estimated.

The organization of this chapter is as follows. Section 4.2 presents a derivation of our model, introducing and discussing the essential underlying assumptions. Section 4.3 presents a unified precise setting, which summarizes the derivation of section 4.2, and investigates the properties of the solutions to our model. Section 4.4 describes a quasi-likelihood methodology to calibrate the model to synthetic and empirical financial time series data and a hypothesis test for bubbles. Section 4.5 concludes. Proofs are given in the appendix in section 4.6.

4.2 Derivation of the momentum driven price model

4.2.1 General formulation of momentum dynamics

Consider a price series $\{S_t, t = 0, 1, 2, ...\}$ in discrete time and its associated returns $\{r_t := \ln(S_t/S_{t-1}), t = 1, 2, ..\}$. The exponential moving average momentum X_t over a typical time horizon $\tau = 1/\theta$, with $\theta \in (0, 1)$, is defined as

$$X_t = (1-\theta)X_{t-1} + \theta\, r_t \, . \tag{4.1}$$

Let L be the lag operator defined by $L[x(t)] = x(t-1)$. Then, equation (4.1) can be reformulated as

$$[1-(1-\theta)L]X_t = \theta r_t = \theta \ln\left(\frac{S_t}{LS_t}\right) \tag{4.2}$$

Its solution reads

$$X_t = [1-(1-\theta)L]^{-1}\left(\theta \ln \frac{S_t}{LS_t}\right) = \theta \sum_{k=0}^{\infty}(1-\theta)^k \ln \frac{S_{t-k}}{S_{t-k-1}} \ . \tag{4.3}$$

This shows that the average momentum X_t is nothing but the geometric moving average of the historical log price increments.

Equation (4.1) can be reformulated in terms of the increments $\Delta X_t := X_t - X_{t-1}$ and $\Delta \ln S_t := \ln(S_t/S_{t-1})$ to obtain

$$\Delta X_t = -\theta X_{t-1} + \theta\,\Delta \ln S_t \ . \tag{4.4}$$

This suggests the following extension to continuous times of the average price momentum, obtained as the solution of the following stochastic differential equation

$$dX_t = -\theta X_t dt + \theta \frac{dS_t}{S_t} \ , \tag{4.5}$$

which can be interpreted as the continuous limit of the difference equation (4.4).

Example. Let us first consider the simplest example of a constant time horizon $\tau = 1/\theta$ and an asset price S that follows a geometrical Brownian motion (GBM),

$$\frac{dS_t}{S_t} = \mu\, dt + \sigma dW_t \ . \tag{4.6}$$

Replacing this process in (4.5) shows that the average price momentum X must be the Ornstein-Uhlenbeck process

$$dX_t = -\theta(X_t - \mu)dt + \theta\sigma dW_t \ , \tag{4.7}$$

whose solution reads

$$X_t = \mu + (x_0 - \mu)e^{-\theta t} + \theta\sigma \int_0^t e^{-\theta(t-s)} dW_s \ . \tag{4.8}$$

This solution can be transformed into the more convenient form

$$X_t = \mu + \sigma\sqrt{\frac{\theta}{2}} e^{-\theta t} \cdot \widetilde{W}(e^{2\theta t}) \ , \tag{4.9}$$

for a Brownian motion $\widetilde{W}$. This shows that, while the GBM price process (4.6) is non-stationary, its moving average X is a stationary process with a stationary distribution given by

$$X_t \sim N\left(\mu, \frac{\theta}{2}\sigma^2\right) \ . \tag{4.10}$$

The longer the time horizon $\tau(\theta) = 1/\theta$ over which the average price momentum is estimated, the smaller is its variance $\theta\sigma^2/2$. This is nothing else but the statement that the average price momentum is a consistent estimator of the price drift μ, which converges with a standard deviation scaling as the inverse square root $1/\sqrt{\tau(\theta)}$ of the time horizon $\tau(\theta)$ used for its estimation.

4.2.2 Preliminary setting

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,\infty)}, \mathbb{P})$ be a filtered probability space with a right-continuous and $\mathbb{P}$-complete filtration, let $(W_t)_{t\in[0,\infty)}$ be a real valued $(\mathcal{F}_t)_{t\in[0,\infty)}$-adapted Brownian motion, let $f : [0,\infty) \times \Omega \to \mathbb{R}$ be adapted and $g : [0,\infty) \times \Omega \to [0,\infty)$ be predictable be *suitably regular*[1] stochastic processes, let $(S_t)_{t\in[0,\infty)}$ be an Itô process with the property that

$$\frac{dS_t}{S_t} = f_t dt + g_t dW_t \,. \tag{4.11}$$

Further, following the reasoning in section 4.2.1 above, let $\theta : [0,\infty) \times \Omega \to \mathbb{R}$ be a stochastic process describing the inverse time horizon and $(X_t)_{t\in[0,\infty)}$ be the average price momentum of S given by

$$dX_t = -\theta_t X_t dt + \theta_t \frac{dS_t}{S_t} \,. \tag{4.12}$$

To impose further structure on our model, we will make two major additional assumptions that are discussed below, a structural assumption on the impact of momentum on the price and an explicit mechanism allowing for positive feedback of increasing momentum on itself.

4.2.3 Assumption I: Impact of momentum on absolute price level

Assuming the preliminary setting 4.2.2 and based on the intuition that investors focus in aggregate on the momentum as the leading indicator of pricing, we introduce

A . There exists a C^2-function $\lambda : [0,\infty) \to [0,\infty)$ with strictly positive first and second derivative such that the price can be written as a function of the momentum X, in the sense that

$$S_t = \lambda(X_t), \quad t \in [0,\infty). \tag{4.13}$$

Lemma 4.6.1 in section 4.6.1 shows that assumption (A) naturally implies that the time-varying horizon θ_t and the volatility represented by g_t^2 satisfy

$$\theta_t = \frac{\lambda(X_t)}{\lambda'(X_t)} \quad \text{and} \quad g_t^2 = \frac{2\lambda(X_t)X_t}{\theta_t^2 \lambda''(X_t)}, \quad \text{for all } t \in [0,\infty). \tag{4.14}$$

[1] Below we will allow for processes with explosion and thus deal with solutions S and coefficient functions f, g being defined only on stochastic intervals $[0,\zeta)$ for a random time $\zeta : \Omega \to [0,\infty)$.

Comments. The impact of momentum on price has been considered in the literature by letting its drift and volatility functions depend on X to arrive at a specification of the form

$$\frac{dS_t}{S_t} = f(X_t)\,dt + g(X_t)\,dW_t\ . \tag{4.15}$$

This kind of structure has been employed in the stochastic volatility model of [132], which builds the price process as depending on an exponentially weighted moving average of some past price reference level, similar to equation (4.15). However, in their model, X_t is the price change surprise, defined as the difference between the current log-price and an exponentially weighted average of past log-prices, rather than the price momentum itself used in our model. [159] and [106] develop models of the impact of trading strategies and portfolio allocation on prices that have similar forms as (4.15).

Note that the above assumption (A) is a somewhat stronger assumption than (4.15) and imposes a lot of structure on our model. To see this, recall the example of Brownian motion discussed in section 4.2.1 above: S is a geometric Brownian motion while its exponential moving average X is an Ornstein-Uhlenbeck process. Consequently, there can be no function λ with the property that $S_t = \lambda(X_t)$ for $t \in [0, \infty)$, indicating that assumption (A) a priori restricts feasible asset price dynamics. Viewed from a different perspective, we start from the inherently non-Markovian structure given by equations (4.11) and (4.12) (the price increments depend on the momentum, which itself depends on the price history) and look for a Markovian special case that still satisfies the momentum equations. Also note that assumption (A) does not impose a specific form of the drift f_t of S.

4.2.4 Assumption II : Positive feedback of momentum on momentum time-horizon

Assuming again the preliminary setting 4.2.2 and based on the possible effect of positive feedback of momentum on its own time horizon $\tau = 1/\theta$ during a bubble, we introduce

B . There exist constants $\theta^* \in (0, \infty)$ and $\eta \in [0, 1)$ such that the random time horizon $(\theta_t)_{t\in[0,\infty)}$ can be written as

$$\theta_t = \theta(X_t) = \theta^* + \eta X_t. \tag{4.16}$$

Together with assumption (A) (in the form of equation (4.14)), this implies that there exists a constant $\underline{S} \in (0, \infty)$ such that

$$S_t = \underline{S}\left(1 + \frac{\eta}{\theta^*}X_t\right)^{\frac{1}{\eta}} = \underline{S}\left(\frac{\theta_t}{\theta^*}\right)^{\frac{1}{\eta}}. \tag{4.17}$$

and

$$g_t = g(X_t) = \sqrt{\frac{2}{1-\eta}X_t}. \tag{4.18}$$

Note that the case $\eta = 0$ is embedded in this formulation via its pointwise limit

$$S_t = \lim_{\eta \to 0} \left(1 + \frac{\eta}{\theta^*} X_t\right)^{\frac{1}{\eta}} = \exp\left(\frac{X_t}{\theta^*}\right). \tag{4.19}$$

Comments. Statistical tests of the performance of momentum or trend-following strategies suggest that they have delivered significant positive performance through very different economic and financial environments and over very long time horizons [136, 171]. This is reflected in their broad use by hedge-funds and professional investors. However, their implementation in general involves time-varying exposures to manage the highly variable risk of momentum [16, 82, 239]. Moreover, they exhibit periods with weak performance [139], as for instance since 2011 for most developed financial markets. Pressured by the need to provide positive risk-adjusted return, many professional investors periodically tinker the parameters of their trend-following strategies to adapt to the recent past. For this purpose machine learning algorithms are often employed, with the rational to approximate best the unknown price generating process by adapting the parameters of the momentum strategies. This tends to create destabilising feedback loops in the form of bubble-corrections cycles [29, 93].

Thus, during a typical financial cycle, the time horizon used in momentum strategies is progressively reduced as an adaptation to the increasing price levels and to the accelerating volatility dynamics. In a *reflexive* race to forerun competitors, trend-followers attempt to adjust their position before the herd. They believe that, by decreasing the time horizon of their momentum indicators, they will be able to forerun their peers[2]. In general, this culminates in a bubble peak, followed by a high volatility regime, possibly a crash and the subsequent detuning of trend-following strategies. And it is only after a recovery period that they become again usable when a new trend appears.

The goal of assumption (B) is to include this effect in the momentum driven price model. To represent the attempt by trend-followers not to be left behind by their peers, we introduce this second-order trend following in the sense that the parameter θ, introduced in section 4.2.1 to estimate the price momentum, is itself a function of the observed momentum X. Specifically, we assume that, for every t, θ_t is a linearly increasing function of X_t. The representative investor thus becomes more and more myopic for increasing momentum, as his time horizon $\tau = 1/\theta$ is shrinking.

4.2.5 Complementary Assumption: linear impact of momentum on the drift

Finally, to complement assumptions (A) and (B) of sections 4.2.3 and 4.2.4 and for practical ease, we add the following assumption on the drift f_t in the preliminary setting 4.2.2. In analogy with earlier assumptions, we assume that the drift can be written as a function of the momentum, specifically

C . *There exist constants $c \in \mathbb{R}$, $d \in (0, \infty)$ such that the drift f of S can be written as the following linear*

[2] This description reflects information provided by professionals through many private communications.

function

$$f_t = f(X_t) = cX_t + d. \tag{4.20}$$

While all arguments and classification obtained below apply to both $c < 0$ and $c \geq 0$, we will focus on the later case $c \geq 0$, corresponding to the more natural positive relationship between momentum and average growth rate.

Collecting assumptions (A), (B) and (C) leads us to the following final model setting for the momentum driven price model.

4.3 Precise model setting and solutions

4.3.1 Setting

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,\infty)}, \mathbb{P})$ be a filtered probability space with a right-continuous and $\mathbb{P}$-complete filtration, let $(W_t)_{t\in[0,\infty)}$ be a real valued $(\mathcal{F}_t)_{t\in[0,\infty)}$-adapted Brownian motion, let $\eta \in [0, 1)$, $\theta^* \in (0, \infty)$, $c \in [0, \infty)$, $d \in (0, \infty)$ be constants, let $f : [0, \infty) \to \mathbb{R}$, $g : [0, \infty) \to [0, \infty)$ and $\theta : [0, \infty) \to (0, \infty)$ be functions given by

$$f(x) = cx + d, \tag{4.21}$$

$$g(x) = \sqrt{\frac{2x}{1-\eta}}, \tag{4.22}$$

$$\theta(x) = \theta^* + \eta\, x, \tag{4.23}$$

and $X_0, \underline{S} \in (0, \infty)$. Then, following the discussion in section 4.2 above, there exists[3] a unique two-dimensional diffusion $(S_t, X_t)_{t\in[0,\infty)}$, describing a stock price S and its momentum X, with the property that

$$dX_t = -\theta(X_t)X_t dt + \theta(X_t)\frac{dS_t}{S_t} \quad \text{and} \quad X(t=0) = X_0, \tag{4.24}$$

$$\frac{dS_t}{S_t} = f(X_t)dt + g(X_t)dW_t, \tag{4.25}$$

$$S_t = \underline{S}\left(1 + \frac{\eta}{\theta^*}X_t\right)^{\frac{1}{\eta}}. \tag{4.26}$$

4.3.2 Classification based on boundary behavior

The setting above allows for diverse behavior of the stock price S and its momentum X. To further differentiate this, we need some terminology.

[3] Up to a random time, at which the processes may explode. For (pathwise) uniqueness and (strong) existence of solutions see, e.g., theorem 3.1 and 3.2 in [141]. For details on explosion see comments below.

Definition 4.3.1. Let the setting in section 4.3.1 be fulfilled and $X : [0,\infty) \times \Omega \to [0,\infty)$ be a strong solution of the SDE (4.24) with initial value $X_0 \in (0,\infty)$ up to its explosion time $\tau = \inf\{t \in (0,\infty] | S_t = \infty\}$. Then X is called

(a) *explosive*, if $\mathbb{P}[\tau < \infty] > 0$,

(b) *diffusive-transient*,[4] if $\mathbb{P}[\tau = \infty] = 1$ and $\lim_{t\to\infty} X_t = \infty$,

(c) *recurrent*, if $\forall a \in (0, X_0)$ it holds that $\mathbb{P}[X_t = a \text{ for some } t \in (0,\infty)] = 1$,

(d) *strictly positive*, if $\mathbb{P}[X_t > 0 \text{ for all } t \in (0,\infty)] = 1$,

(e) *instantaneously reflected at* 0, if $\mathbb{P}[X_t = 0 \text{ for some } t \in (0,\infty)] > 0$ and for all $t, \epsilon \in (0,\infty)$ we have $\mathbb{P}\left[X\mathbb{1}_{[t,t+\epsilon]} \equiv 0\right] = 0$.

The notions (a)-(c) describe the behavior of X at the upper boundary infinity, while the notions (d)-(e) describe the behavior at the lower boundary 0. We will see below that the full parameter range of section 4.3.1 can be covered with these descriptions.

4.3.2.1 The case $\eta = 0$.

In this special case, the characteristic time-horizon used by the representative investor to gauge momentum is constant, $\theta_t = \theta^*$, and the governing equation (4.24) for the momentum X reads

$$dX_t = \theta^* \left(X_t \left(c-1\right) + d\right) dt + \theta^* \sqrt{2X_t} dW_t \,, \tag{4.27}$$

and

$$S_t = \lim_{\eta\to 0} \underline{S} \left(1 + \frac{\eta}{\theta^*} X_t\right)^{\frac{1}{\eta}} = \underline{S} \exp\left(\frac{X_t}{\theta^*}\right), \tag{4.28}$$

Using theorems 5.1, 5.5 (for the case $c = 1$) and theorems 5.3, 5.7 (for the case $c \neq 1$) in [50], the behavior of X can be distinguished by

1. X is diffusive-transient for $c \in [1,\infty)$ and $c = 1 \wedge d > \theta^*$
2. X is recurrent for $c \in [0,1)$ and $c = 1 \wedge d \in (0,\theta^*]$
3. X is strictly positive for $d \geqslant \theta^*$ and instantaneously reflected at 0 if $d \in (0,\theta^*)$.

Comments and classification

- In the recurrent case, as (i) there is no positive feedback of increasing momentum on the momentum time-horizon, and (ii) the momentum is roughly mean-reversal with shrinking volatility, we will refer to this case as the ***non-bubble regime***. Indeed, here X is equal to the well known CIR model [61], which can thus be motivated as the dynamics of a variable X that averages over a fixed memory length the past realizations of an underlying process S.

[4] To avoid confusion with the term *transient* as used in economics for temporary phenomena, we explicitly add *diffusive*.

- In the diffusive-transient case, equation (4.27) suggests that X has a stochastic exponential growth rate bounded from below by $\theta^*(c-1)$ and with shrinking volatility. In particular, ignoring the stochastic term, for large S we have $S_t \sim \exp(\exp(\theta^*(c-1)\ t))$, and S grows double-exponentially. As S shows super-exponential growth but is not explosive, we will refer to this case as ***mild bubble regime***.

4.3.2.2 The case $\eta \in (0,1)$.

Here, the characteristic time-horizon of the representative investor is an decreasing function of the momentum and equation (4.24) reads

$$dX_t = (\theta^* + \eta X_t)\,(X_t\,(c-1)+d)\,dt + (\theta^* + \eta X_t)\sqrt{\frac{2X_t}{1-\eta}}\,dW_t. \tag{4.29}$$

Lemma 4.6.2 in section 4.6.1 shows that

1. X is explosive for $c \in \left(\frac{1}{1-\eta}, \infty\right)$ and recurrent for $c \in \left[0, \frac{1}{1-\eta}\right]$,
2. X is strictly positive for $d \in \left[\frac{\theta^*}{1-\eta}, \infty\right)$ and instantaneously reflected at 0 if $d \in \left(0, \frac{\theta^*}{1-\eta}\right)$.

As the self-referential influence of the momentum on its own time horizon creates the possibility of large hikes of X (and thus S), we will refer to both these cases as the *bubble regimes*.

We can make the following further distinction for two kinds of bubbles, given $\eta \in (0,1)$.

Comments and classification

- In the recurrent case $c \in \left[0, \frac{1}{1-\eta}\right]$, X will experience periodical rallies with rapid growth followed by sharp declines. S, the exponential of X, will thus experience similar but steeper rallies that mimic the pattern of bubbles. For details see section 4.3.3.2 below. We refer to this case as ***recurrent bubble regime***.
- In the explosive case $c \in \left(\frac{1}{1-\eta}, \infty\right)$, X and S both experience super-exponential growth, faster than double exponential, with a stochastic finite-time-singularity. We refer to this case as the ***wild bubble regime***.

4.3.2.3 Interpretation of model parameters

Based on the model derivation in section 4.2, let us discuss the model parameters (θ^*, η, c, d) and their interpretation, before we tackle the calibration of the model.

1. The parameter $1/\theta^*$ determines the characteristic time-horizon of the exponential moving average momentum in the regime $\eta = 0$, cf. section 4.2.1. For $\eta \in (0,1)$, θ^* sets the scale $\sim \theta^*/\eta$ for the minimum momentum that can influence the time horizon with which momentum is estimated by investors.

2. The parameter η represents the impact per unit of momentum on the time-horizon that is used to measure momentum. The term $\dfrac{\theta^*}{\eta}$ is the characteristic momentum necessary to influence the time horizon θ in the bubble regime: ranging from $\sim \infty$ for $\eta \sim 0$ (no feedback of momentum on the time horizon) to θ^* for $\eta \sim 1$ (small momentum creates considerable feedback on time horizon). Consistent with our view of bubbles being characterized by the influence of momentum on its on time-scales, η can be understood as a *bubbliness* parameter.

3. The parameter c quantifies the linear dependence of stock returns on unit momentum. For given $\eta \in [0, 1)$, it allows one to distinguish the four main regimes discussed in sections 3.2.1 and 3.2.2 above, *non-* and *mild*, *recurrent*, or *wild* bubble. As such, c, together with η, determines the bubble type.

4. The parameter d represents an intrinsic return independent of momentum and is thus not related to any notion of bubbliness. However, it determines the nature of price or momentum dynamics near zero, in particular whether the momentum can vanish or not. As such, it has a considerable influence on the stylized dynamics of the process.

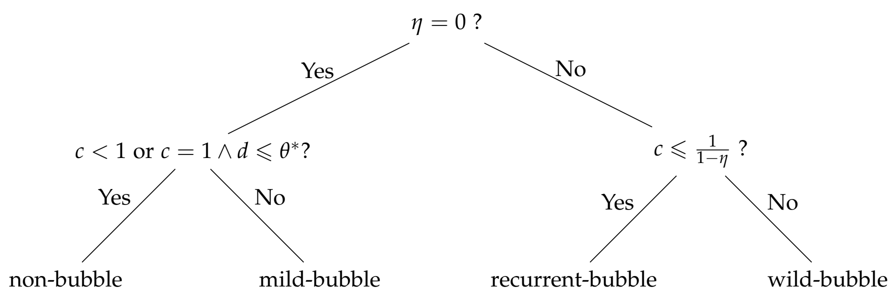


Figure 4.1: Different types of bubbles distinguished by the relationship between η and c

4.3.3 Nondimensionalization

Below we investigate the role of different parameters in our model and propose a reparameterization yielding a standardized set of parameters, which will simplify calibration and further discussion.

4.3.3.1 Change of time horizon

Here, we conduct a time change using the characteristic time horizon of the momentum, $1/\theta^*$. This shows how X can be derived from a *time*-dimensionless process Y and reduces the characteristic parameters (θ^*, η, c, d) to (η, c, d).

Let the setting in section 4.3.1 be fulfilled and (S, X) the corresponding diffusion, additionally let $d^Y \in (0, \infty)$ be a constant and let $f^Y : [0, \infty) \to \mathbb{R}$ and $\theta^Y : [0, \infty) \to (0, \infty)$ be functions given by

$$f^Y(y) = cy + d^Y, \tag{4.30}$$
$$\theta^Y(y) = 1 + \eta y. \tag{4.31}$$

As above, there exists a two-dimensional Itô process $(S_t^Y, Y_t)_{t\in[0,\infty)}$ with initial value (S_0, X_0), describing a stock price S^Y and its momentum Y, with the property that

$$dY_t = -\theta^Y(Y_t)Y_t dt + \theta^Y(Y_t)\,\frac{dS_t^Y}{S_t^Y}\,, \tag{4.32}$$
$$\frac{dS_t^Y}{S_t^Y} = f^Y(Y_t)dt + g(Y_t)dW_t, \tag{4.33}$$
$$S_t^Y = \underline{S}\,(1 + \eta\, Y_t)^{\frac{1}{\eta}}\,. \tag{4.34}$$

Then we have the following result.

Proposition 4.3.1. *Let the setting above be fulfilled and assume $d = d^Y\theta^*$. Then it holds (in the sense of pathwise uniqueness) that*

$$\theta^* Y_{t\theta^*} = X_t \quad \textit{and} \quad S_{t\theta^*}^Y = S_t. \tag{4.35}$$

The proposition follows immediately from corollary 3 in [194], which shows equivalence in distribution, and the fact that the involved stochastic differential equations have strong solutions. We can conclude that the parameter θ^* acts as a time change in our model and the behavior of X can be fully characterized in the reduced parameter space (η, c, d).

4.3.3.2 Reparameterization

Following our classification of bubbles based on boundary behavior in section 4.3.2 above, we introduce two reduced parameters $C \in [0,\infty), D \in (0,\infty)$ as

$$C = c(1-\eta), \quad D = \frac{d(1-\eta)}{\theta^*}\,. \tag{4.36}$$

As mentioned in subsection 4.2.5, we focus on the case $c \in [0,\infty)$ and thus $C \in [0,\infty)$ corresponding to the normal relationship between momentum and average growth rate. By plugging (4.21), (4.22) into (4.24) and (4.25), the dynamics of X and S can be rewritten as follows. For initial values $X(t=0) = X_0$ and $S(t=0) = \underline{S}\left(1 + \frac{\eta}{\theta^*}X_0\right)^{\frac{1}{\eta}}$, we have

- $\eta = 0$:

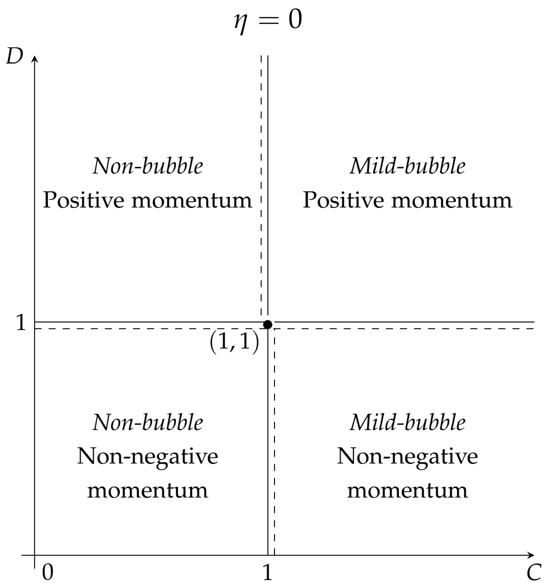
$$dX_t = [\ (C-1)\theta^* X_t + D{\theta^*}^2\ ]\ dt + \theta^*\sqrt{2X_t}\ dW_t, \tag{4.37}$$
$$dS_t = S_t\theta^*(C\ \ln S_t + D)\,dt + S_t\sqrt{2\,\theta^*\ln S_t}\,dW_t\,. \tag{4.38}$$

- $\boldsymbol{\eta \in (0,1)}$:

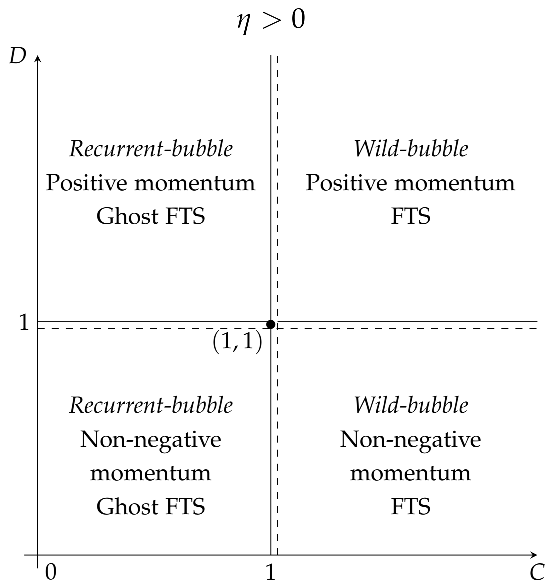
$$dX_t = (\theta^* + \eta X_t)\left(\frac{C-1+\eta}{1-\eta}X_t + \frac{D\theta^*}{1-\eta}\right)dt + (\theta^* + \eta X_t)\sqrt{\frac{2X_t}{1-\eta}}\,dW_t \tag{4.39}$$

$$dS_t = \frac{\theta^*}{\eta(1-\eta)}[\,C\,S^{\eta+1} + (D\eta - C)S_t\,]\,dt + \sqrt{\frac{2\,\theta^*}{(1-\eta)\eta}}\cdot\sqrt{S_t^{\eta+2} - S_t^2}\,dW_t\,. \tag{4.40}$$

This reparameterization allows for a standardized categorization of bubble types in the parameter space (C, D) across different values of η. Figure 4.2 illustrates these categories.



(a) The two types of bubble and two types of non-bubble when $\eta = 0$.



(b) The four types of stronger bubble when $\eta \in (0,1)$.

Figure 4.2: Bubble classification using the standardized parameters C and D. *FTS* refers to a finite-time singularity (explosive). The term *Ghost FTS* refers to transient explosive behavior, see section 4.3.3.3 below for details.

Note that, for fixed parameters $\theta^*, C \in [0,\infty), D \in (0,\infty)$, the growth rate of a bubble increases with $\eta \in [0,1)$. This can be seen directly from the drift terms of equations (4.38) and (4.40), noting that the function $h(\eta) = \frac{S^\eta - 1}{\eta(1-\eta)}$ is monotonically increasing for $S > 0$. This justifies the use of η as a measure of *bubbliness* and confirms the terminology introduced in section 4.3.2,

- *wild* bubbles as opposed to *mild* bubbles for $C \in (1,\infty)$ and
- *recurrent* bubbles as opposed to *non*-bubbles for $C \in (0,1)$.

Below we will show that recurrent bubbles ($\eta \in (0,1)$, $C \in [0,1]$) experience volatility-driven so-called *ghost-finite time singularities* followed by sharp declines. This means that both bubble types with $\eta \in (0,1)$ show a certain type of explosive behavior, which motivates us to test for the hypothesis of $\eta = 0$ in bubble detection, see section 4.4.4.

4.3.3.3 Recurrent bubbles and ghost-finite time singularities

The recurrent bubble case, $\eta \in (0,1)$ and $C \in [0,1]$, is characterized by solutions that are reminiscent of the so-called *ghost finite-time-singularity* introduced in [224]. Indeed, the explosive volatility term leads to transiently explosive growth phases followed by decline phases, these two regimes alternating and repeating themselves endlessly. Section 4.6.2 presents a detailed explanation and description of the properties of the recurrent bubble case in the light of the ghost finite-time-singularity.

4.3.4 Simulation of price trajectories

In order to illustrate the behavior of solutions for different parameters as discussed above, we present several simulations. Each set of simulations is organized in four subfigures, constructed by fixing two of the parameters (η, C, D) and plotting trajectories varying the third parameter. Each set uses a single random seed for all paths, identical observation length of 1000, same initial values $X_0 = 0.001$ and $\underline{S} = 1$ and the same discrete time increment $\Delta t = 0.01$ in the simulation algorithm. Based on the discussion in section 4.3.3.1, without loss of generality we fix the time-horizon θ^* to be 0.002.

Figure 4.3 shows non-bubble ($\eta = 0, C = 0.5$) and mild-bubble ($\eta = 0, C = 1.5$) regimes with positive momentum ($D = 1.5$) and non-negative momentum ($D = 0.5$) for each regime. The trajectories in the top graphs experience recurrent fluctuations while the two trajectories in the bottom graphs show persistent super-exponential growth.

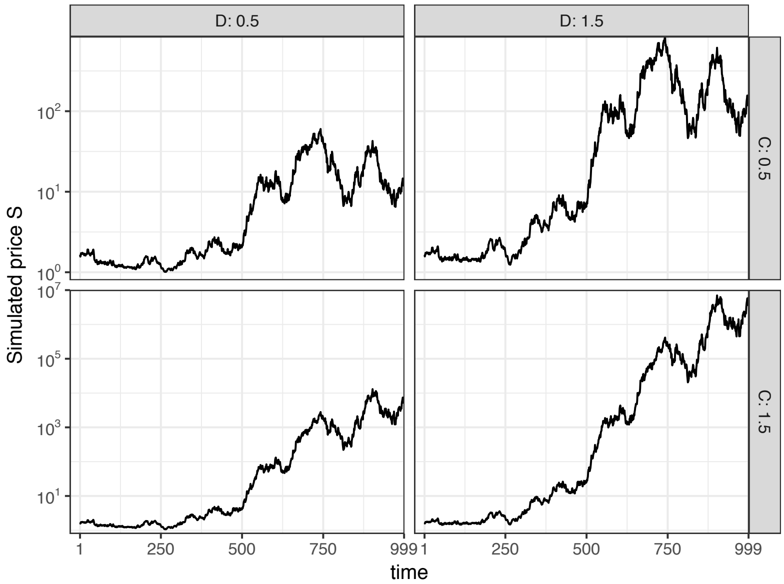


Figure 4.3: Simulated price time series for S with fixed C, D and $\eta = 0$. Note the different vertical scales in the top compared with bottom graphs.

Figure 4.4 demonstrate the four types of bubbles classified in fig. 4.2 (b) with nine simulated trajectories corresponding to different sizes of η. As shown, larger values of η generate stronger bubble for both type of recurrent bubble and wild bubble, indicated by the higher volatility and shorter life for the bubble to be fully pumped up to infinity.

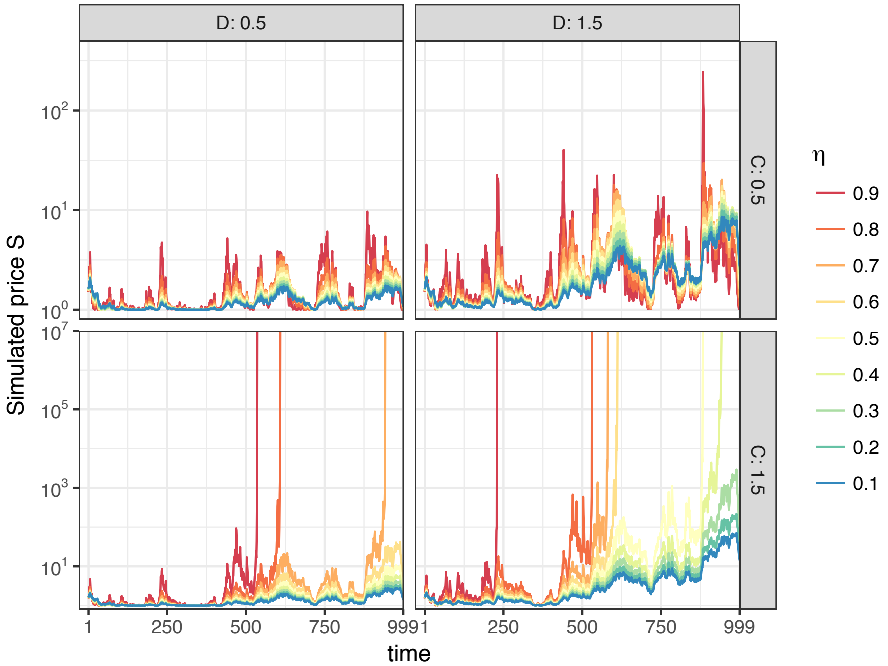


Figure 4.4: Simulated price time series for S with fixed C, D and $\eta \in [0.1, 0.9]$. Note that (increasing) η represents (increasing) *bubbliness* strength, as θ^*/η is the characteristic momentum necessary to influence the time horizon $1/\theta$. Note also the different vertical scales in the top compared with bottom graphs.

Figure 4.5 shows how the shape of bubbles changes for different values of C. The top subfigures show simulated trajectories of bubbles with non-negative momentum, while the bottom subfigures show simulated trajectories of bubbles with strictly positive momentum. The simulated bubbles on the right are stronger (more explosive) than the ones in the left, due to the larger size of η. A distinct phase transition happens for C less than 1 to larger than 1, moving from recurrent to explosive behavior. For larger values of D, the bubble characteristics are more pronounced. Larger values of η, besides earlier explosion times, seems to increase the number of rallies on a given time interval.

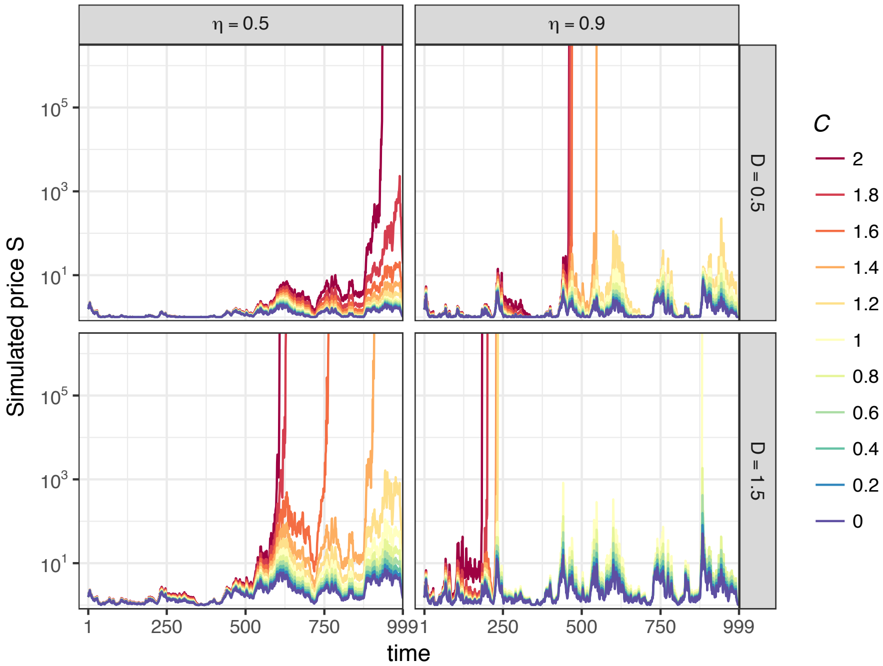


Figure 4.5: Simulated price time series for S with fixed D, η and $C \in [0, 2]$. Note that X (and thus S) is explosive for $C \in (1, \infty)$. Note the different vertical scales in the top compared with bottom graphs.

Figure 4.6 shows how the shape of bubble changes with different values of D. The top subfigures show simulated trajectories of recurrent bubbles, while the bottom subfigures show simulated trajectories of wild bubbles. The simulated bubbles on the right are stronger (more explosive) than ones in the left, due to the larger size of η. There is no distinct phase transition of bubble shape around $D = 1$. However, larger D allows for larger relative price levels of S. We can also confirm the observation that larger values of η lead to shorter time intervals between distinct rallies, as mentioned for fig. 4.5.

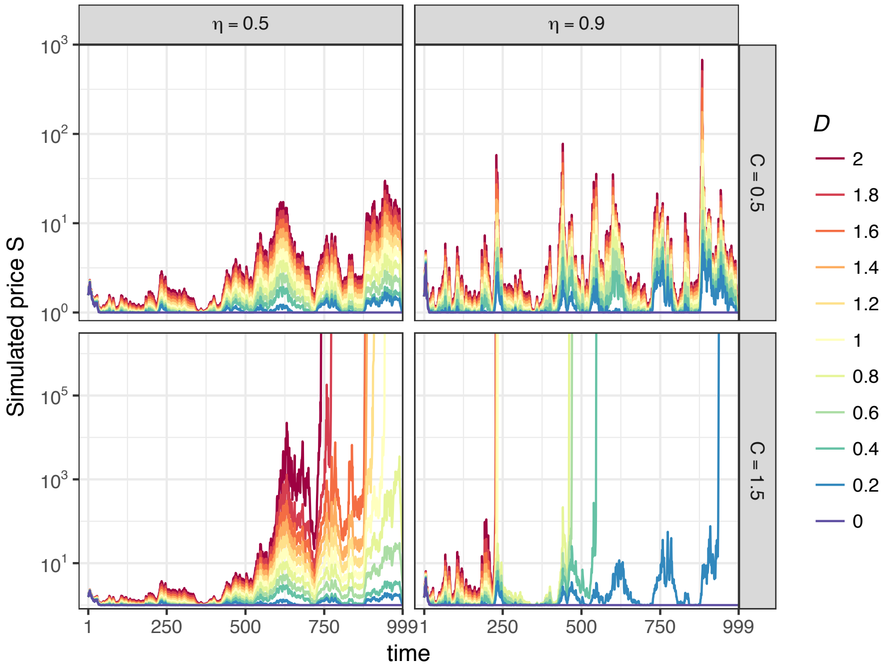


Figure 4.6: Simulated price time series for S with fixed C, η and $D \in [0, 2]$. Note that X is strictly positive (and thus S strictly larger than its lower bound $\underline{S}$) for $D \in [1, \infty)$.

4.4 Statistical inference and calibration for the bubble model

4.4.1 Formulation of the quasi-likelihood calibration method

In the present study, we adopt the maximum quasi-likelihood method to calibrate the model, partly because there is no available closed-form solution for the transition density of our bubble model. As the dynamics specification for $\eta = 0$ is different from that for $\eta \in (0, 1)$, these two cases are treated separately.

- Conditional on $\eta = 0$, the dynamics of S_t follows eq.(4.38). The Euler discretization for such SDE reads

$$S_{t+\Delta t} - S_t = S_t \theta^* (C \ \ln S_t + D) \, \Delta t + S_t \sqrt{2 \, \theta^* \ln S_t} \, (W_{t+\Delta t} - W_t) \ . \tag{4.41}$$

As $W_{t+\Delta t} - W_t \sim \mathcal{N}(0, \Delta t)$, the approximate transition densities in this case is

$$S_{t+\Delta t} \, | \, S_t \sim \mathcal{N}(S_t [1 + \theta^* (C \ \ln S_t + D)] \, \Delta t, \ \ 2 \, \theta^* S_t^2 \ln S_t \, \Delta t \ ) \ . \tag{4.42}$$

Therefore, the conditional log quasi-likelihood for a series of observations $\{S_t\}_{t=0}^{T}$ can be expressed as

$$
\begin{aligned}
\ln \mathcal{L}_\Theta &= \sum_{t=0}^{T-1} \ln \Pr{}^{\Theta} \,(S_{t+1} \,|\, S_t) \\
&= \sum_{t=0}^{T-1} -\frac{1}{2}\ln 2\pi - \ln\sqrt{2\,\theta^* S_t^2 \ln S_t} - \frac{[\,S_{t+1} - S_t - S_t\theta^*(C\,\ln S_t + D)\,]^2}{4\,\theta^* S_t^2 \ln S_t}\,. \qquad (4.43)
\end{aligned}
$$

Thereafter, we denote $\Theta^{ne} = \{\theta^*, C, D\}$ and re-express equation (4.43) as $\ln \mathcal{L}^{ne}_{\Theta^{ne}}$. The superscript *ne* indicate the case where the dynamics of S_t and X_t is of a non-explosive nature ($\eta = 0$).

- Conditional on $\eta \in (0,1)$, the dynamics of S_t follows equation (4.40). In this case, the Euler discretization for such SDE can be written as

$$
\begin{aligned}
S_{t+\Delta t} - S_t &= \frac{\theta^*}{\eta(1-\eta)}[\,C\,S^{\eta+1} + (D\eta - C)S_t\,]\,\Delta t \\
&\quad + \sqrt{\frac{2\,\theta^*}{(1-\eta)\eta}} \cdot \sqrt{S_t^{\eta+2} - S_t^2}\,(W_{t+\Delta t} - W_t)\,. \qquad (4.44)
\end{aligned}
$$

As $W_{t+\Delta t} - W_t \sim \mathcal{N}(0, \Delta t)$, the approximate transition densities in this case is given by

$$
S_{t+\Delta t} \,|\, S_t \sim \mathcal{N}(\,S_t + \frac{\theta^*}{\eta(1-\eta)}[\,C\,S^{\eta+1} + (D\eta - C)S_t\,]\,\Delta t,\;\; \frac{2\,\theta^*(S_t^{\eta+2} - S_t^2)}{\eta(1-\eta)}\,\Delta t\,)\,. \qquad (4.45)
$$

Therefore, the conditional log quasi-likelihood for a series of observations $\{S_t\}_{t=0}^{T}$ can be expressed as ($\Delta t = 1$)

$$
\begin{aligned}
\ln \mathcal{L}_\Theta &= \sum_{t=0}^{T-1} \ln \Pr{}^{\Theta} \,(S_{t+1} \,|\, S_t) \\
&= \sum_{t=0}^{T-1} -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln\frac{2\,\theta^*(S_t^{\eta+2} - S_t^2)}{\eta(1-\eta)} \\
&\qquad - \frac{[\,\eta(1-\eta)(S_{t+1} - S_t) - \theta^*(C\,S^{\eta+1} + (D\eta - C)S_t)\,]^2}{4\,\theta^*\,\eta(1-\eta)(S_t^{\eta+2} - S_t^2)}\,. \qquad (4.46)
\end{aligned}
$$

To distinguish from the case $\eta = 0$, we use $\Theta^e = \{\theta^*, \eta, C, D\}$ for the set of parameters and re-express eq.(4.46) as $\ln \mathcal{L}^e_{\Theta^e}$, The superscript *e* refers to the possibility that the dynamics of S_t and X_t can be explosive with the existence of a (ghost) finite-time-singularity.

The conditional quasi-likelihood maximization problem to determine the optimal parameters in each case can be written respectively

$$\max_{\Theta^{ne}} \ln \mathcal{L}^{ne}_{\Theta^{ne}} = \max_{\theta^*,C,D} \ln \mathcal{L}^{ne}(\{S_t\}_{t=0}^{t=T}, \theta^*, C, D) \qquad \text{conditional on } \eta = 0 \tag{4.47}$$

$$\max_{\Theta^{e}} \ln \mathcal{L}^{e}_{\Theta^{e}} = \max_{,\theta^*,\eta,C,D} \ln \mathcal{L}^{e}(\{S_t\}_{t=0}^{t=T},, \theta^*, \eta, C, D) \qquad \text{conditional on } \eta \in (0,1). \tag{4.48}$$

The corresponding variance for each estimated parameter can be obtained from the diagonal terms of the following Fisher information matrix

$$\mathbb{V}ar(\widehat{\Theta}^{ne}) = \left(\frac{\partial^2 \mathcal{L}^{ne}}{\partial\Theta^{ne}\partial\Theta^{ne}}\right)^{-1}\Bigg|_{\Theta^{ne}=\widehat{\Theta}^{ne}}, \qquad \text{conditional on } \eta = 0 \tag{4.49}$$

$$\mathbb{V}ar(\widehat{\Theta}^{e}) = \left(\frac{\partial^2 \mathcal{L}^{e}}{\partial\Theta^{e}\partial\Theta^{e}}\right)^{-1}\Bigg|_{\Theta^{e}=\widehat{\Theta}^{e}}, \qquad \text{conditional on } \eta \in (0,1). \tag{4.50}$$

Given a financial time series of daily close prices $\{S_t\}_{t=0}^{t=T}$, we use the L-BFGS-B method to optimize the log-likelihood. The limited-memory BFGS optimization algorithm belongs to the family of quasi-Newton methods that approximates the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm using a limited amount of computer memory. The L-BFGS-B extends L-BFGS to handle simple bound constraints in the search of the optimal parameters [39].

4.4.2 Calibration of synthetic price time series

In order to employ the L-BFGS-B optimization algorithm and taking into account computation efficiency, we explicitly assign a prior searching space for all four parameters $\{\theta^*, \eta, C, D\}$ before performing their estimation.

- For parameter θ^*, we assign its search to be in the interval $[10^{-8}, 0.004]$. Recall that θ^* has two roles. In the case $\eta = 0$, it is the inverse time horizon over which the momentum is supposed to be evaluated by traders. For $\eta \in (0,1)$, it is the characteristic momentum value above which a given realized momentum has an impact on the time horizon over which it is itself estimated (see expression (4.23) with (4.24)). Thus, the lower bound encompasses the cases where there is essentially an almost infinite memory (compared to the longest existing historical financial time series) of how past price changes influence momentum. 10^{-8} corresponds approximately to 10 million days≈ 400 thousand years. It also allows for a full feedback of momentum on its time horizon. For the upper bound, we are informed by a large body of literature mentioning that momentum trading typically make profit when using times scales from one to five years (See [121, 189] and the literature therein). We thus assign the upper search bound for θ^* to be $\frac{1}{250} = 0.004$, where one time unit corresponds to one trading day.

- For η, we search in what is essentially the whole possible interval $[10^{-8}, 1 - 10^{-8}]$, with the exclusion of tiny portions at the two boundaries for the sake of computational precision. We indeed want to avoid the divergence of the log likelihood function $\ln \mathcal{L}^e$ at the lower boundary $\eta = 0$.

- The search interval for parameter C is chosen to be $[0, 5]$. First, the condition that C is non-negative just means that a positive momentum tends to self-perpetuate itself (contrarian behaviors are excluded from the model in the present work). Since C has the meaning of an expected return per unit of momentum adjusted by the bubbliness η, the upper bound 5 is likely to be more than large enough: given some momentum, which reflects the belief by investors on the average price trend over a given time horizon modified by the inverse of the bubbliness $1 - \eta$, the coefficient C transforms this momentum into an actual average price that is $C\frac{1}{1-\eta}$ times larger over the same time horizon; it is thus reasonable that C should be larger than a few units. Thus the upper bound of 5 is largely sufficient.

- Last, the search interval for parameter D is chosen to be $[0, 500]$. Given that $D = \frac{d(1-\eta)}{\theta^*}$, the upper bound D is amply sufficient as it would correspond to more than 100% return in one time unit (one day in the subsequent synthetic time series and in the real life applications below) for momentum trading strategy with a typical time-horizon memory of $\frac{1}{0.002} = 200$ days during the bubble with only small bubbliness.

We test our calibration algorithm on synthetic simulated price time series. For $\eta = 0$ and $\eta \in (0, 1)$, tables 4.1 and 4.2 respectively list the true parameters used to generate the price time series, the values of calibrated parameters, as well as the values of the logarithmic quasi-likelihood calculated for each simulated series with those parameters. Table 4.1 gives the results for 4 synthetic time series shown in fig. 4.3. Table 4.2 gives the results for 8 synthetic time series with positive bubbliness ($\eta \in (0, 1)$) shown in fig.4.4, which cover the different regimes for C and D and contrasts the cases $\eta = 0$ and $\eta \in (0, 1)$.

C	$\widehat{C}$	D	$\widehat{D}$	θ^*	$\widehat{\theta^*}$	$-2\ \ln \mathcal{L}^{ne}$	$-2\ \ln \widehat{\mathcal{L}^{ne}}$
0.5	1.003 (0.905)	0.5	1.038 (0.702)	0.002	0.0021 (¡0.0001)	358.6	355.9
0.5	1.068 (0.723)	1.5	1.058 (1.297)	0.002	0.0021 (¡0.0001)	3842	3840
1.5	1.284 (0.587)	0.5	1.119 (0.917)	0.002	0.0022 (¡0.0001)	4696	4693
1.5	1.864 (0.428)	1.5	1.372 (1.282)	0.002	0.0024 (¡0.0001)	10640	10639

Table 4.1: Estimation of the 4 synthetic series shown in figure 4.3 with $\eta = 0$. The true parameter values used to generate the synthetic time series are indicated without the hat ˆ. The estimated parameters are shown with the ˆ. The values in brackets below the estimated values give the corresponding standard errors.

η	$\widehat{\eta}$	C	$\widehat{C}$	D	$\widehat{D}$	θ^*	$\widehat{\theta^*}$	$-2\ \ln \mathcal{L}^{ne}$	$-2\ \ln \widehat{\mathcal{L}^{ne}}$
0.9	0.936	0.5	0.194	0.5	0.127	0.002	0.004	769.5	355.3
	(0.005)		(0.168)		(0.008)		(0.0002)		
0.9	0.873	0.5	0.319	1.5	0.923	0.002	0.004	2094	1996
	(0.010)		(0.171)		(0.186)		(0.0003)		
0.9	0.942	1.5	0.694	0.5	0.084	0.002	0.004	247.3	187.9
	(0.009)		(0.470)		(0.013)		(0.0005)		
0.9	0.893	1.5	0.499	1.5	0.750	0.002	0.004	608.1	564.7
	(0.019)		(0.470)		(0.358)		(0.0007)		
0.5	0.224	0.5	0.294	0.5	0.140	0.002	0.004	2760	2472
	(0.05)		(0.849)		(0.018)		(0.0003)		
0.5	0.464	0.5	0.910	1.5	1.109	0.002	0.002	479.5	480.8
	(0.019)		(0.543)		(0.547)		(¡0.0001)		
0.5	0.384	1.5	0.989	0.5	0.098	0.002	0.004	3357	3243
	(0.022)		(0.228)		(0.029)		(0.0002)		
0.5	0.588	1.5	1.703	1.5	1.224	0.002	0.002	229.0	109.0
	(0.020)		(0.417)		(0.581)		(¡0.0001)		

Table 4.2: Estimation of 8 synthetic series shown in fig. 4.4. The true parameter values used to generate the synthetic time series are indicated without the hat ˆ. The estimated parameters are shown with the ˆ. The values in brackets below the estimated values give the corresponding standard errors.

From the above two tables, we can draw the following conclusions concerning the performance of our conditional quasi-likelihood estimation method.

- The estimation of θ^* conditional on $\eta = 0$ (non-bubble or mild-bubble cases) is quite accurate. The estimation of θ^* conditional on $\eta \in (0, 1)$ (recurrent or wild bubbles) is less precise but still in a reasonable range. In particular, for moderate bubbliness ($\eta = 0.5$) and strictly positive momenta ($D \in [1, \infty)$), the estimation of θ^* is good. For other bubble types, η is overestimated. It compensates the underestimation of both C and D, at the cost of an overestimation of the volatility (recall eq.(4.38) and eq. (4.40)), as we will discuss soon.

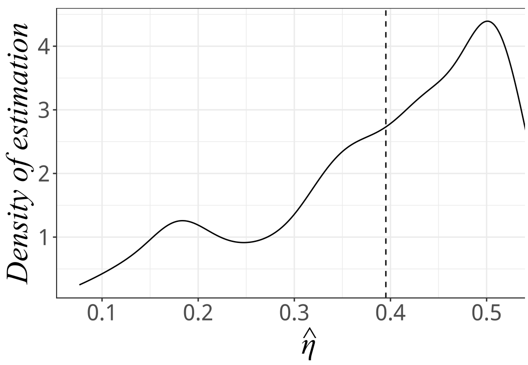
- The estimation of η looks quite reliable for both recurrent or wild bubbles, except for the cases with $\eta = 0.5$ and $D \in (0, 1)$, for which η are underestimated. Nonetheless, the estimator is able to identify correctly the range of the bubbliness, as the estimations of true values $\eta = 0.5$ are all well separated from those with true values $\eta = 0.9$.

- The estimations of C are quite far from their true values for all bubble types. Moreover, the standard deviation of these estimations are quite large, so that the estimations of C fall within two standard deviations of the true value. In other words, given the rather large standard deviations, the estimated values can be deemed compatible with the true values. It is encouraging to observe that the larger the true value of C, the largest its estimator, inter alia. The source of the discrepancy between estimated and true values of C is likely to be found in the Euler discretization scheme, which has used a time step $\Delta t = 1$ to match the daily observation scale. With this choice $\Delta t = 1$, we are quite far from the consistency condition that the Euler discretization converges in the limit $\Delta t \to 0$. For the stronger bubbles ($\eta \in (0, 1)$), the parameters are underestimated (except for $\eta = 0.5$ and $D \in (0, 1)$ for which the parameters are slightly overestimated), which leads to an underestimation of the real strength of the positive feedbacks between momentum and time scale. Actually, as the drift and diffusion coefficients have θ^* as a common parameter, the quasi-likelihood approach tend to give more weight on such a shared parameter, which thus leads to an underestimation of C.

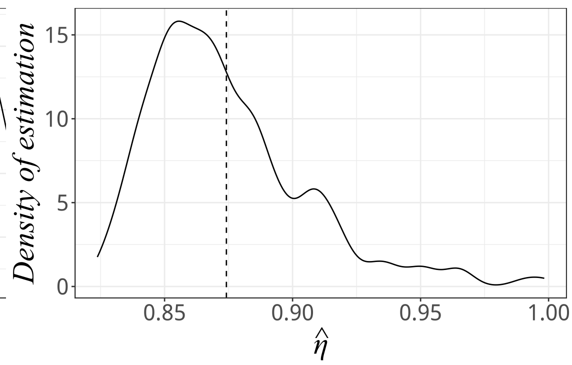
- For the non-bubble and mild-bubble cases, the estimations of D fall within two standard deviations of the true value. As for C, the larger the true value of C, the largest its estimator, inter alia. The origin of the discrepancy between estimated and true values of D is also likely to be found in the Euler discretization scheme, as discussed for C. For the stronger bubbles, the estimation biases are approximately the same, around -0.4 for all cases.

In order to investigate in detail the performance of this estimation method based on Euler quasi-likelihood approximation, the empirical density of the estimated value of each parameter is created by repeating the estimation procedure on an ensemble of 200 synthetic time series generated with fixed specific model parameters but different random seeds. The following figures 4.7 -4.10 summarize the results. The solid line shows the empirical density function, and the vertical dashed line gives the empirical first moment (statistical mean value).

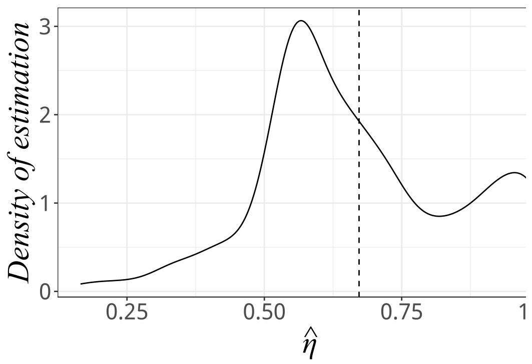
Figure 4.7 confirms the relatively good performance of the estimation of η, with the empirical density function of $\hat{\eta}$ peaking close the true value. However, the empirical density function is quite broad, exhibiting a significant half-width of about $0.1 - 0.2$, which is much larger than the standard error $0.01 \sim 0.03$ estimated on each calibration using the Fisher information matrix, as reported in figure 4.8. Moreover, one can observe a significant skewness and a fat-tailed structure of the density functions. This reflects the strong sensitivity of the estimation of η to the specific realization of the stochastic innovations associated with a given price path trajectory, such that the ensemble standard deviation is much larger than the single realization standard deviation. This phenomenon is typical of time series without ergodicity, which is the case here given the relatively small sizes of time windows and the nature of the bubble dynamics in our model.



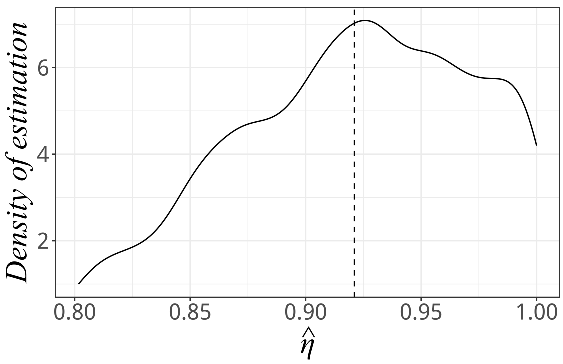
(a) $C = 0.5,\ \eta = 0.5$



(b) $C = 0.5,\ \eta = 0.9$

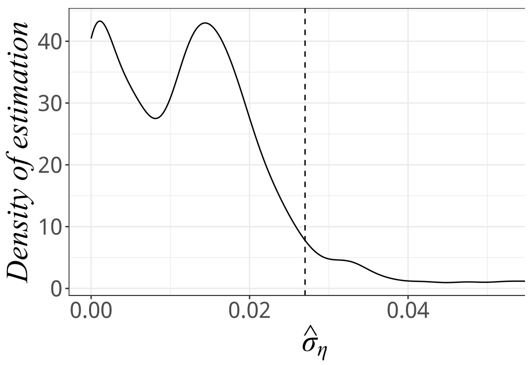


(c) $C = 1.5,\ \eta = 0.5$

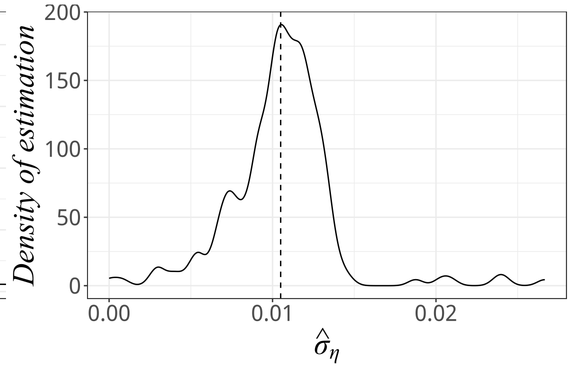


(d) $C = 1.5,\ \eta = 0.9$

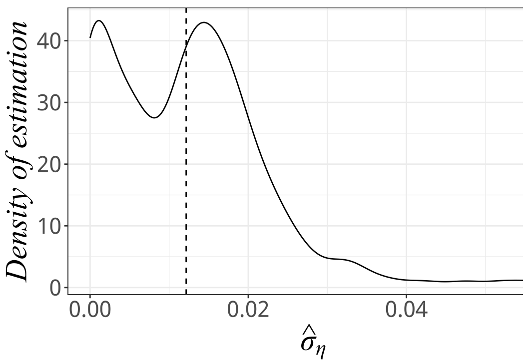
Figure 4.7: Empirical density functions of estimated η on an ensemble of 200 synthetic time series with different random seeds, generated with fixed specific model parameters $D = 1.5$, $\theta^* = 0.002$ and C indicated below each graph. The vertical dashed line gives the empirical first moment (statistical mean value).



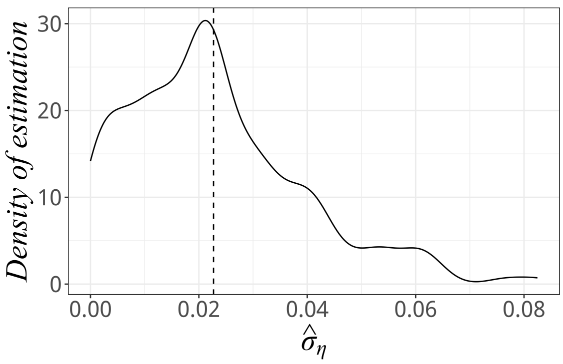
(a) $C = 0.5,\ \eta = 0.5$



(b) $C = 0.5,\ \eta = 0.9$



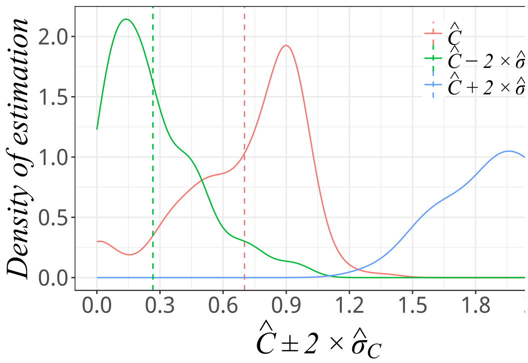
(c) $C = 1.5,\ \eta = 0.5$



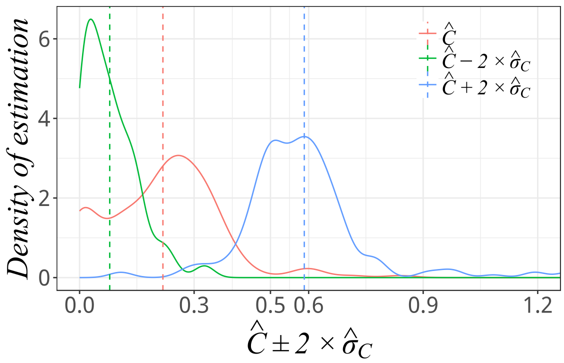
(d) $C = 1.5,\ \eta = 0.9$

Figure 4.8: Empirical density functions of estimated standard deviations of η obtained from the Fisher information matrix for each fit within an ensemble of 200 synthetic time series with different random seeds, generated with fixed specific model parameters $D = 1.5$, $\theta^* = 0.002$ and C indicated below each graph. The vertical dashed line gives the empirical first moment (statistical mean value) of the estimated standard deviations of η.

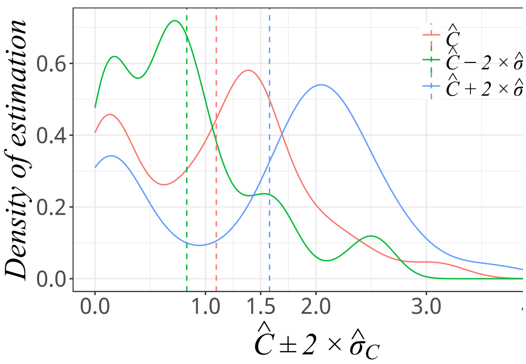
Figure 4.9 shows that smaller (resp. larger) values of bubbliness tend to yield an over-estimation (resp. under-estimation) of C. The shown empirical density functions of C over an ensemble of 200 synthetic time series are quite broad with half-widths comparable to the Fisher matrix based estimation of the standard deviation for each individual time series, as illustrated by the empirical density functions of C, corrected, respectively, by $+$ and $-$ two times the standard deviation of C obtained from the estimated Fisher matrix for each calibration. Figure 4.10 shows that the modes and means of the empirical density functions of D are smaller than the true value for all shown values of C and η parameters. The dispersion of the estimated D is also quite broad and similar across the ensemble of synthetic time series, as estimated for each time series with the Fisher information matrix. In all cases, the estimated values of both C and D fall within two standard deviations of their true values.



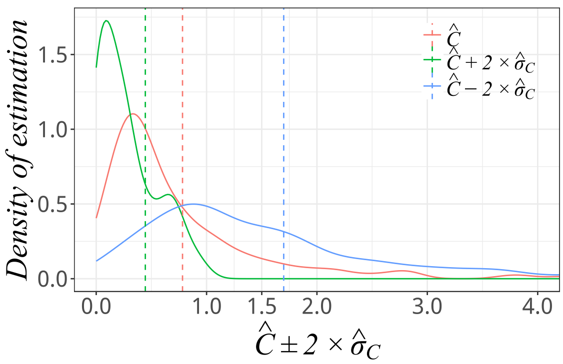
(a) $C = 0.5, \ \eta = 0.5$



(b) $C = 0.5, \ \eta = 0.9$

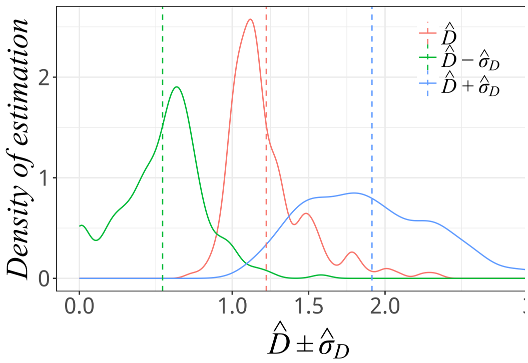


(c) $C = 1.5, \ \eta = 0.5$

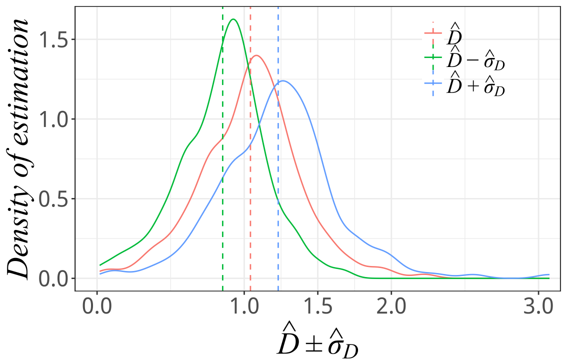


(d) $C = 1.5, \ \eta = 0.9$

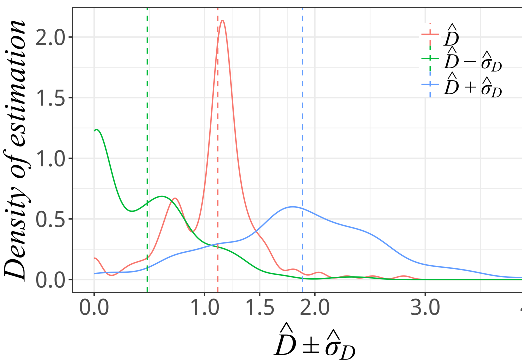
Figure 4.9: Empirical density functions of C over an ensemble of 200 synthetic time series with different random seeds, generated with fixed model parameters $D = 1.5$, $\theta^* = 0.002$ and C and η as indicated below each graph. As indicated in the graphs, the two other curves are the empirical density functions of C corrected respectively by $+$ and $-$ two times the standard deviation of C obtained from the estimated Fisher matrix for each calibration. For each of the three obtained density functions, the vertical dashed line gives its statistical mean value.



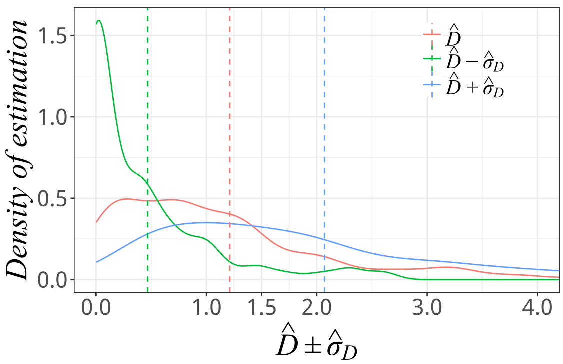
(a) $C = 0.5, \ \eta = 0.5$



(b) $C = 0.5, \ \eta = 0.9$



(c) $C = 1.5, \ \eta = 0.5$



(d) $C = 1.5, \ \eta = 0.9$

Figure 4.10: Empirical density functions of D over an ensemble of 200 synthetic time series with different random seeds, generated with the same fixed specific model parameters $D = 1.5$, $\theta^* = 0.002$ and C and η as indicated below each graph. As indicated in the graphs, the two other curves are the empirical density functions of D corrected respectively by $+$ and $-$ two times the standard deviation of D obtained from the estimated Fisher matrix for each calibration. For each of the three obtained density functions, the vertical dashed line gives its statistical mean value.

4.4.3 Binary classification of recurrent ($C < 1$) versus wild ($C > 1$) bubbles

Given the biases and broad statistical distributions of estimated parameters C for $\eta \in (0, 1)$, we seek to reliably characterize a given time series as being either in the recurrent ($C \in (0, 1)$) or wild ($C \in (1, \infty)$) bubble class, using some combination of estimated parameters.

To achieve this, we construct a classifier metric defined by $\widehat{C} + h \cdot \hat{\sigma}_C$, where h is a parameter to be determined to ensure maximum sensitivity and specificity of the resulting classification, as defined below. The proposed classification rule is shown in figure 4.11.

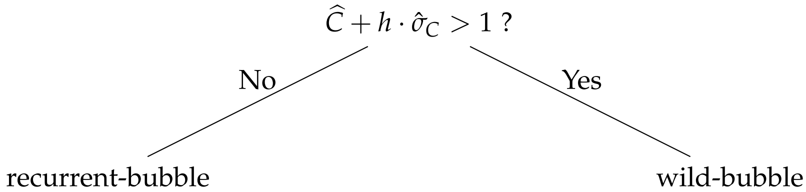


Figure 4.11: Classification scheme to decide if a given time series is in the recurrent or wild bubble regime (conditional of $\eta \in (0, 1)$). The parameter h is chosen to optimize sensitivity and specificity properties (see text).

We consider two populations of 200 time series of 1000 time steps each, with 200 different random seeds. Both populations share the same initial momentum and parameters $\theta^* = 0.002, \eta = 0.9, D = 1.5$. The two populations are distinguished by one having $C = 0.5$ (recurrent bubbles) and the other having $C = 1.5$ (wild bubbles).

For each value of h, we estimate the probability $\Pr_h(\text{No}|R)$ that a time series will be qualified as a recurrent bubble (answer 'No' to the question in figure 4.11), given that the time series has been generated in the recurrent bubble regime $C = 0.5 < 1$. $\Pr_h(\text{No}|R)$ is determined as the fraction among the 200 time series generated with $C = 0.5$ that answer 'No' to the question in figure 4.11 for that given h value. Similarly, for each value of h, we estimate the probability $\Pr_h(\text{No}|W)$ that a time series will be qualified as a recurrent bubble (answer 'No' to the question in figure 4.11), given that the time series has been generated in the wild bubble regime $C = 1.5 > 1$. $\Pr_h(\text{No}|W)$ is determined as the fraction among the 200 time series generated with $C = 1.5$ that answer 'No' to the question in figure 4.11 for that given h value.

Since the classification proposed in figure 4.11 is binary, we have

$$\Pr_h(\text{Yes}|R) = 1 - \Pr_h(\text{No}|R) \quad \text{and} \quad \Pr_h(\text{Yes}|W) = 1 - \Pr_h(\text{No}|W) \,. \tag{4.51}$$

Knowing $\Pr_h(\text{No}|R)$ and $\Pr_h(\text{No}|W)$, for an unknown time series, the probability $\Pr(R|\text{No})$ that it is a recurrent bubble, given that the answer to the question in figure 4.11 is 'No' is obtained as

$$\Pr_h(R|\text{No}) = \frac{\Pr_h(\text{No}|R)\ p(R)}{\Pr_h(\text{No}|R)\ p(R) + \Pr_h(\text{No}|W)\ p(W)} \,, \tag{4.52}$$

where $p(R)$ (resp. $p(W)$ is the unconditional probability that the time series is a recurrent (resp. wild) bubble. The probability $\Pr_h(W|\text{No})$ that it is a wild bubble, given that the answer to the question in figure 4.11 is 'No' is then

$$\Pr_h(W|\text{No}) = 1 - \Pr_h(R|\text{No}) \ , \tag{4.53}$$

since, for a given answer 'No', the state of the time series is binary, either recurrent or wild.

Similarly, using $\Pr_h(\text{Yes}|R)$ and $\Pr_h(\text{Yes}|W)$ given in (4.51), we can derive $\Pr_h(R|\text{Yes})$, the probability that the time series is in a recurrent bubble, given that the answer to the question in figure 4.11 is 'Yes':

$$\Pr_h(R|\text{Yes}) = \frac{\Pr_h(\text{Yes}|R)\ p(R)}{\Pr_h(\text{Yes}|R)\ p(R) + \Pr_h(\text{Yes}|W) p(W)} \ , \tag{4.54}$$

where $\Pr_h(\text{Yes}|R)$ and $\Pr_h(\text{Yes}|W)$ are known from (4.51). The probability $\Pr(W|\text{Yes})$ of being a wild bubble, given that the answer to the question in figure 4.11 is 'Yes', then reads

$$\Pr_h(W|\text{Yes}) = 1 - \Pr_h(R|\text{Yes}) \ , \tag{4.55}$$

since, for a given answer 'Yes', the state of the time series is binary, either recurrent or wild.

As an illustration, assuming that we have no prior information on the relative frequencies of recurrent vs wild bubbles, we assign the no-informative priors $p(R) = p(W) = 1/2$ and expressions (4.52) and (4.54) reduce to

$$\Pr_h(R|\text{No}) = 1 - \Pr_h(W|\text{No}) = \frac{\Pr_h(\text{No}|R)}{\Pr_h(\text{No}|R) + \Pr_h(\text{No}|W)} \ , \tag{4.56}$$

and

$$\Pr_h(R|\text{Yes}) = 1 - \Pr_h(W|\text{Yes}) = \frac{\Pr_h(\text{Yes}|R)}{\Pr_h(\text{Yes}|R) + \Pr_h(\text{Yes}|W)} \ . \tag{4.57}$$

For instance, let us assume $\Pr_h(\text{No}|R) = 0.7$ (true positive rate of 70% for recurrent bubbles) and $\Pr_h(\text{Yes}|R) = 0.3$ (false negative rate of 30% for recurrent bubbles) and $\Pr_h(\text{No}|W) = 0.2$ (false negative rate of 20% for wild bubbles) and $\Pr_h(\text{Yes}|W) = 0.8$ (true positive rate of 80% for wild bubbles). Expressions (4.56) and (4.57) yields

$$\Pr_h(R|\text{No}) = 78\% \ ; \ \Pr_h(R|\text{Yes}) = 27\% \ ; \ \Pr_h(W|\text{No}) = 22\% \ ; \ \Pr_h(W|\text{Yes}) = 73\% \ . \tag{4.58}$$

Thus, if the answer is 'No' and this is the only information we have, we will attribute a probability of 78% that the time series is recurrent and 22% that it is wild. If the answer is 'Yes' and this is the only information we have, we will attribute a probability of 73% that the time series is wild and 27% that it is recurrent.

In sum, the complete information of the binary classifier defined in figure 4.11 for a given h value is contained in just two conditional probabilities $\Pr_h(\text{No}|R)$ and $\Pr_h(\text{No}|W)$, which are plotted as a function of h in figure 4.12. These two probabilities are obtained as explained above from two populations of 200 time series of 1000 time steps each, with 200 different random seeds, one population being in the recurrent bubble regime, while the other one is in the wild bubble regime. One can observe that the true positive rate $\Pr_h(\text{No}|R)$ for recurrent bubbles is all the large, the smaller h is. In contrast, the false negative rate $\Pr_h(\text{No}|W)$ for wild bubble decreases with h. This leads to conflicting requirements: to have the largest possible $\Pr_h(\text{No}|R)$, we need h small, while to have the smallest possible $\Pr_h(\text{No}|W)$, we need to have a large h. We can thus expect that the correct identification rates $\Pr_h(R|\text{No})$ and $\Pr_h(W|\text{Yes})$ depends nonlinearly and non-monotonously on h. Indeed, figure 4.13 shows that choosing $h = 0$ would lead to a perfect classification score $\Pr_h(W|\text{Yes}) = 1$ for the identification of wild bubbles at the cost of a modest performance $\Pr_h(R|\text{No}) = 0.57$ for the identification of recurrent bubbles. Choosing $h = 3.5$ gives the maximum possible score $\Pr_h(R|\text{No}) = 0.8$ for the identification of recurrent bubbles together with a quite honorable $\Pr_h(W|\text{Yes}) = 0.85$ for the identification of a wild bubble.

As h can be varied, the optimal classification scheme applied to an unknown time series is the following. One should first try the value $h = 0$ and check what is the answer to the question in figure 4.11 for $h = 0$. If it is 'Yes', then we know with 100% certainty that the time series is in the wild bubble regime. If the answer is 'No' for $h = 0$, one idea is to choose $h = 3.5$ to ask again the question in figure 4.11. If the answer is 'No' with $h = 3.5$, we conclude that the time series is recurrent with 80% probability. If the answer is 'Yes' with $h = 3.5$, we deduce that the time series is wild with 85% probability. We can actually do better. If the answer is 'No' for $h = 0$, we can scan all the values of h from 0 to 3.5 to check if there is a value $h < 3.5$ for which the answer is 'Yes'. Since $\Pr_h(W|\text{Yes})$ is monotonically decreasing from 1 for $h = 0$ to 0.85 for $h = 3.5$, the first time we get a 'Yes' for the smallest $h < 3.5$ gives us a success probability $\Pr_h(W|\text{Yes})$ larger than 85%. If there is no 'Yes' found for $h < 3.5$, the best conclusion is that the time series is recurrent with a probability of 80%.

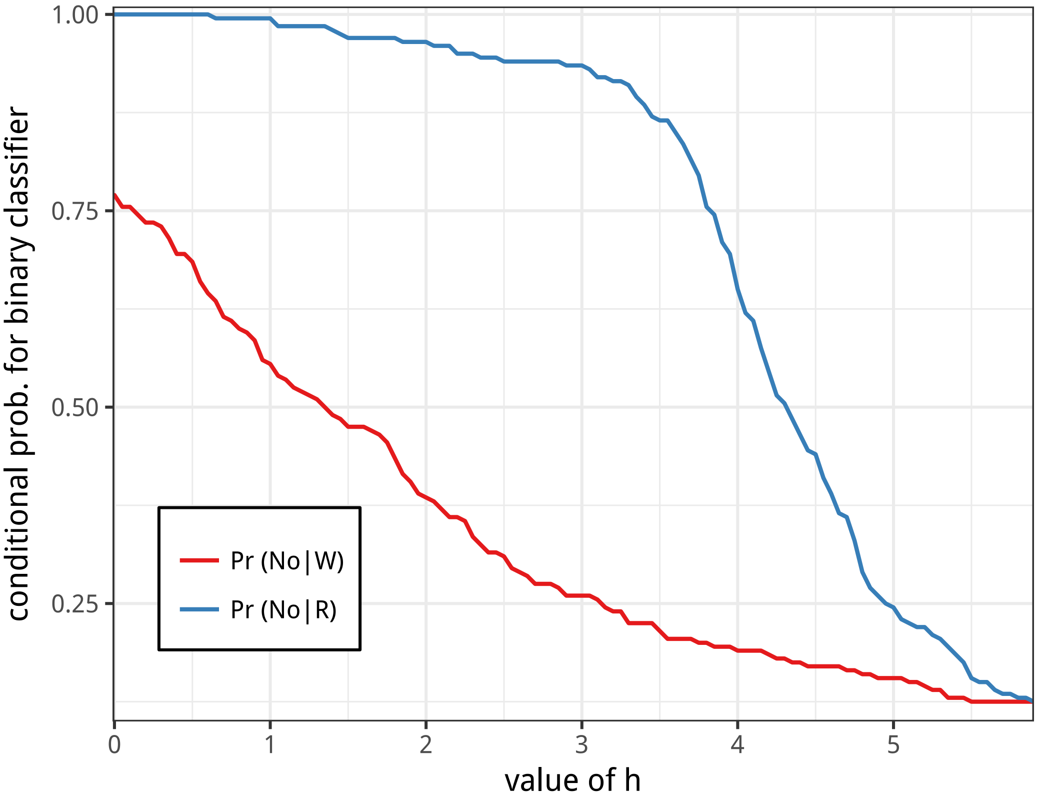


Figure 4.12: The two conditional probabilities, $\Pr_h(\text{No}|W)$ (true positive) and $\Pr_h(\text{No}|R)$ (false negative), as a function of h.

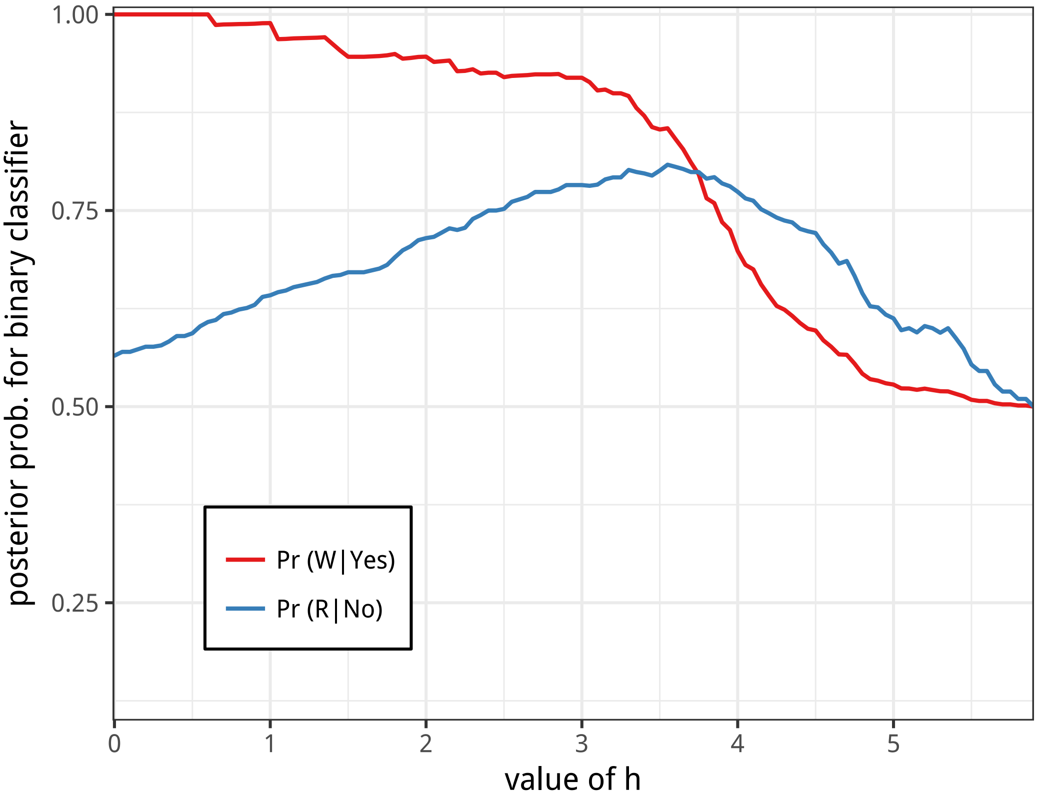


Figure 4.13: The two derived posterior probabilities $\Pr_h(R|\text{No})$ and $\Pr_h(W|\text{Yes})$ given by expressions (4.52) and (4.55) as a function of h, with the assumption that both of the unknown priori probabilities $p(R) = p(W) = 1/2$. $\Pr_h(R|\text{No})$ (resp. $\Pr_h(W|\text{Yes})$) is the probability or success score for the correct identification of a recurrent (resp. wild) bubble, given that the answer to the question figure 4.11 is 'No' (resp. 'Yes') for a given h.

4.4.4 Bubble detection based on Wilks' Test

Wilks' Theorem [251] provides a simple and efficient test to compare two nested models using their log-likelihood ratio. For the proposed time-varying momentum horizon bubble model, the bubble type of $\eta = 0$ with dynamics given by (4.38) is indeed nested into the model characterized by (4.40) with $\eta \in [0, 1)$. Given a price time series, we can thus subject it to the Wilks' test with the following null and alternative hypotheses, based on the :

H_0 : $\eta = 0$, no bubble or mild-bubble (without finite-time-singularity).

H_1 : $\eta \in (0, 1)$, recurrent or wild bubble (with ghost- or true finite-time-singularity and likely decline/regime-change).

The Wilks statistics $T_{ne:e}$ is usually defined as

$$T_{ne:e} := -2 \left(\max_{\theta^*, C, D} \ln \mathcal{L}^{ne} - \max_{\eta, \theta^*, C, D} \ln \mathcal{L}^{e} \right) . \tag{4.59}$$

Under the hypothesis that the null H_0 holds, $T_{ne:e}$ is distributed according to the chi-square distribution with a number of degrees of freedom equal to the number of parameters that needs to be fixed to go from H_1 to H_0, here 1, since the parameter space for H_0, i.e., $(\theta^*, C, D) \in \mathbb{R}^{+3}$ lies on a 3-dimensional hyperplane within the 4 dimensional space represented by $(\eta, \theta^*, C, D) \in [0, 1) \otimes \mathbb{R}^{+3}$. However, there is a correction needed to account for the fact that H_0 is recovered from H_1 by fixing η on the boundary 0 of the interval $\eta \in [0, 1)$. In this case, rather than $T_{ne:e}$ being distributed according to $\chi^2(1)$, the distribution of $T_{ne:e}$ needs to be adjusted (see [49])

$$T_{ne:e} := -2 \left(\max_{\theta^*, C, D} \ln \mathcal{L}^{ne} - \max_{\eta, \theta^*, C, D} \ln \mathcal{L}^{e} \right) \sim Mixed(0, \chi^2(1); \frac{1}{2}) . \tag{4.60}$$

where $Mixed(0, \chi^2(1); \frac{1}{2})$ denote the mixed law such that the statistic is 0 with one chance out of two and follows a $\chi^2(1)$ with the other one chance out of two. Under the null hypothesis H_0, this statistic T should be distributed according to the above mixed distribution: the number of parameters to fix in the general model (4.40) to arrive at the particular model (4.38) in the nested hypothesis testing is 1 (parameter η), while $\eta = 0$ is on the border of $\eta \in [0, 1)$. Thus, the $q-$quantile for $\chi^2(1)$ is equivalent to the $\frac{1+q}{2}-$quantile for the distribution of $T_{ne:e}$.

Figure 4.14 and fig. 4.15 illustrate the results obtained by implementing the Wilks' test on synthetic price time series to examine the capability of differentiating the regimes $\eta = 0$ and $\eta \in (0, 1)$. As shown in fig. 4.14, for all 400 simulated price time series generated with $\eta \in (0, 1)$, the null H_0 ($\eta = 0$) can be rejected with significance level of 2.5%. By and large, for the time series with larger bubbliness (larger η), the null H_0 is easier to reject, regardless of whether they are recurrent or wild bubbles. For bubble with moderate bubbliness (left panel with $\eta = 0.5$), the null H_0 is easier to reject for the wild bubbles (triangle) than for the recurrent bubbles (small circle). For bubbles with strong bubbliness (right panel with $\eta = 0.9$), the easier rejection of the null for wild bubbles than for recurrent bubble can be attributed to the explosive behavior and short duration of some of the wild bubble time series.

As shown in fig. 4.15 for time series generated with $\eta = 0$, the Wilks' test performs well as there is only two false alarm at the 97.5% confidence level and, for all other time series, the null cannot be rejected.

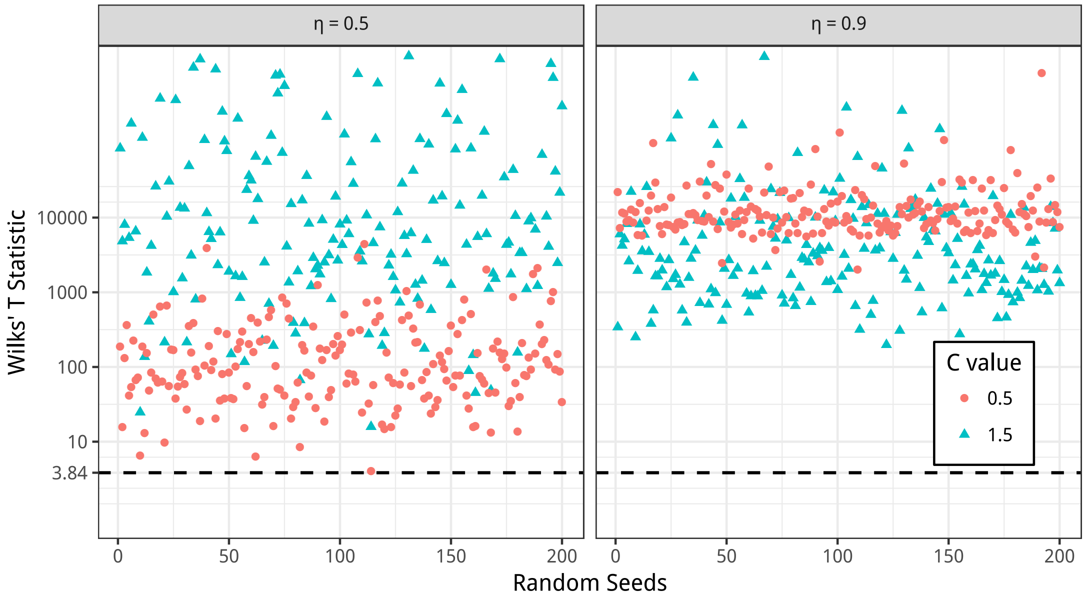


Figure 4.14: Wilks' statistics (4.59) calculated for synthetic price time series generated for two types of strong bubbles, with $\eta \in (0, 1)$. There are 400 time series in total. All time series are generated with the fixed parameter values $\theta^* = 0.002, D = 1.5$. For each pair (η, C), there are 100 synthetic time series to test. The number 3.84 correspond to the 95%$-$quantile for the $\chi^2(1)$ distribution and indicates the 97.5% confidence level to reject H_0.

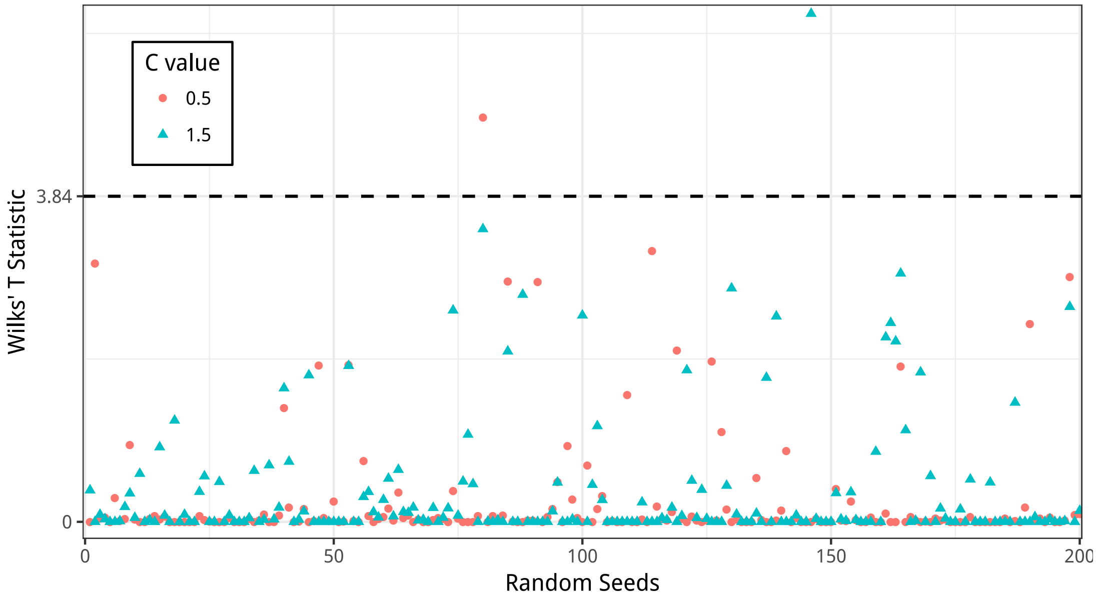


Figure 4.15: Wilks' statistics (4.59) calculated on synthetic price time series generated with $\eta = 0$. There are 200 time series in total. All time series are generated with fixed parameter values $\theta^* = 0.002, D = 1.5$. For each value of C, there are 100 synthetic time series to test. The number 3.84 corresponds to the 95%−quantile for the $\chi^2(1)$ distribution and indicates the 97.5% confidence level to reject H_0.

We implement the test on two historical financial price trajectories. They are indices constructed with the 505 stock prices that entered in the composition of the S&P 500 stock index from the beginning of 1998 to the end of 2002. The first Internet stock index is an equally weighted portfolio of the firms related to the Internet sector according to GICS. The second "brick and mortar" index is an equally weighted portfolio of the remaining "brick and mortar" firms listed as S&P 500 components. Figure 4.16 shows that the "brick and mortar" index fluctuated by not more than ±25% over the 5 years. In contrast, the Internet stock index was multiplied by a factor more than 11 (corresponding to a return of more than > 1100% from 1998 to its peak in the first quarter of 2000). This index then shrunk with a great crash from a value above 150 to below 50 is March-April 2000, followed by a jumpy decay to achieve a level at the end of 2002 below its value in January 1998. The contrast between the behavior of these two indices over the same 5 years horizon cannot be more striking.

Table 4.3 demonstrates that the Wilks' test can cleanly classify the two indices into two different classes. For the Internet index, H_0 is very strongly rejected, while it is not for the "brick and mortar" stock index.

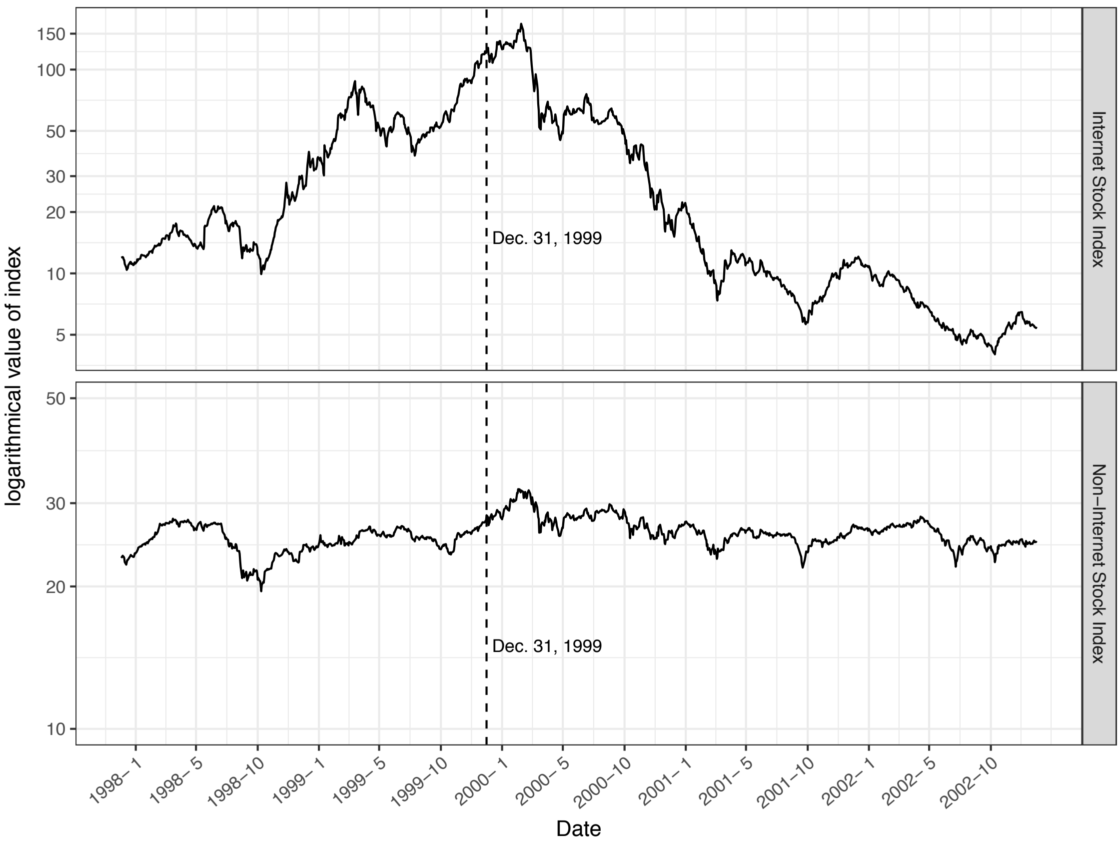


Figure 4.16: Logarithm of the price level of the Internet (top panel) and of the "brick and mortar" (bottom panel) stock indices. The time periods goes from the beginning of 1998 to the end of 2002. Note the difference in vertical scales in the two plots.

	$\eta = 0$				$\eta > 0$					$T_{ne:e}$	$p-$Value
	C	D	θ^*	$-\mathcal{L}^{ne}$	η	C	D	θ^*	$-\mathcal{L}^{e}$		
Internet	0.5	< 0.01	5.32×10^{-4}	1663.4	0.281	0.719	0.538	1.31×10^{-4}	1567.7	191.5	< 0.001
Non-internet	0.5	4.84	2.06×10^{-5}	235.08	< 0.01	0.442	5.98	2.06×10^{-5}	235.08	0.017	0.948

Table 4.3: Estimation of the two models $\eta = 0$ and $\eta \in (0, 1)$ and Wilk's test on two stock indices during the dot-com bubble and following crash from 1998 to 2002.

Table 4.4 is the same as Table 4.3, with the difference of restricting the time interval from the beginning of 1998 to the end of 1999, i.e. only during the run-up of the bubble and before the crash occurs and following bearish regime develops. This is important to disentangle the impact of the bubble regime that has not yet revealed the underlying finite time singularity from the crash regime in distinguishing the two classes of dynamics. Again, we find that, for the Internet index, the null hypothesis H_0 is rejected at the 95% confidence level (with a p-value of 0.018). In contrast, the null is not rejected for the "brick and mortar" stock index.

	The class of $\eta = 0$				The class of $\eta > 0$					$T_{ne:e}$	$p-$Value
	C	D	θ^*	$-\mathcal{L}^{ne}$	η	C	D	θ^*	$-\mathcal{L}^{e}$		
Internet	0.007	10.54	3.7×10^{-4}	858.9	0.099	0.452	22.5	2.4×10^{-4}	856.7	4.42	0.018
Non-internet	5.210	12.33	1.51×10^{-5}	5.653	< 0.001	4.75	10.54	1.51×10^{-5}	5.648	0.01	0.960

Table 4.4: Same as table 4.3 from the beginning of 1998 to the end of 1999.

4.5 Concluding remarks

In this work, we have introduced a reduced form model of financial bubbles, based on a simple specification of the stochastic dynamics of the price momentum of a financial asset. The model assumes that the price is a deterministic function of the instantaneous momentum, a quantity that constitutes one of the most popular technical indicators powering investment techniques used by hedge-funds, mutual funds and individual investors in general. In our model, the instantaneous momentum was constructed by a simple continuous moving average of past prices over a given time horizon. One key feature is to consider that this time horizon is not fixed but tends to shrink as a bullish market develops, reflecting the race of trend-following agents to forerun their competitors. The feedback of momentum onto its time horizon, namely the stronger the momentum, the shorter the time horizon, leads to several interesting market dynamics that can be classified in terms of an index η of bubbliness and of the dependence of stock returns on momentum. We have identified four main classes of market behavior: non-bubble regime (an exponential of a CIR process), a mild bubble regime characterized by a transient exponential of an exponential growth of the price, a recurrent bubble regime with transient stochastic ghost-singularity and a wild bubble regime with a genuine stochastic finite-time singularity. We have shown how to test for the presence of these different regimes through intensive tests on synthetic time series. Finally, we have presented a bubble detection test based on our model, which we have applied to the Internet bubble regime. This has allowed us to confirm the reasonable performance of our proposed calibration method to qualify the Internet stocks as being described by the recurrent bubble regime compared with the non-Internet stocks well characterized by the non-bubble class.

We end by pointing out two limits of our model, which should be addressed in future work. First, since only positive momenta are allowed, the model does not describe negative bubbles characterized by downward accelerating prices spiraling towards a bottom followed by a rebound or rally. Second, the model assumes representative momentum investors with a single typical time scale. This should be improved in the future by accounting for the multiple time scales characterizing the universe of trading strategies, from high-frequency trading, intra-day trading, to long term investment horizons. Re-introducing these multiple time scales can offer in principle the possibility to account for price jumps.

4.6 Appendix

4.6.1 Proofs

Lemma 4.6.1. *Assume the setting in section 4.2.2 and let $\lambda : [0,\infty) \to [0,\infty)$ be a C^2-function with strictly positive first and second derivative such that $S_t = \lambda(X_t), t \in [0,\infty)$. Then it holds that*

$$\theta_t = \frac{\lambda(X_t)}{\lambda'(X_t)} \quad \textit{and} \quad g_t^2 = \frac{2\lambda(X_t)X_t}{\theta_t^2 \lambda''(X_t)}, \quad \textit{for } t \in [0,\infty). \tag{4.61}$$

Proof of lemma 4.6.1. Equations (4.11) and (4.12) imply that the quadratic variation of X is given by

$$[X]_t = \theta_t^2 g_t^2 dt. \tag{4.62}$$

Assume that S satisfies $S_t = \lambda(X_t)$ for $t \in [0,\infty)$. Then Itô's formula shows that

$$\begin{aligned} dS_t &= \lambda'(X_t)dX_t + \frac{1}{2}\lambda''(X_t)d[X]_t \\ &= \lambda'(X_t)dX_t + \frac{1}{2}\lambda''(X_t)\theta_t^2 g_t^2 dt \end{aligned} \tag{4.63}$$

Rearranging equation (4.12) we get

$$dS_t = \frac{S_t}{\theta_t}dX_t + S_t X_t dt, \tag{4.64}$$

and comparing the ($\mathbb{P}$-a.s.unique) Itô-equation terms in (4.63) and (4.64), we conclude that

$$\theta_t = \frac{\lambda(X_t)}{\lambda'(X_t)} \quad \text{and} \quad g_t^2 = \frac{2\lambda(X_t)X_t}{\theta_t^2 \lambda''(X_t)}. \tag{4.65}$$

The proof of lemma 4.6.1 is thus completed. □

Lemma 4.6.2. *Assume the setting in section 4.3.1 and let $\beta < \infty$. Then the process X is given by the SDE*

$$dX_t = \left(\theta^* + \frac{X_t}{\beta}\right)\left(X_t\left(c-1\right)+d\right)dt + \left(\theta^* + \frac{X_t}{\beta}\right)\sqrt{\frac{2\beta X_t}{\beta-1}}dW_t \tag{4.66}$$

and it holds that

(a) X is explosive for $c \in \left(\frac{\beta}{\beta-1}, \infty\right)$ *and recurrent for* $c \in \left[0, \frac{\beta}{\beta-1}\right]$*, and*

(b) X is strictly positive for $d \in \left[\frac{\theta^*\beta}{\beta-1}, \infty\right)$ *and instantaneously reflected at* 0 *if* $d \in \left(0, \frac{\theta^*\beta}{\beta-1}\right)$*.*

Proof of lemma 4.6.2 . The form of X in (4.66) follows immediately from equations (4.21)-(4.25).

For some $a \in (0, \infty)$, define the functions $b\colon [0,\infty) \to (-\infty,\infty)$, $\sigma\colon [0,\infty) \to [0,\infty)$, $\rho\colon [a,\infty) \to [0,\infty]$ and $s\colon [a,\infty) \to [0,\infty]$ by

$$b(x) = \frac{c-1}{\beta}x^2 + \left((c-1)\theta^* + \frac{d}{\beta}\right)x + d\theta^*, \tag{4.67}$$

$$\sigma(x) = \sqrt{\frac{2}{\beta(\beta-1)}}x^{\frac{3}{2}} + \theta^*\sqrt{\frac{2\beta}{\beta-1}}x^{\frac{1}{2}}, \tag{4.68}$$

$$\rho(x) = \exp\left(-\int_a^x \frac{2b(y)}{\sigma(y)^2}dy\right), \tag{4.69}$$

$$s(x) = \int_x^\infty \rho(y)dy. \tag{4.70}$$

Theorem 4.1 in [50] shows that X is recurrent if for some $a \in (0,\infty)$ it holds that $s(a) = \infty$. Observe that there exists a constant $K \in (0,\infty)$ such that, for $(c-1)(\beta-1) \leqslant 1$, we have

$$\int_a^\infty \rho(x)dx \geqslant K\int_a^\infty \exp\left(-\int_a^x \frac{(c-1)(\beta-1)}{y}dy\right)dx = K\int_a^\infty \left(\frac{a}{x}\right)^{(c-1)(\beta-1)}dx = \infty. \tag{4.71}$$

This proves that X is recurrent for $c \leqslant \beta/(\beta-1)$. Theorem 4.3 in [50] shows that X is explosive if for some $a \in (0,\infty)$ it holds that

$$s(a) < \infty \quad \text{and} \quad \int_a^\infty \frac{s(x)}{\rho(x)\sigma^2(x)}dx < \infty. \tag{4.72}$$

To see this, we assume that $c > \beta/(\beta-1)$ and choose

$$\epsilon \in \left(0, \min\left\{\frac{2c}{\beta} - \frac{2}{\beta-1}, \frac{\left(\frac{2}{\beta(\beta-1)}\right)^2}{\frac{2(c-1)}{\beta} - \frac{2}{\beta(\beta-1)}}\right\}\right) \tag{4.73}$$

and $a = a(\epsilon) \in (0,\infty)$ such that

$$\left(\sqrt{\frac{2}{\beta(\beta-1)}}y^{\frac{1}{2}} + \theta^*\sqrt{\frac{2\beta}{\beta-1}}\right)^2 < \left(\epsilon + \frac{2}{\beta(\beta-1)}\right)y^2, \quad \text{for all } y \in [a,\infty). \tag{4.74}$$

This implies that, for an exponent

$$E = \frac{2(c-1)}{\beta\left(\epsilon + \frac{2}{\beta(\beta-1)}\right)} > 1, \tag{4.75}$$

it holds that

$$s(a) = \int_a^\infty \rho(x)dx < \int_a^\infty \exp\left(-\int_a^x \frac{2\frac{c-1}{\beta}y}{\left(\epsilon + \frac{2}{\beta(\beta-1)}\right)y^2}dy\right)dx = \int_a^\infty \left(\frac{a}{x}\right)^E dx < \infty. \tag{4.76}$$

To verify the second part of inequalities (4.72), we first note that, from equation (4.76), we get

$$s(x) \leqslant \int_x^\infty \left(\frac{a}{y}\right)^E dx = \frac{a^E}{E-1}\frac{1}{x^{E-1}}, \tag{4.77}$$

and thus, in combination with the inequality on $\rho(x)$ used in equation (4.71), we can conclude that there exists a constant $C' \in (0,\infty)$ such that

$$\int_a^\infty \frac{s(x)}{\rho(x)\sigma^2(x)}dx \leqslant C' \int_a^\infty \frac{1}{x^{E-1}}\frac{1}{x^3}x^{(c-1)(\beta-1)}dx, \tag{4.78}$$

which is finite, as by choice of ϵ in (4.73) we have

$$E + 2 - (c-1)(\beta-1) > 1. \tag{4.79}$$

The prove of claim (a) is thus completed.

To analyze the behavior of X at 0, define, for $a \in (0,\infty)$, the functions $\tilde{\rho}\colon [a,\infty) \to [0,\infty]$ and $\tilde{s}\colon [a,\infty) \to [0,\infty]$ by

$$\tilde{\rho}(x) = \exp\left(\int_x^a \frac{2b(y)}{\sigma(y)^2}dy\right), \tag{4.80}$$

$$\tilde{s}(x) = \int_x^a \tilde{\rho}(y)dy, \tag{4.81}$$

with b and σ as in equations (4.67)-(4.68). Note that there exist constants $K, K', K'' \in (0,\infty)$ and $a = a(K'') \in (0,\infty)$ such that

$$\tilde{\rho}(x) \leqslant K\left(\frac{a}{x}\right)^{\frac{(\beta-1)d}{\theta^*\beta}} \quad \text{and} \tag{4.82}$$

$$\tilde{\rho}(x) \geqslant K'\left(\frac{a\left(K''x + (\theta^*)^2\frac{2\beta}{\beta-1}\right)}{x}\right)^{\frac{(\beta-1)d}{\theta^*\beta}} \qquad \text{for } x \in (0,a]. \tag{4.83}$$

Theorem 2.12 in [50] shows that X is instantaneously reflected at 0 if there exists some $a \in (0,\infty)$

such that

$$\int_0^a \tilde{\rho}(x)dx < \infty, \quad \int_0^a \frac{1+|b(x)|}{\tilde{\rho}(x)\sigma^2(x)}dx < \infty \quad \text{and} \quad \int_0^a \frac{|b(x)|}{\sigma^2(x)}dx = \infty. \tag{4.84}$$

Assume that $d < \theta^*\beta/(\beta-1)$. Then, the first inequality of (4.84) follows immediately from equation (4.82). For the second inequality of (4.84), we use estimate (4.83) and note that there exists a constant $K \in (0,\infty)$ such that

$$\sigma^2(x) \geqslant (\theta^*)^2\frac{2\beta}{\beta-1}x \quad \text{and} \quad |b(x)| \leqslant Kd\theta^*, \quad x \in (0,a]. \tag{4.85}$$

To derive the third inequality of (4.84), it suffices to note that there exists a constant $K' \in (0,\infty)$ and some $a = a(K') \in (0,\infty)$ such that

$$\int_0^a \frac{|b(x)|}{\sigma^2(x)}dx \geqslant K' \int_0^a \frac{1}{x}dx = \infty. \tag{4.86}$$

Thus the sufficient condition of theorem 2.12 in [50] is fulfilled.

Theorem 2.16 in [50] shows that X is strictly positive if there exists some $a \in (0,\infty)$ such that

$$\int_0^a \tilde{\rho}(x)dx = \infty \quad \text{and} \quad \int_0^a \frac{1+|b(x)|}{\tilde{\rho}(x)\sigma^2(x)}\tilde{s}(x)dx < \infty. \tag{4.87}$$

Assume that $d \geqslant \theta^*\beta/(\beta-1)$. Then, the first condition in (4.87) follows immediately from equation (4.83). To derive the second inequality of (4.87), we use estimates (4.82), (4.83), (4.86) and note that there exists a constant $K'' \in (0,\infty)$ and $a = a(K'') \in (0,\infty)$ such that

$$\tilde{s}(x) \leqslant \frac{K''}{x^{\frac{(\beta-1)d}{\theta^*\beta}-1}}, \quad x \in (0,a]. \tag{4.88}$$

This completes the proof of claim (b) and lemma 4.6.2. □

4.6.2 Explanation of the notion of ghost finite time singularities applied to the recurrent bubble regime

Here, we make precise the concept of a ghost finite-time singularity mentioned in subsection 4.3.3.3 as applied to the recurrent bubble case $\eta \in (0,1)$ and $C \in [0,1]$. In a nutshell, the term ghost finite-time singularity describes processes that follow explosive dynamics temporarily, shadowing a true finite time singularity dynamics up to a time when it departs from it and saturates as a result of an inherent stabilizing mechanism that takes over and prevents the divergence.

Following section 4.3.3.1, we set $\theta^* \equiv 1$ without loss of generality. First, we introduce a bounded process

$$Z_t = \frac{1}{\gamma\eta}\tan^{-1}\left(\sqrt{S_t^\eta - 1}\right), \qquad Z_t \in [\,\underline{Z}, \overline{Z}\,] = [\,0, \frac{\pi}{2\gamma\eta}\,], \tag{4.89}$$

where $\gamma = \sqrt{\dfrac{2}{\eta(1-\eta)}}$. Conversely,

$$S_t = [\tan^2(\gamma\eta Z_t) + 1]^{\frac{1}{\eta}} = \sec^{\frac{2}{\eta}}(\gamma\eta Z_t). \tag{4.90}$$

With Itô's lemma, we can check that Z is a unitary diffusion process with

$$dZ_t = \gamma\left[\frac{2(C-1)-\eta}{4}\tan(\gamma\eta Z_t) + \frac{\eta(2D-1)}{4}\cot(\gamma\eta Z_t)\right]\mathrm{d}t + dW_t\ . \tag{4.91}$$

The drift coefficient of Z depends on the balance of power between the tangent and cotangent terms, being $\dfrac{2(C-1)-\eta}{4}\tan(\gamma\eta Z_t)$ and $\dfrac{\eta(2D-1)}{4}\cot(\gamma\eta Z_t)$, respectively. If $C \leqslant 1$ and $D > \frac{1}{2}$, the tangent term is always negative and the cotangent term is always positive. For this set of parameters, the evolution of S can be described in three consecutive stages.

- When S_t is close to 1^+, i.e., $\gamma\eta Z_t \approx 0^+$, the positive cotangent term dominates the drift of Z_t and contribute a strong impetus to blow up the bubble.
- After this initial pumping up, at some point the effect from the cotangent term will be offset by the negative effect of the tangent term. Then the drift of Z may be neglected, and $Z_t \sim W_t$. Hence,

$$S_t \sim \sec^{\frac{2}{\eta}}(\gamma\eta W_t) \sim \left(\frac{\pi}{2} - \gamma\eta W_t\right)^{-\frac{2}{\eta}}\ , \tag{4.92}$$

  which gives an hyperbolic (stochastic) growth for S_t. The above approximate dynamics for S indeed show a finite-time-singularity with a theoretical random critical time point $\widetilde{T}_c$. If we solve for the first $\frac{\pi}{2}$-passage time of the diffusion $\gamma\eta W$, we get a Levy distribution for $\widetilde{T}_c$, in particular

$$\widetilde{T}_c \sim \text{Levy}(0, \frac{\pi^2}{4\gamma^2\eta^2})\ . \tag{4.93}$$

- When S becomes large, the negative tangent term will eventually dominate the cotangent term and act as a strong repeller to prevent explosion of S. This mechanism drives the recurrent behavior, such that the finite time singularity only serves as an approximate, *ghost*-like reference.

The case $D \leqslant \frac{1}{2}$ evolves analogously, without the initially contributing positive drift term.

Part III

Statistical inference of point processes

Chapter 5

The ARMA point process and its estimation

5.1 Introduction

Here we introduce the *ARMApp* (auto-regressive-moving-average point process), and provide means for its maximum likelihood estimation. Motivation of this model and its use benefit from embedding it in a conceptual framework: Activity in a system can be decomposed into *exo*(genous) – having external origin – and *endo*(genous) – having internal origin. Given observation of the overall activity of the system, the statistical problem is to identify its exo and endo parts. Endo processes concern the memory, conditional dynamics, and mechanisms within the system – which is also subjected to exo processes, which may introduce shocks, cycles, and trends into the dynamics. Mixtures of exo and endo processes pervade science and nature [227]: In seismology after-shocks are triggered endo events [122]. In epidemiology, basic incidence of diseases are followed by contagious outbreaks. In financial markets, there are events that lack (identifiable) precursors as well as positive feedbacks and self-excited trading activity [13, 14, 97, 116]. On the Internet, diverse content – tweets, Youtube videos, ideas/memes, and so on – 'go viral' via epidemic-type spreading [47, 62]. Indeed a general problem, of crucial importance for inference and prediction, is to capture the roles of endo and exo processes in a model or via some other metric or test. We call this the *endo-exo problem*, for which testing for unit-roots, or long memory are clear examples. To further illustrate: an exo impulse applied to a system generates a response, which identifies endo characteristics of the system. Additionally, an endo process may generate bursty behavior on its own, which may appear similar to an exogenously driven event. The problem is to untangle these fundamentally different classes of activity.

In the case of point type observations, e.g, having a sample of points with time and/or space coordinates, burstiness or clustering may be observed and the endo-exo problem consists in identifying and modeling the underlying nature of such clustering.[1] Point-process models featuring

[1] In the following we will use basic terminology from the point process literature, see appendix 5.9.1 for a short introduction.

a specific generating process can be considered, often with a causal structure and interpretation. In the literature, two cluster mechanisms have been very prominent, based on a sensible starting point for exo activity, the memoryless *Poisson process*. The *Hawkes process*, on the one hand, builds on this via self-excitation of the activity level, which is a reasonable endo process where positive feedback exists. It has been widely used to model the role of endo activity in financial markets, with important consequences for market stability and risk management [14, 32, 46, 97]. Alternatively, *shot noise* builds on the Poisson process by having each of its points generate a burst of additional points, forming a *cluster*. This could be a reasonable model for the response to external shocks. Shot noise processes have been the scope of insurance risk-theoretic analysis [4, 65, 163]. As argued in [67], credit default claims and insurance claims are subject to both exo shocks as well as endo contagion. Combinations of bursty and contagious dynamics for information diffusion on Twitter have been proposed [47].

To classify these cluster mechanisms and profit from familiar notions, we exploit analogies between point process and time series models. General parallels were explored in [35]. Here we consider that point processes aggregated on a grid (i.e., taking bin-counts) form a time series of non-negative integers. In fact, the Hawkes process has been shown to be equivalent to the INAR (integer valued auto-regressive) as the coarseness of the grid goes to zero [161]. The same equivalence between shot noise and the INMA (integer valued moving-average) time series model is apparent. We introduce the ARMA point process (ARMApp) as the analogue of the INARMA (integer valued ARMA) time series [99, 182, 246, 247]. Although the analogy is striking, we do not prove equivalence. Rather, we focus on developing statistical estimation of the ARMApp. The richness of the ARMA time series framework is well known. Similarly, the ARMApp contains regular Poisson type flow and both of the fundamental cluster types mentioned above (that is, shot noise bursts and self-excited dynamics), which seem to provide a basic set of elements relevant for endo-exo modeling in practice.

To compare with the literature, the ARMApp can be considered as a special case of the Hawkes process with general immigration, whose spectral measure formula was derived in [34], and further analyzed in [31]. In terms of more concrete applied models, the ARMApp is quite similar to the *dynamic contagion* point process of [67], which also combines flow, bursts, and self-excitation. The dynamic contagion process with exponential memory functions has been studied, and approximate estimation developed by Kalman filter [66]. Our contribution, aside from the intuitive analogy and perhaps incremental novelty of the model, is the advance of allowing for maximum likelihood estimation of the ARMApp with time-varying exo flow, and diverse memory functions. Note that such an approach could also be applied to the dynamic contagion process. The practical value of this is worth emphasizing: As recently shown for the Hawkes process [250], seminally encountered in hydrology [131], and pervasive in econometrics [108], it is crucial to account for variation in the exo process (whether it be flow rate or trend). Indeed, mis-treating the exo part e.g., as constant when actually trending or cyclical is well known to lead to spurious long and strong memory.

Statistical estimation of models with a moving average (shot noise) component with MLE (maximum likelihood estimation) is complicated when innovations are not observed. Less efficient or approximate methods are often relied upon [33, 36, 66, 201, 238]. To overcome this for the ARMApp, we derive an MCEM (Monte-Carlo expectation-maximization) algorithm, extended from the spatial point process literature [187, 188]. In simulation studies, estimation is shown to perform well and select the correct submodel (full ARMA, Hawkes or Neyman-Scott).

In terms of structure: First we define the ARMApp, derive its basic properties, and discuss connections to related models from the point process literature (section 5.3). Then we establish a formal analogy to the INARMA time series (section 5.4) and develop MLE via the MCEM algorithm (section 5.5). A simulation study demonstrates performance of the estimation procedure (section 5.6), and a practical example highlights its strengths and limitations in addressing the endo-exo problem (section 5.7).

5.2 Notation

The following notation and terminology is used throughout this chapter. On a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ we will work with point processes $N : \Omega \times \mathcal{B}(\mathbb{R}) \to \mathbb{N}_0 \cup \{\infty\}$ defined as random measures on $\mathbb{R}$. For a point process N and a Borel set $A \subseteq \mathbb{R}$, let $N(A)$ denote the (random) number of points in A. Throughout we will exclude multiple occurrences of points, that is, $N(\{t\}) \leqslant 1$ for all $t \in \mathbb{R}$. In this case, N can be described by an ordered random sequence $(T_j)_{j\in\mathbb{Z}}$ with $T_0 < 0 \leqslant T_1$, denoting the distinct jump times of N.

In the special case of a half open interval $(a, b]$, we will write $N((a, b]) = N(a, b)$. We will use $N_t = N(0, t)$ for $t \in [0, \infty)$ and $N_t = -N([t, 0))$ for $t \in (-\infty, 0)$. For a process N defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we will denote by $\sigma(N)$ the history of the process N given by $\sigma(N) = (\mathcal{F}_t)_{t\in(-\infty,\infty)}$, where $\mathcal{F}_t \subseteq \mathcal{F}$ is the σ-algebra generated by the evolution of N up to and including t, that is, $\mathcal{F}_t = \sigma(N_s, s \leq t)$.

Moreover, below we assume familiarity with the notions of (in)homogeneous Poisson process, immigrant/offspring point, Poisson cluster process (Neyman-Scott process) and Galton-Watson cluster process (Hawkes process). For a short introduction to these concepts, see appendix 5.9.1.

5.3 The ARMA point process

5.3.1 Setting and definition

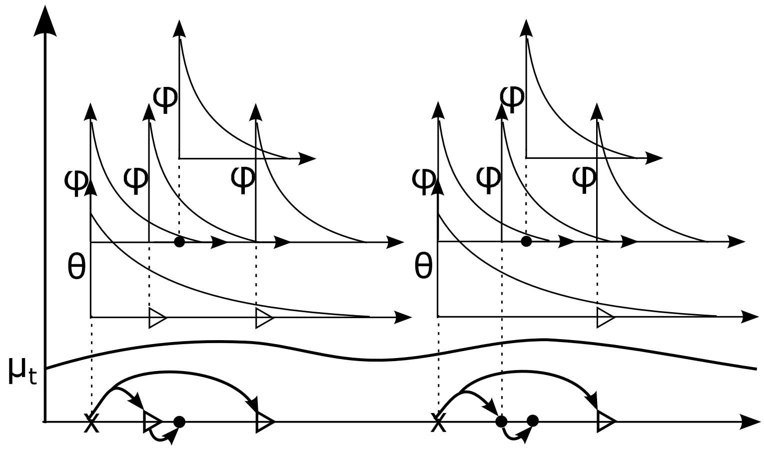


Figure 5.1: A realization of the ARMApp with innovation (immigration) intensity μ, MA (shot noise) intensity θ, and AR (Hawkes) intensity ϕ. Immigrants, θ-offspring and ϕ-offspring are denoted by x, triangle, and dot, respectively. A point is connected to the intensity that it triggers by a vertical dashed line. All points are projected onto the horizontal axis, with parenthood indicated by arrows, forming the full realization. The figure is adapted from [248].

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\mu \in (0, \infty)$, let $N^\mu : \Omega \times \mathcal{B}(\mathbb{R}) \to \mathbb{N}_0$ be a homogeneous Poisson process on $\mathbb{R}$ with rate μ, let $\eta \in [0, 1)$, $\gamma \in [0, \infty)$, let $\theta, \phi : [0, \infty) \to [0, \infty)$ be integrable functions with the property that

$$\int_0^\infty \theta(t)dt = \gamma \quad \text{and} \quad \int_0^\infty \phi(t)dt = \eta\,, \tag{5.1}$$

and let $N^I : \Omega \times \mathcal{B}(\mathbb{R}) \to \mathbb{N}_0$ be a Poisson cluster process with Poisson offspring distribution θ based on immigrants N^μ.[2] Then the ARMA point process (ARMApp) is defined as the Galton-Watson cluster process $N : \Omega \times \mathcal{B}(\mathbb{R}) \to \mathbb{N}_0$ with Poisson offspring distribution ϕ based on immigrants N^I.[3] This process is visualized in Figure 5.1.

5.3.2 Properties

Lemma 6.3.II and Exercise 6.3.5 in [63] ensure the existence and stationarity of N^I and N. The main point here is that, for a stationary immigration process, there exists a stationary cluster process if clusters are i.i.d. distributed (given the cluster center) and the cluster size is finite, which is ensured by $\eta \in [0,1)$ and $\gamma \in [0,\infty)$. The construction of a point process as a collection of Galton-Watson branching processes with Poisson offspring distribution was initially used in the point process representation of the Hawkes process in [120], see also Example 6.3(c) in [63].

For the full filtration $\mathcal{F} = \sigma(N^\mu, N)$, it holds that the $\mathcal{F}$-conditional intensity function is given by

$$\lambda(t) = \mu + \int_{-\infty}^t \theta(t-s)dN_s^\mu + \int_{-\infty}^t \phi(t-s)dN_s. \tag{5.2}$$

With respect to the filtration $\mathcal{F} = (\mathcal{F}_t)_{t\in(-\infty,\infty)}$, the intensity function can be seen as a conditional hazard function in the sense that

$$\lambda(t) = \lim_{\Delta\downarrow 0} \Delta^{-1}\mathrm{E}\left[N(t,t+\Delta)|\mathcal{F}_{t-}\right]. \tag{5.3}$$

For details see Chapter 7 of [63]. Note that the conditional intensity function uniquely defines the probability structure of a point process only if it is measurable with respect to its *internal history*.[4] Thus, unlike the case for the classical Hawkes process, the $\mathcal{F}$-conditional intensity function (5.2) is not a defining property of N.

Alternatively, one can use a random sequence of indicator variables depending on N and N^μ to concisely describe the conditional intensity and the information flow. For $t \in \mathbb{R}$, we can define the random sequence $\mathcal{Z}_t = (Z_i)_{i\in\mathbb{N}}$ where, for $i \in \mathbb{Z} \cap (-\infty, N_t]$, $Z_i = 1$ if the $(N_t - i + 1)$-th last point was an immigrant and $Z_i = 0$ otherwise. Then the full history is generated by N and $\mathcal{Z}$,

[2] That is, N^I is the collection of N^μ and Poisson clusters generated by N^μ, see definition 5.9.4.
[3] That is, N is the collection of Galton-Watson clusters triggered by points in N^I, see definition 5.9.5.
[4] See Chapter 7 and Proposition 7.2.IV. in [63] for details.

that is $\mathcal{F} = \sigma(N^\mu, N) = \sigma(\mathcal{Z}, N)$ and the $\mathcal{F}$-conditional intensity can be written as

$$\begin{aligned} \lambda(t) &= \mu + \sum_{j=-\infty}^{N_t^\mu} \theta(t - T_j^\mu) + \sum_{k=-\infty}^{N_t} \phi(t - T_k) \\ &= \mu + \sum_{j=-\infty}^{N_t} Z_j^\mu \theta(t - T_j) + \sum_{k=-\infty}^{N_t} \phi(t - T_k) \,, \end{aligned} \tag{5.4}$$

where $(T_j^\mu)_{j\in\mathbb{Z}}$ and $(T_j)_{j\in\mathbb{Z}}$ denote the jump times of N^μ and N, respectively.

5.3.3 First and second order statistics

Using stationarity of the process, we can take the expectation of the $\mathcal{F}$-conditional intensity function in (5.2) to get the expected intensity,

$$\bar{\lambda} = \frac{\mu(1+\gamma)}{1-\eta} \,, \tag{5.5}$$

which defines the *first moment measure* of N, that is, for any Borel set $A \subseteq \mathbb{R}$, we have $E[N(A)] = \int_A \bar{\lambda} dt$. By stationarity of the process, the *covariance measure*, if it exists, is fully defined by its density $c(u) = Cov(dN_t, dN_{t-u})$,[5] see 6.1.I. in [63]. The covariance measure has a singular Dirac component at 0 and thus c can be written as

$$c(u) = \overline{\lambda}\delta(u) + \overline{\lambda}h(u) - \overline{\lambda}^2 \tag{5.6}$$

with the symmetric function $h : (-\infty, \infty) \to [0, \infty)$, called *palm-intensity* in [59], given by

$$\begin{aligned} h(u) &= \mathbb{P}\left[dN_{t+u} = 1 | dN_t = 1\right] \frac{1}{du} = \mathbb{E}\left[dN_t dN_{t-u}\right] \frac{1}{\overline{\lambda} dt du}, \quad u \in (0, \infty) \\ h(0) &= 0. \end{aligned} \tag{5.7}$$

To derive an expression for h utilizing equation (5.2), we first need to derive an equation for the function $\tau : (-\infty, \infty) \to [0, \infty)$ given by the defining equation

$$\mathbb{E}\left[dN_t dN_{t-u}^\mu\right] = (\delta(u)\mu + \tau(u))\, dt du. \tag{5.8}$$

We multiply equation (5.2) with dN_{t-u}^μ / du and take expectations to arrive at

$$\tau(u) = \mu^2 + \int_0^t \theta(t-s) \frac{1}{du} \mathbb{E}\left[dN_s^\mu dN_{t-u}^\mu\right] + \int_{-\infty}^t \phi(t-s)\tau(s-t+u) ds, \quad u \neq 0. \tag{5.9}$$

[5]Hereafter, we will use artificial objects like $Cov(dN_t, dN_{t-u})$ for simplicity. What we actually mean here is that c is a density in the sense that for Borel sets $A, B \subseteq \mathbb{R}$ it holds that $Cov(N(A), N(B)) = Cov(\int_A dN_t, \int_B dN_t) = \int_{A\times B} c(u-t) du dt$. For a rigorous definition of moment measures and their densities, we refer to Chapter 5.4 in [63].

Using $\mathbb{E}\left[dN_s^\mu dN_{t-u}^\mu\right] = \left(\mu\delta(s-t+u)+\mu^2\right) dsdu$ we get

$$\begin{aligned}
&\tau(u) = \mu^2(1+\gamma) + \mu(\theta(u)+\phi(u)) + \int_0^\infty \phi(s)\tau(u-s)ds, \quad u \in (0,\infty), \\
&\tau(0) = \overline{\lambda}\mu.
\end{aligned} \tag{5.10}$$

Now we are concerned with calculating the full conditional intensity function h. To this end, let $\rho : [0,\infty) \to [0,\infty)$ be the function given by $\rho = \overline{\lambda}h$. Multiplying equation (5.2) with dN_{t-u}/du, taking expectations and using identities (5.7) and (5.10) gives

$$\begin{aligned}
\rho(u) &= \mu\overline{\lambda} + \mu\theta(u) + \int_{-\infty}^{t-u} \theta(t-s)\tau(t-u-s)ds + \overline{\lambda}\phi(u) + \int_0^\infty \phi(s)\rho(u-s)ds \\
&= \mu\overline{\lambda} + \mu\theta(u) + \overline{\lambda}\phi(u) + \int_0^\infty \theta(s+u)\tau(s)ds + \int_0^u \theta(s)\tau(s-u)ds \\
&\quad + \int_0^\infty \phi(u+s)\rho(s)ds + \int_0^u \phi(s)\rho(u-s)ds.
\end{aligned} \tag{5.11}$$

If equation (5.11) has a solution, the covariance measure exists and has density c defined in (5.6). As ϕ and θ are integrable functions, by L^1-theory for Fredholm integral equations, see, e.g., chapter 2.3 in [200], there exist locally integrable solutions to equations (5.10) and (5.11). For exponential densities, equation (5.11) can be solved explicitly. To this end, let $\theta_0, \theta_1, \phi_0 \in (0,\infty)$, $\phi_1 \in (\phi_0,\infty)$ and let $\theta(t) = \theta_0 e^{-\theta_1 t}$ and $\phi(t) = \phi_0 e^{-\phi_1 t} = \eta\phi_1 e^{-\phi_1 t}$ (from the definition of the branching ratio η given by (5.1), which shows indeed that the condition $\eta < 1$ is equivalent to $\phi_1 = \phi_0/\eta > \phi_0$). Then, as we show in appendix 5.9.2, there exist constants $K_1, K_2 \in (0,\infty)$ depending on $(\theta_0, \theta_1, \phi_0, \phi_1)$ such that the palm intensity h is given by

$$h(t) = \overline{\lambda} + K_1 e^{-(\phi_1-\phi_0)t} + K_2 e^{-\theta_1 t} = \overline{\lambda} + K_1 e^{-(1-\eta)\phi_1 t} + K_2 e^{-\theta_1 t}, \quad t \in (0,\infty). \tag{5.12}$$

This, in particular, shows that estimation based on second order statistics alone will not be satisfactory. The term $e^{-(1-\eta)\phi_1 t}$ recovers the standard renormalization of the characteristic time scale $\frac{1}{\phi_1}$ to $\frac{1}{\phi_1}\frac{1}{1-\eta}$ by counting over all generations of the ϕ-process.

For the special cases of a Hawkes or a Neyman-Scott (NS) processes, the above expression reduces to the respective well-known palm intensities, in particular it holds that $K_1 = 0$ for $\phi_0 = 0$ and $K_2 = 0$ for $\theta_0 = 0$. As for the Hawkes and the NS processes, we have $\lim_{t\to\infty} h(t) = \overline{\lambda}$.

5.3.4 Comments

Branching interpretation. The explicit construction of the ARMApp brings great insight, as visualized in Fig. 5.1: μ introduces *immigrants*, which may then trigger a single generation of *θ-offspring* with intensity $\theta(\cdot)$, and then all existing points (μ-immigrants and their *θ-offspring*) trigger a generation of *ϕ-offspring* with intensity $\phi(\cdot)$, which may, in turn, trigger the subsequent generation of *ϕ-offspring* in the same way. The sum of these independent inhomogeneous Poisson processes provides the $\mathcal{F}$-conditional intensity (5.2), and the set of immigrant and (θ- and ϕ-)offspring points forms the ARMApp realization. The *branching ratio* γ is the expected number of

θ-*offspring* of a single immigrant, and the *branching ratio* η is the expected number of immediate ϕ-*offspring* of any point. Further, counting all generations, a single point is expected to produce $\eta + \eta^2 + \cdots = 1/(1-\eta) - 1$ ϕ-*offspring*. Thus, as in the Hawkes process, η is the expected proportion of all points that are ϕ-*offspring*. From the ARMApp defined above, by setting $\gamma = 0$ one recovers the Hawkes process, and by setting $\eta = 0$ one recovers a modified Neyman-Scott (NS) process[6] (see [191]). It can thus be regarded as an extension of both a self-exciting and an i.i.d. cluster process.

Extensions/Modifications.

1. The ARMApp can be extended by using an inhomogeneous Poisson process N^μ in definition 5.3.1 for a locally integrable rate function $\mu\colon (-\infty, \infty) \to [0, \infty)$. The resulting cluster process of immigrants, N^I, can be extended to marked point process with the marks $(Y_i)_{i\in\mathbb{N}}$ being i.i.d. non-negative valued random variables, such that the cluster generated by the j^{th} point of N^μ at $t_j \in \mathbb{R}$ is distributed according to an inhomogeneous Poisson process with rate $y_j\theta(\cdot - t_j)$ for a realization y_j of Y_j. Moreover, the ARMApp N can be extended to a marked point process with the marks $(Z_i)_{i\in\mathbb{N}}$ being i.i.d. non-negative valued random variables, such that the cluster generated by the k^{th} point of N at $t_k \in \mathbb{R}$ is distributed according to an inhomogeneous Poisson process with rate $z_k\phi(\cdot - t_k)$ for a realization z_k of Z_k. Then the $\mathcal{F}$-conditional intensity of N in the form of (5.4) can be written as

$$\lambda(t) = \mu(t) + \sum_{j=-\infty}^{N_t^\mu} Y_j\theta(t - T_j^\mu) + \sum_{k=-\infty}^{N_t} Z_k\phi(t - T_k)\ . \tag{5.13}$$

2. We will also show how to estimate a variant of the process (5.13) with background immigration, that is, where the immigrants N^μ are not included in the sample. In this case, the $\mathcal{F}$-conditional intensity of the process takes the form (5.13) without the first term $\mu(t)$ in the r.h.s.

$$\lambda(t) = \sum_{j=-\infty}^{N_t^\mu} Y_j\theta(t - T_j^\mu) + \sum_{k=-\infty}^{N_t} Z_k\phi(t - T_k)\ . \tag{5.14}$$

   While the inclusion of the immigrants can be argued to be *natural* due to its similarity with the Hawkes process or INARMA time series (introduced below), which both count immigrants, their exclusion – i.e., the use of background immigration – can be argued to be *natural* by interpreting it as a Hawkes process with Neyman-Scott immigration. In practice the application may clarify which specification makes sense.

Dynamic contagion process. A similar Hawes process with general immigrants has been introduced as the *dynamic contagion process* in [67]. The authors define a point process that has a

[6] I.e., an NS process where the immigrant is included in the counts.

stationary[7] conditional intensity function, with respect to a suitable history,

$$\lambda(t) = \nu + \sum_{j=-\infty}^{N_t^\mu} Y_j \theta(t - T_j^\mu) + \sum_{k=-\infty}^{N_t} Z_k \phi(t - T_k) \ . \tag{5.15}$$

for $\nu \in [0, \infty)$ that corresponds to a ("background") immigration Poisson process N^ν, a Poisson process N^μ defined as above and independent from N^ν and (respectively) identically distributed marks $(Y_j)_{j\in\mathbb{Z}}, (Z_k)_{k\in\mathbb{Z}}$. This process is similar to but different from the ARMApp, where immigration and external shocks to the intensity are identical.[8] The next section provides a justification for our (different) choice of the specific structure of N and the label *ARMApp*.

5.4 The relationship to integer-valued time series

The ideas in this section are based on Kirchner [161], who considers the framework of a classic Hawkes process.

5.4.1 The INARMA model

The discrete valued analogues of classical time series models [37] have seen a flurry of recent development [99, 182, 246, 247] and enjoy many current and potential applications. Here we consider the INARMA(p, q) process, an integer-valued ARMA process for counts. For its definition, we will need the *Poisson thinning* operator $\circ$ to preserve the count value of the process. In all of the following, we assume an underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For some count variable $Z : \Omega \to \mathbb{N}$, a non-negative real number $\alpha \in [0, \infty)$, a sequence of i.i.d. Poisson distributed random variables $(Y_i)_{i\in\mathbb{Z}} \overset{i.i.d}{\sim} \text{Pois}(\alpha)$, we define the count variable $\alpha \circ Z : \Omega \to \mathbb{N}$ by

$$\alpha \circ Z = \sum_{i=1}^{Z} Y_i, \ \alpha > 0, \text{ and } \alpha \circ 0 := 0. \tag{5.16}$$

Definition 5.4.1. Let $p, q \in \mathbb{N}$, let $\widetilde{\mu} \in [0, \infty)$, let $(\epsilon_i)_{i\in\mathbb{Z}}$ be a sequence of i.i.d. random variables $\epsilon_i \overset{i.i.d}{\sim} \text{Poisson}(\widetilde{\mu})$, let $(\widetilde{\theta}_k)_{k=1,\dots,p}$ and $(\widetilde{\phi}_j)_{j=1,\dots,q}$ be real-valued, non-negative sequences with the property that $\sum_{j=1}^{p} \widetilde{\phi}_j < 1$. Then an integer-valued time series $X : \Omega \times \mathbb{Z} \to \mathbb{N}$ is an *INARMA*(p, q) *process* if it satisfies the difference equation

$$X_l = \epsilon_l + \sum_{k=1}^{q} \widetilde{\theta}_k \circ \epsilon_{l-k} + \sum_{j=1}^{p} \widetilde{\phi}_j \circ X_{l-j}, \quad l \in \mathbb{Z}, \tag{5.17}$$

[7] In [67], they work in a non-stationary framework starting from an initial value at λ_0 at 0, whose influence diminishes over time $t \to \infty$.

[8] By using a mixture distribution for the marks $(Y_i)_{i\in\mathbb{N}}$ in (5.13), the ARMApp can mimic the structure that some immigrants (cf. N^ν) do not trigger a cluster but some do (cf. N^μ). However, every immigrant (N^ν and N^μ) will be counted. In this sense, a very similar but not the exact same process can be estimated with the algorithm presented here. For the special case $\nu = 0$, the dynamic contagion process is identical to modification (2) of the ARMApp mentioned above.

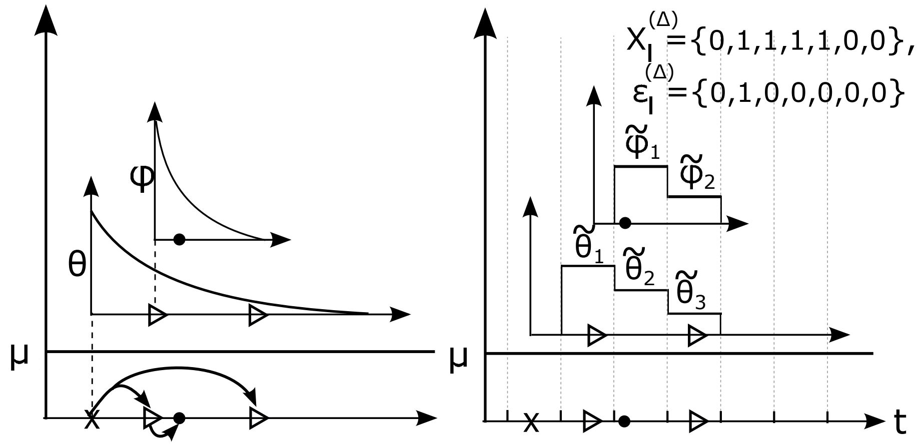


Figure 5.2: Representation of a cluster generated by the ARMApp (left plot), and the INARMA$(2,3)$ process (right plot) which approximates the ARMApp with grid Δ. Only intensities that generated a point are shown. The immigrant and total counts are given which form the INARMA realization rather than the points. The origins of the axes framing the AR and MA triggering coefficients are located at the time values of the points that triggered them to highlight a source of approximation error: the INARMA can only trigger across bins, not within them. The figure is adapted from [248].

where all thinning operations are mutually independent.[9]

5.4.2 Comments

The INARMA process (5.17) exists as a multi-type branching process that is stationary for $\sum_{j=1}^{p} \tilde{\varphi}_j < 1$, see corollary 2 in [79].

The autocovariance function $\gamma : \mathbb{Z} \to [0, \infty)$ of the INARMA process (5.17) is the function $\gamma(u) = Cov(X_l, X_{l-u})$, often equivalently used via the rescaled autocorrelation function (ACF). It is explicitly understood for INAR(p) and INMA(q) processes with Bernoulli thinning, see [80] and [181], respectively. The case of Poisson thinning is covered by a general result in [168] for INAR(p) and can be extended directly from [181] for INMA(q) processes. To our knowledge, an explicit description of the ACF for the full INARMA(p,q) process (5.17) has not been established yet and is left for future research. Given the autocovariance γ, we can find best linear conditional predictors for X in the space of real-valued time series and one might extend the notion of partial autocorrelation (PACF) to INARMA(p,q)-processes.[10]

From the difference equation (5.17), assuming a time distance $\Delta \in [0, \infty)$ between the counts, it is straightforward to deduce a discrete conditional intensity function $\lambda^{(\Delta)}$ given by

$$\lambda^{(\Delta)}(l) = \frac{1}{\Delta} E[X_l | X_{(l-1):(l-p)}, \epsilon_{(l-1):(l-q)}] = \frac{1}{\Delta} \left(\tilde{\mu} + \sum_{k=1}^{q} \tilde{\theta}_k \epsilon_{l-k} + \sum_{j=1}^{p} \tilde{\varphi}_j X_{l-j} \right), \quad l \in \mathbb{Z}. \tag{5.18}$$

[9] There exist other interpretations for the serial dependence of thinning operations in the literature, see, e.g., chapter 5 of [100] for a specific interpretation in the general case or [33] for an overview of interpretations in the INARMA$(0,q)$-case.

[10] See, e.g., sections 3.4 and 5.2 [37] for a derivation of the partial autocorrelation for real-valued time series in terms of optimal linear predictors, using the Durbin-Levinson algorithm.

The branching interpretation of the INARMA process (5.17) is the same as for the ARMApp: the innovation count ϵ_l introduces immigrants, and thinning (5.16) has the interpretation that each of the X_j points in the j^{th} bin is expected to produce $\widetilde{\phi}_{l-j}$ offspring in the l^{th} bin, where $l > j$. Thus the INARMA process introduces a burst of offspring triggered by immigrants, via thinning with $\widetilde{\theta}$ coefficients, and an autoregressive tree of offspring triggered by all past events, via thinning with $\widetilde{\phi}$ coefficients.

Observe that the thinning operator defined in equation (5.16) has the property that $\alpha \circ Z | Z = z$ is a sum of z independent Poisson variables with parameter α, and has distribution $\text{Pois}(\alpha z)$. Thus, given that both all thinnings in (5.17) are independent of each other, and of the Poisson innovation ϵ_l, then the conditional df of $X_l | X_{(l-p):(l-1)}, \epsilon_{(l-q):(l-1)}$ is also Poisson. The unconditional df of X_l, on the other hand, is not Poisson. In this sense, we maintain structural similarities to ARMApp, whose distribution, given the full history, is that of an inhomogeneous Poisson process. It is important to note that the standard thinning used in integer time series is Binomial thinning, where the variable Y has a Bernoulli distribution. In this case, the unconditional distribution of X_l is Poisson, but the conditional one is not. A survey of the different thinning specifications employed within the literature are summarized in [247].

5.4.3 Connection to the ARMApp

As discussed above, the branching interpretation of the INARMA time series and the ARMApp are identical. Next, a formal argument is made to suggest an asymptotic equivalence between the process from section 5.3.1 and INARMA time series, motivating the term *ARMApp*. For this, let N be an ARMApp as defined in section 5.3.1 given by an immigration rate $\mu \in (0, \infty)$ and integrable intensities $\theta, \phi : [0, \infty) \to [0, \infty)$. If one aggregates the ARMApp N on bins of width $\Delta > 0$, one obtains the counting variables $\{N_l^{(\Delta)} = N(\Delta l, \Delta(l+1)),\ l \in \mathbb{Z}\}$ for all points and $\{\epsilon_l^{(\Delta)} = N^\mu(\Delta l, \Delta(l+1)),\ l \in \mathbb{Z}\}$ for innovations, see Fig. 5.2 for an example. Then, for mild assumptions on θ and ϕ, we $\mathbb{P}$-a.s. have the convergence

$$\begin{aligned} \sum_{k=1}^{q} \theta(k\Delta)\epsilon^{(\Delta)}_{\lfloor t/\Delta \rfloor - k} &\xrightarrow[p,q\to\infty,\Delta\to 0]{} \int_{-\infty}^{t} \theta(t-s) dN^\mu_s, \\ \sum_{j=1}^{p} \phi(j\Delta)N^{(\Delta)}_{\lfloor t/\Delta \rfloor - j} &\xrightarrow[p,q\to\infty,\Delta\to 0]{} \int_{-\infty}^{t} \phi(t-s) dN_s. \end{aligned} \tag{5.19}$$

Thus, for the aggregated model $N^{(\Delta)}$, using (5.2) and (5.19), we expect a discrete conditional intensity function

$$\lambda^{(\Delta)}(l) = \mu + \sum_{k=1}^{q} \theta(k\Delta)\epsilon^{(\Delta)}_{l-k} + \sum_{j=1}^{p} \phi(j\Delta) N^{(\Delta)}_{l-j} + \text{err}(\Delta, p, q) \ , \tag{5.20}$$

where $\text{err}(\Delta, p, q)$ consists of approximation of the integral in (5.19) with finite step functions of length p, q and the influence of offspring that are triggered by points N or N^μ in the *same* interval (of length Δ). Thus, we expect $\text{err}(\Delta, p, q) \to 0$ for $p, q \to \infty$ and $\Delta \to 0$ and the aggregated

ARMApp to (approximately) follow an INARMA process. This formal reasoning leads us to the conjecture that the finite dimensional distributions of a sequence of INARMA processes converge to the finite dimensional distributions of the ARMApp.

Conjecture. 1. (The INARMA(∞, ∞) process.) *Let $\widetilde{\mu} \in [0, \infty)$, let $(\epsilon_i)_{i \in \mathbb{Z}}$ be sequence of i.i.d. random variables $\epsilon_i \overset{i.i.d}{\sim}$ Poisson$(\widetilde{\mu})$ and let $(\widetilde{\theta}_k)_{k \in \mathbb{N}}$ and $(\widetilde{\phi}_j)_{j \in \mathbb{N}}$ be real-valued, non-negative sequences with the property that $\sum_{k=1}^{\infty} \widetilde{\theta}_k < \infty$ and $\sum_{j=1}^{\infty} \widetilde{\phi}_j < 1$. Then there exists an integer-valued stationary time series $X : \Omega \times \mathbb{Z} \to \mathbb{N}$ that satisfies the difference equation*

$$X_l = \epsilon_l + \sum_{k=1}^{\infty} \widetilde{\theta}_k \circ \epsilon_{l-k} + \sum_{j=1}^{\infty} \widetilde{\phi}_j \circ X_{l-j}, \quad l \in \mathbb{Z}. \tag{5.21}$$

2. (Approximation of the ARMApp.) *Let N be an ARMApp as defined in section 5.3.1 with a piecewise continuous intensity ϕ, then there exists $\delta \in (0, \infty)$ such that*

   - *for $\Delta \in (0, \delta)$ equation (5.21) with $\widetilde{\mu} = \mu\Delta$, $\widetilde{\theta}_k = \Delta\theta(k\Delta)$, and $\widetilde{\phi}_j = \Delta\phi(j\Delta)$ defines a stationary INARMA process $X^{(\Delta)}$ and*
   - *the family of point processes $(N^{\Delta})_{\Delta \in (0,\delta)}$ given by*

$$N^{\Delta}(A) = \sum_{n:\, n\Delta \in A} X_n^{\Delta}, \quad \text{for a Borel set } A \subseteq \mathbb{R} \tag{5.22}$$

   converges weakly[11] *to N for $\Delta \to 0$.*

A rigorous proof goes beyond the scope of this article. As special cases, the aggregated Hawkes process is approximated by the INAR process, and the aggregated Neyman-Scott process is approximated by an INMA process. The weak convergence of the INAR process to the Hawkes process was established in [161].

5.4.4 Implications

In case the above conjecture, as suggested by the formal argument, turns out to be true, it verifies the already useful analogy between time series models and (binned) point processes.

As an important example, the autocorrelation (ACF) and partial autocorrelation (PACE) functions, which have been thoroughly studied for real-valued time series and are extendible to integer-valued time series (see the comments in section 5.4.2 above), can be used to characterize the ARMApp. The ACF, being defined as the covariance between two lagged random variables for time series, is directly related to the palm intensity (5.7), which defines the covariance measure of the point process. An analogy to the PACF, whose definition involves the notion of linear (best) predictors given a sub-σ-algebra, is not as readily defined for a point process, and leaves room for further research.

Moreover, as is the case for the ARMApp, the INARMA model cannot be directly/ simultaneously estimated by MLE due to missing information – here being the innovation counts

[11] *Weak convergence* of point processes is understood as vague convergence of their induced measures and equivalent to a convergence of their finite dimensional distributions, see, e.g., [207].

$\epsilon_{l-1}^{(\Delta)}, \epsilon_{l-2}^{(\Delta)}, \ldots$ where only the complete counts $X_{l-1}^{(\Delta)}, X_{l-2}^{(\Delta)}, \ldots$ are observed. The EM algorithm provided in section 5.5 to estimate the ARMApp may also be applied to the INARMA model and thus provides a powerful approach to fitting INARMA time series models.

5.5 Estimation of the ARMApp with EM algorithms

5.5.1 Motivations for the EM scheme

Unlike the Hawkes process, the conditional intensity functions of the NS, the ARMA, and marked extensions (5.13) depend on (knowing) the immigrants. These processes are Poisson, given the conditional intensity, which is itself stochastic. For such processes, in general [188], the likelihood is given in terms of an expectation with respect to the unobserved random intensity function. The typical solution in this case is to perform likelihood inference by MCMC (Markov Chain Monte Carlo) sampling [188]. In the case of univariate temporal point processes, this turns out to be rather simple to implement, and works well.

Also, practically speaking, it is crucial in applications to consistently estimate trends in immigration $\mu(t)$ to avoid mistaking deterministic trends for stochastic fluctuations. For reasonable models with such features, moment-based estimation is not useful, and "direct" MLE via numerical maximization of the loglikelihood may perform poorly due to joint estimation of a large number of parameters perhaps along with non-parametric $\mu(t)$. In such a setting, the EM algorithm [183] for maximum likelihood estimation is powerful as it decomposes otherwise unwieldy multi-parameter estimation into sub-estimations – specifically here into very simple problem of density estimations from iid samples with weights.

Regarding scope, this EM framework for ARMA point-processes allows for the estimation of a range of model specifications, e.g.:

- Submodels, including the Hawkes, NS, and SNCP.
- Immigrant points are observed (included in the sample) or unobserved, cf. (5.14).
- Trends in immigration and/or the branching ratios (as well as other parameters).
- Without or with marks (as in (5.13)) that are either observed, or unobserved and thus must be simulated within the EM algorithm.
- And, last but not least, INARMA time series with Poisson thinning (5.17) can be estimated by a simple and obvious modification of the EM algorithm.

Below, the EM algorithm for the case of the ARMA with marks and inhomogeneous immigration will be presented. Section 5.5.2.2 will be concerned with the simpler case where immigrants and marks are observed, section 5.5.2.3 will treat the case of unobserved immigrants. Section 5.5.3 concludes with a step-wise description of the algorithm and section 5.5.4 discusses convergence properties.

5.5.2 Derivation of the the EM algorithm

5.5.2.1 Notations

We are concerned with the ARMApp introduced in section 5.3.1 on a finite observation window $[0,t]$. For ease of presentation, we will look at the induced random sequence $T : \Omega \to \mathbb{R}^{\mathbb{N}}$ consisting of all the times $s \in [0,t]$ where $\Delta N_s \neq 0$. A single realization of T will then be a vector denoted by $T(\omega) =: \boldsymbol{t} = (t_1, \ldots, t_n)$ for a realization $\omega \in \Omega$ and $n = N_t(\omega)$. We will allow for inhomogeneous immigration with intensity $\mu(\cdot)$ and marks $(Y_i)_{i\in\mathbb{N}}$, cf. the discussion before equation (5.13). For ease of presentation, we will introduce some notation.

1. Let C be the random sequence induced by the marked immigrant process N^μ on $[0,t]$, with realization $\boldsymbol{c} = (c_1, \ldots, c_{n_c})$ of length $N^\mu_t = n_c$. Each element has the time and mark $c_j = (s_j, y_j) \in (0,t] \times [0,\infty)$. Separately, denote the points $\boldsymbol{s} = (s_1, \ldots, s_{n_c})$ that are iid on window $(0,t]$ with density $\mu(\cdot)/\mu((0,t])$, and the marks as $\boldsymbol{y} = (y_1, \ldots, y_{n_c})$, where marks are iid with pdf $m(\cdot)$.

2. Similarly, denote by $\boldsymbol{o} = (o_1, \ldots, o_{n_o})$ the realization of the offspring process with $n_o = n - n_c$ and $\boldsymbol{t} = \boldsymbol{c} \cup \boldsymbol{o}$.

3. Finally, let $f, g : [0,\infty) \to [0,\infty)$ be the density function of the AR and MA kernels, that is, $\theta = \gamma g$ and $\phi = \eta f$. Due to the presence of marks, without loss of generality we can assume that $\gamma \equiv 1$.

Then T has a realized $\sigma(N^\mu, N)$-conditional intensity

$$\lambda(s) \approx \mu(s) + \sum_{j=1}^{n_c} y_j \theta(s - s_j) + \sum_{k=1}^{n} \phi(s - t_n), \quad s \in [0,t]. \tag{5.23}$$

5.5.2.2 Description of the EM algorithm for observed immigrants and marks

We will start with the simpler but less realistic case of where we observe a realization of the random sequence $C_t(\omega) = \boldsymbol{c} = (c_1, \ldots, c_{n_c})$ and show how the EM algorithm can be applied in this situation to derive parameter estimates based on a conditional expectation of the full likelihood (E-step) for the parameter vector

$$\boldsymbol{\beta} = (\mu, f, m, g, \eta).$$

The parametric form or non-parametric subclass of the densities μ, f, m, g has to be specified in the maximization step (M-step) to yield a well-defined maximization problem.

Likelihood expectation (E-step) To derive a likelihood function, we will use the fact that, given the full branching structure $\boldsymbol{Z}$, the ARMA process decomposes into three independent inhomogeneous Poisson processes (density p_ρ for intensity ρ) and the density of the marks m. In particular, the density factorizes

$$p(\boldsymbol{t}|\boldsymbol{Z}) = p_\mu(\boldsymbol{s}) p_m(\boldsymbol{y}) p_\theta(\boldsymbol{o}_\theta) p_\phi(\boldsymbol{o}_\phi), \tag{5.24}$$

where $\boldsymbol{o}_\phi$ and $\boldsymbol{o}_\theta$ denote the AR and MA offspring, respectively. The branching structure $\boldsymbol{Z}$ is therefore a highly useful unknown, and will be treated as our EM "missing data". Formally define the missing data by indicator variables

$$\boldsymbol{Z} = \{Z^{\theta}_{i,j}\ ,\ i = 1, ..., n,\ j = 1, ..., n_c\} \cup \{Z^{\phi}_{i,j}\ ,\ i = 1, ..., n,\ j = 1, ..., n\}\ , \tag{5.25}$$

which are zero except $Z^{\theta}_{i,j} = 1$ if t_i is triggered by $\theta(\cdot - s_j)$, and $Z^{\phi}_{i,j} = 1$ if t_i is triggered by $\phi(\cdot - t_j)$.

Given the missing data (5.25), the "complete data" likelihood can be derived by using the missing data variables to rewrite (5.24),

$$\begin{aligned}
L(\boldsymbol{\beta} \mid \boldsymbol{t}, \boldsymbol{c}, \boldsymbol{Z}) = \prod_{i=1}^{n_c} \mu(s_i) \mathrm{Exp}\{-\int_0^t \mu(s)ds\} m(y_i) \times \\
\prod_{i=1}^{n}\prod_{j=1}^{n_c} \Big[y_j \theta(t_i - s_j)\Big]^{Z^{\theta}_{i,j}} \mathrm{Exp}\{-\sum_{j=1}^{n_c} y_j \int_0^t \theta(s - s_j)ds\} \times \\
\prod_{i=1}^{n}\prod_{k=1}^{n} \Big[\phi(t_i - t_k)\Big]^{Z^{\phi}_{i,k}} \mathrm{Exp}\{-\sum_{j=1}^{n} \int_0^t \phi(s - t_j)ds\}\ ,
\end{aligned} \tag{5.26}$$

where an intensity is only evaluated at the times agreeing with the branching structure encoded within the missing data. Instead of optimizing the (inaccessible) complete data likelihood (5.26), the EM algorithm uses the objective function

$$Q(\boldsymbol{\beta} \mid \widehat{\boldsymbol{\beta}}) = \mathbb{E}_{\boldsymbol{Z}|\widehat{\boldsymbol{\beta}},\boldsymbol{t},\boldsymbol{c}}[\log L(\boldsymbol{\beta} \mid \boldsymbol{t}, \boldsymbol{c}, \boldsymbol{Z})]\ , \tag{5.27}$$

the expectation of the log-likelihood over the missing data, given the complete observations $(\boldsymbol{t}, \boldsymbol{c})$, and a parameter estimate $\widehat{\boldsymbol{\beta}}$, where at the $r + 1^{th}$ iteration,

$$\widehat{\boldsymbol{\beta}}^{(r+1)} = \arg\max_{\boldsymbol{\beta}} Q(\boldsymbol{\beta} \mid \widehat{\boldsymbol{\beta}}^{(r)})\ . \tag{5.28}$$

The function Q then contains the probabilities

$$\begin{aligned}
\pi^{\theta}_{i,j} = \mathbb{P}\{Z^{\theta}_{i,j} = 1 | \boldsymbol{t}, \boldsymbol{c}, \boldsymbol{\beta}\} &= \begin{cases} \dfrac{y_j \theta(t_i - s_j)}{\sum_{j=1}^{n_c} y_j \theta(t_i - s_j) + \sum_{j=1}^{n} \phi(t_i - t_j)} & \text{for } t_i \in \boldsymbol{o} \\ 0 & \text{else} \end{cases} \\
\pi^{\phi}_{i,j} = \mathbb{P}\{Z^{\phi}_{i,j} = 1 | \boldsymbol{t}, \boldsymbol{c}, \boldsymbol{\beta}\} &= \begin{cases} \dfrac{\phi(t_i - t_j)}{\sum_{j=1}^{n_c} y_j \theta(t_i - s_j) + \sum_{j=1}^{n} \phi(t_i - t_j)} & \text{for } t_i \in \boldsymbol{o} \\ 0 & \text{else} \end{cases}
\end{aligned} \tag{5.29}$$

in place of the missing data indicator variables (5.25), while decoupling of the components in the likelihood (5.26) is preserved. These probabilities follow from the thinning [63] whereby the probability that t_i comes from one of the independent (sub-)processes is equal to that process' share of the total conditional intensity function at t_i. For specified model components with given parameter estimate $\widehat{\boldsymbol{\beta}}$, these probabilities (5.29) can be computed.

Maximization (M-step) The maximization of the expected log-likelihood (5.27) has the structure of probability density estimation with sample weights (5.29) in place of the indicator variables encoding the missing data. This decoupling into iid density estimation enables the estimation of relatively complex, as well as non-parametric densities. Specifically,

1. g is estimated on iid positive interevent times $\{ t_i - s_j,\ s_j < t_i \}$ with weights $\pi^{\theta}_{i,j}$,
2. f is estimated on iid positive interevent times $\{ t_i - t_j,\ t_j < t_i \}$ with weights $\pi^{\phi}_{i,j}$,
3. $\mu/\mu((0,t])$ is estimated on immigration times $\boldsymbol{s}$,[12]
4. m is estimated on the iid sample $\boldsymbol{y}$.

The branching ratio estimator $\widehat{\eta}$ is

$$\widehat{\eta} = \frac{\sum_{i,j} \pi^{\phi}_{i,j}}{\sum_{j=1}^{n} \int_0^{t-t_j} f(s)ds}, \tag{5.30}$$

with cumulative distribution in the denominator to correct for expected offspring truncated by the end of the observation window.

The E-step and M-step may then be iterated, and the estimates taken when the parameter estimates and log-likelihood have converged.

5.5.2.3 MCMC extension for unknown immigrants and marks

If we want to apply the EM algorithm in the more realistic case where we do not observe a sample $\boldsymbol{c}$ of the immigrant process C – i.e., we neither know which points are immigrants nor the values of their marks – the missing data (5.25) becomes $(\boldsymbol{Z}, C)$ and the objective function (5.27) takes the form

$$\begin{aligned} Q(\boldsymbol{\beta} \mid \widehat{\boldsymbol{\beta}}) &= \mathbb{E}_{\boldsymbol{Z},C|\widehat{\boldsymbol{\beta}},\boldsymbol{t}}[\log L(\boldsymbol{\beta} \mid \boldsymbol{t}, C, \boldsymbol{Z})] \\ &= \mathbb{E}_{C|\widehat{\boldsymbol{\beta}},\boldsymbol{t}}\left[\mathbb{E}_{\boldsymbol{Z}|\widehat{\boldsymbol{\beta}},\boldsymbol{t},C}[\log L(\boldsymbol{\beta} \mid \boldsymbol{t}, C, \boldsymbol{Z})]\right]. \end{aligned} \tag{5.31}$$

Assume we are able to generate samples $\boldsymbol{c}^{(1)}, \dots, \boldsymbol{c}^{(K)}$ from a density $p_C(\cdot \mid \widehat{\boldsymbol{\beta}}, \boldsymbol{t})$, we can then approximate the outer expectation with respect to C to approximate the objective function by

$$Q(\boldsymbol{\beta} \mid \widehat{\boldsymbol{\beta}}) \approx \frac{1}{K} \sum_{k=1}^{K} \left[\mathbb{E}_{\boldsymbol{Z}|\widehat{\boldsymbol{\beta}},\boldsymbol{t},\boldsymbol{c}^{(k)}}[\log L(\boldsymbol{\beta} \mid \boldsymbol{t}, \boldsymbol{c}^{(k)}, \boldsymbol{Z})]\right]. \tag{5.32}$$

To evaluate the inner expectation, we assume immigrants and marks are known ($C = \boldsymbol{c}^{(k)}$) and the procedure of section 5.5.2.2 applies. In particular, for the probabilities $\pi^{\phi,k}_{i,j}$ and $\pi^{\theta,k}_{i,j}$ calculated

[12] To recover the estimated intensity, this density is multiplied by n_c to satisfy $\int_0^{n_c} \mu(s)ds = n_c$.

according to (5.29) for $c^{(k)}$, then

$$Q(\boldsymbol{\beta} \mid \widehat{\boldsymbol{\beta}}) \approx \frac{1}{K}\sum_{k=1}^{K}\sum_{i=1}^{n_c(k)} \log\left(\mu\left(s_i^{(k)}\right)\right) - \int_0^t \mu(s)ds + \frac{1}{K}\sum_{k=1}^{K}\sum_{i=1}^{n_c(k)} \log\left(m\left(y_i^{(k)}\right)\right) +$$
$$\frac{1}{K}\sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j=1}^{n_c(k)} \pi_{i,j}^{\theta,k} \log\left[y_j^{(k)}\theta\left(t_i - s_j^{(k)}\right)\right] - \frac{1}{K}\sum_{k=1}^{K}\sum_{j=1}^{n_c(k)} y_j^{(k)} \int_0^t \theta\left(s - s_j^{(k)}\right) ds + \tag{5.33}$$
$$\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{1}{K}\sum_{k=1}^{K}\pi_{i,j}^{\phi,k}\right)\log\left[\phi(t_i - t_j)\right] - \sum_{j=1}^{n}\int_0^t \phi(s - t_j)ds \; .$$

Thus, as before in section 5.5.2.2, maximization over parameters $\boldsymbol{\beta}$ is decoupled into (weighted) iid density estimation and g, f, μ, m and η can be estimated separately. The estimation of θ involves a pooling of interevent times $\{t_i - s_j^{(k)}\}$ for $k = 1, \ldots, K$; while the estimation of ϕ merely requires weights averaged over the ensemble.. However, the density of the immigrants

$$p(\boldsymbol{c} \mid \boldsymbol{t}) = \frac{p(\boldsymbol{t} \mid \boldsymbol{c})p(\boldsymbol{c})}{p(\boldsymbol{t})} \tag{5.34}$$

is not known analytically due to the lack of an expression for the denominator of (5.34).

Conditional simulation of MCMC sample. Instead, we employ a simple MCMC algorithm to simulate immigrant realizations from this distribution. The algorithm is extended from [187] and [188] (sections 7.1.2 and 10.2.1) and uses a Metropolis Hastings algorithm to generate a Markov chain sample from the unnormalized density $p_{C|T}(\cdot \mid \boldsymbol{t})p_T(\boldsymbol{t})$.

Note that the joint probability $p(\boldsymbol{t}, \boldsymbol{c}) = 0$ if $\boldsymbol{c} \nsubseteq \boldsymbol{t}$ and thus

$$p(\boldsymbol{c} \mid \boldsymbol{t}) = K_1(\boldsymbol{t})p\left(\boldsymbol{t} \mid \boldsymbol{c}\right)\prod_{s_i \in \boldsymbol{s}} \mu(s_i) = K_2(\boldsymbol{t})p\left(\boldsymbol{o} \mid \boldsymbol{c}\right)\prod_{s_i \in \boldsymbol{s}} \mu(s_i), \tag{5.35}$$

with constants K_1, K_2 depending on $\boldsymbol{t}$. Given the marked immigrants, the offspring $\boldsymbol{o}$ are a realization from a modified ARMA (cf. section 5.3.4) with (realized) conditional intensity

$$\lambda_{O|C}(s) \approx \sum_{s_j \in \boldsymbol{s}} y_j \theta(s - s_j) + \sum_{t_j \in \boldsymbol{t}} \phi(s - t_j) \; , \tag{5.36}$$

ignoring the influence of points on $(-\infty, 0)$. Then, using proposition 7.2 III in [63],

$$p\left(\boldsymbol{o}|\boldsymbol{c}\right) = \prod_{t_i \in \boldsymbol{o}} \lambda_{O|C}(t_i)\text{Exp}\{-\int_0^t \lambda_{O|C}(s)ds\}. \tag{5.37}$$

Metropolis-Hastings iteration. Given the current state of the Markov chain, $\boldsymbol{c}^{(k)} = \boldsymbol{c} = (\boldsymbol{s}, \boldsymbol{y})$, our Metropolis-Hastings iteration consists of proposing either the birth or death of an immigrant with probability 1/2. In the case of birth, choose $s^* \in \boldsymbol{o}$ uniformly among offspring $\boldsymbol{o} = \boldsymbol{t} \setminus \boldsymbol{s}$ and generate a mark y^* from $m(\cdot)$. Then, using equations (5.35)-(5.37) and $c^* = (s^*, y^*)$, the

Metropolis-Hastings birth ratio[13] reads

$$
\begin{aligned}
r_b(\boldsymbol{c}, c^*) &= \frac{n - n_c}{n_c + 1} \frac{p(\boldsymbol{c} \cup c^* \mid t)}{p(\boldsymbol{c} \mid t)} \\
&= \frac{n - n_c}{n_c + 1} \mu(s^*) \mathrm{Exp}\{-\int_0^t y^* \theta(s - s^*) ds\} \times \\
&\prod_{\substack{t_i \in \boldsymbol{o} \\ t_i \neq s^*}} \left(1 + \frac{y^* \theta(t_i - s^*)}{\sum_{s_j \in \boldsymbol{s}} y_j \theta(t_i - s_j) + \sum_{t_j \in \boldsymbol{t}} \phi(t_i - t_j)}\right) \Bigg/ \left(\sum_{s_j \in \boldsymbol{s}} y_j \theta(s^* - s_j) + \sum_{t_j \in \boldsymbol{t}} \phi(s^* - t_j)\right),
\end{aligned}
\tag{5.38}
$$

and we accept and move to the new state $\boldsymbol{c}^{(k+1)} = \boldsymbol{c} \cup c^*$ with *acceptance probability* $\min\{1, r_b(\boldsymbol{c}, c^*)\}$. In the case of death, we choose $s^* \in \boldsymbol{s}$ uniformly and arrive at a death ratio

$$
r_d(\boldsymbol{c}, c^*) = \frac{1}{r_b(\boldsymbol{c} \setminus c^*, c^*)} \tag{5.39}
$$

and move to the new state $\boldsymbol{c}^{(k+1)} = \boldsymbol{c} \setminus c^*$ with *acceptance probability* $\min\{1, r_d(\boldsymbol{c}, c^*)\}$.

[13] C.f. equation (7.6) in [188].

5.5.3 Step-wise description of the EM algorithm

We use the notation of section 5.5.2.

0. Start with initial parameter "guess" $\boldsymbol{\beta}^{(0)}$ and a set of immigrants $\boldsymbol{c}^{(0)} \subseteq \boldsymbol{t}$.
For every iteration $r \in \mathbb{N}$, repeat E and M step as follows.
I. E-Step (MCMC)
Generate Markov chain of immigrants $\boldsymbol{c}^{(r,1)}, \ldots, \boldsymbol{c}^{(r,K)}$ given $\boldsymbol{\beta}^{(r-1)}$. In every iteration k
1. Flip a coin to choose birth or death.
2a. In case of birth, choose $\boldsymbol{s}^* \in \boldsymbol{o}^{(r,k-1)}$ uniformly and generate y^* from $m^{(r-1)}$, calculate the birth ratio $r_b(\boldsymbol{c}^{(r,k-1)}, c^*)$ in (5.38) and accept $\boldsymbol{c}^{(r,k)} = \boldsymbol{c}^{(r,k-1)} \cup (s^*, y^*)$ with probability $\min\{1, r_b\}$.
2b. In case of death, choose $c^* \in \boldsymbol{c}^{(r,k-1)}$ uniformly, calculate the death ratio $r_d(\boldsymbol{c}^{(r,k-1)}, c^*)$ in (5.39) and accept $\boldsymbol{c}^{(r,k)} = \boldsymbol{c}^{(r,k-1)} \setminus c^*$ with probability $\min\{1, r_b\}$
3. In case of non-acceptance in 2a. or 2b., set $\boldsymbol{c}^{(r,k)} = \boldsymbol{c}^{(r,k-1)}$
4. Calculate the probabilities $\pi_{i,j}^{\theta,k}, \pi_{i,j}^{\phi,k}$ as defined in (5.29) for $\boldsymbol{c}^{(r,k)}$ and $\boldsymbol{\beta}^{(r-1)}$
5. Set $k \to k+1 \leqslant K$ and return to 1.
II. M-step (Decoupled density estimation)
6. Based on the MCMC-generated set of immigrants and probabilities,
$\left(\boldsymbol{c}^{(r,k)}, \pi_{i,j}^{\theta,k}, \pi_{i,j}^{\phi,k}\right)_{k=1}^{K}$, maximize the decoupled objective function (5.33) to get

$$\boldsymbol{\beta}^{(r)} = \arg\max_{\boldsymbol{\beta}} Q(\boldsymbol{\beta} \mid \boldsymbol{\beta}^{(r-1)})$$

7. Set $r \to r+1$ and return to step 1.

The algorithm may terminate after the parameters have converged, according to the selected criterion. Modifications of the algorithm for the unmarked case, and for immigrants excluded, are relatively straightforward and discussed in Appendix 5.9.3.

5.5.4 Comments

The algorithm requires storage and computation with matrices that are $O(n^2)$. On a standard PC, this makes computation prohibitive for samples with $n > 10^4$. However this implementation is crude as, in this case, even the largest interevent time $t_n - t_1$ is considered as an interevent time by which t_n could be triggered via $\theta(t_n - t_1)$ or $\phi(t_n - t_1)$, despite the fact that the probability of this could be effectively 0. Thus, for window size t large relative to the support of θ and ϕ, one can safely omit interevent times above a certain threshold. The result of this is banded/sparse matrices which reduce storage and computation from n^2 to $n \times m$, potentially with $m << n$. In the case of the INARMA model, storage is less of an issue as only counts in bins need to be stored, rather than the location of each point.
Technically, the use of MCMC in the E-step of the EM algorithm makes it a Monte-Carlo EM (MCEM) algorithm, having slightly weaker properties than the pure EM [245]. Regarding the

convergence of the EM algorithm [76, 209, 252], one first needs that the necessary MLE regularity conditions are satisfied [193], for instance having smooth distributions that are not too heavy tailed. Next, it must be ensured that the sequence of parameter estimates does not reach the boundary of the parameter space. For instance, if estimates μ, η, or γ are equal to zero at any iteration, or equivalently, if the support of $f(.)$ or $g(.)$ is smaller than the smallest interevent time, then the estimates will remain zero (5.30). However, given non-zero starting estimates, the EM algorithm estimates satisfy the constraints of the model parameters. Regarding speed of convergence, there is the general result of [76, 252] that the algorithm will not worsen the likelihood with each iteration. Further, from [209], given that Q (5.27) is differentiable in $\boldsymbol{\beta}$ and the M-step has a unique solution, then the EM algorithm iterates in a positive direction on the true likelihood surface. Finally, when the missing information is small compared to the complete information, EM exhibits approximate Newton behavior with superlinear convergence near the true optimum. In terms of the ARMApp, as well as other mixture type models, this means that, when clusters are overlapping, convergence will be slow, as has been shown for the Hawkes process with exponential offspring distribution [173, 235], as well as other mixture models [209, 254]. In particular for the NS process with immigrants included, since the number of offspring in a cluster is Poisson distributed, it will be difficult to distinguish heavily overlapping clusters from a single homogenous Poisson process. Exactly deriving the convergence properties of the ARMApp model for a given parameterization would add no general insight.

5.6 Simulation study

Here a range of simulation studies are given to demonstrate performance and some relevant issues. Using the notation from section 5.3.1 , simple model specifications are employed: we use both unmarked or exponentially distributed marks, moderate self-excitation, constant immigration and exponential kernels with scale parameters θ_1^{-1} and ϕ_1^{-1}, that is,

$$\theta(t) = \gamma\theta_1 e^{-\theta_1 t} \quad \text{and} \quad \phi(t) = \eta\phi_1 e^{-\phi_1 t} \,. \tag{5.40}$$

Simulation is very fast, with details in appendix 5.9.4.

First is a demonstration of the performance of MCEM estimation of the simplest specification of the marked ARMApp (5.13): exponential kernels with short memory, exponential mark distribution, moderate self-excitation, and constant immigration. The results of the repeated simulation and estimation are summarized in Tab 5.1. The main insight is that the estimation performs well. Further, as a robustness test, allowing for the immigration to be too flexible – having up to 20 degrees of freedom, where the true immigration has only 1 – has little impact on the estimated parameters. However, estimation is likely to be less robust to such an error when the immigration can better approximate the clusters (e.g., when clusters are larger and longer).

p	0	3	5	10	15	20
μ	0.979	0.0999	0.0995	0.1010	0.1014	0.1019
γ	4.21	4.27	4.27	4.16	4.13	4.03
η	0.51	0.51	0.50	0.50	0.51	0.50

Table 5.1: Average parameter estimates of the marked ARMA, excluding memory scale parameters, averaged over 300 independent replications (simulations and estimations) done for a range of degrees of freedom (p) of the estimated immigration, using the R:logspline estimator. The standard deviation of the estimated parameters are about 0.008, 0.38, and 0.02 resp. (for all p). The simulated data is from the marked ARMA with parameters $\mu = 0.1$, $\gamma = 4$, $\theta_1^{-1} = 0.1$, $\eta = 0.5$, $\phi_1^{-1} = 1$, with average sample size of 2000. In this case, γ is the scale parameter of the exponential mark distribution. Estimation is done by the MCEM algorithm, with K=50 MC samples taken from the chain of 300'000 Metropolis Hastings iterations.

Next, acknowledging the lack of accessible likelihood as well as residuals to perform model selection or testing, the ability of the MCEM algorithm to "select" the correct (sub-)model within the ARMA framework is important. E.g., on an AR simulation, the fitted ARMA should converge to have negligible γ and consistently recover the AR part. To test this consistency, pure AR, pure MA, and the full ARMA are fit to AR, MA, and ARMA simulations, respectively, with results summarized in Tab. 5.2. For both AR and MA simulated data, the fitted ARMA converges well to the true sub-model. And for the full ARMA model, both AR and MA parts are well estimated. Of course, the range of misspecified models (e.g., the AR fit to the ARMA simulation) also provide estimates. Without an objective criterion to compare these different fits, it is therefore best to start with the broadest overall model – in this case the ARMA – and allow the MCEM algorithm to converge and select the relevant nested model. However, better methods to compare and test models are highly desirable.

Fitted	μ	γ	θ_1^{-1}	η	ϕ_1^{-1}
ARMA sim.					
ARMA	0.095 (0.00025)	5.52 (0.288)	2.03 (0.12)	0.47 (0.0024)	0.097 (0.00019)
NS/MA	0.21 (0.004)	4.84 (1.62)	0.67 (0.049)	–	–
Hawkes/AR	0.27 (0.003)	–	–	0.77 (0.0014)	0.22 (0.0015)
MA sim.					
ARMA	0.10 (0.00016)	6.98 (0.10)	1.96 (0.017)	0.01 (0.0001)	2.06 (5.55)
NS/MA	0.10 (0.002)	7.06 (0.09)	1.98 (0.016)	–	–
Hawkes/AR	0.16 (0.001)	–	–	0.81 (0.0007)	0.83 (0.0066)
AR sim.					
ARMA	0.09 (0.0002)	0.13 (0.10)	5.58 (54.2)	0.79 (0.003)	0.099 (0.0001)
NS/MA	0.13 (0.002)	2.73 (0.33)	0.18 (0.0006)	–	–
Hawkes/AR	0.10 (0.0001)	–	–	0.79 (0.0023)	0.10 (0.00001)

Table 5.2: Average (and standard deviation) of parameter estimates over 100 independent replications. The first set are for data simulated from the ARMA with parameters ($\mu = 0.1, \gamma = 5, \theta_1^{-1} = 2, \eta = 0.5, \phi_1^{-1} = 0.1$) with the same unmarked ARMA (5.2), and nested MA, and AR sub-models fit respectively. The second set are for data simulated from the MA with parameters ($\mu = 0.1, \gamma = 7, \theta_1^{-1} = 2$), and the final set for data simulated from the AR (Hawkes) with parameters ($\mu = 0.1, \eta = 0.8, \phi_1^{-1} = 0.1$). All simulations contain 500 points, and the MCEM estimation is done with K=50 MC samples taken from a chain of length 300'000 generated by the Metropolis-Hastings algorithm.

Finally, the effect of cluster overlap is briefly examined, by fitting the unmarked ARMA, with exponential kernels, and varying the immigration intensity. It is intuitive that more heavily overlapping clusters are more difficult to distinguish from immigration. As summarized in Table 5.3, as the clusters becomes more overlapping, indeed the immigration intensity becomes increasingly overestimated and in particular γ underestimated.

μ_{sim}	0.10	0.31	0.52	0.73	0.94	1.16	1.37	1.58	1.79	2.00
μ	0.10	0.32	0.57	0.78	1.05	1.33	1.62	1.80	2.04	2.23
γ	6.17	5.97	5.76	5.19	4.76	4.36	4.21	4.23	3.89	3.74
θ_1^{-1}	0.54	0.54	0.52	0.47	0.44	0.42	0.40	0.40	0.37	0.34
η	0.39	0.39	0.40	0.47	0.47	0.49	0.49	0.54	0.55	0.60
ϕ_1^{-1}	0.24	0.24	0.25	0.28	0.29	0.31	0.31	0.31	0.31	0.30
sd(μ)	0.011	0.041	0.085	0.165	0.258	0.342	0.435	0.542	0.605	0.773
sd(γ)	0.426	0.740	0.956	1.288	0.994	1.284	1.091	1.350	1.103	1.734
sd(θ_1^{-1})	0.041	0.060	0.072	0.095	0.092	0.097	0.075	0.078	0.052	0.066
sd(η)	0.059	0.096	0.107	0.115	0.105	0.132	0.113	0.126	0.137	0.134
sd(ϕ_1^{-1})	0.033	0.040	0.042	0.071	0.078	0.104	0.072	0.068	0.072	0.064

Table 5.3: Average (top rows) and standard deviation (lower rows) for estimated parameters, computed over 500 independent replications for each value of immigration intensity used in the simulation, μ_{sim}. The data are simulated from the unmarked ARMA with parameters $(\mu_{sim}, \gamma = 5, \theta_1^{-1} = 0.5, \eta = 0.5, \phi_1^{-1} = 0.25)$ and with each realization containing 1000 points. The MCEM estimation is done with K=50 MC samples taken from a chain of length 300'000 generated by the Metropolis-Hastings algorithm.

5.7 Exemplary case study: endogeneity in financial markets

To briefly illustrate the use and limitations of the ARMApp methodology developed here, we consider modelling the event times of price changes of financial time series. This is an important application of point processes, due to regulatory and risk management interest in the potential of large financial losses due to clustering of extreme price fluctuations. Hawkes processes have been applied here (e.g., see [45]). However, in principle, inhomogeneities, shocks and contagious dynamics may be present, and thus a more general ARMA model worth considering.
In particular, here we analyze price changes of a single trading day for S&P 500 E-mini futures. We can utilize the powerful EM framework and allow for flexible (time-dependent) immigration intensity $\mu(t)$. To estimate μ on interevent times of the immigrant process (see section 5.5 for details), we use the R:logspline estimator with $p = 1, \ldots, 15$ degrees of freedom. A value of $p = 1$ corresponds to constant immigration $\mu(t) \equiv \mu$. Fig. 5.3 shows the fitted parameters for a Hawkes process with conditional intensity function

$$\lambda(t) = \mu(t) + \sum_{k=-\infty}^{N_t} \eta\phi_1 e^{-\phi_1(t-T_k)} . \tag{5.41}$$

Allowing a flexible immigration intensity $\mu(t)$ is highly consequential, as the estimates for η drop substantially. In the Hawkes process framework, this has been discussed in [250]. To illustrate the effect of the ARMA model, fig. 5.4 shows the fitted parameters for an ARMApp with conditional intensity function

$$\lambda(t) = \mu(t) + \sum_{j=-\infty}^{N_t^\mu} Y_j\theta_1 e^{-\theta_1(t-T_j^\mu)} + \sum_{k=-\infty}^{N_t} \eta\phi_1 e^{-\phi_1(t-T_k)} , \tag{5.42}$$

where $(Y_j)_{j\in\mathbb{N}}$ is a sequence of i.i.d. exponentially distributed marks with $\mathbb{E}\left[Y_j\right] = \gamma$. Allowing the immigrant process to trigger shot-noise bursts (that is, moving from Hawkes to ARMApp) further reduces the parameter η and thus the estimated endogeneity. To demonstrate the flexibility of the estimation algorithm, we implement a mixture distribution for the marks, where only some immigrants cause an exogenous burst of points. In particular, fig. 5.5 shows the fitted parameters of an ARMApp with conditional intensity function

$$\lambda(t) = \mu(t) + \sum_{j=-\infty}^{N_t^\mu} Y_j\theta_1 e^{-\theta_1(t-T_j^\mu)} + \sum_{k=-\infty}^{N_t} \eta\phi_1 e^{-\phi_1(t-T_k)} , \tag{5.43}$$

where $(Y_j)_{j\in\mathbb{N}}$ is a sequence of i.i.d. distributed marks that are either zero or equal to an exponentially distributed random variable with mean τ_m, that is,

$$Y_j = \begin{cases} 0 & \text{with propability } 1-q, \\ \mathrm{Exp}(1/\tau_m) & \text{with propability } q. \end{cases} \tag{5.44}$$

Following the discussion in section 5.5.4, the fact that the EM algorithm increases the likelihood of parameter estimates (for p arbitrary but fixed), we can conclude on a maximum-likelihood basis that the ARMApp model is superior (in a descriptive sense) to the Hawkes process and allows for a better estimate of endogeneity in the price evolution in S&P 500 E-mini futures. To assess the size and significance of this improvement and to objectively choose the appropriate flexibility p of immigration, a comparison of (unconditional) likelihoods across models will be necessary. One possibility is to estimate *likelihood ratios* and *observed Fisher information* using MCMC algorithms as discussed in, e.g., chapter 8 of [188], and is left for future research.

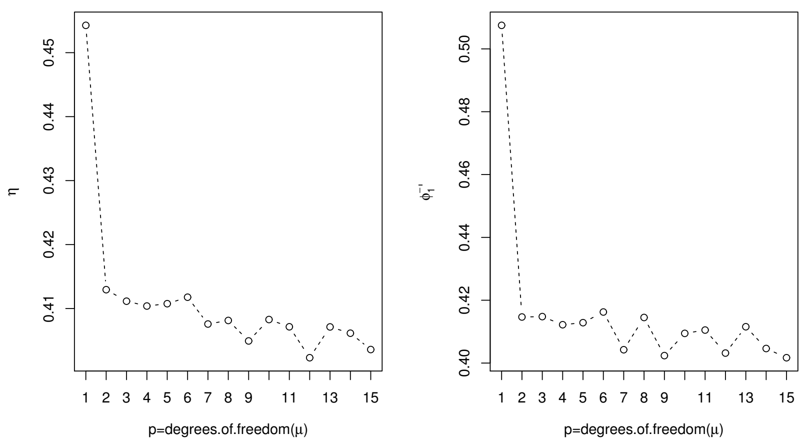


Figure 5.3: Estimated parameter values of the Hawkes point process with exponential kernel, fitted to time points of price changes of S&P 500 E-mini futures. The time scale of the memory kernels is seconds.

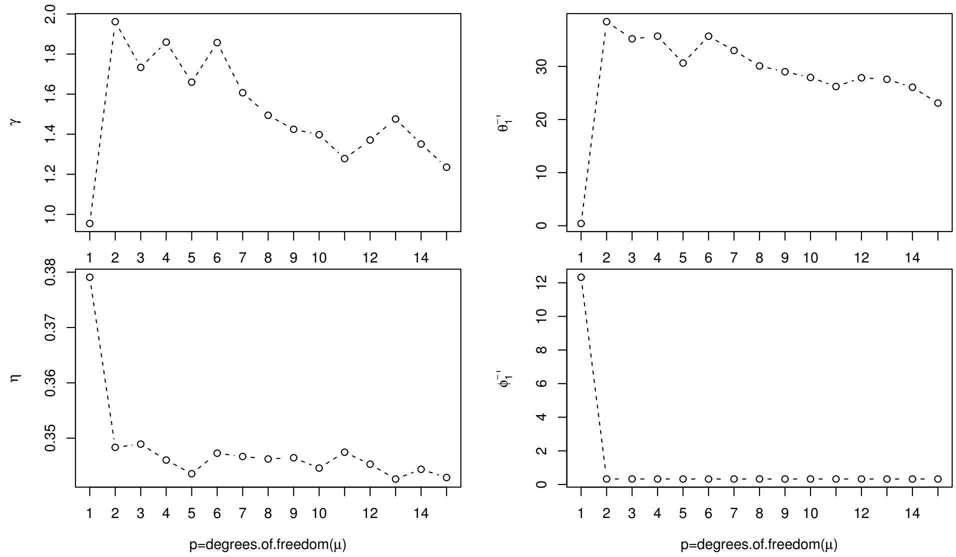


Figure 5.4: Estimated parameter values of the ARMApp with exponential kernels and exponential marks, fitted to time points of price changes of S&P 500 E-mini futures. The time scale of the memory kernels is seconds.

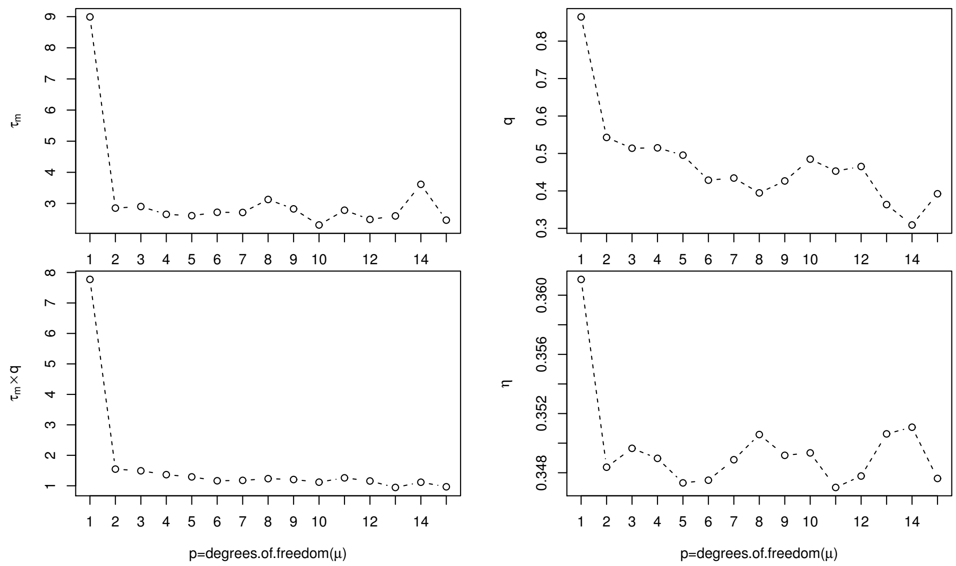


Figure 5.5: Estimated parameter values of the ARMApp with exponential kernels and mixture distribution for marks, fitted to time points of price changes of S&P 500 E-mini futures.

5.8 Discussion

Here the ARMApp was introduced, being the point process analogue of the INARMA time series model, combining bursty (shot noise) and auto-regressive dynamics (the Hawkes process). By simulation study, an MCEM algorithm was shown to reliably perform MLE for this model, also consistently converging to sub-models – performing a selection among the models nested within the ARMA. To assess significant differences and choose a suitable model for immigration, likelihood ratios and confidence intervals can be estimated with MCMC, which is left for future research (cf. the discussion in the section 5.7 above). Further, the estimation algorithm allows for a variety of parametric specifications, which need not be confined to the Markovian case, nor to fixed immigration intensity. Model applicability was argued based on the known presence of both bursty phenomena and auto-regressive self-excitation in various fields, as well as by analogy to the INARMA. This work may benefit from developments on the time series side, such as reversible jump MCMC for model selection [85, 190], as well as techniques for likelihood quantification [7, 242], which has been shown to be important for distinguishing exogenous fluctuations in immigration from bursts and endogenous contagion dynamics [250].

5.9 Appendix

5.9.1 Point process essentials

For convenience of the reader, we introduce basic terminology and processes from the point process literature, using the notation from section 5.2. For a thorough introduction to point processes, we refer to [59] or [63].
The essential building block is the Poisson process, and the only object for which distributional properties are specified.

Definition 5.9.1 (Poisson process). Let $\rho : (-\infty, \infty) \to [0, \infty)$ be a locally integrable function. Then a point process $N : \Omega \times \mathcal{B}(\mathbb{R}) \to \mathbb{N}_0$ is called **Poisson process** if for every finite collection of disjoint, bounded Borel sets $\{A_1, \dots, A_k\}$ and nonnegative integers $\{n_1, \dots, n_k\}$ we have

$$\mathbb{P}\left[N(A_i) = n_i, i = 1, \dots, k\right] = \prod_{i=1}^{k} \frac{\left(\int_{A_i} \rho(s)ds\right)^{n_i}}{n_i!} e^{-\int_{A_i} \rho(s)ds}. \tag{5.45}$$

The function ρ is called **intensity** or **rate function** of N. If $\rho \equiv$const, then N is called **homogeneous**, otherwise **inhomogeneous** Poisson process.

Using this notion, we introduce a Poisson cluster of points that is centered at some (random) cluster center, with subsequent points being distributed according to an inhomogeneous Poisson process with a given intensity.

Definition 5.9.2 (Poisson cluster). Let $\rho : [0, \infty) \to [0, \infty)$ be a locally integrable function and $T : \Omega \to \mathbb{R}$. Then a point process $N : \Omega \times \mathcal{B}(\mathbb{R}) \to \mathbb{N}_0$ is called **(ρ-)Poisson cluster** if

1. $N\left((-\infty, T]\right) = N\left(\{T\}\right) = 1$ $\mathbb{P}$-almost surely, and
2. conditional on T, $(N_{s+T} - N_T)_{s\in[0,\infty)}$ follows an inhomogeneous Poisson process with intensity ρ.

The random time point T is called **parent**, the subsequent points of N are called (ρ-)**offspring triggered by T**.

The expected number of Points in a Poisson cluster is given by $\int_0^\infty \rho(s)ds \in [0, \infty]$.
If, for some random cluster center, every point in a Poisson cluster is itself the center of a Poisson cluster, whose points in turn trigger Poisson clusters (and so forth), the collection of these points form a so-called Galton-Watson cluster. The idea is formalized in the following definition.

Definition 5.9.3 (Galton-Watson cluster). Let $\rho : [0, \infty) \to [0, \infty)$ be an integrable function and $T : \Omega \to \mathbb{R}$. Then a point process $N : \Omega \times \mathcal{B}(\mathbb{R}) \to \mathbb{N}_0$ is called **(ρ-)Galton-Watson cluster** if it is the collection

$$N = \sum_{\substack{n\in\mathbb{N}_0 \\ j=1,\dots,Z_n}} C_j^{(n)} \tag{5.46}$$

for a random counting sequence $(Z_n)_{n\in\mathbb{N}}$ and ρ-Poisson clusters $C_j^{(n)}$ with the property that

1. $Z_0 = 1$, $C_1^{(0)} = \delta_T(\cdot)$ and, for all $n \in \mathbb{N}$,
2. $Z_{n+1} = \sum_{j=1}^{Z_n} C_j^{(n)}(\mathbb{R})$ and
3. every Poisson cluster $C_j^{(n)}$, $j = 1, \ldots, Z_n$, has a unique parent in

$$\sum_{j=1,\ldots,Z_{n-1}} C_j^{(n-1)}. \tag{5.47}$$

The sequence $(Z_n)_{n\in\mathbb{N}}$ counts the number of points in each generation n and Galton-Watson branching process, see [117]. A Galton-Watson cluster is finite if the expected cluster size of a single Poisson cluster satisfies $\int_0^\infty \rho(s)ds < 1$. In this case, the expected number of Points in a Galton-Watson cluster is equal to $1/(1 - \int_0^\infty \rho(s)ds)$.
Combining these basic cluster mechanisms with a process of cluster centers allows one to construct two fundamental cluster processes.

Definition 5.9.4 (Poisson cluster process). Let $\rho : [0, \infty) \to [0, \infty)$ be a locally integrable function and N^I be a point process with jump times $(T_j)_{j\in\mathbb{Z}}$. Then a point process $N : \Omega \times \mathcal{B}(\mathbb{R}) \to \mathbb{N}_0$ is called **(ρ-)Poisson cluster process** if it is the collection

$$N = \sum_{j\in\mathbb{Z}} C_j \tag{5.48}$$

for ρ-Poisson clusters C_j with parent T_j, respectively. The process N^I is called **parent process** or **immigration process**.

If the immigrant process N^I follows a homogeneous Poisson process, the resulting Poisson cluster process is a modified **Neyman-Scott process** ([191]), the modification being that immigrants are included in the count. This type of process can equivalently be considered as a **shot-noise** process, see example 6.2(a) in [63], with the interpretation that each immigrant increases the probability of further points. In the time series literature, this type of behavior is referred to as **moving-average**.

Definition 5.9.5 (Galton-Watson cluster process). Let $\rho : [0, \infty) \to [0, \infty)$ be an integrable function and N^I be a point process with jump time $(T_j)_{j\in\mathbb{Z}}$. Then a point process $N : \Omega \times \mathcal{B}(\mathbb{R}) \to \mathbb{N}_0$ is called **(ρ-)Galton-Watson cluster process** if it is the collection

$$N = \sum_{j\in\mathbb{Z}} C_j \tag{5.49}$$

for ρ-Galton-Watson clusters C_j with parent T_j, respectively. The process N^I is called **parent process** or **immigration process**.

If the immigrant process N^I follows a homogeneous Poisson process, the resulting Galton-Watson cluster process is a **Hawkes process**, see [120] or example 6.3(c) in [63]. As every point increases the probability of further points, the behavior of Galton-Watson processes is referred to as **self-exciting**. In the time series literature, this type of behavior is called **autoregressive**.

5.9.2 Derivation of the palm intensity in the exponential case

Below we will derive an explicit solution for the palm intensity function (5.7) of the ARMApp for exponential densities. In this goal, let the setting in section 5.3.1 be fulfilled, with $\theta_0, \theta_1, \phi_0 \in (0, \infty)$, $\phi_1 \in (\phi_0, \infty)$, $\theta(t) = \theta_0 e^{-\theta_1 t}$, $\phi(t) = \phi_0 e^{-\phi_1 t}$, with h and τ be defined by (5.7) and (5.8), respectively, and $\rho = \overline{\lambda} h$. The Laplace transform of a function $f : [0, \infty) \to \mathbb{R}$ will be denoted by f^*.

To get an explicit solution for τ, we Laplace-transform equation (5.10) to get

$$\tau^*(s) = \frac{\mu^2(1+\gamma)}{s} + \frac{\mu\phi_0 + \overline{\lambda}\mu\eta}{\phi_1 + s} + \frac{\mu\theta_0}{\theta_1 + s} + \frac{\phi_0}{\phi_1 + s}\tau^*(s). \tag{5.50}$$

Rearranging terms to solve for $\tau^*(s)$ and doing a partial fraction decomposition allows us to invert the Laplace transform to arrive at

$$\begin{aligned} \tau(u) &= \overline{\lambda}\mu + \left(L_1 e^{-(\phi_1-\phi_0)u} + L_2 e^{-\theta_1 u}\right)\mathbf{1}\{u > 0\}, \text{ where} \\ L_1 &= \mu\phi_0 \frac{\theta_1 + \theta_0 - \phi_1 + \phi_0}{\theta_1 - \phi_1 + \phi_0}, \\ L_2 &= \mu\theta_0 \frac{\theta_1 - \phi_1}{\theta_1 - \phi_1 + \phi_0}. \end{aligned} \tag{5.51}$$

Using this explicit solution for τ, we get from equation (5.11) that

$$\begin{aligned} \rho(u) = {} & \mu\overline{\lambda}(1+\gamma) + e^{-\theta_1 u}\left[\mu\theta_0 + \frac{L_1\theta_0}{\theta_1 + \phi_1 - \phi_0} + \frac{L_2\gamma}{2}\right] + e^{-\phi_1 u}\overline{\lambda}\phi_0 \\ & + \int_0^u \phi(x)\rho(u-x)dx + \int_0^\infty \phi(x+u)\rho(x)dx. \end{aligned} \tag{5.52}$$

Analogously to the solution for τ, we do a Laplace transformation of equation (5.52), partial fraction fraction decomposition and an inverse transform to arrive at

$$\rho(u) = \overline{\lambda}^2 + K_1 e^{-(\phi_1-\phi_0)u} + K_2 e^{-\theta_1 u}, \tag{5.53}$$

where the constants $K_1, K_2 \in [0, \infty)$ are given by

$$\begin{aligned} K_0 &= \mu\theta_0 + \mu\phi_0 \frac{\theta_1 + \theta_0 - \phi_1 + \phi_0}{\theta_1 - \phi_1 + \phi_0} \frac{\theta_0}{\theta_1 + \phi_1 - \phi_0} + \frac{1}{2} \frac{\mu\theta_0(\theta_1 - \phi_1)\theta_0}{(\theta_1 - \phi_1 + \phi_0)\theta_1} \\ K_1 &= K_0 \left(\frac{\phi_1\phi_0}{\theta_1 + \phi_1} - \frac{\phi_0}{\phi_1 - \phi_0 - \theta_1}\right) + \overline{\lambda}\phi_0 \left(1 + \frac{\phi_0}{2(\phi_1 - \phi_0)}\right) \\ K_2 &= K_0 \frac{\phi_1 - \theta_1}{\phi_1 - \phi_0 - \theta_1}. \end{aligned} \tag{5.54}$$

5.9.3 EM algorithms for the ARMApp

5.9.3.1 Unmarked ARMA

If T is an ARMApp as in section 5.5.2 without marks and $\sigma(N^\mu, N)$-conditional intensity

$$\lambda(s) = \mu(s) + \sum_{j=-\infty}^{N^\mu(s)} \theta(s - T_j^\mu) + \sum_{k=-\infty}^{N(s)} \phi(s - T_k) \ , \tag{5.55}$$

estimation is almost identical. The main difference is that $\gamma = \int_0^\infty \theta(s)ds$ will be included as a parameter instead of the density of marks, m. The complete data likelihood (5.26) becomes

$$\begin{gathered} L(\boldsymbol{\beta} \mid \boldsymbol{t}, \boldsymbol{c}, \boldsymbol{Z}) = \prod_{i=1}^{n_c} \mu(s_i) \mathrm{Exp}\{-\int_0^t \mu(s)ds\} \times \\ \prod_{i=1}^{n} \prod_{j=1}^{n_c} \left[\theta(t_i - s_j)\right]^{Z_{i,j}^\theta} \mathrm{Exp}\{-\sum_{j=1}^{n_c} \int_0^t \theta(s - s_j)ds\} \times \\ \prod_{i=1}^{n} \prod_{k=1}^{n} \left[\phi(t_i - t_k)\right]^{Z_{i,k}^\phi} \mathrm{Exp}\{-\sum_{j=1}^{n} \int_0^t \phi(s - t_j)ds\} \end{gathered} \tag{5.56}$$

and γ is estimated by

$$\widehat{\gamma} = \frac{\sum_{i,j} \pi_{i,j}^\theta}{\sum_{j=1}^{n_c} \int_0^{t-s_j} g(s)ds} . \tag{5.57}$$

All equations of sections 5.5.2.2, 5.5.2.3 and 5.5.3 extend using constant marks $y \equiv 1$.

5.9.3.2 Immigrants not included in the sample

As introduced in section 5.3.4, one may wish to estimate a modification of the ARMApp, where the immigrant process N^μ is not included in the process N and the $\sigma(N^\mu, N)$-conditional intensity is given by

$$\lambda(s) = \sum_{j=-\infty}^{N^\mu(s)} Y_j \theta(s - T_j^\mu) + \sum_{k=-\infty}^{N(s)} \phi(s - T_k) \ . \tag{5.58}$$

To describe the (mostly straightforward) extension of the algorithm above, additional to the notation of section 5.5.2, we will need $\tilde{\boldsymbol{t}} = \boldsymbol{t} \cup \boldsymbol{s}$, the *ordered* union of points and immigrants with $\tilde{n} = n + n_c$ elements. The indicator variables (5.25) have a modified domain

$$\boldsymbol{Z} = \{Z_{i,j}^\theta \ , \ i = 1, ..., n, \ j = 1, ..., n_c\} \cup \{Z_{i,j}^\phi \ , \ i = 1, ..., \tilde{n}, \ j = 1, ..., \tilde{n}\} \ , \tag{5.59}$$

and the complete data likelihood becomes

$$
\begin{aligned}
L(\boldsymbol{\beta} \mid \boldsymbol{t}, \boldsymbol{c}, \boldsymbol{Z}) = &\prod_{i=1}^{n_c} \mu(s_i) \mathrm{Exp}\{-\int_0^t \mu(s) ds\} m(y_i) \times \\
&\prod_{i=1}^{n} \prod_{j=1}^{n_c} \Big[y_j \theta(t_i - s_j)\Big]^{Z^{\theta}_{i,j}} \mathrm{Exp}\{-\sum_{j=1}^{n_c} y_j \int_0^t \theta(s - s_j) ds\} \times \\
&\prod_{i=1}^{n} \prod_{k=1}^{\tilde{n}} \Big[\phi(t_i - \tilde{t}_k)\Big]^{Z^{\phi}_{i,k}} \mathrm{Exp}\{-\sum_{j=1}^{\tilde{n}} \int_0^t \phi(s - \tilde{t}_j) ds\} \ .
\end{aligned}
$$

Moreover, the MCMC algorithm used in the E-step needs to be adapted. Given the current state of the Markov chain $\boldsymbol{c} = (\boldsymbol{s}, \boldsymbol{y})$, in a birth step, we choose an immigrant s^* uniformly in the time window $[0, t]$ (instead of choosing among offspring), while a mark y^* is generated from m. As before, in a death step, we choose an immigrant uniformly among the immigrant set $\boldsymbol{c}$. The Metropolis-Hastings birth ratio becomes

$$
\begin{aligned}
r_b(\boldsymbol{c}, c^*) = &\frac{|t|}{n_c + 1} \mu(s^*) \mathrm{Exp}\{-\int_0^t y^* \theta(s - s^*) + \phi(s - s^*) ds\} \times \\
&\prod_{t_i \in \boldsymbol{t}} \left(1 + \frac{y^* \theta(t_i - s^*) + \phi(t_i - s^*)}{\sum_{s_j \in \boldsymbol{s}} y_j \theta(t_i - s_j) + \sum_{\tilde{t}_j \in \tilde{\boldsymbol{t}}} \phi(t_i - \tilde{t}_j)}\right)
\end{aligned}
\tag{5.60}
$$

and the death ratio $r_d(\boldsymbol{c}, c^*) = 1/r_b(\boldsymbol{c} \setminus c^*, c^*)$.
Then the algorithm described in section 5.5.3 carries over with the obvious modifications of probabilities (5.29) and objective function (5.33).

5.9.4 Simulation of the ARMApp

Below a detailed algorithm for simulating ARMApp is presented. It exploits the fact that, given the branching structure, the innovation, MA, and AR processes are mutually independent inhomogeneous Poisson processes. Efficient simulation algorithms for the Hawkes process follow the same approach. At the end of step II, one has simulated an NS process. By skipping step II, and completing step III, one simulates a Hawkes process. To avoid edge effects, one should simulate on a large window, and discard the *burn in period*. When the kernels can be integrated and inverted, using inverse transform sampling [52] makes the algorithm very fast.

Simulation algorithm

I. Simulate the immigrant points

(i) Simulate Poisson process $\{T_i^{(0)}\}_{i\in 1,\dots,n_c}$ on the window $(0,t]$, where $N_t = n_c$.

II. Simulate MA points

(i) For each immigrant $i = 1,\dots,n_c$: simulate the number of offspring $N_i^{\theta} \overset{i.i.d}{\sim} \text{Pois}(\gamma)$, and then sample the N_i^{θ} inter-event times $S_{i,j},\ j = 1,\dots,N_i^{\theta}$, i.i.d from pdf g. If $N_i^{\theta} = 0$, simulate zero inter-event times for that immigrant.

(ii) The MA points generated by the i^{th} immigrant are then $\{T_{i,j}^{\theta}\}_{j=0:N_i^{\theta}} = T_i^{(0)} + \{S_{i,j}\}_{j=0:N_i^{\theta}}$.

(iii) The immigrant and MA points together are $\{T_i^{(1)}\}_{i\in 1,\dots,n_1} = \{T_i^{(0)}\} \cup \{T_{1,j}^{\theta}\} \cup \cdots \cup \{T_{n_c,j}^{\theta}\}$, where $n_1 = n_0 + \sum_{i=1}^{n_c} N_i^{\theta}$.

III. Simulate AR points by generation

(i) Set the fertile points $A = \{1,\dots,n_1\}$, generation $k = 1$, and zeroth generation points $\{T_i^{\phi[0]}\} = \{T_i^{(1)}\}$.

(ii) For the current generation k, for all fertile points $\forall i \in A$: simulate the number of direct offspring $N_i^{\phi[k]} \overset{i.i.d}{\sim} \text{Pois}(\eta)$, and then the $N_i^{\phi[k]}$ inter-event times $S_{i,j}^{\phi[k]},\ j = 1,\dots,N_i^{\phi[k]}$, i.i.d from pdf f. If $N_i^{\phi[k]} = 0$, simulate zero inter-event times for that point.

(iii) The AR points generated by the i^{th} point are then $\{T_{i,j}^{\phi[k]}\}_{j=0:N_i^{\phi[k]}} = T_i^{\phi[k-1]} + \{S_{i,j}^{\phi[k]}\}_{j=0:N_i^{\phi[k]}}$ and

(iv) the union of these sets, $\{T^{\phi[k]}\} = \{T_{1,j}^{\phi[k]}\} \cup \cdots \cup \{T_{N_i^{\phi[k]},j}^{\phi[k]}\}$, is the offspring of generation k. (v) Update the fertile set $A = \{i : T_i^{\phi[k]} < r\}$ to be all points born in the current generation k that fall within the window $(0,t]$.

(vi) If A is non-empty, then increment the generation ($k = k+1$) and return to (ii), otherwise return the realization formed by joining all generations: $\{T_i\}_{i=1:n} = \{T_i^{(1)}\} \cup \{T^{\phi[1]}\} \cup \cdots \cup \{T^{\phi[k]}\}$.

Bibliography

[1] D. Abreu and M. K. Brunnermeier. Bubbles and crashes. *Econometrica 71 (1), pp. 173-204*, 2003.

[2] Y. Aït-Sahalia. Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. *Econometrica 70 (1), pp. 223-262*, 2002.

[3] H. Albrecher. *Advanced financial modelling*. Berlin : de Gruyter, 2009.

[4] H. Albrecher and S. Asmussen. Ruin probabilities and aggregrate claims distributions for shot noise Cox processes. *Scandinavian Actuarial Journal 2006 (2), pp. 86–110*, 2006.

[5] F. Allen and G. Gorton. Churning bubbles. *The Review of Economic Studies 60 (4), pp. 813-836*, 1993.

[6] F. Allen, S. Morris, and A. Postlewaite. Finite bubbles with short sale constraints and asymmetric information. *Journal of Economic Theory 61, pp. 206-229*, 1993.

[7] N. Alzahrani, P. Neal, S. E. Spencer, T. J. McKinley, and P. Touloupou. Model selection for time series of count data. *Computational Statistics & Data Analysis 122, pp. 33-44*, 2018.

[8] J. V. Andersen and D. Sornette. Fearless versus fearful speculative financial bubbles. *Physica A 337, pp. 565-585*, 2004.

[9] G. Antonacci. *Following the trend (diversified managed future trading)*. Wiley (West Sussex, UK), 2013.

[10] G. Antonacci. *Dual Momentum Investing: An Innovative Strategy for Higher Returns with Lower Risk*. McGraw-Hill Education, 2014.

[11] K. Arrow and G. Debreu. Existence of an equilibrium for a competitive economy. *Econometrica 22 (3), pp. 265–290*, 1954.

[12] C. S. Asness, T. J. Moskowitz, and L. H. Pedersen. Value and momentum everywhere. *The Journal of Finance 68 (3), pp. 929–985*, 2013.

[13] E. Bacry, K. Dayri, and J.-F. Muzy. Non-parametric kernel estimation for symmetric Hawkes processes. Application to high frequency financial data. *Eur. Phys. J. B 85: 157*, 2012.

[14] E. Bacry, I. Mastromatteo, and J.-F. Muzy. Hawkes processes in finance. *Market Microstructure and Liquidity 1 (1): 1550005*, 2015.

[15] N. Barberis, A. Shleifer, and R. Vishny. A Model of Investor Sentiment. *Journal of Financial Economics 49 (3), pp. 307-343*, 1998.

[16] P. Barroso and P. Santa-Clara. Momentum has its moments. *Journal of Financial Economics 116, pp. 111–120*, 2015.

[17] D. S. Bates. The Crash of '87: Was It Expected? The Evidence from Options Markets. *The Journal of Finance 46 (3), pp. 1009-1044*, 1991.

[18] D. S. Bates. Post-'87 crash fears in the S&P 500 futures option market. *Journal of Econometrics*, 2000.

[19] D. S. Bates. U.s. stock market crash risk, 1926–2010. *Journal of Financial Economics*, 2012.

[20] C. M. Bender and S. A. Orszag. *Advanced mathematical methods for scientists and engineers*. McGraw-Hill, Inc., 1978.

[21] F. Biagini, H. Föllmer, and S. Nedelcu. Shifting martingale measures and the birth of a bubble as a submartingale. *Finance and Stochatics 18, pp. 297–326.*, 2014.

[22] F. Biagini and S. Nedelcu. The formation of financial bubbles in defaultable markets. *SIAM J. Finan. Math., 6(1), pp. 530–558*, 2015.

[23] T. R. Bielecki and M. Rutkowski. *Credit Risk: Modeling, Valuation and Hedging*. Springer, 2002.

[24] G. Birkhoff and G.-C. Rota. *Ordinary Differential Equations*. John Wiley & Sons, 1989.

[25] T. Björk and I. Slinko. Towards a general theory of good-deal bounds. *Review of Finance 10, pp. 221-260*, 2006.

[26] F. Black and M. Scholes. The pricing of options and corporate liabilities. *J. Political Economy 81, pp. 637-659*, 1973.

[27] O. J. Blanchard. Speculative bubbles, crashes and rational expectations. *Economic Letters 3, pp. 387–389*, 1979.

[28] C. Borio and M. Drehmann. Towards an operational framework for finacial stability: "Fuzzy" measurements and its consequences. *Financial stability, monetary policy and central banking (Central Bank of Chile), Edition 1, Volume 15, Chapter 4, pp. 63-123*, 2011.

[29] J.-P. Bouchaud and D. Challet. Why have asset price properties changed so little in 200 years. *Available at https://arxiv.org/abs/1605.00634*, 2016.

[30] N. Bouleau and D. Lamberton. Residual Risks and Hedging Strategies in Markovian Markets. *Stochastic Processes and their Applications 33, pp. 131-150*, 1989.

[31] A. Boumezoued. Population viewpoint on Hawkes processes. *Advances in Applied Probability 48, pp. 463-480*, 2016.

[32] C. Bowsher. Modelling security market events in continuous time: Intensity based, multivariate point process models . *Journal of Econometrics 141 (2), pp. 876–912*, 2007.

[33] K. Brännäs and A. Hall. Estimation in integer-valued moving average models. *Applied Stochastic Models in Business and Industry 17 (3), pp. 277–291*, 2001.

[34] P. Brémaud and L. Massoulié. Power spectra of general shot noises and Hawkes point processes with a random excitation. *Advances in Applied Probability 34 (1), pp. 205–222*, 2002.

[35] D. R. Brillinger. Time series, point processes, and hybrids. *The Canadian Journal of Statistics/La Revue Canadienne de Statistique 22 (2), pp. 177–206*, 1994.

[36] D. R. Brillinger. Statistical inference for stationary point processes. In *Selected Works of David Brillinger, pp. 499–543*. Springer, 2012.

[37] P. J. Brockwell and R. A. Davis. *Time series: theory and methods*. Springer Science & Business Media, 2013.

[38] M. Brunnermeier and M. Oehmke. Bubbles, financial crises, and systemic risk. *Handbook of the Economics of Finance, Chapter 18, Elsevier B.V., pp. 1221-1288*, 2013.

[39] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing 16 (5), pp.1190–1208*, 1995.

[40] C. Camerer. Bubbles and fads in asset prices. *Journal of Economic Surveys 3 (1), pp. 3-41*, 1989.

[41] J. Y. Campbell, A. Lo, and A. MacKinlay. *The Econometrics of Financial Markets*. Princeton University Press, 1997.

[42] J. Y. Campbell and R. Shiller. The dividend-price ration and expectations of future dividends and discount factors. *Review of Financial Studies 1, pp. 195-228*, 1989.

[43] M. M. Carhart. On Persistence in Mutual Fund Performance. *The Journal of Finance 52 (1), pp. 57–82*, 1997.

[44] P. Carr and V. Linetsky. A jump to default extended CEV model: An application of Bessel processes. *Finance and Stochastics, Vol. 10, No. 3, pp. 303-330*, 2006.

[45] V. Chavez-Demoulin, A. C. Davison, and A. J. McNeil. Estimating value-at-risk: a point process approach. *Quantitative Finance 5 (2), pp. 227–234*, 2005.

[46] V. Chavez-Demoulin and J. McGill. High-frequency financial data modeling using hawkes processes. *Journal of Banking & Finance*, 36(12):3415–3426, 2012.

[47] F. Chen and W. Tan. Marked self-exciting point process modelling of information diffusion on Twitter. *Annals of Applied Statistics*, 2018.

[48] P. Cheridito, D. Filipovic, and M. Yor. Equivalent and absolutely continuous measure changes for jump-diffusion processes. *The Annals of Applied Probability, Vol. 15, No. 3, pp. 1713-1732*, 2005.

[49] H. Chernoff. On the distribution of the likelihood ratio. *The Annals of Mathematical Statistics 25 (3), pp. 573-578*, 1954.

[50] A. S. Cherny and H.-J. Engelbert. *Singular stochastic differential equations*. Lecture notes in mathematics No. 1858. Berlin : Springer, 2005.

[51] K. Christensen, R. Oomen, and R. Renò. The drift burst hypothesis. *Preprint available at SSRN:https://ssrn.com/abstract=2842535*, 2016.

[52] E. Cinlar. *Introduction to stochastic processes*. Courier Corporation, 2013.

[53] J. Cochrane. *Asset pricing: Revised Edition*. Princeton University Press, 2005.

[54] J. H. Cochrane and J. Saa-Requejo. Beyond arbitrage: Good-deal asset price bounds in incomplete markets. *The Journal of Political Economy, Volume 108, Issue 1, pp. 79-119*, 2000.

[55] R. Cont and P. Tankov. *Financial Modelling With Jump Processes*. Chapman and Hall/CRC, 2004.

[56] F. Corsi and D. Sornette. Follow the money: The monetary roots of bubbles and crashes. *International Review of Financial Analysis 32, pp. 47-59*, 2014.

[57] A. M. G. Cox and D. Hobson. Local martingales, bubbles and option prices. *Finance and Stochastics 9, pp. 477-492*, 2005.

[58] A. M. G. Cox, Z. Hou, and J. Obłój. Robust pricing and hedging under trading restrictions and the emergence of local martingale models. *Finance and Stochastics 20 (3), pp. 669-704*, 2016.

[59] D. Cox and V. Isham. *Point Processes*. CRC Press, 1980.

[60] J. Cox. Notes on option pricing 1: constant elasticity of variance diffusion. *Working paper, Standford University (1975) (reprinted in J. Portfolio Manag. 23 (2), pp. 5-17)*, 1996.

[61] J. Cox, J. Ingersoll, and S. Ross. A theory of the term structure of interest rates. *Econometrica 53, pp. 385–407*, 1985.

[62] R. Crane and D. Sornette. Robust dynamic classes revealed by measuring the response function of a social system. *Proceedings of the National Academy of Sciences 105 (41), pp. 15649–15653*, 2008.

[63] D. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes. Volume I: General theory and structure*, volume 1 of *Probability and Its Applications*. Springer Verlag, 2nd edition edition, 2003.

[64] K. D. Daniel, D. Hirschleifer, and A. Subrahmanyam. Investor psychology and security market under- and overreactions. *Journal of Finance 53 (6), pp. 1839–1885*, 1998.

[65] A. Dassios and J.-W. Jang. Pricing of catastrophe reinsurance and derivatives using the Cox process with shot noise intensity. *Finance and Stochastics 7 (1), pp. 73–95*, 2003.

[66] A. Dassios and J.-W. Jang. Kalman-Bucy filtering for linear systems driven by the Cox process with shot noise intensity and its application to the pricing of reinsurance contracts. *Journal of Applied Probability 42 (1), pp. 93–107*, 2005.

[67] A. Dassios and H. Zhao. A dynamic contagion process. *Advances in Applied Probability 43 (3), pp. 814–846*, 2011.

[68] B. De Long, A. Shleifer, L. Summers, and R. Waldmann. Noise trader risk in financial markets. *The Journal of Political Economy 98 (4), pp. 703-738*, 1990.

[69] B. De Long, A. Shleifer, L. Summers, and R. Waldmann. Positive feedback investment strategies and destabilizing rational speculation. *The Journal of Finance 45 (2), pp. 379-395*, 1990.

[70] F. Delbaen and W. Schachermayer. A general version of the fundamental theorem of asset pricing. *Mathematische Annalen 300, pp. 463-520*, 1994.

[71] F. Delbaen and W. Schachermayer. The fundamental theorem of asset pricing for unbounded stochastic processes. *Mathematische Annalen 312, pp. 215-250*, 1998.

[72] F. Delbaen and W. Schachermayer. A simple counterexample to several problems in the theory of asset pricing. *Mathematical Finance, 8 (1), pp. 1-11*, 1998.

[73] F. Delbaen and H. Shirakawa. No arbitrage condition for positive diffusion price processes. *Asia-Pacific Financial Markets 9, pp. 159-168*, 2002.

[74] R. Demirer, D. Lien, and H. Zhang. Industry herding and momentum strategies. *Pacific-Basin Finance Journal 32, pp. 95-110*, 2015.

[75] G. Demos and D. Sornette. Birth or burst of financial bubbles: which one is easier to diagnose? *Quantitative Finance 17 (5), pp. 657-675*, 2017.

[76] A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journ. of the Royal Statistical Society. Serie. B (Methodological) 39 (1), pp. 1–38*, 1977.

[77] B. T. Diba and H. I. Grossman. Explosive rational bubbles in stock markets? *American Economic Review 78(3), pp. 520-530*, 1988.

[78] B. T. Diba and H. I. Grossman. The theory of rational bubbles in stock prices. *Economic Journal 98, pp. 746-754*, 1988.

[79] J. Dion, G. Gauthier, and A. Latour. Branching processes with immigration and integer-valued time series. *Serdica Mathematical Journal 21 (2), pp. 123–136*, 1995.

[80] J. G. Du and Y. Li. The integer-valued autoregressive (INAR(p)) model. *Journal of Time Series Analysis 12 (2), pp. 129–142*, 1991.

[81] P. H. Dybvig and S. A. Ross. Arbitrage. In S. N. Durlauf and L. E. Blume, editors, *The New Palgrave Dictionary of Economics*. Palgrave Macmillan, 2008.

[82] G. Elaut and P. Erdös. Adaptive time series momentum – benchmark for trend-following funds. *Available at SSRN: http://ssrn.com/abstract=2772047*, 2016.

[83] D. Elworthy, X.-M. Li, and M. Yor. The importance of strictly local martingales; applications to radial Ornstein-Uhlenbeck processes. *Probability Theory Related Fields 155, pp. 325-355*, 1999.

[84] D. Emmanuel and J. MacBeth. Further results on the constant elasticity of variance call option pricing model. *Journal of Financial and Quantitaive Analysis, 17, pp. 533-554*, 1982.

[85] V. Enciso-Mora, P. Neal, and T. Subba Rao. Efficient order selection algorithms for integer-valued arma processes. *Journal of Time Series Analysis 30 (1), pp. 1–18*, 2009.

[86] H. J. Engelbert and W. Schmidt. Strong markov continuous local martingales and solutions of one-dimensional stochastic differential equations (part iii). *Math. Nachr. 151, pp. 149-197*, 1991.

[87] S. N. Ethier and T. G. Kurtz. *Markov processes. Characterization and Convergence.* Wiley, New York., 1986.

[88] G. W. Evans. Pitfalls in testing for explosive bubbles in asset prices. *The American Economic Review 81/4, pp. 922-930*, 1991.

[89] E. Fama. Efficient capital markets: A review of theory and empirical work. *The Journal of Finance, Vol. 25, No. 2, pp. 383–417*, 1970.

[90] E. Fama. Perspectives on October 1987 or what did we learn from the crash? In R. W. Kamphuis, R. C. Kormendi, and J. W. Watson, editors, *Black Monday and the Future of Financial Markets*. Irwin, Homewood, IL, 1989.

[91] E. Fama. Efficient capital markets: II. *The Journal of Finance, Vol. 46, No. 5, pp. 1575-1617*, 1991.

[92] E. Fama and K. French. Size, value, and momentum in international stock returns. *Journal of Financial Economics 105, pp. 457–472*, 2012.

[93] J. D. Farmer. Market force, ecology and evolution. *Industrial and Corporate Change 11 (5), pp. 895–953*, 2002.

[94] P. M. N. Feehan, R. Gong, and J. Song. Feynman-Kac formulas for solutions to degenerate elliptic and parabolic boundary-value and obstacle problems with Dirichlet boundary conditions. *Available on arXiv:1509.03864v1*, 2015.

[95] P. M. N. Feehan and C. A. Pop. Stochastic representation of solutions to degenerate elliptic and parabolic boundary value and obstacle problems with dirichlet boundary conditions. *Trans. Amer. Soc. 367, pp. 981-1031*, 2015.

[96] J. Feigenbaum. A statistical analysis of log-periodic precursors to financial crashes. *Quantitative Finance, 1(3), pp. 346-360*, 2001.

[97] V. Filimonov and D. Sornette. Quantifying reflexivity in financial markets: Toward a prediction of flash crashes. *Physical Review E 85 (5), 056108*, 2012.

[98] R. Flood and P. Garber. Market fundamentals versus price-level bubbles: The first tests. *Journal of Political Economy 88, pp. 745-770*, 1980.

[99] K. Fokianos. Count time series models. *Time Series – Methods and Applications 30, pp. 315–347*, 30, 2007.

[100] K. Fokianos and B. Kedem. *Regression Models for Time Series Analysis*. New York, N.Y : Wiley, 2002.

[101] H. Föllmer and P. Protter. Local martingales and filtration shrinkage. *ESAIM Probab. Stat. 15, pp. S25–S38*, 2011.

[102] H. Föllmer and D. Sondermann. Hedging of Non-Redundant Contingent Claims. In W. Hildenbrand and A. Mas-Colell, editors, *Contributions to Mathematical Economics*. North-Holland, 1986.

[103] Z. Forró, P. Cauwels, and D. Sornette. When games meet reality: is Zynga overvalued? *Journal of Investment Strategies 1 (3), pp. 119-145*, 2012.

[104] A. Friedman. *Stochastic Differential Equations and Applications, Volume 1*. Academic Press, New York, 1975.

[105] A. Friedman. *Stochastic Differential Equations and Applications, Volume 2*. Academic Press, New York, 1976.

[106] N. Garleanu and L. H. Pedersen. Dynamic trading with predictable returns and transaction costs. *Journal of Finance 68(6), pp. 2309-2340*, 2013.

[107] S. Giglio, M. Maggiori, and J. Stroebel. No-bubble condition: Model-free test in housing markets. *Econometrica 84 (3), pp. 1047-1091*, 2016.

[108] C. W. Granger. The typical spectral shape of an economic variable. *Econometrica 34 (1), pp. 150–161*, 1966.

[109] R. Greenwood, A. Shleifer, and Y. You. Bubbles for Fama. *NBER Working Paper No. w23191. Available at SSRN: https://ssrn.com/abstract=2924281*, 2017.

[110] M. Grinblatt and B. Han. Prospect theory, mental accounting, and momentum. *Journal of financial economics 78 (2), pp. 311–339*, 2005.

[111] M. Grinblatt and T. J. Moskowitz. Predicting stock price movements from past returns: the role of consistency and tax-loss selling. *Journal of Financial Economics 71, pp. 541-579*, 2004.

[112] M. Grinblatt, S. Titman, and R. Wermers. Momentum investment strategies, portfolio performance, and herding: a study of mutual fund behavior. *The American Economic Review 85 (5), pp. 1088-1105*, 1995.

[113] P. Guasoni and M. Rásonyi. Fragility of arbitrage and bubbles in local martingale diffusion models. *Finance and Stochastics 19, pp. 215-231*, 2015.

[114] R. S. Gürkaynak. Econometric tests of asset price bubbles: Taking stock. *Journal of Economic Surveys 22 (1), pp. 166-186*, 2008.

[115] L. Hansen and R. Jagannathan. Implications of security market data for models of dynamic economies. *Journal of Political Economy 99, pp. 225-262*, 1991.

[116] S. Hardiman, N. Bercot, and J. Bouchaud. Critical reflexivity in financial markets: a Hawkes process analysis. *European Journal of Physics B 86: 442*, 2013.

[117] T. E. Harris. *The theory of branching processes*. Courier Corporation, 2002.

[118] J. M. Harrison and D. M. Kreps. Investor behavior in a stock market with heterogeneous expectations. *The Quarterly Journal of Economics, Vol. 92, No. 2, pp. 323-336*, 1978.

[119] J. M. Harrison and D. M. Kreps. Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory 20, pp. 381–408*, 1979.

[120] A. Hawkes and D. Oakes. A Cluster Process Representation of a Self Exciting Process. *J. Appl. Prob 11, pp. 497–503*, 1974.

[121] X. Z. He and K. Li. Profitability of time series momentum. *Journal of Banking & Finance 53, pp. 140–157*, 2015.

[122] A. Helmstetter and D. Sornette. Subcritical and supercritical regimes in epidemic models of earthquake aftershocks. *Journal of Geophysical Research: Solid Earth (1978–2012)*, 107(B10):2237, doi:10.1029/2001JB001580, 2002.

[123] M. Herdegen. No arbitrage in a numéraire independent modeling framework. *Mathematical Finance 27(2), pp. 568–603*, 2017.

[124] M. Herdegen and S. Herrmann. A class of strict local martinales. *Swiss Finance Institute Research Paper No. 14-18. Available at SSRN: http://ssrn.com/abstract=2402248 or http://dx.doi.org/10.2139/ssrn.2402248*, 2014.

[125] M. Herdegen and S. Herrmann. Single jump processes and strict local martingales. *Stochastic Processes and Their Applications 126, pp. 337-359*, 2016.

[126] M. Herdegen and S. Herrmann. Strict local martingales and optimal investment in a Black-Scholes model with a bubble. *Mathematical Finance (forthcoming). Available online at https://doi.org/10.1111/mafi.12175.*, 2018.

[127] M. Herdegen and S. Herrmann. Strict Local Martingales and Optimal Investment in a Black-Scholes Model with a Bubble. *Mathematical Finance 29, pp. 285–328*, 2019.

[128] M. Herdegen and M. Schweizer. Strong bubbles and strict local martingales. *International Journal of Theoretical and Applied Finance, 19, 1650022*, 2016.

[129] M. Herdegen and M. Schweizer. Semi-efficient valuations and put-call parity. *Mathematical Finance (forthcoming). Available online at https://doi.org/10.1111/mafi.12162*, 2017.

[130] S. L. Heston, M. Loewenstein, and G. A. Willard. Options and bubbles. *Review of financial studies 20, pp. 359-390*, 2007.

[131] K. W. Hipel and A. I. McLeod. *Time series modelling of water resources and environmental systems*, volume 45. Elsevier, 1994.

[132] D. G. Hobson and L. Rogers. Complete models with stochastic volatility. *Mathematical Finance 8 (1), pp. 27–48*, 1998.

[133] R. Hoitash and M. Krishnan. Herding, momentum and investor over-reaction. *Review of Quantitative Finance and Accounting 30 (1), pp. 25-47*, 30(1):25–47, 2008.

[134] U. Homm and J. Breitung. Testing for speculative bubbles in stock markets: A comparison of alternative methods. *Journal of Financial Econometrics 10 (1), pp. 198-231*, 2012.

[135] H. Hulley and E. Platen. A visual classification of local martingales. *Research Paper Number: 238, Quantitative Finance Research Centre, University of Technology, Sydney*, 2008.

[136] B. Hurst, Y. H. Ooi, and L. H. Pedersen. A century of evidence on trend-following investing. *AQR Capital Management (Fall), pp. 1–12*, 2014.

[137] A. Hüsler, D. Sornette, and C. Hommes. Super-exponential bubbles in lab experiments: Evidence for anchoring over-optimistic expectations on price. *Journal of Economic Behavior and Organization 92, pp. 304-316*, 2013.

[138] A. Hüsler, D. Sornette, and C. H. Hommes. Super-exponential bubbles in lab experiments: evidence for anchoring over-optimistic expectations on price. *Journal Economic Behavior and Organization 92, pp. 304-316*, 2013.

[139] S. Hwang and A. Rubesam. The disappearance of momentum. *The European Journal of Finance 21 (7), pp. 584-607*, 2015.

[140] K. Ide and D. Sornette. Oscillatory Finite-Time Singularities in Finance, Population and Rupture. *Physica A 307 (1-2), pp. 63-106*, 2002.

[141] N. Ikeda and S. Watanabe. *Stochastic differential equations and diffusion processes*. North-Holland Publishing Co., Amsterdam, 1981.

[142] K. Ilinski. Critical crashes. *Int. J. Mod. Phys. C, 10(4), pp. 741-*, 1999.

[143] J. Jacod and A. N. Shiryaev. *Limit Theorems for Stochastic Processes*. Berlin: Springer-Verlag., 2nd edition edition, 2003.

[144] R. Jarrow. Testing for asset price bubbles: three new approaches. *Quantitative Finance Letters 4, pp. 4-9*, 2016.

[145] R. Jarrow, Y. Kchia, and P. Protter. How to detect an asset bubble. *SIAM J. Finan. Math., 2(1), pp. 839–865*, 2011.

[146] R. Jarrow, Y. Kchia, and P. Protter. Is there a bubble in LinkedIn's stock price? *The Journal of Portfolio Management, 38(1), pp. 125–130*, 2011.

[147] R. Jarrow, P. Protter, and K. Shimbo. Asset price bubbles in complete markets. *Advances in Mathematical Finance (Birkhäuser), pp. 97-121.*, 2007.

[148] R. Jarrow, P. Protter, and K. Shimbo. Asset price bubbles in incomplete markets. *Mathematical Finance 20, pp. 145-185*, 2010.

[149] R. A. Jarrow and M. Larsson. The meaning of market efficiency. *Mathematical Finance 22/1, pp. 1-30*, 2012.

[150] M. Jeanblanc and M. Rutkowski. Modelling of default risk: An overview. *Mathematical Finance: Theory and Practice, pp. 171-269*, 2000.

[151] N. Jegadeesh and S. Titman. Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance 48 (1), pp. 65–91*, 1993.

[152] N. Jegadeesh and S. Titman. Profitability of Momentum Strategies: An Evaluation of Alternative Explanations. *The Journal of Finance 56 (2), pp. 699–720*, 2001.

[153] A. Johansen, O. Ledoit, and D. Sornette. Crashes as critical points. *Int. J. Theor. Applied Finance 3 (2), pp. 219-255*, 2000.

[154] A. Johansen and D. Sornette. Critical crashes. *Risk, Vol 12, No. 1, pp. 91-94*, 1999.

[155] A. Johansen, D. Sornette, and O. Ledoit. Predicting financial crashes using discrete scale invariance. *Journal of Risk, Vol 1, No.4, pp. 5-32*, 1999.

[156] T. Kaizoji, M. Leiss, A. Saichev, and D. Sornette. Super-exponential endogenous bubbles in an equilibrium model of rational and noise traders. *Journal of Economic Behavior and Organization 112, pp. 289–310*, 2015.

[157] J. Kallsen and R. Vierthauer. Quadratic hedging in affine stochastic volatility models. *Review of Derivatives Research*, 2009.

[158] I. Karatzas and S. Shreve. *Brownian motion and stochastic calculus*. Springer, 1988.

[159] T. S. Kim and E. Omberg. Dynamic nonmyopic portfolio behavior. *Review of Financial Studies 9 (1), pp. 141–161*, 1996.

[160] C. P. Kindleberger. *Manias, panic and crashes*. Wiley, New York, 1996.

[161] M. Kirchner. Hawkes and INAR(∞) processes. *Stochastic Processes and their Applications 126 (8), pp. 2494–2525*, 2016.

[162] S. Klöppel and M. Schweizer. Dynamic utility-based good deal bounds. *Statistics & Decisions 25, pp. 285-309*, 2007.

[163] C. Klüppelberg and T. Mikosch. Explosive Poisson shot noise processes with applications to risk reserves. *Bernoulli 1 (1-2), pp. 125–147*, 1995.

[164] S. Kotani. On a condition that one-dimensional diffusion processes are martingales. *In memoriam Paul-André Meyer: Séminaire de Probabilités XXXIX, volume 1874 of Lecture Notes in Math., pp. 149-156*, 2006.

[165] D. O. Kramkov. Optional decomposition of supermartingales and hedging contingent claims in incomplete security markets. *Probab. Theory Relat. Fields 105, pp. 459-479*, 1996.

[166] D. Lamberton and B. Lapeyre. *Introduction to Stochastic Calculus Applied to Finance*. Chapman and Hall/CRC, 2007.

[167] M. Larsson and J. Ruf. Notes on stochastic exponential and logarithm. *Available on arXiv:1702.03573*, 2017.

[168] A. Latour. Existence and stochastic structure of a non-negative integer-valued autoregressive process. *Journal of Time Series Analysis 19 (4), pp. 439-455*, 1998.

[169] J. H. Lee and P. C. B. Phillips. Asset pricing with financial bubble risk. *Journal of Empirical Finance 38, pp. 590–622*, 2015.

[170] M. Leiss, H. H. Nax, and D. Sornette. Super-exponential growth expectations and the global financial crisis. *Journal of Economic Dynamics and Control 55, pp. 1-13*, 2015.

[171] Y. Lempérière, C. Deremble, P. Seager, M. Potters, and J.-P. Bouchaud. Two centuries of trend following. *The Journal of Investment Strategies 3 (3), pp. 41–61*, 2014.

[172] S. LeRoy. Efficient capital markets: A comment. *The Journal of Finance, Vol. 31, No. 1, pp. 139-141*, 1976.

[173] E. Lewis and G. Mohler. A Nonparametric EM algorithm for Multiscale Hawkes Processes. *Preprint*, pages 1–16, May 2011.

[174] L. Lin, R. Ren, and D. Sornette. The Volatility-Confined LPPL Model: A Consistent Model of 'Explosive' Financial Bubbles With Mean-Reversing Residuals. *International Review of Financial Analysis 33, pp. 210–225*, 2014.

[175] L. Lin, M. Schatz, and D. Sornette. A simple mechanism for financial bubbles: time-varying momentum horizon. *Quantitative Finance, 19 (6), pp. 937-959*, 2019.

[176] L. Lin and D. Sornette. Diagnostics of Rational Expectation Financial Bubbles with Stochastic Mean-Reverting Termination Times. *The European Journal of Finance 19 (5-6), pp. 344-365*, 2013.

[177] R. S. Liptser and A. N. Shiryaev. *Statistics of Random Processes*. Springer, 2nd edition, 2000.

[178] J. Liu and J. Pan. Dynamic derivative strategies. *Journal of Financial Economics 69, pp. 401-430*, 2003.

[179] M. Loewenstein and G. A. Willard. Rational equilibrium asset-pricing bubbles in continuous trading models. *Journal of Economic Theory 91, pp. 17-58*, 2000.

[180] M. Loewenstein and G. A. Willard. Consumption and bubbles. *Journal of Economic Theory 148, pp. 563-600*, 2013.

[181] E. McKenzie. Some ARMA models for dependent sequences of Poisson counts. *Advances in Applied Probability 20 (4), pp. 822-835*, 1988.

[182] E. McKenzie. Discrete variate time series. *Handbook of Statistics 21, pp. 573–606*, 2003.

[183] G. McLachlan and T. Krishnan. *The EM algorithm and extensions*. John Wiley & Sons, 2007.

[184] A. Mijatović and M. Urusov. On the martingale property of certain local martingales. *Probability Theory and Related Fields, 152 (1), pp. 1-30*, 2012.

[185] H. P. Minsky. Financial instability revisited: the economics of disaster. *Board of Governors, Federal Reserve System: Fundamental Reappraisal of the Federal Reserve Discount Mechanism*, 1972.

[186] M.Leiss, H. H. Nax, and D. Sornette. Super-exponential growth expectations and the global financial crisis. *Journal of Economic Dynamics and Control, 55, C*, 2015.

[187] J. Møller. Shot noise Cox processes. *Advances in Applied Probability 35 (3), pp. 614–640*, 2003.

[188] J. Møller and R. P. Waagepetersen. *Statistical inference and simulation for spatial point processes*. CRC Press, 2003.

[189] T. J. Moskowitz, Y. H. Ooi, and L. H. Pedersen. Time series momentum. *Journal of Financial Economics 104 (2), pp. 228–250*, 2012.

[190] P. Neal and T. Subba Rao. MCMC for Integer-Valued ARMA processes. *Journal of Time Series Analysis 28 (1), pp. 92-110*, 2007.

[191] J. Neyman and E. Scott. Statistical approach to problems of cosmology. *Journal of the Royal Statistical Society, Series B., 20, pp. 1–43*, 1958.

[192] M. Nutz and J. A. Scheinkman. Shorting in speculative markets. *Preprint available at SSRN: https://ssrn.com/abstract=2969112*, 2017.

[193] Y. Ogata. The asymptotic behaviour of maximum likelihood estimators for stationary point processes. *Annals of the Institute of Statistical Mathematics 30 (1), pp. 243–261*, 1978.

[194] B. Øksendal. When is a stochastic integral a time change of a diffusion? *Journal of Theoretical Probability 3 (2), pp. 207–226*, 1990.

[195] A. Orléan. Bayesian interactions and collective dynamics of opinion: herd behavior and mimetic contagion. *Journal of Economic Behavior & Organization 28, pp. 257-274*, 1995.

[196] L. Pástor and P. Veronesi. Was there a Nasdaq bubble in the late 1990s? *Journal of Financial Economics, 81 (1), pp. 61-100*, 2006.

[197] H. Pham. On quadratic hedging in continuous time. *Math Meth Oper Res 51, pp.315-339*, 2000.

[198] P. B. Phillips, S. Shi, and J. Yu. Testing for multiple bubbles: historical episodes of exuberance and collapse in the S&P500. *International Economic Review 56 (4), pp. 1043-1078*, 2015.

[199] P. C. B. Phillips, Y. Wu, and J. Yu. Explosive behaviour in the 1990s NASDAQ: When did exuberance escalate asset values? *International Economic Review 52, pp. 201-226*, 2011.

[200] A. C. Pipkin. *A course on integral equations*. Springer Science & Business Media, 1991.

[201] M. Prokešová and E. B. V. Jensen. Asymptotic palm likelihood theory for stationary point processes. *Annals of the Institute of Statistical Mathematics 65 (2), pp. 387–412*, 2013.

[202] P. Protter. On the existence, uniqueness, convergence and explosions of solutions of systems of stochastic integral equations. *The Annals of Probability 5 (2), pp.243-261*, 1977.

[203] P. Protter. *Stochastic Integration and Differential Equations*. Springer, 1990.

[204] P. Protter. *Stochastic Integration and Differential Equations*. Springer, 2010.

[205] P. Protter. A mathematical theory of financial bubbles. *Paris-Princeton Lectures on Mathematical Finance 2013, pp. 1-108*, 2013.

[206] R. Radner. Existence of equilibrium of plans, prices, and price expectations in a sequence of markets. *Econometrica 40 (2), pp. 289–303*, 1972.

[207] S. Resnick. *Extreme Values, Regular Variation and Point Processes*. New York: Springer, 1987.

[208] J. Ruf. Hedging und Arbitrage. *Mathematical Finance 23(2), pp. 297-317*, 2013.

[209] R. Salakhutdinov, S. Roweis, and Z. Ghahramani. Optimization with EM and expectation-conjugate-gradient. In *ICML, pp. 672–679*, 2003.

[210] P. Samuelson. Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review, 6 (2), pp.41-49*, 2003.

[211] M. Santos and M. Woodford. Rational asset pricing bubbles. *Econometrica 65, pp. 19-57*, 1997.

[212] M. Schatz and D. Sornette. A Nonuniformly Integrable Martingale Bubble with a Crash. *SIAM J. Finan. Math., 10(2), pp. 615–631*, 2019.

[213] M. Schatz and D. Sornette. Inefficient bubbles and efficient drawdowns in financial markets. *Available at SSRN: https://ssrn.com/abstract=3210598.*, 2019.

[214] J. A. Scheinkman and W. Xiong. Overconfidence and speculative bubbles. *Journal of Political Economy 111 (6), pp. 1183-1219*, 2003.

[215] M. Schweizer. On the minimal martingale measure and the Föllmer-Schweizer decomposition. *Stochastic Analysis and Applications 13, pp. 573-599*, 1995.

[216] M. Schweizer. A guided tour through quradatic hedging approaches. In E. Jouini, J. Cvitanić, and M. Musiela, editors, *Option Pricing, Interest Rates and Risk Management*. Cambridge University Press, 2001.

[217] M. Seyrich and D. Sornette. Micro-foundation using percolation theory of the finite-time singular behavior of the crash hazard rate in a class of rational expectation bubbles. *International Journal of Modern Physics C 27 (10), 1650113 (20 pages)*, 2016.

[218] H. Shefrin. *A behavioral approach to asset pricing*. Amsterdam : Elsevier, 2008.

[219] A. Shiryaev, M. Zhitlukhin, and W. Ziemba. When to sell Apple and the NASDAQ? trading bubbles with a stochastic disorder model. *The Journal of Portfolio Management, Vol. 40, No. 2, pp. 54-63.*, 2014.

[220] A. N. Shiryaev and M. V. Zhitlukhin. Optimal stopping problems for a Brownian motion with a disorder on a finite interval. *Available on arXiv, No. 1212.3709.*, 2012.

[221] A. Shleifer. *Inefficient markets: an introduction to behavioral finance*. Oxford University Press, 2000.

[222] A. Shleifer and L. H. Summers. The noise trader approach to finance. *The Journal of Economic Perspectives 4 (2), pp. 19-23*, 1990.

[223] A. Shleifer and R. Vishny. The limits of arbitrage. *The Journal of Finance 52 (1), pp. 57–82*, 1997.

[224] D. Smug, P. Ashwin, and D. Sornette. Predicting financial market crashes using ghost singularities. *PLOS One*, 13(3):e0195265, Mar. 2018.

[225] D. Sornette. Critical market crashes. *Physics Reports 378 (1), pp. 1-98*, 2003.

[226] D. Sornette. *Why Stock Markets Crash*. Princeton University Press, 2004.

[227] D. Sornette. Endogenous versus exogenous origins of crises. *In the monograph entitled "Extreme Events in Nature and Society," Series: The Frontiers Collection, S. Albeverio, V. Jentsch and H. Kantz, eds. (Springer, Heidelberg), pp. 95-119*, 2005.

[228] D. Sornette and J. V. Andersen. A nonlinear super-exponential rational model of speculative financial bubbles. *International Journal of Modern Physics C, Vol. 13, No. 2, pp. 171-187*, 2002.

[229] D. Sornette and P. Cauwels. 1980-2008: The illusion of the perpetual money machine and what it bodes for the future. *Risks 2, pp. 103-131*, 2014.

[230] D. Sornette and P. Cauwels. Financial bubbles: mechanisms and diagnostics. *Review of Behavioral Economics 2 (3), pp. 279- 305*, 2015.

[231] D. Sornette and P. Cauwels. Managing risks in a creepy world. *Journal of Risk Management in Financial Institutions (JRMFI) 8 (1), pp. 83-108*, 2015.

[232] D. Sornette, P. Cauwels, and G. Smilyanov. Can we use volatility to diagnose financial bubbles? lessons from 40 historical bubbles. *Quantitative Finance and Economics, 2(1), pp. 486-594*, 2018.

[233] D. Sornette and A. Johansen. Significance of log-periodic precursors to financial crashes. *Quantitative Finance 1 (4), pp. 452-471*, 2001.

[234] D. Sornette, H. Takayasu, and W.-X. Zhou. Finite-Time Singularity Signature of Hyperinflation. *Physica A 325, pp. 492–506*, 2003.

[235] D. Sornette and S. Utkin. Limits of declustering methods for disentangling exogenous from endogenous events in time series with foreshocks, main shocks, and aftershocks. *Physical Review E 79 (6), 061110*, 2009.

[236] R. L. Stratonovich. A new representation for stochastic integrals and equations. *J. SIAM Control 4 (2), pp. 362-371*, 1966.

[237] C. Stricker. Quasimartingales, martingales locales, semimartingales et filtration naturelle. *Z. Wahrscheinlichkeitstheorie verw Gebiete 39, pp. 55–63*, 1977.

[238] U. Tanaka, Y. Ogata, and D. Stoyan. Parameter estimation and model selection for Neyman-Scott point processes. *Biometrical Journal 50 (1), pp. 43–57*, 2008.

[239] N. Taylor. The rise and fall of technical trading rule success. *Journal of Banking & Finance 40, pp. 286–302*, 2014.

[240] J. Tirole. On the possibility of speculation under rational expectations. *Econometrica, 50 (5), pp. 1163-1181*, 1982.

[241] J. Tirole. Asset bubbles and overlapping generations. *Econometrica 53, pp. 1499-1528*, 1985.

[242] P. Touloupou, N. Alzahrani, P. Neal, S. E. Spencer, T. J. McKinley, et al. Efficient model comparison techniques for models requiring large scale data augmentation. *Bayesian Analysis 13 (2), pp. 437–459*, 2013.

[243] H.-C. van Bothmer and C. Meister. Predicting critical crashes? a new restriction for the free variables. *Physica A 320, pp. 539–547*, 2003.

[244] A. Černý. Generalized Sharpe ratios and asset pricing in incomplete markets. *European Finance Review, 7(2), pp. 191-233*, 2003.

[245] G. C. Wei and M. A. Tanner. A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American statistical Association 85 (411), pp. 699–704*, 1990.

[246] C. H. Weiß. Serial dependence and regression of Poisson INARMA models. *Journal of Statistical Planning and Inference 138 (10), pp. 2975–2990*, 2008.

[247] C. H. Weiß. Thinning operations for modeling time series of counts—a survey. *AStA Advances in Statistical Analysis 92 (3), pp. 319–341*, 2008.

[248] S. Wheatley. *Extending the Hawkes process, a general outlier test, & case studies in extreme risk.* PhD thesis, ETH Zurich, 2016.

[249] S. Wheatley, M. Schatz, and D. Sornette. The ARMA Point Process and its Estimation. *Available on arXiv:1806.09948*, 2018.

[250] S. Wheatley, A. Wehrli, and D. Sornette. The endo–exo problem in high frequency financial price fluctuations and rejecting criticality. *Quantitative Finance 19 (7), pp. 1165 - 1178*, 2019.

[251] S. Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics 9 (1), pp. 60–62*, 1938.

[252] C. Wu. On the convergence properties of the EM algorithm. *The Annals of Statistics 11 (1), pp. 95–103*, 1983.

[253] M. Wyart and J.-P. Bouchaud. Self-referential behavior, overreaction and conventions in financial markets. *Journal of Economic Behavior & Organization 63, pp. 1–24*, 2007.

[254] L. Xu and M. Jordan. On convergence properties of the EM algorithm for Gaussian mixtures. *Neural computation 8, pp. 129-151*, 1995.

[255] W.-X. Zhou and D. Sornette. Fundamental factors versus herding in the 2000–2005 US stock market and prediction. *Physica A 360(2), pp. 459-482*, 2006.

Curriculum vitae

MICHAEL SCHATZ

Department of Management, Technology and Economics, ETH Zurich
Scheuchzerstrasse 7, 8092 Zurich, Switzerland
+41/774706646 | mschatz@ethz.ch
Born 22/05/1990, Linz | Austrian citizen

Reasearch interests

Stochastic Analysis: Point processes and estimation, Single jump diffusions, Martingale theory

Mathematical Finance: Financial bubbles, Momentum trading, Option pricing & replication during bubbles

Education

PhD Candidate, Department of MTEC, ETH Zurich — 01/2016-03/2020
Supervisor: Prof. Didier Sornette

MSc Mathematics, ETH Zurich — 09/2012-03/2015
Master thesis under the supervision of Prof. Arnulf Jentzen
GPA 5.9/6.0, top 10%

Non-award, Australian National University — 02/2013-05/2013

BSc Mathematics, ETH Zurich / JKU Linz — 09/2009-09/2012
GPA 5.3/6.0 (Zurich) | 5.9/6.0 (Linz)

Publications and preprints

A Nonuniformly Integrable Martingale Bubble with a Crash (with Didier Sornette), SIAM Journal of Financial Mathematics 10 (2), 615-631 (2019).

A simple mechanism for financial bubbles: time-varying momentum horizon (with Didier Sornette and Lin Li), Quantitative Finance 19 (6), 937-959 (2019).

Inefficient Bubbles and Efficient Drawdowns in Financial Markets (with Didier Sornette), Submitted.

The ARMA Point Process and its Estimation (with Spencer Wheatley and Didier Sornette), Preprint.

Quadratic hedging during financial bubbles and crashes (with Didier Sornette), Working paper.

Teaching activity

Research project advisor (MSc Mathematics), Option pricing in single jump models — 2019

Teaching certificate (ETH Zurich), Degree in didactics with focus on mathematics — 09/2018-ongoing

Teaching assistant (ETH Zurich), Linear Algebra, Analysis, and Advanced numerical analysis — 10/2012-03/2015

Tutor, employed at Lernwerkstatt Linz — 01/2009-10/2010

Academic activity

Referee service Int. J. Theor. Appl. Finance — 2018/2019

Research stay at UCLA Statistics department, Supervisor: F. P. Schoenberg — 09/2019-12/2019

Conference/seminar talks

XIX Quantitative Finance Workshop, Rome	01/2018
13th German Probability and Statistic days, Freiburg	03/2018
ETH Risk Center Seminar Series, Zurich	04/2018
Conference of Acturial Science and Finance, Samos	05/2018
Bachelier Finance World Congress, Dublin	06/2018
Statistics Seminar Talk, University of Warwick	03/2019
UCLA Statistics Seminar, University of California, Los Angeles	10/2019

Grants

Swiss National Foundation grant (30% acceptance rate) for reasearch proposal *Financial mathematics of positive-feedback bubbles and crashes*

Employment

Bain&Company, Internship, Zurich — 05/2015-08/2015
Customer retention modeling in telecommunications

UBS Switzerland AG, Internship, Zurich — 09/2013-02/2014
Model risk in short-rate and equity valuation models

WorkOut3D, Start-Up (unpaid) — 03/2014-12/2014

Further skills/activities

Programming Skills	C++ | R | Matlab | Python
Software Skills	MS Office | AutoCAD | Mathematica | Bloomberg | Reuters | Latex
Hobbies/ Interests	Sports | Travelling | Movies | Literature
Sports (Selection)	Soccer | Skiing | Surfing | Mountain Biking | Mountaineering | active soccer team member and former long lasting captain of junior soccer team

Languages

German	Mother tongue
English	Fluent