DISS. ETH NO. 30502

FROM FINANCIAL BUBBLES TO EXTREME EVENTS: ECONOMETRIC AND STATISTICAL METHODS

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES
(Dr. sc. ETH Zurich)

presented by

RAN WEI
MSc Financial Mathematics
Tongji University

born on 22 March 1992

accepted on the recommendation of

Prof. Dr. Jan-Egbert Sturm, examiner
Prof. Dr. Didier Sornette, co-examiner
Prof. Dr. Josef Teichmann, co-examiner
Prof. Dr. Felix Kübler, co-examiner

2024

ABSTRACT

Extreme market events, characterized by sudden and significant disruptions in market behavior, pose profound challenges to economic stability and predictability. These events, often exemplified by financial bubbles and crashes, disrupt the normative frameworks and functionalities of financial markets and can lead to drastic economic consequences. Studying bubbles and further extreme events is therefore crucial, despite their long-standing reputation as unpredictable phenomena, which stems from the rarity of these events and the complex, interdependent factors leading to their occurrence. While traditional financial theories and models typically rely on known, usually oversimplified distributions which only work well under stable conditions, specific methods can be developed to capture the transitioning dynamics. Through these methods, even within the unpredictability, patterns and underlying mechanisms can be studied and understood.

The scope of this thesis revolves around the methods, namely econometric models and statistical tests, for the exploration of these rare but impactful events.

The first half of the thesis introduces two time series models that capture the dynamics of market crashes by exploring the crash hazard rate. The first model links the crash hazard rate to a non-local estimation of mispricing, integrating historical price levels and investor expectations to better capture potential market crashes and drawdowns. This model challenges traditional bubble models by describing the mispricing as non-local self-referencing, breaking the instantaneous relationship between risk and return. It quantifies what the investors 'anchor' on, providing a more realistic depiction of financial crashes. Extending this analysis, the second model differentiates the influences on the crash hazard rate into two parts: the baseline intensities from external shocks and the self-excitations from endogenous interactions. This distinction allows for an in-depth analysis of market behaviors, illustrating that market crashes are typically not isolated events but are contagious, especially during periods of market exuberance. This model is exceptionally good at capturing the asymmetric shape in bubble development, where a prolonged period of gradual price increases are followed by swift and steep declines. Moreover, it captures key stylized facts of financial time series such as volatility clustering and fat-tailedness, without incorporating traditional volatility models like GARCH. By inte-

grating principles from behavioral finance such as positive feedback loop and herding, these models help understand the complexities of market dynamics, providing insights into *how* and *why* markets crash.

The second half of the thesis focuses on statistical tests for outlier detection in data with exponential and Pareto tails. In the analysis of financial data, it is often observed that distributions in the tails follow a Pareto law, which indicates the fat-tailed characteristics of financial time series. However, even within these Pareto-distributed tails, there exist significant anomalies. These outliers are not statistical curiosities; on the contrary, they represent critical events such as market crashes that hold substantial relevance. The statistical tests and methods developed in our study are particularly designed to capture these anomalies. Thanks to Extreme Value Theory, the tests can also be applied to more general samples that have approximately exponential or Pareto tails. By discussing the identification of outliers in data with fat-tailed distributions, these statistical tools help define *what* is 'extreme' in financial market.

The dual perspective provided by the innovative econometric models and statistical tests in this thesis regarding bubbles and further extreme events sheds light on the mechanisms and implications of these rare and transformative occurrences in the financial world and beyond.

ZUSAMMENFASSUNG

Extrem finanzielle Ereignisse, gekennzeichnet durch plötzliche und signifikante Störungen im Marktverhalten, stellen tiefe Herausforderungen für die wirtschaftliche Stabilität und Vorhersagbarkeit dar. Diese Ereignisse, oft als Finanzblasen und -crashs bezeichnet, stören die normativen Rahmenwerke und Funktionalitäten der Finanzmärkte und können zu drastischen wirtschaftlichen Konsequenzen führen. Das Studium extremer finanzieller Ereignisse ist daher entscheidend, trotz ihres langjährigen Rufs als unvorhersehbare Phänomene, was aus der Seltenheit dieser Ereignisse und den komplexen, interdependenten Faktoren, die zu ihrem Auftreten führen, resultiert. Während traditionelle finanzielle Theorien und Modelle typischerweise auf bekannten, meist übervereinfachten Verteilungen basieren, die nur unter stabilen Bedingungen gut funktionieren, können spezifische Werkzeuge entwickelt werden, um die übergangsweisen Dynamiken zu erfassen. Mit diesen Werkzeugen können, auch innerhalb der Unvorhersehbarkeit, Muster und zugrundeliegende Mechanismen studiert und verstanden werden.

Der Umfang dieser Arbeit dreht sich um die Werkzeuge, nämlich ökonometrische Modelle und statistische Tests, zur Erforschung dieser seltenen, aber einflussreichen Ereignisse.

Die erste Hälfte der Arbeit führt zwei Zeitreihenmodelle ein, die die Dynamik von Marktzusammenbrüchen durch die Untersuchung der Crash-Gefahrenrate erfassen. Das erste Modell verknüpft die Crash-Gefahrenrate mit einer nicht-lokalen Schätzung der Fehlbewertung, integriert historische Preisniveaus und Investorenerwartungen, um potenzielle Marktzusammenbrüche und -abschwünge besser zu erfassen. Dieses Modell stellt traditionelle Blasenmodelle in Frage, indem es die Fehlbewertung als nicht-lokal selbstreferenzierend beschreibt, wodurch die sofortige Beziehung zwischen Risiko und Rendite aufgebrochen wird. Es quantifiziert, woran sich die Investoren 'anchor', und bietet eine realistischere Darstellung von Finanzkrisen. Die erweiterte Analyse des zweiten Modells differenziert die Einflüsse auf die Crash-Gefahrenrate in zwei Teile: die Baseline-Intensitäten von externen Schocks und die Selbstanregungen von endogenen Interaktionen. Diese Unterscheidung ermöglicht eine tiefgehende Analyse von Marktverhaltensweisen und zeigt, dass Finanzcrashs typischerweise keine isolierten Ereignisse sind, sondern ansteckend, besonders in Zeiten von

Markteuphorie. Dieses Modell ist besonders gut darin, die asymmetrische Form in der Blasenentwicklung zu erfassen, bei der eine verlängerte Periode allmählicher Preisanstiege von schnellen und steilen Rückgängen gefolgt wird. Darüber hinaus erfasst es wichtige stilisierte Fakten von Finanzzeitreihen wie Volatilitätsclusterung und Fett-Schwänzigkeit, ohne traditionelle Volatilitätsmodelle wie GARCH einzubeziehen. Durch die Integration von Prinzipien der Verhaltensfinanzierung wie positive Rückkopplungsschleife und Herdenverhalten helfen diese Modelle, die Komplexität der Marktdynamik zu verstehen und bieten Einblicke in *wie* und *warum* Märkte zusammenbrechen.

Die zweite Hälfte der Arbeit konzentriert sich auf statistische Tests zur Ausreirerkennung in Daten mit exponentiellen und Pareto-Schwänzen. Bei der Analyse von Finanzdaten wird oft beobachtet, dass die Verteilungen in den Schwänzen einem Pareto-Gesetz folgen, was die fett-schwänzigen Eigenschaften von Finanzzeitreihen anzeigt. Doch selbst innerhalb dieser Pareto-verteilen Schwänze gibt es bedeutende Anomalien. Diese Ausreir sind keine statistischen Kuriositäten; im Gegenteil, sie stellen kritische Ereignisse wie Marktkrisen dar, die von erheblicher Bedeutung sind. Die in unserer Studie entwickelten statistischen Tests und Methoden sind speziell dafür ausgelegt, diese Anomalien zu erfassen. Dank der Extremwerttheorie können unsere Tests auch auf allgemeinere Proben angewendet werden, die annähernd exponentielle oder Pareto-Schwänze aufweisen. Indem die Identifizierung von Ausreirn in Daten mit fett-schwänzigen Verteilungen diskutiert wird, helfen diese statistischen Werkzeuge zu definieren, *was* als 'extrem' auf dem Finanzmarkt gilt.

Die doppelte Perspektive, die durch die innovativen ökonometrischen Modelle und statistischen Tests in dieser Arbeit bezüglich extremer finanzieller Ereignisse geboten wird, beleuchtet die Mechanismen und Implikationen dieser seltenen und transformativen Vorkommnisse in der Finanzwelt und darüber hinaus.

ACKNOWLEDGEMENTS

In the solitude of this academic journey, I have been fortunate to be surrounded by individuals who have been both allies and beacons.

I extend my sincere thanks to my initial supervisor, Professor Didier Sornette, who has been instrumental in the realization of this thesis. His guidance and spirit have influenced my academic pursuits and personal endeavors, and his retirement leaves a legacy that continues to inspire me. I must also express my appreciation to Professor Jan-Egbert Sturm, who stepped in during the crucial final stages of my study. Despite the brevity of our acquaintance, I am immensely thankful for his support. I am grateful to my advisor, Professor Josef Teichmann, whose insights refined my work and whose support carried me through the study. I am equally thankful to my co-authors, Professor Yannick Malevergne and Doctor Alexander Wehrli, for their collaboration and contributions to the joint research efforts. Working alongside them has been an enriching experience, and I am grateful for the knowledge and enthusiasm they brought to the projects.

Lastly, I would like to thank my family, especially my partner, for their unconditional and unending support. It is to them that I owe the completion of this work.

CONTENTS

1

INTRODUCTION

When I see a bubble forming, I rush in to buy, adding fuel to the fire. That is not irrational.

— George Soros

Central to the extreme market events are financial bubbles, which are characterized by transient explosive growth followed by abrupt and intense corrections. As we explore the intricate dynamics of financial bubbles, it becomes imperative to look back at the historical bubbles that have shaped our current understanding.

1.1 LESSONS FROM HISTORICAL BUBBLES

The Tulip Mania and the South Sea Bubble

The Tulip Mania (1637) and the South Sea Bubble (1720) are among the earliest recorded financial bubbles, showing how speculative frenzy can detach asset prices from their intrinsic values. During the Tulip Mania, the prices of tulip bulbs soared to extraordinary heights before collapsing, driven by a speculative craze with little real economic activity. Similarly, the South Sea Bubble saw the stock price of the South Sea Company skyrocket based on speculative trading, ultimately ending in a catastrophic crash. Both cases reveal the nature of zero-sum games to a certain extent, where the profits of some investors came at the expense of others, leaving little lasting value for the broader economy.

The British railway bubble and the dot-com bubble

In contrast, the British railway bubble of the 1840s and the dot-com bubble of the late 1990s, while also marked by speculative excesses, left behind substantial infrastructure and technological progress. A few years after the British railway bubble burst, The Economist remarked, 'Mechanically or scientifically, the railways, with all their multiplied conveniences and contrivances, are an honour to our age and country: commercially, they are

great failures.[1]' This comment captures the dual nature of such bubbles: while they can drive significant technological and infrastructural progress, they can also lead to severe financial losses and misallocation of resources. The dot-com bubble had a similar pattern, where the collapse of inflated internet stocks in the early 2000s showed how technological optimism and speculative investments can lead to problems. Despite the financial chaos, it still resulted in important tech innovations and digital infrastructure.

Modern financial crises: 1929, 1987, and 2007-2008

The stock market crashes of 1929, 1987, and the 2007-2008 financial crisis illustrate the consequences of financial bubbles on regional and global economy. The 1929 crash, which was precipitated by over-speculation and over-leveraging among investors, was also deeply influenced by factors such as poor economic policies and bank weaknesses. The market failure led to the Great Depression, a severe worldwide economic depression that lasted through the 1930s.

In contrast, the crash of 1987, known as Black Monday, highlighted the risks associated with computerized trading and automated markets. While automated trading systems and portfolio insurance were intended to handle risks, they actually accelerated the market failure once their limits are exceeded during turmoils (Malliaris and Urrutia, 1992). Despite the severity of the crash, the impact was relatively short-lived and did not lead to a depression.

The 2007-2008 financial crisis, rooted in the housing bubble and exacerbated by complex financial derivatives like mortgage-backed securities, demonstrated how innovations in financial technology can contribute to systematic risks. The financial systems were unprepared for the cascading of defaults and the devaluation of supposedly safe assets. As noted by George Soros in his critiques of the period, this failure of the financial systems highlighted the severe consequence of relying too heavily on poorly understood financial instruments (Soros, 2009). In response, regulations such as the Dodd-Frank Act were enacted to strengthen the financial systems.

The study of financial bubbles has fundamentally changed how we think about economics, from viewing extreme market events as unprecedented abnormal outliers to recognizing them as systematic and cyclical phenomena. This change of view can be traced back to Hyman Minsky's Financial

1 Sept. 15, 1855 issue, pp.1010-11

Instability Hypothesis (Minsky, 1977), which provides a framework for understanding financial bubbles, emphasizing how periods of financial stability can lead to speculative euphoria and increasing risk-taking. Works like *Manias, Panics, and Crashes* (Kindleberger, 1978), and *A Short History of Financial Euphoria* (Galbraith, 1994) also have highlighted the recurring nature of financial instability driven by speculative manias.

1.2 APPROACHES TO UNDERSTANDING BUBBLES

As previously argued, the study of financial bubbles has changed our understanding of economics. This shift in perspective has encouraged researchers and practitioners to develop econometric models to explain the formation, persistence, and bursting of financial bubbles. The various approaches to understanding bubbles can be categorized into rational expectations, behavioral finance, and econophysics models led by Log-Periodic Power Law Singularity (LPPLS) models. Through these approaches, we gain deeper insights into the mechanisms that drive financial bubbles and their impact on markets.

Rational expectation

The Efficient Market Hypothesis (EMH) proposed by Fama (1970) argues that asset prices fully reflect all available information, making it impossible to consistently achieve higher returns without taking on additional risks. According to EMH, bubbles should not occur because any deviation from fundamental values is quickly corrected by rational investors. However, the existence of bubbles suggests that markets are not always perfectly efficient. The pioneers of rational bubble models have provided a more nuanced perspective on market efficiency. This class of models are based on rational expectation theory, introduced by Muth (1961) and further developed by Lucas (1972), suggests that individuals form expectations about the future based on all available information and that these expectations are, on average, accurate. Rational expectation explains that the deviation from intrinsic value persists because investors believe they can sell the overvalued asset to someone else at a higher price. The study of rational bubbles gained significant attention in the 1980s and 1990s. Several models have been developed to explain how rational bubbles can exist even when investors are rational.

Blanchard (1979), Blanchard and Watson (1982) introduced a rational bubble model demonstrating the emergence of rational bubbles due to investors' expectation of future price increases. Tirole (1982) explored the conditions under which bubbles cannot exist in general equilibrium settings, while Santos and Woodford (1997) analyzed the sustainability of rational bubbles under different market conditions. These studies imply that, for finite horizon investors, bubbles cannot arise as long as rational investors can sell their shares without restrictions in any future scenario, and for infinite horizon investors, bubbles become possible when removing the transversality condition.

Tirole (1985) extended his analysis to a more realistic market structure and examined the conditions under which bubbles can exist in a market with overlapping generations of investors, suggesting that bubbles can be sustained through intertemporal trades and expectations. Weil (1987) also provided insights into how different confidence levels among different generations can impact the formation and persistence of bubbles. These models highlight the role of demographic and generational factors on bubble dynamics, suggesting that bubbles can persist if the expectations of future generations continue to support high asset prices. However, these models generally assume stable demographic factors over time. Real-world deviations, such as changes in population growth or investor behavior, can compromise the models' applicability.

Diba and Grossman (1988a) investigated the relationship between rational bubbles and market fundamentals. Further research by Froot and Obstfeld (1991) explored intrinsic bubbles and how they are related to fundamental economic variables. Garber (2000) also analyzed historical bubbles to understand their connection to market fundamentals and investor behavior. These studies highlight the role of market fundamentals and investor expectations in bubble formation, revealing that it is challenging to distinguish between bubbles and fundamental-driven price changes.

Evans (1991) further extended the rational bubbles framework by incorporating stochastic processes, suggesting that prices can follow a stochastic path with occasional deviations from fundamental values due to rational speculation. West (1987) further contributed by investigating the empirical characteristics of asset price movements in the presence of stochastic bubbles. The stochastic nature of bubbles implies that they can be unpredictable, making it difficult for policymakers and investors to identify. While these models account for randomness, they still assume that investors have a consistent understanding of fundamental values.

Empirically validating rational bubbles is challenging due to the need to accurately measure fundamental values and investor expectations. Research has shown that rational bubble models do not reliably comply with empirical tests (see e.g. Adam and Szafarz, 1992; Camerer, 1989a; Flood and Hodrick, 1990b, for reviews). Diba and Grossman (1988b) introduced a test to detect rational bubbles based on the stationary properties of asset prices and dividends, whereas Evans (1991) demonstrated that such a test has limited power in detecting periodically collapsing bubbles. At a more fundamental level, Lux and Sornette (2002) and Malevergne and Sornette (2001) have further showed that the distributions of asset returns predicted by rational bubble models are inconsistent with empirical estimates.

Irrational exuberance

'Irrational exuberance' is a term proposed by former Federal Reserve Chairman Alan Greenspan in a 1996 speech, refers to the phenomenon where investor enthusiasm drives asset prices above their fundamental values. This term became popular with the publication of Robert J. Shiller's book *Irrational Exuberance* in 2000, which scrutinized the market bubbles in the 1990s and early 2000s. The concept of irrational exuberance has its root in behavioral finance, which seeks to understand how psychological and emotional biases influence financial decisions, making it critically diverges from classical financial theories like the EMH.

The theoretical foundations of behavioral finance stem largely from the Prospect Theory and Mental Accounting. Kahneman and Tversky (1979) introduced the Prospect Theory and explained that people value gains and losses differently, leading to decision-making that deviates from the expected utility theory. Thaler (1985) on the other hand explored how individuals classify and treat money differently based on subjective criteria in his work on Mental Accounting.

One of the core principles of behavioral finance is the concept of cognitive biases. These biases, such as overconfidence, loss aversion, and herd behavior, can cause investors to make decisions that deviate from rationality. It also emphasizes the role of emotions in financial decision-making. Emotions such as greed and fear can drive market volatility and create feedback loops that exacerbate price fluctuations. During periods of market exuberance, greed can fuel speculative buying, pushing prices to unsustainable levels. Conversely, during market downturns, fear can lead to panic selling, causing prices to drop below intrinsic values. These emotional re-

sponses question the idea that markets are always rational and emphasize the importance of psychological factors in influencing the market dynamics. Social factors also play a significant role in behavioral finance. Information cascades, where individuals base their decisions on the actions of others rather than on the information they have, can lead to collective irrationality and the formation of bubbles. Additionally, the influence of media and opinion leaders can influence investor sentiment and drive market trend.

The recognition of cognitive biases and emotional responses have led to specific behavioral finance models. The Noise Trader model (Black, 1986a; De Long et al., 1990) introduces irrational traders into market dynamics, illustrating how their non-fundamental trading can result in increased asset price volatility and the creation of bubbles. Similarly, models on Sentiment Indices (Baker and Wurgler, 2006) and Overconfidence (Odean, 1998) explore how excessive optimism and self-assurance among investors can lead to mispricings that significantly deviate from fundamental values. Furthermore, the Limits to Arbitrage model (Shleifer and Vishny, 1997) addresses why rational, profit-seeking investors sometimes cannot exploit mispricings due to risks and other market frictions, thereby allowing bubbles to inflate further and persist. The perspectives adopted by these models are crucial in explaining why speculative bubbles can form and eventually burst.

The behavioral finance models, however, still rely on the framework of rational expectations to some extent (Barberis and Thaler, 2002; Shleifer, 2000). This is because irrational behaviors do not always provide a structured framework. While these models can effectively capture the bubbles dynamics (although sometimes considered retrospective), they struggle to provide systematically applicable theories and therefore need the mathematical rigor from traditional finance theories. The integration of rational frameworks helps make the behavioral finance models analytically tractable, but also constrains the full expression of irrational market dynamics.

LPPLS models

Unlike traditional financial models that primarily analyze trends and cycles in the financial market, the Log-Periodic Power Law Singularity (LPPLS) models (Johansen, Ledoit, and Sornette, 2000; Sornette, 2003a; Sornette, 2003b) are designed to identify the early signals of speculative bubbles. This class of models is inspired by phase transitions and critical phenomena in physics, such as the behavior of materials near their melting points or the alignment of spins in a ferromagnetic material. By viewing financial

markets as complex systems nearing critical phase transition, a slight perturbation can lead to a substantial shift in market dynamics – a bifurcation. This critical state is characterized in LPPLS models by parameters that scale with a log-periodic power law, indicating the susceptibility near a critical point.

The power law component expressed by $K(t_c - t)^m$ captures the super-exponential acceleration in asset prices, where t_c is the critical time expected for the bubble to burst and m between 0 and 1 quantifies the intensity of the bubble, representing a power law singularity. This super-exponential acceleration is distinctive of speculative bubbles and is driven by positive feedback loops where increased investment further drives up asset prices, leading to even more investment. Meanwhile, the log-periodic component involving $\cos(\omega \ln(t_c - t) + \phi)$ adds increasing frequency of oscillations as markets approach t_c, enhancing the model's ability to capture the dynamics at the brink of a crash. The crash hazard rate represents the probability per unit time that a crash will occur, given that it has not yet happened. The hazard rate of the model escalates as the critical time nears, indicating heightened market instability and improving the model's predictive accuracy regarding the timing of significant market corrections. This synthesis of the power law growth and log-periodic fluctuations thus provides a comprehensive framework for understanding and predicting the phases and instabilities of financial bubbles.

The LPPLS models consider the presence of two types of traders within the market: rational traders who operate with rational expectations, and noise traders who exhibit herding behavior and amplify market volatility. One can roughly see the power law part representing the rational acknowledgment of larger risks and hence higher returns, and the oscillation part as reflecting the interactions between different market participants and the growing instability of the market. In fact, Schatz and Sornette (2020) have clarified that there is no such thing as a rational bubble. The behaviors of agents need to depart in one way or another from the standard assumption, i.e., there needs to be at least one agent that is not rational or not utility-maximizing, in the formation of a bubble.

The LPPLS model has been applied extensively to diagnose and predict financial bubbles across various asset classes, including equities (e.g. Zhang et al., 2016), commodities (e.g. Sornette, Woodard, and Zhou, 2009), real estates (e.g. Zhou and Sornette, 2006), and cryptocurrencies (e.g. Wheatley, Sornette, Huber, et al., 2019). Latest development includes the Deep LPPLS model, which utilizes deep neural networks to enhance the precision of pa-

rameter estimation and extend the forecasting capabilities to critical points in various systems (Nielsen, Sornette, and Raissi, 2024).

Having outlined the historical development and milestones in the field, we now introduce the frameworks and methodologies employed in this thesis.

1.3 New bubble models with non-local hazard rate

In the models discussed above, the relationship between risk and return is often depicted as an instantaneous connection, which oversimplifies the market dynamics where inertia exists. Assuming that risks are instantly reflected in returns effectively ignores the cumulative effects of prolonged mispricings built up through a complex interplay between market forces and investor psychology. During bubble periods, the perceived risk is not instantly reflected in returns due to investor exuberance, herd behavior, and various frictions. The assumption of instantaneous relationship can lead to misunderstanding of the timing and the magnitude of bubble dynamics. Therefore, it is natural to consider a non-local crash hazard rate.

Non-local behavioral hazard rate

In financial markets, the concept of 'anchoring' refers to the cognitive bias where individuals rely on previously known information as reference point to make new decisions. This anchoring effect significantly influences investors' perceptions regarding asset prices. Specifically, in the context of market bubbles and crashes, investors tend to anchor on past price levels when assessing current prices and predicting future trends.

Building on the concept of anchoring, we propose the non-local behavioral self-referencing hazard rate, which is determined by a non-local estimation of mispricing that reflects deviations from long-term price trends. In the novel class of models that incorporates such non-local hazard rate, investors rely on relative metrics, namely the ratio of the current price to the price at a previous time adjusted for expected growth. This ratio serves as an indicator of mispricing, denoted as $\delta_{t,\tau}$, where t is the current time point and τ is the time period over which the mispricing is assessed (τ is also among the model parameters to be estimated). This idea is illustrated in Figure 1.1

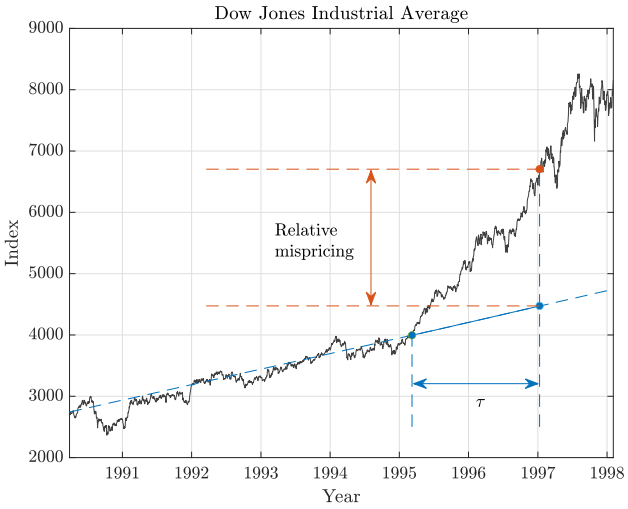

FIGURE 1.1: Illustration of the concept of mispricing captured by the ratio $\delta_{t,\tau}$

When $\delta_{t,\tau}$ exceeds 1, it signals overpricing. The larger $\delta_{t,\tau}$ is, the larger the departure of the price from its inferred long-term level. When $\delta_{t,\tau}$ becomes significantly larger than 1, investors progressively presume that the price trajectory is not sustainable and anticipate a growing probability of a correction, which will terminate the bubble in progress and bring back the price to its long-term growth trend.

The expression for the log-return of the risky asset r_t in the model is given by

$$r_t = \bar{r} + \sigma_t \cdot \varepsilon_t + \kappa_t \cdot (\lambda_t - J_t \cdot I_t),$$

where $\bar{r}$ is the unconditional expected return, κ_t is the average size of a jump, σ_t is the volatility, ε_t is a standard normal Gaussian variable, J_t is a positive standardized random variable, and I_t is a Bernoulli random variable with a conditional success probability λ_t, which is, as we explained earlier, the non-local crash hazard rate.

In this setup, λ_t is modeled using a logistic function of the mispricing measure because it is a straightforward function that ensures small mispricings have minimal impact on crash probability, while large deviations push the probability of a crash close to 1.

The model inherently incorporates a positive feedback mechanism: the higher the price deviates from what can be expected over the long term,

the larger the probability of a crash. This elevated crash probability, in turn, demands a higher compensatory return for investors to undertake the increased risk. As the expected return rises, it can drive prices even higher, further deviating from the fundamental value and increasing the crash hazard rate. This cycle can continue, potentially fueling bubble development until a significant correction occurs in the form of a market crash. When the bubble eventually bursts, the market correction is not just a reflection of current risk but also a consequence of the accumulated imbalances.

Self-exciting hazard rate

The non-local behavioral hazard rate allows for the incorporation of past information over extended periods, offering a robust method for modeling mispricing and market corrections. However, the jump occurrence in the previously proposed model is governed by a Bernoulli process, meaning the jump clustering is a result of the built-up high crash hazard rate due to a deficit of correcting jumps, and each jump within the cluster is independently triggered by some level of mispricing. This independence restricts the model from accounting for the clustering of jumps resulting from the endogenous interplays within the market, as illustrated by e.g. *reflexivity* (Soros, 1988). This limitation is also a common problem for models that use Poisson and Lévy processes to model jumps. Their assumption of event independence prohibits them from accounting for the clustering of jumps identified in financial series.

The remedy we propose is to add a self-exciting term to the crash hazard rate. Given the scope of our study, we conveniently employ the discrete-time Hawkes process, allowing the crash hazard rate to be expressed as the intensity of a Hawkes process

$$\lambda_t = \mu_t + \sum_{j=1}^{t-1} \phi(j\Delta) J_{t-j},$$

where the baseline intensity μ_t is characterized by a simple linear function of the non-local self-referencing mispricing. As mispricing level increases, the probability of immigrant jumps occurring increases, and the offspring jumps also start from a higher initial intensity, making the system more unstable. The higher probability of frequent and closely clustered jumps demands higher expected returns for risk compensation, boosting the bubble development until corrective jumps occur.

Empirically, asset prices seem more volatile than their fundamental values (Shiller, 1981b; Shiller, 1992a). This is known as 'excess volatility', which suggests that price fluctuations are far greater than what can be explained by relevant information. This leads to the understanding that price dynamics are largely endogenous, driven by mechanisms within the market itself. The inclusion of the self-excitation of jump process enhances the positive feedback loop and attributes a significant portion of market movements to endogenous dynamics. This new framework demonstrates how internal market behaviors can scale up to produce critical phenomena, aligning with bubble models like LPPLS, where price movements and speculative bubbles are described as consequences of market sentiments and positive feedback loops. Integrating this mispricing-jump interaction with Hawkes process allows our model to naturally capture the clustering of jumps and the endogenous dynamics of market volatility.

A key feature of the self-exciting hazard rate is that once the system is triggered by external shocks, the subsequent jumps can sustain even after the initial trigger diminishes. This is in contrast to a non-self-exciting hazard rate, where the probability of jumps would decrease once the mispricing or external shock is removed. The self-exciting nature of the model means that past events continue to influence future events, perpetuating the volatility even in the absence of new external triggers.

To quantify mispricing and its impact on the hazard rate, the model uses a measure similar to the one in the previous model, defined as the ratio of the current price to a reference price adjusted for expected growth over a period. This model with self-exciting hazard rate is in particular good at capturing asymmetries in volatility and jumps. It also has the ability to replicate key stylized facts of financial time series, such as volatility clustering, leverage effects, and fat-tailed return distributions. Unlike traditional volatility models like GARCH, which require additional constructs to account for the aforementioned features, the Hawkes process inherently captures them through its self-exciting nature. This makes the model both parsimonious and effective for practical applications in risk management and regulation.

Methodology

The major tasks facing econometric models are to rigorously estimate and correctly interpret their parameters. For our first model, Maximum Likelihood Estimation (MLE) offers a straightforward way to estimate the

parameters that best fit the observed data when the log-likelihood function is tractable. For the second model, a more complex Monte-Carlo Expectation-Maximization (MCEM) is employed to handle the log-likelihood function that does not have a closed form, especially in the presence of latent variables (in our case, the jumps). This method uses Monte Carlo simulations to estimate the conditional expectations required in the E-step of the EM algorithm. By simulating numerous potential outcomes, the log-likelihood function can be effectively evaluated. Bayesian optimization is used in the M-step to improve the accuracy of parameter estimation given it is gradient-free and naturally deals with the noise intrinsic to MC simulations.

The models are validated using synthetic data, where the underlying parameters are known, making it possible to assess the quality of parameter estimation. The validated models are used to simulate asset prices to explore different market phases and their implications. The models are then applied to real-world financial data from major stock indices to diagnose historical price movements, revealing the underlying dynamics of the markets using the estimated parameters. The robustness and applicability of these models are showcased through their ability to simulate and diagnose the dynamics of financial markets.

1.4 STOCK MARKET CRASHES ARE OUTLIERS

Johansen and Sornette (1998a) argue that the largest crashes of the century, such as those in 1929 and 1987, do not merely represent the tails of a probabilistic distribution but are statistical anomalies that occur with a frequency and magnitude not predictable by standard models. This argument is supported by their analysis showing that these extreme events significantly deviate from the exponential decay expected in less extreme drawdowns. Later, Johansen and Sornette (2002b) extend this study to a broader range of financial markets and show that large price drawdowns are considered outliers because these extreme events are characterized by different statistical behaviors compared to more moderate fluctuations, suggesting that they are driven by distinct mechanisms.

These insights into the outliers in financial markets set the stage for the development of more refined statistical tests to identify these outliers. By recognizing the extreme events as beyond the norms, one can better deal with the risks brought by such anomalies.

The tests statistics

We introduce two ratio-based robust test statistics, the Max-Robust-Sum (MRS) and the Sum-Robust-Sum (SRS), which are extensions of established test statistics and are developed to provide robust outlier detection capabilities.

The Max-Robust-Sum (MRS) statistic is a modification of Kimber (1982), which considers the ratio of the maximum value to the sum of all observations. The original statistic is simple and effective in certain settings, but it performs poorly in the presence of multiple outliers due to its susceptibility to masking. The MRS addresses this issue by adjusting the denominator to exclude the top m values where m is a predefined number that potentially represents the maximum number of outliers:

$$T_{j,m}^{MRS} = \frac{x_{(j)}}{\sum_{i=m+1}^{n} x_{(i)}}, \quad m \geq 0 \, .$$

Here $x_{(j)}$ represents the j-th largest value. The adjustment makes the statistic less prone to masking by reducing the disproportionate influence of the largest values, allowing outliers within the dataset to be detected more reliably.

The Sum-Robust-Sum (SRS) statistic is a modification of Lewis and Fieller (1979), which considers the sum of the suspected outliers compared to the sum of the remaining data. Similar to the MRS, the SRS modifies the original statistic by excluding the top m values in the denominator:

$$T_{r,m}^{SRS} = \frac{\sum_{i=1}^{r} x_{(i)}}{\sum_{i=m+1}^{n} x_{(i)}}, \quad m \geq 1 \, .$$

In this formula, the sum of the suspected top r outliers is compared to the sum of the remaining data, preventing the swamping of the test statistic by extreme but not necessarily indicative data points.

We provide the analytical distribution functions of these statistics, both of which are independent of the parameters from the underlying distribution, making them highly adaptable and reducing the risk of parameter estimation errors influencing the test outcomes. This feature is critical in practice, where parameter misestimation can lead to significant biases in outlier detection.

Through empirical analysis using datasets from finance and other social sectors, the MRS and SRS have demonstrated superior performance in identifying outliers.

Extreme Value Theory (EVT)

EVT is primarily concerned with the types of distributions that extreme values can follow given a large sample from a population. The theory categorizes these extremes into three types of distributions: Gumbel, Fréchet, and Weibull, which collectively form the Generalized Extreme Value distribution. Each type corresponds to different tail behaviors, which are crucial for constructing statistical tests to the specific characteristics of the data.

In the context of our outlier detection tests, EVT provides a theoretical justification for focusing on the tail properties of the distribution. By employing EVT, we ensure that our methods are not only robust to outliers but also adapted to the inherent properties of financial data, which typically exhibit fat-tailed distributions. This adaptation is critical, as financial markets are known for their propensity for tail risks, which traditional models often underestimate.

The Pickands-Balkema-de Haan Theorem, a cornerstone of EVT, particularly influences our approach. This theorem states (Embrechts, Klüppelberg, and Mikosch, 1997): For a broad range of distributions, for random variable X, with sufficiently high threshold u, the excess distribution function, $F_u(x) = P\{X - u \leq x | X - u > 0\}$ (i.e., the tail of the distribution function), is approximated by the Generalized Pareto Distribution Function. This approximation is crucial for our testing strategy as it allows us to model the tail behavior effectively and to apply the MRS and SRS tests, which are specifically designed for samples with underlying exponential and Pareto tail distributions. Within this framework, we can detect outliers by evaluating how significantly a data point deviates from this modeled tail behavior.

Employing EVT effectively simplifies statistical models by reducing complex or unknown distribution tails to a more manageable exponential form. Methods based on exponential and Pareto tail approximations can be robust as they rely not on the exact shape of the tail but on its asymptotic behavior as extremes become large. Given the ubiquity of exponential tails in real-world data distributions, our tests gain broader applicability, enhancing their utility across different data types and scenarios.

Methodology

Our methodology begins with comparative analyses using synthetic datasets designed to mimic various practical scenarios. Each dataset is structured to contain no outliers, single outliers, multiple dispersed outliers, and clustered outliers. We run these analyses to test the resilience of our outlier detection methods against common issues like masking and swamping. The MRS and SRS are compared with traditional test statistics and are found to have superior performance in these challenging scenarios.

Applying the statistics to real-world data involves a crucial step of threshold selection, which is guided by EVT. The threshold determines at what point the data behavior is extreme enough to be considered for outlier testing. In our approach, we employ several methods to determine the optimal threshold for applying the outlier tests, including (mainly) the Hill plot, the AIC of exponential and non-parametric fits. And their comparative evaluation helps in selecting a threshold that maximize the power of the test.

Once the threshold is determined, the EVT-based MRS and SRS are applied to detect the extremes by focusing on how data points above the threshold deviate from the modeled tail distribution. To illustrate the practical implications of our methodology, we conduct case studies on financial crashes, nuclear power generation accidents, stock market returns, epidemic fatalities, and city sizes. These case studies from different sectors demonstrate the effectiveness of the proposed approach, and significant outliers are detected and related to the concept of 'Dragon King' events – meaningful extreme outliers that arise from a unique generating mechanism.

1.5 OVERVIEW OF THE THESIS

This thesis is based on a series of three research papers. Chapter 2 is based on the paper co-authored with Yannick Malevergne and Didier Sornette; chapter 3 is based on the paper co-authored with Alexander Wehrli and Didier Sornette; chapter 4 is based on the paper co-authored with Didier Sornette.

The first part of this thesis investigates the econometric modeling of financial bubbles through innovative approaches to crash hazard rate estimation. The models leverage behavioral finance concepts, notably the anchoring effect, to offer a new perspective on the risk-return dynamic in financial markets.

- **Chapter 2, Modeling financial crashes with non-local behavioral self-referencing hazard rate** introduces a time series model where the crash hazard rate is a function of a non-local estimation of mispricing. The model combines a traditional jump-diffusion framework with elements in behavioral finance, enabling it to effectively describe drawdowns and crashes as market conditions characterized by clustered negative jumps over finite periods. This approach addresses a common issue in most crash jump models used in financial time series, which often fail during calibration because they unrealistically assume crashes occur through a single large negative jump. The estimation of the model parameters is performed on both synthetic and real market data. Results from daily series of three stock indices reveal that the hidden expected return generally exceeds the realized return over time, indicating that financial markets tend to be consistently underpriced.

- **Chapter 3, Modeling financial crashes with self-exciting hazard rate** extends the model introduced in the previous chapter by introducing a self-exciting component to the crash hazard rate. The authors demonstrate that the model not only explains the mechanism of market crashes but also captures key stylized facts of financial time series including volatility clustering, leverage effects, fat-tailedness, long memory, without relying on conventional volatility models like GARCH. The model is further applied to major market indices and it is revealed that the distinctive performance trends of the indices are largely driven by underlying differences in their baseline intensities and degree of self-excitation, rather than by their underlying returns or diffusive volatilities.

Transitioning from econometric models, the second part of the thesis focuses on the detection of outliers in financial data using robust test statistics. Based on Extreme Value Theory, this approach can be broadly applied to data having approximately exponential or Pareto tails.

- **Chapter 4, Detecting outliers in samples with exponential and Pareto tails** establishes robust test statistics for detecting multiple outliers in samples with exponential tails. The test statistics are shown to reduce the susceptibility of inward sequential testing to masking, potentially revitalizing the utility of inward approach. A comprehensive comparison of the test statistics is done, considering performance

of the proposed tests in both block and sequential tests, and contrasting their performance with classical test statistics across various data scenarios. Case studies across different social sectors are carried out to showcase the versatility and effectiveness of these methods in practical applications.

Bibliography and appendix follow each chapter.

REFERENCES

Adam, M. and Szafarz, A. (1992). "Speculative Bubbles and Financial Markets". In: *Oxford Economic Papers* 44.4, 626.

Baker, M. and Wurgler, J. (2006). "Investor sentiment and the cross-section of stock returns". In: *The Journal of Finance* 61.4, 1645.

Barberis, N. and Thaler, R. (2002). *A Survey of Behavioral Finance*. Working Paper 9222. National Bureau of Economic Research.

Black, F. (1986a). "Noise". In: *The Journal of Finance* 41.3, 528.

Blanchard, O. (1979). "Speculative bubbles, crashes and rational expectations". In: *Economics Letters* 3.4, 387.

Blanchard, O. and Watson, M. (1982). *Bubbles, Rational Expectations and Financial Markets*. Working paper no. 945. NBER.

Camerer, C. (1989b). "Bubbles and Fads in Asset Prices". In: *Journal of Economic Surveys* 3.1, 3.

De Long, B., Shleifer, A., Summers, L., and Waldmann, R. (1990). "Noise Trader Risk in Financial Markets". In: *Journal of Political Economy* 98.4, 703.

Diba, B. and Grossman, H. (1988a). "Explosive rational bubbles in stock prices". In: *American Economic Review* 78.3, 520.

Diba, B. and Grossman, H. (1988b). "The Theory of Rational Bubbles in Stock Prices". In: *The Economic Journal* 98.392, 746.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). " Modelling extremal events: for insurance and finance. Vol. 33". In: *Springer*.

Evans, G. (1991). "Pitfalls in Testing for Explosive Bubbles in Asset Prices". In: *The American Economic Review* 81.4, 922.

Fama, E. (1970). "Efficient Capital Markets: A Review of Theory and Empirical Work". In: *Journal of Finance* 25.2, 383.

Flood, R. and Hodrick, R. (1990a). "On Testing for Speculative Bubbles". In: *Journal of Economic Perspectives* 4.2, 85.

Froot, K. and Obstfeld, M. (1991). "Intrinsic Bubbles: The Case of Stock Prices". In: *The American Economic Review* 81.5, 1189.

Galbraith, J. K. (1994). *A Short History of Financial Euphoria*. Penguin Publishing Group.

Garber, P. (2000). *Famous First Bubbles: The Fundamentals of Early Manias*. MIT Press.

Johansen, A., Ledoit, O., and Sornette, D. (2000). "Crashes as critical points". In: *International Journal of Theoretical and Applied Finance* 3.2, 219.

Johansen, A. and Sornette, D. (1998a). " Stock market crashes are outliers". In: *Eur. Phys. J. B* 1, 141.
Johansen, A. and Sornette, D. (2002b). " Large stock market price drawdowns are outliers". In: *Journal of Risk* 4, 69.
Kahneman, D. and Tversky, A. (1979). "Prospect Theory: An Analysis of Decision under Risk". In: *Econometrica* 47.2, 263.
Keynes, J. M. (1936). *The General Theory of Employment, Interest, and Money*. Macmillan.
Kimber, A. C. (1982). "Tests for many outliers in an exponential sample". In: *Applied Statistics*, 263.
Kindleberger, C. (1978). *Manias, Panics, and Crashes: A History of Financial Crisis*. New York, NY: Basic Books.
Lewis, T. and Fieller, N. R. J. (1979). "A recursive algorithm for null distributions for outliers: I Gamma Samples". In: *Technometrics* 21, 371.
Lucas, R. E. (1972). "Expectations and the neutrality of money". In: *Journal of Economic Theory* 4.2, 103.
Lux, T. and Sornette, D. (2002). "On Rational Bubbles and Fat Tails". In: *Journal of Money, Credit and Banking* 34.3, 589.
Malevergne, Y. and Sornette, D. (2001). "Multi-dimensional rational bubbles and fat tails". In: *Quantitative Finance* 1, 533.
Malliaris, A. G. and Urrutia, J. L. (1992). "The International Crash of October 1987: Causality Tests". In: *The Journal of Financial and Quantitative Analysis* 27.3, 353.
Minsky, H. (1977). "The Financial Instability Hypothesis: An Interpretation of Keynes and an Alternative to "Standard" Theory". In: *Challenge* 20.1, 20.
Muth, J. (1961). "Rational Expectations and the Theory of Price Movements". In: *Econometrica* 29.3, 315.
Nielsen, J., Sornette, D., and Raissi, M. (2024). "Deep LPPLS: Forecasting of temporal critical points in natural, engineering and financial systems". In: *Swiss Finance Institute Research Paper No. 24-33*.
Odean, T. (1998). "Volume, volatility, price, and profit when all traders are above average". In: *The Journal of Finance* 53.6, 1887.
Santos, M. and Woodford, M. (1997). "Rational asset pricing bubbles". In: *Econometrica* 65, 19.
Schatz, M. and Sornette, D. (2020). "Inefficient bubbles and efficient drawdowns in financial markets". In: *International Journal of Theoretical and Applied Finance* 23.7, 2050047.

Shiller, R. J. (1981a). "Do Stock Prices Move Too Much to Be Justified by Subsequent Changes in Dividends?" In: *American Economic Review* 71, 421.

Shiller, R. J. (1992b). *Market Volatility*. Vol. 1. Cambridge, MA: The MIT Press.

Shleifer, A. (2000). *Inefficient Markets: An Introduction to Behavioural Finance*. Oxford: Oxford University Press.

Shleifer, A. and Vishny, R. (1997). "The limits of arbitrage". In: *The Journal of Finance* 52.1, 35.

Sornette, D. (2003a). "Critical market crashes". In: *Physics Reports* 378.1, 1.

Sornette, D. (2003b). *Why Stock Markets Crash: Critical Events in Complex Financial Systems*. Princeton, NJ: Princeton University Press.

Sornette, D., Woodard, R., and Zhou, W. (2009). "The 2006–2008 oil bubble: Evidence of speculation, and prediction". In: *Physica A: Statistical Mechanics and its Applications* 388.8, 1571.

Soros, G. (1988). *The Alchemy of Finance: Reading the Mind of the Market*. Simon & Schuster.

Soros, G. (2009). *The Crash of 2008 and What it Means: The New Paradigm for Financial Markets*. PublicAffairs.

Thaler, R. (1985). "Mental Accounting and Consumer Choice". In: *Marketing Science* 4.3, 199.

Tirole, J. (1982). "On the Possibility of Speculation under Rational Expectations". In: *Econometrica* 50.5, 1163.

Tirole, J. (1985). "Asset bubbles and overlapping generations". In: *Econometrica* 53.6, 1499.

Weil, P (1987). "Confidence and the Real Value of Money in an Overlapping Generations Economy". In: *The Quarterly Journal of Economics* 102.1, 1.

West, K. (1987). "A Specification Test for Speculative Bubbles". In: *The Quarterly Journal of Economics* 102.3, 553.

Wheatley, S., Sornette, D., Huber, T., Reppen, M., and Gantner, R. N. (2019). "Are Bitcoin bubbles predictable? Combining a generalized Metcalfe's Law and the Log-Periodic Power Law Singularity model". In: *Royal Society Open Science* 6.180538.

Zhang, Q., Sornette, D., Balcilar, M., Gupta, R., Ozdemir, Z. A., and Yetkiner, H. (2016). "LPPLS bubble indicators over two centuries of the SP 500 index". In: *Physica A: Statistical Mechanics and its Applications* 458, 126.

Zhou, W. and Sornette, D. (2006). "Is there a real-estate bubble in the US?" In: *Physica A: Statistical Mechanics and its Applications* 361.1, 297.

2

MODELING FINANCIAL CRASHES WITH NON-LOCAL BEHAVIORAL SELF-REFERENCING HAZARD RATE

In this chapter, we introduce a novel class of models with its crash hazard rate being determined by a function of a *non-local* estimation of mispricing. Rooted in behavioral finance, the non-local estimation embodies in particular the characteristic of 'anchoring' on past price levels and the 'probability judgment' about the likelihood of a crash as a function of the self-referential mispricing, enabling us to disentangle the risk-return relationship from its instantaneous connection. By describing drawdowns and crashes as market regimes with correlated negative jumps clustering over a finite period of time, our model provides a solution to the problem plaguing most crash jump models, which are in general rejected in calibrations of real financial time series because they assume that crashes occur in a single large negative jump, which is counterfactual. The model estimation is implemented on synthetic time series and real markets, shedding light on the estimation of the 'true' expected return, which is usually confounded by the entanglement between volatility and jump risks. Estimated from the daily time series of three stock indexes, the hidden expected return exhibits a secular increase over time and tends to be larger than the realized return, suggesting that financial markets have been overall underpriced.

2.1 INTRODUCTION

Financial bubbles, i.e. transient explosive prices followed by crashes, have intrigued economists for centuries because of their dramatic influence and inherent complexity. The possible origins, underlying mechanisms and dynamical behaviors of bubbles are intensively discussed by researchers, practitioners and policymakers (see e.g. Kaizoji and Sornette, 2010; Stiglitz, 1990, for reviews). In a market populated by fully rational agents and assuming general equilibrium, Tirole (1982) rules out the existence of bubbles, and thus an asset must be valued according to its fundamental. For finite horizon investors, bubbles cannot arise as long as rational investors are unconstrained from selling the desired number of shares in all future contingency. For infinite horizon investors, bubbles seem to be allowed formally

when removing the transversality condition (Blanchard, 1979; Blanchard and Watson, 1982), but Tirole (1985) and Santos and Woodford (1997) show that strong rational expectations bubbles are necessarily driven by market inefficiencies that transcend equilibria in markets populated by rational, utility-maximizing investors. In fact, Schatz and Sornette (2020) have clarified that there is no such thing as a rational (expectation) bubble. To arrive at a price process that follows a strong rational bubble, the behaviors of agents need to depart in one way or another from the standard assumption, i.e., there needs to be at least one agent that is not rational or not utility-maximizing. This is for instance instantiated in the Johansen-Ledoit-Sornette model of rational expectation bubbles, in which a population of rational fundamentalists co-exist with noise traders following social imitation to make their investment decisions (Johansen, Ledoit, and Sornette, 2000; Johansen, Sornette, and Ledoit, 1999a).

In rational bubble models, the risk-return relationship imposes that the conditional expected return is proportional to the contemporaneous crash hazard rate. This condition is usually assumed to hold instantaneously. It poses a serious empirical problem because real crashes are usually large drawdowns developing over weeks to months (Johansen and Sornette, 2001b; Johansen and Sornette, 2010a). We thus argue that the condition matching instantaneously return and risk is not a reasonable building block for a sound model of financial bubbles because it assumes perfect markets and no friction. The condition is even less likely to be true in times of exuberant bubbles and of punishing crashes. Moreover, when the price plateaus for some time at a very large overvaluation level, which does occur empirically, rational expectation imposes that the crash hazard rate vanishes instantaneously, and thus crash risks vanish. This prediction is counterfactual, as many large crashes occur after the price has peaked and gone sideway for weeks or months. In this paper, we thus propose to replace the instantaneous connection between risk and return by a non-local relationship that accounts for behavioral anchoring on past price levels.

A huge literature investigates the theoretical formulation and the empirical detectability of rational bubbles (see e.g. Adam and Szafarz, 1992; Camerer, 1989a; Flood and Hodrick, 1990b, for reviews). It has shown that rational bubble models do not reliably comply with empirical tests. Diba and Grossman (1988a) introduce a test to detect rational bubbles based on the stationary properties of asset prices and dividends, whereas Evans (1991) has proven that such test has limited power in detecting periodically collapsing bubbles. At a more fundamental level, Lux and Sornette (2002)

and Malevergne and Sornette (2001) demonstrate that the distributions of asset returns predicted by rational bubble models are inconsistent with empirical estimates.

Empirically, asset prices seem more volatile than their fundamental values (Shiller, 1981b; Shiller, 1992a). This feature is attributed to behavioral traders who are not fully rational and may exhibit social imitation leading to collective behaviors such as herding (Shleifer, 2000; Thaler, 1994). Additionally, Ma et al. (2021) report that the proportion of noise traders increases during crises even if they leave the market sooner than others because they follow sub-optimal strategies[1]. In models where rational investors interact with noise traders whose trading incentives are biased, the price of an asset can deviate from its fundamental if rational agents are limited in their capacity to eliminate the mispricing (De Long et al., 1990). Abreu and Brunnermeier (2003) further show that rational investors prefer to ride rather than attack the bubbles due to the difficulty in determining the beginning of the bubble and synchronize their market impact. Demos and Sornette (2017) present empirical evidence that diagnosing the beginning of a bubble is in fact much easier than forecasting its demise, suggesting another reason for the lack of synchronization and the robustness of bubbles. Johansen, Ledoit, and Sornette (2000) and Johansen, Sornette, and Ledoit (1999a) introduce a framework that combines the rational expectation theory and behavioral imitation and herding. The main characteristic of this framework that embodies a positive feedback loop is the transient faster-than-exponential growth of the price during the bubble phase (Sornette and Cauwels, 2015). Empirical evidence of such explosive behaviors can be found in (Ardila-Alvarez, Forro, and Sornette, 2021; Johansen and Sornette, 2010a; Lin, Schatz, and Sornette, 2019).

Different from the existing literature, we propose a novel class of asset price processes in which the crash hazard rate is a function of the estimated *non-local* mispricing, embodying a positive feedback mechanisms that leads to a faster-than-exponential growth in asset prices. The rational expectation setting ensures that the expected return conditional on no crash must offset the expected loss in case of a crash. To the best of our knowledge and as mentioned above, in the rational bubble literature, this balance condition is always expressed in terms of contemporaneous returns. We depart from this stance and determine the crash hazard rate by a function of a non-local estimation of mispricing. Specifically, our model assumes that investors

1 See also Bekiros (2010) for evidence of the profitability of behavioral trading rules in highly speculative environment.

estimate the mispricing as an exponential moving average of the difference between the present and the past log-prices over a long time scale (typically one year or more), and the crash hazard rate is characterized by an S-shaped logistic function of the mispricing. The model describes jump occurrences as outcomes of the interplay between the mispricing and the crash hazard rate. In other words, an increase in the mispricing increases the crash probability and, by the condition of no-arbitrage, leads to a larger conditional expected return. This process introduces a positive feedback mechanism in the price dynamics: the higher the price above what can be expected over a long term, the larger the probability of a crash; the larger the probability of a crash, the higher the compensatory rate of return for investors to undertake the increasing risk, further possibly fueling a possible bubble development until a cluster of correcting jumps occur.

Unlike jump clustering traditionally being featured as a sequence of jumps with each increasing the likelihood of subsequent ones, our model views jump clustering as a result of accumulated mispricing, reflecting the market's response to it and its effect over time. When, by the occurrence of stochastic fluctuations, there is a deficit of correcting jumps, mispricing develops in the presence of a finite crash hazard rate. The growing mispricing leads to a further increase in the crash hazard rate, eventually triggering a cluster of negative jumps over a relatively short period of time, i.e. a realistically looking crash.

Our view of behavioral anchoring also resonates with the path-dependent volatility models, which emphasize the role of historical data in informing current volatility, as presented in Hobson and Rogers (1998) and Guyon and Lekeufack (2023). In this respect, in the model we propose, past price trends that increase (resp. decrease) mispricing increase (resp. decrease) the intensity of jumps and thus future realized volatility. Our focus on investor sentiment and decision-making is thus consistent with path-dependent volatility models and shares the premise that past market behavior has a significant impact on current market dynamics. In other words, historical market performance has an important impact in shaping investor expectations and market outcomes. Moreover, incorporating a jump process enables us to capture abrupt market changes, at the cost of introducing additional complexity in model calibration, particularly in estimating the jump process parameters.

In order to obtain a sound model calibration and robust estimation of the parameters, the maximum likelihood estimation (MLE) is performed on the analytically derived log-likelihood function with the help of the Nelder-

Mead algorithm improved by Taboo Search. The parameter estimation is applied to synthetic time series and real markets (DAX, HSI and S&P 500) and complemented by an analysis of bias estimation. Our model estimation provides a novel and improved method for estimating the underlying expected return, whose value is usually obscured by jumps and their economic compensator.

The rest of the paper is organized as follows. Section 2.2 lays out the model with a detailed presentation of the return and jump processes. Section 2.3 presents the statistical properties of the model with simulations and analytical derivations of the stationary conditions and of the log-likelihood function. Section 2.4 (respectively Section 2.5) presents the results of the parameter estimation for synthetic (respectively real) time series. Section 2.6 concludes. Additional derivations, proofs and robustness checks can be found in Appendices.

2.2 PRESENTATION OF THE MODEL

2.2.1 *Definition of the return process and rational expectation condition*

- $r_t = \ln \frac{S_t}{S_{t-1}}$
- μ_t, σ_t and κ_t are $\mathcal{F}_{t-1}$ measurable
- ε_t is a standard normal Gaussian variable independent of $\mathcal{F}_{t-1}$
- J_t is a positive standardized random variable independent of $\mathcal{F}_{t-1}$
- I_t is a Bernoulli random variable with conditional success rate λ_t
- ε_t, J_t and I_t are mutually independent conditional on $\mathcal{F}_{t-1}$

Consider a discrete-time economy where trades take place at time $t = 0, 1, 2, \ldots$ One time step can be regarded as one trading day for the sake of illustration. The considered process is defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ endowed with the filtration $\{\mathcal{F}_t\}_{t\in\{0,1,2,\ldots\}}$. Two assets are traded in the economy: a risk-free asset whose rate of return is r_f and a risky asset whose return is r_t at time t. With the definition $r_t = \ln \frac{S_t}{S_{t-1}}$ where S_t is the risky asset price, the log-return of the risky asset at time t follows the dynamics

$$r_t = \mu_t + \sigma_t \cdot \varepsilon_t - \kappa_t \cdot J_t \cdot I_t \,, \tag{2.1}$$

where μ_t, σ_t and κ_t are $\mathcal{F}_{t-1}$ measurable, ε_t is a standard normal Gaussian variable independent of $\mathcal{F}_{t-1}$, J_t is a positive standardized random variable independent of $\mathcal{F}_{t-1}$, and I_t is a Bernoulli random variable with conditional success probability $\Pr(I_t | \mathcal{F}_{t-1}) = \lambda_t$. The random variables ε_t, J_t and I_t are mutually independent conditional on $\mathcal{F}_{t-1}$. Notice that the model specification encompasses several standard models such as the (geometric) random walk model ($\mu_t = \mu$, $\sigma_t = \sigma$, $\kappa_t = 0$), the GARCH model ($\mu_t = \mu$, $\sigma_t = \text{GARCH}(p,q)$, $\kappa_t = 0$) and most of the models that incorporate rational bubbles ($\mu_t = \mu$, $\sigma_t = \sigma$, $\kappa_t = \kappa$, $\lambda_t = \lambda$).

The expected return conditional on $\mathcal{F}_{t-1}$ is given by

$$\mathbb{E}_{t-1}[r_t] = \mu_t - \kappa_t \cdot \lambda_t , \tag{2.2}$$

where $\mathbb{E}_t[\cdot]$ stands for $\mathbb{E}[\cdot | \mathcal{F}_t]$, as usual. Under the constraint that the conditional expected return is constant and equal to $\bar{r}$, we have $\mathbb{E}_{t-1}[r_t] = \bar{r}$, hence

$$\mu_t = \bar{r} + \kappa_t \cdot \lambda_t, \tag{2.3}$$

which is nothing but the conditional expected return given no crash occurs at time t: $\mathbb{E}_{t-1}[r_t | I_t = 0] = \mu_t$. By substitution in equation (2.1), we obtain

$$r_t = \bar{r} + \kappa_t \cdot \lambda_t + \sigma_t \cdot \varepsilon_t - \kappa_t \cdot J_t \cdot I_t . \tag{2.4}$$

The condition $\mathbb{E}_{t-1}[r_t] = \bar{r}$ translates into the no-arbitrage condition or martingale condition for the process $r_t - \bar{r}$ to be associated with rational expectations, and $\bar{r}$ is interpreted as the benchmark risk-free rate r_f plus the risk premium of the risky asset.

2.2.2 *Definition of the mispricing and jump probability*

In order to specify λ_t, the conditional probability of a jump at time t, we assume that the investors acknowledge the difficulties and uncertainties in estimating the fundamental value of the risky security, in the spirit of Black (1986b) famous quote 'an efficient market is one in which price is within a factor 2 of value'. Given the difficulties in quantifying mispricing as a deviation from an absolute (fundamental) price level, we consider that the investors rely rather on a relative metric, namely they assess the amplitude of the price change over a finite time period. They thus compare the observed price at time t to that at an earlier time $t - \tau$ (τ is for instance one calendar year), and form their judgment on the presence of a bubble

based on whether the growth of the price from $t-\tau$ to t is 'reasonable' in comparison to their expectations shaped for instance by some average growth rate over a long-term, in this paper proxied by $\bar{r}$. This idea, is captured by the ratio

$$\delta_{t,\tau} = \frac{S_t}{S_{t-\tau} \cdot e^{\bar{r}\cdot\tau}}\,, \tag{2.5}$$

whose deviation from 1 quantifies the departure of the current price S_t from what would have been the asset price had it grown at the expected rate $\bar{r}$ during the period under consideration. The price $S_{t-\tau} \cdot e^{\bar{r}\cdot\tau}$ can be seen as the *anchoring price* at time t. The difference $\delta_{t,\tau} - 1$ quantifies the over/under-pricing with respect to the expected long-term trend. The larger $\delta_{t,\tau} - 1$ is, the larger the departure of the price from its inferred long-term level. When $\delta_{t,\tau}$ becomes significantly larger than 1, investors progressively presume that the price trajectory is not sustainable and anticipate a growing probability of a correction, which will terminate the bubble in progress and bring back the price to its long-term growth trend.

Expression (3.5) can be rewritten as

$$\frac{1}{\tau}\ln\delta_{t,\tau} = \frac{1}{\tau}\sum_{k=1}^{\tau}\ln\frac{S_{t-\tau+k}}{S_{t-\tau+k-1}} - \bar{r} = \frac{1}{\tau}\sum_{k=1}^{\tau}\left(r_{t-\tau+k} - \bar{r}\right)\,, \tag{2.6}$$

which defines $\frac{1}{\tau}\ln\delta_{t,\tau}$ as the average daily excess return over the long-term growth rate $\bar{r}$. For estimation purpose, it is natural to introduce a better-behaved smooth version of expression (3.6) as an exponential moving average of the returns over $\tau \simeq 1/(1-a)$ time steps

$$\ln\delta_{t,a} = (1-a)\left(r_t - \bar{r}\right) + a\cdot\ln\delta_{t-1,a}\,, \tag{2.7}$$

whose solution reads

$$\ln\delta_{t,a} = (1-a)\sum_{k=0}^{\infty} a^k\left(r_{t-k} - \bar{r}\right)\,. \tag{2.8}$$

By the correspondence $1-a \approx \frac{1}{\tau}$, one can see that expression (3.8) is the exponentially smoothed version of (3.6) and thus

$$\ln\delta_{t,a} \approx \frac{1}{\tau}\ln\delta_{t,\tau}, \quad \text{with } a \approx 1 - \frac{1}{\tau}\,. \tag{2.9}$$

In summary, we assume that investors form an estimation $\delta_{t,a}$ of the possible existence of some overpricing via formula (3.8). Then, their collec-

tive anticipation of a correction leads to the formation of an effective jump probability at time step t that we postulate to be given by

$$\lambda_t = L\left(\frac{\ln \delta_{t-1,a} - \ln \delta_{\text{ref}}}{s}\right), \tag{2.10}$$

where

$$L(x) := \frac{1}{1+e^{-x}} \tag{2.11}$$

denotes the (standard) logistic function, δ_{ref} is the average daily mispricing ratio $\delta_{t-1,a}$ beyond which the probability for a jump to occur at time t is larger than one-half, and the scale factor $s > 0$ controls the reactivity with which λ_t varies from 0 to 1 as the mispricing develops, i.e. s controls the steepness of the logistic function.[1] Defining

$$X_t := \frac{1}{s} \ln \frac{\delta_{t-1,a}}{\delta_{\text{ref}}}, \tag{2.12}$$

expression (2.10) becomes

$$\lambda_t = L(X_t) \tag{2.13}$$

with

$$X_t = (1-a)\cdot \bar{X} + a \cdot X_{t-1} + \eta \cdot (r_{t-1} - \bar{r}), \tag{2.14}$$

which is derived from (3.7) with

$$\bar{X} := -\frac{\ln \delta_{\text{ref}}}{s} \quad \text{and} \quad \eta := \frac{1-a}{s}. \tag{2.15}$$

Expression (2.14) is an AR(1) equation for the variable $(X_t - \bar{X})$, so that $\mathbb{E}[X_t] = \bar{X}$. The introduction of X_t, η along with the function $L(\cdot)$ is for simplicity in notation and estimation.

Notice that the jump probability λ_t introduces a positive feedback onto the asset price dynamics. The larger the current price S_t, the larger the probability λ_{t+1} of a jump at the next time step, and the larger the conditional expected return given no jump will occur as seen in equation (2.3). Hence, as long as no or few jumps occur, the asset price exhibits a faster-than-exponential growth due to increasing conditional expected returns (Sornette and Cauwels, 2015).

1 The choice of the logistic function 2.10 for the crash hazard rate is mainly for illustration purpose but also because of the standard use of the logistic function to represent decision functions (called logit in McFadden, 1974).

2.2.3 *Summary of the model equations*

Putting all the above together, and assuming additionally a GARCH (1,1) dynamics for the volatility σ_t, our model reads

$$\begin{aligned} r_t &= \bar{r} + \kappa_t \cdot \lambda_t + \sigma_t \cdot \varepsilon_t - \kappa_t \cdot J_t \cdot I_t \,, \\ \sigma_t^2 &= \bar{\sigma}^2 \cdot (1-\alpha-\beta) + \alpha \cdot (r_{t-1} - \bar{r})^2 + \beta \cdot \sigma_{t-1}^2 \,, \\ X_t &= (1-a) \cdot \bar{X} + a \cdot X_{t-1} + \eta \cdot (r_{t-1} - \bar{r}) \,, \\ \text{with} \;\; \lambda_t &= \Pr\left[I_t \,|\mathcal{F}_{t-1}\right] = L\left(X_t\right) \quad \text{and} \quad L\left(x\right) = \frac{1}{1+e^{-x}} \,, \end{aligned} \tag{2.16}$$

where the first equation is the dynamics of the log-return of the risky asset, the second equation writes σ_t as following a standard GARCH (1,1) process, the latter ones describe the measurement of the mispricing with the jump probability.

In the present work, we do not consider the complication when the amplitude κ_t is time-dependent, instead we restrict our attention to the simplest case where $\kappa_t = \kappa$ being constant.[1] The jump probability λ_t is determined via the logistic function by the variable X_t. Parameter $\bar{X}$ controls the small background probability of jumps $\lambda_{\text{background}} := L\left(\bar{X}\right)$ in the absence of mispricing. Indeed, absent mispricing, $\delta_t \simeq 1$ and $X_t \simeq \bar{X}$, so that $\lambda_{\text{background}} \ll 1$ requires $e^{-\bar{X}} \gg 1$, meaning that reasonable values of $\bar{X}$ are negative. For instance, for $\bar{X} = -4$, $\lambda_{background} \approx 0.018$, which means that there is a 1.8% probability for a negative jump of amplitude κ to occur in a given time step. Parameter η quantifies how the mispricing impacts the jump probability. Parameter a encodes the time $\tau \approx 1/(1-a)$ over which the mispricing of the risky asset away from the long term return $\bar{r}$ is influencing the jump probability λ_t.

Parameters $\bar{X}$, η and a can be considered as resulting from a re-parameterization of the intuitive derivation motivating the presented model. Given (2.15), $\ln \delta_{\text{ref}}$ can be expressed as a function of a, $\bar{X}$ and η as

$$\ln \delta_{\text{ref}} = -\frac{1-a}{\eta} \bar{X} \,. \tag{2.17}$$

1 As our model is discrete in time, with one time step taken to represent one trading day, one can expect that there is no jump on some days and several jumps on other days. Hence, it would be justified to consider that κ_t is distributed in sizes or even that κ_t is itself proportional to the jump rate, i.e. $\kappa_t = \kappa \cdot \lambda_t$ for some constant κ. This would amount to consider that the jump J_t at a given discrete time step t is actually the aggregation of N_t jumps of average size κ drawn from a Poisson process with intensity λ_t. Hence the average size of the aggregated jump would be $\kappa \cdot \mathbb{E}_{t-1}\left[N_t\right] = \kappa \cdot \lambda_t$.

As mentioned above, a realized average excess return $\frac{1}{\tau}\sum_{k=1}^{\tau}(r_{t-1}-\bar{r}) \approx \ln \delta_{\text{ref}}$ makes the jump probability grow from $\lambda_{\text{background}}$ to 0.5. For $a = 0.996$ (corresponding to $\tau = 250$ financial trading time steps, i.e. one calendar year), $\bar{X} = -4$ and $\eta = 8$, equation (2.17) gives $\ln \delta_{\text{ref}} = 0.2\%$ corresponding to a yearly 65% excess return. If we switch the narrative from excess return to mispricing level, this would amount to say that, when the average daily mispricing level reaches $\delta_{\text{ref}} - 1 = 0.2\%$ corresponding to a mispricing level of $\delta_{\text{ref}}^{250} - 1 = 65\%$ over one year, the probability that a jump occurs is approximately 0.5 per time step.

2.3 STATISTICAL PROPERTIES OF THE MODEL

2.3.1 *Synthetic example*

As a visual illustration of the model properties, Figures 2.1 and 2.2 show respectively a typical synthetic price trajectory and its returns over 40 years (10000 trading days), generated with the following parameters:

- long-term expected rate of return $\bar{r} = 7\%$ per annum;
- $J_t \sim \mathcal{E}(1)$ follows a standard exponential distribution and the average size $\kappa = 2\%$;
- GARCH parameters $\alpha = 0.05$, $\beta = 0.94$ so that $(\alpha + \beta < 1)$; and GARCH volatility $\bar{\sigma} = 15\%$ per annum;
- $\bar{X} = -4$, $\eta = 8$, $a = 0.996$ so that the reference period $\tau = \frac{1}{1-a}$ is 250 days (one calendar year).

As discussed at the end of the last section, with these parameter values, the background probability of jumps is $\lambda_{\text{background}} \approx 0.018$ and a jump probability $\lambda_t = 0.5$ per time step is associated with an average daily mispricing level of 0.2%. In Figure 2.1, one can observe three large rises of the jump probability λ_t that peak around day 2000, day 3200 and day 8100 with the jump probability λ_t around 0.5 (the first two) and 0.2 (the third one), associated with average daily mispricing levels around respectively 0.2% and 0.13%, which correspond to mispricing levels around respectively 65% and 38% over one year. The rises of λ_t lead to clustering of negative jumps associated with subsequent crashes or drawdowns, producing shapes in the asset price of usual speculative bubbles, except that, in the third bubble case, the price level quickly rebounds and plateaus after a rather

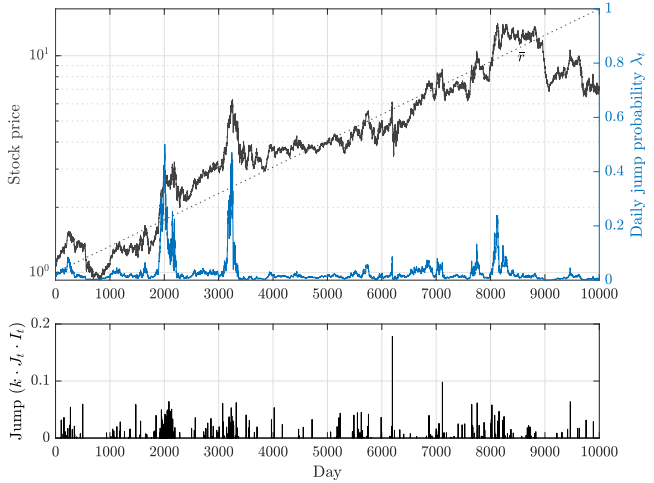

FIGURE 2.1: Synthetic price trajectory generated with parameters $(\bar{r}, \bar{\sigma}, \alpha, \beta, \kappa, \bar{X}, \eta, a) = (0.07$ (ann.), 0.15 (ann.), 0.05, 0.94, 0.02, -4, 8, 0.996) in semi-logarithmic scale (upper panel, left axis) with the price growing over the long term at rate $\bar{r}$ shown with the dotted line, jump probability (upper panel, right axis) and jump occurrence (lower panel). The average annual realized return is 4.7% and the realized volatility is 22.7%.

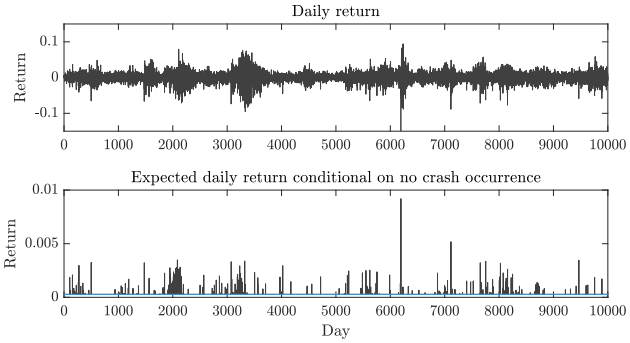

FIGURE 2.2: Daily returns of the synthetic price trajectory shown in Figure 2.1 (upper panel) and the expected daily return conditional on no crash occurrence at the next time step (lower panel) with the horizontal line representing the level of $\bar{r}$ (corresponding to 7%/250 = 0.028%=0.00028 per day).

small correction. As a consequence of the rebound and the long-lived plateau lasting over two years, the jump probability also remains markedly above the background level until the mispricing level diminishes. This illustrates the main mechanisms of the model that large values of the jump probability λ_t, which in turn produce large conditional expected returns given by expression (2.3), are not necessarily associated with crashes. The phenomenology of bubble formation, peak and drawdown is much richer than hitherto modeled in most previous works in which a single large negative jump brings a bubble to an end. Moreover, one can notice the occurrences of transient upward convex log-price trajectories, which qualify the presence of faster-than-exponential growth (recall that, in the linear-log representation, a straight line qualifies an exponential growth), representing in particular herding behavior among investors.

Comparing Figure 2.1 and 2.2, one can also observe the intermittent bursts of the volatility associated with the clustering of jumps. The seemingly paradoxical behavior that bubbles are often associated with small volatility, as observed here, has been empirically documented by Sornette, Cauwels, and Smilyanov (2018) in an extensive empirical study of 40 historical bubbles. Persistent and sizable mispricing occurs when the jumps remain insufficient to correct it, leading to the long-term existence of large jump probabilities and hence of excess return.

In addition to the visual realizations in Figures 2.1 and 2.2, Appendix 2.A presents a more quantitative study contributing to the conclusion that the model generates synthetic time series that obey the standard stylized facts of asset prices with the occurrence of bubbles. The synthetic price trajectory with its returns presented in Figure 2.19 and 2.20 is an instance of our systematic study on the influence of the parameters κ, $\bar{X}$ and η. In particular, We vary κ, $\bar{X}$ and η with the theoretical volatility and $\ln \delta_{\text{ref}}$ (2.17) almost unchanged. In Figure 2.19, the average jump size κ is smaller and the background jump probability is larger ($|\bar{X}|$ is smaller) compared to Figures 2.1. The comparison illustrates the property that, when the jumps are on average smaller in size and more frequent while the reference level of the mispricing and the theoretical volatility remain the same, the dynamics of the asset price as well as the jump probability is less exuberant. Meanwhile the clustering of jumps and the burst of returns are less noticeable, resulting in smaller bubbles. These qualitative features also help determine the relevant ranges of the parameters to generate time series with bubbles that mimic those occurring in real financial markets. Specifically, a reasonable

range of parameters we have found is: $\kappa \in [0.001, 0.1]$, $\bar{X} \in [-10, -1]$ and $\eta \in (0, 10]$.

2.3.2 *Stationary distribution of returns and its moments*

In this section, we analyze the first moments of the stationary distribution of the returns of the model defined in subsection 2.2.3. The formal proof that the process of returns r_t (2.16) admits a stationary distribution with finite second moment is given in Appendix 2.C.

The unconditional expected return is indeed $\mathbb{E}[r_t] = \bar{r}$ and the unconditional variance reads

$$\mathrm{Var}[r_t] = \mathbb{E}\left[\sigma_t^2\right] + \mathbb{E}\left[\kappa_t^2\left(\lambda_t \cdot \mathbb{E}\left[J_t^2\right] - \lambda_t^2\right)\right]. \tag{2.18}$$

By Jensen's inequality, $\mathbb{E}\left[J_t^2\right] \geq \mathbb{E}[J_t]^2 = 1$, therefore $\mathrm{Var}[r_t] \geq \mathbb{E}\left[\sigma_t^2\right]$. In particular, if we assume that J_t is a standard exponential random variable, we have $\mathbb{E}\left[J_t^2\right] = 2$ and

$$\mathrm{Var}[r_t] = \mathbb{E}\left[\sigma_t^2\right] + \mathbb{E}\left[\kappa_t^2\left(2\lambda_t - \lambda_t^2\right)\right]. \tag{2.19}$$

The first expectation term in the right hand side of expression (2.19) can be evaluated from the GARCH volatility with the condition $(\alpha + \beta < 1)$ for a finite second moment to exist. Taking the expectation of both sides of the GARCH equation in (2.16), we get

$$\mathbb{E}\left[\sigma_t^2\right] = \bar{\sigma}^2 + \frac{\alpha \cdot \mathbb{E}\left[\kappa_t^2\left(2\lambda_t - \lambda_t^2\right)\right]}{1 - \alpha - \beta} \tag{2.20}$$

and

$$\mathrm{Var}[r_t] = \bar{\sigma}^2 + \frac{(1-\beta) \cdot \mathbb{E}\left[\kappa_t^2\left(2\lambda_t - \lambda_t^2\right)\right]}{1 - \alpha - \beta}. \tag{2.21}$$

With the assumption that $\kappa_t = \kappa$ constant, we obtain

$$\mathrm{Var}[r_t] = \bar{\sigma}^2 + \frac{(1-\beta) \cdot \kappa^2 \cdot \mathbb{E}\left[2\lambda_t - \lambda_t^2\right]}{1 - \alpha - \beta}. \tag{2.22}$$

The skewness of the distribution of returns is also easy to express. With κ constant in order to avoid a too cumbersome equation, we have

$$s^3 = \frac{\mathbb{E}\left[(r_t - \bar{r})^3\right]}{\mathbb{E}\left[(r_t - \bar{r})^2\right]^{3/2}} \tag{2.23}$$

$$= -2\left(1 - \frac{\alpha}{1-\beta}\right)^{3/2} \frac{\mathbb{E}\left[\left(1-(1-\lambda_t)^3\right)\right]}{\mathbb{E}\left[\left(1-(1-\lambda_t)^2\right)\right]^{3/2}} \left(1 + \frac{(1-\alpha-\beta)\,\bar{\sigma}^2}{(1-\beta)\,\kappa^2 \mathbb{E}\left[\left(1-(1-\lambda_t)^2\right)\right]}\right)^{-3/2}, \tag{2.24}$$

which gives large negative skewness provided that α remains small compared to $1-\beta$. Using expression (2.22) and the definition of X_t in (2.16), we obtain

$$\mathrm{Var}\left[X_t\right] = \frac{\eta^2}{1-a^2}\left(\bar{\sigma}^2 + \frac{(1-\beta)\cdot\kappa^2\cdot\mathbb{E}\left[\left(2\lambda_t - \lambda_t^2\right)\right]}{1-\alpha-\beta}\right). \tag{2.25}$$

As the time scale $\tau \simeq 1/(1-a)$ over which the mispricing is evaluated is expected to be large (one year or more), the coefficient a is close to 1 and the stochastic process X_t is thus close to a unit-root AR(1) process. In this regime, the approximate stationary distribution of X_t is known to be well-approximated by (Cumberland and Sykes, 1982)

$$X_t \sim \mathcal{N}\left(\bar{X}, \frac{\eta^2}{1-a^2}\cdot\mathrm{Var}\left[r_t\right]\right). \tag{2.26}$$

A more accurate approximation, which accounts for the tail behavior of the distribution, can be found in (Olvera-Cravioto, 2010). As a consequence, the asymptotic distribution of $\lambda_t \overset{\text{law}}{=} L(X_t)$ is the logistic-normal distribution (Aitchison and Shen, 1980). Note that the logistic-normal distribution admits either one single mode in the bulk of the distribution or two modes located at both ends (see Figure 2.3). Denoting the mean and the standard deviation of X_t in (2.26) by μ_{logit} and σ_{logit}, the location of the mode(s) is given by the solution(s) to equation: $\ln\left(\frac{x}{1-x}\right) = \sigma_{\text{logit}}^2\cdot(2x-1) + \mu_{\text{logit}}$.

Hence, the logistic-normal density is symmetric when $\mu_{\text{logit}} = 0$, skewed to the right (respectively left) when $\mu_{\text{logit}} > 0$ (respectively $\mu_{\text{logit}} < 0$). It admits one single mode whenever $\sigma_{\text{logit}} \leq \sqrt{2}$ or $\left|\mu_{\text{logit}}\right| > \sigma_{\text{logit}}^2\sqrt{1 - \frac{2}{\sigma_{\text{logit}}^2}} -$

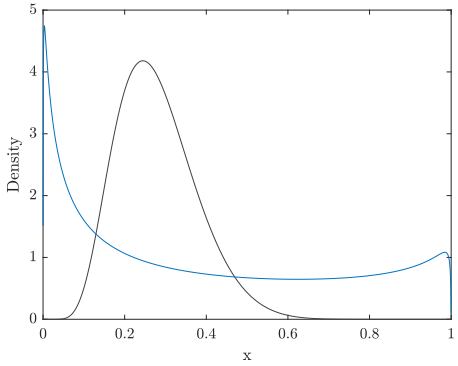

FIGURE 2.3: Density of the logistic-normal distribution with $\mu_{\text{logit}} = -1$ and $\sigma_{\text{logit}} = 0.5$ (black curve) and $\mu_{\text{logit}} = -0.75$ and $\sigma_{\text{logit}} = 2.25$ (blue curve).

$\ln \frac{1+\sqrt{1-\frac{2}{\sigma^2_{\text{logit}}}}}{1-\sqrt{1-\frac{2}{\sigma^2_{\text{logit}}}}}$ and two modes otherwise. This means that, as long as the condition

$$\text{Var}\left[r_t\right] \leq \frac{2\left(1-a^2\right)}{\eta^2} \tag{2.27}$$

holds, σ_{logit} is such that the logistic-normal density is unimodal. Substituting a with $1-\frac{1}{\tau}$ and taking the square root of both sides in (2.27), the condition becomes

$$\sqrt{\text{Var}\left[r_t\right]} \leq \frac{\sqrt{2\cdot\frac{2\tau-1}{\tau}}}{\sqrt{\tau}\cdot\eta} \simeq \frac{2}{\sqrt{\tau}\cdot\eta}, \tag{2.28}$$

which requires the theoretical volatility to not exceed $\frac{2}{\eta}$ per annum. Given the value e.g. $\eta = 4$ in our simulation above, $\frac{2}{\eta} = 50\%$ is much larger than any usual level of realized volatility of indexes and stocks. With the relevant range of parameter $\eta \in (0, 10]$, $\frac{2}{\eta}$ is equal to or above 20% and with the parameters of the GARCH process and of κ, $\bar{X}$ and η chosen accordingly, the logistic-normal density remains unimodal.

While the moments of the logistic-normal distribution have no close form, in the limit where σ_{logit} remains small, we have

$$\mathbb{E}\left[\lambda_t\right] = L\left(\mu_{\text{logit}}\right) + \frac{\sigma^2_{\text{logit}}}{2} \cdot L''\left(\mu_{\text{logit}}\right) + O\left(\sigma^4_{\text{logit}}\right), \tag{2.29}$$

$$\mathbb{E}\left[\lambda_t^2\right] = L\left(\mu_{\text{logit}}\right)^2 + \sigma^2_{\text{logit}}\left(L\left(\mu_{\text{logit}}\right) \cdot L''\left(\mu_{\text{logit}}\right) + L'\left(\mu_{\text{logit}}\right)^2\right) + O\left(\sigma^4_{\text{logit}}\right). \tag{2.30}$$

Equation (2.29) shows that, up to first order in σ_{logit}, the average jump probability is given by $L\left(\bar{X}\right)$ so that, with the range $\bar{X} \in [-10, -1]$, the average time between jumps is the order of 1 to 10^3 time steps. Expressions (2.29) and (2.30) lead to

$$\begin{aligned} 2\mathbb{E}\left[\lambda_t\right] - \mathbb{E}\left[\lambda_t^2\right] &= L\left(\mu_{\text{logit}}\right)\left(2 - L\left(\mu_{\text{logit}}\right)\right) \\ &\quad + \sigma^2_{\text{logit}}\left(\left(1 - L\left(\mu_{\text{logit}}\right)\right) L''\left(\mu_{\text{logit}}\right) - L'\left(\mu_{\text{logit}}\right)^2\right) + O\left(\sigma^4_{\text{logit}}\right), \end{aligned} \tag{2.31}$$

and, by substitution in equation (2.25), we get

$$\sigma^2_{\text{logit}} \simeq \frac{\eta^2}{1-a^2} \cdot \frac{\bar{\sigma}^2 \cdot (1-\alpha-\beta) + (1-\beta) \cdot \kappa^2 \cdot L\left(\mu_{\text{logit}}\right)\left(2 - L\left(\mu_{\text{logit}}\right)\right)}{1 - \alpha - \beta - \frac{\eta^2}{1-a^2} \cdot (1-\beta) \cdot \kappa^2 \left(\left(1 - L\left(\mu_{\text{logit}}\right)\right) L''\left(\mu_{\text{logit}}\right) - L'\left(\mu_{\text{logit}}\right)^2\right)}. \tag{2.32}$$

This approximation is valid as long as σ^2_{logit} is small enough and the stationary distribution of λ_t remains unimodal.

2.3.3 *Time-reversal asymmetry*

Time reversal asymmetry is a well-known feature of asset prices and, in particular, of stock prices. It manifests itself through the leverage effect (Black, 1976; Christie, 1982) and the asymmetric dependence between squared-returns and conditional variance – what Blanc, Donier, and Bouchaud (2017) call the *Zumbach effect*.

The simple GARCH formulation chosen in Equation (16) to model the volatility does not directly incorporate the leverage effect, as it is the case in the EGARCH, JGR-GARCH, or TARCH models, to name a few (Glosten, Jagannathan, and Runkle, 1993; Nelson, 1991; Rabemananjara and Zakoian,

1993). Nevertheless, the effect is present in the model as a consequence of the asymmetry of the return distribution. Indeed, it can be shown that

$$\mathbb{E}\left[\left(\sigma_t^2 - \mathbb{E}\left[\sigma_t^2\right]\right)\cdot\left(r_{t-1}-\bar{r}\right)\right] = \alpha\mathbb{E}\left[\left(r_{t-1}-\bar{r}\right)^3\right], \tag{2.33}$$

$$= -2\alpha\mathbb{E}\left[\kappa_t^3\left(1-\left(1-\lambda_t\right)^3\right)\right], \tag{2.34}$$

which is negative. Thus, as a first approximation, the cross-correlation between σ_t^2 and r_{t-1} is equal to α times the skewness of the return distribution, which means that the model can generate cross-correlation between σ_t^2 and r_{t-1}that can reach -30%, which is sufficiently large to be consistent with the literature.

The model is also consistent with the Zumbach effect, *i.e.* the fact that past squared returns predict future volatilities better than past volatilities predict future squared returns. To illustrate this effect, let us define the cross-correlation

$$\rho^{(2)}(\tau) = \frac{\mathrm{Cov}\left(\sigma_t^2, \left(r_{t-\tau}-\bar{r}\right)^2\right)}{\sqrt{\mathrm{Var}\left(\sigma_t^2\right)\mathrm{Var}\left(\left(r_t-\bar{r}\right)^2\right)}}, \tag{2.35}$$

as in Euch et al. (2020), and its integrated difference

$$\Delta(\tau) = \sum_{i=1}^{\tau}\left(\rho^{(2)}(i) - \rho^{(2)}(-i)\right). \tag{2.36}$$

Simple, but tedious, algebraic manipulations yield

$$\lim_{\tau\to\infty}\Delta(\tau) = \frac{1}{\alpha}\sqrt{\frac{\mathrm{Var}\left(\sigma_t^2\right)}{\mathrm{Var}\left(\left(r_t-\bar{r}\right)^2\right)}}\cdot\left(1+\mathrm{Corr}\left(\sigma_t^2,\sigma_{t-1}^2\right)\right), \tag{2.37}$$

$$= \sqrt{\frac{(1+\alpha+\beta)^2}{1+\beta(\beta-2\alpha)}\left(1-\frac{2\beta\kappa^2}{\alpha}\frac{\mathrm{Cov}\left(\sigma_t^2,\left(1-\lambda_t\right)^2\right)}{\mathrm{Var}\left(\left(r_t-\bar{r}\right)^2\right)}\right)} \times\left(1-\frac{\alpha\kappa^2\mathrm{Cov}\left(\sigma_t^2,\left(1-\lambda_t\right)^2\right)}{(1+\alpha+\beta)\,\mathrm{Var}\left(\sigma_t^2\right)}\right), \tag{2.38}$$

which is positive, and show that, at least on average, $\rho^{(2)}(\tau) > \rho^{(2)}(-\tau)$ for $\tau > 0$, *i.e.* the correlation between past squared returns and future volatilities is greater than the correlation between past volatilities and future squared returns. For a GARCH model, *i.e.* $\kappa = 0$, the relationship is simplified:

$$\lim_{\tau \to \infty} \Delta(\tau) = \frac{1 + \alpha + \beta}{\sqrt{1 + \beta(\beta - 2\alpha)}}. \tag{2.39}$$

Empirically, $\Delta(\infty)$ is found to be of the order of one, which is consistent with the usual values of the parameters α close to zero and β close to one.

2.3.4 *The log-likelihood derivation and parameter estimation*

The likelihood of the model can be expressed in closed form when J_t is exponentially distributed and κ_t is constant. Given expression (2.1), conditional on $\mathcal{F}_{t-1}$, J_t and I_t, we know that

$$r_t \,|\mathcal{F}_{t-1}, J_t, I_t \overset{\text{law}}{=} \mathcal{N}\left(\mu_t - \kappa \cdot J_t \cdot I_t, \sigma_t^2\right). \tag{2.40}$$

Taking the expectation with respect to I_t, we obtain a mixture of two Gaussian distributions

$$r_t \,|\mathcal{F}_{t-1}, J_t \overset{\text{law}}{=} \begin{cases} \mathcal{N}\left(\mu_t - \kappa \cdot J_t, \sigma_t^2\right) & \text{with probability } \lambda_t, \\ \mathcal{N}\left(\mu_t, \sigma_t^2\right) & \text{with probability } 1 - \lambda_t. \end{cases} \tag{2.41}$$

Averaging the first Gaussian distribution with respect to J_t yields the exponentially modified Gaussian distribution $\text{EMG}(\mu_t, \sigma_t, \kappa)$ with density (see appendix 2.D for details)

$$f_{\text{EMG}}(x \,|\mu, \sigma, \kappa) = \frac{1}{|\kappa|} e^{\frac{x-\mu}{\kappa} + \frac{\sigma^2}{2\kappa^2}} \left(1 - \Phi\left(\text{sgn}(\kappa)\left(\frac{x-\mu}{\sigma} + \frac{\sigma}{\kappa}\right)\right)\right), \tag{2.42}$$

$$= \frac{\sigma_t}{|\kappa|} \cdot \varphi(x \,|\mu_t, \sigma_t) \cdot R\left(\text{sgn}(\kappa)\left(\frac{x-\mu}{\sigma} + \frac{\sigma}{\kappa}\right)\right), \tag{2.43}$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal law, $\varphi(x \,|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ is the Gaussian density and $R(x) = \frac{1-\Phi(x)}{\varphi(x|0,1)}$, which can be easily evaluated as pointed out by Marsaglia (2004).

Hence the conditional distribution of the log-returns reads

$$r_t \,|\mathcal{F}_{t-1} \overset{\text{law}}{=} \begin{cases} \text{EMG}\,(\mu_t, \sigma_t, \kappa_t) & \text{with probability } \lambda_t \\ \mathcal{N}\left(\mu_t, \sigma_t^2\right) & \text{with probability } 1-\lambda_t \end{cases} \tag{2.44}$$

and the log-likelihood function of the model becomes

$$\begin{aligned} &\ln \mathcal{L}\left(\theta \,\middle|\, \{S_t\}_{t=0}^{T}\right) \\ &= \sum_{t=1}^{T} \ln\left[\lambda_t \cdot f_{\text{EMG}}\left(r_t \,|(\mu_t, \sigma_t, \kappa_t)\right) + (1-\lambda_t) \cdot \varphi\left(r_t \,|\mu_t, \sigma_t\right)\right] \\ &= \sum_{t=1}^{T} \ln \varphi\left(r_t \,|\mu_t, \sigma_t\right) + \sum_{t=1}^{T} \ln\left[1 - \lambda_t + \lambda_t \frac{\sigma_t}{|\kappa|} \cdot R\left(\kappa\left(\frac{r_t - \mu_t}{\sigma_t} + \frac{\sigma_t}{\kappa}\right)\right)\right] \end{aligned}$$

where $\theta = (\bar{r}, \bar{\sigma}, \alpha, \beta, \kappa, \bar{X}, \eta, a)$ denotes the vector of parameters.

In the following, we carry out MLE using Nelder and Mead (1965) algorithm, which is a direct search method commonly applied to optimization problems and can be easily applied to nonlinear situations for which derivatives may be unknown. The taboo search method based on Cvijovic and Klinowski (1995) is introduced to randomize the initial points for the optimization process. As a result, the algorithm avoids entrapment in local minima and ensures the continuation of the exploration of the parameter space. In general, this search gives a near-optimal final solution. The choices of the search parameters are such that an increase in any of these parameters does not help improve the result of the optimization. Indeed, the objective function (2.45) may be *sloppy*, in the word of Waterfall et al. (2006). This means that the eigenvalues of the Fisher matrix of the log-likelihood function at the point of its maximum exhibit a broad range of values, with the smallest eigenvalues being associated with quasi-degenerate directions in the space of parameters. Along these directions, the corresponding parameters are highly undetermined. In addition to the danger of being misled towards incorrect parameter values, the direct maximization of the likelihood is also likely to exhibit a slow convergence. As a tentative solution, along with the help of the search method, the parameter estimations are always verified with profile likelihood.

2.4 PARAMETER ESTIMATION FOR SYNTHETIC DATA

2.4.1 *Methodology of parameter estimation*

$\theta = (\bar{r}, \bar{\sigma}, \alpha, \beta, \kappa, \bar{X}, \eta, a)$,
$(\bar{r}, \kappa, \bar{X}) \in \mathbb{R}^3$, $(\bar{\sigma}, \eta) \in \mathbb{R}^2_+$, $(\alpha, \beta) \in \{(x, y) \in [0,1)\,,\ x + y < 1\}$, $a \in [0,1)$

Our goal is to explore the model performance in explaining real time series of financial returns. In order to efficiently perform the MLE, we introduce a multi-step methodology to estimate the parameters $(\bar{r}, \kappa, \bar{X}) \in \mathbb{R}^3$, $(\bar{\sigma}, \eta) \in \mathbb{R}^2_+$, $(\alpha, \beta) \in \{(x, y) \in [0,1)\,,\ x + y < 1\}$ and $a \in [0,1)$. Note that we introduced the model with the restriction $\kappa > 0$ in the asset price dynamics in order to capture downward jumps and crashes. Nonetheless, the model remains well-defined for $\kappa \leq 0$. Therefore it is reasonable to set the parameter space as $\kappa \in \mathbb{R}$.

Step 1: GARCH model

$$r_t = \bar{r} + \sigma_t \cdot \varepsilon_t \tag{2.45}$$

$$\sigma_t^2 = \bar{\sigma}^2 \cdot (1 - \alpha - \beta) + \alpha \cdot (r_{t-1} - \bar{r})^2 + \beta \cdot \sigma_{t-1}^2 \tag{2.46}$$

Step 2: Constant intensity model (CI model)

$$r_t = \bar{r} + \cdot\kappa \cdot \lambda + \sigma_t \cdot \varepsilon_t - \kappa \cdot J_t \cdot I_t \tag{2.47}$$

$$\sigma_t^2 = \bar{\sigma}^2 \cdot (1 - \alpha - \beta) + \alpha \cdot (r_{t-1} - \bar{r})^2 + \beta \cdot \sigma_{t-1}^2 \tag{2.48}$$

$$\text{with } \lambda = L\,(\bar{X}) \tag{2.49}$$

Step 3: Complete model

Finally, we deal with the complete model (2.16) and adopt the same strategy as in the previous step. We first define a mesh for the parameters $(\eta, a) \in \mathbb{R}_+ \times [0,1)$ and maximize the likelihood of the complete model with respect to the six parameters $(\bar{r}, \bar{\sigma}, \alpha, \beta, \kappa, \bar{X})$ at fixed η and a and with the initial value $X_0 = \bar{X}$. We use the obtained parameters to initialize the Nelder-Mead algorithm to maximize the likelihood of the complete model with respect to its eight parameters. Note that, for $\eta = 0$, the complete model reduces to the constant intensity model.

One can go further and ask which of these models are sufficient to describe the data. Here the complete model encompasses the GARCH and CI models. The two nested models will be compared with the complete model using Wilks' theorem on maximum log-likelihood ratios, which allows us to test nested hypotheses and compute the corresponding p-values. The Wilks statistics is given by $LR = 2(\ln \mathcal{L}(\hat{\theta}) - \ln \mathcal{L}(\hat{\theta_0}))$, where $\hat{\theta}$ and $\hat{\theta_0}$ are the MLEs for the alternative and the null hypotheses.

2.4.2 *Parameter estimation for the synthetic time series*

For the synthetic financial time series shown in the last section in Figure 2.1, the realized average return is 4.7% per annum and is lower than the underlying long-term rate $\bar{r} = 7\%$. The realized volatility is 22.70% per annum, close to the theoretical volatility 20.96% given by (2.22). Table 2.2 presents the estimates obtained by the methodology described in the previous subsection and their standard errors derived from the Hessian matrix of the log-likelihood function, as well as the Wilks statistics and the corresponding p-values. We find that our estimators effectively capture the underlying values of the parameters. One should however note the differences between the standard errors of the different parameters. The standard errors are small for $\bar{\sigma}, \alpha, \beta, a$, relatively large for $\kappa, \bar{X}$ and particularly large for $\bar{r}$ and η. Such differences reflect the intrinsic stochasticity of the time series of finite duration.

The quality of the estimations is further validated in Figure 2.4, where the estimated jump probability is compared with the true jump probability. One can see that the estimated jump probability captures well the dynamics of the underlying jump probability. Taking a further look at the difference between the estimated and the true jump probabilities, we find on the lower panel of Figure 2.4 that the value of the estimated jump probability is almost always larger than the true value in bubble regimes, implying that the estimated jump probability tends to slightly overestimate the emergence of a bubble and its growth compared with the true jump probability. In contrast, when the end of the bubble approaches, the estimated jump probability is slightly smaller than the true value. See also Table 2.12 and Figure 2.21 for the parameter estimations of the synthetic price trajectory in Appendix 2.A.

We are particularly gratified to observe that the estimated value for $\bar{r}$ is within 10% from the true value, while the realized average return, at 4.7% per annum, is quite smaller. Indeed, if we did not use the knowledge about

θ	$\bar{r}$ (ann.)	$\bar{\sigma}$ (ann.)	α	β	κ	$\bar{X}$	η	a	Wilks	p-value
	0.07	0.15	0.05	0.94	0.02	-4	8	0.996		
$\hat{\theta}_{\text{GARCH}}$	0.044	0.215	0.055	0.939						
	(0.028)	(0.014)	(0.0040)	(0.0044)						
$\hat{\theta}_{\text{CI}}$	0.065	0.141	0.050	0.943	0.020	-3.9			507.2	$< 10^{-4}$
	(0.029)	(0.009)	(0.0033)	(0.0038)	(0.0029)	(0.25)				
$\hat{\theta}_{\text{Complete}}$	0.077	0.138	0.048	0.943	0.017	-3.9	9.2	0.9952	46.5	$< 10^{-4}$
	(0.027)	(0.008)	(0.0034)	(0.0039)	(0.0025)	(0.27)	(1.4)	(0.0015)		

TABLE 2.2: Parameter estimation and associated standard errors for the synthetic time series in Figure 2.1, with the underlying values of the parameters. The Wilks statistics and the corresponding p-values are given for the test of the GARCH model nested in the Constant Intensity model and the test of the Constant Intensity model nested in the Complete model.

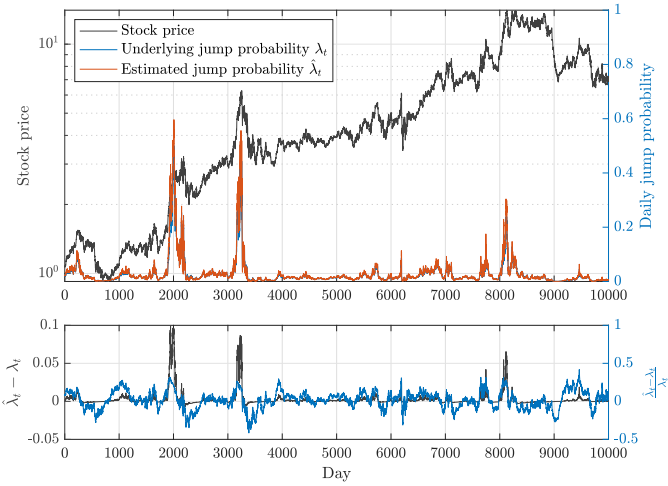

FIGURE 2.4: Synthetic price trajectory (as in Figure 2.1) in semi-logarithmic scale (upper panel, left axis) with the estimated and underlying jump probability (upper panel, right axis), the difference between the estimated and the underlying jump probabilities (lower panel, left axis) and the ratio of the difference to the underlying value (lower panel, right axis).

the generating process, we would use the standard sufficient statistics for $\bar{r}$, equal to the realized average return, which here gives an incorrect value for the expected return $\bar{r}$. This example suggests that correctly accounting for the structure of the generating process allows us to recover the expected return $\bar{r}$ which is otherwise hidden in the realized average return.

2.4.3 *Distribution of parameter estimates over realizations with fixed underlying parameters*

Having studied the parameter estimation for a single realization of the price process, we further explore the robustness of the estimation procedure by constructing the distribution of the parameter estimates over an ensemble of realizations generated with the same model parameters. We perform the parameter estimation over 100 realizations, each with 10000 time steps, generated with the parameters $(\bar{r}, \bar{\sigma}, \alpha, \beta, \kappa, \bar{X}, \eta, a) = (0.07\ \text{(ann.)}, 0.15\ \text{(ann.)}, 0.05, 0.94, 0.02, -4.0, 8, 0.996)$. The mean value and the standard deviation (within parenthesis) of the parameter estimates over these 100 realizations are as follow:

$$\begin{aligned}
&\mathbb{E}[\hat{\bar{r}}] = 0.070\,(0.025)\,\text{(ann.)}, && \mathbb{E}[\hat{\kappa}] = 0.020\,(0.002), \\
&\mathbb{E}[\hat{\bar{\sigma}}] = 0.151\,(0.008)\,\text{(ann.)}, && \mathbb{E}[\hat{\bar{X}}] = -4.0\,(0.3), \\
&\mathbb{E}[\hat{\alpha}] = 0.050\,(0.004), && \mathbb{E}[\hat{\eta}] = 8.3\,(1.5), \\
&\mathbb{E}[\hat{\beta}] = 0.940\,(0.004), && \mathbb{E}[\hat{a}] = 0.9960\,(0.0014),
\end{aligned} \tag{2.50}$$

and the distributions of the parameter estimates are presented in Figure 2.5. One can observe that the mean and mode values are essentially identical to the true values. Importantly, the close match between the standard deviations and the model-derived standard errors in Table 2.2 reinforces the credibility of the model's estimates. This alignment suggests that the model is effectively capturing the inherent variability in data.

These phenomena are observed with alternative sets of parameters, as demonstrated e.g. in Appendix 2.A. This further confirms the reliability of the model and the robustness of the estimation procedure in providing *consistent* estimates under varying conditions.

One more study of the distribution of parameter estimates is carried out here in order to investigate the impact of the GARCH effect on the model and examine how the heteroskedasticity captured by the GARCH process affects the other parameters estimation. For this purpose, another 100 synthetic time series are generated with the restricted model where

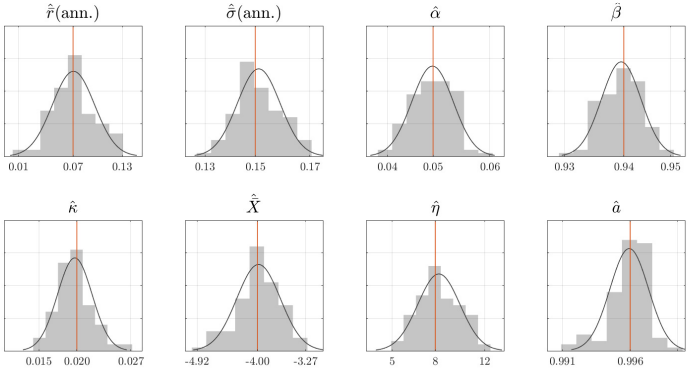

FIGURE 2.5: Distributions of the parameter estimates over 100 realizations, each with 10000 time steps, generated with the same parameters $(\bar{r}, \bar{\sigma}, \alpha, \beta, \kappa, \bar{X}, \eta, a) = (0.07 \text{ (ann.)}, 0.15 \text{ (ann.)}, 0.05, 0.94, 0.02, -4.0, 8, 0.996)$. The underlying parameters are shown with the vertical red lines with values readable on the x-axis and fits of normal distributions are shown as black smooth curves.

the GARCH parameters α and β are equal to 0 i.e. no GARCH effect. In particular, the values of the other underlying parameters remain the same except for $\bar{\sigma}$ which is chosen such that the restricted and the unrestricted models have the same theoretical volatility. Figure 2.6 shows that there is practically no difference in the means of the parameter estimates (except for $\bar{\sigma}$) and little difference in the distributions of $\kappa, \bar{X}, \eta$ and a, which can be expected to be less impacted or even not impacted by the GARCH estimation.

A systematic study of the convergence of the estimators as a function of the time series size is beyond the scope of the present study.

2.4.4 *Dependence of the expected return estimation on the average realized return*

In subsection 2.4.2, we have alluded to the fact that the estimated expected return $\bar{r}$ improves on the naive sufficient statistics for a Geometric Brownian motion in the form of the average realized return. To quantify how general is this result, Figure 2.7 presents the dependence of $\hat{\bar{r}}$ as a function of the average realized return (or average annual growth rate of the price), over the 100 realizations used in the previous subsection. We can observe that

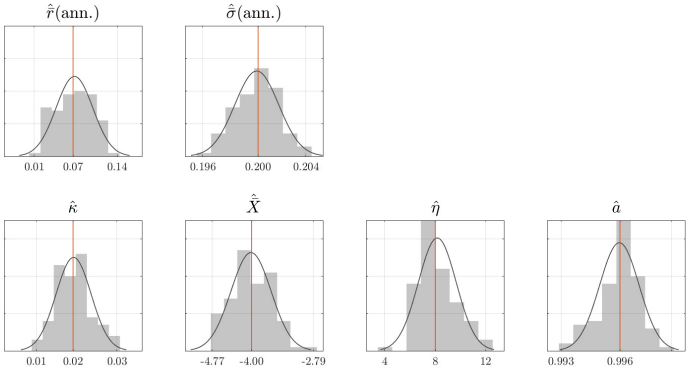

FIGURE 2.6: Distributions of the parameter estimates over 100 realizations, each with 10000 time steps, generated with the parameters $(\bar{r}, \bar{\sigma}, \kappa, \bar{X}, \eta, a) = (0.07 \text{ (ann.)}, 0.20 \text{ (ann.)}, 0.02, -4.0, 8, 0.996)$ without GARCH effect. The underlying parameters are shown with the vertical red lines with values readable on the x-axis and fits of normal distributions are shown as black smooth curves.

the average realized return does exert an influence on $\hat{\bar{r}}$, with a correlation coefficient equal to 0.7. However, the fact that the correlation coefficient is smaller than 1 confirms the value of accounting for the structure of the generating process in order to better estimate the expected return $\bar{r}$. Unfortunately, we cannot avoid a contamination of $\hat{\bar{r}}$ by the average realized return. Quantitatively, the mean values of $\hat{\bar{r}}$ and the average realized returns are almost equally close to the underlying value $\bar{r}$, nonetheless the standard deviation of the model estimates of $\bar{r}$ is a sizable 40% below the standard error of the average realized returns.

The correlation found in Figure 2.7 can be exploited to correct the bias introduced by the realized return. The regression $\hat{\bar{r}} = 0.45 r_{\text{growth}} + 0.04$, where r_{growth} is the average annual realized growth rate of the price, can be rewritten $\Delta\hat{\bar{r}} := \hat{\bar{r}} - \bar{r} = 0.45(r_{\text{growth}} - \bar{r})$ where $\bar{r} = 0.07$ is the true value for the set of parameters corresponding to Figure 2.7. This provides a corrected estimation of $\bar{r}$ given by

$$\hat{\bar{r}}_{\text{corrected}} = \hat{\bar{r}} - \Delta\hat{\bar{r}} = \hat{\bar{r}} - 0.45 r_{\text{growth}} + 0.03, \tag{2.51}$$

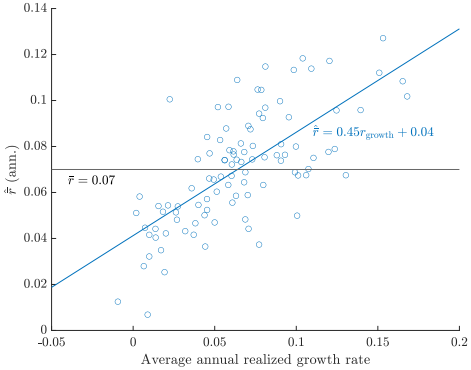

FIGURE 2.7: Estimates of the expected return $\bar{r}$ as a function of the annual mean realized growth rate over 100 realizations, each with 10000 time steps, generated with the same parameters $(\bar{r}, \bar{\sigma}, \alpha, \beta, \kappa, \bar{X}, \eta, a) =$ (0.07 (ann.), 0.15 (ann.), 0.05, 0.94, 0.02, -4.0, 8, 0.996), with the regression line of the scatter diagram. The correlation coefficient is equal to 0.7.

General application of this correction to unknown time series would require constructing an analogue of Figure 2.7 for every possible set of parameters and is beyond the scope of the present study.

2.5 PARAMETER ESTIMATION FOR REAL FINANCIAL TIME SERIES

In this section, we estimate the parameters of the model (2.16) on the daily time series of three stock indexes: the DAX Performance Index, the Hang Seng Index (HSI) and the S&P 500. The lengths of these time series are approximately 50 to 60 years. For the HSI, the estimation procedure is separately applied to the pre- and post-transition periods where changes of society and policies in 1997 caused fundamental changes of regimes in the indexes. For the S&P 500, the estimation procedure is applied to rolling windows as well as to the entire time span. After each estimation, we generate an ensemble of realizations with the estimated parameters of that index and perform the same estimation procedure on the realizations. The distributions of the estimated parameters of the simulated time series are presented with violin plots and bivariate distributions in comparison with the generating values, further proving the robustness of the model and the estimation procedure. Additional parameter estimation conducted on the

Nikkei 225, the Nasdaq Composite and the Dow Jones Industrial Average are presented in Appendix 2.B. (Note that the time series are presented with the starting point rescaled to one.)

2.5.1 *DAX Performance Index*

Figure 2.8 shows the DAX Performance Index time series from 28 September 1959 to 31 December 2020 with the estimated jump probability. The average annual realized return of the sample price trajectory is 5.9% per annum, larger than the estimate of the expected return $\bar{r}$ shown in Table 2.4.

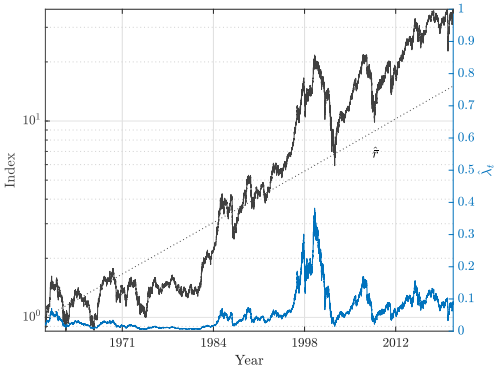

FIGURE 2.8: The DAX Performance Index daily time series from 28 September 1959 to 31 December 2020 in semi-logarithmic scale (left axis) compared with an exponential growth at the rate $\hat{\bar{r}}$ (ann.) = 4.4% (dotted line) and estimated daily jump probability (right axis). The average annual realized return is 5.9%, the realized volatility is 19.6% and the theoretical volatility is 16.7%.

As the estimation of the self-referential memory determined by the parameter $\hat{a}$ is more than 30 years, other parameter estimates are influenced by the early part of the historical data that is accompanied by very low jump probability and flat volatile prices. Given the estimates of the three parameters controlling the amplitude of the self-referencing propensity in the model, we have $\delta_{ref} = (e^{-(1-\hat{a})\frac{\hat{\bar{X}}}{\hat{\eta}}} | \hat{\bar{X}} = -3.7, \hat{\eta} = 2.5, \hat{a} = 0.99988) = 1.0002$ hence the reference of the mispricing level is $\delta_{ref}^{250\times 30} - 1 \approx 3$ over 30 years, indicating that a jump probability of 0.5 per day requires a mispricing

θ	$\bar{r}$ (ann.)	$\bar{\sigma}$ (ann.)	α	β	κ	$\bar{X}$	η	a	Wilks	p-value
$\hat{\theta}_{\text{GARCH}}$	0.105	0.215	0.117	0.868						
	(0.020)	(0.015)	(0.0062)	(0.0068)						
$\hat{\theta}_{\text{CI}}$	0.094	0.156	0.095	0.887	0.015	-3.9			71.7	$< 10^{-4}$
	(0.021)	(0.008)	(0.0057)	(0.0068)	(0.0019)	(0.24)				
$\hat{\theta}_{\text{Complete}}$	0.044	0.142	0.096	0.882	0.011	-3.7	2.5	0.99988	45.9	$< 10^{-4}$
	(0.008)	(0.006)	(0.0055)	(0.0066)	(0.0011)	(0.39)	(0.53)	(0.0001)		

TABLE 2.4: Parameter estimation and associated standard errors for the DAX daily time series with the Wilks statistics of the pairwise comparisons of the three models described in subsection 2.4.1 and the corresponding p-values. To present the results of the estimation, the same structure as Table 2.2 is used.

level of approximately 300% over 30 years, or 3.7% excess return per year. Notice that the most remarkable peak ≈ 0.4 of the jump probability in Figure 2.8 corresponds to the dot-com bubble in 2000. The model estimation suggests that this peak is associated with a mispricing level of about 230% over 30 years.

We generate 100 time series with the estimated parameters of the DAX time series in Table 2.4 and estimate the parameters of the simulated time series. Figures 2.9 and 2.10 exhibit the estimates for the simulations in violin plot and bivariate distribution. The fact that the underlying values (estimates for the DAX time series) are almost always at the center of the distributions of both plots indicates the consistency in the estimation.

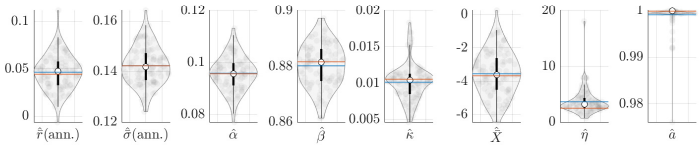

FIGURE 2.9: Violin plots of the parameter estimates over the 100 simulated time series generated with the estimated parameters of the DAX time series, with blue lines indicating the mean values of the distributions and red lines the estimates for the DAX time series.

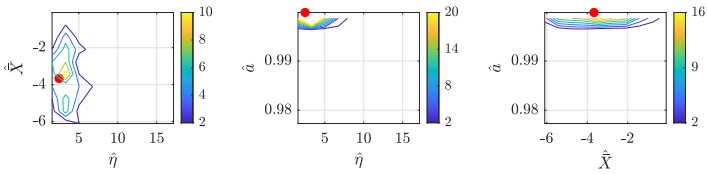

FIGURE 2.10: Isolines of bivariate distributions of the three parameters controlling the amplitude of the self-referencing propensity in the model over the 100 simulated time series generated with the estimated parameters of the DAX time series, with the red dots indicating the estimates for the DAX time series.

2.5.2 *Hang Seng Index*

The realized yearly return of the Hang Seng Index from 24 November 1969 to 31 December 2020 is 10.1%, quite smaller than the estimate of the expected return $\bar{r}$ shown in Table 2.6. One can observe in Figure 2.11 that the price followed approximately an exponential growth at the rate $\bar{r}$ until 1997 but thereafter leveled off, departing from the dotted line.

Although the distributions of estimates for the simulated time series shown in Figures 2.12 and 2.13 suggest consistency in estimation and validate the results, it is justified to distinguish different epochs and estimate the parameters separately. Indeed, the price dynamics of the HSI exhibits a change of regime in 1997, probably in part associated with the official transfer of sovereignty from the United Kingdom to the People's Republic of China on 1 July 1997 and linked up with the 1997 Asian financial crisis, which started also in July 1997 in Thailand and spread by contagion over most of East Asia and Southeast Asia. By mid-August 1998, the HSI had fallen by more than 60% to 6660 points from its previous peak and, despite a rapid recovery in the following year, has grown at a slower pace ever since, compared to the period before 1997. We therefore apply the parameter estimation to the pre- and post-transition regimes, where the transitional point is set on 13 August 1998, when the index reached its lowest during the Asian crisis. The estimations and visualizations of the two regimes are presented in Table 2.8 and Figure 2.14.

The parameter estimation reveals expected returns $\bar{r}$ around 20.7% and 12% for the two regimes respectively, both quite larger than the average realized returns of 12.9% and 6.4% respectively. Notably, the first regime, impacted by the pronounced 1973-1974 global shock, exhibits an estimate of the average jump size κ nearly triple that of the second regime and exceeds

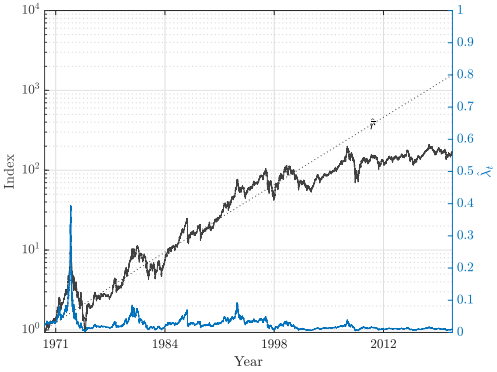

FIGURE 2.11: The Hang Seng Index from 24 November 1969 to 31 December 2020 in semi-logarithmic scale (left axis) compared with an exponential growth at the rate $\hat{\bar{r}}$ (ann.) = 14.4% (dotted line) and the estimated daily jump probability (right axis). The average annual realized return is 10.1%, the realized volatility is 28.1% and the theoretical volatility is 28.6%.

the overall estimated average jump size of the index. For the entire sample, we have $\delta_{ref} = (e^{-(1-\hat{a})\frac{\hat{X}}{\hat{\eta}}} | \hat{X} = -3.6, \hat{\eta} = 1.9, \hat{a} = 0.9991) = 1.0017$, hence the mispricing reference level is $\delta_{ref}^{250\times 4.5} - 1 = 5.6$ over 4.5 years, indicating that a jump probability of 0.5 per day requires a mispricing level of about 560% over four and a half years (i.e. 146% per year). This high reference level explains the low jump probability of the sample.

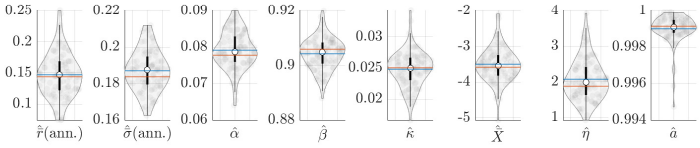

FIGURE 2.12: Violin plots of the parameter estimates over the 100 simulated time series generated with the estimated parameters of the HSI time serie, with blue lines indicating the mean values of the distributions and red lines the estimates for the HSI time series.

θ	$\bar{r}$ (ann.)	$\bar{\sigma}$ (ann.)	α	β	κ	$\bar{X}$	η	a	Wilks	p-value
$\hat{\theta}_{\text{GARCH}}$	0.209	0.312	0.117	0.872						
	(0.033)	(0.026)	(0.0058)	(0.0060)						
$\hat{\theta}_{\text{CI}}$	0.151	0.188	0.079	0.905	0.025	-3.9			560.7	$< 10^{-4}$
	(0.032)	(0.009)	(0.0049)	(0.0054)	(0.0031)	(0.22)				
$\hat{\theta}_{\text{Complete}}$	0.144	0.184	0.078	0.906	0.025	-3.6	1.9	0.9991	18.5	$< 10^{-4}$
	(0.030)	(0.008)	(0.0048)	(0.0054)	(0.0032)	(0.26)	(0.45)	(0.0004)		

TABLE 2.6: Parameter estimation and associated standard errors for the HSI daily time series with the Wilks statistics of the pairwise comparisons of the three models described in subsection 2.4.1 and the corresponding p-values.

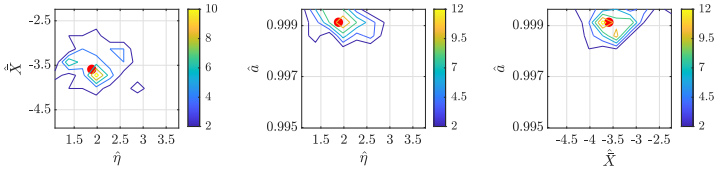

FIGURE 2.13: Isolines of bivariate distributions of the three parameters controlling the amplitude of the self-referencing propensity in the model over the 100 simulated time series generated with the estimated parameters of the HSI time series, with the red dots indicating the estimates for the HSI time series.

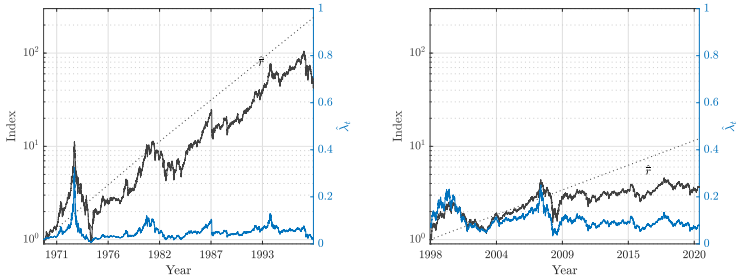

FIGURE 2.14: The HSI from 24 November 1969 to 13 August 1998 (left panel, left axis) and from 14 August 1998 to 31 December 2020 (right panel, left axis) compared with exponential growths at the rate $\hat{\bar{r}}$ (ann.) = 20.7% and 12% respectively (dotted lines) and the estimated daily jump probabilities (right axes). The average annual realized returns are respectively 12.9% and 6.4%.

θ	$\bar{r}$ (ann.)	$\bar{\sigma}$ (ann.)	α	β	κ	$\bar{X}$	η	a
$\hat{\theta}^{\text{1st}}_{\text{Complete}}$	0.207	0.212	0.109	0.873	0.033	-3.8	1.6	0.9989
(till 13 August 1998)	(0.047)	(0.016)	(0.0083)	(0.0087)	(0.0055)	(0.30)	(0.54)	(0.0007)
$\hat{\theta}^{\text{2nd}}_{\text{Complete}}$	0.120	0.169	0.054	0.936	0.012	-2.6	1.9	0.9988
(from 14 August 1998)	(0.043)	(0.015)	(0.0053)	(0.0059)	(0.0020)	(0.48)	(1.1)	(0.0008)

TABLE 2.8: Parameter estimation and associated standard errors for the two epochs of HSI.

2.5.3 *S&P 500*

The S&P 500 time series from 4 March 1957 to 31 December 2020 with the estimated daily jump probability is shown in Figure 2.15. The average realized return of the index is 6.9% per annum while the estimated expected return is $\hat{\bar{r}} = 10.8\%$ (ann.), leading to a rather large discrepancy between the price trajectory and the exponential growth with rate $\hat{\bar{r}}$ depicted by the dotted line. The distributions of the estimates for the simulated time series generated with the estimated parameters of the S&P 500 time series are presented in Figures 2.16 and 2.17, showing consistency of the estimation.

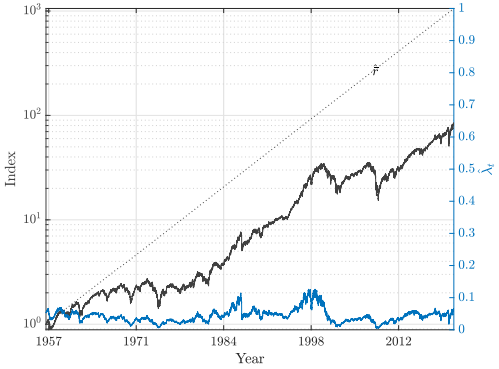

FIGURE 2.15: The S&P 500 from from 4 March 1957 to 31 December 2020 in semi-logarithmic scale (left axis) compared with an exponential growth at the rate $\hat{\bar{r}}$ (ann.) = 10.8% (dotted line) and the estimated daily jump probability (right axis). The average annual realized return is 6.9%, the realized volatility is 16.1% and the theoretical volatility is 16.3%.

θ	$\bar{r}$ (ann.)	$\bar{\sigma}$ (ann.)	α	β	κ	$\bar{X}$	η	a	Wilks	p-value
$\hat{\theta}_{\text{GARCH}}$	0.131	0.165	0.097	0.893						
	(0.016)	(0.011)	(0.0048)	(0.0052)						
$\hat{\theta}_{\text{CI}}$	0.110	0.093	0.077	0.910	0.009	-2.8			598.4	$< 10^{-4}$
	(0.016)	(0.0059)	(0.0041)	(0.0046)	(0.0009)	(0.24)				
$\hat{\theta}_{\text{Complete}}$	0.108	0.097	0.078	0.909	0.009	-2.9	3.1	0.9986	6.9	0.0314
	(0.015)	(0.0055)	(0.0040)	(0.0044)	(0.0009)	(0.23)	(0.72)	(0.0004)		

TABLE 2.10: Parameter estimation and associated standard errors for the S&P 500 daily time series with the Wilks statistics of the pairwise comparisons of the three models described in subsection 2.4.1 and the corresponding p-values.

Given the observed discrepancy between the price trajectory and the exponential growth at the rate $\hat{\bar{r}}$ (ann.), we consider the possibility for a non-stationarity of the model parameters by performing the estimation in time windows of duration of 5000 days and with increment of 2500 days. Figure 2.18 shows the results with each straight segment representing the estimated expected return of the corresponding rolling window. The estimated expected returns span a range from the low of 5.4% for 1971-1990

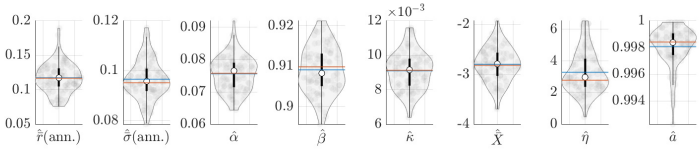

FIGURE 2.16: Violin plots of the parameter estimates over the 100 simulated time series generated with the estimated parameters of the S&P 500 time series, with blue lines indicating the mean values of the distributions and red lines the estimates for the S&P 500 time series.

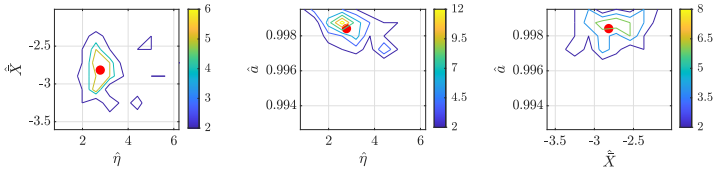

FIGURE 2.17: Isolines of bivariate distributions of the three parameters controlling the amplitude of the self-referencing propensity in the model over the 100 simulated time series generated with the estimated parameters of the S&P 500 time series, with the red dots indicating the estimates for the S&P 500 time series.

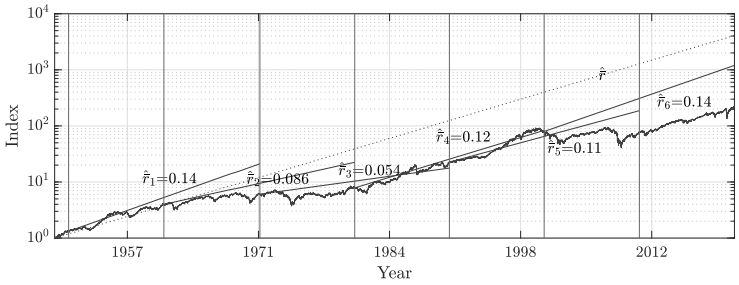

FIGURE 2.18: The S&P 500 time series compared with an exponential growth at the rate $\hat{\bar{r}}$ (ann.) = 10.8 (dotted line) estimated over the whole time interval and with the exponential growth at the different risk premia estimated in each 5000 days long windows (solid lines) as explained in the text. Each straight segment represents an exponential growth at the rate of the estimated expected return of the corresponding rolling window. The rolling window size is 5000 time steps and the increment (between neighboring vertical lines) is 2500 time steps.

to a high of 14% both for 1950-1971 and 2000-2020. Overall, they bracket nicely the expected return estimated for the whole period. Even at times of plateauing and highly volatile markets with no clear trend, the parameter estimation gives robust positive values for the expected return.

2.6 CONCLUSION

We have proposed a novel class of models in which the crash hazard rate is determined by a logit function of a non-local estimation of mispricing expressed in excess return. We have argued that, given the large uncertainties surrounding the determination of fundamental values, investors develop a kind of heuristic approach to diagnose the existence of over-pricing and of bubbles, by looking at the deviations between the current price and what it would have been if it had grown at an average long term rate, the deviation being estimated over a time of scale of one year or larger. This brings in a kind of self-referencing which is further elaborated by introducing a jump or crash hazard rate that depend on this self-referencing mispricing. The model with such crash hazard rate incorporates a positive feedback associated with the fact that a larger mispricing leads to a larger crash hazard rate which itself is compensated by a larger expected return as a compensation for the risks incurred by investors. In addition, the model also includes the behavioral anchoring on the past price levels over a long time scale. This formulation improves on previous rational expectation bubble models whose risk and return are in general instantaneously matched.

We have documented that the model produces synthetic price time series that exhibit structures and properties that are comparable to those of real financial time series, with also the existence of transient bubbles looking similar to real financial bubbles with transient faster-than-exponential growth of the price. Rather than crashing in a single jump, our model allows for multiple jumps and the synthetic time series exhibit quite realistically looking drawdowns. We have also quantified the performance of the parameter estimation for the synthetic times series, with a good recovery of the true parameters and well-understood standard deviations and inter-dependencies between the estimated parameters.

The model has been calibrated to the daily time series of three stock indexes, which has allowed us to characterize their bubbles via the determination of the time dynamics of the jump probabilities. As a key feature of our model, the usually hidden expected return (providing a guidance to what should be the long-term expected return) has been estimated for each

of the indexes over the entire time domain and over pre- and post-transition periods or rolling windows. We have found that when, the self-referencing memory is very long (e.g. more than 20 years) and underpricing exists in historical data, the estimated expected return can be close to or lower than the average realized return. Otherwise the estimated expected return is more likely to exceed the average realized return, implying that the model deems the financial time series overall to be underpriced.

APPENDIX

2.A ADDITIONAL SYNTHETIC EXAMPLE

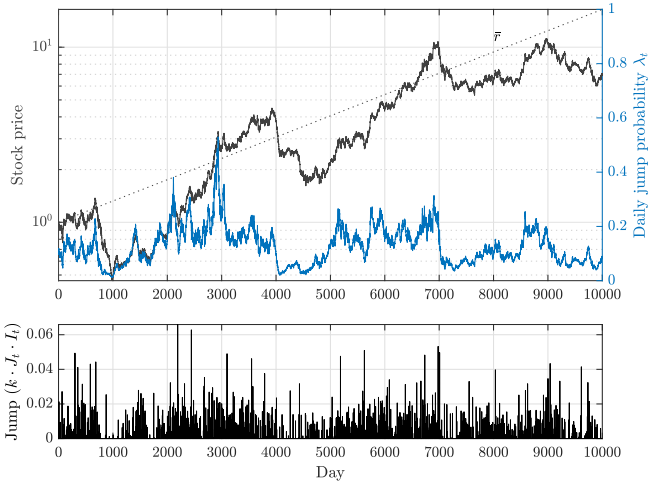

FIGURE 2.19: Synthetic price trajectory generated with parameters $(\bar{r}, \bar{\sigma}, \alpha, \beta, \kappa, \bar{X}, \eta, a)$ = (0.07 (ann.), 0.15 (ann.), 0.05, 0.94, 0.008, −2, 4, 0.996) in semi-logarithmic scale (upper panel, left axis) with an exponential growth at the rate $\bar{r}$ shown with the dotted line, jump probability (upper panel, right axis) and jump occurrence (lower panel).

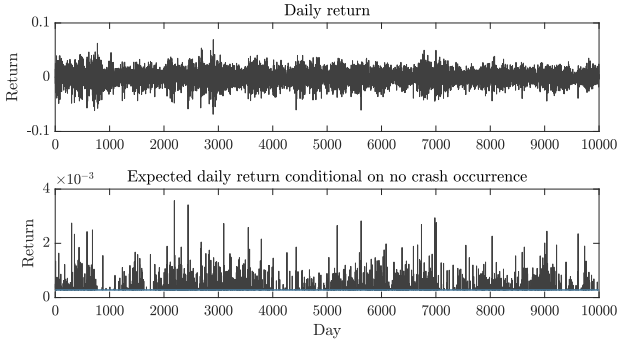

FIGURE 2.20: Daily returns of the price trajectory shown in figure 2.19 (upper panel) and the expected daily returns conditional on no crash occurrence at the next time step (lower panel) with the blue line representing the level of $\bar{r}$.

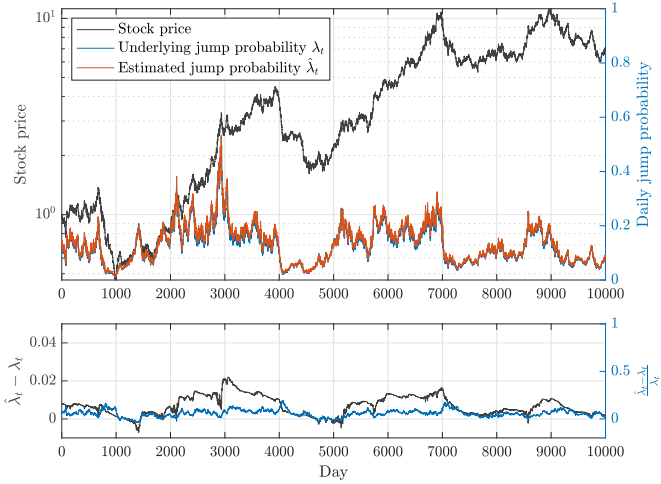

FIGURE 2.21: Synthetic price trajectory (as in Figure 2.19) in semi-logarithmic scale (upper panel, left axis) with underlying jump probability and estimated jump probability (upper panel, right axis), the difference of the estimated and the underlying jump probabilities (lower panel, left axis) and the ratio of the difference to the underlying value (lower panel, right axis). The average annual realized return is 5% and the realized volatility is 20.7%.

θ	$\bar{r}$ (ann.)	$\bar{\sigma}$ (ann.)	α	β	κ	$\bar{X}$	η	a	Wilks	p-value
	0.07	0.15	0.05	0.94	0.008	-2	4	0.996		
$\hat{\theta}_{\text{GARCH}}$	0.060	0.210	0.051	0.939						
	(0.031)	(0.010)	(0.0042)	(0.0053)						
$\hat{\theta}_{\text{CI}}$	0.063	0.156	0.050	0.939	0.0082	-2.0			109.0	$< 10^{-4}$
	(0.031)	(0.012)	(0.0052)	(0.0061)	(0.0012)	(0.45)				
$\hat{\theta}_{\text{Complete}}$	0.072	0.153	0.049	0.939	0.0081	-1.9	3.7	0.9964	13.1	0.0014
	(0.031)	(0.012)	(0.0044)	(0.0054)	(0.0012)	(0.56)	(1.2)	(0.0024)		

TABLE 2.12: Parameter estimation and associated standard errors for the synthetic time series in Figure 2.19 given by the three models described in subsection 2.4.1 with the Wilks statistics of the pairwise comparisons of the three models and the corresponding p-values. To present the results of the estimation, the same structure as Table 2.2 is used.

Parameter estimates are performed over 100 realizations, each with 10000 time steps, generated with the same underlying parameters $(\bar{r}, \bar{\sigma}, \alpha, \beta, \kappa, \bar{X}, \eta, a) =$ (0.07 (ann.), 0.15 (ann.), 0.05, 0.94, 0.008, -2.0, 4, 0.996). The mean value and the standard deviation (within parenthesis) of the parameters estimates over the 100 realizations are as follow

$$\begin{aligned}
\mathbb{E}\left[\hat{\bar{r}}\right] &= 0.073\,(0.030)\,\text{(ann.)}\,, & \mathbb{E}\left[\hat{\kappa}\right] &= 0.008\,(0.001)\,, \\
\mathbb{E}\left[\hat{\bar{\sigma}}\right] &= 0.152\,(0.014)\,\text{(ann.)}\,, & \mathbb{E}\left[\hat{\bar{X}}\right] &= -2.0\,(0.6)\,, \\
\mathbb{E}\left[\hat{\alpha}\right] &= 0.050\,(0.004)\,, & \mathbb{E}\left[\hat{\eta}\right] &= 4.6\,(2.2)\,, \\
\mathbb{E}\left[\hat{\beta}\right] &= 0.940\,(0.005)\,, & \mathbb{E}\left[\hat{a}\right] &= 0.9960\,(0.0030)\,.
\end{aligned} \tag{2.52}$$

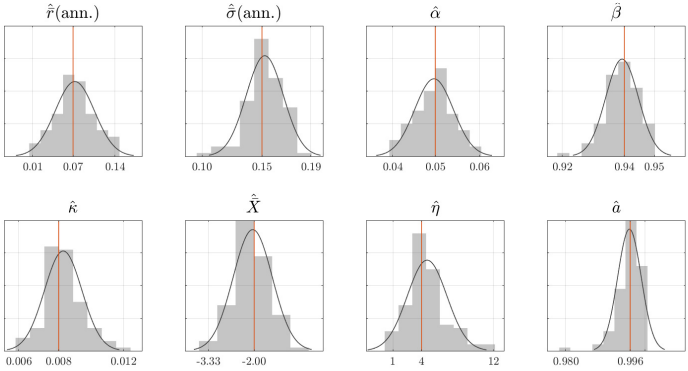

FIGURE 2.22: Distributions of the parameter estimates over 100 realizations, each with 10000 time steps, generated with the same parameters $(\bar{r}, \bar{\sigma}, \alpha, \beta, \kappa, \bar{X}, \eta, a) = (0.07$ (ann.), 0.15 (ann.), 0.05, 0.94, 0.008, -2.0, 4, 0.996). The underlying parameters are shown with the vertical red lines with values readable on the x-axis and fits of normal distributions are shown as black smooth curves.

2.B ADDITIONAL PARAMETER ESTIMATION FOR REAL DATA

Nikkei 225

The realized average return of the Nikkei 225 from 16 May 1949 to 31 December 2020 is 7.1% per annum, again lower than the estimate of the expected return $\bar{r}$, see Table 2.14. Similarly to the HSI case presented in the main body of the paper, the estimated expected return $\hat{\bar{r}}$ for the Nikkei time series is approximately equal to the average realized growth rate in earlier years, from 1950s to 1980s, but is found much larger than the average realized growth rate from 1990s onward, see Figure 2.23. The distributions of the estimates for the simulated time series shown in Figures 2.24 and 2.25 suggest that the parameters estimated from the Nikkei time series are consistent.

The transition of the price trajectory of the Nikkei 225 index is associated with the burst of the real-estate and financial bubbles in 1990 to 1991, which have been followed by the so-called 'Japan's lost decades' (of growth) (Funabashi and Kushner, 2015). We estimate the parameters separately for the two regimes divided at the date of 18 August 1992 where the index reached its historical low subsequent to the asset price bubble. The

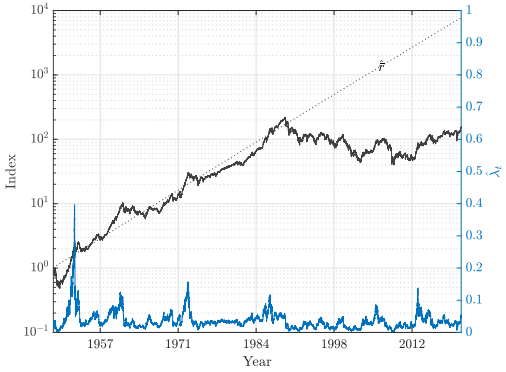

FIGURE 2.23: The Nikkei 225 from 16 May 1949 to 31 December 2020 in semi-logarithmic scale (left axis) compared with an exponential growth at the rate $\hat{\bar{r}}$ (ann.) = 12.6% (dotted line) and the estimated daily jump probability (right axis). The average annual realized return is 7.1%, the realized volatility is 19.5% and the theoretical volatility is 20.4%

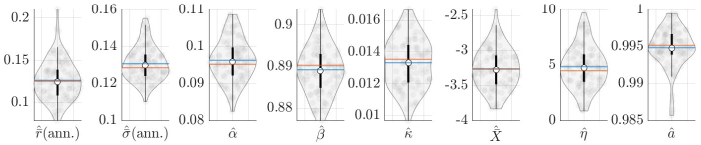

FIGURE 2.24: Violin plots of the parameter estimates over the 100 simulated time series generated with the estimated parameters of the Nikkei time series, with blue lines indicating the mean values of the distributions and red lines the estimates for the Nikkei time series.

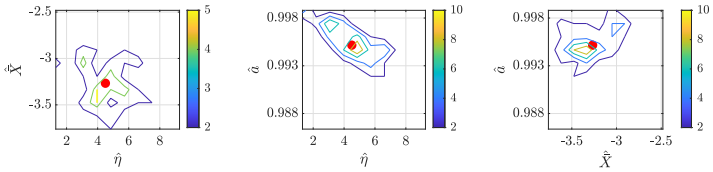

FIGURE 2.25: Isolines of bivariate distributions of the three parameters controlling the amplitude of the self-referencing propensity in the model over the 100 simulated time series generated with the estimated parameters of the Nikkei time series, with the red dots indicating the estimates for the Nikkei time series.

θ	$\bar{r}$ (ann.)	$\bar{\sigma}$ (ann.)	α	β	κ	$\bar{X}$	η	a	Wilks	p-value
$\hat{\theta}_{\text{GARCH}}$	0.172	0.311	0.140	0.854						
	(0.019)	(0.045)	(0.0063)	(0.0063)						
$\hat{\theta}_{\text{CI}}$	0.122	0.124	0.095	0.889	0.012	-3.0			957.4	$< 10^{-4}$
	(0.019)	(0.007)	(0.0055)	(0.0064)	(0.0019)	(0.15)				
$\hat{\theta}_{\text{Complete}}$	0.126	0.129	0.095	0.890	0.014	-3.3	4.5	0.9952	17.9	0.0001
	(0.019)	(0.008)	(0.0050)	(0.0053)	(0.0015)	(0.24)	(0.89)	(0.0010)		

TABLE 2.14: Parameter estimation and associated standard errors for the Nikkei daily time series with the Wilks statistics of the pairwise comparisons of the three models described in subsection 2.4.1 and the corresponding p-values.

parameter estimates for the two regimes are presented in Table 2.16 and the time series together with the jump probabilities are shown in Figure 2.26. The obtained expected returns of 14.2% and 9% are quite larger than the realized average yearly growth rate of 10.1% and 2.3% respectively. In the second period, the parameter estimation deems the jump probability to be constant ($\hat{\eta} = 0$). This means that the constant intensity model (2.47–2.49) calibrates the time series well enough. This can be interpreted as the model recognizing an absence of bubbles with positive feedbacks in the second period.

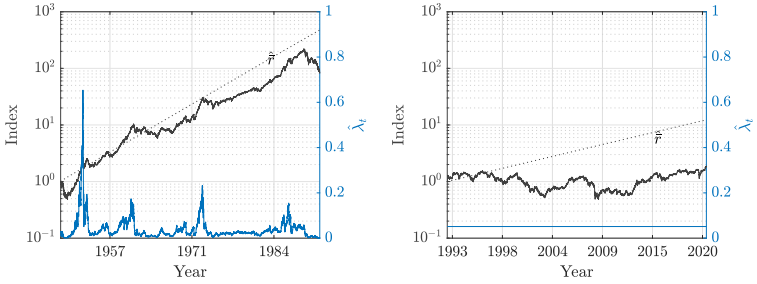

FIGURE 2.26: The Nikkei 225 from 16 May 1949 to 18 August 1992 (left panel, left axis) and from 19 August 1992 to 31 December 2020 (right panel, left axis) compared with an exponential growth at the rate $\hat{\bar{r}}$ (ann.) = 14.2% and 9% respectively (dotted lines) and the estimated daily jump probabilities (right axes). The average annual realized returns are respectively 10% and 2.3%.

θ	$\bar{r}$ (ann.)	$\bar{\sigma}$ (ann.)	α	β	κ	$\bar{X}$	η	a
$\hat{\theta}^{\text{1st}}_{\text{Complete}}$	0.142	0.114	0.121	0.846	0.014	-3.5	6.3	0.9958
(till 18 August 1992)	(0.022)	(0.005)	(0.0080)	(0.0095)	(0.0014)	(0.22)	(1.1)	(0.0010)
$\hat{\theta}^{\text{2nd}}_{\text{CI}}$	0.090	0.183	0.081	0.899	0.013	-2.9	-	-
(from 19 August 1992)	(0.039)	(0.010)	(0.0066)	(0.0082)	(0.0020)	(0.32)		

TABLE 2.16: Parameter estimation and associated standard errors for the Nikkei's two epochs

Given the estimates of the three parameters controlling the amplitude of the self-referencing propensity in the model over the whole time interval, we have $\delta_{ref} = (e^{-(1-\hat{a})\frac{\hat{X}}{\hat{\eta}}} | \hat{X} = -3.3, \hat{\eta} = 4.5, \hat{a} = 0.9952) = 1.0035$, hence the reference of the mispricing level is $\delta_{ref}^{250\times 0.83} - 1 \approx 1$ over 0.83 year, indicating that a jump probability of 0.5 per day requires a mispricing level of approximately 100% over ten months. This reference level is larger than that of the first period, so that the same price amplitudes in Figure 2.23 are associated with smaller jump probabilities compared to those in the first subplot of Figure 2.26. In the second regime, the jump probability remains constant at $L(\hat{\hat{X}}^{\text{2nd}}) = 5.2\%$. Although the estimated expected return is smaller than that for the whole period, the exponential growth at the rate $\hat{r}^{\text{2nd}}$ is still continuously above the real price shown in the second subplot of Figure 2.26, implying that the model deems the Nikkei seriously underpriced after 1992.

Nasdaq Composite

The Nasdaq Composite time series from 5 February 1971 to 31 December 2020 has an average realized return 9.7% per annum while the estimated expected return $\hat{\bar{r}}$ is 14.5%. The difference of these two returns is clearly visible in Figure 2.27 where the exponential growth at the rate $\hat{\bar{r}}$ remains above the real price, especially after 2000. However, from 1976 to 1998 and from 2009 to 2020, the Nasdaq index grows at a yearly rate approximately in line with the estimated expected return. As a result, the daily jump probabilities during these time intervals remain close to $L(\hat{\hat{X}}) = 17\%$ (recall that it is the growth rate of the price in comparison with the expected return that determines if a jump probability exceeds the average jump probability so that a drawdown is more likely to happen). The discrepancy between

the dotted line and the price trajectory can be mainly attributed to the 1973–1974 stock market crash and to the dot-com bubble drawdown and subsequent abnormal behavior until the great financial crisis of 2008.

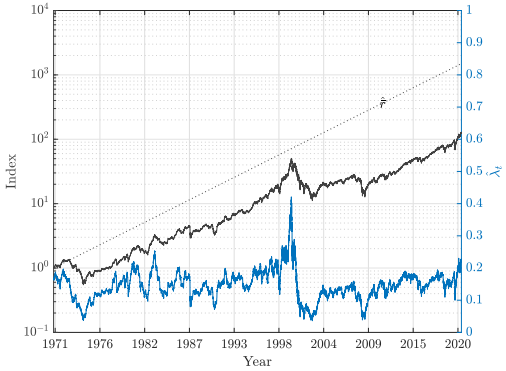

FIGURE 2.27: The Nasdaq Composite from 5 February 1971 to 31 December 2020 in semi-logarithmic scale (left axis) compared with an exponential growth at the rate $\hat{\bar{r}}$ (ann.)=14.5% (dotted line) and the estimated daily jump probability (right axis). The average annual realized return is 9.7%, the realized volatility is 19.2% and the theoretical volatility is 16.1%

θ	$\bar{r}$ (ann.)	$\bar{\sigma}$ (ann.)	α	β	κ	$\bar{X}$	η	a	Wilks	p-value
$\hat{\theta}_{\text{GARCH}}$	0.179	0.192	0.116	0.872						
	(0.021)	(0.015)	(0.0059)	(0.0061)						
$\hat{\theta}_{\text{CI}}$	0.148	0.065	0.089	0.890	0.007	-1.6			664.7	$< 10^{-4}$
	(0.022)	(0.007)	(0.0054)	(0.0060)	(0.0005)	(0.17)				
$\hat{\theta}_{\text{Complete}}$	0.145	0.066	0.089	0.892	0.007	-1.6	2.1	0.9977	5.7	0.0577
	(0.022)	(0.008)	(0.0055)	(0.0061)	(0.0007)	(0.23)	(0.60)	(0.0011)		

TABLE 2.18: Parameter estimation and associated standard errors for the Nasdaq daily time series with the Wilks statistics of the pairwise comparisons of the three models described in subsection 2.4.1 and the corresponding p-values.

We take 9 October 2002, when the index reached its historical low subsequent to the dot-com bubble, as the transition point and estimated the

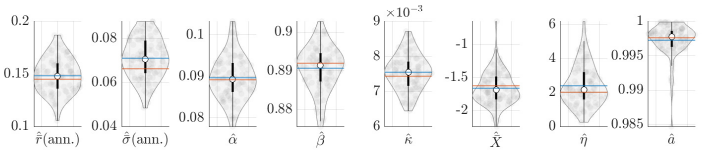

FIGURE 2.28: Violin plots of the parameter estimates over the 100 simulated time series generated with the estimated parameters of the Nasdaq time series, with blue lines indicating the mean values of the distributions and red lines the estimates for the Nasdaq time series.

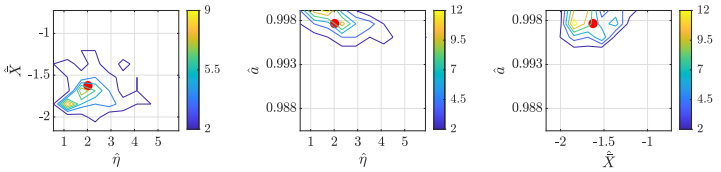

FIGURE 2.29: Isolines of bivariate distributions of the three parameters controlling the amplitude of the self-referencing propensity in the model over the 100 simulated time series generated with the estimated parameters of the Nasdaq time series, with the red dots indicating the estimates for the Nasdaq time series.

parameters for the pre- and post-transition regimes separately. The results are presented in Table 2.20 and the time series together with the jump probabilities are shown in Figure 2.30. In the first regime, the estimated expected return of 7.5% is equal to the average realized yearly return. In the second period, the model reduces to the constant intensity model with $\hat{\eta} = 0$ and the estimated expected return is 20.8%, much higher than the average realized yearly return 13.1%. In the first regime, the crash and drawdown ending in October 2002 can thus be interpreted as en efficient correction to the long term growth trend.

Given the estimates of the three parameters controlling the amplitude of the self-referencing propensity in the model over the whole time interval, we have $\delta_{ref} = (e^{-(1-\hat{a})\frac{\hat{X}}{\hat{\eta}}} | \hat{X} = -1.6, \hat{\eta} = 2.0, \hat{a} = 0.9977) = 1.0018$, hence the all-time reference of the mispricing level is $\delta_{ref}^{250\times 1.73} - 1 = 1.2$ over 1.73 years, indicating a jump probability of 0.5 per day requires a mispricing level of 120% over about 21 months (which corresponds to a 111% annual growth rate). This reference level is higher than that for the first regime. For this reason, and also because of a longer self-referencing memory, the

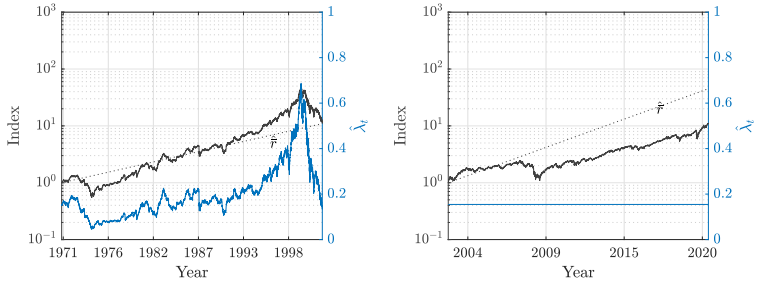

FIGURE 2.30: The Nasdaq Composite from 5 February 1971 to 9 October 2002 (left panel, left axis) and from 10 October 2002 to 31 December 2020 (right panel, left axis) compared with exponential growths at the rate $\hat{\bar{r}}$ (ann.) = 7.5% and 20.8% respectively (dotted lines) and the estimated daily jump probabilities (right axes). The average annual realized returns are respectively 7.5% and 13.1%.

θ	$\bar{r}$ (ann.)	$\bar{\sigma}$ (ann.)	α	β	κ	$\bar{X}$	η	a
$\hat{\theta}^{\text{1st}}_{\text{Complete}}$	0.075	0.064	0.090	0.883	0.007	-1.7	1.5	0.99990
(till 9 October 2002)	(0.041)	(0.008)	(0.0074)	(0.0110)	(0.0008)	(0.43)	(0.84)	(0.0007)
$\hat{\theta}^{\text{2nd}}_{\text{CI}}$	0.208	0.121	0.099	0.875	0.008	-1.7	-	-
(from 10 October 2002)	(0.042)	(0.009)	(0.0093)	(0.0102)	(0.0008)	(0.23)		

TABLE 2.20: Parameter estimation and associated standard errors for the two separate time intervals of the Nasdaq index.

jump probability estimated from the first regime of the index has a different behavior and a more pronounced peak for the dot-com bubble compared to what are given by the estimation for the entire sample.

Dow Jones Industrial Average

The Dow Jones Industrial Average time series from 14 July 1896 to 31 December 2020 with the estimated daily jump probability is shown in Figure 2.31. The discrepancy between the price trajectory with the average realized annual return 9.7% and an exponential growth at the rate of 14.5% has a pattern similar to that discussed for the S&P 500 time series.

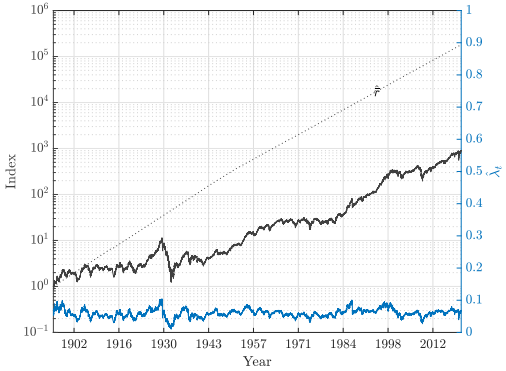

FIGURE 2.31: The Dow Jones Industrial time series from 14 July 1896 to 31 December 2020 in semi-logarithmic scale (left axis) with price growing at the rate $\hat{\bar{r}}$ (ann.) = 9% (dotted line) and the estimated daily jump probability (right axis). The average annual realized return is 5.0%, the realized volatility is 17.5% and the theoretical volatility is 16%.

θ	$\bar{r}$ (ann.)	$\bar{\sigma}$ (ann.)	α	β	κ	$\bar{X}$	η	a	Wilks	p-value
$\hat{\theta}_{\text{GARCH}}$	0.118	0.189	0.101	0.889						
	(0.012)	(0.010)	(0.0033)	(0.0036)						
$\hat{\theta}_{\text{CI}}$	0.087	0.095	0.073	0.910	0.010	-2.7			1619.1	$< 10^{-4}$
	(0.012)	(0.003)	(0.0029)	(0.0033)	(0.0005)	(0.10)				
$\hat{\theta}_{\text{Complete}}$	0.090	0.096	0.073	0.910	0.010	-2.7	1.3	0.9982	5.9	0.0512
	(0.024)	(0.010)	(0.0046)	(0.0037)	(0.0017)	(0.26)	(0.85)	(0.0018)		

TABLE 2.22: Parameter estimation and associated standard errors for the DJIA daily time series with the Wilks statistics of the pairwise comparisons of the three models described in subsection 2.4.1 and the corresponding p-values.

Due to a very high reference level of mispricing determined by the parameter estimation, the jump probability stays close to its average value $L(\hat{\bar{X}}) = 6\%$. Nevertheless, the Great Crash in 1929 as well as the crashes in 1987 and 2000 are characterized by small but visible peaks in the jump probability. Given the length of the time series spanning 125 years, with arguably many impactful geopolitical (two world wars), economic (great depression) and financial regimes (different interest rate regimes), we apply

the same rolling window approach as for the S&P 500 time series in the previous subsection. The results are shown in Figure 2.34. Two observations stand out. First, one can observe a secular tendency for the expected return to increase over time from a low value of 2.7% from 1902 to 1917, and the largest value of 14.7% observed for the last window from 1999 to 2020. A possible interpretation is the increasing expectations of investors of the positive impacts of the second, third and now fourth industrial revolutions in creating value that translated in expectations revealed by the Dow Jones index valuation. A complementary interpretation is the growing role of finance in funding the economy and in shaping expectations, as well as the increasing impact of central bank interventions arguably propping up the markets (Sornette and Cauwels, 2014). The second interesting observation is that, even in the time period covering the Great Depression or the Great Recession, the estimated expected return remains positive with a minimum of 6.7% during the Great Depression while the realized average return of the same time window is -0.8%.

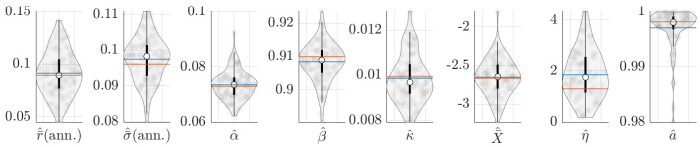

FIGURE 2.32: Violin plots of the parameter estimates over the 100 simulated time series generated with the estimated parameters of the DJIA time series, with blue lines indicating the mean values of the distributions and red lines the estimates for the DJIA time series.

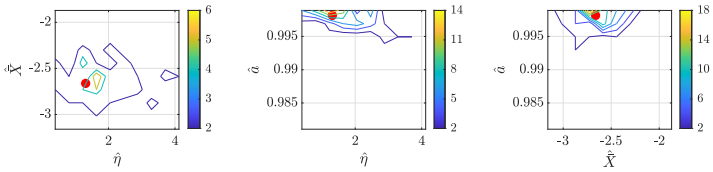

FIGURE 2.33: Isolines of bivariate distributions of the three parameters controlling the amplitude of the self-referencing propensity in the model over the 100 simulated time series generated with the estimated parameters of the DJIA time series, with the red dots indicating the estimates for the DJIA time series.

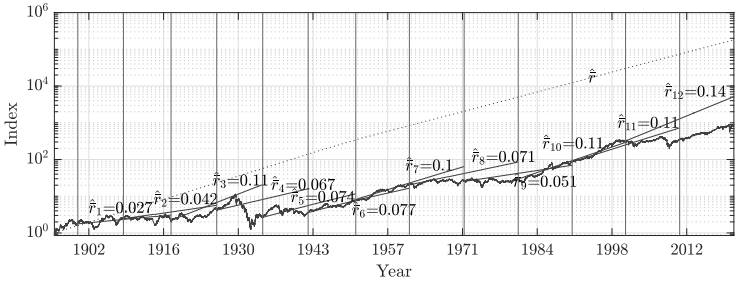

FIGURE 2.34: The DJIA time series with prices growing at the overall rate $\hat{\bar{r}}$ (ann.) = 9% (dotted line) and at the expected rate of each rolling window (solid lines). The rolling window size is 5000 time steps and the increment (between neighboring vertical lines) is 2500 time steps.

2.C STATIONARITY AND EXISTENCE OF THE MOMENTS

Following the proof in (Embrechts, Samorodnitsky, et al., 1998), we establish the existence of a unique stationary distribution for $Z_t = \left(r_t, \sigma_t^2, X_t\right)$ defined in model (2.16) and the existence of the second moment of r_t.

The process $Z_t = \left(r_t, \sigma_t^2, X_t\right)$ is a three-dimensional Markov chain with state space $E = \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}$ that can be expressed as

$$r_t = \kappa_t \cdot \lambda_t + \sigma_t \cdot \varepsilon_t - \kappa_t \cdot J_t \cdot I_t \,, \tag{2.53}$$

$$\sigma_t^2 = \omega + \alpha \cdot r_{t-1}^2 + \beta \cdot \sigma_{t-1}^2 \,, \tag{2.54}$$

$$X_t = -\frac{1-a}{s} \cdot \ln d + a \cdot X_{t-1} + \frac{1-a}{s} \cdot r_{t-1} \,, \tag{2.55}$$

where $\varepsilon_t \sim \mathcal{N}(0,1)$, $J_t \sim \mathcal{E}(1)$ and $\Pr\left[I_t = 1 | \mathcal{F}_{t-1}\right] = \lambda_t$, with $\lambda_t = L\left(X_t\right)$ constituting a continuous increasing mapping from $\mathbb{R}$ to $[0,1]$. We here make explicit that $\eta := \frac{1-a}{s}$ and $\bar{X} := -\frac{\ln \delta_{\text{ref}}}{s}$ (see expression (2.15)) and define $\ln d := -s\bar{X}$ (see expression (2.14)). Without loss of generality, we have assumed $\bar{r} = 0$.

The parameter space is such that $(\omega, \alpha, \beta, d, s) \in \mathbb{R}_+^5$, $a \in [0,1)$ and $\mathbb{E}\left[\kappa_t^2\right] < \infty$. The transition probabilities is denoted by $P(x, A)$, with $x \in E$ and A a measurable set.

Necessary condition

Let us first establish a necessary condition for the existence of a second moment of r_t assuming the existence of a stationary distribution. The second moment of r_t reads

$$\mathbb{E}\left[r_t^2\right] = \mathbb{E}\left[\left(\kappa_t \cdot \lambda_t + \sigma_t \cdot \varepsilon_t - \kappa_t \cdot J_t \cdot I_t\right)^2\right], \tag{2.56}$$

$$= \mathbb{E}\left[\sigma_t^2\right] + \mathbb{E}\left[\kappa_t^2 \cdot \left(2\lambda_t - \lambda_t^2\right)\right]. \tag{2.57}$$

The first expectation can be evaluated from the dynamic of the volatility:

$$\mathbb{E}\left[\sigma_t^2\right] = \omega + \alpha \cdot \mathbb{E}\left[r_{t-1}^2\right] + \beta \cdot \mathbb{E}\left[\sigma_{t-1}^2\right], \tag{2.58}$$

$$= \omega + (\alpha + \beta) \cdot \mathbb{E}\left[\sigma_{t-1}^2\right] + \alpha \cdot \mathbb{E}\left[\kappa_{t-1}^2 \cdot \left(2\lambda_{t-1} - \lambda_{t-1}^2\right)\right], \tag{2.59}$$

which yields

$$\mathbb{E}\left[\sigma_t^2\right] = \omega + (\alpha + \beta) \cdot \mathbb{E}\left[\sigma_t^2\right] + \alpha \cdot \mathbb{E}\left[\kappa_t^2 \cdot \left(2\lambda_t - \lambda_t^2\right)\right] \tag{2.60}$$

if the Markov chain is stationary, and

$$\mathbb{E}\left[\sigma_t^2\right] = \frac{\omega + \alpha \cdot \mathbb{E}\left[\kappa_t^2 \left(2\lambda_t - \lambda_t^2\right)\right]}{1 - \alpha - \beta}, \tag{2.61}$$

and

$$\mathbb{E}\left[r_t^2\right] = \frac{\omega + (1 - \beta) \cdot \mathbb{E}\left[\kappa_t^2 \left(2\lambda_t - \lambda_t^2\right)\right]}{1 - \alpha - \beta}, \tag{2.62}$$

provided that the *necessary condition* $\alpha + \beta < 1$ holds.

Sufficient condition

Taking the necessary condition $\alpha + \beta < 1$ for granted, let us now prove that it is sufficient for the existence of a stationary distribution of Z_t with a finite second moment for r_t.

Given the density of the noise variables $(\varepsilon_t, J_t, I_t)$, the Markov chain $\{Z_t, t \geq 0\}$ is irreducible. As a result of part (iii) of Theorem 6.0.1 in (Meyn and Tweedie, 1993), it is even an irreducible T-chain. Indeed, according to this theorem, the property of being an irreducible T-chain is equivalent to the weak Feller property, which is equivalent to the weak continuity property:

> The conditional distribution of Z_t, given $Z_{t-1} = y_k$, converges weakly to that of Z_t given $Z_{t-1} = y$ if $y_k \to y$.

Therefore, to prove that the Markov chain $\{Z_t\}$ is an irreducible T-chain, we just need to establish the weak continuity property, which follows from the continuous mapping theorem (Billingsley, 2008, Theorem 5.1, p. 30) because the dynamics of $\{Z_t\}$ is such that every new state is described by analytic functions of the previous state and the new random variables.

Now, given $\{Z_t\}$ is an irreducible T-chain, we rely on the following lemma from Embrechts, Samorodnitsky, et al. (1998):

i) Suppose that there exists a real-valued mapping $V \geq 0$ with $V \not\equiv 0$ and a compact set C such that

$$V(Z_t) - PV(Z_t) \geq 0, \quad \text{for} \quad Z_t \notin C, \tag{2.63}$$

where P expresses the expectation value at the following step of the Markov chain, i.e. $PV(x) = \mathbb{E}(V(Z_t)|Z_{t-1} = x) = \int_E V(y) P(x, dy)$. Then there exists an invariant measure and the Markov chain is Harris recurrent.

ii) Suppose that there exists a real-valued mapping $V \geq 0$ and $V \not\equiv 0$ and a compact set C such that

$$V(Z_t) - PV(Z_t) \geq -b \cdot 1_C(Z_t) + 1, \tag{2.64}$$

where 1_C is the indicator function of the set C (1 if in C, 0 if outside), and b is a finite positive number. Then there exists a unique stationary distribution π and $\{Z_t\}$ is positive Harris recurrent.

iii) Suppose that $\{Z_t\}$ is positive Harris recurrent and there exist nonnegative measurable functions f and V so that

$$V(Z_t) - PV(Z_t) \geq f(Z_t) - b, \tag{2.65}$$

with a finite $b \geq 0$. Then the π-expectation of f, $\pi f = \int f(x) \pi(dx)$, is finite, specifically: $\pi f < b < \infty$.

Hence, we have to find a mapping V that satisfies to previous conditions to establish our result. Given the necessary condition $\alpha + \beta < 1$ for the existence of the second moment of r_t, there exits a real u such that

$$\frac{\alpha}{1-\alpha} < u < \frac{1-\beta}{\beta}. \tag{2.66}$$

Applying the previous lemma to $\{Z_t\}$ with the mapping V:

$$\begin{aligned} \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R} &\to \mathbb{R}_+ , && (2.67) \\ (x, y, z) &\mapsto V(x, y, z) = u \cdot x^2 + y + L(z)^c , && (2.68) \end{aligned}$$

with $c \geq 0$, we have

$$\begin{aligned} & PV\left(r, \sigma^2, X\right) \\ =& \mathbb{E}\left[u \cdot r_t^2 + \sigma_t^2 + L(X_t)^c \middle| r_{t-1} = r, \sigma_{t-1}^2 = \sigma^2, X_{t-1} = X\right] , && (2.69) \\ =& (1+u) \cdot \left(\omega + \alpha \cdot r^2 + \beta \cdot \sigma^2\right) + L\left(-\frac{1-a}{s} \cdot \ln d + a \cdot X + \frac{1-a}{s} \cdot r\right)^c \\ & + \kappa^2 \cdot L\left(-\frac{1-a}{s} \cdot \ln d + a \cdot X + \frac{1-a}{s} \cdot r\right)^3 \\ & \times \left(2 - L\left(-\frac{1-a}{s} \cdot \ln d + a \cdot X + \frac{1-a}{s} \cdot r\right)\right) , && (2.70) \end{aligned}$$

Hence we get

$$\begin{aligned} & V\left(r_t, \sigma_t^2, X_t\right) - PV\left(r_t, \sigma_t^2, X_t\right) \\ =& (u - (1+u) \cdot \alpha) \cdot r_t^2 + (1 - (1+u) \cdot \beta) \cdot \sigma_t^2 + L(X_t)^c \\ & - (1+u) \cdot \omega - L\left(-\frac{1-a}{s} \cdot \ln d + a \cdot X + \frac{1-a}{s} \cdot r\right)^c \\ & - \kappa^2 \cdot L\left(-\frac{1-a}{s} \cdot \ln d + a \cdot X + \frac{1-a}{s} \cdot r\right)^3 \\ & \times \left(2 - L\left(-\frac{1-a}{s} \cdot \ln d + a \cdot X + \frac{1-a}{s} \cdot r\right)\right) . && (2.71) \end{aligned}$$

Given $(\alpha, \beta) \in [0,1]$, the coefficients before r_t^2 and σ_t^2 are positive so that the sum of the three terms in the first line is positive and can be made as large as desired. The sum of all the remaining terms is negative and its absolute value can be bounded by $(1+u) \cdot \omega + 1 + 2 \cdot \kappa$, since $L \in [0,1]$. Hence, V satisfies the points *(i)* and *(ii)* of the lemma. Now, defining

$$f(x, y, z) = \xi \cdot \left(x^2 + y\right) + L(z)^c , \tag{2.72}$$

with $0 < \xi < \min\{u - (1+u)\cdot\alpha, 1 - (1+u)\cdot\beta\}$, we satisfy the last point of the lemma and prove that the condition $\alpha + \beta < 1$ is sufficient for the existence of a stationary distribution of $\{Z_t\}$ that admits a finite second moment for r_t.

2.D EXPONENTIALLY MODIFIED GAUSSIAN DISTRIBUTION

The joined law of the triple $\left(r_t = \ln \frac{S_t}{S_{t-1}}, J_t, I_t\right)$ conditional on $\mathcal{F}_{t-1}$ is

$$(r_t, J_t, I_t) \overset{\text{law}}{=} \mathcal{N}\left(\mu_t - \kappa\cdot J_t\cdot I_t, \sigma_t^2\right)\cdot\mathcal{E}(1)\cdot\lambda_t^{I_t}\cdot(1-\lambda_t)^{1-I_t}\,, \tag{2.73}$$

with $\kappa \in \mathbb{R}$.

The marginal density of r_t conditional on $\mathcal{F}_{t-1}$ reads

$$g_t(x) = \sum_{I=0}^{1}\lambda_t^I\cdot(1-\lambda_t)^{1-I}\cdot\int_0^\infty dJ \frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-\frac{1}{2}\left(\frac{x-\mu_t+\kappa\cdot J\cdot I}{\sigma_t}\right)^2} e^{-J}\,, \tag{2.74}$$

$$= \lambda_t\cdot f_{\text{EMG}}(x\,|\mu_t,\sigma_t,\kappa) + (1-\lambda_t)\cdot\varphi(x\,|\mu_t,\sigma_t)\,, \tag{2.75}$$

where

$$\varphi(x\,|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(x-\mu)^2}{2\sigma^2}}\,, \tag{2.76}$$

and f_{EMG} is the exponentially modified Gaussian distribution with density

$$f_{\text{EMG}}(x\,|\mu,\sigma,\kappa) = \frac{1}{|\kappa|} e^{\frac{x-\mu}{\kappa}+\frac{\sigma^2}{2\kappa^2}}\cdot\left(1-\Phi\left(\text{sgn}(\kappa)\cdot\left(\frac{x-\mu}{\sigma}+\frac{\sigma}{\kappa}\right)\right)\right), \quad \kappa\neq 0\,, \tag{2.77}$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal law, that is

$$f_{\text{EMG}}(x\,|\mu,\sigma,\kappa) = \begin{cases} \frac{1}{\kappa} e^{\frac{x-\mu}{\kappa}+\frac{\sigma^2}{2\kappa^2}}\cdot\left(1-\Phi\left(\frac{x-\mu}{\sigma}+\frac{\sigma}{\kappa}\right)\right), & \kappa>0\,, \\ \frac{1}{|\kappa|} e^{\frac{x-\mu}{\kappa}+\frac{\sigma^2}{2\kappa^2}}\cdot\Phi\left(\frac{x-\mu}{\sigma}+\frac{\sigma}{\kappa}\right), & \kappa<0\,, \end{cases} \tag{2.78}$$

so that

$$f_{\text{EMG}}(x\,|\mu,\sigma,\kappa) = f_{\text{EMG}}(-x\,|-\mu,\sigma,\kappa)\,. \tag{2.79}$$

Of course when $\kappa = 0$, we have $f_{\text{EMG}}(x\,|\mu,\sigma,0) = \varphi(x\,|\mu,\sigma)$.

REFERENCES

Abreu, D. and Brunnermeier, M. (2003). "Bubbles and Crashes". In: *Econometrica* 71.1, 173.

Aitchison, J. and Shen, S. (1980). "Logistic-normal Distributions: Some Properties and Uses". In: *Biometrika* 67.2, 261.

Ardila-Alvarez, D., Forro, Z., and Sornette, D. (2021). "The acceleration effect and Gamma factor in asset pricing". In: *Physica A* 569, 125367.

Barndorff-Nielsen, O. (1983). "On a Formula for the Distribution of the Maximum Likelihood Estimator". In: *Biometrika* 70.2, 343.

Bekiros, Stelios D. (2010). "Fuzzy adaptive decision-making for boundedly rational traders in speculative stock markets". In: *European Journal of Operational Research* 202.1, 285.

Berntsen, P. (2015). "Particle Filter Adapted to Jump-diffusion Model of Bubbles and Crashes with Non-local Crash-hazard Rate Estimation". MA thesis. ETH Zurich.

Beyaztas, U. and Beyaztas, B. H. (2019). "On Jackknife-After-Bootstrap Method for Dependent Data". In: *Computational Economics* 53, 1613.

Billingsley, P. (2008). *Convergence of Probability Measures*. 2nd. Wiley-Interscience.

Black, F. (1976). "Studies of Stock Price Volatility Changes". In: *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economic Statistics Section*. Washington DC, 177.

Black, F. (1986b). "Noise". In: *Journal of Finance* 41, 529.

Blanc, P., Donier, J., and Bouchaud, J.-P. (2017). "Quadratic Hawkes processes for financial prices". In: *Quantitative Finance* 17.2, 171.

Camerer, C. (1989a). "BUBBLES AND FADS IN ASSET PRICES". In: *Journal of Economic Surveys* 3.1, 3.

Cameron, A. and Trivedi, P. (2005). *Microeconometrics: Methods and Applications*. Cambridge: Cambridge University Press.

Christie, A. (1982). "The Stochastic Behavior of Common Stock Variances: Value, Leverage and Interest Rate Effects". In: *Journal of Financial Economics* 10, 407.

Cumberland, W. G. and Sykes, Z. M. (1982). "Weak convergence of an autoregressive process used in modeling population growth". In: *Journal of Applied Probability* 19.2, 450.

Cvijovic, D. and Klinowski, J. (1995). "Taboo Search: An Approach to the Multiple Minima Problem". In: *Science* 267.5198, 664.

Demos, G. and Sornette, D. (2017). "Birth or burst of financial bubbles: which one is easier to diagnose?" In: *Quantitative Finance* 17.5, 657.

Embrechts, P., Samorodnitsky, G., Dacorogna, M., and Muller, U. (1998). "How Heavy are the Tails of a Stationary HARCH(k) Process? A Study of the Moments". In: *Stochastic Processes and Related Topics*. Ed. by Karatzas, I., Rajput, B., and Taqqu, M. Boston, MA: Birkhauser, 69.

Eraker, B., Johannes, M., and Polson, N. (2003). "The Impact of Jumps in Volatility and Returns". In: *The Journal of Finance* 58.3, 1269.

Euch, O. El, Gatheral, J., Radoičić, R., and Rosenbaum, M. (2020). "The Zumbach effect under rough Heston". In: *Quantitative Finance* 20.2, 235.

Filimonov, V., Demos, G., and Sornette, D. (2017). "Modified Profile Likelihood Inference and Interval Forecast of the Burst of Financial Bubbles". In: *Quantitative Finance* 17.8, 1167.

Flood, R. and Hodrick, R. (1990b). "On Testing for Speculative Bubbles". In: *The Journal of Economic Perspectives* 4.2, 85.

Franchetto, M. (2016). "Stochastic Online and Offline Calibration of a Jump-diffusion Model for Financial Bubbles and Crashes with Non-local Self-referencing Mispricing and an Application to Risk Measure Estimation". MA thesis. ETH Zurich.

Funabashi, Y. and Kushner, B. (2015). *Examining Japan's Lost Decades*. Milton Park, Abingdon, Oxfordshire: Routledge.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). "On the relation between the expected value and the volatility of the nominal excess return on stocks". In: *The Journal of Finance* 48.5, 1779.

Guyon, J. and Lekeufack, J. (2023). "Volatility is (mostly) path-dependent". In: *Quantitative Finance* 23.9, 1221.

Hobson, D. and Rogers, L. (1998). "Complete models with stochastic volatility". In: *Mathematical Finance* 8.1, 27.

Husler, A., Sornette, D., and Hommes, C. (2013). "Super-exponential Bubbles in Lab Experiments: Evidence for Anchoring Over-optimistic Expectations on Price". In: *Journal of Economic Behavior & Organization* 92, 304.

Jiang, Z., Zhou, W., Sornette, D., Woodard, R., Bastiaensen, K., and Cauwels, P. (2010). "Bubble diagnosis and prediction of the 2005-2007 and 2008-2009 Chinese stock market bubbles". In: *Journal of Economic Behavior & Organization* 74.3, 149.

Johansen, A. and Sornette, D. (1999). "Financial "anti-bubbles": Log-periodicity in gold and Nikkei collapses". In: *Int. J. Mod. Phys. C* 10.4, 563.

Johansen, A. and Sornette, D. (2001a). "Finite-time singularity in the dynamics of the world population and economic indices". In: *Physica A* 294.3-4, 465.

Johansen, A. and Sornette, D. (2001b). "Large Stock Market Price Drawdowns Are Outliers". In: *Journal of Risk* 4.2, 69.

Johansen, A. and Sornette, D. (2010a). "Shocks, Crashes and Bubbles in Financial Markets". In: *Brussels Economic Review (Cahiers economiques de Bruxelles)* 53.2, 201.

Johansen, A., Sornette, D., and Ledoit, O. (1999a). "Predicting Financial Crashes Using Discrete Scale Invariance". In: *Journal of Risk* 1.4, 5.

Kaizoji, T. and Sornette, D. (2010). "Bubbles and crashes". In: *Encyclopedia of Quantitative Finance*. Ed. by Cont, R. Wiley.

Lahiri, S. N. (2002). "On the jackknife-after-bootstrap method for dependent data and its consistency properties". In: *Econometric Theory* 18.1, 79.

Lin, L., Schatz, M., and Sornette, D. (2019). "A simple mechanism for financial bubbles: time-varying momentum horizon". In: *Quantitative Finance* 19.6, 937.

Ma, T., Fraser-Mackenzie, P.A.F., Sung, M., Kansara, A.P., and Johnson, J.E.V. (2021). "Are the least successful traders those most likely to exit the market? A survival analysis contribution to the efficient market debate". In: *European Journal of Operational Research*.

Malevergne, Y. and Sornette, D. (2014). "Jump-diffusion Model of Bubbles and Crashes with Non-local Behavioral Self-referencing". Unpublished.

Marsaglia, G. (2004). "Evaluating the Normal Distribution". In: *Journal of Statistical Software* 11.4, 1.

McFadden, D. (1974). "Conditional logit analysis of qualitative choice behavior". In: *Frontiers in Econometrics*. Ed. by Zarembka, P. Academic Press, 105.

McNeil, A., Frey, R., and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton, NJ: Princeton University Press.

Meese, R. (1986). "Testing for Bubbles in Exchange Markets: A Case of Sparkling Rates?" In: *Journal of Political Economy* 94.2, 345.

Meyn, S. and Tweedie, R. (1993). *Markov Chains and Stochastic Stability*. London: Springer-Verlag.

Minsky, H. (1986). *Stabilizing an Unstable Economy*. New Haven, CT: Yale University Press.

Nelder, J. and Mead, R. (1965). "A Simplex Method for Function Minimization". In: *The Computer Journal* 7.4, 308.

Nelson, D. B. (1991). "Conditional heteroskedasticity in asset returns: A new approach". In: *Econometrica* 59, 347.

Olvera-Cravioto, M. (2010). "On the Distribution of the Nearly Unstable AR(1) Process with Heavy Tails". In: *Advances in Applied Probability* 42.1, 106.

Osborne, M.F.M. (1959). "Brownian motion in the stock market". In: *Operations research* 7.2, 145.

Phillips, P.C.B., Shi, S-P., and Yu, J. (2015). "Testing for Multiple Bubbles: Historical Episodes of Exuberance and Collapse in the S&P 500". In: *International Economic Review* 56.4, 1043.

Phillips, P.C.B., Wu, Y., and Yu, J. (2011). "Explosive behavior in the 1990s Nasdaq: when did exuberance escalate asset values?" In: *International Economic Review* 52.1, 201.

Phillips, Peter C.B. and Yu, Jun (2011). "Dating the Timeline of Financial Bubbles During the Subprime Crisis". In: *Quantitative Economics* 2.3, 455.

Rabemananjara, R. and Zakoian, J. M. (1993). "Threshold ARCH models and asymmetries in volatility". In: *Journal of Applied Econometrics* 8, 31.

Shiller, R. J. (1981b). "Do stock prices move too much to be justified by subsequent changes in dividends?" In: *American Economic Review* 71, 421.

Shiller, R. J. (1992a). *Market Volatility*. MIT Press Books. Cambridge, MA: The MIT Press.

Shiller, R. J. (2000). *Irrational Exuberance*. Princeton, NJ: Princeton University Press.

Sornette, D. and Andersen, J. (2002). "A Nonlinear Super-Exponential Rational Model of Speculative Financial Bubbles". In: *Int. J. Mod. Phys. C* 13.2, 171.

Sornette, D. and Cauwels, P. (2014). "1980-2008: The Illusion of the Perpetual Money Machine and what it bodes for the future". In: *Risks* 2, 103.

Sornette, D. and Cauwels, P. (2015). "Financial bubbles: Mechanisms and diagnostics". In: *Review of Behavioral Economics* 2.3, 279.

Sornette, D., Cauwels, P., and Smilyanov, G. (2018). "Can We Use Volatility to Diagnose Financial Bubbles? Lessons from 40 Historical Bubbles". In: *Quantitative Finance and Economics* 2.1, 486.

Sornette, D. and Johansen, A. (1997). "Large financial crashes". In: *Physica A* 245.3, 411.

Sornette, D. and Malevergne, Y. (2001). "From rational bubbles to crashes". In: *Physica A: Statistical Mechanics and its Applications* 299.1, 40.

Sornette, D. and Woodard, R. (2010). "Financial Bubbles, Real Estate bubbles, Derivative Bubbles, and the Financial and Economic Crisis". In: *Econophysics Approaches to Large-Scale Business Data and Financial Crisis*. Ed. by Takayasu, M., Watanabe, T., and Takayasu, H. Springer Japan, 101.

Sornette, D., Woodard, R., Yan, W., and Zhou, W. (2013). "Clarifications to questions and criticisms on the Johansen-Ledoit-Sornette financial bubble model". In: *Physica A* 392.19, 4417.

Sornette, D. and Zhou, W. (2002). "The US 2000-2002 market descent: How much longer and deeper?" In: *Quantitative Finance* 2.6, 468.

Stiglitz, J. E. (1990). "Symposium on Bubbles". In: *Journal of Economic Perspectives* 4.2, 13.

Thaler, H. (1994). *Quasi Rational Economics*. New York, NY: Russell Sage Foundation.

Vogel, H. (2018). *Financial Market Bubbles and Crashes, Second Edition: Features, Causes, and Effects*. Cham: Palgrave Macmillan.

Waterfall, J., Casey, F., Gutenkunst, R., Brown, K., Myers, C., Brouwer, P., Elser, V., and Sethna, J. (2006). "Sloppy-model Universality Class and the Vandermonde Matrix". In: *Physical Review Letters* 97.15, 150601.

Xiong, W. (2013). *Bubbles, Crises, and Heterogeneous Beliefs*. Working paper no. 18905. NBER.

3

MODELING FINANCIAL CRASHES WITH SELF-EXCITING HAZARD RATE

This chapter presents an extended model of the previous model, which includes a crash hazard rate consisting of two components: the baseline intensity, which accounts for the exogenous shocks, and is determined by a function of mispricing level that reflects the propensity to anchor on past price levels; and the self-excitation, which accounts for the endogenous shocks. Adhering to the basic tenet of economic theory, rational expectations, our model expands previous frameworks by linking the crash hazard rate to a non-local self-referencing estimation of mispricing, and on correcting jumps. This linkage particularly encodes the tendency for stronger herding behaviors, both in frequency and magnitude, when prices are exuberant. We show that the model not only explains the mechanism of market crashes but also captures key stylized facts of financial time series including volatility clustering, leverage effects, fat-tailedness, long memory, without the need for traditional volatility models like GARCH. We further apply the model to major market indices and find that the distinctive performance trends of the indices are largely driven by underlying differences in their baseline intensities and degree of self-excitation, rather than by their underlying returns or diffusive volatilities.

3.1 INTRODUCTION

Traditional financial econometrics has predominantly treated market jumps as isolated events. Since Merton (1976) and Cont and Tankov (2004), Poission and Lévy processes have been extensively employed in option pricing, portfolio optimization, and beyond. However, their assumption of event independence prohibits them from accounting for the clustering of jumps identified in financial series. The introduction of Hawkes process (Hawkes (1971b) and Hawkes (1971a)) marks a shift in addressing this limitation. Originally developed for seismology, Hawkes process has enjoyed great popularity in finance econometrics in recent decades because it naturally accounts for the clustering phenomena by describing that the occurrence of an event increases the likelihood of subsequent ones. Across its broad appli-

cation in finance econometrics, Hawkes process is used to describe market events like trades, order flows, news releases, etc. (see e.g. Hawkes (2018) and Hawkes (2020) for reviews), and is particularly preferred for modelling in high-frequency environment (see e.g. Bacry, Mastromatteo, and Muzy (2015) for review) because the microstructure in high-frequency data is aligned with the event-based nature of Hawkes process. For coarser time scales, Hawkes process is jointly used with diffusion process (Aït-Sahalia, Cacho-Diaz, and Laeven (2015), Errais, Giesecke, and Goldberg (2010)) to additionally account for the integration of information over longer period.

Hawkes process is particularly useful for its structural distinction between the *exogenous* shocks such as market and geopolitical news, and *endogenous* interactions driven by internal processes, such as trader behavior and market sentiment that can lead to herding and feedback loops. Hawkes process defines a clear endogenous term understood in terms of a branching process, and allows for the classification of different types of events and the study of pertinent concepts like *reflexivity* (Soros (1988), Filimonov and Sornette (2012a)), which implies that perceptions of investors can influence market outcomes and in turn affect perceptions, creating a feedback loop that can deviate the asset price significantly from its fundamentals, and *excess volatility* (Shiller (1981a) and Shiller (1992b), Wyart and Bouchaud (2007), Wehrli and Sornette (2021)), which suggests that price fluctuations are far greater than can be explained by relevant information. This leads to the understanding that price dynamics are largely endogenous, driven by mechanisms within the market itself, which diverges from the Efficient Market Hypothesis.

Grounded in the studies of the endogenous and exogenous origins of financial crises (Sornette, Malevergne, and Muzy (2004), Sornette (2005), Johansen and Sornette (2010b)), we employ the model framework introduced by Malevergne, Sornette, and Wei (2023) and propose to link the crash hazard rate to an endogenous term. This framework combines the rational expectation theory and positive feedback loops by defining the crash hazard rate as a function of mispricing level. It extends the literature of rational bubble models, where the risk-return condition holds instantaneously, by describing the mispricing as non-local self-referencing. Despite the advances, the jump occurrence in this framework is governed by a Bernoulli process, meaning the jump clustering is a result of the built-up high crash hazard rate due to a deficit of correcting jumps, and each jump within the cluster is independently triggered by some level of mispricing. This independence restricts the model from capturing asymmetries in volatility

and jumps. The solution we propose in this paper is to add a self-exciting term to the crash hazard rate. As the scope of this paper focuses on daily time series, we conveniently employ the discrete-time Hawkes process, allowing the crash hazard rate to be expressed as the intensity of a Hawkes process

$$\lambda_t = \mu_t + \sum_{j=1}^{t-1} \phi(j\Delta) J_{t-j}.$$

The baseline intensity μ_t is characterized by a simple linear function of the non-local self-referencing mispricing. As mispricing level increases, the probability of immigrant jumps occurring increases, and the offspring jumps also start from a higher initial intensity, making the system more unstable. The higher probability of frequent and closely clustered jumps demands higher expected returns for risk compensation, boosting the bubble development until corrective jumps occur. The inclusion of the self-excitation of jump process enhances this positive feedback loop and attributes a significant portion of market movements to endogenous dynamics. This new framework demonstrates how internal market behaviors can scale up to produce critical phenomena, aligning with bubble models like LPPLS (Johansen, Ledoit, and Sornette (2000) and Johansen, Sornette, and Ledoit (1999b)), where price movements and speculative bubbles are described as consequences of market sentiments and positive feedback loops. Integrating this mispricing-jump interaction with Hawkes process allows our model to naturally capture the clustering of jumps and the endogenous dynamics of market volatility: the market instability and corrections are not only due to a high perceived risk, but also the internal market interactions.

More critically, the adoption of the Hawkes process enables the model to effectively replicate the effects traditionally achieved by GARCH-family models. Traditionally, models that incorporate independent jump processes rely on volatility processes like GARCH and its numerous variations to capture the volatility clustering and further characteristics such as fat-tailedness, leverage effect, long memory, etc. These models often fall short in considering the intricate dynamics of financial bubbles and crashes because these epochs cannot be simply viewed as epiphenomenon resulting from volatility properties or from isolated jumps, rather, they are largely due to the direct interaction of the jumps themselves, i.e. the endogenous effect. The self-exciting feature inherent in the Hawkes process provides a parsimonious and elegant solution for modeling volatility clustering, potentially eliminating the need for additional constructs to account for time-varying volatility.

While extensions such as the quadratic Hawkes (Blanc, Donier, and Bouchaud (2017)) process have also been developed to better simulate the stylized facts like fat-tailedness, we demonstrate that a well-calibrated simple Hawkes process is sufficient. By focusing on the core mechanisms of market interactions and the basic properties of Hawkes processes, our approach simplifies the modeling and calibration process without compromising the effectiveness in replicating real financial time series.

In order to obtain a sound model calibration and robust estimation of the parameters, we perform a Monte-Carlo Expectation-Maximization (MCEM) algorithm given the expected log-likelihood function does not have a closed form. The parameter estimation is then applied to synthetic time series and major market indices and complemented by an analysis of bias estimation.

The rest of the paper is organized as follows. Section 3.2 lays out the model with its properties. Section 3.3 demonstrates the stylized fact of model-generated data. Section 3.4 derives the expected log-likelihood for the MCEM algorithm. Section 3.5 presents the results of the parameter estimation for synthetic and empirical time series. Section 3.6 concludes and propose further extension of the model.

3.2 THE MODEL WITH SELF-EXCITED JUMPS

$$r_t = \gamma_t + \sigma_t \varepsilon_t - \kappa J_t$$

- $r_t = \log S_{t\Delta} - \log S_{(t-1)\Delta}$
- γ_t and σ_t are measurable with respect to the filtration $\{\mathcal{F}_t\}_{t=1,2,...,T/\Delta}$
- ε_t is an iid standard normal random variable
- J_t counts the number of jumps occurred in time step t

Assume we observe the price of a risky asset S over a long time interval $[0, T]$. We discretize the time interval so that the time period $t = 1, 2, ..., T/\Delta$ lasts from a continuous time $(t-1)\Delta$ to $t\Delta$. In this setting, the log-return $r_t = \log S_{t\Delta} - \log S_{(t-1)\Delta}$ of a risky asset follows the dynamics

$$r_t = \gamma_t + \sigma_t \varepsilon_t - \kappa J_t, \tag{3.1}$$

where γ_t and σ_t are measurable with respect to the filtration $\{\mathcal{F}_t\}_{t=1,2,...,T/\Delta}$ of the underlying probability space, and ε is an i.i.d. standard normal

random variable. The random variable J_t counts the number of jumps that occurred in the time step t and κ denotes their average jump size.

We can consider J_t to be the aggregate of an underlying point process N, which itself is defined as a random measure on $\mathbb{R}$, so that $N((a,b]) = N(a,b)$ counts the random number of points – or specifically the intra-period jumps – on the half open interval $(a,b]$. Thus we have $J_t = N((t-1)\Delta, t\Delta)$. If we assume the intra-period jumps to follow a generalized Hawkes point process, then the discrete conditional intensity

$$J_t = N((t-1)\Delta, t\Delta)$$

$$\lambda_t = \lim_{\Delta \to 0} \Delta^{-1} \mathbb{E}\left[J_t | \mathcal{F}^J_{t-1}\right] = \mu_t + \sum_{j=1}^{t-1} \eta_{t-j} h(j\Delta) J_{t-j} \tag{3.2}$$

with $\mathcal{F}^J_{t-1} = \sigma(J_k : k \leq t-1) \subset \mathcal{F}_t$ ensures (weak) convergence to the conditional intensity of N (Kirchner, 2016). The η_{t-j} are the fertilities, defining the expected number of jumps in some time step $t-j$ that is reproduced in all future time steps. The memory kernel $h_j := h(j\Delta)$ defines the probability that a jump is retained in the j-th time step and thus defines an *offspring probability mass function* with corresponding cumulative distribution $H_l := H(l\Delta) = \sum_{j \leq l} h_j$. The jump process is finally sub-critical if the branching ratio $\eta := \mathbb{E}[\eta_j] < 1$.

Using this specification, the conditional expected return is given by

$$\begin{aligned} \mathbb{E}\left[r_t | \mathcal{F}_{t-1}\right] &= \gamma_t - \kappa \mathbb{E}\left[J_t | \mathcal{F}_{t-1}\right] \\ &= \gamma_t - \kappa \lambda_t \end{aligned} \tag{3.3}$$

and when constrained to be constant, $\mathbb{E}[r_t | \mathcal{F}_{t-1}] = \bar{r}$, we recover

$$r_t = \bar{r} + \sigma_t \varepsilon_t - \kappa \left(J_t - \mathbb{E}\left[J_t | \mathcal{F}_{t-1}\right]\right). \tag{3.4}$$

As in Malevergne, Sornette, and Wei (2023), we want the conditional jump intensity to depend on the mispricing of the asset, i.e. its deviation from growth at rate $\bar{r}$. We can achieve this by having the background (or exogenous) jump intensity μ_t, defined in (3.2), depend on the ratio

$$\delta_{t,\tau} = \frac{S_t}{S_{t-\tau} \cdot e^{\bar{r} \cdot \tau}}, \tag{3.5}$$

whose deviation from 1 quantifies the departure of the current price S_t from what would have been the asset price had it grown at the expected

rate $\bar{r}$ during the period under consideration. The price $S_{t-\tau} \cdot e^{\bar{r} \cdot \tau}$ can be seen as the *anchoring price* at time t. The difference $\delta_{t,\tau} - 1$ quantifies the over/under-pricing with respect to the expected long-term trend.

If we take the perspective of *anchoring return*, expression (3.5) can be further rewritten as

$$\frac{1}{\tau} \ln \delta_{t,\tau} = \frac{1}{\tau} \sum_{k=1}^{\tau} \ln \frac{S_{t-\tau+k}}{S_{t-\tau+k-1}} - \bar{r} = \frac{1}{\tau} \sum_{k=1}^{\tau} \left(r_{t-\tau+k} - \bar{r} \right) , \tag{3.6}$$

which defines the average daily excess return over the long-term growth rate $\bar{r}$. For estimation purpose, it is natural to introduce a better-behaved smooth version of expression (3.6) as an exponential moving average of the returns over $\tau \simeq 1/(1-a)$ time steps,

$$\ln \delta_{t,a} = (1-a) \left(r_{t-1} - \bar{r} \right) + a \cdot \ln \delta_{t-1,a} , \tag{3.7}$$

whose solution reads

$$\ln \delta_{t,a} = (1-a) \sum_{k=0}^{\infty} a^k \left(r_{t-k} - \bar{r} \right) , \tag{3.8}$$

which is the exponentially smoothed version of (3.6) and thus

$$\ln \delta_{t,a} \approx \frac{1}{\tau} \ln \delta_{t,\tau}, \quad \text{with} \;\; a \approx 1 - \frac{1}{\tau} . \tag{3.9}$$

Subsequently, we can formulate the dependence between the level of mispricing and the probability of jump occurrence using a monotonically increasing function. For computational efficiency, we consider a piecewise linear function [1]

$$\mu_t = \max(0, \nu \ln \delta + \bar{\mu}), \tag{3.10}$$

where ν measures the sensitivity of jump intensity to changes in mispricing and $\bar{\mu}$ adjust the base rate allowing for the existence of jumps even when there is no mispricing.

In this setup, *exogenous* jumps due to mispricing at time t occur as $I_t \sim \text{Pois}(\mu_t) = \text{Pois}(\bar{\mu} F(\delta_{t-1,\tau}))$ and then we have additional *endoge-*

1 the internal dynamics and interactions of past jumps play a much more significant role in determining the jump intensity than the external factor of mispricing. Diminished Sensitivity to Mispricing: When the self-exciting component overwhelmingly influences the jump intensity, the specific choice of function connecting mispricing to the external jump intensity μ might have a less critical impact on the model's overall behavior. This is because the model's dynamics are primarily driven by the history of jumps rather than the current level of mispricing.

nous jumps occurring as $O_t \sim \text{Pois}(\sum_{j=1}^{t-1} \phi_{t-j} J_j)$ with $\phi_j := \eta_{t-j} h(j\Delta)$ and $J_t = I_t + O_t$. For the jump offspring probability h, we choose a geometric probability mass function $h_j = (1-p)^{j-1}p$, which corresponds to exponentially decreasing memory in discrete times.

3.2.1 *Model summary*

Combining the equations, we obtain the full model as

$$r_t = \bar{r} + \sigma\varepsilon_t - \kappa(J_t - \lambda_t) \tag{3.11}$$

$$J_t|\mathcal{F}_{t-1} \sim \text{Pois}(\lambda_t) \tag{3.12}$$

$$\lambda_t = \mu_t + \eta \sum_{j=1}^{t-1} h_j J_{t-j} \tag{3.13}$$

$$\text{where} \quad \mu_t = \max(0, \nu \ln \delta_{t,a} + \bar{\mu}) \tag{3.14}$$

$$\ln \delta_{t,a} = (1-a)(r_{t-1} - \bar{r}) + a \ln \delta_{t-1,a} \tag{3.15}$$

$$\text{and} \quad h_j = (1-p)^{j-1}p \tag{3.16}$$

with parameter set $\theta = (\bar{r}, \sigma, \kappa, a, \nu, \bar{\mu}, \eta, p)$, where $\sigma, \kappa, \bar{\mu} > 0$ and $0 < a, \eta, p < 1$ to ensure the practical applicability and stationarity of the model. The discrete time setting aggregates the effects of multiple jumps occurring within a single time interval, with λ_t representing the jump intensity at time t and κ the average size of the aggregated jumps. A constant κ assumes that the average impact of the aggregated jumps is stable over time while λ_t scales this impact according to the expected number of jumps. The paper mainly focuses on scenarios when κ is positive, that is, when jumps are negative and capture the dynamics of market crashes and drawdowns. Although positive jumps are not explicitly included, the positive feedback loops created by the buildup of perceived risks are capable of pushing the bubble formation and exhibiting faster-than-exponential growth.

For parsimonious purpose, σ and η, p are also held constant, assuming that the diffusive volatility does not react to market dynamics, and that the influence of past jumps on the current condition is uniform. We show in the next sections that a time-varying baseline intensity that is responsive to the mispricing level is sufficient to account for dynamics of jumps and volatility observed in financial time series, provided the branching ratio η is close to 1.

3.2.2 *Moments of returns and stationarity*

The unconditional expected return is given by

$$\mathbb{E}[r_t] = \mathbb{E}[\mathbb{E}[r_t|\mathcal{F}_{t-1}]] = \bar{r}, \tag{3.17}$$

and since J_t at each time point is a Poisson random variable with mean λ_t, the unconditional variance reads

$$\begin{aligned}\mathrm{Var}[r_t] =& \mathbb{E}\left[\sigma^2\right] + \mathbb{E}\left[\kappa^2 (J_t - \lambda_t)^2\right]. && (3.18)\\ =& \sigma^2 + \kappa^2 \mathbb{E}\left[\mathbb{E}\left[(J_t - \lambda_t)^2 |\mathcal{F}_{t-1}\right]\right] && (3.19)\\ =& \sigma^2 + \kappa^2 \mathbb{E}[\lambda_t]. && (3.20)\end{aligned}$$

The expectation of intensity, given its expression (3.13), reads

$$\begin{aligned}\mathbb{E}[\lambda_t] =& \tilde{\bar{\mu}} + \eta \mathbb{E}\left[\sum_{j=1}^{t-1} h_j J_{t-j}\right] && (3.21)\\ =& \tilde{\bar{\mu}} + \eta \sum_{j=1}^{t-1} h_j \mathbb{E}\left[\lambda_{t-j}\right], && (3.22)\end{aligned}$$

where $\mathbb{E}[\max(0, \nu \ln \delta_{t,a} + \bar{\mu})]$ is denoted as $\tilde{\bar{\mu}}$, and the solution reads

$$\mathbb{E}[\lambda_t] = \frac{\tilde{\bar{\mu}}}{1-\eta}, \tag{3.23}$$

which corresponds to the classic results of the first moment of a classic Hawkes process.

Given $\mathbb{E}[\ln \delta_{t,a}] = 0$ provided by (3.14), a parameterization regarding ν and $\bar{\mu}$ can be chosen such that $\tilde{\bar{\mu}}$ is equal to or can be approximated by $\bar{\mu}$. Under this circumstance, the unconditional variance of returns (3.20) can be expressed as

$$\mathrm{Var}[r_t] = \sigma^2 + \frac{\kappa^2 \bar{\mu}}{1-\eta}. \tag{3.24}$$

Given the Efficient Market Hypothesis states that markets efficiently reflect all available information in asset prices almost instantaneously, so price changes are just reflecting exogenous news. Therefore, price formation must be efficient enough, so that no endogenous processes should be present in any observation. Such a criterion for market efficiency is translated into

the requirement of the branching ratio $\eta = 0$ in the framework of Hawkes processes. Within the framework of our model, the excess volatility can be therefore recognized as approximately

$$\frac{\eta}{1-\eta} \cdot \kappa^2 \bar{\mu}. \tag{3.25}$$

Next, we derive the skewness of the returns using the results of Poisson central moments,

$$\gamma_1 = \frac{\mathbb{E}\left[(r_t - \bar{r})^3\right]}{\left(\mathbb{E}\left[(r_t - \bar{r})^2\right]\right)^{3/2}} = \frac{-\kappa^3 \mathbb{E}\left[(J_t - \lambda_t)^3\right]}{\left(\sigma^2 + \kappa^2 \mathbb{E}\left[(J_t - \lambda_t)^2\right]\right)^{3/2}} = -\frac{\mathbb{E}\left[\lambda_t\right]}{\left(\frac{\sigma^2}{\kappa^2} + \mathbb{E}\left[\lambda_t\right]\right)^{3/2}}, \tag{3.26}$$

and given (3.23), the skewness can be expressed by

$$\gamma_1 = -\frac{\frac{\bar{\mu}}{1-\eta}}{\left(\frac{\sigma^2}{\kappa^2} + \frac{\bar{\mu}}{1-\eta}\right)^{3/2}}. \tag{3.27}$$

We observe intriguing behavior here as (3.27) indicates that the skewness decreases when η increases from 0 and reaches its minimum at $\eta* = 1 - \frac{\bar{\mu}\kappa^2}{2\sigma^2}$, then increases (decreases in magnitude) and finally vanishes as η approaches 1. Initially, a larger η emphasizes negative skewness, but the strong clustering and strong self-excitation at criticality will eventually remove the skewness. The existence of a minimum at $\eta*$ can be explained by the competition between the standard volatility and the jumps.

We further derive the kurtosis of returns

$$\gamma_2 = \frac{\mathbb{E}\left[(r_t - \bar{r})^4\right]}{\left(\mathbb{E}\left[(r_t - \bar{r})^2\right]\right)^2} = \frac{3\sigma^4 + (6\sigma^2\kappa^2 + \kappa^4)\mathbb{E}\left[\lambda_t\right] + 3\kappa^4 \mathbb{E}\left[\lambda_t^2\right]}{\left(\sigma^2 + \kappa^2 \mathbb{E}\left[\lambda_t\right]\right)^2}. \tag{3.28}$$

To derive $\mathbb{E}\left[\lambda_t^2\right]$, we first recall Kirchner (2016) Proposition 5, which derives the second moment of INAR(∞) sequence. Let $R(j) := \mathrm{Cov}(J_t, J_{t+j})$, we have

$$R(j) = \frac{\mu}{1-\eta} \sum_{k=0}^{\infty} \beta_k \beta_{k+|j|}, \tag{3.29}$$

where $\beta_0 := 1$, $\beta_k := \eta \sum_{i=1}^{k} h_i \beta_{k-i}$.

Given this, we can calculate the second moment of Hawkes intensity

$$\mathbb{E}\left[\lambda_t^2\right] = \mathbb{E}\left[\mu_t^2 + 2\mu\eta\sum_{j=1}^{t-1} h_j J_{t-j} + \eta^2\left(\sum_{j=1}^{t-1} h_j J_{t-j}\right)^2\right] \tag{3.30}$$

$$= \mathbb{E}\left[\mu_t^2\right] + 2\mu\eta\sum_{j=1}^{t-1} h_j \mathbb{E}\left[\lambda_t\right] + \eta^2\left(\sum_{j=1}^{t-1} h_j^2 R(0) + 2\sum_{k<j} h_j h_k R(j-k) + \left(\sum_{j=1}^{t-1} h_j\right)^2 \mathbb{E}\left[\lambda_t\right]^2\right). \tag{3.31}$$

In fact, given the conclusion in Kirchner (2016) that INAR(∞) is (strictly) stationary and exists unique solution, the same properties can propagate to r_t given the return equation linearly combines a stationary EWMA and an INAR(∞), suppose the model parameters are well-specified.

3.3 STYLIZED PROPERTIES OF SYNTHETIC DATA

3.3.1 *Synthetic example*

The requirement of stationary and the results of the moments, in particular (3.24) and (3.27) have imposed certain limitations in the parameterization. As a visual illustration of the model properties, Fig. 3.1 show a typical synthetic price trajectory and its returns over 40 years (10000 trading days), generated with the following parameterization:

- $\bar{r} = 7\%$ per annum, $\sigma = 8\%$ per annum
- $J_t|\mathcal{F}_{t-1} \sim \text{Pois}(\lambda_t)$, $\kappa = 1\%$
- $\bar{\mu} = 0.1$, $\nu = 2$, $a = 0.99$
- $\eta = 0.93$, $p = 0.4$.
- long-term expected rate of return $\bar{r} = 7\%$ per annum, diffusive volatility $\sigma = 8\%$ per annum;
- $J_t|\mathcal{F}_{t-1} \sim \text{Pois}(\lambda_t)$ follows Poisson distribution and the scale factor of jump components $\kappa = 1\%$;

- parameters of the baseline intensity μ_t are $\bar{\mu} = 0.1$, $\nu = 2$; and the memory for the EWMA is $a = 0.99$ about 100 trading days;
- parameters of the self-exciting part are branching ratio $\eta = 0.93$ and memory decay $p = 0.4$.

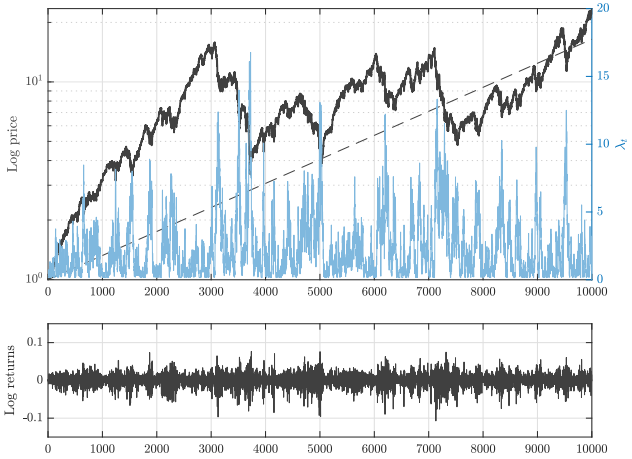

FIGURE 3.1: Synthetic log-price generated with parameters $(\bar{r}, \sigma, \kappa, a, \nu, \bar{\mu}, \eta, p) = (0.07(ann.), 0.08(ann.), 0.01, 0.99, 2, 0.1, 0.93, 0.4)$ (upper panel, left axis) with the price growing at rate $\bar{r}$ over long term shown in straight line, jump probability (upper panel, right axis) and log-returns (lower panel).

3.3.2 *Stylized properties for market profiles*

In order to inform the basic adequacy of our model to describe properties of actual bubbles and crashes, we first look at some empirical examples. To identify where bubbles occurred in a price series, we will follow the simple approach e.g. also used in Westphal and Sornette (2020) and define the end of a bubble (or more generally, a drawup) as a peak in the time series.

A peak occurs at time-step t_i if

$$S_{t_i} \geq S_{t_j} \qquad \forall t_j \in [t_i - \tau^d, t_i + \tau^d] \tag{3.32}$$

$$\tau^d = 63$$

for some time scale τ^d, defining the minimum distance between two peaks. In the following, we will choose $\tau^d = 63$, corresponding approximately one quarter of a calendar year. This time scale seems to satisfactorily identify some well-known bubbles like the dot-com bubble, or the one ending in October 1987, without additionally adding too many small/irrelevant drawdowns (except maybe for the case of the Dow Jones Industrial Index). The trough of the crash following the end of the bubble is then found at the time at which the price is minimal between two consecutive peaks.

We present a summary of the resulting drawups and drawdowns characteristics with the moments of returns for major indices in Table 3.2 and for sample data in Table 3.4. The resulting start and end times of the drawups are visualized in Figure 3.7 in the Appendix. The analysis of the sample data reveals a close match with the stylized properties of empirical markets.

	$\langle r \rangle$ (ann.)	std(r) (ann.)	skew(r)	kurt(r)	N	$\langle B \rangle$/Max.	$\langle C \rangle$/Max.	$\langle \tau_B \rangle$	$\langle \tau_C \rangle$	slope ratio
S&P	0.087	0.182	-1.21	28.4	40	31%/114%	15%/53%	0.74	0.25	2.5
Nasdaq	0.105	0.221	-0.33	11.2	40	44%/256%	20%/62%	0.68	0.29	2.1
Nikkei	0.037	0.222	-0.30	10.7	48	31%/88%	20%/51%	0.51	0.33	2.2
HSI	0.051	0.259	-2.12	56.5	42	42%/145%	24%/58%	0.52	0.33	2.0
DAX	0.078	0.219	-0.30	9.81	39	37%/111%	19%/60%	0.63	0.30	2.2
DJIA	0.077	0.175	-0.41	15.9	36	27%/98%	14%/50%	0.66	0.25	2.4

TABLE 3.2: Overview of the moments of returns and price dynamics for stock indices, obtained with $\tau^d = 63$ trading days. $\langle r \rangle$ the average annual return, std(r) its standard deviation, skew(r) the skewness and kurt(r) the kurtosis. N denotes the number of drawups in the sample, $\langle B \rangle$ their average size, accompanied by the maximum size, and $\langle \tau_B \rangle$ the corresponding average duration. $\langle C \rangle$ and $\langle \tau_C \rangle$ denote the same quantities for the drawdowns, slope ratio is the average ratio of the slope of a drawup to that of the corresponding drawdown. All durations are expressed in years. See Figure 3.7 for a visualization of the drawups start and end times.

	$\langle r \rangle$ (ann.)	std(r) (ann.)	skew(r)	kurt(r)	N	$\langle B \rangle / \text{Max.}$	$\langle C \rangle / \text{Max.}$	$\langle \tau_B \rangle$	$\langle \tau_C \rangle$	slope ratio
Sample	0.073	0.205	-0.67	10.5	34	$45\%/_{214\%}$	$22\%/_{62\%}$	0.88	0.27	2.9

TABLE 3.4: Overview of the moments of returns and price dynamics for the sample time series as seen in Fig. 3.1. Notations are as defined in Table 3.2.

3.3.3 *Stylized properties for returns*

Volatility clustering and long-range dependence

In analyzing the synthetic sample time series, we first demonstrate volatility clustering and long-range dependence by plotting the autocorrelation of log-returns and squared log-returns in Fig. 3.2. Both autocorrelation functions exhibit patterns consistent with those observed in empirical data.

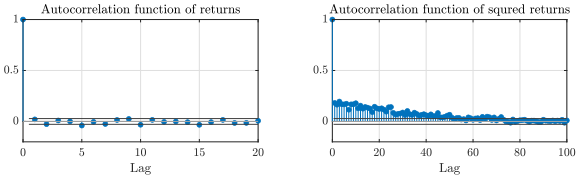

FIGURE 3.2: Autocorrelation functions of log-returns (left panel) and squared log-returns (right panel) for the sample data as in Fig.3.1.

Leverage effect

The leverage effect (Black, 1976; Christie, 1982) is an acknowledged feature of financial time series, referring to the negative correlation between asset returns and changes in volatility. Figure 3.3 demonstrates the leverage effect of the sample time series, which has similar pattern and greater intensity compared to empirical time series.

Given the presence of leverage effect, we further study the level volatility persistency employing (Glosten-Jagannathan-Runkle) GJR model, which is a GARCH variant that accounts for leverage effect, with the following method:

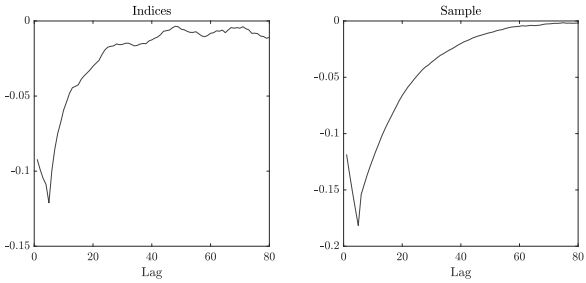

FIGURE 3.3: Cross correlation between log-returns and squared log-returns, averaged over indices S&P500, Nasdaq, Nikkei, HSI, DAX, and DJIA(left panel); and of sample data as in Fig.3.1 (right panel).

First, conduct Engle's ARCH test. The number of lagged terms to include in the test statistic calculation is determined by fitting ARCH(q) models for q = 1, ..., 50 and selecting the best fitting model q* using AIC. If the p-value for the null of no ARCH effects from these tests indicates rejection of the null, then the bubble model produces volatility clustering.

Second, perform a model selection test using the GJR model. In the GJR formulation, large negative changes are more likely to be clustered than positive changes. Since the GARCH model is nested in the GJR model (GJR reduces to GARCH if all leverage coefficients are zero), we can test the GARCH model against a GJR alternative using a likelihood ratio test. The procedure is as follows: since GARCH(p,q) processes are locally equivalent to ARCH(p + q) processes, consider all the possible GARCH(p,q) models with p + q = q* and then select the best fitting GARCH model using AIC. Subsequently, fit a GJR model with the same p,q like the AIC-optimal GARCH model and then perform a likelihood ratio test for the null of the GARCH model. If the p-value from this likelihood ratio test indicates rejection of the null, then the bubble model produces asymmetric volatility clustering.

We use the sum of the GARCH and ARCH coefficients and the leverage parameter to indicate the volatility persistency. The results obtained from the AIC optimal GARCH models fitted to the synthetic time series with different parameterizations are shown in Fig.3.4. Given where the 0.95 level is placed on the isoline maps, the volatility persistency remains close to one as in empirical data only when η is also close to 1.

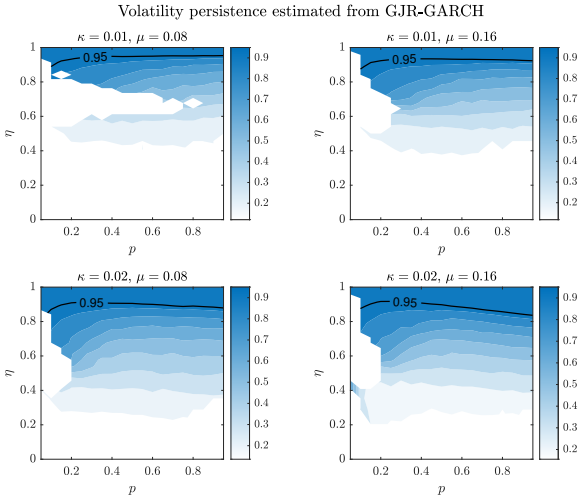

FIGURE 3.4: Isoline map of the volatility persistency, calculated as the sum of the GARCH and ARCH coefficients and the leverage parameter obtained from the AIC optimal GARCH models fitted to the synthetic time series. The colored areas indicates rejection of the null (the GARCH model without leverage terms) at the 5% level. The values shown are averages over 100 realizations.

Zumbach's effects

The model is also consistent with the Zumbach's effect (Blanc, Donier, and Bouchaud, 2017), *i.e.* the fact that past squared returns predict future volatilities better than past volatilities predict future squared returns. This effect is more nuanced and often neglected in time series models. To illustrate the effect, let us define the cross-correlation

$$\rho^{(2)}(\tau) = \frac{\mathrm{Cov}\left(\sigma_t^2, (r_{t-\tau} - \bar{r})^2\right)}{\sqrt{\mathrm{Var}\left(\sigma_t^2\right)\mathrm{Var}\left((r_t - \bar{r})^2\right)}}, \tag{3.33}$$

as in Euch et al. (2020), and its integrated difference

$$\Delta(\tau) = \sum_{i=1}^{\tau} \left(\rho^{(2)}(i) - \rho^{(2)}(-i)\right). \tag{3.34}$$

Figure 3.5 illustrates the Zumbach's effect in empirical and synthetic sample time series, presented by the integrated difference (3.34). The consistent patters further indicate the relevance of the model.

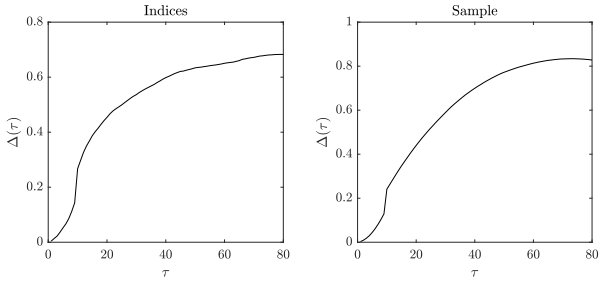

FIGURE 3.5: Integrated difference $\Delta(\tau)$, averaged over indices S&P500, Nasdaq, Nikkei, HSI, DAX, and DJIA(left panel); and of sample data as in Fig.3.1. (right panel)

Fat-tailedness

We demonstrate the fat-tail characteristic of synthetic data by plotting complementary cumulative distribution function (CCDF) of log returns on log-log scale and compare it to two empirical data. Fig.3.6 show how well our model is capable of generating fat tails.

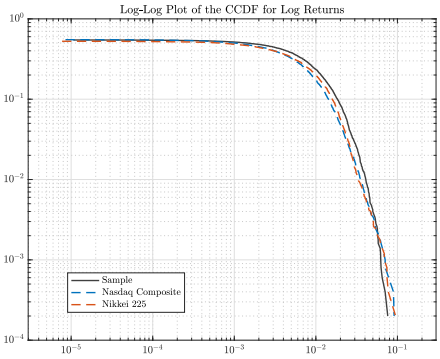

FIGURE 3.6: Tail behavior

3.4 MCEM ALGORITHM FOR MODEL CALIBRATION

In general, an expectation-maximization (EM) algorithm is an iterative method to find maximum likelihood or maximum a posteriori estimates of parameters in statistical models, where the model depends on unobserved latent variables.

Given the statistical model which generates a set X of observed data, a set of unobserved latent data or missing values Z, and a vector of unknown parameters θ, along with a likelihood function $L(\theta; X, Z) = p(X, Z|\theta)$, the maximum likelihood estimate (MLE) of the unknown parameters is determined by maximizing the marginal likelihood of the observed data

$$L(\theta; X) = \int p(X, Z|\theta) dZ = \int p(X|Z, \theta) p(Z|\theta) dZ, \tag{3.35}$$

This quantity is often intractable since Z is unobserved and the distribution of Z is unknown before attaining θ. The EM algorithm seeks to find the MLE of the marginal likelihood by iteratively applying these two steps:

- Expectation step (E-step): Define $Q(\theta|\theta^{(t)}) := \mathbb{E}_{Z|X,\theta^{(t)}}[\ln L(\theta; X, Z)]$
- Maximization step (M-step): $\theta^{(t+1)} := \arg\max_\theta Q(\theta|\theta^t)$

In our model, the arrival times of the jumps are unobservable. With the discrete-time setting, the arrival times of the jumps are interpreted as the numbers of jumps in different time interval. Let j_t be the latent variable that determines the number of jumps in time interval $((t-1)\Delta, t\Delta)$, then $\Pr(J_t = l) = \frac{\lambda_t^l e^{-\lambda_t}}{l!}$, where $l = 0, 1, 2, ...$ Here we need to impose $\sup\{J_t | 1 \le t \le T/\Delta\} = m$ and m being finite in order to close the likelihood function. Let $\pi_{t,l} = \Pr(J_t = l)$, $l = 0, 1, 2, ..., m$, we have $\sum_{l=0}^m \pi_{t,l} = 1$ and the distribution of π_t can be described by a truncated Poisson distribution.

Given the jumps $J = \{J_t\}_{t=1}^T$, the returns $r = \{r_t\}_{t=1}^T$ are described by

$$r_t \sim \mathcal{N}(\bar{r} + \kappa\lambda_t - \kappa J_t, \sigma^2). \tag{3.36}$$

The aim is to estimate the unknown parameters

$$\boldsymbol{\theta} = (\bar{r}, \bar{\sigma}, \kappa, \nu, \bar{\mu}, a_\tau, \eta, p).$$

One can factorize the likelihood conditional on the log returns r, such that the likelihood reads

$$p(\boldsymbol{r}, \boldsymbol{J}|\boldsymbol{\theta}) = p(\boldsymbol{J}|\boldsymbol{r}, \boldsymbol{\theta})\, p(\boldsymbol{r}|\boldsymbol{\theta}) = p(\boldsymbol{J}|\boldsymbol{r}, \boldsymbol{\theta}) \prod_{t=1}^{T} p(r_t|r_{t-1}, ..., r_1, \boldsymbol{\theta}) \tag{3.37a}$$

$$= p(\boldsymbol{J}|\boldsymbol{r}, \boldsymbol{\theta}) \prod_{t=1}^{T} \sum_{\tilde{\boldsymbol{J}} \in \mathbb{N}_0^t} p(r_t, \tilde{\boldsymbol{J}}_{1:t}|r_{t-1}, ..., r_1, \boldsymbol{\theta}) \tag{3.37b}$$

$$= p(\boldsymbol{J}|\boldsymbol{r}, \boldsymbol{\theta}) \prod_{t=1}^{T} \sum_{\tilde{\boldsymbol{J}} \in \mathbb{N}_0^t} p(r_t|\tilde{\boldsymbol{J}}_{1:t}, \boldsymbol{\theta}_r) p(\tilde{\boldsymbol{J}}_{1:t}|r_{t-1}, ..., r_1, \boldsymbol{\theta}_J) \tag{3.37c}$$

$$= p(\boldsymbol{J}|\boldsymbol{r}, \boldsymbol{\theta}) \prod_{t=1}^{T} \mathbb{E}_{\tilde{\boldsymbol{J}}_{1:t}|r_{t-1},...,r_1,\boldsymbol{\theta}}[p(r_t|\tilde{\boldsymbol{J}}_{1:t}, \boldsymbol{\theta}_r)], \tag{3.37d}$$

where $\tilde{J}_{1:t}$ is all possible realizations of jumps up to time t, θ_r refers to the subset of model parameters that are specifically related to the return process and θ_J refers to that of the jump process.

To perform the EM algorithm, the entire expected log-likelihood is derived as

$$\mathbb{E}_{\boldsymbol{J}|\boldsymbol{r},\boldsymbol{\theta}'}[\ln p(\boldsymbol{r}, \boldsymbol{J}|\boldsymbol{\theta})] \tag{3.38a}$$

$$= \mathbb{E}_{\boldsymbol{J}|\boldsymbol{r},\boldsymbol{\theta}'}\Big[\ln p(\boldsymbol{J}|\boldsymbol{r}, \boldsymbol{\theta}) + \sum_{t=1}^{T} \ln \mathbb{E}_{\tilde{\boldsymbol{J}}_{1:t}|r_{t-1},...,r_1,\boldsymbol{\theta}}[p(r_t|\tilde{\boldsymbol{J}}_{1:t}, \boldsymbol{\theta}_r)] \Big] \tag{3.38b}$$

$$= \mathbb{E}_{\boldsymbol{J}|\boldsymbol{r},\boldsymbol{\theta}'}\Big[\sum_{t=1}^{T} \ln p(J_t|J_{t-1}, ..., J_1, \boldsymbol{r}, \boldsymbol{\theta}) \Big] + \sum_{t=1}^{T} \ln \mathbb{E}_{\tilde{\boldsymbol{J}}_{1:t}|r_{t-1},...,r_1,\boldsymbol{\theta}}[p(r_t|\tilde{\boldsymbol{J}}_{1:t}, \boldsymbol{\theta}_r)] \tag{3.38c}$$

$$= \mathbb{E}_{\boldsymbol{J}|\boldsymbol{r},\boldsymbol{\theta}'}\Big[\sum_{t=1}^{T} \ln \frac{\lambda_t(\boldsymbol{J}, \boldsymbol{r}, \boldsymbol{\theta}')^{J_t} e^{-\lambda_t(\boldsymbol{J},\boldsymbol{r},\boldsymbol{\theta}')}}{J_t!} \Big] \tag{3.38d}$$

$$+ \sum_{t=1}^{T} \ln \mathbb{E}_{\tilde{\boldsymbol{J}}_{1:t}|r_{t-1},...,r_1,\boldsymbol{\theta}}\Big[\frac{1}{\sigma\sqrt{2\pi}} \exp\Big\{ - \frac{(r_t - \bar{r} - \kappa(\lambda_t - \tilde{J}_t))^2}{2\sigma^2} \Big\} \Big], \tag{3.38e}$$

where θ' denotes the parameter estimates from the previous iteration of the EM algorithm, and (3.38d) and (3.38e) can be further derived as

$$\mathbb{E}_{\boldsymbol{J}|\boldsymbol{r},\boldsymbol{\theta}'}\Big[\sum_{t=1}^{T} (J_t \ln \lambda_t - \ln(J_t!) - \lambda_t) \Big] \tag{3.39}$$

and

$$-\frac{T}{2}\ln 2\pi\sigma^2+\sum_{t=1}^{T}\ln \mathbb{E}_{\tilde{J}_{1:t}|\theta,r_{1:t-1}}\Big[\exp\Big\{-\frac{\left(r_t-\bar{r}-\kappa(\lambda_t-\tilde{J}_t)\right)^2}{2\sigma^2}\Big\}\Big]. \quad (3.40)$$

3.5 PARAMETER ESTIMATION

In our study, the expectation of the log-likelihood function is calculated using the Monte-Carlo simulations, hence the overall approach is a Monte-Carlo Expectation-Maximization (MCEM) algorithm. In each iteration of the E-step, the Monte-Carlo simulation is employed, generating 24,000 samples to estimate the conditional expectation of the latent variables given the observed data. Given the stochastic nature of the expectation calculation, each iteration of the MCEM may yield slightly different parameter estimates, influenced by the inherent randomness of the MC. Given the challenges of intractable log-likelihood function and the non-smooth nature of the Monte-Carlo surface, traditional methods for estimating standard errors, such as those derived from the Hessian matrix, are not feasible. To address the variability introduced by this stochasticity and to estimate the uncertainty of the obtained parameter estimates, we conduct 100 complete MCEM runs with different starting points, use the mean of the estimates as the final estimates, and report the empirical standard deviation of the estimates across the runs. This method provides a straightforward and practical estimation of the parameter uncertainties in the context of MCEM, and is robust in our context.

To align the model parameters more closely with the observed data, a penalty term is incorporated into the optimization in order to minimize the discrepancy between the theoretical variance of the returns (3.24) and the realized variance observed in the data. The penalty linearly increases as the divergence between these two variances grows and encourages parameter estimates to reflect the empirical characteristics of the data.

3.5.1 *Parameter estimations for single synthetic data and ensemble distribution*

The estimation of the sample time series as shown in Fig.3.1 is presented in Table.3.6. On one hand, parameters are estimated with reasonable accuracy, suggesting that the model captures the dynamics well, especially those related to $\bar{\sigma}, \kappa, \eta$, even with a single dataset. On the other hand, the high standard deviations for parameters such as ν and $\bar{\mu}$ indicate a higher level

of uncertainty in the estimates, suggesting that single synthetic data may not be sufficient to capture the full complexity of the parameter dynamics.

	$\bar{r}$ (ann.)	$\bar{\sigma}$ (ann.)	κ	a	ν	$\bar{\mu}$	η	p
θ	0.070	0.080	0.01	0.99	2	0.1	0.93	0.4
$\hat{\theta}$	0.088	0.094	0.012	0.991	2.3	0.122	0.926	0.49
Emp. SD	(0.0298)	(0.0199)	(0.0053)	(0.0015)	(1.93)	(0.0435)	(0.0376)	(0.214)

Table 3.6: Parameter estimation and associated empirical standard deviation for the sample synthetic data.

In order to account for a complete manifestation of the characteristics necessary for a more precise parameter estimation and to average out anomalies, we perform the parameter estimation over 100 synthetic time series generated with the same parameterization as in the sample data. The mean and standard derivation of the estimates of the parameters are reported in Table 3.6. In general, the mean of the estimates from an ensemble of data is more consistent with the underlying value compared to the estimates from single data, which is quite intuitive given the law of large numbers. However, we also notice that ν, the scale factor of mispricing in the exogenous measure violates this pattern as its mean of the estimates is even larger than the already high estimate from single data. This discrepancy does not necessarily indicate an overestimation, instead it reflects the inherent difficulty in estimating ν because small changes in ν do not produce significant changes in the model's output. To understand this, we must acknowledge the limitation in treating ν as a constant. During non-bubble regimes, a minimal ν is required because the market condition is relatively stable, and a large ν could lead to overreactions to normal market noise. Conversely, during bubble regimes, market dynamics are characterized by strong positive feedback loops and jump intensity is highly responsive to changes in the mispricing, quickly accumulating potential energy for sharp corrections. Indeed, the estimation of ν may improve in performance if only the bubble regimes are considered. However, it is beyond the scope of this time series study given that identifying bubble regimes is another significant challenge in econometrics.

	$\bar{r}$ (ann.)	$\bar{\sigma}$ (ann.)	κ	a	ν	$\bar{\mu}$	η	p
θ	0.070	0.080	0.01	0.99	2	0.1	0.93	0.4
mean ($\hat{\theta}$)	0.073	0.088	0.011	0.989	2.5	0.113	0.937	0.42
std. ($\hat{\theta}$)	0.0272	0.0134	0.0039	0.0053	1.76	0.0241	0.0254	0.164

TABLE 3.8: The mean and standard deviation for the estimates of 100 synthetic time series generated with the same parameterization as in the sample data.

3.5.2 *Parameter estimations for empirical time series*

In this subsection, we calibrate the model on the daily time series of six major global indices: S&P 500, Nasdaq Composite, Nikkei 225, Hang Seng Index, DAX, and Dow Jones Industrial Average. For each indices, a time window of 5000 time points from 12-May-2004 to 29-December-2023 is selected. Parameter estimation and associated empirical standard errors are reported in Table 3.10.

	$\hat{\bar{r}}$ (ann.)	$\hat{\bar{\sigma}}$ (ann.)	$\hat{\kappa}$	$\hat{a}$	$\hat{\nu}$	$\hat{\bar{\mu}}$	$\hat{\eta}$	$\hat{p}$
S&P 500	0.077	0.088	0.013	0.990	3.1	0.063	0.912	0.50
Emp. SD	(0.0206)	(0.0172)	(0.0055)	(0.0046)	(1.68)	(0.0390)	(0.0425)	(0.144)
Nasdaq	0.073	0.089	0.013	0.990	3.2	0.114	0.920	0.47
Emp. SD	(0.0229)	(0.0174)	(0.0047)	(0.0050)	(2.05)	(0.0547)	(0.0433)	(0.143)
Nikkei	0.072	0.097	0.012	0.991	4.0	0.109	0.906	0.47
Emp. SD	(0.0234)	(0.0173)	(0.0041)	(0.0045)	(2.44)	(0.0585)	(0.0437)	(0.141)
HSI	0.069	0.094	0.012	0.991	3.5	0.093	0.917	0.47
Emp. SD	(0.0232)	(0.0171)	(0.0044)	(0.0045)	(2.25)	(0.0429)	(0.0433)	(0.154)
DAX	0.070	0.093	0.013	0.990	3.4	0.084	0.908	0.48
Emp. SD	(0.0213)	(0.0166)	(0.0048)	(0.0046)	(2.06)	(0.0473)	(0.0439)	(0.158)
DJIA	0.077	0.089	0.013	0.990	3.0	0.059	0.913	0.51
Emp. SD	(0.0205)	(0.0159)	(0.0057)	(0.0046)	(1.64)	(0.0325)	(0.0443)	(0.146)

TABLE 3.10: Parameter estimation and associated empirical standard errors for stock indices daily time series of a 5000-point time window.

While it is tempting to attribute market performance to expected returns, long-term growth rate may decouple itself from the performance trend. This is because, even though factors like expected returns provide a general picture of market behaviors, different market can react differently to the news and trigger different degrees of cascades, resulting in diverse trends and bubble dynamics. In fact, as shown in Table 3.10, the major contributions to the distinctive dynamics of different indices are $\bar{\mu}$ and η, not only due to the differences in these parameters across indices, but also because variations in these parameters have a more substantial impact compared to other parameters (e.g. $\bar{\sigma}$ and ν).

First and foremost, the near-critical branching ratios estimated from empirical data are in line with the model calibration on synthetic data, suggesting high endogeneity, i.e. strong self-reinforcing mechanisms. While branching ratios for high frequency data are found to be around 0.8 for equity futures and around 0.6 for foreign exchange (Wehrli, Wheatley, and Sornette, 2021a), higher branching ratios at low frequencies are expected given the greater influence of behavioral mechanisms.

The two benchmarks, S&P 500 and DJIA, have the lowest estimated $\bar{\mu}$, indicating a relatively low sensitivity to external shocks. This aligns with the fact that the two indices are considered stable and diversified. Their estimated η are at moderate levels, suggesting a balance between investor stability and speculative trading. They are more prone to feedback loops than, e.g. the Nikkei, but less so than e.g. the Nasdaq. The Nasdaq is characterized by the highest estimated $\bar{\mu}$ and η, which implies that it is more susceptible to volatility clustering and feedback loops. The Nasdaq tends to attract a larger volume of speculative trading partly due to the growth-oriented nature of the companies listed. This speculative behavior can amplify reactions to news and economic data, leading to increased momentum and more abrupt market changes. The high-growth technology stocks also tend to be sensitive to innovations and shifts in investor sentiment, further amplifying market instability.

For regional indices like the DAX (Germany), Nikkei (Japan), and HSI (Hong Kong), the Nikkei has a higher estimated $\bar{\mu}$ and a lower estimated η, possibly due to geopolitical factors and currency fluctuations. Its special financial environment is characterized by a prolonged period of deflation and generally conservative strategies. The resultant risk aversion can lead to, for instance, more pronounced reactions to negative shocks but less aggressive buying during recoveries. On the other hand, HSI has a higher estimated η, which appears aligned with Hong Kong's market characteris-

tics, known for its high level of retail and speculative trading, which can amplify market reactions and create feedback loops.

Additionally, Table 3.12 displays the theoretical volatility (3.24), excess volatility (3.25) and skewness (3.27) calculated from the parameter estimation, in comparison with the realized volatility and skewness. The excess volatility again confirms the significant role of endogenous interactions, and in turn justifies a small σ in the parameterization and calibration given the set-up of our model.

Furthermore, the theoretical skewness is consistently observed to be negative and smaller than the realized volatility. This is determined by the model framework, which accounts only for negative jumps as opposed to the reality where there exist both negative and positive jumps. The theoretical skewness is not necessarily aimed to match empirical skewness but rather to give an insight into the asymmetry of the inherent distribution of returns. Moreover, The empirical skewness of a particular time series segment may reflect specific, non-recurring events that are not characteristic of the typical behavior of the market (refer to the skewness for the indices for longer period in Table 3.2).

	Realized Vol.	Theoretical Vol.	Theoretical Ex. Vol.	Realized Skew	Theoretical Skew
S&P 500	19.2%	19.5%	16.6%	-0.52	-0.84
Nasdaq	21.5%	26.1%	23.5%	-0.42	-0.70
Nikkei	22.5%	22.6%	19.4%	-0.45	-0.68
HSI	23.1%	22.2%	19.2%	0.05	-0.70
DAX	20.8%	21.7%	18.7%	-0.22	-0.77
DJIA	19.1%	19.1%	16.2%	-0.38	-0.84

TABLE 3.12: Theoretical volatility, excess volatility and skewness calculated from the parameter estimation for stock indices over a 5000-point time window, compared with their realized volatility and skewness.

3.6 CONCLUSION

We have developed a novel way of modeling financial time series with crash hazard rate being influenced by both exogenous shocks and endoge-

nous dynamics. It combines two fundamental mechanisms – the propensity of agents to anchor on past price levels, and self-excitation of market crashes – to improve the description and diagnosis of speculative bubbles and their corrections.

Unlike traditional models that primarily consider exogenous factors, our model attributes a significantly portion of market dynamics to endogenous factors, accounting for how trader expectations and the clustering of market events can self-excite and propagate through time, thereby influencing market volatility and leading to crashes. The model effectively captures various stylized facts in financial markets such as volatility clustering, leverage effects, and fat-tailed distributions of returns, without relying on conventional volatility models like GARCH.

Parameter estimation through the MCEM algorithm provides insights into market dynamics. Model calibration on empirical data indicates that the branching ratio operates near critical levels, suggesting that markets are generally dominated by endogenous interactions. The calibration results also reveal that the baseline intensity of market jumps and the branching ratio are the primary drivers behind the distinctive behavioral patterns observed across various financial indices.

APPENDIX

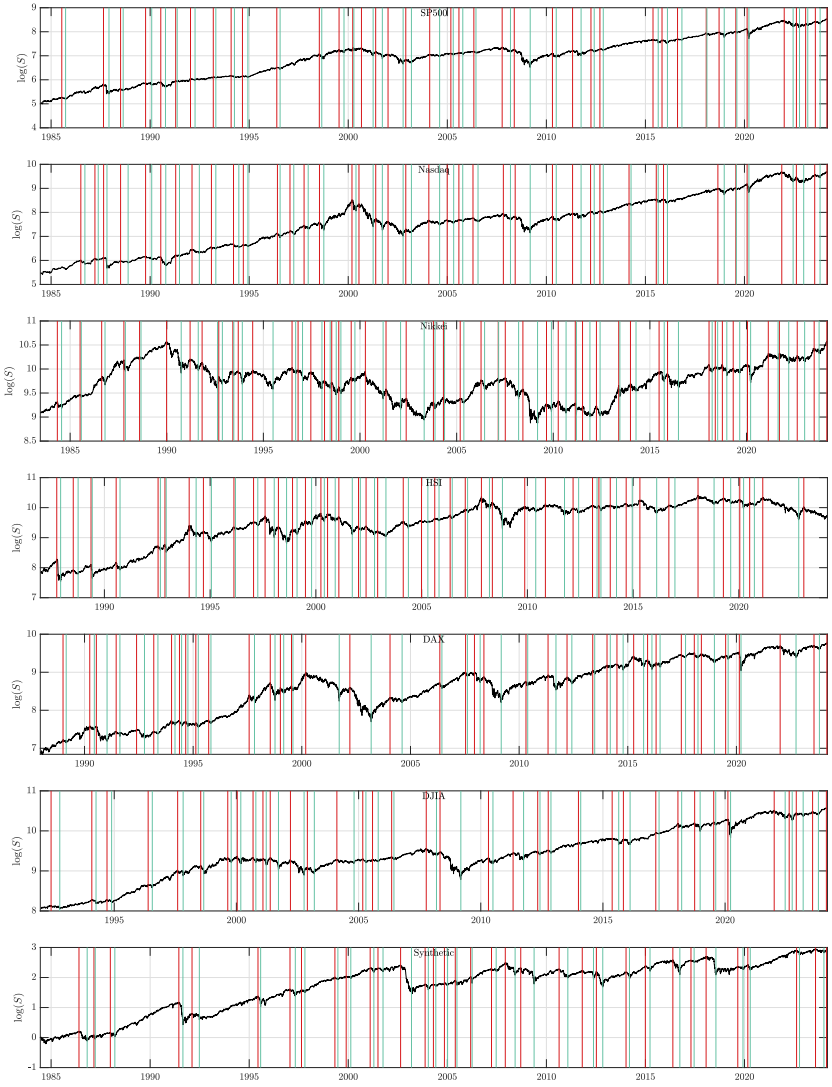

FIGURE 3.7: The start (green) and end (red) times of the drawups identified in the daily log-price series. The last panel shows the dynamics of price drawups for synthetic data in comparison with those for empirical data from major indices in the previous panels.

REFERENCES

Aït-Sahalia, Y., Cacho-Diaz, J., and Laeven, R. (2015). "Modeling financial contagion using mutually exciting jump processes". In: *Journal of Financial Economics* 117.3, 585.

Bacry, E., Dayri, K., and Muzy, J.-F. (2012). "Non-parametric kernel estimation for symmetric Hawkes processes. Application to high frequency financial data". In: *Eur. Phys. J. B* 85, 157.

Bacry, E., Mastromatteo, I., and Muzy, J.-F. (2015). "Hawkes processes in finance". In: *Market Microstructure and Liquidity* 1, 1550005.

Bacry, E. and Muzy, J.-F. (2016). "First- and Second-Order Statistics Characterization of Hawkes Processes and Non-Parametric Estimation". In: *IEEE Transactions on Information Theory* 62, 2184.

Cont, R. and Tankov, P. (2004). *Financial Modelling with Jump Processes*. Boca Raton, Florida: Chapman & Hall/CRC.

Da Fonseca, J. and Zaatour, R. (2014). "Hawkes process: Fast calibration, application to trade clustering, and diffusive limit". In: *Journal of Futures Markets* 34, 548.

Diggle, P. J. (2006). *Spatio-temporal point processes: methods and applications*. Vol. 107. Monographs on Statistics and Applied Probability, 1.

Errais, E., Giesecke, K., and Goldberg, L. R. (2010). "Affine point processes and portfolio credit risk". In: *SIAM Journal on Financial Mathematics* 1.1, 642.

Filimonov, V. and Sornette, D. (2012a). "Quantifying Reflexivity in Financial Markets: Toward a Prediction of Flash Crashes". In: *Physical Review E* 85.5.

Filimonov, V. and Sornette, D. (2015a). "Apparent criticality and calibration issues in the Hawkes self-excited point process model: application to high-frequency financial data". In: *Quantitative Finance* 15, 1293.

Filimonov, V. and Sornette, D. (2015b). "Power law scaling and "Dragon-Kings" in distributions of intraday financial drawdowns". In: *Chaos, Solitons & Fractals* 74, 27.

Hardiman, S. J., Bercot, N., and Bouchaud, J.-P. (2013). "Critical reflexivity in financial markets: A Hawkes process analysis". In: *European Physical Journal B* 86, 442.

Hardiman, S. J. and Bouchaud, J.-P. (2014). "Branching-ratio approximation for the self-exciting Hawkes process". In: *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 90, 1.

Hawkes, A. G. (2018). "Hawkes processes and their applications to finance: a review". In: *Quantitative Finance* 18, 193.
Hawkes, A. G. (2020). "Hawkes jump-diffusions and finance: a brief history and review". In: *The European Journal of Finance*, 1.
Hawkes, A. G. (1971b). "Point Spectra of Some Mutually-Exciting Point Processes". In: *Journal of the Royal Statistical Society B* 33, 438.
Hawkes, A. G. (1971a). "Spectra of Some Self-Exciting and Mutually-Exciting Point Processes". In: *Biometrika* 58, 83.
Johansen, A. and Sornette, D. (2010b). "Shocks, Crashes and Bubbles in Financial Markets". In: *Brussels Economic Review* 53.2, 201.
Johansen, A., Sornette, D., and Ledoit, O. (1999b). "Predicting Financial Crashes Using Discrete Scale Invariance". In: *Journal of Risk* 1.4, 5.
Kirchner, M. (2016). "Hawkes and INAR(∞) processes". In: *Stochastic Processes and their Applications* 126.8, 2494.
Malevergne, Y., Sornette, D., and Wei, R. (2023). "A model of financial bubbles and drawdowns with non-local behavioral self-referencing". In: *Working Paper*.
Merton, R. C. (1976). "Option Pricing When Underlying Stock Returns are Discontinuous". In: *Journal of Financial Economics* 3, 125.
Schatz, M., Wheatley, S., and Sornette, D. (2018). "The ARMA Point Process and its Estimation". In: *arXiv preprint arXiv:1806.09948*.
Sornette, D. (2005). "Endogenous Versus Exogenous Origins of Crises". In: *Extreme Events in Nature and Society*. Ed. by Albeverio, S., Jentsch, V., and Kantz, H. The Frontiers Collection. Heidelberg: Springer, 95.
Sornette, D., Malevergne, Y., and Muzy, J.-F. (2004). "Volatility Fingerprints of Large Shocks: Endogenous Versus Exogenous". In: *The Application of Econophysics*. Ed. by Takayasu, H. Tokyo: Springer Japan, 91.
Wehrli, A. and Sornette, D. (2020). *Classification of flash crashes using the Hawkes(p,q) framework*. Tech. rep. 20-92. Swiss Finance Institute.
Wehrli, A. and Sornette, D. (2021). *Excess Financial Volatility Explained by Endogenous Excitations Revealed by EM Calibrations of a Generalized Hawkes Point Process*. Research Paper Series 21-35. Swiss Finance Institute.
Wehrli, A., Wheatley, S., and Sornette, D. (2021a). "Scale-, time- and asset-dependence of Hawkes process estimates on high frequency price changes". In: *Quantitative Finance* 21.5, 729.
Westphal, R. and Sornette, D. (2020). "Market impact and performance of arbitrageurs of financial bubbles in an agent-based model". In: *Journal of Economic Behavior & Organization* 171, 1.

Wheatley, S., Wehrli, A., and Sornette, D. (2019b). "The endo-exo problem in high frequency financial price fluctuations and rejecting criticality". In: *Quantitative Finance* 19, 1165.

Wyart, M. and Bouchaud (2007). "Self-referential behaviour, overreaction and conventions in financial markets". In: *Journal of Economic Behavior & Organization* 63, 1.

4

MULTIPLE OUTLIERS DETECTION IN SAMPLES WITH EXPONENTIAL AND PARETO TAILS

This chapter introduces a novel class of ratio-based robust test statistics, designed to enhance the robustness of outlier detection in samples with exponential or Pareto tails. We also reintroduce the inward sequential testing method – formerly relegated since the introduction of outward testing – and show that MRS and SRS tests reduce susceptibility of inward approach to masking, making the inward test as powerful as, and potentially less error-prone than, outward tests. Moreover, inward testing does not require the complicated type I error control of outward tests. A comprehensive comparison of the test statistics is done, considering performance of the proposed tests in both block and sequential tests, and contrasting their performance with classical test statistics across various data scenarios. In the case studies across financial crashes, nuclear power generation accidents, stock market returns, epidemic fatalities, and city sizes – significant outliers are detected and related to the concept of 'Dragon King' events, defined as meaningful outliers that arise from a unique generating mechanism.

4.1 INTRODUCTION

Anomalous observations, while sometimes dismissed as nuisances, can be of primary relevance in applications ranging from medical diagnosis to climate science (see Aggarwal, 2013 for examples). The statistical analysis of these outliers has been extensively documented in literature (e.g., Barnett and Lewis, 1994; Hawkins, 1980 are classic references). Traditionally, statistical methods focuses on testing outliers relative to a *null* model, i.e., a model without outliers, which in many cases assumes an underlying Gaussian distribution. Common approaches include the contaminated normal models that assume Gaussian sample with Gaussian outliers (Sec. 3.4 of Hawkins, 1980), and Gaussian mixture models that are often used to detect anomalous sub-populations (Aitkin and Wilson, 1980; Hodge and Austin, 2004). However, empirical data in many fields often do not follow Gaussian distributions. In many scenarios, data can be better described by distributions with fatter tails such as exponential or power law distributions.

In particular, Pareto (power law) distributions are prevalent across a broad spectrum of phenomena, ranging from natural hazards like earthquakes, landslides, floods, and tsunamis, to industrial catastrophes such as chemical spills, nuclear accidents, and power blackouts, and extend to social systems and geopolitical events, including the distribution of wars and conflicts intensities measured by human losses (Laherrère and Sornette, 1998; Mitzenmacher, 2004; Newman, 2005; Sornette, 2006).

This paper thus focuses on the detection of outliers in samples having approximately exponential or Pareto tails. Notably, through a simple transformation, outlier tests designed for exponential samples can also be applied to Pareto samples. This places the test statistics with exponential underlying distribution at the center of our study. Moreover, Extreme Value Theory (EVT) suggests further generality of the exponential distribution by providing that general 'well-behaved' distribution functions asymptotically exhibit either exponential or Pareto tails (Embrechts, Klüppelberg, and Mikosch, 1997). Although the exponential null models have been covered in the literature (Balakrishnan, 1996 provides a review), we show that its application extends far beyond common usage. In addition, the tests we consider are independent of the parameter of the exponential distribution function, avoiding the danger of estimating the parameters of the distribution function in the presence of outliers, which can lead to strong biases.

Grounded in classical testing approaches, we introduce the two ratio-based test statistics, *max-robust-sum* (MRS) and *sum-robust-sum* (SRS), which are respectively modifications of the *max-sum* (MS) (Kimber, 1982) and the *sum-sum* (SS) test statistics (Chikkagoudar and Kunchur, 1983; Lewis and Fieller, 1979). The modifications involve altering the total sum in the ratio to a partial sum, ensure 'robust' detection of outliers by recalibrating the weight given to potential outliers in the calculation.

Historically, block testing and sequential testing have been utilized in outlier detection. The masking and swamping issues in block testing have led to the development of sequential testing. While inward sequential tests are particularly vulnerable to masking, outward sequential tests were developed to mitigate this issue (Kimber, 1982; Rosner, 1975), and have since become the standard approach. Despite their widespread adoption, outward tests are substantially more complicated as they require multiple testing corrections that control the Type I error rate. In this study, we show that the MRS and SRS test statistics effectively addresses the masking problem of inward tests, making them competitive with outward tests. Our

method opens new avenues for using inward tests without the extensive complications associated with outward methods. Furthermore, we conduct a comprehensive comparison of different statistics performed under these testing approaches, providing useful practical insights that – in the opinion of the authors – go beyond the existing literature (e.g., Balasooriya and Gadag, 1994; Chikkagoudar and Kunchur, 1983; Lin and Balakrishnan, 2009; Lin and Balakrishnan, 2014).

We also offer five interesting and extensive case studies including financial crashes, nuclear power generation accidents, stock market returns, epidemics fatalities, and city sizes. These studies shifts the outlier detection from reliability/failure applications (the exponential case) towards applications in risk modeling (the Pareto case). A number of studies have found suggestive evidence that there are extreme events 'beyond' the Pareto sample (Sornette, 2009; Sornette and Ouillon, 2012a). This brings into play the concept of 'Dragon Kings' (Sornette, 2009). We show that the proposed outlier detection method effectively identifies these 'Dragon Kings'.

The paper is organized as follows. Section 4.2 proposes the test statistics and provides their analytical distribution functions. Section 4.3 presents a variety of comparative studies across different scenarios, evaluating the performance of various tests with both dispersed and clustered outliers and their susceptibility to masking and swamping. Section 4.4 describes the general methodology in outlier detection and the argument, based on EVT, that supports the generality of the exponential outlier test. Section 4.5 explains the Dragon King (DK) concept and gives case studies on financial crashes, nuclear power generation accidents, stock market returns, epidemic fatalities, and city sizes. Section 4.6 concludes.

4.2 TEST STATISTICS

The setup is an ordered sample $x_{(1)} > x_{(2)} > ... > x_{(n)}$ where $n-k$ of the observations are i.i.d. realizations of a random variable, $X \overset{\text{iid}}{\sim} \text{Exp}(\alpha)$, with the distribution function,

$$F_X(x) = 1 - \exp\{-\alpha x\},\ x \geq 0,\ \alpha > 0\,, \tag{4.1}$$

and the remaining k points are outliers, also i.i.d. with some different distribution function, and independent of X. It is unknown which points are outliers, and the objective is to detect them. Moreover, if X is exponentially

distributed, then $Y = u \exp\{X\} \overset{\text{iid}}{\sim} \text{Pareto}(\alpha, u)$. That is, the exponential of an exponential random variable has the Pareto distribution function,

$$F(x) = 1 - (x/u)^{-\alpha},\ x \geq u,\ \alpha > 0\ . \tag{4.2}$$

Therefore, one can take the logarithm of Pareto samples and apply outlier tests intended for exponential samples.

4.2.1 *Gallery of test statistics*

We now propose the MRS and SRS statistics, and review other standard test statistics for outlier detection in exponential samples. In general, outlier test statistics compare the 'outlyingness' of the suspected outliers by contrasting them against some measure of dispersion within another subset of the data. Some of the measures are based on spacings or maxima, others on the sums of observation sizes.

The *max-robust-sum* (MRS) statistic for the j-th rank,

$$T_{j,m}^{MRS} = \frac{x_{(j)}}{\sum_{i=m+1}^{n} x_{(i)}},\ \ m \geq 0\ , \tag{4.3}$$

is a modification of a classic statistic (Kimber, 1982), here referred to as *max-sum* (MS) statistic. m is a pre-specified maximal number of outliers, and the MS statistic is recovered when $m = 0$. Index j allows the test to be used in sequential procedures for $j = 1, ..., m$. Having $m > 0$ in the denominator prevents masking: when the true number of outliers is $r > 0$, for $m < r$, there will be $m - r$ outliers in the denominator that will make $x_{(j)}$ appear less outlying. This consideration becomes crucial in inward testing, and is similar to using robust scale estimates in the case of outliers relative to a normal population (Iglewicz and Martinez, 1982). Thus, the choice of m is a tradeoff between sample size (power) and sample purity (masking avoidance). The classic MS statistic has optimal properties in the presence of a single outlier (Hawkins, 1980) as it uses a single value in the numerator. MRS/MS does not cause swamping because $x_{(j-1)}$ being outlying has no influence on the test for its smaller neighbour $x_{(j)}$. However, the limitation of this statistic is that it loses effectiveness when outliers are clustered together.

The *sum-robust-sum* (SRS) test statistic for r upper outliers,

$$T_{r,m}^{SRS} = \frac{\sum_{i=1}^{r} x_{(i)}}{\sum_{i=m+1}^{n} x_{(i)}}, \quad m \geq 1 \,, \tag{4.4}$$

is a modification of another classic test statistic (Chikkagoudar and Kunchur, 1983; Lewis and Fieller, 1979), here referred to as *sum-sum* (SS) statistic. m is again a pre-specified maximal number of outliers, and when $m = 0$, the SS statistic is recovered. In the classical form ($m = 0$), the test is equivalent to a likelihood ratio test when the outliers also come from an exponential (Balakrishnan, 1996). Due to the sum over r in the numerator, SRS/SS suffers from swamping. Nevertheless, it is not susceptible to masking because it uses the observation magnitude rather than differences; i.e., it does not compare $x_{(1)}$ versus $x_{(2)}$, which may be close to each other, but far from the rest of the sample. These test are particularly effective when the outliers are clustered.

Another classic test statistic for r upper outliers is the *Dixon* (D) statistic (Dixon, 1950),

$$T_r^D = \frac{x_{(1)}}{x_{(r+1)}} \,, \tag{4.5}$$

whose distribution function under the null is given by (Likeš, 1967). In the outward testing case, the joint distribution function was given by (Lin and Balakrishnan, 2014). It is often regarded as an inferior alternative to the SS as it considers only a limited number of points in the dataset. This statistic is less robust and also susceptible to masking compared to the SS.

We also include a test from the literature of complex systems on detecting 'Dragon King' (DK) outliers (Pisarenko and Sornette, 2012). The statistic for r upper outliers,

$$T_r^{DK} = \frac{\sum_{i=1}^{r} z_i}{\sum_{i=r+1}^{n} z_i} \sim F_{2r,2(n-r)}, \tag{4.6}$$

uses the weighted spacings $z_i = i(x_{(i)} - x_{(i+1)})$, $i = 1, ..., n-1$, $z_n = nx_{(n)}$. The statistic follows an F-distribution under the exponential null distribution. It suffers from both masking and swamping, and is less effective in the presence of multiple clustered outliers because it counts spacings rather than absolutes. This statistic is thus mainly advantageous due to the simplicity of its distribution function under the null.

Under the exponential underlying distribution (4.1), the distribution functions of all the test statistics above enjoy the pleasant property of being invariant to α. This stems from the Rényi representation of spacings

(Balakrishnan, 1996; Rényi, 1953), where, for $E_i \overset{\text{iid}}{\sim} \text{Exp}(\alpha)$, the spacings $S_i = X_{(i)} - X_{(i-1)}$ are equal in distribution to $(\alpha i)^{-1} E_i$ where $E_i \overset{\text{iid}}{\sim} \text{Exp}(1)$. Consequently, α cancels out in the ratios of sums of spacings or order statistics (which are themselves sum of spacings). This property is particularly valuable as it avoids the need to estimate α in the presence of outliers.

In addition to the test statistics mentioned above, a mixture model is considered as a benchmark,

$$f(x) = (1-\pi)\alpha \exp\{-\alpha x\} + \pi \phi(x; \mu, \sigma) \ , \ \alpha, \sigma > 0 \ , \tag{4.7}$$

where the Gaussian density $\phi(x; \mu, \sigma)$ accounts for the outlier regime, and $0 \leq \pi \leq 1$ is the weight. It is common and natural to consider Gaussian distributions to model outliers in the mixture model Aitkin and Wilson, 1980; Hodge and Austin, 2004; Verdinelli and Wasserman, 1991. It is in particular beneficial to use the mixture model when numerous outliers together form a distinct, well-defined distribution of their own. The classification of points as either outliers or not is based on the relative weights of the components. The model parameters are estimated using the EM (Expectation-Maximization) algorithm (Redner and Walker, 1984), and a likelihood ratio test against the null ($\pi = 0$) is used to generate p-values and estimate the number of outliers ($n\hat{\pi}$). The major advantage of this approach is that it does not require sequential testing, thereby naturally avoiding masking and swamping. Moreover, the model can be extended beyond the exponential, e.g., to a Weibull or gamma distribution function, without complicating the procedure. It is important to note that this method does not distinguish between inliers and outliers – i.e., the density ϕ can be significant both within and beyond where the null distribution function has substantial mass.

4.2.2 *Distribution function of test statistics*

Let X_i, $i = 1, \ldots, n$ be i.i.d realizations of a random variable X with exponential distribution

$$F(x) = 1 - e^{-\alpha x}, \quad x \geq 0. \tag{4.8}$$

Define the order statistics as the ordered sample $X_{(1)} > X_{(2)} > \cdots > X_{(n)}$, and the trimmed sum $S_{m,n} = X_{(m+1)} + \cdots + X_{(n)}$, $0 \leq m < n$, hence the test statistic $T_{j,m} = X_{(j)} / S_{m,n}, j = 1, \ldots, m$. Also define the spacing

$D_{j,m} = X_{(j)} - X_{(m)}$, and use the triple $\left(X_{(m)}, D_{j,m}, S_{m,n}\right)$ to derive the distribution of $T_{j,m}$.

First, we derive the distribution of $D_{j,m}$ and the joint distribution of $\left(X_{(m)}, S_{m,n}\right)$

$$f_{D_{j,m}}(d) \propto \left(1 - e^{-\alpha d}\right)^{m-j-1} e^{-j\alpha d}, \quad d \geq 0 \tag{4.9}$$

and

$$f_{\left(X_{(m)}, S_{m,n}\right)}(x,s) \propto e^{-\alpha(s+mx)} \sum_{k=0}^{n-m-1} (-1)^k \binom{n-m}{k} (s-kx)_+^{n-m-1}, \quad 0 \leq s \leq (n-m)\, x. \tag{4.10}$$

The joint distribution of the triple is then derived as the product of (4.9) and (4.10), presented by

$$\begin{aligned} f_{\left(X_{(m)}, D_{j,m}, S_{m,n}\right)}(x,d,s) &= f_{D_{j,m}}(d)\, f_{\left(X_{(m)}, S_{m,n}\right)}(x,s) \\ &\propto \left(1 - e^{-\alpha d}\right)^{m-j-1} e^{-\alpha(jd+s+mx)} \sum_{k=0}^{n-m-1} (-1)^k \binom{n-m}{k} (s-kx)_+^{n-m-1} \end{aligned} \tag{4.11}$$

over the region $0 \leq s \leq (n-m)\, x,\ d \geq 0$.

Next, apply the change of variable $z = x + d$ and derive $f_{\left(X_{(j)}, D_{j,m}, S_{m,n}\right)}(z = x+d, d, s)$ from (4.11), then marginalize over d to obtain $f_{\left(X_{(j)}, S_{m,n}\right)}(z,s)$, which reads

$$\begin{aligned} f_{\left(X_{(j)}, S_{m,n}\right)}(z,s) \propto \sum_{i=1}^{m-j} (-1)^{m-j-i} \binom{m-j-1}{i-1} e^{-\alpha(s+mz)} \Bigg\{ \sum_{k=1}^{n-m-1} (-1)^k \binom{n-m}{k} k^{n-m-1} \\ \times \frac{(n-m-1)!}{(-\alpha i)^{n-m}} (A_0 + B_0 - C_0) + \frac{s^{n-m-1}}{\alpha i} \left(e^{\alpha i\left(z - \frac{s}{n-m}\right)} - 1\right) \Bigg\}, \quad s \leq (n-m)\, z \end{aligned} \tag{4.12}$$

where

$$A_0 = e^{\alpha i\left(z - \frac{s}{k}\right)_+}, \tag{4.13}$$

$$B_0 = \sum_{l=1}^{n-m-1} (-1)^l \frac{\left(\alpha i \left(\frac{s}{k} - z\right)_+\right)^l}{l!}, \tag{4.14}$$

$$C_0 = e^{\alpha i\left(z - \frac{s}{n-m}\right)} \sum_{l=0}^{n-m-1} (-1)^l \frac{\left(\alpha i \left(\frac{s}{k} - \frac{s}{n-m}\right)\right)^l}{l!}. \tag{4.15}$$

Proceed with another change of variable $t = z/s$ to derive $f_{\left(T_{j,m},S_{m,n}\right)}(t = z/s, s)$, then integrate out s to get the null distribution function of the MRS (for $j \leq m-1$),

$$f_{T_{j,m}}(t) \propto \sum_{i=1}^{m-j}(-1)^{m-j-i}\binom{m-j-1}{i-1}\left\{\sum_{k=1}^{n-m-1}(-1)^k\binom{n-m}{k}k^{n-m-1}\frac{A_1+B_1-C_1}{(-i)^{n-m}}+\right.$$
$$\left.\frac{n-m}{i}\left(\frac{1}{\left(1+i\left(\frac{1}{n-m}-t\right)+mt\right)^{n-m+1}}-\frac{1}{(1+mt)^{n-m+1}}\right)\right\}, \quad t \geq \frac{1}{n-m} \tag{4.16}$$

where

$$A_1 = \frac{1}{(1+mt-i(t-\frac{1}{k})_+)^2}, \tag{4.17}$$

$$B_1 = \sum_{l=1}^{n-m-1}(-1)^l(1+l)\frac{\left(i\left(\frac{1}{k}-t\right)_+\right)^l}{(1+mt)^{l+2}}, \tag{4.18}$$

$$C_1 = \sum_{l=0}^{n-m-1}(-1)^l(1+l)\frac{\left(i\left(\frac{1}{k}-\frac{1}{n-m}\right)\right)^l}{\left(1+i\left(\frac{1}{n-m}-t\right)+mt\right)^{l+2}}. \tag{4.19}$$

When $j = m$, the null distribution function of the MRS reads

$$f_{T_{m,m}}(t) \propto \sum_{k=0}^{n-m-1}(-1)^k\binom{n-m}{k}\frac{(1-kt)_+^{n-m-1}}{(1+mt)^{n-m-1}}, \quad t \geq \frac{1}{n-m}. \tag{4.20}$$

Next, in order to derive the distribution function of $T_{r,m} = \frac{\sum_{i=1}^{r}x_{(i)}}{\sum_{i=m+1}^{n}x_{(i)}} = \frac{S_{0,r}}{S_{m,n}}, r = 1,\ldots,m$, we are interested in the conditional distribution of $\left(S_{0,r}|X_{(r+1)}\right)$, which reads

$$f_{\left(S_{0,r}|X_{(r+1)}\right)}(s|x) \propto e^{-\alpha(s-rx)}(s-rx)^{r-1}, \quad s \geq rx. \tag{4.21}$$

The conditional distribution helps us to derive the joint distribution of the triple $\left(X_{(r+1)}, S_{0,r}, S_{m,n}\right)$ given the joint distribution of $\left(X_{(r+1)}, S_{m,n}\right)$, which can be conveniently obtained from (4.12). By performing transformation of variables and marginalization on the distribution function of

the triple, one can obtain the null distribution function of the SRS. Both it and the null distribution function of the MRS, which is detailed in the paper, are provided as Matlab code and are available for access on GitHub (`github.com/ranwei-ethz/dfs-appendix`) for direct implementation and verification of the procedures discussed.

4.3 OUTLIER TEST PERFORMANCE

4.3.1 *Set-up of synthetic tests and issues*

In this section, we compare the performance of the different tests through simulation studies. First, we examine the robustness of test statistics in block tests and contrast them with the mixture test, and study the issues of masking and swamping. Then, we draw comparisons between inward, outward, and mixture tests. Additionally, we examine how misspecification of the null affects test performance. The setup uses a standard exponential sample across four outlier scenarios: (0) no outliers, (I) a single outlier, (II) multiple dispersed outliers, and (III) a cluster of multiple outliers. These scenarios are plotted in Fig. 4.1.

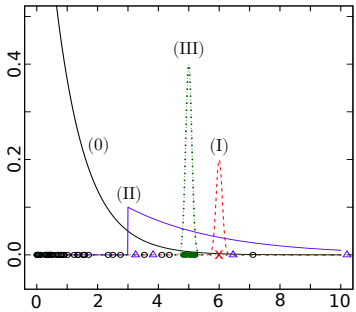

FIGURE 4.1: **Outlier cases**. The null case (0) is the standard exponential for which a realization of 50 points are plotted as open circles. Three outlier cases are considered in addition to the null: (I) a normal distribution function with mean $\mu = 6$ and $\sigma = 0.1$, presented by a dashed red line, and its single outlier is the red x mark; (II) multiple dispersed outliers $Y_i \sim 3+\text{Exp}(1/\beta)$, $i = 1, ..., 5$, plotted with a solid blue line for $\beta = 4$, with blue triangles indicating (a realization of) the outliers; (III) multiple clustered outliers $Y_i \sim \text{Norm}(\mu, 0.1)$, $i = 1, ..., 5$, plotted with a green dotted line for $\mu = 5$, with green dots indicating the outliers.

Let us first revisit the testing methods and the issues of masking and swamping for the sake of clarity.

The block test examines a fixed set r of suspected outliers in a single test. It determines if all r points are outliers or none. Its success critically depends on accurately choosing r, with risks of masking if r is set below the actual number of outliers, and swamping if set above. However, if well specified, it is powerful because of the simultaneous usage of all data.

The inward test sequentially tests the most extreme data point and removes it if identified as an outlier. The test continues to the next extreme value until a point is not recognized as an outlier. The estimated number of outliers $\hat{k}$ is the number of rejected (marginal) tests. Clearly, this test can suffer from both masking and swamping, and the weaknesses of the inward procedure were cited as motivation for the *outward* test (Hawkins, 1980; Kimber, 1982; Rosner, 1975):

The outward test specifies a maximum number of outliers r, and starts by testing if the r-th largest value $x_{(r)}$ is an outlier by excluding the larger values $x_{(r-1)}, x_{(r-2)}, ..., x_{(2)}, x_{(1)}$. If this test is rejected, then r outliers are identified. If not, the test progresses to the $(r-1)$-th largest point $x_{(r-1)}$. The test continues until the first detection of outlier. The outward test minimizes the probability and magnitude of both masking and swamping, and has therefore been claimed superior over the inward (Balasooriya and Gadag, 1994; Chikkagoudar and Kunchur, 1983; Kimber, 1982) and received more subsequent development (Lin and Balakrishnan, 2009; Lin and Balakrishnan, 2014).

However, control of the type I error (the probability of a false detection of an outlier) is difficult in the outward test. The test considers the null hypothesis H_0 that there are no outliers, with multiple alternatives, H_j that there are j outliers $j = 1, ..., r$, with test statistic T_j. A single rejection of the r tests rejects the null H_0. Thus, to achieve an overall type I error level of $0 \leq a \leq 1$, e.g., the common level of 0.05 or 0.1, the marginal tests need to have a lower level. The larger r is, the larger the correction will be, and thus the lower the power of the test. This 'multiple testing correction' requires knowing the joint and marginal distribution function of, generally dependent, $T_j,\ j = 1, ..., r$. More specifically, one defines all marginal tests to have equal level b, i.e., $Pr\{T_j > t_j\} = b,\ j = 1, ..., r$, and the level b is determined such that $Pr\{T_j \leq t_j,\ j = 1, ..., r | H_0\} = 1 - a$. Clearly $a^r \leq b \leq a$, where the lower bound corresponds to the case of independent tests (the Bonferroni bound), and the upper bound to perfect dependence.

For the specific test statistic (4.3) discussed below, the joint and marginal distribution functions have been derived as described in 4.2.2.

In contrast, for the inward method, the type I error level is equal to the marginal level ($a = b$) because a rejection of the null only happens when the first marginal test (for the largest point, $x_{(1)}$) is rejected. This is a major advantage over the outward procedure in terms of computation and also because no power is lost due to a multiple testing correction.

4.3.2 *Performance of block tests*

Here, we examine the power (at level 0.1) of the range of test statistics employed in block tests, where the block size r, and the robustness value m are set to the true number of outliers k. We consider three scenarios involving different types of outliers: (I) $n = 20$, $k = 1$, $X_i \sim$Exp(1), $i = 1, ..., 19$, $X_{20} \sim$Norm$(\mu, 0.1)$, $\mu = 3, ..., 10$; (II) $n = 50$, $k = 5$, $X_i \sim 3+$Exp$(1/\beta)$, $i = 46, ..., 50$, $\beta = 1, 2, ..., 6$; (III) $n = 50$, $k = 5$, $X_i \sim$Exp(1), $i = 1, 2, ..., 45$, $X_i \sim$Norm$(\mu, 0.1)$, $i = 46, ..., 50$, $\mu = 3, 4, ..., 10$. The mixture model (4.7) is only estimated in the cases with multiple outliers.

Fig. 4.2 shows the power curves of the test statistics, for a range of outlier parameters being computed over 10'000 independent simulations. For a single outlier (case I), most of the tests are exactly identical (by definition), with the exception of the DK and D tests, which are weaker. For multiple dispersed outliers (case II), the SS and the SRS tests outperform others and are equivalently effective, followed by the MRS test. The mixture is poorly specified and is thus weakest. For clustered outliers (case III), the performance of the tests varies greatly. Indeed, the test statistics with the sum in the numerator often identifies the cluster of outliers. However, the well specified mixture model is most powerful, also identifying the 'outliers' when they are not really outlying but rather a contamination well within the sample (i.e., 'inliers'). In detecting multiple outliers, the SS and SRS tests perform almost identically, largely because they both operate under optimal conditions where the block size r and the robustness value m are set to the true number of outliers k. The robustness of the SRS statistic is further demonstrated subsequently in context involving masking and swamping.

We now present simulation studies to expose the degree to which the different test statistics suffer from masking and swamping in block tests – that is, how accurately they estimate the number of outliers. This is done by performing the tests on synthetic data for a range of block sizes. The three scenarios considered are: (I) swamping due to a single outlier,

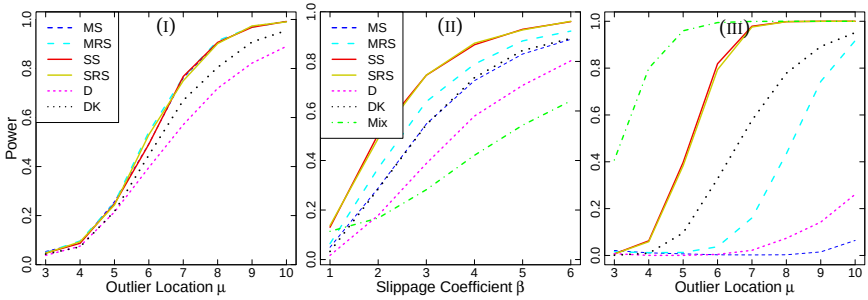

FIGURE 4.2: Power curves (at level 0.1) for the MS/MRS (4.3), SS/SRS (4.4), D (4.5), DK (4.6) statistics, employed in block tests, with the mixture model (4.7) estimated in the cases with multiple outliers. The power curves are plotted against the outlier location μ and the slippage coefficient β, which are specified in the three outlier cases: (I) $n = 20$, $k = 1$, $X_i \sim$Exp(1), $i = 1, ..., 19$, $X_{20} \sim$Norm$(\mu, 0.1)$; (II) $n = 50$, $k = 5$, $X_i \sim 3+$Exp$(1/\beta)$, $i = 46, ..., 50$; (III) $n = 50$, $k = 5$, $X_i \sim$Norm$(\mu, 0.1)$, $i = 46, ..., 50$.

$n = 30$, $k = 1$, $X_i \sim$Exp(1), $i = 1, ..., 29$, $X_{30} \sim$Norm$(8, 0.1)$; (II) swamping without masking due to dispersed outliers, $n = 30$, $k = 5$, $X_i \sim$Exp(1), $i = 1, ..., 25$, $X_i \sim 3+$Exp$(1/5)$, $i = 26, ..., 30$; and (III) swamping with masking due to clustered outliers, $n = 30$, $k = 5$, $X_i \sim$Exp(1), $i = 1, ..., 25$, $X_i \sim$Norm$(8, 0.1)$, $i = 26, ..., 30$.

Our simulation study determines the frequency at which the tests are rejected, at level 0.1, in 10'000 independent samples, for a range of block sizes ($b = 1, 2, ..., 10$). The results are in Fig. 4.3. The MS and MRS tests are not affected by block size since the maximum is always the largest point. In the next section, the inward test will apply the MRS statistic to the largest point, then the second largest, and so on. In that case, the MRS will not cause swamping. As anticipated, masking is problematic for the MS statistic, especially when large observations are densely clustered. Further, as intended, the MRS suffers from masking less than the MS. The SS and SRS tests suffer less from masking and swamping than those based on spacings and maxima. Swamping is pervasive in block testing, even when there is only a single large outlier. That the rejection rate decays slowly as the block size surpasses the true block size indicates that the minimal p-value in the sequence of estimates will not reliably indicate the true block size. These problems motivate sequential testing.

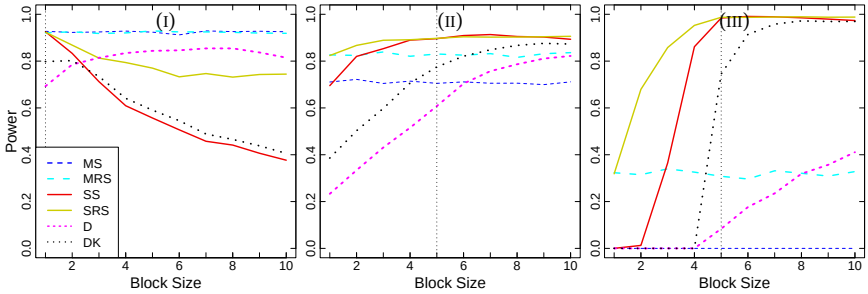

FIGURE 4.3: Power curves (at level 0.1) for the MS/MRS (4.3), SS/SRS (4.4), D (4.5), DK (4.6) statistics, with different block sizes. The data generating processes for the cases are: (I) $n = 30$, $k = 1$, $X_i \sim$Exp(1), $i = 1, ..., 29$, $X_{30} \sim$Norm$(8, 0.1)$; (II) $n = 30$, $k = 5$, $X_i \sim 3+$Exp$(1/5)$, $i = 26, ..., 30$; and (III) $n = 30$, $k = 5$, $X_i \sim$Norm$(8, 0.1)$, $i = 26, ..., 30$. In each case, the true number of outliers is given by the vertical dotted line.

4.3.3 *Performance of sequential tests*

Here, inward and outward sequential procedures are compared, along with the mixture test. Again the four outlier scenarios visualized in Fig. 4.1 are considered. The tests used are: (i) the outward test with MS, MRS, SS, and SRS statistics; (ii) the inward test with only the MRS statistic, which is necessary to avoid masking and swamping; (iii) the mixture model (4.7); and (iv) the SRS block test, given the correct number of outliers. This last option, which was the best performing block test in Fig. 4.2, provides a benchmark.

The distribution functions for the test statistics were simulated with 50'000 samples from the null model. All tests were done with a level of 0.1. For the outward test, the level of the marginal tests b was lowered to obtain the overall level of $a = 0.1$. For each test, this was done by applying the test on 10'000 independent samples generated from the null, for multiple values of b, and selecting b such that $a(b) = 0.1 \pm 0.005$. The resultant marginal levels are in Table 4.1. Note how large of an adjustment is needed in the outward test, whereas in the inward test there is no adjustment: $b^{\text{Inward}} = a = 0.1$.

The results, for slightly different specifications of the four cases, and in order of decreasing sample size, are in Tables 4.2, 4.3, and 4.4. In case (0), where there are no outliers, the inward and mixture procedures have

n	r	MS	SS	MRS	SRS
50	10	0.018	0.05	0.025	0.049
30	5	0.028	0.055	0.0345	0.0575
15	5	0.025	0.06	0.036	0.056

TABLE 4.1: Marginal levels (b) for outward tests for different sample sizes (n), maximal number of outliers (r), and robustness value ($m = r$) to obtain an overall type I error level of $a = 0.1$

Case	Quantity	MS Out	SS Out	MRS Out	SRS Out	MRS In	Mix	SRS Block
(o)	Rej. Rate	0.11	0.10	0.11	0.10	0.10	0.14	0.10
(o)	$\widehat{k}$	(3,6,9)	(5,9,10)	(3,6,9)	(5,9,10)	(1,1,3)	(2,2,4)	
(I)	Rej. Rate	0.30	0.22	0.30	0.22	0.64	0.09	0.69
(I)	$\widehat{k}$	(2,3,6)	(2,5,10)	(2,3,7)	(2,5,10)	(1,1,2)	(2,2,2)	$= 1$
(II)	Rej. Rate	0.91	0.75	0.89	0.75	0.04	0.95	0.38
(II)	$\widehat{k}$	(5,7,8)	(5,7,10)	(5,7,9)	(5,7,10)	(1,9,10)	(5,5,6)	$= 5$
(III)	Rej. Rate	0.96	0.96	0.97	0.96	0.95	0.63	0.98
(III)	$\widehat{k}$	(5,6,8)	(4,6,10)	(5,6,9)	(4,6,10)	(6,7,10)	(3,10,10)	$= 5$

TABLE 4.2: $n = 50$ (sample size), $m = 10$ (robustness value). Summary of tests over 5000 repeated simulations of four cases: (o) the null case ($X \sim$Exp(1)), (I) a single large outlier ($X_i \sim$Exp(1), $i = 1, ..., 49$; $X_{50} \sim$Norm$(7, 0.1)$), (II) a cluster of multiple outliers ($X_i \sim$Exp(1), $i = 1, ..., 45$; $X_i \sim$Norm$(5, 0.1)$, $i = 46, ..., 50$); (III) multiple dispersed outliers ($X_i \sim$Exp(1), $i = 1, ..., 45$; $X_i \sim \max(\{X_i : i = 1, ..., 45\})$+Exp$(1/5)$, $i = 46, ..., 50$). The rejection rate and the median $\widehat{k}$ and quartiles of the estimated number of outliers (in the event of a rejection) are given in alternating rows.

false positive events that estimate a small number of outliers, whereas the outward procedures falsely identify large numbers of outliers. In case (I) of the sequential procedures, the inward test is most powerful at identifying the single outlier, even matching the power of the block test. The outward tests are substantially weakened, even with relatively small $m = 5$. The inward test provides superior estimation of outliers, whereas the other tests tend to overestimate. In case (II), with a cluster of outliers, both the benchmark (the block test) and the inward test perform poorly. They are outperformed by the outward test, which is less susceptible to masking, by design. However, here the mixture approach is both the most powerful

and accurate in estimating outlier numeracy. In case (III), with multiple dispersed outliers, all of the inward and outward approaches are similarly competitive, while being slightly dominated by the block test. The mixture approach is weak since the outlier component is poorly specified. For the outward procedure, the MS/MRS statistic dominates the SS statistic.

In summary, the inward procedure with the MRS test statistic is more computationally convenient than the outward procedure, commits less severe false positives, and can even be more powerful when identifying single or multiple dispersed outliers. In the event of a dense cluster of outliers, a mixture approach can be more computationally convenient and powerful than the outward approach. Within the outward approach, the MS/MRS statistic is found superior to the SS/SRS statistic, and robust modifications performed similarly.

Case	Quantity	MS Out	SS Out	MRS Out	SRS Out	MRS In	Mix	SRS Block
(0)	Rej. Rate	0.11	0.11	0.11	0.11	0.11	0.16	0.10
(0)	$\widehat{k}$	(2,3,5)	(4,5,5)	(2,4,5)	(3,5,5)	(1,1,3)	(2,2,5)	
(I)	Rej. Rate	0.45	0.32	0.43	0.33	0.72	0.08	0.75
(I)	$\widehat{k}$	(1,2,3)	(1,3,5)	(1,2,3)	(1,2,5)	(1,1,2)	(2,2,2)	$= 1$
(II)	Rej. Rate	0.72	0.63	0.73	0.64	0.08	0.96	0.36
(II)	$\widehat{k}$	(3,4,5)	(3,4,5)	(3,4,5)	(3,4,5)	(4,5,5)	(3,3,3)	$= 3$
(III)	Rej. Rate	0.87	0.86	0.89	0.86	0.88	0.50	0.90
(III)	$\widehat{k}$	(2,4,4)	(2,4,5)	(2,4,5)	(3,4,5)	(3,4,5)	(2,5,7)	$= 3$

TABLE 4.3: $n = 30$ (sample size), $m = 5$ (robustness value). Summary of tests over 5000 repeated simulations of four cases: (0) the null case ($X_i \sim$Exp(1)), (I) a single large outlier ($X_i \sim$Exp(1), $i = 1,...,29$; $X_{30} \sim$Norm$(7, 0.1)$), (II) a cluster of multiple outliers ($X_i \sim$Exp(1), $i = 1,...,27$; $X_i \sim$Norm$(5, 0.1)$, $i = 28, 29, 30$), (III) multiple dispersed outliers ($X_i \sim$Exp(1), $i = 1,...,27$; $X_i \sim \max(\{X_i : i = 1,...,27\})+Exp(1/5)$, $i = 28, 29, 30$). The rejection rate and the median $\widehat{k}$ and quartiles of the estimated number of outliers (in the event of a rejection) are given in alternating rows.

Case	Quantity	MS Out	SS Out	MRS Out	SRS Out	MRS In	Mix	SRS Block
(o)	Rej. Rate	0.11	0.11	0.11	0.11	0.08	0.16	0.10
(o)	$\widehat{k}$	(2,3,4)	(3,5,5)	(2,3,5)	(3,5,5)	(1,2,4)	(2,3,5)	
(I)	Rej. Rate	0.25	0.22	0.23	0.20	0.30	0.14	0.30
(I)	$\widehat{k}$	(2,3,4)	(2,4,5)	(2,3,4)	(2,4,5)	(1,2,3)	(2,2,4)	$= 1$
(II)	Rej. Rate	0.42	0.42	0.43	0.41	0.04	0.93	0.13
(II)	$\widehat{k}$	(3,4,5)	(4,5,5)	(3,4,5)	(3,5,5)	(3,4,5)	(3,3,3)	$= 3$
(III)	Rej. Rate	0.63	0.62	0.64	0.62	0.63	0.37	0.66
(III)	$\widehat{k}$	(2,3,4)	(2,4,5)	(2,3,4)	(2,4,5)	(2,3,5)	(2,3,4)	$= 3$

TABLE 4.4: $n = 15$ (sample size), $m = 5$ (robustness value). Summary of tests over 5000 repeated simulations of four cases: (o) the null case ($X_i \sim$Exp(1)), (I) a single large outlier ($X_i \sim$Exp(1), $i = 1, ..., 14$; $X_{15} \sim$Norm$(4, 0.1)$), (II) a cluster of multiple outliers ($X_i \sim$Exp(1), $i = 1, ..., 12$; $X_i \sim$Norm$(4, 0.1)$, $i = 13, 14, 15$), (III) multiple dispersed outliers ($X_i \sim$Exp(1), $i = 1, ..., 12$, $X_i \sim \max(\{X_i : i = 1, ..., 12\}) +Exp(1/5)$, $i = 13, 14, 15$). The rejection rate and the median $\widehat{k}$ and quartiles of the estimated number of outliers (in the event of a rejection) are given in alternating rows.

4.3.4 *Robustness to null mis-specification*

In practice, the correct specification of the null/main model is of considerable importance. Here, the sensitivity of the rate of false positives (level / type I error), and true positives (power), to the degree of misspecification of the null are exposed via a simulation study, for the battery of test statistics implemented in block tests. We consider simulating data from a Weibull distribution,

$$F(x) = 1 - \exp\{-(x/\tau)^\kappa\},\ x \geq 0,\ \tau, \kappa > 0\,, \tag{4.22}$$

which is exponential ($\alpha = \tau^{-1}$) when $\kappa = 1$, is fat tailed for $\kappa < 1$, and becomes concentrated at τ as κ becomes large. The results of the simulation study are presented in Fig. 4.4 and can be described as follows.

Panel (b) concerns the rate of false positives where $r = 3$ outliers are tested, with level $a = 0.1$, in a Weibull (4.22) sample of size $n = 30$, for a range of shape parameters κ, without outliers. When $\kappa < 1$, the distribution function is fat tailed, having many events that are large, and thus the tests falsely identify many points as outliers. This is problematic in practice (with small to moderate sample sizes), because one does not know what the true

null model is. For instance, with $n = 30$, even when the true distribution function is considered as an alternative model versus the exponential, and using the powerful likelihood ratio test, 50 percent of the time (for $\kappa \approx 0.6$), one will not reject the exponential model at a level of 0.1. In this case, when falsely retaining the exponential model, the type I error will be between 0.3 and 0.5, depending on the selected test statistic. The KS test of compatibility of the data with the exponential distribution function is even less powerful, allowing for more severe false positives.

Case (c) considers the frequency of true positives (power). The setup is the same as above, but 3 dispersed outliers are included. When the Weibull distribution function becomes less fat tailed, the power of the SRS and MRS tests decreases whereas the power of the D and DK tests increases. Here, with $n = 30$, for the tests of the Weibull versus the exponential, including the outliers in the sample, there is a high probability (0.6-0.8) of not rejecting the exponential model when $1 < \kappa < 1.5$, where the power of some of the tests is weakened.

It is clear that the power, and especially the level, are highly sensitive to the validity of the exponential model, and misspecification of the null can lead to erroneous inference. This has important implications for the practical application of the tests. In particular, one should have a sufficiently large sample to diagnose the validity of the null, and not blindly accept/reject the result of the test and its diagnostics.

4.4 GENERALITY OF EXPONENTIAL DISTRIBUTION

It is important to note that outlier tests with both the Pareto and exponential underlying distributions are generally applicable to data having approximately Pareto or exponential tails. This follows from the well known Pickands-Balkema-de Haan theorem of Extreme Value Theory (EVT), that states (Embrechts, Klüppelberg, and Mikosch, 1997): For a broad range of distributions, for random variable X, with sufficiently high threshold u, the excess distribution function, $F_u(x) = P\{X - u \leq x | X - u > 0\}$ (i.e., the tail of the distribution function), is approximated by the GPD (Generalized Pareto Distribution Function),

$$GPD(x; \xi, \beta, \mu) = \begin{cases} 1 - (1 - \xi(x - \mu)/\beta)^{-1/\xi}, & \text{if } \xi \neq 0 \\ 1 - \exp(-(x - \mu)/\beta), & \text{if } \xi = 0\,, \end{cases} \tag{4.23}$$

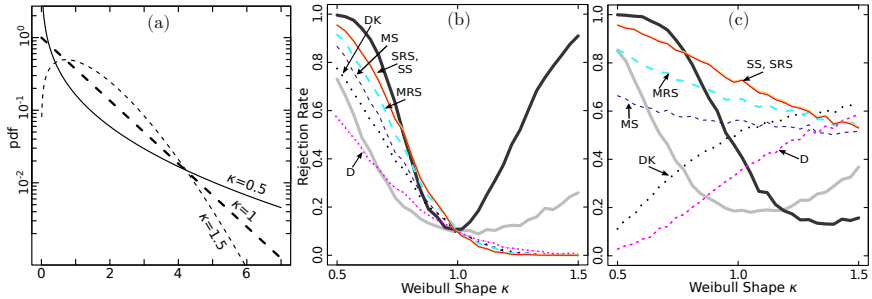

FIGURE 4.4: **Test robustness** Panel (a): The Weibull PDF (4.22) plotted for parameters (κ, β) equal to $(0.5, 0.4)$, $(1, 1)$ and $(1.5, 1.5)$. Panel (b): The frequency of rejection of the null of no outliers, at level 0.1, in the presense of no outliers, for block tests for $r = 3$ outliers, assuming an exponential null model, when the data is generated from a Weibull for a range of shape parameters κ. Panel (c): The frequency of rejection of the null using a level 0.1, of the block tests for $r = 3$ outliers, with the same setup as frame (b), except that 3 outliers are truly present. The models for the cases are: (b) $X_i \sim$Weibull$(\kappa, 1)$, $i = 1, ..., 30$; (c) $X_i \sim$Weibull$(\kappa, 1)$, $i = 1, ..., 27$, $X_i \sim \max(\{X_i : i = 1, ..., 27\}) + \text{Exp}(1/3)$, $i = 28, 29, 30$). For each case, simulation and testing were performed 1000 times for κ sweeping 0.5 to 1.5. The tests are colour coded: SS (red solid), SRS with $m = r$ (yellow solid), MS (blue dashed), MRS with $m = r$ (turquoise heavy dashed), D (magenta light dotted), DK (black dotted). In both frames, the black heavy solid line is the power of the likelihood ratio test of the Weibull versus the exponential on the data (including outliers). Similarily the grey heavy solid line is for the Kolmogorov-Smirnov test.

in the sense that,

$$\lim_{u \to \infty} \sup_{0 \le x} | F_u(x) - GPD(x | \xi, \beta(u), \mu)| = 0 \,, \quad \beta(u) > 0, \forall u. \quad (4.24)$$

If $\xi = 0$ (the Gumbel case), then the GPD (4.23) is exponential with lower truncation $\mu = u$ and scale parameter $\beta = 1/\alpha$. This case includes common distributions such as the exponential (obviously), the Normal, and even some fat-tailed ones such as the Lognormal. If $\xi > 0$ (the Fréchet case), the GPD (4.23) is (generalized) Pareto with $\mu = u$, $\sigma = u/\alpha$, and $\xi = 1/\alpha$. This case includes heavy tailed distributions such as the Pareto and Log-gamma. The only other case ($\xi < 0$: the Weibull case) is where the distribution function has a finite upper endpoint, which is of less interest in outlier

detection. Therefore, since a Pareto tail can be transformed to an exponential one, outlier testing in exponential samples is (asymptotically) extremely general!

Since the GPD approximation (4.24) is only asymptotically valid, one must select a sufficiently large lower threshold u before applying outlier tests. The problem of threshold selection is a tradeoff between bias and variance, and is the primary statistical issue in the EVT literature, where it is referred to as sample fraction selection. In the physics literature, threshold selection and goodness of fit diagnostics are important for the interpretation of mechanisms underlying power laws found in datasets. There are a variety of tools available for this task.

The classic 'Hill plot' method Hill, 1975 for threshold selection consists of estimating the model for a range of thresholds and selecting the lowest threshold (the largest sample fraction) where the estimate is 'stable' – i.e., consistent with values of the estimate for larger thresholds. See Fig. 4.5 for an example. Of course, one can also look for statistically significant changes in the estimated parameter relative to the hopefully stable value obtained deeper in the tail Bauke, 2007; Hall and Welsh, 1985; Hill, 1975, however more powerful principled methods exist (see e.g., Beirlant et al., 2006; Gomes and Oliveira, 2001 for a review). For instance, let us mention the methods based on minimizing the asymptotic mean square error of the estimate. This requires assuming the (class of) distribution beyond the power law tail Hall and Welsh, 1985, or using bootstrap methods Danielsson et al., 2001; Gomes and Oliveira, 2001.

These methods have not been extensively adopted outside of the EVT literature. For instance, the most highly cited paper on the estimation of power laws and sample fractions Clauset, Shalizi, and Newman, 2009 does not mention the sample fraction estimation literature. However a subsequent work Virkar and Clauset, 2014, extending the method to binned/aggregated data, does provide such references. The popular work Clauset, Shalizi, and Newman, 2009 suggests choosing the pair of u and α that have the smallest KSD (Kolmogorov-Smirnov distance). The KSD criterion penalizes error, and rewards sample size. However, as noted by Corral, Font, and Camacho, 2011; Deluca and Corral, 2013, comparing KSD across samples of different size is not necessarily consistent as the KSD simply scales with growing sample size like $\sim 1/n^{0.5}$. Further, in Clauset, Shalizi, and Newman, 2009, no argument was given why this is optimal. In Corral, Font, and Camacho, 2011; Deluca and Corral, 2013, it was shown that the method fails when the distribution has a power tail whose parameter changes from one value

to another. Originally, Hall and Welsh, 1985; Hill, 1975 proposed applying a test for decreasing u, and selecting u at the value before the first value where the test is rejected. In Deluca and Corral, 2013, a similar approach was proposed based on the KS test, where instead one would select the largest sample that could not be rejected, regardless of if rejection occurs at higher thresholds.

These methods can be thought of as outlier tests, where 'lower outliers' are points below the tail threshold u, that are discordant with the tail. However, instead of elaborating on this, a more general automatic approach is recommended: One should fit both the exponential, and a more complicated density to the range of upper samples, and identify the threshold at which the complicated density is not significantly better. If the more complicated density is sufficiently flexible, this should determine that the exponential provides a good approximation and is sufficient to describe the data above the threshold. One could consider comparing nested models with the likelihood ratio test, however this is only a comparison with a specific alternative model. For a more general alternative model, one can use a non-parametric estimator, such as the logspline estimator (available in R:locfit) Kooperberg and Stone, 1991. One can then compare the null with this alternative with the Akaike Information Criterion (AIC). In the presense of clear outliers, one may wish to use estimators that censor, or are robust to the outliers.

Concerning outlier testing, it is useful to estimate the sample fraction to have an idea of where the tail approximation begins to apply. However, tests can often accept a model for a larger sample but reject it in the tail! Thus, one should apply outlier tests for a range of lower thresholds and look for stability in outlier test results for data that do not violate the null. That is, letting $n_u \leq n$ be the size of the largest upper sample that can be defended based on the methods discussed above, an outlier test should be applied to the upper samples consisting of the $n_u, n_{u-1}, ..., 10 + r$ largest points, where r is the expected number of outliers, and where one should certainly not consider samples of size smaller than ten. Consistent identification of outliers in these upper subsamples, where the GPD approximation (4.24) is most relevant, and where the null model cannot be rejected, should be interpreted as a robust result. This algorithm involves $c = n_u - (10 + r)$ consecutive dependent tests, which gives multiple chances for a false positive. However, under the null, the probability of rejecting $c > 1$ consecutive tests, decreases as c increases. Based on simulation studies with the range of models considered within this work, we offer as a rough rule of thumb, that

for a sample of size $10 < n < 100$, one should require a run of $c = n/10$ tests to be rejected to maintain control of the type I error.

4.5 CASE STUDY AND 'DRAGON KINGS'

4.5.1 *Pareto distributions and beyond: the Dragon-King hypothesis*

Outlier detection with exponential underlying distribution has been primarily motivated by reliability engineering applications. Switching perspective from reliability to risk, the exponential of an exponential variable has the (heavy-tailed) Pareto distribution (4.2) that is typically used for modeling extremes in both natural and social sciences: earthquake energies, stock prices, claims in non-life insurance, etc. (Embrechts, Klüppelberg, and Mikosch, 1997; Mitzenmacher, 2004; Newman, 2005; Sornette, 2006).

The Pareto distribution is unique in that it is scale invariant (Dubrulle, Graner, and Sornette, 1998; Lesne and Laguës, 2011), suggesting that events of all sizes – including extremely large ones – are generated by a single mechanism operating at different scales. This feature allows this single parsimonious distribution function to generate a broad range of event sizes. Thus, if a phenomenon is scale invariant, then extreme events are not predictable and there is nothing anomalous about them as there is nothing to distinguish these events from their smaller siblings, other than their resultant size. This reasoning has been advanced to explain the extreme difficulties in forecasting large earthquakes (Geller et al., 1997): according to the approximate scale invariance of the Gutenberg-Richter law, large earthquakes are just earthquakes that started small... and did not stop growing.

However, a number of studies have found either strong or, in other cases, suggestive evidence that there are extreme events 'beyond' the Pareto sample (Sornette, 2009; Sornette and Ouillon, 2012a), i.e., outliers, inspiring the concept of the 'Dragon King' (DK) (Sornette, 2009) event. DK embody a double metaphor implying that an event is both extremely large (a king (Laherrère and Sornette, 1998)), and generated from a unique mechanism/origin (a dragon) relative to other events in the system/sample. The hypothesis advanced in (Sornette, 2009; Sornette and Ouillon, 2012a) is that DK events are generated by a distinct mechanism (e.g., positive feedback) that intermittently amplifies extreme events, leading to the generation of runaway disasters as well as extraordinary opportunities/successes. Due to the uniqueness of such events, there is hope that such extremes may exhibit

precursory signs, disclosing some predictability. The identification of the existence of such phenomena is also clearly important – for example, with applications in risk management. Examples of such DK events have been proposed to include failures of material systems, landslides (Lei et al., 2023) and some large earthquakes in geophysics, financial crashes in economics (Filimonov and Sornette, 2015c; Johansen and Sornette, 1998b; Johansen and Sornette, 2002a), and epileptic seizures and human parturition in biology (Sornette, 2009; Sornette and Ouillon, 2012a). Identifying DKs with convincing statistical significance is a prerequisite to the investigation of their origin, understanding their generating mechanisms, and developing forecasting methods, controls, and resilient system designs. Motivated by these considerations, and to provide pedagogical examples, five case studies are considered where DK events are tested as statistical outliers.

4.5.2 *Financial crashes*

It is well known that crashes in financial markets occur frequently and can have a significant effect not only on market participants, but also on the broader economy. It is often thought that financial markets are unpredictable – i.e., they are scale invariant / fractal (Mandelbrot and Hudson, 2014; Sornette, 2003c) (Pareto distributed). However, in (Filimonov and Sornette, 2015c; Johansen and Sornette, 1998b; Johansen and Sornette, 2002a) it was found that the sample of crash sizes – measured from the peak to the valley of the event (so-called drawdowns) – contained outliers (defined below). However, the statistical test used in (Filimonov and Sornette, 2015c) contains an error in the distribution function of the marginal test statistics, and (Johansen and Sornette, 1998b; Johansen and Sornette, 2002a) did not use standard outlier tests. To correct this, and provide an example, this problem is revisited with the same data. The data are the drawdowns computed for the eleven most actively traded Futures Contracts on the American and European Indices[1], from January 1, 2005 to December 30, 2011.

A peak-to-valley measure of the size of intra-day financial crashes is considered: an *ϵ-drawdown* (hereforth referred to simply as a drawdown) is the total cumulative return of a negative run in price over time, with some specified tolerance for small positive changes along the way (Johansen and

1 US: 1) ES, S&P 500, E-mini; 2) NQ, NASDAQ, E-mini ; 3) DJ, Dow Jones, E-mini. European: 4) AEX, Netherlands; 5) CAC, CAC40, France; 6) DAX, Germany; 7) FTSE, UK; 8) IBEX, Spain; 9) OMX, OMX Stockholm 30, Sweden; 10) SMI, Switzerland; 11) STOXX, Euro STOXX, Europe.

Sornette, 2002a). A *drawup* is its positive counterpart. This is an interesting measure of risk because it captures the transient dependence of price changes in time, whereas studying the unconditional distribution of returns does not. More specifically, considering one trading day $[t_0, t_1]$, prices taken at intervals of width Δ are $p_i = p(t_0 + i\Delta)$, $i = 1, ..., n = \lfloor (t_1 - t_0)/\Delta \rfloor$. The *returns* are then $r_i = \log(p_i/p_{i-1})$. One starts at the first negative return $i_0 = \min\{i : r_i < 0\}$. Then, the cumulative return,

$$r_{i_0,i} = \sum_{j=i_0}^{i} r_j = \log(p_i/p_{i_0}),\ i > i_0\ ,$$

tracks the negative growth of the drawdown, continuing for $i = i_0, i_0 + 1 \ldots$ until the first value of i, say i_2, such that the cumulative return has appeared to reverse direction, relative to its lowest point:

$$r_{i_0,i_2} - \min_{i_0 \le j \le i_2} r_{i_0,j} > \epsilon\sigma\ .$$

Parameter $\epsilon \ge 0$ tunes the tolerance of moves in the opposite direction, and σ is the standard deviation of the returns from the previous trading day. The inclusion of σ makes the tolerance adaptive, which allows for volatility regimes. Finally, stepping backwards from i_2, which is the index of a positive change, the drawdown is defined to have occurred from the start i_0 to the lowest point, which occurs at $i_1 = \text{argmin}_{j \in (i_0 \le j \le i_2)} r_{i_0,j}$. From the next index, $i_1 + 1$, a drawup is defined to begin and computed in a similar way. Drawdowns and drawups alternate in this contiguous way, for the entire trading day.

In panel (a) of Fig. 4.5, for the eight contracts thought to contain an outlier, the largest 5000 drawdowns are plotted according to their empirical CCDF (complementary CDF, i.e., $1 - F(x)$). The empirical CCDF appear approximately linear in the double logarithmic scale, indicating a qualitatively good fit, with the exception of some outliers. There are also some additional differences in the tail. For instance, the tail of the CCDF drops beneath the Pareto fit before crossing back to form the outlying empirical tail. This could suggest an amplification mechanism operating above a threshold size. In panel (c), the Hill plot is given, where the MLE for the tail exponent α is plotted for a range of upper sample sizes. The parameters tend to have an increasing trend, indicating slight convexity in the CCDF in panel (a), and thus a loss of outlier testing power for large sample sizes (Sec. 4.3.4). Based on the Hill plot, the estimator for the top 1'000 points appears to be approximately stable for most of the contracts. For systematic

threshold / sample fraction selection, three methods are used: 1. comparing the AIC of exponential and nonparametric (R:logspline) fits, 2. selecting the smallest KSD as recommended in Clauset, Shalizi, and Newman, 2009, and 3. selecting the smallest threshold where the KS test p-value is above 0.10. The results of these methods are given in Tab. 4.5. All but one of the 24 tests select at least the top 1'000 points, thus upper samples of this size and smaller will be considered for outlier testing.

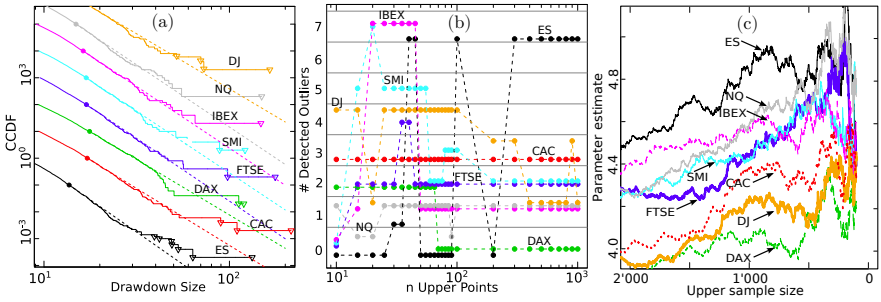

FIGURE 4.5: **Financial Market Crashes. (a):** The 5000 largest drawdowns for each of the 8 futures contracts thought to contain outliers, plotted according to their empirical CCDF in double logarithmic scale. For clarity, each CCDF above the black one is multiplied by 10 relative to the one beneath it. The Pareto distribution function with MLE parameter for the top 500 points is given by the dashed lines, starting at the solid dot. The triangles identify the points that were identified as outliers based on the interpretation of panel (b). **(b)** The number of identified outliers is plotted against sample size where the MRS test (4.3) with level $a = 0.1$ has been applied inward with $m = r = 10$, for a range of sample sizes n, for each contract in (a) with the same colour coding. **(c)** Hill plot: The estimated tail exponent is plotted for a range of upper sample sizes. (see online version for colour)

In panel (a) of Fig. 4.5, the apparent outliers are large and dispersed. Thus, the MRS test statistic (4.3) should be powerful (Sec. 4.3.3) and can be applied inward for a range of thresholds, requiring a fraction of the computation of outward testing. For each dataset, the inward test was performed – with MRS, $m = 10$, level $a = 0.1$, and upper sample size ranging from $n = 10$ to $n = 1000$. For all contracts, excluding AEX, OMX, and STOXX, at least 1 outlier was found and are indicated in Panel (b) of Fig. 4.5. For some of the contracts, the results are quite stable across sample size (e.g., CAC and FTSE). For others, the impurity of the distribution function plays a

Test	ES	CAC	DAX	FTSE	SMI	IBEX	NQ	DJ
AIC	1′184	1′214	2′290	2′734	2′704	2′055	1′154	3′757
KSD	1′049	1′115	1′520	3′144	609	1′501	1′074	1′134
p	1′985	1′323	2′403	3′714	2′701	2′255	3′123	4′000

TABLE 4.5: The selected number of points in the upper sample based on comparing the AIC of the null and the nonparametric model, minimizing the KSD, and the largest upper sample with KS test p value greater than 0.1.

role in the interpretation. For instance, for DAX, two outliers are detected once the test is restricted to the bent-down tail. For ES, choosing between zero and seven outliers is more subjective – are there multiple outliers, or does the tail grow heavier? For IBEX, it is clear that the identification of seven outliers is due to the dip in the empirical CCDF occurring between drawdown size of twenty and thirty. The alternative choice of 1 outlier is more stable with respect to a broad range of values of n. The interpreted outliers are indicated in panel (a).

The largest outliers coincide with major news events: The 07 July 2005 London bombings coincided with the largest outliers of CAC, DAX, FTSE, SMI, and IBEX – all being based on European indices. Further, DAX and CAC each have an outlier corresponding to the 'Mini Flash Crash' of 27 Dec. 2010 (e.g., see (Bundesbank, Accessed on 29-07-2015)). All American contracts (ES, DJ, and NQ) have their largest outliers coinciding with the infamous '2010 Flash Crash' of 6 May 2010. We thus observe that outliers occur either due to some exogenous impacts (London bombings) or as a result of an endogenous transiently unstable dynamics (flash crash). Indeed, in (Filimonov, Bicchetti, et al., 2014; Filimonov and Sornette, 2012b; Wehrli and Sornette, 2022a; Wehrli and Sornette, 2022b; Wehrli, Wheatley, and Sornette, 2021b; Wheatley, Wehrli, and Sornette, 2019a), it was suggested that financial markets exhibit a significant endogeneity or 'reflexivity', in the sense that nowadays up to 70-80% of trades occurring at the time scales of fractions of seconds to tens of minutes are motivated (or triggered) by previous trades. In this framework (Filimonov, Bicchetti, et al., 2014; Filimonov and Sornette, 2012b), Dragon Kings emerge when the market dynamics become critical and super-critical, that is when the future trades are triggered only by previous trades and not by news, making the financial markets essentially self-referential in these periods. Thus, some of the outliers can be classified as Dragon-King drawdowns.

4.5.3 *Stock returns*

An issue of debate is if the 1987 stock market crash (Black Monday) was an outlier. We focus on (Schluter and Trede, 2008), which is the most recent study on this problem. In (Schluter and Trede, 2008), considering daily returns on the Dow Jones Industrial Index, from 3 January 1977 to 31 January 2005, it was claimed that Black Monday is not an outlier. In further detail, the returns were whitened by taking the residuals of a standard AR(1)-GARCH(1,1) model estimated on the returns. Next, the two largest whitened returns $X_{(2)}$ and $X_{(1)}$ were tested as outlying. The test used relies on the GPD approximation (4.23) of the tail of the sample, and requires an estimate of the tail parameter α. A sample size of $n = 732$ was used to estimate α. The test statistic $T_r = X_{(r)}/X_{(r+1)}$, comparing $X_{(r)}$ to the previous (next largest) order statistic $X_{(r+1)}$, was used to test if $X_{(2)}$ and $X_{(1)}$ were outlying. Testing outward, with a level of 0.05, neither of these points were identified as outliers.

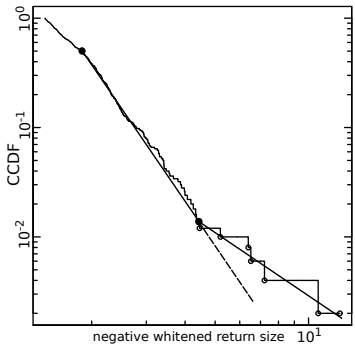

FIGURE 4.6: **Stock Returns:** The rough line provides the empirical CCDF of the magnitude of the 500 largest whitened returns of the Dow Jones Industrial Index from 3 January 1977 to 31 January 2005. The solid lines between solid dots provide Pareto model estimates for two magnitude layers. The dashed line extends the slope of the first layer for comparison with that of the second.

To evaluate the approach taken in (Schluter and Trede, 2008), we first plot the CCDF of the 500 largest negative whitened returns in Fig. 4.6. This plot was not provided in (Schluter and Trede, 2008), but is clearly essential to assessing above which threshold the GPD approximation (4.23) is sound. A

few important points are apparent from the figure: Firstly, the CCDF above the 200 largest observations is shallow/concave, and thus considering more than 200 points (i.e., 732 in (Schluter and Trede, 2008)) in the sample will weaken the test (i.e., the estimate of α will be too small). Secondly, the second largest point is similar in magnitude to the largest. Thus, the test using $T_1 = x_{(1)}/x_{(2)}$ will be masked by $x_{(2)}$, and not rejected. Finally, the top 6 or 7 points seem to follow a heavier tailed distribution. Thus, 6 or 7 points should be tested as outlying, rather than only 2, and a sum test statistic, measuring the cumulative departure of the empirical tail, could be more powerful.

First, we consider estimating a Pareto distribution with two layers. The first layer, containing 193 points, covers $1.97 < x \leq 4.45$ and has MLE $\hat{\alpha}_1 = 3.8$. The second layer, containing the 7 largest points, covers $4.45 < X$ and has MLE $\hat{\alpha}_2 = 1.8$. Given that the first layer model is true, there is a $p = 0.02$ probability of observing such an extreme difference between the estimated parameters. This two layer model appears to describe the empirical CCDF well (Fig. 4.6). Next, a single layer model for the top 200 points, covering $4.45 < X$ was estimated with MLE $\hat{\alpha}_0 = 3.9$. The likelihood ratio test of the two layer versus one layer model is rejected in favour of the two layer with p-value 0.07. Further, applying the SS test for $r = 6$ with the top 200 points rejects that there are no outliers with $p = 0.04$. Finally, applying the DK test for 6 outliers, for upper sample sizes ranging from 20 to 200, all tests have $p < 0.04$. Thus it appears that the 6 largest points are outlying.

The largest one is, unsurprisingly, 'Black Monday' Oct. 19, 1987, which is unambiguously classified as an outlier. An enormous literature has dwelled on its possible origin with a lot of confusion as no simple proximate cause can explain its occurrence. We find more compelling the story that it marked the end of a large financial bubble and thus corresponded to its burst (Johansen and Sornette, 2010c; Sornette, 2003c; Sornette, Johansen, and Bouchaud, 1996). The second largest event occurred on 'Black Friday' Oct. 13, 1989 and is usually associated with a fall of the junk bond market (`https://en.wikipedia.org/wiki/Friday_the_13th_mini-crash`). The third largest loss corresponds to the first day of reopening of the US stock markets on Sept. 17, 2001 after Sept. 11, 2001. It is not clear to us how to interpret the fourth largest loss that happened on Nov. 15, 1991. The fifth largest loss on Oct. 27, 1997 is analyzed in details in (Sornette, 2003c), which paints a picture much richer than the usual story that this was a global stock market crash caused by an economic crisis in Asia. This

loss can actually be seen also as a partial burst of a bubble that had been surging in the few previous years (recall the famous quip on the 'irrational exuberance' of the stock markets by Alan Greenspan, then the Chairman of the US Federal Reserve, on Dec. 5, 1996 (`http://www.federalreserve.gov/boarddocs/speeches/1996/19961205.htm`)). The sixth largest loss on Nov. 9, 1986 is not clearly associated with any exogenous cause, to the best of our knowledge. These six outliers are part of the list found by other researchers (e.g. (Fortune, 1993)).

4.5.4 *Nuclear accidents*

We consider as events accidents occurring at nuclear power plants, studied in (Ayoub and Sornette, 2023; Ayoub, Stankovski, et al., 2021; Wheatley, Sornette, and Sovacool, 2017). For this two measures of severity are considered: the cost measured in 2011 US Dollars, for which there are 173 values over the period of 1960 to 2015; and a logarithmic measure of radiation released called the Nuclear Accident Magnitude Scale (NAMS) (Smythe, 2011), for which there are 33 values over the same period. Since the disaster at Fukushima in 2011, Nuclear power has come under major public scrutiny. Further, the level of risk that the nuclear industry claims is consistently much lower than statistical analysis of past events indicates (Sornette, Maillart, and Kröger, 2013). Thus, it is crucial to arrive at a better understanding of the true risk level in this critical application.

The disasters occurring at Chernobyl (1986) and Fukushima (2011) are the most costly accidents thus far, and together are estimated to have caused damage costing 430 Billion 2011 US dollars. This is roughly equal to five times the cost of all 173 other events together. These events, together with TMI (Three Mile Island, 1979), are also the largest radiation release events. These events are thus extremely large. It is instructive to ask whether a heavy Pareto tail is sufficient to account for these extreme risks or, alternatively, if the tests discussed here can identify outliers / DKs in this data.

In Fig. 4.7 the empirical CCDF (complementary CDF i.e., $Pr\{X > x\}$) for NAMS and the log cost are plotted. For log cost only the 114 events occurring post TMI are included due to an abrupt change in distribution after TMI. For NAMS, the largest three events form a cluster, and appear outlying relative to the exponential distribution with $\hat{\alpha}_{NAMS} = 0.7\ (0.3)$ fit by MLE to the top 15 points. Not surprisingly the distribution of NAMS and the log cost are similar, as they are certainly related. For log cost, the two or three largest events appear to be outlying relative to the exponential

distribution with $\hat{\alpha}_{\$} = 0.6$ (0.14) fit by MLE to the top 50 points. As shown in the Hill plot, inset in Fig. 4.7, when comparing the AIC of the logspline nonparametric fit with the exponential one, the exponential cannot be rejected for samples smaller than the 60 largest points. Further, when performing the KS test, the exponential fit cannot be rejected (at a level of 0.05) for samples smaller than the 80 largest points. Thus the exponential approximation for the tail, and thus the outlier test, should only be applied to not more than the upper 60 points.

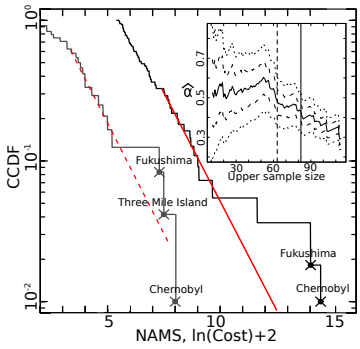

FIGURE 4.7: **Nuclear Power Plant Accidents:** the CCDF (cumulative complementary distribution function, i.e., $Pr\{X > x\}$) of the log of the 56 largest cost events, in millions of US Dollars, (black solid) shifted by 2 units for visibility, and the CCDF of all 33 log radiation release (NAMS) values, in solid grey. The fitted lines are exponential MLE fits. The inner panel is the Hill plot for the cost values. The solid rough line is the MLE for the exponential distribution for the tail of cost events, for multiple upper sample sizes. It is bracketed by lines indicating one and two standard deviations of the estimator. The vertical dashed line indicates the largest sample at which the exponential cannot be rejected (based on AIC) as being as good as the logspline nonparametric fit. The vertical solid line indicates the largest sample at which the exponential cannot be rejected by the KS test.

We now test the outliers with a number of the aforementioned tests. The results are presented in Tab. 4.6 and summarized below. First considering NAMS, in Fig. 4.7 the CCDF is visibly concave until the top 15 points or so, causing a decrease of test power for tests applied to larger sample fractions (Sec. 2.7). Since the outliers are clustered, (from Sections 2.4 and 2.6) the mixture approach is most powerful, and inward tests the weakest.

Despite the small sample size, the mixture test consistently identifies 2 or 3 outliers over a range of upper samples. This confirms that the cluster of large events is a significant feature, however this cluster of large events is not far enough beyond the tail for the other tests to reject the null. It is also important to note that the sample size is very small, and thus our ability to diagnose the validity of the null is weak! Next, cost values are considered for which a larger sample is available. The outward test and the mixture test consistently identify the two largest points as significant outliers. The SRS block test fluctuates around a value of about 0.1. It is not surprising that tests based on the MRS fail to reject due to lack of power when the largest point is not extremely large. Thus Fukushima and Chernobyl appear to be outliers in both radiation released (NAMS) and cost. This is compatible with our understanding of these accidents, where the disaster escalated beyond the threshold of control, leading to an unmitigated proliferation of damage. That these points are outlying in both (dependent) samples would give higher significance if a bi-variate outlier test were performed.

It is worth mentioning that there is a positive relationship between NAMS and cost: Considering the 30 events with substantial radiation release (NAMS>0), a linear regression of the logarithm of cost (the response) versus NAMS (the explanatory variable) yields an intercept of 2.33 (0.7), $p = 0.003$ and a slope of 0.97 (0.17), $p < 10^{-5}$, with coefficient of determination $R^2 = 0.5$. Further, the same regression can be done for the 16 events that have occurred at Sellafield, in the UK. The result of this is an intercept of 2.30 (1.0), $p = 0.04$ and a slope of 1.17 (0.39), $p = 0.001$, with coefficient of determination $R^2 = 0.4$. Thus, there is a significant relationship between radiation release and cost, where we have simply considered a linear relationship. Of course the regression parameters for different plants will depend on the value of property development around the plant.

4.5.5 *Fatalities in Epidemics*

We now study the number of fatalities caused by outbreaks of bacterial, viral, and parasitic diseases (epidemics). A dataset for this, with 1,285 events covering the period from 1900 to 2024, was provided by The International Disaster Database (EM-DAT). The dataset excludes, and in some case provides only national fatalities for, pandemic events (spanning multiple countries). Thus the dataset was complemented with Spanish (1918), Asian (1957), and Hong Kong (1968) influenza pandemics, which each caused in excess of 1 million fatalities (Potter, 2006); the 2009 H1N1 'Swine' influenza

Data	n	$r = m$	MRS	SRS	MS Out	MRS In	Mix	DK
NAMS	20	3	0.62	0.35	0, 0.09 $>$ 0.04	0, 0.62	**2, 0.03**	0.21
NAMS	15	3	0.60	0.32	0, 0.07 $>$ 0.04	0, 0.60	**2, 0.025**	0.20
NAMS	10	3	0.37	0.15	**3, 0.025 $<$ 0.04**	0, 0.37	**3, 0.025**	0.14
Damage	50	2	0.17	**0.08**	**2, 0.03 $<$ 0.06**	0, 0.17	**2, 0.05**	0.18
Damage	40	2	0.23	0.11	**2, 0.04 $<$ 0.06**	0, 0.23	**2, 0.06**	0.22
Damage	20	2	0.25	0.14	**2, 0.05 $<$ 0.055**	0, 0.25	**2, 0.07**	0.25
Damage	15	2	0.17	**0.07**	**2, 0.02 $<$ 0.04**	0, 0.17	**2, 0.03**	0.21
Damage	10	2	**0.06**	**0.02**	**2, 0.01 $<$ 0.04**	**2, 0.01**	0, 0.18	0.16

TABLE 4.6: Summary of outlier tests for NAMS and cost data for the upper n points, for r outliers (with robustness value $m = r$). Bold values indicate significance at a level of $a = 0.1$. Block tests performed include: MRS (4.3), SRS (4.4), mixture likelihood ratio (4.7), and the DK test (4.6). Further, the MS test was applied outward (MS Out), with the number of identified outliers, the p-value, and the adjusted level (to achieve $a = 0.1$) given. For instance, in the first row for MS Out there are zero outliers because the p-value of 0.09 is above the adjusted level of 0.04. Finally, the MRS test was applied inward (MRS In), with the number of identified outliers, and the p-value of the test for the largest point given.

pandemic, which was estimated to cause upwards of 150,000 fatalities (Simonsen et al., 2013); and the COVID-19 pandemic, which has led to more than 7 million fatalities globally according to the World Health Orgnization (WHO), as of the writing of this paper. All epidemics and pandemics will be simply referred to as events. From Panel (a) of Fig. 4.8, it is clear that over time the dataset has become more complete, in particular for small event sizes. Further, in the period from 1900-1980, 15 events had more than 10,000 fatalities (0.19 per year), whereas in the period from 1980-2024, only 2 such event (H1N1 and COVID) occurred (less than 0.05 per year). Notwithstanding potential changes in the true frequency of events, this is obviously a highly significant difference. These historical extreme events – Influenzas, Bubonic plagues, Cholera, etc. – have largely been eradicated through sanitation, vaccines and antibiotics.

Considering the period from 1900 onwards, many changes have occurred that should have influenced both the incidence and severity of events. Due to data incompleteness, the rate of events cannot be studied.

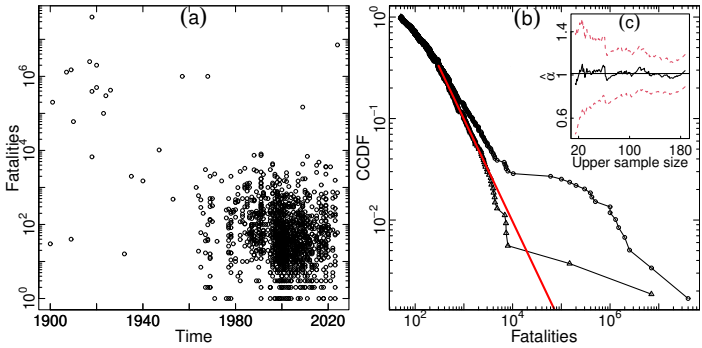

FIGURE 4.8: *Epidemic Fatalities:* **(a)** scatterplot of epidemic fatalities from 1900 to 2024. **(b)** the CCDF of the 593 events in excess of 50 fatalities from 1900-2024 (circles) and the CCDF of the 537 events in excess of 50 fatalities from 1980-2024 (triangles). The latter is fitted with a Pareto tail using a lower threshold $u = 300$, and its MLE $\hat{\alpha} = 1.01$ (0.03) is given by the Hill plot in in panel **(c)**, where the estimate and associated standard deviation are indicated by the horizontal line and the dash lines.

Despite this, the sample in excess of 50 fatalities from 1980 onwards, containing 537 points, is roughly stationary in severity. For instance, when repeatedly (1000 times) sampling 100 points from the 537 points, splitting the 100 points into two equal subsamples, and testing their distributions for equivalence with the KS test, only 10.9% of p-values were less than 0.1. Thus, the modern sample – spanning the 44 years – may be used as a proxy to evaluate the outlyingness of the historical extremes, or at least to evaluate how outlying they would be if they were to occur now.

The events in excess of 50 fatalities from both 1900 onwards and 1980 onwards are plotted according to their CCDF in Panel (b) of Fig. 4.8. The sample from 1980 approximately has a Pareto tail (see Panel (c)) with parameter around 1.01 (0.03) for the 178 points above the lower threshold of 300. With increasing lower truncations, the estimated parameter increases (as the CCDF bends down), however this is not a significant departure from the estimated tail. For instance, the Anderson-Darling test for the fit of the top 178 points gives a p-value of 0.75. The tail of the sample from 1900

onwards is skewed both by the inclusion of historic large events, and also by the absence of their smaller siblings, which were not recorded.

The value of the exponent $\alpha \approx 1$ is reminiscent of Zipf's law, which is known to derive quite robustly from the interplay between three simple ingredients (Saichev, Malevergne, and Sornette, 2009): birth, proportional growth (also known as 'preferential attachment' in network theory) and death. If the variance of the proportional growth component is large, the distribution of event sizes converges to a power law with exponent $\alpha \approx 1$. These ingredients are arguably minimum constituents of epidemic processes and rationalize our finding $\alpha = 1.01\ (0.03)$. What is really surprising is the detection of outliers that we present below, which, in some cases, suggests the activation of strong amplification processes beyond the proportional growth mechanisms.

We turn our attention to the detection of outliers relative to the approximately stationary data from 1980 onwards. The 17 events in excess of 10,000 fatalities – 15 of which happened before 1980 – are considered. The two events closest to the threshold are the Cholera outbreak with 10,276 fatalities (Egypt, 1947) and the pneumonic plague with 60,000 fatalities (China, 1910). Given the smallest event is very close to the threshold, it is treated as a non-outlier for the purposes of this analysis. We begin with the second smallest event, considering as a sample the 176 points with between 300 and 10,000 fatalities occurring since 1980, plus the aforementioned Cholera and pneumonic outbreak. Testing for a single outlier with the DK test (4.6) gives a p-value of 0. Thus any of the other suspected outliers would be identified as significant outliers too. And, including multiple of these outliers in the sample, and testing them together, would provide equally high significance.

With respect to the mechanisms at the origin of these outliers, it is likely that each case may be associated with specific catalysing processes. For one of the largest Dragon Kings, the Spanish flu of 1918, serves as a clear example with an identified amplification mechanism. This epidemic infected 500–600 million people, a third of the world's population at that time, and claimed an estimated 40-50 million lives, about five times the toll of the First World War. The first cases of the unknown disease were registered in Kansas, America, in January 1918. By March 1918, more than 100 soldiers fell ill at the US army camp in Funston, Haskell County, where more than 5000 recruits were training for further military operations on the European battlefronts of the First World War. Most of the recruits were farmers, had regular contact with domestic animals and were less resistant to viruses than recruits from cities. The high concentration of personnel in the camp

simplified human-to-human transmission. At that time, viruses were not known to medicine, and some doctors had not even accepted the idea that microorganisms could cause disease. Later, the personnel of Funston camp were transferred to Europe by ship, and during the long transatlantic crossing, the virus spread among soldiers coming from other parts of the USA. Upon arriving in Europe, American soldiers infected British and French forces, which in their turn infected German forces in hand-to-hand combat. When Woodrow Wilson, President of the United States from 1913 to 1921, began to receive reports about a severe epidemic among American forces, he made no public acknowledgement of the disease (Barry, 2004). Moreover, other governments involved in the war made similar decisions – censorship, lies, and even active propaganda – to keep up morale, allowing the disease to continue to spread without any preventive measures. The pandemic was named 'Spanish flu' because Spain was a neutral country during the First World War and did not suppress the media, so it was only Spanish newspapers that published honest articles about the severity of the disease – despite the fact that it had originated in the USA and spread initially among American soldiers in the absence of a proper response by the US government. This lack of response was probably due to the US strategic goal of developing a strong political influence in the post-WWI peace process that was to shape international politics in the following decades. In summary, the amplification of the pandemic can be attributed to two key mechanisms: the highly efficient transmission facilitated by the movement of soldiers and the absence of any preventive measures due to the war's priority.

Similarly, the COVID-19 pandemic reflects the characteristics of an amplification mechanism, where specific catalyzing factors led to an unprecedented global impact. The virus, SARS-CoV-2, first emerged in Wuhan, China, in late 2019 and rapidly spread worldwide. The pandemic was exacerbated by factors such as global interconnectedness, delays in implementing public health measures, and varying levels of preparedness among countries. By mid-2024, the pandemic had resulted in over 7 million confirmed deaths, with actual figures likely higher due to underreporting and discrepancies in data collection. Economies around the world were severely impacted as measures such as lockdowns and social distancing, designed to contain the virus, led to unprecedented disruptions in global supply chains and labor markets. Unlike the Spanish flu, where war efforts took precedence, the initial underestimation of COVID-19 severity by governments and the slow international response played significant roles

in its spread. The COVID-19 pandemic highlights the potential for future events in the realm of global health, where specific conditions – such as increased travel, urbanization, and environmental degradation – can lead to catastrophic outcomes. This underscores the importance of recognizing and mitigating the factors that contribute to such extreme events, particularly in the context of ongoing threats like antimicrobial resistance and climate change.

We thus conclude that we found evidence of Dragon Kings in the database of epidemic events, with the 2009 swine flu pandemic and the more extreme COVID-19 being notable examples in the more recent post-1980 period. Although the AIDS epidemic was not included in this analysis, it represents another significant outlier in the realm of public health. In 2014, 1.2 million [1 million–1.5 million] people died from AIDS-related illnesses, a marked improvement from the peak in 2015, when 2.3 million [2.1 million–2.6 million] deaths were recorded. Since its identification, an estimated $\sim$ 36 million people have died from AIDS-related causes (UNAIDS, 2012; UNAIDS, 2015). The concept of Dragon King in epidemic dynamics suggests that while such extreme events are rare, they are not beyond the realm of possibility, especially in a world where human activities increasingly intersect with natural processes in unpredictable ways.

4.5.6 *City sizes*

Within the disciplines of economics, geography and geopolitics (among others), the distribution of city and of agglomeration sizes is of particular interest, due to the importance of urban primacy, and because it constitutes one of the key stylized facts. There is a large literature documenting that the distributions of city and agglomeration sizes follows a Pareto distribution with parameter close to one (Zipf's Law) (see e.g. (Saichev, Malevergne, and Sornette, 2009) and references therein). There has been some debate over if the distribution would be better represented by a log-normal (Eeckhout, 2004; Eeckhout, 2009; Levy, 2009), however the debate has been clearly settled in favour of the Pareto for the 1000 largest cities (Malevergne, Pisarenko, and Sornette, 2011). Note that both the Pareto and log-normal are generally taken as result from Gibrat's principle of proportional growth (Gibrat, 1931) (see (Saichev, Malevergne, and Sornette, 2009) for a general derivation).

In (Pisarenko and Sornette, 2012), the DK test (4.6) was used to identify outlying population agglomerations for a number of countries, assuming a

Pareto tail. Here we consider city sizes rather than agglomerations since this data is available for more countries. We only consider agglomeration sizes for the case of Paris, France for comparison with (Pisarenko and Sornette, 2012). Data for 14 large countries[2] were taken from *CITY POPULATION: Population Statistics for Countries, Administrative Areas, Cities and Agglomerations - Interactive Maps and Charts* Accessed on 01-01-2015. All tests use the SRS block test statistic for testing the largest point as an outlier, with the exception or Russia where two outliers are tested.

In Fig. 4.9, the 35 largest cities of each country are plotted according to their empirical CCDF, rescaled in a way to make the largest cities comparable. Since not all of the samples appear to follow a pure Pareto, results on robustness and testing the tail (Sections 4.3.4 and 4.4) are relevant here.

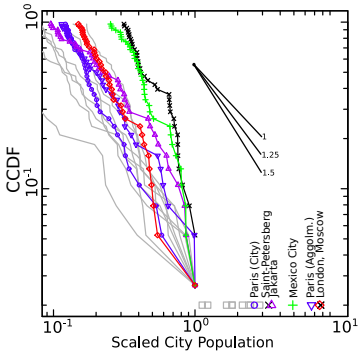

FIGURE 4.9: **City sizes**: plot of the CCDF for the 35 largest cities (and also agglomerations for France) in each of the 14 countries: Brazil, China, France, India, Indonesia, Japan, Korea, Mexico, Nigeria, Pakistan, Philippines, Russia, the UK, and the USA. The sizes were scaled such that the second largest point (third largest for Russia) in each country has size 1. The scaled largest point (two largest for Russia) are plotted in the bottom right. Each country that is suspected of having outliers is in colour: France (blue circles for cities, blue downward triangles for agglomerations), Russia (black x marks), Indonesia (purple triangles), Mexico (green crosses), and England (red squares).

First considering French cities, for upper sample sizes of $5 < n \leq 35$, the p-value fluctuates in a range of $0.1 - 0.2$. Thus, there is only marginal

2 Brasil, China, France, India, Indonesia, Japan, Korea, Mexico, Nigeria, Pakistan, Phillipines, Russia, the UK, and the USA.

evidence that the city of Paris is an outlier. However, the agglomeration of Paris is relatively larger, and for $5 < n \leq 25$ the p-value fluctuates between 0.02 and 0.15, providing stronger evidence of the uniqueness of Paris. The CCDF of Indonesia is concave. Thus, if too large of a sample is considered in the test, Jakarta will not be detected as an outlier. For instance, if one draws a line that best interpolates all points of the empirical CCDF, the line will be so shallow that the Jakarta point falls beneath it, essentially masking the outlier. For this reason, Jakarta, Indonesia has $p < 0.1$ only for the upper most points $5 < n < 11$. Mexico is an even more extreme case of the above, having $p < 0.1$ for $5 < n < 20$ for Mexico City. London, UK, is the most significant, having $0.001 < p < 0.05$ for all $5 < n \leq 35$. Finally, testing both Moscow and Saint-Petersburg as outliers, the p-value is in $0.01 < p < 0.15$, with a mean of 0.09 for all $5 < n \leq 35$. In conclusion, it is absolutely clear that London is an outlier, and the largest city/cities of five of the remaining fourteen countries considered have moderate/suggestive evidence that they are outlying. A plausible mechanism for the outlier status of London (and other cities) can be attributed to the positive feedback loops that have bolstered the outsized importance of these imperial power centers over the past centuries and decades, coupled with their self-reinforcing economic attractiveness.

4.6 DISCUSSION

We have provided a comprehensive study of outlier detection in the highly general case of samples with exponential and Pareto tails. By considering a variety of test statistics and outlier scenarios, many useful insights are made available to for practitioners. Further, a simple yet novel modification of classical test statistics was shown to make the convenient inward test competitive with the relatively arduous outward test.

Insights include that one should select the correct test statistics based on the nature of the suspected outliers. For instance, a mixture model can be very useful for clustered outliers, whereas an inward test with a MS type statistic will be powerless. Next, the power and level of outlier tests are highly sensitive to the correct specification of the null. For robust results, it may be better to focus on the tail of the sample, where EVT provides that the best approximation is attained. If the approximation is poor even in the tail, then one should choose a better null model to avoid spurious inference. Further, tests should be applied for a range of upper samples

sample (growing lower threshold) and consistent rejection required for a robust rejection to be verified.

In the case studies, the concept of Dragon King events was introduced. This stresses that some outliers are meaningful, and perhaps special. Further, one should certainly not simply discard these outliers but rather focus on understanding them. Significant outliers were found in the sizes of financial returns and crashes, epidemic fatalities, nuclear power generation accidents, and city sizes within countries. In the cases of financial crashes and nuclear accidents, the existence of Dragon Kings should be considered in the assessment of risk.

ACKNOWLEDGEMENT

We thank Doctor Spencer Wheatley for his contribution to the preliminary version of this paper. His assistance in the initial analysis especially on the case studies was crucial for the work.

REFERENCES

Aggarwal, C. (2013). "Outlier analysis". In: *Springer Science & Business Media* 12.

Aitkin, M. and Wilson, G. T. (1980). "Mixture models, outliers, and the EM algorithm". In: *Technometrics* 22.3, 325.

Ayoub, A. and Sornette, D. (2023). "The Power of Precursors: An Empirical Assessment of Nuclear Power Risks". In: *Progress in Nuclear Energy* 164.104878, 1.

Ayoub, A., Stankovski, A., Kröger, W., and Sornette, D (2021). "Precursors and startling lessons: Statistical analysis of 1250 events with safety significance from the civil nuclear sector". In: *Reliability Engineering and System Safety* 214.107820.

Balakrishnan, K. (1996). "Exponential distribution: theory, methods and applications". In: *CRC press*, 228.

Balasooriya, U. and Gadag, V. (1994). "Tests for upper outliers in the two-parameter exponential distribution". In: *Journal of statistical computation and simulation* 50.3-4, 249.

Barnett, V. and Lewis, T. (1994). "Outliers in Statistical Data. 3rd ed." In: *John Wiley*, 285.

Barry, J. M. (2004). "The Great Influenza: The Epic Story of the Deadliest Plague in History". In: *New York, Penguin*.

Bauke, H. (2007). "Parameter estimation for power-law distributions by maximum likelihood methods". In: *The European Physical Journal B* 58.2, 167.

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2006). *Statistics of extremes: theory and applications*. John Wiley & Sons.

Bundesbank, Deutsche (Accessed on 29-07-2015). *High-frequency trading and market implications*.

Chikkagoudar, M. S. and Kunchur, S. H. (1983). "Distributions of test statistics for multiple outliers in exponential samples". In: *Communications in Statistics—Theory and Methods* 12, 2127.

CITY POPULATION: Population Statistics for Countries, Administrative Areas, Cities and Agglomerations - Interactive Maps and Charts (Accessed on 01-01-2015). `https://www.citypopulation.de`.

Clauset, A., Shalizi, C. R., and Newman, M. E. J. (2009). "Power-law distributions in empirical data". In: *SIAM review* 51.4, 661.

Cohen, M. (1992). "Epidemiology of drug resistance: implications for a post—antimicrobial era". In: *Science* 257.5073, 1050.

Corral, A., Font, F., and Camacho, J. (2011). "Noncharacteristic half-lives in radioactive decay". In: *Physical Review E* 83.6, 066103.

Danielsson, J., Haan, L. de, Peng, L., and Vries, C. G. de (2001). "Using a bootstrap method to choose the sample fraction in tail index estimation". In: *Journal of Multivariate analysis* 76.2, 226.

Davies, J. A. (2002). "The individual success of musicians, like that of physicists, follows a stretched exponential distribution". In: *The European Physical Journal B-Condensed Matter and Complex Systems* 27.4, 445.

Deemer, W. and Votaw, D. (1955). "Estimation of parameters of truncated or censored exponential distributions". In: *The Annals of Mathematical Statistics, 498-504*.

Deluca, A. and Corral, Á. (2013). "Fitting and goodness-of-fit test of non-truncated and truncated power-law distributions". In: *Acta Geophysica* 61.6, 1351.

Dixon, W. J. (1950). "Analysis of Extreme Values". In: *The Annals of Mathematical Statistics*, 488.

Dubrulle, B., Graner, F., and Sornette, D. (1998). "Scale Invariance and Beyond". In: *Les Houches Workshop, March 10-14, 1997 (Centre de Physique des Houches), Springer*.

Eeckhout, J. (2004). "Gibrat's law for (all) cities". In: *American Economic Review* 94, 1429.

Eeckhout, J. (2009). "Gibrat's law for (all) cities: Reply". In: *American Economic Review* 99, 1676.

Feller, W. (1971). "An Introduction to Probability Theory and its Applications". In: *vol. II (Wiley, New York)*.

Filimonov, V., Bicchetti, D., Maystre, N., and Sornette, D. (2014). "Quantification of the High Level of Endogeneity and of Structural Regime Shifts in Commodity Prices". In: *The Journal of International Money and Finance* 42.5, 174.

Filimonov, V. and Sornette, D. (2012b). "Quantifying reflexivity in financial markets: Toward a prediction of flash crashes". In: *Physical Review E* 85.5, 056108.

Filimonov, V. and Sornette, D. (2015c). "Power law scaling and "Dragon-Kings" in distributions of intraday financial drawdowns". In: *Chaos, Solitons & Fractals* 74, 27.

Fortune, P. (1993). "Stock market crashes: what have we learned from October 1987". In: *New England Economic Review* March/April, 3.

Geller, R. J., Jackson, Kagan, Y. Y., and Mulargia, F. (1997). "Cannot earthquakes be predicted?" In: *Science* 278.5337, 487.

Gibrat, R. (1931). "Les Inégalités Economiques; Applications aux inégalitiés des richesses, à la concentration des entreprises, aux populations des villes, aux statistiques des familles, etc., d'une loi nouvelle, la loi de l'effet proportionnel". In: *American Economic Review)* Paris: Librarie du Recueil Sirey.

Gomes, M. I. and Oliveira, O. (2001). "The bootstrap methodology in statistics of extremes—choice of the optimal sample fraction". In: *Extremes* 4.4, 331.

Guha-Sapir, D., Below, R., and Hoyois, Ph. (Accessed on 02-07-2015). *EM-DAT: The CRED/OFDA International Disaster Database, www.emdat.be, Université Catholique de Louvain, Brussels, Belgium*. www.emdat.be.

Hall, P. and Welsh, A. H. (1985). "Adaptive estimates of parameters of regular variation". In: *The Annals of Statistics* 13.1, 331.

Hawkins, D. M. (1980). "Identification of outliers. Vol. 11". In: *Chapman and Hall*.

Hill, B. M. (1975). "A simple general approach to inference about the tail of a distribution". In: *The annals of statistics* 3.5, 1163.

Hodge, V. J. and Austin, J. (2004). "A survey of outlier detection methodologies". In: *Artificial Intelligence Review* 22.2, 85.

Iglewicz, B. and Martinez, J. (1982). "Outlier detection using robust measures of scale". In: *Journal of Statistical Computation and Simulation* 15.4, 285.

Johansen, A. and Sornette, D. (1998b). "Stock market crashes are outliers". In: *European Physical Journal B* 1, 141.

Johansen, A. and Sornette, D. (2002a). "Large Stock Market Price Drawdowns Are Outliers". In: *Journal of Risk* 4.2. Winter 2001/02, 69.

Johansen, A. and Sornette, D. (2010c). "Shocks, crashes and bubbles in financial markets". In: *Brussels Economic Review (Cahiers Economiques de Bruxelles)* 53.2, 201.

Kooperberg, C. and Stone, C. J. (1991). "A study of logspline density estimation". In: *Computational Statistics & Data Analysis* 12.3, 327.

Laherrère, J. and Sornette, D. (1998). "Stretched exponential distributions in nature and economy: "fat tails" with characteristic scales". In: *The European Physical Journal B-Condensed Matter and Complex Systems* 2.4, 525.

Lalitha, S. and Kumar, N. (2012). "Multiple outlier test for upper outliers in an exponential sample". In: *Journal of Applied Statistics* 39.6, 1323.

Lei, Q., Sornette, D., Yang, H., and Loew, S. (2023). "Real-Time Forecast of Catastrophic Landslides via Dragon-King Detection". In: *Geophysical Research Letters* 50. e2022GL100832, 1.

Lesne, A. and Laguës, M. (2011). "Scale Invariance: From Phase Transitions to Turbulence". In: *Springer (2012 edition)*.

Levy, M. (2009). "Gibrat's law for (all) cities: A comment". In: *American Economic Review* 99, 1672.

Likeš, J. (1967). "Distribution of Dixon's statistics in the case of an exponential population". In: *Metrika* 11.1, 46.

Lin, C. T. and Balakrishnan, N. (2009). "Exact computation of the null distribution of a test for multiple outliers in an exponential sample". In: *Computational Statistics & Data Analysis* 53.9, 3281.

Lin, C.T. and Balakrishnan, N. (2014). "Tests for Multiple Outliers in an Exponential Sample". In: *Communications in Statistics – Simulation and Computation* 43.4, 706.

Longford, N. T. (2013). "Searching for contaminants". In: *Journal of Applied Statistics* 40.9, 2010.

Malevergne, Y., Pisarenko, V., and Sornette, D. (2011). "Testing the Pareto against the lognormal distributions with the uniformly most powerful unbiased test applied to the distribution of cities". In: *Physical Review E* 83.036111.

Mandelbrot, B. and Hudson, R. (2014). "The Misbehavior of Markets: A fractal view of financial turbulence". In: *Basic Books*.

Mitzenmacher, M. (2004). "A brief history of generative models for power law and lognormal distributions". In: *Internet Mathematics* 1.2, 226.

Newman, M. E. J. (2005). " Power laws, Pareto distributions and Zipf's law". In: *Contemp. Physics* 46.5, 323.

Pisarenko, V. and Sornette, D. (2012). "Robust statistical tests of Dragon-Kings beyond power law distributions". In: *The European Physical Journal-Special Topics* 205.1, 95.

Potter, C. W. (2006). "A History of Influenza". In: *J Appl Microbiol.* 91.4, 572.

Redner, R. and Walker, H. (1984). "Mixture densities, maximum likelihood and the EM algorithm". In: *SIAM review* 26.2, 195.

Rényi, A. (1953). "On the theory of order statistics". In: *Acta Mathematica Hungarica* 4.3, 191.

Rosner, B. (1975). "On the detection of many outliers". In: *Technometrics* 17.2, 221.

Saichev, A., Malevergne, Y., and Sornette, D. (2009). "Theory of Zipf's law and beyond". In: *Lecture Notes in Economics and Mathematical Systems* 632, Springer, ISBN: 978-3-642-02945.

Schluter, C. and Trede, M. (2008). "Identifying multiple outliers in heavy-tailed distributions with an application to market crashes". In: *Journal of Empirical Finance* 15.4, 700.

Simonsen, L., Spreeuwenberg, P., Lustig, R., Taylor, R.J., Fleming, D.M., Kroneman, M., Van Kerkhove, M.D., Mounts, A.W., and Paget, W.J. (2013). "(the GLaMOR Collaborating Teams), Global mortality estimates for the 2009 influenza pandemic from the GLaMOR project: a modeling study". In: *PLOS Medicine* 10.11, e1001558.

Smythe, D (2011). "An objective nuclear accident magnitude scale for quantification of severe and catastrophic events". In: *Physics Today: Points of View*.

Sornette, D. (2003c). *Why Stock Markets Crash: Critical Events in Complex Financial Systems*. Re-edited 2017 with new preface. Princeton University Press.

Sornette, D. (2006). "Critical phenomena in natural sciences: chaos, fractals, selforganization and disorder: concepts and tools". In: *Springer Science & Business*.

Sornette, D. (2009). "Dragon-Kings, Black Swans and the Prediction of Crises". In: *International Journal of Terraspace Science and Engineering)* 2.1, 1.

Sornette, D., Johansen, A., and Bouchaud, J.-P. (1996). "Stock market crashes, Precursors and Replicas". In: *J.Phys.I France* 6.1, 167.

Sornette, D., Maillart, T., and Kröger, W. (2013). "Exploring the limits of safety analysis in complex technological systems". In: *International Journal of Disaster Risk Reduction* 9, 59.

Sornette, D. and Ouillon, G. (2012a). "Dragon-kings: mechanisms, statistical methods and empirical evidence". In: *Eur. Phys. J. Special Topics* 5.1, 1.

Sornette, D. and Ouillon, G. (2012b). "editors of the special issue on "Discussion and debate: from black swans to dragon-kings – Is there life beyond power laws?"" In: *Eur. Phys. J. Special Topics* 5.1, 1.

UNAIDS (2012). *Report on the global AIDS epidemic 2012*. https://www.unaids.org/sites/default/files/media_asset/20121120_UNAIDS_Global_Report_2012_with_annexes_en_1.pdf. Accessed on 29-07-2015.

UNAIDS (2015). *MDG 6: 15 years, 15 lessons of hope from the AIDS response*. https://www.unaids.org/sites/default/files/media_asset/MDG6Report_en.pdf. Accessed on 29-07-2015.

Verdinelli, I. and Wasserman, L. (1991). "Bayesian analysis of outlier problems using the Gibbs sampler". In: *Statistics and Computing* 1.2, 105.

Virkar, Y. and Clauset, A. (2014). "Power-law distributions in binned empirical data". In: *The Annals of Applied Statistics* 8.1, 89.

Wehrli, A. and Sornette, D. (2022a). "Classification of flash crashes using the Hawkes(p,q) framework". In: *Quantitative Finance* 22.2, 213.

Wehrli, A. and Sornette, D. (2022b). "The excess volatility puzzle explained by financial noise amplification from endogenous feedbacks". In: *Scientific Reports* 12. 1-13, 18895.

Wehrli, A., Wheatley, S., and Sornette, D. (2021b). "Scale-, time- and asset-dependence of Hawkes process parameters estimated on high frequency price change data". In: *Quantitative Finance* 21.5, 729.

Wheatley, S, Wehrli, A., and Sornette, D. (2019a). "The endo-exo problem in high frequency financial price fluctuations and rejecting criticality". In: *Quantitative Finance* 19.7, 1165.

Wheatley, S., Sornette, D., and Sovacool, B. (2017). "Of Disasters and Dragon Kings: A Statistical Analysis of Nuclear Power Incidents & Accidents". In: *Risk Analysis* 37.1, 99.

5

CONCLUSION AND OUTLOOK

The thesis has explored into both econometric modeling and statistical analysis that are developed to study the dynamics of financial bubbles and extreme market events.

In terms of econometric modeling, two models from a novel class of models are presented, both centered around the heuristic approach investors develop due to large uncertainties in determining fundamental values. Recognizing the challenge of accurately assessing fundamentals, investors estimate mispricing by comparing the current price to a long-term average growth rate and shape investment decisions based on the mispricing level. The first model captures this by determining the crash hazard rate through a non-local estimation of mispricing expressed in excess returns. Building upon this foundation, the second model incorporates elements of self-excitation in the crash hazard rate. The models with such crash hazard rates also incorporate positive feedback loops associated with the fact that a larger mispricing leads to a larger crash hazard rate which itself is compensated by a larger expected return as a compensation for the risks incurred by investors.

We have documented that the models produce synthetic price time series that exhibit structures and properties that are comparable to those of real financial time series, with also the existence of transient bubbles looking similar to real financial bubbles with transient faster-than-exponential growth of the price. Rather than crashing in a single jump, our models account for multiple jumps in the first model, and contagious clustered jumps in the second model. We have also quantified the performance of the parameter estimation for the synthetic times series, with a good recovery of the true parameters and well-understood standard deviations of the estimated parameters.

Reflecting on our initial motivation to incorporate non-local self-referencing crash hazard rate and the further self-excitations in the model, it was primarily to capture the transient dynamics and 'abnormality' of market bubbles. One of the essential aspects of market dynamics during a bubble is the increasing vulnerability as mispricing escalates before corrections. However, this growing vulnerability is not fully expressed in the current model set-ups. This is largely due to the static parameters used for non-local mis-

pricing, which is also a direct cause of the difficulty in precisely estimating other parameters (e.g. ν as explained in section 3.5.1). To address this, there are two ingredients that can aid for future model iterations:

1. A short-term mispricing calculation over e.g. 1-2 weeks, in addition to the existing long-term mispricing.

2. A dependency where μ_t is non-zero only for negative values of the short-term mispricing (loss side), and increases as losses accumulate.

This transient μ_t, consist of both a long-term measure and a short-term trigger, prevents non-bubbles regimes from averaging out the effects that are active only during bubble regimes. It will provide a more accurate and responsive tool for risk management and financial analysis.

In terms of statistical analysis, we have provided a comprehensive study of outlier detection in the highly general case of samples with exponential and Pareto tails. Specifically, the Max-Robust-Sum (MRS) and Sum-Robust-Sum (SRS) statistics have been adapted to enhance sensitivity to outlier in a variety of scenarios. Empirical validations across various data types demonstrate superior performance over conventional methods, especially in complex outlier scenarios. Furthermore, the simple yet novel modifications of test statistics are shown to make the convenient inward test competitive with the relatively arduous outward test. The statistics and testing methods presented in the study offer useful tools for practitioners, facilitating more accurate and efficient analysis of data anomalies in diverse settings.

To further enhance the models introduced in this thesis, particularly in the areas of parameter estimation, model optimization, and prediction, machine learning offers a way to handle real-world scenarios more effectively. On one hand, neural networks such as deep feedforward or recurrent networks can be employed to improve parameter estimation by uncovering non-linear relationships and interactions in the data, resulting in more accurate and reliable parameter optimization. On the other hand, traditional econometric models are often calibrated using fixed datasets and parameterization that limits their ability to adapt to new information, whereas machine learning can be applied to implement adaptive calibration, allowing the models to adapt in real time to changing market conditions, ensuring that the models remain accurate and relevant.