DIPARTIMENTO MATEMATICA
DIPARTIMENTO DI MATEMATICA "TULLIO LEVI-CIVITA"

UNIVERSITÀ DEGLI STUDI DI PADOVA

SEDE AMMINISTRATIVA: UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI MATEMATICA "TULLIO LEVI-CIVITA"

CORSO DI DOTTORATO DI RICERCA IN SCIENZE MATEMATICHE
CURRICULUM MATEMATICA COMPUTAZIONALE
XXXVIII CICLO

Hawkes and Affine Processes in Risk Modeling: Theory and Applications in Finance and Cybersecurity

COORDINATORE
Prof. Giovanni Colombo
Università degli Studi di Padova

SUPERVISORE
Prof. Giorgia Callegaro
Università degli Studi di Padova

CO-SUPERVISORE
Prof. Claudio Fontana
Università degli Studi di Padova

DOTTORANDA
Beatrice Ongarato
Matricola 2082878

ANNO ACCADEMICO
2024-2025

Abstract

The following Ph.D. thesis consists of three chapters and explores various applications of Hawkes processes and affine models within the field of risk modeling. Each chapter corresponds to a distinct paper.

In the first project, we address an hedging problem, specifically we look for a *semi-static variance-optimal* strategy. We minimize the variance of the hedging error, combining static and dynamic positions in different market instruments. The problem is analyzed in an affine modeling framework, featuring stochastic volatility and self-exciting jumps in the log-price. The optimal strategy is characterized analytically through multi-dimensional complex integrals and computed numerically. We also perform a parameter sensitivity analysis and examine the impact of incorporating jumps on the hedging error.

The second work focuses on a stochastic control problem applied to cyber-risk mitigation. A continuous-time stochastic model is developed, incorporating Hawkes processes to describe the arrival of cyberattacks targeting a specific entity. We formulate a control problem which is solved via dynamic programming, and we determine the optimal investment strategy. We then perform numerical experiments to highlight the role that attack modeling plays in determining the optimal response and resource allocation.

The third chapter addresses a theoretical problem. Given an affine process under a certain probability measure, we characterize the family of all stable measure transformations that preserve the affine structure of the process. This theoretical insight is fundamental for applications such as pricing and risk management, ensuring that the affine properties are maintained under different probability measures.

Sommario

La seguente tesi di dottorato è composta da tre capitoli ed esplora diverse applicazioni dei processi di Hawkes e dei modelli affini nell'ambito della modellizzazione del rischio. Ad ogni capitolo corrisponde un articolo scientifico.

Nel primo progetto affrontiamo un problema di copertura ottimale, in particolare studiamo una strategia *semi-static variance-optimal*. Minimizziamo la varianza dell'errore di copertura, combinando posizioni statiche e dinamiche in diversi strumenti di mercato. Il problema viene analizzato nel caso di un modello affine che presenta volatilità stocastica e salti autoeccitanti nel prezzo. La strategia ottimale è caratterizzata analiticamente attraverso integrali multipli su domini complessi e calcolata numericamente. Eseguiamo anche un'analisi di sensibilità dei parametri ed esaminiamo l'impatto dell'incorporazione dei salti sull'errore di copertura.

Il secondo lavoro si incentra su un problema di controllo stocastico applicato al contenimento del rischio informatico. Sviluppiamo un modello stocastico a tempo continuo, che utilizza i processi di Hawkes per descrivere l'arrivo degli attacchi informatici diretti ad una specifica entità. Formuliamo un problema di controllo che risolviamo tramite programmazione dinamica, determinando la strategia di investimento ottimale. Eseguiamo poi alcuni esperimenti numerici per evidenziare il ruolo che la modellazione degli attacchi svolge nel determinare la risposta ottimale e l'allocazione delle risorse.

Nel terzo capitolo affrontiamo un problema teorico. Definito un processo affine sotto una certa misura di probabilità, caratterizziamo la famiglia di cambiamenti di misura che preservano la struttura affine del dato processo. Questa analisi è fondamentale per applicazioni quali il *pricing* e la gestione del rischio, garantendo che le proprietà affini siano mantenute sotto varie misure di probabilità.

Contents

List of Figures

List of Tables

Introduction

Quantitative finance researchers often face a natural question: *what happens when we introduce jumps?* This consideration reflects a broader challenge in stochastic modeling: diffusion processes, despite their elegance and mathematical tractability, are often unable to capture the complexity of real-world systems.

Poisson processes constitute the most elementary tools for modeling discontinuities, describing the occurrence of independent events under the assumption of a constant arrival rate. Combined with diffusion, in a simple jump-diffusion setting, they can reproduce several realistic market features. Nevertheless, they remain too simplistic, as they cannot incorporate the dependence between event arrivals and the clustering that may arise in empirical observations. These features emerge in different contexts and cannot be overlooked: financial markets might face cascades of defaults during a crisis, insurers may experience waves of claims after a natural disaster, and cyberattacks often occur in bursts due to the interconnection of digital infrastructures. Capturing such behaviors requires richer classes of processes.

Hawkes processes naturally extend the Poisson framework by allowing the probability of a new event occurring to depend on past arrivals, making them suitable to describe self-exciting effects. In a more general setting, one can rely on affine processes to reproduce analogous features. Their mathematical structure allows jump intensities to be stochastic and state-dependent, making them suitable for modeling dependencies across different risk factors. Clearly, this increased realism comes at a cost: moving from purely continuous to jump-driven dynamics, and from deterministic to stochastic intensities, entails significant mathematical challenges, requiring the development of advanced analytical and computational techniques.

In this thesis, we employ Hawkes and affine processes to develop a risk modeling framework which incorporates jumps and cross-excitation between factors. The dissertation is organized in three chapters, each one addressing a distinct problem.

In Chapter 1, we tackle a hedging problem in an incomplete affine market model with self-exciting jumps. We compute the *semi-static variance optimal* strategy in this framework and analyze how the introduction of contagion-type jumps affect the hedging error.

Chapter 2 develops a stochastic setting for modeling cyberattacks using Hawkes processes. Within this setting, we address an optimal control problem aimed at quantifying the optimal cybersecurity investment when attacks arrive in clusters.

In Chapter 3, we investigate a more theoretical subject. We study probability measure transformations for affine models, establishing necessary and sufficient conditions under which the affine property is preserved.

Before delving into the three research questions addressed in this thesis, we provide an overview of Hawkes and affine processes.

A brief tour on Hawkes and affine processes

As mentioned in the title, the two central classes of stochastic processes employed throughout the thesis are Hawkes and affine processes. In this section, we present the main concepts underlying these mathematical objects, explaining their connections, and providing an overview of their key properties and applications.

Hawkes processes Hawkes processes, introduced in the seminal work of Hawkes (1971), are *self-exciting* point processes. They are counting processes characterized by a stochastic intensity, which represents the instantaneous rate at which new events occur given the past. Unlike standard Poisson processes, where events occur independently at a constant rate, Hawkes processes capture interdependent arrivals: each occurrence of an event increases the likelihood of future events, leading to *clustered events*. While several processes have stochastic intensities, such as Cox processes in D. R. Cox (1955), Hawkes explicitly model endogenous self-excitation through their structure.

Mathematically, we denote by $(N_t)_{t\geq 0}$ the counting process, where N_t represents the total number of event arrivals in a system up to time t, and by $(T_n)_{n\geq 1}$ the jump times. The Hawkes intensity process is defined by

$$\lambda_t = \lambda_0 + \sum_{n=1}^{N_t} K(t - T_n) \quad \text{for all } t \geq 0,$$

where $\lambda_0 > 0$ is the baseline intensity and $K\colon [0,\infty) \to [0,\infty)$ is the excitation kernel. Each jump of N increases the intensity, with size and persistence depending on K. The choice of the kernel K is crucial on determining the properties of the process. For a general kernel K, Hawkes processes are non-Markovian, as the intensity can depend on the entire past history. In this thesis, as usual in the applications of Hawkes processes, we focus on the exponential kernel case, $K(t) = \xi e^{-\beta t}$. This particular choice ensures the Markov property and will be central to our analysis. The memoryless property of exponential Hawkes processes makes them a natural choice for the applied problems addressed in the first two chapters. Moreover, only with this kernel choice Hawkes processes can be embedded in the class of affine processes, which are analyzed in Chapter 3. Extensions such as marked and multi-dimensional Hawkes processes (see Laub et al. (2021, Section 3.6) for an overview) exist and are discussed in the following chapters of the thesis.

Due to their self-exciting nature, Hawkes processes have been applied in a wide range of fields. In seismology, they capture aftershocks following an earthquake, see Ogata (1978); in neuroscience, they describe the neuronal spike train activity, see Reynaud-Bouret et al. (2013); in insurance, they simulate the arrival of aggregated claims, see Stabile and Torrisi (2010). In this dissertation, we concentrate on applications in finance and cyber-risk, respectively in the first and second chapter. Chapter 1 applies Hawkes processes to model contagion in the market price, similarly to Aït-Sahalia et al. (2015) and Filimonov et al. (2014). The focus is on the implications of jump incorporation in risk assessment. Beyond this specific case, Hawkes processes have wide applicability in finance, including high-frequency trading, market microstructure modeling, and contagion mechanisms underlying credit defaults (an overview is provided in Bacry et al. (2015)). Chapter 2 concentrates on cyber-risk, employing Hawkes processes to describe the arrival of cyberattacks. Cyber threats often arrive in bursts and propagate through networks in a self-exciting manner, making Hawkes processes well-suited to model attack dynamics, see Baldwin et al. (2017), and Bessy-Roland et al. (2021). We study how this cluster modeling influences the optimal investment strategy in cybersecurity. The third chapter is mainly theoretical, so it does not tackle any specific application, but the obtained results can be exploited in all the above mentioned applicative fields.

Affine processes General affine processes were first introduced in Duffie, Filipović et al. (2003), and are a well-established class of stochastic processes in probability and mathematical finance. They are Markov processes whose characteristic function is exponentially affine in the initial state. Mathematically, this property can be expressed as follows. Let $(X_t)_{t\geq 0}$ be an affine process taking values in a suitable state space. Then, the conditional expectation of the exponential transform of X_t admits the closed-form representation

$$\mathbb{E}\left[e^{\langle u, X_t\rangle} | X_0 = x\right] = \exp\left(\phi(t,u) + \langle \psi(t,u), x\rangle\right),$$

where u is a complex vector in the domain of definition of the transform, $\langle \cdot, \cdot \rangle$ is an inner product, and ϕ, ψ are deterministic functions solving a system of generalized Riccati ordinary differential equations. This elegant mathematical structure simplifies the computation of many crucial quantities, such as moments, in a theoretical setting, or option and bond prices, from an applied point of view.

A major strength of affine processes lies on their ability to unify different models within one theoretical framework. They can incorporate a wide range of stylized features such stochastic volatility, mean reversion, heavy tails and jumps. Due to our interest on discontinuous models, it is worth highlighting that affine processes potentially reproduce quite sophisticated jumps behavior: they allow for finite, but also for infinite intensity, simultaneous jumps in multiple components, and stochastic intensities. In particular, the intensity can depend in an affine way on any non-negative component, allowing for cross-excitement and self-excitement effects between factors. Exponential Hawkes processes can be embedded into the class of affine models, as they can be interpreted as point processes with affine drift and intensity which is linearly dependent on the process itself.

Affine processes have emerged as highly attractive class of processes across a wide range of fields, offering a balance between flexible modeling and analytical tractability. Interestingly, even before their formal definition in Duffie, Filipović et al. (2003), many widely used models were in fact affine in nature. For instance, the CBI (continuously branching with immigration) processes introduced by Kawazu and Watanabe (1971) belong to the affine class. Similarly, classical interest rate models, such as Vasicek (1977) and J. C. Cox et al. (1985) are special instances of affine processes. The same holds for major financial models, such as Black-Scholes, Heston (1993) and Bates (1996). Subsequently, affine processes have been also employed in credit risk, see e.g. intensity-based models in Duffie (2005), and in longevity and mortality risk modeling, see Biffis (2005), Schrager (2006) and Luciano and Vigna (2008). The examples above are by no means exhaustive. Thanks to their generality, affine processes provide a versatile framework that can be tailored to model a broad spectrum of phenomena.

The first and third chapters of this dissertation rely heavily on affine processes. In Chapter 1, we study an affine generalization of the Heston model with self-exciting jumps. The characterization of its Laplace transform function through ordinary differential equations will be key in the study of the problem. Chapter 3 takes a broader perspective, focusing on structural properties of affine processes under measure changes. This analysis ensures that the affine structure can simultaneously be exploited under different probability measures.

Structure of the Ph.D. thesis

The dissertation consists of three chapters, each corresponding to a paper. The problems addressed in each chapter are self-contained, but they are connected to risk modeling via Hawkes and affine processes. The papers are presented in chronological order.

Chaper 1: ***Semi-static variance-optimal hedging with self-exciting jumps*** The first chapter is based on Callegaro, Di Tella et al. (2025). This work tackles one of the key problems in mathematical finance: hedging against financial risk. With the increasing complexity of financial instruments, and the introduction of more sophisticated models, hedging remains a relevant challenge. The inclusion of additional sources of randomness in market models, such as stochastic intensity and volatility, improves model realism at the cost of market completeness. In literature, multiple approaches to hedging in incomplete markets have been proposed (see Björk (2009, Section 15.8) for an overview, and Dana and Jeanblanc (2003, Chapter 8) for an introduction on the topic). In this work, we focus on a *semi-static variance-optimal* hedging approach. Specifically, we hedge a *variance swap* by minimizing the residual variance under the risk-neutral measure, combining a dynamic trading strategy with a static position in a fixed basket of European options. The underlying model includes a stochastic, Heston-like volatility component, and self-exciting jumps in the log-asset price, driven by a marked Hawkes process with exponential kernel. This setup allows to capture stylized features observed in real markets by several authors (see, e.g., Aït-Sahalia et al. (2015) and Herrera and González (2014)), and nowadays regarded as well-established properties of price dynamics. The work advances both the theoretical understanding and practical implementation of semi-static hedging strategies in affine models with self-exciting jumps. We highlight the principal contributions as follows.

- **Theoretical contribution.** The main theoretical advancement of the work concerns the characterization of the model's Laplace transform. In the context of semi-static variance-optimal hedging, discounted asset prices are required to be square-integrable martingales under a risk-neutral measure, and this assumption is tightly linked with the Laplace transform domain. The investigated model belongs to the class of affine processes, meaning that its Laplace transform can be described via a system of Riccati equations. Since a closed-form solution of this system is not available, we provide an analytical study of the Laplace transform domain (see Section 1.3 and Proposition 1.7). Differing from similar works in the literature (e.g. Brachetta et al. (2024, Lemma B.1)), our setting incorporates exponential marks in the intensity, leading to nontrivial technical challenges connected to the non-existence of all exponential moments.

- **Computational contribution.** A major contribution of our work consists in the development of numerical techniques to evaluate the hedging error. The optimal hedging strategy requires the computation of four-dimensional complex integrals, see Theorem 1.21. It is also needed to implement Monte Carlo simulations, and to numerically solve and evaluate the Riccati system. We develop computational techniques for simulating the model efficiently and for dealing with high-dimensional integration.

- **Applicative relevance.** From an applied perspective, our study provides an explicit derivation of the semi-static variance-optimal hedging strategy for a variance swap hedged through a basket of European options. The strategy is expressed via a multi-dimensional integral representation. Although these integrals must be evaluated numerically, all the key coefficients are derived in closed form, ensuring analytical tractability for practical applications. Moreover, we conduct a sensitivity analysis on the model parameters, assessing their impact on the hedging error.

Chapter 2: ***A stochastic Gordon-Loeb model for optimal cybersecurity investment under clustered attacks*** In the second chapter, we present the results of Callegaro, Fontana et al. (2025). In this work, we address a stochastic optimal control problem in a Hawkes setting, with an application to cyber-risk. In the last five years years, cyber-risk has emerged as one of the most relevant sources of risk, attracting increasing attention from both academia and industry. Due to the interconnected nature of IT systems, a successful breach may trigger a cascade of subsequent attacks, as highligted in the empirical analysis of Baldwin et al. (2017). To reproduce this clustered effect, Hawkes processes have been recently applied to the modeling of cyber-risks in Bessy-Roland et al. (2021) and Hillairet et al. (2023). Our study aims at applying stochastic control techniques to determine the optimal risk mitigation strategy for an entity facing cyberattacks. While previous studies have mostly addressed this problem under deterministic settings (see Gordon and Loeb (2002) and Krutilla et al. (2021)) or within diffusion-based frameworks (see Tatsumi and Goto (2010)), we focus instead on the stochastic nature of cyberattacks and, in particular, on their contagion effects. Our work contributes in different directions to the emerging literature on the application of stochastic control methods to cybersecurity.

- **Modeling contribution.** Our proposed model is a continuous-time, dynamic, and stochastic generalization of a well-established model in cybersecurity, introduced in Gordon and Loeb (2002). It is the first formulation of the Gordon-Loeb model that incorporates Hawkes processes to describe the arrival of cyberattacks. Gordon and Loeb model assumes that cyberattacks which thread an entity are filtered by its security system, depending on a certain vulnerability. The cybersecurity investments increase the level of information security, reducing the vulnerability of the system and, consequently, the number of cyberattacks which penetrate it. We model the number of attacks with an Hawkes process and represent the cumulated losses through a compound process. In this setting, the control acts on the vulnerability of the system, in an attempt to contrast the clustering arrivals of attacks. To determine the optimal investment in cybersecurity, we build upon the criterion proposed by Gordon and Loeb (2002). We maximize a utility function, which represents the benefit-cost tradeoff obtained by the entity when investing.

- **Theoretical contribution.** The problem is framed as a stochastic control problem with jumps and is addressed using dynamic programming techniques. The optimal value is characterized by a partial-integro differential equation. Although the value function cannot be computed analytically, we analyze its main properties. We investigate the rate of growth of the solution, and study its convexity and Lipschitz properties with respect to the state variables of the problem. Lastly, we prove a verification theorem under suitable regularity assumptions.

- **Computational contribution.** As mentioned above, the control problem formulation is not analytically tractable. To study the behavior of the value function and of the optimal control, we employ different numerical techniques (finite differences, method of lines) to solve the PIDE and characterize the optimal strategy. The literature on PIDEs under Hawkes dynamics is limited (see e.g. Gaïgi et al. (2025) and Houssard et al. (2025) for some recent results), and mostly develops methods tailored to specific contexts. Moreover, unlike most of the existing literature on control with Hawkes processes, our analysis considers a three-dimensional setting, which significantly increases the complexity of the problem.

- **Applicative relevance.** Finally, one of the aims of this work is to highlight the practical implications of considering clustered cyberattacks. For this purpose, we develop a series of numerical experiments showing that accounting for attack clustering leads to more responsive and effective investment policies. We conclude that ignoring the possibility of clusters of cyberattacks might result in a severe underestimation of cyber-risk and suboptimal response strategies.

Chapter 3: ***Stable measure transformations for affine jump-diffusions*** The third chapter is part of an ongoing research project with Prof. Claudio Fontana. We provide a complete characterization of structure-preserving measure changes within the class of affine processes. As highlighted above, such processes provide a powerful framework for modeling risk across different areas.

A key question is whether the affine structure is maintained when changing the probability measure. Moving from one probability to a locally equivalent one is a standard procedure in finance. For example, statistical estimation and risk management are usually performed under the real-world probability, while pricing derivatives requires working under a risk-neutral measure. In general, changes of measure are not structure-preserving. Retaining the affine property would be highly desirable, as it allows to continue exploiting the analytical advantages of these models. This provides a natural motivation for our theoretical investigation. Previous papers have examined this research question in the affine setting (see, e.g., Cheridito, Filipović and Kimmel (2007), Fontana (2012) and Kallsen and Muhle-Karbe (2010)). Our contribution goes beyond these works by establishing necessary and sufficient conditions to identify stable transformations, thereby providing a full characterization in the general affine jump–diffusion setting. The chapter advances the understanding of structure-preserving measure changes in an affine setting, by combining a rigorous theoretical analysis with applicative insights. The main contributions can be summarized as follows.

- **Theoretical contribution.** In the work, we provide necessary and sufficient conditions to characterize the admissible structure-preserving transformations. We consider an affine process under a certain probability measure and we establish a criterion that guarantees, when fulfilled, that under a locally equivalent probability, the given process remains affine. Conversely, any measure change preserving the affine structure must meet our conditions. Achieving a full characterization is nontrivial. The key challenge consists in verifying that the considered density process is a true martingale. By deriving necessary and sufficient conditions, we completely identify the class of admissible transformations in a general affine jump–diffusion setting, providing the sharpest criterion possible. Compared to prior works on related problems in an affine context, our study addresses a more general affine setting than Cheridito, Filipović and Kimmel (2007) and Fontana (2012), and provides a more comprehensive characterization than the one in Kallsen and Muhle-Karbe (2010). A complete comparison with the existing literature is provided in Section 3.4.

- **Applicative relevance.** Our analysis goes beyond a purely theoretical investigation by establishing conditions that are explicit and manageable, enhancing the practical relevance of our result. In Section 3.5, we illustrate the applicative impact of our contribution. We provide various examples, showing how our characterization can be implemented in widely used jump-diffusion models. We also focus on Hawkes processes, analyzing their construction via measure changes and identifying which transformations preserve their structure. These applications demonstrate the practical relevance of our findings, simplifying calculations and supporting tractable modeling in tasks such as risk management, pricing, and credit assessment.

CHAPTER 1

Semi-static variance-optimal hedging with self-exciting jumps

This chapter is based on Callegaro, Di Tella et al. (2025) and it is a joint collaboration with Prof. Giorgia Callegaro, Dr. Paolo Di Tella and Prof. Carlo Sgarra. Submitted in November 2024, the corresponding paper is published on Mathematics of Operations Research.

The aim of this work is to investigate a quadratic, i.e., *variance-optimal*, semi-static hedging problem in an incomplete market model where the underlying log-asset price is driven by a diffusion process with stochastic volatility and a self-exciting jump process of Hawkes type. More precisely, we aim at hedging a claim at time $T > 0$ by using a portfolio of available contingent claims, so to minimize the variance of the residual hedging error at time T. In order to improve the replication of the claim, we look for a hybrid hedging strategy of semi-static type, in which some assets are continuously rebalanced (the dynamic hedging component) and for some other assets a buy-and-hold strategy (the static component) is performed. We discuss in detail a specific example in which the approach proposed is applied, i.e., a variance swap hedged by means of European options, and we provide a numerical illustration of the results obtained.

1.1 Introduction

Two of the main targets of financial mathematics are pricing and hedging of contingent claims. The aim of this work is to investigate a quadratic, i.e., *variance-optimal*, semi-static hedging problem in an incomplete market model where the underlying log-asset price is driven by a diffusion process with stochastic volatility and a self-exciting jump process of Hawkes type. Before introducing our framework, we briefly summarize some key findings in the variance-optimal hedging literature.

In Kallsen and Vierthauer (2009) the authors determine the variance-optimal hedge for a subset of affine processes including a number of popular stochastic volatility models. They obtain semi-explicit formulas for the optimal hedging strategy and the minimal hedging error by applying general structural results and Laplace transform techniques. In Černý and Christoph (2023), they investigate quadratic hedging in a semimartingale market that does not necessarily contain a risk-free asset and they establish an equivalence result for hedging with and without change of numéraire.

Lim (2004) deals with the problems of quadratic hedging and pricing, and mean-variance portfolio selection in an incomplete market setting with continuous trading, multiple assets, and Brownian information. In particular, the author assumes that the parameters describing the market model may be random processes.

In Schal (1994) an option with maturity time T corresponding to a contingent claim H in an incomplete market is considered and the work investigates what would be a fair hedging price for H by taking into account an optimal dynamical hedging plan against H.

Here we denote by S a stochastic process modeling the price of some financial asset traded in the market. We want to hedge a contingent claim η^0 written on S by means of a basket of other contingent claims $\boldsymbol{\eta} = (\eta^1 \dots, \eta^d)^\top$, by adopting a semi-static strategy. *Semi-static* hedging consists in taking a dynamic (i.e., continuously rebalanced) position in S, denoted by ϑ, and a static (i.e., buy-and-hold) position in the fixed basket of contingent claims, $\boldsymbol{\nu}$. For certain hedging problems, semi-static strategies allow for perfect replication even in incomplete markets: see, e.g., the semi-static replication of variance swaps in Carr and Madan (2001), Neuberger (1994). We require the strategy to be *variance-optimal*, meaning that we will perform a minimization of the variance of the residual hedging error at a terminal time $T > 0$:

$$\varepsilon^2 = \min_{\boldsymbol{\nu}\in\mathbb{R}^d,\vartheta\in L^2(S),c\in\mathbb{R}} \mathbb{E}\left[\left(\underbrace{c}_{\text{initial capital}} - \overbrace{\mathbb{E}[\boldsymbol{\nu}^\top\boldsymbol{\eta}]}^{\text{cost of static portfolio}} + \underbrace{\int_0^T \vartheta_s \, dS_s}_{\text{dynamic position}} - (\eta^0 - \overbrace{\boldsymbol{\nu}^\top\boldsymbol{\eta}}^{\text{static position}}) \right)^2\right], \qquad (1.1)$$

given $L^2(S)$ a suitable space where the stochastic integral is well defined.

A procedure for finding the solution to problem (1.1) is provided in Di Tella et al. (2019) and Di Tella et al. (2020), in a general setting. The main idea is to rewrite the problem as an *inner* and an *outer problem*, which can be solved separately. In particular, the inner problem is a classic variance-optimal hedging problem which can be solved with standard techniques, see, e.g. Föllmer and Sondermann (1986). On the other hand, the outer problem turns out to be a finite-dimensional quadratic optimization problem. The solution of the outer problem is then a function of three coefficients A, B, C, which depend on a particular decomposition (Galtchouk-Kunita-Watanabe decomposition, see Kunita and Watanabe (1967)) of the claims η^0, $\boldsymbol{\eta}$. The semistatic hedging strategies have been computed already in Di Tella et al. (2019) for different models for S: Heston, the 3/2 and a model driven by Lévy jumps. Note that the numerical analysis in Di Tella et al. (2020) has been performed only for the Heston model.

The purpose of this work is to solve the semi-static variance optimal hedging problem in a new market model and for a variance swap η^0, by means of a basket of European options $\boldsymbol{\eta}$. The model considered is proved to be affine and the stock price process S encompasses both a Hawkes-type jump component, which describes the self-exciting features, and a stochastic volatility model of Heston type. Research into models with jumps, especially of self-exciting type, is significant as it has been observed that prices the financial markets, see, e.g., Filimonov et al. (2014), and for energy market, see, e.g., Herrera and González (2014) - exhibit spikes displaying a clustering behavior. Hawkes-based jump diffusion models have been used to describe the dynamics of asset prices across several different classes, see Aït-Sahalia et al. (2015) and Hainaut and Moraux (2018) for equities, Brignone et al. (2024) and Gonzato and Sgarra (2021) for commodity markets, Errais et al. (2010) for credit risk derivatives, Hainaut (2016) for interest rates, Rambaldi et al. (2015) for foreign exchange rates.

As a first contribution, we introduce a new stochastic setting, by studying its properties as an affine semimartingale model. We characterize the Laplace transform, studying its existence under suitable conditions on the parameters. This latter analysis is mainly based on an investigation of the Laplace transform domain, as, although the transform can be described through the solution of a system of generalized Riccati equations, determining the explosion time of this system is non-trivial. We remark that characterizing the Laplace transform is relevant, since we want to express contingent claims via a Fourier transform representation, in the spirit of Kallsen and Pauwels (2010). This, as we shall show, allows to obtain more explicit results for the quadratic hedging strategy.

As a second contribution we provide, as an explicit example, the computation of the semi-static hedging strategy of a variance swap by means of a portfolio of European options written on the underlying. Variance swaps contracts are commonly traded in equity markets, typically on S&P index, but a remarkable interest in variance swaps has grown in recent years in commodity markets and, in particular, in energy markets. In Prokopczuk et al. (2017), e.g., the authors analyze the variance risk of commodity markets, constructing synthetic variance swaps and finding significantly negative realized variance swap payoffs in most markets. By following the general methodology exposed in Di Tella et al. (2019, Theorem 3.2) we obtain a semi-explicit expression of A, B, C appearing in the Galtchuk-Kunita-Watanabe decomposition. The inclusion of self-exciting jumps in the model gives rise to some non-trivial difficulties, especially in determining the existence of the expectation appearing in the expressions for A, B, C. We analyze this aspect in detail in Appendix 1.D, where we also provide some explicit formulas for the moments.

The third contribution of the present work is the numerical computation of the optimal strategies. The task in non-trivial as in our specific case, the quantities A, B, C can be explicitly written in terms of integrals over time, expectations of random variables and multiple integrals on strips of the complex plane. Moreover, the integrands cannot be expressed explicitly and depend on the numerical solutions of the Riccati equations, hence making it necessary to simulate the model via Monte Carlo and to apply quadrature rules.

The chapter is organized as follows: In Section 1.2 we introduce the self-exciting jump-diffusion model with stochastic volatility, focusing on the SDEs that describes the price dynamics, on the jump measure and its compensator. In Section 1.3 we study the existence of exponential moments for the the main stochastic processes involved in our model. In Section 1.4, we investigate the affine structure of the model, identifying the domain of existence of its exponential moments and characterizing the Laplace transform in terms of a system of generalized Riccati equations. In Section 1.5, we then introduce the general hedging problem, by highlighting its connection with the Fourier representation of the contingent claims. In Section 1.6 we shall solve the semi-static hedging problem in the case of a variance swap hedged by a basket of European options. Finally, in Section 1.7 we provide some numerical results related to our specific example. Appendices 1.A, 1.B and 1.D contain auxiliary technical results and Appendix 1.C the proof of one of the main results used in the work.

1.2 The model

We fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $E := \mathbb{R} \times \mathbb{R}_+$ and we consider the E-valued marked point-process $(T_n, Y_n)_n$ which is completely characterized by the discrete-random measure

$$\mu(\mathrm{d}t, \mathrm{d}x, \mathrm{d}y) = \sum_{n=1}^{\infty} \mathbb{1}_{\{T_n < +\infty\}} \delta_{(T_n, Y_n)}(\mathrm{d}t, \mathrm{d}x, \mathrm{d}y) \tag{1.2}$$

where $Y_n = (\eta_n^X, \eta_n^\lambda)$, $n \geq 1$, $T_0 = 0$, $Y_0 = 0$, $0 < T_n \uparrow +\infty$ as $n \to +\infty$, $T_n < T_{n+1}$ on $\{T_n < +\infty\}$ and $Y_n \neq 0$ if and only if $T_n < +\infty$ and δ_a is the Dirac measure located at point a. We assume that the marks $(\eta_n^X)_n$ are i.i.d. and Gaussian with $\eta_1^X \sim \mathcal{N}(\gamma, \delta^2)$, while the marks $(\eta_n^\lambda)_n$ are independent of $(\eta_n^X)_n$, i.i.d. and exponentially distributed with $\eta_1^\lambda \sim Exp(\zeta)$.

We denote by $N = \int_0^\cdot \int_E \mathbb{1}_{\mathbb{R}^2}(x, y)\mu(\mathrm{d}s, \mathrm{d}x, \mathrm{d}y)$ the point-process associated with μ, so that the following identity holds:

$$N_t = \sum_{n \geq 1} \mathbb{1}_{\{T_n \leq t\}}, \quad t \geq 0.$$

We denote by $\mathbb{F}^\mu$ the smallest filtration satisfying the usual conditions such that μ is an optional integer-valued random measure. The random times $(T_n)_n$ are then $\mathbb{F}^\mu$-stopping times and the process N is a point process with respect to $\mathbb{F}^\mu$. The process λ is then defined by

$$\mathrm{d}\lambda_t = \beta_\lambda(\alpha_\lambda - \lambda_{t-})\,\mathrm{d}t + \mathrm{d}J_t^\lambda, \quad \lambda_0 \in \mathbb{R}_+ \tag{1.3}$$

where $J_t^\lambda = \int_0^t \int_E y \mathbb{1}_{\mathbb{R}}(x)\mu(\mathrm{d}s, \mathrm{d}x, \mathrm{d}y)$, $t \geq 0$. Equivalently, in integral form, we have:

$$\lambda_t = \alpha_\lambda + (\lambda_0 - \alpha_\lambda)e^{-\beta_\lambda t} + \sum_{n=1}^{N_t} e^{-\beta_\lambda(t - T_n)}\eta_n^\lambda. \tag{1.4}$$

We also introduce $J_t^X = \int_0^t \int_E x \mathbb{1}_{\mathbb{R}_+}(y)\mu(\mathrm{d}s, \mathrm{d}x, \mathrm{d}y)$ and we have the identities

$$J_t^\lambda = \sum_{i=1}^{N_t} \eta_i^\lambda, \qquad J_t^X = \sum_{i=1}^{N_t} \eta_i^X.$$

The marked-point process μ is the jump-measure of the $\mathbb{R}^2$-valued $\mathbb{F}^\mu$-semimartingale (J^λ, J^X).

We assume that the $\mathbb{F}^\mu$-dual predictable projection of μ under $\mathbb{P}$ is given by

$$\nu(\mathrm{d}t, \mathrm{d}x, \mathrm{d}y) = \lambda_{t-}\theta(\mathrm{d}x, \mathrm{d}y)\mathrm{d}t, \tag{1.5}$$

where $\theta(\mathrm{d}x, \mathrm{d}y) = \theta^X(\mathrm{d}x)\theta^\lambda(\mathrm{d}y)$, with θ^X and θ^λ denoting the distribution functions of $\mathcal{N}(\gamma, \delta^2)$ and $Exp(\zeta)$, respectively, that is,

$$\theta^X(\mathrm{d}x) = \frac{1}{\sqrt{2\pi\delta^2}} \exp\left(-\frac{(x-\gamma)^2}{2\delta^2}\right)\mathrm{d}x \qquad\qquad \theta^\lambda(\mathrm{d}y) = \zeta e^{-\zeta y}\mathbb{1}_{\mathbb{R}_+}(y)\mathrm{d}y.$$

Clearly, the process $N^p_t = \int_0^t \int_E \mathbb{1}_{\mathbb{R}^2}(x,y)\nu(\mathrm{d}s, \mathrm{d}x, \mathrm{d}y) = \int_0^t \lambda_{s-}\mathrm{d}s$, $t \geq 0$, is the $\mathbb{F}^\mu$-predictable compensator of N. In other words, N is a *self-exciting* counting process, also called *Hawkes* point-process.

This model can be obtained (at least over a fixed-time horizon $[0,T]$, $T > 0$) starting from a probability measure $\mathbb{Q}$ under which μ is a Poisson random measure with $\mathbb{F}^\mu$-dual predictable projection $\nu^{\mathbb{Q}}(\mathrm{d}t, \mathrm{d}x, \mathrm{d}y) = \theta(\mathrm{d}x, \mathrm{d}y)\mathrm{d}t$ and then defining the measure $\mathbb{P}$ by

$$L_T := \left.\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}\right|_{\mathcal{F}^\mu_T} = \mathcal{E}\Big(\int_0^\cdot (\lambda_{s-} - 1)\mathrm{d}(N_s - s)\Big)_T,$$

the Doléans-Dade exponential $\mathcal{E}(\int_0^\cdot (\lambda_{s-} - 1)\mathrm{d}(N_s - s)) > 0$ being a $(\mathbb{Q}, \mathbb{F}^\mu)$ martingale with mean one (see Brachetta et al. (2024, Proposition 2.6) and Brémaud (1981, Theorems VIII.6 and VIII.10) for details). In particular, under $\mathbb{Q}$ the marks $(Y_n)_n$ are assumed to fulfill the same properties as under $\mathbb{P}$. We also have an explicit expression for L_T:

$$L_T = e^{-\int_0^T (\lambda_s - 1)\mathrm{d}s + \int_0^T \log(\lambda_{s-})\mathrm{d}N_s}. \tag{1.6}$$

Remark 1.1. Notice that under $\mathbb{Q}$ the random variables $(T_n, \eta^X_n, \eta^\lambda_n)_n$ are independent (and so also the $\mathbb{Q}$-Poisson process N is independent of $(\eta^X_n, \eta^\lambda_n)_n$), while this is not the case for N under $\mathbb{P}$, where $(\eta^X_n)_n$ and $(\eta^\lambda_n)_n$ remain independent, but $(T_n)_n$ is not independent of (η^λ_n). Under $\mathbb{P}$, independence holds between the marks $(\eta^X_n)_n$ and $(N, (\eta^\lambda_n)_n)$. What is more, the law of the random variables $(\eta^X_n, \eta^\lambda_n)_n$ is the same both under $\mathbb{P}$ and $\mathbb{Q}$.

We now consider a two dimensional Gaussian process $W = (W^{(1)}, W^{(2)})$, where $W^{(i)}$ is a Brownian motion, $i = 1, 2$, with respect to the smallest filtration $\mathbb{F}^W$ satisfying the usual conditions such that W is adapted. We assume that $\langle W^{(1)}, W^{(2)}\rangle_t = \rho t$, where $\rho \in [-1, 1]$, and that $\mathbb{F}^\mu$ and $\mathbb{F}^W$ are independent. We fix a time-horizon $T > 0$ and assume that

$\mathcal{F} = \mathcal{F}_T^{\mu} \vee \mathcal{F}_T^{W}$. The reference filtration $\mathbb{F}$ is the enlargement of $\mathbb{F}^{\mu}$ by $\mathbb{F}^{W}$, that is,

$$\mathbb{F} = \mathbb{F}^{\mu} \vee \mathbb{F}^{W}.$$

Because of the independence between $\mathbb{F}^{\mu}$ and $\mathbb{F}^{W}$, $\mathbb{F}$ is right-continuous too, hence it satisfies the usual conditions. Furthermore, the $\mathbb{F}$-dual predictable projection $\nu^{\mathbb{F}}$ of μ coincides with ν in equation (1.5).

The market model is given by a stock $S = \mathrm{e}^X$ where, for real parameters $\kappa_1, \beta_v, \alpha_v, \sigma_v, \beta_\lambda, \alpha_\lambda$, we assume

$$\begin{cases} \mathrm{d}X_t = \Big(-\dfrac{1}{2}V_t - (\kappa_1 - 1)\lambda_{t-}\Big)\mathrm{d}t + \sqrt{V_t}\mathrm{d}W_t^{(1)} + \mathrm{d}J_t^X, & X_0 = x_0 \in \mathbb{R}, \qquad (1.7a) \\ \mathrm{d}V_t = \beta_v(\alpha_v - V_t)\mathrm{d}t + \sigma_v\sqrt{V_t}\mathrm{d}W_t^{(2)}, & V_0 = v_0 \in \mathbb{R}_+, \qquad (1.7b) \\ \mathrm{d}\lambda_t = \beta_\lambda(\alpha_\lambda - \lambda_{t-})\,\mathrm{d}t + \mathrm{d}J_t^\lambda, & \lambda_0 \in \mathbb{R}_+. \qquad (1.7c) \end{cases}$$

In the present setting, we fix the risk-free rate $r = 0$, so no discounting is required. If we choose the parameters in such a way as to ensure that the price process $S = \mathrm{e}^X$ is a local martingale with respect to $\mathbb{P}$, the NFLVR arbitration requirement is fulfilled due to the First Fundamental Theorem of Asset Pricing. Since the model is incomplete, by the second fundamental theorem of asset pricing, there are infinitely many equivalent local martingale measures. In the present setting we choose to stand on a risk-neutral modeling approach by specifying the dynamics directly under the risk-neutral probability measure.

For notational convenience, we will denote the triplet as $Z = (X, V, \lambda)$. The model proposed encompasses several models available in the literature: Heston, Jump-diffusion models both of Lévy and Hawkes type. It exhibits both jumps and stochastic volatility features, with jumps clustering. As already mentioned in the Introduction, these features have been observed in the market by several authors and nowadays are considered well-established properties of prices dynamics.

Remark 1.2. The choice of a Gaussian probability distribution for the log-return's jumps has a long tradition dating back to Merton (1976). As pointed out in Cont and Tankov (2003, Section 4), the simplest jump-diffusion models that properly capture the log-returns dynamics are those with Gaussian distributed jump sizes and Kou-type models in Kou (2002), where they assume a double exponential distribution for jump sizes. On the other hand, the choice for the exponential distribution of the jump's intensity is motivated by the non-negative support of the probability density. A similar model assuming Kou-type jumps for both the log-returns and the intensity jumps' size has been proposed in Hainaut and Moraux (2018). Also Brignone et al. (2024) and Liu and Zhu (2019) address a similar structure, however, they do not incorporate marks in the intensity. The methodology presented in this work can be adapted to different settings, considering different probability distributions for the marks. However, important properties must be preserved. For example, the joint Laplace transform must exist finite in an open neighborhood of the origin. Furthermore, to ensure that $S = \mathrm{e}^X$ is a square-integrable martingale under $\mathbb{P}$, we need to ensure that $(1,0,0)$ and $(2,0,0)$ belong to this open neighborhood (see Corollary 1.10 below). Clearly, this has to be traduced in terms of the parameters on which the chosen distribution depends. These conditions have to be treated case by case.

In order to ensure the strict positivity of V, $V > 0$, we assume that $v_0, \alpha_v, \beta_v, \sigma_v > 0$ and that the Feller condition $2\alpha_v\beta_v \geq \sigma_v^2$ holds. We choose $\kappa_1 = e^{\gamma+\delta^2/2}$ and we assume that $\kappa_1 > 1/2$ and $\zeta > T\mathbb{E}(e^{2|\eta_1^X|})$, so that S is a square-integrable martingale. This is a crucial assumption as it implies that $\mathbb{P}$ is a risk-neutral measure. Further details on this will be given below. Here we notice that this special choice of κ_1 ensures that S is a local martingale. Indeed, applying Itô formula to e^X we get

$$\begin{aligned} S_t = S_0 &+ \int_0^t S_{s-}\lambda_{s-}(\mathrm{e}^{(\gamma+\frac{\delta^2}{2})} - \kappa_1)\mathrm{d}s + \int_0^t \sqrt{V_s}S_{s-}\mathrm{d}W_s^{(1)} \\ &+ \int_0^t \int_{\mathbb{R}^2} S_{s-}(\mathrm{e}^x - 1)\mathbb{1}_{\mathbb{R}_+}(y)(\mu - \nu)(\mathrm{d}s, \mathrm{d}x, \mathrm{d}y) \end{aligned}$$

and the drift vanishes if and only if $\kappa_1 = e^{(\gamma+\frac{\delta^2}{2})}$.

We assume $\alpha_\lambda, \beta_\lambda > 0$ and $\beta_\lambda > \int_0^\infty x\theta^\lambda(\mathrm{d}x)$, that imply the stationarity of λ and the existence of all the moments of λ_t, $t \in [0,T]$, see, e.g., Dassios and Zhao (2011, Subsection 3.4) . We also require, for technical simplifications in dealing with exponential moments of the random measure μ, that $\lambda_0 > \alpha_\lambda > 1$, which implies $\lambda_t > 1$ $\mathbb{P}$-almost surely and for every $t \in [0,T]$.

We recap all the assumptions on model parameters in Table 1.1.

From (1.7a), the integral form of X is

$$X_t = X_0 - \frac{1}{2}\int_0^t V_s\mathrm{d}s - (\kappa_1 - 1)\int_0^t \lambda_{s-}\mathrm{d}s + \int_0^t \sqrt{V_s}\mathrm{d}W_s^{(1)} + J_t^X.$$

Stochastic process	#	Assumption	Implication
X	$i)$	$\kappa_1 = e^{\gamma+\delta^2/2}$	S martingale
	$ii)$	$\kappa_1 > 1/2$	S square-integrable
		$\zeta > T\mathbb{E}(e^{2\lvert\eta_1^X\rvert})$	
V	$iii)$	$v_0, \alpha_v, \beta_v, \sigma_v > 0$	Positivity
	$iv)$	$2\alpha_v\beta_v \geq \sigma_v^2$	Feller condition
λ	$v)$	$\alpha_\lambda, \beta_\lambda > 0$	Stationarity + moments
	$vi)$	$\beta_\lambda > \int_0^\infty x\theta^\lambda(\mathrm{d}x)$	Stationarity + moments
	$vii)$	$\lambda_0 > \alpha_\lambda > 1$	Existence of exp moments

Table 1.1: All the model parameters' assumptions at a glance.

1.3 Moments

We gather in this section key results on the moments of the random building blocks of our model. Whenever an expectation $\mathbb{E}[\cdot]$ is written without a superscript, it is computed under the probability measure $\mathbb{P}$.

Proposition 1.3. *Given $T > 0$ and λ as in equation (1.4), if $a \in \mathbb{R}$, $a < \frac{\zeta}{T}$ we have:*

$$\mathbb{E}^{\mathbb{Q}}[e^{a\int_0^T \lambda_s \mathrm{d}s}] < \infty. \tag{1.8}$$

Moreover, if $a < 1$ and $\zeta > T$, then

$$\mathbb{E}\left[e^{a\int_0^T \lambda_s \mathrm{d}s}\right] < \infty.$$

Proof. Since λ is positive, the result holds trivially for $a \leq 0$. Therefore, we will focus on the case $a > 0$ from now on. First of all notice that, from equation (1.4) and for $c > 0$, we have:

$$\lambda_s \leq \max\{\lambda_0, \alpha_\lambda\} + \sum_{n=1}^{N_s} \eta_n^\lambda =: c + \sum_{n=1}^{N_s} \eta_n^\lambda \tag{1.9}$$

so that, using the fact that, under $\mathbb{Q}$, N is independent of $(\eta_n^\lambda)_n$ (recall Remark 1.1) and that the Laplace transform for an exponential random variable $\eta_1^\lambda \sim Exp(\zeta)$ is known, we find

$$\mathbb{E}^{\mathbb{Q}}\left[e^{a\int_0^T \lambda_s \mathrm{d}s}\right] \leq e^{acT}\mathbb{E}^{\mathbb{Q}}\left[e^{a\int_0^T \sum_{n=1}^{N_T}\eta_n^\lambda \mathrm{d}s}\right] = e^{acT}\mathbb{E}^{\mathbb{Q}}\left[\prod_{n=1}^{N_T} E^{\mathbb{Q}}\left(e^{aT\eta_1^\lambda}\right)\right] = e^{acT}\mathbb{E}^{\mathbb{Q}}\left[\left(\frac{\zeta}{\zeta - aT}\right)^{N_T}\right]$$

which is finite if $aT < \zeta$. Indeed, under $\mathbb{Q}$, the random variable $N_T \sim Po(T)$ and we have

$$\mathbb{E}^{\mathbb{Q}}\left[\left(\frac{\zeta}{\zeta - aT}\right)^{N_T}\right] = \sum_{\ell=0}^{+\infty} \mathbb{Q}(N_T = \ell)\left(\frac{\zeta}{\zeta - aT}\right)^{\ell} = \sum_{\ell=0}^{+\infty} e^{-T}\frac{T^{\ell}}{\ell!}\left(\frac{\zeta}{\zeta - aT}\right)^{\ell} = e^{-T}e^{\frac{\zeta T}{\zeta - aT}}$$

and so

$$\mathbb{E}^{\mathbb{Q}}\left[e^{a\int_0^T \lambda_s \mathrm{d}s}\right] \leq e^{acT}e^{-T}e^{\frac{\zeta T}{\zeta - aT}} \leq e^{acT}e^{\frac{\zeta T}{\zeta - aT}}$$

which is finite since ζ and T are positive and fixed and $aT < \zeta$.

We now aim at investigating the Laplace transform of $\int_0^T \lambda_s \mathrm{d}s$ under the measure $\mathbb{P}$. Exploiting the change of measure in equation (1.6), we pass under the measure $\mathbb{Q}$:

$$\mathbb{E}\left[e^{a\int_0^T \lambda_s \mathrm{d}s}\right] = \mathbb{E}^{\mathbb{Q}}\left[L_T e^{a\int_0^T \lambda_s \mathrm{d}s}\right] = \mathbb{E}^{\mathbb{Q}}\left[e^{(a-1)\int_0^T \lambda_s \mathrm{d}s + T + \int_0^T \log(\lambda_{s-})\mathrm{d}N_s}\right].$$

Since $a < 1$, exploiting equation (1.41) and the convergence result in equation (1.8) for $\zeta > T$

$$\mathbb{E}\left[e^{a\int_0^T \lambda_s \mathrm{d}s}\right] \leq C\mathbb{E}^{\mathbb{Q}}\left[e^{\int_0^T \log(\lambda_{s-})\mathrm{d}N_s}\right] = C\mathbb{E}^{\mathbb{Q}}\left[e^{\int_0^T (\lambda_{s-} - 1)\mathrm{d}s}\right] < \infty.$$

□

Remark 1.4. Notice that Assumption $ii)$ in Table 1.1 implies that $\zeta > T$.

We now compute the exponential moments $\mathbb{E}[e^{H*\mu_T}]$, with μ the random measure in equation (1.2), $H : [0,T] \times \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}$ and where

$$H * \mu_T := \int_0^T \int_{\mathbb{R}\times\mathbb{R}_+} H(t,x,y)\mu(\mathrm{d}t, \mathrm{d}x, \mathrm{d}y).$$

Proposition 1.5. *Given $c_1, c_2 \geq 0$ and H defined as*

$$H(t,x,y) = c_1 \mathbb{1}_{\mathbb{R}_+}(y)|x| + c_2 \mathbb{1}_{\mathbb{R}}(x) y,$$

if the intensity process λ satisfies $\lambda_t > 1$ $\mathbb{P}$-almost surely and for every $t \in [0,T]$ and if

$$T < \frac{\zeta - c_2}{\mathbb{E}(e^{c_1|\eta_1^X|})}$$

then $\mathbb{E}[e^{H*\mu_T}] < \infty$.

Proof. First of all notice that $H * \mu_T = c_1 \sum_{i=1}^{N_T} |\eta_i^X| + c_2 J_T^\lambda$. We have, passing under $\mathbb{Q}$ and via equation (1.6):

$$\mathbb{E}[e^{H*\mu_T}] = \mathbb{E}^{\mathbb{Q}}[L_T e^{H*\mu_T}] = \mathbb{E}^{\mathbb{Q}}[e^{-\int_0^T (\lambda_s - 1)\mathrm{d}s + \int_0^T \log(\lambda_{s-}) dN_s} e^{c_1 \sum_{i=1}^{N_T} |\eta_i^X| + c_2 J_T^\lambda}]$$

Now, being the intensity process $\lambda_t > 1$ $\mathbb{P}$ and $\mathbb{Q}$ almost surely and for every $t \in [0, T]$, we immediately find:

$$\mathbb{E}^{\mathbb{Q}}[e^{-\int_0^T (\lambda_s - 1)\mathrm{d}s + \int_0^T \log(\lambda_{s-}) dN_s} e^{c_1 \sum_{i=1}^{N_T} |\eta_i^X| + c_2 J_T^\lambda}] \leq \mathbb{E}^{\mathbb{Q}}[e^{\int_0^T \log(\lambda_{s-}) dN_s} e^{c_1 \sum_{i=1}^{N_T} |\eta_i^X| + c_2 J_T^\lambda}]$$

If we introduce $\widetilde{H}(t, x, y) := \mathbb{1}_{\mathbb{R}\times\mathbb{R}_+}(x, y) \log(\lambda_{t-}) + H(t, x, y)$ we can use equation (1.42) in Appendix 1.B (recall that $\nu^{\mathbb{Q}}(\mathrm{d}t, \mathrm{d}x, \mathrm{d}y) = \theta(\mathrm{d}x, \mathrm{d}y)\mathrm{d}t = \theta^X(\mathrm{d}x)\theta^\lambda(\mathrm{d}y)\mathrm{d}t$) and prove that

$$\begin{aligned}
\mathbb{E}[e^{H*\mu_T}] &\leq \mathbb{E}^{\mathbb{Q}}\left[e^{\int_0^T \log(\lambda_{s-}) dN_s} e^{c_1 \sum_{i=1}^{N_T} |\eta_i^X| + c_2 J_T^\lambda}\right] = \mathbb{E}^{\mathbb{Q}}\left[e^{\widetilde{H}*\mu_T}\right] \\
&= \mathbb{E}^{\mathbb{Q}}\left[e^{\int_0^T \int_{\mathbb{R}\times\mathbb{R}_+} \left(e^{\widetilde{H}(t,x,y)} - 1\right)\theta^X(\mathrm{d}x)\theta^\lambda(\mathrm{d}y)\mathrm{d}t}\right] \\
&= \mathbb{E}^{\mathbb{Q}}\left[e^{\int_0^T \int_{-\infty}^{+\infty} \int_0^{+\infty} \left(e^{\mathbb{1}_{\mathbb{R}\times\mathbb{R}_+}(x,y)\log(\lambda_{t-}) + c_1 \mathbb{1}_{\mathbb{R}_+}(y)|x| + c_2 \mathbb{1}_{\mathbb{R}}(x) y} - 1\right)\theta^X(\mathrm{d}x)\theta^\lambda(\mathrm{d}y)\mathrm{d}t}\right] \\
&\leq \mathbb{E}^{\mathbb{Q}}\left[\exp\left(\int_0^T \int_{-\infty}^{+\infty} \int_0^{+\infty} e^{\mathbb{1}_{\mathbb{R}\times\mathbb{R}_+}(x,y)\log(\lambda_{t-}) + c_1 \mathbb{1}_{\mathbb{R}_+}(y)|x| + c_2 \mathbb{1}_{\mathbb{R}}(x) y}\theta^X(\mathrm{d}x)\theta^\lambda(\mathrm{d}y)\mathrm{d}t\right)\right] \\
&= \mathbb{E}^{\mathbb{Q}}\left[\exp\left(\int_{-\infty}^{+\infty} \int_0^{+\infty} e^{c_1|x| + c_2 y}\left(\int_0^T e^{\log(\lambda_{t-})} dt\right)\theta^X(\mathrm{d}x)\theta^\lambda(\mathrm{d}y)\right)\right] \\
&= \mathbb{E}^{\mathbb{Q}}\left[\exp\left(\int_0^T \lambda_t \mathrm{d}t \cdot \mathbb{E}^{\mathbb{Q}}(e^{c_1|\eta_1^X|})\mathbb{E}^{\mathbb{Q}}(e^{c_2 \eta_1^\lambda})\right)\right] = \mathbb{E}^{\mathbb{Q}}\left[\exp\left(\int_0^T \lambda_t \mathrm{d}t \cdot \mathbb{E}(e^{c_1|\eta_1^X|})\frac{\zeta}{\zeta - c_2}\right)\right]
\end{aligned}$$

where we used the moment generating function for an exponential random variable for $\zeta - c_2 > 0$. It remains now to use the moment generating function for the folded normal $|\eta_1^X|$, with $\eta_1^X \sim \mathcal{N}(\gamma, \delta^2)$:

$$\mathbb{E}[e^{c_1|\eta_1^X|}] = e^{\frac{\delta^2 c_1^2}{2} - c_1\gamma}\Phi\left(\frac{c_1\delta^2 - \gamma}{\delta}\right) + e^{\frac{\delta^2 c_1^2}{2} + c_1\gamma}\Phi\left(\frac{c_1\delta^2 + \gamma}{\delta}\right)$$

where Φ is the cumulative distribution function of a standard Gaussian. So, we finally have:

$$\mathbb{E}[e^{H*\mu_T}] \leq \mathbb{E}^{\mathbb{Q}}\left[\exp\left(\int_0^T \lambda_t \mathrm{d}t \cdot \mathbb{E}(e^{c_1|\eta_1^X|})\frac{\zeta}{\zeta - c_2}\right)\right]$$

which is finite, in virtue of Proposition 1.3, using $a = \mathbb{E}(e^{c_1|\eta_1^X|})\frac{\zeta}{\zeta - c_2}$. □

Remark 1.6. Assumption $vii)$ in Table 1.1, i.e., $\lambda_0 > \alpha_\lambda$ and $\alpha_\lambda > 1$, implies that the condition $\lambda_t > 1$ $\mathbb{P}$-almost surely and for every $t \in [0, T]$ is satisfied (recall equation (1.4))

1.4 From the affine structure to the Laplace transform of (X, V, λ)

It is possible to prove that $Z = (X, V, \lambda)$, as defined in system (1.7), is an affine process. Exploiting this property, we will then compute its Fourier-Laplace transform.

We start by referring to the definition of semimartingale characteristics given in Jacod and Shiryaev (2013, Definition II.2.6), working with the standard truncation function $h(x) = x\mathbb{1}_{\{|x|\leq 1\}}$.

The triplet we are interested in is (b, c, K), the *local (or differential) characteristics of* Z, as defined in Eberlein and Kallsen (2019, Definition 4.3). It can be easily computed as:

$$b_t = \begin{pmatrix} -\frac{1}{2}V_t - (\kappa_1 - 1)\lambda_{t-} + \int_{\mathbb{R}} h(x)\theta^X(\mathrm{d}x)\lambda_{t-} \\ \beta_v(\alpha_v - V_t) \\ \beta_\lambda(\alpha_\lambda - \lambda_{t-}) + \int_{\mathbb{R}_+} h(x)\theta^\lambda(\mathrm{d}x)\lambda_{t-} \end{pmatrix}, \tag{1.10}$$

$$c_t = \begin{pmatrix} V_t & \rho\sigma_v V_t & 0 \\ \rho\sigma_v V_t & \sigma_v^2 V_t & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{1.11}$$

$$K_t(\mathrm{d}x) = \theta^X(\mathrm{d}x_1)\delta_0(\mathrm{d}x_2)\theta^\lambda(\mathrm{d}x_3)\lambda_{t-}. \tag{1.12}$$

The local characteristics (b, c, K) are clearly affine functions of (X, V, λ). Up to a permutation of the indices, the triplet (b, c, K) is admissible in the sense of Eberlein and Kallsen (2019, Section 6.1, Equation (6.26)) and it follows by Theorem 6.6 in Eberlein and Kallsen (2019) that Z is an affine multivariate process. In particular, Z is Markov and time-homogeneous.

By exploiting the affine property of the model, we can provide a characterization of its Laplace transform. We fix a time horizon $0 \leq T < T^*$, where T^* is the explosion time associated to the Heston model, see Andersen and Piterbarg (2007, Proposition 3.1). First, we characterize the domain of existence $\mathcal{D}_{\mathcal{L}(Z_T)}$ of the real Laplace transform of Z_T:

$$\mathcal{D}_{\mathcal{L}(Z_T)} = \{u \in \mathbb{R}^3 : \mathbb{E}[\exp(u^\top Z_T)] < \infty\}. \tag{1.13}$$

Proposition 1.7. *Under the Assumptions in Table 1.1 , and for $T < T^*$, the existence domain for the Laplace transform of Z_T satisfies*

$$\mathcal{D}_{\mathcal{L}(Z_T)} \supseteq \mathcal{E} := \left\{(u_1, u_2, u_3) \in \mathbb{R}^3 : u_1(\kappa_1 - 1) + 1 > 0, u_2 \in \mathbb{R}, u_3 < \zeta - T\mathbb{E}(e^{u_1|\eta_1^X|})\right\}.$$

Proof. Notice that negative values of u_3 are trivially included in the Laplace domain, so that in the proof we focus on positive values of u_3. First of all notice that, using $X_T^{\text{Heston}} = X_0 - \frac{1}{2}\int_0^T V_s\,\mathrm{d}s + \int_0^T \sqrt{V_s}\mathrm{d}W_s^{(1)}$, we find

$$\mathbb{E}[\exp(u_1 X_T + u_2 V_T + u_3\lambda_T)] = \mathbb{E}\left[\exp\left(u_1 X_T^{\text{Heston}} - u_1(\kappa_1 - 1)\int_0^T \lambda_s \mathrm{d}s + u_1 J_T^X + u_2 V_T + u_3 \lambda_T\right)\right]$$

so that, due to the independence of $\mathbb{F}^W$ and $\mathbb{F}^\mu$, the above expectation is finite if and only if

$$\mathbb{E}[\exp(u_1 X_T^{\text{Heston}} + u_2 V_T)] < \infty, \tag{1.14}$$

$$\mathbb{E}\left[\exp(-u_1(\kappa_1 - 1)\int_0^T \lambda_t \mathrm{d}t + u_1 J_T^X + u_3\lambda_T)\right] < \infty. \tag{1.15}$$

The expected value in (1.14) is finite for every $u_1, u_2 \in \mathbb{R}$, since we have assumed that $T < T^*$. Focusing now on the expected value in (1.15), via inequality (1.9) and since $\lambda_0 > \alpha_\lambda$ by Assumption $vii)$ in Table 1.1, we find:

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(-u_1(\kappa_1 - 1)\int_0^T \lambda_t \mathrm{d}t + u_1 J_T^X + u_3\lambda_T\right)\right] \\
&\quad\le \mathbb{E}\left[\exp\left(-u_1(\kappa_1 - 1)\int_0^T \lambda_t \mathrm{d}t + u_1 J_T^X + u_3(\lambda_0 + J_T^\lambda)\right)\right] \\
&\quad\le \widetilde{c}\,\mathbb{E}\left[\exp\left(-u_1(\kappa_1 - 1)\int_0^T \lambda_t \mathrm{d}t + u_1 J_T^X + u_3 J_T^\lambda\right)\right]
\end{aligned}$$

with $\widetilde{c} = e^{u_3\lambda_0}$. We now pass under $\mathbb{Q}$ through the change of measure in equation (1.6), to exploit the richer independence (recall Remark 1.1):

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(-u_1(\kappa_1 - 1)\int_0^T \lambda_t \mathrm{d}t + u_1 J_T^X + u_3 J_T^\lambda\right)\right] \\
&\quad= \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_0^T(\lambda_t - 1)\mathrm{d}t + \int_0^T \log(\lambda_{t-})dN_t} e^{-u_1(\kappa_1-1)\int_0^T \lambda_t \mathrm{d}t + u_1 J_T^X + u_3 J_T^\lambda}\right] \\
&\quad= e^T \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_0^T \lambda_t[u_1(\kappa_1 - 1)+1]\mathrm{d}t} e^{\int_0^T \log(\lambda_{t-})dN_t} e^{u_1 J_T^X + u_3 J_T^\lambda}\right].
\end{aligned}$$

Under the assumption $u_1(\kappa_1 - 1) + 1 > 0$, being $\lambda_0 > 1$ and the positive stochastic process λ (see Remark 1.6), we find

$$\mathbb{E}\left[\exp\left(-u_1(\kappa_1 - 1)\int_0^T \lambda_t \mathrm{d}t + u_1 J_T^X + u_3 J_T^\lambda\right)\right] \leq e^T \mathbb{E}^{\mathbb{Q}}\left[e^{\int_0^T \log(\lambda_{t-}) dN_t} e^{u_1 J_T^X + u_3 J_T^\lambda}\right] = e^T \mathbb{E}^{\mathbb{Q}}[e^{\overline{H} * \mu_T}]$$

with $\overline{H}(t, x, y) := \mathbb{1}_{\mathbb{R}\times\mathbb{R}_+}(x, y)\log(\lambda_{t-}) + u_1 \mathbb{1}_{\mathbb{R}_+}(y)x + u_3 \mathbb{1}_{\mathbb{R}}(x)y$. Since we clearly have:

$$\overline{H}(t, x, y) \leq \mathbb{1}_{\mathbb{R}\times\mathbb{R}_+}(x, y)\log(\lambda_{t-}) + u_1 \mathbb{1}_{\mathbb{R}_+}(y)|x| + u_3 \mathbb{1}_{\mathbb{R}}(x)y$$

which is $\widetilde{H}$ in the proof of Proposition 1.5 for $c_1 = u_1, c_2 = u_3$, by exploiting the same ideas there we find that $\mathbb{E}^{\mathbb{Q}}[e^{\overline{H} * \mu_T}] < \infty$ if $\mathbb{E}(e^{u_1|\eta_1^X|})\frac{\zeta}{\zeta - u_3} < \frac{\zeta}{T}$. To conclude, the expectation $\mathbb{E}[\exp(u_1 X_T + u_2 V_T + u_3 \lambda_T)]$ is finite if

$$\begin{cases} u_3 < \zeta - T\mathbb{E}(e^{u_1|\eta_1^X|}) \\ u_1(\kappa_1 - 1) + 1 > 0. \end{cases}$$

□

Finally, we characterize the complex, conditional Laplace transform of Z_T.

Theorem 1.8. *Given $u \in \mathcal{S}(\mathcal{D}_{\mathcal{L}(Z_T)}) := \{u = (u_1, u_2, u_3) \in \mathbb{C}^3 : \mathfrak{Re}(u) \in \mathcal{D}_{\mathcal{L}(Z_T)}\}, T < T^*$, the conditional Laplace transform of Z_T given $\mathcal{F}_t, t \in [0, T]$, can be written as follows:*

$$\mathbb{E}[\exp(u^\top Z_T)|\mathcal{F}_t] = \exp\left(\phi(T - t, u) + u_1 X_t + \psi(T - t, u)V_t + \chi(T - t, u)\lambda_t\right) \tag{1.16}$$

where φ , ψ and χ are the unique solutions of the extended Riccati system:

$$\begin{cases} \frac{\partial \phi}{\partial t}(t, u) = \alpha_v \beta_v \psi(t, u) + \alpha_\lambda \beta_\lambda \chi(t, u), & \phi(0, u) = 0 \\ \frac{\partial \psi}{\partial t}(t, u) = -\frac{1}{2}u_1 + \frac{1}{2}u_1^2 - \beta_v \psi(t, u) + \rho\sigma_v u_1 \psi(t, u) + \frac{1}{2}\sigma_v^2 \psi(t, u)^2, & \psi(0, u) = u_2, \\ \frac{\partial \chi}{\partial t}(t, u) = -\beta_\lambda \chi(t, u) - (\kappa_1 - 1)u_1 + e^{\gamma u_1 + \delta^2 u_1^2/2}\frac{\zeta}{\zeta - \chi(t, u)} - 1, & \chi(0, u) = u_3, \end{cases} \tag{1.17}$$

and ϕ, ψ, χ are C^1 complex functions such that

$$\phi\colon t \mapsto \phi(t, u), \quad \psi\colon t \mapsto \psi(t, u), \quad \chi\colon t \mapsto \chi(t, u), \mathfrak{Re}(\chi(t, u)) < \zeta.$$

Proof. The proof of this proposition is provided below in Section 1.4.1. □

Remark 1.9. The explicit expression for the function $\psi(t, u)$ in the Riccati system (1.17) is known, see Alfonsi (2015, Prop. 4.2.1). In particular, for $u \in \mathcal{S}(\mathcal{D}_{\mathcal{L}(Z_t)})$, $\psi(t, u)$ it is given by

$$\psi(t,u) := \begin{cases} u_2 + (r_- - u_2)\frac{1-\exp\left(-t\sqrt{\Delta}\right)}{1-g\exp(-t\sqrt{\Delta})}, & \Delta(u_1) \neq 0, \\ u_2 + (r_- - u_2)^2 \frac{\sigma_v^2 t}{2+\sigma_v^2 t(r_- - u_2)}, & \Delta(u_1) = 0\,. \end{cases}$$

where $\Delta(u_1) = (\rho\sigma_v u_1 - \beta_v)^2 - \sigma_v^2(u_1^2 - u_1)$, $r_\pm = r_\pm(u) := \frac{1}{\sigma_v^2}\left(\beta_v - \rho\sigma_v u_1 \pm \sqrt{\Delta(u_1)}\right)$, $g = g(u) = \frac{r_- - u_2}{r_+ - u_2}$. The following convention holds

$$\frac{\exp(-t\sqrt{\Delta}) - g}{1-g} := 1, \qquad \frac{1-\exp(t\sqrt{\Delta})}{1-g\exp(t\sqrt{\Delta})} := 0$$

whenever the denominator of g is equal to zero.

Corollary 1.10. *Under the Assumptions in Table 1.1, the stochastic process $S = e^X$ with X defined in equation (1.7a) and $\kappa_1 = e^{\gamma+\delta^2/2}$ is a square-integrable martingale.*

Proof. The process S is square-integrable if $(2,0,0) \in \mathcal{E} \subseteq \mathcal{D}_{\mathcal{L}(Z_T)}$. Recalling Proposition 1.7, the condition holds if $\kappa_1 > \frac{1}{2}$, $\zeta > T\mathbb{E}(e^{2|\eta_1^X|})$, which correspond with Assumptions $ii)$ in Table 1.1. Moreover, using once more Proposition 1.7, we also have that $(1,0,0) \in \mathcal{E} \subseteq \mathcal{D}_{\mathcal{L}(Z_T)}$. Thus, we can write equations (1.16), (1.17) for $u = (1,0,0)$ and observe that the martingality condition for $S = e^X$, i.e., $\mathbb{E}[\exp(X_T)|\mathcal{F}_t] = \exp(X_t), 0 \le t \le T$, holds since if $\kappa_1 = e^{\gamma+\delta^2/2}$ the unique solution (ϕ,ψ,χ) of the following system is exactly $(0,0,0)$

$$\begin{cases} \frac{\partial\phi}{\partial t}(t,u) = \alpha_v\beta_v\psi(t,u) + \alpha_\lambda\beta_\lambda\chi(t,u), & \phi(0,u) = 0, \\ \frac{\partial\psi}{\partial t}(t,u) = -\beta_v\psi(t,u) + \rho\sigma_v\psi(t,u) + \frac{1}{2}\sigma_v^2\psi(t,u)^2, & \psi(0,u) = 0, \\ \frac{\partial\chi}{\partial t}(t,u) = -\beta_\lambda\chi(t,u) - (\kappa_1 - 1) + e^{\gamma+\delta^2/2}\frac{\zeta}{\zeta-\chi(t,u)} - 1, & \chi(0,u) = 0. \end{cases}$$

□

1.4.1 Proof of Theorem 1.8

It is enough to prove formula (1.16) for $t = 0$, in the case of the non-conditional Laplace transform. Indeed, assume that

$$\mathbb{E}[\exp(u^\top Z_T)] = \exp\left(\phi(T,u) + u_1 X_0 + \psi(T,u)V_0 + \chi(T,u)\lambda_0\right), \tag{1.18}$$

where (ϕ,ψ,χ) is the solution to system (1.17). Since Z is a time-homogeneous Markov process, equation (1.16) follows as direct consequence of the Markov property.

Let us then focus on equation (1.18). The proof is based on the application of Keller-Ressel and Mayerhofer (2015, Theorem 2.26), where a formula for the complex Laplace transform of an affine process is provided. Let us state the main ideas of the theorem in the case of a general affine process $Z = (Z_t^1, Z_t^2, Z_t^3)$ having state space $D \subseteq \mathbb{R}^3$ and affine characteristics (b, c, K). Without loss of generality, we can assume that $K_t(\mathrm{d}x)$ can be written as $G(Z_t, \mathrm{d}x)$ where

$$G(z, \mathrm{d}x) = G(z_1, z_2, z_3, \mathrm{d}x) = z_1 G_1(\mathrm{d}x) + z_2 G_2(\mathrm{d}x) + z_3 G_3(\mathrm{d}x).$$

First of all, consider the convex set

$$\mathcal{Y} = \bigcap_{z \in D} \left\{ y \in \mathbb{R}^3 : \int_{||x|| \geq 1} e^{y^\top x} G(z, \mathrm{d}x) < \infty \right\} \tag{1.19}$$

and the strip $\mathcal{S}(\mathcal{Y}^\circ) = \{u \in \mathbb{C}^3 : \mathfrak{Re}(u) \in \mathcal{Y}^\circ\}$, where $\mathcal{Y}^\circ$ is the interior of $\mathcal{Y}$.

Following the statement of the theorem, one has to take $u \in \mathcal{S}(\mathcal{Y}^\circ)$ such that the extended Riccati system, see Keller-Ressel and Mayerhofer (2015, Definition 2.10), has solution for initial value $\mathfrak{Re}(u)$ up to time T. If (p, q_1, q_2, q_3) is the solution of the extended Riccati system, then one must verify that $(q_1, q_2, q_3)(t, \mathfrak{Re}(u)) \in \mathcal{Y}^\circ$. Under the latter condition, also the complex Riccati system, defined in Keller-Ressel and Mayerhofer (2015, Definition 2.22), has unique solution $(\Phi, \Psi_1, \Psi_2, \Psi_3)$ with initial value u and up to time T. The theorem concludes that if all the previous conditions are satisfied, then for all $t \in [0, T]$, $\mathbb{E}[|\exp(u^\top Z_t)|] < \infty$ and

$$\mathbb{E}[\exp(u^\top Z_t)] = \exp\left(\Phi(T, u) + \Psi_1(t, u) X_0 + \Psi_2(t, u) V_0 + \Psi_3(t, u) \lambda_0\right).$$

We will apply the theorem to our specific case, where $Z = (X, V, \lambda)$ is the affine process in equation (1.7) having state space $D = \mathbb{R} \times \mathbb{R}_+^2 \subseteq \mathbb{R}^3$ and affine characteristics (b, c, K) as in equations (1.10), (1.11), (1.12). We write

$$K_t(\mathrm{d}x) = G(z, \mathrm{d}x) = z_3 G_3(\mathrm{d}x), \qquad G_3(\mathrm{d}x) = \theta^X(\mathrm{d}x_1)\delta_0(\mathrm{d}x_2)\theta^\lambda(\mathrm{d}x_3) \tag{1.20}$$

noticing that $G_3(\mathrm{d}x)$ is a Lévy measure. We will proceed in steps: first we will investigate the shape of $\mathcal{Y}$, introduced in equation (1.19), and of its interior in our specific model. Then, we will identify a subset of $\mathcal{Y}$ such that the extended Riccati system has solution when taking initial value in that subset. We will then verify that the solution stays in $\mathcal{Y}^\circ$. The extended and the complex Riccati systems will be characterized and we will see that the latter completely coincide with the system in equation (1.17). Finally, we will assemble the steps and verify formula (1.16).

Step 1) Characterizing the convex set $\mathcal{Y}$ and its interior $\mathcal{Y}^\circ$.

Proposition 1.11. *Consider $Z = (X, V, \lambda)$ as in equations (1.7) and $D = \mathbb{R} \times \mathbb{R}_+^2$. The following equality holds*

$$\mathcal{Y} = \bigcap_{z \in D} \left\{ y \in \mathbb{R}^3 : \int_{\mathbb{R}^3} e^{y^\top x} G(z, \mathrm{d}x) < \infty \right\} = \mathbb{R} \times \mathbb{R} \times (-\infty, \zeta).$$

Proof. Let us focus on the first identity. By equation (1.20), we have that $G(z, \mathrm{d}x)$ is the multiplication of z_3 with the Lévy measure $G_3(\mathrm{d}x)$. Then, for every z fixed, it follows by in Sato (1999, Theorem 25.17) that when dealing with Levy measure

$$\int_{||x|| \geq 1} e^{y^\top x} G_3(\mathrm{d}x) < \infty \quad \text{if and only if} \quad \int_{\mathbb{R}^3} e^{y^\top x} G_3(\mathrm{d}x) < \infty,$$

which is equivalent to

$$\int_{||x|| \geq 1} z_3 e^{y^\top x} G_3(\mathrm{d}x) < \infty \quad \text{if and only if} \quad \int_{\mathbb{R}^3} z_3 e^{y^\top x} G_3(\mathrm{d}x) < \infty.$$

Recalling the link between $G(z, \mathrm{d}x)$ and $G_3(\mathrm{d}x)$ in equation (1.20), the first equality is then proved. The second identity follows writing explicitly F for our specific model, as expressed in equation (1.20). Fix $z = (z_1, z_2, z_3) \in D$, then

$$\begin{aligned} \int_{\mathbb{R}^3} e^{y^\top x} G(z, \mathrm{d}x) &= \int_{\mathbb{R}^3} e^{y_1 x_1 + y_2 x_2 + y_3 x_3} z_3 \theta^X(\mathrm{d}x_1) \delta_0(\mathrm{d}x_2) \theta^\lambda(\mathrm{d}x_3) \\ &= z_3 \int_{\mathbb{R}} e^{y_1 x_1} \theta^X(\mathrm{d}x_1) \cdot \int_{\mathbb{R}} e^{y_3 x_3} \theta^\lambda(\mathrm{d}x_3). \end{aligned} \tag{1.21}$$

Referring to the properties of the exponential moments of the Gaussian and Exponential distributions, one can conclude that the quantity (1.21) is finite only if $y_3 < \zeta$, where ζ is the parameter of the exponential random variable. As there are no restrictions on y_1, y_2, the second equality in the statement follows. □

Remark 1.12 (On the interior of $\mathcal{Y}$). By Proposition 1.11, one observes that $\mathcal{Y}$ is open, then $\mathcal{Y}^\circ = \mathcal{Y}$.

Step 2) Finding a subset of $\mathcal{Y}$ such that the extended Riccati system has solution when choosing an initial value in it. As stated in Keller-Ressel and Mayerhofer (2015, Theorem 2.14), $\mathcal{D}_{\mathcal{L}(Z_T)} \subseteq \mathcal{Y}$, where $\mathcal{D}_{\mathcal{L}(Z_T)}$ is the domain of the exponential moments of Z_T, as defined in equation (1.13). In Keller-Ressel and Mayerhofer (2015, Theorem 2.14), they also state that if $y \in \mathcal{D}_{\mathcal{L}(Z_T)}$, i.e., $\mathbb{E}[\exp(y^\top Z_T)] < \infty$, then the extended Riccati system has solution $(p, q_1, q_2, q_3)(t, y)$ up to time T. The subset we are looking for is exactly $\mathcal{D}_{\mathcal{L}(Z_T)}$.

Step 3) Proving that if a solution of the extended Riccati system exists, then it stays in $\mathcal{Y}^\circ$. Recall that we are working with an affine process having state space $D \subseteq \mathbb{R}^3$. In this case, the extended Riccati system, as defined in Keller-Ressel and Mayerhofer (2015, Definition 2.10), is a system of four generalized Riccati differential equations whose solution is given by (p, q_1, q_2, q_3). In particular, fixed $y \in \mathcal{Y}$ such that the solution exists and $t \in [0, T]$, by definition we have that

$$p \colon t \mapsto p(t, y) \in \mathbb{R}, \quad q_i \colon t \mapsto q_i(t, y) \in \mathcal{Y}$$

are C^1-functions for $i = 1, 2, 3$. Since $\mathcal{Y}$ is open, as stated in Remark 1.12 then $q(t, y) \in \mathcal{Y} \equiv \mathcal{Y}^\circ$.

Step 4) Extended and complex Riccati systems. We now characterize the extended and complex Riccati systems, introduced in Keller-Ressel and Mayerhofer (2015, Definition 2.10, Definition 2.22), respectively.

Proposition 1.13 (Extended Riccati system)**.** *Let $T \geq 0$ and $y = (y_1, y_2, y_3) \in \mathcal{Y}$. The extended Riccati system associated to $Z = (X, V, \lambda)$ is the following:*

$$\begin{cases} \frac{\partial p}{\partial t}(t, y) = \alpha_v \beta_v q_2(t, y) + \alpha_\lambda \beta_\lambda q_3(t, y), & p(0, y) = 0, \\ \frac{\partial q_1}{\partial t}(t, y) = 0, & q_1(0, y) = y_1, \\ \frac{\partial q_2}{\partial t}(t, y) = -\frac{1}{2} q_1(t, y) + \frac{1}{2} q_1(t, y)^2 - \beta_v q_2(t, y) + \rho \sigma_v q_1(t, y) q_2(t, y) + \frac{1}{2} \sigma_v^2 q_2(t, y)^2, & q_2(0, y) = y_2, \\ \frac{\partial q_3}{\partial t}(t, y) = -(\kappa_1 - 1) q_1(t, y) - \beta_\lambda q_3(t, y) + e^{\gamma q_1(t, y) + \delta^2 q_1(t, y)^2/2} \frac{\zeta}{\zeta - q_3(t, y)} - 1, & q_3(0, y) = y_3 \end{cases}$$

where $t \in [0, T]$ and

$$p \colon t \mapsto p(t, y) \in \mathbb{R}, \quad q_1 \colon t \mapsto q_1(t, y) \in \mathbb{R}, \quad q_2 \colon t \mapsto q_2(t, y) \in \mathbb{R}, \quad q_3 \colon t \mapsto q_3(t, y) \in (-\infty, \zeta)$$

are C^1-functions.

Proof. Fixed $y \in \mathcal{Y}$ and $T \geq 0$, in Keller-Ressel and Mayerhofer (2015, Definition 2.10) it is stated that the extended Riccati system is given by

$$\begin{cases} \frac{\partial p}{\partial t}(t,y) = F(q(t,y)), & p(0,y) = 0 \\ \frac{\partial q_i}{\partial t}(t,y) = R_i(q(t,y)), & q_i(0,y) = y_i, \end{cases}$$

for $i = 1,2,3$, where $F, R_i \colon \mathcal{Y} \to \mathbb{R}$ are functions whose shape depends on the affine structure of Z, see Keller-Ressel and Mayerhofer (2015, Proposition 2.8). In our specific case, recalling that the characteristics of Z are given by (1.10), (1.11), (1.12), we have

$$F(y) = \beta_v \alpha_v y_2 + \beta_\lambda \alpha_\lambda y_3 \tag{1.22}$$

$$R_1(y) = 0 \tag{1.23}$$

$$R_2(y) = \frac{1}{2}y_1^2 + \rho\sigma_v y_1 y_2 + \frac{1}{2}\sigma_v^2 y_2^2 - \frac{1}{2}y_1 - \beta_v y_2 \tag{1.24}$$

$$\begin{aligned} R_3(y) = &-(\kappa_1 - 1)y_1 + \int_{\mathbb{R}} h(x)\theta^X(\mathrm{d}x)y_1 - \beta_\lambda y_3 + \int_{\mathbb{R}_+} h(x)\theta^\lambda(\mathrm{d}x)y_3 \\ &+ \int_{\mathbb{R}^3\setminus\{0\}} (e^{y^\top x} - 1 - h(x)^\top y)\theta^X(\mathrm{d}x_1)\delta_0(\mathrm{d}x_2)\theta^\lambda(\mathrm{d}x_3). \end{aligned} \tag{1.25}$$

Writing explicitly the scalar product, formula (1.25) becomes

$$\int_{\mathbb{R}^3\setminus\{0\}} (e^{x_1y_1}e^{x_2y_2}e^{x_3y_3} - 1 - h(x_1)y_1 - h(x_2)y_2 - h(x_3)y_3)\theta^X(\mathrm{d}x_1)\delta_0(\mathrm{d}x_2)\theta^\lambda(\mathrm{d}x_3).$$

To rewrite the integral of the sum as the sum of the integrals we need to verify that the integral of every addendum converges:

- Since $y \in \mathcal{Y}$, $y_3 < \zeta$. It follows that

$$\begin{aligned} &\int_{\mathbb{R}^3\setminus\{0\}} e^{x_1y_1}e^{x_2y_2}e^{x_3y_3}\theta^X(\mathrm{d}x_1)\delta_0(\mathrm{d}x_2)\theta^\lambda(\mathrm{d}x_3) \\ &= \int_{\mathbb{R}} e^{x_1y_1}\theta^X(\mathrm{d}x_1)\cdot\int_{\mathbb{R}} e^{x_2y_2}\delta_0(\mathrm{d}x_2)\cdot\int_{\mathbb{R}} e^{x_3y_3}\theta^\lambda(\mathrm{d}x_3) = e^{\gamma y_1 + \delta^2 y_1^2/2}\frac{\zeta}{\zeta - y_3} < \infty \end{aligned}$$

  from the properties of Gaussian and Exponential probability random variables.

- Clearly $\int_{\mathbb{R}^3\setminus\{0\}} 1\cdot\theta^X(\mathrm{d}x_1)\delta_0(\mathrm{d}x_2)\theta^\lambda(\mathrm{d}x_3) = 1 < \infty$.

- We also have that, for every $j = 1,2,3$,

$$\int_{\mathbb{R}^3\setminus\{0\}} h(x_j)y_j\theta^X(\mathrm{d}x_1)\delta_0(\mathrm{d}x_2)\theta^\lambda(\mathrm{d}x_3) = y_j\int_{\mathbb{R}^3\setminus\{0\}} \mathbb{1}_{|x_j|\le 1}x_j\theta^X(\mathrm{d}x_1)\delta_0(\mathrm{d}x_2)\theta^\lambda(\mathrm{d}x_3) < \infty.$$

Then, we get

$$R_3(y) = -(\kappa_1 - 1)y_1 - \beta_\lambda y_3 + e^{\gamma y_1 + \delta^2 y_1^2/2} \frac{\zeta}{\zeta - y_3} - 1. \tag{1.26}$$

Replacing q_i to y_i for $i = 1, 2, 3$ in equations (1.22), (1.23), (1.24), (1.26), one obtains exactly the system in the statement. □

Proposition 1.14 (Complex Riccati system)**.** *Let $T \geq 0$ and $u = (u_1, u_2, u_3) \in \mathcal{S}(\mathcal{Y}^\circ)$. The complex Riccati system associated to Z is* (1.17).

Proof. In Keller-Ressel and Mayerhofer (2015, Definition 2.22), the complex Riccati system is given by

$$\begin{cases} \frac{\partial \Phi}{\partial t}(t, u) = \bar{F}(\Psi(t, u)), & \Phi(0, u) = 0 \\ \frac{\partial \Psi_i}{\partial t}(t, u) = \bar{R}_i(\Psi(t, u)), & \Psi(0, u) = u, \end{cases}$$

where $\bar{F}, \bar{R}_i \colon \mathcal{S}(\mathcal{Y}^\circ) \longrightarrow \mathbb{C}$ are the analytic extensions of F, R_i, see equations (1.22), (1.23), (1.24), (1.26), to $\mathcal{S}(\mathcal{Y}^\circ)$. Note that, the analytic extensions of F, R_i have exactly the same analytical form of F, R_i. The complex Riccati is then given, for $u \in \mathcal{S}(\mathcal{Y}^\circ)$, by

$$\begin{cases} \frac{\partial \Phi}{\partial t}(t, u) = \alpha_v \beta_v \Psi_2(t, u) + \alpha_\lambda \beta_\lambda \Psi_3(t, u), & \Phi(0, u) = 0, \\ \frac{\partial \Psi_1}{\partial t}(t, u) = 0, & \Psi_1(0, u) = u_1, \\ \frac{\partial \Psi_2}{\partial t}(t, u) = -\frac{1}{2}\Psi_1(t, u) + \frac{1}{2}\Psi_1(t, u)^2 - \beta_v \Psi_2(t, u) + \rho \sigma_v y_1 \Psi_2(t, u) + \frac{1}{2}\sigma_v^2 \Psi_2(t, u)^2, & \Psi_2(0, u) = u_2, \\ \frac{\partial \Psi_3}{\partial t}(t, u) = -(\kappa_1 - 1)\Psi_1(t, u) - \beta_\lambda \Psi_3(t, u) + e^{\gamma \Psi_1(t,u) + \delta^2 \Psi_1(t,u)^2/2} \frac{\zeta}{\zeta - \Psi_3(t,u)} - 1, & \Psi_3(0, u) = u_3, \end{cases}$$

where $t \in [0, T]$ and

$$\Phi \colon t \mapsto \Phi(t, u) \in \mathbb{C}, \quad \Psi_{1,2} \colon t \mapsto \Psi_{1,2}(t, y) \in \mathbb{C}, \quad \Psi_3 \colon t \mapsto \Psi_3(t, y) \in \{u \in \mathbb{C} : \Re\mathfrak{e}(u) < \zeta\}$$

are C^1-functions. To see that it corresponds to (1.17) it is enough to notice that the solution to the ODE

$$\frac{\partial \Psi_1}{\partial t}(t, u) = 0, \qquad \Psi_1(0, u) = u_1$$

is $\Psi_1 \equiv u_1$ and to rename $\Phi = \phi, \Psi_2 = \psi, \Psi_3 = \chi$. □

Step 5) Conclusion. Let us sum up all the previous steps to obtain $\mathbb{E}[\exp(u^\top Z_T)]$ for $u \in \mathcal{S}(\mathcal{D}_{\mathcal{L}(Z_T)})$. If we take $u \in \mathcal{S}(\mathcal{D}_{\mathcal{L}(Z_T)}) \subset \mathcal{S}(\mathcal{Y})$, then by definition $\Re\mathfrak{e}(u) \in \mathcal{D}_{\mathcal{L}(Z_T)} \subseteq \mathcal{Y}$. By Step 2 we know that the extended Riccati system having as starting point $\Re\mathfrak{e}(u) \in \mathcal{D}_T$ has solution (p, q_1, q_2, q_3). By Step 3 we ensure that $q(t, \Re\mathfrak{e}(u)) \in \mathcal{Y}^\circ$. Applying Keller-Ressel and Mayerhofer (2015, Theorem 2.26), we get that also the complex Riccati system

(1.17), see Step 4, has unique solution $(\phi, \psi, \chi)(t, u)$ up to time $0 \le T < T^*$, then

$$\mathbb{E}[\exp(u^\top Z_T)] = \exp\left(\phi(T, u) + u_1 X_0 + \psi(T, u) V_0 + \chi(T, u)\lambda_0\right).$$

1.5 The hedging problem

Notation Let us introduce here some useful notation.

- For $u \in \mathbb{C}^n$, we denote by $\bar{u}$ its complex conjugate.
- $L^2(\mathbb{P})$ is the set of random variables with finite second order moment under $\mathbb{P}$.
- $\mathcal{H}^2_{\mathbb{C}}$ (resp. $\mathcal{H}^2$) is the *Hilbert space of càdlàg complex-valued (resp. real-valued)* $\mathbb{F}$*-adapted square-integrable martingales.*
- If $X, Y \in \mathcal{H}^2$, then $\langle X, Y\rangle$ denotes the predictable covariation of X and Y. Notice that X and Y are orthogonal if and only if $X_0 Y_0 = 0$ and $\langle X, Y\rangle = 0$.
- For $Z^j = X^j + iY^j \in \mathcal{H}^2_{\mathbb{C}}, j = 1, 2$

$$\langle Z^1, Z^2\rangle = (\langle X^1, X^2\rangle - \langle Y^1, Y^2\rangle) + i(\langle X^1, Y^2\rangle + \langle Y^1, X^2\rangle).$$

- For $X \in \mathcal{H}^2_{\mathbb{C}}$ (resp., $X \in \mathcal{H}^2$) we define the *space of complex-valued (resp. real-valued) integrands for* X as

$$L^2_{\mathbb{C}}(X) := \left\{\vartheta \text{ predictable and complex-valued: } \mathbb{E}\left[\int_0^T |\vartheta_s|^2 \, d\langle X, \bar{X}\rangle_s\right] < \infty\right\}$$

  (resp., $L^2(X)$).

1.5.1 Semi-static hedging variance-optimal strategies

Recall that S models the price process of a tradable asset. Consider a given claim $\eta^0 \in L^2(\mathbb{P})$, which we want to hedge, and a fixed basket of contingent claims $\boldsymbol{\eta} := (\eta^1, \dots, \eta^d)^\top$, $\eta^j \in L^2(\mathbb{P})$ for $j = 1, \dots, d$, that we want to use, together with S, to hedge η^0. Our aim is finding a semi-static variance-optimal hedging strategy in order to hedge η^0. Before introducing the actual optimization problem, let us give a brief explanation of what a *semi-static variance-optimal* strategy is.

In Föllmer and Sondermann (1986), the authors introduce variance-optimal hedging for the first time as a method to hedge contingent claims in incomplete markets. The underlying idea is to find a self-financing strategy, for a given claim η^0, which minimizes the risk-neutral variance of the residual hedging error at a terminal time $T > 0$. Always in Föllmer and Sondermann (1986), they show that the solution of this optimization problem is given by the Galtchouk-Kunita-Watanabe (GKW) decomposition of η^0 with respect to the price S, see Kunita and Watanabe (1967) for further details. This decomposition will be a key tool also in the more general case of semi-static variance-optimal hedging, as we will see later in this section.

We also require for a semi-static strategy, meaning that we aim at combining a dynamic (i.e., continuously rebalanced) position in S and a static (i.e., buy-and-hold) position in the other assets $\boldsymbol{\eta}$. Being a generalization of fully dynamic strategies, semi-static variance-optimal strategies typically allow for a reduction of the quadratic hedging error. Moreover, no rebalancing costs or liquidity risks are associated with the static part of the strategy, allowing the use of assets with limited liquidity as static hedging instruments. Semi-static strategies have appeared in mathematical finance in different contexts, see e.g. Carr (2011), Beiglböck et al. (2013). Since the ultimate goal of this work is the hedge of a variance swap, we bring attention to Neuberger (1994) and Carr and Madan (2001), where they show the semi-static replication of variance swaps by Neuberger's formula.

Definition 1.15. A *semi-static* strategy is a pair $(\vartheta, \boldsymbol{\nu}) \in L^2(S) \times \mathbb{R}^d$. Together with an initial wealth $c \in \mathbb{R}$ it is called variance-optimal hedging strategy for the square-integrable contingent claim η^0 if it is a solution to the *semi-static variance-optimal hedging problem*

$$\varepsilon^2 = \min_{\boldsymbol{\nu}\in\mathbb{R}^d, \vartheta\in L^2(S), c\in\mathbb{R}} \mathbb{E}\left[\left(c - \mathbb{E}[\boldsymbol{\nu}^\top \boldsymbol{\eta}] + \int_0^T \vartheta_s \, \mathrm{d}S_s - (\eta^0 - \boldsymbol{\nu}^\top \boldsymbol{\eta})\right)^2\right].$$

A semi-static strategy is a strategy which has both a dynamic position in S, denoted by ϑ, and a static position in the fixed basket $\boldsymbol{\eta} = (\eta^1, \dots, \eta^d)^\top$, which is denoted by $\boldsymbol{\nu}$.

In Di Tella et al. (2019), Di Tella et al. (2020), they show that the semi-static hedging problem can be split into an inner and an outer optimization problem, respectively

$$\varepsilon^2(\boldsymbol{\nu}) = \min_{\vartheta\in L^2(S), c\in\mathbb{R}} \mathbb{E}\left[\left(c - \mathbb{E}[\boldsymbol{\nu}^\top \boldsymbol{\eta}] + \int_0^T \vartheta_s \, \mathrm{d}S_s - (\eta^0 - \boldsymbol{\nu}^\top \boldsymbol{\eta})\right)^2\right] \tag{1.27}$$

$$\varepsilon^2 = \min_{\boldsymbol{\nu}\in\mathbb{R}^d} \varepsilon^2(\boldsymbol{\nu}). \tag{1.28}$$

The two problems can be solved separately. The inner problem, equation (1.27), is a classic variance-optimal hedging problem, see Föllmer and Sondermann (1986). As we have mentioned before, the solution of the problem is linked to the GKW decomposition.

Thus, before writing the solution of equation (1.27), let us introduce this key technical tool, referring to Kunita and Watanabe (1967).

Definition 1.16 (GKW decomposition)**.** Let $X, Y \in \mathcal{H}^2_{\mathbb{C}}$. Then there exist a unique $\vartheta \in L^2_{\mathbb{C}}(X)$ and a unique $L \in \mathcal{H}^2_{\mathbb{C}}$, $L_0 = 0$, such that $\langle X, \bar{L}\rangle = 0$ and the following decomposition holds:

$$Y_\cdot = Y_0 + \int_0^\cdot \vartheta_s \, \mathrm{d}X_s + L_\cdot.$$

The couple (L, ϑ) is called the *GKW decomposition of Y with respect to X*. From a financial point of view, the decomposition is made of: an initial capital Y_0, an investment strategy ϑ which helps dealing with the hedgeable risk and the residual risk L, the *orthogonal component*. Notice that $\vartheta = \mathrm{d}\langle Y, \bar{X}\rangle / \mathrm{d}\langle X, \bar{X}\rangle$.

The optimal solution of equation (1.27) is given, for a fixed $\boldsymbol{\nu} = (\nu_1, \nu_2, \dots, \nu_d)^\top$, by $(\vartheta^{\boldsymbol{\nu},*}, c^*)$ defined as

$$\vartheta^{\boldsymbol{\nu},*} = \vartheta^0 - \sum_{j=1}^d \nu_j \vartheta^j, \quad c^* = \mathbb{E}[\eta^0],$$

where (ϑ^j, L^j) are the GKW decomposition of η^j with respect to S for $j = 0, \dots, d$.

The outer problem in equation (1.28) is a finite-dimensional quadratic optimization problem. In the proof of Di Tella et al. (2020, Theorem 2.3) they show that $\varepsilon^2(\boldsymbol{\nu})$ can be written in the following form

$$\varepsilon^2(\boldsymbol{\nu}) = A - 2\boldsymbol{\nu}^\top B + \boldsymbol{\nu}^\top C \boldsymbol{\nu}, \tag{1.29}$$

where the coefficients A, B, and C are given by

$$A := \mathbb{E}[\langle L^0, L^0\rangle_T], \quad B^j := \mathbb{E}[\langle L^0, L^j\rangle_T], \quad C^{ij} := \mathbb{E}[\langle L^i, L^j\rangle_T], \quad i, j = 1, \dots, d. \tag{1.30}$$

Again in Di Tella et al. (2020, Theorem 2.3), they show that if C is invertible, then the optimal strategy $(\vartheta^*, \boldsymbol{\nu}^*, c^*)$ for the semi-static variance-optimal hedging problem (2.1) is then given by

$$\boldsymbol{\nu}^* = C^{-1}B, \quad \vartheta^* = \vartheta^0 - \sum_{j=1}^d \nu_j^* \vartheta^j, \quad c^* = \mathbb{E}[\eta^0] \tag{1.31}$$

and the minimal squared hedging error is given by

$$\varepsilon^2(\boldsymbol{\nu}^*) = A - B^T C^{-1} B. \tag{1.32}$$

Remark 1.17. Note that if C is non-invertible, then any $\nu \in \mathbb{R}^d$ which is solution of the linear system $C\nu = B$, together with $c = \mathbb{E}[\eta_T^0]$ and $\vartheta^\nu = \vartheta^0 - \sum_{j=1}^d \nu_j \vartheta^j$, is a solution of the semistatic hedging problem. The solution that minimizes the Euclidian norm of ν can be obtained by setting $\nu = C^* B$, where C^* denotes the Moore–Penrose pseudo-inverse of C.

Thus, to solve the semi-static variance-optimal hedging problem, it is necessary to:

a) calculate the predictable covariations of the residuals L^j in the GKW decomposition of η^j with respect to S for $j = 0, \ldots, d$;

b) take the expectations of these predictable covariations to obtain A, B, C as by equation (1.30).

On this regard, we recall in Appendix 1.A the key Corollary 3.1 from Di Tella et al. (2019), which provides the GKW and the predictable covariations in the case where the claims are functions of semimartingales with affine characteristics. In what follows we will see how the Fourier representation of the claims is a key tool to compute the three quantities above and how it can be combined with Corollary 3.1 from Di Tella et al. (2019).

1.5.2 Solving the semi-static hedging problem via Fourier representation

Variance-optimal hedging has been historically combined with Fourier methods and we refer the reader to Di Tella et al. (2019), where this idea is exploited in a very general factor model, and when dealing with semi-static hedging.

Consider a general model $\mathbf{Z} = (\mathbf{Z}^1, \ldots, \mathbf{Z}^n)$. We should be cautious not to confuse Z, our model specified in Section 1.2, with $\mathbf{Z}$, a general model taking values in $\mathbb{R}^n$. Assume $\mathbf{Z}^1$ to be the log-price process of the underlying stock, i.e., $S = e^{\mathbf{Z}^1}$. Consider a European option with payoff $\eta = h(\mathbf{Z}_T^1)$, $\eta \in L^2(\mathbb{P})$, for some *real-valued function* h with domain in $\mathbb{R}$. Assume the two-sided Laplace transform of h, denoted $\widehat{h}$, exists in $R \in \mathbb{R}$ and that it is integrable on the strip $\mathcal{S}(R) := \{u \in \mathbb{C}^n : \mathfrak{Re}(u) = R\}$. Then, h has the following representation:

$$h(z) = \frac{1}{2\pi i} \int_{R-i\infty}^{R+i\infty} \exp(u^\top z) \widehat{h}(u)\, \mathrm{d}u = \int_{\mathcal{S}(R)} \exp(u^\top z) \zeta(\mathrm{d}u), \tag{1.33}$$

where ζ is the finite complex-valued absolutely continuous measure on $\mathcal{S}(R)$ defined by

$$\zeta(\mathrm{d}u) := \frac{1}{2\pi i} \widehat{h}(u)\, \mathrm{d}u. \tag{1.34}$$

If the integrability condition $\mathbb{E}[e^{R\mathbf{Z}_T^1}] < \infty$ holds, then the risk-neutral price of the claim η at time $t \in [0, T]$ can be recovered by the Fourier-type integral

$$\eta_t = \int_{\mathcal{S}(R)} \eta_t(u)\zeta(\mathrm{d}u) \tag{1.35}$$

where

$$\eta_t(u) = \mathbb{E}[e^{u\mathbf{Z}_T^1}|\mathcal{F}_t].$$

Remark 1.18. In the case where η is an European Call (resp. Put), i.e., for a fixed a strike $K > 0$, $h(z) = (e^z - K)^+$ (resp. $h(z) = (K - e^z)^+$), we have that

$$\widehat{h}(u) = \frac{K^{1-u}}{u(u-1)}, \tag{1.36}$$

for $R > 1$ (resp. $R < 0$). See Hubalek et al. (2006, Section 4) for further details.

Now that we have highlighted how the Fourier method is applied for the pricing of European options, we can explain its application to the semi-static hedging case. Recall that the solution of a general semi-static hedging optimization problem, given by equation (1.31), is characterized by the GKW representation of the claims. The core idea is that the Fourier representation of the claim in equation (1.35) and the GKW decomposition of the claim with respect to S can be interchanged.

Similarly to before, consider a factor process $\mathbf{Z} = (\mathbf{Z}^1, \ldots, \mathbf{Z}^n)$, such that the underlying stock $S = e^{\mathbf{Z}^1} \in \mathcal{H}^2$. Let $R_j \in \mathbb{R}$ be such that $\mathbb{E}[e^{2R_j\mathbf{Z}_T^1}] < \infty$ and define $\mathcal{S}^j := \{u \in \mathbb{C} : \mathfrak{Re}(u) = R_j\}$, $j = 1, \ldots, d$, and the strip $\mathcal{S} := \bigcup_{j=1}^d \mathcal{S}^j$. Assume to be working with a family of real-valued square-integrable payoffs of European options $\eta^j = h^j(\mathbf{Z}_T^1)$, $j = 1, \ldots, d$, where h^j admits a Fourier representation, as in equations (1.33), (1.34), on $\mathcal{S}_j$ and w.r.t. the measures ζ^j.

In Di Tella et al. (2019, Theorem 4.1, Theorem 4.2), they show how to write the GKW representation of the claims η^j in terms of the GKW representation of a more general object

$$\eta_t(u) = \mathbb{E}[e^{u\mathbf{Z}_T^1}|\mathcal{F}_t], \quad t \in [0, T], u \in \mathcal{S}. \tag{1.37}$$

We denote by (ϑ^j, L^j) the GKW decomposition of η^j with respect to S and by $(\vartheta(u), L(u))$, $u \in \mathcal{S}$, those of $\eta_t(u)$ (introduced in equation (1.37)) with respect to S, for $u \in \mathcal{S}$. The following formulas hold:

- From Di Tella et al. (2019, Theorem 4.1, equation (4.10)) we get the hedging strategy

$$\vartheta^j = \int_{\mathcal{S}^j} \vartheta(u)\, \zeta^j(\mathrm{d}u).$$

- From Di Tella et al. (2019, Theorem 4.2, equation (4.12)) we have, for $i, j = 1, \dots, d$,

$$\langle L^i, L^j \rangle_T = \int_0^T \int_{\mathcal{S}^i} \int_{\mathcal{S}^j} \mathrm{d}\langle L(u_i), L(u_j)\rangle_t \zeta^j(\mathrm{d}u_j)\zeta^i(\mathrm{d}u_i).$$

Proposition 1.19. *Consider a factor process $\mathbf{Z} = (\mathbf{Z}^1, \dots, \mathbf{Z}^n)$, such that the underlying stock $S = e^{\mathbf{Z}^1} \in \mathcal{H}^2$. Let $R_j \in \mathbb{R}$ be such that $\mathbb{E}[e^{2R_j \mathbf{Z}^1_T}] < \infty$ and define $\mathcal{S}^j := \{u \in \mathbb{C} : \Re\mathfrak{e}(u) = R_j\}, j = 1, \dots, d$ and the strip $\mathcal{S} := \bigcup_{j=1}^d \mathcal{S}^j$. Consider some real-valued square-integrable European options $\eta^j = h^j(\mathbf{Z}^1_T)$, $j = 1, \dots, d$ where h^j has a Fourier representation as in equations (1.33), (1.34) on $\mathcal{S}_j$ w.r.t. measures ζ^j. Then A, B, C in equations (1.30) are obtained via integration as follows*

$$\begin{aligned}
A &= \int_0^T \mathbb{E}\left[\frac{\mathrm{d}\langle L^0, L^0\rangle_t}{\mathrm{d}t}\right] \mathrm{d}t,\\
B^j &= \int_0^T \int_{\mathcal{S}^j} \mathbb{E}\left[\frac{\mathrm{d}\langle L^0, L(u_j)\rangle_t}{\mathrm{d}t}\right] \zeta^j(\mathrm{d}u_j)\mathrm{d}t,\\
C^{i,j} &= \int_0^T \int_{\mathcal{S}^i} \int_{\mathcal{S}^j} \mathbb{E}\left[\frac{\mathrm{d}\langle L(u_i), L(u_j)\rangle_t}{\mathrm{d}t}\right] \zeta^j(\mathrm{d}u_j)\zeta^i(\mathrm{d}u_i)\mathrm{d}t,
\end{aligned}$$

for $i, j = 1, \dots, d$.

Proof. Refer to Di Tella et al. (2019, Theorem 4.1, Theorem 4.2). □

Remark 1.20. Exploiting the Fourier transform of the payoff of the contingent claims η^j to compute the semi-static hedging strategies has numerous advantages:

- Instead of computing the GKW decomposition for each claim η^j, it is enough to compute it once for $\eta_t(u)$ and then to obtain A, B, C by integration as in Proposition 1.19.

- To compute the GKW decomposition for each claim η^j in the case of an affine model, one might exploit Di Tella et al. (2019, Corollary 3.1), recalled in Appendix 1.A. However, to apply the Corollary, one needs the payoff h^j to be at least two times differentiable, which is not true for example in the case of European Calls and Puts.

On the other hand, for many affine models the Laplace transform $\eta_t(u)$ is a smooth function.

- Note that $\eta_t(u)$ in equation (1.37) is the complex, conditional Laplace transform of $\mathbf{Z}^1$. This quantity is available in closed form in many models (such as affine models).

1.6 Application: hedging variance swaps

We now would like to solve the semi-static hedging problem introduced in Section 1.5 in the case where η^0 is a variance swap written on the stock $S = e^X$, where X is the log-price defined in equation (1.7a). More precisely, the payoff at maturity $T > 0$ of the variance swap is defined as

$$\eta^0_T = [X, X]_T - k, \tag{1.38}$$

where k is the so-called swap rate, i.e., $k = \mathbb{E}[[X,X]_T]$ so that $\mathbb{E}[\eta^0_T] = 0$ and the contract is zero at inception. The set of contingent claims used to hedge is a basket $(\eta^1, \dots, \eta^d)$ of European options written on S.

As we stated in Section 1.5, the hedging error uniquely depends on the quantities A, B, C in equation (1.30). In the next proposition, these key three quantities will be characterized for this specific derivative.

Theorem 1.21. *Let $Z = (X, V, \lambda)$ be given by the model in equations (1.7) and let the claim η^0 be a variance swap on $S = e^X$ and the contingent claims $\boldsymbol{\eta} = (\eta^1 \dots, \eta^d)^\top$ be European puts and calls with payoffs $h^j(S_T)$, $j = 1, \dots, d$. We denote by $\widehat{h^j}$ the two-sided Laplace transforms of h^j as in equation (1.36), which are integrable along strips $\mathcal{S}^j$, where $\mathcal{S}^j = \{z \in \mathbb{C} : \Re\mathfrak{e}(z) = R_j\}$, such that $\mathbb{E}[e^{2R_j X_T}] < \infty$. Fix $0 < T < T^*$ where T^* is the explosion time in the Heston model. Then the dynamic hedging strategies ϑ^0 and ϑ^j, $j = 1, \dots, d$ and the coefficients A, B, C are given by*

$$\vartheta^0_t = \frac{1}{S_{t-}(V_t + \lambda_t \bar{\kappa})} \left(\Theta^0_1(t) V_t + \Theta^0_2(t) \lambda_{t-}\right),$$

$$\vartheta^j_t = \frac{1}{S_{t-}(V_t + \lambda_{t-} \bar{\kappa})} \int_{\mathcal{S}^j} \left(\Theta^j_1(t, u_j) f_{t-}(u_j) V_t + \Theta^j_2(t, u_j) f_{t-}(u_j) \lambda_{t-}\right) \zeta^j(\mathrm{d}u_j),$$

$$A = \int_0^T A_1(t) \mathbb{E}\left[\frac{V_t^2}{V_t + \lambda_{t-}\bar{\kappa}}\right] + A_2(t) \mathbb{E}\left[\frac{\lambda_t V_t}{V_t + \lambda_{t-}\bar{\kappa}}\right] + A_3(t) \mathbb{E}\left[\frac{\lambda_t^2}{V_t + \lambda_{t-}\bar{\kappa}}\right] \mathrm{d}t,$$

$$B^j = \int_0^T \int_{\mathcal{S}^j} B_1(t, u_j) \mathbb{E}\left[\frac{f_{t-}(u_j) V_t^2}{V_t + \lambda_{t-}\bar{\kappa}}\right] + B_2(t, u_j) \mathbb{E}\left[\frac{f_{t-}(u_j) \lambda_t V_t}{V_t + \lambda_{t-}\bar{\kappa}}\right]$$
$$+ B_3(t, u_j) \mathbb{E}\left[\frac{f_{t-}(u_j) \lambda_t^2}{V_t + \lambda_{t-}\bar{\kappa}}\right] \zeta^j(\mathrm{d}u_j)\mathrm{d}t,$$

$$C^{i,j} = \int_0^T \int_{\mathcal{S}^i} \int_{\mathcal{S}^j} C_1(t, u_i, u_j)\mathbb{E}\left[\frac{f_{t-}(u_i)f_{t-}(u_j)V_t^2}{V_t + \lambda_{t-}\bar{\kappa}}\right] + C_2(t, u_i, u_j)\mathbb{E}\left[\frac{f_{t-}(u_i)f_{t-}(u_j)\lambda_t V_t}{V_t + \lambda_{t-}\bar{\kappa}}\right]$$
$$+ C_3(t, u_i, u_j)\mathbb{E}\left[\frac{f_{t-}(u_i)f_{t-}(u_j)\lambda_t^2}{V_t + \lambda_{t-}\bar{\kappa}}\right] \zeta^i(\mathrm{d}u_i)\zeta^j(\mathrm{d}u_j)\mathrm{d}t,$$

where $\bar{\kappa} = (e^{2(\gamma+\delta^2)} - 2e^{\gamma+\delta^2/2} + 1)$ *and*

$$f_{t-}(u_j) = \exp(\phi_{T-t}(u_j) + u_j X_{t-} + \psi_{T-t}(u_j)V_{t-} + \chi_{T-t}(u_j)\lambda_{t-}) \tag{1.39}$$

with $\Theta_k^0, \Theta_k^j, A_k, B_k, C_k$ *deterministic functions of* t, u_i, u_j*, which are made precise in the proof and* $g_t(u_j) = g(t, u_j, 0, 0)$*, for* $g = \phi, \psi, \chi$ *as defined as in Theorem 1.8.*

Proof. The proof is provided in Appendix 1.C. □

1.7 Numerical results

In this section, we focus on the computation of A, B, C and of the hedging error $\varepsilon^2(\boldsymbol{\nu}^*)$ for a variance swap hedged by a set of European Call options. In particular, we want to perform some parameter-sensitivity tests. We executed the code via the cluster of the Department of Mathematics of University of Padova, using a configuration of 10 cores and 32GB of RAM. We report that the average time for computing A, B, C for $d = 21$ options was about 40 minutes.

The computations involved in this task are far from being numerically trivial, in that they involve:

- The computation of the solutions ϕ, ψ, χ to the generalized Riccati system in (1.17), useful to obtain $f_{t-}(u_j)$ as in equation (1.39) and the deterministic coefficients A_k, B_k, C_k (see Section 1.C),
- The expectations in A, B, C in Theorem 1.21,
- Multiple integrals over the complex strips $\mathcal{S}^j := \{u \in \mathbb{C} : \mathfrak{Re}(u) = R_j\}, j = 1, \ldots, d$, where d is the number of vanilla options used to hedge.

We highlight that ψ is computed in closed form, recall Remark 1.9. On the other hand, ϕ and χ must be computed numerically and for this we used the built-in Python ODE solver, `scipy.integrate.solve_ivp`. The explicit Runge-Kutta method of order 5(4) (RK45)—the default method—was employed, along with the default tolerances: a relative tolerance (`rtol`) of 1e-3 and an absolute tolerance (`atol`) of 1e-6. The expectations appearing in A, B, C in Theorem 1.21 are computed via Monte-Carlo with a relatively small number of trajectories, equal to $M = 10^4$ (we observed no significant improvement when increasing to $M = 10^5$, so we chose $M = 10^4$ in order to reduce the computational effort). We choose $T = 1$, taking $N = 100$ equispaced timesteps.

Remark 1.22. Despite the availability of closed formulas for the moments, obtained in Section 1.D, we decided to use Monte Carlo for the sake of computational time. Indeed, closed formulas, yet requiring integration of partial derivatives of ϕ, ψ, χ, produce results close to the benchmark, obtained with Monte Carlo with $M = 10^5$, but with higher computational time.

To simulate the trajectories of λ, V and then X, we use the following algorithm, partially inspired by Brignone et al. (2024):

- For a given λ_0, simulate trajectories of λ_t and J_t^λ, as in Dassios and Zhao (2013). From (1.3), compute $\int_0^t \lambda_s \mathrm{d}s = -\frac{\lambda_t - \lambda_0 - \beta_\lambda \alpha_\lambda - J_T^\lambda}{\beta_\lambda}$.

- Being V a CIR process, for $t \in [0, T]$ and $u \in [0, t]$, V_u is given and so $V_t = \frac{\sigma_v^2(1-e^{-\beta_v(t-u)})}{4\beta_v} \chi_d'^2(k)$ where $\chi_d'^2(k)$ is a non-central chi-squared distribution with $d := 4\alpha_v \beta_v / \sigma_v^2$ degrees of freedom and non-centrality parameter $k := \frac{4\beta_v e^{-\beta_v(t-u)} V_u}{\sigma_v^2(1-e^{-\beta_v(t-u)})}$.

- Simulate X using a Euler scheme.

Observe that in B and C we are interested in calculating the expected value of quantities depending on u_j and (u_i, u_j), respectively, where $u_j \in \mathcal{S}^j = \{z \in \mathbb{C} : \Re\mathrm{e}(z) = R_j\}$, as defined in Theorem 1.21. Since $\mathcal{S}^j$ is an unbounded domain, it must be approximated by a bounded one, that we will denote with $R_j + iS_{N^j}$. Since we are dealing with Call options only, we can consider only one strip $\mathcal{S}^C$. We choose $R^C = 2$, see Remark 1.18 and choose as S_{N^C} the interval $[-30, 30]$ partitioned via $N^C = 40$ equispaced points. We also tested the algorithm using various subintervals and different levels of grid's refinement to approximate $\mathcal{S}^C$ and the choice of S_{N^C} indicated above ultimately provided the best trade-off between precision and computation effort. Notice that one has to compute:

- $3 \cdot N^C \cdot N$ expectations for A.

- $3 \cdot N^C \cdot N$ expectations for B.

- $3 \cdot N^C \cdot N^C \cdot N$ expectations for C.

Then one integrates over the appropriate domain with respect to the measure $\zeta^C(\mathrm{d}u) = \frac{1}{2\pi i}\frac{K^{1-u}}{u(u-1)}\mathrm{d}u$, in particular we observe that the measure changes if we consider options having different strikes. To compute the integrals, we used the trapezoidal rule via `numpy.trapz`.

We refer to the parameters in Table 1.2 as the *standard set of parameters*:

X_0	γ	δ	ρ	V_0	α_v	β_v	σ_v	λ_0	α_λ	β_λ	ζ
4.605	0.06	0.02	-0.7165	0.0174	0.0354	1.3253	0.3877	3	2	1	2.5

Table 1.2: Standard set of parameters.

This corresponds to a model where the log-spot price jumps with intensity λ and the jumps' sizes are $\eta_1^X \sim \mathcal{N}(0.06, 0.0004)$. The intensity is itself stochastic, starting at $\lambda_0 = 3$, with jumps of size $\eta_1^\lambda \sim Exp(2.5)$, and shot-noise decay.

The assumptions in Table 1.1 are satisfied, together with the additional condition required in Theorem 1.21, $\mathbb{E}[e^{2R^C X_T}] < \infty$, with $R^C = 2$. We notice, in the light of Proposition 1.7, that the latter condition corresponds to verifying:

$$\begin{cases} 0 < \zeta - T\mathbb{E}(e^{2R^C|\eta_1^X|}) \\ 2R^C(\kappa_1 - 1) + 1 > 0. \end{cases}$$

The parameters of V are chosen consistently with those proposed for the Heston model in Gatheral (2006) and, for comparison, they are also consistent with those in Di Tella et al. (2020). In Figure 1.1, we compare trajectories of the asset S depicted in our model (right) and in the Heston one (left). We observe a clear presence of positive jumps in the right-hand-side picture. Moreover, compared to the left panel, the trajectories on the right span a wider range of values, reflecting the impact of self-exciting jumps.

Then, we compute the error $\varepsilon^2(\boldsymbol{\nu}^*)$, as defined in equation (1.32). We select a set of European call options as contingent claims, with 21 strikes evenly spaced between $K = 100$ and $K = 200$. The dimension d of the basket of contingent claims is varied: specifically, for $d = 1$, the basket includes only the call option with strike 100; for $d = 2$, the call with strike 105 is added; and so on, progressively including options with higher strikes as d increases. We compare the hedging error in four different scenarios:

1. The *standard set of parameters* in Table 1.2,

2. $\beta_\lambda = 8$, and the other parameters as in Table 1.2,

3. $\zeta = 1000$, and the other parameters as in Table 1.2. With this choice, we want to study the case where the intensity is close to the deterministic function $\alpha_\lambda + (\lambda_0 - \alpha_\lambda)e^{-\beta_\lambda t}$.

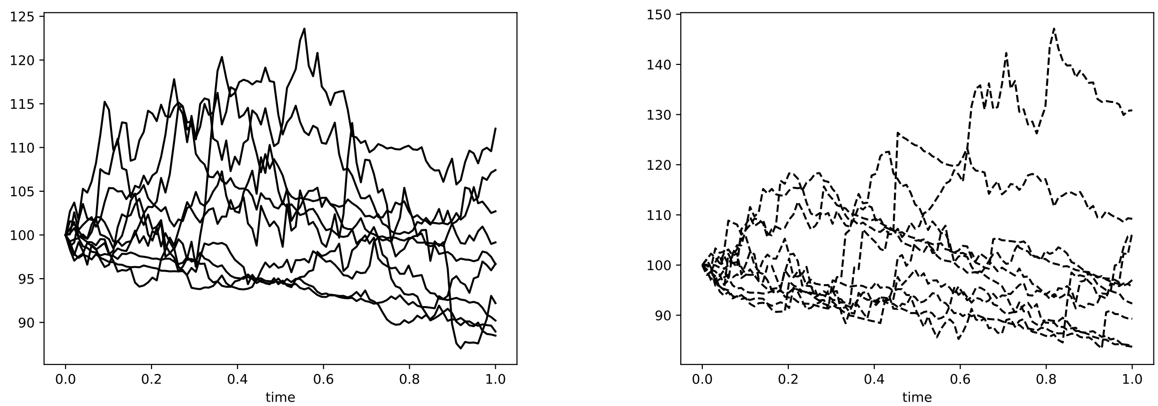


Figure 1.1: Trajectories in the Heston model (left) and in our model (right), equation (1.7).

4. A proxy of Heston model: $\gamma, \delta \approx 0$, and the other parameters as in Table 1.2.

Notice that for every analyzed scenario, the assumptions in Table 1.1 are satisfied. First, we compute the quantity A (recall (1.29)), which corresponds to the hedging error when the strategy is fully dynamic (no static hedging with contingent claims, $\boldsymbol{\nu} = 0$). We report the results in Table 1.3.

A	1	2	3	4 - Heston
$\varepsilon^2(0)$	$4.486 \cdot 10^{-4}$	$3.946 \cdot 10^{-4}$	$4.069 \cdot 10^{-4}$	$2.278 \cdot 10^{-4}$

Table 1.3: Hedging error when $\boldsymbol{\nu} = 0$ in the four scenarios.

We notice that the smallest error $\varepsilon^2(0)$ occurs in the Heston's case (case 4): as expected, introducing jumps of Hawkes type, hence more randomness, leads to a larger hedging error. We observe that the largest $\varepsilon^2(0)$ error appears in case 1, when we consider a stochastic intensity λ which starts at $\lambda_0 = 3$, decays with $\beta_\lambda = 1$ and whose jumps have size with law $Exp(2.5)$, hence, roughly speaking, when the intensity is bigger. This is reasonable as we expect that an asset with an unstable behavior is harder to hedge. Comparing cases 1 and 2, we observe that considering a larger decaying factor results in a smaller error. As the intensity decays more rapidly, the impact of jumps decreases quickly, reducing the overall error. The same observations holds for case 3: here the intensity's jumps sizes are smaller, hence the hedging error is smaller . We notice that the errors in case 2 and 3 are really close.

We then compute the hedging error $\varepsilon(\boldsymbol{\nu}^*)^2$ with European Calls, switching on the static hedging, hence in a semi-static setting. We highlight that in all the four cases, the symmetric matrix C was ill-conditioned. The same issue was spotted in Di Tella et al.

(2020). Instead of inverting the matrix directly, we use its Moore-Penrose pseudo-inverse (via `numpy.linalg.pinv` in Python). The error is graphically depicted in Figure 1.2, as a function of $d \in \{1, 2, \ldots, 21\}$. We also consider a relative gain $|\varepsilon(\boldsymbol{\nu}^*)^2 - A|/A$, which represents the relative improvement we have with respect to the fully dynamic error A. In all scenarios, we observe a general decreasing behavior of the absolute error $\varepsilon^2(\boldsymbol{\nu}^*)$ as

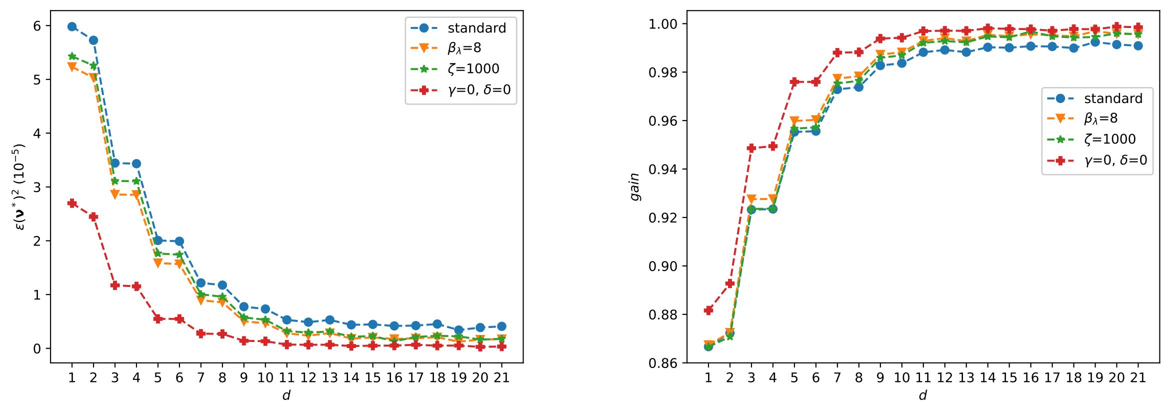


Figure 1.2: The absolute error (left) and relative gain (right), respectively.

a function of d, the number of contingent claims in the hedging basket, while the relative gain increases as d increases. The error for the Heston's case is the lowest: this is reasonable since we have perfect semi-static hedging of the variance swap in the Heston model, see Neuberger (1994). Consistently with Table 1.3, the other scenarios incorporate self-exciting jumps, which introduce additional risk and, consequently, higher hedging errors. We highlight that considering a static component in the hedging strategy reduces the error also in the presence of jumps, as it is visible in Figure 1.2 (left). Remarkably, see Figure 1.2 (right), incorporating just one option in the basket leads to an improvement of approximately 85%. As more options are included, the error continues to decrease and stabilizes after $d \geq 10$. In both the standard Heston case and Heston with self-exciting jumps, we achieve the same order of magnitude for the hedging error. These results confirm that effective semi-static hedging remains achievable even when self-exciting jumps are present. This strategy provides significant error reduction and improved hedging performance with potentially only a small number of additional contingent claims.

1.8 Conclusions

We investigated the affine structure of the model proposed, which exhibits several features including stochastic volatility (of Heston type) and self-exciting jumps (of Hawkes type with exponential memory kernel), by computing the conditional Laplace transform and providing conditions granting the existence of finite moments. Starting from these results, we computed a semistatic hedging strategy for a variance swap, minimizing the variance of the replication error at maturity, with a portfolio of European Call options. The explicit computation of the strategy, which is in close form, requires a non-trivial numerical approach. We therefore provided some numerical results on a specific example.

Although in the example provided we chose, for simplicity of illustration, to use a portfolio composed only by European Call options, we aim at investigating a similar numerical procedure using both Call and Put options. Interesting next research directions could be: the computation of hedging strategies for derivatives with different and more general payoffs; dealing with different maturities and more general model settings, eventually including non-Markovian features. A deep learning approach will certainly allow to overcome the non-trivial numerical issues raising in these more general settings.

1.A A key result

We recall below the crucial result Di Tella et al. (2019, Corollary 3.1).

Corollary 1.A.1. *Consider* $\mathbf{Z} = (\mathbf{Z}^1, \dots, \mathbf{Z}^n)^\top$ *a quasi-left continuous locally square-integrable semimartingale with state space* $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ *and denote by* $(\mathbf{b}, \mathbf{c}, \mathbf{K})$ *its predictable differential characteristics. Define* $S := e^{\mathbf{Z}^1}$ *and assume* $S \in \mathcal{H}^2$*. Then, if* $Y^i \in \mathcal{H}^2_{loc}$*, for* $i = 1, 2$*, is such that* $Y^i_t = f^i(t, \mathbf{Z}_t)$ *for functions* $f^i \in C^{1,2}(\mathbb{R}_+ \times \mathbb{R}^n)$*, the GKW decomposition* (ϑ^i, L^i) *of* Y^i *w.r.t.* S *is given by* $L^i = Y^i - Y^i_0 - \int_0^\cdot \vartheta^i_s \, \mathrm{d}S_s$*. Setting*

$$\xi_t := \mathbf{c}^{11}_t + \int_{\mathbb{R}^n} (e^{x_1} - 1)^2 \mathbf{K}_t(\mathrm{d}x), \quad t \in [0, T], \tag{1.40}$$

we get, Lebesgue-a.e.,

$$\vartheta^i_t = \frac{1}{S_{t-}\xi_t} \Big(\sum_{j=1}^n \partial_{x_j} f^i(t, \mathbf{Z}_{t-}) \mathbf{c}^{1j}_t + \int_{\mathbb{R}^n} (e^{x_1} - 1) W^i(t, x) \mathbf{K}_t(\mathrm{d}x) \Big), \quad i = 1, 2$$

and

$$\begin{aligned}\frac{\mathrm{d}\langle L^1, L^2\rangle_t}{\mathrm{d}t} = &\sum_{j=1}^n \sum_{k=1}^n \partial_{x_j} f^1(t, \mathbf{Z}_{t-}) \partial_{x_k} f^2(t, \mathbf{Z}_{t-}) \Big(\mathbf{c}^{jk}_t - \frac{\mathbf{c}^{1j}_t}{\xi_t} \mathbf{c}^{1k}_t \Big) \\ &- \frac{1}{\xi_t} \sum_{j=1}^n \partial_{x_j} f^1(t, \mathbf{Z}_{t-}) \mathbf{c}^{j1}_t \int_{\mathbb{R}^n} (e^{x_1} - 1) W^2(t, x) \mathbf{K}_t(\mathrm{d}x)\end{aligned}$$

$$- \frac{1}{\xi_t} \sum_{j=1}^{n} \partial_{x_j} f^2(t, \mathbf{Z}_{t-}) \mathbf{c}_t^{j1} \int_{\mathbb{R}^n} (e^{x_1} - 1) W^1(t, x) \mathbf{K}_t(\mathrm{d}x) + \int_{\mathbb{R}^n} W^1(t, x) W^2(t, x) \mathbf{K}_t(\mathrm{d}x)$$
$$- \frac{1}{\xi_t} \Big(\int_{\mathbb{R}^n} (e^{x_1} - 1) W^1(t, x) \mathbf{K}_t(\mathrm{d}x) \Big) \Big(\int_{\mathbb{R}^n} (e^{x_1} - 1) W^2(t, x) \mathbf{K}_t(\mathrm{d}x) \Big)$$

where

$$W^i(t, x) := f^i(t, x + \mathbf{Z}_{t-}) - f^i(t, \mathbf{Z}_{t-}), i = 1, 2.$$

1.B A useful lemma

We start with a general result on the exponential moment of the integral of a Poisson process.

Lemma 1.B.1. *On a given filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, where $\mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}$ satisfies the usual hypotheses, consider a Poisson process N with $\mathbb{F}$-intensity $\lambda > 0$ and a non-negative and $\mathbb{F}$-predictable stochastic process $(b_t)_{t\in[0,T]}$. We have*

$$\mathbb{E}\left[e^{\int_0^T b_t dN_t}\right] = \mathbb{E}\left[e^{\int_0^T (e^{b_t}-1)\lambda \mathrm{d}t}\right]. \tag{1.41}$$

The identity holds also for general $\mathbb{F}$-predictable stochastic process $(b_t)_{t\in[0,T]}$, provided that $\mathbb{E}\left[e^{\int_0^T e^{|b_t|} \mathrm{d}t}\right] < \infty$.

Proof. If b is bounded and $\mathbb{F}$-predictable process, it is enough to prove the result for any arbitrary process $b_t = \mathbb{1}_{(t_1,t_2]}(t)\mathbb{1}_A, 0 \le t_1 < t_2 \le T, A \in \mathcal{F}_{t_1}$. So, consider $0 \le t_1 < t_2 \le T$ and $A \in \mathcal{F}_{t_1}$ and denote by A^c the complementary set of A. We have:

$$\mathbb{E}\left[e^{\int_0^T b_t\, dN_t}\right] = \mathbb{E}\left[e^{\int_{t_1}^{t_2} \mathbb{1}_A\, dN_t}\right] = \mathbb{E}\left[e^{\mathbb{1}_A(N_{t_2}-N_{t_1})}(\mathbb{1}_A + \mathbb{1}_{A^C})\right] = \mathbb{E}\left[\mathbb{E}[e^{(N_{t_2}-N_{t_1})} \mid \mathcal{F}_{t_1}]\mathbb{1}_A + \mathbb{1}_{A^C}\right]$$
$$= \mathbb{E}\left[\mathbb{E}[e^{(N_{t_2}-N_{t_1})}]\mathbb{1}_A + \mathbb{1}_{A^C}\right] = \mathbb{E}\left[e^{(e-1)\lambda(t_2-t_1)}\mathbb{1}_A + \mathbb{1}_{A^C}\right] = \mathbb{E}\left[e^{(e-1)\lambda(t_2-t_1)\mathbb{1}_A}\right].$$

On the other hand, we notice that, almost surely,

$$e^{b_t} - 1 = e^{\mathbb{1}_{(t_1,t_2]}(t)\mathbb{1}_A} - 1 = e \cdot \mathbb{1}_{(t_1,t_2]}(t)\mathbb{1}_A - \mathbb{1}_{(t_1,t_2]}(t)\mathbb{1}_A = (e-1)\mathbb{1}_{(t_1,t_2]}(t)\mathbb{1}_A$$

and so

$$\mathbb{E}\left[e^{\int_0^T (e^{b_t}-1)\lambda\, \mathrm{d}t}\right] = \mathbb{E}\left[e^{\int_0^T (e-1)\mathbb{1}_{(t_1,t_2]}(t)\mathbb{1}_A \lambda\, \mathrm{d}t}\right] = \mathbb{E}\left[e^{(e-1)\lambda(t_2-t_1)\mathbb{1}_A}\right],$$

which proves the statement for any bounded $\mathbb{F}$-predictable process.

- In case when b is non-negative and $\mathbb{F}$-predictable but unbounded, let us define, for $n \ge 0$, the bounded sequence $b_t^n = b_t \wedge n$. We clearly have that $b_t^n \le b_t^{n+1}$ a.s. and when n goes to $+\infty$, $b_t^n \to b_t$ a.s. for every $t \in [0,T]$. Now define the sequences of random variables

$$X_n = e^{\int_0^T b_t^n \mathrm{d}N_t}, \qquad Y_n = e^{\int_0^T (e^{b_t^n}-1)\lambda \mathrm{d}t}, \quad n \geq 0.$$

We have $X_n \leq X_{n+1}$ a.s. and $X_n \to X = e^{\int_0^T b_t \mathrm{d}N_t}$ a.s. and $Y_n \leq Y_{n+1}$ a.s. and $Y_n \to Y = e^{\int_0^T (e^{b_t}-1)\lambda \mathrm{d}t}$ a.s. Moreover, since b_t^n is bounded for every $t \in [0,T]$ and for every $n \geq 0$, we have, by applying the statement in the bounded case, $\mathbb{E}[e^{\int_0^T b_t^n \mathrm{d}N_t}] = \mathbb{E}[e^{\int_0^T (e^{b_t^n}-1)\lambda \mathrm{d}t}]$ and by monotone convergence it follows that $\mathbb{E}[e^{\int_0^T b_t \mathrm{d}N_t}] = \mathbb{E}[e^{\int_0^T (e^{b_t}-1)\lambda \mathrm{d}t}]$.

- For a general, $\mathbb{F}$-predictable b, we define for $n \geq 0$, the bounded sequence $b_t^n = b_t \mathbb{1}_{|b_t \leq n|}$. We observe that $|b_t^n| \leq |b_t|$ for every n. Similarly to the positive case, $X_n = e^{\int_0^T b_t^n \mathrm{d}N_t}$, $Y_n = e^{\int_0^T (e^{b_t^n}-1)\lambda \mathrm{d}t}$ for $n \geq 0$, and we can state $\mathbb{E}[e^{\int_0^T b_t^n \mathrm{d}N_t}] = \mathbb{E}[e^{\int_0^T (e^{b_t^n}-1)\lambda \mathrm{d}t}]$. Notice that $X_n = e^{\int_0^T b_t^n \mathrm{d}N_t} \to X = e^{\int_0^T b_t \mathrm{d}N_t}$ almost surely. Moreover,

$$|e^{\int_0^T b_t^n \mathrm{d}N_t}| \leq e^{\int_0^T |b_t^n| \mathrm{d}N_t} \leq e^{\int_0^T |b_t| \mathrm{d}N_t}.$$

  The latter quantity is integrable because $|b_t|$ is non-negative, and we have $\mathbb{E}[e^{\int_0^T |b_t| \mathrm{d}N_t}] = \mathbb{E}[e^{\int_0^T (e^{|b_t|}-1)\lambda \mathrm{d}t}]$, which is finite by assumption. It holds $Y_n = e^{\int_0^T (e^{b_t^n}-1)\lambda \mathrm{d}t} \to Y = e^{\int_0^T (e^{b_t}-1)\lambda \mathrm{d}t}$ almost surely and $|e^{\int_0^T (e^{b_t}-1)\lambda \mathrm{d}t}| \leq e^{\int_0^T (e^{|b_t|}-1)\lambda \mathrm{d}t}$, for every n, where the latter quantity is integrable by hypothesis. The lemma's statement follows by dominated convergence.

□

Remark 1.B.2. As a generalization of the previous result, if μ is a Poisson random measure on $\mathbb{R} \times \mathbb{R}_+$ with compensator $F(\mathrm{d}x, \mathrm{d}y)\mathrm{d}t$ and H is a positive $\mathbb{F}$-predictable function, then if
$H * \mu_T = \int_0^T \int_{\mathbb{R}\times\mathbb{R}_+} H(t,x,y)\mu(\mathrm{d}t, \mathrm{d}x, \mathrm{d}y)$, we have

$$\mathbb{E}[e^{H * \mu_T}] = \mathbb{E}[e^{\int_0^T \int_{\mathbb{R}\times\mathbb{R}_+} (e^{H(t,x,y)}-1)F(\mathrm{d}x,\mathrm{d}y)\mathrm{d}t}]. \tag{1.42}$$

1.C Proving Theorem 1.21

Notation. From now on, for simplicity, we will occasionally denote $g(t, u_j, 0, 0)$ by $g_t(u_j)$, for $g = \phi, \psi, \chi$, as defined in Theorem 1.8.

The proof of Theorem 1.21 is mainly based on the application of Proposition 1.19 and Corollary 1.A.1. The calculation of A, B, and C proceeds through the following steps:

- Identify a suitable factor process $\mathbf{Z}$ and determine its Laplace transform.
- Check that $\mathbf{Z}$ and the claims $\eta^0_t = \mathbb{E}[\eta^0_T|\mathcal{F}_t]$, $\eta_t(u_j) = \mathbb{E}[e^{u_j X_T}|\mathcal{F}_t]$ satisfy the assumptions of Corollary 1.A.1.
- Express η^0_t and $\eta_t(u_j)$ as functions of $\mathbf{Z}$.
- Use Corollary 1.A.1 to compute the predictable covariations between the GKW residuals of η^0 and $\eta(u_j)$.
- Compute their expectations and integrate to obtain the final result.

The choice of the factor process and of the functionals for the claims are closely linked: $\mathbf{Z}$ depends on the structure of the claims η^0 and $\eta(u_j)$. In the case of a variance swap, see equation (1.38), it is natural to choose as $\mathbf{Z}$ the model (X, V, λ) augmented with the quadratic variation of X. This preserves an affine structure and simplifies the functional representation. We highlight that further details on the expectations of the covariations will be given later in Proposition 1.D.4.

Before starting with the actual proof, let us recall some properties of Z.

Remark 1.C.1 (On the moments of Z). In the light of Proposition 1.7 and Assumptions $i)$, $ii)$ in Table 1.1, we observe that the Laplace transform of Z exists in a open neighborhood of $(0, 0, 0)$. This ensures the existence of all the moments of (X_t, V_t, λ_t), for $t \in [0, T]$. We can make a similar observation for $\int_0^t \lambda_s \mathrm{d}s$, whose Laplace transform was studied in Proposition 1.3. Moreover, referring to Drimus (2012, Proposition 2.1), we can state that the integral process $\int_0^t V_s \mathrm{d}s$ admits all moments, whenever we are working before the Heston explosion time T^*. We can conclude that the vector $(X_t, V_t, \lambda_t, \int_0^t V_s \mathrm{d}s, \int_0^t \lambda_s \mathrm{d}s)$ admits moments of all orders.

Proposition 1.C.2. *Let ν be the dual predictable projection of the jump measure μ, see equation* (1.5)*, then*

$$\int_0^t \int_{\mathbb{R}^2} x^2 \mathbb{1}_{\mathbb{R}}(y) \nu(\mathrm{d}s, \mathrm{d}x, \mathrm{d}y) = (\delta^2 + \gamma^2) \int_0^t \lambda_{s-} \mathrm{d}s.$$

As a consequence, the stochastic process $\left(\int_0^t \int_{\mathbb{R}^2} x^2 \mathbb{1}_{\mathbb{R}}(y) \nu(\mathrm{d}s, \mathrm{d}x, \mathrm{d}y)\right)_{t \in [0,T]}$ *is integrable.*

Proof. Writing explicitly ν as in equation (1.5), we have

$$\int_0^t \int_{\mathbb{R}^2} x^2 \mathbb{1}_{\mathbb{R}}(y) \nu(\mathrm{d}s, \mathrm{d}x, \mathrm{d}y) = \int_0^t \int_{\mathbb{R}^2} x^2 \mathbb{1}_{\mathbb{R}}(y) \theta^X(\mathrm{d}x) \theta^\lambda(\mathrm{d}y) \lambda_{s-} \mathrm{d}s$$

$$= \int_{\mathbb{R}} x^2 \theta^X(\mathrm{d}x) \int_0^t \lambda_{s-} \mathrm{d}s = (\delta^2 + \gamma^2) \int_0^t \lambda_{s-} \mathrm{d}s,$$

where the last quality follows from the fact that $\theta^X(\mathrm{d}x)$ is the probability density of a Normal with parameters (γ, δ^2). The stochastic process $\left(\int_0^t \int_{\mathbb{R}^2} x^2 \mathbb{1}_{\mathbb{R}}(y) \nu(\mathrm{d}s, \mathrm{d}x, \mathrm{d}y)\right)_{t \in [0,T]}$ admits expectation as $\int_0^t \lambda_s \mathrm{d}s$ does, as explained in Remark 1.C.1. □

We can now start with the proof. In Step 1 we will introduce an auxiliary semimartingale $\mathbf{Z}$ and in Step 2 we will verify that all the needed assumptions to apply Corollary 1.A.1 hold. In Step 3 we will compute some key common quantities to A, B, C and finally in Step 4 we will proceed with the computation of the predictable covariations.

Step 1) The factor process $\mathbf{Z}$ Both in Proposition 1.19 and Corollary 1.A.1, the statement depends on the choice of a factor process $\mathbf{Z}$. In our case we take $\mathbf{Z} = (X, V, \lambda, [X, X])$, i.e., the vector formed by the three model components together with the quadratic variation of X. Recalling that X is defined via equation (1.7a), its quadratic variation is given by

$$[X, X]_t = \int_0^t V_s \mathrm{d}s + [J^X, J^X]_t = \int_0^t V_s \mathrm{d}s + \sum_{s<t} (\Delta J_s^X)^2$$
$$= \int_0^t V_s \mathrm{d}s + \int_0^t \int_{\mathbb{R}^2} x^2 \mathbb{1}_{\mathbb{R}}(y) \mu(\mathrm{d}s, \mathrm{d}x, \mathrm{d}y). \tag{1.43}$$

We now verify that $\mathbf{Z} = (X, V, \lambda, [X, X])$ satisfies all the hypotheses required in Proposition 1.19 and Corollary 1.A.1.

- $\mathbf{Z} = (X, V, \lambda, [X, X])$ is a quasi-left continuous locally square-integrable semimartingale. Quasi-left continuity follows from the fact that we are working in a jump-diffusion setting. The square-integrability of (X, V, λ) has been noticed in Remark 1.C.1. Focusing on $[X, X]$, we have that $\int_0^t V_s \mathrm{d}s$ is square-integrable due to Remark 1.C.1, as L^2 is closed w.r.t. summation. It is then enough to prove that $\sum_{s<t} (\Delta J_s^X)^2$ is square-integrable. We have:

$$\mathbb{E}\left[\left(\sum_{s<t} (\Delta J_s^X)^2\right)^2\right] = \mathbb{E}\left[\left(\int_0^t \int_{\mathbb{R}^2} x^2 \mathbb{1}_{\mathbb{R}}(y) \mu(\mathrm{d}s, \mathrm{d}x, \mathrm{d}y)\right)^2\right]$$
$$\leq 2\mathbb{E}\left[\left(\int_0^t \int_{\mathbb{R}^2} x^2 \mathbb{1}_{\mathbb{R}}(y) (\mu - \nu)(\mathrm{d}s, \mathrm{d}x, \mathrm{d}y)\right)^2\right] + 2\mathbb{E}\left[\left(\int_0^t \int_{\mathbb{R}^2} x^2 \mathbb{1}_{\mathbb{R}}(y) \nu(\mathrm{d}s, \mathrm{d}x, \mathrm{d}y)\right)^2\right]$$

$$= 2\mathbb{E}\left[\int_0^t \int_{\mathbb{R}^2} x^4 \mathbb{1}_{\mathbb{R}}(y)\nu(\mathrm{d}s, \mathrm{d}x, \mathrm{d}y)\right] + 2\mathbb{E}\left[\left((\delta^2+\gamma^2)\int_0^t \lambda_s \mathrm{d}s\right)^2\right]$$

$$= 2\int_{\mathbb{R}} x^4 \theta^X(\mathrm{d}x)\mathbb{E}\left[\int_0^t \lambda_s \mathrm{d}s\right] + 2(\delta^2+\gamma^2)^2\mathbb{E}\left[\left(\int_0^t \lambda_s \mathrm{d}s\right)^2\right]$$

The first inequality follows from the standard inequality $a^2 = (a-b+b)^2 \leq 2(a-b)^2 + 2b^2$. The next-to-last passage is a consequence of Ito's isometry and of Proposition 1.C.2. The last one is just a matter of computations. We conclude recalling that $\int_0^t \lambda_s \mathrm{d}s$ admits all the moments as explained in Remark 1.C.1 and that Gaussian random variables admit fourth order moment.

- As stated in Eberlein and Kallsen (2019, Proposition 6.18), $\mathbf{Z} = (X, V, \lambda, [X, X])$ is affine (the quadratic variation of components of an affine process retains the affine structure) and has differential characteristics, for $G \in \mathcal{B}(\mathbb{R}^4)$:

$$\mathbf{b}_t = \begin{pmatrix} -\frac{1}{2}V_t - (\kappa_1 - 1)\lambda_{t-} + \int_{\mathbb{R}} h(x)\theta^X(\mathrm{d}x)\lambda_{t-} \\ \beta_v(\alpha_v - V_t) \\ \beta_\lambda(\alpha_\lambda - \lambda_{t-}) + \int_{\mathbb{R}} h(x)\theta^\lambda(\mathrm{d}x)\lambda_{t-} \\ V_t + \int_{\mathbb{R}} h(x^2)\theta^X(\mathrm{d}x)\lambda_{t-} \end{pmatrix},$$

$$\mathbf{c}_t = \begin{pmatrix} V_t & \rho\sigma_v V_t & 0 & 0 \\ \rho\sigma_v V_t & \sigma_v^2 V_t & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \tag{1.44}$$

$$\mathbf{K}_t(G) = \int_{\mathbb{R}^3} \mathbb{1}_G(x_1, x_2, x_3, x_1^2)\theta^X(\mathrm{d}x_1)\delta_0(\mathrm{d}x_2)\theta^\lambda(\mathrm{d}x_3)\lambda_{t-}. \tag{1.45}$$

Step 2) Assumptions for the GKW decomposition and consequences Referring to Section 1.5 and in particular to Section 1.5.2, we recall that computing the semi-static variance-optimal strategy means computing the GKW decomposition of the following two objects $\eta_t^0 = \mathbb{E}[\eta_T^0|\mathcal{F}_t]$ and $\eta_t(u_j) = \mathbb{E}[e^{u_j X_T}|\mathcal{F}_t]$ for $u_j \in \mathcal{S}^j$. In order for the decomposition to exist, we must verify that $\eta_t^0 \in \mathcal{H}^2$ and $\eta_t(u_j) \in \mathcal{H}^2_{\mathbb{C}}$. The two objects are martingale by definition, hence one only needs to investigate their square-integrability.

- In order to verify the square-integrability of η_t^0, it is enough to show that it can be written as an affine function of $\mathbf{Z}$; in particular, as all the components of $\mathbf{Z}$ are be square-integrable, also η_t^0 is.

Proposition 1.C.3. *$\eta_t^0 = f^0(t, \mathbf{Z})$ where $f^0(t, x_1, x_2, x_3, x_4) = \widetilde{\alpha}(t) + \widetilde{\gamma}(t) + \alpha(t)x_2 + \gamma(t)x_3 + x_4 - k$, with*

* $\widetilde{\alpha}(t) = \alpha_v \beta_v^{-1}(\beta_v(T-t) - 1 + e^{-\beta_v(T-t)})$,
* $\alpha(t) = \beta_v^{-1}(1 - e^{-\beta_v(T-t)})$,
* $\gamma(t) = -\frac{\gamma^2+\delta^2}{\beta_\lambda - \frac{1}{\zeta}}\left(e^{-\left(\beta_\lambda - \frac{1}{\zeta}\right)(T-t)} - 1\right)$,
* $\widetilde{\gamma}(t) = (\gamma^2+\delta^2)\frac{\alpha_\lambda \beta_\lambda}{\beta_\lambda - \frac{1}{\zeta}}\left((T-t) + \frac{1}{\beta_\lambda - \frac{1}{\zeta}}\left(e^{-\left(\beta_\lambda - \frac{1}{\zeta}\right)(T-t)} - 1\right)\right)$.

Moreover, note that: $\partial_{x_1} f^0 = 0, \quad \partial_{x_2} f^0 = \alpha(t), \quad \partial_{x_3} f^0 = \gamma(t), \quad \partial_{x_4} f^0 = 1.$

Proof. Recalling the quadratic variation of X in equation (1.43) and the definition of η_T^0 in equation (1.38), we have that

$$\begin{aligned} \eta_t^0 &= \mathbb{E}[[X,X]_T - k|\mathcal{F}_t] = [X,X]_t + \mathbb{E}\left[[X,X]_T - [X,X]_t|\mathcal{F}_t\right] - k \\ &= [X,X]_t + \mathbb{E}\left[\int_t^T V_s \, \mathrm{d}s \middle| \mathcal{F}_t\right] + \mathbb{E}\left[\sum_{t \le s \le T} (\Delta J_s^X)^2 \middle| \mathcal{F}_t\right] - k. \end{aligned} \tag{1.46}$$

The conditional expectation of $\int_t^T V_s \mathrm{d}s$ has already been computed in Di Tella et al. (2019, Section 5, Proposition 5.1)

$$\mathbb{E}\left[\int_t^T V_s \, \mathrm{d}s \middle| \mathcal{F}_t\right] = \widetilde{\alpha}(t) + \alpha(t)V_t,$$

where $\tilde{\alpha}(t), \alpha(t)$ are those in the statement. Focusing now on the second conditional expectation in equation (1.46), we use $\mu(\mathrm{d}s, \mathrm{d}x) = (\mu(\mathrm{d}s, \mathrm{d}x) - \nu(\mathrm{d}s, \mathrm{d}x)) + \nu(\mathrm{d}s, \mathrm{d}x)$ and proceed as in the proof of Proposition 1.C.2,

$$\begin{aligned} \mathbb{E}\left[\sum_{t \le s \le T} (\Delta J_s^X)^2 \middle| \mathcal{F}_t\right] &= \mathbb{E}\left[\int_t^T \int_{\mathbb{R}^2} x^2 \mathbb{1}_{\mathbb{R}}(y) \mu(\mathrm{d}s, \mathrm{d}x, \mathrm{d}y) \middle| \mathcal{F}_t\right] \\ &= (\gamma^2+\delta^2)\mathbb{E}\left[\int_t^T \lambda_s \, \mathrm{d}s \middle| \mathcal{F}_t\right] = (\gamma^2+\delta^2)\mathbb{E}\left[\int_t^T \lambda_s \, \mathrm{d}s \middle| \mathcal{F}_t\right] \\ &= (\gamma^2+\delta^2)\int_t^T \mathbb{E}[\lambda_s|\mathcal{F}_t] \, \mathrm{d}s. \end{aligned}$$

Since λ is the intensity of an exponential Hawkes process, it is Markov and time-homogeneous. From Dassios and Zhao (2011, Theorem 3.6, equation (3.16)), it follows that, for $s \geq t$

$$\mathbb{E}[\lambda_s|\mathcal{F}_t] = \frac{\alpha_\lambda \beta_\lambda}{\beta_\lambda - \frac{1}{\zeta}} + e^{-\left(\beta_\lambda - \frac{1}{\zeta}\right)(s-t)}\left(\lambda_t - \frac{\alpha_\lambda \beta_\lambda}{\beta_\lambda - \frac{1}{\zeta}}\right)$$

and so the integral between t and T of $\mathbb{E}[\lambda_s|\mathcal{F}_t]$ is given by

$$\begin{aligned}\int_t^T \mathbb{E}[\lambda_s|\mathcal{F}_t]\,\mathrm{d}s &= \frac{\alpha_\lambda \beta_\lambda}{\beta_\lambda - \frac{1}{\zeta}}(T-t) - \frac{1}{\beta_\lambda - \frac{1}{\zeta}}\left(\lambda_t - \frac{\alpha_\lambda \beta_\lambda}{\beta_\lambda - \frac{1}{\zeta}}\right)\left(e^{-\left(\beta_\lambda - \frac{1}{\zeta}\right)(T-t)} - 1\right) \\ &= -\frac{1}{\beta_\lambda - \frac{1}{\zeta}}\left(e^{-\left(\beta_\lambda - \frac{1}{\zeta}\right)(T-t)} - 1\right)\lambda_t + \frac{\alpha_\lambda \beta_\lambda}{\beta_\lambda - \frac{1}{\zeta}}\left((T-t) + \frac{1}{\beta_\lambda - \frac{1}{\zeta}}\left(e^{-\left(\beta_\lambda - \frac{1}{\zeta}\right)(T-t)} - 1\right)\right) \\ &= \frac{1}{\gamma^2 + \delta^2}\left(\gamma(t)\lambda_t + \widetilde{\gamma}(t)\right).\end{aligned}$$

We can conclude that $\eta_t^0 = \widetilde{\alpha}(t) + \widetilde{\gamma}(t) + \alpha(t)V_t + \gamma(t)\lambda_t + [X,X]_t - k$. $\square$

- For the moment, we assume that $\eta_t(u_j) = \mathbb{E}[e^{u_j X_T}|\mathcal{F}_t]$ is square-integrable. It will be clear in Proposition 1.D.4, why this condition is implied by the hypothesis $\mathbb{E}[e^{2R_j X_T}] < \infty$. In this section we investigate the consequences of the square-integrability condition has on the image of χ, where the latter has been introduced in Theorem 1.8.

Proposition 1.C.4. *Consider $u_j \in \mathcal{S}^j := \{u \in \mathbb{C} : \Re\mathrm{e}(u) = R_j\}$. Under the assumptions that the stochastic process $\eta_t(u_j) = \mathbb{E}[e^{u_j X_T}|\mathcal{F}_t], t \in [0,T]$, belongs to $\mathcal{H}^2_{\mathbb{C}}$, we get that $\Re\mathrm{e}(\chi_{T-t}(u_j)) < \zeta/2$. In particular, taking $u_i \in \mathcal{S}^i, u_j \in \mathcal{S}^j$, $\eta_t(u_i), \eta_t(u_j) \in \mathcal{H}^2_{\mathbb{C}}$, we have that $\Re\mathrm{e}(\chi_{T-t}(u_j) + \chi_{T-t}(u_i)) < \zeta$.*

Proof. The stochastic process $\eta(u_j)$ is by definition a $(\mathbb{P}, \mathbb{F})$-martingale. Exploiting Theorem 1.8 with $u = (u_j, 0, 0)$, we obtain that $\eta_t(u_j) = \exp(\phi_{T-t}(u_j) + u_j X_t + \psi_{T-t}(u_j)V_t + \chi_{T-t}(u_j)\lambda_t)$. The square-integrability assumption reads

$$\mathbb{E}[|\eta_t(u_j)|^2] = \mathbb{E}[e^{\Re\mathrm{e}(2\phi_{T-t}(u_j)) + \Re\mathrm{e}(2u_j)X_t + \Re\mathrm{e}(2\psi_{T-t}(u_j))V_t + \Re\mathrm{e}(2\chi_{T-t}(u_j))\lambda_t}] < \infty.$$

and since $\Re\mathrm{e}(2\phi_{T-t}(u_j))$ is deterministic, the latter inequality is equivalent to

$$\mathbb{E}[e^{\Re\mathrm{e}(2u_j)X_t + \Re\mathrm{e}(2\psi_{T-t}(u_j))V_t + \Re\mathrm{e}(2\chi_{T-t}(u_j))\lambda_t}] < \infty. \tag{1.47}$$

Equation (1.47) can be seen as $\mathbb{E}[e^{y^\top Z_t}] < \infty$ with $y = (\Re\mathrm{e}(2u_j), \Re\mathrm{e}(2\psi_{T-t}(u_j)), \Re\mathrm{e}(2\chi_{T-t}(u_j))$, $Z_t = (X_t, V_t, \lambda_t)$. From Keller-Ressel and Mayerhofer (2015, Theorem 2.14(a)), we have that if $\mathbb{E}[e^{y^\top Z_t}] < \infty$, then $y \in \mathcal{Y} = \mathbb{R} \times \mathbb{R} \times (-\infty, \zeta)$, as in Proposition 1.11. In particular $2\Re\mathfrak{e}(\chi_{T-t}(u_j)) < \zeta$. The second inequality in the statement follows. □

We now show that $\eta_t(u_j)$ can be written as a function of $\mathbf{Z}_t = (X_t, V_t, \lambda_t, [X,X]_t)$. This result will be key in Step 4.

Proposition 1.C.5. *The random variable $\eta_t(u_j) = \mathbb{E}[e^{u_j X_T}|\mathcal{F}_t]$ can be written as $\eta_t(u_j) = f(t, \mathbf{Z}_t, u_j)$ where $f(t, x_1, x_2, x_3, x_4, u_j) = \exp(\phi_{T-t}(u_j) + u_j x_1 + \psi_{T-t}(u_j)x_2 + \chi_{T-t}(u_j)x_3)$, ϕ, ψ, χ are the solutions of the ODEs system (1.17). Note that*

$$\frac{\partial f_t(u_j)}{\partial x_1} = u_j f_t(u_j), \quad \frac{\partial f_t(u_j)}{\partial x_2} = \psi_{T-t}(u_j) f_t(u_j), \quad \frac{\partial f_t(u_j)}{\partial x_3} = \chi_{T-t}(u_j) f_t(u_j), \quad \frac{\partial f_t(u_j)}{\partial x_4} = 0,$$

where for simplicity $f_t(u_j) = f(t, X_t, V_t, \lambda_t, [X,X]_t, u_j)$.

Proof. Notice that $[X,X]$ has no role in the dynamics of X, given by equation (1.7a). Thus, one can compute the conditional Laplace transform of X_T without considering the impact of its quadratic variation $[X,X]$: that corresponds to applying Theorem 1.8. □

Step 3) Computation of common quantities In Corollary 1.A.1 there appear some quantities which only depend on the factor process $\mathbf{Z} = (X, V, \lambda, [X,X])$. In particular, we refer to the differential characteristics $\mathbf{c}, \mathbf{K}$ and to $\left(\mathbf{c}_t^{jk} - \frac{\mathbf{c}_t^{1j}}{\xi_t}\mathbf{c}_t^{1k}\right)_{j,k=1,\dots,4}$, that we recall in Table 1.4. Another key quantity is ξ, see equation (1.40), that we compute below, recalling $\mathbf{c}$ and $\mathbf{K}$ in equation (1.44), (1.45):

$$\xi_t := V_t + \int_{\mathbb{R}^4} (e^{x_1} - 1)^2 \mathbf{K}_t(\mathrm{d}x) = V_t + \lambda_t(e^{2(\gamma+\delta^2)} - 2e^{\gamma+\delta^2/2} + 1) = V_t + \lambda_{t-}\bar{\kappa}. \tag{1.48}$$

where (1.48) follows from the properties of the Gaussian distribution and we define $\bar{\kappa} := e^{2(\gamma+\delta^2)} - 2e^{\gamma+\delta^2/2} + 1$.

Step 4) Predictable covariations We need to specify all the needed mathematical objects to apply Corollary 1.A.1. For the affine model, we refer to $\mathbf{Z} = (X, V, \lambda, [X,X])$, whose main characteristics are summarized in Table 1.4. We report all the key quantities in Tables 1.5 and 1.6.

Matrix $\mathbf{c}$	$\begin{pmatrix} V_t & \rho\sigma_v V_t & 0 & 0 \\ \rho\sigma_v V_t & \sigma_v^2 V_t & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$
Matrix $\left(\mathbf{c}_t^{jk} - \frac{\mathbf{c}_t^{1j}}{\xi_t}\mathbf{c}_t^{1k}\right)_{j,k=1,\dots,4}$	$\begin{pmatrix} \frac{\lambda_{t-}V_t\bar{\kappa}}{V_t+\lambda_{t-}\bar{\kappa}} & \frac{\rho\sigma_v\lambda_{t-}V_t\bar{\kappa}}{V_t+\lambda_{t-}\bar{\kappa}} & 0 & 0 \\ \frac{\rho\sigma_v\lambda_{t-}V_t\bar{\kappa}}{V_t+\lambda_{t-}\bar{\kappa}} & \frac{(1-\rho^2)\sigma_v^2V_t^2+\sigma_v^2\lambda_{t-}V_t\bar{\kappa}}{V_t+\lambda_{t-}\bar{\kappa}} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$
Compensator $\mathbf{K}_t$	$\mathbf{K}_t(G) = \int_{\mathbb{R}^3} \mathbb{1}_G(x_1, x_2, x_3, x_1^2)\theta^X(\mathrm{d}x_1)\delta_0(\mathrm{d}x_2)\theta^\lambda(\mathrm{d}x_3)\lambda_{t-}$, $G \in \mathcal{B}(\mathbb{R}^4)$
ξ_t	$V_t + \lambda_{t-}\bar{\kappa}$

Table 1.4: Key quantities for the application of Corollary 1.A.1, associated to the process $\mathbf{Z}$.

	$\langle L^0, L^0\rangle_t$	$\langle L^0, L(u_j)\rangle_t$	$\langle L(u_i), L(u_j)\rangle_t$
Y_t^1	η_t^0	η_t^0	$\eta_t(u_i)$
Y_t^2	η_t^0	$\eta_t(u_j)$	$\eta_t(u_j)$
f^1	$f^0(t, \mathbf{Z}_t)$	$f^0(t, \mathbf{Z}_t)$	$f(t, \mathbf{Z}_t, u_i)$
f^2	$f^0(t, \mathbf{Z}_t)$	$f(t, \mathbf{Z}_t, u_j)$	$f(t, \mathbf{Z}_t, u_j)$
W^1	$f^0(t, x + \mathbf{Z}_{t-}) - f^0(t, \mathbf{Z}_{t-})$	$f^0(t, x + \mathbf{Z}_{t-}) - f^0(t, \mathbf{Z}_{t-})$	$f(t, x + \mathbf{Z}_{t-}, u_i) - f(t, \mathbf{Z}_{t-}, u_i)$
W^2	$f^0(t, x + \mathbf{Z}_{t-}) - f^0(t, \mathbf{Z}_{t-})$	$f(t, x + \mathbf{Z}_{t-}, u_j) - f(t, \mathbf{Z}_{t-}, u_j)$	$f(t, x + \mathbf{Z}_{t-}, u_j) - f(t, \mathbf{Z}_{t-}, u_j)$

where:
- $f^0(t, x) = \widetilde{\alpha}(t) + \widetilde{\gamma}(t) + \alpha(t)x_2 + \gamma(t)x_3 + x_4 - k$, as determined in Proposition 1.C.3,
- $f(t, x, u_j) = \exp(\phi_{T-t}(u_j) + u_j x_1 + \psi_{T-t}(u_j)x_2 + \chi_{T-t}(u_j)x_3)$, as determined in Proposition 1.C.5,
- $f^0(t, x + \mathbf{Z}_{t-}) - f^0(t, \mathbf{Z}_{t-}) = \alpha(t)x_2 + \gamma(t)x_3 + x_4$,
- $f(t, x + \mathbf{Z}_{t-}, u_j) - f(t, \mathbf{Z}_{t-}, u_j) = f_{t-}(u_j)[\exp(x_1 u_j + x_2\psi_{T-t}(u_j) + x_3\chi_{T-t}(u_j)) - 1]$,

with $f_{t-}(u_j) = \exp(\phi_{T-t}(u_j) + u_j X_{t-} + \psi_{T-t}(u_j)V_{t-} + \chi_{T-t}(u_j)\lambda_{t-})$.

Table 1.5: Claims Y_t^1, Y_t^2, and corresponding functions.

	∂_{x_1}	∂_{x_2}	∂_{x_3}	∂_{x_4}
$f^0(t, \mathbf{Z}_t)$	0	$\alpha(t)$	$\gamma(t)$	1
$f(t, \mathbf{Z}_t, u_j)$	$u_j f_t(u_j)$	$\psi_{T-t}(u_j)f_t(u_j)$	$\chi_{T-t}(u_j)f_t(u_j)$	0

Table 1.6: Key derivatives of f^0, f as defined in Propositions 1.C.3 and 1.C.5, respectively.

I) **Compute** $\langle L^0, L^0\rangle_t$. We refer to Corollary 1.A.1, to the first column of Table 1.5 and on the first row of Table 1.6. We observe that $W^1 = W^2$, thus:

$$\begin{aligned}\int_{\mathbb{R}^4}(e^{x_1}-1)W^1(t,x)\mathbf{K}_t(\mathrm{d}x) &= \int_{\mathbb{R}^4}(e^{x_1}-1)(\alpha(t)x_2+\gamma(t)x_3+x_4)\mathbf{K}_t(\mathrm{d}x)\\ &= \lambda_{t-}\left(\gamma(t)\int_{\mathbb{R}}(e^{x_1}-1)\theta^X(\mathrm{d}x_1)\cdot\int_{\mathbb{R}}x_3\theta^\lambda(\mathrm{d}x_3)+\int_{\mathbb{R}}(e^{x_1}-1)x_1^2\theta^X(\mathrm{d}x_1)\right)\\ &= \lambda_{t-}\left(\frac{\gamma(t)}{\zeta}(e^{\gamma+\delta^2/2}-1)+(\delta^2+(\gamma+\delta^2)^2)e^{\gamma+\delta^2/2}-\gamma^2-\delta^2\right). \qquad (1.49)\end{aligned}$$

The, we compute:

$$\begin{aligned}\int_{\mathbb{R}^4}(W^1(t,x))^2\mathbf{K}_t(\mathrm{d}x) &= \int_{\mathbb{R}^4}(\alpha(t)x_2+\gamma(t)x_3+x_4)^2\mathbf{K}_t(\mathrm{d}x)\\ &= \lambda_{t-}\left(\gamma(t)^2\int_{\mathbb{R}}x_3^2\theta^\lambda(\mathrm{d}x_3)+2\gamma(t)\int_{\mathbb{R}}x_3\theta^\lambda(\mathrm{d}x_3)\cdot\int_{\mathbb{R}}x_1^2\theta^X(\mathrm{d}x_1)+\int_{\mathbb{R}}x_1^4\theta^X(\mathrm{d}x_1)\right)\\ &= \lambda_{t-}\left(\gamma(t)^2\frac{2}{\zeta^2}+2\gamma(t)\frac{\gamma^2+\delta^2}{\zeta}+\gamma^4+6\gamma^2\delta^2+3\delta^4\right). \qquad (1.50)\end{aligned}$$

The identities above mainly rely on the properties of Gaussian and exponential random variables. Finally, we can write the predictable covariation as:

$$\begin{aligned}\frac{d\langle L^0,L^0\rangle_t}{\mathrm{d}t} &= \alpha(t)^2\frac{(1-\rho^2)\sigma_v^2V_t^2+\sigma_v^2\lambda_{t-}V_t\bar{\kappa}}{V_t+\lambda_{t-}\bar{\kappa}}\\ &- \alpha(t)\frac{2\rho\sigma_v\lambda_{t-}V_t}{V_t+\lambda_{t-}\bar{\kappa}}\left(\frac{\gamma(t)}{\zeta}(e^{\gamma+\delta^2/2}-1)+(\delta^2+(\gamma+\delta^2)^2)e^{\gamma+\delta^2/2}-\gamma^2-\delta^2\right)\\ &+ \lambda_{t-}\left(\gamma(t)^2\frac{2}{\zeta^2}+2\gamma(t)\frac{\gamma^2+\delta^2}{\zeta}+\gamma^4+6\gamma^2\delta^2+3\delta^4\right)\\ &- \frac{\lambda_{t-}^2}{V_t+\lambda_{t-}\bar{\kappa}}\left(\frac{\gamma(t)}{\zeta}(e^{\gamma+\delta^2/2}-1)+(\delta^2+(\gamma+\delta^2)^2)e^{\gamma+\delta^2/2}-\gamma^2-\delta^2\right)^2.\\ &= A_1(t)\frac{V_t^2}{V_t+\lambda_{t-}\bar{\kappa}}+A_2(t)\frac{\lambda_{t-}V_t}{V_t+\lambda_{t-}\bar{\kappa}}+A_3(t)\frac{\lambda_{t-}^2}{V_t+\lambda_{t-}\bar{\kappa}}\end{aligned}$$

where

$$\begin{aligned}A_1(t) &= \alpha(t)^2(1-\rho^2)\sigma_v^2,\\ A_2(t) &= \alpha(t)^2\sigma_v^2\bar{\kappa}-2\alpha(t)\rho\sigma_v\left(\frac{\gamma(t)}{\zeta}(e^{\gamma+\delta^2/2}-1)+(\delta^2+(\gamma+\delta^2)^2)e^{\gamma+\delta^2/2}-\gamma^2-\delta^2\right)\\ &+\left(\gamma(t)^2\frac{2}{\zeta^2}+2\gamma(t)\frac{\gamma^2+\delta^2}{\zeta}+\gamma^4+6\gamma^2\delta^2+3\delta^4\right),\end{aligned}$$

$$A_3(t) = \bar{\kappa}\left(\gamma(t)^2\frac{2}{\zeta^2} + 2\gamma(t)\frac{\gamma^2+\delta^2}{\zeta} + \gamma^4 + 6\gamma^2\delta^2 + 3\delta^4\right) - \left(\frac{\gamma(t)}{\zeta}(e^{\gamma+\delta^2/2}-1) + (\delta^2+(\gamma+\delta^2)^2)e^{\gamma+\delta^2/2} - \gamma^2 - \delta^2\right)^2$$

and

$$\vartheta_t^0 = \frac{1}{S_{t-}(V_t + \bar{\kappa}\lambda_{t-})}\left(\Theta_1^0(t)V_t + \Theta_2^0(t)\lambda_{t-}\right),$$
$$\Theta_1^0(t) = \alpha(t)\rho\sigma_v,$$
$$\Theta_2^0(t) = \frac{\gamma(t)}{\zeta}(e^{\gamma+\delta^2/2}-1) + (\delta^2+(\gamma+\delta^2)^2)e^{\gamma+\delta^2/2} - \gamma^2 - \delta^2.$$

II) **Compute** $\langle L^0, L(u_j)\rangle_t$. We refer to Corollary 1.A.1, to the second column of Table 1.5 and both the rows of Table 1.6. We observe that the main quantities related to W^1 has already been computed on equations (1.49), (1.50). We mainly focus on the term related to W^2 and to the mixed terms, thus:

$$\int_{\mathbb{R}^4}(e^{x_1}-1)W^2(t,x)\mathbf{K}_t(\mathrm{d}x) = \int_{\mathbb{R}^4}(e^{x_1}-1)f_{t-}(u_j)[e^{x_1u_j+x_2\psi_{T-t}(u_j)+x_3\chi_{T-t}(u_j)}-1]\mathbf{K}_t(\mathrm{d}x)$$
$$= f_{t-}(u_j)\lambda_{t-}\Big(\int_{\mathbb{R}} e^{u_jx_1}(e^{x_1}-1)\theta^X(\mathrm{d}x_1)\int_{\mathbb{R}} e^{x_3\chi_{T-t}(u_j)}\theta^\lambda(\mathrm{d}x_3) - \int_{\mathbb{R}}(e^{x_1}-1)\theta^X(\mathrm{d}x_1)\Big)$$
$$= f_{t-}(u_j)\lambda_{t-}\Big(\Big(e^{\gamma(u_j+1)+\delta^2(u_j+1)^2/2} - e^{\gamma u_j+\delta^2u_j^2/2}\Big)\frac{\zeta}{\zeta-\chi_{T-t}(u_j)} - e^{\gamma+\delta^2/2}+1\Big). \tag{1.51}$$

The equation above is obtained exploiting the properties of Exponential and Gaussian random variables. With similar techniques we compute:

$$\int_{\mathbb{R}^4} W^1(t,x)W^2(t,x)\mathbf{K}_t(\mathrm{d}x)$$
$$= \int_{\mathbb{R}^4}(\alpha(t)x_2+\gamma(t)x_3+x_4)\cdot(f_{t-}(u_j)[e^{(x_1u_j+x_2\psi_{T-t}(u_j)+x_3\chi_{T-t}(u_j)}-1])\mathbf{K}_t(\mathrm{d}x)$$
$$= \int_{\mathbb{R}}\int_{\mathbb{R}}\gamma(t)x_3f_{t-}(u_j)[e^{x_1u_j+x_3\chi_{T-t}(u_j)}-1]\theta^X(\mathrm{d}x_1)\theta^\lambda(\mathrm{d}x_3)\lambda_{t-}$$
$$+ \int_{\mathbb{R}}\int_{\mathbb{R}}x_1^2f_{t-}(u_j)[e^{x_1u_j+x_3\chi_{T-t}(u_j)}-1]\theta^X(\mathrm{d}x_1)\theta^\lambda(\mathrm{d}x_3)\lambda_{t-}$$
$$= f_{t-}(u_j)\lambda_{t-}\gamma(t)\Big(\int_{\mathbb{R}} e^{x_1u_j}\theta^X(\mathrm{d}x_1)\int_{\mathbb{R}} x_3e^{x_3\chi_{T-t}(u_j)}\theta^\lambda(\mathrm{d}x_3) - \int_{\mathbb{R}} x_3\theta^\lambda(\mathrm{d}x_3)\Big)$$

$$
\begin{aligned}
&+ f_{t-}(u_j)\lambda_{t-}\Big(\int_{\mathbb{R}} x_1^2 e^{x_1 u_j}\theta^X(\mathrm{d}x_1)\int_{\mathbb{R}} e^{x_3\chi_{T-t}(u_j)}\theta^\lambda(\mathrm{d}x_3) - \int_{\mathbb{R}} x_1^2\theta^X(\mathrm{d}x_1)\Big)\\
&= f_{t-}(u_j)\lambda_{t-}\Big(\gamma(t)\Big(e^{\gamma u_j+\delta^2u_j^2/2}\frac{\zeta}{(\zeta-\chi_{T-t}(u_j))^2}-\frac{1}{\zeta}\Big)\\
&+ (\delta^2+(\gamma+\delta^2u_j)^2)e^{\gamma u_j+\delta^2u_j^2/2}\frac{\zeta}{\zeta-\chi_{T-t}(u_j)}-(\gamma^2+\delta^2)\Big).
\end{aligned}
$$

The predictable covariation is given by:

$$
\begin{aligned}
\frac{d\langle L^0, L(u_j)\rangle}{\mathrm{d}t} &= \alpha(t)u_j f_{t-}(u_j)\frac{\rho\sigma_v\lambda_{t-}V_t\bar{\kappa}}{V_t+\lambda_{t-}\bar{\kappa}} + \alpha(t)\psi_{T-t}(u_j)f_{t-}(u_j)\frac{(1-\rho^2)\sigma_v^2V_t^2+\sigma_v^2\lambda_{t-}V_t\bar{\kappa}}{V_t+\lambda_{t-}\bar{\kappa}}\\
&- \frac{1}{V_t+\lambda_{t-}\bar{\kappa}}\alpha(t)\rho\sigma_v V_t f_{t-}(u_j)\lambda_{t-}\\
&\cdot\Big(\Big(e^{\gamma(u_j+1)+\delta^2(u_j+1)^2/2}-e^{\gamma u_j+\delta^2u_j^2/2}\Big)\frac{\zeta}{\zeta-\chi_{T-t}(u_j)}-e^{\gamma+\delta^2/2}+1\Big)\\
&- \frac{1}{V_t+\lambda_{t-}\bar{\kappa}}\Big(u_j f_{t-}(u_j)V_t+\psi_{T-t}(u_j)f_{t-}(u_j)\rho\sigma_v V_t\Big)\lambda_{t-}\cdot\\
&\Big(\frac{\gamma(t)}{\zeta}(e^{\gamma+\delta^2/2}-1)+(\delta^2+(\gamma+\delta^2)^2)e^{\gamma+\delta^2/2}-\gamma^2-\delta^2\Big)\\
&+ f_{t-}(u_j)\lambda_{t-}\Big(\gamma(t)\Big(e^{\gamma u_j+\delta^2u_j^2/2}\frac{\zeta}{(\zeta-\chi_{T-t}(u_j))^2}-\frac{1}{\zeta}\Big)\\
&+ (\delta^2+(\gamma+\delta^2u_j)^2)e^{\gamma u_j+\delta^2u_j^2/2}\frac{\zeta}{\zeta-\chi_{T-t}(u_j)}-(\gamma^2+\delta^2)\Big)\\
&- \frac{1}{V_t+\lambda_{t-}\bar{\kappa}}\lambda_{t-}\Big(\frac{\gamma(t)}{\zeta}(e^{\gamma+\delta^2/2}-1)+(\delta^2+(\gamma+\delta^2)^2)e^{\gamma+\delta^2/2}-\gamma^2-\delta^2\Big)\\
&\cdot f_{t-}(u_j)\lambda_{t-}\Big(\Big(e^{\gamma(u_j+1)+\delta^2(u_j+1)^2/2}-e^{\gamma u_j+\delta^2u_j^2/2}\Big)\frac{\zeta}{\zeta-\chi_{T-t}(u_j)}-e^{\gamma+\delta^2/2}+1\Big)\\
&= B_1(t,u_j)\frac{f_{t-}(u_j)V_t^2}{V_t+\lambda_{t-}\bar{\kappa}} + B_2(t,u_j)\frac{f_{t-}(u_j)\lambda_{t-}V_t}{V_t+\lambda_{t-}t\bar{\kappa}} + B_3(t,u_j)\frac{f_{t-}(u_j)\lambda_{t-}^2}{V_t+\lambda_{t-}\bar{\kappa}}
\end{aligned}
$$

where

$$
\begin{aligned}
B_1(t,u_j) &= \alpha(t)\psi_{T-t}(u_j)(1-\rho^2)\sigma_v^2\\
B_2(t,u_j) &= \alpha(t)u_j\rho\sigma_v\bar{\kappa}+\alpha(t)\psi_{T-t}(u_j)\sigma_v^2\bar{\kappa}\\
&- \alpha(t)\rho\sigma_v\Big(\Big(e^{\gamma(u_j+1)+\delta^2(u_j+1)^2/2}-e^{\gamma u_j+\delta^2u_j^2/2}\Big)\frac{\zeta}{\zeta-\chi_{T-t}(u_j)}-e^{\gamma+\delta^2/2}+1\Big)\\
&- (u_j+\psi_{T-t}(u_j)\rho\sigma_v)\Big(\frac{\gamma(t)}{\zeta}(e^{\gamma+\delta^2/2}-1)+(\delta^2+(\gamma+\delta^2)^2)e^{\gamma+\delta^2/2}-\gamma^2-\delta^2\Big)
\end{aligned}
$$

$$
\begin{aligned}
&+ \Big(\gamma(t)\Big(e^{\gamma u_j+\delta^2 u_j^2/2}\frac{\zeta}{(\zeta-\chi_{T-t}(u_j))^2}-\frac{1}{\zeta}\Big)\\
&+(\delta^2+(\gamma+\delta^2 u_j)^2)e^{\gamma u_j+\delta^2 u_j^2/2}\frac{\zeta}{\zeta-\chi_{T-t}(u_j)}-(\gamma^2+\delta^2)\Big)\\
B_3(t,u_j) = \bar{\kappa}&\Big(\gamma(t)\Big(e^{\gamma u_j+\delta^2 u_j^2/2}\frac{\zeta}{(\zeta-\chi_{T-t}(u_j))^2}-\frac{1}{\zeta}\Big)\\
&+(\delta^2+(\gamma+\delta^2 u_j)^2)e^{\gamma u_j+\delta^2 u_j^2/2}\frac{\zeta}{\zeta-\chi_{T-t}(u_j)}-(\gamma^2+\delta^2)\Big)\\
&-\Big(\frac{\gamma(t)}{\zeta}(e^{\gamma+\delta^2/2}-1)+(\delta^2+(\gamma+\delta^2)^2)e^{\gamma+\delta^2/2}-\gamma^2-\delta^2\Big)\\
&\cdot\Big(\Big(e^{\gamma(u_j+1)+\delta^2(u_j+1)^2/2}-e^{\gamma u_j+\delta^2 u_j^2/2}\Big)\frac{\zeta}{\zeta-\chi_{T-t}(u_j)}-e^{\gamma+\delta^2/2}+1\Big).
\end{aligned}
$$

and

$$
\begin{aligned}
\vartheta_t^j &= \frac{1}{S_{t-}(V_t+\bar{\kappa}\lambda_{t-})}\int_{\mathcal{S}^j}(\Theta_1^j(t,u_j)f_{t-}(u_j)V_t+\Theta_2^j(t,u_j)f_{t-}(u_j)\lambda_t)\zeta^j(\mathrm{d}u_j),\\
\Theta_1^j(t,u_j) &= u_j+\psi_{T-t}(u_j)\rho\sigma_v,\\
\Theta_2^j(t,u_j) &= \Big(e^{\gamma(u_j+1)+\delta^2(u_j+1)^2/2}-e^{\gamma u_j+\delta^2 u_j^2/2}\Big)\frac{\zeta}{\zeta-\chi_{T-t}(u_j)}-e^{\gamma+\delta^2/2}+1.
\end{aligned}
$$

III) **Compute** $\langle L(u_i), L(u_j)\rangle_t$. We refer to Corollary 1.A.1, to the third column of Table 1.5 and the second row of Table 1.6. We observe that one of the main quantities has already been computed in equation (1.51), thus we only focus on the remaining one:

$$
\begin{aligned}
&\int_{\mathbb{R}^4} W^1(t,x)W^2(t,x)\mathbf{K}_t(\mathrm{d}x)\\
&= \int_{\mathbb{R}^4}(f_{t-}(u_i)(e^{x_1u_i+x_2\psi_{T-t}(u_i)+x_3\chi_{T-t}(u_i))}-1)\cdot(f_{t-}(u_j)(e^{x_1u_j+x_2\psi_{T-t}(u_j)+x_3\chi_{T-t}(u_j))}-1)\mathbf{K}_t(\mathrm{d}x)\\
&= f_{t-}(u_i)f_{t-}(u_j)\lambda_{t-}\int_{\mathbb{R}}\int_{\mathbb{R}}(e^{x_1u_i+x_3\chi_{T-t}(u_i)}-1)(e^{x_1u_j+x_3\chi_{T-t}(u_j)}-1)\theta^X(\mathrm{d}x_1)\theta^\lambda(\mathrm{d}x_3)\\
&= f_{t-}(u_i)f_{t-}(u_j)\lambda_{t-}\Big(\int_{\mathbb{R}}\int_{\mathbb{R}}e^{x_1(u_i+u_j)+x_3(\chi_{T-t}(u_i)+\chi_{T-t}(u_j))}-e^{x_1u_i+x_3\chi_{T-t}(u_i)}\\
&-e^{x_1u_j+x_3\chi_{T-t}(u_j)}+1\Big)\theta^X(\mathrm{d}x_1)\theta^\lambda(\mathrm{d}x_3)\\
&= f_{t-}(u_i)f_{t-}(u_j)\lambda_{t-}\Big(\int_{\mathbb{R}}e^{x_1(u_i+u_j)}\theta^X(\mathrm{d}x_1)\int_{\mathbb{R}}e^{x_3(\chi_{T-t}(u_i)+\chi_{T-t}(u_j))}\theta^\lambda(\mathrm{d}x_3)\\
&-\int_{\mathbb{R}}e^{x_1u_i}\theta^X(\mathrm{d}x_1)\int_{\mathbb{R}}e^{x_3\chi_{T-t}(u_i)}\theta^\lambda(\mathrm{d}x_3)-\int_{\mathbb{R}}e^{x_1u_j}\theta^X(\mathrm{d}x_1)\int_{\mathbb{R}}e^{x_3\chi_{T-t}(u_j)}\theta^\lambda(\mathrm{d}x_3)+1\Big)
\end{aligned}
$$

$$= f_{t-}(u_i)f_{t-}(u_j)\lambda_{t-}\Big(e^{\gamma(u_i+u_j)+\delta^2(u_i+u_j)^2/2}\frac{\zeta}{\zeta-\chi_{T-t}(u_i)-\chi_{T-t}(u_j)} - e^{\gamma u_i+\delta^2 u_i^2/2}\frac{\zeta}{\zeta-\chi_{T-t}(u_i)}$$
$$- e^{\gamma u_j+\delta^2 u_j^2/2}\frac{\zeta}{\zeta-\chi_{T-t}(u_j)} + 1\Big).$$

The predictable covariation is given by:

$$\frac{d\langle L(u_i), L(u_j)\rangle}{\mathrm{d}t} = u_i u_j f_{t-}(u_i) f_{t-}(u_j)\frac{\lambda_{t-}V_t\bar{\kappa}}{V_t+\lambda_{t-}\bar{\kappa}}$$
$$+ \frac{\rho\sigma_v\lambda_{t-}V_t\bar{\kappa}}{V_t+\lambda_{t-}\bar{\kappa}} f_{t-}(u_i)f_{t-}(u_j)(u_i\psi_{T-t}(u_j)+u_j\psi_{T-t}(u_i))+$$
$$+ \psi_{T-t}(u_i)\psi_{T-t}(u_j)f_{t-}(u_i)f_{t-}(u_j)\frac{(1-\rho^2)\sigma_v^2 V_t^2+\sigma_v^2\lambda_{t-}V_t\bar{\kappa}}{V_t+\lambda_{t-}\bar{\kappa}}$$
$$- \frac{1}{V_t+\lambda_{t-}\bar{\kappa}} f_{t-}(u_i)\Big(u_j V_t + \psi_{T-t}(u_j)\rho\sigma_v V_t\Big) f_{t-}(u_j)$$
$$\cdot \lambda_{t-}\Big(\Big(e^{\gamma(u_i+1)+\delta^2(u_i+1)^2/2} - e^{\gamma u_i+\delta^2 u_i^2/2}\Big)\frac{\zeta}{\zeta-\chi_{T-t}(u_i)} - e^{\gamma+\delta^2/2}+1\Big)$$
$$- \frac{1}{V_t+\lambda_{t-}\bar{\kappa}} f_{t-}(u_j)\Big(u_i V_t + \psi_{T-t}(u_i)\rho\sigma_v V_t\Big) f_{t-}(u_i)$$
$$\cdot \lambda_{t-}\Big(\Big(e^{\gamma(u_j+1)+\delta^2(u_j+1)^2/2} - e^{\gamma u_j+\delta^2 u_j^2/2}\Big)\frac{\zeta}{\zeta-\chi_{T-t}(u_j)} - e^{\gamma+\delta^2/2}+1\Big)$$
$$+ f_{t-}(u_i)f_{t-}(u_j)\lambda_{t-}\Big(e^{\gamma(u_i+u_j)+\delta^2(u_i+u_j)^2/2}\frac{\zeta}{\zeta-\chi_{T-t}(u_i)-\chi_{T-t}(u_j)} - e^{\gamma u_i+\delta^2 u_i^2/2}\frac{\zeta}{\zeta-\chi_{T-t}(u_i)}$$
$$- e^{\gamma u_j+\delta^2 u_j^2/2}\frac{\zeta}{\zeta-\chi_{T-t}(u_j)} + 1\Big)$$
$$- \frac{1}{V_t+\lambda_{t-}\bar{\kappa}} f_{t-}(u_i)f_{t-}(u_j)\lambda_t^2\Big(\Big(e^{\gamma(u_i+1)+\delta^2(u_i+1)^2/2} - e^{\gamma u_i+\delta^2 u_i^2/2}\Big)\frac{\zeta}{\zeta-\chi_{T-t}(u_i)} - e^{\gamma+\delta^2/2}+1\Big)$$
$$\cdot \Big(\Big(e^{\gamma(u_j+1)+\delta^2(u_j+1)^2/2} - e^{\gamma u_j+\delta^2 u_j^2/2}\Big)\frac{\zeta}{\zeta-\chi_{T-t}(u_j)} - e^{\gamma+\delta^2/2}+1\Big)$$
$$= C_1(t,u_i,u_j)\frac{f_{t-}(u_i)f_{t-}(u_j)V_t^2}{V_t+\lambda_{t-}\bar{\kappa}} + C_2(t,u_i,u_j)\frac{f_{t-}(u_i)f_{t-}(u_j)\lambda_{t-}V_t}{V_t+\lambda_{t-}\bar{\kappa}}$$
$$+ C_3(t,u_i,u_j)\frac{f_{t-}(u_i)f_{t-}(u_j)\lambda_{t-}^2}{V_t+\lambda_{t-}\bar{\kappa}}$$

where

$$C_1(t,u_i,u_j) = \psi_{T-t}(u_i)\psi_{T-t}(u_j)(1-\rho^2)\sigma_v^2$$
$$C_2(t,u_i,u_j) = u_i u_j\bar{\kappa} + \rho\sigma_v\bar{\kappa}(u_i\psi_{T-t}(u_j)+u_j\psi_{T-t}(u_i)) + \psi_{T-t}(u_i)\psi_{T-t}(u_j)\sigma_v^2\bar{\kappa}$$

$$
\begin{aligned}
&- (u_i + \psi_{T-t}(u_i)\rho\sigma_v)\Big(\Big(e^{\gamma(u_j+1)+\delta^2(u_j+1)^2/2} - e^{\gamma u_j+\delta^2 u_j^2/2}\Big)\frac{\zeta}{\zeta - \chi_{T-t}(u_j)} - e^{\gamma+\delta^2/2} + 1\Big) \\
&- (u_j + \psi_{T-t}(u_j)\rho\sigma_v)\Big(\Big(e^{\gamma(u_i+1)+\delta^2(u_i+1)^2/2} - e^{\gamma u_i+\delta^2 u_i^2/2}\Big)\frac{\zeta}{\zeta - \chi_{T-t}(u_i)} - e^{\gamma+\delta^2/2} + 1\Big) \\
&+ \Big(e^{\gamma(u_i+u_j)+\delta^2(u_i+u_j)^2/2}\frac{\zeta}{\zeta - \chi_{T-t}(u_i) - \chi_{T-t}(u_j)} \\
&- e^{\gamma u_i+\delta^2 u_i^2/2}\frac{\zeta}{\zeta - \chi_{T-t}(u_i)} - e^{\gamma u_j+\delta^2 u_j^2/2}\frac{\zeta}{\zeta - \chi_{T-t}(u_j)} + 1\Big) \\
C_3(t, u_i, u_j) = \bar{\kappa}\Big(&e^{\gamma(u_i+u_j)+\delta^2(u_i+u_j)^2/2}\frac{\zeta}{\zeta - \chi_{T-t}(u_i) - \chi_{T-t}(u_j)} \\
&- e^{\gamma u_i+\delta^2 u_i^2/2}\frac{\zeta}{\zeta - \chi_{T-t}(u_i)} - e^{\gamma u_j+\delta^2 u_j^2/2}\frac{\zeta}{\zeta - \chi_{T-t}(u_j)} + 1\Big) \\
&- \Big(\Big(e^{\gamma(u_i+1)+\delta^2(u_i+1)^2/2} - e^{\gamma u_i+\delta^2 u_i^2/2}\Big)\frac{\zeta}{\zeta - \chi_{T-t}(u_i)} - e^{\gamma+\delta^2/2} + 1\Big) \\
&\cdot \Big(\Big(e^{\gamma(u_j+1)+\delta^2(u_j+1)^2/2} - e^{\gamma u_j+\delta^2 u_j^2/2}\Big)\frac{\zeta}{\zeta - \chi_{T-t}(u_j)} - e^{\gamma+\delta^2/2} + 1\Big).
\end{aligned}
$$

1.D Existence and computations of moments

Lemma 1.D.1. *Let* $\mathcal{D}_{\mathcal{L}(Z_T)} = \{u \in \mathbb{R}^3 : \mathbb{E}[\exp(u^\top Z_T)] < \infty\}$ *and* $\mathcal{S}(\mathcal{D}_{\mathcal{L}(Z_T)}) := \{u \in \mathbb{R}^3 : \mathfrak{Re}(u) \in \mathcal{D}_{\mathcal{L}(Z_T)}\}$*, where* $Z = (X, V, \lambda)$*. Let the functions* χ, ψ, χ *be defined as in Theorem 1.8. Then the following properties hold:*

a) The set $\mathcal{D}_{\mathcal{L}(Z_T)}$ *is open.*

b) The set $\mathcal{D}_{\mathcal{L}(Z_T)}$ *is convex.*

c) If $(u_1, u_2, u_3) \in \mathcal{S}(\mathcal{D}_{\mathcal{L}(Z_T)})$*, then* $(u_1, \psi(T-t, u_1, u_2, u_3), \chi(T-t, u_1, u_2, u_3)) \in \mathcal{S}(\mathcal{D}_{\mathcal{L}(Z_t)})$*.*

d) The functions $\phi(t, u_1, u_2, u_3), \psi(t, u_1, u_2, u_3), \chi(t, u_1, u_2, u_3)$ *are analytic on* $\mathcal{S}(\mathcal{D}_{\mathcal{L}(Z_t)})$*.*

e) If $(a, b, c) \in \mathcal{S}(\mathcal{D}_{\mathcal{L}(Z_T)})$ *, then* $(a, b', c') \in \mathcal{S}(\mathcal{D}_{\mathcal{L}(Z_T)})$ *for all* $b' \le b, c' \le c$*.*

f) $\mathfrak{Re}\psi(t, u_1, u_2, u_3) \le \psi(t, \mathfrak{Re}u_1, \mathfrak{Re}u_2, \mathfrak{Re}u_3), \mathfrak{Re}\chi(t, u_1, u_2, u_3) \le \chi(t, \mathfrak{Re}u_1, \mathfrak{Re}u_2, \mathfrak{Re}u_3)$ *for all* $(u_1, u_2, u_3) \in \mathcal{S}(\mathcal{D}_{\mathcal{L}(Z_t)})$*.*

Proof. a) In the spirit of Filipovic and Mayerhofer (2009, Lemma 2.3), we can prove the first point referring to Amann (2011, Theorem 7.6, Theorem 8.3).

b) Let $(u_1, u_2, u_3), (w_1, w_2, w_3) \in \mathcal{D}_{\mathcal{L}(Z_T)}$, and $h \in (0, 1)$. Then by Hölder inequality

$$\mathbb{E}[\exp(h(u_1X_T + u_2V_T + u_3\lambda_T) + (1-h)(w_1X_T + w_2V_T + w_3\lambda_T))]$$

$$
\begin{aligned}
&= \mathbb{E}[\exp(h(u_1X_T+u_2V_T+u_3\lambda_T))\cdot\exp((1-h)(w_1X_T+w_2V_T+w_3\lambda_T))] \\
&\le \mathbb{E}[\exp(u_1X_T+u_2V_T+u_3\lambda_T)]^h\cdot\mathbb{E}[\exp(w_1X_T+w_2V_T+w_3\lambda_T)]^{1-h}<\infty.
\end{aligned}
$$

It follows that $h(u_1,u_2,u_3)+(1-h)(w_1,w_2,w_3)\in\mathcal{D}_{\mathcal{L}(Z_T)}$.

c) Let $(u_1,u_2,u_3)\in\mathcal{D}_{\mathcal{L}(Z_T)}$. Then, exploiting Theorem 1.8, we can state that all the following quantities are finite

$$
\begin{aligned}
\mathbb{E}[e^{u_1X_T+u_2V_T+u_3\lambda_T}] &= \mathbb{E}[\mathbb{E}[e^{u_1X_T+u_2V_T+u_3\lambda_T}|\mathcal{F}_t]] \\
&= \mathbb{E}[e^{\phi(T-t,u_1,u_2,u_3)+u_1X_t+\psi(T-t,u_1,u_2,u_3)V_t+\chi(T-t,u_1,u_2,u_3)\lambda_t}].
\end{aligned}
$$

d) In the spirit of Filipovic and Mayerhofer (2009, Lemma 2.3), the regularity of ϕ,ψ,χ follows by Dieudonné (1960, Theorem 10.8.2).

e), f) The last two points can be proved analogously to Di Tella et al. (2020, Lemma A.1).

□

Lemma 1.D.2. *Let $(u_1,u_2,u_3)\in\mathcal{S}(\mathcal{D}_{\mathcal{L}(Z_t)})$ and denote*

$$
h(t,u_1,u_2,u_3,V_0,\lambda_0):=\phi(t,u_1,u_2,u_3)+\psi(t,u_1,u_2,u_3)V_0+\chi(t,u_1,u_2,u_3)\lambda_0.
$$

The following identities hold

$$
\begin{aligned}
\mathbb{E}[e^{u_1X_t+u_2V_t+u_3\lambda_t}V_t] &= e^{u_1X_0}e^{h(t,u_1,u_2,u_3,V_0,\lambda_0)}\partial_{u_2}h(t,u_1,u_2,u_3,V_0,\lambda_0),\\
\mathbb{E}[e^{u_1X_t+u_2V_t+u_3\lambda_t}\lambda_t] &= e^{u_1X_0}e^{h(t,u_1,u_2,u_3,V_0,\lambda_0)}\partial_{u_3}h(t,u_1,u_2,u_3,V_0,\lambda_0),\\
\mathbb{E}[e^{u_1X_t+u_2V_t+u_3\lambda_t}V_t^2] &= e^{u_1X_0}e^{h(t,u_1,u_2,u_3,V_0,\lambda_0)}((\partial_{u_2}h(t,u_1,u_2,u_3,V_0,\lambda_0))^2+\partial^2_{u_2}h(t,u_1,u_2,u_3,V_0,\lambda_0)),\\
\mathbb{E}[e^{u_1X_t+u_2V_t+u_3\lambda_t}\lambda_t^2] &= e^{u_1X_0}e^{h(t,u_1,u_2,u_3,V_0,\lambda_0)}((\partial_{u_3}h(t,u_1,u_2,u_3,V_0,\lambda_0))^2+\partial^2_{u_3}h(t,u_1,u_2,u_3,V_0,\lambda_0)),\\
\mathbb{E}[e^{u_1X_t+u_2V_t+u_3\lambda_t}\lambda_tV_t] &= e^{u_1X_0}e^{h(t,u_1,u_2,u_3,V_0,\lambda_0)}(\partial_{u_2}h(t,u_1,u_2,u_3,V_0,\lambda_0)\cdot\partial_{u_3}h(t,u_1,u_2,u_3,V_0,\lambda_0)\\
&\quad+\partial_{u_2}\partial_{u_3}h(t,u_1,u_2,u_3,V_0,\lambda_0)).
\end{aligned}
$$

In particular, all the expectations above are finite.

Proof. The proof is a modification of the one in Di Tella et al. (2020, Lemma A.2). Fix $(x_1,x_2,x_3)\in\mathcal{D}_{\mathcal{L}(Z_t)}$ and consider $(u_1,u_2,u_3)\in\mathcal{S}(\mathcal{D}_{\mathcal{L}(Z_t)})$ of the form $u_j=x_j+iy_j$, for $j=1,2,3$. By assumption $K=\mathbb{E}[e^{x_1X_t+x_2V_t+x_3\lambda_t}]$ exists and belongs to $(0,+\infty)$. Define a probability measure $\mathbb{M}$ on $(\Omega,\mathcal{F}_t)$ by $\frac{\mathbb{M}}{\mathbb{P}}\big|_{\mathcal{F}_t}=\exp(x_1X_t+x_2V_t+x_3\lambda_t)/K$, i.e., by exponential tilting of $\mathbb{P}$. The characteristic function of (X_t,V_t,λ_t) under $\mathbb{M}$ is given by

$$
K\mathbb{E}^{\mathbb{M}}[e^{iy_1X_t+iy_2V_t+iy_3\lambda_t}]=\mathbb{E}[e^{u_1X_t+u_2V_t+u_3\lambda_t}]
$$

$$= \exp\left(\phi(t,u_1,u_2,u_3) + u_1 X_0 + \psi(t,u_1,u_2,u_3)V_0 + \chi(t,u_1,u_2,u_3)\lambda_0\right).$$

Due to the analyticity properties of $\phi(t,u_1,u_2,u_3)$, $\psi(t,u_1,u_2,u_3)$ and $\chi(t,u_1,u_2,u_3)$ proved in 1.D.1, c), all partial derivatives of the left hand side with respect to (y_1,y_2,y_3) exist. Applying some standard results on differentiability of characteristic functions (see e.g., Lukacs (1970, Section 2.3)) we obtain

$$\begin{aligned} K\mathbb{E}^{\mathbb{M}}[e^{iy_1X_t+iy_2V_t+iy_3\lambda_t}V_t] &= -i\frac{\mathrm{d}}{\mathrm{d}x_2}K\mathbb{E}^{\mathbb{M}}[e^{iy_1X_t+iy_2V_t+iy_3\lambda_t}] \\ &= \frac{\mathrm{d}}{\mathrm{d}u_2}\exp\left(\phi(t,u_1,u_2,u_3) + u_1X_0 + \psi(t,u_1,u_2,u_3)V_0 + \chi(t,u_1,u_2,u_3)\lambda_0\right). \end{aligned}$$

Transforming the left hand side back to measure $\mathbb{P}$ yields the desired result. The other formulas are obtained analogously. □

Proposition 1.D.3. *Let $\phi|_{\mathbb{R}_+\times\mathbb{R}^3}, \psi|_{\mathbb{R}_+\times\mathbb{R}^3}, \chi|_{\mathbb{R}_+\times\mathbb{R}^3}$ be the restriction of the functions ϕ,ψ,χ to the real domain. Then $\phi|_{\mathbb{R}_+\times\mathbb{R}^3}, \psi|_{\mathbb{R}_+\times\mathbb{R}^3}, \chi|_{\mathbb{R}_+\times\mathbb{R}^3} : (t,x_1,x_2,x_3) \mapsto \mathbb{R}$ are convex functions in x.*

Proof. The convexity property follows from a modification of Sarychev (1996, Theorem 2). □

Proposition 1.D.4. *Under the assumptions of Theorem 1.21, in particular $\mathbb{E}[e^{2R_jX_T}] < \infty$, all the expected values appearing in Theorem are finite and can be computed explicitly. Let*

$$h(t,u_1,u_2,u_3,V_0,\lambda_0) := \phi(t,u_1,u_2,u_3) + \psi(t,u_1,u_2,u_3)V_0 + \chi(t,u_1,u_2,u_3)\lambda_0,$$

$$f_{t-}(u_j) = \exp(\phi_{T-t}(u_j) + u_jX_{t-} + \psi_{T-t}(u_j)V_{t-} + \chi_{T-t}(u_j)\lambda_{t-}),$$

with $g(t,u_1,0,0) = g_t(u_1)$, for $g = \phi,\psi,\chi$, we have

I)
$$\begin{aligned} \mathbb{E}\left[\frac{V_t^2}{V_t+\lambda_t\bar{\kappa}}\right] &= \int_{-\infty}^0 e^{h(t,0,s,s\bar{\kappa},V_0,\lambda_0)}((\partial_{u_2}h(t,0,s,s\bar{\kappa},V_0,\lambda_0))^2 + \partial_{u_2}^2 h(t,0,s,s\bar{\kappa},V_0,\lambda_0))\mathrm{d}s, \\ \mathbb{E}\left[\frac{\lambda_t^2}{V_t+\lambda_t\bar{\kappa}}\right] &= \int_{-\infty}^0 e^{h(t,0,s,s\bar{\kappa},V_0,\lambda_0)}((\partial_{u_3}h(t,0,s,s\bar{\kappa},V_0,\lambda_0))^2 + \partial_{u_3}^2 h(t,0,s,s\bar{\kappa},V_0,\lambda_0))\mathrm{d}s, \\ \mathbb{E}\left[\frac{\lambda_tV_t}{V_t+\lambda_t\bar{\kappa}}\right] &= \int_{-\infty}^0 e^{h(t,0,s,s\bar{\kappa},V_0,\lambda_0)}(\partial_{u_2}h(t,0,s,s\bar{\kappa},V_0,\lambda_0)\partial_{u_3}h(t,0,s,s\bar{\kappa},V_0,\lambda_0) \\ &\quad + \partial_{u_2}\partial_{u_3}h(t,0,s,s\bar{\kappa},V_0,\lambda_0))\mathrm{d}s, \end{aligned}$$

II) Moreover, for $\bar{u}^j_{s,t} = (u_j, \psi_{T-t}(u_j)+s, \chi_{T-t}(u_j)+s\bar{\kappa})$, $u_j \in \mathcal{S}^j = \{z \in \mathbb{C} : \Re\mathrm{e}(z) = R_j\}$

$$\mathbb{E}\left[\frac{f_{t-}(u_j)V_t^2}{V_t+\lambda_t\bar{\kappa}}\right] = \int_{-\infty}^0 e^{u_jX_0}e^{h(\bar{u}^j_{s,t},V_0,\lambda_0)}((\partial_{u_2}h(\bar{u}^j_{s,t},V_0,\lambda_0))^2 + \partial_{u_2}^2 h(\bar{u}^j_{s,t},V_0,\lambda_0))\mathrm{d}s,$$

$$\mathbb{E}\left[\frac{f_{t-}(u_j)\lambda_t^2}{V_t+\lambda_t\bar{\kappa}}\right] = \int_{-\infty}^{0} e^{u_j X_0} e^{h(\bar{u}^j_{s,t},V_0,\lambda_0)}((\partial_{u_3}h(\bar{u}^j_{s,t},V_0,\lambda_0))^2+\partial^2_{u_3}h(\bar{u}^j_{s,t},V_0,\lambda_0))\mathrm{d}s,$$

$$\mathbb{E}\left[\frac{f_{t-}(u_j)\lambda_t V_t}{V_t+\lambda_t\bar{\kappa}}\right] = \int_{-\infty}^{0} e^{u_j X_0} e^{h(\bar{u}^j_{s,t},V_0,\lambda_0)}(\partial_{u_2}h(\bar{u}^j_{s,t},V_0,\lambda_0)\partial_{u_3}h(\bar{u}^j_{s,t},V_0,\lambda_0)+ \partial_{u_2}\partial_{u_3}h(\bar{u}^j_{s,t},V_0,\lambda_0))\mathrm{d}s.$$

III) For $\bar{u}^{i,j}_{s,t}=(u_i+u_j,\psi_{T-t}(u_i)+\psi_{T-t}(u_j)+s,\chi_{T-t}(u_i)+\chi_{T-t}(u_j)+s\bar{\kappa})$, $u^k\in\mathcal{S}^k=\{z\in\mathbb{C}:\mathfrak{Re}(z)=R_k\}$, $k=i,j$

$$\mathbb{E}\left[\frac{f_{t-}(u_i)f_{t-}(u_j)V_t^2}{V_t+\lambda_t\bar{\kappa}}\right] = \int_{-\infty}^{0} e^{(u_i+u_j)X_0} e^{h(\bar{u}^{i,j}_{s,t},V_0,\lambda_0)}((\partial_{u_2}h(\bar{u}^{i,j}_{s,t},V_0,\lambda_0))^2+\partial^2_{u_2}h(\bar{u}^{i,j}_{s,t},V_0,\lambda_0))\mathrm{d}s,$$

$$\mathbb{E}\left[\frac{f_{t-}(u_i)f_{t-}(u_j)\lambda_t^2}{V_t+\lambda_t\bar{\kappa}}\right] = \int_{-\infty}^{0} e^{(u_i+u_j)X_0} e^{h(\bar{u}^{i,j}_{s,t},V_0,\lambda_0)}((\partial_{u_3}h(\bar{u}^{i,j}_{s,t},V_0,\lambda_0))^2+\partial^2_{u_3}h(\bar{u}^{i,j}_{s,t},V_0,\lambda_0))\mathrm{d}s,$$

$$\mathbb{E}\left[\frac{f_{t-}(u_i)f_{t-}(u_j)\lambda_t V_t}{V_t+\lambda_t\bar{\kappa}}\right] = \int_{-\infty}^{0} e^{(u_i+u_j)X_0} e^{h(\bar{u}^{i,j}_{s,t},V_0,\lambda_0)}(\partial_{u_2}h(\bar{u}^{i,j}_{s,t},V_0,\lambda_0)\partial_{u_3}h(\bar{u}^{i,j}_{s,t},V_0,\lambda_0) + \partial_{u_2}\partial_{u_3}h(\bar{u}^{i,j}_{s,t},V_0,\lambda_0))\mathrm{d}s.$$

Furthermore, we can show that if $\mathbb{E}[e^{2R_jX_T}]<\infty$, then $\eta_t(u_j)=\mathbb{E}[e^{u_jX_T}|\mathcal{F}_t]$ is square-integrable, for $u_j\in\mathcal{S}^j$.

Proof. First of all, we prove that the absolute values of the random variables above admit finite expectations. Then, the explicit computation of the expectations follows by the application of Fubini's theorem.

I) The random variables satisfy the following inequalities:

$$\frac{V_t^2}{V_t+\lambda_t\bar{\kappa}} = V_t\cdot\frac{V_t}{V_t+\lambda_t\bar{\kappa}} < V_t,\quad \frac{\lambda_t V_t}{V_t+\lambda_t\bar{\kappa}} < \lambda_t,\quad \frac{\lambda_t^2}{V_t+\lambda_t\bar{\kappa}} = \frac{\lambda_t}{\bar{\kappa}}\frac{\bar{\kappa}\lambda_t}{V_t+\lambda_t\bar{\kappa}} < \frac{\lambda_t}{\bar{\kappa}},$$

being $V_t,\lambda_t,\bar{\kappa}$ all positive quantities. Thus all the above quantities are integrable since all the moments of V_t and λ_t exist, as explained in Remark 1.C.1.

II) Since we have

$$\frac{V_t}{V_t+\lambda_t\bar{\kappa}}<1,\quad \frac{\lambda_t}{V_t+\lambda_t\bar{\kappa}}<\frac{1}{\bar{\kappa}} \tag{1.52}$$

to verify that all the random variables in group II) admit finite expectation, it is enough to verify that $|f_{t-}(u_j)V_t|, |f_{t-}(u_j)\lambda_t|$ do. It means we need to prove that

$$\mathbb{E}[e^{\Re e\phi_{T-t}(u_j)+\Re e u_j X_{t-}+\Re e\psi_{T-t}(u_j)V_{t-}+\Re e\chi_{T-t}(u_j)\lambda_{t-}}V_t] < \infty \tag{1.53}$$

$$\mathbb{E}[e^{\Re e\phi_{T-t}(u_j)+\Re e u_j X_{t-}+\Re e\psi_{T-t}(u_j)V_{t-}+\Re e\chi_{T-t}(u_j)\lambda_{t-}}\lambda_t] < \infty. \tag{1.54}$$

By the assumption $\mathbb{E}[e^{2R_jX_T}] < \infty$ and $u_j \in \mathcal{S}^j$, we have that $\Re e u_j = R_j$ and $(R_j, 0, 0) \in \mathcal{D}_{\mathcal{L}(Z_T)}$ by Jensen's inequality. Thus, by Lemma 1.D.1 c), $(R_j, \psi_{T-t}(R_j), \chi_{T-t}(R_j)) \in \mathcal{D}_{\mathcal{L}(Z_t)}$. Combining Lemma 1.D.1 e, f), we get that

$$(\Re e u_j, \Re e\psi_{T-t}(u_j), \Re e\chi_{T-t}(u_j)) \in \mathcal{D}_{\mathcal{L}(Z_t)}.$$

Applying Lemma 1.D.2, we prove the finiteness of the expected value in (1.53), (1.54).

III) Using the same estimates as in (1.52), we claim that the random variables in group III) admit expectation if the following processes do

$$|f_{t-}(u_i)f_{t-}(u_j)V_t|, \quad |f_{t-}(u_i)f_{t-}(u_j)\lambda_t|$$

i.e.,

$$\mathbb{E}[e^{\Re e\phi_{T-t}(u_i)+\Re e\phi_{T-t}(u_j)+(\Re e u_i+\Re e u_j)X_{t-}+(\Re e\psi_{T-t}(u_i)+\Re e\psi_{T-t}(u_j))V_{t-}+(\Re e\chi_{T-t}(u_i)+\Re e\chi_{T-t}(u_j))\lambda_{t-}}V_t] < \infty,$$

$$\mathbb{E}[e^{\Re e\phi_{T-t}(u_i)+\Re e\phi_{T-t}(u_j)+(\Re e u_i+\Re e u_j)X_{t-}+(\Re e\psi_{T-t}(u_i)+\Re e\psi_{T-t}(u_j))V_{t-}+(\Re e\chi_{T-t}(u_i)+\Re e\chi_{T-t}(u_j))\lambda_{t-}}\lambda_t] < \infty.$$

By assumptions $\mathbb{E}[e^{2R_kX_T}] < \infty$, $u_k \in \mathcal{S}^k$, we have that $\Re e(u_k) = R_k$ and $(2R_k, 0, 0) \in \mathcal{D}_{\mathcal{L}(Z_T)}$, for $k = i, j$. Applying the same techniques as in point II), one can prove that $(2R_k, \psi_{T-t}(2R_k), \chi_{T-t}(2R_k)) \in \mathcal{D}_{\mathcal{L}(Z_t)}$. By convexity of the set $\mathcal{D}_{\mathcal{L}(Z_t)}$, see Lemma 1.D.1, we have that

$$(R_i + R_j, \frac{1}{2}(\psi_{T-t}(2R_i) + \psi_{T-t}(2R_j)), \frac{1}{2}(\chi_{T-t}(2R_i) + \chi_{T-t}(2R_j))) \in \mathcal{D}_{\mathcal{L}(Z_t)}.$$

Moreover, by convexity of ψ and χ, see Proposition 1.D.3, we have that

$$(R_i + R_j, \psi_{T-t}(R_i) + \psi_{T-t}(R_j), \chi_{T-t}(R_i) + \chi_{T-t}(R_j)) \in \mathcal{D}_{\mathcal{L}(Z_t)}.$$

Combining Lemma 1.D.1 e, f), we get that $(\Re e u_i + \Re e u_j, \Re e\psi_{T-t}(u_i) + \Re e\psi_{T-t}(u_j), \Re e\chi_{T-t}(u_i) + \Re e\chi_{T-t}(u_j)) \in \mathcal{D}_{\mathcal{L}(Z_t)}$. Applying Lemma 1.D.2, we prove the finiteness of the required expected values.

To compute explicitly the expectations, one should exploit the following integral representation, as in Kallsen and Pauwels (2010, Lemma 5.5). Since for Lemma 1.D.1 e), $(0, s, s\bar{\kappa}) \in \mathcal{S}(\mathcal{D}_{\mathcal{L}(Z_t)})$ when $s < 0$, we can exploit Lemma 1.D.2 and write

$$\begin{aligned}\mathbb{E}\left[\frac{V_t^2}{V_t + \lambda_t\bar{\kappa}}\right] &= \int_{-\infty}^{0} \mathbb{E}[e^{s(V_t+\bar{\kappa}\lambda_t)}V_t^2]\mathrm{d}s \\ &= \int_{-\infty}^{0} e^{h(t,0,s,s\bar{\kappa},V_0,\lambda_0)}((\partial_{u_2}h(t,0,s,s\bar{\kappa},V_0,\lambda_0))^2 + \partial^2_{u_2}h(t,0,s,s\bar{\kappa},V_0,\lambda_0))\mathrm{d}s.\end{aligned}$$

Analogous computations hold for the other processes appearing in Proposition 1.D.4.

We conclude the statement noticing that by repeating the same machinery in point III) for $R_j = R_i$, we can conclude that

$$(2\Re\mathfrak{e}u_j, 2\Re\mathfrak{e}\psi_{T-t}(u_j), 2\Re\mathfrak{e}\chi_{T-t}(u_j)) \in \mathcal{D}_{\mathcal{L}(Z_t)},$$

meaning that

$$\mathbb{E}[|\eta_t(u_j)|^2] = \mathbb{E}[e^{2\Re\mathfrak{e}\phi_{T-t}(u_j)+2\Re\mathfrak{e}u_jX_t+2\Re\mathfrak{e}\psi_{T-t}(u_j)V_t+2\Re\mathfrak{e}\chi_{T-t}(u_j)\lambda_t}] < \infty.$$

□

1.E Convexity properties

This appendix provides a rigorous proof of Proposition 1.D.3. We begin by stating and proving the one-dimensional version of Sarychev (1996, Proposition 3.1). In particular, we emphasize that the one-dimensional case holds under weaker assumptions than the multi-dimensional setting. This is because, as noted in Cuchiero (2011, Remark 2.3.6), every function in a one-dimensional vector space is automatically quasi-monotone. This observation explains why, in Proposition 1.E.1, we do not impose quasi-monotonicity on the function g.

Proposition 1.E.1. *Let a function $g(t,x) : D \subseteq \mathbb{R}^2 \to \mathbb{R}$ be continuous in an open domain $G \subset \mathbb{R}_+ \times \mathbb{R}$ together with its first partial derivatives $\frac{\partial g}{\partial x}$ for any fixed $t \in [0,T]$. For two C^1-smooth functions $y(t), z(t)$ defined on $[0,T]$ assume that $(t, y(t)), (t, z(t)) \in G$ for $t \in [0,T]$. If $y(t), z(t)$ satisfy*

$$\begin{aligned} &y'(t) = g(t, y(t)), \quad y(0) = \xi_0 = z(0), \\ &z'(t) \geq g(t, z(t)), \end{aligned}$$

then $y(t) \leq z(t)$ on $[0,T]$.

Proof. The proof is analogous to Sarychev (1996, Proposition 3.1). Consider for $\varepsilon > 0$, $y_\varepsilon'(t) = g(t, y_\varepsilon(t)) - \varepsilon, y_\varepsilon(0) = \xi_0 - \varepsilon$. It is enough to prove $y_\varepsilon(t) < z(t)$ and then make $\varepsilon \to 0_+$. We assume that the inequality fails for a certain t_0 leading to $y_\varepsilon(t) < z(t)$ for $t < t_0$ and $y_\varepsilon(t_0) = z(t_0)$. Then, we derive

$$\lim_{t \to t_0 -} \frac{(z(t) - y_\varepsilon(t)) - (z(t_0) - y_\varepsilon(t_0))}{t - t_0} \leq 0,$$

i.e. $z'(t_0) \leq y_\varepsilon'(t_0)$. It follows that

$$y_\varepsilon'(t_0) \geq z'(t_0) \geq g(t_0, z(t_0)) = g(t_0, y_\varepsilon(t_0)) > g(t_0, y_\varepsilon(t_0)) - \varepsilon = y_\varepsilon'(t_0).$$

This is a contradiction. □

We also rewrite a modification for the one dimensional parameter case of Sarychev (1996, Theorem 2). The main difference between Theorem 1.E.2 and the one in Sarychev (1996, Theorem 2) is that they consider f to be continuous for all x, while we only ask continuity and differentiability in a subdomain. In Keller-Ressel, Mayerhofer and Smirnov (2010), the authors address the more general case where f is regular only on a subdomain; however, they do not explicitly consider the parameter case directly.

Theorem 1.E.2. *Consider the following parametric ODE*

$$x'(t) = f(t, x, \mu), \quad x(0) = \xi,$$

and we denote its solution at time t by $x_t(\xi, \mu)$. We assume that:

1. *$U = \Omega \times \Lambda \subset \mathbb{R} \times \mathbb{R}$ is an open domain and $f(t, x, \mu)\colon [0, T] \times U \to \mathbb{R}$.*
2. *$f(t, x, \mu)$ is continuous in x, μ in U together with its partial derivatives in x and μ.*
3. *$f(t, x, \mu)$ is convex in (x, μ).*
4. *Let be $D^\mu \subset U$ a convex domain for which, for every $(\xi, \mu) \in D^\mu$, the solution of the ODE with initial data ξ and parameter μ exists on $[0, T]$ and $x_t(\xi, \mu) \in \Omega$ for every $t \in [0, T]$.*

Then, $x_T(\xi, \mu)$ is convex in $(\xi, \mu) \in D^\mu$.

Proof. The proof is a modification for the one dimensional parameter case of Sarychev (1996, Theorem 2). Consider $(\hat{\xi}, \hat{\mu}), (\hat{\xi} + \delta\xi, \hat{\mu} + \delta\mu) \in D^\mu$ and define

$$\Delta x(t; \delta\xi, \delta\mu) = x(t; \hat{\xi} + \delta\xi, \hat{\mu} + \delta\mu) - x(t; \hat{\xi}, \hat{\mu}),$$

where we denote $\hat{x}(t) := x(t;\hat{\xi},\hat{\mu}), \hat{x}_\delta(t) := x(t;\hat{\xi}+\delta\xi,\hat{\mu}+\delta\mu), \Delta x(t;\delta\xi,\delta\mu) := \Delta x(t)$. We notice that Δx satisfies

$$\Delta x'(t) = f(t,\hat{x}(t)+\Delta x,\hat{\mu}+\delta\mu) - f(t,\hat{x}(t),\hat{\mu}), \quad \Delta x(0) = \delta\xi.$$

By convexity and regularity of the function f,

$$\Delta x'(t) \geq f_x(t,\hat{x}(t),\hat{\mu})\Delta x + f_\mu(t,\hat{x}(t),\hat{\mu})\delta\mu.$$

We consider the following ODE

$$\delta x'(t) = f_x(t,\hat{x}(t),\hat{\mu})\delta x + f_\mu(t,\hat{x}(t),\hat{\mu})\delta\mu, \quad \delta x(0) = \delta\xi.$$

Notice that δx is homogeneous in the initial value. Due to Proposition 1.E.1, we get that $\Delta x_T \geq \delta x_T$, thus

$$x_T(\hat{\xi}+\delta\xi,\hat{\mu}+\delta\mu) \geq x_T(\hat{\xi},\hat{\mu}) + \delta x_T(\delta\xi,\delta\mu).$$

In particular, take $\alpha \in (0,1), (\xi,\mu), (\eta,\nu) \in D^\mu$ and $\xi^\alpha = \alpha\xi + (1-\alpha)\eta, \mu^\alpha = \alpha\mu + (1-\alpha)\nu$.

- $x_T(\xi,\mu) \geq x_T(\xi^\alpha,\mu^\alpha) + (1-\alpha)\delta x_T(\xi-\nu,\mu-\nu)$;
- $x_T(\eta,\nu) \geq x_T(\xi^\alpha,\mu^\alpha) - \alpha\delta x_T(\xi-\nu,\mu-\nu)$.

Multiplying the first for α, and the latter for $1-\alpha$, we prove the convexity of x_T. □

Finally, we employ the previous propositions to prove rigorously Proposition 1.D.3.

Proof. From now on, whenever we write ϕ,ψ,χ, we refer to their restriction to the real domain. We observe that if we prove that ψ and χ are convex in x, then also ϕ is. Indeed by the Riccati system in Proposition 4.2 we observe that

$$\phi(t,x_1,x_2,x_3) = \int_0^t \alpha_v\beta_v\psi(s,x_1,x_2,x_3) + \alpha_\lambda\beta_\lambda\chi(s,x_1,x_2,x_3)\mathrm{d}s,$$

for $\alpha_v,\beta_v,\alpha_\lambda,\beta_\lambda > 0$.

The convexity of ψ follows by Theorem 1.E.2. In particular, we recall that by the system in Proposition 4.2 that ψ solves

$$\frac{\mathrm{d}\psi(t,x_1)}{\mathrm{d}t} = f_\psi(\psi,x_1), \quad \psi(0,x_1) = x_2,$$

where

$$f_\psi(y,x_1) = -\frac{1}{2}x_1 + \frac{1}{2}x_1^2 - \beta_v y + \rho\sigma_v x_1 y + \frac{1}{2}\sigma_v^2 y^2.$$

We have to interpret x_1 as a parameter. $f_\psi(y,x_1)$ is continuous and differentiable for every $(y,x_1) \in \mathbb{R}^2$. Moreover, the Hessian of f_ψ is given by

$$\begin{pmatrix} \sigma_v^2 & \rho\sigma_v \\ \rho\sigma_v & 1 \end{pmatrix},$$

thus f_ψ is convex in (y,x_1) since $\rho \in [-1,1]$. The convexity of $\psi(t,x_1,x_2,x_3)$ follows by applying Theorem 1.E.2.

As for χ, we observe that it solves

$$\frac{\mathrm{d}\chi(t,x_1)}{\mathrm{d}t} = f_\chi(\chi,x_1), \quad \chi(0,x_1) = x_3,$$

where

$$f_\chi(y,x_1) = -\beta_\lambda y - (\kappa_1 - 1)x_1 + e^{\gamma x_1 + \delta^2 x_1^2/2}\frac{\zeta}{\zeta - y} - 1.$$

Referring to the notations of Theorem 1.E.2, let $U = \Omega \times \Lambda$, where $\Omega = (-\infty, \zeta)$, $\Lambda = \mathbb{R}$. We observe that f_χ is continuous and differentiable in U. The Hessian f_χ is given by

$$\begin{pmatrix} e^{\gamma x_1 + \delta^2 x_1^2/2} 2\frac{\zeta}{(\zeta-y)^3} & e^{\gamma x_1 + \delta^2 x_1^2/2}(\gamma + \delta^2 x_1)\frac{\zeta}{(\zeta-y)^2} \\ e^{\gamma x_1 + \delta^2 x_1^2/2}(\gamma + \delta^2 x_1)\frac{\zeta}{(\zeta-y)^2} & e^{\gamma x_1 + \delta^2 x_1^2/2}((\gamma + \delta^2 x_1)^2 + \delta^2)\frac{\zeta}{(\zeta-y)} \end{pmatrix}.$$

The elements in the diagonal are positive and the determinant is given by

$$e^{2\gamma x_1 + \delta^2 x_1^2}\frac{\zeta^2}{(\zeta - y)^4}((\gamma + \delta^2 x_1)^2 + 2\delta^2) > 0,$$

thus the Hessian is positive semi-definite and the function f_χ is convex in U. D^{x_1} (as D^μ in Theorem 1.E.2) can be obtained by Proposition 4.1 and it is given by (x_3, x_1) which satisfy

$$x_3 < \zeta - T\mathbb{E}(e^{x_1|\eta_1^X|}) \tag{1.55}$$

$$x_1(\kappa_1 - 1) + 1 > 0. \tag{1.56}$$

The domain D^{x_1} is convex since first we choose an half line for x_1 using equation (1.56), and then x_3 is the under-graph of a concave function in x_1, see equation (1.55) Moreover, by the definition of D^{x_1} (see Proposition 4.1), we have that for every $(x_3,x_1) \in D^{x_1}$, $\mathbb{E}[e^{x_1 X_T + x_3 \lambda_T}] < \infty$. It follows by Keller-Ressel and Mayerhofer (2015, Theorem 2.14, a)), $\chi(t,x_1,0,x_3) < \zeta$ for every $t \in [0,T]$, thus $\chi_t(x_1,x_3) \in \Omega$ for every t. The convexity of χ follows by applying Theorem 1.E.2. □

CHAPTER 2

A stochastic Gordon-Loeb model for optimal cybersecurity investment under clustered attacks

This is a joint work with Prof. Giorgia Callegaro, Prof. Claudio Fontana and Prof. Caroline Hillairet. The corresponding paper was submitted in May 2025 and is available on arxiv.

We develop a continuous-time stochastic model for optimal cybersecurity investment under the threat of cyberattacks. The arrival of attacks is modeled using a Hawkes process, capturing the empirically relevant feature of clustering in cyberattacks. Extending the Gordon-Loeb model, each attack may result in a breach, with breach probability depending on the system's vulnerability. We aim at determining the optimal cybersecurity investment to reduce vulnerability. The problem is cast as a two-dimensional Markovian stochastic optimal control problem and solved using dynamic programming methods. Numerical results illustrate how accounting for attack clustering leads to more responsive and effective investment policies, offering significant improvements over static and Poisson-based benchmark strategies. Our findings underscore the value of incorporating realistic threat dynamics into cybersecurity risk management.

2.1 Introduction

Cyber-risk is nowadays widely acknowledged as one of the major sources of operational risk for organizations worldwide. The 2024 ENISA Threat Landscape Report (see ENISA (2024)) documents "a notable escalation in cybersecurity attacks, setting new benchmarks in both the variety and number of incidents, as well as their consequences". According to the AON 9th Global Risk Management Survey[1], cyberattacks and data breaches represent the foremost source of global risk faced by organizations, with the second biggest risk being business interruption, which is itself often a consequence of cyber-incidents. A recent poll on Risk.net confirms information security and IT disruption as the top two sources of operational risk for 2025.[2] According to IBM, the global average cost of a data breach has reached nearly 5M USD in 2024, an increase of more than 10% over the previous year.[3]

The rapid and widespread emergence of cyber-risk as a key source of operational risk has led to a significant increase in cybersecurity spending. In the 2025 ICS/OT cybersecurity budget survey of the SANS Institute (see SANS Institute (2025)) 55% of the respondents reported a substantial rise in cybersecurity budgets over the previous two years. This trend underscores the importance of adopting effective cybersecurity investment policies that balance risk mitigation with cost efficiency.

The problem of optimal cybersecurity investment has been first addressed in the seminal work of Gordon and Loeb (2002). In their model, reviewed in Section 2.2.1 below, the decision maker can reduce the vulnerability to cyberattacks by investing in cybersecurity. The optimal expenditure in cybersecurity is determined by maximizing the expected net benefit of reducing the breach probability. The Gordon-Loeb model laid the foundations for a rigorous quantitative analysis of cybersecurity investments and has been the subject of numerous extensions and generalizations: we mention here only some studies that are closely related to our context, referring to Fedele and Roner (2022) for a comprehensive overview. The key ingredient of the Gordon-Loeb model is represented by the security breach probability function (see Section 2.2.1), which has been further analyzed in Huang and Behara (2013) and Mazzoccoli and Naldi (2022). The risk-neutral assumption of Gordon and Loeb (2002) has been relaxed to accommodate risk-averse preferences in Miaoui and Boudriga (2019).

The original Gordon-Loeb model is a static model and, therefore, does not allow to

[1]Source: https://www.aon.com/en/insights/reports/global-risk-management-survey.

[2]Source: https://www.risk.net/risk-management/7961268/top-10-operational-risks-for-2025.

[3]Source: https://www.ibm.com/reports/data-breach.

address the crucial issue of the optimal timing of investment decisions. Adopting a real-options approach, Gordon, Loeb and Lucyshyn (2003) and Tatsumi and Goto (2010) have proposed dynamic versions of the model that allow analyzing the optimal timing and level of cybersecurity investment. Closer to our setup, a dynamic extension of the Gordon-Loeb model has been developed in Krutilla et al. (2021), considering the problem of optimal cybersecurity investment over an infinite time horizon and assuming that cybersecurity assets are subject to depreciation over time, while future net benefits of cybersecurity investment are discounted.

An effective cybersecurity investment policy must be adaptive and evolve in response to changing threat environments. As noted by Zeller and Scherer (2022), a key feature of cyber-risk is its dynamic nature, due to the rapid technological transformation and the evolution of threat actors. Similarly, Balzano and Marzi (2025) emphasize the need for adaptable and responsive cybersecurity policies in order to face the challenge of dynamic cyberattacks. The framework of Krutilla et al. (2021) is based on a deterministic model and, therefore, cannot capture the dynamic behavior of cyber-risk. Addressing this need, the main contribution of this work consists in proposing a modeling framework for optimal cybersecurity investment in a dynamic stochastic setup, allowing for investment policies which respond in real time to randomly occurring cyberattacks. Our work therefore contributes both to cyber-risk modeling and to cyber-risk management, as categorized in the recent survey by He et al. (2024). Moreover, our stochastic modeling framework takes into account the empirically relevant feature of temporally clustered cyberattacks.

A distinctive feature of our modeling framework, which will be described in Section 2.2.2, is indeed the use of a Hawkes process to model the arrival of cyberattacks. First introduced by Alan G. Hawkes in Hawkes (1971), these stochastic processes are used to model event arrivals over time and are particularly suited to situations where the occurrence of one event increases the likelihood of subsequent events (self-excitation), thereby generating temporally clustered events. This modeling choice is particularly relevant in the context of cyber-risk. Cyberattacks frequently occur in bursts, for instance following the discovery of a vulnerability or due to the propagation of malware across interconnected systems (see Nguyen et al. (2024)). Such clustered patterns are not adequately captured by memoryless models, such as those based on Poisson processes. Empirical evidence supports the appropriateness of the Hawkes framework for modeling cyber-risk. A contagious behavior in cyberattacks has been documented in Baldwin et al. (2017), analyzing the threats to key internet services using data from the SANS Institute. Using data from from the Privacy Rights Clearinghouse, it has been empirically demonstrated in Bessy-Roland et al. (2021) that Hawkes-based models provide a more realistic representation of the interdependence of data breaches compared to Poisson-based models. The recent work Boumezoued et al. (2023) reinforces this perspective by calibrating a two-phase Hawkes model to cyberattack data taking into account publication of cyber-vulnerabilities. These studies provide strong support for modeling cyberattacks via Hawkes processes, as described in more detail in Section 2.2.2.

In this work, we address the challenge of optimal cybersecurity investment under temporally clustered cyberattacks, in line with the empirical evidence reported above. In particular, we aim at studying the adaptive investment policy that best responds in real time to the random arrival of cyberattacks, within a framework that balances realism with analytical tractability. To this end, we develop a continuous-time stochastic extension of the classical Gordon-Loeb model, describing attack arrivals with a Hawkes process. The model incorporates key operational features such as technological obsolescence and the decreasing marginal effectiveness of large investments. The resulting optimization problem is cast as a stochastic optimal control problem and solved via dynamic programming methods. We develop efficient numerical schemes to compute the optimal policy and we quantify the benefits of dynamic investment strategies under clustered attacks. By integrating risk dynamics into the cybersecurity investment problem, our framework provides new insights into how organizations can better allocate resources to mitigate cyber-risk.

The chapter is organized as follows: In Section 2.2, we recall the original Gordon-Loeb model and introduce our continuous-time stochastic extension. In Section 2.3, we formulate the cybersecurity investment problem and characterize the optimal policy, proving some regularity properties of the value function, and a verification theorem. Section 2.4 details the model parameters and the numerical methods used in our analysis. Section 2.5 presents the results of our numerical analysis and discusses their practical implications for cyber-risk management. Section 2.6 concludes.

2.2 The model

We study the decision problem faced by an entity (a public administration or a large corporation) that is threatened by a massive number of randomly occurring cyberattacks with a temporally clustered pattern. As in the Gordon-Loeb model (reviewed in Section 2.2.1), not all cyberattacks result in successful breaches of the entity's system. The success rate of each attack depends on the system's vulnerability, which the entity can mitigate by investing in cybersecurity.

2.2.1 The Gordon-Loeb model

The Gordon-Loeb model, introduced in 2002 in the seminal work Gordon and Loeb (2002), provides a static framework for determining the optimal investment in cybersecurity to protect a given information set under the threat of cyberattacks. In our context, the information set corresponds to the entity's IT infrastructure. In the Gordon-Loeb model, the information set is characterized by three key parameters, all assumed to be constant:

- $p \in [0,1]$: the probability that a cyberattack occurs;
- $v \in [0,1]$: the probability that a cyberattack successfully breaches the information set (vulnerability);
- $\ell \geq 0$: the loss incurred when a breach occurs.

Without any cybersecurity investment, the expected loss is $vp\ell$. To mitigate its vulnerability, the entity may invest an amount $z \geq 0$ in cybersecurity. The effectiveness of this investment is measured by a security breach probability function $S(z,v)$, which represents the probability that an attack successfully breaches the information set, given investment level z and initial vulnerability v. The resulting expected loss is thus $S(z,v)p\ell$. Gordon and Loeb require the function S to satisfy the properties listed in the following assumption.

Assumption A.

(A1) $S(z,0) = 0$, for all $z \geq 0$, i.e., an invulnerable information set always remains invulnerable;

(A2) $S(0,v) = v$, i.e., in the absence of any investment, the information set retains its baseline vulnerability;

(A3) S is decreasing and convex in z, so that $S_z(z,v) < 0$ and $S_{zz}(z,v) > 0$, for all $z \geq 0$, i.e., cybersecurity investment reduces breach probability with diminishing marginal effectiveness.

Gordon and Loeb consider two classes of security breach probability functions, which satisfy Assumption A:

$$S_I(z,v) = \frac{v}{(az+1)^b} \qquad \text{and} \qquad S_{II}(z,v) = v^{az+1}, \tag{2.1}$$

for some parameters $a, b > 0$.

In Gordon and Loeb (2002), the optimal investment in cybersecurity is determined by maximizing the Expected Net Benefit of Investment in information Security (ENBIS), defined as follows:

$$\text{ENBIS}(z) := \big(v - S(z, v)\big)p\ell - z. \tag{2.2}$$

The ENBIS function quantifies the net trade-off between the benefit (captured by the reduction in the expected loss due to the investment in cybersecurity) and the direct cost of investing. Under Assumption A, the optimal investment level z^* is determined by the following first-order condition:

$$-S_z(z^*, v)p\ell - 1 = 0.$$

Remark 2.1. For both classes of security breach functions in (2.1), Gordon and Loeb show that the optimal cybersecurity investment never exceeds $1/e \approx 37\%$ of the expected loss:

$$z^* < \frac{vp\ell}{e}.$$

2.2.2 A continuous-time model driven by a Hawkes process

We now introduce a continuous-time model for randomly occurring cyberattacks. We want to capture the empirically relevant feature of clustering of cyberattacks, while retaining the key elements of the original Gordon-Loeb model reviewed in Section 2.2.1.

The arrival of cyberattacks is described by a Hawkes process $N = (N_t)_{t\geq 0}$, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with N_t representing the number of cyberattacks up to time t, for all $t \geq 0$. The process N is characterized by a self-exciting stochastic intensity $(\lambda_t)_{t\geq 0}$ solving the following stochastic differential equation:

$$\mathrm{d}\lambda_t = \beta(\alpha - \lambda_t)\,\mathrm{d}t + \xi \mathrm{d}N_t, \quad \lambda_0 > 0, \tag{2.3}$$

where

- $\alpha > 0$ is the long-term mean intensity;
- $\lambda_0 > 0$ is the intensity at the initial time $t = 0$;
- $\beta > 0$ is the exponential decay rate;
- $\xi > 0$ determines the magnitude of self-excitation;
- $(T_i)_{i\in\mathbb{N}^*}$ are the random times at which attacks occur.

The intensity explicit solution is given by

$$\lambda_t = \alpha + (\lambda_0 - \alpha)e^{-\beta t} + \xi \sum_{i=1}^{N_t} e^{-\beta(t-T_i)}.$$

Figure 2.1 shows a simulated trajectory of N and λ, showing the clustering behavior induced by the self-exciting mechanism described above. General presentations of the theory and the applications of Hawkes processes can be found in Laub et al. (2021) and Lima (2023).

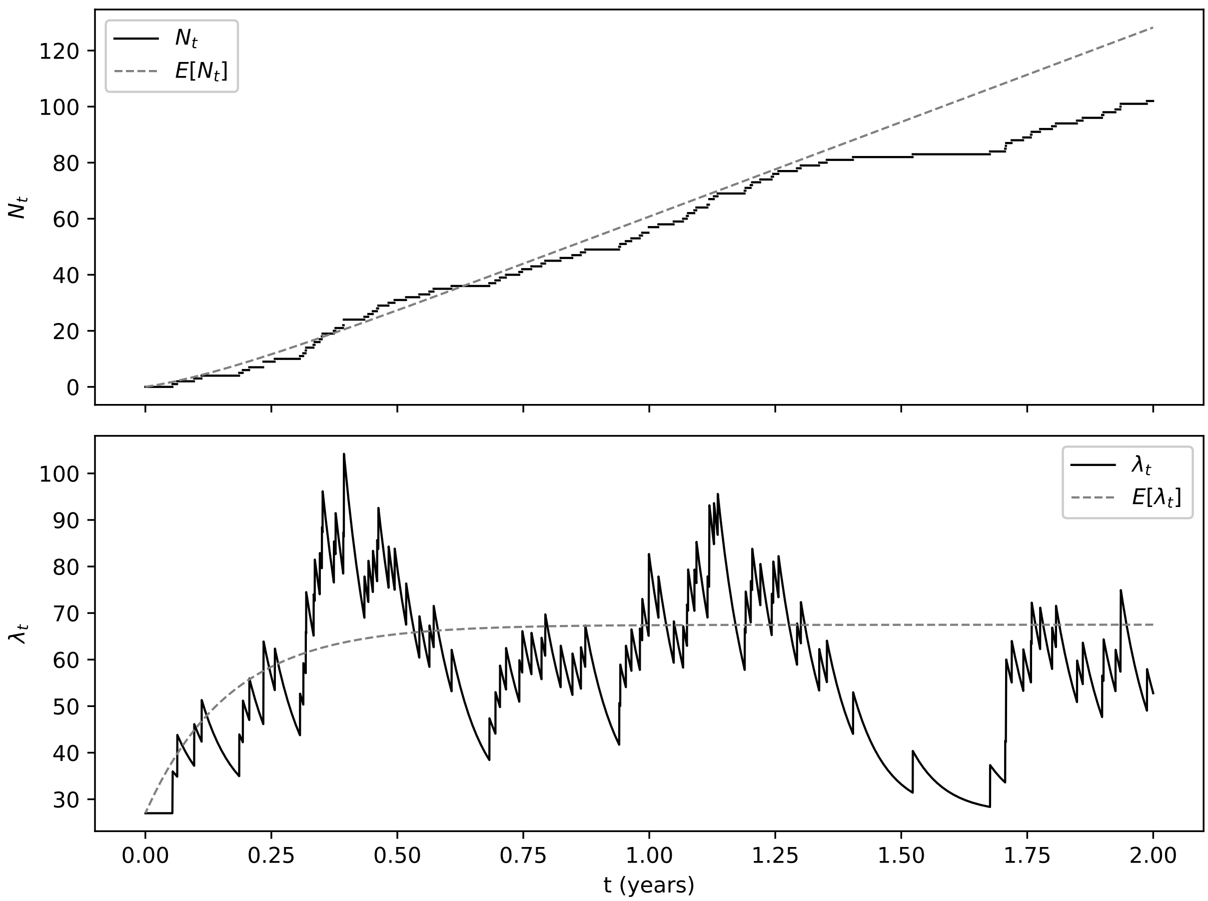


Figure 2.1: One simulated trajectory of N (top) and λ (bottom), for $\alpha = 27$, $\lambda_0 = 27$, $\beta = 15$, $\xi = 9$.

We assume throughout the chapter that $\xi < \beta$. The latter condition ensures that the L^1-norm of the self-excitation kernel of the Hawkes process is strictly less than one. This guarantees that the process is non-explosive, meaning that it generates almost surely a finite number of events over any finite time interval. The same condition also corresponds to the stationarity condition, widely adopted in the theory of Hawkes processes since the seminal work of Hawkes and Oakes (1974).

In the next proposition (adapted from Dassios and Zhao (2013)), we compute the expectation of some basic quantities which will be used later.

Proposition 2.2. *Let $(N_t)_{t\geq 0}$ be a Hawkes process with intensity $(\lambda_t)_{t\geq 0}$ given in (2.3).Then, for all $t \geq 0$,*

$$\mathbb{E}[\lambda_t] = \frac{\alpha\beta}{\beta-\xi} + e^{-(\beta-\xi)t}\left(\lambda_0 - \frac{\alpha\beta}{\beta-\xi}\right),$$
$$\mathbb{E}[N_t] = \int_0^t \mathbb{E}[\lambda_s]\mathrm{d}s = \frac{\alpha\beta}{\beta-\xi}t - \frac{1}{\beta-\xi}\left(\lambda_0 - \frac{\alpha\beta}{\beta-\xi}\right)\left(e^{-(\beta-\xi)t} - 1\right).$$

We denote by $(T_i)_{i\in\mathbb{N}^*}$ the random jump times of the process N, representing the arrival times of cyberattacks. For each $t \geq 0$, we denote by $\mathcal{F}_t := \sigma(N_s; s \leq t)$ the natural filtration generated by the Hawkes process N, representing the information generated by the history of the attack timings up to time t. The natural filtration of N is right-continuous, see Brémaud (1981, Theorem III.T1).

In the absence of cybersecurity investment, each attack is assumed to breach the entity's IT system with fixed probability v (vulnerability). For each $i \in \mathbb{N}^*$, we introduce a Bernoulli random variable B_i^v of parameter v, where the event $\{B_i^v = 1\}$ corresponds to a successful breach caused by the i-th attack. In the event of a breach, the entity incurs a random monetary loss η_i, realized at the attack time T_i. Otherwise, if $B_i^v = 0$, the attack is blocked and no loss occurs at time T_i.

The families of random variables $(B_i^v)_{i\in\mathbb{N}^*}$ and $(\eta_i)_{i\in\mathbb{N}^*}$ are assumed to satisfy the following standing assumption.

Assumption B. The family $(\eta_i)_{i\in\mathbb{N}^*}$ is composed by i.i.d. positive random variables in $L^1(\mathbb{P})$. The families $(\eta_i)_{i\in\mathbb{N}^*}$ and $(B_i^v)_{i\in\mathbb{N}^*}$ are mutually independent and independent of N.

The cumulative loss incurred over a planning horizon $[0, T]$, in the absence of any cybersecurity investment, is given by:

$$L_T^0 := \sum_{i=1}^{N_T} \eta_i B_i^v. \tag{2.4}$$

In our dynamic model, the entity can react to the evolving threat environment by investing in cybersecurity, in order to mitigate its vulnerability to cyberattacks. Investment occurs continuously throughout the planning horizon $[0, T]$ and is described by a non-negative investment rate process $(z_t)_{t\in[0,T]}$. For each $t < T$, the quantity z_t represents the increase in the level of cybersecurity over the infinitesimal time interval $[t, t + \mathrm{d}t]$. We require that the control process $(z_t)_{t\in[0,T]}$ be predictable with respect to the filtration $(\mathcal{F}_t)_{t\geq 0}$ generated by N. We point out that, in our setup, the outcomes of previous attacks (i.e., whether breaches have occurred or not) do not carry any relevant informational content for decision making, as they do not affect the dynamics of future attack arrivals.

Investment in cybersecurity is subject to rapid technological obsolescence (see, e.g., Hayes and Bodhani (2013)). In line with Krutilla et al. (2021), we take into account this significant aspect in our model by introducing a depreciation rate $\rho > 0$. The cybersecurity level reached at time t is then defined as follows, for all $t \in [0, T]$:

$$H_t = H_0 e^{-\rho t} + \int_0^t e^{-\rho(t-s)} z_s \mathrm{d}s, \tag{2.5}$$

which equivalently, in differential form, reads as follows:

$$\mathrm{d}H_t = (z_t - \rho H_t)\mathrm{d}t, \qquad H_0 \geq 0.$$

As in Krutilla et al. (2021), we interpret the cybersecurity level as an aggregated asset, which can be thought of as a combination of technological infrastructures, software, and human expertise.

In our continuous-time framework, we let the breach probability evolve dynamically with the current cybersecurity level. More specifically, suppose that the decision maker adopts an investment policy $z = (z_t)_{t\in[0,T]}$. At each attack time T_i, a breach is assumed to occur with probability

$$S(H_{T_i}, v), \tag{2.6}$$

where H_{T_i} is given by (2.5) evaluated at $t = T_i$ and S is a security breach probability function satisfying Assumption A, as in the original Gordon-Loeb model. Hence, the probability that the i-th attack successfully breaches the IT system depends on the cybersecurity level H_{T_i} reached at the attack's time T_i. In turn, H_{T_i} is determined by the investment realized over the time period $[0, T_i]$, taking into account technological obsolescence. If the i-th attack breaches the IT system, then the entity incurs into a loss of η_i, otherwise the attack is blocked and the entity does not suffer any loss at time T_i.

Remark 2.3. The proposed model allows for adaptive real-time cybersecurity investment. More specifically, the arrival of an attack triggers an increased likelihood of further attacks within a short timeframe, due to the form (2.3) of the intensity. The decision maker can respond in real-time by increasing cybersecurity investment, which in turn reduces future breach probabilities through the function S in (2.6). The optimal investment policy will be determined in Section 2.3, while the practical importance of allowing for an adaptive real-time investment strategy - rather than a static policy as in the original Gordon-Loeb model - will be empirically analyzed in Section 2.5.3.

Analogously to the case without investment in cybersecurity, we can write as follows the cumulative losses L_T^z incurred on the time interval $[0,T]$ when investing in cybersecurity according to a generic rate $z = (z_t)_{t\in[0,T]}$:

$$L_T^z := \sum_{i=1}^{N_T} \eta_i B_i^{S(H_{T_i},v)}, \tag{2.7}$$

where $(B_i^{S(H_{T_i},v)})_{i\in\mathbb{N}^*}$ is a family of random variables taking values in $\{0,1\}$ and satisfying the following assumption.

Assumption C. For any process $(z_t)_{t\in[0,T]}$, it holds that

$$\mathbb{P}\Big(B_i^{S(H_{T_i},v)} = 1 \Big| \mathcal{F}_T\Big) = S(H_{T_i}, v), \qquad \text{for all } i \in \mathbb{N}^*,$$

where $(H_t)_{t\in[0,T]}$ is determined by $(z_t)_{t\in[0,T]}$ as in (2.5). Moreover, for each $i \in \mathbb{N}^*$, the random variables $B_i^{S(H_{T_i},v)}$ and η_i are conditionally independent given $\mathcal{F}_T$.

Remark 2.4. The cumulative loss process $(L_t^z)_{t\in[0,T]}$ defined as in (2.7) constitutes a *marked Hawkes process*, in the terminology of point processes (see Brémaud (1981)). In our modeling framework, the marks (losses) are endogenous and depend on the dynamically evolving cybersecurity level $(H_t)_{t\in[0,T]}$.

For strategic decision making, a key quantity is represented by the expected losses due to cyberattacks over the time interval $[0,T]$ when adopting a suitable cybersecurity policy. This is the content of the following proposition, which will be fundamental for addressing the optimal investment problem in Section 2.3. We denote by $\bar{\eta} := \mathbb{E}[\eta_i]$ the expected loss resulting from a successful breach, for all $i \in \mathbb{N}^*$.

Proposition 2.5. *Under Assumptions B and C, it holds that*

$$\mathbb{E}[L_T^0] = \bar{\eta}\, v\, \mathbb{E}\left[\int_0^T \lambda_t \mathrm{d}t\right],$$

$$\mathbb{E}[L_T^z] = \bar{\eta}\, \mathbb{E}\left[\int_0^T S(H_t, v)\lambda_t \mathrm{d}t\right].$$

Therefore, the expected net benefit of investment is

$$\mathbb{E}[L_T^0 - L_T^z] = \bar{\eta}\,\mathbb{E}\left[\int_0^T \big(v - S(H_t, v)\big)\lambda_t \mathrm{d}t\right].$$

Proof. Let $z = (z_t)_{t\in[0,T]}$ be an arbitrary cybersecurity investment rate process. Applying the tower property of conditional expectation and making use of Assumptions B and C, we can compute

$$\begin{aligned}\mathbb{E}[L_T^z] &= \mathbb{E}\left[\sum_{i=1}^{N_T} \eta_i B_i^{S(H_{T_i},v)}\right] = \mathbb{E}\left[\sum_{i=1}^{N_T} \mathbb{E}\left[\eta_i B_i^{S(H_{T_i},v)} \middle| \mathcal{F}_T\right]\right] = \bar{\eta}\,\mathbb{E}\left[\sum_{i=1}^{N_T} S(H_{T_i}, v)\right] \\ &= \bar{\eta}\,\mathbb{E}\left[\int_0^T S(H_t, v)\mathrm{d}N_t\right] = \bar{\eta}\,\mathbb{E}\left[\int_0^T S(H_t, v)\lambda_t \mathrm{d}t\right],\end{aligned}$$

where the last step follows by definition of intensity (see, e.g., Brémaud (1981, Definition II.D7)), together with the continuity of the process H. The first equation in the statement of the proposition follows as a special case by taking $z \equiv 0$. □

2.3 Optimal cybersecurity investment

In this section, we determine the optimal cybersecurity investment policy, in the model setup introduced in Section 2.2.2. In the spirit of the original Gordon-Loeb model, we aim at characterizing the investment rate process $z^* = (z_t^*)_{t\in[0,T]}$ which maximizes the net trade-off between the benefits and the costs of cybersecurity over a planning horizon $[0, T]$.

To ensure the well-posedness of the optimization problem, we constrain the admissible investment policies to a suitably defined admissible set $\mathcal{Z}$.

Definition 2.6. The admissible set $\mathcal{Z}$ is defined as the set of all non-negative, $\mathcal{F}_t$-predictable processes $(z_t)_{t\in[0,T]}$ such that $\mathbb{E}[\int_0^T z_t^2 \mathrm{d}t] < \infty$.

Remark 2.7. We point out that, as a direct consequence of the Cauchy-Schwarz inequality, the integral in (2.5) is always well-defined for every process $z \in \mathcal{Z}$.

We now formulate the central optimization problem, which generalizes the benefit-cost trade-off function in (2.2) to a dynamic stochastic setting. The objective is to maximize the expected net benefit of cybersecurity investments:

$$\sup_{z\in\mathcal{Z}} \mathbb{E}\left[L_T^0 - L_T^z - \int_0^T \left(\delta z_t + \frac{\gamma}{2} z_t^2\right) \mathrm{d}t + U(H_T)\right], \tag{2.8}$$

where L_T^0 and L_T^z are defined in (2.4) and (2.7), respectively, and the state variables λ and H satisfy the dynamics

$$\mathrm{d}\lambda_t = \beta(\alpha - \lambda_t)\mathrm{d}t + \xi \mathrm{d}N_t, \tag{2.9}$$

$$\mathrm{d}H_t = (z_t - \rho H_t)\mathrm{d}t. \tag{2.10}$$

In the objective functional (2.8), the term $\mathbb{E}[L_T^0 - L_T^z]$ represents the reduction in the expected losses due to the investment in cybersecurity. Differently from (2.2), we consider in (2.8) a non-linear cost function $z \mapsto \delta z + \gamma z^2/2$, for $\delta, \gamma > 0$. The non-linearity penalizes irregular or highly concentrated investment strategies, reflecting real-world constraints and incentivizing smoother, more sustained cybersecurity efforts (e.g., continuous IT updates versus abrupt large-scale interventions). The term $U(H_T)$ accounts for the residual utility of the cybersecurity level reached at the end of the planning horizon. This accounts for the fact that cybersecurity investment carries a long-term benefit, since the entity does not cease to exist after the planning horizon. The function $U : \mathbb{R}_+ \to \mathbb{R}_+$ is assumed to be a non-negative, increasing and concave utility function.

Up to a rescaling of the model parameters, there is no loss of generality in taking $\delta = 1$. Hence, making use of Proposition 2.5, we can equivalently rewrite problem (2.8) as follows:

$$\sup_{z \in \mathcal{Z}} \mathbb{E}\left[\int_0^T \left(\bar{\eta}\big(v - S(H_t, v)\big)\lambda_t - z_t - \frac{\gamma}{2} z_t^2\right) \mathrm{d}t + U(H_T)\right]. \tag{2.11}$$

Problem (2.11) is a bi-dimensional stochastic optimal control problem, where the stochastic intensity process $(\lambda_t)_{t\in[0,T]}$ acts as an additional state variable beyond the controlled process $(H_t)_{t\in[0,T]}$. Due to the Markovian structure of the system, dynamic programming techniques can be applied for the solution of (2.11). To this end, we first introduce the following notation, for any $(t, \lambda, h) \in [0,T] \times (0,\infty) \times \mathbb{R}_+$:

- $$H_s^{t,h,z} := h e^{-\rho(s-t)} + \int_t^s e^{-\rho(s-v)} z_v \mathrm{d}v, \tag{2.12}$$

  for all $s \in [t,T]$, representing the cybersecurity level reached at time s when starting from level $H_t = h$ at time t and investing according to a process $z \in \mathcal{Z}$;

- $$\lambda_s^{t,\lambda} := \alpha + (\lambda - \alpha)e^{-\beta(s-t)} + \xi \sum_{i=N_t+1}^{N_s} e^{-\beta(s-T_i)}, \tag{2.13}$$

  for all $s \in [t,T]$, representing the stochastic intensity at time s when starting from value $\lambda_t = \lambda$ at time t.

For any stopping time τ taking values in $[0,T]$, we denote by $\mathcal{Z}_\tau$ the set of all processes $(z_t)_{t\in[0,T]}$ such that $(z_{\tau\vee t} - z_\tau)_{t\in[0,T]} \in \mathcal{Z}$.

We define as follows the benefit-cost trade-off functional J associated to a given investment rate process z:

$$J(t,\lambda,h;z) := \mathbb{E}\left[\int_t^T \bar{\eta}(v - S(H_s^{t,h,z}, v))\lambda_s^{t,\lambda}\mathrm{d}s - \int_t^T \left(z_s + \frac{\gamma}{2}z_s^2\right)\mathrm{d}s + U(H_T^{t,h,z})\right].$$

Consequently, the value function associated to the stochastic optimal control problem (2.11) is given by

$$V(t,\lambda,h) := \sup_{z\in\mathcal{Z}_t} J(t,\lambda,h;z). \tag{2.14}$$

In our dynamic model, the value function $V(t,\lambda,h)$ encodes the benefit-cost trade-off of cybersecurity investment over the residual planning horizon $[t,T]$, when considered at time t with current cybersecurity level h and intensity λ. In the next propositions, we examine some properties of the function V.

Proposition 2.8. *For $(t,\lambda,h) \in [0,T)\times(0,\infty)\times\mathbb{R}_+$, $V(t,\lambda,h)$ is non-negative and has linear growth, i.e.*

$$V(t,\lambda,h) \le C(1+\lambda+h),$$

for some positive constant C.

Proof. By definition, $V(t,\lambda,h) \ge J(t,\lambda,h;0) = \mathbb{E}\left[\int_t^T \bar{\eta}(v - S(h,v))\lambda_s^{t,\lambda}\mathrm{d}s + U(h)\right] \ge 0$.

To prove the linear growth, we recall that $v - S(\cdot, v)$ is bounded, so we can write

$$\begin{aligned}
V(t,\lambda,h) &\le \sup_{z\in\mathcal{Z}_t}\mathbb{E}\left[\int_t^T 2v\bar{\eta}\lambda_s^{t,\lambda}\mathrm{d}s - \int_t^T\left(z_s + \frac{\gamma}{2}z_s^2\right)\mathrm{d}s + U(H_T^{t,h,z})\right] \\
&\le 2v\bar{\eta}\left(\frac{\lambda}{\beta-\xi}\left(1 - e^{-(\beta-\xi)(T-t)}\right) + \frac{\alpha\beta}{\beta-\xi}\left((T-t) - \frac{1}{\beta-\xi}\left(1 - e^{-(\beta-\xi)(T-t)}\right)\right)\right) \\
&\quad + \sup_{z\in\mathcal{Z}_t}\mathbb{E}\left[-\int_t^T\left(z_s + \frac{\gamma}{2}z_s^2\right)\mathrm{d}s + U(H_T^{t,h,z})\right] \\
&\le C(1+\lambda) + \sup_{z\in\mathcal{Z}_t}\mathbb{E}\left[-\int_t^T\left(z_s + \frac{\gamma}{2}z_s^2\right)\mathrm{d}s + U(H_T^{t,h,z})\right],
\end{aligned}$$

The inequality uses the expectation of N in Proposition 2.2, and the positivity and monotonicity properties of $x \in [0,\infty) \mapsto x - \frac{1}{\beta-\xi}(1 - e^{-(\beta-\xi)x})$. Since U is a one variable concave function, it exists $x_0 > 0$ such that the derivative $U'(x_0)$ is finite and

$U(x) \le U(x_0) + U'(x_0)(x - x_0)$, hence

$$
\begin{aligned}
U(H_T^{t,h,z}) &\le U(x_0) + U'(x_0)(H_T^{t,h,z} - x_0) \le U(x_0) + U'(x_0)\left(h + \int_t^T z_s \mathrm{d}s - x_0\right) \\
&\le U(x_0) + U'(x_0)\left(h + \int_t^T z_s \mathrm{d}s\right) - U'(x_0)x_0 \le C\left(1 + h + \int_t^T z_s \mathrm{d}s\right),
\end{aligned}
$$

where we exploit the integral expression of H in equation (2.12), and the fact that U is non-negative and increasing. We can assume C is a positive constant. It follows that

$$
\begin{aligned}
V(t,\lambda,h) &\le C(1+\lambda+h) + \sup_{z\in\mathcal{Z}_t} \mathbb{E}\left[-\int_t^T \left(\tilde{C} z_s + \frac{\gamma}{2} z_s^2\right) \mathrm{d}s\right] \\
&\le C(1+\lambda+h) + \frac{\max(0,-\tilde{C})^2}{2\gamma} T \le C(1+\lambda+h).
\end{aligned}
$$

where $\tilde{C}$ is a real number which depends on the previous estimates, and $C > 0$ denotes a positive constant whose value may change from one occurrence to the next. □

Proposition 2.9. *i) For every $(t,h) \in [0,T) \times \mathbb{R}_+$, the map $\lambda \mapsto V(t,\lambda,h)$ is strictly increasing and globally Lipschitz.*

ii) For every $(t,\lambda) \in [0,T] \times (0,+\infty)$, the map $h \mapsto V(t,\lambda,h)$ is strictly increasing and concave. Moreover, uder the assumption that U and $S(\cdot, v)$ are uniformly Lipschitz in h, the map is Lipschitz on $\mathbb{R}_+$, with a Lipschitz constant depending linearly on λ.

Proof. i) Let $\lambda_1 < \lambda_2$. For $s > t$, $\lambda_s^{t,\lambda_2} - \lambda_s^{t,\lambda_1}$ is a strictly positive process, see Gaïgi et al. (2025, Proposition 3.1). It follows that $\lambda_s^{t,\lambda_2} > \lambda_s^{t,\lambda_1}$ almost everywhere. For every fixed $z, h \ge 0$, the quantity $(v - S(H_s^{t,h,z}, v))\bar{\eta}$ is non-negative, thus

$$
\begin{aligned}
J(t,\lambda_1,h;z) &= \mathbb{E}\left[\int_t^T \left[(v - S(H_s^{t,h,z}, v))\bar{\eta}\lambda_s^{t,\lambda_1} - z_s - \frac{\gamma}{2} z_s^2\right] \mathrm{d}s + U(H_T^{t,h,z})\right] \\
&< \mathbb{E}\left[\int_t^T \left[(v - S(H_s^{t,h,z}, v))\bar{\eta}\lambda_s^{t,\lambda_2} - z_s - \frac{\gamma}{2} z_s^2\right] \mathrm{d}s + U(H_T^{t,h,z})\right] \le V(t,\lambda_2,h).
\end{aligned}
$$

Taking z as the optimal control for the initial values (t,λ_1,h), we get that V is increasing in λ.

To prove the Lipschitz property, we fix t and z, admissible control, and write for two general initial conditions $\lambda_1, \lambda_2 > 0$

$$
|J(t,\lambda_1,h;z) - J(t,\lambda_2,h;z)|
$$

$$
\begin{aligned}
&= \left| \mathbb{E}\left[\int_t^T \left[(v - S(H_s^{t,h,z}, v))\bar{\eta}\lambda_s^{t,\lambda_1} - (v - S(H_s^{t,h,z}, v))\bar{\eta}\lambda_s^{t,\lambda_2} \right] \mathrm{d}s \right.\right. \\
&\left.\left. + U(H_T^{t,h,z}) - U(H_T^{t,h,z}) \right] \right| \\
&\leq \mathbb{E}\left[\int_t^T |v - S(H_s^{t,h,z})||\lambda_s^{t,\lambda_1} - \lambda_s^{t,\lambda_2}| \mathrm{d}s \right] \leq C\mathbb{E}\left[\int_t^T |\lambda_s^{t,\lambda_1} - \lambda_s^{t,\lambda_2}| \mathrm{d}s \right],
\end{aligned}
$$

where we exploit that $v - S(\cdot, v)$ is non-negative and bounded. Assume $\lambda_2 > \lambda_1$, which leads to $\lambda_s^{t,\lambda_2} > \lambda_s^{t,\lambda_1}$. Then, applying the formula for the expectation of N in Proposition 2.2, we get

$$
\mathbb{E}\left[\int_t^T \lambda_s^{t,\lambda_2} - \lambda_s^{t,\lambda_1} \mathrm{d}s \right] = \frac{1}{\beta - \xi}(\lambda_2 - \lambda_1)\left(1 - e^{-(\beta-\xi)(T-t)}\right) \leq C|\lambda_2 - \lambda_1|.
$$

The same estimate holds choosing $\lambda_1 > \lambda_2$. Thus it follows

$$
|J(t, \lambda_1, h; z) - J(t, \lambda_2, h; z)| \leq C\mathbb{E}\left[\int_t^T |\lambda_s^{t,\lambda_1} - \lambda_s^{t,\lambda_2}| \mathrm{d}s \right] \leq C|\lambda_1 - \lambda_2|.
$$

Now consider $V(t, \lambda_1, h)$, for any $\varepsilon > 0$, there exists z_1^* such that

$$
J(t, \lambda_1, h; z_1^*) > V(t, \lambda_1, h) - \varepsilon.
$$

We can then write

$$
\begin{aligned}
V(t, \lambda_1, h) - V(t, \lambda_2, h) &< J(t, \lambda_1, h; z_1^*) - J(t, \lambda_2, h; z_1^*) + \varepsilon \\
&\leq |J(t, \lambda_1, h; z_1^*) - J(t, \lambda, h_2; z_1^*)| + \varepsilon \\
&\leq C|\lambda_1 - \lambda_2| + \varepsilon.
\end{aligned}
$$

We repeat the same reasoning swapping (λ_1, h), (λ_2, h) and since ε is arbitrary, we conclude that $\lambda \mapsto V(t, \lambda, h)$ is Lipschitz.

ii) Let $h_1 < h_2$, then for every fixed control z, for every time $s > t$, $H_s^{t,h_1,z} < H_s^{t,h_2,z}$ almost everywhere. Since $S(h, v)$ is decreasing in h, see Assumption (A3), $-S(h, v)$ is increasing in h. The function U is increasing in h by hypothesis, thus it follows:

$$
\begin{aligned}
J(t, \lambda, h_1; z) &= \mathbb{E}\left[\int_t^T \left(v - S(H_s^{t,h_1,z}, v))\bar{\eta}\lambda_s^{t,\lambda} - z_s - \frac{\gamma}{2}z_s^2 \right) \mathrm{d}s + U(H_T^{t,h_1,z}) \right] \\
&< \mathbb{E}\left[\int_t^T \left((v - S(H_s^{t,h_2,z}, v))\bar{\eta}\lambda_s^{t,\lambda} - z_s - \frac{\gamma}{2}z_s^2 \right) \mathrm{d}s + U(H_T^{t,h_2,z}) \right] \leq V(t, \lambda, h_2).
\end{aligned}
$$

In particular, taking z as the optimal control for the initial values (λ, h_1) at time t, we get the result. To prove the concavity, we consider $\ell \in (0,1)$ and two general h_1, h_2. For fixed z_1, z_2, recalling the integral expression of H in equation (2.12), we write

$$\begin{aligned} H_s^{t,\ell h_1+(1-\ell)h_2,\ell z_1+(1-\ell)z_2} &:= (\ell h_1 + (1-\ell)h_2)e^{-\rho(s-t)} + \int_t^s e^{-\rho(s-v)}(\ell z_{1,v} + (1-\ell)z_{2,v})\mathrm{d}v \\ &= \ell H_s^{t,h_1,z_1} + (1-\ell)H_s^{t,h_2,z_2}. \end{aligned}$$

By assumption, $-S(h,v)$ and U are concave in h, and the quadratic cost is concave in z thus it follows

$$J(t,\lambda,\ell h_1 + (1-\ell)h_2; \ell z_1 + (1-\ell)z_2) \geq \ell J(t,\lambda,h_1;z_1) + (1-\ell)J(t,\lambda,h_2;z_2).$$

To prove the concavity of V, we recall that by definition of V for any $\varepsilon > 0$, there exists z_1^*, z_2^* such that

$$\begin{aligned} J(t,\lambda,h_1;z_1^*) &> V(t,\lambda,h_1) - \varepsilon, \\ J(t,\lambda,h_2;z_2^*) &> V(t,\lambda,h_2) - \varepsilon. \end{aligned}$$

By choosing $z_1 = z_1^*$ and $z_2 = z_2^*$, we get

$$\begin{aligned} V(t,\lambda,\ell h_1 + (1-\ell)h_2) &\geq J(t,\lambda,\ell h_1 + (1-\ell)h_2; \ell z_1^* + (1-\ell)z_2^*) \\ &> \ell J(\lambda,h_1;z_1^*) + (1-\ell)J(\lambda,h_2;z_2^*) \\ &> \ell V(t,\lambda,h_1) + (1-\ell)V(t,\lambda,h_2) - 2\varepsilon. \end{aligned}$$

We now focus on the Lipschitz property. Recall the integral formula of H in equation (2.12) and observe that for z fixed, $s > t$:

$$|H_s^{t,h_1,z} - H_s^{t,h_2,z}| = |(h_1 - h_2)e^{-\rho(s-t)}| \leq |h_1 - h_2|.$$

We write

$$\begin{aligned} &|J(t,\lambda,h_1;z) - J(t,\lambda,h_2;z)| \\ &= \left| \mathbb{E}\left[\int_t^T \left((v - S(H_s^{t,h_1,z},v))\bar{\eta}\lambda_s^{t,\lambda} - (v - S(H_s^{t,h_2,z},v))\bar{\eta}\lambda_s^{t,\lambda} \right) \mathrm{d}s \right.\right. \\ &\left.\left. + U(H_T^{t,h_1,z}) - U(H_T^{t,h_2,z}) \right] \right| \end{aligned}$$

$$\begin{aligned}
&\leq \mathbb{E}\left[\int_t^T \left(|(v - S(H_s^{t,h_1,z}, v))\bar{\eta}\lambda_s^{t,\lambda} - (v - S(H_s^{t,h_2,z}, v))|\bar{\eta}\lambda_s^{t,\lambda}\right) \mathrm{d}s\right.\\
&\left.+|U(H_T^{t,h_1,z}) - U(H_T^{t,h_2,z})|\right]\\
&= \mathbb{E}\left[\int_t^T |S(H_s^{t,h_2,z}, v) - S(H_s^{t,h_1,z}, v)|\bar{\eta}\lambda_s^{t,\lambda}\mathrm{d}s + C|H_T^{t,h_1,z} - H_T^{t,h_2,z}|\right]\\
&= \mathbb{E}\left[\int_t^T \tilde{C}|H_s^{t,h_2,z} - H_s^{t,h_1,z}|\lambda_s^{t,\lambda}\mathrm{d}s + C|H_T^{t,h_1,z} - H_T^{t,h_2,z}|\right]\\
&\leq |h_1 - h_2|\left(\tilde{C}\mathbb{E}\left[\int_t^T \lambda_s^{t,\lambda}\mathrm{d}s\right] + C\right)\\
&\leq |h_1 - h_2|\left(\tilde{C}\frac{\alpha\beta}{\beta-\xi}(T-t) - \frac{1}{\beta-\xi}\left(\lambda - \frac{\alpha\beta}{\beta-\xi}\right)\left(e^{-(\beta-\xi)(T-t)} - 1\right) + C\right)\\
&\leq |h_1 - h_2|(\tilde{C}\lambda + C).
\end{aligned}$$

With analogous techniques as before, we can prove that also $h \mapsto V(t, \lambda, h)$ is Lipschitz with constant linearly dependent on λ.

□

Remark 2.10. The fact that V is increasing in λ indicates that the benefit of cybersecurity investment is greater in the presence of a greater risk of cyberattacks. Its increasing and concave dependence on the current cybersecurity level h shows that raising h always improves the expected future benefit, but the marginal value of additional protection decreases as the cybersecurity level h grows.

Remark 2.11. S_I and S_{II} as in equation (2.1) are uniformly Lipschitz in the investment variable.

Proposition 2.12. *For every $(t, \lambda, h) \in [0, T] \times (0, \infty) \times \mathbb{R}_+$, it holds that*

$$V(t, \lambda, h) \geq J(t, \lambda, h; \rho h), \tag{2.15}$$

where

$$\begin{aligned}
J(t, \lambda, h; \rho h) = U(h) - \rho h\left(1 + \frac{\gamma}{2}\rho h\right)(T - t)&\\
+ \bar{\eta}\big(v - S(h, v)\big)\left(\frac{\alpha\beta}{\beta-\xi}(T-t) - \frac{1}{\beta-\xi}\left(\lambda - \frac{\alpha\beta}{\beta-\xi}\right)\left(e^{-(\beta-\xi)(T-t)} - 1\right)\right)&.
\end{aligned}$$

Proof. By definition of the value function (2.14), it holds that $V(t, \lambda, h) \geq J(t, \lambda, h, z)$, for any given $z \in \mathcal{Z}_t$. In particular, the constant process $\bar{z} \equiv \rho h$ belongs to $\mathcal{Z}_t$ and, therefore, inequality (2.15) holds. In view of equation (2.12), we have that

$$H_s^{t,h,\rho h} = he^{-\rho(s-t)} + \int_t^s e^{-\rho(s-v)}\rho h \mathrm{d}v = h,$$

for all $s \in [t,T]$. Therefore, we obtain that

$$J(t,\lambda,h;\rho h) = \bar{\eta}\big(v - S(h,v)\big)\mathbb{E}\left[\int_t^T \lambda_s^{t,\lambda}\mathrm{d}s\right] - \rho h\left(1+\frac{\gamma}{2}\rho h\right)(T-t) + U(h).$$

The expectation $\mathbb{E}[\int_t^T \lambda_s^{t,\lambda}\mathrm{d}s]$ can be computed by a straightforward adaptation of Proposition 2.2 (compare also with Dassios and Zhao (2011, Theorem 3.6)), thus completing the proof. □

Remark 2.13. The lower bound obtained in Proposition 2.12 is associated to a fixed investment rate which offsets technological obsolescence by maintaining the cybersecurity level constant over time (this follows directly from equation (2.10)). In Section 2.5.3, we numerically show that the optimal dynamic investment policy characterized in Theorem 2.14 consistently outperforms any constant investment strategy, highlighting the value of real-time adaptability in cybersecurity investment.

We now proceed to characterize the value function V as the solution to a Hamilton-Jacobi-Bellman (HJB) partial integro-differential equation (PIDE). To this end, we introduce the following standard assumption.

Assumption D. The dynamic programming principle holds: for all (t,λ,h) in $[0,T)\times(0,\infty)\times\mathbb{R}_+$ and for every stopping time τ taking values in $[t,T]$, it holds that

$$V(t,\lambda,h) = \sup_{z\in\mathcal{Z}_t}\mathbb{E}\left[\int_t^\tau \bar{\eta}\big(v - S(H_s^{t,h,z},v)\big)\lambda_s^{t,\lambda}\mathrm{d}s - \int_t^\tau\left(z_s + \frac{\gamma}{2}z_s^2\right)\mathrm{d}s + V(\tau,\lambda_\tau^{t,\lambda},H_\tau^{t,h,z})\right].$$

Theorem 2.14. *For brevity of notation, in the statement and in the proof of this theorem, we omit to denote explicitly the dependence of V on its arguments (t,λ,h). Suppose that Assumption D holds. Assume furthermore that the value function V defined in* (2.14) *is of class $C^{1,1,1}$ (i.e., continuously* *differentiable in all its arguments). Then, the function V solves the following HJB-PIDE:*

$$\begin{aligned}&\frac{\partial V}{\partial t}+\beta(\alpha-\lambda)\frac{\partial V}{\partial \lambda}-\rho h\frac{\partial V}{\partial h}+\lambda(V(t,\lambda+\xi,h)-V(t,\lambda,h))\\&\quad+\bar{\eta}(v-S(h,v))\lambda+\frac{\left(\left(\frac{\partial V}{\partial h}-1\right)^{+}\right)^{2}}{2\gamma}=0,\\&V(T,\lambda,h)=U(h).\end{aligned}\tag{2.16}$$

In addition, the optimal investment rate process z^* is given by

$$z^*=\frac{\left(\frac{\partial V}{\partial h}-1\right)^{+}}{\gamma}.\tag{2.17}$$

Proof. In view of Assumption D and the assumption that V is of class $C^{1,1,1}$, standard arguments based on Itô's formula together with (2.9) and (2.10) imply that V satisfies the following HJB equation (see, e.g., Bensoussan and Chevalier-Roignant (2024, Section 5.2)):

$$\begin{aligned}0&=\sup_{z\geq 0}\Bigg(\frac{\partial V}{\partial t}+\beta(\alpha-\lambda)\frac{\partial V}{\partial \lambda}-\rho h\frac{\partial V}{\partial h}+z\frac{\partial V}{\partial h}+\lambda\big(V(t,\lambda+\xi,h)-V(t,\lambda,h)\big)\\&\qquad+\bar{\eta}(v-S(h,v))\lambda-z-\frac{\gamma}{2}z^2\Bigg)\\&=\frac{\partial V}{\partial t}+\beta(\alpha-\lambda)\frac{\partial V}{\partial \lambda}-\rho h\frac{\partial V}{\partial h}+\lambda\big(V(t,\lambda+\xi,h)-V(t,\lambda,h)\big)+\bar{\eta}(v-S(h,v))\lambda\\&\qquad+\sup_{z\geq 0}\left(z\frac{\partial V}{\partial h}-z-\frac{\gamma}{2}z^2\right).\end{aligned}$$

The supremum in the last line is given by

$$\sup_{z\geq 0}\left(z\frac{\partial V}{\partial h}-z-\frac{\gamma}{2}z^2\right)=\begin{cases}0, & \text{if } \frac{\partial V}{\partial h}\leq 1,\\ \frac{1}{2\gamma}\left(\frac{\partial V}{\partial h}-1\right)^2, & \text{otherwise,}\end{cases}$$

and is reached by the optimal control given in equation (2.17). If U and $S(\cdot,v)$ are Lipschitz, then the optimal control is admissible. Indeed, by Proposition 2.9, it follows that

$$z_t^*=\frac{\left(\frac{\partial V(t,\lambda_t,H_t)}{\partial h}-1\right)^{+}}{\gamma}<\frac{1}{\gamma}\left|\frac{\partial V(t,\lambda_t,H_t)}{\partial h}\right|+\frac{1}{\gamma}\leq\frac{1}{\gamma}\left(\tilde{C}\lambda_t+C\right),$$

thus the optimal control is admissible due to the integrability property of the intensity λ, see Dassios and Zhao (2013, Proposition 2.3). □

A numerical method for the solution of the PIDE (2.16) will be presented in Section 2.4 and then applied in Section 2.5.

The optimal investment rate z_t^* in equation (2.17) depends on current time t, on the current level λ_t of the stochastic intensity and on the current cybersecurity level H_t. In particular, the dependence on λ_t makes z_t^* adaptive, meaning that it reacts to the random arrival of cyberattacks. Since the arrival of an attack increases the likelihood of further attacks, due to the self-exciting behavior of the Hawkes process, this enables the decision maker to strategically increase the cybersecurity investment in order to raise the cybersecurity level as a defense for the incoming attacks. This important feature will be numerically illustrated in Section 2.5.4.

Remark 2.15. The optimal policy described in equation (2.17) admits a clear economic interpretation: it is worth investing in cybersecurity whenever the marginal benefit of the investment is greater than its marginal cost. This insight aligns with the earlier results of Krutilla et al. (2021) in a dynamic but deterministic setup.

We conclude this section by proving a verification theorem, under additional regularity assumptions on the value function.

Theorem 2.16 (Verification theorem). *Let w be a non-negative function in $C^{1,1,1}([0,T)\times(0,+\infty)\times\mathbb{R}_+)$ and $C^0([0,T]\times(0,+\infty)\times\mathbb{R}_+)$. Moreover, assume w has linear growth*

$$w(t,\lambda,h) \le C(1+\lambda+h).$$

i) Suppose that for $(t,\lambda,h)\in[0,T)\times(0,+\infty)\times\mathbb{R}_+$,

$$\begin{aligned} 0 \ge\ & \frac{\partial w}{\partial t} + \beta(\alpha-\lambda)\frac{\partial w}{\partial \lambda} - \rho h \frac{\partial w}{\partial h} + \lambda\big(w(t,\lambda+\xi,h) - w(t,\lambda,h)\big) \\ & + \bar{\eta}(v - S(h,v))\lambda + \sup_{z\in\mathcal{Z}}\left(z\frac{\partial w}{\partial h} - z - \frac{\gamma}{2}z^2\right) \\ w(T,\lambda,h) \ge\ & U(h). \end{aligned}$$

Then $w \ge V$ on $[0,T]\times(0,+\infty)\times\mathbb{R}_+$.

ii) Suppose that $w(T,\lambda,h) = U(h)$, and there exists a measurable non-negative function $z^*(t,\lambda,h)$, $(t,\lambda,h)\in[0,T)\times(0,+\infty)\times\mathbb{R}_+$ such that

$$0 = \frac{\partial w}{\partial t} + \beta(\alpha-\lambda)\frac{\partial w}{\partial \lambda} - \rho h \frac{\partial w}{\partial h} + \lambda\big(w(t,\lambda+\xi,h) - w(t,\lambda,h)\big)$$

$$
\begin{aligned}
&+ \bar{\eta}(v - S(h,v))\lambda + \sup_{z \in \mathcal{Z}} \left(z\frac{\partial w}{\partial h} - z - \frac{\gamma}{2} z^2 \right) \\
&= \frac{\partial w}{\partial t} + \beta(\alpha - \lambda)\frac{\partial w}{\partial \lambda} - \rho h \frac{\partial w}{\partial h} + \lambda\big(w(t,\lambda+\xi,h) - w(t,\lambda,h)\big) \\
&+ \bar{\eta}(v - S(h,v))\lambda + z^* \frac{\partial w}{\partial h} - z^* - \frac{\gamma}{2}(z^*)^2,
\end{aligned}
$$

the SDE (2.10) evaluated in $z = z^*$ admits a unique solution and it is denoted by H^*, the process $\{z^*(t, \lambda_s^{t,\lambda}, H_s^{*t,h}), t \le s \le T\}$ lies in $\mathcal{Z}_t$. Then $w = V$ on $[0,T] \times (0,+\infty) \times \mathbb{R}_+$, and z^* is an optimal Markovian control.

Proof. i) Suppose $w \in C^{1,1,1}$, $t < T$, $z \in \mathcal{Z}_t$ and $s \in [t,T)$. Let τ be a stopping time valued in $[t,T)$. We apply Ito's formula to $w(s \wedge \tau, \lambda_{s\wedge\tau}^{t,\lambda}, H_{s\wedge\tau}^{t,h})$. For simplicity, we write w and λ_t, H_t, z_t, omitting the dependence on the initial state (t,λ,h)

$$
\begin{aligned}
&w(s\wedge\tau, \lambda_{s\wedge\tau}, H_{s\wedge\tau}) \\
&\quad = w(t,\lambda,h) + \int_t^{s\wedge\tau} \left(\frac{\partial w}{\partial t} + \frac{\partial w}{\partial \lambda}\beta(\alpha - \lambda_u) + \frac{\partial w}{\partial h}(z_u - \rho H_u) \right) \mathrm{d}u \\
&\quad + \int_t^{s\wedge\tau} (w(\cdot + \xi) - w)\mathrm{d}N_u \\
&\quad = w(t,\lambda,h) + \int_t^{s\wedge\tau} \left(\frac{\partial w}{\partial t} + \frac{\partial w}{\partial \lambda}\beta(\alpha - \lambda_u) + \frac{\partial w}{\partial h}(z_u - \rho H_u) + (w(\cdot+\xi) - w)\lambda_u \right) \mathrm{d}u \\
&\quad + \int_t^{s\wedge\tau} (w(\cdot + \xi) - w)\mathrm{d}M_u,
\end{aligned}
$$

where $M_t = N_t - \int_0^t \lambda_u \mathrm{d}u$, i.e. the compensated jump measure. We then take a localizing sequence $(\tau_n)_n$, $\tau_n = \inf\{s \ge 0 : |H_s| \vee |\lambda_s| \ge n\} \wedge n$ and we write:

$$
\begin{aligned}
&\mathbb{E}[w(s\wedge\tau_n, \lambda_{s\wedge\tau_n}, H_{s\wedge\tau_n})] \\
&\quad = w(t,\lambda,h) + \mathbb{E}\left[\int_t^{s\wedge\tau_n} \left(\frac{\partial w}{\partial t} + \frac{\partial w}{\partial \lambda}\beta(\alpha - \lambda_u) - \rho\frac{\partial w}{\partial h}H_u + (w(\cdot+\xi) - w)\lambda_u \right) \mathrm{d}u \right] \\
&\quad + \mathbb{E}\left[\int_t^{s\wedge\tau_n} \frac{\partial w}{\partial h} z_u \mathrm{d}u \right] \\
&\quad \le w(t,\lambda,h) + \mathbb{E}\left[\int_t^{s\wedge\tau_n} \left(-\bar{\eta}(v - S(H_u,v))\lambda_u + z_u + \frac{\gamma}{2}(z_u)^2 \right) \mathrm{d}u \right]
\end{aligned}
$$

where the latter inequality follows by the assumption on w. We now take the limit $n \to +\infty$: note that $v - S(H_u, v) \le v$ and $\mathbb{E}\left[\int_0^T \lambda_u \mathrm{d}u\right] < \infty$. If $z \in \mathcal{Z}_t$, we can apply dominated convergence and take the limit inside the expectation. For the left-hand side, we use the linear growth of w

$$w(s\wedge\tau_n,\lambda_{s\wedge\tau_n},H_{s\wedge\tau_n})\leq C(1+\lambda_{s\wedge\tau_n}+H_{s\wedge\tau_n})\leq C(1+\sup_{s\in[t,T]}(\lambda_{s\wedge\tau_n}+H_{s\wedge\tau_n})).$$

By the integral expression of λ in equation (2.13), we can estimate

$$\sup_{s\in[t,T]}\lambda_{s\wedge\tau_n}\leq\alpha+\xi\sup_{s\in[t,T]}N_t\leq\alpha+\xi N_T,$$

since N is a counting process. By the properties of N, see Proposition 2.2, we can conclude

$$\mathbb{E}\Big[\sup_{s\in[t,T]}\lambda_{s\wedge\tau_n}\Big]\leq C(1+\mathbb{E}[N_T])<+\infty.$$

Similarly, for H, we can write $\sup_{s\in[t,T]}H_{s\wedge\tau_n}\leq h+\int_0^T z_v\mathrm{d}v$. It follows that $\mathbb{E}\Big[\sup_{s\in[t,T]}H_{s\wedge\tau_n}\Big]<+\infty$ since $\mathbb{E}\left[\int_0^T z_v\mathrm{d}v\right]<+\infty$ by the admissible hypothesis on $\mathcal{Z}_t$. We can apply dominated convergence and obtain

$$\lim_{n\to\infty}\mathbb{E}[w(s\wedge\tau_n,\lambda_{s\wedge\tau_n},H_{s\wedge\tau_n})]=\mathbb{E}[w(s,\lambda_s,H_s)].$$

Since w is continuous on its domain, by sending s to T and by the hypothesis $w(T,\lambda,h)\geq U(h)$, we obtain

$$w(t,\lambda,h)\geq\mathbb{E}\left[\int_t^T\left(\bar{\eta}(v-S(H_u,v)\lambda_u-z_u-\frac{\gamma}{2}(z_u)^2\right)\mathrm{d}u\right]+\mathbb{E}[U(H_T)].$$

For the arbitrariness of z, we deduce that $w(t,\lambda,h)\geq V(t,\lambda,h)$ for all $(t,\lambda,h)\in[0,T]\times(0,+\infty)\times\mathbb{R}_+$.

ii) We repeat the same reasoning as before, taking $z^*\in\mathcal{Z}$. By applying Ito's formula, we can write for $s\in[t,T)$ and for τ stopping time valued in $[t,T)$:

$$\begin{aligned}&w(s\wedge\tau,\lambda_{s\wedge\tau},H^*_{s\wedge\tau})\\&\quad=w(t,\lambda,h)+\int_t^{s\wedge\tau}\left(\frac{\partial w}{\partial t}+\frac{\partial w}{\partial\lambda}\beta(\alpha-\lambda_u)+\frac{\partial w}{\partial h}(z^*_u-\rho H^*_u)+(w(\cdot+\xi)-w)\lambda_u\right)\mathrm{d}u\\&\quad+\int_t^{s\wedge\tau}(w(\cdot+\xi)-w)\mathrm{d}M_u.\end{aligned}$$

We then take a localizing sequence $(\tau_n)_n$, $\tau_n=\inf\{s\geq0:|H^*_s|\vee|\lambda_s|\geq n\}\wedge n$ and we exploit the fact that w solves the PIDE. This leads to

$$\begin{aligned}&\mathbb{E}[w(s \wedge \tau_n, \lambda_{s\wedge\tau_n}, H^*_{s\wedge\tau_n})]\\&\quad = w(t,h,\lambda) + \mathbb{E}\left[\int_t^{s\wedge\tau_n} \left(-\bar{\eta}(v - S(H^*_u, v)\lambda_u + z^*_u + \frac{\gamma}{2}(z^*_u)^2\right) \mathrm{d}u\right].\end{aligned}$$

Applying dominated convergence, analogously to before, and taking $s \to T$, we write

$$\begin{aligned}w(T,\lambda,h) &= \mathbb{E}\left[\int_t^T \left(\bar{\eta}\,(v - S(H^*_u, v))\,\lambda_u - z^*_u - \frac{\gamma}{2}(z^*_u)^2\right) \mathrm{d}u + w(T, \lambda_T, H^*_T)\right]\\&= \mathbb{E}\left[\int_t^T \left(\bar{\eta}\,(v - S(H^*_u, v))\,\lambda_u - z^*_u - \frac{\gamma}{2}(z^*_u)^2\right) \mathrm{d}u + U(H^*_T)\right] = J(t,\lambda,h;z^*),\end{aligned}$$

where we use that $w(T,\lambda_T,H^*_T) = U(H^*_t)$. This shows that $V(t,\lambda,h) \geq J(t,\lambda,h;z^*) \geq w(t,\lambda,h) \geq V(t,\lambda,h)$, i.e. $w = V$ on $[0,T] \times (0,+\infty) \times \mathbb{R}_+$, and that z^* is an optimal Markovian control.

□

2.4 Numerical methods

In this section, we describe the parameters' choice and the numerical methods adopted for the solution of the optimization problem introduced in Section 2.3.

Specification of the model parameters We report in Tables 2.1, 2.2, 2.3 the *standard set* of the model parameters. Unless mentioned otherwise, the numerical analysis will be performed using the standard set of parameters.

function type	v	a	b
S_I	0.65	10^{-1}	1

Table 2.1: Specification of the security breach function.

α	β	ξ	λ_0
27	15	9	27

Table 2.2: Parameters of the stochastic intensity.

δ	γ	$\bar{\eta}$(k$)	$U(h)$	ρ	T
1	0.05	10	$\sqrt{h}$	0.2	1

Table 2.3: Parameters of the optimization problem.

We employ a security breach probability function S of class I, as defined in equation (2.1). The parameters values in Table 2.1 are consistent with those in Skeoch (2022), and also in line with previous works (see Gordon and Loeb (2002) and Mazzoccoli and Naldi (2020)). Taking v, a, b as in Table 2.1, a plot of the function $h \mapsto S_I(h, v)$ is shown in Figure 2.2.

The parameters of the stochastic intensity of the Hawkes process (see Table 2.2) are chosen to generate on average approximately 60 cyberattacks per year. We believe that this is a reasonable figure, in the absence of reliable estimates of the number of cyberattacks targeting a single entity.[4]

Remark 2.17. For the standard set of parameters, we have $\lambda_0 = \alpha$ and so the stochastic intensity λ_t can be expressed as follows:

$$\lambda_t = \lambda_0 + \xi \sum_{i=1}^{N_t} e^{-\beta(t-T_i)}. \tag{2.18}$$

We consider a one-year planning horizon ($T = 1$) and set an average loss of 10k$ for each successful breach, resulting in a total expected annual loss of approximately 390k$ without cybersecurity investments, which is in the same order of magnitude of Skeoch (2022). The depreciation rate is set at $\rho = 0.2$, consistently with the technological depreciation rates considered in Krutilla et al. (2021). The parameter γ is set at a rather low value, in order to avoid an excessive penalization of large investment rates. Finally, we choose $U(h) = \sqrt{h}$, representing a strictly increasing and concave CRRA utility function.

Numerical solution of the HJB-PIDE As shown in Section 2.3, determining the optimal cybersecurity investment requires the solution of the non-linear PIDE (2.16). Due to the complexity of the problem, one cannot expect to find explicitly an analytical solution and, hence, numerical methods are required. We opt for the *method of lines*, as described in Yuan (1999). This technique consists in discretizing the PIDE in the spatial domain $(\lambda, h) \in (0, \infty) \times \mathbb{R}_+$ but not in time, and then in integrating the semi-discrete problem as a system of ODEs. In our setting, we discretize the (λ, h) dimensions with a central

[4]Empirical estimates of the intensity of cyberattacks can be found in the recent works Bessy-Roland et al. (2021), Boumezoued et al. (2023), Li and Mamon (2023)). However, these estimates are not suitable for our purposes, since they are based on the number of attacks at a worldwide scale, while our model takes the viewpoint of a single entity.

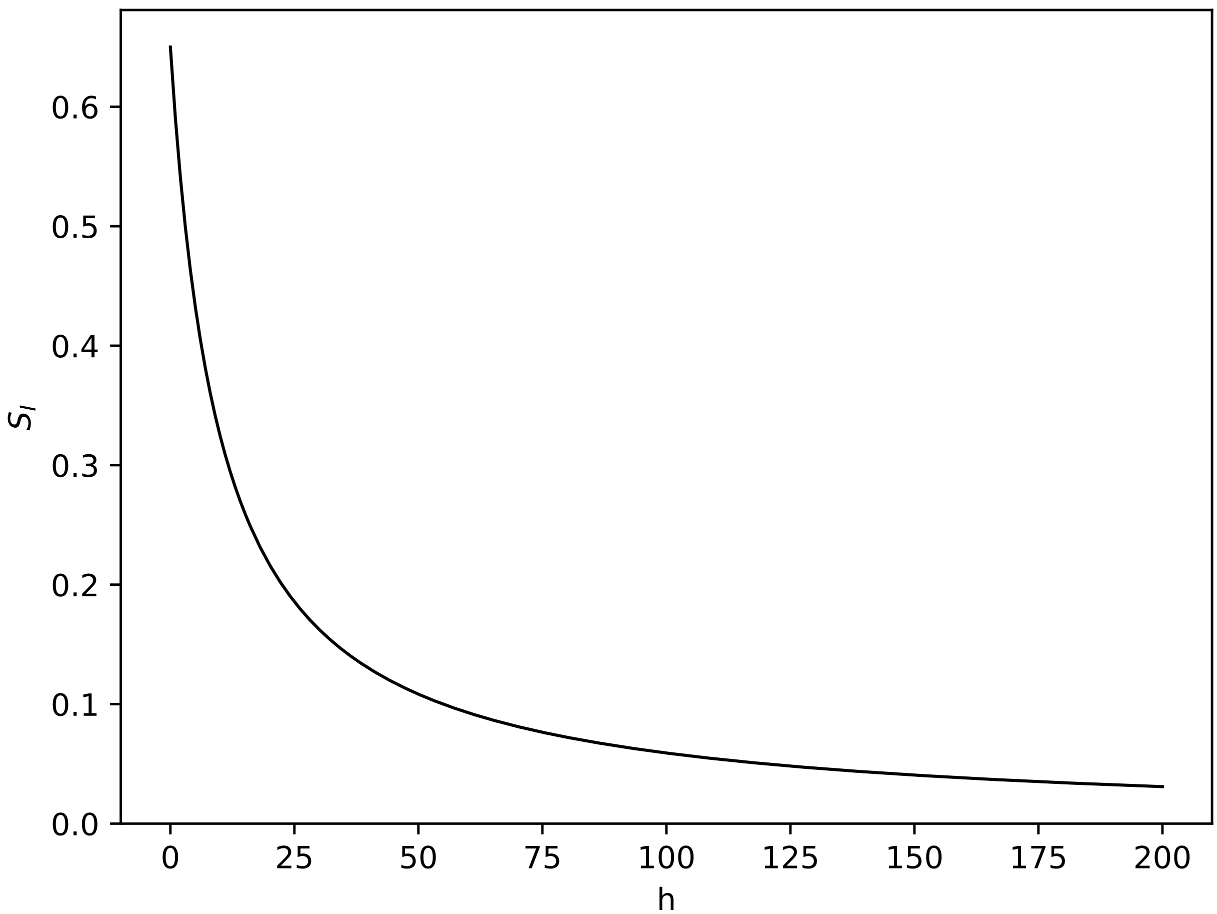


Figure 2.2: Security breach function (parameters as in Table 2.1).

difference and then numerically solve the resulting ODE system. Similarly to the case of PIDEs arising in Lévy models (see, e.g., Cont and Voltchkova (2005)), the unbounded space domain $(0, \infty) \times \mathbb{R}_+$ is localized into a bounded domain $[\lambda_{\min}, \lambda_{\max}] \times [h_{\min}, h_{\max}]$. We refer to Algorithm 1 for a precise description of the implementation of this method.

In our implementation, we specify as follows the algorithm's meta-parameters:

$\lambda_{\min}$	$\lambda_{\max}$	$\Delta\lambda$	$h_{\min}$	$h_{\max}$	Δh
27	216	1	0	50	0.5

Table 2.4: Meta-parameters for Algorithm 1.

The value $h_{\min} = 0$ corresponds to the absence of cybersecurity investment, while $h_{\max} = 50$ represents an upper bound which is rarely achieved in our setup under the standard parameter set. We choose $\lambda_{\min} = \lambda_0$, which coincides with the lower bound of the stochastic intensity λ_t, see equation (2.18). We set $\lambda_{\max} = \mathbb{E}[\lambda_T] + 7\sqrt{\mathrm{Var}[\lambda_T]} \approx 216$, in order to ensure that the truncation of the intensity domain does not have any material impact on our numerical results. The value function V is extrapolated beyond $[\lambda_{\min}, \lambda_{\max}]$ by setting

$$V(t, \lambda, h) = V(t, \lambda_{\max}, h), \qquad \text{for all } \lambda > \lambda_{\max},$$

analogously to the scheme implemented in Gaïgi et al. (2025, Section 5.1). When plotting the function V in Section 2.5, we shall consider a subinterval of $[\lambda_{\min}, \lambda_{\max}]$: intensity values close to $\lambda_{\max}$ are rarely achieved and might lead to numerical instabilities.

We have implemented Algorithm 1 in Python, using the built-in ODE solver `scipy.integrate.solve\_ivp`. We make use of an implicit Runge-Kutta method of the Radau IIA family of order 5 (see Hairer et al. (1993) for further details).

Optimal cybersecurity investment rate Besides determining the optimal net benefit of cybersecurity investments, we aim at computing the real-time adaptive strategy that best responds to the arrival of cyberattacks. To this end, after solving the PIDE (2.16) via Algorithm 1, we compute numerically the optimal investment rate given in equation (2.17) along a simulated sequence of cyberattacks. This entails simulating a trajectory of the stochastic intensity $(\lambda_t(\omega))_{t\in[t_{\text{init}},T]}$, starting from an initial cybersecurity level H_{init} at time t_{init}. Our numerical method for the computation of the optimal investment rate is described in Algorithm 2 and will be numerically implemented in Sections 2.5.4 and 2.5.4.

Algorithm 1 Numerical solution of the PIDE (2.16)

1: Set $\lambda_{\min}, \lambda_{\max}, h_{\min}, h_{\max}$.
2: Discretize $[\lambda_{\min}, \lambda_{\max}]$, with $\lambda_0 = \lambda_{\min}, \lambda_N = \lambda_{\max}$ and $\lambda_n - \lambda_{n-1} = \Delta\lambda$, for $n = 1, \dots, N$.
3: Discretize $[h_{\min}, h_{\max}]$, with $h_0 = h_{\min}, h_M = h_{\max}$ and $h_m - h_{m-1} = \Delta h$, for $m = 1, \dots, M$.
4: Set $V_{n,m}(t) := V(t, \lambda_n, h_m)$, for all n and m.
5: Approximate the partial derivatives w.r.t. λ:

$$\frac{\partial V}{\partial \lambda}(t, \lambda_n, h_m) \approx \frac{V_{n+1,m}(t) - V_{n-1,m}(t)}{2\Delta\lambda},$$
$$\frac{\partial V}{\partial \lambda}(t, \lambda_0, h_m) \approx \frac{V_{1,m}(t) - V_{0,m}(t)}{\Delta\lambda},$$
$$\frac{\partial V}{\partial \lambda}(t, \lambda_N, h_m) \approx \frac{V_{N,m}(t) - V_{N-1,m}(t)}{\Delta\lambda}.$$

6: Approximate the partial derivatives w.r.t. h:

$$\frac{\partial V}{\partial h}(t, \lambda_n, h_m) \approx \frac{V_{n,m+1}(t) - V_{n,m-1}(t)}{2\Delta h},$$
$$\frac{\partial V}{\partial h}(t, \lambda_n, h_0) \approx \frac{V_{n,1}(t) - V_{n,0}(t)}{\Delta h},$$
$$\frac{\partial V}{\partial h}(t, \lambda_n, h_M) \approx \frac{V_{n,M}(t) - V_{n,M-1}(t)}{\Delta h}.$$

7: Let $\tilde{n} = \frac{\lfloor \xi \rfloor}{\Delta\lambda}$ and set

$$V(t, \lambda_n + \xi, h_m) \approx V_{(n+\tilde{n})\wedge N,m}(t).$$

8: Solve the ODE system given for all n, m by

$$V'_{n,m}(t) = \beta(\lambda_n - \alpha)\frac{V_{n+1,m}(t) - V_{n-1,m}(t)}{2\Delta\lambda} + \rho h \frac{V_{n,m+1}(t) - V_{n,m-1}(t)}{2\Delta h}$$
$$- \lambda_n(V_{n+\tilde{n}\wedge N,m}(t) - V_{n,m}(t)) - \bar{\eta}(v - S(h_m, v))\lambda_n - \frac{\left(\left(\frac{V_{n,m+1}(t) - V_{n,m-1}(t)}{2\Delta h} - 1\right)^+\right)^2}{2\gamma},$$
$$V_{n,m}(T) = U(h_m).$$

Algorithm 2 Numerical computation of the optimal control

Set $t_{\min}, t_{\max}, \lambda_{\min}, \lambda_{\max}, h_{\min}, h_{\max}$.
Discretize $[t_{\min}, t_{\max}]$, with $t_0 = t_{\min}, t_I = t_{\max}$ and $t_i - t_{i-1} = \Delta t$, for $i = 1, \dots, I$.
Discretize $[\lambda_{\min}, \lambda_{\max}]$, with $\lambda_0 = \lambda_{\min}, \lambda_N = \lambda_{\max}$ and $\lambda_n - \lambda_{n-1} = \Delta\lambda$, for $n = 1, \dots, N$.
Discretize $[h_{\min}, h_{\max}]$, with $h_0 = h_{\min}, h_M = h_{\max}$ and $h_m - h_{m-1} = \Delta h$, for $m = 1, \dots, M$.
Compute $V(t_i, \lambda_n, h_m)$ and $z^*(t_i, \lambda_n, h_m)$, for $i = 0, \dots, I$, $n = 0, \dots N$, $m = 0, \dots, M$.
Simulate a trajectory $\lambda_{t_i}(\omega)$, $i = 0, \dots, I$.
For the initial time $t_{\text{init}} \geq t_{\min}$, set $\bar{i} := \text{argmin}_i\{|t_i - t_{\text{init}}|\}$.
Consider the initial state $H_{t_{\bar{i}}} = H_{\text{init}}$:
for i in $\bar{i}, \dots, I$, **do**
    set $k := \text{argmin}_\ell\{|\lambda_\ell - \lambda_{t_i}(\omega)|\}$;
    set $j := \text{argmin}_m\{|h_m - H_{t_i}|\}$;
    let $z^*_{t_i} = z^*(t_i, \lambda_k, h_j)$;
    $H^{z^*}_{t_{i+1}} := H^{z^*}_{t_i} - \rho H^{z^*}_{t_i}\Delta t + z^*_{t_i}\Delta t$.
end for

2.5 Results and discussion

In this section, we report some numerical results that illustrate the key properties and implications of the model. In particular, we are interested in assessing the benefit of adopting the optimal dynamic cybersecurity investment policy.

2.5.1 Value function and optimal cybersecurity policy

Figure 2.3 displays the value function V and the optimal cybersecurity investment rate z^*. In panels 2.3a and 2.3b we plot, respectively, V and z^* for fixed intensity $\lambda = 27$, varying t and h. Coherently with Remark 2.10, we observe that the value function is increasing in h, while the optimal investment rate is decreasing. This behavior reflects the fact that higher cybersecurity levels yield greater benefits and reduce the need for further cybersecurity investments. In panels 2.3c and 2.3d we plot, respectively, V and z^* for fixed $h = 0$, varying t and λ. Coherently with Remark 2.10, we observe that both the value function and the optimal investment rate are increasing in λ. This is explained by the fact that, in the presence of a higher risk of cyberattacks, investing in cybersecurity becomes more valuable due to the larger potential of mitigating expected losses. As can be seen from panels 2.3e and 2.3f, both the value function and the optimal investment rate decrease over time. This is due to the fact that, under the standard parameter configuration (see Section 2.4), the residual utility $U(H_T)$ of cybersecurity plays a relatively minor role and, therefore, the value of additional cybersecurity investment declines as the end of the planning horizon $[0, T]$ approaches.

2.5.2 Parameter sensitivity

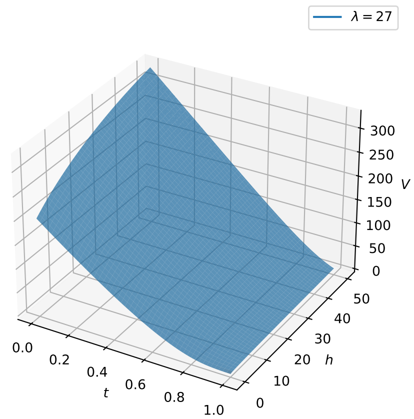
In the proposed model, the parameter β determines the clustering behavior of cyberattacks. Specifically, higher values of β correspond to a more rapid decay of the intensity following each attack, thereby reducing the likelihood of temporally clustered attacks. To evaluate the impact of clustered cyberattacks, we compare in Figure 2.4 the value function and the optimal investment rate under two scenarios: $\beta = 15$ (more clustered attacks) and $\beta = 50$ (less clustered attacks). We observe that both the value function and the optimal investment rate are substantially greater in the case $\beta = 15$: if cyberattacks occur in clustered patterns, it is optimal to invest more in cybersecurity in order to mitigate the risk of large cumulative losses arising from rapid attack sequences. This finding underscores the critical importance of accounting for clustering dynamics in the optimal management of cyber-risk.

We also analyze the role of obsolescence in cybersecurity investment decisions, motivated by the analysis in Krutilla et al. (2021), which highlights its significance in a dynamic setup. Figure 2.5 displays the value function and the optimal investment rate under two contrasting depreciation scenarios: $\rho = 0$ (no obsolescence) and $\rho = 1$ (high obsolescence).[5] We observe that $\rho = 1$ leads to smaller values for V and z^*, in line with the findings of Krutilla et al. (2021) in a deterministic setup. Our results confirm that a rapid depreciation of cybersecurity effectiveness reduces both the expected net benefit and the incentive to invest.

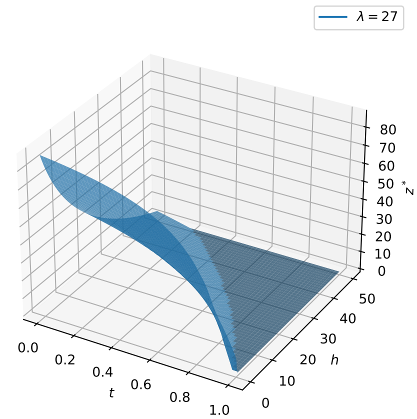
2.5.3 Comparison with a static investment strategy

The optimal investment rate z^* characterized in Theorem 2.14 represents the real-time adaptive cybersecurity policy that best responds to the arrival of cyberattacks. In order to assess whether the adoption of an adaptive dynamic strategy provides a tangible benefit, we compare it against the best constant investment strategy, i.e., the strategy $z_t = \bar{z}$, for all $t \in [0, T]$, that maximizes the benefit-cost trade-off functional J. When investing according

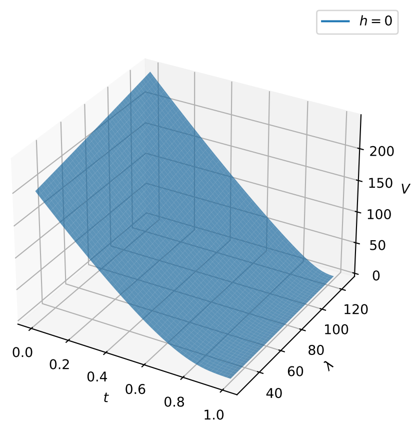
[5] In practical terms, a depreciation rate of $\rho = 1$ implies that a given initial cybersecurity level H_0 depreciates by over 73% over a one-year period.



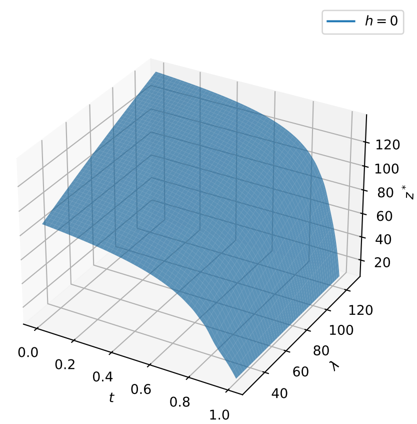
(a) Value function $V(t, \lambda, h)$ for $\lambda = 27$.



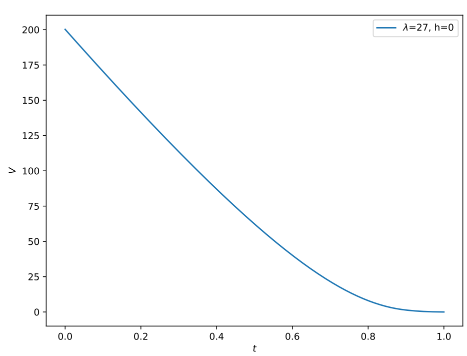
(b) Optimal control $z_t^*(\lambda, h)$ for $\lambda = 27$.



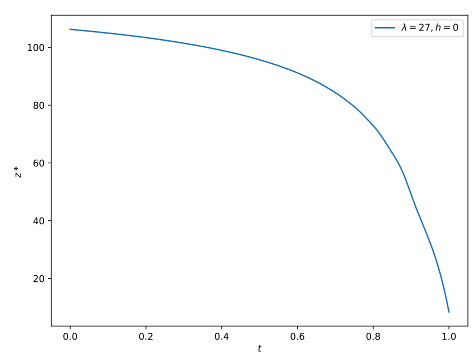
(c) Value function $V(t, \lambda, h)$ for $h = 0$.



(d) Optimal control $z_t^*(\lambda, h)$ for $h = 0$.

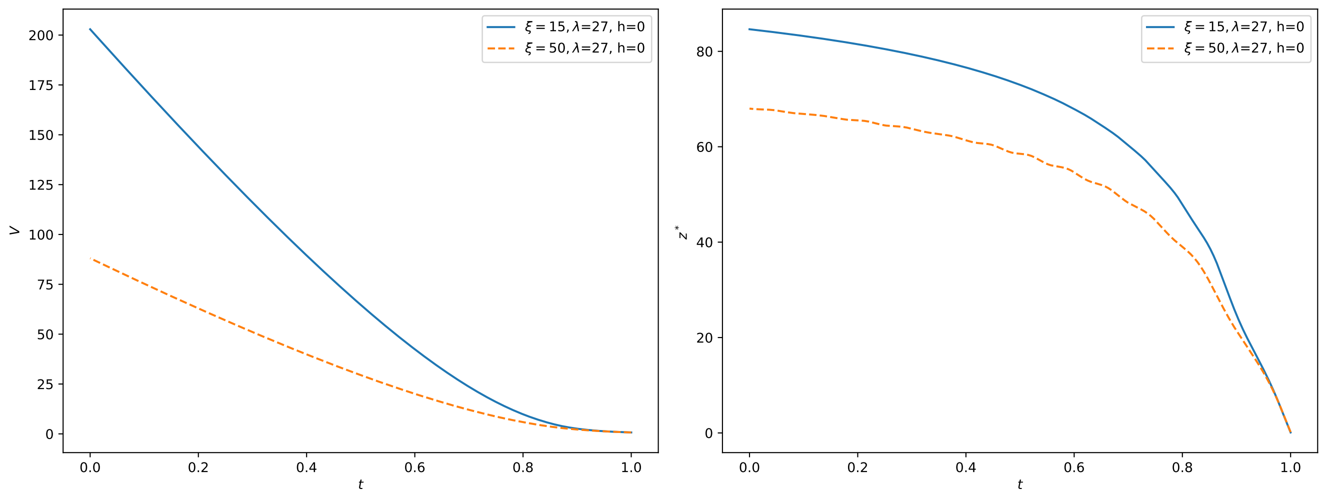


(e) Value function $V(t, \lambda, h)$ for $\lambda = 27, h = 0$.



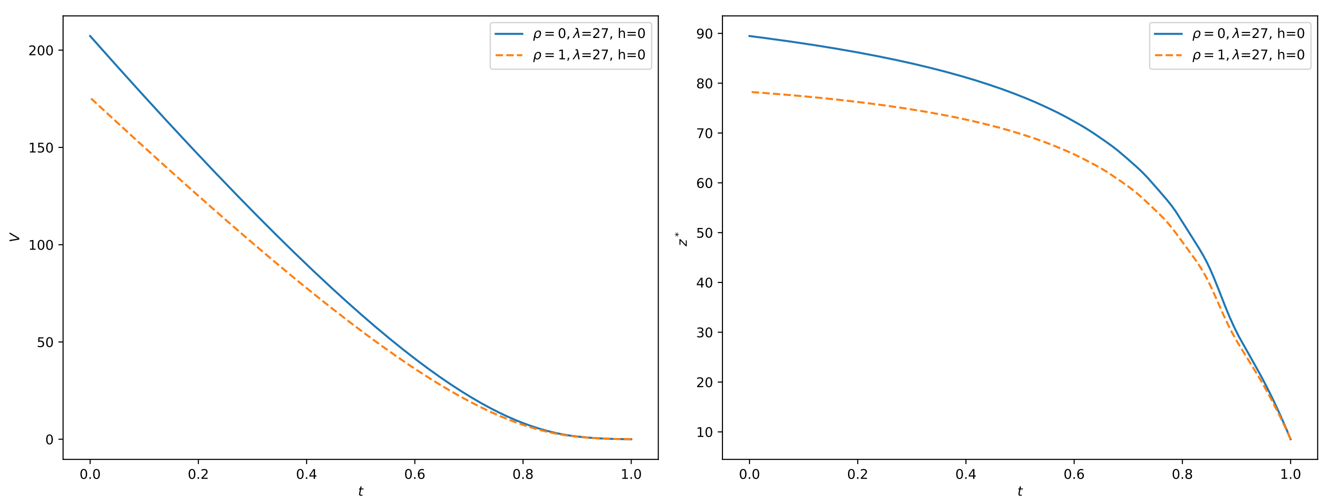
(f) Optimal control $z_t^*(\lambda, h)$ for $\lambda = 27, h = 0$.

Figure 2.3: Value function and optimal investment rate computed under the standard parameters set.



(a) Value function for different β, $\lambda = 27, h = 0$. (b) Optimal control for different β, $\lambda = 27, h = 0$.

Figure 2.4: Value function and optimal investment rate for $\beta = 15$ and $\beta = 50$, for fixed h and λ.



(a) Value function $V(t, \lambda, h)$ for $\lambda = 27, h = 0$. (b) Optimal control $z_t^*(\lambda, h)$ for $\lambda = 27, h = 0$.

Figure 2.5: Value function and optimal investment rate for $\rho = 0$ and $\rho = 1$, for fixed h and λ.

to a costant rate $\bar{z}$, the benefit-cost trade-off functional J takes the following form:

$$J(t,\lambda,h;\bar{z}) = \int_t^T \bar{\eta}\big(v - S(H_s^{t,h,\bar{z}}, v)\big)\mathbb{E}[\lambda_s^{t,\lambda}]\mathrm{d}s - (T-t)(\bar{z} + \frac{\gamma}{2}\bar{z}^2) + U(H_T^{t,h,\bar{z}}), \tag{2.19}$$

where the cybersecurity level $H^{t,h,\bar{z}}$ is given by

$$H_s^{t,h,\bar{z}} = he^{-\rho(s-t)} + \frac{\bar{z}}{\rho}(1 - e^{-\rho(s-t)}).$$

The optimal constant investment rate $\bar{z}^*$ solves the problem

$$J(t,\lambda,h;\bar{z}^*) := \sup_{\bar{z}\in\mathbb{R}_+} J(t,\lambda,h;\bar{z}). \tag{2.20}$$

In view of Proposition 2.2, the expectation $\mathbb{E}[\lambda_s^{t,\lambda}]$ in (2.19) can be computed in closed form. Therefore, the optimization problem (2.20) reduces to a deterministic maximization with respect to a scalar variable, which can be easily solved numerically. To this effect, we adopt the built-in global scalar optimizer scipy.optimize.differential_evolution in Python.

We quantify the relative gain obtained by investing according to the optimal dynamic policy z^* versus the constant policy $\bar{z}^*$ by computing the following quantity:

$$\%\text{gain}(t,\lambda,h) := 100 \times \frac{V(t,\lambda,h) - J(t,\lambda,h;\bar{z}^*)}{J(t,\lambda,h;\bar{z}^*)}. \tag{2.21}$$

Figure 2.6 displays the relative gain over time for varying cybersecurity levels h, for $\lambda = 27$ fixed. At the initial time $t = 0$, the gain reaches 15% for $h = 0.5$, 14% for $h = 1$, 12% for $h = 2$, while it is 9.04% for $h = 5$, 5.7% for $h = 10$ and 2.6% for $h = 20$. These results show that the optimal dynamic investment strategy z^* consistently outperforms the best constant strategy $\bar{z}^*$, underscoring the importance of adaptive and responsive cybersecurity investments. The fact that the gain is rather small for large initial cybersecurity levels is coherent with the findings in Section 2.5.1: when the initial cybersecurity level is already high, the benefit of further investments diminishes, thereby reducing the relative advantage of the optimal policy. Moreover, a further analysis shows that the gain increases monotonically with respect to λ, indicating that the advantage of adopting the dynamic optimal policy (2.17) becomes more pronounced in high-risk scenarios.

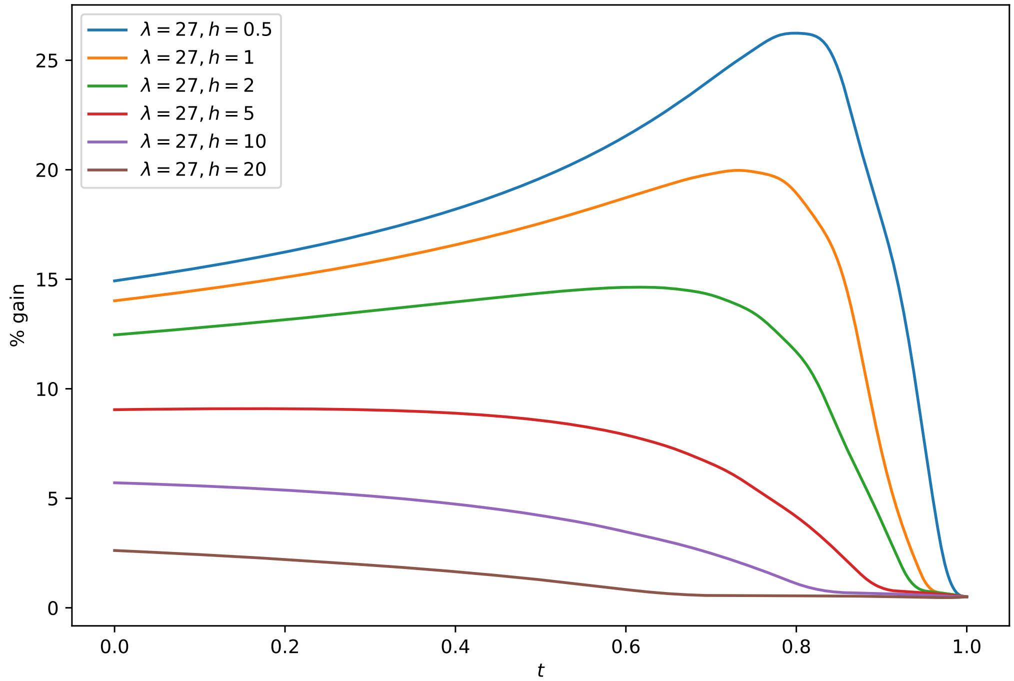


Figure 2.6: Relative gain with respect to the optimal constant investment rate.

2.5.4 Comparison with a standard Poisson model

To further assess the impact of clustered cyberattacks, we compare our model, which features a self-exciting Hawkes process, with a simplified version based on a standard Poisson process. A Poisson process $P = (P_t)_{t\in[0,T]}$ is characterized by a constant intensity λ^P and does not capture any temporal dependence in the arrival of attacks. Indeed, conditionally on $P_T = n$, the attack times are distributed as the order statistics of n i.i.d. random variables uniformly distributed on $[0,T]$, for every $n \in \mathbb{N}$. This setup can be recovered as a special case of the model introduced in Section 2.2.2 by setting $\xi = 0$ and $\alpha = \lambda_0$ in the intensity dynamics (2.3).

We consider the same optimization problem as in Section 2.3 and we replace the Hawkes process N with a Poisson process P of constant intensity λ^P and denote the resulting optimal investment rate by z^{P*}. A key observation is that, in this case, problem (2.11) reduces to a deterministic optimal control problem. The associated value function, $V^P(t,h)$, solves the following PDE:

$$\begin{aligned}&\frac{\partial V^P}{\partial t} - \rho h \frac{\partial V^P}{\partial h} + \lambda^P \bar{\eta}\,(v - S(h,v)) + \frac{\left((\frac{\partial V^P}{\partial h} - 1)^+\right)^2}{\gamma} = 0,\\ &V^P(T,h) = U(h).\end{aligned} \tag{2.22}$$

This PDE can be numerically solved using a scheme similar to Algorithm 1. Analogously to Theorem 2.14, the optimal investment rate z^{P*} in the Poisson model is given by

$$z^{P*} = \frac{\left(\frac{\partial V^P}{\partial h} - 1\right)^+}{\gamma}.$$

Remark 2.18. The optimal policy z^{P*} is deterministic. This is due to the fact that, in the Poisson model, the occurrence of a cyberattack does not carry any informational content.

We compare the Hawkes-based model with two Poisson-based benchmarks:

(i) a Poisson model with intensity λ_b^P chosen as

$$\lambda_b^P = \lambda_0 = 27; \tag{2.23}$$

(ii) a Poisson model with intensity λ_e^P chosen as

$$\lambda_e^P = \frac{\lambda_0 \beta}{\beta - \xi} + \frac{1 - e^{-\beta T}}{T(\beta - \xi)} \left(\lambda_0 - \frac{\lambda_0 \beta}{\beta - \xi} \right) \approx 61. \tag{2.24}$$

The first case corresponds to a Poisson process with the same baseline intensity of the Hawkes process. This scenario can be thought of as the situation where the entity underestimates the likelihood of cyberattacks (possibly due to relying on a limited or unrepresentative dataset) and considers it to be constant over time. In the second case, in view of Proposition 2.5, the value λ_e^P is chosen so that $\mathbb{E}[P_T] = \mathbb{E}[N_T]$, ensuring that the Hawkes-based model and the Poisson model with intensity λ_e^P generate the same expected number of cyberattacks over the planning horizon $[0, T]$. This reflects a case where the average attack frequency is estimated correctly, but the clustering dynamics are ignored.

We shall make use of the following notation:

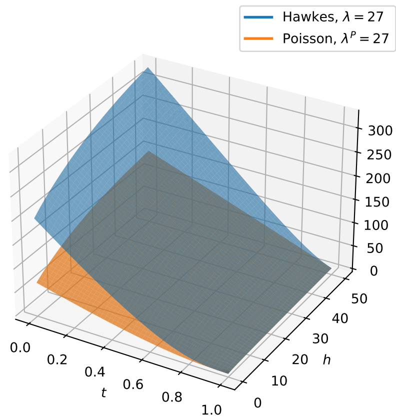
- $V_b^P(t, h)$ is the value function associated to the PDE (2.22) for the Poisson model with intensity λ_b^P specified in (2.23) and $z_t^{P*,b}(h)$ is the associated optimal control;
- $V_e^P(t, h)$ is the value function associated to the PDE (2.22) for the Poisson model with intensity λ_e^P specified in (2.24) and $z_t^{P*,e}(h)$ is the associated optimal control;
- $V(t, \lambda, h)$ is the value function associated to the PIDE (2.16) and $z_t^*(\lambda, h)$ is the associated optimal control.

Value functions and optimal cybersecurity policies Figure 2.7 displays the results of the comparison with the Poisson model (i) with intensity λ_b^P. We observe that both the value function and the optimal cybersecurity investment rate under the Hawkes-based model consistently dominate their counterparts in the Poisson model (i) across the entire planning horizon. This is a direct consequence of the fact that $\lambda_t \geq \lambda_b^P$, for all $t \in [0, T]$. In other words, the Poisson model (i) not only disregards the temporal clustering of cyberattacks, but also systematically underestimates their frequency. As a result, the perceived benefit of cybersecurity investment is lower, leading in turn to a suboptimal investment strategy.

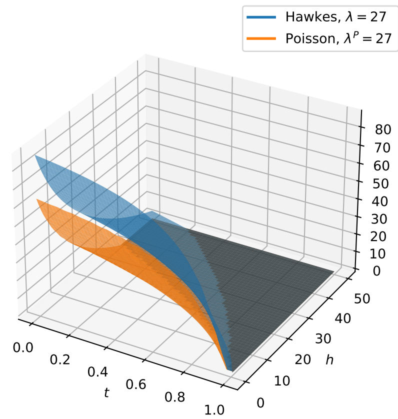
Figure 2.8 reports the comparison with the Poisson model (ii) with intensity λ_e^P. Panels 2.8a and 2.8b show that the benefit of cybersecurity investment and the optimal investment rate are slightly greater in the presence of clustered attacks (Hawkes-based model). This finding is confirmed in Panels 2.8e and 2.8f, which compare the value functions and optimal investment rates for fixed values $\lambda = \lambda_e^P$ and $h = 0$. Further insight is provided by panels 2.8c and 2.8d, which display respectively the value functions and the optimal investment rates at the initial time $t = 0$, across varying intensity levels. We can observe that the difference between the Hawkes and the Poisson models is negligible for small values of λ, while it becomes increasingly pronounced at higher values of λ. Interestingly, panel 2.8d shows that the optimal investment rate under the Hawkes model may be either higher or lower than that in the Poisson model, depending on whether the current intensity λ exceeds λ_e^P or not. This feature will be analyzed in more detail in Section 2.5.4 below. Overall, these findings indicate that even when the average attack intensity is correctly estimated, neglecting the temporal clustering of cyberattacks can lead to suboptimal cybersecurity investment decisions.

Relative gain Proceeding similarly to Section 2.5.3, we now evaluate the additional benefit derived from implementing the optimal adaptive policy z^*, as defined in (2.17), relative to the dynamic but deterministic policy z^{P*} derived in a Poisson-based model. We assume that the underlying model is the one introduced in Section 2.2.2 and compute the value function V given in (2.14) via Algorithm 1, using the standard parameter set described in Section 2.4. When employing the deterministic strategy z^{P*}, the expected net benefit from cybersecurity investment is quantified as follows:

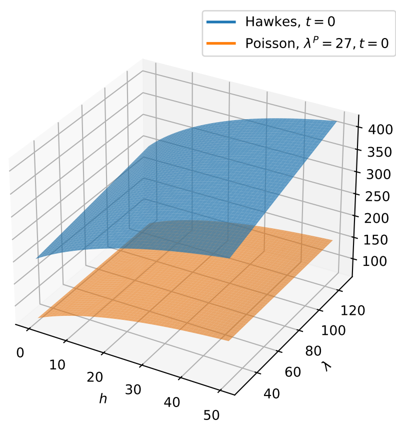
$$
\begin{aligned}
J(t, \lambda, h; z^{P*}) = &\int_t^T \left(\bar{\eta}(v - S(H_s^{t,h,z^{P*}}, v)) \mathbb{E}[\lambda_s^{t,\lambda}] \right) \mathrm{d}s \\
&- \int_t^T \left(z_s^{P*} + \frac{\gamma}{2} (z_s^{P*})^2 \right) \mathrm{d}s + U\left(H_T^{t,h,z^{P*}}\right),
\end{aligned}
$$



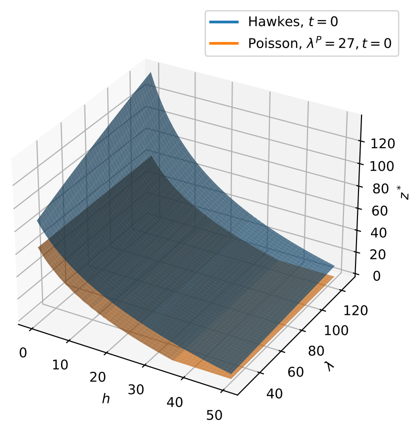
(a) Value functions $V(t, \lambda_b^P, h)$ and $V_b^P(t, h)$.



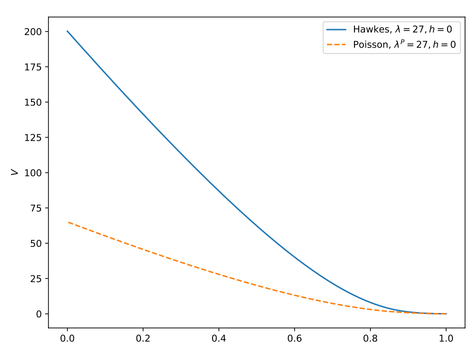
(b) Optimal controls $z_t^*(\lambda_b^P, h)$ and $z_t^{P*,b}(h)$.



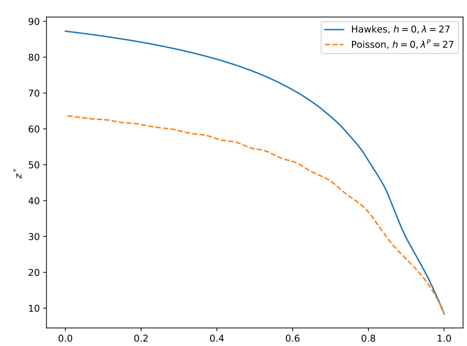
(c) Value functions $V(0, \lambda, h)$ and $V_b^P(0, h)$.



(d) Optimal controls $z_0^*(\lambda_b^P, h)$ and $z_0^{P*,b}(h)$.

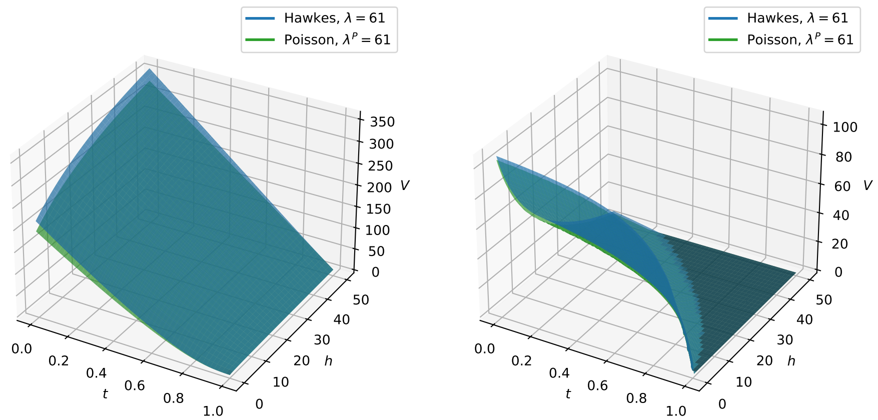


(e) Value functions $V(t, \lambda_b^P, 0)$ and $V_b^P(t, 0)$.



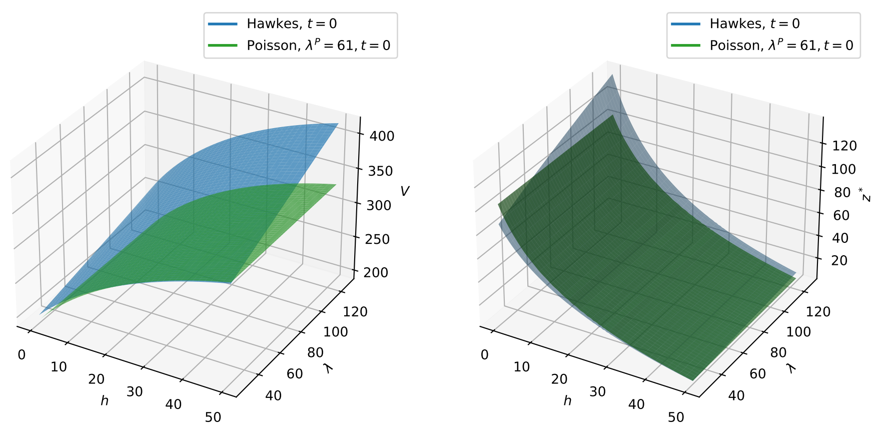
(f) Optimal controls $z_t^*(\lambda_b^P, 0)$ and $z_t^{P*,b}(0)$.

Figure 2.7: Comparison with a Poisson model with constant intensity $\lambda_b^P = 27$.



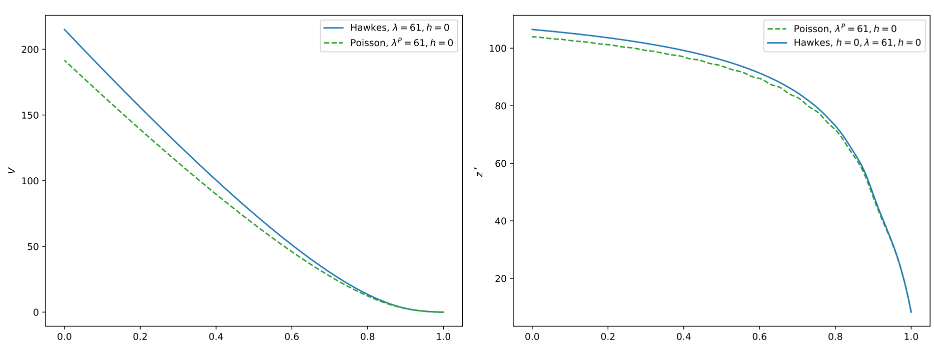
(a) Value functions $V(t, \lambda_e^P, h)$ and $V_e^P(t, h)$.

(b) Optimal controls $z_t^*(\lambda_e^P, h)$ and $z_t^{P^*,e}(h)$.



(c) Value functions $V(0, \lambda, h)$ and $V_e^P(0, h)$.

(d) Optimal controls $z_0^*(\lambda, h)$ and $z_0^{P^*,e}(h)$.



(e) Value functions $V(t, \lambda_e^P, 0)$ and $V_e^P(t, 0)$.

(f) Optimal controls $z_t^*(\lambda_e^P, 0)$ and and $z_t^{P^*,e}(0)$.

Figure 2.8: Comparison with a Poisson model with constant intensity $\lambda_e^P = 61$.

where $H_s^{t,h,z^{P*}}$ is defined as in (2.12) with $z = z^{P*}$ and $\mathbb{E}[\lambda_s^{t,\lambda}]$ can be computed explicitly by Proposition 2.2. Once the PDE (2.22) is numerically solved, z^{P*} can be computed via Algorithm 2 taking λ constant. Our numerical implementation of the Poisson-based model adopts the following specification:

$h_{\min}$	$h_{\max}$	Δh	$\lambda_{\min}$	$\lambda_{\max}$	$\Delta\lambda$	t_{init}	H_{init}	$\lambda_t(\omega)$
0	50	0.5	27	216	1	t	h	λ^P

Table 2.5: Meta-parameters for Algorithm 2.

The gain of the optimal investment policy z^* with respect to z^{P*} is computed as follows, in analogy to (2.21):

$$\%\text{gain}^P(t,\lambda,h) := 100 \times \frac{V(t,\lambda,h) - J(t,\lambda,h;z^{P*})}{J(t,\lambda,h;z^{P*})}. \tag{2.25}$$

Figure 2.9 reports the quantity $\%\text{gain}^P(t,\lambda,h)$ comparing the Hawkes-based model against two Poisson-based benchmarks with intensities λ_b^P and λ_e^P, as considered above. For $h = 0$, the gain increases with λ, ranging between 7.6% and 11.4% for λ_b^P, and between 0.04% and 0.6% for λ_e^P. For $h = 20$, the gain becomes nearly constant in λ. The fact that the gain for λ_e^P is limited can be explained by the fact that the objective functional (2.8) is linear with respect to the losses and λ_e^P is chosen in such a way that the Poisson-based model generates the same expected losses of our Hawkes-based model.

Adaptive dynamics of the optimal investment policy Finally, we illustrate the adaptive behavior of the optimal investment policy given in equation (2.17). While the overall improvement over a Poisson-based strategy may appear limited in terms of overall gain (see Figure 2.9), the key strength of our approach lies in its capacity to dynamically adjust the cybersecurity investment in response to the arrival of cyberattacks. To this effect, panels 2.10a and 2.11a display two simulated paths of the Hawkes intensity $(\lambda_t)_{t\in[0,T]}$, alongside the constant intensities λ_b^P and λ_e^P defined in Section 2.5.4. The corresponding optimal investment policies are shown in panels 2.10b and 2.11b. Consistent with the analysis in Section 2.5.4, the optimal investment rate z_t^* is always larger than the Poisson-based benchmark z_b^{P*}, due to the fact that $\lambda_t \geq \lambda_b^P$, for all $t \in [0,T]$. In contrast, the comparison with the benchmark strategy z_e^{P*} is more nuanced. We highlight in cyan the time intervals during which $\lambda_t > \lambda_e^P$. Our simulations reveal that when $\lambda_t \leq \lambda_e^P$, the adaptive strategy z_t^* closely aligns with z_e^{P*}. However, when $\lambda_t > \lambda_e^P$, especially during extended periods resulting from clusters of cyberattacks, the investment rate z_t^*

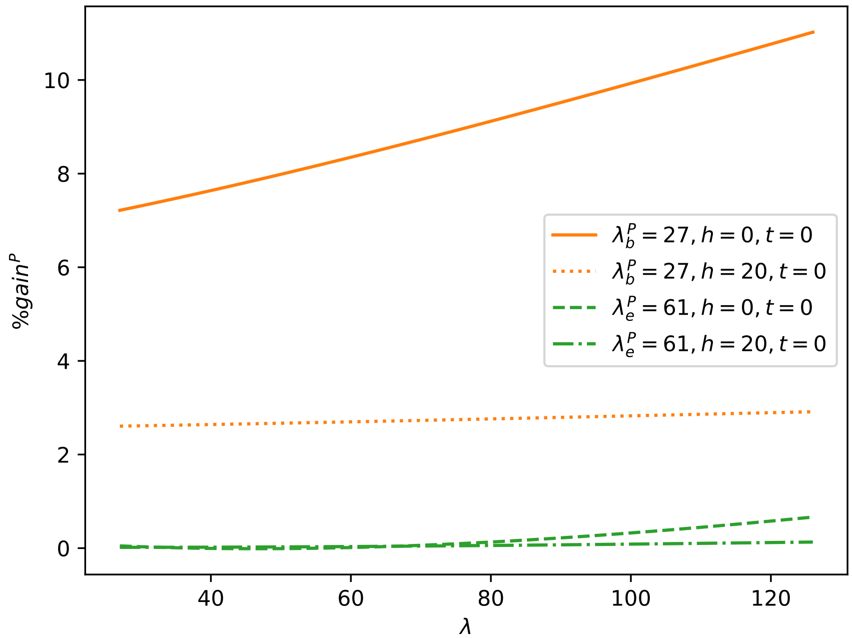


Figure 2.9: Relative gain with respect to the Poisson deterministic strategy, as defined in (2.25).

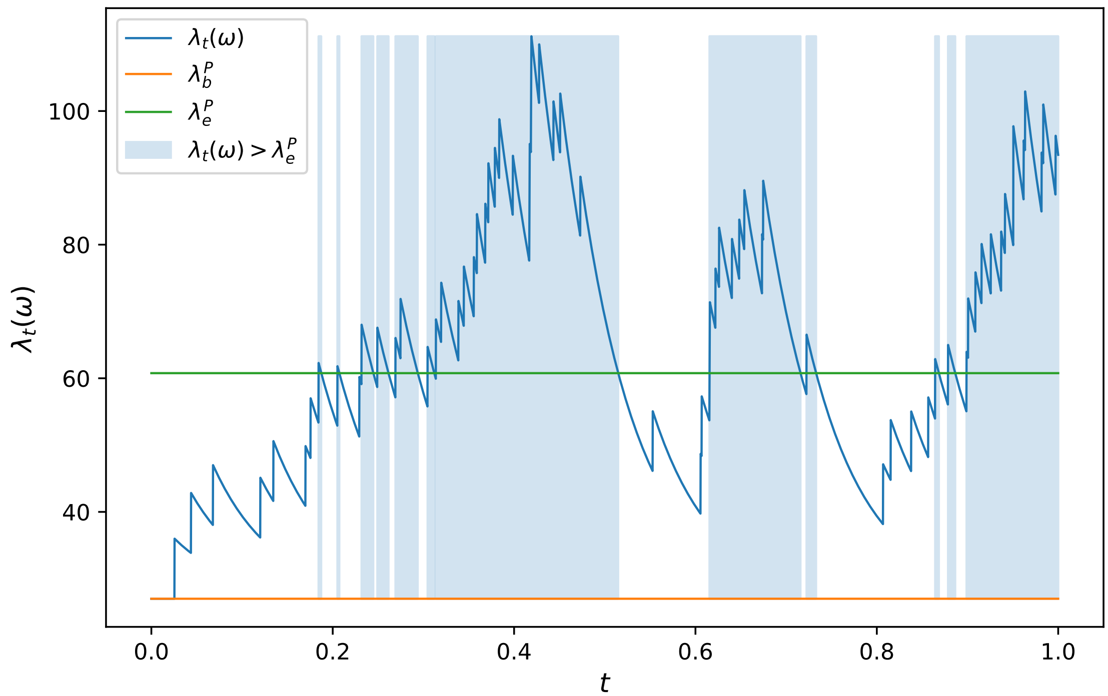
increases markedly, exceeding the corresponding deterministic strategy. This shows that the optimal cybersecurity investment policy z_t^* can react in real-time to rapid sequences of cyberattacks. Finally, under the standard parameter set, the investment rate naturally declines toward the end of the planning horizon $[0, T]$, as the accumulated cybersecurity level suffices to mitigate future risk.

2.6 Conclusions

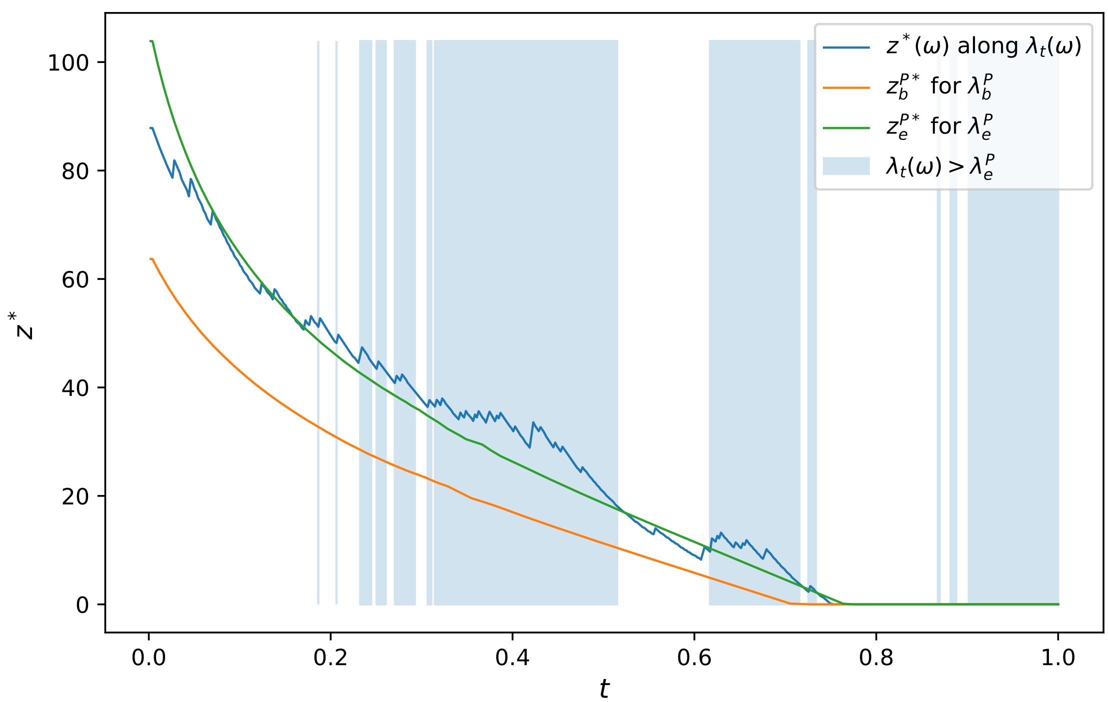
In this work, we introduce a dynamic and stochastic extension of the Gordon–Loeb model Gordon and Loeb (2002) for optimal cybersecurity investment, incorporating temporally clustered cyberattacks via a Hawkes process. Our modeling framework captures the empirically observed phenomenon of attack bursts, thus offering a more realistic representation of the current cyber-risk environment. We formulate the cybersecurity investment decision problem as a two-dimensional stochastic optimal control problem, maximizing the expected net benefit of cybersecurity investments. We allow for adaptive investment policies that respond in real-time to the arrival of cyberattacks.

Our numerical results demonstrate that the optimal cybersecurity investment policy consistently outperforms both static benchmarks and Poisson-based models that ignore clustering. In particular, even when Poisson models are calibrated to match the expected attack frequency, they fail to capture the implications of attack clustering on investment timing and magnitude, thus leading to suboptimal investment decisions. Our findings indicate that the optimal dynamic strategy is able to react promptly to attack clusters, offering substantial improvements in expected net benefit in high-risk scenarios. Overall, our results underscore the importance of accounting for dynamic and stochastic threat patterns in cybersecurity planning. The proposed framework supports risk managers and policymakers in designing responsive cybersecurity investment strategies tailored to the evolving cyber-risk landscape.

Future research directions include a rigorous theoretical investigation of the control problem. In this work, most results have been derived under the assumption of sufficient regularity of the value function. Nevertheless, such regularity is not guaranteed in general. A natural next step is to study the viscosity solutions of the associated PIDE. In the case of control with Hawkes processes, only partial results on viscosity solutions are available, see Bensoussan and Chevalier-Roignant (2024), Houssard et al. (2025) and Gaïgi et al. (2025), and, to the best of our knowledge, none of them is directly applicable in our case. Advancing this line of research could provide a more solid theoretical foundation for control problems with self-exciting jumps and rigorously justify the assumptions employed in this work. On the more applicative side, further developments might regard the empirical calibration to sector-specific cyber incident data, the consideration of risk-aversion with respect to losses resulting from cyberattacks, and the integration of cyber-insurance as a complementary tool for risk mitigation (see Awiszus et al. (2023), Dou et al. (2020), Mazzoccoli and Naldi (2020), Miaoui and Boudriga (2019), Öğüt et al. (2011) and Skeoch (2022) for some recent studies in this direction). Our framework can also be applied from the viewpoint of an insurance firm which provides insurance against losses due to cyberattacks, thus laying the foundations for the development of Cramér-Lundberg-type models (see, e.g., Mikosch (2009)) for cyber-insurance. Finally, our modeling setup can also be extended to multivariate Hawkes processes (as considered in Embrechts et al. (2011), or in the more general versions of Bielecki et al. (2022) and Bielecki et al. (2023)) to differentiate among multiple types of cyberattacks (see Bentley et al. (2020) for a multivariate generalization of the static Gordon-Loeb model).

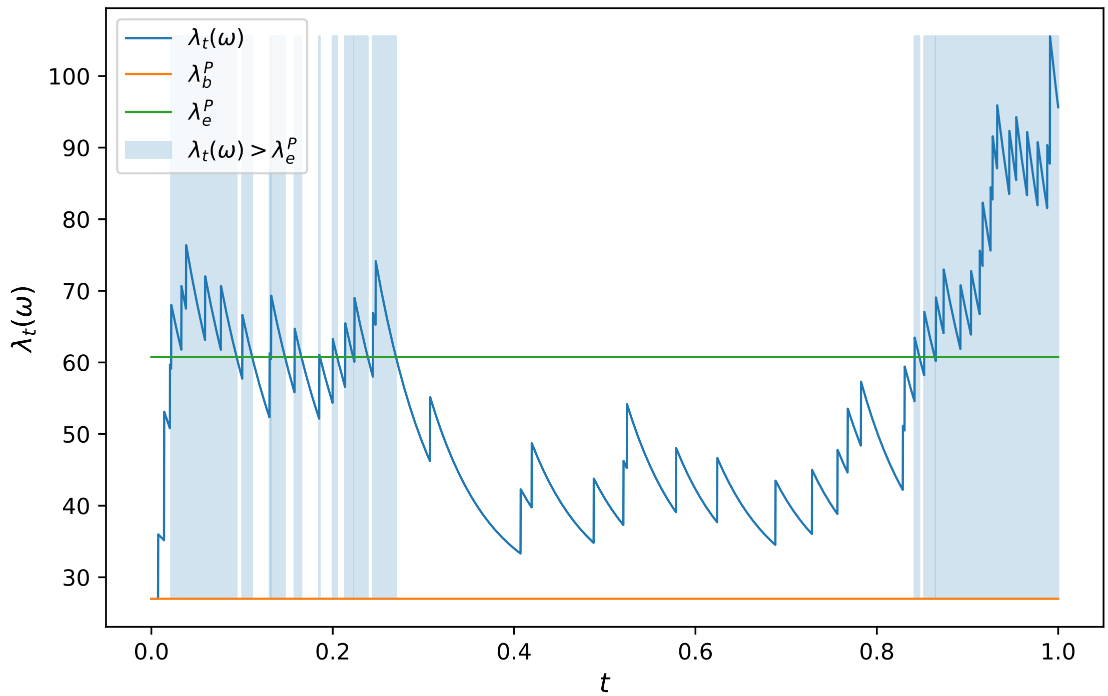


(a) Intensity path 1.

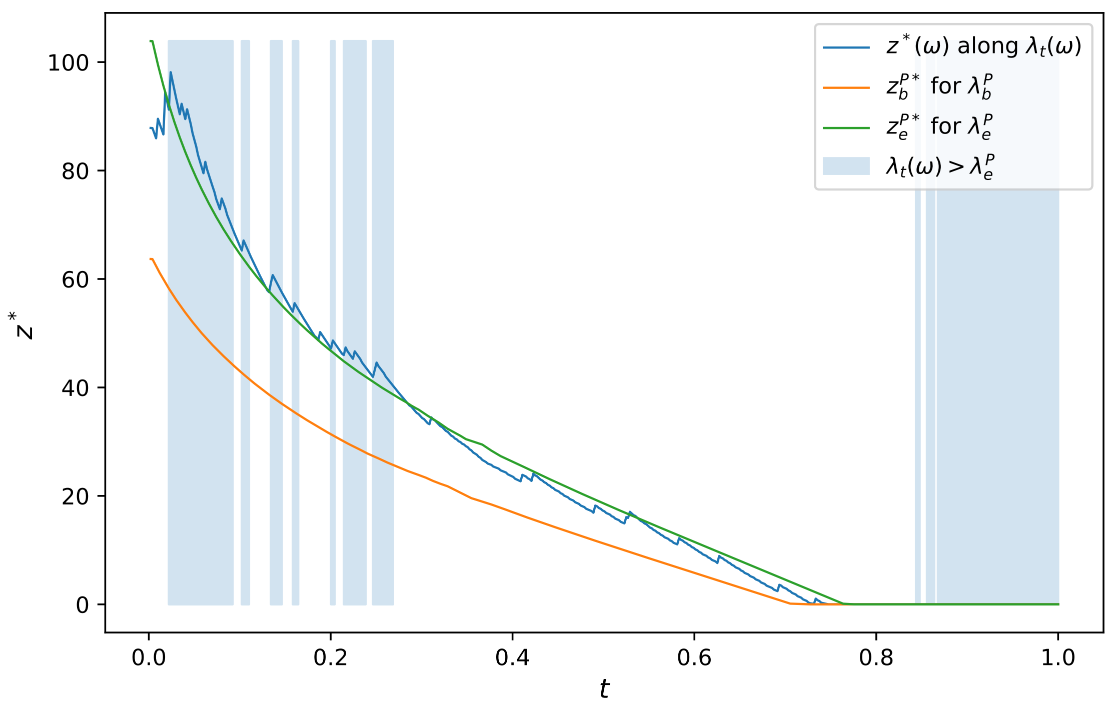


(b) Optimal strategy 1.

Figure 2.10: Simulated intensity path and optimal strategy (trajectory 1).



(a) Intensity path 2.



(b) Optimal strategy 2.

Figure 2.11: Simulated intensity path and optimal strategy (trajectory 2).

CHAPTER 3

Stable measure transformations for affine jump-diffusions

The work is an ongoing research project with Prof. Claudio Fontana.

Affine processes are Markov processes whose characteristic functions admit an exponentially affine dependence on the initial state. This structural property underlies their widespread use in applications, as it ensures a balance between model flexibility and analytical tractability. This naturally raises the question of whether the affine property can be preserved under equivalent changes of probability measures. In this work, we provide a full characterization of the class of locally equivalent probability measures that preserve the affine structure in a general jump–diffusion setting. We generalize existing results in the literature by providing necessary and sufficient conditions for admissible transformations, yielding an explicit and verifiable criterion that enhances the applicability of our work across multiple areas.

3.1 Introduction

Affine processes are a class of Markov processes whose characteristic function has exponentially affine dependence on the initial state. While special cases of affine processes had been studied earlier in the literature, a general definition was first formalized in Duffie, Filipović et al. (2003). Due to their analytical tractability and modeling flexibility, affine processes were largely employed in mathematical finance and probability, see e.g. Cuchiero et al. (2011), Duffie, Pan et al. (2000), Filipovic and Mayerhofer (2009), Keller-Ressel (2008) and Keller-Ressel and Mayerhofer (2015). Their key feature is that their characteristic function can be expressed in terms of a system of generalized Riccati differential equations, which allows for semi-closed form solutions. This mathematical structure makes affine processes a powerful tool across various fields. They have been extensively used interest rates modeling, with classical examples including the models in Vasicek (1977), J. C. Cox et al. (1985) and Dai and Singleton multivariate extension, Dai and Singleton (2000). In credit risk, they underlie intensity-based models, see Duffie (2005). In asset pricing, many widely used models are affine, such as the Black–Scholes model, exponential Lévy models, Cont and Tankov (2003), Heston (1993), and Bates (1996). They have also found applications in insurance, particularly in longevity and mortality risk modeling, see Biffis (2005), Schrager (2006) and Luciano and Vigna (2008).

The aforementioned mathematical structure enables efficient calculations, and it is therefore desirable to be maintained when performing changes of measure. In many financial and actuarial applications, it is necessary to move from the real-world measure $\mathbb{P}$ to a risk-neutral measure $\mathbb{Q}$. Usually, statistical estimation and risk management is performed under the real-world measure, while pricing is carried under the the risk-neutral probability. In the context of credit risk, the survival probabilities are computed under $\mathbb{P}$, while the arbitrage-free valuation of financial derivatives is performed under $\mathbb{Q}$. Similarly, in interest rates modeling, statistical estimation is performed under the physical measure, and financial products are evaluated under the risk-neutral one. In life insurance, models are estimated using demographic data under the real-world measure, whereas the valuation of insurance products or longevity-related securities are performed under $\mathbb{Q}$. In arbitrage theory, the change of measure from $\mathbb{P}$ to an equivalent martingale measure $\mathbb{Q}$ is fundamental to ensure the absence of arbitrage opportunities and to express discounted asset prices as $\mathbb{Q}$-martingales. Beyond finance and insurance, the preservation of the affine structures under changes of measure is also highly desirable in other fields, such as stochastic control. We refer, for instance, to risk-sensitive control problems, see Fleming and Soner (2006, Section VI.2), where changing measure allows to solve a more tractable control problem. Preservation of the affine structure under measure changes is therefore highly desirable, both for tractability and practical implementation.

The aim of this chapter is to characterize the family of all locally equivalent probability measures under which the affine structure is preserved. The task is not trivial and a complete characterization is not found in the literature, although several papers address related issues. In Palmowski and Rolski 2002, the authors discuss absolute continuity for general classes of Markov processes. They employ an exponential martingale as the density process and show that, under mild assumptions, the process remains Markov under the new equivalent measure. In Cheridito, Filipović and Yor (2005), the authors study the sufficient conditions for two processes to be absolutely continuous. The problem is analyzed in a general framework where the considered processes are defined as solutions to martingale problems, without imposing semimartingale assumptions. This work provides a general and useful result, nevertheless we suppose that weaker and more explicit conditions can be found when considering an affine structure, and most importantly, necessary conditions can be derived. A first contribution is this direction is provided by Cheridito, Filipović and Kimmel (2007), where they characterize the equivalent measure changes that preserve affine structure for affine diffusive processes. Their argumentation, particularly Theorem 1, refines and contextualizes the results of Cheridito, Filipović and Yor (2005, Theorem 2.4) in a more structured context. We aim at establishing an analogous result in general jump-diffusion setting.

A similar problem is tackled in Fontana (2012), where the author studies an analogous characterization in the context of credit risk. In particular, the work considers an intensity-based model in which the default intensity is a linear function of an affine diffusion process and studies which changes of measures maintains the affine structure. Also the preservation of the immersion property under a change of measure is addressed. The work represents an initial example of structure-preserving measure changes for affine jump processes, and the approach can be generalized to a broader class of affine processes.

In a general affine semimartingale setting, sufficient conditions for the preservation of the affine structure under a change of measure have been obtain in Kallsen and Muhle-Karbe (2010, Section 4). The authors show that, under certain conditions, the stochastic exponential of an affine process can serve as the Radon–Nikodym derivative for a structure-preserving change of measure. While this approach identifies some valid changes of measure, it does not provide a full characterization of all such transformations.

To give a broader point of view on the topic, we recall that this structure-preserving change of measure characterization is connected to the question of whether a positive exponential local martingale is a true martingale and, hence, can be used as the density process for an equivalent change of measure. This problem has been central in stochastic calculus, with foundational contributions by Novikov (1973) and Kazamaki (1977), and generalizations to semimartingales with jumps made by Lépingle and Mémin (1978), Mémin and Shiryaev (1979) and Mémin (2006). Comprehensive treatments on the topic are provided in Protter (2005, Section III.8), Revuz and Yor (2013, Chapter VIII). The problem has also been largely studied in the context of mathematical finance, see e.g. Kallsen and Shiryaev (2002), Cheridito, Filipović and Yor (2005), Protter and Shimbo (2008), Blei and Engelbert (2009), Kallsen and Muhle-Karbe (2010), Mijatović and Urusov (2012) and Mayerhofer et al. (2011).

Our contribution lies in characterizing all stable measure transformations in a multi-dimensional affine setting, assuming some boundary non-attainment condition on the process. The chapter is structured as follows: In Section 3.2, we introduce the main notions connected to affine processes and discuss some sufficient boundary non-attainment conditions. In Section 3.3, we state and prove the main theorems, which characterize all the structure-preserving measure transformations. In Section 3.4, we discuss how our results relate the existing literature. In Section 3.5, we apply of our main findings to key jump-diffusion models.

3.1.1 Notation

Throughout the chapter, we use the following notation, consistent with Jacod and Shiryaev (2013).

- $L^2_{\text{loc}}(W)$ is the set of all predictable processes H such that the process $H \cdot \langle W, W \rangle$ is locally integrable, where W is a d-dimensional Brownian motion.

- $\mathcal{M}(\mathbb{P})$ is the set of uniformly integrable martingales with respect to a probability measure $\mathbb{P}$. $\mathcal{M}_{\text{loc}}(\mathbb{P})$ denotes the class of local martingales, obtained by localizing $\mathcal{M}(\mathbb{P})$.

- $\mathcal{A}_{\text{loc}}(\mathbb{P})$ is the set of adapted processes with locally integrable variation with respect to a probability measure $\mathbb{P}$.

- $G_{\text{loc}}(\mu)$ denotes the set of predictable functions locally integrable with respect to the compensated random measure $\mu - \nu$ in the sense of Jacod and Shiryaev (2013, Definition II.1.27).

3.2 Affine processes and boundary non-attainment conditions

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $W^{\mathbb{P}}$ be a d-dimensional Brownian motion and μ a random measure supported on the probability space. We call an *affine jump-diffusion* a solution to the SDE having the following structure

$$\mathrm{d}X_t = (b + AX_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t^{\mathbb{P}} + \int_E \xi\mu(\mathrm{d}t, \mathrm{d}\xi), \quad X_0 \in E \tag{3.1}$$

where $E \subseteq \mathbb{R}^d$ is the state space of the process, $b \in \mathbb{R}^d$, $A \in \mathbb{R}^{d\times d}$, the covariation is given by

$$\sigma(x)\sigma(x)^\top = \Sigma_0 + \sum_{i=1}^{d} \Sigma_j x_j,$$

for $\Sigma_j \in \mathbb{R}^{d\times d}$, $j \in 0, \dots, d$, and the random measure μ is associated to the following compensating measure

$$\nu^{\mathbb{P}}(\mathrm{d}t, \mathrm{d}\xi) = \theta_0(\mathrm{d}\xi) + \sum_{j=1}^{d} \theta_j(\mathrm{d}\xi)X_{t-}, \tag{3.2}$$

for some measures $\theta_j(\mathrm{d}\xi)$ on E. Equation (3.1) does not admit strong solution in general, nor is uniqueness guaranteed. Moreover, not every subset E of $\mathbb{R}^d$ is an admissible state space. In Duffie, Filipović et al. (2003), the authors study a set of *admissibility conditions*, i.e. they provide some conditions on the parameters in order for the problem to admit strong solution and uniqueness in law. In this work, we focus only on the solutions of affine SDEs having values in $E = \mathbb{R}^m_{++} \times \mathbb{R}^{d-m}$, for a certain $m \in \{1, \dots, d\}$, where $\mathbb{R}^m_{++} := \{x \in \mathbb{R}^m : X^i > 0, \forall i = 1, \dots, m\}$. In particular, we highlight in Assumption A the sufficient conditions for the non-negative components of the X process to never reach the boundary.

Assumption A. Consider:

(A1) A set of $d + 1$ matrices $\Sigma_j \in \mathbb{R}^{d\times d}$, $j = 0, \dots, d$ such that:

- (a) Σ_j for $j = 0, \dots, m$ are symmetric positive semi-definite, with Σ_j positive definite for at least one j.
- (b) $\Sigma_j = 0$ if $j \geq m + 1$.

(A2) $\sigma(x)\sigma(x)^T = \Sigma_0 + \sum_{j=1}^{m} \Sigma_j x_j$, $\sigma^{ij}(x) = \delta_{ij}\sqrt{\Sigma_i^{ii} x_i}$ for $i \in \{1,\dots,m\}$ and $j \in \{1,\dots,d\}$, where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise.

(A3) A d-dimensional vector b such that $b_i \geq \frac{1}{2}\Sigma_i^{ii}$, for $i \in \{1,\dots,m\}$.

(A4) A matrix A such that $A_{ij} = 0$, for $i \in \{1,\dots,m\}$ and $j \in \{m+1,\dots,d\}$, and $A^{ij} \geq 0$ for $i, j \in \{1,\dots,m\}$ with $i \neq j$.

(A5) A Borel measure m on E and $d+1$ non-negative measurable functions $m_j \colon E \longrightarrow \mathbb{R}_+$, $j = 0,\dots m$ such that:

(a) $(m_0(\xi),\dots,m_m(\xi)) \in \mathbb{R}_+^m \setminus \{(0,\dots,0)\} \forall \xi \in E$.

(b) $m_j(\xi) \equiv 0$ for $j \geq m+1$.

(c) Let $K(x, \mathrm{d}\xi) = \left(m_0(\xi) + \sum_{j=1}^{m} m_j(\xi)x_j\right) m(\mathrm{d}\xi)$,

$$\int_E (1 \wedge ||\xi||) K(x, \mathrm{d}\xi) < \infty \quad \forall x \in E. \tag{3.3}$$

Remark 3.1. In Assumption (A2), we require σ to have a diagonal structure. This allows to consider simpler Feller conditions, see Assumption (A3). In Assumption (A5)(c), we ask for the compensating measure to be absolutely continuous with respect to a Borel measure m, which satisfies appropriate integrability conditions. This structure can be obtained starting from equation (3.2), defining $m := \sum_{j=0}^{m} \theta_j(\mathrm{d}\xi)$. By construction, θ_j, $j = 0,\dots,m$, are absolutely continuous with respect to m. We also highlight condition (3.3), which allows us to consider no truncation function, see Sato (1999, Remark 8.4).

Consider a tuple

$$\left(b + Ax, \Sigma_0 + \sum_{j=1}^{m} \Sigma_j x_j, \left(m_0(\xi) + \sum_{j=1}^{m} m_j(\xi)x_j\right) m(\mathrm{d}\xi)\right)$$

which satisfies Assumption A. Exploiting the key result provided in Duffie, Filipović et al. (2003, Theorem 2.7), we conclude that the SDE (3.1) associated with this set of parameters admits a strong solution on the given probability space, provided that the jump measure μ has compensator

$$\nu^{\mathbb{P}}(\mathrm{d}t, \mathrm{d}\xi) = \left(m_0(\xi) + \sum_{j=1}^{m} m_j(\xi)x_j\right) m(\mathrm{d}\xi)\mathrm{d}t.$$

The same theorem also guarantees uniqueness in law for the process X under $\mathbb{P}$. We can also establish that, under Assumption A, X is strictly positive process. Consider the Cox–Ingersoll–Ross process Y given by

$$Y_t = X_0 + \int_0^t (b + AY_s)\,\mathrm{d}s + \int_0^t \sigma(Y_s)\,\mathrm{d}W_s^{\mathbb{P}}.$$

By Assumption (A3), it follows from J. C. Cox et al. (1985, Section 3) that $Y_t^i > 0$ for all $t \geq 0, i \in \{1, \ldots, m\}$. Since the jump measure μ introduces only non-negative jumps due to Assumption (A5), $X_t^i \geq Y_t^i > 0$ for all $t \geq 0$, $i \in \{1, \ldots, m\}$. We observe that Assumptions (A3) and (A5) are sufficient boundary non-attainment conditions and guarantee that the first m components of X are always strictly positive.

3.3 Stable measure transformation

Notation. Whenever we refer to *two sets of parameters*, we mean two tuples

$$\left(b^{\mathbb{P}} + A^{\mathbb{P}}x, \Sigma_0 + \sum_{j=1}^{m} \Sigma_j x_j, \left(m_0^{\mathbb{P}}(\xi) + \sum_{j=1}^{m} m_j^{\mathbb{P}}(\xi) x_j\right) m(\mathrm{d}\xi)\right),$$

$$\left(b^{\mathbb{Q}} + A^{\mathbb{Q}}x, \Sigma_0 + \sum_{j=1}^{m} \Sigma_j x_j, \left(m_0^{\mathbb{Q}}(\xi) + \sum_{j=1}^{m} m_j^{\mathbb{Q}}(\xi) x_j\right) m(\mathrm{d}\xi)\right).$$

The first one is associated to a probability measure $\mathbb{P}$, while the second to a measure $\mathbb{Q}$. We will not indicate the probability measure for covariation terms, since these remain unchanged under an equivalent measure change.

Consider a finite horizon $T > 0$. Let $(X_t)_{t\in[0,T]}$ denote the affine process given as the unique strong solution on $[0, T]$ of the SDE (3.1) for the set of parameters $\left(b^{\mathbb{P}} + A^{\mathbb{P}}x, \Sigma_0 + \sum_{j=1}^{m} \Sigma_j x_j, \left(m_0^{\mathbb{P}}(\xi) + \sum_{j=1}^{m} m_j^{\mathbb{P}}(\xi) x_j\right) m(\mathrm{d}\xi)\right)$ which satisfies Assumption A. We aim to characterize all the locally equivalent probability measures $\mathbb{Q} \sim^{\mathrm{loc}} \mathbb{P}$, under which X preserves its affine structure in the sense of Definition 3.2. We show that X maintains its affine structure under such a measure $\mathbb{Q}$ if and only if the associated density process Z satisfies Property B, precisely stated below.

Definition 3.2. Let $\mathbb{Q}$ be a probability measure on $(\Omega, \mathcal{F})$. We say that the process X has an affine structure with respect to $\mathbb{Q}$ if it satisfies on $(\Omega, \mathcal{F}, \mathbb{Q})$ an affine SDE of the type

$$X_t = X_0 + \int_0^t (b^{\mathbb{Q}} + A^{\mathbb{Q}} X_s)\mathrm{d}s + \int_0^t \sigma(X_s)\mathrm{d}W_s^{\mathbb{Q}} + \int_0^t \int_E \xi\mu(\mathrm{d}s, \mathrm{d}\xi), \quad t \in [0, T], \tag{3.4}$$

where $\left(b^{\mathbb{Q}} + A^{\mathbb{Q}}x, \Sigma_0 + \sum_{j=1}^m \Sigma_j x_j, \left(m_0^{\mathbb{Q}}(\xi) + \sum_{j=1}^m m_j^{\mathbb{Q}}(\xi)x_j\right) m(\mathrm{d}\xi)\right)$ are some parameters satisfying Assumption A, $W^{\mathbb{Q}}$ is a $\mathbb{Q}$-Brownian motion, μ a random measure having compensator $\nu^{\mathbb{Q}}(\mathrm{d}t, \mathrm{d}\xi) = \left(m_0^{\mathbb{Q}}(\xi) + \sum_{j=1}^m m_j^{\mathbb{Q}}(\xi)x_j\right) m(\mathrm{d}\xi)\mathrm{d}t$.

Property B. Consider two sets of parameters which satisfy Assumption A. We consider a density process

$$Z_t = \mathcal{E}\left(\int_0^t \phi(X_{s-})^\top \mathrm{d}W_s^{\mathbb{P}} + \int_0^t \int_E \psi(X_{s-}, \xi) \cdot (\mu(\mathrm{d}s, \mathrm{d}\xi) - \nu^{\mathbb{P}}(\mathrm{d}s, \mathrm{d}\xi))\right), \tag{3.5}$$

where the $\mathbb{R}^d$-valued function $\phi \colon E \to \mathbb{R}^d$ has the following form

$$\phi(X_{s-}) = \sigma(X_{s-})^{-1}(b^{\mathbb{Q}} - b^{\mathbb{P}} + (A^{\mathbb{Q}} - A^{\mathbb{P}})X_{s-}), \tag{3.6}$$

while the function $\psi \colon E \times E \to (-1, +\infty)$ must be of the following kind

$$\psi(X_{s-}, \xi) = \frac{m_0^{\mathbb{Q}}(\xi) + \sum_{j=1}^m m_j^{\mathbb{Q}}(\xi)X_{s-}^j}{m_0^{\mathbb{P}}(\xi) + \sum_{j=1}^m m_j^{\mathbb{P}}(\xi)X_{s-}^j} - 1. \tag{3.7}$$

The functions ϕ and ψ are well-defined thanks to Assumption A.

In Theorem 3.3 we prove that, when the density process Z is constructed to satisfy Property B, the resulting probability measure $\mathbb{Q}$ is locally equivalent to $\mathbb{P}$ and preserves the affine structure of X.

Theorem 3.3. *Consider two sets of parameters which satisfy Assumption A. Let X be the strong unique solution of the SDE (3.1) under the measure $\mathbb{P}$, with the corresponding parameters, and Z a process which satisfies Property B. Assume that the following condition holds:*

$$\int_E \left(\sqrt{m_0^{\mathbb{Q}}(\xi) + \sum_{j=1}^m m_j^{\mathbb{Q}}(\xi)x_j} - \sqrt{m_0^{\mathbb{P}}(\xi) + \sum_{j=1}^m m_j^{\mathbb{P}}(\xi)x_j}\right)^2 m(\mathrm{d}\xi) < \infty \tag{3.8}$$

for every $x \in E$. Then, Z is the density process with respect to $\mathbb{P}$ of a locally equivalent probability measure $\mathbb{Q}$ and X satisfies the SDE (3.4) under $\mathbb{Q}$.

Proof. We consider the process Z as defined in equation (3.5) and assume that it satisfies Property B. If we show that Z is a $\mathbb{P}$-martingale, then the measure $\mathbb{Q} := Z_T \cdot \mathbb{P}$ defines a probability measure equivalent to $\mathbb{P}$. Consequently, it follows by Jacod and Shiryaev (2013, Theorem III.3.24) that:

- The process

$$W_t^{\mathbb{Q}} = W_t^{\mathbb{P}} - \int_0^t \phi(X_s)\mathrm{d}s = W_t^{\mathbb{P}} - \int_0^t \sigma(X_s)^{-1}(b^{\mathbb{Q}} - b^{\mathbb{P}} + (A^{\mathbb{Q}} - A^{\mathbb{P}})X_s)\mathrm{d}s$$

  is a Brownian motion under $\mathbb{Q}$.

- The compensator of the jump measure under $\mathbb{Q}$ is given by

$$\nu^{\mathbb{Q}}(\mathrm{d}s, \mathrm{d}\xi) = (\psi(X_{s-}, \xi) + 1) \cdot \nu^{\mathbb{P}}(\mathrm{d}s, \mathrm{d}\xi) = \left(m_0^{\mathbb{Q}}(\xi) + \sum_{j=1}^{m} m_j^{\mathbb{Q}}(\xi) X_{s-}^j \right) m(\mathrm{d}\xi)\mathrm{d}s.$$

- Under $\mathbb{Q}$, X solves equation (3.4) for every $t \in [0, T]$.

We consider the auxiliary process $\tilde{X}$ under the probability measure $\mathbb{P}$.

$$\tilde{X}_t = X_0 + \int_0^t (b^{\mathbb{Q}} + A^{\mathbb{Q}}\tilde{X}_s)\mathrm{d}s + \int_0^t \sigma(\tilde{X}_s)\mathrm{d}W_s^{\mathbb{P}} + \int_0^t \int_E \xi \tilde{\mu}(\mathrm{d}s, \mathrm{d}\xi),$$

where $\tilde{\mu}$ is a jump measure having compensator $\nu^{\mathbb{Q}}$ under $\mathbb{P}$. Exploiting the existence and uniqueness result of Duffie, Filipović et al. (2003, Theorem 2.7), presented in Section 3.2, we can state that the distribution of $\tilde{X}$ is unique under $\mathbb{P}$ and that $\tilde{X}^j$ for $j \in \{1, \dots, m\}$ is strictly positive. For each $n \geq 1$, we define

$$U_n := \left(\frac{1}{n}, n\right)^m \times (-n, n)^{d-m}$$

and two families of stopping times as

$$\begin{aligned} \tau_n &= \inf\{t > 0 : X_t \notin U_n \text{ or } X_{t-} \notin U_n\} \wedge n \wedge T, \\ \tilde{\tau}_n &= \inf\{t > 0 : \tilde{X}_t \notin U_n \text{ or } \tilde{X}_{t-} \notin U_n\} \wedge n \wedge T. \end{aligned} \tag{3.9}$$

The stopping times satisfy

$$\lim_{n\to\infty} \mathbb{P}(\tau_n = T) = \lim_{n\to\infty} \mathbb{P}(\tilde{\tau}_n = T) = 1. \tag{3.10}$$

We observe that the function ϕ as defined in equation (3.6), is continuous in U_n thus bounded in the set. We can state that for all $n > 0$, it exists $K_n \in \mathbb{R}_+$ such that

$$\phi(x)^\top \phi(x) < K_n$$

for all $x \in U_n$. Therefore, by Jacod and Shiryaev (2013, Theorem I.4.40 b)), the stochastic integral $\int_0^\cdot \phi(X_{t-})^\top \mathrm{d}W_t^{\mathbb{P}}$ is well-defined as a local martingale under $\mathbb{P}$. We can also state that for every $x \in U_n$, it exists $M_n > 0$ such that

$$\int_E \left(\sqrt{m_0^{\mathbb{Q}}(\xi) + \sum_{j=1}^m m_j^{\mathbb{Q}}(\xi) x_j} - \sqrt{m_0^{\mathbb{P}}(\xi) + \sum_{j=1}^m m_j^{\mathbb{P}}(\xi) x_j} \right)^2 m(\mathrm{d}\xi) < M_n.$$

The latter property will be proved in Lemma 3.4. This inequality implies that

$$\begin{aligned} &\int_0^{T\wedge\tau_n} \int_E (1 - \sqrt{\psi(X_{t-},\xi)+1})^2 \nu^{\mathbb{P}}(\mathrm{d}t, \mathrm{d}\xi) \\ &\quad = \int_0^{T\wedge\tau_n} \int_E \left(\sqrt{m_0^{\mathbb{Q}}(\xi) + \sum_{j=1}^m m_j^{\mathbb{Q}}(\xi) x_j} - \sqrt{m_0^{\mathbb{P}}(\xi) + \sum_{j=1}^m m_j^{\mathbb{P}}(\xi) x_j} \right)^2 m(\mathrm{d}\xi)\mathrm{d}t \leq M_n T \end{aligned} \tag{3.11}$$

for all n. Applying Jacod and Shiryaev (2013, Theorem II.1.33, d)), this implies that $\psi(X_{t-},\xi) \in G_{\mathrm{loc}}(\mu)$ under $\mathbb{P}$ and therefore the stochastic integral

$$\int_0^\cdot \int_E \psi(X_{t-},\xi)(\mu(\mathrm{d}t, \mathrm{d}\xi) - \nu^{\mathbb{P}}(\mathrm{d}t, \mathrm{d}\xi))$$

is well-defined as a local martingale under $\mathbb{P}$. Exploiting Jacod and Shiryaev (2013, Theorem I.4.61), we can conclude that Z as defined in equation (3.5), is a well-defined, strictly positive local martingale under $\mathbb{P}$, hence a supermartingale. To prove that Z is a true martingale, it suffices to verify that $\mathbb{E}^{\mathbb{P}}[Z_T] = 1$. We define the process

$$\begin{aligned} Z_t^n = \mathcal{E}\Bigg(&\int_0^t \phi(X_{s-})^\top \mathbb{1}_{\{s\leq\tau_n\}} \, \mathrm{d}W_s^{\mathbb{P}} \\ &+ \int_0^t \int_E \psi(X_{s-},\xi) \mathbb{1}_{\{s\leq\tau_n\}} \cdot (\mu(\mathrm{d}s, \mathrm{d}\xi) - \nu^{\mathbb{P}}(\mathrm{d}s, \mathrm{d}\xi)) \Bigg). \end{aligned} \tag{3.12}$$

It can be proved that Z_T^n is a $\mathbb{P}$-martingale, see Lemma 3.5, and that $\mathbb{Q}^n = Z_T^n \cdot \mathbb{P}$ is a probability measure equivalent to $\mathbb{P}$. Under $\mathbb{Q}^n$:

- $W_t^n = W_t^{\mathbb{P}} - \int_0^t \phi(X_{s-})\mathbb{1}_{\{s\leq\tau_n\}}\mathrm{d}s$ is a Brownian motion.
- The compensating measure is given by $\nu^n(\mathrm{d}t,\mathrm{d}\xi) = (\psi(X_{t-},\xi)\mathbb{1}_{\{t\leq\tau_n\}} + 1)\nu^{\mathbb{P}}(\mathrm{d}t,\mathrm{d}\xi)$.

We can write the process X under the measure $\mathbb{Q}^n$, stopped at time τ_n as

$$X_{t\wedge\tau_n} = X_0 + \int_0^{t\wedge\tau_n}(b^{\mathbb{Q}} + A^{\mathbb{Q}}X_s)\mathrm{d}s + \int_0^{t\wedge\tau_n}\sigma(X_s)\mathrm{d}W_s^n + \int_0^{t\wedge\tau_n}\int_E \xi\mu'(\mathrm{d}s,\mathrm{d}\xi),$$

where μ' is a jump measure having compensating measure given by $\nu^n(\mathrm{d}t,\mathrm{d}\xi)$. Due to Assumption A, the distribution of $\tilde{X}$ under $\mathbb{P}$ is unique, as application of Duffie, Filipović et al. (2003, Theorem 2.7). In particular, the martingale problem associated to $\tilde{X}$ under $\mathbb{P}$ is well-posed and has unique solution. It follows by Jacod and Shiryaev (2013, Theorem III.2.40) that there exists a unique solution to the stopped martingale problem associated to the characteristics of $\tilde{X}$ under $\mathbb{P}$. We conclude that the distribution of $\tilde{X}_{\cdot\wedge\tau_n}$ under $\mathbb{P}$ is unique. We observe that the process $X_{\cdot\wedge\tau_n}$ under $\mathbb{Q}^n$ has the same martingale characteristics of $\tilde{X}$, thus under $\mathbb{Q}^n$, the stopped process $(X_{t\wedge\tau_n})_{t\geq 0}$ has the same distribution as the stopped process $(\tilde{X}_{t\wedge\tilde{\tau}_n})_{t\geq 0}$ under $\mathbb{P}$. Therefore

$$\mathbb{E}^{\mathbb{P}}[Z_T] = \lim_n \mathbb{E}^{\mathbb{P}}[Z_T^n \mathbb{1}_{\{\tau_n = T\}}] = \lim_n \mathbb{Q}^n(\tau_n = T) = \lim_n \mathbb{P}(\tilde{\tau}_n = T) = 1.$$

The first equality follows from monotone convergence, while the second follows by applying the definition of measure $\mathbb{Q}^n$. The third step follows from the fact that the distribution of $(X_{t\wedge\tau_n})_{t\geq 0}$ under $\mathbb{Q}$ is the same distribution of $(\tilde{X}_{t\wedge\tilde{\tau}_n})_{t\geq 0}$ under $\mathbb{P}$. The last equality follows from equation (3.10). □

Lemma 3.4. *Consider two sets of parameters which satisfy Assumption A and assume that inequality (3.8) holds for every $x \in \mathbb{R}^m_{++}$. Then*

$$\int_E \left(\sqrt{m_0^{\mathbb{Q}}(\xi) + \sum_{j=1}^m m_j^{\mathbb{Q}}(\xi)x_j} - \sqrt{m_0^{\mathbb{P}}(\xi) + \sum_{j=1}^m m_j^{\mathbb{P}}(\xi)x_j}\right)^2 m(\mathrm{d}\xi) < M_n$$

for every $x \in \prod^m \left(\frac{1}{n}, n\right)$.

Proof. For every $x \in \mathbb{R}^m_{++}$, we define the function $G(x)$ as

$$x \mapsto G(x) := \int_E \left(\sqrt{m_0^{\mathbb{Q}}(\xi) + \sum_{j=1}^m m_j^{\mathbb{Q}}(\xi)x_j} - \sqrt{m_0^{\mathbb{P}}(\xi) + \sum_{j=1}^m m_j^{\mathbb{P}}(\xi)x_j}\right)^2 m(\mathrm{d}\xi) < \infty.$$

By hypothesis, $G(x) < \infty$ for every $x \in \mathbb{R}^m_{++}$, we aim to prove that it exists $M_n > 0$ such that $G(x) < M_n$ when $x \in \prod^m \left(\frac{1}{n}, n\right)$. We define $\kappa(u,v) := u + v - 2\sqrt{uv}$, for $u, v > 0$. We observe that κ is convex, since its Hessian its given by

$$\begin{pmatrix} \frac{\sqrt{v}}{2u^{3/2}} & -\frac{1}{2\sqrt{uv}} \\ -\frac{1}{2\sqrt{uv}} & \frac{\sqrt{u}}{2v^{3/2}} \end{pmatrix},$$

which is a positive semi-definite matrix. It follows that also

$$x \mapsto \kappa\left(m_0^{\mathbb{P}}(\xi) + \sum_{j=1}^m m_j^{\mathbb{P}}(\xi)x_j, m_0^{\mathbb{Q}}(\xi) + \sum_{j=1}^m m_j^{\mathbb{Q}}(\xi)x_j\right)$$

is convex in x for very ξ, since convexity is invariant under affine maps. We have

$$G(x) = \int_E \kappa(m_0^{\mathbb{P}}(\xi) + \sum_{j=1}^m m_j^{\mathbb{P}}(\xi)x_j, m_0^{\mathbb{Q}}(\xi) + \sum_{j=1}^m m_j^{\mathbb{Q}}(\xi)x_j)m(\mathrm{d}\xi).$$

Due to the convexity of κ in x, we can conclude that also G is convex, thus continuous in its thus continuous in the interior of its effective domain $\mathbb{R}^m_{++}$, i.e. $\mathbb{R}^m_{++}$ itself, see Mordukhovich and Nam (2022, Corollary 2.153). Then being $\prod^m \left(\frac{1}{n}, n\right)$ a bounded set, we can write $G(x) \leq M_n$ for $x \in \prod^m \left(\frac{1}{n}, n\right)$. □

The following lemma plays a crucial role in the proof of Theorem 3.3. It establishes that the stopped process associated with our density Z is a true martingale under $\mathbb{P}$, which is a key step to conclude that the original density Z itself is a true martingale under $\mathbb{P}$.

Lemma 3.5. *In the setting of Theorem 3.3, let Z^n be defined as in equation (3.12), where ϕ and ψ are given by equations (3.6), (3.7) respectively and the stopping times $(\tau_n)_n$ are defined as in equation (3.9). Then, Z^n is a true martingale under $\mathbb{P}$.*

Proof. Consider

$$N = \int_0^\cdot \phi(X_{t-})^\top \mathrm{d}W_t^{\mathbb{P}} + \int_0^\cdot \int_E \psi(X_{t-}, \xi)(\mu(\mathrm{d}t, \mathrm{d}\xi) - \nu^{\mathbb{P}}(\mathrm{d}t, \mathrm{d}\xi))$$

and recall that Z as defined in equation (3.5) is the Doléans-Dade exponential of N, i.e. $Z = \mathcal{E}(N)$. We define the process

$$N^1 := \int_0^\cdot \phi(X_{t-})^\top \mathrm{d}W_t^{\mathbb{P}} + \int_0^\cdot \int_E \psi(X_{t-}, \xi)\mathbb{1}_{\{|\psi(X_{t-},\xi)|\leq 1\}}(\mu(\mathrm{d}t, \mathrm{d}\xi) - \nu^{\mathbb{P}}(\mathrm{d}t, \mathrm{d}\xi))$$

and

$$N^2 := N - N^1 = \int_0^\cdot \int_E \psi(X_{t-},\xi)\mathbb{1}_{\{|\psi(X_{t-},\xi)|>1\}}(\mu(\mathrm{d}t,\mathrm{d}\xi) - \nu^{\mathbb{P}}(\mathrm{d}t,\mathrm{d}\xi)).$$

By definition, N^1 and N^2 do not have common jumps and N^2 is a pure jump local martingale, hence $Z = \mathcal{E}(N^1 + N^2) = \mathcal{E}(N^1)\mathcal{E}(N^2)$, see Protter (2005, Theorem II.8.38). Our aim is to prove that $Z_{\cdot\wedge\tau_n}$ is a $\mathbb{P}$-martingale. The proof proceeds in the following steps:

- First we show that $\mathcal{E}(N^1)_{\cdot\wedge\tau_n}$ is a uniformly integrable martingale under $\mathbb{P}$.
- We introduce a new probability measure $\mathbb{P}^n$ via $\frac{\mathrm{d}\mathbb{P}^n}{\mathrm{d}\mathbb{P}} = \mathcal{E}(N^1)_{\tau_n}$.
- We prove that $\mathcal{E}(N^2)_{\cdot\wedge\tau_n}$ is a uniformly integrable martingale under $\mathbb{P}^n$. As a consequence, $Z_{\cdot\wedge\tau_n}$ is a uniformly integrable martingale under $\mathbb{P}$.

Let us first focus on N^1. By Ito's formula, we compute the square of $\mathcal{E}(N^1)$

$$\begin{aligned}\mathcal{E}(N^1)_t^2 = 1 + 2\int_0^t \mathcal{E}(N^1)_{s-}^2 \mathrm{d}N_s^1 + \int_0^t \mathcal{E}(N^1)_{s-}^2 \phi(X_{s-})^\top \phi(X_{s-})\mathrm{d}s \\ + \int_0^t \int_E \mathcal{E}(N^1)_{s-}^2 \psi^2(X_{s-})\mathbb{1}_{\{|\psi(X_{t-},\xi)|\le 1\}}\mu(\mathrm{d}s,\mathrm{d}\xi).\end{aligned}$$

Exploiting an analogous inequality to Jacod and Shiryaev (2013, Theorem II.1.33 d)), we note that $x^2\mathbb{1}_{|x|\le 1} + |x|\mathbb{1}_{|x|>1} \le c(1-\sqrt{1+x})^2$. Therefore, since inequality (3.11) holds, we have that

$$\int_0^{\tau_n\wedge T} \int_E \psi^2(X_{s-},\xi)\mathbb{1}_{\{|\psi(X_{s-},1)|\le 1\}}\mu^{\mathbb{P}}(\mathrm{d}s,\mathrm{d}\xi) \le C_n. \tag{3.13}$$

We now rewrite $\mathcal{E}(N^1)^2$ in terms of the compensating measure $\nu^{\mathbb{P}}$, i.e.

$$\begin{aligned}\mathcal{E}(N^1)_t^2 = 1 + 2\int_0^t \mathcal{E}(N^1)_{s-}^2 \mathrm{d}N_s^1 + \int_0^t \mathcal{E}(N^1)_{s-}^2 \phi(X_{s-})^\top \phi(X_{s-})\mathrm{d}s \\ + \int_0^t \int_E \mathcal{E}(N^1)_{s-}^2 \psi^2(X_{s-})\mathbb{1}_{\{|\psi(X_{t-},\xi)|\le 1\}}(\mu(\mathrm{d}s,\mathrm{d}\xi) - \nu^{\mathbb{P}}(\mathrm{d}s,\mathrm{d}\xi)) \\ + \int_0^t \int_E \mathcal{E}(N^1)_{s-}^2 \psi^2(X_{s-})\mathbb{1}_{\{|\psi(X_{t-},\xi)|\le 1\}}\nu^{\mathbb{P}}(\mathrm{d}s,\mathrm{d}\xi).\end{aligned}$$

Since the compensating measure $\nu^{\mathbb{P}}(\mathrm{d}s,\mathrm{d}\xi)$ does not jump, we state

$$\begin{aligned}\mathcal{E}(N^1)_{t\wedge\tau_n}^2 = \mathcal{E}\left(2N^1 + \int_0^{t\wedge\tau_n} \int_E \psi^2(X_{s-},\xi)\mathbb{1}_{\{|\psi(X_{s-},\xi)|\le 1\}}(\mu(\mathrm{d}s,\mathrm{d}\xi) - \nu^{\mathbb{P}}(\mathrm{d}s,\mathrm{d}\xi))\right)_{t\wedge\tau_n} \\ \cdot \exp\left(\int_0^{t\wedge\tau_n} \phi(X_{s-})^\top \phi(X_{s-})\mathrm{d}s + \int_0^{t\wedge\tau_n} \int_E \psi^2(X_{s-},\xi)\mathbb{1}_{\{|\psi(X_{s-},\xi)|\le 1\}}\nu^{\mathbb{P}}(\mathrm{d}s,\mathrm{d}\xi)\right)\end{aligned}$$

$$\leq e^{C_n}\mathcal{E}\left(2N^1 + \int_0^{t\wedge\tau_n}\int_E \psi^2(X_{s-},\xi)\mathbb{1}_{\{|\psi(X_{s-},\xi)|\leq 1\}}(\mu(\mathrm{d}s,\mathrm{d}\xi) - \nu^{\mathbb{P}}(\mathrm{d}s,\mathrm{d}\xi))\right)_{t\wedge\tau_n}, \tag{3.14}$$

where the latter inequality follows by the definition of τ_n, the continuity of ϕ and equation (3.13). Consider the stochastic exponential appearing in the last line of equation (3.14). It is a positive local martingale, where positivity follows by $2\Delta N^1 + \psi^2\mathbb{1}_{\{|\psi|\leq 1\}} = (2\psi + \psi^2)\mathbb{1}_{\{|\psi|\leq 1\}} > -1$ since $\psi > -1$, see Jacod and Shiryaev (2013, Theorem I.4.61). It hence follows that it is a supermartingale and

$$\mathbb{E}[\mathcal{E}(N^1)^2_{t\wedge\tau_n}] \leq e^{C_n}.$$

We have thus proved that $\mathcal{E}(N^1)_{\cdot\wedge\tau_n}$ is a square-integrable martingale hence uniformly integrable.

For each n, we can hence define the probability $\mathbb{P}^n$ by $\frac{\mathrm{d}\mathbb{P}^n}{\mathrm{d}\mathbb{P}} = \mathcal{E}(N^1)_{\tau_n}$. Let $\mathbb{E}^n$ the expectation under $\mathbb{P}^n$. In Theorem 3.3, we have proved that $\mathcal{E}(N^1)\mathcal{E}(N^2) = Z$ is a local martingale under $\mathbb{P}$ and by definition of $\mathbb{P}^n$ this implies $\mathcal{E}(N^2)_{\cdot\wedge\tau_n}$ is a local martingale under $\mathbb{P}^n$ for all n. It is then enough to prove that $\mathcal{E}(N^2)_{\cdot\wedge\tau_n}$ is a uniformly integrable martingale under $\mathbb{P}^n$, i.e. $\mathbb{E}^n[\mathcal{E}(N^2)_{T\wedge\tau_n}] = 1$, this would imply that also $Z_{\cdot\wedge\tau_n}$ is a is a uniformly integrable martingale under $\mathbb{P}$.

Note that $\mathcal{E}(N^2)_{\cdot\wedge\tau_n}$ is of finite variation and in $\mathcal{M}_{\mathrm{loc}}(\mathbb{P}^n)$, therefore also in $\mathcal{A}_{\mathrm{loc}}(\mathbb{P}^n)$ by Jacod and Shiryaev (2013, Lemma I.3.11). Let $(\sigma_m)_m$ be a localizing sequence of stopping times such that $\mathbb{E}^n[\sum_{s\leq\sigma_m}\Delta\mathcal{E}(N^2)_{s\wedge\tau_n}] < \infty$, for all m. We can compute:

$$\begin{aligned}
&\mathbb{E}^n\left[\sum_{s\leq\sigma_m}\Delta\mathcal{E}(N^2)_{s\wedge\tau_n}\right]\\
&= \mathbb{E}^n\left[\int_0^{\tau_n\wedge\sigma_m}\mathcal{E}(N^2)_{s-}\int_E \psi(X_{s-},\xi)\mathbb{1}_{\{|\psi(X_{s-},\xi)|>1\}}\mu(\mathrm{d}s,\mathrm{d}\xi)\right]\\
&= \mathbb{E}^n\left[\int_0^{\tau_n\wedge\sigma_m}\mathcal{E}(N^2)_{s-}\int_E \psi(X_{s-},\xi)\mathbb{1}_{\{|\psi(X_{s-},\xi)|>1\}}\nu^{\mathbb{P}}(\mathrm{d}s,\mathrm{d}\xi)\right]\\
&= \mathbb{E}^n\left[\mathcal{E}(N^2)_{\tau_n\wedge\sigma_m}\int_0^{\tau_n\wedge\sigma_m}\int_E \psi(X_{s-},\xi)\mathbb{1}_{\{|\psi(X_{s-},\xi)|>1\}}\nu^{\mathbb{P}}(\mathrm{d}s,\mathrm{d}\xi)\right] \leq \mathbb{E}^n[\mathcal{E}(N^2)_{\tau_n\wedge\sigma_m}]C_n = C_n.
\end{aligned}$$

The third equality follows by the application of Jacod and Shiryaev (2013, Lemma I.3.12). We conclude using that
$\int_0^{\tau_n\wedge\sigma_m}\int_E \psi(X_{s-},\xi)\mathbb{1}_{\{|\psi(X_{s-},\xi)|>1\}}\nu^{\mathbb{P}}(\mathrm{d}s,\mathrm{d}\xi) \leq \int_0^{\tau_n}\int_E(1-\sqrt{1+\psi(X_{s-},\xi)})^2\nu^{\mathbb{P}}(\mathrm{d}s,\mathrm{d}\xi) \leq C_n$,

by Lemma 3.4. The last equality uses the fact that $\mathcal{E}(N^2)_{\cdot\wedge\tau_n\wedge\sigma_m} \in \mathcal{M}(\mathbb{P}^n)$. Consequently,

$$\mathbb{E}^n\left[\sum_{s>0}\Delta\mathcal{E}(N^2)_{s\wedge\tau_n}\right] = \lim_{m\to+\infty}\mathbb{E}^n\left[\sum_{s\le\sigma_m}\Delta\mathcal{E}(N^2)_{s\wedge\tau_n}\right] \le C_n < \infty.$$

By Jacod and Shiryaev (2013, Theorem I.4.56 b)), this implies that $\mathcal{E}(N^2)_{\cdot\wedge\tau_n}$ is of integrable variation under $\mathbb{P}^n$ and, as a consequence, $\mathcal{E}(N^2)_{\cdot\wedge\tau_n}$ is a uniformly integrable martingale under $\mathbb{P}^n$. □

Remark 3.6. The example presented in Cheridito, Filipović and Yor (2005, Section 6) can be seen as a special case of Theorem 3.3. In the example the authors consider a one-dimensional setting ($d = 1$), in which under the measure $\mathbb{P}$, the jump component follows a compound Poisson process with $m_0^{\mathbb{P}}(\xi) = \lambda$ and $m_1^{\mathbb{P}}(\xi) = 0$.

Remark 3.7. Observe that, in the statement of Theorem 3.3, we require the compensating measures under $\mathbb{P}$ and $\mathbb{Q}$ to be absolutely continuous with respect to a certain Borel measure m. Similarly to Remark 3.1, given

$$\nu^{\mathbb{P}}(\mathrm{d}t,\mathrm{d}\xi) = \theta_0^{\mathbb{P}}(\mathrm{d}\xi) + \sum_{j=1}^{m}\theta_j^{\mathbb{P}}(\mathrm{d}\xi)X_{t-}^j\mathrm{d}t, \quad \nu^{\mathbb{Q}}(\mathrm{d}t,\mathrm{d}\xi) = \theta_0^{\mathbb{Q}}(\mathrm{d}\xi) + \sum_{j=1}^{m}\theta_j^{\mathbb{Q}}(\mathrm{d}\xi)X_{t-}^j\mathrm{d}t$$

it is enough to take $m(\mathrm{d}\xi) := \sum_{j=0}^m \theta_j^{\mathbb{P}} + \sum_{j=0}^m \theta_j^{\mathbb{Q}}$, to obtain the required structure. Clearly, this does not imply that for arbitrary compensators $\nu^{\mathbb{P}}$ and $\nu^{\mathbb{Q}}$, one can construct an equivalent change of measure. A simple counterexample arises when the jump measures have different supports. For instance, consider $\nu^{\mathbb{P}}(\mathrm{d}t,\mathrm{d}\xi) = \delta_1(\mathrm{d}\xi)$ and $\nu^{\mathbb{Q}}(\mathrm{d}t,\mathrm{d}\xi) = e^{-\xi}\mathbb{1}_{\{\xi>0\}}\,\mathrm{d}\xi$. Define the reference measure as

$$m(\mathrm{d}\xi) := \delta_1(\mathrm{d}\xi) + \mathbb{1}_{\{\xi>0\}}\,\mathrm{d}\xi.$$

Then, the compensators can be expressed with respect to m as

$$\begin{aligned}\nu^{\mathbb{P}}(\mathrm{d}t,\mathrm{d}\xi) &= \mathbb{1}_{\{1\}}(\xi)\,m(\mathrm{d}\xi)\mathrm{d}t,\\ \nu^{\mathbb{Q}}(\mathrm{d}t,\mathrm{d}\xi) &= e^{-\xi}\,\mathbb{1}_{\{\xi>0\}}\,m(\mathrm{d}\xi)\mathrm{d}t.\end{aligned}$$

Function $\mathbb{1}_{\{1\}}(\xi)$ does not satisfy the assumptions of the theorem, in particular the requirement $(m_0^{\mathbb{P}}(\xi), m_1^{\mathbb{P}}(\xi)) \in \mathbb{R}_+^2 \setminus \{(0,0)\}$.

Remark 3.8. Let us note a few properties and implications of condition (3.8):

a) Condition (3.8) is equivalent to

$$\int_E (1 - \sqrt{\psi(x,\xi) + 1})^2 \left(m_0^{\mathbb{P}}(\xi) + \sum_{j=1}^m m_j^{\mathbb{P}}(\xi) x_j \right) m(\mathrm{d}\xi) < \infty$$

for ψ as in equation (3.7).

b) We observe that for all $x > 0$

$$(1 - \sqrt{x})^2 \leq x \log(x) - x + 1.$$

This inequality implies that the integrability condition for the jump component of the change of measure required in Cheridito, Filipović and Yor (2005, Remark 2.5) is stronger and hence implies our condition.

c) Assume that for every $x > 0$

$$\int_E \left(m_0^{\mathbb{P}}(\xi) + \sum_{j=1}^m m_j^{\mathbb{P}}(\xi) x_j \right) m(\mathrm{d}\xi) < \infty \text{ and } \int_E \left(m_0^{\mathbb{Q}}(\xi) + \sum_{j=1}^m m_j^{\mathbb{Q}}(\xi) x_j \right) m(\mathrm{d}\xi) < \infty,$$

which is clearly satisfied whenever for $j = 0, \ldots, m$

$$\int_E m_j^{\mathbb{P}}(\xi) m(\mathrm{d}\xi) < \infty \text{ and } \int_E m_j^{\mathbb{Q}}(\xi) m(\mathrm{d}\xi) < \infty.$$

The integrability assumptions above imply Condition (3.8). Indeed, by Cauchy-Schwarz inequality

$$\begin{aligned}
&\int_E \left(\sqrt{m_0^{\mathbb{Q}}(\xi) + \sum_{j=1}^m m_j^{\mathbb{Q}}(\xi) x_j} - \sqrt{m_0^{\mathbb{P}}(\xi) + \sum_{j=1}^m m_j^{\mathbb{P}}(\xi) x_j} \right)^2 m(\mathrm{d}\xi) \\
&\quad = \int_E \left(m_0^{\mathbb{Q}}(\xi) + \sum_{j=1}^m m_j^{\mathbb{Q}}(\xi) x_j + m_0^{\mathbb{P}}(\xi) + \sum_{j=1}^m m_j^{\mathbb{P}}(\xi) x_j \right) m(\mathrm{d}\xi) \\
&\quad - 2 \int_E \left(\sqrt{m_0^{\mathbb{P}}(\xi) + \sum_{j=1}^m m_j^{\mathbb{P}}(\xi) x_j} \sqrt{m_0^{\mathbb{P}}(\xi) + \sum_{j=1}^m m_j^{\mathbb{P}}(\xi) x_j} \right) m(\mathrm{d}\xi) \\
&\quad \leq \int_E \left(m_0^{\mathbb{Q}}(\xi) + \sum_{j=1}^m m_j^{\mathbb{Q}}(\xi) x_j \right) m(\mathrm{d}\xi) + \int_E \left(m_0^{\mathbb{P}}(\xi) + \sum_{j=1}^m m_j^{\mathbb{P}}(\xi) x_j \right) m(\mathrm{d}\xi)
\end{aligned}$$

$$+2\left(\int_E \left(m_0^{\mathbb{Q}}(\xi)+\sum_{j=1}^m m_j^{\mathbb{Q}}(\xi)x_j\right)m(\mathrm{d}\xi)\right)^{\frac{1}{2}}\left(\int_E \left(m_0^{\mathbb{P}}(\xi)+\sum_{j=1}^m m_j^{\mathbb{P}}(\xi)x_j\right)m(\mathrm{d}\xi)\right)^{\frac{1}{2}}<\infty.$$

We now state and prove the following theorem, which shows that if a locally equivalent measure $\mathbb{Q}$ maintains the affine structure, then its associated density Z must satisfy Property B up to an orthogonal component. This theorem can be interpreted as a converse result, providing the necessary conditions for Z.

Theorem 3.9. *Let X be the process as defined in equation (3.1) on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathbb{Q}$ be a probability measure on $(\Omega, \mathcal{F})$ with $\mathbb{Q} \sim^{\mathrm{loc}} \mathbb{P}$. If $\mathbb{Q}$ preserves the affine structure of X in the sense of Definition 3.2 then its density process Z with respect to $\mathbb{P}$ can be represented as*

$$Z_t = \mathcal{E}\left(\int_0^t \phi(X_{s-})^\top \mathrm{d}W_s^{\mathbb{P}} + \int_0^t \int_E \psi(X_{s-},\xi)\cdot(\mu(\mathrm{d}s,\mathrm{d}\xi)-\nu^{\mathbb{P}}(\mathrm{d}s,\mathrm{d}\xi)) + N'\right)$$

where ϕ and ψ satisfy respectively equations (3.6), (3.7) for some parameters $\left(b^{\mathbb{Q}} + A^{\mathbb{Q}}x, \Sigma_0 + \sum_{j=1}^m \Sigma_j x_j, \left(m_0^{\mathbb{Q}}(\xi) + \sum_{j=1}^m m_j^{\mathbb{Q}}(\xi)x_j\right) m(\mathrm{d}\xi)\right)$ which satisfy Assumption A, N' is a $\mathbb{P}$-local martingale with $N'_0 = 0, \langle N', W^{\mathbb{P}}\rangle = 0$ and $\Delta N'$ orthogonal with respect to μ in the sense of Jacod and Shiryaev (2013, Lemma III.4.24). In particular, up to an orthogonal component, Z must satisfy Property B. Moreover, the following condition holds

$$\int_E \left(\sqrt{m_0^{\mathbb{Q}}(\xi)+\sum_{j=1}^m m_j^{\mathbb{Q}}(\xi)X_{t-}^j} - \sqrt{m_0^{\mathbb{P}}(\xi)+\sum_{j=1}^m m_j^{\mathbb{P}}(\xi)X_{t-}^j}\right)^2 m(\mathrm{d}\xi) < \infty$$

for almost every $t \in [0, T]$ and $\mathbb{P}$-almost surely.

Proof. Let $\mathbb{P}$ and $\mathbb{Q}$ be two locally equivalent probability measures such that $\mathbb{Q}$ preserves the affine structure of X. Since the measures are locally equivalent, the density process

$$Z_t := \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\Big|_{\mathcal{F}_t}$$

is a strictly positive $\mathbb{P}$-martingale, see Jacod and Shiryaev (2013, Proposition III.3.5 a)). Due to Jacod and Shiryaev (2013, Theorem II.8.3), the stochastic integral $L_t = \int_0^t (Z_{s-})^{-1} \mathrm{d}Z_s$ is well-defined as a local martingale with $L_0 = 0$. Applying the martingale representation theorem, see Jacod and Shiryaev (2013, Lemma III.4.24), L can be written as a stochastic integral of the type

$$L_t = \int_0^t \Phi_s^\top \mathrm{d}W_s^{\mathbb{P}} + \int_0^t \int_E \Psi_s(\xi) \cdot (\mu(\mathrm{d}s, \mathrm{d}\xi) - \nu^{\mathbb{P}}(\mathrm{d}s, \mathrm{d}\xi)) + N',$$

for a predictable process $\Phi_s \in L^2_{\mathrm{loc}}(W^{\mathbb{P}})$ and a predictable function $\Phi_s(\xi) \in G_{\mathrm{loc}}(\mu)$. Since $Z_t = 1 + \int_0^t Z_{s-}\mathrm{d}L_s$, the process Z_t can be represented as $Z_t = \mathcal{E}(L_t)$. Due to Girsanov's theorem, the process $W^{\mathbb{Q}}$ defined as:

$$W_t^{\mathbb{Q}} := W_t^{\mathbb{P}} - \int_0^t \frac{1}{Z_{s-}} \mathrm{d}\langle Z, W^{\mathbb{P}}\rangle_s = W_t^{\mathbb{P}} - \int_0^t \Phi_s \mathrm{d}s$$

is a $\mathbb{Q}$-Brownian motion. Moreover, the characteristics of X under $\mathbb{Q}$ are given by, see Jacod and Shiryaev (2013, Theorem III.3.24):

$$b^{\mathbb{Q}} + A^{\mathbb{Q}} X_{t-} = b^{\mathbb{P}} + A^{\mathbb{P}} X_{t-} + \sigma(X_{t-})\Phi_{t-} \tag{3.15}$$

$$(m_0^{\mathbb{Q}}(\xi) + \sum_{j=1}^m m_j^{\mathbb{Q}}(\xi) X_{t-}^j) = (\Psi_{t-}(\xi) + 1)(m_0^{\mathbb{P}}(\xi) + \sum_{j=1}^m m_j^{\mathbb{P}}(\xi) X_{t-}^j). \tag{3.16}$$

The only functions Φ, Ψ which satisfies equations (3.15), (3.16) are $\Phi_{t-} = \phi(X_{t-}), \Psi_{t-}(\xi) = \psi(X_{t-}, \xi)$ where ϕ, ψ are those in equations (3.6), (3.7). The following integrability condition must be satisfied

$$\int_0^T \int_E \left(1 - \sqrt{1 + \psi(X_{t-}, \xi)}\right)^2 \nu^{\mathbb{P}}(\mathrm{d}t, \mathrm{d}\xi) < \infty$$

almost surely since it is necessary and sufficient condition for the well-posedness of the stochastic integral with respect to the compensated jump measure, see Jacod and Shiryaev (2013, Theorem II.1.33, d)). Due to the preservation of the affine structure, it reads as the condition in the statement. □

Theorem 3.3 shows that, given an affine SDE X satisfying certain boundary non-attainment conditions under a reference probability measure $\mathbb{P}$, one can construct a locally equivalent measure $\mathbb{Q}$ through a density Z which satisfies Property B. Under $\mathbb{Q}$, X is affine. Conversely, Theorem 3.9 establishes that any locally equivalent measure under which X maintains its affine structure is associated to a density Z which necessarily satisfies Property B. In this sense, we fully characterize all structure-preserving measure transformations for affine jump-diffusions.

Remark 3.10. We highlight that condition (3.8) as stated in Theorem 3.3 do not follows as a direct consequence of Theorem 3.9. The latter theorem only implies a weaker version of the condition. The two corresponds when $m_j^{\mathbb{P},\mathbb{Q}} \equiv 0$, for $j \geq 1$ or when X has full support on its state space.

3.4 Comparison with the literature

In this section, we discuss the relation of our results with the existing literature.

In Palmowski and Rolski 2002, the authors develop a general framework for exponential changes of measure for Markov processes. Their construction relies on assuming that the associated density process is defined so as to be a martingale. This allows to introduce a new probability measure under which the process remains Markov, under mild regularity conditions. By contrast, in our work, the martingale property of the density process is proved directly for the specific class of measure changes considered (see Lemma 3.5), rather than being assumed. Moreover, our analysis is conducted within a particular subclass of Markov processes, namely affine processes.

A key result in the literature is provided in Cheridito, Filipović and Yor (2005, Theorem 2.4), where the authors study two jump-diffusion processes, which are generally not semimartingales, and provide sufficient conditions for their distributions to be equivalent or absolutely continuous. The problem is formulated in a general setting, whereas we focus specifically on processes with an affine structure. In particular, in Cheridito, Filipović and Yor (2005, Remark 2.5), they ask for the compensating measure ν and the density process ψ to satisfy a condition of the following kind

$$\int_E (\psi(x,\xi)\log(\psi(x,\xi)) - \psi(x,\xi) + 1)\nu^{\mathbb{P}}(x, \mathrm{d}\xi) < M_n,$$

for x in a certain finite set of the state space. In Remark 3.8 b), our condition on the compensating measure, inequality (3.8), is in fact a weaker integrability conditions than the one in Cheridito, Filipović and Yor (2005). Moreover, we believe that our inequality is easier to verify in models with sufficient structure, see for example Remark 3.8 c).

As noted in the introduction, an initial result on structure-preserving transformations for affine processes is provided in Cheridito, Filipović and Kimmel (2007). In particular, in Cheridito, Filipović and Kimmel (2007, Theorem 1), the authors study stable measure transformations of a class of affine diffusion processes which do not attain the boundary. Their result coincides with Theorem 3.3 in the specific case of an affine diffuse process. Their main contribution lies in analyzing a class of market prices of risk that are inversely proportional to the square root of the state variable. Theorem 3.3 employs similar techniques to those in Cheridito, Filipović and Kimmel (2007, Theorem 1).

A closely related work to ours is Fontana (2012, Chapter 2), where the author studies an intensity-based model involving an affine diffusion process X and a random default time whose intensity depends affinely on X. The work characterizes the class of all locally equivalent probability measures that preserve the affine structure of a reduced-form credit risk model. More precisely, necessary and sufficient conditions on the density process are provided to ensure that the default time remains a doubly stochastic random time under both measures, and that the diffusion driving the default intensity maintains its affine structure under both. This represents an initial extension of Cheridito, Filipović and Kimmel (2007) to models with jumps, although intensity-based models can be seen as simplified cases of general affine jump-diffusion frameworks. In Fontana (2012, Chapter 2), the author also investigates some immersion properties, which are of particular interest in the case of credit risk.

Finally, we mention Kallsen and Muhle-Karbe (2010), where the authors investigate the conditions under which the stochastic exponential of a multivariate affine process is a martingale. This problem is closely related to ours, as such stochastic exponentials can be used to define changes of measure. In fact, in Kallsen and Muhle-Karbe (2010, Theorem 4.1), they examine when two parameter sets of affine processes correspond to the same process under equivalent probability measures. This problem is strongly connected to ours, and is formulated in a general affine jump-diffusion setting. The authors also allows for time-inhomogeneous coefficients and the presence of truncation functions. In Kallsen and Muhle-Karbe (2010, Corollary 4.2), they focus on the time-homogeneous case and derive integrability conditions on the compensator that are analogous to inequality (3.8). It is crucial to point out that, the transformations identified by Kallsen and Muhle-Karbe (2010), are not all the admissible ones. For affine processes that may reach the boundary, the only admissible transformations are those given by stochastic exponentials of affine processes. In contrast, under our assumption of boundary non-attainment, Assumption A, we can also consider changes of measure involving the inverse of the volatility or the presence of the state process in the denominator. With Theorem 3.9, we indeed prove that all the admissible stable transformations are those identified by Property B.

3.5 Examples and applications

Example 3.11 (One-dimensional case)**.** As a first example, we consider the one-dimensional non-negative affine case, where under a certain measure $\mathbb{P}$,

$$X_t = X_0 + \int_0^t (b^{\mathbb{P}} + a^{\mathbb{P}} X_s)\,\mathrm{d}s + \int_0^t \sigma\sqrt{X_s}\,\mathrm{d}W_s^{\mathbb{P}} + \int_0^t \int_0^\infty \xi\,\mu(\mathrm{d}s,\mathrm{d}\xi), \quad t \geq 0,$$

with compensating measure

$$\nu^{\mathbb{P}}(\mathrm{d}t,\mathrm{d}\xi) = (m_0^{\mathbb{P}}(\xi) + m_1^{\mathbb{P}}(\xi)X_{t-})m(\mathrm{d}\xi)\,\mathrm{d}t.$$

In the one-dimensional non-negative case Assumption A reads: $\sigma > 0, b^{\mathbb{P}} \geq \frac{1}{2}\sigma^2, (m_0(\xi), m_1(\xi)) \in \mathbb{R}_+^2 \setminus \{(0,0)\} \forall\xi$ and for all $x > 0$

$$\int_0^\infty (1 \wedge |\xi|)(m_0(\xi) + m_1(\xi)x)m(\mathrm{d}\xi) < \infty.$$

In the light of Theorem 3.3 and Theorem 3.9, all the stable transformations which maintain the affine structure in the sense of Definition 3.2 are given by

$$\left.\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\right|_{\mathcal{F}_t} = Z_t = \mathcal{E}\left(\int_0^t \phi(X_{s-})\mathrm{d}W_s^{\mathbb{P}} + \int_0^t \int_0^\infty \psi(X_{s-},\xi)\cdot(\mu(\mathrm{d}s,\mathrm{d}\xi) - \nu^{\mathbb{P}}(\mathrm{d}s,\mathrm{d}\xi))\right)$$

where

$$\phi(X_{s-}) = \frac{b^{\mathbb{Q}} - b^{\mathbb{P}} + (a^{\mathbb{Q}} - a^{\mathbb{P}})X_{s-}}{\sigma\sqrt{X_{s-}}}$$
$$\psi(X_{s-},\xi) = \frac{m_0^{\mathbb{Q}}(\xi) + m_1^{\mathbb{Q}}(\xi)X_{s-}}{m_0^{\mathbb{P}}(\xi) + m_1^{\mathbb{P}}(\xi)X_{s-}} - 1,$$

given that for all $x > 0$,

$$\int_0^\infty \left(\sqrt{m_0^{\mathbb{Q}}(\xi) + m_1^{\mathbb{Q}}(\xi)x} - \sqrt{m_0^{\mathbb{P}}(\xi) + m_1^{\mathbb{P}}(\xi)x}\right)^2 m(\mathrm{d}\xi) < \infty.$$

On the other hand, if we consider a general affine process which can take values in all $\mathbb{R}$, the only possible structure is the following

$$X_t = X_0 + \int_0^t (b^{\mathbb{P}} + a^{\mathbb{P}} X_s)\,\mathrm{d}s + \int_0^t \sigma\mathrm{d}W_s^{\mathbb{P}} + \int_0^t \int_0^\infty \xi\,\mu(\mathrm{d}s,\mathrm{d}\xi), \quad t \geq 0,$$

with compensating measure given by $\nu^{\mathbb{P}}(\mathrm{d}t, \mathrm{d}\xi) = m_0^{\mathbb{P}}(\xi)m(\mathrm{d}\xi)\,\mathrm{d}t$. Indeed, being X potentially negative it cannot have a role in the volatility nor intensity dynamics. In this case the only admissible transformations are those given by

$$\left.\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\right|_{\mathcal{F}_t} = Z_t = \mathcal{E}\left(\int_0^t \phi(X_{s-})\mathrm{d}W_s^{\mathbb{P}} + \int_0^t \int_0^\infty \psi(X_{s-},\xi)\cdot(\mu(\mathrm{d}s,\mathrm{d}\xi) - \nu^{\mathbb{P}}(\mathrm{d}s,\mathrm{d}\xi))\right)$$

where

$$\phi(X_{s-}) = \frac{b^{\mathbb{Q}} - b^{\mathbb{P}} + (a^{\mathbb{Q}} - a^{\mathbb{P}})X_{s-}}{\sigma}, \quad \psi(X_{s-},\xi) = \frac{m_0^{\mathbb{Q}}(\xi)}{m_0^{\mathbb{P}}(\xi)} - 1,$$

given that,

$$\int_0^\infty \left(\sqrt{m_0^{\mathbb{Q}}(\xi)} - \sqrt{m_0^{\mathbb{P}}(\xi)}\right)^2 m(\mathrm{d}\xi) < \infty.$$

Example 3.12 (Hawkes intensity)**.** A direct application of the one dimensional case is the Hawkes intensity. Let λ be the intensity of a marked Hawkes process N under the measure $\mathbb{P}$

$$\mathrm{d}\lambda_t = (b^{\mathbb{P}} + a^{\mathbb{P}}\lambda_t)\mathrm{d}t + \sum_i^{N_t} \eta_i,$$

for $b^{\mathbb{P}} > 0$, $a^{\mathbb{P}} < 0$ and where $(\eta_i)_i$ i.i.d. positive random variable, $\eta_i \sim m_1^{\mathbb{P}}(\xi)m(\mathrm{d}\xi)$, where m is a measure on $\mathbb{R}_{++}$ and $m_1^{\mathbb{P}}(\xi) > 0$. We ask for the usual integrability conditions of Assumption (A5)(c). The intensity λ is an affine process since its drift is affine and its compensator is given by

$$\nu^{\mathbb{P}}(\mathrm{d}t, \mathrm{d}\xi) = \lambda_{t-} m_1^{\mathbb{P}}(\xi)m(\mathrm{d}\xi)\mathrm{d}t.$$

Exploiting Theorem 3.3, we can prove that it exists an equivalent probability measure $\mathbb{Q}$ under which λ is a jump process having the same drift and compensator

$$\nu^{\mathbb{Q}}(\mathrm{d}t, \mathrm{d}\xi) = (m_0^{\mathbb{Q}}(\xi) + m_1^{\mathbb{Q}}(\xi)\lambda_{t-})m(\mathrm{d}\xi),$$

whenever condition (3.8) is satisfied for every $x > 0$

$$\int_0^\infty \left(\sqrt{m_0^{\mathbb{Q}}(\xi) + m_1^{\mathbb{Q}}(\xi)x} - \sqrt{m_1^{\mathbb{P}}(\xi)x}\right)^2 m(\mathrm{d}\xi) < \infty.$$

and the parameters for $\mathbb{Q}$ satisfy Assumption A. The measure $\mathbb{Q}$ is constructed through the following density process

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\Big|_{\mathcal{F}_t} = Z_t = \mathcal{E}\left(\int_0^t \int_0^\infty \psi(\lambda_{s-}, \xi)\cdot(\mu(\mathrm{d}s, \mathrm{d}\xi) - \nu^{\mathbb{P}}(\mathrm{d}s, \mathrm{d}\xi))\right),$$
$$\psi(X_{s-}, \xi) = \frac{m_0^{\mathbb{Q}}(\xi) + m_1^{\mathbb{Q}}(\xi)\lambda_{s-}}{m_1^{\mathbb{P}}(\xi)\lambda_{s-}} - 1.$$

In particular, Remark 3.8 implies that one can transform a Hawkes intensity with a given mark distribution into another Hawkes intensity with a different mark distribution (with $m_0^{\mathbb{Q}} \equiv 0$), provided both are $m_1^{\mathbb{P}}, m_1^{\mathbb{Q}}$ are probability density functions and share the same support. Similarly, taking $m_1^{\mathbb{Q}} \equiv 0$, we can transform an Hawkes intensity to a compound Poisson with drift. Notice that no admissible change of measure can modify the drift.

Example 3.13 (Jump-diffusion)**.** We consider a simplified version of our affine model, i.e. a jump-diffusion model as the one introduced in Duffie, Pan et al. (2000). The authors focus on SDE of the following kind

$$\mathrm{d}X_t = (b^{\mathbb{P}} + A^{\mathbb{P}}X_t)\mathrm{d} + \sigma(X_t)\mathrm{d}W_t^{\mathbb{P}} + \mathrm{d}J_t,$$

where J is a pure jump process whose jumps have a fixed distribution m on $\mathbb{R}^n$ and arrive with intensity $\lambda(X_t) = \gamma^{\mathbb{P}} + \Gamma^{\mathbb{P}^\top} X_t$, for $(\gamma^{\mathbb{P}}, \Gamma^{\mathbb{P}}) \in \mathbb{R} \times \mathbb{R}^n$. We can write

$$J_t = \int_0^t \int_{\mathbb{R}^d} \xi\mu(\mathrm{d}t, \mathrm{d}\xi),$$

with μ random measure having compensator $\nu^{\mathbb{P}}(\mathrm{d}t, \mathrm{d}\xi) = (\gamma^{\mathbb{P}} + \Gamma^{\mathbb{P}^\top} X_{t-})m(\mathrm{d}\xi)\mathrm{d}t$. We are in the case where $m_0^{\mathbb{P}}(\xi) = \gamma^{\mathbb{P}}$, $m_j^{\mathbb{P}}(\xi) = \Gamma_j^{\mathbb{P}}$ for every $j \in \{1, \dots, m\}$. Assuming that X takes values in $E = \mathbb{R}_{++}^m \times \mathbb{R}^{d-m}$, Assumptions (A1)-(A4) read as usual, while we ask that $\gamma^{\mathbb{P}} \in \mathbb{R}_+$, $\Gamma_j^{\mathbb{P}} \in \mathbb{R}_+$ for $j \in \{0, \dots, m\}$ and $\Gamma_j^{\mathbb{P}} \equiv 0$ for $j \in \{m+1, \dots, d\}$. Moreover, $\gamma^{\mathbb{P}} + \sum_{j=1}^m \Gamma_j^{\mathbb{P}} > 0$. We ask for m to be a Radon measure on E such that

$$\int_E (1 \wedge ||\xi||)m(\mathrm{d}\xi) < \infty. \tag{3.17}$$

The changes of measure which maintain the affine structure inside the class of jump-diffusion processes are those given by

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\Big|_{\mathcal{F}_t} = Z_t = \mathcal{E}\left(\int_0^t \phi(X_{s-})^\top \mathrm{d}W_s^{\mathbb{P}} + \int_0^t \int_E \psi(X_{s-}, \xi)\cdot(\mu(\mathrm{d}s, \mathrm{d}\xi) - \nu^{\mathbb{P}}(\mathrm{d}s, \mathrm{d}\xi))\right),$$
$$\phi(X_{s-}) = \sigma(X_{s-})^{-1}(b^{\mathbb{Q}} - b^{\mathbb{P}} + (A^{\mathbb{Q}} - A^{\mathbb{P}})X_{s-}),$$

$$\psi(X_{s-}, \xi) = \frac{\gamma^{\mathbb{Q}} + \sum_{j=1}^{m} \Gamma_j^{\mathbb{Q}} X_{s-}^j}{\gamma^{\mathbb{P}} + \sum_{j=1}^{m} \Gamma_j^{\mathbb{P}} X_{s-}^j} - 1,$$

where

$$\nu^{\mathbb{Q}}(\mathrm{d}t, \mathrm{d}\xi) = \left(\gamma^{\mathbb{Q}} + \sum_{j=1}^{m} \Gamma_j^{\mathbb{Q}} x_j\right) m(\mathrm{d}\xi),$$

and Assumption A are satisfied also for the set of parameters under $\mathbb{Q}$. In this case condition (3.8) reads as

$$\int_E \left(\sqrt{(\gamma^{\mathbb{Q}} + \Gamma^{\mathbb{Q}^\top} x)} - \sqrt{(\gamma^{\mathbb{P}} + \Gamma^{\mathbb{P}^\top} x)}\right)^2 m(\mathrm{d}\xi) < \infty,$$

which is clearly finite whenever $\int_E m(\mathrm{d}\xi) < \infty$, in the light of Remark 3.8. Notice that if the measure m is finite, then also inequality (3.17) holds.

Example 3.14 (Hawkes processes construction)**.** We introduce a change of measure technique used to construct Hawkes processes from a Poisson process, see e.g. Bernis and Scotti (2020). This transformation can also be interpreted as a particular case of Example 3.13.

Let N be a compound Poisson process with constant intensity γ under the probability measure $\mathbb{P}$, and let λ be a process that shares the same jump times as N. The marks of the two processes follow a joint distribution m. The two dimensional process (N, λ) is affine and can be written as

$$\begin{aligned}
\mathrm{d}N_t &= \int_{\mathbb{R}^2_{++}} \xi_1 \mu(\mathrm{d}t, \mathrm{d}\xi), \\
\mathrm{d}\lambda_t &= (b^{\mathbb{P}} + a^{\mathbb{P}} \lambda_t)\,\mathrm{d}t + \int_{\mathbb{R}^2_{++}} \xi_2 \mu(\mathrm{d}t, \mathrm{d}\xi)
\end{aligned}$$

where μ is a jump measure and the associated compensating measure is given by $\nu^{\mathbb{P}}(\mathrm{d}t, \mathrm{d}\xi) = m(\mathrm{d}\xi)\gamma \mathrm{d}t$.

Our aim is to characterize a locally equivalent measure $\mathbb{Q}$ under which the couple (N, λ) has the distribution of a Hawkes process along with its corresponding intensity, meaning $\nu^{\mathbb{Q}}(\mathrm{d}t, \mathrm{d}\xi) = m(\mathrm{d}\xi)\lambda_{t-}\mathrm{d}t$. Condition (3.8) reads as

$$\int_{\mathbb{R}^2_{++}} \left(\sqrt{x_2} - \sqrt{\gamma}\right)^2 m(\mathrm{d}\xi) = \int_{\mathbb{R}^2_{++}} (x_2 + \gamma - 2\sqrt{x_2}\sqrt{\gamma}) m(\mathrm{d}\xi) < \infty,$$

and holds if the measure m is finite, i.e. $\int_{\mathbb{R}^2_{++}} m(\mathrm{d}\xi) < \infty$. This implies that also Assumption (A5)(c) holds. The corresponding change of measure is given by

$$\begin{aligned}
\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\Big|_{\mathcal{F}_t} = Z_t &= \mathcal{E}\left(\int_0^t \int_{\mathbb{R}^2_{++}} \left(\frac{\lambda_{s-}}{\gamma} - 1\right) \cdot (\mu(\mathrm{d}s, \mathrm{d}\xi) - \gamma m(\mathrm{d}\xi)\mathrm{d}s)\right) \\
&= \exp\left(\int_0^t \int_{\mathbb{R}^2_{++}} -\left(\frac{\lambda_{s-}}{\gamma} - 1\right) \cdot \gamma m(\mathrm{d}\xi)\mathrm{d}s + \int_0^t \int_{\mathbb{R}^2_{++}} \log\left(\frac{\lambda_{s-}}{\gamma}\right) \mu(\mathrm{d}s, \mathrm{d}\xi)\right) \\
&= \exp\left(\int_0^t \int_{\mathbb{R}^2_{++}} -(\lambda_{s-} - \gamma) \cdot m(\mathrm{d}\xi)\mathrm{d}s\right) \exp\left(\int_0^t \int_{\mathbb{R}^2_{++}} \log\left(\frac{\lambda_{s-}}{\gamma}\right) \mu(\mathrm{d}s, \mathrm{d}\xi)\right) \\
&= \exp\left(-\int_0^t (\lambda_{s-} - \gamma)\, \mathrm{d}s \cdot \int_{\mathbb{R}^2_{++}} m(\mathrm{d}\xi)\right) \cdot \prod_{T_i < t} \frac{\lambda_{T_i -}}{\gamma}.
\end{aligned}$$

This transformation is equivalent to the one presented in Bernis and Scotti (2020, Theorem 3.6).

Example 3.15 (Intensity-based models)**.** Intensity-based models are a key applications of affine models in credit risk. We consider an example of stable transformation in this context, referring to Fontana (2012, Chapter 2). Let X be a d-dimensional diffusive process in $E = \mathbb{R}^m_{++} \times \mathbb{R}^{d-m}$, which satisfies Assumption A. Let τ be a random default time and assume it is a doubly stochastic random time, having intensity $\lambda^{\mathbb{P}}$ which depends on X

$$\lambda^{\mathbb{P}}_t = \gamma^{\mathbb{P}} + \Gamma^{\mathbb{P}\top} X_t,$$

for $\gamma^{\mathbb{P}} \in \mathbb{R}, \Gamma^{\mathbb{P}} \in \mathbb{R}^d$. We denote by H the default indicator process $H_t := \mathbb{1}_{\{\tau \le t\}}$ for $t \ge 0$. The pair (X, H) has affine structure, in particular τ is a pure jump component, and the couple has compensating measure given by

$$\nu^{\mathbb{P}}(\mathrm{d}t, \mathrm{d}\xi) = (\gamma^{\mathbb{P}} + \sum_{j=1}^m \Gamma^{\mathbb{P}}_j X^j_{t-})\delta_{(0,\dots,0,1)}(\mathrm{d}\xi_1, \dots, \mathrm{d}\xi_d, \mathrm{d}\xi_{d+1}).$$

In this case, Assumption A reads as those in Fontana (2012, Definition 2.2.5, (ii)), in particular we ask for $\gamma^{\mathbb{P}} \in \mathbb{R}_+, \Gamma^{\mathbb{P}} \in \mathbb{R}^d_+$, with $\Gamma^{\mathbb{P}}_j = 0$ for $j \ge m+1$ and $\gamma^{\mathbb{P}} + \sum_{j=1}^m \Gamma^{\mathbb{P}}_j > 0$. We observe that the following integrability condition always holds

$$\int_{E \times \mathbb{R}_{++}} \delta_{(0,\dots,0,1)}(\mathrm{d}\xi_1, \dots, \mathrm{d}\xi_d, \mathrm{d}\xi_{d+1}) < \infty,$$

thus also condition (3.8) is always satisfied

$$\int_{E\times\mathbb{R}_{++}}\left(\sqrt{(\gamma^{\mathbb{Q}}+\Gamma^{\mathbb{Q}\top}x)}-\sqrt{(\gamma^{\mathbb{P}}+\Gamma^{\mathbb{P}\top}x)}\right)^2\delta_{(0,\dots,0,1)}(\mathrm{d}\xi_1,\dots,\mathrm{d}\xi_d,\mathrm{d}\xi_{d+1})<\infty.$$

Stable measures transformations are those identified by Fontana (2012, Theorem 2.3.12), i.e.

$$\begin{aligned}
\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\Big|_{\mathcal{F}_t} &= Z_t=\mathcal{E}\left(\int_0^t\phi(X_{s-})^\top\mathrm{d}W_s^{\mathbb{P}}+\int_0^t\psi(X_{s-})\cdot\mathrm{d}M_s^{\mathbb{P}}\right),\\
\phi(X_{s-}) &= \sigma(X_{s-})^{-1}(b^{\mathbb{Q}}-b^{\mathbb{P}}+(A^{\mathbb{Q}}-A^{\mathbb{P}})X_{s-}),\\
\psi(X_{s-}) &= \frac{\gamma^{\mathbb{Q}}+\sum_{j=1}^m\Gamma_j^{\mathbb{Q}}X_{s-}^j}{\gamma^{\mathbb{P}}+\sum_{j=1}^m\Gamma_j^{\mathbb{P}}X_{s-}^j}-1,
\end{aligned}$$

where $M_t^{\mathbb{P}}:=H_t-\int_0^{t\wedge\tau}\lambda_u^{\mathbb{P}}\mathrm{d}u$.

Example 3.16 (α-stable subordinator)**.** We now investigate an equivalent change of measure presented in Chen and Filipović (2005). The authors consider a multi-dimensional process whose compensating measure is given by

$$\nu^{\mathbb{P}}(\mathrm{d}t,\mathrm{d}\xi)=m_0^{\mathbb{P}}(\mathrm{d}\xi)\mathrm{d}t=\frac{\theta}{\Gamma(1-\theta)}\frac{1}{\xi^{1+\theta}}\mathrm{d}\xi\mathrm{d}t,$$

where $\theta\in(0,1)$. We observe that $m_0^{\mathbb{P}}$ satisfies condition (3.3), indeed

$$\int_0^\infty(1\wedge\xi)\frac{\theta}{\Gamma(1-\theta)}\frac{1}{\xi^{1+\theta}}\mathrm{d}\xi=\int_0^1\frac{\theta}{\Gamma(1-\theta)}\frac{\xi}{\xi^{1+\theta}}\mathrm{d}\xi+\int_1^\infty\frac{\theta}{\Gamma(1-\theta)}\frac{1}{\xi^{1+\theta}}\mathrm{d}\xi<\infty.$$

The authors of the paper explore whether it is possible to modify the parameter θ of the jump distribution via an equivalent change of measure. They conjecture that this is not possible, since the proposed integrand fails to satisfy the integrability conditions required by the main theorem of Cheridito, Filipović and Yor (2005). We prove that their conjecture is indeed correct: condition (3.8) is not satisfied in this case, and in light of Theorem 3.9, this shows that such a change of measure is impossible. Let

$$m_0^{\mathbb{Q}}(\mathrm{d}\xi)=\frac{\tilde{\theta}}{\Gamma(1-\theta)}\frac{1}{\xi^{1+\tilde{\theta}}}\mathrm{d}\xi,$$

where $\tilde{\theta} \in (0,1)$. Assume without loss of generality that $\tilde{\theta} > \theta$.

$$\int_0^\infty \left(\sqrt{m_0^{\mathbb{Q}}(\xi)} - \sqrt{m_0^{\mathbb{P}}(\xi)}\right)^2 \mathrm{d}\xi = \int_0^\infty \left(\tilde{C}\xi^{-\frac{1}{2}-\frac{\tilde{\theta}}{2}} - C\xi^{-\frac{1}{2}-\frac{\theta}{2}}\right)^2 \mathrm{d}\xi$$

$$= \int_0^\infty \left(\tilde{C}\xi^{-\frac{1}{2}-\frac{\tilde{\theta}}{2}}\left(1 - \frac{C}{\tilde{C}}\xi^{\frac{\tilde{\theta}}{2}-\frac{\theta}{2}}\right)\right)^2 \mathrm{d}\xi = \int_0^\infty \tilde{C}^2\xi^{-1-\tilde{\theta}}\left(1 - \frac{C}{\tilde{C}}\xi^{\frac{\tilde{\theta}}{2}-\frac{\theta}{2}}\right)^2 \mathrm{d}\xi.$$

Near zero, the integrand behaves like $\frac{1}{\xi^{1+\tilde{\theta}}}$, which is not integrable, thereby causing the integral to diverge.

3.6 Conclusions

In this work, we establish a criterion to characterize all structure-preserving measure changes within the class of affine processes. In Theorem 3.3, we show that our criterion ensures that the measure transformation preserves the affine structure, while in Theorem 3.9, we demonstrate that any transformation that maintains the affine structure must satisfy our conditions. Condition (3.8) plays a central role in establishing the sufficient conditions, while Theorem 3.9 shows that an analogous, though weaker, condition is also necessary. Compared with the existing literature, our findings offer a more comprehensive view on stable measure transformations in general affine jump-diffusion models, providing a complete characterization not fully established in prior works. Moreover, we illustrate the practical relevance of our approach through various examples, allowing the exploitation of structure-preserving transformations in various applications. Future research directions include an extension of the considered affine setting to allow for infinite-activity jumps, thereby relaxing some integrability conditions of the random measure in Assumption A. Furthermore, it would be meaningful to develop a similar stable measure characterization to include jumps occurring at predictable or predetermined times, as in the case of affine semimartingales as considered by Keller-Ressel, Schmidt et al. (2019).

References

Aït-Sahalia, Y., Cacho-Diaz, J. and Laeven, R. J. (2015). 'Modeling financial contagion using mutually exciting jump processes'. In: *Journal of Financial Economics* 117.3, pp. 585–606.

Alfonsi, A. (2015). *Affine Diffusions and Related Processes: Simulation, Theory and Applications*. Springer Cham.

Amann, H. (2011). *Ordinary Differential Equations: An Introduction to Nonlinear Analysis*. Vol. 13. Walter de Gruyter.

Andersen, L. B. and Piterbarg, V. V. (2007). 'Moment explosions in stochastic volatility models'. In: *Finance and Stochastics* 11, pp. 29–50.

Awiszus, K., Knispel, T., Penner, I., Svindland, G., Voß, A. and Weber, S. (2023). 'Modeling and pricing cyber insurance: idiosyncratic, systematic, and systemic risks'. In: *European Actuarial Journal* 13.1, pp. 1–53.

Bacry, E., Mastromatteo, I. and Muzy, J.-F. (2015). 'Hawkes Processes in Finance'. In: *Market Microstructure and Liquidity* 1.01, p. 1550005.

Baldwin, A., Gheyas, I., Ioannidis, C., Pym, D. and Williams, J. (2017). 'Contagion in cyber security attacks'. In: *Journal of the Operational Research Society* 68.7, pp. 780–791.

Balzano, M. and Marzi, G. (2025). 'At the cybersecurity frontier: key strategies and persistent challenges for business leaders'. In: *Strategic Change* 34.2, pp. 181–192.

Bates, D. S. (1996). 'Jumps and stochastic volatility: exchange rate processes implicit in Deutsche mark options'. In: *The Review of Financial Studies* 9.1, pp. 69–107.

Beiglböck, M., Henry-Labordère, P. and Penkner, F. (2013). 'Model-independent bounds for option prices: a mass transport approach.' In: *Finance and Stochastics* 17, pp. 477–501.

Bensoussan, A. and Chevalier-Roignant, B. (2024). 'Stochastic control for diffusions with self-exciting jumps: an overview'. In: *Mathematical Control and Related Fields* 14.4, pp. 1452–1476.

Bentley, M., Stephenson, A., Toscas, P. and Zhu, Z. (2020). 'A multivariate model to quantify and mitigate cybersecurity risk'. In: *Risks* 8.2, p. 61.

Bernis, G. and Scotti, S. (2020). 'Clustering effects via Hawkes processes'. In: *From Probability to Finance: Lecture Notes of BICMR Summer School on Financial Mathematics*. Springer, pp. 145–181.

Bessy-Roland, Y., Boumezoued, A. and Hillairet, C. (2021). 'Multivariate Hawkes process for cyber insurance'. In: *Annals of Actuarial Science* 15.1, pp. 14–39.

Bielecki, T. R., Jakubowski, J. and Niewęgłowski, M. (2022). 'Construction and simulation of generalized multivariate Hawkes processes'. In: *Methodology and Computing in Applied Probability* 24, pp. 2865–2896.

Bielecki, T. R., Jakubowski, J. and Niewęgłowski, M. (2023). 'Multivariate Hawkes processes with simultaneous occurrence of excitation events coming from different sources'. In: *Stochastic Models* 39.3, pp. 537–565.

Biffis, E. (2005). 'Affine processes for dynamic mortality and actuarial valuations'. In: *Insurance: Mathematics and economics* 37.3, pp. 443–468.

Björk, T. (2009). *Arbitrage Theory in Continuous Time*. Oxford University Press.

Blei, S. and Engelbert, H.-J. (2009). 'On exponential local martingales associated with strong Markov continuous local martingales'. In: *Stochastic processes and their Applications* 119.9, pp. 2859–2880.

Boumezoued, A., Cherkaoui, Y. and Hillairet, C. (2023). 'Cyber risk modeling using a two-phase Hawkes process with external excitation'. In: *arXiv preprint arXiv:2311.15701*.

Brachetta, M., Callegaro, G., Ceci, C. and Sgarra, C. (2024). 'Optimal reinsurance via BSDEs in a partially observable model with jump clusters'. In: *Finance and Stochastics* 28.2, pp. 453–495.

Brémaud, P. (1981). *Point Processes and Queues: Martingale Dynamics*. Springer.

Brignone, R., Gonzato, L. and Sgarra, C. (2024). 'Commodity Asian option pricing and simulation in a 4-factor model with jump clusters'. In: *Annals of Operations Research* 336, pp. 275–306.

Callegaro, G., Di Tella, P., Ongarato, B. and Sgarra, C. (2025). 'Semistatic variance-optimal hedging with self-exciting jumps'. In: *Mathematics of Operations Research* 0.0.

Callegaro, G., Fontana, C., Hillairet, C. and Ongarato, B. (2025). 'A stochastic Gordon-Loeb model for optimal cybersecurity investment under clustered attacks'. In: *arXiv preprint arXiv:2505.01221*.

Carr, P. (2011). 'Semi-static hedging of barrier options under Poisson jumps'. In: *International Journal of Theoretical and Applied Finance* 14.07, pp. 1091–1111.

Carr, P. and Madan, D. B. (2001). 'Towards a theory of volatility trading'. In: *Option pricing, interest rates and risk management, Handbooks Mathematical Finance*. Ed. by Jouini, E., Cvitanic, J. and Musiela, M. Cambridge: Cambridge University Press, pp. 458–476.

Černý Aleš, C. and Christoph Kallsen, J. (2023). 'Numéraire-invariant quadratic hedging and mean variance portfolio allocation'. In: *Mathematics of Operations Research* 49.2, pp. 752–781.

Chen, L. and Filipović, D. (2005). 'A simple model for credit migration and spread curves'. In: *Finance and Stochastics* 9, pp. 211–231.

Cheridito, P., Filipović, D. and Kimmel, R. L. (2007). 'Market price of risk specifications for affine models: theory and evidence'. In: *Journal of Financial Economics* 83.1, pp. 123–170.

Cheridito, P., Filipović, D. and Yor, M. (2005). 'Equivalent and absolutely continuous measure changes for jump-diffusion processes'. In: *The Annals of Applied Probability*, pp. 1713–1732.

Cont, R. and Tankov, P. (2003). *Financial Modelling with Jump Processes*. Chapman and Hall/CRC Financial Mathematics Series. CRC Press.

Cont, R. and Voltchkova, E. (2005). 'A finite difference scheme for option pricing in jump diffusion and exponential Lévy models'. In: *SIAM Journal on Numerical Analysis* 43.4, pp. 1596–1626.

Cox, D. R. (1955). 'Some statistical methods connected with series of events'. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 17.2, pp. 129–157.

Cox, J. C., Ingersoll, J. E. and Ross, S. A. (1985). 'A theory of the term structure of interest rates'. In: *Econometrica* 53.2, pp. 385–407.

Cuchiero, C. (2011). 'Affine and polynomial processes'. PhD dissertation. ETH Zurich.

Cuchiero, C., Filipović, D., Mayerhofer, E. and Teichmann, J. (2011). 'Affine processes on positive semidefinite matrices'. In: *The Annals of Applied Probability* 21, pp. 397–463.

Dai, Q. and Singleton, K. J. (2000). 'Specification analysis of affine term structure models'. In: *The Journal of Finance* 55.5, pp. 1943–1978.

Dana, R.-A. and Jeanblanc, M. (2003). *Financial Markets in Continuous Time*. Springer Berlin, Heidelberg.

Dassios, A. and Zhao, H. (2011). 'A dynamic contagion process'. In: *Advances in Applied Probability* 43.3, pp. 814–846.

Dassios, A. and Zhao, H. (2013). 'Exact simulation of Hawkes process with exponentially decaying intensity'. In: *Electronic Communications in Probability* 18, pp. 1–13.

Di Tella, P., Haubold, M. and Keller-Ressel, M. (2019). 'Semi-static variance-optimal hedging in stochastic volatility models with Fourier representation'. In: *Journal of Applied Probability* 56.3, pp. 787–809.

Di Tella, P., Haubold, M. and Keller-Ressel, M. (2020). 'Semistatic and sparse variance-optimal hedging'. In: *Mathematical Finance* 30.2, pp. 403–425.

Dieudonné, J. (1960). *Foundations of Modern Analysis. Pure and Applied Mathematics*. New York: Academic Press.

Dou, W., Tang, W., Wu, X., Qi, L. and Xu, X. (2020). 'An insurance theory based optimal cyber-insurance contract against moral hazard'. In: *Information Sciences* 527, pp. 576–589.

Drimus, G. G. (2012). 'Options on realized variance by transform methods: a non-affine stochastic volatility model'. In: *Quantitative Finance* 12.11, pp. 1679–1694.

Duffie, D. (2005). 'Credit risk modeling with affine processes'. In: *Journal of Banking & Finance* 29.11, pp. 2751–2802.

Duffie, D., Filipović, D. and Schachermayer, W. (2003). 'Affine processes and applications in finance'. In: *The Annals of Applied Probability* 13.3, pp. 984–1053.

Duffie, D., Pan, J. and Singleton, K. (2000). 'Transform analysis and asset pricing for affine jump-diffusions'. In: *Econometrica* 68.6, pp. 1343–1376.

Eberlein, E. and Kallsen, J. (2019). *Mathematical Finance*. Cham, Switzerland: Springer Finance.

Embrechts, P., Liniger, T. and Lin, L. (2011). 'Multivariate Hawkes processes: an application to financial data'. In: *Journal of Applied Probability* 48(A).3, pp. 367–378.

ENISA (2024). 'ENISA Threat Landscape 2024 Report'. Available at `https://www.enisa.europa.eu/publications/enisa-threat-landscape-2024`.

Errais, E., Giesecke, K. and Goldberg, L. R. (2010). 'Affine point processes and portfolio credit risk'. In: *SIAM J. Fin. Math.* 1.1, pp. 642–665.

Fedele, A. and Roner, C. (2022). 'Dangerous games: a literature review on cybersecurity investments'. In: *Journal of Economic Surveys* 36.1, pp. 157–187.

Filimonov, V., Bicchetti, D., Maystre, N. and Sornette, D. (2014). 'Quantification of the high level of endogeneity and of structural regime shifts in commodity markets'. In: *Journal of international Money and Finance* 42, pp. 174–192.

Filipovic, D. and Mayerhofer, E. (2009). 'Affine diffusion processes: theory and applications'. In: *Advanced Financial Modelling* 8, pp. 1–40.

Fleming, W. H. and Soner, H. M. (2006). *Controlled Markov Processes and Viscosity Solutions*. Springer.

Föllmer, H. and Sondermann, D. (1986). 'Hedging of Non-Reduntant Contingent Claims'. In: *Contributions to mathematical economics*. Ed. by Hildenbrand, W. and MasColell, A. North Holland, The Netherlands: Elsevier Science Publishing, pp. 205–223.

Fontana, C. (2012). 'Four Essays in Financial Mathematics'. PhD thesis. Università degli Studi di Padova.

Gaïgi, M., Ly Vath, V. and Scotti, S. (2025). ‘Optimal harvesting under uncertain environment with clusters of catastrophes’. In: *Journal of Economic Dynamics and Control* 179, p. 105165.

Gatheral, J. (2006). *The Volatility Surface: A Practitioner's Guide*. Wiley.

Gonzato, L. and Sgarra, C. (2021). ‘Self-exciting jumps in the oil market: Bayesian estimation and dynamic hedging’. In: *Energy Economics* 99, p. 105279.

Gordon, L. A. and Loeb, M. P. (2002). ‘The economics of information security investment’. In: *ACM Transactions on Information and System Security* 5.4, pp. 438–457.

Gordon, L. A., Loeb, M. P. and Lucyshyn, W. (2003). ‘Information security expenditures and real options: a wait-and-see approach’. In: *Computer Security Journal* 19.2, pp. 1–7.

Hainaut, D. (2016). ‘A model for interest rates with clustering effects’. In: *Quantitative Finance* 16.8, pp. 1203–1218.

Hainaut, D. and Moraux, F. (2018). ‘Hedging of options in the presence of jump clustering’. In: *Journal of Computational Finance* 22.3, pp. 1–35.

Hairer, E., Nørsett, S. P. and Wanner, G. (1993). *Solving Ordinary Differential Equations II: Stiff and Differential-Algebraic Problems*. Springer.

Hawkes, A. G. (1971). ‘Spectra of some self-exciting and mutually exciting point processes’. In: *Biometrika* 58.1, pp. 83–90.

Hawkes, A. G. and Oakes, D. (1974). ‘A cluster process representation of a self-exciting process’. In: *Journal of Applied Probability* 11, pp. 493–503.

Hayes, J. and Bodhani, A. (2013). ‘Cyber security: small firms under fire’. In: *Engineering & Technology* 8.6, pp. 80–83.

He, R., Jin, Z. and Li, J. S.-H. (2024). ‘Modeling and management of cyber risk: a cross-disciplinary review’. In: *Annals of Actuarial Science* 18.2, pp. 270–309.

Herrera, R. and González, N. (2014). ‘The modeling and forecasting of extreme events in electricity spot markets’. In: *International Journal of Forecasting* 30.3, pp. 477–490.

Heston, S. L. (1993). ‘A closed-form solution for options with stochastic volatility with applications to bond and currency options’. In: *The Review of Financial Studies* 6.2, pp. 327–343.

Hillairet, C., Réveillac, A. and Rosenbaum, M. (2023). ‘An expansion formula for Hawkes processes and application to cyber-insurance derivatives’. In: *Stochastic Processes and their Applications* 160, pp. 89–119.

Houssard, A., Ly Vath, V. and Scotti, S. (2025). ‘Optimal investment and consumption with transaction costs in a Hawkes jump-diffusion model’. In: *Available at SSRN.*

Huang, D. C. and Behara, R. S. (2013). ‘Economics of information security investment in the case of concurrent heterogeneous attacks with budget constraints’. In: *International Journal of Production Economics* 141.1, pp. 255–268.

Hubalek, F., Kallsen, J. and Krawczyk, L. (2006). ‘Variance-optimal hedging for processes with stationary independent increments’. In: *The Annals of Applied Probability* 16.2, pp. 853–885.

Jacod, J. and Shiryaev, A. N. (2013). *Limit Theorems for Stochastic Processes*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg.

Kallsen, J. and Muhle-Karbe, J. (2010). ‘Exponentially affine martingales, affine measure changes and exponential moments of affine processes’. In: *Stochastic Processes and their Applications* 120.2, pp. 163–181.

Kallsen, J. and Pauwels, A. (2010). ‘Variance-optimal hedging in general affine stochastic volatility models’. In: *Advances in Applied Probability* 42.1, pp. 83–105.

Kallsen, J. and Shiryaev, A. N. (2002). ‘The cumulant process and Esscher’s change of measure’. In: *Finance and Stochastics* 6, pp. 397–428.

Kallsen, J. and Vierthauer, R. (2009). ‘Quadratic hedging in affine stochastic volatility models’. In: *Review of Derivatives Research* 12, pp. 3–27.

Kawazu, K. and Watanabe, S. (1971). ‘Branching processes with immigration and related limit theorems’. In: *Theory of Probability & Its Applications* 16.1, pp. 36–54.

Kazamaki, N. (1977). ‘On a problem of Girsanov’. In: *Tohoku Mathematical Journal* 29.4, pp. 597–600.

Keller-Ressel, M. (2008). ‘Affine processes: theory and applications in finance’. PhD thesis. Technische Universität Wien.

Keller-Ressel, M. and Mayerhofer, E. (2015). ‘Exponential moments of affine processes’. In: *The Annals of Applied Probability* 25.2, pp. 714–752.

Keller-Ressel, M., Mayerhofer, E. and Smirnov, A. G. (2010). ‘On convexity of solutions of ordinary differential equations’. In: *Journal of Mathematical Analysis and Applications* 368.1, pp. 247–253.

Keller-Ressel, M., Schmidt, T. and Wardenga, R. (2019). ‘Affine processes beyond stochastic continuity’. In: *The Annals of Applied Probability* 29.6, pp. 3387–3437.

Kou, S. G. (2002). ‘A jump-diffusion model for option pricing’. In: *Management Science* 48.8, pp. 1086–1101.

Krutilla, K., Alexeev, A., Jardine, E. and Good, D. (2021). ‘The benefits and costs of cybersecurity risk reduction: a dynamic extension of the Gordon and Loeb model’. In: *Risk Analysis* 41.10, pp. 1795–1808.

Kunita, H. and Watanabe, S. (1967). 'On square integrable martingales'. In: *Nagoya Mathematical Journal* 30, pp. 209–245.

Laub, P. J., Lee, Y. and Taimre, T. (2021). *The Elements of Hawkes Processes*. Springer.

Lépingle, D. and Mémin, J. (1978). 'Sur l'intégrabilité uniforme des martingales exponentielles'. In: *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 42, pp. 175–203.

Li, Y. and Mamon, R. (2023). 'Modelling health-data breaches with application to cyber insurance'. In: *Computers and Security* 124, p. 102963.

Lim, A. E. (2004). 'Quadratic hedging and mean-variance portfolio selection with random parameters in an incomplete market'. In: *Mathematics of Operations Research* 29.1, pp. 3–27.

Lima, R. (2023). 'Hawkes processes modeling, inference, and control: an overview'. In: *SIAM Review* 65.2, pp. 331–374.

Liu, W. and Zhu, S.-P. (2019). 'Pricing variance swaps under the Hawkes jump-diffusion process'. In: *Journal of Futures Markets* 39.6, pp. 635–655.

Luciano, E. and Vigna, E. (2008). 'Mortality risk via affine stochastic intensities: calibration and empirical relevance'. In: *Belgian Actuarial Journal* 8.1.

Lukacs, E. (1970). *Characteristic Functions*. New York: Hafner Publishing Co.

Mayerhofer, E., Muhle-Karbe, J. and Smirnov, A. G. (2011). 'A characterization of the martingale property of exponentially affine processes'. In: *Stochastic Processes and their Applications* 121.3, pp. 568–582.

Mazzoccoli, A. and Naldi, M. (2020). 'Robustness of optimal investment decisions in mixed insurance/investment cyber risk management'. In: *Risk Analysis* 40.3, pp. 550–564.

Mazzoccoli, A. and Naldi, M. (2022). 'An overview of security breach probability models'. In: *Risks* 10.11:220.

Mémin, J. and Shiryaev, A. N. (1979). 'Un critère prévisible pour l'uniforme intégrabilité des semimartingales exponentielles'. In: *Séminaire de Probabilités XIII* 721, pp. 147–161.

Mémin, J. (2006). 'Décompositions multiplicatives de semimartingales exponentielles et applications'. In: *Séminaire de Probabilités XII*. Vol. 721. Springer, pp. 35–46.

Merton, R. C. (1976). 'Option pricing when underlying stock returns are discontinuous'. In: *Journal of Financial Economics* 3.1–2, pp. 125–144.

Miaoui, Y. and Boudriga, N. (2019). 'Enterprise security economics: a self-defense versus cyber-insurance dilemma'. In: *Applied Stochastic Models in Business and Industry* 35.3, pp. 448–478.

Mijatović, A. and Urusov, M. (2012). 'Deterministic criteria for the absence of arbitrage in one-dimensional diffusion models'. In: *Finance and Stochastics* 16, pp. 225–247.

Mikosch, T. (2009). *Non-Life Insurance Mathematics. An Introduction with the Poisson Process*. Springer Berlin, Heidelberg.

Mordukhovich, B. S. and Nam, N. M. (2022). *Convex Analysis and Beyond. Volume I: Basic Theory*. Vol. 1. Springer Cham.

Neuberger, A. (1994). 'The log contract'. In: *Journal of Portfolio Management* 20.2, pp. 74–80.

Nguyen, D. D. A., Alain, P., Autrel, F., Bouabdallah, A., François, J. and Doyen, G. (2024). 'How fast does malware leveraging EternalBlue propagate? The case of WannaCry and NotPetya'. In: *2024 IEEE 10th International Conference on Network Softwarization (NetSoft)*, pp. 399–404.

Novikov, A. (1973). 'On an identity for stochastic integrals'. In: *Theory of Probability & Its Applications* 17.4, pp. 717–720.

Ogata, Y. (1978). 'The asymptotic behaviour of maximum likelihood estimators for stationary point processes'. In: *Annals of the Institute of Statistical Mathematics* 30, pp. 243–261.

Öğüt, H., Raghunathan, S. and Menon, N. (2011). 'Cyber security risk management: public policy implications of correlated risk, imperfect ability to prove loss, and observability of self-protection'. In: *Risk Analysis* 31.3, pp. 497–512.

Palmowski, Z. and Rolski, T. (2002). 'A technique for exponential change of measure for Markov processes'. In: *Bernoulli* 8.6, pp. 767–785.

Prokopczuk, M., Symeonidis, L. and Wese Simen, C. (2017). 'Variance risk in commodity markets'. In: *Journal of Banking and Finance* 81, pp. 136–149.

Protter, P. E. (2005). *Stochastic Integration and Differential Equations*. Springer Berlin, Heidelberg.

Protter, P. E. and Shimbo, K. (2008). 'No Arbitrage and General Semimartingales'. In: *Markov processes and related topics: a Festschrift for Thomas G. Kurtz*. Vol. 4. IMS Lecture Notes–Monograph Series, pp. 267–284.

Rambaldi, M., Pennesi, P. and Lillo, F. (2015). 'Modeling foreign exchange market activity around macroeconomic news: Hawkes-process approach'. In: *Phys Rev E Stat Nonlin Soft Matter Phys.* 91, p. 012819.

Revuz, D. and Yor, M. (2013). *Continuous Martingales and Brownian Motion*. Vol. 293. Springer Science & Business Media.

Reynaud-Bouret, P., Rivoirard, V. and Tuleau-Malot, C. (2013). 'Inference of functional connectivity in Neurosciences via Hawkes processes'. In: *2013 IEEE Global Conference on Signal and Information Processing*, pp. 317–320.

SANS Institute (2025). '2025 ICS/OT Cybersecurity Budget: Spending Trends, Challenges, and the Future'. Available at https://www.sans.org/white-papers/2025-ics-ot-cybersecurity-budget-spending-trends-challenges-future/.

Sarychev, A. V. (1996). 'On equation $x^{(n+1)} = f(t, x, \dot{x}, \ldots, x^{(n)})$ with convex quasi-monotone righthand side'. In: *Nonlinear Analysis: Theory, Methods & Applications* 27.7, pp. 785–792.

Sato, K.-i. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Vol. 68. Cambridge University Press.

Schal, M. (1994). 'On quadratic cost criteria for option hedging'. In: *Mathematics of Operations Research* 19.1, pp. 121–131.

Schrager, D. F. (2006). 'Affine stochastic mortality'. In: *Insurance: Mathematics and Economics* 38.1, pp. 81–97.

Skeoch, H. R. (2022). 'Expanding the Gordon-Loeb model to cyber-insurance'. In: *Computers and Security* 112, p. 102533.

Stabile, G. and Torrisi, G. L. (2010). 'Risk processes with non-stationary Hawkes claims arrivals'. In: *Methodology and Computing in Applied Probability* 12, pp. 415–429.

Tatsumi, K.-i. and Goto, M. (2010). 'Optimal timing of information security investment: a real options approach'. In: *Economics of information security and privacy*. Ed. by Moore, T., Pym, D. and Ioannidis, C. Springer, pp. 211–228.

Vasicek, O. (1977). 'An equilibrium characterization of the term structure'. In: *Journal of Financial Economics* 5.2, pp. 177–188.

Yuan, S. (1999). 'ODE-oriented semi-analytical methods'. In: *Computational Mechanics in Structural Engineering*, pp. 375–388.

Zeller, G. and Scherer, M. (2022). 'A comprehensive model for cyber risk based on marked point processes and its application to insurance'. In: *European Actuarial Journal* 12, pp. 33–85.

Acknowledgments

When I started my Bachelor's degree, I was far from imagining that I would one day complete a Ph.D. It has been an intense and exciting journey, but that would never have happened without all the people who supported, encouraged, and inspired me throughout these years.

First of all, I am deeply indebted to my supervisors, Prof. Giorgia Callegaro and Prof. Claudio Fontana. I met Giorgia during my Master's degree, and she introduced me to the fascinating world of research. With her curiosity and enthusiasm, she accompanied me throughout the entire Ph.D., from application to thesis submission. Her guidance shaped me both as a researcher and as a person, encouraging me to take risks and not fear failure. I also thank Claudio, for agreeing to supervise me through this Ph.D. He guided me with great patience, always showing me new and interesting viewpoints in our discussions, and teaching me to approach both mathematical and non-mathematical problems with clarity and perspective. Throughout the years, Giorgia and Claudio supported all my ideas, whether research-related, conference planning, visits abroad, or any other matters. Moreover, they guided me through delicate decisions. I will always be thankful for their trust and mentorship.

I acknowledge the referees for taking the time to read my manuscript and for providing their valuable feedback, as well as the members of the Ph.D. defense committee.

My sincere thanks go to my co-authors, who helped bring the papers included in this thesis to life. In particular, I am grateful to Prof. Caroline Hillairet, who hosted me at CREST during my second year and helped me feel part of the research community in Paris. I also thank Dr. Paolo Di Tella and Prof. Carlo Sgarra for the many hours spent on Zoom discussing research problems, checking computations, and sharing a good laugh along the way.

I would like to acknowledge Prof. Keller-Ressel for offering me the opportunity to continue my academic career as a Postdoc at TU Dresden.

A warm thanks goes to my Ph.D. colleagues of the 38th Cycle in Padova for sharing the daily ups and downs of doctoral life, and many after-lunch crosswords. Special thanks to the Probability group of the University of Padova, which provided me with a great environment to explore different topics and research lines. I am especially grateful to Giacomo, Ofelia, and Alekos, who helped me navigate this new academic world from day one (and even before that), and made every conference we attended together much more fun. I am also thankful to Loretta and Marinella for their immense patience through administrative matters. I would like to thank my colleagues at CREST and the Finance and Actuarial Science group for making me feel welcome and for making my visiting period so scientifically stimulating.

I owe so much to all my friends. To Lucia: I was looking for a room in Paris, but I found a friend. Being your sous-chef is always an honour. To Ludovico and Francesco: you made my time in Padova so much fun, sharing countless dinners, board games, and Mario Kart nights. To my university friends, Eva, Gloria, Leonardo, and Nicolò: even though meeting up isn't as easy as it once was, every time we do, you warm my heart. To all my friends from my hometown (there are way too many of you to list!): thank you for always being there, no matter what, and no matter where I was in the world.

I owe my deepest gratitude to my parents and my brother Leonardo. You never let me doubt my value, supporting me in every experience, even when research took me far from home.

Finally, thank you Pietro. It is hard to find the words to express how much your support and constant encouragement have meant to me throughout this Ph.D. journey, and in everyday life. Thank you for being by my side every step of the way.