City Research Online

City, University of London Institutional Repository

On sources of risk in quadratic hedging and incomplete markets

Juraj Špilda

Cass Business School

City, University of London

Contents

List of Figures

Acknowledgements

First and foremost, I would like to thank my supervisor, Prof. Aleš Černý, for his continual support throughout my PhD research and for all his patience with me.

I would like to acknowledge and thank City, University of London for its financial support, which allowed me to undertake this PhD. Part of this research was also made possible thanks to the financial support of the Tatra Banka Foundation, for which I am extremely grateful.

Many thanks to my fellow doctoral colleagues Andrew Hunt, Andres Villegas and Natalia Matanova for the pleasant conversations over lunch or tea in the office kitchen, and for making the office a place one looks forward to going to.

Thanks goes to Juraj Hledík, for spotting my mistakes in Chapter 1.

Finally, I would like to thank my family for their continual love and support, even when they weren't sure what I was doing and why I was doing it. Most importantly, I would like to thank my wife Funda Üstek Špilda, for all her love and for giving me the strength to continue my research even during the rainy days, as is so often the case in London. I look forward to sharing all my future adventures with her.

Declaration

I, declare that this thesis titled, 'On sources of risk in quadratic hedging and incomplete markets' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.

- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.

- Where I have consulted the published work of others, this is always clearly attributed.

- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.

- I have acknowledged all main sources of help.

- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

- I grant powers of discretion to the University Librarian to allow this thesis to be copied in whole or in part without further reference to the author. This permission covers only simple copies made for study purpose, subject to the normal conditions of acknowledgement.

Signed:

_____

Date: 26th April 2017

_____

Abstract

This thesis is divided into three chapters, each dealing with a different aspect of market incompleteness and its consequences on quadratic hedging strategies and hedging errors.

The first chapter studies the effects of market incompleteness due to discrete time trading. We derive the asymptotics (in trading frequency) of the quadratic hedging error of a digital option and obtain a correction to the classical granularity formula, showing that for discontinuous payoffs, the second order term driven by the Cash Gamma remains highly significant. We also show that the discrete-time quadratic hedging strategy generates the same asymptotic error as a continuous-time Black-Scholes delta-hedging strategy used on a discrete set of times.

The second chapter studies the effects of market incompleteness due to jumps in cases when the discretization error from Chapter 1 is predictable. We compute the hedging error under an exponential Lévy model for a general 'Lévy contract' that encompasses log contracts, variance swaps and higher order moment swaps. We compare two utility-based pricing approaches for incomplete markets: quadratic hedging (corresponding to quadratic utility) and exponential utility. We show that for small jumps, numerically difficult exponential utility results are well-estimated via closed-form quadratic hedging formulas. We use our results on hedging errors to obtain 'good-deal bounds' for variance and skewness swaps.

The third chapter studies the effects of market incompleteness due to uncertainty in the exact specification of the data generating process. We conduct quadratic hedging under a regime-switching Lévy model, which switches between a finite set of distributions based on the value of a (hidden) state variable. We solve the quadratic hedging problem in two steps. First we compute a stochastic differential equation for the filtered estimate of the hidden state. We then use it to solve the quadratic hedging problem with this additional observable variable via classic techniques. We provide Fourier Transform formulas for the mean-value process and hedging strategy, and a recursive scheme for the hedging error.

Introduction - an overview of contributions

This thesis is divided into three main chapters, each dealing with a different aspect of market incompleteness and its consequences on variance-optimal prices, hedging strategies and hedging errors.

Chapter 1

The first chapter is a study of the effects of market incompleteness due to discrete time hedging. We specifically investigate the asymptotics of the hedging error of a digital option as we rebalance our hedging position more frequently. Building on the results of Bertsimas et al. [2000] for vanilla options and Gobet and Temam [2001a] for digital options, we show that when the underlying is a martingale and we consider a digital option, the variance-optimal hedging strategy, designed under a discrete time incomplete market setting, generates the same asymptotic hedging error as that of a continuous-time Black-Scholes Δ-hedging strategy used on a discrete set of times. This brings into relation the discrete-time quadratic hedging error and the "tracking error" of following a continuous-time strategy on a discrete set of trading dates. We develop a more precise, second-order formula to compute the hedging error asymptotics. We show that this second order term, usually ignored in the literature, is a modified variant of the term obtained by Bertsimas et al. [2000] for vanilla options, with an additional compensating for the explosive, divergent nature of the Cash Gamma of a digital option at maturity. We show that the Cash Gamma remains the main driver for the hedging error of a digital option even though it does not appear in the first order asymptotic term as derived by Gobet and Temam [2001a], by showing that for sensible, realistic numerical values for parameters in the model, the second-order approximation of the hedging error of a digital option is significantly

more precise than the first-order approximation.

Chapter 2

The second chapter builds on the ideas of the first chapter by considering a contract for which the Cash Gamma, and hence the discretization error, is completely predictable. This is the case of the log contract, which serves as a building block for variance swaps. Since the discretization error for such contracts is no longer stochastic, we focus on investigating what additional sources of market incompleteness may impact these contracts. Specifically, we look at the impact of market incompleteness due to jumps in stock returns on the log contract and variance swap. We introduce jump risk into our model via exponential Lévy processes and compute the price, hedging strategy and hedging error for a new, generic type of contract, which we label the 'Lévy contract' - this encompasses log contracts, variance swaps and higher order moment swaps. This adds to the literature on pricing variance swaps with jumps by not only computing prices, but also hedging strategies and hedging errors. We use an incomplete-market utility maximization approach to calculate these quantities. We consider and compare two utility functions to measure gain and loss: on the one hand, mean-variance preferences; on the other hand, exponential utility. We show that the former is equivalent to solving the variance-optimal hedging problem, and we use quadratic hedging quantities introduced in Chapter 1 to express prices, hedging strategies and hedging errors for the 'Lévy contract'. We also show that the latter essentially leads to calculating an exponential compensator, which we can explicitly calculate using well-established results when the payoff of the derivative is given by the realisation of a Lévy process.

We connect results on indifference pricing with our results on hedging errors to find economically sensible price ranges (so-called 'good-deal bounds') for variance and skewness swaps. In addition to this being a new result, these bounds are also tighter than the no-arbitrage bounds typically derived in the literature. We show that asymptotically, as jumps in our driving Lévy process become small and the skew and kurtosis decay to zero, exponential utility pricing results (which can only be obtained implicitly) are well-estimated via simple closed-form formulas from variance-optimal hedging, using the first four moments of the returns distribution given by the model. We find that variance swap prices should contain an adjustment for the skewness of returns, whereas skewness swaps should contain an adjustment for the kurtosis of returns. We find that the width of price bounds on variance swaps and skewness

swaps is driven by moments of up to the 4th order and 6th order respectively.

Chapter 3

The third chapter is motivated by the results of the second chapter, which strongly depend on knowledge of the moments of the distribution of our returns. Therefore, chapter 3 focuses on how pricing and hedging is further impacted if we introduce uncertainty into the knowledge of the exact specifications of the "true" underlying model generating returns. In contrast to Chapter 2, which assumes a fixed model constant in time and relies on this fact to provide results, in the third chapter we conduct variance-optimal hedging when our returns are driven by a regime-switching Lévy process, allowing the returns distribution to switch between a finite set of distributions, based on the value of a state variable controlling the current regime. We also make this regime state variable unobservable (putting this model into the category of models often referred to as Hidden Markov Models), requiring us to filter out an estimate of the current state based on observed returns. We solve the variance-optimal hedging problem in a martingale setting in two steps: first, we derive an explicit stochastic differential equation for the filtered estimate of the true regime driving returns, extending and clarifying results from Ceci and Colaneri [2012]. Having obtained a stochastic dynamics for the filtered estimate, we proceed to solve the variance-optimal hedging problem via classic techniques with an additional observable state variable (the filtered estimate). We provide Fourier Transform-based formulas for the mean-value process and hedging strategy, and a recursive scheme to compute the expected hedging error. To the best of our knowledge, no-one has previously calculated the hedging error in such a model. We implement our theoretical results numerically and illustrate the difference between the regime-switching model and a simple weighted average of models. We run Monte Carlo simulations to verify the significance of the impact of the regime-switching hedging strategy as opposed to simpler approaches. Finally, we show how the Hidden Markov Model degenerates into multiple simpler models, and we compare it against these less complex models.

Chapter 1

On the hedging error asymptotics of a digital option

1.1 Hedging errors - motivation and literature review

The discovery of no-arbitrage option pricing (Black and Scholes [1973]) and the concept of replication of derivatives via their underlying assets (Merton [1973]) has led to a revolution in the world of finance. The notional value of all the outstanding derivatives now overshadows the value of the underlying assets several times over. The use of derivative contracts permeates the financial world in many ways, affecting businesses and governments alike: airlines fix their costs by buying jet fuel futures, farmers ensure a steady price for their wheat harvest by selling wheat futures, oil-rich states buy put options on oil to ensure a floor on annual revenues, pension funds buy interest-rate swaps to cover their ongoing future liabilities. They are only able to do so, however, because there are counter-parties confident they can manage the risk in these derivative contracts, either because they believe they can foresee market activity and want exposure to the asset class (e.g. hedge funds or pension funds), or because they feel confident they can manage their risk properly by trading in the underlying asset to replicate the final payoff. Therefore it is of utmost importance that we understand what risks this concept entails when deployed in practice.

As presented in Merton [1973], the theory of replication includes several assumptions

which can never be satisfied in reality: no transaction costs, continuous trading, the ability to borrow infinite amounts of cash, an agent whose trading does not influence stock prices, lognormally distributed returns. This brings great doubt as to whether the practices it encourages are truly safe. Therefore people have been researching the problems that arise in practice when these assumptions are broken.

In this first chapter, we will focus on the assumption of continuous trading (as in practice continuous trading is infeasible, not least due to the transaction costs involved). We will break the assumption and consider a discretely hedged contingent claim. We will analyze the expected hedging errors we obtain for vanilla and digital options if we follow a hedging strategy optimal in discrete time. We will be particularly interested in the asymptotic behaviour of these errors for the case of digital options as we increase the frequency of rebalancing, since in this case the order of convergence of the error changes as we approach maturity. Our ultimate goal is to extend the existing formulas in the literature for the asymptotics, showing that not only the first order, but also the second order asymptotic term is significant. We will also compare these asymptotics to those of a hedging strategy optimal in continuous time but applied at a discrete set of times.

This chapter is organized as follows: in the first section, we will discuss the evolution of the concept of contract replication and hedging. We will review the standard Black-Scholes-Merton approach and then introduce the concept of mean-variance (a.k.a. variance-optimal or quadratic) hedging, a more versatile hedging strategy optimal either in discrete or continuous time in terms of minimizing hedging errors in the L^2 sense and applicable to a general semimartingale underlying. In regard to it we also introduce the related locally optimal risk-minimizing hedging strategy, which minimizes hedging error in the L^2 sense over a single (potentially infinitesimal) timestep. We will then review the literature that has analyzed the so-called "tracking error" - the error made when using a continuous-time trading strategy on a discrete set of times; we will focus in particular on the literature analyzing the asymptotic behaviour of these errors. In the second section, we will perform our own heuristic analysis of the variance-optimal strategy and its hedging errors to gain intuition into how these errors evolve asymptotically for derivatives with regular and discontinuous payoffs, showing that their asymptotic behaviour differs only near maturity. We will perform this analysis on the examples of a standard call option and a digital call option. In the third chapter, we will investigate how the asymptotic hedging error of a digital option evolves over time and we will show how the rate of decay of error changes as

the value of our derivative approaches maturity; we will contrast our results with the standard results for a vanilla call option and show that the change is caused by the explosive behaviour of the Cash Gamma of a digital option at maturity. In the fourth section, we will show how this all relates back to the tracking error of a Black-Scholes strategy and to previously known results.

1.1.1 The evolution of hedging

The stock market has been around for centuries, and derivative securities have been around at least since the 17th century (see Schaede [1989]). Some may argue derivatives appeared even as early as ancient Greece, where, according to an account from Aristotle [1999, Book 1, section 1259a], the famous Thales of Miletus entered a forward-type agreement on olive oil presses. Traders who operated in these early markets usually used heuristic "rules of thumb" to protect their open positions and make profits, these rules coming from years of experience. Only in the 20th century did more scientific approaches to handling risk in the markets appear. The idea of securing arbitrage profits by hedging positions in a derivative contract, specifically an option, by buying/selling a specific proportion of the underlying stock, was first publicly presented in Thorp and Kassouf [1967]. In this system, the seller of an option would go on to buy a particular quantity of stocks to ensure that no matter which direction the market moved, the total portfolio value would remain constant, with losses in options being replaced by gains in the stock and vice versa. This was inspired by the work in Samuelson [1965], where a rational price for an option was derived, which in turn used the much older result in Bachelier [1900] of modeling stock prices as what would later be named by Norbert Wiener as Brownian motion.

1.1.1.1 Black-Scholes-Merton Δ-hedging

Despite this early research, it was only the papers Black and Scholes [1973] and Merton [1973] that managed to bring about an explosion of activity in the derivatives market. Black and Scholes [1973] provided a closed-form solution for the no-arbitrage price of a vanilla call option, and Merton [1973] mathematically formalized the idea of option pricing via a replicating portfolio, using the tools provided by stochastic calculus, which had progressed significantly since the time of Bachelier [1900]. The fundamental idea of replication is similar to that presented in Thorp and Kassouf

[1967]: by trading options and stocks in particular proportions you can make a risk-less profit, i.e. get the return of a risk-free investment such as a money market account. Conversely, you can trade stocks and hold the remaining cash in a money market account in such proportions that you end up with the same payoff as you would obtain with an option, for any possible stock price in the future. Therefore, a trader in a bank who sells an option to a client can "hedge" his open position by trading stocks and investing in a money market account, thus reducing the riskiness of his book and ensuring he can meet his client's demands no matter the future scenario. The proportion of stocks to be bought, as dictated by the Black-Scholes theory, is referred to as the Δ-hedge. The initial cash needed to engage in this trading also uniquely determines the price of the option: by a no-arbitrage argument, it has to cost as much as the trading strategy does; if the price were higher (lower), one could sell (buy) the option and follow the Δ-hedging strategy to obtain the same payoff at a lower cost, resulting in an arbitrage profit.

These reasonably simple arguments managed to completely transform the financial industry and led to a boom in the derivatives market. However, this system only works for a narrow class of models with many assumptions built in, as was already highlighted in the introduction. Some papers, such as Haug and Taleb [2011], still recommend sticking to the simpler heuristics developed by traders in the past, as the trader ends up relying too much on the model and forgets its differences from reality. Nevertheless, the Black-Scholes-Merton argument remains highly popular amongst researchers and practitioners alike, and its robustness has been thoroughly scrutinized (see e.g. Forde [2003], Karoui et al. [1998] and references therein).

1.1.1.2 Mean-variance hedging

One of the contested assumptions of the Black-Scholes model is that of market completeness, which among other things implies that options are obsolete since they can be perfectly replicated by trading the underlying stock and putting cash into a money account. This is obviously false since options continue to be traded, but the fact that Δ-hedging based on the Black-Scholes model does reduce the risk of issuing an option remains true. To better understand why this holds, a new incomplete market approach to hedging arose from Hodges and Neuberger [1989], where the option price is given as a solution to a utility maximization problem. Though originally this approach was studied under standard utility functions used in economics and produced

non-linear pricing rules, a new strand of literature arose from using this concept with a "utility function" $-x^2$ (which does not satisfy the standard Inada conditions required - see e.g. Hugonnier et al. [2005]) where at every time step, the economic agent chooses a hedging strategy so as to maximize his utility, which he obtains from having minimal expected squared loss at maturity from his hedged option position. The price of the replicated claim and the optimal hedging strategy is then formally given as the solution to the optimization problem

$$\inf_{\{\vartheta_t\}_{t=0,1,\ldots,T-1}} \mathbb{E}[(G_T^x(\vartheta) - H)^2] \tag{1.1}$$

$$s.t. \ G_T^x(\vartheta) = R_f^T x + \sum_{t=0}^{T-1} R_f^{T-t-1}\vartheta_t S_t(R_{t+1} - R_f)$$

where H is the payoff of the contingent claim at expiry and $G_T^x(\vartheta)$ is the terminal value of the gains process of the self-financing portfolio that holds ϑ_t stocks with returns of $\{R_{t+1}\}_{t=0,\ldots,T-1}$, and the remainder of cash in risk-free money account with return R_f in an attempt to replicate the claim, with initial capital x. The trading strategy ϑ_t and initial capital x must satisfy technical conditions as given in Černý and Kallsen [2009, Definition 2.2] to be admissible; importantly, ϑ_t must be a predictable, i.e. $\mathcal{F}_t$-measurable process. In plain terms, we logically need to decide our strategy ϑ_t before we realize stock gains $\Delta S_{t+1} = S_{t+1} - S_t$.

As the reader may notice, problem (1.1) can be seen as a sort of ordinary least squares minimization problem with a constraint. Early work on this regression-based technique can be found in Föllmer and Schweizer [1989] and Föllmer and Sondermann [1986] and has been generalized to a great extent over the years; most recently in Černý and Kallsen [2007], which provides a general semimartingale framework encompassing both discrete and continuous time models.

The strategy $\varphi(x, H)$ that solves problem (1.1) is usually referred to as the dynamically, globally optimal strategy, or the mean-variance or variance-optimal hedge. It is the strategy that gives an expected error with mean zero and minimal variance, as described in Schweizer [1995], and is closely related to the classical result of one-period mean-variance portfolio optimization as developed by Markowitz [1952]. It is also closely related to a suboptimal strategy, the so-called locally optimal strategy, in which the optimization for each ϑ_t minimizes the *conditional* squared hedging error over a single time-step and is oblivious to how well the hedging strategy performed in previous and future timesteps. It is sometimes also referred to as the risk-minimizing

strategy. Mathematically, at each time-step, the hedging strategy ϑ_t is given by solving the problem (as in Černý and Kallsen [2009, eqn. (4.1)])

$$\min_{\vartheta_t, x} \mathbb{E}_{t-1}[(x + \vartheta_t \Delta S_t - V_t)^2] \qquad s.t. \ V_T := H. \tag{1.2}$$

We use the short-hand notation $\mathbb{E}_t[\cdot] = \mathbb{E}[\cdot|\mathcal{F}_t]$ throughout. We will refer to V as the mean-value process, which coincides with the option payoff at maturity T. In a complete market model (such as Black-Scholes), the mean-value process coincides with the price of the option; otherwise, it has only mathematical meaning with no real-world interpretation. If we model the underlying S as a semimartingale, the optimal initial endowment x at time $t - 1$ turns out to be $x = V_{t-1} = \mathbb{E}_{t-1}^{\mathbb{Q}}[V_t] = \mathbb{E}_{t-1}^{\mathbb{Q}}[\mathbb{E}_t^{\mathbb{Q}}[V_{t+1}]] = \cdots = \mathbb{E}_{t-1}^{\mathbb{Q}}[H]$, where $\mathbb{Q}$ is the so-called variance-optimal (σ-)martingale measure. For the special case of the discounted price process being a martingale, this measure $\mathbb{Q}$ coincides with the physical measure $\mathbb{P}$. We can notice that in locally optimal hedging, we always assume that we managed to hedge perfectly the change in price between time-steps $t-1$ and t(i.e. locally) and at t we compute the hedging strategy ϑ_t as if we had an initial endowment equal to V_t, i.e. the value we had set out to obtain by hedging. We denote the locally optimal hedging coefficient by ξ and in Černý and Kallsen [2009, eqn. (4.6)] it is explicitly given in discrete time as

$$\xi_t = \frac{\mathrm{Cov}_{t-1}(V_t, \Delta S_t)}{\mathrm{Var}_{t-1}(\Delta S_t)}. \tag{1.3}$$

We can see the hedging coefficient not only as a solution to the one-period problem (1.2), but also as part of the so-called Föllmer-Schweizer decomposition of the payoff H, first introduced in Föllmer and Schweizer [1990]:

$$H = H_0 + \int_0^T \xi_{t-} \, dS_t + N_T,$$

where H_0 is some constant (in our problem it can be seen as the initial capital) and N is a local martingale orthogonal to S under physical measure $\mathbb{P}$, i.e. $\langle N, S \rangle = 0$, or more plainly $\mathrm{Cov}_{t-1}(\Delta N_t, \Delta S_t) = 0$. In a discrete time setting, the integral converts to a sum. Given this decomposition, the price of the contingent claim is computed as

$$V_t = \mathbb{E}_t^{\hat{\mathbb{P}}}[H] = \mathbb{E}^{\hat{\mathbb{P}}}[H] + \int_0^t \xi_{t-} \, dS_t + N_t,$$

where $\hat{\mathbb{P}}$ is the so-called minimal martingale measure, which in certain situations (e.g.

when S is continuous) coincides with the variance-optimal martingale measure $\mathbb{Q}$.

The existence of a solution to the local minimization problem (1.2) is, in light of this new perspective, equivalent to the existence of the Föllmer-Schweizer decomposition. A sufficient condition for the existence is the assumption that stock returns are IID, allowing the solutions to (1.2) at various time-steps to be independent of each other. Without the IID assumption, a locally optimal strategy may or may not be well-defined, since the distribution of returns under the physical, historical measure may not have all the moments the theory requires. Černý and Kallsen [2009, (Example 8.9)] provides an example of such a situation. However, a globally optimal solution to (1.1) always exists under the general conditions provided in Černý and Kallsen [2007]. If both strategies do exist and the underlying price is assumed to be continuous (i.e. without jumps), they are closely related. In the case of IID returns their relation is specifically given in Černý and Kallsen [2009, eqn. (4.16)] as

$$\varphi_t(x) = \xi_t + \tilde{\lambda}_t(V_{t-1} - G_{t-1}^x(\varphi(x, H))).$$

where $\tilde{\lambda}_t = \mathbb{E}_{t-1}[\Delta S_t]/\mathbb{E}_{t-1}[(\Delta S_t)^2]$. For non-IID returns this formula slightly alters, as the $\tilde{\lambda}$ terms have to be obtained under a different, so-called opportunity-neutral, measure.

If we look back, we can notice that the entire exposition on mean-variance hedging did not require us to assume a specific model for the underlying and most of the analysis refers to a situation where the underlying is a semimartingale, which is currently the most general setting available that is still mathematically tractable. Thus its main strength is that it does not require either market completeness or the assumption of log-normal returns and is therefore better suited to handle kurtosis in a stock return distribution (see Černý [2007]). Another level of flexibility originates from the possibility to interpret the expectation in problem (1.1) to be either under the historical physical probability, as in e.g. Hubalek et al. [2006], or under a martingale probability, as in e.g. Föllmer and Sondermann [1986]. Cont et al. [2007] presents a case in support of using the martingale probability. The choice of martingale measure, in essence, allows us to choose whether we want to calibrate our prices based on the historical distribution of the underlying or the current prices of other derivatives in the market. The decision as to which approach to use remains an open question.

1.1.2 Hedging errors over time

In this section we will discuss the consequences of breaking one of the assumptions in the Black-Scholes model that lead to market incompleteness, particularly that of continuous trading. Instead of rebalancing our replicating portfolio continuously, we will only rebalance at a discrete set of times. This is an error that all traders implementing their trading strategies via replication experience. The first paper to consider the discretization of replicating hedging strategies is Boyle and Emanuel [1980], but the first breakthrough was in Leland [1985], where the discretization occurs due to the introduction of transaction costs, which explode when hedging becomes continuous. In this paper, there are the first signs that the quantity

$$\frac{\partial^2 V}{\partial S^2} S^2 = \Gamma S^2,$$

the so-called Cash Gamma of an option, is of particular importance to the hedging error made. Here, V denotes the value of the contingent claim being replicated, S denotes the underlying asset. The computations in this paper are then made more explicit by Toft [1996], where exact formulas for hedging errors due to discretization are derived.

There was then a growing interest in the asymptotics of the hedging error and how exactly it decays as we increase hedging frequencies. Independently from each other, Zhang [1999] and Bertsimas et al. [2000] derived the exact order of decay of the hedging error when following a Black-Scholes type Δ-hedging strategy with the underlying driven by the stochastic differential equation

$$dS = \mu(t, S)dt + \sigma(t, S)dW.$$

Both arrived at the fact that if the payoff of the contingent claim is sufficiently smooth, the error will decay in a particular manner. Specifically, if we consider equidistant trading intervals of length δ on a time interval $[0, T]$, then Zhang [1999] and Bertsimas et al. [2000] conclude that the total squared hedging error (in an L^2 sense) $\varepsilon_0^2 = \varepsilon_0^2(\delta) = \mathbb{E}_0[(H - V_T)^2]$, will have the following form:

$$\varepsilon_0^2(\delta) = \left(\frac{1}{2} \int_0^T \mathbb{E}_0 \left[\left(\sigma^2(t, S_t) S_t^2 \Gamma_t \right)^2 \right] dt \right) \delta + \mathcal{O}(\delta^{3/2}) = g^2 \delta + \mathcal{O}(\delta^{3/2}) \qquad (1.4)$$

The authors in Bertsimas et al. [2000] coin a new name for the variable g: granularity,

as it refers to how "granular" time is, i.e. how much error we experience when assuming continuous time in a discrete time reality. However, we must stress that the result only holds in two cases:

- The payoff of the contingent claim is 6 times continuously differentiable w.r.t. the underlying, the driving diffusion process has coefficients $\mu(t, S)$, $\sigma(t, S)$ that are differentiable once w.r.t. to time and 3 times w.r.t. to the underlying, and $S\sigma(t, S)$ is six times differentiable. Moreover, all of these derivatives have to be bounded.

- The payoff of the contingent claim is continuous and piece-wise linear, all of the differentiability and boundedness conditions above on the diffusion parameters are satisfied and moreover, $S^2 \frac{\partial^\alpha \sigma(t,S)}{\partial S^\alpha}$ is bounded for all $2 \leq \alpha \leq 6$.

Thus, the result does not hold for derivatives with discontinuous payoffs. Furthermore, the results were only derived for a standard Δ-hedging strategy. Despite all the assumptions, the result is significant, because it has a very direct application to trading: if the trader wants to reduce his tracking error ε_0 by a half, he has to trade four times as often, i.e. reduce his rebalancing interval δ by a factor of four.

The literature then breaks the assumption on payoff regularity and observes that when this occurs, the first order of decay of the L^2 squared error is no longer $\mathcal{O}(\delta)$, but $\mathcal{O}(\sqrt{\delta})$ and hence the granularity changes. This is observed in Temam [2001] and Gobet and Temam [2001b] from a mathematical finance perspective. This property was also studied from a purely mathematical perspective, as a problem of approximating a stochastic integral in L^2-space in Geiss [2002], where the convergence rate is dependent on the so-called fractional regularity of the terminal condition, i.e. the payoff. The authors show that error decay of order $\mathcal{O}(\delta)$ can be regained by taking non-uniform time steps.

Further recent research in Gobet and Makhlouf [2012] shows that for both continuous and discontinuous payoffs, replacing the Δ-hedging strategy with a $\Delta - \Gamma$-hedging strategy reduces the total squared error but does not alter the order of convergence for a regular call, and does not even change the total error for a digital call. The former implies that although we are able to reduce the error introduced into hedging via the (path-dependent) Cash Gamma term, we are never able to completely eliminate it, i.e. hedge out all the Gamma risk.

The literature also extends the analysis in Bertsimas et al. [2000] to more complicated and realistic processes, from stochastic volatility models in Hayashi and Mykland [2005], exponential Lévy models in Denkl et al. [2013] and Broden and Tankov [2011], and general Itô processes with jumps in Tankov and Voltchkova [2009]. Broden and Tankov [2011] and Tankov and Voltchkova [2009] analyze the order of convergence of the tracking error in presence of jumps and show that it may depend on the finer structure of the jump measure around zero, and L^2 convergence may be of a different order than convergence in probability. Tankov and Voltchkova [2009] also shows that the hedging error (not the squared hedging error) of a digital option is of the same order as that of a regular call option. Combining this with the result from Gobet and Temam [2001b], we see that the digital option has a mean error of the same order as claims with continuously differentiable payoffs, but its variance is of a different order.

1.1.3   Connecting granularity, tracking errors of the Δ-hedge and quadratic hedging

As we have seen in the previous section, tracking errors have been widely examined mostly just for a Black-Scholes Δ-hedging strategy, although there are also a few results concerning mean-variance hedging errors. Tankov and Voltchkova [2009] finds the convergence of the mean of the error ε_0 to zero for general Lévy-Itô processes to be independent of the hedging strategy used. Denkl et al. [2013] investigates and compares the tracking error for the Δ-hedging strategy and a mean-variance optimal strategy. Neither of these, however, do any asymptotic analysis of ε_0^2 for a digital option in terms of finding the exact rate of decay of the error. The closest paper to that issue is Broden and Tankov [2011]. However, there the focus is more on the influence of adding jumps to any asymptotic results. Importantly, all the above studies study the tracking error of a continuous-time strategy followed in discrete time, instead of a strategy directly computed as optimal in discrete time.

We will now show how all three concepts introduced in the previous sections - granularity, tracking errors and quadratic hedging - relate to each other.

By Toft [1996] we know that the tracking error of a continuous-time Black-Scholes strategy on a discrete set of trading times can be decomposed into single-step errors between trading dates, as can its variance, and the single-step squared tracking error

has the form

$$\mathbb{E}_t[(V_t + \Delta_t \Delta S_t - C_{t+1})^2] = \left(\frac{1}{2}\Gamma_t S_t^2 \operatorname{Var}_t(R_{t+1})\right)^2 \delta$$

where V_t is the replicating portfolio held at time t, Δ_t is the Black-Scholes hedging strategy and C_{t+1} is the theoretical continuous-time price. Summing up over all timesteps, we get a total squared tracking error of

$$\varepsilon_0^2(\delta) = \sum_{t=0}^{T-1} \left(\frac{1}{2}\Gamma_t S_t^2 \operatorname{Var}_t(R_{t+1})\right)^2 \delta$$

By comparing the above to the granularity formula (1.4) one can already see where the granularity formula originates from.

In contrast to that, we know that the squared hedging error of the locally optimal strategy can also be decomposed into single-step errors. By Černý and Kallsen [2009, eqns (4.5), (4.9),(4.10)] we know the total error is

$$\mathbb{E}_0[(V_T - H)^2] = (x - V_0)^2 + \sum_{t=0}^{T-1} \mathbb{E}_0[\psi_t] \tag{1.5}$$

where

$$\psi_t = \mathbb{E}_t[(V_t + \xi_t \Delta S_t - V_{t+1})^2] \tag{1.6}$$

is the one-step conditional squared hedging error when the underlying S is discounted. Here x denotes initial capital, V the mean-value process and ξ the locally optimal hedge.

We know, due to the way the risk-minimizing locally optimal hedging strategy is constructed, that in discrete time it will minimize variance of the single-step hedging error and therefore should have a total error variance lower than the tracking error variance. However, we will show numerically that asymptotically the single-step tracking error of the Black-Scholes hedge is very close to the single-step conditional hedging error of the locally optimal hedging strategy, i.e.

$$\psi_t \approx \left(\frac{1}{2}\Gamma_t S_t^2 \operatorname{Var}_t(R_{t+1})\right)^2 \delta$$

and therefore, the granularity coefficients of both strategies will be the same.

15

The dynamically optimal mean-variance strategy has a "granularity" coefficient of its own, which is necessarily smaller than the one for the Black-Scholes hedge and the locally optimal strategy, due to it being a superior strategy by construction. The superiority to the Black-Scholes hedge is numerically demonstrated for exponential Lévy processes in Denkl et al. [2013]. Further, when the locally optimal hedge exists, whether under the IID return assumption or otherwise, the errors of the dynamic and local mean-variance strategy are closely linked, the total unconditional squared error of the dynamic being given as a (possibly stochastically) weighted sum of conditional local hedging errors, as given in Černý and Kallsen [2009, eqn. 4.21]:

$$\mathbb{E}_0[(G_T^x(\varphi(x, H)) - V_T)^2] = L_0(x - V_0)^2 + \sum_{t=1}^{T} \mathbb{E}_0[L_t \psi_t].$$

Here, L_t is the so-called opportunity process; for IID returns, it is deterministic. For non-IID returns, the situation becomes more complicated, since L_t becomes stochastic.

1.1.4 Our setup and the research question

In our analysis we will consider a locally optimal quadratic hedging strategy ξ over a fixed time period $[0, T]$ and trading on a set of times $t = 0, 1, ..., T - 1$. We consider a market with only two assets - a risk-free and a risky one. The risk-free asset will be a continuously compounded money market account with risk-free rate r, i.e. $\hat{S}_t^0 = e^{rt}$. To model the dynamics of the risky underlying under the physical measure $\mathbb{P}$, we will assume for simplicity that the discounted price process $S_t := e^{-rt}\hat{S}_t$ is a martingale, its evolution given by:

$$S_t = S_0 \exp\left(-\frac{1}{2}\sigma^2 t + \sigma W_t\right) = S_0 \mathcal{E}(\sigma W_t). \tag{1.7}$$

Here $\mathcal{E}(\cdot)$ denotes the stochastic exponential of Dóleans-Dade. We model the discounted price process as a martingale to retain elegant solutions throughout; an explicit solution with drift can be obtained, but it leads to unnecessarily lengthy algebra and would not provide much additional insight, since Hubalek et al. [2006] illustrates that the drift rate does not significantly affect variance-optimal hedging strategies and errors - therefore a simple martingale model is a good proxy for the full model. Already in Bertsimas et al. [2000, Figure 3], we see that under a ge-

ometric Brownian motion, the drift does not have any significant influence on the granularity of the discrete-time hedging error when doing Black-Scholes Δ-hedging. By Denkl et al. [2013, Lemma 3.5] we know that in our martingale setting the classic continuous-time Δ-hedging strategy and the continuous-time locally optimal hedging strategy coincide, which is why the afore-mentioned fact also holds for locally optimal (risk-minimizing) hedging.

We also look only at Brownian motion as the driver of randomness and omit any jump processes, because previous work by Tankov and Voltchkova [2009] and Broden and Tankov [2011] already shows that including jumps does not influence the order of convergence, although it does increase the absolute magnitude of the (squared) hedging error.

Since the dynamics are driven by a geometric Brownian motion, we are assuming that stock returns are IID, and thus the locally optimal strategy we are considering is well-defined. The strategy $(\xi_t)_{t=0\ldots T-1}$ will be $\mathcal{F}_t$-measurable, where $\mathcal{F}$ is the natural filtration of the Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Furthermore, it is in close relation to the dynamically optimal strategy, as we have mentioned before. Thus, the analysis on the locally optimal strategy can be a good proxy for results on the dynamically optimal strategy.

In the previous sections, we introduced the reader to the concepts of replication, Δ-hedging and mean-variance (or quadratic) hedging. We also discussed the recent progress made in the analysis of the asymptotics of tracking errors when using these continuous-time hedging strategies at discrete time intervals. Importantly, we saw that the order of decay of the error with decreasing time intervals between trades was dependent on the smoothness of the payoff function of the contingent claim we are replicating, as was demonstrated in Gobet and Temam [2001b].

Not only is there a difference in decay for a claim with a discontinuous payoff, but the standard granularity (1.4) is infinite, and hence ill-defined. On the other hand, we know that, for some small fixed $\eta > 0$, up to time $T - \eta$ before maturity, the result by Bertsimas et al. [2000] has to hold, since the price (in a Black-Scholes setting) or the mean-value process (in a quadratic hedging setting) is sufficiently smooth to satisfy all the conditions required for the standard formula to work. The aim of this first chapter is to put these two facts together and obtain a description of how hedging errors behave for digital options over the entire time interval $[0, T]$. Specifically, we will investigate how the one-step hedging errors accumulate and by doing so derive a more general understanding of "granularity".

The contribution of this chapter is two-fold. First, we will show that although the granularity formula (1.4) is infinite for a digital call option, we can control this explosive behaviour by a minor correction to the original formula. Furthermore, thanks to our direct computational approach, it becomes clear that the granularity formula is still highly relevant to risk management of the option position over its lifetime even in the case of a digital call. Specifically, we will show that the worse overall order of convergence $\mathcal{O}(\sqrt{\delta})$ of the total squared hedging error for a derivative with a discontinuity in its payoff is caused by portfolio rebalancing in the time interval nearing maturity $[T - \eta, T]$ for some fixed η, and the order of decay over the previous life of the derivative was the original $\mathcal{O}(\delta)$, with granularity (1.4). Secondly, we will show that for the martingale case, the granularity of the locally optimal quadratic hedging error is the same as that of the Black-Scholes tracking error in the case of a digital option.

In the following sections we will proceed as follows: in section 1.2, we will investigate and compare the order of decay of the one-step locally optimal mean-variance hedging error at the very last time-step for the cases of a vanilla and digital call. We will observe various rates of decay, which establish ground for our hypothesis. In section 1.3, we will go deeper into analyzing the hedging errors of a digital option and show that up to a cut-off time $T - \eta$, the digital option retains the same granularity as a vanilla call. We will also show that this relation breaks down for sufficiently small η, when the term of order $\mathcal{O}(\sqrt{\delta})$ will begin to dominate. This will then lead us to an asymptotic formula for the total squared hedging error of a digital call. In section 1.4, we will relate the results to the Black-Scholes tracking error, contrast our results with those of Gobet and Temam [2001b] (correcting their original formula) and illustrate the importance of the second-order term numerically.

1.2 Convergence of errors - heuristic analysis of last step

In this section we will perform a heuristic analysis of the behaviour of hedging errors at the very last rebalancing date when using a locally optimal mean-variance hedging strategy for the cases of a vanilla and digital call. We will see that the vanilla and digital call errors behave differently for $\delta \to 0$, with the vanilla call error diminishing to zero much faster than that of a digital call. This result will serve as motivation to further pursue the puzzling behaviour of a digital option, since we observe that at the final timestep, the convergence rate is of order $\mathcal{O}(\sqrt{\delta})$; we already know from the literature, however, that it is of order $\mathcal{O}(\delta)$ anywhere before this last time-step.

1.2.1 Last step error of a call option

In this section we will compute the last-step hedging error of a vanilla call. By equation (1.3) and Černý and Kallsen [2009, eqn. (4.9)], we know that the one-step conditional hedging error (1.6) can be written as

$$\psi_t = \mathrm{Var}_{t-1}(V_t) - \left(\frac{\mathrm{Cov}_{t-1}(V_t, \Delta S_t)}{\mathrm{Var}_{t-1}(\Delta S_t)} \right)^2 \mathrm{Var}_{t-1}(\Delta S_t),$$

where V_t is the mean-value process at time t of the claim with payoff H at time T. For the last time-step, $t = N\delta = T$, we know that the mean value process of the vanilla call option has value $V_N = H = f(S_N) = (S_N - K)^+ = (S_N - K)1_{S_N > K}$.

In this section, we will, for simplicity, drop index N; any variable X that is to be understood at time $t_{N-1} = (N-1)\delta$ will be denoted X_-. In this notation, $V = f(S) = (S - K)1_{S > K}$. The hedging error at the last step will be

$$\psi = \mathrm{Var}_-(f(S)) - \left(\frac{\mathrm{Cov}_-(f(S), S)}{\mathrm{Var}_-(S)} \right)^2 \mathrm{Var}_-(S). \tag{1.8}$$

Furthermore, we will denote the stock price at time $T - \delta$ as $S_- = Ke^y$, where y measures the log deviation of the stock price from the strike. The stock price at time T will then be $S = K \exp\left(y - \frac{1}{2}s^2 + sZ\right)$, where $Z \sim N(0, 1)$ and $s = \sigma\sqrt{\delta}$.

Finally, we introduce one more piece of notation. We will denote a standard normal

cumulative distribution function (CDF) by $\Phi(\cdot)$. Furthermore, We will write $\Phi(\cdot, \cdot; \rho)$ to denote a bivariate normal CDF with correlation coefficient ρ, zero mean and unit variances. Throughout, we will use the shorthand notation

$$\Phi(x; \rho) := \Phi(x, x; \rho), \qquad \Phi(x) := \Phi(x, x; 1), \tag{1.9}$$

and we will note that in this notation,

$$\Phi^2(x) = \Phi(x, x; 0) = \Phi(x; 0). \tag{1.10}$$

The probability distribution function (PDF) of a standard normal variable will be denoted as $\varphi(\cdot)$.

Theorem 1.1. *Under the stock price model* (1.7), *the quadratic hedging error ψ at time $T - \delta$ for a vanilla call option is given as*

$$\psi = g\left(\frac{\log(S_-/K)}{\sigma\sqrt{\delta}}, \sigma\sqrt{\delta}\right)$$

where

$$
\begin{aligned}
g(z, s) := K^2 \Bigg[& e^{2sz+s^2}\Phi\left(z + \frac{3s}{2}\right) - 2e^{sz}\Phi\left(z + \frac{s}{2}\right) + \Phi\left(z - \frac{s}{2}\right) - \\
& e^{2sz}\Phi^2\left(z + \frac{s}{2}\right) + 2e^{sz}\Phi\left(z + \frac{s}{2}\right)\Phi\left(z - \frac{s}{2}\right) - \Phi^2\left(z - \frac{s}{2}\right) - \\
& \frac{1}{e^{s^2} - 1}\left(e^{sz+s^2}\Phi\left(z + \frac{3s}{2}\right) - (e^{sz} + 1)\Phi\left(z + \frac{s}{2}\right) + \Phi\left(z - \frac{s}{2}\right)\right)^2 \Bigg].
\end{aligned}
\tag{1.11}
$$

Proof. To compute the hedging error ψ explicitly, we will evaluate formula (1.8). To do that, we need to compute 3 terms: $i)\ \mathrm{Var}_-(S), ii)\ \mathrm{Var}_-(f(S)), iii)\ \mathrm{Cov}_-(f(S), S)$. The first item is easily evaluated directly, and after some integral computations we get that

$$
\begin{aligned}
\mathrm{Var}_-(S) &= \mathbb{E}_-[(S - S_-)^2] = \mathbb{E}[(Ke^{y-s^2/2+sZ} - Ke^y)^2] \\
&= K^2 e^{2y}\mathbb{E}[(e^{-s^2/2+sZ} - 1)^2] = K^2 e^{2y}(e^{s^2} - 1).
\end{aligned}
$$

For the second item, we first observe that

$$\mathrm{Var}_-(f(S)) = \mathbb{E}_-[f^2(S)] - \mathbb{E}_-^2[f(S)] = \mathbb{E}_-[(S - K)^2 1_{S>K}] - \mathbb{E}_-^2[(S - K)1_{S>K}].$$

20

We now compute each of the two parts separately, using lemma A.3:

$$\mathbb{E}_-[(S-K)^2 1_{S>K}] = \mathbb{E}_-[S^2 1_{S>K}] - 2K\mathbb{E}_-[S1_{S>K}] + K^2\mathbb{E}_-[1_{S>K}]$$

$$= K^2 \left[e^{2y+s^2} \Phi\left(\frac{y}{s} + \frac{3s}{2}\right) - 2e^y \Phi\left(\frac{y}{s} + \frac{s}{2}\right) + \Phi\left(\frac{y}{s} - \frac{s}{2}\right) \right]$$

$$\mathbb{E}_-[(S-K)1_{S>K}] = \mathbb{E}_-[S1_{S>K}] - K\mathbb{E}_-[1_{S>K}]$$

$$= K \left[e^y \Phi\left(\frac{y}{s} + \frac{s}{2}\right) - \Phi\left(\frac{y}{s} - \frac{s}{2}\right) \right].$$

Finally, using lemma A.3 again, we evaluate the covariance:

$$\text{Cov}_-(f(S), S) = \mathbb{E}_-[f(S)(S - S_-)]$$

$$= \mathbb{E}_-[S^2 1_{S>K}] - (S_- + K)\mathbb{E}_-[S1_{S>K}] + S_- K\mathbb{E}_-[1_{S>K}]$$

$$= K^2 \left[e^{2y+s^2} \Phi\left(\frac{y}{s} + \frac{3s}{2}\right) - (e^y + 1)e^y \Phi\left(\frac{y}{s} + \frac{s}{2}\right) + e^y \Phi\left(\frac{y}{s} - \frac{s}{2}\right) \right]$$

When we put all of the above together, we obtain the desired result. □

We see that the hedging error is a function of $s = \sigma\sqrt{\delta}$, the standard deviation of the stock price over one time step, and $z = y/s$, the number of standard deviations the log stock price is away from the log option strike (a particular scaling of log moneyness).

We can gain insight into the properties of the complicated formula (1.11) graphically. When plotting ψ in variable $z = y/(\sigma\sqrt{\delta})$, we can see what this function looks like and how it behaves for $s \to 0$, i.e. $\delta \to 0$ numerically, as seen in figure 1.1. We observe that the function ψ decays to 0 as we send $\delta \to 0$. But ultimately, we want to know the behaviour of ψ as seen from time $t = 0$, i.e. we wish to compute $\mathbb{E}_0[\psi]$, as we are evaluating the *expected* hedging error. In the next section we provide a general theorem that gives us a clear list of assumptions on $g(\cdot, \cdot)$ under which we can easily compute the expected hedging error $\mathbb{E}_0[\psi]$ at some time close to maturity. We are interested in times close to maturity, because as we will later show, these are the times where the asymptotics of a digital option differ from a regular call option.

1.2.1.1 Towards computing $\mathbb{E}_0[\psi]$ - a Taylor expansion

Here we provide a general theorem that allows us to compute $\mathbb{E}_0[\psi]$ when ψ has the functional form $g(\cdot, \cdot)$ we showed in the previous section.

Theorem 1.2. *At fixed time $t = n\delta$ and taking a fixed $\eta \geq 0$ s.t. $T-t \leq \eta$ ($T = N\delta$),*
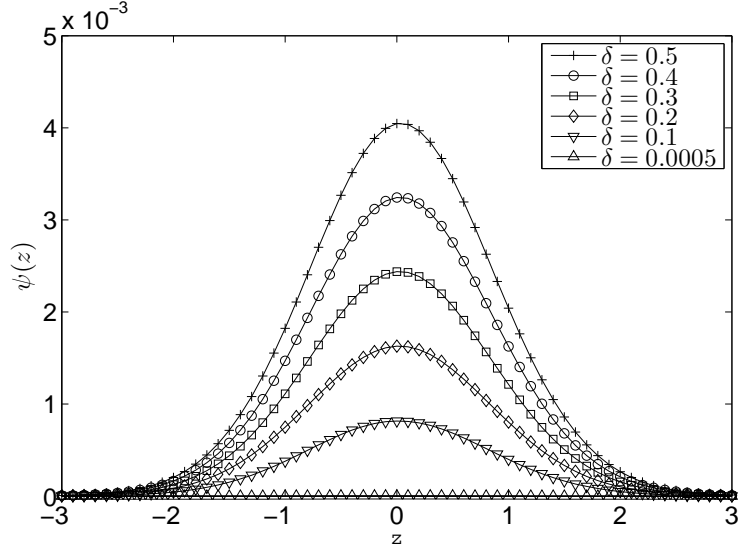
Figure 1.1: The diminishing error ψ for $\delta \to 0$. $K = 1$, $\sigma = 0.3$.

we are given a function $g(z,s)$ that has a Taylor expansion (as given by lemma (A.4)) in the form:

$$g(z,s) = g(z,0) + \frac{\partial}{\partial s}g(z,0)s + \frac{\partial^2}{\partial s^2}g(z,a)s^2; 0 < a < s$$

with derivatives uniformly bounded in z, and furthermore:

- $f(z) := g(z,0)$ is well defined and is an even function,

- $\int_{\mathbb{R}} z^2 f(z)\,dz < \infty$,

- $\frac{\partial}{\partial s}g(z,s)_{|s=0} = 0$,

- $g(z,s)$ is uniformly bounded, i.e. $|g(z,s)| < \infty \; \forall (z,s) : z \in \mathbb{R}, s \in [0, \sigma\eta]$

Let $y_n \sim N\left(y_0 - \frac{1}{2}\sigma^2 t, \sigma^2 t\right)$ be a normally distributed random variable, where $y_n = \log \frac{S_n}{K}$. Then it holds that

$$\mathbb{E}_0\left[g\left(\frac{y_n}{\sigma\sqrt{\delta}}, \sigma\sqrt{\delta}\right)\right] = \int_{\mathbb{R}} f(z)\,dz \frac{1}{\sqrt{2\pi\sigma^2 T}} \exp\left(-\frac{1}{2}\frac{\left(\log \frac{S_0}{K} - \frac{1}{2}\sigma^2 T\right)^2}{\sigma^2 T}\right)\sigma\sqrt{\delta}$$
$$+ \mathcal{O}(\delta^{3/2}) + \mathcal{O}(\eta\sqrt{\delta}).$$

Proof. We are attempting to compute

$$\mathbb{E}_0\left[g\left(\frac{y_n}{\sigma\sqrt{\delta}}, \sigma\sqrt{\delta}\right)\right] = \int_{\mathbb{R}} g\left(\frac{x}{\sigma\sqrt{\delta}}, \sigma\sqrt{\delta}\right) p\left(t, x\right) \, dx$$

where

$$p(t,x) := \frac{1}{\sqrt{2\pi\sigma^2 t}} \exp\left(-\frac{1}{2}\frac{\left(x - y_0 + \frac{1}{2}\sigma^2 t\right)^2}{\sigma^2 t}\right).$$

First, we transform variables:

$$\mathbb{E}_0\left[g\left(\frac{y_n}{\sigma\sqrt{\delta}}, \sigma\sqrt{\delta}\right)\right] = \int_{\mathbb{R}} g\left(\frac{x}{\sigma\sqrt{\delta}}, \sigma\sqrt{\delta}\right) p\left(t, x\right) \, dx = \left| z = \frac{x}{\sigma\sqrt{\delta}} \right|$$

$$= \int_{\mathbb{R}} g\left(z, \sigma\sqrt{\delta}\right) p(t, \sigma\sqrt{\delta}z) \, dz \, \sigma\sqrt{\delta}$$

By Taylor expansion and the assumption $\frac{\partial}{\partial s}g(z,s)_{|s=0} = 0$, it holds that

$$g(z, \sigma\sqrt{\delta}) = g(z, 0) + \frac{\partial^2}{\partial s^2}g(z, s)_{|s=a}\sigma^2\delta; 0 < a < \sigma\sqrt{\delta}$$

Using the fact that $g(z, 0) = f(z)$, the derivatives of g are bounded and the function $p(t, z)$ is uniformly bounded and decays to 0 as $z \to 0$, we can substitute into the previous equation to get:

$$\mathbb{E}_0\left[g\left(\frac{y_n}{\sigma\sqrt{\delta}}, \sigma\sqrt{\delta}\right)\right] = \int_{\mathbb{R}} f(z)p(t, \sigma\sqrt{\delta}z) \, dz \, \sigma\sqrt{\delta} + \mathcal{O}(\delta^{3/2})$$

Our goal now is to make the density function $p(t, z)$ independent of the integral and centered around T, i.e. expand the function around the point $(T, 0)$. Taylor's theorem A.4 and the fact that the derivatives of density functions are bounded gives us:

$$p(t, \sigma\sqrt{\delta}z) = p(T, \sigma\sqrt{\delta}z) - \frac{\partial}{\partial t}p(t, \sigma\sqrt{\delta}z)_{|t=b}(T - t); t < b < T$$

We have assumed that $(T-t) \leq \eta$. Together with the uniform boundedness of density functions and their derivatives, we can make a worst-case estimate of the order of the Taylor expansion error and write

$$p(t, \sigma\sqrt{\delta}z) = p(T, \sigma\sqrt{\delta}z) + R(t, z),$$

where the remainder $R(t, z)$ is uniformly bounded by a fixed $C\eta$, i.e.

$$\forall (t, z) : R(t, z) \le C\eta.$$

Now we expand $p(T, \sigma\sqrt{\delta}z)$ in variable z around point 0:

$$p(T, \sigma\sqrt{\delta}z) = p(T, 0) + \frac{\partial}{\partial z}p(T, 0)\,\sigma\sqrt{\delta}z + \frac{1}{2}\frac{\partial^2}{\partial z^2}p(T, z)_{|z=c}\,\sigma^2\delta z^2; 0 < c < z.$$

Computing the derivatives, we find that

$$\frac{\partial}{\partial z}p(t, z) = p(t, z)\left(-\frac{z - y_0 + \frac{1}{2}\sigma^2 t}{\sigma^2 t}\right), \quad \frac{\partial^2}{\partial z^2}p(t, z) = p(t, z)\left[\frac{(z - y_0 + \frac{1}{2}\sigma^2 t)^2}{\sigma^4 t^2} - \frac{1}{\sigma^2 t}\right].$$

Thus for all z it holds that

$$p(T, \sigma\sqrt{\delta}z) = p(T, 0)\left[1 + \frac{y_0 - \frac{1}{2}\sigma^2 T}{\sigma T}\sqrt{\delta}z\right] + \frac{1}{2}p(T, c)\left[\frac{(c - y_0 + \frac{1}{2}\sigma^2 T)^2}{\sigma^2 T^2} - \frac{1}{T}\right]\delta z^2.$$

Putting it all together:

$$p(t, \sigma\sqrt{\delta}z) = p(T, \sigma\sqrt{\delta}z) + R(t, z)$$
$$= p(T, 0)\left[1 + \frac{y_0 - \frac{1}{2}\sigma^2 T}{\sigma T}\sqrt{\delta}z\right] + \frac{1}{2}p(T, c)\left[\frac{(c - y_0 + \frac{1}{2}\sigma^2 T)^2}{\sigma^2 T^2} - \frac{1}{T}\right]\delta z^2 + R(t, z)$$

Therefore

$$\int_{\mathbb{R}} f(z)p(t, \sigma\sqrt{\delta}z)\,dz\,\sigma\sqrt{\delta} = \int_{\mathbb{R}} f(z)\,dz\,p(T, 0)\sigma\sqrt{\delta} + \frac{y_0 - \frac{1}{2}\sigma^2 T}{T}\int_{\mathbb{R}} zf(z)\,dz\,\sigma\delta$$
$$+ \frac{1}{2}p(T, c)\left[\frac{(c - y_0 + \frac{1}{2}\sigma^2 T)^2}{\sigma^2 T^2} - \frac{1}{T}\right]\int_{\mathbb{R}} z^2 f(z)\,dz\,\sigma\delta^{3/2} + \int_{\mathbb{R}} f(z)R(t, z)\,dz\sigma\sqrt{\delta}.$$

The integral of the remainder is easily addressed, as it holds that

$$\int_{\mathbb{R}} f(z)R(t, z)\,dz\,\sigma\sqrt{\delta} \le \int_{\mathbb{R}} f(z)\,dz C\eta\sigma\sqrt{\delta},$$

and hence the term is, in the worst possible case, of order $\mathcal{O}(\sqrt{\delta}\eta)$.

Next, we note that the second term vanishes, as the integral $\int_{\mathbb{R}} zf(z)\,dz = 0$, since we have assumed that $f(z)$ is an even function. Finally, we have assumed that

$\int_{\mathbb{R}} z^2 f(z) \, dz < \infty$, which ensures that the term of order $\mathcal{O}(\delta^{3/2})$ is finite.

Therefore the approximation of $\mathbb{E}_0 \left[g \left(y_n/(\sigma\sqrt{\delta}), \sigma\sqrt{\delta} \right) \right]$ is given as

$$
\begin{aligned}
\mathbb{E}_0 \left[g \left(\frac{y_n}{\sigma\sqrt{\delta}}, \sigma\sqrt{\delta} \right) \right] &= \int_{\mathbb{R}} f(z) p(t, \sigma\sqrt{\delta}z) \, dz \, \sigma\sqrt{\delta} \\
&= \int_{\mathbb{R}} f(z) \, dz \, p(T, 0) \sigma\sqrt{\delta} + \mathcal{O}(\delta^{3/2}) + \mathcal{O}(\sqrt{\delta}\eta).
\end{aligned}
$$

$\square$

1.2.1.2 Asymptotic expansion for vanilla call hedging error

As we showed in theorem 1.2, the functional form $g(z, s)$ of the error $\psi = g(z, s)$ in terms of variable z from theorem 1.1 has a significance when computing the asymptotics of the hedging error near maturity in the sense that $f(z) = \lim_{s \to 0+} g(z, s)$ contributes to the multiplier on the term of order $\mathcal{O}(\sqrt{\delta})$. Since the functional form $g(z, s)$ from 1.1 satisfies all the assumptions of theorem 1.2, the time $t = 0$ expectation of the one-step hedging error at terminal time T can be written in the form

$$
\mathbb{E}_0[\psi] = C \left(\int_{\mathbb{R}} f(z) dz \right) \sigma\sqrt{\delta} + \mathcal{O}(\delta^{3/2}),
$$

where C is a variable independent of δ and $f(z) = \lim_{s \to 0+} g(s, z)$. Here we note that this formula is not dependent on time to maturity η because we apply it for $\eta = 0$.

Therefore, if $f(z) \equiv 0$, the integral in the formula above will be 0 and hence the hedging error will decay at a rate $\mathcal{O}(\delta^{3/2})$:

$$
\mathbb{E}_0[\psi] = \mathcal{O}(\delta^{3/2})
$$

From the numerical results depicted in figure 1.1, we would expect this will hold for a standard call option.

In the theorem that follows, we not only find that the function (1.11) asymptotically decays to 0 as $s \to 0$, but we also find that it does so at a speed of the order $\mathcal{O}(s^2)$ (which implies a rate of $\mathcal{O}(\delta)$ when we substitute $s = \sigma\sqrt{\delta}$). It provides us with a Taylor expansion approximation of $\psi = g(z, s)$ in variable $s = \sigma\sqrt{\delta}$.

Theorem 1.3. *Under model (1.7), for the last-step mean-variance hedging error*

$\psi = g\left(\sigma\sqrt{\delta}, \frac{y_n}{\sigma\sqrt{\delta}}\right)$ *of a call option it holds that*

$$\lim_{s \to 0} s^{-2} g(z, s) = K^2 \left[\Phi(z) + z\varphi(z) + z^2\Phi(z) - (\varphi(z) + z\Phi(z))^2\right] \tag{1.12}$$

where z and s are defined as in Theorem 1.1 and $\varphi(x) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}x^2)$ is the probability density function of a standard normal distribution. Furthermore,

$$f(z) = \lim_{s \to 0^+} s^0 g(z, s) \equiv 0.$$

Proof. What we ultimately want to compute is a Taylor expansion of $g(z, s)$ in variable s around point $(z, 0)$ up to order $\mathcal{O}(s^2)$. To Taylor expand $g(z, s)$ as given by equation (1.11), we need to use the following Taylor expansions of the building blocks:

$$\Phi\left(z - \frac{s}{2}\right) = \Phi(z) - \frac{s}{2}\varphi(z) + \frac{1}{2}\varphi'(z)\frac{s^2}{4} + o(s^2),$$

$$\Phi\left(z + \frac{s}{2}\right) = \Phi(z) + \frac{s}{2}\varphi(z) + \frac{1}{2}\varphi'(z)\frac{s^2}{4} + o(s^2),$$

$$\Phi\left(z + \frac{3s}{2}\right) = \Phi(z) + \frac{3s}{2}\varphi(z) + \frac{1}{2}\varphi'(z)\frac{9s^2}{4} + o(s^2),$$

$$e^{s^2} - 1 = s^2 + o(s^2).$$

Indeed, we substitute these asymptotic estimates into (1.11) and after some algebraic manipulations, we get that

$$g(z, s) = s^2 K^2 \left[\Phi(z) + z\varphi(z) + z^2\Phi(z) - (\varphi(z) + z\Phi(z))^2\right] + o(s^2),$$

which directly leads us to our desired results. $\qquad\square$

We verify the results of this theorem numerically, to make sure that the asymptotic form for $g(z, s)$ given by (1.12) is a good approximation for the full formula (1.11). We can see in figure 1.2 that the approximation is very exact indeed even for reasonably large values of δ - the value $\delta = 0.1$ corresponds to rebalancing every 25 days (i.e. roughly monthly) when measured in years and considering a year with 252 business days. Using the above theorem we now compute $\mathbb{E}_0[\psi] = C\left(\int_{\mathbb{R}} f(z)\,dz\right)\sigma\sqrt{\delta} + \mathcal{O}(\delta^{3/2}) = 0 + \mathcal{O}(\delta^{3/2}) = \mathcal{O}(\delta^{3/2})$, to confirm that asymptotically, the hedging error at maturity as seen from time $t = 0$ will decay at a rate $\mathcal{O}(\delta^{3/2})$. In the next section we will see how this differs from results for a digital option.
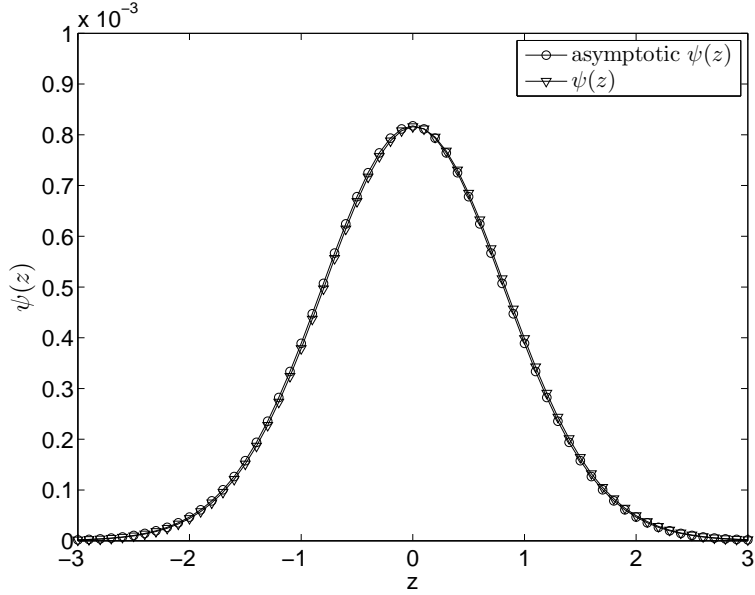
Figure 1.2: Comparison of the asymptotic approximation of ψ (1.12) and full function (1.11). $K = 1$, $\sigma = 0.3$, $\delta = 0.1$.

1.2.2 Last step error of a digital option

In this section we will derive the hedging error at the last rebalancing date for a digital option in the same spirit as we did for a vanilla call and we will point out how it differs from that of a vanilla call option. Just as before, we want to compute (1.8), but this time taking $f(S) = 1_{S>K}$.

Theorem 1.4. *Under the stock price model* (1.7), *the mean-variance hedging error at time* $T - \delta$ *for a digital call option is given as* $\psi = g\left(\frac{\log(S_-/K)}{\sigma\sqrt{\delta}}, \sigma\sqrt{\delta}\right)$, *where*

$$g(z, s) = \Phi\left(z - \frac{s}{2}\right) - \Phi^2\left(z - \frac{s}{2}\right) - \frac{1}{e^{s^2} - 1}\left(\Phi\left(z + \frac{s}{2}\right) - \Phi\left(z - \frac{s}{2}\right)\right)^2. \quad (1.13)$$

Proof. Just as for the vanilla call, we need to compute $\text{Var}_-(S)$, $\text{Var}_-(f(S))$ and $\text{Cov}_-(f(S), S)$ by formula 1.8. Since $\text{Var}_-(S)$ is the same in both cases, we needn't compute it again.

To compute $\text{Var}_-(f(S))$, we again make use of the decomposition:

$$\text{Var}_-(f(S)) = \mathbb{E}_-[f^2(S)] - \mathbb{E}_-^2[f(S)] = \mathbb{E}_-[1_{S>K}] - \mathbb{E}_-^2[1_{S>K}]$$

Using lemma A.3, we easily see that

$$\mathrm{Var}_-(1_{S>K}) = \Phi\left(\frac{y}{s} - \frac{s}{2}\right) - \Phi^2\left(\frac{y}{s} - \frac{s}{2}\right)$$

The covariance term is similarly first decomposed

$$\mathrm{Cov}_-(f(S), S) = \mathbb{E}_-[f(S)(S - S_-)] = \mathbb{E}_-[S1_{S>K}] - Ke^y\mathbb{E}_-[1_{S>K}],$$

and by virtue of A.3 we again easily obtain the result

$$\mathrm{Cov}_-(f(S), S) = Ke^y\left(\Phi\left(\frac{y}{s} + \frac{s}{2}\right) - \Phi\left(\frac{y}{s} - \frac{s}{2}\right)\right).$$

Now we have all we need to put together the last step hedging error (1.8) for a digital option. □

We will again analyze this result numerically to get better insight. Figure 1.3 shows us that asymptotically, as $\delta \to 0$, the total error does not decay to 0 for all values of z, but instead converges to a function of z which is non-zero for approximately 3 standard deviations of the stock price from the strike.
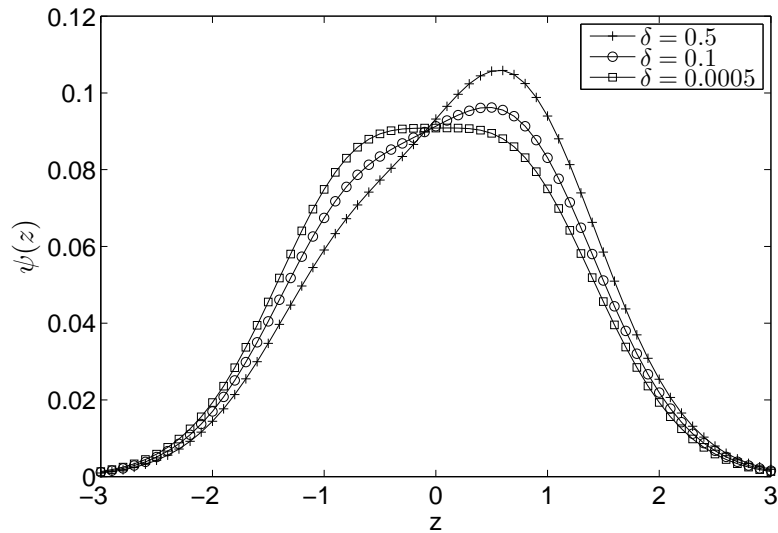
Figure 1.3: The error ψ for $\delta \to 0$. $\sigma = 0.3$.

Theorem 1.5. *For the last-step mean-variance hedging error (1.13) of a digital call option $\psi = g\left(\frac{\log S_-/K}{\sigma\sqrt{\delta}}, \sigma\sqrt{\delta}\right)$ it holds that*

$$f(z) = \lim_{s \to 0+} s^0 g(z, s) = \Phi(z) - \Phi^2(z) - \varphi^2(z)$$

28

where $\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}x^2)$ *is the probability density function of a standard normal distribution.*

Proof. The proof is analogous to that of theorem 1.3. To compute the limit $f(z)$, we can directly take the limit in the first two terms of (1.13) and are left with the fraction. There, we use the same Taylor expansions as in theorem 1.3 to compute the limit of the fraction in formula (1.13) and after some algebraic manipulations we obtain our result. $\square$

We again verify numerically that this asymptotic approximation $f(z)$ for $\delta \approx 0$ is correct and uniformly converges to the full function ψ for small δ, as seen in figure 1.4.
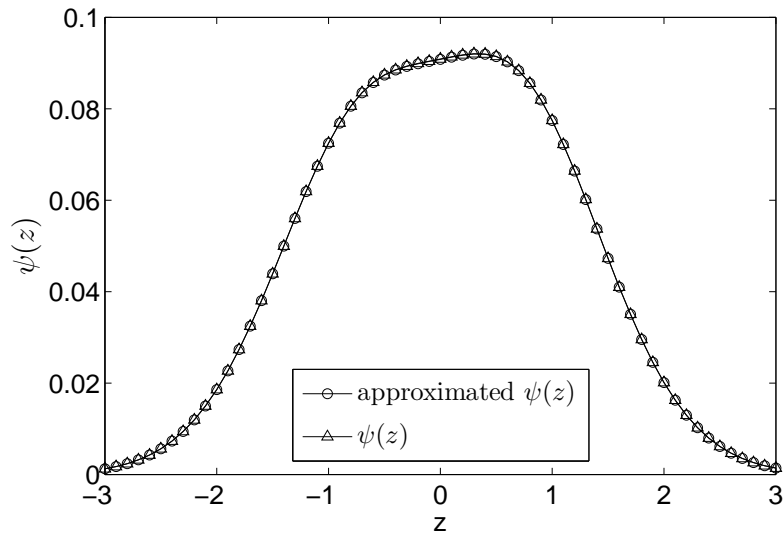
Figure 1.4: Comparison of the asymptotic approximation via $f(z)$ and full function ψ. $\sigma = 0.3$, $\delta = 0.01$.

We thus obtain that because $f(z) \neq 0$, the leading term in the asymptotic expansion of the expectation of the terminal error is of order $\mathcal{O}(\sqrt{\delta})$, since

$$\mathbb{E}_0[\psi] = C \int_{\mathbb{R}} f(z) \, dz \sigma \sqrt{\delta} + \mathcal{O}(\delta^{3/2}) = \mathcal{O}(\sqrt{\delta}).$$

Thus we obtain results in line with those from the literature on tracking errors in a Δ-hedging setup: the overall error will decay at a rate of $\mathcal{O}(\sqrt{\delta})$. At the same time we stated that for most of the life of the option, however, the digital call error behaves in the same fashion as a vanilla call. In this section, we have managed to pinpoint

at which point the difference in convergence appears. Specifically, we have verified that the slower convergence for a digital option is caused by rebalancing errors very close to expiry, where the assumptions of Bertsimas et al. [2000] fail to hold and the mean-value process loses its smoothness.

1.3 Granularity of claims with discontinuous pay-offs

In the previous section, we saw an analysis of the last step hedging errors of a vanilla and digital call. We concluded that the error diminishes at a slower rate for a digital option. In this section we will analyze the one-step hedging errors of a digital option in more depth, looking at the error made at any rebalancing date, not just the last one. We will investigate the sum of these errors as seen from time $t = 0$ and we will show how this sum relates to granularity (1.4). We will see that although formula (1.4) does not apply over the entire time interval $[0, T]$ of the duration of the contract, we can choose a parameter η such that the formula applies over the time interval $[0, T - \eta]$. This will mean that the error decay of order $\mathcal{O}(\sqrt{\delta})$ found in Gobet and Temam [2001b] is only caused by a few last rebalancing periods over period $[T - \eta, T]$. Finally, we will derive a correction term for the granularity formula in the case of a digital option and show how it is connected to the overall asymptotics of the digital option.

1.3.1 Hedging error for a digital call at one time step

We will now explicitly compute the one-step quadratic hedging error within the model given by (1.7). Let us consider an option with maturity $T = N\delta$, where N is the number of (equidistant) trading dates at which we rebalance our hedged portfolio and δ is the time interval between dates. We will denote the trading dates $t_n, n \in \{0, ..., N - 1\}$. In our discrete time setting, we will utilize the subscript n when referring to variables given at time t_n. The mean-value process $V_n = \mathbb{E}_n^{\mathbb{Q}}[1_{S>K}]$ of the binary option at any trading time t_n under our simple model is easily computed. In the martingale setting the variance-optimal measure $\mathbb{Q}$ coincides with the physical measure $\mathbb{P}$ and by lemma A.3 we find that

$$V_n = \Phi\left(\frac{\log \frac{S_n}{K} - \frac{1}{2}\sigma^2(T - t_n)}{\sigma\sqrt{T - t_n}}\right),$$

where $\Phi(\cdot)$ denotes the CDF of a standard normal random variable. Using a mean-variance strategy, we know that the one-step hedging error, as given in Černý [2007,

eqn. (3.11)], is

$$\psi_n = \mathrm{Var}_n(V_{n+1}) - \left(\frac{\mathrm{Cov}_n(S_{n+1}, V_{n+1})}{\mathrm{Var}_n(S_{n+1})} \right)^2 \mathrm{Var}_n(S_{n+1}). \qquad (1.14)$$

Conditionally on the value at time t_n, the stock price at time t_{n+1} will be given as

$$S_{n+1} = S_n \exp\left(-\frac{\sigma^2 \delta}{2} + \sigma\sqrt{\delta} Z_{n+1} \right),$$

where Z_{n+1} is a standard normally distributed random variable ($Z_{n+1} \sim N(0,1)$). Similarly as in section 1.2, we will use two variable transformations to simplify notation. First, we will be interested in the log moneyness $y_n = \log(S_n/K)$. Secondly, we will abbreviate the volatility over one time-step to the variable $s = \sigma\sqrt{\delta}$. In light of this new notation, our stock is given as

$$y_{n+1} = \log \frac{S_{n+1}}{K} = y_n - \frac{1}{2}s^2 + sZ_{n+1}$$

Finally, we will use $t_n = t_{N-k} = (N-k)\delta$ to rewrite the value of the digital option at times t_{N-k}:

$$V_{N-k} = \Phi\left(\frac{y_{N-k}}{s\sqrt{k}} - \frac{1}{2}s\sqrt{k} \right); k \in \{1, ..., N\},$$

where $k = N - n$. We will use the abbreviated notation

$$d_{N-k} := \frac{y_{N-k}}{s\sqrt{k}} - \frac{1}{2}s\sqrt{k}.$$

We remind the reader that $\Phi(\cdot, \cdot; \rho)$ denotes a bivariate standard normal CDF with correlation coefficient ρ, with shorthand notation

$$\Phi(x; \rho) := \Phi(x, x; \rho), \qquad \Phi(x) := \Phi(x, x; 1), \qquad \Phi^2(x) = \Phi(x, x; 0) = \Phi(x; 0).$$

In the next theorem, we compute the single-step hedging error ψ_n for any timestep t_n.

Theorem 1.6. *Under the stock price model* (1.7), *the mean-variance hedging error* ψ_n *at time* $t_n = n\delta$ *for a digital call option is given as*

$$\psi_n = \psi_{N-k} = \Phi\left(x; \frac{1}{k} \right) - \Phi^2(x) - \frac{1}{e^{s^2} - 1}\left(\Phi\left(x + \frac{s}{\sqrt{k}} \right) - \Phi(x) \right)^2 \quad k = 1, ..., N$$

$$(1.15)$$

where

$$x = \frac{y_{N-k}}{s\sqrt{k}} - \frac{1}{2}s\sqrt{k}, s = \sigma\sqrt{\delta}$$

Proof. Similarly to the previous section, this requires us to use formula (1.14) and hence evaluate three terms: i) $\text{Var}_n(S_{n+1})$, ii)$\text{Var}_n(V_{n+1}) = \mathbb{E}_n[V_{n+1}^2] - \mathbb{E}_n^2[V_{n+1}]$, iii) $\text{Cov}_n(V_{n+1}, S_{n+1}) = \mathbb{E}_n[V_{n+1}(S_{n+1} - S_n)]$.

We needn't compute the first term again, as it is identical to the result in section 1.2, theorem 1.1; let us recall that $\text{Var}_n(S_{n+1}) = K^2 \exp(2y_n)(\exp(s^2)-1)$. The remaining two items are less trivial. We know that $\mathbb{E}_n[V_{n+1}] = V_n$, since the underlying (and hence the mean-value process) is a martingale. The other expression $\mathbb{E}_n[V_{n+1}^2] = \mathbb{E}_{N-k}[V_{N-k+1}^2]$ can be evaluated thanks to Toft [1996, eqn. 52]:

$$\mathbb{E}_{N-k}[V_{N-k+1}^2] = \int_{\mathbb{R}} \Phi^2\left(\frac{x}{\sqrt{k-1}} + \frac{y_{N-k}}{s\sqrt{k-1}} - \frac{1}{2}s\frac{k}{\sqrt{k-1}}\right) \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}\,dx$$

$$= \Phi\left(\frac{y_{N-k}}{s\sqrt{k}} - \frac{1}{2}s\sqrt{k}, \frac{y_{N-k}}{s\sqrt{k}} - \frac{1}{2}s\sqrt{k}; \frac{1}{k}\right).$$

The covariance term can be converted into two simpler blocks:

$$\mathbb{E}_n[V_{n+1}(S_{n+1} - S_n)] = Ke^{y_n}\left(e^{-\frac{1}{2}s^2}\mathbb{E}_n[\Phi(d_{n+1})e^{sZ_{n+1}}] - \mathbb{E}_n[V_{n+1}]\right)$$

The first expectation can be evaluated thanks to Toft [1996, eqn. 49]:

$$\mathbb{E}_n[\Phi(d_{n+1})e^{sZ_{n+1}}] = \mathbb{E}_{N-k}[\Phi(d_{N-k+1})e^{sZ_{N-k+1}}]$$

$$= \int_{\mathbb{R}} \Phi\left(\frac{x}{\sqrt{k-1}} + \frac{y_{N-k}}{s\sqrt{k-1}} - \frac{1}{2}s\frac{k}{\sqrt{k-1}}\right) \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x^2-2sx)}\,dx$$

$$= e^{\frac{1}{2}s^2}\Phi\left(\frac{y_{N-k}}{s\sqrt{k}} - \frac{1}{2}s\sqrt{k} + \frac{s}{\sqrt{k}}\right)$$

The second expectation uses the martingale property as was the case before, hence $\mathbb{E}_n[V_{n+1}] = V_n$. When we plug in all the partial results into equation (1.14), we obtain our result. □

Let us notice that for $k = 1$, i.e. the last rebalancing $n = N-1$, $\Phi(x; 1/k) = \Phi(x; 1) = \Phi(x)$, i.e. the bivariate distribution will collapse to a one-dimensional standard normal distribution and the result will convert to what we obtained in the

previous section:

$$\psi_{N-1} = \Phi(x) - \Phi^2(x) - \frac{(\Phi(x+s) - \Phi(x))^2}{e^{s^2} - 1}$$

In figure 1.5 we see the shape of the one-step hedging error ψ_n with respect to the number of standard deviations away from the strike, i.e. $z_n = y_n/(\sigma\sqrt{\delta})$.
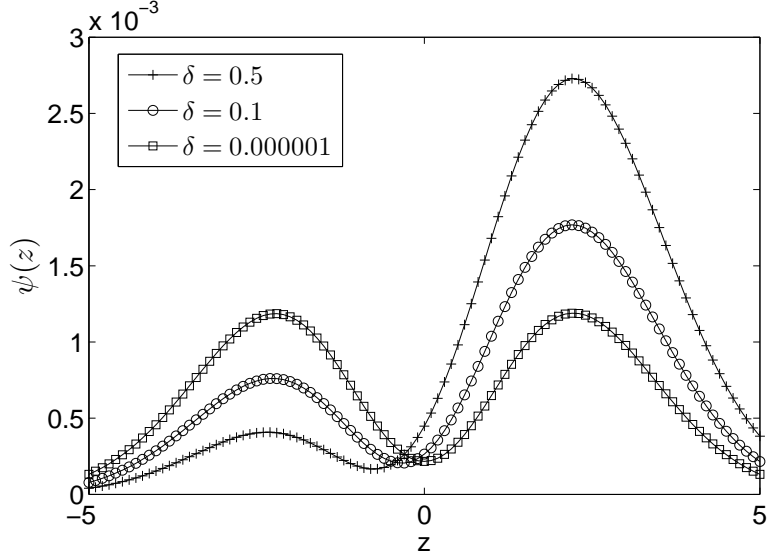
Figure 1.5: The error ψ_n for $\delta \to 0$. $\sigma = 0.3$, $n = 5$, $N = 10$.

1.3.1.1 Computing and analyzing $\mathbb{E}_0[\psi_n]$

Ultimately, we are interested in the total squared hedging error:

$$\varepsilon_0^2 = \sum_{n=0}^{N-1} \mathbb{E}_0[\psi_n]. \tag{1.16}$$

Because we are interested in computing (1.16), we need to know $\mathbb{E}_0[\psi_n]$, not just ψ_n. The random variable in ψ_n, as seen from time $t = 0$, is y_n, for which we know it holds that, unconditionally, $y_n \sim N(y_0 - \frac{1}{2}\sigma^2 t, \sigma^2 t)$.

Theorem 1.7. *For a digital call with maturity $T = N\delta$, the expected value at time*

$t = 0$ *of the mean-variance hedging error at time $t = n\delta$ is*

$$\mathbb{E}_0[\psi_n] = \Phi\left(x; \frac{n+1}{N}\right) - \Phi\left(x; \frac{n}{N}\right) - \qquad (1.17)$$

$$\frac{1}{e^{s^2} - 1}\left(\Phi\left(x + \frac{s}{\sqrt{N}}; \frac{n}{N}\right) - 2\Phi\left(x + \frac{s}{\sqrt{N}}, x; \frac{n}{N}\right) + \Phi\left(x; \frac{n}{N}\right)\right)$$

where $x = y_0/(s\sqrt{N}) - \frac{1}{2}s\sqrt{N}$, $s = \sigma\sqrt{\delta}$.

Proof. To compute $\mathbb{E}_0[\psi_n]$, we take expectations in formula (1.15) to get

$$\mathbb{E}_0[\psi_n] = \mathbb{E}_0[\psi_{N-k}]$$

$$= \mathbb{E}_0\left[\Phi\left(x; \frac{1}{k}\right)\right] - \mathbb{E}_0\left[\Phi^2(x)\right]$$

$$- \frac{1}{e^{s^2} - 1}\left(\mathbb{E}_0\left[\Phi\left(x + \frac{s}{\sqrt{k}}\right)^2\right] - 2\mathbb{E}_0\left[\Phi(x)\Phi\left(x + \frac{s}{\sqrt{k}}\right)\right] + \mathbb{E}_0\left[\Phi^2(x)\right]\right)$$

$$= A - B - \frac{1}{e^{s^2} - 1}\left(C - 2D + B\right),$$

where

$$x = \frac{y_0}{s\sqrt{k}} + \frac{1}{2}\frac{n}{\sqrt{k}}s + \sqrt{\frac{n}{k}}Z_n = aZ_n + b; \quad Z_n \sim N(0,1).$$

We now separately compute terms A, B, C, D. To compute them, we first use Toft [1996, eqn. 51] to obtain that

$$\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) \Phi\left(ax + b; \rho\right) dx = \Phi\left(\sqrt{\frac{1}{1+a^2}}b; \frac{a^2 + \rho}{1 + a^2}\right). \qquad (1.18)$$

For A we use equation (1.18) with $\rho = 1/k$ to obtain

$$A = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) \Phi\left(ax + b; \frac{1}{k}\right) dx$$

$$= \Phi\left(\sqrt{\frac{1}{1+a^2}}b; \frac{a^2 + \frac{1}{k}}{1+a^2}\right) = \Phi\left(\frac{y_0}{s\sqrt{N}} - \frac{1}{2}s\sqrt{N}; \frac{n+1}{N}\right).$$

For B, we recall the notation (1.10), and thus we set $\rho = 0$ in equation (1.18) to obtain

$$B = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) \Phi^2 (ax + b) \, dx$$

$$= \Phi\left(\sqrt{\frac{1}{1+a^2}}b; \frac{a^2}{1+a^2}\right) = \Phi\left(\frac{y_0}{s\sqrt{N}} - \frac{1}{2}s\sqrt{N}; \frac{n}{N}\right).$$

For C, we again set $\rho = 0$ in (1.18) to get

$$C = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) \Phi^2 \left(ax + b + \frac{s}{\sqrt{k}}\right) \, dx$$

$$= \Phi\left(\sqrt{\frac{1}{1+a^2}}\left(b + \frac{s}{\sqrt{k}}\right); \frac{a^2}{1+a^2}\right) = \Phi\left(\frac{y_0}{s\sqrt{N}} - \frac{1}{2}s\sqrt{N} + \frac{s}{\sqrt{N}}; \frac{n}{N}\right).$$

For D we require a slightly different formula, specifically Toft [1996, eqn. 52], to obtain that

$$D = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) \Phi\left(ax + b + \frac{s}{\sqrt{k}}\right) \Phi(ax + b) \, dx$$

$$= \Phi\left(\sqrt{\frac{1}{1+a^2}}\left(b + \frac{s}{\sqrt{k}}\right), \sqrt{\frac{1}{1+a^2}}b; \frac{a^2}{1+a^2}\right)$$

$$= \Phi\left(\frac{y_0}{s\sqrt{N}} - \frac{1}{2}s\sqrt{N} + \frac{s}{\sqrt{N}}, \frac{y_0}{s\sqrt{N}} - \frac{1}{2}s\sqrt{N}; \frac{n}{N}\right).$$

Thus we have managed to compute all the separate building blocks, giving us a complete analytic formula for the one-step hedging error as seen from time $t = 0$. $\quad\square$

We will now try to capture the asymptotic behaviour of this expectation. To compute the expectation asymptotically, we will use the fact that ψ_n can be seen as a function of the form $\psi_n = g_n\left(y_n/(\sigma\sqrt{\delta}), \sigma\sqrt{\delta}\right)$. As we saw in theorem 1.2, to compute asymptotics of the unconditional hedging error near maturity, we are interested in the form of the function $f_n(z) = \lim_{s\to 0} g_n(z, s)$. Let us first write the function $g_n(z, s)$ that gives us the one-step error formula (1.15) when we give it inputs $(y_n/(\sigma\sqrt{\delta}), \sigma\sqrt{\delta})$:

$$g_n(z, s) = g_{N-k}(z, s) = \Phi\left(\frac{z}{\sqrt{k}} - \frac{1}{2}s\sqrt{k}; \frac{1}{k}\right) - \Phi^2\left(\frac{z}{\sqrt{k}} - \frac{1}{2}s\sqrt{k}\right) \qquad (1.19)$$

$$- \frac{\left(\Phi\left(\frac{z}{\sqrt{k}} - \frac{1}{2}s\sqrt{k} + \frac{s}{\sqrt{k}}\right) - \Phi\left(\frac{z}{\sqrt{k}} - \frac{1}{2}s\sqrt{k}\right)\right)^2}{e^{s^2} - 1}.$$

We now want to take the limit $s \to 0$ to obtain $f_{N-k}(z)$.

Theorem 1.8. *Given the function $g_n(z,s)$ (1.19) connected to the mean-variance hedging error ψ_n (1.15) at time $t_n = n\delta$ of a digital call option, then*

$$f_{N-k}(z) = \lim_{s \to 0} s^0 g_{N-k}(z,s) = \Phi\left(\frac{z}{\sqrt{k}};\frac{1}{k}\right) - \Phi^2\left(\frac{z}{\sqrt{k}}\right) - \frac{1}{k}\varphi^2\left(\frac{z}{\sqrt{k}}\right) \qquad (1.20)$$

Proof. We see that the first two terms of (1.19) are not problematic and we can directly take the limit. For the last term, the limit is of the form $0/0$, thus we will have to use asymptotic arguments to estimate its form as $s \to 0$. Similarly as in section 1.2, we can use a Taylor expansion of the terms around the point $z/\sqrt{k}$ (see Theorem A.4):

$$\Phi\left(\frac{z}{\sqrt{k}} - \frac{1}{2}s\sqrt{k}\right) = \Phi\left(\frac{z}{\sqrt{k}}\right) - \varphi\left(\frac{z}{\sqrt{k}}\right)\frac{1}{2}s\sqrt{k} + \frac{1}{2}\varphi'\left(\frac{z}{\sqrt{k}}\right)\frac{1}{4}s^2 k + R(z)s^3$$

$$\Phi\left(\frac{z}{\sqrt{k}} - \frac{1}{2}s\sqrt{k} + \frac{s}{\sqrt{k}}\right) = \Phi\left(\frac{z}{\sqrt{k}}\right) + \varphi\left(\frac{z}{\sqrt{k}}\right)\left(\frac{s}{\sqrt{k}} - \frac{1}{2}s\sqrt{k}\right)$$

$$+ \frac{1}{2}\varphi'\left(\frac{z}{\sqrt{k}}\right)\left(\frac{s}{\sqrt{k}} - \frac{1}{2}s\sqrt{k}\right)^2 + R(z)s^3$$

$$e^{s^2} - 1 = s^2 + o(s^2).$$

Here $R(z)$ is a bounded function, as it is of the form $const \cdot \partial^2 \varphi(z)/\partial z^2$ and the derivatives of a normal distribution are bounded. We see that therefore asymptotically

$$\frac{\left(\Phi\left(\frac{z}{\sqrt{k}} - \frac{1}{2}s\sqrt{k} + \frac{s}{\sqrt{k}}\right) - \Phi\left(\frac{z}{\sqrt{k}} - \frac{1}{2}s\sqrt{k}\right)\right)^2}{e^{s^2} - 1}$$

$$= \frac{\left(\frac{s}{\sqrt{k}}\varphi\left(\frac{z}{\sqrt{k}}\right) + o(s^2)\right)^2}{s^2 + o(s^2)} = \frac{1}{k}\varphi^2\left(\frac{z}{\sqrt{k}}\right) + o(s^2).$$

We can also compute the asymptotics of the first two terms to see the complete

asymptotics of $g_{N-k}(z, s)$ in s:

$$\Phi\left(\frac{z}{\sqrt{k}} - \frac{1}{2}s\sqrt{k}; \frac{1}{k}\right) = \Phi\left(\frac{z}{\sqrt{k}}; \frac{1}{k}\right) - s\sqrt{k}\int_{-\infty}^{\frac{z}{\sqrt{k}}} \varphi\left(x, \frac{z}{\sqrt{k}}; \frac{1}{k}\right) dx$$

$$+ \frac{1}{4}s^2 k\left[\varphi\left(\frac{z}{\sqrt{k}}, \frac{z}{\sqrt{k}}; \frac{1}{k}\right) - \frac{1}{1-(1/k)^2}\int_{-\infty}^{\frac{z}{\sqrt{k}}}\left(\frac{z}{\sqrt{k}} - \frac{x}{k}\right)\varphi\left(x, \frac{z}{\sqrt{k}}; \frac{1}{k}\right) dx\right]$$

$$+ o(s^2)$$

$$\Phi^2\left(\frac{z}{\sqrt{k}} - \frac{1}{2}s\sqrt{k}\right) = \Phi^2\left(\frac{z}{\sqrt{k}}\right) - s\sqrt{k}\Phi\left(\frac{z}{\sqrt{k}}\right)\varphi\left(\frac{z}{\sqrt{k}}\right)$$

$$+ \frac{1}{4}s^2 k\left[\varphi^2\left(\frac{z}{\sqrt{k}}\right) - \frac{1}{2}z\varphi\left(\frac{z}{\sqrt{k}}\right)\Phi\left(\frac{z}{\sqrt{k}}\right)\right] + o(s^2)$$

Therefore we get that

$$g_{N-k}(z, s) = \Phi\left(\frac{z}{\sqrt{k}}; \frac{1}{k}\right) - \Phi^2\left(\frac{z}{\sqrt{k}}\right) - \frac{1}{k}\varphi^2\left(\frac{z}{\sqrt{k}}\right)$$

$$+ s\sqrt{k}\left(\Phi\left(\frac{z}{\sqrt{k}}\right)\varphi\left(\frac{z}{\sqrt{k}}\right) - \int_{-\infty}^{\frac{z}{\sqrt{k}}}\varphi\left(x, \frac{z}{\sqrt{k}}; \frac{1}{k}\right) dx\right)$$

$$+ s^2\frac{k}{4}\left(\varphi\left(\frac{z}{\sqrt{k}}, \frac{z}{\sqrt{k}}; \frac{1}{k}\right) - \varphi^2\left(\frac{z}{\sqrt{k}}\right)\right)$$

$$+ s^2\frac{k}{4}\left(\frac{1}{2}z\varphi\left(\frac{z}{\sqrt{k}}\right)\Phi\left(\frac{z}{\sqrt{k}}\right) - \frac{1}{1-(1/k)^2}\int_{-\infty}^{\frac{z}{\sqrt{k}}}\left(\frac{z}{\sqrt{k}} - \frac{x}{k}\right)\varphi\left(x, \frac{z}{\sqrt{k}}; \frac{1}{k}\right) dx\right)$$

$$+ o(s^2).$$

and the asymptotic approximation of the curve in figure 1.5 for $s \to 0$, i.e. $f_{N-k}(z)$, is

$$f_{N-k}(z) = \Phi\left(\frac{z}{\sqrt{k}}; \frac{1}{k}\right) - \Phi^2\left(\frac{z}{\sqrt{k}}\right) - \frac{1}{k}\varphi^2\left(\frac{z}{\sqrt{k}}\right).$$

$\square$

In figure 1.6 we verify whether our asymptotic function $f_n(z)$ is a good approximation of the full function $\psi_n = g_n\left(y_n/(\sigma\sqrt{\delta}), \sigma\sqrt{\delta}\right)$; we see that the approximation is very close to the full analytic expression as we take $\delta \to 0$.

Now that we have the function $f_{N-k}(z)$, we can use it to compute the expectation $\mathbb{E}_0[\psi_{N-k}]$ near maturity; as was shown in theorem 1.2, the expectation can, for small
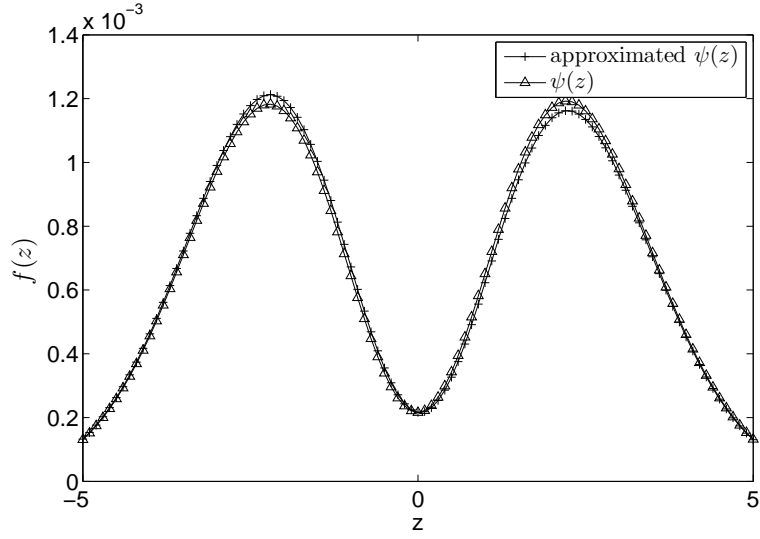
Figure 1.6: Comparison of the asymptotic approximation of ψ_n and full function. $\sigma = 0.3$, $\delta = 0.0001$, $n = 5$, $N = 10$.

values of k, be asymptotically approximated as

$$\mathbb{E}_0[\psi_{N-k}] = \int_{\mathbb{R}} f_{N-k}(z)\, dz \frac{1}{\sqrt{2\pi\sigma^2 T}} \exp\left(-\frac{1}{2}\frac{\left(\log\frac{S_0}{K} - \frac{1}{2}\sigma^2 T\right)^2}{\sigma^2 T}\right) \sigma\sqrt{\delta} + \mathcal{O}(\delta^{3/2}) + \mathcal{O}(\sqrt{\delta}\eta).$$

We see that the unconditional one-step hedging error, up to order $\mathcal{O}(\sqrt{\delta})$, is a product of the total volume under the curve from figure 1.5 given by the integral term $\int_{\mathbb{R}} f_{N-k}(z)\, dz$, the unconditional transitional density of the strike at time T as seen from time $t = 0$, and the standard deviation of returns over one time-step.

Therefore, to complete our asymptotic computation of $\mathbb{E}_0[\psi_n]$ at some time step near maturity on the interval $[T - \eta, T]$, we need to integrate over $f_{N-k}(z)$. In the next theorem, we compute the asymptotic hedging error when we evaluate that integral over $f_{N-k}(z)$.

Theorem 1.9. *Given a fixed $\eta > 0$, the expected value of the hedging error ψ_n of a digital call with maturity $T = N\delta$ at a time $t_n = n\delta$ s.t. $T - t_n \leq \eta$ is asymptotically*

given as

$$\mathbb{E}_0[\psi_{N-k}] = \left[\sqrt{k}\left(\frac{2}{2\sqrt{\pi}}\left(1 - \sqrt{1 - \frac{1}{k}}\right)\right) - \frac{1}{\sqrt{k}}\frac{1}{2\sqrt{\pi}}\right] \tag{1.21}$$

$$\times \frac{1}{\sqrt{2\pi\sigma^2 T}}\exp\left(-\frac{1}{2}\frac{\left(\log\frac{S_0}{K} - \frac{1}{2}\sigma^2 T\right)^2}{\sigma^2 T}\right)\sigma\sqrt{\delta} + \mathcal{O}(\delta^{3/2}) + \mathcal{O}(\sqrt{\delta}\eta)$$

$$= \left[\frac{1}{\sqrt{\pi}}(\sqrt{k} - \sqrt{k-1}) - \frac{1}{2\sqrt{\pi}}\frac{1}{\sqrt{k}}\right]$$

$$\times \frac{1}{\sqrt{2\pi\sigma^2 T}}\exp\left(-\frac{1}{2}\frac{\left(\log\frac{S_0}{K} - \frac{1}{2}\sigma^2 T\right)^2}{\sigma^2 T}\right)\sigma\sqrt{\delta} + \mathcal{O}(\delta^{3/2}) + \mathcal{O}(\sqrt{\delta}\eta).$$

Proof. First we prove we can use theorem 1.2 with a digital call option; we recall that the function $f_n(z)$ is given as

$$f_n(z) = f_{N-k}(z) = \Phi\left(\frac{z}{\sqrt{k}}; \frac{1}{k}\right) - \Phi^2\left(\frac{z}{\sqrt{k}}\right) - \frac{1}{k}\varphi^2\left(\frac{z}{\sqrt{k}}\right).$$

We know that $\varphi^2(x)$ is an even function, thus we have to show that a function of the form $\Phi(x; \rho) - \Phi^2(x)$ is even, i.e.

$$\Phi(x; \rho) - \Phi^2(x) = \Phi(-x; \rho) - \Phi^2(-x),$$

or rearranged,

$$\Phi(x; \rho) - \Phi(-x; \rho) = \Phi^2(x) - \Phi^2(-x) = \Phi^2(x) - (1 - \Phi(x))^2 \tag{1.22}$$

$$= 2\Phi(x) - 1.$$

But by Abramowitz and Stegun [1972, eqn. 26.3.9.], we know that relation (1.22) holds, which means that $f(z)$ is truly an even function. It is trivial to verify that the conditions on $g_n(z, s)$ hold, and so we can use theorem 1.2.

By that theorem, we now need to compute $\int_{\mathbb{R}} f_n(z)\, dz$, i.e. we need to compute the

integral

$$\int_{\mathbb{R}} \left(\Phi \left(\frac{z}{\sqrt{k}}; \frac{1}{k} \right) - \Phi^2 \left(\frac{z}{\sqrt{k}} \right) - \frac{1}{k} \varphi^2 \left(\frac{z}{\sqrt{k}} \right) \right) dz$$
$$= \sqrt{k} \int_{\mathbb{R}} \Phi \left(y; \frac{1}{k} \right) - \Phi(y; 0) \, dy - \frac{1}{\sqrt{k}} \int_{\mathbb{R}} \varphi^2(y) \, dy$$
$$= \sqrt{k} \left(\int_{-\infty}^{0} \Phi \left(y; \frac{1}{k} \right) - \Phi(y; 0) \, dy + \int_{0}^{\infty} \Phi \left(y; \frac{1}{k} \right) - \Phi(y; 0) \, dy \right)$$
$$\quad - \frac{1}{\sqrt{k}} \int_{\mathbb{R}} \varphi^2(y) \, dy.$$

Using lemma A.2 we know that

$$\int_{-\infty}^{0} \Phi(y; \rho) - \Phi(y; 0) \, dy = \sqrt{\frac{2}{\pi}} \left(\frac{1}{2} - \frac{\sqrt{2(1-\rho)}}{4} - \left(\frac{1}{2} - \frac{\sqrt{2}}{4} \right) \right)$$
$$= \sqrt{\frac{2}{\pi}} \left(\frac{\sqrt{2}}{4} \left(1 - \sqrt{(1-\rho)} \right) \right).$$

In our case, $\rho = \frac{1}{k}$ and the result simplifies to $\frac{1}{2\sqrt{\pi}} \left(1 - \sqrt{1 - \frac{1}{k}} \right)$. The integral over the positive half-line can be computed by using the fact that

$$\int_{0}^{\infty} \Phi(y; \rho) - \Phi(y; 0) \, dy = \int_{0}^{\infty} (1 - \Phi(y; 0)) - (1 - \Phi(y; \rho)) \, dy,$$

and observing that Lemma A.1 gives us its value in terms of an integral over the negative half-line, for which we can again use Lemma A.2. It turns out that the integral over the positive half-line has the same value as that over the negative half-line. The last integral $\int_{\mathbb{R}} \varphi^2(y) \, dy$ can easily be computed directly and equals to $1/(2\sqrt{\pi})$. Putting results together and using Theorem 1.2, we get the desired result. $\qquad \square$

1.3.2   Computing Cash Gamma squared for a digital call

In this section we will aim to compute the expectation of the Cash Gamma squared $\left(\frac{\partial^2 V}{\partial S^2} S^2 \right)^2 = (\Gamma S^2)^2$ of a digital call and compare it to the results in the previous section, since by the hypothesis in Černý [2009, eqn. 13.78], the Cash Gamma squared and the locally optimal hedging error should be closely related:

$$\psi_t = \left(\frac{1}{2} \Gamma_t S_t^2 \sigma^2 \delta \right)^2 (\text{Kurt}_t(R_{t+1}) - 1) + \mathcal{O}(\delta^{5/2}). \tag{1.23}$$

Here $\text{Kurt}_t(R_{t+1})$ is the kurtosis of returns on the stocks (in our case, $\text{Kurt}_t(R_{t+1}) = 3$) and $\sigma^2 \delta$ is the instantaneous variance of returns. Let us remind ourselves that the value of the digital call option at time t and with expiry date T is

$$V_t = \Phi\left(\frac{\log \frac{S_t}{K} - \frac{1}{2}\sigma^2(T-t)}{\sigma\sqrt{T-t}}\right),$$

or abbreviated,

$$V_t = \Phi(d); d = \frac{\log \frac{S_t}{K} - \frac{1}{2}\sigma^2(T-t)}{\sigma\sqrt{T-t}}. \tag{1.24}$$

The Delta of the digital option is

$$\Delta = \varphi(d)\frac{1}{S\sigma\sqrt{T-t}},$$

where $\varphi(x) = \partial\Phi(x)/\partial x$, and hence the Gamma is

$$\Gamma = \frac{-1}{S^2\sigma\sqrt{T-t}}\varphi(d)\left(1 + \frac{d}{\sigma\sqrt{T-t}}\right)$$

The Cash Gamma squared is then given by the formula

$$\left(\Gamma S^2\right)^2 = \frac{\varphi^2(d)}{\sigma^2(T-t)}\left(1 + \frac{d}{\sigma\sqrt{T-t}}\right)^2, \tag{1.25}$$

or in a more extended form:

$$\left(\Gamma S^2\right)^2 = \frac{\varphi^2(d)}{(\sigma^2(T-t))^3}\left(\left(\log \frac{S}{K}\right)^2 + \sigma^2(T-t)\log\frac{S}{K} + \frac{1}{4}\sigma^4(T-t)^2\right). \tag{1.26}$$

In figure 1.7 we can see that the estimate from (1.23) for ψ_n using the explicit formula (1.25) coincides nicely with the one-step hedging error as computed in (1.15).

1.3.2.1 Towards granularity - computing expectations of the Cash Gamma squared

As we saw in equation (1.4), to compute the granularity (as defined by Bertsimas et al. [2000]) of the digital call option, we first have to compute

$$\mathbb{E}_0\left[\left(\Gamma_t S_t^2\right)^2\right].$$
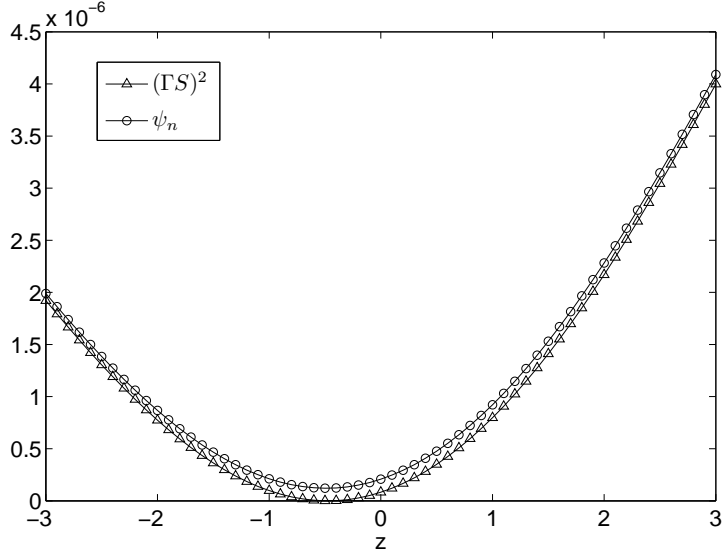
Figure 1.7: Comparison of the approximation of $\psi_n(z)$ via Cash Gamma squared formula (1.23) and the full function ψ_n. $\sigma = 0.3$, $\delta = 1/360$, $n = 300$, $N = 360$. Independent variable $z = \frac{y_n}{s}$.

In the next theorem, we provide such a result.

Theorem 1.10. *Under the stock price model (1.7), it holds that for $t < T$:*

$$
\mathbb{E}_0[(\Gamma S^2)^2] = \frac{1}{(\sigma^2(T-t))^3}\left[\mathbb{E}_0[\varphi^2(d)]\left(\frac{1}{2}\sigma^2(T-t) + \left(y_0 - \frac{\sigma^2 t}{2}\right)\right)^2 \right. \tag{1.27}
$$
$$
+ \mathbb{E}_0[\varphi^2(d)Z_n]\left(2\left(y_0 - \frac{\sigma^2 t}{2}\right)\sigma\sqrt{t} + \sigma^2(T-t)\sigma\sqrt{t}\right)
$$
$$
\left. + \mathbb{E}_0[\varphi^2(d)Z_n^2]\sigma^2 t \right].
$$

where

$$
\mathbb{E}_0[\varphi^2(d)] = \frac{1}{2\pi}\sqrt{\frac{T-t}{T+t}}\exp\left(-\frac{(y_0 - \frac{1}{2}\sigma^2 T)^2}{\sigma^2(T+t)}\right)
$$
$$
\mathbb{E}_0[Z_n\varphi^2(d)] = -\frac{\sqrt{t}}{\pi}\sqrt{\frac{T-t}{T+t}}\frac{y_0 - \frac{1}{2}\sigma^2 T}{\sigma(T+t)}\exp\left(-\frac{(y_0 - \frac{1}{2}\sigma^2 T)^2}{\sigma^2(T+t)}\right)
$$
$$
\mathbb{E}_0[Z_n^2\varphi^2(d)] = \left(\frac{T-t}{T+t}\right)^{3/2}\frac{1}{2\pi}\left(1 + \frac{4t(y_0 - \frac{1}{2}\sigma^2 T)^2}{\sigma^2(T+t)(T-t)}\right)\exp\left(-\frac{(y_0 - \frac{1}{2}\sigma^2 T)^2}{\sigma^2(T+t)}\right).
$$

Furthermore, it holds that $\mathbb{E}_0[(\Gamma_t S_t^2)^2] = \mathcal{O}((T-t)^{-3/2})$.

Proof. To compute the expectations of Cash Gamma squared, we first recall that under our model (1.7) we have for $y_n = \log(S_n/K)$ that:

$$y_n = \log \frac{S_n}{K} = y_0 - \frac{1}{2}\sigma^2 t + \sigma\sqrt{t}Z_n; Z_n \sim N(0,1).$$

If we plug the above into equation (1.26) for the Cash Gamma squared and take expectations, we get (1.27). We can see that we will have to compute three expectations: $\mathbb{E}_0[\varphi^2(d)]$, $\mathbb{E}_0[\varphi^2(d)Z_n]$ and $\mathbb{E}_0[\varphi^2(d)Z_n^2]$, where d is defined in (1.24) (the remainder of the terms are deterministic and can be taken out of the expectations). To compute the expectations, we will first note that in general

$$\varphi^2(x) = \frac{1}{\sqrt{2\pi}}\varphi(\sqrt{2}x).$$

We see that the expectations we are computing are a particular example of the general formulas $\mathbb{E}[\varphi(aZ + b)], \mathbb{E}[Z\varphi(aZ + b)], \mathbb{E}[Z^2\varphi(aZ + b)]; Z \sim N(0,1)$, with specific parameters a, b. Therefore, let us evaluate the general formulas and then plug in our parameters a, b. We evaluate the first expression:

$$
\begin{aligned}
\mathbb{E}[\varphi(aZ + b)] &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(ax + b)^2\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx \\
&= \int_{\mathbb{R}} \frac{1}{2\pi} \exp\left(-\frac{1}{2}\left((1 + a^2)x^2 + 2abx + b^2\right)\right) dx = |\sqrt{1 + a^2}x = z| \\
&= \frac{1}{2\pi}\frac{1}{\sqrt{1 + a^2}} \int_{\mathbb{R}} \exp\left(-\frac{1}{2}\left(z^2 + \frac{2ab}{\sqrt{1 + a^2}}z + b^2\right)\right) dz \\
&= \frac{1}{2\pi}\frac{1}{\sqrt{1 + a^2}} \exp\left(\frac{a^2b^2}{2(1 + a^2)} - \frac{b^2}{2}\right) \int_{\mathbb{R}} \exp\left(-\frac{1}{2}\left(z + \frac{ab}{\sqrt{1 + a^2}}\right)^2\right) dz \\
&= \frac{1}{\sqrt{2\pi(1 + a^2)}} \exp\left(-\frac{b^2}{2(1 + a^2)}\right).
\end{aligned}
$$

The remaining two expressions are evaluated in a similar fashion:

$$
\begin{aligned}
\mathbb{E}[Z\varphi(aZ+b)] &= \int_{\mathbb{R}} \frac{1}{2\pi} x \exp\left(-\frac{1}{2}\left((1+a^2)x^2 + 2abx + b^2\right)\right) dx = |\sqrt{1+a^2}\,x = z| \\
&= \frac{1}{2\pi}\frac{1}{1+a^2} \int_{\mathbb{R}} z \exp\left(-\frac{1}{2}\left(z^2 + \frac{2ab}{\sqrt{1+a^2}}z + b^2\right)\right) dz \\
&= \frac{1}{2\pi(1+a^2)} \exp\left(-\frac{b^2}{2(1+a^2)}\right) \int_{\mathbb{R}} z \exp\left(-\frac{1}{2}\left(z + \frac{ab}{\sqrt{1+a^2}}\right)^2\right) dz \\
&= \frac{1}{\sqrt{2\pi}(1+a^2)} \exp\left(-\frac{b^2}{2(1+a^2)}\right) \frac{-ab}{\sqrt{1+a^2}}
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}[Z^2\varphi(aZ+b)] &= \int_{\mathbb{R}} \frac{1}{2\pi} x^2 \exp\left(-\frac{1}{2}\left((1+a^2)x^2 + 2abx + b^2\right)\right) dx = |\sqrt{1+a^2}\,x = z| \\
&= \frac{1}{2\pi(1+a^2)^{3/2}} \exp\left(-\frac{b^2}{2(1+a^2)}\right) \underbrace{\int_{\mathbb{R}} z^2 \exp\left(-\frac{1}{2}\left(z + \frac{ab}{\sqrt{1+a^2}}\right)^2\right) dz}_{\mathbb{E}[Z^2]=Var[Z]+\mathbb{E}^2[Z]} \\
&= \frac{1}{\sqrt{2\pi}(1+a^2)^{3/2}} \left(1 + \frac{a^2 b^2}{1+a^2}\right) \exp\left(-\frac{b^2}{2(1+a^2)}\right)
\end{aligned}
$$

We can now substitute the correct parameters to get the expressions we need. To find them, we observe that

$$
\begin{aligned}
\varphi(\sqrt{2}d) &= \varphi\left(\frac{\sqrt{2}}{\sigma\sqrt{T-t}}\left(y_0 - \frac{\sigma^2 t}{2} + \sigma\sqrt{t}Z - \frac{\sigma^2(T-t)}{2}\right)\right) \\
&= \varphi\left(\frac{\sqrt{2}}{\sigma\sqrt{T-t}}\left(\sigma\sqrt{t}Z + y_0 - \frac{\sigma^2 T}{2}\right)\right).
\end{aligned}
$$

Hence the right parameters are

$$
a = \frac{\sqrt{2t}}{\sqrt{T-t}}, \, b = \frac{\sqrt{2}}{\sigma\sqrt{T-t}}\left(y_0 - \frac{1}{2}\sigma^2 T\right).
$$

When we plug these parameters into our general formulae, we obtain the following results:

$$\mathbb{E}_0[\varphi^2(d)] = \frac{1}{2\pi}\sqrt{\frac{T-t}{T+t}}\exp\left(-\frac{(y_0 - \frac{1}{2}\sigma^2 T)^2}{\sigma^2(T+t)}\right)$$

$$\mathbb{E}_0[Z\varphi^2(d)] = -\frac{\sqrt{t}}{\pi}\sqrt{\frac{T-t}{T+t}}\frac{y_0 - \frac{1}{2}\sigma^2 T}{\sigma(T+t)}\exp\left(-\frac{(y_0 - \frac{1}{2}\sigma^2 T)^2}{\sigma^2(T+t)}\right)$$

$$\mathbb{E}_0[Z^2\varphi^2(d)] = \left(\frac{T-t}{T+t}\right)^{3/2}\frac{1}{2\pi}\left(1 + \frac{4t(y_0 - \frac{1}{2}\sigma^2 T)^2}{\sigma^2(T+t)(T-t)}\right)\exp\left(-\frac{(y_0 - \frac{1}{2}\sigma^2 T)^2}{\sigma^2(T+t)}\right).$$

We have thus obtained all we need to fully evaluate $\mathbb{E}_0[(\Gamma S^2)^2]$. The last thing we need to do is to show that $\mathbb{E}_0[(\Gamma S^2)^2] = \mathcal{O}((T-t)^{-3/2})$, since this is not evident from the way we have noted the solution. Specifically, we need to verify that the terms

$$\frac{1}{(\sigma^2(T-t))^3}\left[\mathbb{E}_0[\varphi^2(d)]\left(y_0 - \frac{\sigma^2 t}{2}\right)^2 + \mathbb{E}_0[\varphi^2(d)Z_n]2\left(y_0 - \frac{\sigma^2 t}{2}\right)\sigma\sqrt{t} + \mathbb{E}_0[\varphi^2(d)Z_n^2]\sigma^2 t\right]$$

are of the desired order. Substituting

$$y_0 - \frac{1}{2}\sigma^2 t = y_0 - \frac{1}{2}\sigma^2 T + \frac{1}{2}\sigma^2(T-t)$$

we can simplify the expression to get the desired result. $\qquad\square$

We numerically verify that the asymptotic formula (1.23) holds in expectations, i.e. that

$$\mathbb{E}_0[\psi_n] = \mathbb{E}_0\left[\left(\frac{1}{2}\Gamma_n S_n^2 \sigma^2 \delta\right)^2\right](\mathrm{Kurt}_n(R_{n+1}) - 1) + \mathcal{O}(\delta^{5/2})$$

when we plug in our result for $\mathbb{E}_0[(\Gamma_t S_t^2)^2]$; we truly obtain the same values as we get from the full formula (1.17). In figure 1.8 we see that for a time sufficiently far away from maturity T, this relation holds. However, we can also see that this relation breaks down and the estimate (1.23) is no longer accurate in expectations as we come close to maturity, as illustrated in figure 1.9.
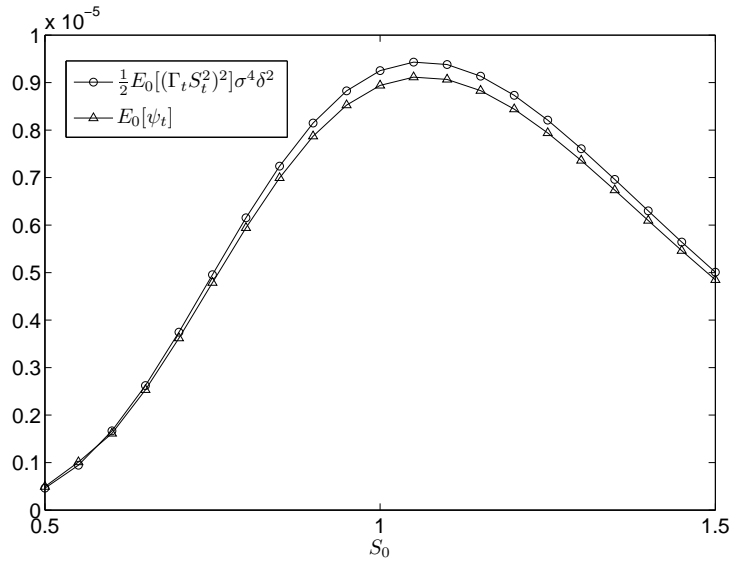
Figure 1.8: Comparison of the results of formula (1.27) and formula (1.17). $\sigma = 0.3$, $K = 1$, $\delta = 1/360$, $n = 330$, $N = 360$.
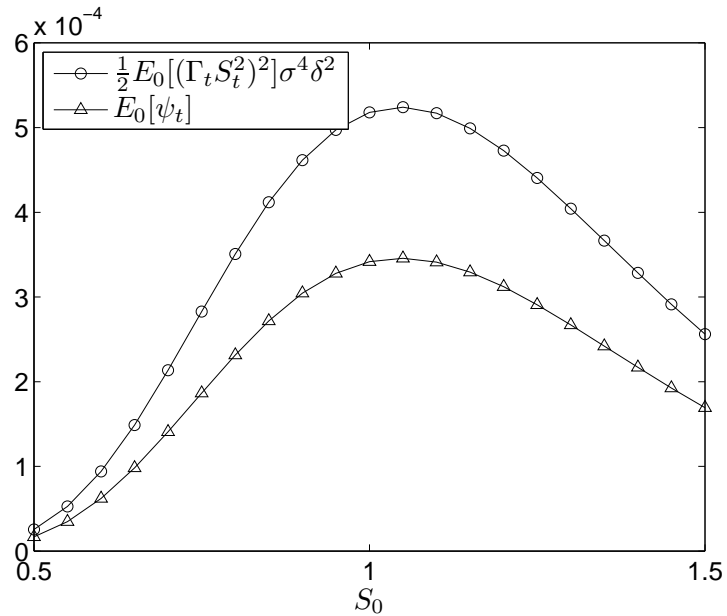
Figure 1.9: Comparison of the results of formula (1.27) and formula (1.17). $\sigma = 0.3$, $K = 1$, $\delta = 1/360$, $n = 358$, $N = 360$.

1.3.3 Summing up the errors - a more general formula for granularity

Now that we have computed the one-step error and seen that it corresponds well to the Cash Gamma squared (both without and with expectations) when far from maturity, we can proceed to sum up all those errors to obtain the total squared hedging error ε_0^2 as given by (1.16).

We will first show how $\varepsilon_0^2 = \varepsilon_0^2(\delta)$ is connected to granularity (1.4). By taking expectations in (1.23) and summing over all the errors, we see that

$$\sum_{n=0}^{N-1} \mathbb{E}_0[\psi_n] = \sum_{n=0}^{N-1} \frac{1}{2}\sigma^4\delta^2 \mathbb{E}_0[(\Gamma_n S_n^2)^2] + \mathcal{O}(\delta^{5/2})$$

$$= \frac{1}{2}\sigma^4\delta \left(\sum_{n=0}^{N-1} \mathbb{E}_0[(\Gamma_n S_n^2)^2]\delta \right) + \mathcal{O}(\delta^{5/2})$$

We can see that as we increase the number of rebalancing dates N (and thus decrease the length of the interval δ), the term in parentheses above would seem to tend to the standard granularity integral

$$\sum_{n=0}^{N-1} \mathbb{E}_0[(\Gamma_n S_n^2)^2]\delta \xrightarrow{\delta \to 0} \int_0^T \mathbb{E}_0[(\Gamma_t S_t^2)^2]\, dt.$$

This, however, only holds for derivatives with continuous payoffs. As we saw above, illustrated in figure 1.9, the relation between the Cash Gamma squared and the one-step local hedging error breaks down when nearing maturity. For a derivative with a discontinuous payoff, the Cash Gamma will no longer be integrable over the interval $[0, T]$ as it contains a singularity at terminal time T and hence the sum will not converge to a Riemann integral. Therefore, the convergence to an integral above will only hold for some time interval $[0, T - \eta]$, where $\eta > 0$ is a fixed parameter. Thus, if we denote $T - \eta = N_\eta\delta$, then the following will hold:

$$\sum_{n=0}^{N_\eta-1} \mathbb{E}_0[(\Gamma_n S_n^2)^2]\delta \xrightarrow{\delta \to 0} \int_0^{T-\eta} \mathbb{E}_0[(\Gamma_t S_t^2)^2]\, dt$$

The question then remains about the behaviour in the limit $\delta \to 0$ of the remainder of the sum, i.e. $\sum_{n=N_\eta}^{N} \mathbb{E}_0[\psi_n]$. The following theorem will, by treating the last terms differently, provide the full asymptotics up to order $\mathcal{O}(\delta)$ for the total squared tracking

error $\varepsilon^2(\delta)$ of a digital call option. We will use $\varphi(\cdot, \mu, \nu)$ to denote the probability density function of a normal random variable with mean μ and variance ν.

Theorem 1.11. *The total squared mean-variance hedging error for a digital call option under model (1.7) is asymptotically given as*

$$\varepsilon_0^2(\delta) = \left(\frac{1}{\sqrt{\pi}} \tilde{\varphi} \sigma \tilde{\zeta} \right) \sqrt{\delta} + \left(\int_0^T \frac{1}{2} \sigma^4 \mathbb{E}_0[(\Gamma_t S_t^2)^2] - \frac{1}{8\sqrt{\pi}} \tilde{\varphi} \sigma \frac{1}{(T-t)^{3/2}} \, dt + \frac{\tilde{\varphi}\sigma}{4\sqrt{\pi T}} \right) \delta + \mathcal{O}(\delta^{3/2})$$

$$(1.28)$$

where

$$\tilde{\zeta} = \sum_{k=1}^{\infty} \left(\sqrt{k} - \sqrt{k-1} - \frac{1}{2\sqrt{k}} \right)$$

and

$$\tilde{\varphi} = \varphi \left(y_0; \mu = \frac{1}{2} \sigma^2 T; \nu = \sigma^2 T \right)$$

Proof. We wish to get an asymptotic expression for

$$\varepsilon_0^2(\delta) = \sum_{n=0}^{N-1} \mathbb{E}_0[\psi_n].$$

We can divide the sum into two parts. For that we take a fixed parameter $\eta > 0$ s.t. $T - \eta = N_\eta \delta$ and consider the sum

$$\varepsilon_0^2(\delta) = \varepsilon_0^2(\delta; \eta) = \sum_{n=0}^{N_\eta-1} \mathbb{E}_0[\psi_n] + \sum_{n=N_\eta}^{N-1} \mathbb{E}_0[\psi_n]$$

Note that this division by parameter η does not influence the total sum, i.e. $\varepsilon_0^2(\delta; \eta_1) = \varepsilon_0^2(\delta; \eta_2)$. For the first part of the sum, we can use the asymptotic estimate (1.23) of $\mathbb{E}_0[\psi_n]$ to get an asymptotic estimate of the total hedging error:

$$\sum_{n=0}^{N_\eta-1} \mathbb{E}_0[\psi_n] = \frac{1}{2} \sigma^4 \delta \left(\sum_{n=0}^{N_\eta-1} \mathbb{E}_0[(\Gamma_n S_n^2)^2] \delta \right) + \mathcal{O}(\delta^{5/2}).$$

Since $\mathbb{E}_0[(\Gamma_n S_n^2)^2]$ is an integrable function for timesteps $t < T - \eta$ with finite first derivative over $[0, T - \eta]$, we know that the term in parentheses - a Riemann sum - approximates a Riemann integral for $\delta \to 0$, with a finite approximation error of order $\mathcal{O}(\delta)$:

$$\sum_{n=0}^{N_\eta-1} \mathbb{E}_0[(\Gamma_n S_n^2)^2] \delta = \int_0^{T-\eta} \mathbb{E}_0[(\Gamma_t S_t^2)^2] \, dt + \mathcal{O}(\delta).$$

If we are interested in the approximation of the total error up to order $\mathcal{O}(\delta)$, the integral approximation is negligible, because we get a term of order $\mathcal{O}(\delta)\delta = \mathcal{O}(\delta^2) < \mathcal{O}(\delta^{3/2})$.

For the second part of the summation, we use the fact that for times $\eta \leq t \leq T$, the individual errors are asymptotically given by (1.21), so computing their sum leads to the problem of computing the sum

$$\sum_{n=N_\eta}^{N-1} \mathbb{E}_0[\psi_n] = \sum_{k=1}^{N-N_\eta} \mathbb{E}_0[\psi_{N-k}]$$

$$= \sum_{k=1}^{N_\eta} \frac{1}{\sqrt{\pi}} \left(\sqrt{k} - \sqrt{k-1} - \frac{1}{2\sqrt{k}} \right) \varphi \left(y_0; \mu = \frac{1}{2}\sigma^2 T; \nu = \sigma^2 T \right) \sigma\sqrt{\delta} + \mathcal{O}(\delta^{3/2})$$

where $y_0 = \log(S_0/K)$ and $\varphi(\cdot, \mu, \nu)$ is the probability density function of a normal random variable with mean μ and variance ν. Using abbreviated notation $\tilde{\varphi} := \varphi\left(y_0; \mu = \frac{1}{2}\sigma^2 T; \nu = \sigma^2 T\right)$, we end up requiring to compute a sum of the form

$$\frac{1}{\sqrt{\pi}}\tilde{\varphi}\sigma\sqrt{\delta} \sum_{k=1}^{N-N_\eta} \left(\sqrt{k} - \sqrt{k-1} - \frac{1}{2\sqrt{k}} \right).$$

To compute the sum above, we will consider its infinite sum counterpart

$$\tilde{\zeta} = \sum_{k=1}^{\infty} \left(\sqrt{k} - \sqrt{k-1} - \frac{1}{2\sqrt{k}} \right). \tag{1.29}$$

Numerically, we find that the partial sums of the infinite series converge and hence the sum is finite, its value being $\tilde{\zeta} \approx 0.7302$. Then we find the asymptotics of our finite sum by estimating the "tail" of the infinite sum:

$$\sum_{k=1}^{N-N_\eta} \left(\sqrt{k} - \sqrt{k-1} - \frac{1}{2\sqrt{k}} \right) = \tilde{\zeta} - \sum_{k=N-N_\eta+1}^{\infty} \left(\sqrt{k} - \sqrt{k-1} - \frac{1}{2\sqrt{k}} \right)$$

$$= \tilde{\zeta} - \frac{1}{2} \sum_{k=N-N_\eta+1}^{\infty} \left(2(\sqrt{k} - \sqrt{k-1}) - \frac{1}{\sqrt{k}} \right) \tag{1.30}$$

It is easy to see that

$$\sum_{k=N-N_\eta+1}^{\infty} 2(\sqrt{k} - \sqrt{k-1}) = \sum_{k=N-N_\eta+1}^{\infty} \int_{k-1}^{k} \frac{dx}{\sqrt{x}} = \int_{N-N_\eta}^{\infty} \frac{dx}{\sqrt{x}}$$

and therefore

$$\sum_{k=N-N_\eta+1}^{\infty} \left(2(\sqrt{k} - \sqrt{k-1}) - \frac{1}{\sqrt{k}} \right) = \int_{N-N_\eta}^{\infty} \frac{dx}{\sqrt{x}} - \sum_{k=N-N_\eta+1}^{\infty} \frac{1}{\sqrt{k}}$$

$$= \int_{N-N_\eta}^{\infty} \frac{dx}{\sqrt{x}} - \sum_{k=N-N_\eta}^{\infty} \frac{1}{\sqrt{k}} + \frac{1}{\sqrt{N-N_\eta}}.$$

We can see that the sum is some type of discretization of the integral. At this point, we recall a fitting version of the Euler-MacLaurin summation formula (see Kac and Cheung [2002, eqn. 25.9]):

$$\sum_{n=a}^{\infty} f(n) = \int_a^{\infty} f(t)dt + \frac{1}{2}f(a) + \mathcal{O}(f'(a))$$

In our case, $f(n) = \frac{1}{\sqrt{n}}, a = N - N_\eta$ and $\mathcal{O}(f'(a)) = \mathcal{O}((N - N_\eta)^{-3/2})$ and therefore we see that

$$\sum_{k=N-N_\eta}^{\infty} \frac{1}{\sqrt{k}} - \int_{N-N_\eta}^{\infty} \frac{dx}{\sqrt{x}} = \frac{1}{2} \frac{1}{\sqrt{N-N_\eta}} + \mathcal{O}((N-N_\eta)^{-3/2})$$

Finally, we notice that $N - N_\eta = \eta/\delta$ to gain the asymptotic behaviour of the partial sum (1.30):

$$\sum_{k=1}^{N-N_\eta} \left(\sqrt{k} - \sqrt{k-1} - \frac{1}{2\sqrt{k}} \right) = \tilde{\zeta} - \frac{1}{2} \left(-\frac{1}{2} \frac{1}{\sqrt{N-N_\eta}} + \frac{1}{\sqrt{N-N_\eta}} + \mathcal{O}((N-N_\eta)^{-3/2}) \right)$$

$$= \tilde{\zeta} - \frac{1}{4}\sqrt{\frac{\delta}{\eta}} + \mathcal{O}\left(\frac{\delta^{3/2}}{\eta^{3/2}} \right)$$

and therefore

$$\sum_{n=N_\eta}^{N-1} \mathbb{E}_0[\psi_n] = \frac{1}{\sqrt{\pi}} \tilde{\varphi} \tilde{\zeta} \sigma \sqrt{\delta} - \frac{1}{4\sqrt{\pi}} \tilde{\varphi} \sigma \frac{\delta}{\sqrt{\eta}} + \mathcal{O}\left(\frac{\delta^{5/2}}{\eta^{3/2}} \right)$$

We now have all we need to complete the asymptotics and the corrected granularity

formula of a digital option:

$$\varepsilon_0^2(\delta; \eta) = \left(\frac{1}{\sqrt{\pi}} \tilde{\varphi} \sigma \tilde{\zeta}\right) \sqrt{\delta} + \left(\frac{1}{2}\sigma^4 \int_0^{T-\eta} \mathbb{E}_0[(\Gamma_t S_t^2)^2]\, dt - \frac{1}{4\sqrt{\pi}} \tilde{\varphi} \sigma \frac{1}{\sqrt{\eta}}\right) \delta + \mathcal{O}\left(\frac{\delta^{5/2}}{\eta^{3/2}}\right)$$

$$= \left(\frac{1}{\sqrt{\pi}} \tilde{\varphi} \sigma \tilde{\zeta}\right) \sqrt{\delta}$$

$$+ \left[\int_0^{T-\eta} \left(\frac{1}{2}\sigma^4 \mathbb{E}_0[(\Gamma_t S_t^2)^2] - \frac{1}{8\sqrt{\pi}} \tilde{\varphi}\sigma \frac{1}{(T-t)^{3/2}}\right) dt + \frac{\tilde{\varphi}\sigma}{4\sqrt{\pi}T}\right]\delta + \mathcal{O}\left(\frac{\delta^{5/2}}{\eta^{3/2}}\right)$$

The function that is under the integral is now integrable on $[0, T]$. However, we cannot take the limit $\eta \to 0$ directly due to the term of order $\mathcal{O}\left(\frac{\delta^{5/2}}{\eta^{3/2}}\right)$. Instead we will show that we can make the difference between the sum $\varepsilon_0^2(\delta) = \sum_{n=0}^{N-1} \mathbb{E}_0[\psi_n]$ and the analytic formula given in the theorem as small as we wish. We can write

$$\left|\frac{\varepsilon_0^2(\delta) - \left(\frac{1}{\sqrt{\pi}} \tilde{\varphi}\sigma\tilde{\zeta}\right)\sqrt{\delta} - \int_0^T \frac{\sigma^4}{2} \mathbb{E}_0[(\Gamma_t S_t^2)^2] - \frac{\tilde{\varphi}\sigma}{8\sqrt{\pi}} \frac{1}{(T-t)^{3/2}}\, dt}{\delta}\right| \leq$$

$$\leq \left|\frac{\sum_{n=N_\eta}^{N-1} \mathbb{E}_0[\psi_n] - \left(\frac{1}{\sqrt{\pi}} \tilde{\varphi}\sigma\tilde{\zeta}\right)\sqrt{\delta} + \delta \int_0^{T-\eta} \frac{\tilde{\varphi}\sigma}{8\sqrt{\pi}} \frac{1}{(T-t)^{3/2}}\, dt - \frac{\tilde{\varphi}\sigma}{4\sqrt{\pi}T}}{\delta}\right|$$

$$+ \left|\frac{\sum_{n=0}^{N_\eta-1} \mathbb{E}_0[\psi_n] - \frac{\sigma^4}{2}\delta \int_0^{T-\eta} \mathbb{E}_0[(\Gamma_t S_t^2)^2]\, dt}{\delta}\right|$$

$$+ \left|\frac{\delta \int_{T-\eta}^T \frac{\sigma^4}{2} \mathbb{E}_0[(\Gamma_t S_t^2)^2] - \frac{\tilde{\varphi}\sigma}{8\sqrt{\pi}} \frac{1}{(T-t)^{3/2}}\, dt}{\delta}\right|$$

We now take the limit $\delta \to 0$ on both sides of the inequality. We know that numerator of the first of the 3 terms is exactly 0, since we previously showed that the sum and the analytic expression are exactly equal. We know that numerator of the second term is equal to 0 up to an approximation error of order $\mathcal{O}(\delta^2)$ and hence the term will decay to 0 in the limit $\delta \to 0$. The final term will be a function dependent on η:

$$h(\eta) = \int_{T-\eta}^T \frac{\sigma^4}{2} \mathbb{E}_0[(\Gamma_t S_t^2)^2] - \frac{\tilde{\varphi}\sigma}{8\sqrt{\pi}} \frac{1}{(T-t)^{3/2}}\, dt$$

We know that the function within the integral is integrable over $[0, T]$ and hence $h(\eta)$ is well-defined. Now we can use a $(\Delta - \epsilon)$ approach to computing the limit. We can always find an η which will satisfy the following:

$$\forall \epsilon > 0 \ \exists \Delta > 0 : (\eta < \Delta \Rightarrow h(\eta) < \epsilon)$$
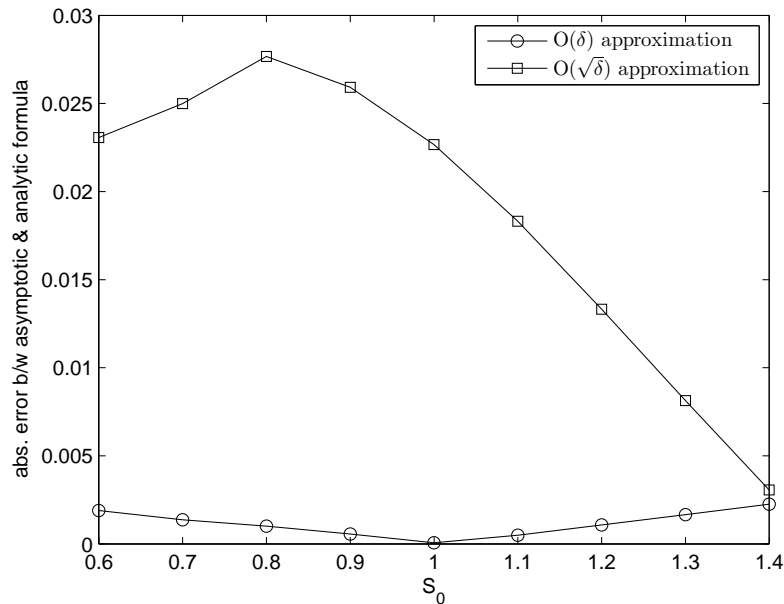
Figure 1.10: Absolute error between analytic formula for ε_0^2 and asymptotic expansions of order $\mathcal{O}(\sqrt{\delta})$ and $\mathcal{O}(\delta)$. $\delta = 1/180$, $K = 1$, $\sigma = 0.3$, $T = 2$.

Therefore, the overall difference between the sum and the analytic formula will be at most ϵ and we can make this ϵ as small as we want. $\square$

We thus see that for a digital option, the leading asymptotic term will always be of order $\mathcal{O}(\sqrt{\delta})$. Moreover, we have found the right correction term that compensates for the explosion in the classical granularity term, thus leading to a new, discontinuous version of the granularity formula. From the proof we see, however, that it is only the very few last terms that deform the order of error decay, meaning that risk management and mark to model rules for handling digital options should be time-varying and adjust for how close to maturity the digital option is, maintaining the importance of the Cash Gamma until very close to maturity. In figure 1.10, we can see that the new extended formula provides significantly greater accuracy in estimating the overall error than just the estimate of order $\mathcal{O}(\sqrt{\delta})$.

1.4 Connecting results to the Black-Scholes tracking error

In this section we will connect the results we have obtained for the discrete-time hedging error of a variance-optimal hedging strategy to the Black-Scholes tracking error (i.e. following a continuous-time strategy on a discrete set of times) for a digital option as discussed in Gobet and Temam [2001b]. We will show that in the case of a martingale underlying, these two different types of hedging error have identical asymptotic behaviour.

In Gobet and Temam [2001b], the authors find that the term of the digital option tracking error that causes the asymptotic behaviour of order $\mathcal{O}(\sqrt{\delta})$ is, when applied to the Black-Scholes model (with or without drift), of the form

$$\mathbb{E}_0 \int_0^T ds \int_{\varphi(s)}^s dt \left(\Gamma_t S_t^2\right)^2 \sigma^4,$$

where $\varphi(s) = \inf\{t_i | s > t_i\}$. We now relate this to our own computations. To do so, we will first split the integral into two parts:

$$\mathbb{E}_0 \int_0^T dt \int_{\varphi(s)}^s dt \left(\Gamma_t S_t^2\right)^2 \sigma^4 = \mathbb{E}_0 \int_0^{T-\eta} ds \int_{\varphi(s)}^s dt \left(\Gamma_t S_t^2\right)^2 \sigma^4 + \mathbb{E}_0 \int_{T-\eta}^T ds \int_{\varphi(s)}^s dt \left(\Gamma_t S_t^2\right)^2 \sigma^4$$

We now rewrite the first double integral above (abbreviating notation, $f(t) := \mathbb{E}_0[\left(\Gamma_t S_t^2\right)^2 \sigma^4]$) as a single integral with a periodic convolution kernel $p_n(t) = k(t/\delta - \lfloor t/\delta \rfloor \delta)$ in the integrated function:

$$\int_0^T ds \int_{\varphi(s)}^s dt f(t) = \sum_{k=0}^{n-1} \int_{t_k}^{t_{k+1}} ds \int_{t_k}^s dt f(t) = \sum_{k=0}^{n-1} \int_{t_k}^{t_{k+1}} f(t) dt \int_t^{t_{k+1}} ds$$

$$= \sum_{k=0}^{n-1} \int_{t_k}^{t_{k+1}} f(t)(t_{k+1} - t) dt = \sum_{k=0}^{n-1} \frac{T}{n} \int_{t_k}^{t_{k+1}} f(t) \frac{n}{T}(t_{k+1} - t) dt$$

$$= \frac{T}{n} \int_0^T f(t) p_n(t) dt = \delta \int_0^T f(t) p_n(t) dt$$

In our case, it can be shown that the convolution kernel is of the form $k(x) = 1 - x$. We now provide a theorem that shows that as long as the function f is integrable, the periodic convolution kernel can be taken outside the integral.

Theorem 1.12. *Given a $[0, T]$-integrable real function g and a periodic convolution kernel $p_n(t) = k(t/\delta - \lfloor t/\delta \rfloor \delta)$, where $k(\cdot)$ is an integrable function defined on $[0, 1]$ and $\delta = T/n$, it holds that*

$$\lim_{n \to \infty} \int_0^T g(t) p_n(t) \, dt = \int_0^1 k(u) \, du \int_0^T g(t) \, dt.$$

Proof. First we will prove that the above holds for a uniformly continuous function f.

In the proof we will use the definition of uniform continuity:

Definition 1.13. Given metric spaces (X, d_1) and (Y, d_2), a function $f : X \to Y$ is called uniformly continuous if for every real number $\varepsilon > 0$ there exists $\delta > 0$ such that for every $x, y \in X$ with $d_1(x, y) < \delta$, we have that $d_2(f(x), f(y)) < \varepsilon$. If X and Y are subsets of the real numbers, d_1 and d_2 can be the standard Euclidean norm, $|| \cdot ||$, yielding the definition: for all $\varepsilon > 0$ there exists a $\delta > 0$ such that for all $x, y \in X, |x - y| < \delta$ implies $|f(x) - f(y)| < \varepsilon$.

Thus, we can find an n such that on every interval $[t_k, t_{k+1}]$ of length $\frac{T}{n}$ we know that $|f(t) - f(t_k)| < \varepsilon$. Therefore, if we replace $f(t)$ with $f(t_k)$, we make an error of at most $\pm \varepsilon$.

To compute the integral, we first separate it into several subintervals:

$$\int_0^T f(t) p_n(t) \, dt = \int_0^T f(t) k \left(\frac{t}{\delta} - \lfloor \frac{t}{\delta} \rfloor \delta \right) \, dt$$

$$= \sum_{k=0}^{n-1} \int_{t_k}^{t_{k+1}} f(t) k \left(\frac{t}{\delta} - \lfloor \frac{t}{\delta} \rfloor \delta \right) \, dt$$

We know due to uniform continuity that we can choose n large enough for it to hold that

$$\int_{t_k}^{t_{k+1}} [f(t_k) - \varepsilon] k \left(\frac{t}{\delta} - \lfloor \frac{t}{\delta} \rfloor \delta \right) \, dt < \int_{t_k}^{t_{k+1}} f(t) k \left(\frac{t}{\delta} - \lfloor \frac{t}{\delta} \rfloor \delta \right) \, dt$$

$$< \int_{t_k}^{t_{k+1}} [f(t_k) + \varepsilon] k \left(\frac{t}{\delta} - \lfloor \frac{t}{\delta} \rfloor \delta \right) \, dt,$$

which will also hold for the sums. We will now investigate the upper and lower bound sums and find that they both converge to the same integral for $n \to \infty$, hence the

middle term must converge to it as well. We will treat the upper bound, the lower bound can be handled analogously.

$$\sum_{k=0}^{n-1} \int_{t_k}^{t_{k+1}} [f(t_k) + \varepsilon] k \left(\frac{t}{\delta} - \lfloor \frac{t}{\delta} \rfloor \delta \right) dt = \underbrace{\sum_{k=0}^{n-1} f(t_k) \int_{t_k}^{t_{k+1}} k \left(\frac{t}{\delta} - \lfloor \frac{t}{\delta} \rfloor \delta \right) dt}_{\text{I}}$$

$$+ \underbrace{\varepsilon \sum_{k=0}^{n-1} \int_{t_k}^{t_{k+1}} k \left(\frac{t}{\delta} - \lfloor \frac{t}{\delta} \rfloor \delta \right) dt}_{\text{II}}$$

We can transform the integral of $k(\cdot)$ over $[t_k, t_{k+1}]$:

$$\int_{t_k}^{t_{k+1}} k \left(\frac{t}{\delta} - \lfloor \frac{t}{\delta} \rfloor \delta \right) dt = |u = \frac{t}{\delta} - \lfloor \frac{t}{\delta} \rfloor \delta, \delta du = dt| = \int_0^1 k(u) du \, \delta$$

Combining this with the fact that $k(\cdot)$ is integrable on $[0, 1]$ and hence its integral has a finite value, $\int_0^1 k(u) \, du = M < \infty$, we find that

$$\text{II} = \varepsilon \sum_{k=0}^{n-1} \int_{t_k}^{t_{k+1}} k(\frac{t}{\delta} - \lfloor \frac{t}{\delta} \rfloor \delta) \, dt = \sum_{k=0}^{n-1} \delta \int_0^1 k(u) \, du \, \varepsilon = TM\varepsilon$$

Using the same transformation, we now compute I:

$$\text{I} = \sum_{k=0}^{n-1} f(t_k) \int_{t_k}^{t_{k+1}} k \left(\frac{t}{\delta} - \lfloor \frac{t}{\delta} \rfloor \delta \right) dt = |u = \frac{t}{\delta} - \lfloor \frac{t}{\delta} \rfloor \delta, \delta du = dt| = \sum_{k=0}^{n-1} f(t_k) \int_0^1 k(u) \, du\delta$$

$$= \int_0^1 k(u) \, du \sum_{k=0}^{n-1} f(t_k)\delta$$

Putting our estimate for II and transformation of I together, we get that

$$\sum_{k=0}^{n-1} \int_{t_k}^{t_{k+1}} f(t)k \left(\frac{t}{\delta} - \lfloor \frac{t}{\delta} \rfloor \delta \right) dt < \int_0^1 k(u) \, du \sum_{k=0}^{n-1} f(t_k)\delta + M\varepsilon$$

By uniform continuity, we know we can make the error ε arbitrarily small by choosing a larger n. At the same time, as we increase n, we see that the sum, thanks to the fact that function f is continuous, will converge to a Riemann integral:

$$\sum_{k=0}^{n-1} f(t_k)\delta \xrightarrow{n \to \infty} \int_0^T f(t) \, dt$$

We can see that the lower bound will also converge to that same integral with an arbitrarily small error. Therefore the term between the two bounds will also converge to that integral and we obtain the desired final result.

Finally we note that the above sum converges to a Riemann integral, but whenever the Riemann integral exists, it coincides with the Lebesgue integral (see e.g. Billingsley [1995, pg. 222]), ensuring the existence of the Lebesgue integral.

We have thus proved the theorem for a continuous function, now all we need to state is the well-known fact that the set of continuous functions with compact support is dense in L^1 space, and hence any integrable function g can be approximated to arbitrary precision with a continuous function f in L^1, or mathematically $\forall \varepsilon > 0 :$ $\int_0^T |g(t) - f(t)| \, dt < \varepsilon.$ $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ $\square$

In our case, where $k(x) = 1 - x$, we have $\displaystyle\int_0^1 k(x) \, dx = \frac{1}{2}$. Thus overall,

$$\mathbb{E}_0 \int_0^{T-\eta} dt \int_{\varphi(t)}^t d\theta \left(\Gamma_\theta S_\theta^2\right)^2 \sigma^4 = \left(\frac{1}{2}\sigma^4 \int_0^{T-\eta} \mathbb{E}_0 \left[\left(\Gamma_t S_t^2\right)^2\right] dt\right)\delta,$$

which is exactly what we obtained in our analysis for errors over $[0, T - \eta]$.

Let us now consider the integral we are computing from Gobet and Temam [2001b] over time period $[T-\eta, T]$. Based on theorem 1.10 we know that $\mathbb{E}_0[(\Gamma S^2)^2]$ is of order $\mathcal{O}((T - t)^{-3/2})$; the term of this order is the source of the singularity and explosive behaviour of the Cash Gamma squared when nearing maturity. From the form of the solution of the expected Cash Gamma squared in equation (1.27) we can see that there exists a function g continuous on $[T - \eta, T]$ s.t.

$$\sigma^4 \mathbb{E}_0[(\Gamma_t S_t^2)^2] = \frac{g(t)}{(T - t)^{3/2}}.$$

Let us note that this defines the same function $g(t)$ as is given by Gobet and Temam [2001b, eqn. 18]. In this new notation, we thus want to investigate the behaviour of the integral

$$\int_{T-\eta}^T ds \int_{\varphi(s)}^s \frac{g(t)}{(T - t)^{3/2}} \, dt \qquad\qquad (1.31)$$

Since $g(t)$ is continuous it does not cause explosive behaviour and $\forall \varepsilon > 0$ we can always choose our fixed η in such a way that $|g(t) - g(T)| < \varepsilon$, i.e. we can make the approximation error made by using $g(T)$ as small as we want. The value of the

constant $g(T)$ can be directly computed to be

$$g(T) = \lim_{t \to T} \mathbb{E}_0[(\Gamma_t S_t^2)^2](T - t)^{3/2} = \frac{\sigma}{4\sqrt{\pi}}\tilde{\varphi},$$

where $\tilde{\varphi}$ is the same as in theorem 1.28 in the previous section. Here we should point out that Gobet and Temam [2001b] incorrectly computes this limit to be

$$\frac{\sigma^3}{4\sqrt{\pi}}\tilde{\varphi},$$

and a correction to this article is provided in Černý and Špilda [2012].

We are now mainly interested in the asymptotic behaviour of the function $(T - t)^{-3/2}$ when we integrate it, as this will provide us with asymptotic behaviour of our integral. The following general theorem provides us with the asymptotic behaviour of the integral.

Theorem 1.14. *For $\alpha > 1$, $\eta > 0$, $\delta = t_{k+1} - t_k = \frac{T}{n}$, and assuming that η/δ is a whole number, it holds that*

$$\int_{T-\eta}^{T} ds \int_{\varphi(s)}^{s} \frac{dt}{(T - t)^\alpha} = \delta^{2-\alpha} \sum_{j=1}^{\eta/\delta} \left(\frac{1}{(1-\alpha)j^{\alpha-1}} - \frac{1}{(1-\alpha)(2-\alpha)}\left(\frac{1}{j^{\alpha-2}} - \frac{1}{(j-1)^{\alpha-2}} \right) \right)$$

Furthermore, the integral has an asymptotic form in terms of δ given as

$$\delta^{2-\alpha}\zeta_\alpha - \frac{1}{2(\alpha-1)}\frac{1}{\eta^{\alpha-1}}\delta + O(\delta^2)$$

where

$$\zeta_\alpha = \lim_{\eta/\delta \to \infty} \sum_{j=1}^{\eta/\delta} \left(\frac{1}{(1-\alpha)j^{\alpha-1}} - \frac{1}{(1-\alpha)(2-\alpha)}\left(\frac{1}{j^{\alpha-2}} - \frac{1}{(j-1)^{\alpha-2}} \right) \right)$$

Proof. We want to compute

$$\int_{T-\eta}^{T} ds \int_{\varphi(s)}^{s} \frac{dt}{(T - t)^\alpha}$$

where $\alpha > 1$. Let k^* be defined by the relation $T - \eta = k^*\delta$. We can rewrite the

problem:

$$
\int_{T-\eta}^{T} ds \int_{\varphi(s)}^{s} \frac{dt}{(T-t)^\alpha} = \sum_{k=k^*}^{n-1} \int_{t_k}^{t_{k+1}} ds \int_{t_k}^{s} \frac{dt}{(T-t)^\alpha}
$$

$$
= \sum_{k=k^*}^{n-1} \int_{t}^{t_{k+1}} ds \int_{t_k}^{t_{k+1}} \frac{dt}{(T-t)^\alpha}
$$

$$
= \sum_{k=k^*}^{n-1} \int_{t_k}^{t_{k+1}} \frac{t_{k+1}-t}{(T-t)^\alpha} dt
$$

$$
= \left| u = \frac{t_{k+1}-t}{\delta} \right| = \sum_{k=k^*}^{n-1} \int_{0}^{1} \delta \frac{\delta u}{(T-t_{k+1}+\delta u)^\alpha} du
$$

$$
= \sum_{k=k^*}^{n-1} \delta^2 \int_{0}^{1} \frac{u\, du}{(n-1-k+u)^\alpha \delta^\alpha}
$$

$$
= \sum_{k=k^*}^{n-1} \delta^{2-\alpha} \int_{0}^{1} \frac{u\, du}{(n-1-k+u)^\alpha}
$$

$$
= |j = n-k| = \sum_{j=1}^{n-k^*} \delta^{2-\alpha} \int_{0}^{1} \frac{u\, du}{(j-1+u)^\alpha}
$$

We note that due to the fact that $T-\eta = k^*\delta$, we can write the summation as being from 1 to η/δ, since

$$
n-k^* = n - \frac{T-\eta}{\delta} = n - n + \frac{\eta}{\delta} = \eta/\delta.
$$

We now compute

$$
\int_{0}^{1} \frac{u\, du}{(j-1+u)^\alpha}.
$$

We can solve this problem through integration by parts:

$$
\int_{0}^{1} \frac{u\, du}{(j-1+u)^\alpha} = |p.p., f(u) = u, g'(u) = (u+j-1)^{-\alpha}|
$$

$$
= \left[\frac{u}{(1-\alpha)(u+j-1)^{\alpha-1}} \right]_{0}^{1} - \frac{1}{1-\alpha} \int_{0}^{1} \frac{du}{(j-1+u)^{\alpha-1}}
$$

$$
= \frac{1}{(1-\alpha)j^{\alpha-1}} - \frac{1}{1-\alpha} \left[\frac{1}{2-\alpha} \frac{1}{(j-1+u)^{\alpha-2}} \right]_{0}^{1}
$$

$$
= \left(\frac{1}{(1-\alpha)j^{\alpha-1}} - \frac{1}{(1-\alpha)(2-\alpha)} \left(\frac{1}{j^{\alpha-2}} - \frac{1}{(j-1)^{\alpha-2}} \right) \right)
$$

If we now substitute back into the original problem, we get the desired result from

the first part of the theorem.

We now continue to find the asymptotic expansion of the sum. We assume that for $\alpha > 1$, the sum converges for $\eta/\delta \to \infty$ (this can be verified numerically).We will denote the value it converges to as ζ_α. Then we can write the sum as follows:

$$S = \sum_{j=1}^{\eta/\delta} \left(\frac{1}{(1-\alpha)j^{\alpha-1}} - \frac{1}{(1-\alpha)(2-\alpha)} \left(\frac{1}{j^{\alpha-2}} - \frac{1}{(j-1)^{\alpha-2}} \right) \right)$$

$$= \zeta_\alpha - \sum_{j=\eta/\delta+1}^{\infty} \left(\frac{1}{(1-\alpha)j^{\alpha-1}} - \frac{1}{(1-\alpha)(2-\alpha)} \left(\frac{1}{j^{\alpha-2}} - \frac{1}{(j-1)^{\alpha-2}} \right) \right)$$

We now try to estimate the end "tail" of the infinite sum. First we note that

$$\frac{1}{(2-\alpha)} \left(\frac{1}{j^{\alpha-2}} - \frac{1}{(j-1)^{\alpha-2}} \right) = \int_{j-1}^{j} \frac{dx}{x^{\alpha-1}}.$$

Thus

$$S = \zeta_\alpha - \left[\frac{1}{1-\alpha} \sum_{j=\eta/\delta+1}^{\infty} \frac{1}{j^{\alpha-1}} - \frac{1}{1-\alpha} \int_{\eta/\delta}^{\infty} \frac{dx}{x^{\alpha-1}} \right]$$

$$= \zeta_\alpha - \frac{1}{1-\alpha} \left[-\frac{1}{\left(\frac{\eta}{\delta} \right)^{\alpha-1}} + \sum_{j=\eta/\delta}^{\infty} \frac{1}{j^{\alpha-1}} - \int_{\eta/\delta}^{\infty} \frac{dx}{x^{\alpha-1}} \right]$$

We now use the Euler-MacLaurin summation formula (see Kac and Cheung [2002, eqn. 25.9]), which states that if the function $f(x)$ and its derivatives decay to 0 as $x \to 0$, then it holds that

$$\sum_{n=a}^{\infty} f(n) = \int_{a}^{\infty} f(x)\,dx + \frac{1}{2}f(a) + \mathcal{O}(f'(a)).$$

Using this, we can see that

$$S = \zeta_\alpha - \frac{1}{1-\alpha} \left[-\left(\frac{\delta}{\eta} \right)^{\alpha-1} + \frac{1}{2} \left(\frac{\delta}{\eta} \right)^{\alpha-1} + O(\delta^\alpha) \right] = \zeta_\alpha - \frac{1}{2(\alpha-1)} \left(\frac{\delta}{\eta} \right)^{\alpha-1} + O(\delta^\alpha)$$

Now all we need to do is multiply by $\delta^{2-\alpha}$ to get the asymptotic value of the integral

60

we are computing:

$$\int_{T-\eta}^{T} ds \int_{\varphi(s)}^{s} \frac{dt}{(T-t)^{\alpha}} = \delta^{2-\alpha} S = \delta^{2-\alpha} \left(\zeta_{\alpha} - \frac{1}{2(\alpha-1)} \left(\frac{\delta}{\eta} \right)^{\alpha-1} + O(\delta^{\alpha}) \right)$$

$$= \delta^{2-\alpha} \zeta_{\alpha} - \frac{1}{2(\alpha-1)} \frac{1}{\eta^{\alpha-1}} \delta + O(\delta^2)$$

$\square$

Let us notice that when we plug $\alpha = 3/2$ into the result obtained above, we get that

$$\int_{T-\eta}^{T} ds \int_{\varphi(s)}^{s} \frac{dt}{(T-t)^{3/2}} = \sqrt{\delta} \zeta_{3/2} - \frac{\delta}{\sqrt{\eta}} + \mathcal{O}(\delta^2)$$

where the infinite sum

$$\zeta_{3/2} = \sum_{k=1}^{\infty} \left(\frac{-2}{\sqrt{k}} + 4(\sqrt{k} - \sqrt{k-1}) \right) = 4 \sum_{k=1}^{\infty} \left((\sqrt{k} - \sqrt{k-1}) - \frac{1}{2\sqrt{k}} \right) = 4\tilde{\zeta}$$

directly connects to our variable $\tilde{\zeta}$ in equation (1.29). Therefore the asymptotics of the integral (1.31) we wanted to compute are

$$g(T) \int_{T-\eta}^{T} ds \int_{\varphi(s)}^{s} \frac{dt}{(T-t)^{3/2}} = \frac{\sigma}{4\sqrt{\pi}} \tilde{\varphi} 4\tilde{\zeta} \sqrt{\delta} - \frac{\delta}{\sqrt{\eta}} \frac{\sigma}{4\sqrt{\pi}} \tilde{\varphi}$$

$$= \left(\frac{\sigma}{\sqrt{\pi}} \tilde{\varphi} \tilde{\zeta} \right) \sqrt{\delta} - \left(\frac{\sigma}{4\sqrt{\pi}} \tilde{\varphi} \right) \frac{\delta}{\sqrt{\eta}}$$

We can put all these results together, and find that

$$\mathbb{E}_0 \int_0^T dt \int_{\varphi(t)}^{t} d\theta \left(\Gamma_{\theta} S_{\theta}^2 \right)^2 \sigma^4 = \mathbb{E}_0 \int_0^{T-\eta} dt \int_{\varphi(t)}^{t} d\theta \left(\Gamma_{\theta} S_{\theta}^2 \right)^2 \sigma^4 + \mathbb{E}_0 \int_{T-\eta}^{T} dt \int_{\varphi(t)}^{t} d\theta \left(\Gamma_{\theta} S_{\theta}^2 \right)^2 \sigma^4$$

$$= \left(\frac{1}{2} \sigma^4 \int_0^{T-\eta} \mathbb{E}_0 \left[\left(\Gamma_t S_t^2 \right)^2 \right] dt \right) \delta + \left(\frac{\sigma}{\sqrt{\pi}} \tilde{\varphi} \tilde{\zeta} \right) \sqrt{\delta} - \left(\frac{\sigma}{4\sqrt{\pi}} \tilde{\varphi} \right) \frac{\delta}{\sqrt{\eta}}$$

We showed previously in theorem 1.11 that this value converges for $\eta \to 0$. Therefore, we have shown that in the martingale setting of model (1.7), the asymptotics of the first-order term that contributes to the hedging error of a Black-Scholes Δ-hedging strategy as computed in Gobet and Temam [2001b] (given the correction in Černý and Špilda [2012]) is exactly the same as that obtained when following a discrete-time variance-optimal hedging strategy. Furthermore, since we showed for the variance-optimal case that the second-order δ^2 term is in fact significant due to

the explosive behaviour of the Cash Gamma at maturity, the same will apply for the Black-Scholes tracking error, which we have shown is given by the same formula.

1.5   Conclusion

In this chapter we set out to investigate the asymptotic behaviour of the quadratic hedging error for a digital call option with respect to the (increasing) frequency δ of the discrete-time variance-optimal hedging strategy. We first contrasted the asymptotic behaviour of a single-step hedging error of a vanilla and digital call at the last step prior to maturity, and showed that these have asymptotics of different orders. We showed a connection between the single-step hedging error and the Cash Gamma of a digital option. Using those results we showed how to sum up all the individual single-step errors and compute the asymptotics of the overall *expected* quadratic hedging error. Finally, we compared the δ asymptotics of the variance optimal hedging strategy (optimal in discrete time), against those of the discretization of the continuous-time Black-Scholes Δ-hedging. We conclude that in our model, the asymptotics of the two hedging errors coincide.

Chapter 2

Good-deal bounds of variance swaps and the Lévy contract

In the first chapter, we saw that the asymptotic hedging error due to discrete-time trading of an optimally hedged contingent claim in the Black-Scholes model is determined primarily by the path-dependent Cash Gamma risk and that in the case of a digital option, the properties of the Cash Gamma near maturity make it hard to deal with. The main focus was on market incompleteness and hedging errors caused by hedging our position at a finite set of dates. In this chapter, we investigate what other incomplete market risks we may still have in a contract which no longer contains an unpredictable Cash Gamma risk. If we were to investigate a specific contract - the log contract - we would find that its Cash Gamma is constant and predictable. This makes the expected size of the discretization error completely predictable.

In fact we will look at a more general contract, which we will refer to as a "Lévy contract". The payoff of this contract will not only encompass the log contract, but also the more practical case of the variance swap, whose theoretical value is strongly related to that of the log contract - when continuously sampled, it is the payoff of a Δ-hedged log contract. It will also allow us to look at higher order moment swaps such as skewness and kurtosis swaps.

The outstanding risks we set out to assess in such a contract are those due to jumps and the higher order moments of the distribution. For this purpose, we will work in the setting of exponential Lévy models, for which the current mathematical tools still allow us to obtain reasonably explicit formulas and computations, while allowing

more general distributions, uncovering risks that were previously ignored within a diffusion model. From Broden and Tankov [2011] it follows that the (here constant) Cash Gamma is the main driver of asymptotic discretization hedging error even when adding jumps to the underlying, so we have not added any additional discretization risk by making our model more general. To keep a focus on jumps as the source of market incompleteness, we will revert to a continuous-time setting. We will also limit ourselves to a model with constant diffusion volatility to exclusively gauge the uncertainty in pricing due to the presence of jumps in the model. This means that in the small jump limit when our model has no jumps, a variance swap will have a deterministic fixed price.

We will investigate the impact of jump risk on the price of the "Lévy contract" (and the contracts derived from it) by looking at 'good-deal' price bounds within an incomplete market where perfect replication is not possible; these price bounds reflect possible hedging errors due to higher moments of the returns distribution, and more generally uncertainty regarding the price of the contract. We will produce these price bounds using two closely related (yet mostly separately handled in the literature) methodologies: exponential utility-based pricing and pricing via variance-optimal hedging.

2.1 Motivation, literature review and research question

The idea of introducing a contract that would have a constant Cash Gamma first arose from the working paper Neuberger [1990] (and accompanying journal article Neuberger [1994]), and independently from that in Dupire [1992]. It was shown that Δ-hedging a log contract provided the holder with a payoff that was (nearly) perfectly correlated with realized volatility of the underlying. The contract would thus allow traders to trade realized volatility in the market. In practice the log contract never traded (possibly due to its negative payout for low values of the underlying); in 1993, however, the CBOE introduced a new product called the VIX, which was intended for similar use, i.e. to trade volatility. The first large interest in the index was shown after the LTCM crash, when many institutions found themselves lacking protection against the sudden burst of volatility. Originally, the index was a weighted average of options on the S&P 500 for a few options near the ATM position and resembled a

slightly more complex straddle. The index was later redefined in 2003 in reaction to publications such as Demeterfi et al. [1999] and Carr and Madan [2002] showing the connection between the (continuous-time) variance swap and its replication via a Δ-hedged position in a log contract, which in turn is replicated by a weighted portfolio of (an infinite number of) vanilla calls and puts. The redefined VIX still essentially remained a weighted basket of options, but now considering a wider range of vanilla options in its computation. The exact payoff of the VIX is defined in CBOE [2009] as $\sigma \times 100$, where:

$$\sigma^2 = \frac{2}{T} \sum_i \frac{\Delta K_i}{K_i^2} e^{rT} Q(K_i) - \frac{1}{T} \left[1 - \frac{F}{K_0} \right]^2,$$

with K_i denoting strike prices, $Q(K_i)$ the corresponding call/put option price and F the forward price. The payoff is the discretized version of the theoretical replication of a log contract via vanilla options. The popularity of the VIX has led to equivalent products being introduced on other underlying indices, e.g. the VDAX for the DAX Index or the VSTOXX for the STOXX 50 Index. Let us note, however, that a majority of variance swaps are bespoke products that are traded OTC between investment firms. Its popularity has also spurred research looking into the possibilities of trading higher order moments of the distribution, e.g. Neuberger [2012], Schoutens [2005], Corcuera et al. [2005], Nadtochiy and Obłój [2017].

2.1.1 Motivating the log contract

The reason for introducing a log contract is quite simple. If we consider a diffusive model of the stock S with (potentially time and spot-dependent) volatility $\sigma(t, S)$ and apply Itô's lemma, we find:

$$d \log S_t = \frac{dS_t}{S_t} - \frac{1}{2} \sigma^2(t, S_t) dt.$$

Rearranged and integrated, we get that the model's total integrated variance is equal to selling 2 log contracts and holding a continuously rebalanced Δ-hedge of $2/S_t$ units of the stock:

$$\frac{1}{T} \int_0^T \sigma^2(t, S_t) dt = \frac{2}{T} \left(\int_0^T \frac{dS_t}{S_t} - \log(S_T/S_0) \right)$$

We can illustrate in this diffusive model how trading a log contract gives us exposure to realized volatility. If we consider a general Δ-hedged portfolio consisting of a (generic) claim with current value $V(t, S)$, then such a portfolio experiences a daily P&L of

$$\text{P\&L}_{\Delta t} = \frac{\partial V}{\partial t} \Delta t + \frac{1}{2} \frac{\partial^2 V}{\partial S^2} (\Delta S)^2$$

However, using classic portfolio replication arguments, we know a generalized Black-Scholes *forward* PDE holds:

$$\frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2(t, S) S^2 \frac{\partial^2 V}{\partial S^2} = 0,$$

so substituting for the time derivative into our P&L, we get that over time Δt, the P&L is

$$\text{P\&L}_{\Delta t} = \frac{1}{2} S^2 \frac{\partial^2 V}{\partial S^2} \left(\frac{(\Delta S)^2}{S^2} - \sigma^2(t, S) \Delta t \right).$$

We can see that the P&L on a hedged position is highly path-dependent due to the cash Gamma. However, if we now consider a contingent claim for which

$$\frac{\partial^2 V}{\partial S^2} = \frac{1}{S^2} \tag{2.1}$$

exactly, we could get rid of any path dependency. Then the P&L of the portfolio summed over all time intervals Δt would be

$$\text{P\&L}_T = \sum_t \frac{1}{2} \left(\frac{(\Delta S_t)^2}{S_t^2} - \sigma(t, S_t)^2 \Delta t \right)$$

which is the difference between (a particular definition of) realized variance over the duration of the contract and the model-assumed variance. The contract that satisfies condition (2.1) is one that has terminal value $V_T = a - \log S_T + b S_T$, where a, b are constants of integration. Thus, by selling and Δ-hedging two log contracts and holding $2a$ in cash and $2b$ units of the underlying, we obtain a portfolio whose value is exactly the difference between (a particular definition of) realized and model-assumed volatility.

2.1.2   Our research question

Although in the beginning of the line of research on log contracts and variance swaps, all the computations appeared to be sufficiently model-independent, over time impor-

tant flaws have been found in the theory. It has been acknowledged that the relation between the log contract and the replicating strip of options is imperfect, and there is no unique market-enforceable price for a variance swap based purely on available option prices.

The literature has therefore at several points in time attempted to provide (somewhat) model-independent arbitrage *bounds* for the price of variance swaps. The approach usually taken is a static hedging one, where you find a portfolio of vanilla options super- and subreplicating the log contract payoff that would provide a bound. Davis et al. [2014] derive these under the assumption of a continuous semimartingale model (i.e. the underlying has no jumps). They find their lower bound to be remarkably close to actual variance swap prices traded in the market, and the upper bound is potentially infinite. In contrast, Hobson and Klimmek [2011] develop model-independent no-arbitrage bounds that allow for jumps, but in return require a continuous range of options available for hedging. The bounds they find are not particularly tight, the price easily being allowed to double without causing arbitrage.

The importance of including jumps to model variance swap prices has been shown in Crosby and Davis [2011] and Carr et al. [2012], who report a divergence of actually traded prices of variance swaps from their theoretical replication value based on diffusive models. They show that you typically need more than 2 Δ-hedged log contracts to replicate a variance swap, as implied by time-changed Lévy models calibrated to market option prices. The practice of using diffusions for pricing variance swaps has also been strongly criticized in more practitioner-oriented literature (Ayache [2006]).

In this context, the question we attempt to answer is whether we can find useful, potentially tighter bounds on the price of a variance swap (and related contracts) while recognizing the importance of jumps in their pricing and hedging.

Thus, the first contribution of this chapter is to extend the limited research into incomplete market pricing of variance swaps when the underlying includes jumps, as typically the pricing is done in a diffusion setting, with market incompleteness stemming from stochastic volatility, as in e.g. Grasselli and Hurd [2007]. We will do our analysis in the established framework of exponential Lévy processes - these are analytically tractable but provide a richer set of distributions, thus letting us see the influence of higher order moments such as skewness and kurtosis via the Lévy jump density, but do not account for stochastic volatility. We provide appropriate skew (and kurtosis) adjustments to the price of a variance swaps as our model departs

from the Black-Scholes framework.

Secondly, instead of wide no-arbitrage bounds, we aim to provide tighter, economically rational bounds on the price of the "Lévy contract". We will obtain tighter price ranges by not only eliminating prices that imply pure arbitrage opportunities, but also those that would lead to "good deals", i.e. attractive investments with high risk-adjusted returns. The concept of "good deal bounds" was first introduced in Cochrane and Saá-Requejo [2000], its main idea being that we can obtain upper and lower price limits by setting an upper bound to the maximal available Sharpe ratio attainable in the market that contains the derivative and the underlying asset. Further developments on good deal bounds then came in Černý and Hodges [2002], Černý [2003], Björk and Slinko [2006], Klöppel and Schweizer [2008], where the authors show the relation of good-deal bounds and maximizing Sharpe ratios to the theory of quadratic hedging. Unlike sub/superreplication bounds, this approach can yield useful price bounds without considering the existence of any options traded in the market.

We will use utility-based pricing (akin to Grasselli and Hurd [2007], see Henderson and Hobson [2009] for an introduction) and the concept of certainty equivalents to rule out prices that are too high or low, by setting an upper bound to the normalized certainty equivalent, the so-called investment potential. We will do this for two different utility functions: exponential and quadratic utility. These allow for analytically explicit (but not necessarily closed-form in the case of exponential utility) solutions. We will show that for the case of the quadratic utility function, utility-based good-deal bounds can be related to the theory of quadratic hedging as described in the previous chapter.

Due to the structure of the solution under exponential utility, where the optimal hedging strategy can only be computed *implicitly*, we also compute an asymptotic approximation as jumps in our model diminish in size and intensity to gain better insight into the small-jump properties of the solution. We obtain approximate closed-form formulas based on the higher-order moments of the distribution when the jumps we experience are reasonably small in the sense of Černý et al. [2013].

Thirdly, we will compare results for exponential and quadratic utility, and we will find that the asymptotic solution for exponential utility yields bounds that differ only minorly from the full explicit solution of quadratic utility. This will show that in settings where our underlying distribution has fat tails but these are not too severe, we lose little by using well-known simple formulae from quadratic hedging instead

of dealing with complex implicit systems of equations from exponential utility pricing. In other words, we show that despite some of its theoretical shortcomings (a non-monotonic utility function), in most practical situations quadratic hedging is a suitable substitute for other more theoretically sound but also less practical and more cumbersome pricing systems.

Connections between the two utilities have previously been made in terms of the asymptotic behaviour of prices when buying in small quantities. Kramkov and Sîrbu [2007] show that for a small number of claims, any standard utility indifference price should asymptotically correspond to a mean-variance hedging price, albeit under a new martingale measure and numéraire. Further work on investigating this connection between exponential and quadratic utility, in various degrees of generality, has been done in Becherer [2006], Mania and Schweizer [2005],Kallsen and Rheinländer [2011]. In contrast, we will be able to compare utility indifference pricing with mean-variance hedging under physical measure, letting us see that a close relation between the two still persists under that measure for the case of variance swaps (or more generally "Lévy contracts").

Finally, we contribute by showing the usefulness of the general "Lévy contract" under investigation, which encapsulates log contracts, variance swaps and higher order swaps simultaneously, akin to a generalization made in Carr and Lee [2013]. Whereas that paper directly considers polynomial transformations of jumps of the Lévy measure (which give moment swap payoffs), we extend their results by first maintaining a more general approach in our analysis, and only reverting to polynomial transformations later on, when it is essential to gain insight into the asymptotics.

2.2 General utility-based pricing theory

In this section we will introduce concepts from general utility-based pricing theory, and how it relates to mean-variance hedging.

Pricing and risk management based on utility theory in general considers a functional u which models an agent's preferences between different (potentially random) payoffs in the future. Mathematically speaking, an agent will prefer random payoff X over Y only if $u(X) > u(Y)$. We then set additional assumptions on the properties of u to model an agent's (rational) behaviour. The two most commonly made assumptions

are the following:

1. u is monotone in a $\mathbb{P}$-a.s. sense, i.e. $u(X) > u(Y) \Leftrightarrow X > Y\,\mathbb{P}$-a.s. In words, the agent always prefers more to less.

2. u is a concave function, i.e. $u(\lambda X + (1 - \lambda)Y) \geq \lambda u(X) + (1 - \lambda)u(Y)$. Economically speaking, an agent does not lose utility from diversification.

In recent years, a wide body of research has shown (see Ben-Tal and Teboulle [2007a], Filipović and Kupper [2007], Filipović and Kupper [2008], Cheridito and Kupper [2009], Černý et al. [2012], Cheridito et al. [2015] and references therein) that it is fruitful to add a third condition to the previous two:

3. u is translation invariant. $u(X + m) = u(X) + m, m \in \mathbb{R}$. which in words means that an agent prefers the payout X over another payout Y if and only if he prefers $X - m$ to $Y - m$ for all m. Economically speaking, the agent's utility of an uncertain payoff is not altered by any additional holdings in cash, i.e. his investment is only dependent on the risk/reward profile of the payoff X, and is independent of any cashflow considerations (for that reason this property is sometimes also referred to as cash invariance).

A utility function u that has all three properties is referred to as a monetary utility function (Filipović and Kupper [2008]). Such a utility function can be in turn related to convex risk measures ρ as defined by Föllmer and Schied [2002]: $\rho(X) = -u(X)$.

Recent research, such as Filipović and Kupper [2007] and Cherny and Kupper [2007], has shown a connection between these general functionals $u(X)$ and the more classic situation of expected utility $\mathbb{E}[U(X)]$, as first used for pricing derivatives in Hodges and Neuberger [1989]. These papers show that the monetary utility function u corresponds to the translation-invariant hull of a classical expected utility. This means it is the smallest translation-invariant functional that dominates the expected utility value. Filipović and Kupper [2007] show that this connection even applies to the non-monotone quadratic utility function $U(x) = x - \frac{\alpha}{2}x^2$, whose translation-invariant hull is shown to be the mean-variance preference functional $u(X) = \mathbb{E}[X] - \frac{\alpha}{2}\text{Var}(X)$ (this can be extended to monotone truncated quadratic utility, as shown in Černý et al. [2012]). We will later see that this, in turn, relates to the problem of mean-variance hedging.

Our goal will be to use utility-based pricing methods to compute good-deal bounds on a "Lévy contract" (which encompasses variance swaps) in an incomplete market model to investigate the price impact of jumps. We will do so both under the theoretically unsound non-monotone mean-variance preferences (i.e. the quadratic utility function) and those implied by the theoretically more amenable exponential utility function.

For both utility functions, we will investigate the translation-invariant version of the expected utility maximization problem, to be able to operate under a united framework. Specifically, we set out to find the maximal translation-invariant expected utility for a portfolio holding $\vartheta_t, t \leq T$ in the stock and q units of a contingent claim with payout H at expiry T:

$$u_\gamma(p,q) = \max_{\vartheta \in \Theta_\gamma(p,q)} \max_{\eta \in \mathbb{R}} \eta + \mathbb{E}[f_\gamma(\vartheta \cdot S_T + q(H-p) - \eta)] \qquad (2.2)$$

In our problem, we consider a normalized HARA utility function f_γ as in Brooks et al. [2012]:

$$f_\gamma(x) = \begin{cases} \frac{(1+x/\gamma)^{1-\gamma}-1}{1/\gamma-1}, & \text{for } \gamma > 0 \\ \ln(1+x), & \text{for } \gamma = 1 \\ \frac{|1+x/\gamma|^{1-\gamma}-1}{1/\gamma-1} & \text{for } \gamma < 0 \\ 1-e^{-x} & \text{for } \gamma = \pm\infty \end{cases}$$

We will focus on the cases $\gamma = -1$ and $\gamma = \infty$, which correspond to mean-variance preferences and exponential utility respectively. We will motivate our reason for normalizing the utility function to have a risk aversion equal to 1 below. As the problem is cash-invariant (by property 3 above), we do not consider any initial wealth.

Let us note that as a consequence of the translation (or cash) invariance of our preferences, we can decompose the utility function into two parts:

$$u_\gamma(p,q) = u_\gamma(0,q) - pq$$

(a simple way to see this is by e.g. defining $X_T = \vartheta \cdot S_T + qH - \eta$ and applying the translation invariance: $u(X_T - pq) = u(X_T) - pq$)

We will choose a set of admissible strategies $\Theta_\gamma(p,q)$ that encompasses both exponential utility and mean-variance preferences, specifically the definition of admissible strategies from [Biagini and Černý, 2011, Definition 1.1] (see definition B.4 in the appendix for details). This set of admissible strategies is well-defined for Lévy models

with finite moments and exponential moments, which will be our setting later on. Under this set of admissible strategies, the optimizer is also within the set, allowing us to consider maxima instead of suprema.

Once we have the maximal utility u_γ, we can compute the optimal quantity to purchase at a given price p:

$$\hat{q}_\gamma(p) = \arg\max_q u_\gamma(p, q),$$

From that we can define the so-called investment potential (IP) of the market with both the underlying and the derivative at price p to measure how utility an agent can gain from investing in the underlying and derivative claim optimally:

$$IP_\gamma(p) := u_\gamma(p, \hat{q}_\gamma(p)).$$

The investment potential (IP) is a normalized version of the certainty equivalent (CE) gain, stating the percentage gain in certainty equivalent wealth per unit of risk aversion. It corresponds to what Ben-Tal and Teboulle [2007b] refer to as the Optimized Certainty Equivalent (see [Černý, 2009, Section 3.5] for a further introduction on the investment potential). The classical certainty equivalent gain CE is an amount of risk-free cash we need to add to our current wealth to obtain the same level of utility as an investment in the risky asset:

$$u_\gamma(CE_\gamma(p, q), 0) = u_\gamma(p, q)$$

It provides a gauge for the lucrativeness of investing in a particular asset for the agent. The only parameter it depends on is the shape of the normalized utility, determined by γ. As a normalized measure, it is invariant on the risk aversion of a particular agent and allows for comparison across agents. An agent who already holds some of the derivative being priced but with lower risk aversion will gain the same amount of utility from the claim at price p as another agent who holds less of the derivative but is more risk averse. It can be shown (see Černý et al. [2012] and the references within) that in the case of mean-variance preferences ($\gamma = -1$), $IP_\gamma(p)$ is directly related to the maximal Sharpe ratio of the market with the derivative and underlying:

$$IP_{-1}(p) = \frac{1}{2}SR^2$$

As we want to assess the specific investment potential of the derivative in addition to the pre-existing investment opportunity in the underlying, we will primarily be

interested in the difference between the IP of a market with and without the derivative (where the derivative is available at price p):

$$\Delta IP_\gamma(p) = u_\gamma(p, \hat{q}_\gamma(p)) - u_\gamma(\cdot, 0) \tag{2.3}$$

As we now have a measure connecting lucrativeness of investment to the price of a contingent claim in the market, we can invert this relationship to obtain good-deal bounds of the derivative $\hat{p}_\gamma^\pm(\Delta IP)$ (the superscript $\pm$ highlights the fact that there is an upper and lower bound, as $\Delta IP_\gamma(p)$ will be a function convex in p with two different values of p giving the same IP). For a given level of IP added into the market, we will obtain a lower and upper price bound for the derivative.

Inversely, if we consider q to be given and adjust p accordingly to solve our optimisation problem, we get the optimal price at a given quantity. From first order conditions and the translation invariance equality above we have

$$\hat{p}_\gamma(q) = \frac{d}{dq} u_\gamma(0, q)$$

and the related IP

$$IP_\gamma(q) := u_\gamma(\hat{p}_\gamma(q), q) = u_\gamma(0, q) - q \frac{d}{dq} u_\gamma(0, q)$$

We will refer to $\hat{p}_\gamma(q)$ as the **utility-based price**, in line with Kramkov and Sîrbu [2007], which can also be understood as a marginal utility price for an investor who already holds q contingent claims as his initial wealth. This can be contrasted with the **indifference price** $p_\gamma^I(q)$ which is defined as a price at which the agent is indifferent between receiving the contingent claim or a lump sum of cash now:

$$p_\gamma^I(q) : u_\gamma(p_\gamma^I(q), q) = u_\gamma(\cdot, 0)$$

If we revisit the definition of ΔIP_γ in equation (2.3), we see that it ultimately measures the distance in utility between the utility-based and utility indifference price:

$$\Delta IP_\gamma(p) = u_\gamma(p, \hat{q}_\gamma(p)) - u_\gamma(\cdot, 0) = u_\gamma(p, \hat{q}_\gamma(p)) - u_\gamma(p^I, q^I(p))$$

2.2.1 Mean-variance preferences

We mentioned previously that the utility maximization with mean-variance preferences and quadratic hedging are closely linked. In the following theorem and its corollary, we show this connection and how it can then be used to compute good-deal bounds by setting bounds to the added investment potential ΔIP_{-1}. We will find that the width of the good-deal bounds for quadratic utility depends primarily on the size of the expected root-mean-square hedging error ε_0 (as defined in the first chapter).

Theorem 2.1. *For mean-variance preferences ($\gamma = -1$), the indirect utility function can be expressed in terms of variables from mean-variance hedging:*

$$u_{-1}(p, q) = \frac{1}{2}(L_0^{-1} - 1) + q(V_0 - p) - \frac{1}{2}q^2 \varepsilon_0^2$$

$$L_0 = \min_{\vartheta} \mathbb{E}[(1 - \vartheta \cdot S_T)^2]$$

$$\varepsilon_0^2 = \min_{\vartheta} \mathbb{E}[(V_0 + \vartheta \cdot S_T - H)^2]$$

Proof. First, we complete the square for the normalized quadratic utility function $f_{-1}(x)$:

$$f_{-1}(x) = x - \frac{1}{2}x^2 = \frac{1}{2} - \frac{1}{2}(1 - x)^2$$

Using this, we can rewrite the indirect utility:

$$u_{-1}(p, q) = \max_{\vartheta \in \Theta_{-1}(p,q)} \max_{\eta \in \mathbb{R}} \eta + \mathbb{E}[f_{-1}(\vartheta \cdot S_T + q(H - p) - \eta)]$$

$$= \max_{\vartheta \in \Theta_{-1}(p,q)} \max_{\eta \in \mathbb{R}} \frac{1}{2} + \eta - \frac{1}{2}\mathbb{E}[(\vartheta \cdot S_T - 1 + q(H - p) - \eta)^2]$$

The term in expectations can be understood as a mean-variance hedging problem for a derivative claim with payoff $\tilde{H} = -q(H - p) + 1 + \eta$ and initial capital $\tilde{v}_0 = 0$. Thus we can write the result in terms of variables of mean-variance hedging of Černý and Kallsen [2007]:

$$u_{-1}(p, q) = \max_{\eta_\eta \in \mathbb{R}} \frac{1}{2} + \eta - \frac{1}{2}\left(L_0(\tilde{V}_0 - \tilde{v}_0)^2 + \varepsilon_0^2(\tilde{H})\right)$$

$$= \max_{\eta_\eta \in \mathbb{R}} \frac{1}{2} + \eta - \frac{1}{2}\left(L_0(-q(V_0 - p) + 1 + \eta)^2 + q^2\varepsilon_0^2(H)\right)$$

Optimizing over η, we find $\eta^* = L_0^{-1} + q(V_0 - p) - 1$. Plugging into the previous equation, we get our result. □

Corollary 2.2. *For mean-variance preferences, good-deal bounds have the form*

$$\hat{p}_{-1}^{\pm}(\Delta IP) = V_0 \pm \varepsilon_0(H)\sqrt{2\Delta IP}$$

Proof. We first find the utility-based quantity $\hat{q}_{-1}(p)$ by computing $\frac{\partial}{\partial q}u_{-1}(p,q) = 0$. We get

$$\hat{q}_{-1}(p) = \frac{V_0 - p}{\varepsilon_0^2(H)}$$

Then we can relate the added investment potential to the market price p of the derivative:

$$\Delta IP_{-1}(p, \hat{q}_{-1}(p)) = u_{-1}(p, \hat{q}_{-1}(p)) - u_{-1}(\cdot, 0)$$

$$= \hat{q}(p)(V_0 - p) - \frac{1}{2}\hat{q}_{-1}(p)^2\varepsilon_0^2(H) = \frac{1}{2}\frac{(V_0 - p)^2}{\varepsilon_0^2(H)}$$

Inverting this relationship, we get our desired result. □

2.2.2 Exponential utility

In a general semimartingale underlying, if we search for optimal η^* in the primal maximization problem (2.2) we find

$$\eta^* = -\log(\max_{\vartheta \in \Theta} \mathbb{E}[\exp(-\vartheta \cdot S_T - qH)]) - qp$$

$$u_\infty(p, q) = -qp - \log(\max_{\vartheta \in \Theta} \mathbb{E}[\exp(-\vartheta \cdot S_T - qH)])$$

Without additional assumptions on the underlying model, we cannot simplify this further. Using results from Delbaen et al. [2002], we have a connection to the dual solution of the problem over a space of martingale measures and we can express the utility function $u_\infty(p, q)$ in terms of minimal entropy. However, this is mainly a theoretical result, with limited opportunity for explicit numerical implementation.

The structure provided by Lévy processes later allows us to get concrete results for exponential utility in terms of quantities of the primal problem (2.2) as shown above. Specifically, for an appropriate payoff H, the term in expectations will correspond

to an exponential compensator, which is well-defined for most Lévy processes and straightforward to compute.

2.3 The Lévy contract: utility pricing under a Lévy model

In the first chapter we used a simple Black-Scholes model to evaluate discrete time hedging errors. However, as discussed earlier, a diffusion model is not sufficient to capture the full economic price of a variance swap (or higher order swaps). Therefore, in this chapter we will move to a more general framework to evaluate hedging errors and the corresponding price ranges, so as to accommodate for the jumps and risk from higher order moments that are observed in real markets.

From now on, we model the forward price $S = e^{-rt}\hat{S}$ via a one-dimensional exponential Lévy process

$$S = S_0 \exp X, \tag{2.4}$$

where X, the cumulative log-return, is a Lévy process with characteristics $(b(h), c, F)$ relative to some truncation function h (typically, $h(x) = x\mathbf{1}_{|x|<1}$). Without loss of generality, we set $X_0 = 0$. The corresponding rate of return will be denoted $\tilde{X}$, with characteristics $(\tilde{b}, \tilde{c}, \tilde{F})$. It relates to the underlying via the stochastic exponential, i.e. $S = S_0\mathcal{E}(\tilde{X})$ (for further details, see Kallsen and Shiryaev [2002]). The Lévy density $\tilde{F}(dx)$ satisfies the usual condition $\int_{\mathbb{R}}(1 \wedge x^2)\tilde{F}(dx) < \infty$. We automatically assume that the first four moments of the rate of return are finite, and label them $\tilde{\mu}, \tilde{\sigma}^2, \tilde{\mathrm{Sk}}, \tilde{\mathrm{EK}}$.

Under such a setting, we look at a general contract, which we will refer to as a "Lévy contract", with payoff

$$H = \alpha T + \beta\tilde{X}_T(h) + (W(x) - \beta h(x)) * J_T^{\tilde{X}} \tag{2.5}$$

where $\alpha, \beta \in \mathbb{R}$ are constants and $W(x)$ is some transformation function of the jumps of $\tilde{X}$. The asterisk operator $*$ indicates a double integral:

$$(f(t, x) * J^X)_t = \int_0^t \int_{\mathbb{R}} f(t, x)J^X(dt, dx)$$

76

The notation $\tilde{X}(h) := \tilde{X} - \tilde{X}_0 - (x - h(x)) * J^{\tilde{X}}$ denotes the compensated part of $\tilde{X}$. Our motivation for such a payoff is its ability to provide us with prices for all moment swaps. If we e.g. choose the function $W(x) = x^2$ and $\alpha = \tilde{c}, \beta = 0$, we get a claim whose payoff is quadratic variation, i.e. the fixed leg of a variance swap. Similarly we can get skewness swaps, kurtosis swaps, or the log contract $\log(S)$, the last of which is obtained by exploiting the relationship between X and $\tilde{X}$ (see Kallsen and Shiryaev [2002]) to fit it to the format of the payoff H:

$$X = \tilde{X} - \frac{1}{2}\tilde{c} + (\log(1+x) - x) * J^{\tilde{X}}.$$

However, for H to be a well-defined Lévy process, parameters $\beta, W(\cdot)$ will need to satisfy certain properties, which we discuss in the next lemma (the constant αT can be added without any consequence).

Lemma 2.3. *Let X be a Lévy process. For a given function $W(\cdot)$ which satisfies the conditions $\int_{\mathbb{R}}(1 \wedge W^2(x))F^X(dx) < \infty$ and $W(0) = 0$ it holds that there exists $\beta \in \mathbb{R}$ s.t.*

$$Y = \beta X(h) + (W(x) - \beta h(x)) * J^X$$

is a Lévy process. Furthermore, if X has infinite total variation $(TV(X) = \infty)$, this β is unique. If $TV(X) < \infty$, then $\beta = Y(0)/X(0)$.

Proof. First, we observe that if Y is well-defined, then its Lévy density is of the form

$$F^Y(G) = \int_{\mathbb{R}} 1_G(W(x))F^X(dx)$$

We know that for a Lévy density it must hold that

$$\int_{\mathbb{R}}(1 \wedge y^2)F^Y(dy) < \infty$$

But by the structure of the Lévy density of Y and our assumption on W,

$$\int_{\mathbb{R}}(1 \wedge y^2)F^Y(dy) = \int_{\mathbb{R}}(1 \wedge W^2(x))F^X(dx) < \infty$$

A Lévy process also requires that there is zero weight on jumps of size 0, which is satisfied, as $W(0) = 0$.

Now we look at the matter of uniqueness. Let us say there are two well-defined

Lévy processes Y, Y' with different coefficients β, β'. Since the difference of two Lévy processes is also necessarily a well-defined Lévy process, it must hold that

$$Y' - Y = (\beta' - \beta)X(h) + (\beta' - \beta)h(x) * J^X$$

is a Lévy process. However, we can see that this can only be a well-defined Lévy process if one of the following two holds: either $TV(X) < \infty$ (in which case $h(x) * J^X < \infty$, because Lévy densities of finite variation processes satisfy $\int_{\mathbb{R}}(1 \wedge |x|)F^X(dx) < \infty$) or it must be true that β is unique, i.e. $\beta' = \beta$.

If $TV(X) < \infty$, then Y can be decomposed into a continuous process $Y(0)$ and a sum of jumps $y * J^Y$:

$$Y = Y(0) + y * J^Y = Y(0) + W(x) * J^X$$

At the same time, from the definition of Y and choosing truncation function $h \equiv 0$,

$$Y = \beta X(0) + W(x) * J^X$$

Therefore $\beta = Y(0)/X(0)$, meaning that pathwise the continuous portion of our constructed process Y will always be a constant multiple of X. $\qquad\square$

2.3.1 The Lévy contract good-deal bounds: mean-variance preferences

As we saw in the general theory, for mean-variance preferences all we need to get good-deal bounds is to compute the mean-value process and quadratic hedging error of the contingent claim. In the case of the "Lévy contract" (2.5), they are given in the following theorem.

Theorem 2.4. *Under an exponential Lévy model, mean-variance ($\gamma = -1$) good-deal price bounds for the contract (2.5) are given as*

$$\hat{p}^{\pm}_{-1}(\Delta IP) = V_0 \pm \varepsilon_0 \sqrt{2\Delta IP}$$

where

$$V_0 = T(\alpha + \beta \left(\tilde{b}(h) + \mu^* \tilde{c} \right) + \int_{x>-1} ((1 + \mu^* x) W(x) - \beta h(x)) \tilde{F}(dx))$$

$$\varepsilon_0^2 = \frac{1 - \exp(-\frac{\tilde{\mu}^2}{\tilde{\sigma}^2} T)}{\tilde{\mu}^2/\tilde{\sigma}^2} \left(\int_{x>-1} W^2(x) \tilde{F}(dx) - 2\beta \int_{x>-1} xW(x) \tilde{F}(dx) - \frac{(\int_{x>-1} xW(x) \tilde{F}(dx))^2}{\tilde{\sigma}^2} \right)$$

$$\mu^* = \frac{-\tilde{\mu}}{\tilde{\sigma}^2},$$

the locally optimal hedging strategy ξ is

$$\xi_t = \frac{1}{S_{t-}} \frac{\beta \tilde{\sigma}^2 + \int_{x>-1} xW(x) \tilde{F}(dx)}{\tilde{\sigma}^2}$$

and the variance-optimal strategy φ is given by the recursive equation

$$\varphi_t = \xi_t - \frac{\tilde{\mu}}{\tilde{\sigma}^2 S_{t-}} (V_0 + \int_0^t \varphi_{u-} dS_u - V_{t-})$$

Proof. We know that under the variance optimal martingale measure (VOMM) $\hat{\mathbb{Q}}$ of mean-variance hedging, we obtain new characteristics for our driving process $\tilde{X}$ (see Kassberger and Liebmann [2011], Miyahara et al. [2007]):

$$\hat{b}(h) = \tilde{b}(h) + \mu^* c + \int_{\mathbb{R}} h(x) \mu^* x \tilde{F}(dx)$$

$$\hat{c} = \tilde{c}$$

$$\int_{\mathbb{R}} f(x) \hat{F}(dx) = \int_{\mathbb{R}} f(x)(1 + \mu^* x) \tilde{F}(dx)$$

$$\mu^* = \frac{-b(h) - \frac{1}{2}c - \int_{\mathbb{R}} ((e^x - 1) - h(x)) F(dx)}{c + \int_{\mathbb{R}} (e^x - 1)^2 F(dx)} = \frac{-\tilde{\mu}}{\tilde{\sigma}^2}$$

We can compute the mean-variance process directly:

$$\begin{aligned}
V_t &= \hat{\mathbb{E}}_t[H_T] = \alpha T + \beta \hat{\mathbb{E}}_t[\tilde{X}_T(h)] + \hat{\mathbb{E}}_t[(W(x) - \beta h(x)) * J^{\tilde{X}}] \\
&= \alpha T + \beta \tilde{X}_t(h) + (W(x) - \beta h(x)) * J_t^{\tilde{X}} \\
&\quad + \beta(T - t)\hat{b}(h) + (T - t) \int_{x>-1} (W(x) - \beta h(x))(1 + \mu^* x) \tilde{F}(dx) \\
&= \alpha T + \beta \tilde{X}_t(h) + (W(x) - \beta h(x)) * J_t^{\tilde{X}} \\
&\quad + \beta(T - t) \left(\tilde{b}(h) + \mu^* \tilde{c} \right) + (T - t) \int_{x>-1} ((1 + \mu^* x)W(x) - \beta h(x)) \tilde{F}(dx)
\end{aligned}$$

79

In our Lévy process setting with a deterministic opportunity set, the hedging error ε_0^2 is given via the quadratic variation of the mean-value process and underlying, and the mean-variance tradeoff process K (see Černý [2007], Černý and Kallsen [2009]):

$$\varepsilon_0^2 = (1 - \exp(-K_T))\frac{T}{K_T}\mathbb{E}[\langle V \rangle_T - \xi^2 \cdot \langle S \rangle_T]$$

Here ξ denotes the locally optimal strategy

$$\xi_t = \frac{d\langle V, S \rangle_t}{d\langle S \rangle_t}$$

Using the canonical martingale decomposition of S, it is straightforward to compute the required quadratic variations and co-variations:

$$
\begin{aligned}
\langle V \rangle_t &= \langle \beta \tilde{X}(h) + (W(x) - \beta h(x)) * J^{\tilde{X}} \rangle_t \\
&= t(\beta^2 \tilde{c} + \int_{x>-1} (\beta^2 x^2 + W^2(x))\tilde{F}(dx)) \\
&= t(\beta^2 \tilde{\sigma}^2 + \int_{x>-1} W^2(x)\tilde{F}(dx)) \\
\langle S \rangle_t &= S_{t-}^2 t(\tilde{c} + \int_{x>-1} x^2 \tilde{F}(dx)) = S_{t-}^2 t\tilde{\sigma}^2 \\
\langle V, S \rangle_t &= \langle \beta \tilde{X}(h) + (W(x) - \beta h(x)) * J^{\tilde{X}}, S \rangle_t = S_{t-}(\beta\tilde{c} + \int_{x>-1} \beta x^2 + xW(x))\tilde{F}(dx)
\end{aligned}
$$

This gives us the locally optimal hedging strategy. Proceeding to compute the second term in the hedging error, we get

$$
\begin{aligned}
\xi^2 \cdot \langle S \rangle_T &= \int_0^T \xi_t^2 d\langle S \rangle_t = \int_0^T \xi_t^2 S_{t-}^2 (\tilde{c} + \int_{x>-1} x^2 \tilde{F}(dx))\, dt \\
&= \frac{T}{\tilde{\sigma}^2}(\beta\tilde{c} + \int_{x>-1} (\beta x^2 + xW(x))\tilde{F}(dx))^2 = \frac{T(\beta\tilde{\sigma}^2 + \int_{x>-1} xW(x)\tilde{F}(dx))^2}{\tilde{\sigma}^2}
\end{aligned}
$$

The last ingredient we need is the mean-variance tradeoff process K_t, for which we know the explicit formula for Lévy models from Hubalek et al. [2006]:

$$K_t = \frac{\kappa^2(1)t}{\kappa(2) - 2\kappa(1)} = \frac{\tilde{\mu}^2 t}{\tilde{\sigma}^2}$$

where $\kappa(\cdot)$ is the cumulant-generating function. Joining all these results gives us the hedging error and locally optimal hedge. Finally, the form of the variance-optimal

hedging strategy follows from [Hubalek et al., 2006, Theorem 3.1]. $\qquad\qquad\square$

2.3.2 The Lévy contract: exponential utility

For exponential (and power) utility it is well-known that when the underlying is an exponential Lévy process, the optimal investment strategy is a constant-proportion strategy, i.e. $\vartheta^*_{t,\infty} = \zeta^*/S_{t-}$ (see e.g. Kallsen [2000], Kardaras [2009], Nutz [2012], Temme [2012] and the references within). This implies that also for hedging, our optimal strategy will be of the same form.

Lemma 2.5. *The optimal strategy ϑ^*_∞ is a constant-dollar strategy of the form $\vartheta^*_\infty = \zeta^*/S_-$, where $\zeta^* \in \mathbb{R}$ is a constant.*

Proof. Using results from Delbaen et al. [2002], we know the derivative contract H in the exponential utility defines a change of measure $d\mathbb{P}^H/d\mathbb{P}$, which allows us to convert the hedging problem to a problem of optimal investment without the derivative, under a different, non-physical measure. We can now use e.g. [Fujiwara, 2006, Theorem 4.1], which completes the proof. $\qquad\qquad\square$

Let us note that the strategy above is dependent on the variables of our utility maximization problem (2.2). Specifically, we will note explicitly as follows the dependence of ζ^* and ϑ^*_∞ on the quantity q bought of the contract: $\zeta^* = \zeta^*(q), \vartheta^*_\infty = \vartheta^*_\infty(q)$.

As a consequence of lemma 2.5, we can rewrite $\vartheta^*_\infty \cdot S_T$ as a multiple of the stochastic logarithm of S , written $\mathcal{L}(S)$, which is equal to the rate of return $\tilde{X}$:

$$\vartheta^*_\infty \cdot S_T = \int_0^T \vartheta^*_{\infty,t} dS_t = \int_0^T \frac{\zeta^*(q)}{S_{t-}} dS_t = \zeta^*(q)\mathcal{L}(S)_T = \zeta^*(q)\tilde{X}_T$$

We note that the stochastic logarithm $\mathcal{L}(S)$ is the inverse operation of the Doléans-Dade stochastic exponential, i.e. $\mathcal{E}(\mathcal{L}(S)) = S$.

The fact that the optimal hedging strategy is of constant proportion will allow us to rewrite the hedging portfolio as another Lévy process, as both the contract H and the continuously rebalanced hedge $\vartheta^*_\infty(q) \cdot S_T = \zeta^*(q)\tilde{X}_T$ are Lévy processes, their linear combination therefore also being a Lévy process. The utility maximization problem will reduce to computing an exponential compensator (see Kallsen and Shiryaev [2002] for a definition). The next lemma will provide us with a general result that al-

lows us to get the exponential compensator for our portfolio process $V_0 + \vartheta_\infty^*(q) \cdot S_T - H$.

Lemma 2.6. *Let X be a Lévy process with characteristics $(b^X(h), c^X, F^X)$ associated with truncation function $h(x)$. Define Lévy process Y:*

$$Y_t := \alpha t + \beta X_t(h) + (W(x) - \beta h(x)) * J^X$$

where $X(h) := X - X_0 - (x - h(x)) * J^X$, α and β are constants and $W(\cdot)$ is a function corresponding to those in lemma 2.3. Then if $\int_{|x|>1} e^{W(x)} F(dx) < \infty$ (i.e. Y is exponentially special), Y has an exponential compensator of the form

$$\kappa = \beta b^X(h) + \frac{1}{2}\beta^2 c^X + \alpha + \int_{\mathbb{R}} \left(e^{W(x)} - 1 - \beta h(x) \right) F^X(dx)$$

Proof. We are looking for the exponential compensator, i.e. a value κ such that:

$$\mathbb{E}\left[\exp\{Y_t - \kappa t\} \right] = 1,$$

in other words a value which would make the compensated exponential Lévy process a martingale. We compute κ by computing the drift of $Z_t = \exp(Y_t - \kappa t)$ and setting it to 0.

First we recall the canonical decomposition of $X(h)$ into a drift and a local martingale part $M(h)$:

$$X(h) = b^X(h)t + M^X(h) = b^X(h)t + X^c + h(x) * (J^X - \nu^X)$$

where ν^X compensates the jump measure J^X.

Using Itô's lemma for Lévy processes, we find:

$$\begin{aligned}
dZ &= Z_-(dY(h) - \kappa dt + \frac{1}{2}(dY)^2 + (e^y - 1 - h(y))dJ^Y) \\
&= Z_-(dY - (W(x) - h(W(x)))dJ^X - \kappa dt + \frac{1}{2}\beta^2 c^X dt + (e^{W(x)} - 1 - h(W(x)))dJ^X) \\
&= Z_-(\alpha dt + \beta dX(h) + (W(x) - \beta h(x))dJ^X - (W(x) - h(W(x)))dJ^X - \kappa dt \\
&\quad + \frac{1}{2}\beta^2 c^X dt + (e^{W(x)} - 1 - h(W(x)))dJ^X) \\
&= Z_-((\alpha + \beta b^X(h))dt + \beta dM^X(h) + \frac{1}{2}\beta^2 c^X dt + (e^{W(x)} - 1 - \beta h(x))dJ^X - \kappa dt)
\end{aligned}$$

Setting $b^{Z(h)} = 0$, we get that

$$\alpha + \beta b^X(h) + \frac{1}{2}\beta^2 c^X + \int_{\mathbb{R}} (e^{W(x)} - 1 - \beta h(x))F^X(dx) - \kappa = 0,$$

or equivalently

$$\kappa = \alpha + \beta b^X(h) + \frac{1}{2}\beta^2 c^X + \int_{\mathbb{R}} (e^{W(x)} - 1 - \beta h(x))F^X(dx)$$

The fact that Y is exponentially special ensures the finiteness of the Lévy integral. $\square$

We now have a general theorem to compute the exponential compensator of Lévy contract (2.5). Therefore, we have everything needed to compute optimal exponential utility and the related constant proportion hedging strategy $\zeta^*(q)$, which we do in the following theorem.

Theorem 2.7. *Under Lévy model* (2.4), *the indirect utility* $u_\infty(p, q)$ *for the Lévy contract* (2.5) *is given as:*

$$u_\infty(p, q) = T\left[q\alpha + (q\beta + \zeta^*(q))\tilde{b}(h) - \frac{1}{2}(q\beta + \zeta^*(q))^2\tilde{c} \right.$$
$$\left. - \int_{x>-1} (e^{-qW(x)-\zeta^*(q)x} - 1 + (q\beta + \zeta^*(q))h(x))\tilde{F}(dx)\right] - pq$$

where $\zeta^*(q)$ *is the solution to the equation*

$$0 = \tilde{b}(h) - (q\beta + \zeta^*(q))\tilde{c} + \int_{x>-1} (xe^{-qW(x)-\zeta^*(q)x} - h(x))\tilde{F}(dx) \qquad (2.6)$$

Proof. We recall that for normalized exponential utility $f_\infty(x) = 1 - \exp(-x)$, the indirect utility function for utility-based hedging is defined as

$$u_\infty(p, q) = \max_{\vartheta \in \Theta_\infty(p,q)} \max_{\eta \in \mathbb{R}} \eta + \mathbb{E}[1 - \exp(-\vartheta \cdot S_T - q(H_T - p) + \eta)]$$

We recall that the optimal strategy $\vartheta_\infty^*(q)$ has a constant proportion representation (recall Lemma 2.5) and that we can rewrite $\vartheta_\infty^*(q) \cdot S_T$ as a multiple of the rate of return $\tilde{X}$:

$$\vartheta_\infty^*(q) \cdot S_T = \zeta^*(q)\tilde{X}_T$$

Given our specific contract (2.5), the sum of the derivative and underlying hedging

strategy together form another Lévy process

$$Y_T := -\zeta^*(q)\tilde{X}_T - qH_T$$

In this light, we can interpret the computation as that of computing the exponential compensator of Y:

$$u_\infty(p,q) = \max_{\eta \in \mathbb{R}} \eta + 1 - \exp(\eta + pq)\mathbb{E}[\exp(Y_T)]$$

Using theorem 2.6, we find that the compensator has form

$$\kappa_Y = -q\alpha + (-q\beta - \zeta^*(q))\tilde{b}(h) + \frac{1}{2}(q\beta + \zeta^*(q))^2\tilde{c}$$
$$+ \int_{x>-1}(e^{-qW(x)-\zeta^*(q)x} - 1 + (q\beta + \zeta^*(q))h(x))\tilde{F}(dx) \qquad (2.7)$$

Thus,

$$u_\infty(p,q) = \max_{\eta \in \mathbb{R}} \eta + 1 - \exp(\eta + pq + \kappa_Y T)$$

Optimizing over η, we find $\eta^* = -\kappa_Y T - pq$ and thus

$$u_\infty(p,q) = -\kappa_Y T - pq.$$

Substituting for κ_Y we get our expression for $u_\infty(p,q)$.

To get the optimal hedging strategy, we simply minimize the compensator over the hedging proportion ζ, i.e. compute $\partial \kappa_Y / \partial \zeta = 0$. This gives us the equation for the optimal hedging strategy. □

We can now proceed to compute good-deal bounds on prices. For this, we need an optimal quantity $\hat{q}_\infty(p)$ and consequently the added investment potential $\Delta IP_\infty(p) = u_\infty(p, \hat{q}_\infty(p)) - u_\infty(\cdot, 0)$. A quick computation gives us that we can obtain optimal $\hat{q}_\infty(p)$ only implicitly from the equation

$$T(\alpha + \beta(\tilde{b}(h) - (q\beta + \zeta^*(q))\tilde{c}) + \int_{x>-1}(W(x)e^{-qW(x)-\zeta^*(q)x} - \beta h(x))\tilde{F}(dx)) - p = 0.$$

(We can see that inversely, the utility-based price can be computed directly). Moreover, this implicit equation is dependent on our optimal hedging strategy $\zeta^*(q)$, which is also only given by an implicit equation, computed in theorem 2.7. In both cases the target variable is within an integral over the Lévy density, making numerical so-

lutions sensitive and unstable. We therefore wish to find an approximate solution which would capture the main spirit of the hedging strategy and price bounds, but via an easily computable, closed-form formula. This would provide more insight into the properties of the price in relation to the structure of the Lévy process.

2.3.3 Exponential utility: asymptotic approximation for small jumps

As discussed in the previous section, we wish to find approximations to the exponential utility hedge and optimal quantity in such a way that would allow us to capture important features of the model but give tractable results. We will do so by considering asymptotically small amounts of jumps in our model. Specifically, to get closed-form approximations of exponential good-deal bounds, we will model the rate of return via a family of Lévy processes $\tilde{X}^\lambda$ with Lévy densities $\tilde{F}^\lambda$ and fixed mean and variance $\tilde{\mu}, \tilde{\sigma}^2$ such that the $\tilde{X}^\lambda$ converges to a Brownian diffusion with that given mean and variance as $\lambda \to 0$. Our motivation for considering such a sequence is to see how introducing small amounts of market incompleteness via a random jump measure and fat tails in our driving process alter our optimal hedging strategy and pricing rule. We will focus on the third and fourth moment of the distribution, i.e. the skewness and kurtosis, and consider moments of higher orders to be negligible.

One particular series with these properties can be obtained via a parametrization

$$\tilde{X}_t^\lambda = (1 - \frac{1}{\lambda})\tilde{\mu}t + \lambda\tilde{X}_{t/\lambda^2} \tag{2.8}$$

from Černý et al. [2013]. For this particular parametrization, we know that

$$\int_{x>-1} f(x)\tilde{F}^\lambda(dx) = \frac{1}{\lambda^2} \int_{x>-1} f(\lambda x)\tilde{F}(dx)$$

This leads to a particular scaling of the original skewness and kurtosis:

$$\tilde{\mathrm{Sk}}^\lambda = \lambda\tilde{\mathrm{Sk}}$$
$$\tilde{\mathrm{EK}}^\lambda = \lambda^2\tilde{\mathrm{EK}}$$

We now provide a theorem which gives conditions under which the Lévy contract H is well-defined in the limit $\lambda \to 0$ for the family of models (2.8).

Theorem 2.8. *Under the family of models (2.8), for the Lévy contract*

$$H^\lambda = \beta \tilde{X}_T^\lambda(h) + (W(x) - \beta h(x)) * J_T^{\tilde{X}^\lambda},$$

(with W satisfying conditions in theorem 2.3) to have a well-defined limit $H^0 = \lim_{\lambda \to 0+} H^\lambda$ it is sufficient if $W \in \mathcal{C}^1$ at $x = 0$ and $\beta = W'(0)$. Furthermore, if $W \in \mathcal{C}^1 \ \forall x \in \mathbb{R}$, then $\beta = W'(0)$ is also a necessary condition.

Proof. For the family of processes (2.8), the contract H takes the following form:

$$
\begin{aligned}
H^\lambda &= \beta \tilde{X}_T^\lambda(h) + (W(x) - \beta h(x)) * J_T^{\tilde{X}^\lambda} \\
&= \beta \left[\left(1 - \frac{1}{\lambda} \right) \tilde{\mu} T + \lambda \tilde{X}_{T/\lambda^2} - (x - h(x)) * J_T^{\tilde{X}^\lambda} \right] + (W(x) - \beta h(x)) * J_T^{\tilde{X}^\lambda} \\
&= \beta \left[\left(1 - \frac{1}{\lambda} \right) \tilde{\mu} T + \lambda \tilde{X}_{T/\lambda^2} \right] + (W(x) - \beta x) * J_T^{\tilde{X}^\lambda} \\
&= \beta \left[\left(1 - \frac{1}{\lambda} \right) \tilde{\mu} T + \lambda \tilde{X}_{T/\lambda^2} \right] + \frac{1}{\lambda^2} (W(\lambda x) - \beta \lambda x) * J_T^{\tilde{X}}
\end{aligned}
$$

Our goal is to investigate the limit $\lim_{\lambda \to 0} H^\lambda$. By [Černý et al., 2013, Lemma 2.6] we know the term in square brackets will converge to $\tilde{\mu} T + \tilde{\sigma} B_T$, where B is a standard Brownian motion. Therefore we need to investigate the limit

$$\lim_{\lambda \to 0} \frac{1}{\lambda^2} (W(\lambda x) - \beta \lambda x)$$

Using a change of variable $y = \lambda x$, we can also rewrite this as

$$x^2 \lim_{y \to 0} \frac{1}{y^2} (W(y) - \beta y)$$

Let us denote the limit

$$L = \lim_{y \to 0} \frac{1}{y^2} (W(y) - \beta y)$$

For H^0 to be well-defined, we need $L < \infty$ and the Lévy density over the limiting function must be finite:

$$\int_{x > -1} L x^2 F^{\tilde{X}}(dx) < \infty.$$

Due to the minimal integrability requirements of Lévy processes and the finite second moment, the density integral will be automatically finite as long as L is a finite-valued limit.

Assuming $W(y) \in \mathcal{C}^1$ on some small interval around 0 (and $W(0) = 0$), we can prove the sufficient and necessary condition for finiteness is $W'(0) = \beta$. Taylor expanding $W(y)$, we have:

$$W(y) = W(0) + W'(0)y + \mathcal{O}(y^2) = W'(0)y + \mathcal{O}(y^2)$$

Thus

$$\lim_{y \to 0} \frac{1}{y^2}(W(y) - \beta y) = \lim_{y \to 0} \frac{(W'(0) - \beta)y + \mathcal{O}(y^2)}{y^2}$$

Trivially $\lim_{y \to 0} \frac{\mathcal{O}(y^2)}{y^2} < \infty$. As a consequence, the limit will only be finite iff $\lim_{y \to 0} \frac{(W'(0) - \beta)}{y} < \infty$. The only way for this to be finite is iff $W'(0) = \beta$.

$\square$

Having established requirements on the form of the Lévy contract for the limit $\lambda \to 0$ to be well-defined, we can progress to getting asymptotic approximations of the indirect utility function $u_\infty(p, q)$. Let us recall that the main term of the indirect exponential utility in theorem 2.7 was the following exponential compensator, where we are interested in an approximation of the Lévy density integral:

$$\kappa_Y^\lambda = - q\alpha + (-q\beta - \zeta^*(q))\tilde{\mu} + \frac{1}{2}(q\beta + \zeta^*(q))^2 \tilde{c}^\lambda$$
$$+ \int_{x > -1} \left(e^{-qW(x) - \zeta^*(q)x} - 1 + (q\beta + \zeta^*(q))x \right) \tilde{F}^\lambda(dx) \qquad (2.9)$$

Transforming the Lévy integral to the original density $\tilde{F}$, that means

$$\int_{x > -1} \left(e^{-qW(x) - \zeta^*(q)x} - 1 + (q\beta + \zeta^*(q))x \right) \tilde{F}^\lambda(dx) =$$
$$\frac{1}{\lambda^2} \int_{x > -1} \left(e^{-qW(\lambda x) - \zeta^*(q)\lambda x} - 1 + (q\beta + \zeta^*(q))\lambda x \right) \tilde{F}(dx)$$

From this perspective, we see that a Taylor expansion of the exponential function is going to provide a good approximation, as λ is going to be small.

Therefore, let us proceed by Taylor-expanding the exponential in the density term (keeping notation in the $\tilde{F}^\lambda$ density). At this point, we could continue conducting the analysis for a general function $W(x)$, but this would lead to a need to do multiple Taylor expansions and would generate complicated expressions providing little new insight. Therefore, we choose a specific function $W(x)$ - a 4th order polynomial, as this allows us to capture moments of the returns distribution:

Assumption 2.1.
$$W(x) = c_1 x + c_2 x^2 + c_3 x^3 + c_4 x^4 \tag{2.10}$$

Such a function $W(x)$ allows us to straightforwardly get results for variance, skewness and kurtosis swaps (it does not allow for a log contract, however). Furthermore, since we will only be interested in the impact of the first four moments of the distribution on prices, we will make approximations of all our terms up to order $\mathcal{O}(\lambda^2)$ (as we saw previously, this is the λ order of kurtosis scaling for the parametrization (2.8)), which means we only consider polynomial terms up to order 4 to be significant (because $\int_{x>-1} x^n \tilde{F}^\lambda(dx) = \lambda^{n-2} \int_{x>-1} x^n \tilde{F}(dx)$). Thus our results also hold for any weight function with higher order polynomial terms, e.g. $W(x) + \mathcal{O}(x^5)$ (allowing us to reclaim log contract asymptotics).

For the Taylor expansion to be valid, or more specifically, for the remainder term in the Lévy density to be well-defined, we will make the following assumption on the moments.

Assumption 2.2. We require that the following moments of the Lévy density are finite:
$$\int_{x>-1} (qW(x) + \zeta^*(q)x)^4 \tilde{F}^\lambda(dx) < \infty$$

We are now ready to state the lemma giving us an approximation for the exponential compensator.

Lemma 2.9. *Under assumption 2.2 and for a weight function $W(x) + \mathcal{O}(x^5)$, where $W(x)$ has form (2.10), the exponential compensator (2.9) can be approximated via the first four moments of the returns of $\tilde{X}$ as*

$$
\begin{aligned}
\kappa_Y^\lambda ={}& -q\left(\alpha + c_2 \int x^2 \tilde{F}^\lambda(dx)\right) + (-qc_1 - \zeta)\tilde{\mu} \\
&+ \frac{1}{2}(qc_1 + \zeta)^2 \tilde{\sigma}^2 \\
&+ \tilde{Sk}^\lambda \tilde{\sigma}^3 \left(c_3 q + c_2 \zeta q + c_1 c_2 q^2 - \frac{1}{6}(\zeta + c_1 q)^3\right) \\
&+ \tilde{EK}^\lambda \tilde{\sigma}^4 \left(\frac{1}{2}c_2^2 q^2 + c_1 c_3 q^2 + c_3 \zeta q - c_4 q - \frac{1}{2}c_2\zeta^2 q - c_1 c_2 \zeta q^2 - \frac{1}{2}c_1^2 c_2 q^3 + \frac{1}{24}(qc_1 + \zeta)^4\right) \\
&+ \int_{x>-1} (R_5(x) + \mathcal{O}(x^5))\tilde{F}^\lambda(dx)
\end{aligned}
\tag{2.11}
$$

where

$$R_5(x) = -\frac{1}{24}(qW(x) + \zeta^*(q)x)^5 \int_0^1 (1-s)^4 \exp\{(-qW(x) - \zeta^*(q)x)s\}\, ds \quad (2.12)$$

Proof. Simply Taylor-expanding the exponential function of the exponential compensator κ_Y^λ for a general function $W(x)$, we obtain that the compensator can be approximated as

$$\begin{aligned}
\kappa_Y^\lambda = &- q\alpha + (-q\beta - \zeta^*(q))\tilde{\mu} + \frac{1}{2}(q\beta + \zeta^*(q))^2\tilde{c}^\lambda \\
&+ \int_{x>-1}\bigg(-qW(x) - \zeta^*(q)x + \frac{1}{2}(q^2W(x)^2 + 2q\zeta^*(q)W(x)x + \zeta^*(q)^2x^2) \\
&\quad - \frac{1}{6}(q^3W(x)^3 + 3q^2W(x)^2x\zeta^*(q) + 3qW(x)x^2\zeta^*(q)^2 + x^3\zeta^*(q)^3) \\
&\quad + \frac{1}{24}(q^4W(x)^4 + 4q^3W(x)^3x\zeta^*(q) + 6q^2W(x)^2x^2\zeta^*(q)^2 + 4qW(x)x^3\zeta^*(q)^3 + x^4\zeta^*(q)^4) \\
&\quad + (q\beta + \zeta^*(q))x + R_5(x) \bigg)\tilde{F}^\lambda(dx) \\
= &- q\alpha + (-q\beta - \zeta^*(q))\tilde{\mu} + \frac{1}{2}(q\beta + \zeta^*(q))^2\tilde{c}^\lambda \\
&+ \int_{x>-1}\bigg(-q(W(x) - \beta x) + \frac{1}{2}(qW(x) + \zeta^*(q)x)^2 - \frac{1}{6}(qW(x) + x\zeta^*(q))^3 \\
&\quad + \frac{1}{24}(qW(x) + x\zeta^*(q))^4 + R_5(x) \bigg)\tilde{F}^\lambda(dx)
\end{aligned}$$

where

$$R_5(x) = -\frac{1}{24}(qW(x) + \zeta^*(q)x)^5 \int_0^1 (1-s)^4 \exp\{(-qW(x) - \zeta^*(q)x)s\}\, ds \quad (2.13)$$

is the Taylor expansion remainder term (see Abramowitz and Stegun [1972, eqn 3.6.3]).

We notice that if we now plug in (2.10) for $W(x)$ and use theorem 2.8, it is necessary that $\beta = W'(0) = c_1$ for $\lim_{\lambda \to 0} H^\lambda$ to be well-defined. Therefore if we Taylor-expand $W(x)$ around 0, we eliminate the outstanding term of order $\mathcal{O}(x)$ in the integral and are left with terms of at least order $\mathcal{O}(x^2)$.

Using our weighting function (2.10), we can multiply terms out and simplify the

approximate compensator κ_Y^λ to get:

$$
\begin{aligned}
\kappa_Y^\lambda = {}& -q\alpha + (-qc_1 - \zeta)\tilde{\mu} \\
& + \frac{1}{2}(qc_1 + \zeta)^2 \tilde{c}^\lambda + \int x^2 \tilde{F}^\lambda(dx)(\frac{1}{2}(\zeta + c_1 q)^2 - c_2 q) \\
& + \int x^3 \tilde{F}^\lambda(dx)(c_3 q + c_2 \zeta q + c_1 c_2 q^2 - \frac{1}{6}(\zeta + c_1 q)^3) \\
& + \int x^4 \tilde{F}^\lambda(dx)(\frac{1}{2}c_2^2 q^2 + c_1 c_3 q^2 + c_3 \zeta q - c_4 q - \frac{1}{2}c_2 \zeta^2 q - c_1 c_2 \zeta q^2 - \frac{1}{2}c_1^2 c_2 q^3 + \frac{1}{24}(qc_1 + \zeta)^4) \\
& + \int_{x>-1} (R_5(x) + \mathcal{O}(x^5))\tilde{F}^\lambda(dx)
\end{aligned}
$$

We can express the Lévy integrals via the moments of returns and rearrange:

$$
\begin{aligned}
\kappa_Y^\lambda = {}& -q(\alpha + c_2 \int x^2 \tilde{F}^\lambda(dx)) + (-qc_1 - \zeta)\tilde{\mu} \\
& + \frac{1}{2}(qc_1 + \zeta)^2 \tilde{\sigma}^2 \\
& + \tilde{\mathrm{Sk}}^\lambda \tilde{\sigma}^3 (c_3 q + c_2 \zeta q + c_1 c_2 q^2 - \frac{1}{6}(\zeta + c_1 q)^3) && (2.14) \\
& + \tilde{\mathrm{EK}}^\lambda \tilde{\sigma}^4 (\frac{1}{2}c_2^2 q^2 + c_1 c_3 q^2 + c_3 \zeta q - c_4 q - \frac{1}{2}c_2 \zeta^2 q - c_1 c_2 \zeta q^2 - \frac{1}{2}c_1^2 c_2 q^3 + \frac{1}{24}(qc_1 + \zeta)^4) \\
& + \int_{x>-1} (R_5(x) + \mathcal{O}(x^5))\tilde{F}^\lambda(dx)
\end{aligned}
$$

Finally, we are interested in the conditions which must hold for the remainder to be well-defined. For this, we can compute the integral in equation (2.13) for $R_5(x)$ explicitly. Repeatedly using by parts integration, we find

$$
\int_0^1 (1-s)^4 e^{As}\, ds = -\frac{1}{A^5}(A^4 + 4A^3 + 12A^2 + 24A - 24e^A + 24)
$$

where in our case, $A = -qW(x) - \zeta^*(q)x$. Thus, we get a full closed-form expression for the remainder:

$$
\begin{aligned}
R_5(x) = {}& -\frac{1}{24}(qW(x) + \zeta^*(q)x)^4 + \frac{1}{6}(qW(x) + \zeta^*(q)x)^3 - \frac{1}{2}(qW(x) + \zeta^*(q)x)^2 \\
& + (qW(x) + \zeta^*(q)x) + e^{-qW(x) - \zeta^*(q)x} - 1
\end{aligned}
$$

This makes it easy to see the conditions we need to set on our distribution for every integral to be finite. We have already assumed the exponential $\int_{x>-1} e^{-qW(x)-\zeta^*(q)x} \tilde{F}(dx)$ is finite, since this was a condition required for the full exponential compensator

to exist. The other condition is on the existence of co-moments $\int_{x>-1}(qW(x) + \zeta^*(q)x)^4\tilde{F}(dx)$, for which we have made assumption 2.2. $\quad\square$

2.3.4 The approximate hedging strategy

Having the approximation of the exponential compensator of the full exponential utility $u_\infty(p,q))$, allows us to compute the approximate hedging strategy for small jumps.

Theorem 2.10. *The full hedging strategy of theorem 2.7 for the family of Lévy processes (2.8) can be expressed as a polynomial in λ in the form*

$$\zeta^\lambda = a_0 + a_1\lambda + a_2\lambda^2 + o(\lambda^2) \tag{2.15}$$

where

$$a_0 = \frac{\tilde{\mu}}{\tilde{\sigma}^2} - qc_1$$

$$a_1 = \tilde{\sigma}\tilde{S}k(\frac{1}{2}\frac{\tilde{\mu}^2}{\tilde{\sigma}^4} - c_2q)$$

$$a_2 = c_2q\tilde{\mu}(\tilde{E}K - \tilde{S}k^2) + \frac{\tilde{\mu}^3}{\tilde{\sigma}^4}(\frac{1}{2}\tilde{S}k^2 - \frac{1}{6}\tilde{E}K) - c_3\tilde{\sigma}^2q\tilde{E}K$$

Proof. The approximate optimal hedging strategy $\zeta^\lambda(q)$ is found by minimizing the exponential compensator, i.e. solving

$$\frac{\partial}{\partial\zeta}\kappa_Y^\lambda = 0$$

Writing out the equation in full using the approximate form of the compensator (2.14), we get:

$$0 = -\tilde{\mu} + \tilde{\sigma}^2(qc_1 + \zeta)$$
$$\tilde{S}k^\lambda\tilde{\sigma}^3(c_2q - \frac{1}{2}(\zeta + c_1q)^2) \tag{2.16}$$
$$\tilde{E}K^\lambda\tilde{\sigma}^4(c_3q - c_2\zeta q - c_1c_2q^2 + \frac{1}{6}(qc_1 + \zeta)^3) + \int_{x>-1}\frac{\partial}{\partial\zeta}R_5(x) + \mathcal{O}(x^5)\tilde{F}^\lambda(dx)$$

The above can generally be seen as an equation of form $G(\zeta,\lambda) = 0$. Now we look for the approximate optimal hedging strategy ζ^λ. Using the implicit function method on

equation (2.16), we look for a hedging strategy of the form

$$\zeta^{\lambda} = f(\lambda) = a_0 + a_1\lambda + a_2\lambda^2 + o(\lambda^2)$$

If we ignore the remainder term $R_5(x)$ (we will show below that we can do this) and solve

$$G(\zeta^0, 0) = 0$$
$$\frac{\partial}{\partial\lambda}G(\zeta^0, 0) = 0$$
$$\frac{\partial^2}{\partial\lambda^2}G(\zeta^0, 0) = 0$$

we identify coefficients a_0, a_1, a_2 to be

$$a_0 = \frac{\tilde{\mu}}{\tilde{\sigma}^2} - qc_1$$

$$a_1 = \tilde{\sigma}\tilde{S}k(\frac{1}{2}\frac{\tilde{\mu}^2}{\tilde{\sigma}^4} - c_2 q)$$

$$a_2 = c_2 q\tilde{\mu}(\text{E}\tilde{\text{K}} - \tilde{S}k^2) + \frac{\tilde{\mu}^3}{\tilde{\sigma}^4}(\frac{1}{2}\tilde{S}k^2 - \frac{1}{6}\text{E}\tilde{\text{K}}) - c_3\tilde{\sigma}^2 q\text{E}\tilde{\text{K}}$$

We now show why the remainder term $\rho_5(\zeta, \lambda) := \int_{x>-1}(\frac{\partial}{\partial\zeta}R_5(x) + \mathcal{O}(x^5))\tilde{F}^{\lambda}(dx)$ from equation (2.16) does not influence our approximate hedging strategy. We can differentiate $R_5(x)$ wrt to ζ for a general $W(x)$:

$$\frac{\partial}{\partial\zeta}R_5(x) = -\frac{5}{24}(qW(x) + \zeta x)^4 x \int_0^1 (1-s)^4 e^{(-qW(x)-\zeta x)s}\,ds$$
$$+ \frac{1}{24}(qW(x) + \zeta x)^5 x \int_0^1 (1-s)^4 s e^{(-qW(x)-\zeta x)s}\,ds$$

If we use $W(x)$ from 2.10, we can compute the Lévy density integral over this remainder term:

$$\rho_5(\zeta^*(q), \lambda) = \int_{x>-1} \frac{\partial}{\partial\zeta^*(q)}R_5(x)\tilde{F}^{\lambda}(dx)$$
$$= \int_{x>-1} \left(-\frac{5}{24}\lambda^3 x^5 (q\sum_{i=1}^{4} c_i\lambda^{i-1}x^{i-1} + \zeta^*(q))^4 \int_0^1 (1-s)^4 e^{(-qW(\lambda x)-\zeta^*(q)\lambda x)s}\,ds \right.$$
$$\left. + \frac{1}{24}\lambda^4 x^6 (q\sum_{i=1}^{4} c_i\lambda^{i-1}x^{i-1} + \zeta^*(q))^5 \int_0^1 (1-s)^4 s e^{(-qW(\lambda x)-\zeta^*(q)\lambda x)s}\,ds \right)$$

92

Since the leading λ terms are at least of order λ^3, it follows that:

$$\rho_5(\zeta^*(q), \lambda)_{|\lambda=0} = 0$$
$$\frac{\partial}{\partial\lambda}\rho_5(\lambda, \zeta^*(q))_{|\lambda=0} = 0$$
$$\frac{\partial^2}{\partial\lambda^2}\rho_5(\lambda, \zeta^*(q))_{|\lambda=0} = 0$$

which means the $\rho_5(\zeta, \lambda)$ term does not enter into the computation of coefficients a_0, a_1, a_2 of the polynomial hedge approximation. The same logic applies if we use a function $W(x)$ which has an additional $\mathcal{O}(x^5)$ term above the polynomial (2.10), as any higher-order term would be subsumed into the $\mathcal{O}(x^5)$ part of the remainder term $\rho_5(\zeta, \lambda)$. □

2.3.5 Asymptotic exponential good-deal bounds for small jumps

Now that we have an approximation to the exponential compensator (and hence the indirect utility) and we have an approximation for the hedging strategy (i.e. it is no longer given by an implicit equation), we can get our desired final result, an approximation to the full exponential utility good-deal bounds under the assumption that our model jumps are small.

Theorem 2.11. *For $\lambda \to 0$ in the family of processes (2.8), the exponential good-deal bounds of the Lévy contract (2.5) with weighting function $W(x) + \mathcal{O}(x^5)$ (where $W(x)$ is given by (2.10)) are*

$$p_\infty^{\lambda,\pm}(\Delta IP) = N_\infty^\lambda \pm \sqrt{2D_\infty^\lambda \Delta IP} + \mathcal{O}(\lambda^3)$$

where

$$N_\infty^\lambda = T(\alpha + c_2 \int_{x>-1} x^2 \tilde{F}^\lambda(dx) - \tilde{S}k^\lambda(c_2\tilde{\mu}\tilde{\sigma} - c_3\tilde{\sigma}^3)$$
$$+ \frac{1}{2}c_2\tilde{\mu}^2(\tilde{E}K^\lambda - (\tilde{S}k^\lambda)^2) - \tilde{E}K^\lambda(c_3\tilde{\mu}\tilde{\sigma}^2 - c_4\tilde{\sigma}^4))$$
$$D_\infty^\lambda = Tc_2^2\tilde{\sigma}^4(\tilde{E}K^\lambda - (\tilde{S}k^\lambda)^2),$$

Proof. Plugging in the approximately optimal hedging strategy $\zeta^\lambda(q)$ from (2.15) into the approximation of the exponential compensator (2.14) we find that

$$
\begin{aligned}
\kappa_Y^\lambda(q) = &-\frac{1}{2}\frac{\tilde{\mu}^2}{\tilde{\sigma}^2} - \frac{1}{6}\tilde{S}k^\lambda\frac{\tilde{\mu}^3}{\tilde{\sigma}^3} + \frac{\tilde{\mu}^4}{\tilde{\sigma}^4}\left(-\frac{1}{8}(\tilde{S}k^\lambda)^2 + \frac{1}{24}\tilde{E}K^\lambda\right) \\
&+ q\bigg(-\alpha - c_2\int_{x>-1}x^2\tilde{F}^\lambda(dx) + \tilde{S}k^\lambda(c_2\tilde{\mu}\tilde{\sigma} - c_3\tilde{\sigma}^3) + \frac{1}{2}c_2\tilde{\mu}^2((\tilde{S}k^\lambda)^2 - \tilde{E}K) \\
&\qquad + \tilde{E}K(c_3\tilde{\mu}\tilde{\sigma}^2 - c_4\tilde{\sigma}^4)\bigg) \\
&+ q^2\frac{1}{2}c_2^2\tilde{\sigma}^4(\tilde{E}K^\lambda - (\tilde{S}k^\lambda)^2) + \mathcal{O}(\lambda^3)
\end{aligned}
$$

From there we can compute the approximate indirect utility $u_\infty^\lambda(p,q) = -T\kappa_Y^\lambda(q) - pq$ and equally the change in investment potential $\Delta IP_\infty^\lambda(p) = u_\infty^\lambda(p, \hat{q}_\infty^\lambda(p)) - u_\infty^\lambda(\cdot, 0) = -T(\kappa_Y^\lambda(\hat{q}_\infty^\lambda(p)) - \kappa_Y^\lambda(0)) - p\hat{q}_\infty^\lambda(p)$. To get the optimal quantity, we solve $\frac{\partial}{\partial q}\Delta IP_\infty^\lambda = 0$ to find:

$$
\hat{q}_\infty^\lambda(p) = \frac{N_\infty^\lambda - p + \mathcal{O}(\lambda^3)}{D_\infty^\lambda},
$$

where

$$
\begin{aligned}
N_\infty^\lambda := &\ T\Big(\alpha + c_2\int_{x>-1}x^2\tilde{F}^\lambda(dx) - \tilde{S}k^\lambda(c_2\tilde{\mu}\tilde{\sigma} - c_3\tilde{\sigma}^3) \\
&+ \frac{1}{2}c_2\tilde{\mu}^2(\tilde{E}K^\lambda - (\tilde{S}k^\lambda)^2) - \tilde{E}K^\lambda(c_3\tilde{\mu}\tilde{\sigma}^2 - c_4\tilde{\sigma}^4)\Big) \\
D_\infty^\lambda := &\ Tc_2^2\tilde{\sigma}^4(\tilde{E}K^\lambda - (\tilde{S}k^\lambda)^2),
\end{aligned}
$$

Plugging the optimal quantity into the indirect utility leads to the investment potential, which is then

$$
\Delta IP_\infty^\lambda(p) = \frac{1}{2}\frac{(N_\infty^\lambda - p + \mathcal{O}(\lambda^3))^2}{D_\infty^\lambda}
$$

Inverting this relationship, we get approximate good-deal bounds up to order λ^2 for the Lévy contract (2.5):

$$
p_\infty^{\lambda,\pm}(\Delta IP) = N_\infty^\lambda \pm \sqrt{2D_\infty^\lambda\Delta IP} + \mathcal{O}(\lambda^3)
$$

$\square$

2.4 Application 1: good-deal bounds of a variance swap

We can use our theoretical results on the Lévy contract to investigate the good-deal bounds of variance swaps. We achieve this by picking the contract variables

$$\alpha = \tilde{c}, \qquad \beta = 0, \qquad W(x) = x^2$$

making the payoff $H = T(\tilde{c} + \int_{\mathbb{R}} x^2 \tilde{F}(dx)) = T\sigma^2 = \int_0^T \sigma^2 dt$.

For mean-variance hedging, by way of results from theorem 2.4 we get the following good-deal bounds:

$$\hat{p}^{\pm}_{-1}(\Delta IP) = V_0 \pm \sqrt{2\varepsilon_0 \Delta IP}$$
$$V_0 = \tilde{\sigma}^2 T - \tilde{\mu}\tilde{\sigma}\tilde{S}kT$$
$$\varepsilon_0^2 = \frac{1 - \exp(-\frac{\tilde{\mu}^2}{\tilde{\sigma}^2}T)}{\tilde{\mu}^2/\tilde{\sigma}^2}\tilde{\sigma}^4(\tilde{\mathrm{EK}} - \tilde{Sk}^2)$$

and a locally optimal hedging strategy

$$\xi_t = \frac{\tilde{\sigma}\tilde{S}k}{S_{t-}}$$

For comparison, the asymptotic exponential good-deal bounds (for small λ) are

$$\hat{p}^{\lambda,\pm}_{\infty}(\Delta IP) = N^{\lambda}_{\infty} \pm \sqrt{2D^{\lambda}_{\infty}\Delta IP} + \mathcal{O}(\lambda^3)$$
$$N^{\lambda}_{\infty} = T(\sigma^2 - \tilde{\mu}\tilde{\sigma}\tilde{S}k^{\lambda} + \frac{1}{2}\tilde{\mu}^2(\tilde{\mathrm{EK}}^{\lambda} - (\tilde{Sk}^{\lambda})^2))$$
$$D^{\lambda}_{\infty} = T\tilde{\sigma}^4(\tilde{\mathrm{EK}}^{\lambda} - (\tilde{Sk}^{\lambda})^2)$$

The approximate pure unit hedge for exponential utility (i.e. the exponential equivalent of the locally optimal hedge for mean-variance preferences), is

$$\frac{1}{S_{t-}}\frac{\zeta^{\lambda}(q) - \zeta^{\lambda}(0)}{q} = \frac{1}{S_{t-}}(\tilde{\sigma}\tilde{S}k^{\lambda} - \tilde{\mu}(\tilde{\mathrm{EK}}^{\lambda} - (\tilde{Sk}^{\lambda})^2) + o(\lambda^2))$$

We can see that in both cases, up to order $\mathcal{O}(\lambda^2)$, the price of the variance swap is adjusted for the skewness of the returns process, and the size of the adjustment depends on the mean return. The exponential utility case has an additional adjust-

ment for the kurtosis of the returns. We interpret this difference to be due to the non-symmetric shape of exponential utility, which, unlike mean-variance preferences, does not have a cut-off point on the investment opportunities it sees.

We can further see that the width of the exponential good-deal bounds is very similar to the mean-variance case - if we Taylor-expand the exponential in the mean-variance case, we will find that up to first order the exponential and mean-variance bounds are identical. On the other hand, there is a minor kurtosis adjustment in the approximate price of the derivative for exponential utility. The difference $\tilde{\text{EK}}^\lambda - (\tilde{Sk}^\lambda)^2$ serves here as an imperfect measure of deviation from normality. Some unpublished results (Harremoës [2000]) also indicate that this difference is related to minimal entropy under restrictions on the first two moments of a distribution.

Comparing the approximate exponential hedging strategy with the mean-variance result, we see again that to the order $\mathcal{O}(\lambda)$ they are identical. We see again a correction based on the difference between kurtosis and squared skewness at the order $\mathcal{O}(\lambda^2)$.

We now look numerically at the difference between the good-deal bounds obtained via mean-variance preferences, full exponential and asymptotic exponential utility. In figure 2.1 we provide a comparison of for the case of an NIG model whose parameters are set to fit the first four moments of the returns distribution (in annual terms), with $\mu = 0.1, \sigma = 0.3, \text{Sk} = -0.5, \text{EK} = 0.7$, with a time horizon of 1 year, $T = 1$.

For the full exponential solution, to achieve numerical stability of the solution to the nonlinear equation (2.6) (solved via MATLAB's `fsolve` function), we truncate the jump sizes of the Lévy density, eliminating rates of return above 100%.

For the asymptotic solution, we consider a solution where $\lambda = 1$. Since λ is scaled in Černý et al. [2013] to the range $[0, 1]$ only for pure convenience and the family of processes X^λ could equally well be reparameterized to be within any other range of values, the parameter's value only gives us an indication of how close we are to Brownian motion and how far away we are from the original Lévy process through the value of the re-scaled skewness and kurtosis. If we find that for our chosen λ, the full and asymptotic price bounds are only a few basis points away from each other at a particular level of ΔIP, it means our exposure to jumps via the derivative is small enough for us to handle them in an asymptotic sense. In other words, a world where returns are generated by $\tilde{X}^\lambda$ is a world with small jumps from the perspective of the hedger of the derivative contract.
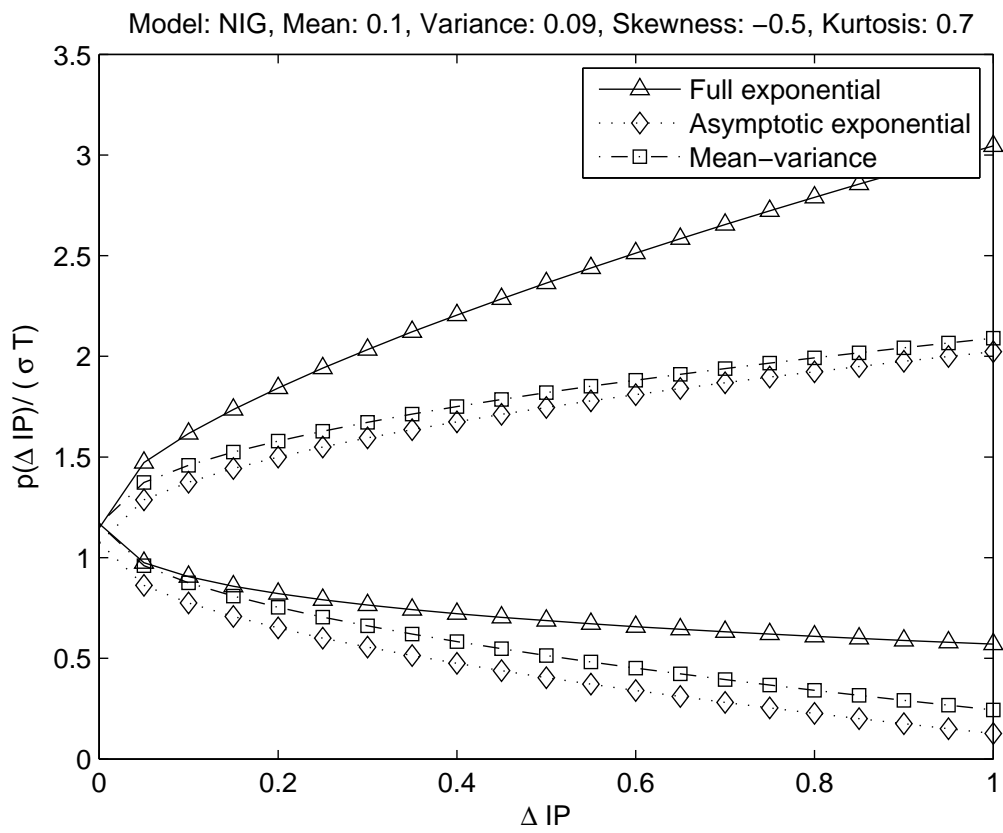
Figure 2.1: Comparison of mean-variance utility, exponential utility and asymptotic utility good-deal bounds relative to the unique complete market price.

From figure 2.1 we can see that we have to misprice the swap quite significantly to introduce interesting trading opportunities into the market for our competitors. Obviously, for large exposures and extreme market opportunities, the asymptotics do not do a very good job of capturing the properties of the full exponential solution. However, the main question we are trying to answer is the following: how much can we misprice without offering other traders a lucrative opportunity? Let us use the connection between ΔIP and the Sharpe ratio for the case of mean-variance preferences ($\gamma = -1$), where $\Delta IP_{-1} = \frac{1}{2} SR_{SH}^2$, as a gauge of lucrativeness. Under the assumption that traders will not consider it worth the effort to potentially improve their annualized Sharpe ratio by less than 0.2, we're really only interested in price differences up to $\Delta IP < 0.02$. The difference between the full exponential solution and the approximation on that interval is of the order 10-15% relative to the Black-Scholes price, where most of the difference is caused by the shifted mean of the asymptotic approximation. This difference is large enough to indicate that we are probably not in a world with completely small jumps. On the other hand, the mean-variance solution only differs from the full exponential utility solution by at most 5%, showing it is a worthy substitute. In terms of the deviations of the good-deal bounds from their respective central case prices (i.e. $p_\gamma(\Delta IP)_{\Delta IP=0}$), all three solutions allow a deviation of at most 12%, which we consider to be tight bounds relative to previous results in the literature.

2.5 Application 2: good-deal bounds of a skewness swap

Similarly to a variance swap, we can get possible price ranges of skewness swaps by picking contract variables

$$\alpha = 0, \qquad \beta = 0, \qquad W(x) = \frac{x^3}{\tilde{\sigma}^3}$$

making the payoff $H = T\frac{\int_{\mathbb{R}} x^3 \tilde{F}(dx)}{\tilde{\sigma}^3} = T\tilde{Sk}$. This ties into previous research on higher-order moment swaps (Schoutens [2005]) and investigations into skewness risk premia (Neuberger [2012], Kozhan et al. [2013] and references therein).

For mean-variance preferences, by way of results from theorem 2.4 we get

$$V_0 = T(\tilde{S}k - \frac{\tilde{\mu}}{\tilde{\sigma}}\tilde{E}K)$$

$$\varepsilon_0^2 = \frac{1 - \exp(-\frac{\tilde{\mu}^2}{\tilde{\sigma}^2}T)}{\tilde{\mu}^2/\tilde{\sigma}^2} \left(\frac{\int_{x>-1} x^6 \tilde{F}(dx)}{\tilde{\sigma}^6} - \tilde{E}K^2 \right)$$

$$\xi_t = \frac{\tilde{E}K}{\tilde{\sigma} S_{t-}}$$

For exponential utility, the approximation up to order $\mathcal{O}(\lambda^2)$ is insufficient to get non-trivial good-deal bounds on the contract, because $D_\infty^\lambda = 0$. As a consequence, ignoring moments of higher order than 4, the optimal $\hat{q}_\infty^\lambda(p)$ automatically strays to $\pm\infty$ in case of any deviation from the unique price for the contract,

$$\hat{p}^\lambda(\Delta IP) = T \left(\tilde{S}k^\lambda - \tilde{E}K^\lambda \frac{\tilde{\mu}}{\tilde{\sigma}} \right)$$

We can notice that at this level of approximation, the mean-value process coincides with the exponential utility price. Furthermore, the approximate hedging strategy also coincides:

$$\frac{1}{S_{t-}} \frac{\zeta^\lambda(q) - \zeta^\lambda(0)}{q} = \frac{\tilde{E}K}{\tilde{\sigma} S_{t-}}$$

To get separate upper and lower good-deal bounds on a skewness swap in the asymptotic exponential sense, we would need to look at higher order terms. Specifically, looking at the mean-variance hedging error, we hypothesize that we would need to go up to the 6th moment (i.e. order $\mathcal{O}(\lambda^4)$) of the distribution to introduce uncertainty on the approximate utility-based price.

2.6   Conclusion

In this chapter, we set out to look at the factors and risks that can further influence the price of log contracts and variance swaps. Specifically we focused on the impact of jumps on price uncertainty, as the literature suggests this is an important factor to consider. We applied utility-based pricing methodology to get good-deal bounds on a general family of contracts, which we labelled the Lévy contract, that encompass the log contract, variance swaps and higher order moment swaps. We were able to compute them for exponential utility and mean-variance preferences. For the former,

we also found a good closed-form approximation via the first four moments of the distribution, which is applicable when the jumps we observe are not too large. We derived explicit formulas for the situation when the Lévy contract has a specific, polynomial structure to its jump transformation. Finally, we showed on the example of a variance swap how good-deal bounds provide more granular information about price uncertainty than standard no-arbitrage bounds. Setting a limit to the investment opportunity we're willing to inject into the market (i.e. an opportunity other traders would react to), we obtain tight bounds on the price of the contract.

Looking critically at the results, however, we see that they are highly dependent on our knowledge of the moments of the returns distribution. That is why in the next chapter, we will try to allow for further uncertainty in this respect, by assuming the returns are driven by a Hidden Markov Model (HMM), where we consider several candidate distributions, but we are uncertain which of them is the "correct" one. To decide that, we will attempt to let the data speak for itself by filtering out the most probable distribution and adjusting our trading strategy based on this estimate.

Chapter 3

Mean-variance hedging for regime-switching Lévy processes

3.1 Motivation and literature review

In the previous chapter we derived, among other things, the mean-value process and hedging error for a variance swap when pricing in an incomplete market setting. We saw that it was completely determined by the first four moments of the underlying returns distribution. Our price estimate will therefore be heavily reliant on us having specified the "correct" distribution of returns, i.e. having chosen the right model. In this chapter we will consider how incorporating uncertainty on these parameters impacts pricing and hedging of contingent claims. There are currently multiple ways of achieving this, all with their strengths and weaknesses.

We could take a highly risk-averse approach and acknowledge the uncertainty in our use of the physical distribution completely by considering all the possible physical measures that are attainable via distortion of the physical measure without adding too many new assumptions as measured by entropy. This is an approach pioneered in many papers by the authors of Hansen and Sargent [2008] and later used for obtaining robust no-good-deal bounds in a discrete time setting in Boyarchenko et al. [2014]. The downside to the robustness is that this leads to rather involved computations even for a simple binomial tree model.

A more tractable approach is for us to define stochastic dynamics of our uncertain

parameters, thus exogenously imposing structure on how our parameter uncertainty develops over time. Under well-chosen dynamics, this can lead to more straightforward and explicit calculations. An example of this approach is the broad class of stochastic volatility models, such as the celebrated Heston model (Heston [1993]). These incomplete market models acknowledge that the volatility of returns is an unobservable and time-varying quantity, and impose a particular dynamic on its behaviour. The enforced dynamics allow us to then compute prices explicitly. However, it may be unclear how to choose the most appropriate model for volatility, as apart from the Heston model we can choose the CEV, SABR, GARCH, 3/2 model (or define our own brand new model), each of them leading to slightly different behaviour of our volatility and prices. Furthermore, the majority of these models focus solely on the uncertainty of the volatility, and consider all other moments of our returns distribution to be deterministic. In a similar vein, we could use the uncertain volatility model of Avellaneda et al. [1995], where volatility σ does not have any dynamics but is simply assumed to lie somewhere within a range $[\sigma_{min}, \sigma_{max}]$ (based on e.g. historical observations), and all other parameters are considered to be known.

An alternative approach is to limit our uncertainty to a finite set of possible returns distributions corresponding to various market states (e.g. highly volatile bear market, slowly rising bull market etc.). We then stipulate a dynamic according to which the market switches between these different states. This is referred to as a regime-switching (RS) model. When calibrated to historical data, it allows us to simultaneously consider empirical historical distributions in several distinct historical periods. As a consequence, it also allows us to vary all the moments of the distribution, not just volatility. This can add realism particularly to the pricing of long-dated claims (e.g. 20-30 year options), which may have to endure several different market states until expiry.

Most commonly, the dynamics of the market state are described by a Markov process - due to this Markovian structure these models are sometimes also referred to as Markov-modulated or Markov-additive processes. They come in two flavours. The first is a model where we assume that the state is observable throughout the lifetime of the claim and thus we know the distribution from which returns are being drawn - here the main new feature is the possibility of switching between regimes. The second is a model where in addition, the state is unobservable and we must infer the most probable state via filtering of the observed returns, taking into account the probability of switching between regimes. This second type of model is sometimes also referred to

as a hidden Markov model (HMM), and it bears resemblance to stochastic volatility models, since the instantaneous volatility of our returns is not observable and our estimate of it changes due to the changing probability of each regime filtered from observed returns.

Regime-switching modelling in finance and economics goes back to Hamilton [1989], but the most prolific author on the topic of regime-switching is Robert Elliott, starting with the general textbook Elliott et al. [1995] and with many articles focused on applications to options pricing, e.g. Elliott et al. [2005], Elliott et al. [2007], Elliott and Siu [2008], Elliott et al. [2010], Elliott and Siu [2012], Elliott and Lian [2013], Elliott and Siu [2013]. Most of these (just as most papers by other researchers on the topic) deal with continuous returns processes driven by a Brownian motion, and the regimes are usually ones with low, medium, or high volatility. More recently, research has appeared on extending regime-switching to encompass returns driven by Lévy processes, starting with Chourdakis [2005], with more recent contributions by Elliott and Osakwe [2006], Elliott et al. [2013], Siu [2014], Hainaut [2011], Hainaut and Robert [2014], Hainaut and Colwell [2014], Swishchuk et al. [2014], Kim et al. [2011]. These papers limit themselves to using the Esscher martingale measure for pricing, which relates to maximisation of exponential utility (see e.g. [Fujiwara and Miyahara, 2003, Section 4]). Furthermore, most of them only provide a price for the derivative, but do not investigate hedging strategies or hedging errors.

For HMMs, it is also important to be able to filter the hidden state out of the data. For Lévy processes, there are only limited results, restricted to either a pure jump process or a jump diffusion process. The filtering is usually done via the so-called Zakai filtering equation, giving a non-normalized version of the filter. Elliott and Royal [2008], Elliott and Siu [2013] provide a Zakai equation for the case of pure jump processes, Siu [2014] provides a Zakai equation for the case of jump-diffusion processes. Ceci and Colaneri [2014], Ceci and Colaneri [2012] provide both a Zakai equation for the unnormalized filter and a Kushner-Stratonovich equation for the normalized filter in the case of a jump-diffusion driving both the state and the price process. Schmidt and Frey [2012] provide a Kushner-Stratonovich equation in a similar setting for filtering default intensities in credit derivatives.

Our goal in this chapter is to derive variance-optimal hedging results under a HMM where the returns in different regimes are driven by Lévy processes. In doing so, we will compute the impact of model uncertainty in our modelling on the quadratic hedg-

ing error (and as we saw in the previous chapter, hence the good-deal bounds). The existing literature on this topic is relatively sparse. Mean-variance portfolio selection with regime-switching has been investigated in Elliott et al. [2010] for continuous processes, in this case the regime was assumed to be known. For mean-variance hedging, Pham [2001] provides continuous-time results for a diffusion process with unobserved drift. For regime-switching diffusion processes, we have mean-variance portfolio results in Elliott and Siu [2008]. An application to pricing and hedging credit derivatives using regime-switching compound Poisson processes is described in Schmidt and Frey [2012].

The works closest to ours, dealing with mean-variance hedging for regime-switching Lévy processes, are Pelsser and Delong [2015], Momeya and Pamen [2011], Ceci et al. [2015], Goutte et al. [2014]. The only paper to evaluate the hedging error of a regime-switching Lévy process is Goutte et al. [2014], but does so only in the case when the regime is observable. Our contribution is threefold:

1. We separate the filtering task from the hedging task, i.e. we obtain all quantities of interest by applying general semimartingale quadratic hedging formulae of Černý and Kallsen [2007] to the filtered dynamics of stock prices, where the filtered state serves as an additional state variable;

2. we obtain more explicit dynamics of the posterior estimate of the unobserved state;

3. and we will evaluate the mean-value process, the hedging strategy and the hedging error for the case when the regime is unobservable throughout the life of the contingent claim.

3.2 Setup

We work on fixed probability space $(\Omega, \mathcal{H}, \mathbb{P})$, which will be a product space generated by $M+1$ random variables X and $L^k, k = 1...M$. Here L^k will be $M \geq 2$ independent (in the sense of [Rogers and Williams, 2000, Vol 1, II.22]) Lévy processes $L^k, k = 1...M$, with local differential characteristics (b^k, c^k, F^k) and characteristic exponentials ϕ^k. Each of the component Lévy processes L^k generates its own filtration $\{\hat{\mathcal{H}}_t^k\}$. We denote the enlarged filtration containing all the components $\hat{\mathcal{H}}_t = \bigvee_{k=1}^{M} \hat{\mathcal{H}}_t^k$. If

we stack these individual Lévy processes into a vector, we obtain a vector process $\boldsymbol{L} = (L^1, L^2, ..., L^M)^\top$ with differential characteristics $(\boldsymbol{b}, \boldsymbol{\Sigma}, \boldsymbol{F})$ and characteristic exponential $\boldsymbol{\phi}$, where $\boldsymbol{\Sigma}$ is the variance-covariance matrix. Since the component Lévy processes are independent and hence uncorrelated, we will simplify notation of the differential characteristics to $(\boldsymbol{b}, \boldsymbol{c}, \boldsymbol{F})$, only recording the vector of variances $\boldsymbol{c}$ instead of the entire variance-covariance matrix $\boldsymbol{\Sigma}$, as all off-diagonal entries of $\boldsymbol{\Sigma}$ are zero. Throughout, $\boldsymbol{b}$ denotes the instantaneous drift that relates to the truncation function $h(x) = x$.

The stock price process S is modelled as a special semimartingale $S = S_0 \exp(Y)$, i.e. the cumulative log-return Y is an exponentially special semimartingale (as defined in Kallsen and Shiryaev [2002]; for a definition of a special and exponentially special semimartingale, see appendix definitions C.1,C.2). More specifically, Y is a regime-switching Lévy process which switches between the M Lévy processes L^k defined above. The observable stock price process S (or equivalently Y) generates a filtration $\hat{\mathcal{F}}_t = \sigma(Y_u | 0 \leq u \leq t)$ - we assume this filtration $\hat{\mathcal{F}} \subset \hat{\mathcal{H}}$ captures all the observable information in our model.

The current state of Y is controlled by an unobservable (M-dimensional) finite-state, continuous-time Markov chain $\boldsymbol{X}$. This Markov chain takes values from a finite set of vectors $\{\boldsymbol{e}_k\}_{k=1...M}$ (here $\boldsymbol{e}_k = (0, ..., 0, 1, 0, ..., 0)^\top$ has a value of 1 at its k-th element), with the intensity of transitions between these values controlled by an (transition) intensity matrix $\boldsymbol{A} = \{a_{ij}\}_{i,j=1...M}$. This matrix is sometimes referred to as the infinitesimal generator of $\boldsymbol{X}$ and its elements satisfy $a_{ii} = -\sum_{j \neq i} a_{ji}$. Process $\boldsymbol{X}$ generates a filtration $\{\mathcal{F}_t\}$ independent from Y. We denote the joint (G)lobal filtration of processes $(\boldsymbol{X}, Y)$ as $\{\mathcal{G}_t\} = \{\mathcal{F}_t \vee \hat{\mathcal{F}}_t\}$. The joint process $(\boldsymbol{X}, Y)$ is $\mathcal{G}_t$-Markov.

The process $\boldsymbol{X}$ belongs into the group of so-called point processes (see Brémaud [1981] for an exposition) and the apriori probability $p_t^i = \mathbb{P}(\boldsymbol{X}_t = \boldsymbol{e_i})$ of any given state can be described via a forward Kolmogorov equation:

$$\frac{d\boldsymbol{p}_t}{dt} = \boldsymbol{A}\boldsymbol{p}_t$$

This can easily be solved to give

$$\boldsymbol{p}_T = \boldsymbol{p}_t \exp(\boldsymbol{A}(T - t))$$

105

where $\exp(\cdot)$ is understood to be the matrix exponential. The process $\boldsymbol{X}$ is a càdlàg semimartingale, because it can be constructed as a sum of independent Poisson random measures controlling each of the transitions between different states (see Brémaud [1999, Theorem 9.1.2], Kella and Yor [2017, eqn. 48] for such a construction). [Elliott et al., 1995, Lemma 2.1] show that $\boldsymbol{X}$ has the following canonical semimartingale decomposition:

$$\boldsymbol{X}_t = \boldsymbol{X}_0 + \int_0^t \boldsymbol{A}\boldsymbol{X}_{u-}\, du + \boldsymbol{M}_t$$

As a consequence, $\boldsymbol{M}_t := \boldsymbol{X}_t - \boldsymbol{X}_0 - \int_0^t \boldsymbol{A}\boldsymbol{X}_{u-}\, du$ is a $(\mathcal{F}_t, \mathbb{P})$-martingale.

Given process $\boldsymbol{X}$, the value of Y at any time t is determined by the following equation:

$$Y_t = Y_0 + \int_0^t \boldsymbol{X}_{s-}^\top d\boldsymbol{L}_s = Y_0 + \boldsymbol{X}_- \cdot \boldsymbol{L}_t$$

Here, we can assume without loss of generality $Y_0 = 0$. Assuming that we can only glean information about $\boldsymbol{X}$ by observing Y (i.e. there are no other observables that carry information about $\boldsymbol{X}$), we will consider our best estimate of unobservable $\boldsymbol{X}$ to be the so-called *optional projection* of $\boldsymbol{X}$ onto filtration $\{\hat{\mathcal{F}}_t\}$, denoted $\hat{\boldsymbol{X}}$. In general, an optional projection is a projection of process $\boldsymbol{X}$ onto a filtration to which it is not adapted. [Rogers and Williams, 2000, Theorem 7.1] (equivalently [Bain and Crisan, 2009, Theorem 2.7]) provides a definition of the optional projection (and simultaneously theorem proving uniqueness).

Theorem, Definition 3.1. *Let X be a bounded measurable process, then there exists an optional process $\hat{X}$ called the optional projection of X such that for every stopping time τ:*

$$\hat{X}_\tau 1_{\tau < \infty} = \mathbb{E}[X_\tau 1_{\tau < \infty} | \hat{\mathcal{F}}_\tau]$$

This process is unique up to indistinguishability, i.e. any processes which satisfy these conditions will be indistinguishable.

In practical terms, the optional projection can, for any finite t, be more simply computed as follows:

$$\hat{\boldsymbol{X}}_t := \mathbb{E}[\boldsymbol{X}_t | \hat{\mathcal{F}}_t]$$

and the paths of this process will be unique (up to a null set - see [Protter, 2004, Chapter I.1] for details of indistinguishability). Each element $\hat{X}^i$ of vector $\hat{\boldsymbol{X}}$ is

the posterior probability of being in state e_i given observations of process Y, which therefore means the elements must sum to 1, i.e.

$$\mathbf{1}^\top \hat{\boldsymbol{X}} = 1$$

Furthermore, since $\boldsymbol{X}$ is a bounded $(\mathcal{G}, \mathbb{P})$-semimartingale, it follows from [Föllmer and Protter, 2010, Theorem 9] that $\hat{\boldsymbol{X}}$ is an $(\hat{\mathcal{F}}, \mathbb{P})$-semimartingale.

Having defined how we will handle the unobservability of $\boldsymbol{X}$, we now proceed to describe the properties of Y under filtration $\{\hat{\mathcal{F}}_t\}$. We will use the following lemma to obtain the characteristic function of Y.

Lemma 3.2. *For each $t \in [0, T]$, let $\boldsymbol{J}(t, T) := (J_1(t, T), J_2(t, T), ..., J_M(t, T)) \in [0, T - t]^{\otimes M}$ where $J_k(t, T)$ is the occupation time of the chain $\boldsymbol{X}$ in state $\boldsymbol{e}_k$ in the interval $[t, T](i.e. \boldsymbol{J}(t, T) := \int_t^T \boldsymbol{X}_{s-}\, ds)$. Suppose for each $\boldsymbol{\lambda} := (\lambda_1, \lambda_2, ..., \lambda_M)^\top \in \mathbb{R}^M$, and $\Phi_{\boldsymbol{J}(t,T)|\mathcal{G}_t}(\boldsymbol{\lambda})$ is the conditional moment-generating function of the vector of occupation times $\boldsymbol{J}(t, T)$ given $\mathcal{G}_t$ under $\mathbb{P}$ evaluated at the vector $\boldsymbol{\lambda}$. That is,*

$$\Phi_{\boldsymbol{J}(t,T)|\mathcal{G}_t}(\boldsymbol{\lambda}) = \mathbb{E}[\exp(\boldsymbol{\lambda}^\top \boldsymbol{J}(t, T))|\mathcal{G}_t]$$

Let $diag(\boldsymbol{\lambda})$ be the diagonal matrix with diagonal elements given by the components of $\boldsymbol{\lambda}$. Then

$$\Phi_{\boldsymbol{J}(t,T)|\mathcal{G}_t}(\boldsymbol{\lambda}) = \mathbf{1}^\top \exp\{(\boldsymbol{A} + diag(\boldsymbol{\lambda}))(T - t)\}\boldsymbol{X}_t.$$

Proof. See [Elliott and Siu, 2013, Lemma 5.1]. $\qquad\square$

Using the lemma above, we can prove that the (conditional) characteristic function of Y is given by the formula from the following theorem.

Theorem 3.3. *The conditional characteristic function of regime-switching Lévy process Y under filtration $\{\hat{\mathcal{F}}_t\}$ is:*

$$\begin{aligned}\Phi_{Y_T|\hat{\mathcal{F}}_t}(u) &= \mathbb{E}[\exp(iuY_T)|\hat{\mathcal{F}}_t] \\ &= \exp(iuY_t)\mathbf{1}^\top \exp\{(\boldsymbol{A} + diag(\boldsymbol{\phi}(u)))(T - t)\}\hat{\boldsymbol{X}}_t\end{aligned}$$

where $\boldsymbol{\phi}(u) = (\phi^1(u), ..., \phi^M(u))^\top$.

Proof. Here we will follow the proof of [Elliott and Siu, 2013, Theorem 5.1]. Using

Itô's formula on $\exp(iuY_t)$ and denoting Y^c the continuous part of Y, we have:

$$d(e^{iuY_t}) = e^{iuY_{t-}}(iudY_t - \frac{1}{2}u^2 d\langle Y^c \rangle_t + (e^{iuy} - 1 - iuy)dJ^Y)$$

$$= e^{iuY_{t-}}(iu\boldsymbol{X}_{t-}^\top d\boldsymbol{L}_t - \frac{1}{2}u^2 \boldsymbol{X}_{t-}^\top d\langle \boldsymbol{L}^c \rangle_t \boldsymbol{X}_{t-} + (e^{iuy} - 1 - iuy)\boldsymbol{X}_{t-}^\top d\boldsymbol{J}^L)$$

Integrating and using the fact that $\boldsymbol{X}^2 = \boldsymbol{X}$, we have

$$e^{iuY_T} = e^{iuY_t} + \int_t^T e^{iuY_{s-}} iu\boldsymbol{X}_{s-}^\top d\boldsymbol{L}_s - \frac{1}{2}u^2 \int_t^T e^{iuY_{s-}} \boldsymbol{X}_{s-}^\top \boldsymbol{c} ds$$

$$+ \int_t^T e^{iuY_{s-}} \int_{\mathbb{R}} (e^{iuy} - 1 - iuy)\boldsymbol{X}_{s-}^\top \boldsymbol{J}^L(dy, ds)$$

Since Y is exponentially special, the Lévy process $\boldsymbol{L}$ has the following decomposition:

$$d\boldsymbol{L} = \boldsymbol{b}dt + \text{diag}(\sqrt{\boldsymbol{c}})d\boldsymbol{W}_t + \int_{\mathbb{R}} y(\boldsymbol{J}^L(dy, dt) - \boldsymbol{F}(dy)dt)$$

We can plug this into our integral representation of $\exp(iuY_T)$:

$$e^{iuY_T} = e^{iuY_t} + \int_t^T e^{iuY_{s-}} iu\boldsymbol{X}_{s-}^\top \left(\boldsymbol{b}ds + \text{diag}(\sqrt{\boldsymbol{c}})d\boldsymbol{W}_s + \int_{\mathbb{R}} y(\boldsymbol{J}^L(dy, ds) - \boldsymbol{F}(dy)ds) \right)$$

$$- \frac{1}{2}u^2 \int_t^T e^{iuY_{s-}} \boldsymbol{X}_{s-}^\top \boldsymbol{c}ds + \int_t^T e^{iuY_{s-}} \int_{\mathbb{R}} (e^{iuy} - 1 - iuy)\boldsymbol{X}_{s-}^\top \boldsymbol{J}^L(dy, ds)$$

Rearranging, we get

$$e^{iuY_T} = e^{iuY_t} + \int_t^T e^{iuY_{s-}} \boldsymbol{X}_{s-}^\top (\boldsymbol{b}iu - \frac{1}{2}u^2 \boldsymbol{c})ds + \int_t^T e^{iuY_{s-}} \text{diag}(\sqrt{\boldsymbol{c}})d\boldsymbol{W}_s$$

$$+ \int_t^T e^{iuY_{s-}} \int_R (e^{iuy} - 1)\boldsymbol{X}_{s-}^\top (\boldsymbol{J}^L(dy, ds) - \boldsymbol{F}(dy)ds)$$

$$+ \int_t^T e^{iuY_{s-}} \boldsymbol{X}_{s-}^\top \int_{\mathbb{R}} (e^{iuy} - 1 - iuy)\boldsymbol{F}(dy)ds$$

Taking expectations under the combined filtration $\mathcal{G}_t \vee \mathcal{F}_T$ (i.e. we know the path of $\boldsymbol{X}$ up to T but the path of Y only up to t), we find

$$\mathbb{E}[e^{iuY_T}|\mathcal{G}_t \vee \mathcal{F}_T] = e^{iuY_t} + \int_t^T \mathbb{E}[e^{iuY_{s-}}|\mathcal{G}_t \vee \mathcal{F}_{s-}]\boldsymbol{X}_{s-}^\top (\boldsymbol{b}iu - \frac{1}{2}\boldsymbol{c}u^2)ds$$

$$+ \int_t^T \mathbb{E}[e^{iuY_{s-}}|\mathcal{G}_t \vee \mathcal{F}_{s-}]\boldsymbol{X}_{s-}^\top \int_{\mathbb{R}} (e^{iuy} - 1 - iuy)\boldsymbol{F}(dy)ds$$

We notice that this is an ODE for variable $Z_s := \mathbb{E}[e^{iuY_s}|\mathcal{G}_t \vee \mathcal{F}_s]$ which is readily solved to give the solution

$$\mathbb{E}[e^{iuY_T}|G_t \vee \mathcal{F}_T] = e^{iuY_t} \exp\left(\int_t^T \boldsymbol{X}_{s-}^\top (\boldsymbol{b}iu - \frac{1}{2}\boldsymbol{c}u^2 + \int_\mathbb{R} (e^{iuy} - 1 - iuy)\boldsymbol{F}(dy))ds\right)$$

Observing that $\boldsymbol{\phi}(u) = \boldsymbol{b}iu - \frac{1}{2}\boldsymbol{c}u^2 + \int_\mathbb{R}(e^{iuy} - 1 - iuy)\boldsymbol{F}(dx)$ is the characteristic exponent of Lévy process $\boldsymbol{L}$, we can re-write this as:

$$\mathbb{E}[e^{iuY_T}|G_t \vee \mathcal{F}_T] = e^{iuY_t} \exp\left(\int_t^T \boldsymbol{X}_{s-}^\top \boldsymbol{\phi}(u)ds\right) = e^{iuY_t} \exp\left(\boldsymbol{\phi}(u)^\top \int_t^T \boldsymbol{X}_{s-}ds\right)$$

Now we recall that $\boldsymbol{J}(t,T) = \int_t^T \boldsymbol{X}_{s-}ds$ is the vector of occupation times that process X spends in each of its M states. Thus we can rewrite this as follows:

$$\mathbb{E}[\exp(iuY_T)|\mathcal{G}_t \vee \mathcal{F}_T] = \exp(iuY_t + \boldsymbol{\phi}(u)^\top \boldsymbol{J}(t,T))$$

Then, under global information $\mathcal{G}_t$, we know from the previous lemma the conditional characteristic function of $\boldsymbol{J}(t,T)$. The tower law of expectations gives us

$$\begin{aligned}
\mathbb{E}[\mathbb{E}[\exp(iuY_T)|\mathcal{G}_t \vee \mathcal{F}_T]|\mathcal{G}_t] &= \exp(iuY_t)\mathbb{E}[\exp(\boldsymbol{\phi}(u)^\top \boldsymbol{J}(t,T))|\mathcal{G}_t] \\
&= \exp(iuY_t)\boldsymbol{1}^\top \Phi_{\boldsymbol{J}(t,T)|\mathcal{G}_t}(\boldsymbol{\phi}(u)) \\
&= \exp(iuY_t)\boldsymbol{1}^\top \exp\{(\boldsymbol{A} + \mathrm{diag}(\boldsymbol{\phi}(u)))(T-t)\}\boldsymbol{X}_t
\end{aligned}$$

Finally, continuing to use the tower law and taking expectations of the above under $\hat{\mathcal{F}}_t \subset \mathcal{G}_t$, all that is left is an optional projection of X contained in a linear term:

$$\begin{aligned}
\mathbb{E}[\exp(iuY_T)|\hat{\mathcal{F}}_t] &= \mathbb{E}[\exp(iuY_t)\boldsymbol{1}^\top \exp\{(\boldsymbol{A} + \mathrm{diag}(\boldsymbol{\phi}(u)))(T-t)\}\boldsymbol{X}_t|\hat{\mathcal{F}}_t] \\
&= \exp(iuY_t)\boldsymbol{1}^\top \exp\{(\boldsymbol{A} + \mathrm{diag}(\boldsymbol{\phi}(u)))(T-t)\}\hat{\boldsymbol{X}}_t
\end{aligned}$$

$\square$

From the characteristic function, we can directly compute the moments of underlying price S (assuming they are finite).

Corollary 3.4. *For any $z \in \mathbb{R}$ such that $\mathbb{E}[\exp(zY_T)] < \infty$ it holds that*

$$\mathbb{E}[S_T^z|\hat{\mathcal{F}}_t] = S_t^z \boldsymbol{1}^\top \exp\{(\boldsymbol{A} + diag(\boldsymbol{\phi}(-iz)))(T-t)\}\hat{\boldsymbol{X}}_t$$

Proof. This follows directly from the previous theorem, as $S_T^z = S_0^z \exp(zY_T)$ and so $\mathbb{E}[S_T^z|\hat{\mathcal{F}}_t] = S_0^z \mathbb{E}[\exp(zY_T)|\hat{\mathcal{F}}_t] = S_0^z \Phi_{Y_T|\hat{\mathcal{F}}_t}(-iz)$. $\qquad\qquad\qquad\square$

We can also use theorem 3.3 to state conditions under which S is a $(\mathcal{G}, \mathbb{P})$ and $(\hat{\mathcal{F}}, \mathbb{P})$-martingale.

Corollary 3.5. *Assuming S has finite absolute first moment, i.e. $\mathbb{E}[|S_T|] < \infty$, then it is a $(\hat{\mathcal{F}}, \mathbb{P})$-martingale (and a $(\mathcal{G}, \mathbb{P})$-martingale) iff*

$$\forall k \in \{1...M\} : \phi^k(-i) = 0$$

in other words, when each of the M component Lévy processes is a martingale.

Proof. See [Hainaut, 2011, Proposition 6.2]. $\qquad\qquad\qquad\qquad\qquad\qquad\square$

3.3   Stochastic dynamics for $\hat{X}$

To use standard mean-variance hedging formulas of Hubalek et al. [2006], Černý [2007] in our setting, we will require the dynamics of $\hat{X}$. Here, we will derive the unnormalized SDE for $\hat{X}$ (as opposed to the normalized dynamics expressed by the so-called Zakai equation, discussed in appendix C.1). This will give us its $\{\hat{\mathcal{F}}_t\}$-characteristics. In this section we provide a theorem that provides an SDE for $\hat{X}$ under $\{\hat{\mathcal{F}}_t\}$ from which it follows that for a Lévy process Y, the joint process $(\hat{X}, Y)$ is $\hat{\mathcal{F}}$-Markov. We will make an assumption on the structure of volatility across the various Lévy processes - we assume the volatility of their diffusive term is identical for all of them, with variation in volatility across different models being caused by differences in the Lévy measure. This assumption could be weakened and for more general non-Lévy models, the diffusive volatility could be dependent on time or the level of Y - we only require it to be independent of state X, an assumption commonly made across the literature (see e.g. Siu [2014] and references therein for a discussion on the necessity of this assumption).

Assumption 3.1. All processes L^k have the same diffusive characteristic c^k:

$$\forall k \in \{1 \cdots M\} : c^k = \bar{c}$$

Under the above assumption, $\boldsymbol{c}^\top \boldsymbol{X}_t = \bar{c}$ and we can define an innovations process I,

which is a $(\mathbb{P}, \hat{\mathcal{F}}_t)$-Wiener process:

$$I_t := W_t + \int_0^t \frac{\boldsymbol{b}^\top (\boldsymbol{X}_s - \hat{\boldsymbol{X}}_s)}{\sqrt{\bar{c}}} ds \qquad (3.1)$$

We will make an additional technical assumption on the Lévy processes we consider.

Assumption 3.2. The Lévy densities F^k of Lévy processes $L^k, k = 1...M$ are absolutely continuous with respect to Lebesgue measure.

As a consequence of the above assumption we have that the Lévy measure F^k has a well defined density function, i.e. there exists such an integrable function f that $F^k(dx) = f^k(x)dx$. Furthermore, the Lévy densities F^k all have the same null sets. For processes with non-integrable densities that have explosive limiting behaviour for very small jumps i.e. $\lim_{x \to 0} f^k(x) = \infty$, we will consider their truncated variants, where we discard any jumps smaller (in absolute value) than a fixed small $\epsilon > 0$ (see Elliott and Royal [2008] for such a truncation in the case of the Variance Gamma process).

In what follows we provide an explicit form of the SDE for $\hat{\boldsymbol{X}}$, using lemma C.6 about the structure of local martingales.

Theorem 3.6. *Let assumptions 3.1 and 3.2 hold and assume that X and Y have no common jumps. Then the optional projection $\hat{\boldsymbol{X}}_t = \mathbb{E}[\boldsymbol{X}_t | \hat{\mathcal{F}}_t]$ can be represented as follows:*

$$\hat{\boldsymbol{X}}_t = \hat{\boldsymbol{X}}_0 + \int_0^t \boldsymbol{A} \hat{\boldsymbol{X}}_{s-} ds + \int_0^t \int_{\mathbb{R}} w(\hat{\boldsymbol{X}}_{s-}, y)(J^Y(ds, dy) - \hat{\boldsymbol{X}}_{s-}^\top \boldsymbol{F}(dy)ds) + \int_0^t a(\hat{\boldsymbol{X}}_{s-})dI_s \qquad (3.2)$$

where I_t is the $(\hat{\mathcal{F}}_t, \mathbb{P})$-Wiener process defined in (3.1) and $w(\boldsymbol{x}, y), a(\boldsymbol{x})$ are vector functions:

$$w(\boldsymbol{x}, y) = diag(\boldsymbol{x}) \left(\frac{d\boldsymbol{F}(y)}{\boldsymbol{x}^\top d\boldsymbol{F}(y)} - 1 \right)$$
$$=: diag(\boldsymbol{x})\tilde{w}(\boldsymbol{x}, y)$$
$$a(\boldsymbol{x}) = diag(\boldsymbol{x}) \frac{(\boldsymbol{b} - \boldsymbol{x}^\top \boldsymbol{b})}{\sqrt{\bar{c}}} =: diag(\boldsymbol{x})\tilde{a}(\boldsymbol{x})$$

Proof. For the proof, see the appendix. $\qquad \qquad \square$

Remark 3.7. If we denote the semimartingale decomposition of our special semi-martingale Y as

$$Y = B^Y + Y^c + y * (J^Y - \nu^Y)$$

and define $Z := (B^Y + Y^c)/\sqrt{\bar{c}}$, then we can rewrite I as follows:

$$I_t := W_t + \frac{1}{\sqrt{\bar{c}}} \int_0^t (\boldsymbol{X}_{s-} - \hat{\boldsymbol{X}}_{s-})^\top \boldsymbol{b}\, ds$$

$$= Z_t - \frac{1}{\sqrt{\bar{c}}} \int_0^t \hat{\boldsymbol{X}}_{s-}^\top \boldsymbol{b}\, ds$$

to write

$$\hat{\boldsymbol{X}}_t = \hat{\boldsymbol{X}}_0 + \int_0^t \boldsymbol{A}\hat{\boldsymbol{X}}_{s-} ds + \int_0^t \int_{\mathbb{R}} w(\hat{\boldsymbol{X}}_{s-}, y)(J^Y(ds, dy) - \hat{\boldsymbol{X}}_{t-}^\top \boldsymbol{F}(dy)ds)$$

$$+ \int_0^t a(\hat{\boldsymbol{X}}_{s-})(dZ_s - \hat{\boldsymbol{X}}_{s-}^\top \boldsymbol{b}\, ds)$$

Here Z is to be understood as the continuous part of the process, i.e. the (compensated) drift and the continuous diffusion. This is useful if the continuous and jump components of Y are observable separately.

Remark 3.8. Note if the process Y being used for filtering has differential characteristic $\boldsymbol{b} \equiv 0$, i.e. it has no local drift, the diffusive term disappears.

Looking at the form of the SDE for $\hat{\boldsymbol{X}}$, we can read off the semimartingale differential characteristics of the process under $\{\hat{\mathcal{F}}_t\}$.

Corollary 3.9. *The semimartingale predictable differential characteristics of $\hat{X}$ under filtration $\{\hat{\mathcal{F}}_t\}$ are:*

$$\boldsymbol{b}_t^{\hat{\boldsymbol{X}}} = \boldsymbol{A}\hat{\boldsymbol{X}}_{t-}$$

$$\boldsymbol{c}_t^{\hat{\boldsymbol{X}}} = a(\hat{\boldsymbol{X}}_{t-})a(\hat{\boldsymbol{X}}_{t-})^\top$$

$$F_t^{\hat{\boldsymbol{X}}}(dy) = w(\hat{\boldsymbol{X}}_{t-}, y)\hat{\boldsymbol{X}}_{t-}^\top \boldsymbol{F}(dy)$$

Example 3.1 (Pure jump process with no switching). Let us assume $\boldsymbol{A} = 0$ and there is no diffusive term. Then the SDE for $\hat{\boldsymbol{X}}$ can be written as

$$\hat{\boldsymbol{X}}_t = \hat{\boldsymbol{X}}_0 + \int_0^t \int_{\mathbb{R}} \mathrm{diag}(\hat{\boldsymbol{X}}_{s-}) \left(\frac{d\boldsymbol{F}(y)}{\hat{\boldsymbol{X}}_{s-}^\top d\boldsymbol{F}(y)} - 1 \right) (J^Y(dy, ds) - \hat{\boldsymbol{X}}_{s-}^\top \boldsymbol{F}(dy)ds)$$

We note that the jump term represents the continuous-time version of Bayes theorem. The update in reaction to a jump here is done in relation to the compensating jumps instead of just the jump term to take into account the effect of 'non-events', i.e. to take into account the drift of the Bayesian update for the moments when no jump arrives (we will show in a discrete time analogy in the next example that such a 'non-event' bears information and impacts the filter).

Example 3.2 (Discrete-time analogy when no jump arrives)**.** One might consider the case of no jumps arriving, i.e. what happens for $w(\hat{\boldsymbol{X}}_{t-}, 0)$. In a pure jump case, different jump processes may have different intensities of arrival. Thus even the fact that no jump has arrived may be informative. Here we will illustrate that this information is used by the filter. First, we provide an analogy in discrete time.

In general we have observable variable Y and a prior estimate $\boldsymbol{p}(t) = \mathbb{P}(\boldsymbol{X}_t | Y^t)$ of our state $\boldsymbol{X}$ where $Y^t = \{Y_s\}_{s=1\ldots t}$ is the set of observations up to time t. By Bayes' Theorem we have:

$$\boldsymbol{p}(t + \Delta t) = \mathbb{P}(\boldsymbol{X}_{t+\Delta t} | Y^{t+\Delta t}) = \frac{\mathrm{diag}(\mathbb{P}(Y_{t+\Delta t} | \boldsymbol{X}_t)) \boldsymbol{p}(t)}{\mathbb{P}(Y_{t+\Delta t} | \boldsymbol{X}_t)^\top \boldsymbol{p}(t)}$$

Here $\mathbb{P}(Y_{t+\Delta t} | \boldsymbol{X}_t)$ denotes a vector of probabilities $[\mathbb{P}(Y_{t+\Delta t} | \boldsymbol{X}_t = \boldsymbol{e}_1), \ldots, \mathbb{P}(Y_{t+\Delta t} | \boldsymbol{X}_t = \boldsymbol{e}_M)]^\top$. Let us now assume we only have two states, with the Lévy processes in each state being compound Poisson processes with arrival intensities λ_1 and λ_2 respectively, and the transition intensity matrix is $\boldsymbol{A} \equiv 0$. Let the observation be that there was no jump i.e. $Y_{t+\Delta t} - Y_t = 0$. Denoting p_1 the first element of vector $\boldsymbol{p}$, that will mean

$$p_1(t + \Delta t) = \frac{(1 - \lambda_1 \Delta t) p_1(t)}{(1 - \lambda_1 \Delta t) p_1(t) + (1 - \lambda_2 \Delta t) p_2(t)}$$

Thus, the change in probability can be written as:

$$p_1(t + \Delta t) - p_1(t) = \frac{(1 - \lambda_1 \Delta t) p_1(t) - (1 - \lambda_1 \Delta t) p_1^2(t) - (1 - \lambda_2 \Delta t) p_1(t) p_2(t)}{(1 - \lambda_1 \Delta t) p_1(t) + (1 - \lambda_2 \Delta t) p_2(t)}$$

Looking at the zero order $\mathcal{O}(1)$ effect, we have

$$p_1(t) - p_1^2(t) - p_1(t) p_2(t) = p_1(t)(1 - p_1(t) - p_2(t)) = 0$$

In other words, at zero order, the value of the filter does not change. In the first order

$\mathcal{O}(\Delta t)$ term, the discrete time probability change is given as

$$-\lambda_1 p_1(t) + \lambda_1 p_1^2(t) + \lambda_2 p_1(t) p_2(t) = (\lambda_2 - \lambda_1) p_1(t) p_2(t)$$

In words, the probability change will depend on the difference of arrival intensities of the two processes. This illustrates that even when no jump arrives, this information has an impact on the value of the filter and is thus informative about the current state.

In the continuous time case, we know that the compensator ν of the Poisson process is $\nu^k(dy, dt) = 1_{y=\alpha} F^k(dy) dt = \lambda_k dt$. The function $w(\boldsymbol{x}, y)$ will simplify to

$$w(\boldsymbol{x}, y) = \text{diag}(\boldsymbol{x}) \left(\frac{\boldsymbol{\lambda}}{\boldsymbol{x}^\top \boldsymbol{\lambda}} - 1 \right)$$

If we observe a jump of size zero (i.e. $J^Y(dy, dt) = 0$) then the filter behaves as follows:

$$d\hat{\boldsymbol{X}}_t = \text{diag}(\hat{\boldsymbol{X}}_{t-}) \left(\frac{\boldsymbol{\lambda}}{\hat{\boldsymbol{X}}_{t-}^\top \boldsymbol{\lambda}} - 1 \right) (0 - \hat{\boldsymbol{X}}_{t-}^\top \boldsymbol{\lambda} dt)$$

$$= \text{diag}(\hat{\boldsymbol{X}}_{t-})(-\boldsymbol{\lambda} + \hat{\boldsymbol{X}}_{t-}^\top \boldsymbol{\lambda}) dt$$

With some simple algebra and using $\hat{X}_{t-}^2 = 1 - \hat{X}_{t-}^1$ we find that the first element $d\hat{X}_t^1$ of the vector $d\hat{\boldsymbol{X}}_t$ is

$$d\hat{X}_t^1 = -\hat{X}_{t-}^1 \lambda_1 + (\hat{X}_{t-}^1)^2 \lambda_1 + \hat{X}_{t-}^1 \hat{X}_{t-}^2 \lambda_2$$

$$= (\lambda_2 - \lambda_1) \hat{X}_{t-}^1 \hat{X}_{t-}^2$$

In other words, we retrieve the same result as in the case of the discrete-time asymptotic approximation.

Example 3.3 (Pure diffusion process with no transitions). Let us assume $\boldsymbol{A} = 0$ and the vector of Lévy processes $\boldsymbol{L}$ contains only diffusions. Then

$$\hat{\boldsymbol{X}}_t = \hat{\boldsymbol{X}}_0 + \frac{1}{\sqrt{\bar{c}}} \int_0^t \text{diag}(\hat{\boldsymbol{X}}_{s-})(\boldsymbol{b} - \boldsymbol{b}^\top \hat{\boldsymbol{X}}_{s-})(dY_s - \boldsymbol{b}^\top \hat{\boldsymbol{X}}_{s-} \, ds)$$

(note here that the Z previously defined in Remark 3.8 is now equal to Y as we have no discontinuous part of the process). In words the new value of the filter $\hat{\boldsymbol{X}}$ is given by the difference of the drift in each state relative to the mean estimated drift based

on past observations, multiplied by the observed move and its deviation from the mean estimated drift, scaled by the volatility and the probability of each state in the past step.

3.4 Mean-value process

We now proceed to apply the same mean-variance hedging theory from the previous chapters in the setting of a regime-switching model. We will solve the classic mean-variance optimization problem

$$\varepsilon_0^2 = \min_{v \in \mathbb{R}} \min_{\vartheta \in \Theta} \mathbb{E}[(v + \vartheta \cdot S_T - H)^2]$$

for the case when S is a $(\mathbb{P}, \hat{\mathcal{F}})$-martingale (Corollary 3.5 stated conditions under which this holds for our model), using Fourier transform methods from Hubalek et al. [2006],Černý [2007] to obtain the mean-value process V and the variance optimal hedging strategy ξ. Unlike these previous papers, we will resort to an approximative backward iteration scheme to numerically compute the quadratic hedging error ε_0^2.

Assumption 3.3. S is a $(\mathbb{P}, \hat{\mathcal{F}})$-martingale.

For the case when S is a martingale, we know from Föllmer and Sondermann [1986] that the mean-value process is simply the expectation of the payoff under physical measure, i.e.

$$V_t = \mathbb{E}[H | \hat{\mathcal{F}}_t]$$

As we intend to compute the mean-value process and variance-optimal hedging strategy via Fourier transform techniques, we will assume that the payoff has an integral representation, as defined in [Kallsen et al., 2009, Assumption 3.1]:

Assumption 3.4. The payoff is of the form $H = f^H(S_T)$ for some function $f^H : (0, \infty) \to \mathbb{R}$, s.t.

$$f^H(s) = \int_{R-i\infty}^{R+i\infty} s^z l(z) \, dz$$

for $l : \mathbb{C} \to \mathbb{C}$ and $R \in \mathbb{R}$ such that $x \to l(R+ix)$ is integrable and $\mathbb{E}[\exp(2RY_T)] < \infty$

For examples of such integral representations of common payoffs, see Hubalek et al. [2006]. Note that the condition of integrability of l may cause restrictions to the values

R can take for certain payoffs (e.g. vanilla puts and calls have the same function $l(z)$ but different offsets R in the complex plane).

Theorem 3.10. *If the payoff H has the form given in assumption 3.4 and S is a martingale (assumption 3.3), then the mean-value process of the regime-switching model can be written as a function v:*

$$V_t = v(t, \hat{\boldsymbol{X}}_t, S_t).$$

such that

$$v(t, \boldsymbol{x}, s) = \int_{R-i\infty}^{R+i\infty} l(z) s^z \mathbf{1}^\top \boldsymbol{M}(z, T-t) \boldsymbol{x} dz$$

and

$$\boldsymbol{M}(z, T-t) := \exp\{(\boldsymbol{A} + diag(\boldsymbol{\phi}(-iz)))(T-t)\}.$$

Proof. Under assumption 3.4, using Fubini's theorem we can write the mean-value process via the Laplace transform as

$$V_t = \mathbb{E}[H|\hat{\mathcal{F}}_t] = \mathbb{E}[f^H(S_T)|\hat{\mathcal{F}}_t] = \mathbb{E}\left[\int_{R-i\infty}^{R+i\infty} S_T^z l(z)\, dz|\hat{\mathcal{F}}_t\right] = \int_{R-i\infty}^{R+i\infty} V(z)_t l(z)\, dz$$

where

$$V(z)_t := \mathbb{E}[S_T^z|\hat{\mathcal{F}}_t]$$

From corollary 3.4, we know:

$$V(z)_t = S_t^z \mathbf{1}^\top \exp\{(\boldsymbol{A} + diag(\boldsymbol{\phi}(-iz)))(T-t)\}\hat{\boldsymbol{X}}_t$$

By defining $\boldsymbol{M}(z, T-t) := \exp\{(\boldsymbol{A} + diag(\boldsymbol{\phi}(-iz)))(T-t)\}$ we get the abbreviated form

$$V(z)_t = S_t^z \mathbf{1}^\top \boldsymbol{M}(z, T-t)\hat{\boldsymbol{X}}_t$$

Thus, the mean-value process is ultimately a function of the current estimated state and the spot price:

$$V_t = v(t, \hat{\boldsymbol{X}}_t, S_t).$$

where

$$v(t, \boldsymbol{x}, s) = \int_{R-i\infty}^{R+i\infty} l(z) S^z \mathbf{1}^\top M(z, T-t) \boldsymbol{x} dz$$

$\square$

116

Let us remark that the regime-switching mean-value process is close to, but not completely equal to (due to impact of transition matrix $\boldsymbol{A}$), the weighted average of two single-regime mean-value processes.

Corollary 3.11. *Under assumptions of theorem 3.10, the mean-value process of the regime-switching model can be written as a function f:*

$$V_t = f(t, \hat{\boldsymbol{X}}_t, Y_t) = f(t, \hat{\boldsymbol{X}}_t, Y_t | S_0, T)$$

such that

$$f(t, \boldsymbol{x}, y) = \int_{R-i\infty}^{R+i\infty} S_0^z l(z) \exp(yz) \mathbf{1}^\top \boldsymbol{M}(z, T-t) \boldsymbol{x} \, dz$$

Proof. This follows directly from the fact that: $V_t = f(t, \hat{\boldsymbol{X}}_t, Y_t) = v(t, \hat{\boldsymbol{X}}_t, S_0 \exp(Y_t))$

$\square$

3.5 Variance optimal hedging strategy

As established in Černý [2007] and references therein, for a martingale S the variance optimal hedging strategy ξ (and in the martingale case also the variance-optimal strategy) can be expressed as a ratio of the predictable quadratic covariation of V and S and the predictable quadratic variation of S:

$$\xi = \frac{d\langle V, S \rangle}{d\langle S, S \rangle}$$

Here the quadratic covariation and variation will always be under the filtration of S (or equivalently, of Y), i.e. under $\hat{\mathcal{F}}_t$, unless specified otherwise.

We know $d\langle S \rangle_t = S_{t-}^2(d\langle Y^c \rangle_t + \int_R (e^z - 1)^2 \hat{\boldsymbol{X}}_{t-}^\top \boldsymbol{F}(dz) dt)$. Under assumption 3.1 this is

$$d\langle S \rangle_t = S_{t-}^2 \sigma^2(\hat{\boldsymbol{X}}_{t-}) dt \tag{3.3}$$

where we have defined $\sigma^2(\hat{\boldsymbol{X}}_{t-}) := \bar{c} + \hat{\boldsymbol{X}}_{t-} \int_R (e^y - 1)^2 \boldsymbol{F}(dy)$.

For the quadratic covariation term $\langle V, S \rangle$, we can move the covariation inside the Laplace transform of V. Using the definition of covariation and assuming that all the appropriate integrals, sums and limits below are finite, we can write (where $t_k \in [0, t]$

are equally spaced times with distance Δt):

$$
\begin{aligned}
\langle V, S \rangle_t &= \lim_{\Delta t \to 0} \sum_{k=1}^{n} (V_{t_k} - V_{t_{k-1}})(S_{t_k} - S_{t_{k-1}}) \\
&= \lim_{\Delta t \to 0} \sum_{k=1}^{n} \left(\int_{R-i\infty}^{R+i\infty} V(z)_{t_k} l(z) dz - \int_{R-i\infty}^{R+i\infty} V(z)_{t_{k-1}} l(z) dz \right) (S_{t_k} - S_{t_{k-1}}) \\
&= \int_{R-i\infty}^{R+i\infty} \lim_{\Delta t \to 0} \sum_{k=1}^{n} (V(z)_{t_k} - V(z)_{t_{k-1}})(S_{t_k} - S_{t_{k-1}}) l(z) dz \\
&= \int_{R-i\infty}^{R+i\infty} \langle V(z), S \rangle_t l(z) \, dz
\end{aligned}
$$

Then for our specific model, we have the following result.

Theorem 3.12. *Under the assumptions of theorem 3.6, assumption 3.3 and assumption 3.4, the variance optimal hedging strategy is given as*

$$
\xi_t = \int_{R-i\infty}^{R+i\infty} \xi_t(z) l(z) dz
$$

where

$$
\begin{aligned}
\xi_t(z) = \frac{S_{t-}^{z-1}}{\sigma^2(\hat{\boldsymbol{X}}_{t-})} \mathbf{1}^{\top} \boldsymbol{M}(T-t, z) \times \\
\times \left(z^2 \bar{c} + z\sqrt{\bar{c}} a(\hat{\boldsymbol{X}}_{t-}) + \int_{\mathbb{R}} (e^y - 1)(e^{zy}(1 + w(\hat{\boldsymbol{X}}_{t-}, y)) - 1) \hat{\boldsymbol{X}}_{t-}^{\top} \boldsymbol{F}(dy) \right)
\end{aligned}
$$

Furthermore, we can write $\xi_t = k(t, \hat{\boldsymbol{X}}_{t-}, S_{t-})$, where

$$
\begin{aligned}
k(t, \boldsymbol{x}, s) = \int_{R-i\infty}^{R+i\infty} \frac{s^{z-1}}{\sigma^2(\boldsymbol{x})} \mathbf{1}^{\top} \boldsymbol{M}(T-t, z) \times \\
(z^2 \bar{c} + z\sqrt{\bar{c}} a(\boldsymbol{x}) + \int_{\mathbb{R}} (e^y - 1)(e^{zy}(1 + w(\boldsymbol{x}, y)) - 1) \boldsymbol{x}^{\top} \boldsymbol{F}(dy)) l(z) dz
\end{aligned}
$$

Proof. Computing an SDE for $dV(z)$, we get:

$$dV(z) = S_0^z \left[\mathbf{1}^\top \boldsymbol{M}(z)\hat{\boldsymbol{X}}_- d(e^{zY}) + e^{zY-}\mathbf{1}^\top \boldsymbol{M}(z)d\hat{\boldsymbol{X}} + \mathbf{1}^\top \boldsymbol{M}(z)d[e^{zY}, \hat{\boldsymbol{X}}] \right.$$

$$\left. + (...)d\boldsymbol{M} + (...)d[\boldsymbol{M}, Y] + (...)d[\boldsymbol{M}, \hat{\boldsymbol{X}}] \right]$$

$$= S_0^z \left[\mathbf{1}^\top \boldsymbol{M}(z)\hat{\boldsymbol{X}}_- e^{zY-}(zdY + \frac{1}{2}z^2 d\langle Y^c \rangle + (e^{zy} - 1 - zy)dJ^Y) \right.$$

$$+ e^{zY-}\mathbf{1}^\top \boldsymbol{M}(z)(\boldsymbol{A}\hat{\boldsymbol{X}}_- dt + w(\hat{\boldsymbol{X}}_-, y)(J^Y(dy, dt) - \hat{\boldsymbol{X}}_-^\top \boldsymbol{F}(dy)dt) + a(\hat{\boldsymbol{X}}_-)dI)$$

$$\left. + \mathbf{1}^\top \boldsymbol{M}(z)e^{zY-}(z\sqrt{\bar{c}}a(\hat{\boldsymbol{X}}_-))dt + (e^z y - 1)w(\hat{\boldsymbol{X}}_-, y)(J^Y(dy, dt) - \hat{\boldsymbol{X}}_-^\top \boldsymbol{F}(dy)dt)) \right]$$

Thus the quadratic covariation with the underlying is:

$$d[V(z), S] = S_-^{z+1}\mathbf{1}^\top \boldsymbol{M}(z)\left[z\sqrt{\bar{c}}(z\sqrt{\bar{c}} + a(\hat{\boldsymbol{X}}_-))dt \right.$$

$$\left. + (e^{zy} - 1 - zy + \text{diag}(\tilde{w}) + (e^{zy} - 1)\text{diag}(\tilde{w}))(e^y - 1)\hat{\boldsymbol{X}}_- dJ^Y \right]$$

$$= S_-^{z+1}\mathbf{1}^\top \boldsymbol{M}(z)\left[(z^2\bar{c} + z\sqrt{\bar{c}}a(\hat{\boldsymbol{X}}_-))dt \right.$$

$$\left. + (e^y - 1)(e^{zy}(1 + \text{diag}(\tilde{w})) - 1))\hat{\boldsymbol{X}}_- dJ^Y \right]$$

Together with the form of the quadratic variation of S from (3.3), we can use the above to compute

$$\xi_t = \int_{R-i\infty}^{R+i\infty} \frac{d\langle V(z), S \rangle_t}{d\langle S \rangle_t} l(z)dz$$

and get the stated result. □

We can also compute the value of the hedging strategy without the use of Fourier transforms, as long as the mean-value process V_t is sufficiently differentiable to apply Itô's lemma, i.e. $V_t = f(t, \hat{\boldsymbol{X}}_t, Y_t) \in \mathcal{C}^{1,2,2}$. In what follows we will denote the partial derivatives of f with subscripts, e.g. $f_x = \partial f / \partial x$, where x is multidimensional where appropriate and thus f_x may be a vector.

Theorem 3.13. *Let $V_t = f(t, \hat{\boldsymbol{X}}_t, Y_t)$ from corollary 3.11 be such that $f \in \mathcal{C}^{1,2,2}$. Then under the assumptions of theorem 3.6 and assumption 3.3, the mean-variance optimal hedging strategy is*

$$\xi_t = f^\xi(t-, \hat{\boldsymbol{X}}_{t-}, Y_{t-})$$

such that

$$f^\xi(t, \boldsymbol{x}, y) = \frac{f_x^\top a(\boldsymbol{x}) + f_y \bar{c} + \boldsymbol{x}^\top \int_{\mathbb{R}} (e^z - 1)(f(t, \boldsymbol{x} + w(\boldsymbol{x}, z), y + z) - f(t, \boldsymbol{x}, y + z))\boldsymbol{F}(dz)}{S_0 \exp(y)\sigma^2(\boldsymbol{x})}$$

Proof. We compute $\xi(t, \hat{\boldsymbol{X}}_{t-}, Y_{t-}) = \frac{d\langle V, S\rangle}{d\langle S\rangle}$ directly. Using Itô's lemma, we have

$$dV = f_t dt + f_x d\hat{\boldsymbol{X}}(h_1) + f_y dY(h_2) \tag{3.4}$$
$$+ \frac{1}{2}(f_{xx} d[\hat{\boldsymbol{X}}^c] + f_{yy} d[Y^c] + 2f_{xy} d[X^c, Y^c])$$
$$+ (f(t, \hat{\boldsymbol{X}}_- + x, Y_- + y) - f(t, \hat{\boldsymbol{X}}_-, Y_-) - f_x h_1(x) - f_y h_2(y)) dJ^{\hat{\boldsymbol{X}}, Y}$$

where h_1, h_2 are truncation functions. Equally,

$$dS = S_-(dY(h) + \frac{1}{2}\bar{c} dt + (e^y - 1 - y)dJ^Y)$$

The predictable quadratic variation of S is given as

$$d\langle S\rangle = S_-^2 \sigma^2(\hat{\boldsymbol{X}}_-)dt$$

The covariation of V and S is

$$d[V, S] = S_-\left[f_x d\langle \hat{\boldsymbol{X}}^c\rangle + f_y d\langle Y^c\rangle + \left((e^y - 1)(f(t, \hat{\boldsymbol{X}}_- + x, Y_- + y) - f(t, \hat{\boldsymbol{X}}_-, Y_-))\right)dJ^{\hat{\boldsymbol{X}}, Y}\right]$$
$$= S_-\left[f_x d\langle \hat{\boldsymbol{X}}^c\rangle + f_y d\langle Y^c\rangle + \left((e^y - 1)(f(t, \hat{\boldsymbol{X}}_- + w(\hat{\boldsymbol{X}}_-, y), Y_- + y) - f(t, \hat{\boldsymbol{X}}_-, Y_-))\right)dJ^Y\right]$$

Its $\hat{\mathcal{F}}$-predictable variant can be written as:

$$d\langle V, S\rangle = S_-\left[f_x a(\hat{\boldsymbol{X}}_-) + f_y \bar{c}\right. \tag{3.5}$$
$$\left. + \hat{\boldsymbol{X}}_-^\top \int_{\mathbb{R}} (e^y - 1)(f(t, \hat{\boldsymbol{X}}_- + w(\hat{\boldsymbol{X}}_-, y), Y_- + y) - f(t, \hat{\boldsymbol{X}}_-, Y_-))\boldsymbol{F}(dy)\right]dt$$

Then the mean-variance optimal hedging strategy is

$$\xi_t = f^\xi(t, \hat{\boldsymbol{X}}_-, Y_-) = \frac{d\langle V, S\rangle}{d\langle S\rangle}$$
$$= \frac{f_x a(\hat{\boldsymbol{X}}_-) + f_y \bar{c} + \hat{\boldsymbol{X}}_-^\top \int_{\mathbb{R}} (e^y - 1)(f(t, \hat{\boldsymbol{X}}_- + w(\hat{\boldsymbol{X}}_-, y), Y_- + y) - f(t, \hat{\boldsymbol{X}}_-, Y_-))\boldsymbol{F}(dy)}{S_- \sigma^2(\hat{\boldsymbol{X}}_-)}$$

$\square$

Remark 3.14. We can re-write $\hat{\boldsymbol{X}}_- + w(\hat{\boldsymbol{X}}_-, y)$ more explicitly as follows:

$$\hat{\boldsymbol{X}}_- + w(\hat{\boldsymbol{X}}_-, y) = \hat{\boldsymbol{X}}_- + \mathrm{diag}(\hat{\boldsymbol{X}}_-)\left(\frac{d\boldsymbol{F}(y)}{\hat{\boldsymbol{X}}_-^\top d\boldsymbol{F}(y)} - 1\right)$$
$$= \mathrm{diag}(\hat{\boldsymbol{X}}_-)\left(\frac{d\boldsymbol{F}(y)}{\hat{\boldsymbol{X}}_-^\top d\boldsymbol{F}(y)}\right)$$

3.6 Quadratic hedging error - a recursive approximation

The final element we need to have a complete picture about the variance optimal hedging strategy is the quadratic hedging error ε_0^2. We saw in the previous chapter that it simplifies significantly when returns are IID (Theorem 2.4) Furthermore, when the underlying is a martingale, the mean-variance tradeoff process $K \equiv 0$, and it simplifies further to the following expectation under physical measure (simplify [Černý and Kallsen, 2007, Theorem 4.12] using [Černý and Kallsen, 2007, Proposition 3.28]):

$$\varepsilon_0^2 := \mathbb{E}[(V_0 + \xi \cdot S_T - H)^2] = \mathbb{E}[\langle V, V\rangle_T - \xi^2 \cdot \langle S, S\rangle_T]$$

We will proceed with providing a solution via a recursive approximation scheme (with the sketch of a partial integral differential equation solution in the appendix). We can write the time t expected quadratic hedging error as a sum of infinitesimal single-step hedging errors $\psi(t, \hat{\boldsymbol{X}}_{t-}, S_{t-})$:

$$\varepsilon_t^2 = \mathbb{E}_t\left[\int_t^T \psi(u, \hat{\boldsymbol{X}}_{u-}, S_{u-})du\right]$$

where

$$\psi(t, \hat{\boldsymbol{X}}_{t-}, S_{t-}) := \frac{d\langle V\rangle_t}{dt} - \xi_{t-}^2 \frac{d\langle S\rangle_t}{dt} \tag{3.6}$$

Denoting ε_t^2 as a function of the state estimate and the underlying price, i.e. $\varepsilon_t^2 = g(t, \hat{\boldsymbol{X}}_t, S_t)$, we can define a discrete-time recursive numerical approximation scheme to compute it for the case when the Lévy process is a pure jump process.

121

Theorem 3.15. *Under martingale assumption 3.3, assuming $\boldsymbol{b} \equiv 0$ and assuming $V_t = f(t, \hat{\boldsymbol{X}}_t, Y_t) \in \mathcal{C}^{1,2,2}$, we can compute $\varepsilon_0^2 = g(0, \hat{\boldsymbol{X}}_0, S_0)$ using the backward recursion:*

$$g(T, \hat{\boldsymbol{X}}_T, S_T) = 0$$

$$g(t, \hat{\boldsymbol{X}}_t, S_t) = \int_{\mathbb{R}} \hat{\boldsymbol{X}}_t^\top \boldsymbol{p}(y) g(t + \Delta t, \hat{\boldsymbol{X}}_{t+\Delta t}(y), S_t \exp(y)) dy + \psi(t, \hat{\boldsymbol{X}}_t, S_t) \Delta t \quad (3.7)$$

$$\hat{\boldsymbol{X}}_{t+\Delta t}(\Delta Y) = \hat{\boldsymbol{X}}_t + \boldsymbol{A}\hat{\boldsymbol{X}}_t \Delta t + diag(\hat{\boldsymbol{X}}_t) \left(\frac{d\boldsymbol{F}(\Delta Y)}{\hat{\boldsymbol{X}}_t^\top d\boldsymbol{F}(\Delta Y)} - 1 \right) \Delta Y$$

$$- diag(\hat{\boldsymbol{X}}_t)(\boldsymbol{F}(\mathbb{R}) - \hat{\boldsymbol{X}}_t^\top \boldsymbol{F}(\mathbb{R})) \Delta t$$

$$(3.8)$$

Here we denote $\hat{\boldsymbol{X}}_{t+\Delta t}(\Delta Y)$ as a function of the realization of the jump ΔY, $\boldsymbol{p}(y)$ is the vector of probability density functions of the M Lévy distributions, and ψ is a function of the form

$$\psi(t, \boldsymbol{x}, s) = \bar{c}(f_y - k^2(t, \boldsymbol{x}, s)s^2) + f_x a^2(x)$$

$$+ \boldsymbol{x}^\top \int_{\mathbb{R}} \left[(f(t, diag(\boldsymbol{x}) \frac{d\boldsymbol{F}(y)}{\boldsymbol{x}^\top d\boldsymbol{F}(y)}, \log(s/S_0) + y) - f(t, \boldsymbol{x}, \log(s/S_0)))^2 \right.$$

$$\left. - k(t, \boldsymbol{x}, y)^2 s^2 (e^y - 1)^2 \right] \boldsymbol{F}(dy).$$

Proof. The recursion comes from the following integral approximation:

$$\varepsilon_t^2 = \mathbb{E}_t \left[\int_t^{t+\Delta t} \psi(u, \boldsymbol{X}_u, S_u) du + \int_{t+\Delta t}^T \psi(u, \boldsymbol{X}_u, S_u) du \right]$$

$$\approx \psi(t, \boldsymbol{X}_t, S_t) \Delta t + \mathbb{E}_t[\varepsilon_{t+\Delta t}^2]$$

$$= \psi(t, \boldsymbol{X}_t, S_t) \Delta t + \mathbb{E}_t[g(t + \Delta t, X_{t+\Delta t}, S_{t+\Delta t})]$$

In the recursive scheme, the time t expectation of g at time $t + \Delta t$ can be computed as an integral over the weighted average (weighted by posterior probability $\hat{\boldsymbol{X}}$) of probability density functions $p_k(y)$ of returns over a single time-step Δt under all the available states:

$$\mathbb{E}_t[g(t + \Delta t, \hat{\boldsymbol{X}}_{t+\Delta t}, S_{t+\Delta t})] = \int_{\mathbb{R}} \hat{\boldsymbol{X}}_t^\top \boldsymbol{p}(y) g(t + \Delta t, \hat{\boldsymbol{X}}_{t+\Delta t}(y), S_t \exp(y)) dy$$

To get the next value $\hat{\boldsymbol{X}}$ at timestep $t + \Delta t$ dependent on the next incoming jump of size ΔY, we discretize the SDE from theorem 3.6 without a diffusive component to get

$$\hat{\boldsymbol{X}}_{t+\Delta t}(\Delta Y) = \hat{\boldsymbol{X}}_t + \boldsymbol{A}\hat{\boldsymbol{X}}_t \Delta t + \text{diag}(\hat{\boldsymbol{X}}_t)\left(\frac{d\boldsymbol{F}(\Delta Y)}{\hat{\boldsymbol{X}}_t^\top d\boldsymbol{F}(\Delta Y)} - 1\right)\Delta Y$$
$$- \text{diag}(\hat{\boldsymbol{X}}_t)(\boldsymbol{F}(\mathbb{R}) - \hat{\boldsymbol{X}}_t^\top \boldsymbol{F}(\mathbb{R}))\Delta t$$

To compute $\psi(t, \hat{\boldsymbol{X}}_t, S_t)$ from equation (3.6) explicitly, we need quantities $\langle V, S \rangle$, $\langle S, S \rangle$ and $\langle V, V \rangle$. We already know $\langle V, S \rangle$ and $\langle S, S \rangle$ from our computations of the optimal hedging strategy (equations (3.5) and (3.3) respectively), so we will primarily focus on the computation of $\langle V, V \rangle$. Assuming $V_t = f(t, \hat{\boldsymbol{X}}_t, Y_t)$ is sufficiently regular to apply Itô's lemma, we can see directly from our previous computations of dV in equation (3.4) that

$$d\langle V \rangle / dt = f_x d\langle \hat{\boldsymbol{X}}_-^c \rangle / dt$$
$$+ f_y^2 \bar{c} + \hat{\boldsymbol{X}}_-^\top \int_{\mathbb{R}} (f(t, \text{diag}(\hat{\boldsymbol{X}}_-)\frac{d\boldsymbol{F}(y)}{\hat{\boldsymbol{X}}_-^\top d\boldsymbol{F}(y)}, Y_- + y) - f(t, \hat{\boldsymbol{X}}_-, Y_-))^2 \boldsymbol{F}(dy)$$

Recalling again that

$$d\langle S \rangle / dt = S_-^2 \sigma^2(\hat{\boldsymbol{X}}_-) = S_-^2 (\bar{c} + \hat{\boldsymbol{X}}_-^\top \int_{\mathbb{R}} (e^y - 1)^2 \boldsymbol{F}(dy)),$$

we can evaluate $\psi(t, \hat{\boldsymbol{X}}_{t-}, S_{t-})$ explicitly:

$$\psi(t, \hat{\boldsymbol{X}}_{t-}, S_{t-}) = \frac{d\langle V \rangle_t}{dt} - \xi_{t-}^2 \frac{d\langle S, S \rangle_t}{dt}$$
$$= \bar{c}(f_y - \xi_{t-}^2 S_-^2) + f_x \frac{d\langle \hat{\boldsymbol{X}}^c \rangle_t}{dt}$$
$$+ \hat{\boldsymbol{X}}_{t-}^\top \int_{\mathbb{R}} \left[(f(t, \text{diag}(\hat{\boldsymbol{X}}_{t-})\frac{d\boldsymbol{F}(y)}{\hat{\boldsymbol{X}}_{t-}^\top d\boldsymbol{F}(y)}, Y_{t-} + y) - f(t, \hat{\boldsymbol{X}}_{t-}, Y_{t-}))^2 \right.$$
$$\left. - \xi_{t-}^2 S_{t-}^2 (e^y - 1)^2 \right] \boldsymbol{F}(dy)$$

Written purely as a function of its inputs, we can re-write ψ:

$$\psi(t, \boldsymbol{x}, s) = \bar{c}(f_y - k^2(t, \boldsymbol{x}, s)s^2) + f_x a^2(x)$$

$$+ \boldsymbol{x}^\top \int_{\mathbb{R}} \left[(f(t, \operatorname{diag}(\boldsymbol{x}) \frac{d\boldsymbol{F}(y)}{\boldsymbol{x}^\top d\boldsymbol{F}(y)}, \log(s/S_0) + y) - f(t, \boldsymbol{x}, \log(s/S_0)))^2 \right.$$

$$\left. - k(t, \boldsymbol{x}, s)^2 s^2 (e^y - 1)^2 \right] \boldsymbol{F}(dy)$$

$\square$

Now we have everything needed to evaluate the hedging error numerically. In section 3.8, we will compute all the components of our regime-switching hedging framework (mean-value process, variance optimal hedge, quadratic hedging error) in a numerical example.

3.7 Filtered variance - relation to stochastic volatility and variance swaps

We now show a few results for the properties of the filtered variance of the regime-switching model. If we define $\sigma_k^2 := c^k + \int_R (e^x - 1)^2 F^k(dx)$ and define the vector $\boldsymbol{\sigma}^2 := [\sigma_1^2, ..., \sigma_M^2]^\top$, then we know that at any time, the instantaneous variance σ_t^2 of process S under filtration $\{\mathcal{G}_t\}$ is simply $\sigma_t^2 = (\boldsymbol{\sigma}^2)^\top \boldsymbol{X}_{t-}$. Therefore its dynamics are $d\sigma_t^2 = (\boldsymbol{\sigma}^2)^\top d\boldsymbol{X}_t$. Restricting ourselves to the filtration of observable returns $\{\hat{\mathcal{F}}_t\}$ (i.e. considering the filtered estimate of instantaneous variance), we find that $\mathbb{E}[d\sigma_t^2|\hat{\mathcal{F}}_{t-}] = (\boldsymbol{\sigma}^2)^\top d\hat{\boldsymbol{X}}_t$. We saw in theorem 3.6 that $\hat{\boldsymbol{X}}$ is of a dynamic nature and follows an SDE. Thus, a Hidden Markov Model will exhibit similar behaviour to a model of stochastic volatility where volatility is considered to be observable.

The resemblance to stochastic volatility models is further strengthened by the existence of a long-term average variance to which the model naturally mean-reverts. Specifically, from e.g. [Brémaud, 1999, Theorem 6.1] we know that non-degenerate Markov chains with transition probability matrix $\boldsymbol{A}$ have a long-term steady state $\boldsymbol{\pi} = \lim_{t\to\infty} \boldsymbol{X}_t$ which is readily available as a solution to the set of equations $\boldsymbol{\pi} \boldsymbol{A} = 0$ and $\boldsymbol{\pi}^\top \mathbf{1} = 1$. As a consequence, if we want to compute the long-run mean of the variance $\lim_{t\to\infty} \mathbb{E}[\sigma_t^2]$, then it is simply the corresponding weighted average of the

variances of the component Lévy processes:

$$\lim_{t\to\infty} \mathbb{E}[\sigma_t^2] = \boldsymbol{\pi}^\top \boldsymbol{\sigma}^2$$

Despite the stochastic nature of filtered volatility, it is straightforward to price a variance swap. Using our previous results from lemma 3.2, we can compute the formula for a variance swap implied by the model. Since the payoff of the continuously sampled variance swap is an expectation of integrated variance, in our model we can write:

$$\mathbb{E}\left[\int_0^T \sigma_t^2 dt\right] = \mathbb{E}\left[\int_0^T (\boldsymbol{\sigma}^2)^\top \boldsymbol{X}_{t-} dt\right] = (\boldsymbol{\sigma}^2)^\top \mathbb{E}[\boldsymbol{J}(0,T)] = \sum_k \mathbb{E}[J_k(0,T)]\sigma_k^2$$

where $J_k(t,T)$ is the occupation time of the Markov chain $\boldsymbol{X}$ in state k over the time interval $[t,T]$. In lemma 3.2 we got the moment-generating function of occupation times. This allows us to directly compute the mean vector of occupation times under filtration $\hat{\mathcal{F}}_t$:

$$\mathbb{E}[\boldsymbol{J}(t,T)|\hat{\mathcal{F}}_t] = \exp(\boldsymbol{A}(T-t))\hat{\boldsymbol{X}}_t(T-t)$$

This gives us that the expected integrated variance (and hence the fair strike of a variance swap) is

$$\mathbb{E}\left[\int_0^T \sigma_t^2 dt\right] = (\boldsymbol{\sigma}^2 T)^\top \exp(\boldsymbol{A}T)\hat{\boldsymbol{X}}_0$$

i.e. a weighted average of variances of the individual Lévy processes modulated by the transition probability between the various states. Let us note that the results above can be considered to be computed under a risk-neutral pricing measure (say defined by market prices of traded instruments), but we do not need to specify anything about it as we know that $\boldsymbol{X}$ and its dynamics are independent of the probability measure of S and the vol of S is independent of risk-neutral measure change.

In figure 3.1 we can observe the evolution of filtered volatility for 3 sample paths when our regime-switching model consists of two NIG models (high vol and low vol)with 2% probability of transferring between states at any given time. Our initial estimate for the probability of each state is $\hat{X}_0^k = 0.5$.
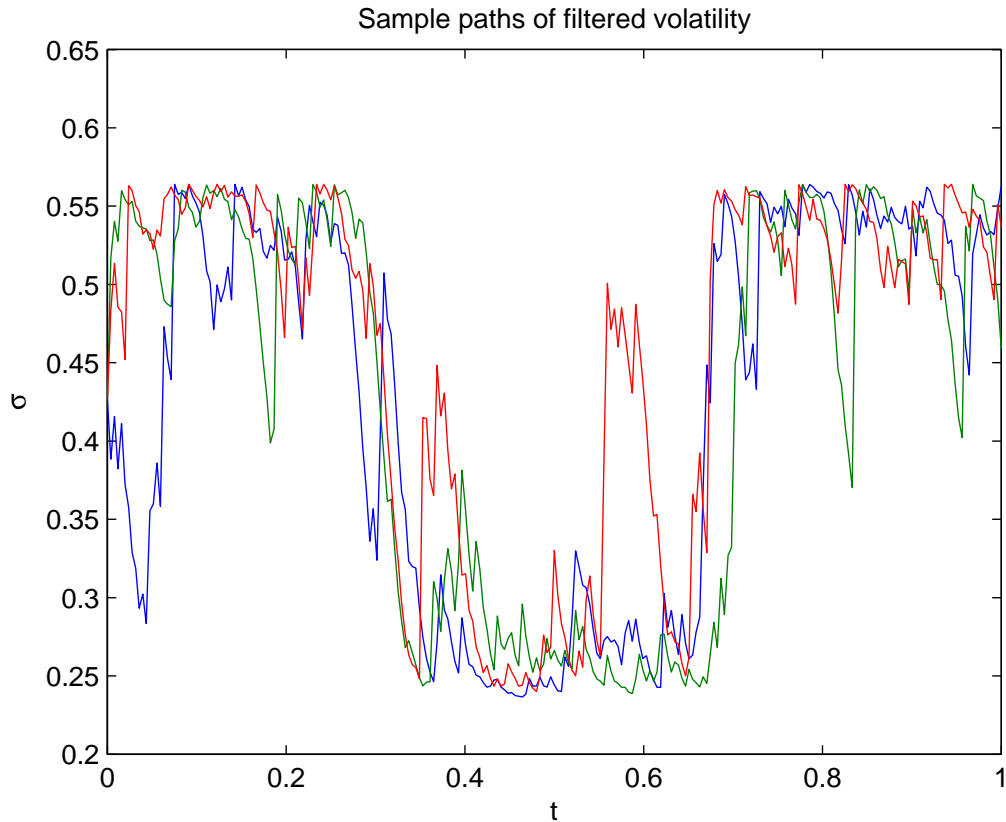
Figure 3.1: Evolution of filtered volatility in 2-state NIG model with volatilities $\sigma_1 = 20\%$, $\sigma_2 = 57\%$. 3 simulated paths.

3.8   Numerical results

In the following section we will implement the previously obtained formulae for the mean-value process, hedging strategy and hedging error numerically. First, we review results from the literature regarding calibration of parameters for regime-switching models.

3.8.1   Calibration and parameters - from the literature

In the literature there are only few examples of calibrated regime-switching Lévy processes, with implementations of Brownian regime-switching models being more common. However, both for the case of regime-switching Lévy processes and regime-switching pure Brownian motions, the number of states is typically limited to only 2 or 3, and they correspond to either states with varying volatility (high/medium/low

vol states) or with varying drift (bullish/bearish/stagnant market states). This is partly due to a practical limitation on the feasible number of states, as the number of parameters in the transition probability matrix to be calibrated grows quadratically with the number of states, leading to potentially unstable parameters and overfitting.

For purely Gaussian models, two-state models of returns have been fitted in Liew and Siu [2010], Hardy [2001], Goutte and Zou [2013], Di Graziano and Rogers [2009]. In a high-frequency setting with large amounts of data, Cartea and Jaimungal [2013] find the optimal number of regimes varies between 2 and 7 for a regime-switching Brownian motion calibrated to intraday single stock trading data.

For jump models, Elliott and Royal [2008] provide a calibration of a 2-state truncated variance-gamma process to the returns of the S&P 500 between 1986 and 2006 via the EM algorithm. They find evidence for the two states corresponding to a high and low volatility state. Chevallier and Goutte [2015, 2014] provide calibrations of a 2-state mean-reverting OU process driven by NIG returns to various global market indices. Hainaut and Courtois [2013] calibrates a mean-reverting 2-state regime-switching NIG process, VG process and a Brownian motion to 6-month CDS premiums when modelling credit default intensities. Haidinger and Warnung [2012] calibrate indices from multiple asset classes to a 2-state generalized hyperbolic distribution. Rogers and Zhang [2011] calibrate a 2-state VG and symmetric generalized hyperbolic distribution simultaneously to multiple stock indices, finding a good fit to most stylized facts of asset returns, including autocorrelation of absolute log returns (a feature typically captured by stochastic volatility models).

3.8.2   Mean-value process - numerical results

We will now illustrate the mean-value process and hedging strategy for a standard call option. We will consider a 2-regime model, switching between two martingale models with NIG jump measures with varying volatility. The first (low vol regime) has a volatility of $\sigma_1 = 20\%$, the other a significantly higher volatility $\sigma_2 = 57\%$. We recall that the NIG characteristic exponent is given as (following notation in Kienitz and Wetterau [2012])

$$\phi_{NIG}(z) = -\delta t(\sqrt{\alpha^2 - (\beta + iz)^2} - \sqrt{\alpha^2 - \beta^2})$$

and the Lévy density is

$$F(dy) = \frac{\delta\alpha}{\pi} \frac{\exp(\beta y) K_1(\alpha|y|)}{|y|} dy,$$

where $K_1(\cdot)$ is the modified Bessel function of the second kind. The low volatility martingale NIG model has parameters $\alpha = 75, \beta = -4.089, \delta = 3.024$, the high volatility NIG model has parameters $\alpha = 17.5, \beta = -10.089, \delta = 3.524$. We recall (see Kienitz and Wetterau [2012]) that the first four moments of the NIG distribution are given via those parameters as follows (with the abbreviation $\gamma = \sqrt{\alpha^2 - \beta^2}$):

$$\mathbb{E}[Y_t] = \delta\beta t/\gamma$$
$$Var[Y_t] = t\alpha^2\delta\gamma^{-3}$$
$$sk = \frac{3\beta}{\alpha\sqrt{\delta t\gamma}}$$
$$kurt = 3\left(1 + \frac{\alpha^2 + 4\beta^2}{\alpha^2 + \delta t\gamma}\right)$$

Note that we apply the martingale adjustment technique of Kienitz and Wetterau [2012] to correct the mean of the distribution and set it to zero (i.e. to be a martingale). We find our low-vol state has lower annualized skewness and kurtosis $Skew_{lowvol} = -0.01, Kurt_{lowvol} = 5.9$ compared to the high-vol state, for which the values are $Skew_{highvol} = -0.24, Kurt_{highvol} = 9$. In figure 3.2 we can see that the probability distributions these processes draw from are significantly different. We compute a regime-switching mean-value process where the two underlying models are those above. We set $\hat{X}_0 = [0.5, 0.5]^\top$, and choose the intensity matrix A to be

$$A = \begin{bmatrix} -0.0204 & 0.0204 \\ 0.0204 & -0.0204 \end{bmatrix}.$$

This corresponds to a 2% transition probability between regimes on any given day, and the average number of days within a regime before switching is $1/a_{12} \approx 50$. Its symmetrical nature implies that transitions from regime 1 to regime 2 are equally likely as transitions in the opposite direction. We note that this is not a requirement and real-world calibrations may result in asymmetrical transition matrices, where one state is more persistent than the other.

We consider a call option with a maturity of $T = 1$ year struck at $K = 100$ for a range of spot prices S_0. The result of the computation can be seen in figure 3.3.
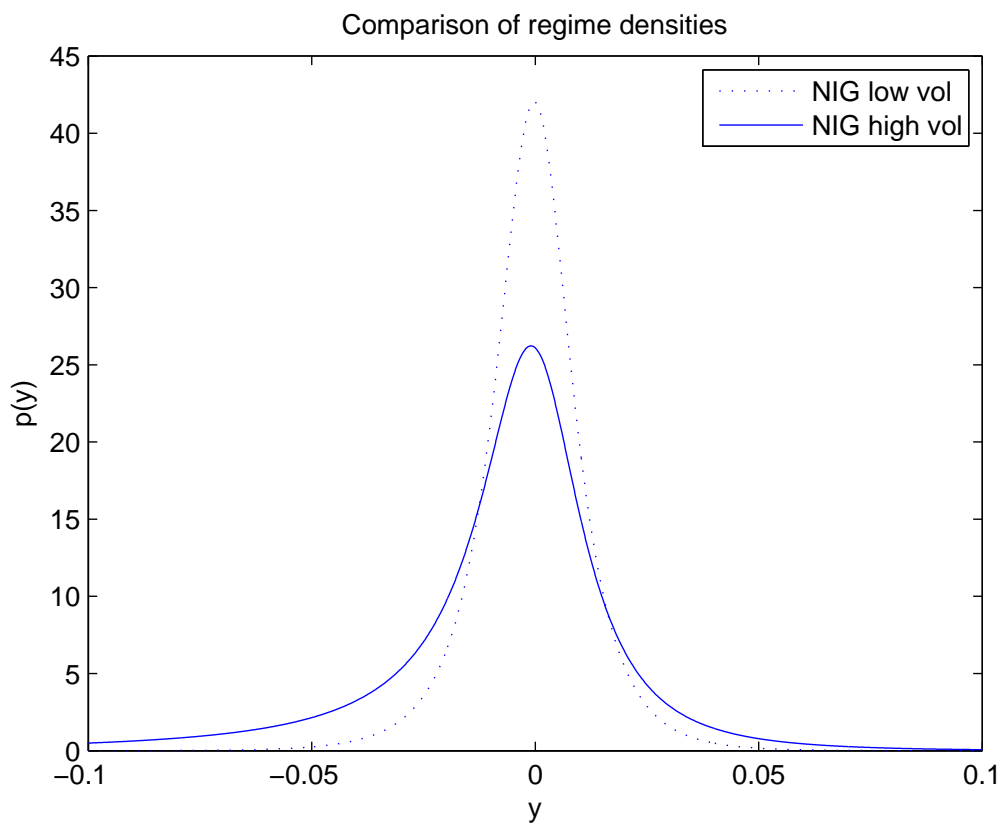
Figure 3.2: Comparisons of low and high vol NIG distribution used to generate daily log returns.
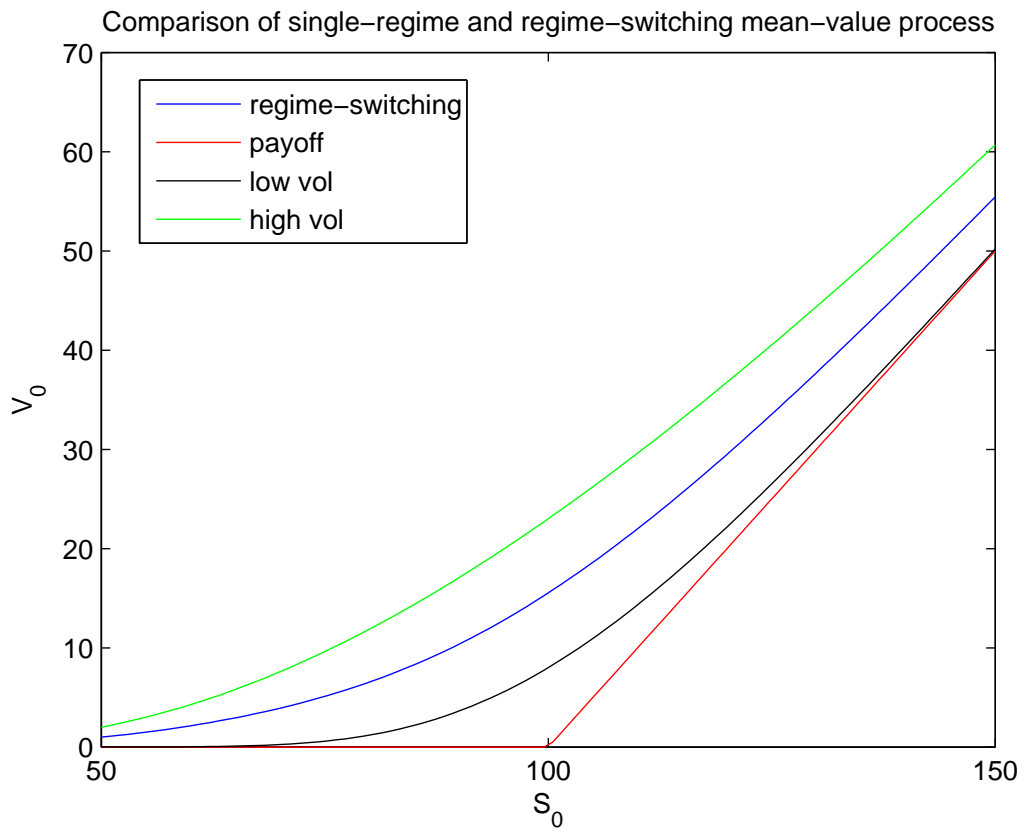
Figure 3.3: Regime switching model mean-value process vs mean-value processes from constituent low-vol ($\sigma_1 = 20\%$ and high-vol $\sigma_2 = 57\%$ models. $K = 100, T = 1$

We can look at the dependence of the mean-value process on the estimated probability of being in a particular state, i.e. how much variation there is between different states. In figure 3.4 we illustrate how the mean-value process of the ATM call option differs across time and different estimates $\hat{X}_t^1$ of being in the low volatility state. We can see that the variation is quite significant, and the dependence on $\hat{X}_t^1$ is (for any fixed time t) nearly linear. As expected, this difference in mean-value process across the state estimate value vanishes as we near maturity.
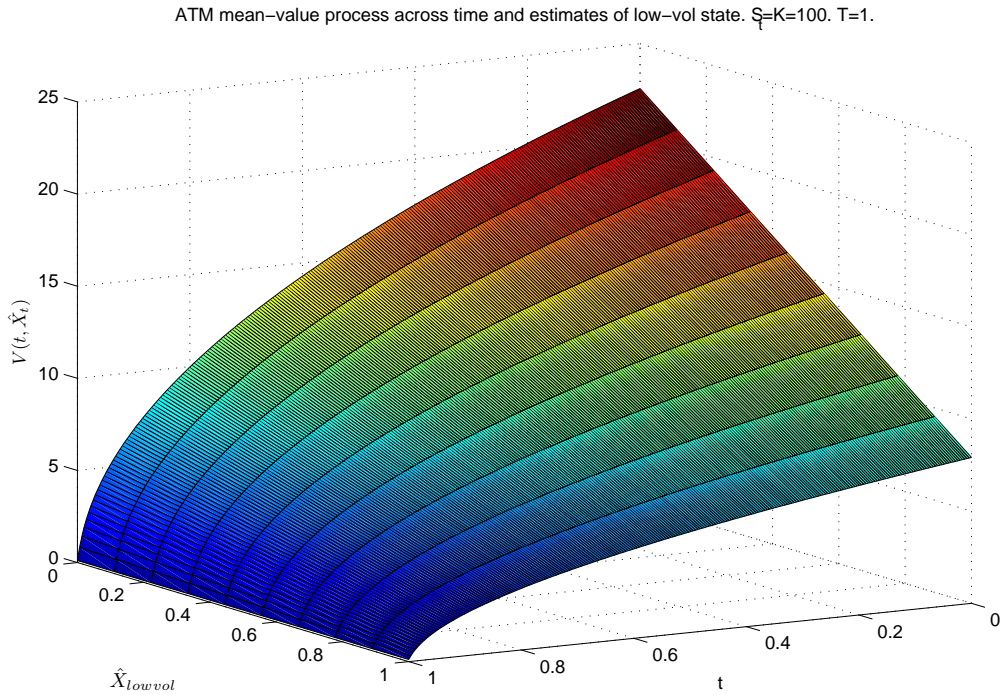
Figure 3.4: Regime-switching mean-value process for a range of times $t \in [0, T]$ and state estimates $\hat{X}_{lowvol} \in [0, 1]$.

In figure 3.5 we show the regime-switching mean-value process for a fixed maturity and varying estimates of the state $\hat{\boldsymbol{X}}$. In this figure we also compare the regime-switching mean-value process for a certain state $\hat{X}^1 \in \{0, 1\}$ against the mean-value process coming from a single-regime model, in effect illustrating the impact of the transition intensity matrix. As the two are very close to each other and hard to distinguish visually, we look closer at the exact size of these differences. In figure 3.6 plot the difference between the regime-switching mean-value process with a given state estimate $\hat{\boldsymbol{X}}$ and a weighted average of single-regime mean-value processes $\hat{X}_t^1 V_t^1 + (1 - \hat{X}_t^1) V_t^2$ for varying state estimates. We see that overall, the price difference

Figure 3.5: Regime-switching mean-value process for a range of state estimates $\hat{X}^1 \in \{0, 0.25, 0.5, 0.75, 1\}$, and single-regime mean-value processes for the low vol and high vol state. Maturity $T = 1$, option strike $K = 100$.
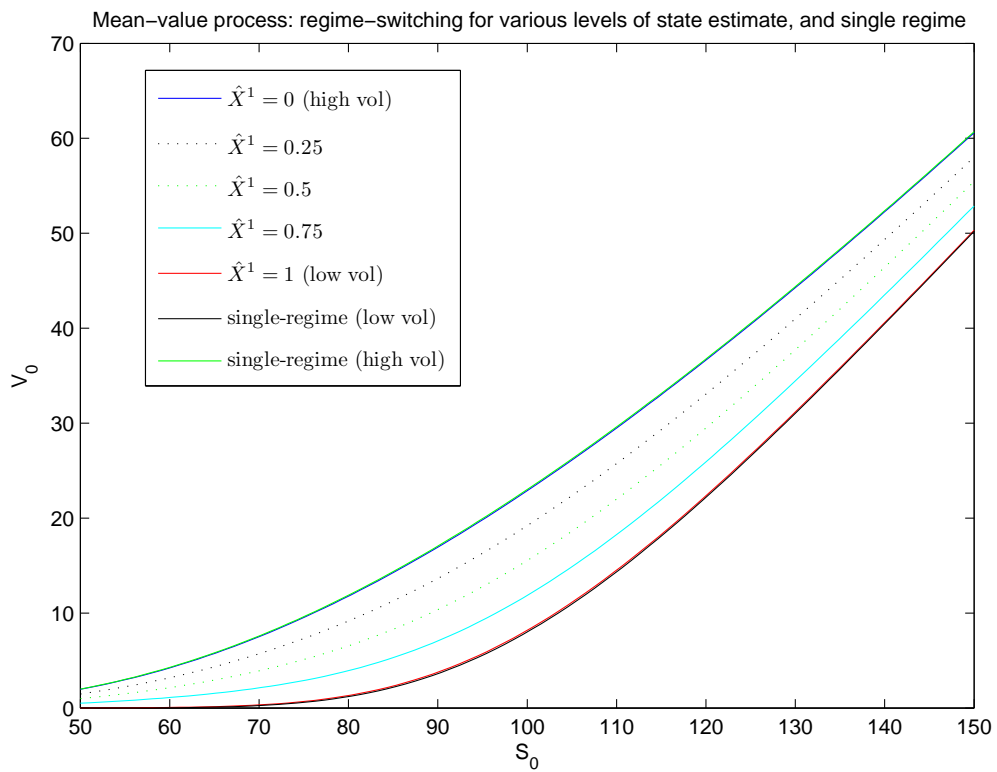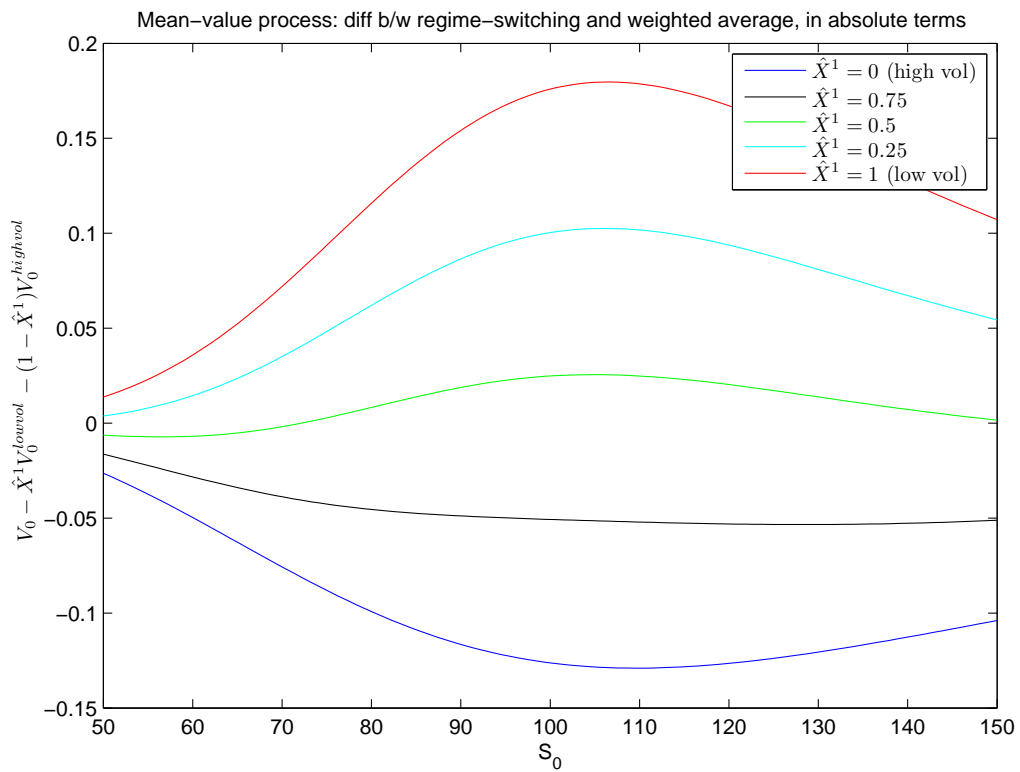
Figure 3.6: Difference between regime-switching mean-value process for a range of state estimates $\hat{X}^1 \in \{0, 0.25, 0.5, 0.75, 1\}$, and a weighted average of single-regime mean-value processes. Maturity $T = 1$, option strike $K = 100$.

is largest when the option is at-the-money and we are very certain about our state estimate. However, the absolute size of the difference is small, and is only significant in percentage terms in the area deep OTM, where the value of the contract is already very small.

3.8.3 Hedging strategy - numerical results

Next, we provide the hedging strategy for the regime-switching model with the parameters described above. Figure 3.7 illustrates the regime-switching hedging strategy and how it compares to the strategies implied by the constituent models. Just as with
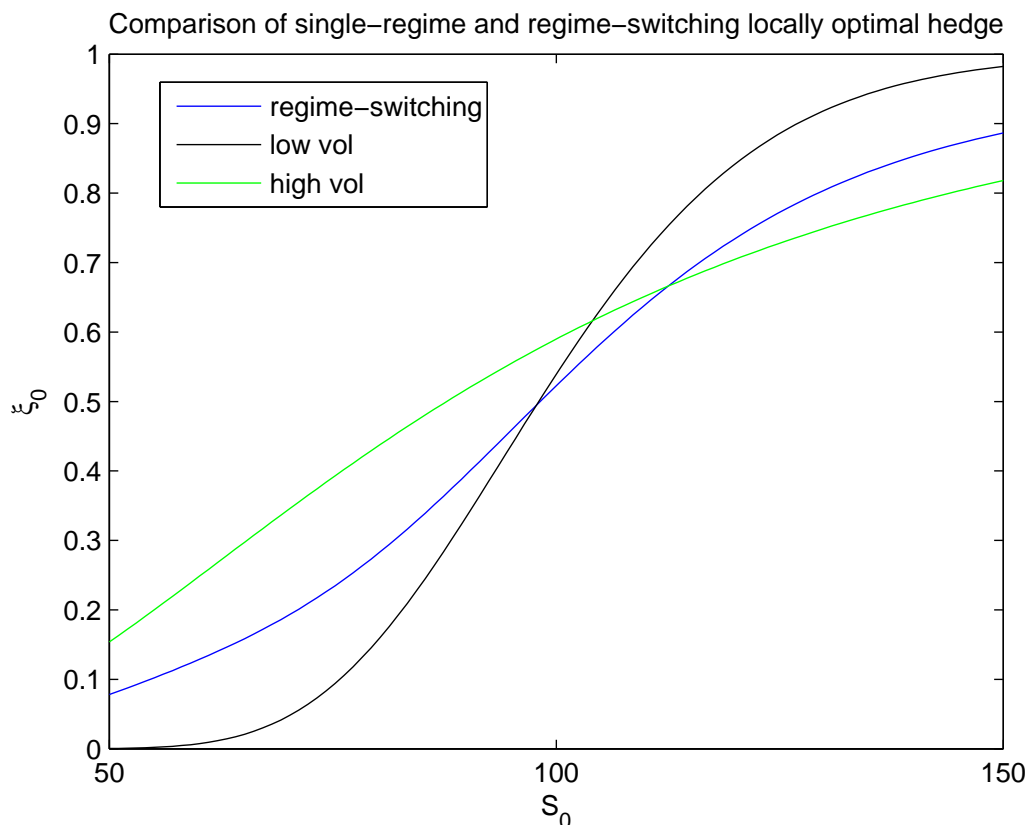
Figure 3.7: Regime-switching variance optimal hedge vs hedging strategies for constituent low-vol ($\sigma_1 = 20\%$ and high-vol $\sigma_2 = 57\%$ models. $K = 100, T = 1$.

the mean-value process, we look at the variation of the hedging strategy dependent on the estimated probability of being in a particular state. In figure 3.8 we illustrate how the hedging strategy of the ATM call option differs across time and different estimates $\hat{X}^1$ of being in state 1. We can see that the variation is quite significant, with

a rather non-linear dependence on the state estimate. Just as with the mean-value

Variance–optimal hedge at S=100 for varying time t and estimate of low–vol state. T=1, K=1.
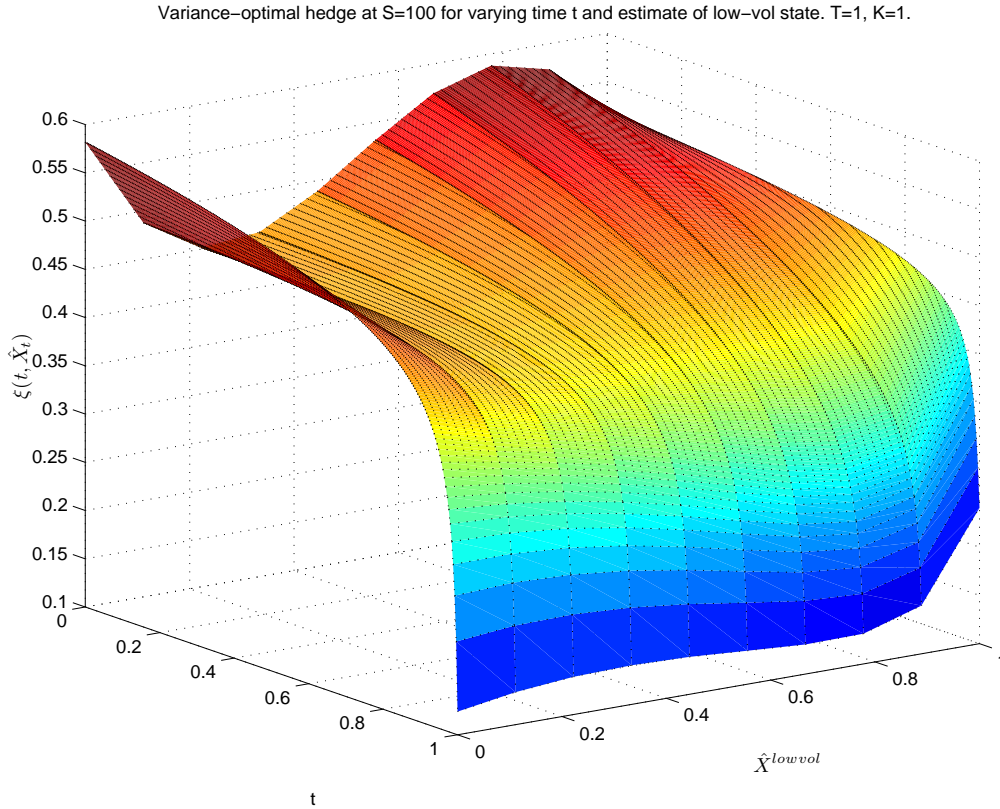
Figure 3.8: Regime-switching variance optimal hedging strategy at $K = S_0 = 100$ for a range of times $t \in [0, T]$ and state estimates $\hat{X}^1 \in [0, 1]$.

process, we can compare the hedging strategy for various levels of $\hat{\boldsymbol{X}}$ at a fixed maturity, and compare the limiting cases where $\hat{X}^1 \in \{0, 1\}$ with the single-regime cases. We provide such a comparison in figure 3.9. We again see that the limiting cases are quite close to the single-regime cases. In figure 3.10 we compare the regime-switching optimal hedge against a simple weighted average of single-regime optimal hedges. We can see that unlike the mean-value process, the difference between the weighted average and the regime-switching hedge can be relatively significant, especially for levels of positions that are slightly ITM, with the hedges being up to 5% different from each other. Again, in percentage terms (i.e. relative to the weighted average) the differences are more significant in the deep OTM level, where the absolute level of the hedge is quite small already.
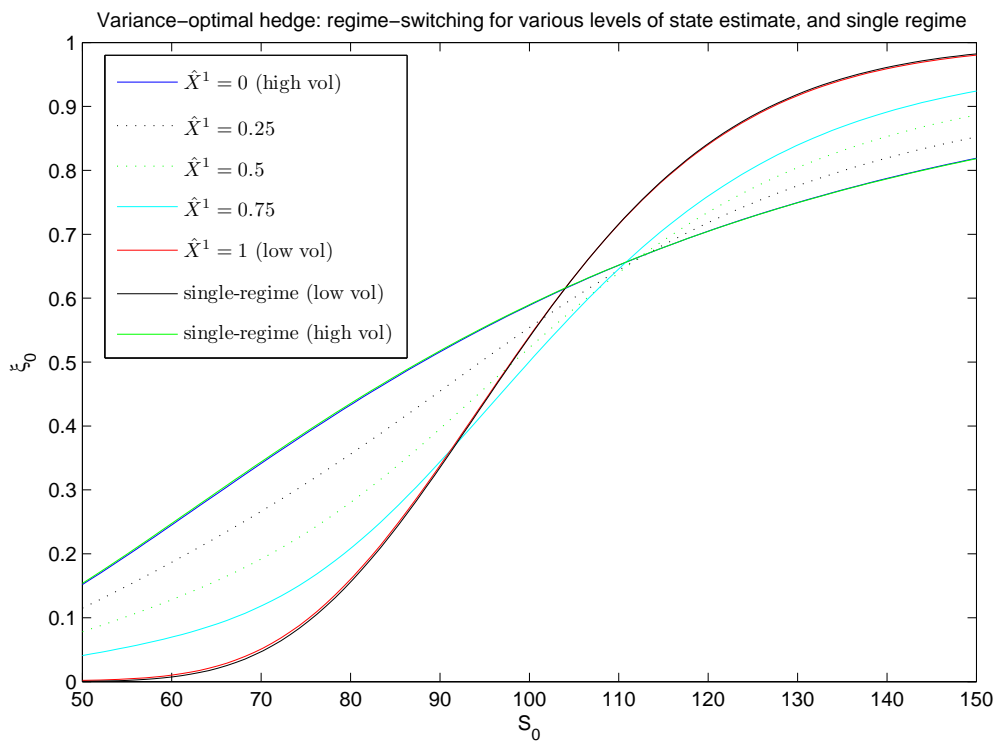
Figure 3.9: Regime-switching variance optimal hedge ξ for a range of state estimates $\hat{X}^1 \in \{0, 0.25, 0.5, 0.75, 1\}$, and single-regime variance optimal hedge ξ for the low vol and high vol state. Maturity $T = 1$, option strike $K = 100$.
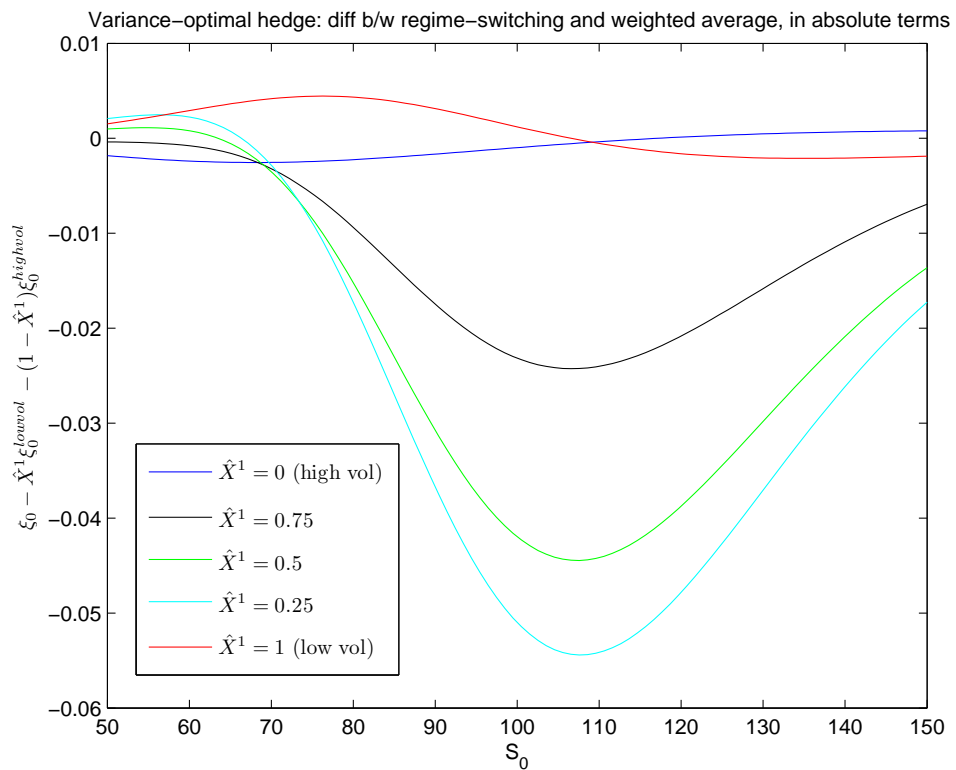
Figure 3.10: Difference between regime-switching variance optimal hedge for a range of state estimates $\hat{X}^1 \in \{0, 0.25, 0.5, 0.75, 1\}$, and a weighted average of single-regime variance optimal hedges. Maturity $T = 1$, option strike $K = 100$.

3.8.4 Expected hedging error - numerical results

Here we show the results we obtain for computing the hedging error via the recursive scheme (3.7). In figure 3.11 we show the error for different initial estimates of the
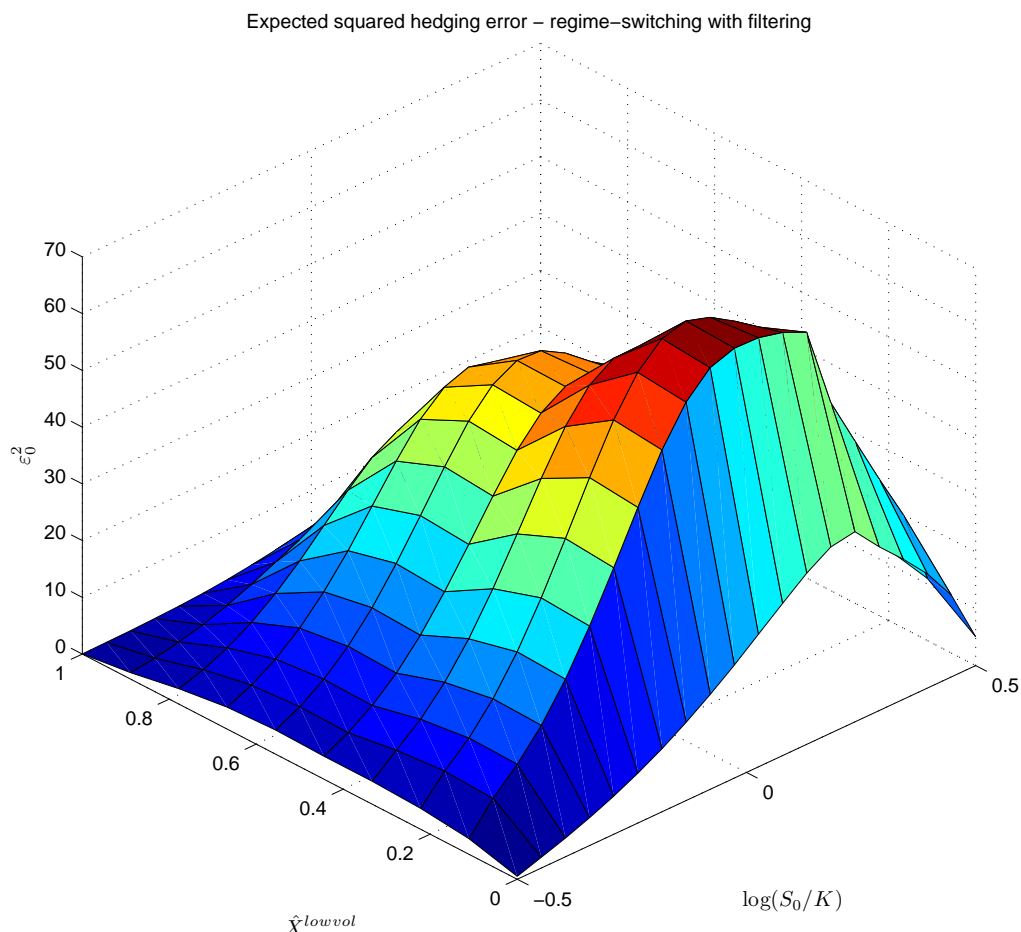
Figure 3.11: Regime switching model expected quadratic hedging error for constituent low-vol ($\sigma_1 = 20\%$) and high-vol ($\sigma_2 = 57\%$) models. Call option strike $K = 100$, across varying initial estimates of the low-vol state $\hat{X}_0^1$ and varying initial log-moneyness $\log\left(\frac{S_0}{K}\right)$.

low-volatility state $\hat{X}_0^1$ and different levels of initial log-moneyness $\log\left(\frac{S_0}{K}\right)$ (varying S_0, as the strike is fixed to $K = 100$). The peak hedging error is (predictably) at-the-money, where we run the greatest risk of having to frequently and significantly readjust our hedge, especially when nearing maturity. The highest value we observe is $\max \varepsilon_0^2 = 67.76$, implying the highest mean square root error (MSRE) $\varepsilon_0 = \$8.23$.

We note that the estimated squared hedging error not only increases as we increase our estimated probability of the high-volatility state, but also with the increasing uncertainty of our current regime estimate - the peak is achieved when we have a certain degree of certainty ($\hat{X}_0^{lowvol} \in [0.2, 0.4]$), i.e. we are in the high-volatility state but are not completely certain of it. Thus the volatility of our filtered estimate contributes significantly to the expected hedging error.

3.8.4.1   A note on numerical precision

We note that due to the format of the equation for the hedging error ε_0^2, there is little point in going very far in terms of log-moneyness, as numerical precision becomes an issue. Consider that the format of the single step hedge error is:

$$\psi(t, \hat{\boldsymbol{X}}_{t-}, S_{t-}) = (d\langle V\rangle_t - \xi_{t-}^2 d\langle S\rangle_t)/dt$$

where quadratic variation of S will be of order $\mathcal{O}(S^2)$. Let our Fourier transform scheme for ξ have an error of ϵ. To the first order of error, $(\xi_{t-} + \epsilon)^2 = \xi_{t-}^2 + 2\epsilon + o(\epsilon)$. We need to ensure that the difference of the quadratic variation of the mean-value process and the scaled quadratic variation of the underlying is positive. Thus

$$(d\langle V\rangle - (\xi_-^2 + 2\epsilon)d\langle S\rangle)/dt > 0$$

This only holds when the hedging error is large enough relative to the scale of the price:

$$\psi(t, \hat{\boldsymbol{X}}_{t-}, S_{t-}) > 2\epsilon d\langle S\rangle_t/dt$$

We know $d\langle S\rangle/dt = S_-^2 \sigma^2(\hat{\boldsymbol{X}}_-)$, with the scale of the number being determined by the leading factor S_-^2:

$$\psi(t, \hat{\boldsymbol{X}}_{t-}, S_{t-}) > 2\epsilon S_{t-}^2 \sigma^2(\hat{\boldsymbol{X}}_{t-})$$

This is a problem for deep ITM/OTM options where the single-step hedging error is quite small. E.g. If our underlying has a level of $S_{t-} = 100$ and the true single-step hedging error ψ deep ITM/OTM has a level of 10^{-2}, we will have to compute our hedging strategy ξ_t to a precision of at least $\epsilon = 10^{-7}$. This issue is amplified further for underlyings with large absolute values, e.g. with the S&P 500 at $S_{t-} = 2000$, we would have to get our hedging strategy to a precision of $\epsilon = 10^{-9}$ deep ITM/OTM. As a consequence we do not investigate hedging errors very far away

from the strike. This is warranted as most options trade within a small range of the ATM level (with moneyness levels greater than 110% or smaller than 90% considered to be deep ITM/OTM). For example for a different problem involving numerical solutions of PIDEs, Cont and Voltchkova [2005] limit their range of log-moneyness to the interval $[\log(2/3), \log(2)] \approx [-0.18, 0.3]$, as this range already covers most market-traded options. From a practitioner's perspective, a majority of trades will be struck at the current spot or forward level at inception, thus the ATM expected hedging error is also the most relevant for practical purposes.

3.8.5 Realized hedging error via simulations

We now investigate the economic impact of using a regime-switching strategy instead of a Black-Scholes delta hedge or a variance optimal strategy with just one underlying Lévy process. Specifically, we will look at the variance of the terminal hedging error when run on simulated paths.

As a simple example, we simulate paths with log-returns drawn from the two log-return distributions compared in figure 3.2. We generate 1000 paths and a simulated realized state process for $T = 1$ business year, with daily timesteps $\Delta t = 1/252$, as shown in figure 3.12. As we assume our underlying is a martingale, both our distributions of returns have zero mean. At the start, we consider both models to be equally probable (i.e. $\hat{X}_0^k = 0.5, k = 1, 2$). We propagate the filter $\hat{\boldsymbol{X}}_t$ forward with the intensity matrix $\boldsymbol{A}$ specified previously. For one of the sample paths, we show the evolution of the filtered state $\hat{X}_t$, as shown in figure 3.13. We can see that the filter quickly adjusts and distinguishes with a reasonable degree of accuracy which model is generating returns. We note that in order to be able to identify the state sufficiently quickly, the two assumed states must be sufficiently different from each other, otherwise the filter will not gain strong certainty on the current state.

We now run portfolio simulations. At the start of the simulation, we buy 1 unit of an ATM call option with expiry $T = 1$ year, and hedge it by dynamically selling short Δ units of stock, rebalancing daily (i.e. using 252 timesteps) without any transaction costs. For simplicity, we assume lending is interest-free and equally excess cash earns no interest. We consider three cases:

1. the trader books the option value via the Black-Scholes formula ($\sigma = 30\% =$
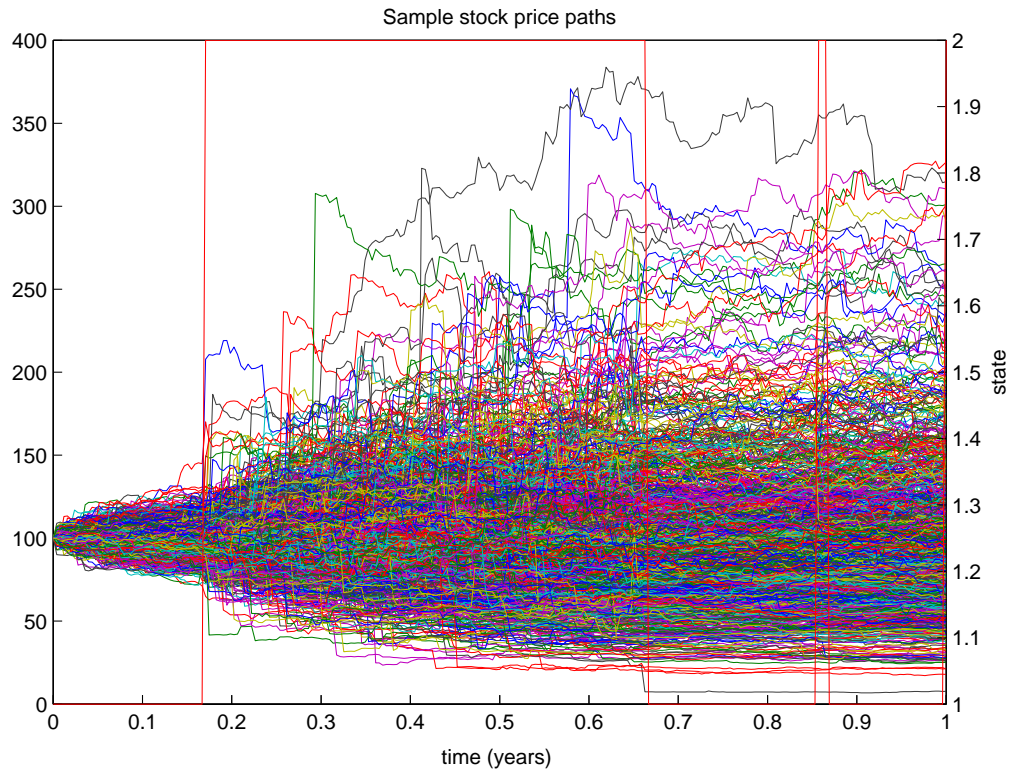
Figure 3.12: Sample stock price paths for a 2-state NIG regime-switching model with high ($\sigma = 57\%$) and low ($\sigma = 20\%$) volatility.
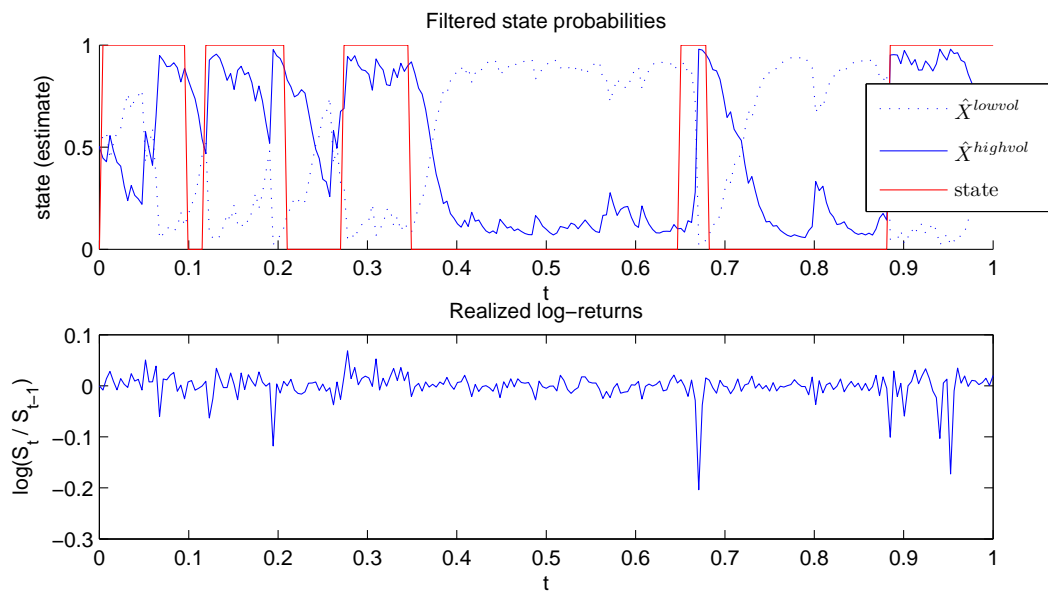
Figure 3.13: Filtered state $\hat{\boldsymbol{X}}_t$ and the corresponding sample log-return path. $\hat{\boldsymbol{X}}_0 = [0.5, 0.5]^\top$.

$\sqrt{0.5^2\sigma_{lowvol}^2 + 0.5^2\sigma_{highvol}^2}$) and uses the Black-Scholes Delta (at that volatility) to dynamically hedge throughout.

2. the trader books the option value as the initial average of the NIG-based mean-value processes under the two regimes and hedges via the same average of the two variance optimal strategies throughout.

3. the trader books the option value as the regime-switching (between normal and NIG returns) mean-value process and hedges via the regime-switching optimal strategy (which continually updates based on the value of filtered state $\hat{X}$).

The distribution of the terminal P&L of trading on 1000 simulated stock paths based on the three different hedging strategies can be seen in figure 3.14. From the figure we
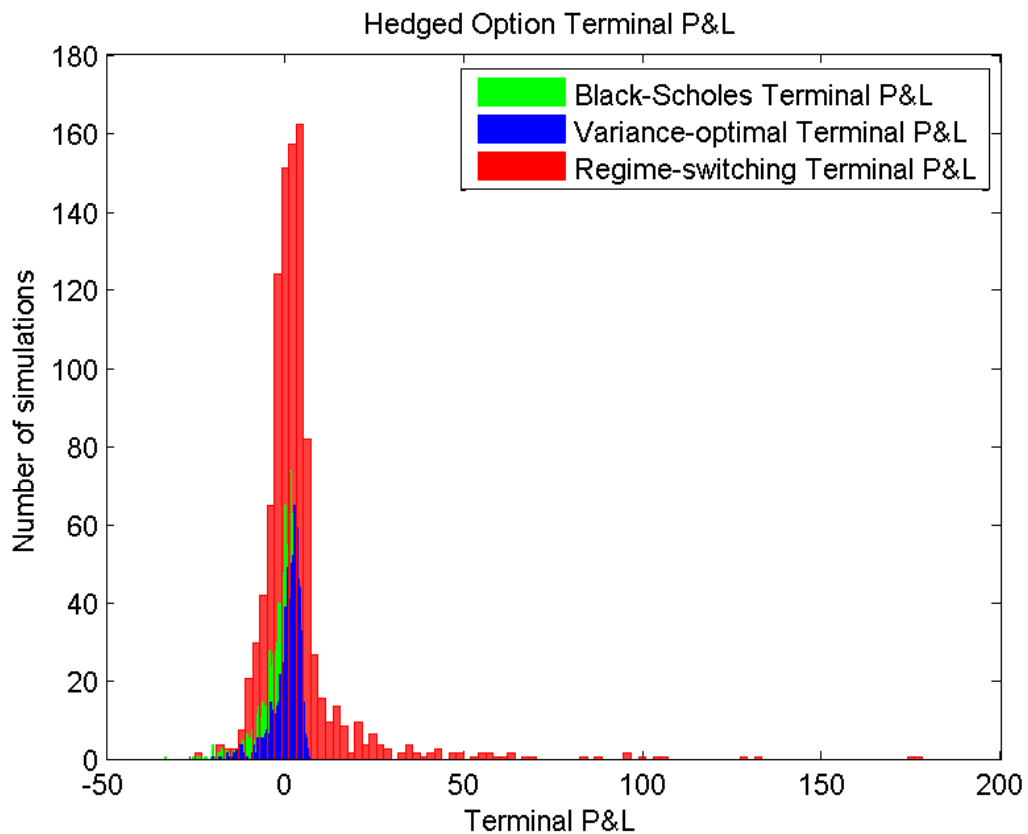
Figure 3.14: Empirical distribution of terminal P&L when following three different hedging strategies. 1000 simulated paths.

can see that the regime-switching hedging strategy fares the best, the plain variance-optimal strategy coming in second and the Black-Scholes strategy faring the worst,

142

exhibiting a somewhat fatter tail than the previous two cases. We see that the regime-switching strategy has a significant edge over regular variance-optimal hedging with a fixed weighted average. Unlike the other two strategies, its distribution actually appears to be positively skewed, whereas the other two are distinctly negatively skewed. The adaptive filter that can identify the regime we are experiencing allows the trader to anticipate periods of volatility shifts and thus appropriately inflating/deflating his hedge in anticipation of larger/smaller movements in the spot.

3.9 Comparison to alternative hedging strategies

Just like the *realized* hedging error, the mean-value process, variance optimal hedging strategy and the corresponding *expected* hedging error of our regime-switching model can be compared to several alternative hedging strategies and corresponding hedging errors. These allow us to identify the changes to the mean-value process/hedge and the additional error or improvement our complex model brings about compared to simpler models. We will compare against the following:

1. The variance optimal strategy and error under a regime-switching model with the assumption that the current state $\boldsymbol{X}$ is observable, i.e. under the global filtration $\{\mathcal{G}_t\}$;

2. The variance optimal strategy under a regime-switching model where we do not attempt to filter and improve our initial regime estimate $\hat{\boldsymbol{X}}_0$, using it throughout for pricing and hedging;

3. The variance optimal strategy and error of a mixture model (i.e. does not assume any regime-switching behaviour) where our initial estimate $\hat{\boldsymbol{X}}_0$ determines the constant proportions with which the two Lévy models contribute to the returns process;

4. The approximation of the hedging error via the asymptotic Cash Gamma squared estimate discussed in Chapter 1, estimated via the volatility and kurtosis of our underlying.

We will find most of the settings listed above are a special limiting case of the more generic regime-switching model.

3.9.1 Observable regime case

The case of observable $\boldsymbol{X}$ corresponds to a situation when our estimate is absolutely certain, i.e. $\hat{X}^1 \in \{0, 1\}$, allowing us to re-use our computations from the hidden state case. The **mean-value process and hedging strategy** are readily read off from charts 3.4 and 3.8. For the mean-value process, this is because it only depends on the expected occupation time of $\boldsymbol{X}$ in each regime, which does not change whether we observe the process or not. For the hedging strategy, if X^1 is constantly either 1 or 0, the filtering equation jump weighting processes $w(\boldsymbol{x}, y) = a(\boldsymbol{x}) \equiv 0$ and the dynamics of the filtered state estimate become the same as that of the state itself.

We expect the **hedging error** to be significantly smaller than our baseline as we observe more information about the model and can therefore hedge more exactly. To compute the hedging error, we must make a minor adjustment to our computations, as the quadratic variation of $\boldsymbol{X}$ is different to that of $\hat{\boldsymbol{X}}$. Specifically, [Cohen and Elliott, 2008, Section 2] show that

$$d\langle \boldsymbol{X} \rangle_t / dt = [\mathrm{diag}(\boldsymbol{A}\boldsymbol{X}_{t-}) - \mathrm{diag}(\boldsymbol{X}_{t-})\boldsymbol{A}^\top - \boldsymbol{A}\mathrm{diag}(\boldsymbol{X}_{t-})].$$

This leads to a modified quadratic variation of the mean-value process:

$$
\begin{aligned}
d\langle V \rangle_t / dt &= f_x(t, \boldsymbol{X}_{t-}, Y_{t-})(\mathrm{diag}(\boldsymbol{A}\boldsymbol{X}_{t-}) - \mathrm{diag}(\boldsymbol{X}_{t-})\boldsymbol{A}^\top - \boldsymbol{A}\mathrm{diag}(\boldsymbol{X}_{t-})) \\
&\quad + f_y \bar{c} + \boldsymbol{X}_{t-}^\top \int_{\mathbb{R}} (f(t, \boldsymbol{X}_{t-}, Y_{t-} + y) - f(t, \boldsymbol{X}_{t-}, Y_{t-}))^2 \boldsymbol{F}(dy)
\end{aligned}
$$

This in turn affects the values of $\psi(t, \boldsymbol{X}_{t-}, S_{t-})$ in the hedging error recursion. However, the general structure of the recursive scheme remains very similar, with a minor change to how we compute expectations of g for the next step:

$$\mathbb{E}_t[g(t + \Delta t, \boldsymbol{X}_{t+\Delta t}, S_{t+\Delta t})] = \sum_{j=1}^{M} X_t^j \sum_{k=1}^{M} P_{kj} \int_{\mathbb{R}} p_k(y) g(t + \Delta t, \boldsymbol{e}_k, S_t \exp(y)) dy$$

Here $P_{kj} = \mathbb{P}(\boldsymbol{X}_{t+\Delta t} = \boldsymbol{e}_k | \boldsymbol{X}_t = \boldsymbol{e}_j)$ is retrieved from the intensity matrix $\boldsymbol{A}$: $\boldsymbol{P} = \exp(\boldsymbol{A}\Delta t)$ and $p_k(y)$ is the probability density function of the k-th log-return distribution. We can see the result we get for our numerical setting in figure 3.15. We see that as we have perfect information about current and future state of the regime, there is only a limited difference between the expected hedging error over 1 year when we start off in a high-vol or low-vol regime, as we can always adapt to the

current regime perfectly.
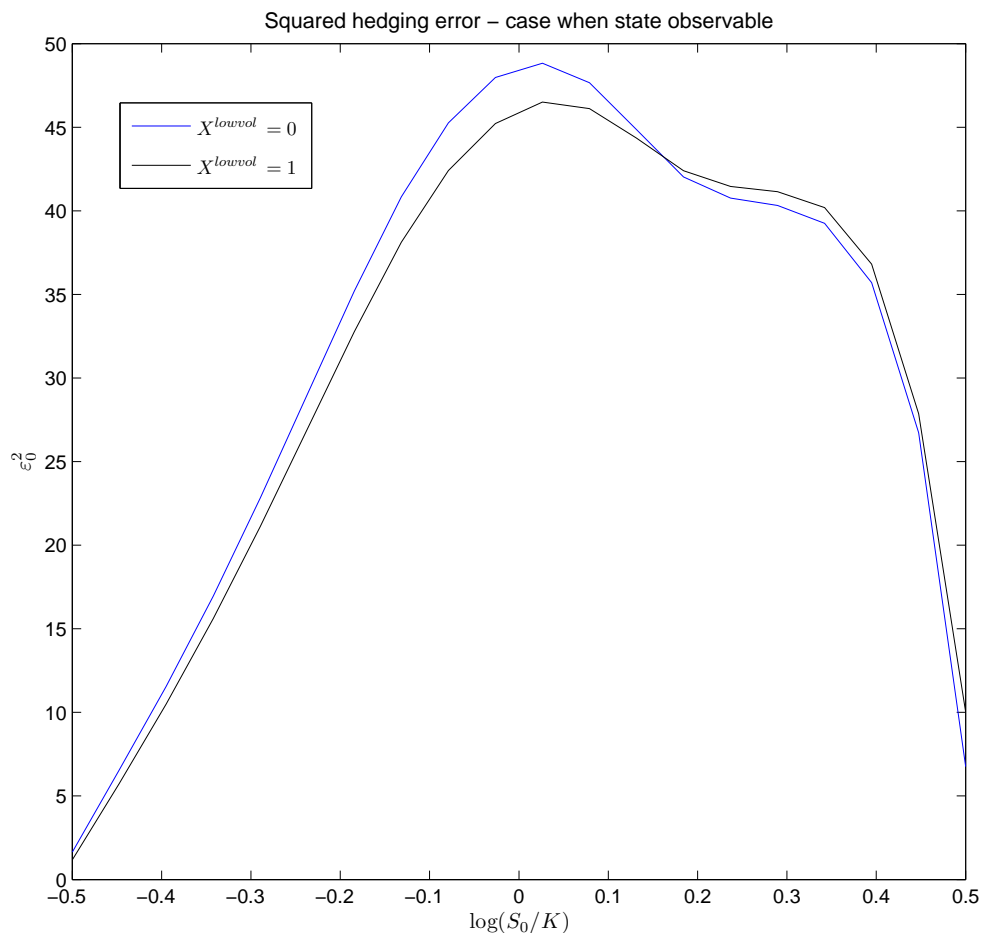
Figure 3.15: Regime switching model expected quadratic hedging error for constituent low-vol ($\sigma_1 = 20\%$) and high-vol ($\sigma_2 = 57\%$) models when $\boldsymbol{X}$ observable. Call option strike $K = 100$, across varying initial estimates of the low-vol state $\hat{X}_0^1$ and varying initial log-moneyness $\log\left(\frac{S_0}{K}\right)$.

3.9.2 Regime-switching model with unobservable state but without filtering

We can alternatively put ourselves into a position where we acknowledge that the returns process undergoes multiple regime shifts throughout the lifetime of the contingent claim, but we do not attempt to estimate the current regime from observations

of the returns process. There are multiple cases when this can be a legitimate stance to take: for example, we may believe that the change between economic regimes is rather subtle and difficult to detect using noisy returns; it may be the case we believe the regimes change very frequently, and we are not able to detect the change in time before another change happens.

In this setup, we consider the true underlying model to still be a regime-switching model and we simply do not update our estimate of $\boldsymbol{X}$. We can re-use the results we have computed for the **mean-value process** from the general case, as the only difference will be that the estimate $\hat{\boldsymbol{X}}_t$ will be replaced with $\hat{\boldsymbol{X}}_0$. In terms of our numerical results this means that from our 2D grid $(\hat{X}^1, Y)$ in figure 3.4, we will consider only a slice across different values of Y for one fixed value $\hat{X}^1 = \hat{X}_0^1$.

The **hedging strategy** will subtly change, as it shall now become independent of filtering variables, losing dependence on the SDE for $\hat{\boldsymbol{X}}$ (and its filtering functions $a(\boldsymbol{x})$ and $w(\boldsymbol{x}, y)$). In such a setting the quadratic covariation between V and S under filtration $\hat{\mathcal{F}}_t$ becomes

$$d\langle V, S \rangle / dt = S_- \left(f_y \bar{c} + \hat{\boldsymbol{X}}_0 \int_{\mathbb{R}} (e^y - 1)(f(t, \hat{\boldsymbol{X}}_0, Y_- + y) - f(t, \hat{\boldsymbol{X}}_0, Y_-)) \boldsymbol{F}(dy) \right)$$

and the quadratic covariation within the Fourier transform becomes

$$d\langle V(z), S \rangle / dt = S_-^{z+1} \boldsymbol{1}^\top \boldsymbol{M}(z, T - t) \left(z^2 \bar{c} + \hat{\boldsymbol{X}}_0 \int_R (e^y - 1)(e^{zy} - 1) \hat{\boldsymbol{X}}_0^\top \boldsymbol{F}(dy) \right)$$

The quadratic variation of $V_t = f(t, \hat{\boldsymbol{X}}_0, Y_t)$ will be

$$d\langle V \rangle_t / dt = f_y^2 \bar{c} + \hat{\boldsymbol{X}}_0^\top \int_{\mathbb{R}} (f(t, \hat{\boldsymbol{X}}_0, Y_{t-} + y) - f(t, \hat{\boldsymbol{X}}_0, Y_{t-}))^2 \boldsymbol{F}(dy)$$

Although our initial pricing will lead to the same price as in the setting where we do filter the current state, we expect our hedging error to be larger, as the underlying model is the same but our hedging strategy will presumably be suboptimal more often, as we do not improve the estimate of the hidden state. We can see the resultant hedging error in figure 3.16.

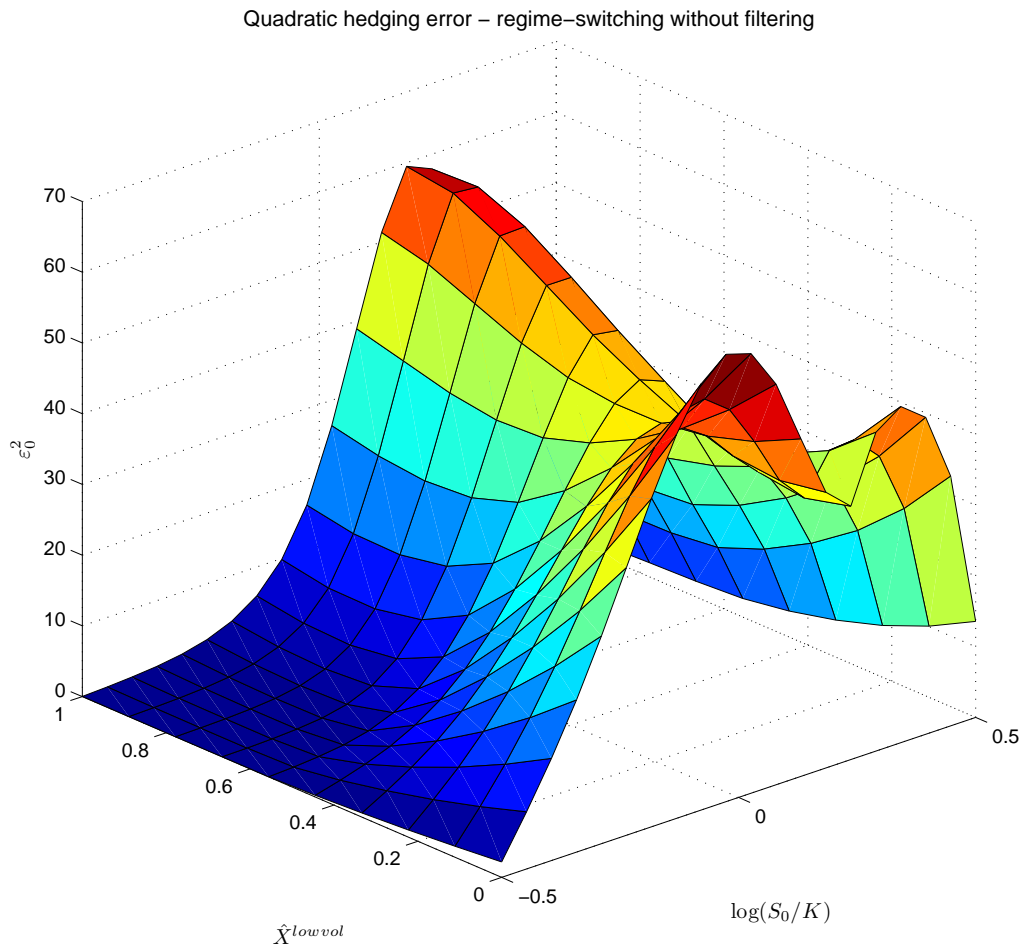Figure 3.16: Expected quadratic hedging error in regime-switching model without filtering for constituent low-vol ($\sigma_1 = 20\%$) and high-vol ($\sigma_2 = 57\%$) models. Call option strike $K = 100$, across varying initial estimates of the low-vol state $\hat{X}_0^1$ and varying initial log-moneyness $\log\left(\frac{S_0}{K}\right)$.

3.9.3 Lévy mixture model

We can consider the classic hedging strategy for either the individual constituent Lévy processes, or for a fixed mixture model - a linear combination of the constituent Lévy processes. In both cases we simply re-use the classical quadratic hedging results of Černý [2007]. A mixture model can also be seen as a particular case of the regime-switching model when the transition matrix $\boldsymbol{A} \equiv \boldsymbol{0}$ and thus we assume no switching happens.

We have already compared the **mean-value process** and **hedging strategy** of the regime-switching model with those of the component models in figures 3.4,3.6, and 3.9, 3.10 respectively. Assuming the two component Lévy processes are independent, the case of the fixed mixture model is simply a weighted average of the two, both for the mean-value process and the variance-optimal hedging strategy.

Regarding the **hedging error**, we could use the Fourier transform results of Černý [2007], but for consistency of numerical stability and precision, we instead use the same backward-iterative scheme as we do in the regime-switching case. With the exception of the mean-value process itself, the formulas for computing the hedging error via the iterative scheme are identical to the case of a regime-switching model where we do not filter the state.

Relative to the regime-switching model, we expect the anticipated hedging error to be smaller, as the model itself never counts on us being caught out by some sudden change in the market.

We can see the resultant hedging error in figure 3.17. Indeed, we can confirm that the mixture model reports a significantly lower expected squared hedging error.

3.9.4 Approximation of the hedging error via the Cash Gamma

Finally, we can consider using a simplified approximation to estimate the quadratic hedging error and see good an estimate this much simpler approach is. We established in the first chapter that the variance optimal strategy relates to the Cash Gamma of the derivative. We can use the approximate discrete-time formulas from [Černý,
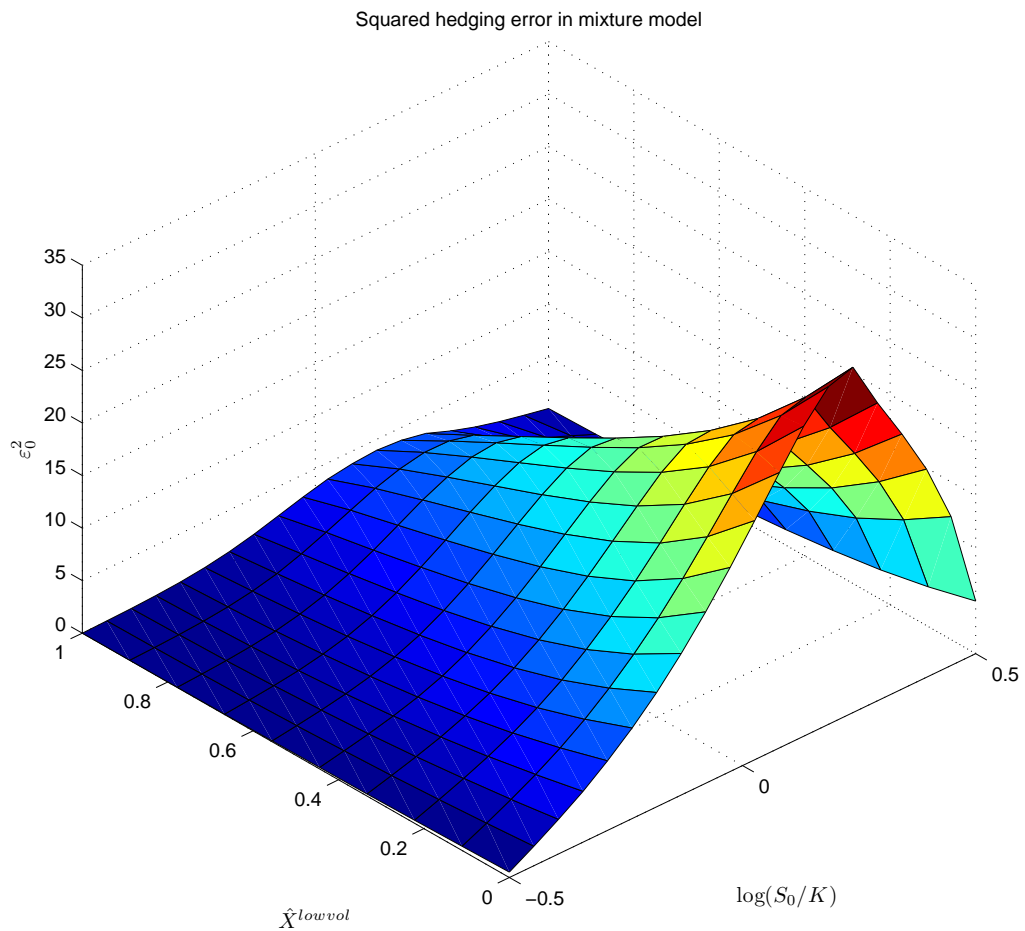
Figure 3.17: Expected quadratic hedging error in a mixture model for constituent low-vol ($\sigma_1 = 20\%$) and high-vol ($\sigma_2 = 57\%$) models with weightings ($\hat{X}^{lowvol}, 1 - \hat{X}^{lowvol}$). Call option strike $K = 100$, across varying initial estimates of the low-vol state $\hat{X}_0^1$ and varying initial log-moneyness $\log\left(\frac{S_0}{K}\right)$.

2009, Chapter 13], simplified for the martingale case:

$$\varepsilon_0^2 = \left(\frac{Kurt - 1}{4}\right)\sigma^4 \sum_{t=0}^{T-1} E_0[(\Gamma_t S_t^2)^2]$$

$$E_0[(\Gamma_t S_t)^2] = \frac{S_0^2}{2\pi\sigma^2\sqrt{T^2 - t^2}}\exp\left(-\frac{(\log(S_0/K) + 0.5\sigma^2 T)^2 + 2\sigma^2 t \log(S_0/K)}{\sigma^2(T+t)}\right)$$

Figure 3.18 shows the estimated hedging error using the formulas above with the weighted average variance and kurtosis of the two underlying models of the regime-switching model. We see that the height of the peak of this Cash Gamma approximation roughly corresponds to the height of the peak of our regime-switching hedging error, suggesting that the approximation is quite good, though it does not have the asymmetry the full solution exhibits with respect to the log-moneyness.
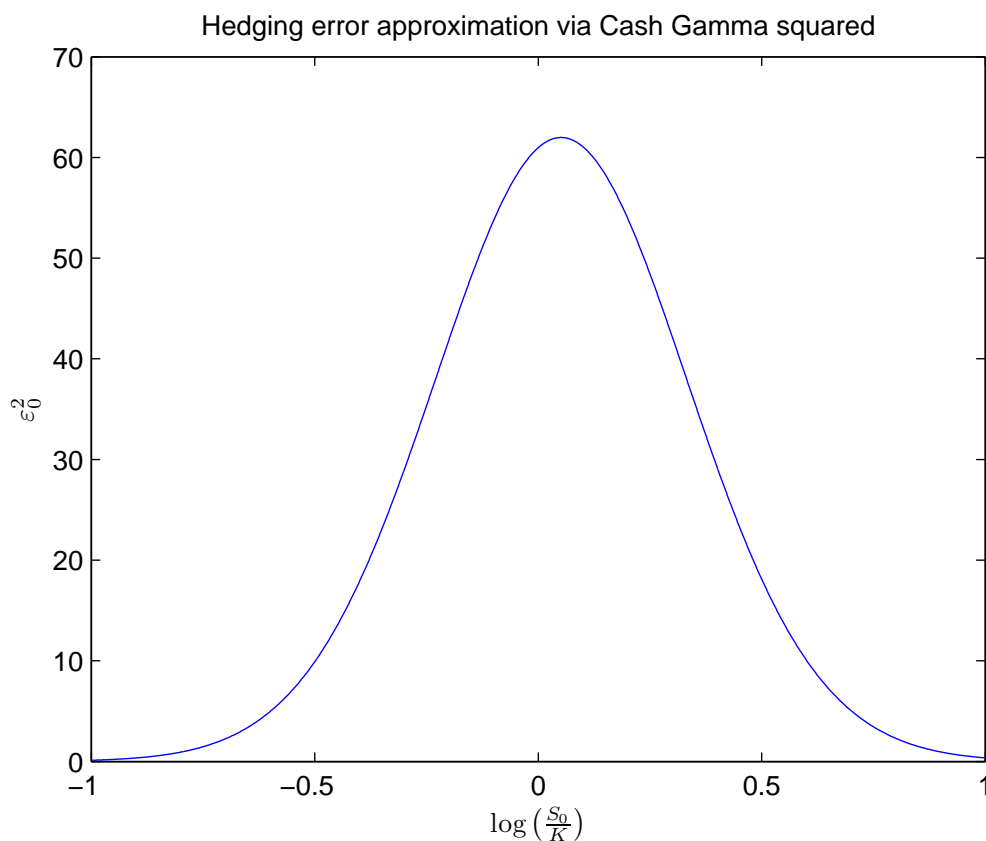
Figure 3.18: Cash Gamma approximation of expected quadratic hedging error. Call option with strike $K = 100$, maturity $T = 1$, across various levels of log-moneyness $\log\left(\frac{S_0}{K}\right)$.

3.10 Conclusion

In this chapter we set out to use a regime-switching methodology to incorporate parameter uncertainty into our pricing and hedging strategy. We provided an SDE that drives the evolution of our filtered estimate (an optional projection) of the unobservable state. With this additional stochastic state variable, we computed the relevant mean-variance hedging quantities (mean-value process, variance-optimal hedging strategy, quadratic hedging error) under such a regime-switching model. We found that the regime-switching mean-value process resembles a weighted average of two independent mean-value processes, the dependence on $\hat{\boldsymbol{X}}$ being nearly linear. For the hedging strategy, the impact of the filtered state $\hat{\boldsymbol{X}}$ is more complex and is determined partly by the dynamics (with respect to the observations Y) of the filtered state. We found that it can deviate from a weighted average case when the option drifts somewhat OTM/ITM during its lifetime. We developed a simple approximative iterative scheme to compute the hedging error. We ran Monte Carlo simulations of the terminal P&L of a delta-hedged option using the regime-switching strategy indicate that when the filter is good at identifying the current state, the regime-switching strategy performs significantly better than a simple single-regime variance optimal strategy. For this reason, we conclude that using a regime-switching model for pricing and hedging appears preferable when our constituent regimes are sufficiently distinctive. We also compared the expected hedging error of the regime-switching model against a range of alternatives. We found that adding a regime-switching component increases the size of the hedging error we expect (in essence making us aware of the additional risk a more realistic model would carry), and can be reduced by either progressively filtering the current state or (if possible) observing it directly.

Conclusion - a summary and a view to the future

In this thesis, we have dealt with various risks causing market incompleteness and how quadratic hedging techniques can take these additional risks into account.

In chapter 1, we dealt with the risk coming from an inability to act out any hedging strategy continuously in a real-world setting. We looked at the asymptotics of hedging errors coming from discrete-time trading as we increase the frequency at which we rebalance our hedge of a digital option. We saw that we can demonstrate the importance of the Cash Gamma for the hedging error of the digital option by including a compensating term in the granularity integral that balances out the explosive nature of the Cash Gamma at maturity and ensures convergence of the integral to a finite value. We found that this second order term significantly improves the quality of the approximation of the full hedging error, especially for ATM strikes. An interesting question this raises and is to be answered by future research is whether this behaviour is specific to a digital option or whether we could obtain the same result for contracts with different types of discontinuities in their payoffs. Specifically, we could look at whether the absorbing behaviour of barriers in barrier options changes the rate of explosion of the Cash Gamma at maturity, and thus changes the nature of the discontinuity.

In chapter 2, we dealt with the risk coming from an inability to protect ourselves from jumps in the price of the underlying, and how this impacts variance swaps and skewness swaps. We did so via a generic "Lévy contract", which encapsulates log contracts, variance swaps and higher order moment swaps. We computed the hedging errors for these contracts using two different utility functions: quadratic utility a.k.a. mean-variance preferences, leading to solving a quadratic hedging problem; and exponential utility, which leads to calculating an exponential compensator. We

obtained closed-form solutions for the quadratic utility case. We found that the exponential utility case required asymptotic small-jump approximations to obtain fully closed-form solutions for the desired quantities (price, hedging strategy, hedging error). Under such an approximation, the exponential utility solution strongly resembles the quadratic hedging case, signifying that when adding small jumps to our model, we can safely use quadratic utility and its easily computable closed-form solutions as good approximations to the more theoretically sound exponential utility solution. The results of this chapter raise several questions that could be addressed. First, we could investigate how our results would differ if we stopped relying strongly on the translation invariance property, allowing an impact of initial wealth and relative risk aversion. Secondly, for true market applications of these results, we would need to extend the results to allow hedging our variance swap not only by trading the underlying, but also by trading a replicating portfolio of vanilla options, which is the standard approach taken by variance swap traders.

In chapter 3, we considered adding an additional risk to the model framework from chapter 2. Not only do we maintain jumps in our underlying price process, but also add the possibility of switching between different distributions of returns during the lifetime of the trade, considering different potential market states (e.g. a bull market and bear market). Under this more complicated model (but under the simplifying assumption of a martingale underlying), we derived all three quantities central to quadratic hedging, namely the mean-value process, the variance-optimal hedging strategy, and the quadratic hedging error. For the first two, we use a Fourier transform approach to obtain explicit, implementable formulas. We contrasted the regime-switching mean-value process and hedging strategy against a simple weighted average of single-regime Lévy models. For the hedging error, we find a backward recursive scheme to numerically calculate it. We provided Monte Carlo simulation results that demonstrate the positive impact it has on the P&L of pricing and hedging options in a regime-switching market. However, for more conclusive evidence of the quality of this model, in future research we would require doing these simulations on realizations of true market paths (e.g. rolling 3-month windows over which we hedge 3-month options). We would also need to calibrate our model to observed market prices of vanilla options at each of these points, testing the quality and stability of fit to the market. From a theoretical perspective, it would also be helpful to extend the results provided beyond a martingale case.

In summary, we have shown that the quadratic hedging approach can be very useful

for dealing with a multitude of risks that lead to incomplete markets. Moreover, unlike most incomplete market approaches, the quadratic hedging approach always inherently carries with it a measure of error that we expect to experience in our hedging, whether this is due to discrete-time trading, jumps in the underlying, or uncertainty about the distribution that generates the log-returns we observe. By contributing to the literature on quadratic hedging and its applications in various circumstances, we hope to further popularize this framework, and slowly but surely approach a situation when it is practical to implement it on a broader scale. The ideal future is one that draws a parallel from Monte Carlo simulations - just as one never considers the mean of a Monte Carlo simulation without considering its standard deviation, we hope one will in the future never consider a derivative price without considering its expected hedging error.

Appendix A

Asymptotics of hedging errors - technical details

Lemma A.1. *For $0 \leq \rho \leq 1$ it holds that*

$$\int_{-\infty}^{0} \Phi(x, x, \rho) \, dx + \int_{0}^{\infty} 1 - \Phi(x, x, \rho) \, dx = \sqrt{\frac{2}{\pi}}$$

where $\Phi(\cdot, \cdot, \rho)$ is the cumulative distribution function of the bivariate normal distribution with zero means, unit variances and correlation ρ.

Proof. Using the notation in Abramowitz and Stegun [1972], we can rewrite our equation as

$$\int_{-\infty}^{0} L(-x, -x; \rho) \, dx + \int_{0}^{\infty} 1 - L(-x, -x; \rho) \, dx = \sqrt{\frac{2}{\pi}}.$$

By Abramowitz and Stegun [1972, eqn. 26.3.9], we know that

$$L(-x, -x; \rho) - L(x, x; \rho) = 2\Phi(x) - 1.$$

We can integrate the above over the interval $[0, \infty]$ to get

$$\int_{0}^{\infty} L(-x, -x; \rho) - L(x, x; \rho) \, dx = \int_{0}^{\infty} 2\Phi(x) - 1 \, dx,$$

or rearranged,

$$\int_0^\infty L(-x, -x; \rho)\,dx + \int_0^\infty 1 - L(x, x; \rho)\,dx = 2\int_0^\infty \Phi(x)\,dx,$$

We see that by a simple change of variable $y = -x$, it holds that:

$$\int_0^\infty 1 - L(x, x; \rho)\,dx = \int_{-\infty}^0 1 - L(-x, -x; \rho)\,dx.$$

Finally, we evaluate the integral on the right-hand side:

$$2\int_0^\infty \Phi(x)\,dx = 2\left[x\Phi(x) + \varphi(x)\right]_0^\infty = 2\frac{1}{\sqrt{2\pi}} = \sqrt{\frac{2}{\pi}}.$$

$\square$

Lemma A.2. *For* $0 \le \rho \le 1$ *it holds that*

$$\sqrt{\frac{\pi}{2}} \int_{-\infty}^0 \Phi_2(x, x, \rho)\,dx = \frac{1}{2} - \frac{\sqrt{2(1-\rho)}}{4},$$

where $\Phi_2(\cdot, \cdot, \rho)$ *is as given in Lemma (A.1).*

Proof. Using $\varphi(x, y; \rho)$ to denote the cumulative density function of a bivariate normal distribution with unit variances and correlation ρ, we can write

$$\int_{-\infty}^0 \Phi(z; \rho)\,dz = \int_{-\infty}^0 \int_{-\infty}^z \int_{-\infty}^z \varphi(x, y; \rho)\,dx\,dy\,dz.$$

We can see that area over which we want to integrate, $\{-\infty \le x \le z; -\infty \le y \le z; -\infty \le z \le 0\}$, can equivalently be written as $\{-\infty \le x \le 0; -\infty \le y \le 0; \max\{x, y\} \le z \le 0\}$. Using this, we write

$$
\begin{aligned}
\int_{-\infty}^0 \int_{-\infty}^z \int_{-\infty}^z \varphi(x, y; \rho)\,dx\,dy\,dz &= \int_{-\infty}^0 \int_{-\infty}^0 \varphi(x, y; \rho)\left(\int_{\max\{x,y\}}^0 dz\right)\,dx\,dy \\
&= \int_{-\infty}^0 \int_{-\infty}^0 \varphi(x, y; \rho) - \max\{x, y\}\,dx\,dy \\
&= -\int_{-\infty}^0 dx \int_{-\infty}^x x\varphi(x, y; \rho)\,dy - \int_{-\infty}^0 dy \int_{-\infty}^y y\varphi(x, y; \rho)\,dx.
\end{aligned}
$$

156

Now we make use of the fact that $\varphi(x, y; \rho) = \varphi(y, x; \rho)$, so we can write

$$\int_{-\infty}^{0} \Phi(z; \rho) \, dz = -2 \int_{-\infty}^{0} \int_{-\infty}^{x} x\varphi(x, y; \rho) \, dy \, dx$$

By virtue of Abramowitz and Stegun [1972, eqn. 26.3.2]:

$$\int_{-\infty}^{0} \int_{-\infty}^{x} x\varphi(x, y; \rho) \, dy \, dx = \int_{-\infty}^{0} \int_{-\infty}^{x} (1 - \rho^2)^{-1/2} x\varphi(x)\varphi\left(\frac{y - \rho x}{\sqrt{1 - \rho^2}}\right) dy$$

$$= \int_{-\infty}^{0} x\varphi(x)\Phi(Ax) \, dx,$$

where $A = \frac{1-\rho}{\sqrt{1-\rho^2}}$. We are left with a standard integration problem that requires the use of integration by parts. Using this technique, we get

$$\int_{-\infty}^{0} x\varphi(x)\Phi(Ax) \, dx = [-\varphi(x)\Phi(Ax)]_{-\infty}^{0} + A \int_{-\infty}^{0} \varphi(x)\varphi(Ax) \, dx$$

$$= -\frac{1}{2\sqrt{2\pi}} + A \int_{-\infty}^{0} \varphi(x)\varphi(Ax) \, dx$$

By a simple change of variables, we obtain that

$$\int_{-\infty}^{0} \varphi(x)\varphi(Ax) \, dx = \frac{1}{2\sqrt{2\pi}\sqrt{A^2 + 1}}.$$

We now have everything needed to get our result. Plugging in for A and putting terms together, we find that

$$\int_{-\infty}^{0} \Phi(z; \rho) \, dz = -2\left(-\frac{1}{2\sqrt{2\pi}} + \frac{\sqrt{1-\rho}}{2\sqrt{2}\sqrt{2\pi}}\right) = \frac{1}{\sqrt{2\pi}}\left(1 - \frac{\sqrt{1-\rho}}{\sqrt{2}}\right).$$

$\square$

Lemma A.3. *For a process S given by dynamics (1.7) evaluated at times $t \in [0, T]$, it holds that $\forall \alpha \geq 0$*

$$E_t[S_T^{\alpha} 1_{S_T > K}] = S_t^{\alpha} \exp\left(\frac{1}{2}(\alpha^2 - \alpha)\sigma^2(T - t)\right) \Phi\left(\frac{y_t}{s} - \left(\frac{1}{2} - \alpha\right)s\right),$$

where $s = \sigma\sqrt{T - t}$ and $y_t = \log\frac{S_t}{K}$ and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution.

Proof. We prove the lemma via a change of measure and Girsanov theorem. The change of measure will be defined by

$$\frac{dP^*}{dP}_{|\mathcal{F}_T} = Z_T = \frac{S_T^\alpha}{E_0[S_T^\alpha]}$$

We can compute Z_T explicitly by using the knowledge of the distribution of log returns, $\log S_T \sim N(\log S_0 - \frac{1}{2}\sigma^2 T, \sigma^2 T)$ and the fact that S^α is martingale:

$$Z_T = \frac{\exp\left(\alpha \log S_T\right)}{E_0[S_T^\alpha]} = \frac{S_0^\alpha \exp\left(-\frac{1}{2}\alpha\sigma^2 T + \alpha\sigma W_T\right)}{S_0^\alpha} = \exp\left(\int_0^T \alpha\sigma \, dW_t - \frac{1}{2}\int_0^T \alpha^2\sigma^2 \, dt\right),$$

where W is a Brownian motion. As it is written, we see that Z_T perfectly fits Girsanov theorem (see e.g. Shreve [2004, Theorem 5.2.3]), and thus we know that under the new measure P^*, Brownian motion is given by

$$W_t^* = W_t - \alpha\sigma t$$

We can now write the expectation we wish to compute as follows:

$$E_t[S_T^\alpha 1_{S_T > K}] = E_t\left[\underbrace{\frac{S_T^\alpha}{E_0[S_T^\alpha]}}_{Z_T} E_0[S_T^\alpha] 1_{S_T > K}\right]$$

By Shreve [2004, Lemma 5.2.2], we know that

$$\frac{1}{Z_t} E_t[Z_T 1_{S_T > K}] = E_t^*[1_{S_T > K}].$$

Moreover, by Girsanov theorem (e.g. Shreve [2004, Theorem 5.2.3]), $Z_t = E_t[Z_T]$, thus leading us to the result

$$E_t[S_T^\alpha 1_{S_T > K}] = Z_t E_0[S_T^\alpha] E_t^*[1_{S_T > K}] = E_t[S_T^\alpha] E_t^*[1_{S_T > K}].$$

By straightforward computations, we obtain that

$$E_t[S_T^\alpha] = S_t^\alpha \exp\left(\frac{1}{2}(\alpha^2 - \alpha)\sigma^2(T - t)\right)$$

Now all we have to do is evaluate $E_t^*[1_{S_T>K}]$:

$$E_t^*[1_{S_T>K}] = P^*(\log S_T > \log K) = P^*\left(\log S_t - \frac{1}{2}\sigma^2(T-t) + \sigma W_{T-t} > \log K\right)$$

$$= P^*\left(\sigma(W_{T-t}^* + \alpha\sigma(T-t)) > \frac{1}{2}\sigma^2(T-t) - \log\frac{S_t}{K}\right)$$

$$= P^*\left(\sigma W_{T-t}^* > \left(\frac{1}{2} - \alpha\right)\sigma^2(T-t) - y_t\right)$$

Finally, we normalize $W_{T-t}^* = \sqrt{T-t}Z^*$ to a random variable $Z^* \sim N(0,1)$ and use the fact that $P^*(Z^* > x) = P^*(Z^* < -x) = \Phi(-x)$ to obtain that

$$E_t^*[1_{S_T>K}] = P^*\left(Z^* > \left(\frac{1}{2} - \alpha\right)s - \frac{y_t}{s}\right) = \Phi\left(\frac{y_t}{s} - \left(\frac{1}{2} - \alpha\right)s\right)$$

□

Lemma A.4 (Taylor's Theorem). *Let $k \geq 1$ be an integer and let the function $f : \mathbb{R} \to \mathbb{R}$ be $k+1$ times differentiable at point $a \in \mathbb{R}$. Then*

$$f(x) = f(a) + f'(a)(x-a) + \frac{f^{(2)}(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(k)}(a)}{k!}(x-a)^k + R_k(x)$$

where $R_k(x)$ is a remainder term of the form

$$R_k(x) = \int_a^x f^{(k+1)}(t)\frac{(x-t)^k}{k!}\,dt = \frac{(x-a)^{k+1}}{(k+1)!}f^{(k+1)}(\xi) \qquad (a < \xi < x)$$

Proof. See Whittaker and Watson [1996]. □

Appendix B

Good deal bounds of variance swaps - technical details

As we will be dealing with Lévy processes that allow for exotic distributions whose moments are not all necessarily finite, we provide lemmas that provide us with sufficient and equivalent conditions for the existence of moments and exponential moments.

We now state a general theorem from the literature that provides an equivalent condition for the finiteness of any $E[g(X_t)]$, as long as g is a so-called submultiplicative function.

Definition B.1. A function g is called submultiplicative if it has the following properties:

- $g > 0$

- $\exists a > 0 \ \forall x, y : g(x + y) \leq ag(x)g(y)$

Theorem B.2. *Let g be a submultiplicative, locally bounded, measurable function. Let $(X_t)_{t \geq 0}$ be a Lévy process with characteristic triplet $(b(h), \sigma^2, F)$. Then*

$$\forall t : \ E[g(X_t)] < \infty \Leftrightarrow \int_{|x| > 1} g(x) F(dx) < \infty$$

Proof. See Sato [1999, Theorem 25.3]. $\qquad \square$

Corollary B.3. *Let $(X_t)_{t \geq 0}$ be a Lévy process with characteristic triplet $(b(h), \sigma^2, F)$. Then*

$$\forall n \forall t : E[|X_t|^n] < \infty \Leftrightarrow \int_{|x|>1} |x|^n F(dx) < \infty$$

$$\forall n \forall t : E[e^{nX_t}] < \infty \Leftrightarrow \int_{|x|>1} e^{nx} F(dx) < \infty$$

Proof. We notice that functions $g(x) = |x|^n$ and $g(x) = e^{nx}$ are submultiplicative and apply Theorem B.2. $\qquad\square$

Finally we know that $E[|X|^n] < \infty$ implies that $E[X^n] < \infty$, thus giving us a sufficient condition for the finiteness of X^n.

For reference, we reproduce here the definition of the set of admissible strategies from Biagini and Černý [2011] which we will consider in the utility maximization problem.

Definition B.4. $H \in L(S)$ is an admissible integrand if $U(H \cdot S_T) \in L^1(\mathbb{P})$ and if there exists an approximate sequence $(H^n)_n$ in $\mathcal{H}$ such that:

1. $H^n \cdot S_t \to H \cdot S_t$ in probability for all $t \in [0, T]$;

2. $U(H^n \cdot S_T) \to U(H \cdot S_T) \in L^1(\mathbb{P})$.

The set of all admissible integrands is denoted by $\bar{\mathcal{H}}$.

Appendix C

Regime-switching quadratic hedging - technical details

The next two definitions are reproduced for reference from Kallsen and Shiryaev [2002].

Definition C.1. A real-valued semimartingale is called *special* if it can be written as $X = X_0 + M + V$ for some local martingale M and some predictable process V of finite variation, both starting at 0.

Definition C.2. Let X be a real-valued semimartingale. X is called *exponentially special* if $\exp(X - X_0)$ is a special semimartingale.

The following are helpful lemmas establishing martingale properties of the Markov process X and optional projections.

Lemma C.3. $\hat{\boldsymbol{X}}_t^M := \exp(-\boldsymbol{A}t)\hat{\boldsymbol{X}}_t$ *is a* $(\hat{\mathcal{F}}_t, \mathbb{P})$-*martingale.*

Proof.

$$\mathbb{E}[\hat{\boldsymbol{X}}_T|\hat{\mathcal{F}}_t] = \mathbb{E}[\mathbb{E}[\boldsymbol{X}_T|\hat{\mathcal{F}}_T]|\hat{\mathcal{F}}_t] = \mathbb{E}[\boldsymbol{X}_T|\hat{\mathcal{F}}_t]$$
$$= \mathbb{E}[\mathbb{E}[\boldsymbol{X}_T|\mathcal{G}_t]|\hat{\mathcal{F}}_t] = \mathbb{E}[\exp(\boldsymbol{A}(T-t))\boldsymbol{X}_t|\hat{\mathcal{F}}_t] = \exp(\boldsymbol{A}(T-t))\hat{\boldsymbol{X}}_t$$

Consequently,

$$\mathbb{E}[\hat{\boldsymbol{X}}_T^M|\hat{\mathcal{F}}_t] = \mathbb{E}[\exp(-\boldsymbol{A}T)\hat{\boldsymbol{X}}_T|\hat{\mathcal{F}}_t] = \exp(-\boldsymbol{A}t)\hat{\boldsymbol{X}}_t = \hat{\boldsymbol{X}}_t^M$$

□

Lemma C.4. *If $\boldsymbol{X}_t$ is a $(\mathcal{G}_t, \mathbb{P})$-martingale, then $\hat{\boldsymbol{X}}_t = \mathbb{E}[\boldsymbol{X}_t|\hat{\mathcal{F}}_t]$ is a $(\hat{\mathcal{F}}_t, \mathbb{P})$-martingale.*

Proof. The $(\mathcal{G}_t, \mathbb{P})$-martingale property of $\boldsymbol{X}$ dictates that $\mathbb{E}[\boldsymbol{X}_t|\mathcal{G}_s] = \boldsymbol{X}_s$. Then it follows directly that

$$\mathbb{E}[\hat{\boldsymbol{X}}_t|\hat{\mathcal{F}}_s] = \mathbb{E}[\mathbb{E}[\boldsymbol{X}_t|\hat{\mathcal{F}}_t]|\hat{\mathcal{F}}_s] = \mathbb{E}[\boldsymbol{X}_t|\hat{\mathcal{F}}_s]$$
$$= \mathbb{E}[\mathbb{E}[\boldsymbol{X}_t|\mathcal{G}_s]\hat{\mathcal{F}}_s] = \mathbb{E}[\boldsymbol{X}_s|\hat{\mathcal{F}}_s] = \hat{\boldsymbol{X}}_s$$

□

Lemma C.5. *For any progressively measurable process Z with $\mathbb{E}\int_0^T |Z_t|dt < \infty$, it holds for optional projections onto filtration $\hat{\mathcal{F}}_t)$ that*

$$\xi_T := \widehat{\int_0^T Z_t dt} - \int_0^T \hat{Z}_t dt$$

is a $(\mathbb{P}, \mathcal{F}_t)$-martingale.

Proof. Rewriting using expectations and using Fubini's theorem and the Tower law,

$$\xi_T = \widehat{\int_0^T Z_t dt} - \int_0^T \hat{Z}_t dt$$
$$= \mathbb{E}[\int_0^T Z_t dt|\hat{\mathcal{F}}_T] - \int_0^T \hat{Z}_t dt$$
$$= |Fubini| = \int_0^T \mathbb{E}[Z_t|\hat{\mathcal{F}}_T] - \mathbb{E}[Z_t|\hat{\mathcal{F}}_t]dt$$
$$= \int_0^T \mathbb{E}[Z_t|\hat{\mathcal{F}}_T] - \mathbb{E}[\mathbb{E}[Z_t|\hat{\mathcal{F}}_T]|\hat{\mathcal{F}}_t]dt$$

Then for any fixed $u \leq T$:

$$
\begin{aligned}
\mathbb{E}[\xi_T|\hat{\mathcal{F}}_u] &= \int_0^u \mathbb{E}[\mathbb{E}[Z_t|\hat{\mathcal{F}}_T]|\hat{\mathcal{F}}_u] - \mathbb{E}[\mathbb{E}[Z_t|\hat{\mathcal{F}}_t]|\hat{\mathcal{F}}_u]dt \\
&\quad + \int_u^T \mathbb{E}[\mathbb{E}[Z_t|\hat{\mathcal{F}}_T]|\hat{\mathcal{F}}_u] - \mathbb{E}[\mathbb{E}[Z_t|\hat{\mathcal{F}}_t]|\hat{\mathcal{F}}_u]dt \\
&= \int_0^u \mathbb{E}[Z_t|\hat{\mathcal{F}}_u] - \mathbb{E}[Z_t|\hat{\mathcal{F}}_t]dt \\
&\quad + \int_u^T \mathbb{E}[Z_t|\hat{\mathcal{F}}_u] - \mathbb{E}[Z_t|\hat{\mathcal{F}}_u]dt \\
&= \int_0^u \mathbb{E}[Z_t|\hat{\mathcal{F}}_u] - \mathbb{E}[Z_t|\hat{\mathcal{F}}_t]dt = \xi_u
\end{aligned}
$$

$\square$

C.1 Dynamics of $\hat{\boldsymbol{X}}$ via Zakai equation

We can obtain an SDE of a normalized version of the filtered state $\hat{\boldsymbol{X}}$ over time. We shall do so in this section by following results from Elliott and Royal [2008] and Siu [2014]. Define a reference probability measure $\bar{\mathbb{P}}$ and a Lévy measure $\bar{F}$ (of our choosing) under $\bar{\mathbb{P}}$. Further, we define the following processes:

$$
\begin{aligned}
L_k(y) &:= \frac{dF^k}{d\bar{F}}(y) \\
\bar{U}_t &:= \int_0^t \sum_k X_{s-}^k \int_{\mathbb{R}} (L_k(y) - 1)(J^Y(dy, ds) - \bar{F}(dy)ds) \\
\bar{\Lambda}_t^1 &:= 1 + \int_0^t \bar{\Lambda}_{s-}^1 d\bar{U}_s \\
&= \exp\left(\int_0^t \sum_k X_{s-}^k \int_{\mathbb{R}} \log(L_k(y))J(dy, ds) - \int_0^t \sum_k X_{s-}^k \int_{\mathbb{R}} (L_k(y) - 1)\bar{F}(dy)ds\right)
\end{aligned}
$$

The process $\bar{\Lambda}^1$ will be used as a change of measure of the jump density, leading to a Lévy measure compensator $\bar{F}$ independent of X under measure $\bar{\mathbb{P}}$. We also define a process $\bar{\Lambda}^2$ for the diffusion part of the process Y, making that part independent of X under $\bar{\mathbb{P}}$:

$$
\bar{\Lambda}_t^2 := \exp\left(\int_0^t \frac{\boldsymbol{X}_s^\top \boldsymbol{b}}{\boldsymbol{X}_s^\top \boldsymbol{c}} dW_s - \frac{1}{2} \int_0^t \left(\frac{\boldsymbol{X}_s^\top \boldsymbol{b}}{\boldsymbol{X}_s^\top \boldsymbol{c}}\right)^2 ds\right)
$$

Define

$$\bar{\Lambda} := \bar{\Lambda}^1 \bar{\Lambda}^2$$

and consider the change of measure

$$\frac{d\mathbb{P}}{d\bar{\mathbb{P}}}_{|\mathcal{G}_t} = \bar{\Lambda}_t$$

Let us note that for us to use the reference measure, we require the change of measure to be locally absolutely continuous with respect to the historical measure $\mathbb{P}$. This holds under certain integrability conditions listed in Jacod and Shiryaev [2003, Theorem IV.4.39]. For some infinite activity processes (such as the VG process), these conditions are not satisfied, but are satisfied for their truncated versions, where jumps smaller than ϵ are ignored. We shall therefore limit ourselves to only processes whose Lévy densities that satisfy the following assumption.

Assumption C.1. All the Lévy measures $F^k, k = 1...M$ are locally absolutely continuous with respect to the reference measure $\bar{\mathbb{P}}$.

Using the change of measure, we obtain an alternative representation of the filter $\hat{X}$ using the continuous-time version of Bayes' Theorem:

$$\hat{\boldsymbol{X}}_t = \mathbb{E}[\boldsymbol{X}_t|\hat{\mathcal{F}}_t] = \frac{\bar{\mathbb{E}}[\bar{\Lambda}_t \boldsymbol{X}_t|\hat{\mathcal{F}}_t]}{\bar{\mathbb{E}}[\bar{\Lambda}_t|\hat{\mathcal{F}}_t]}$$

We will now get the dynamics for the (M-dimensional) unnormalized estimate $\boldsymbol{q}_t(\boldsymbol{X}) = \bar{\mathbb{E}}[\bar{\Lambda}_t \boldsymbol{X}_t|\hat{\mathcal{F}}_t]$. Using Ito's lemma, we have

$$\begin{aligned}
\bar{\Lambda}_t \boldsymbol{X}_t = {}& \bar{\Lambda}_0 \boldsymbol{X}_0 + \int_0^t \bar{\Lambda}_{s-} \boldsymbol{A} \boldsymbol{X}_s ds + \int_0^t \bar{\Lambda}_{s-} d\boldsymbol{M}_s \\
& + \int_0^t \bar{\Lambda}_{s-} \sum_k \boldsymbol{e}_k X_{s-}^k \frac{b^k}{c^k} dW_s \\
& + \int_0^t \bar{\Lambda}_{s-} \sum_k \boldsymbol{e}_k X_{s-}^k \int_{\mathbb{R}} (L_k(y) - 1)(J^Y(dy, ds) - \bar{F}(dy)ds)
\end{aligned}$$

It can be shown that because $\boldsymbol{M}$ is a $(\mathbb{P}, \mathcal{G}_t)$-martingale, it is also a $(\bar{\mathbb{P}}, \mathcal{G}_t)$-martingale. Furthermore, we note that $\bar{\Lambda}_s \boldsymbol{X}_s$ is independent of observations Y_u at times $u \in [s, t]$.

Consequently:

$$\boldsymbol{q}_t(\boldsymbol{X}) = \mathbb{E}[\bar{\Lambda}_t \boldsymbol{X}_t | \hat{\mathcal{F}}_t]$$

$$= \boldsymbol{q}_0(\boldsymbol{X}) + \int_0^t \boldsymbol{A}\boldsymbol{q}_s(\boldsymbol{X})ds + \int_0^t \sum_k \boldsymbol{e}_k \frac{b^k}{c^k} q_s^k(\boldsymbol{X})dW_s$$

$$+ \int_0^t \sum_k \boldsymbol{e}_k q_s^k(\boldsymbol{X}) \int_{\mathbb{R}} (L_k(y) - 1)(J^Y(dy, ds) - \bar{F}(dy)ds)$$

We can further transform this into an ODE via the so-called Gauge transformation by considering an inverse change of Lévy measure:

$$U_t^k = \int_0^t \int_{\mathbb{R}} (\frac{1}{L^k(y)} - 1)(J^Y(dy, ds) - F^k(dy)ds)$$

$$\lambda_t^k = 1 + \int_0^t X_{s-}^k \frac{b^k}{c^k} dW_s + \int_0^t \lambda_{s-}^k dU_s^k$$

$$\boldsymbol{\Gamma}_t = \mathrm{diag}(\lambda_t)$$

where $\boldsymbol{\Gamma}$ is a M-dimensional matrix with elements λ_t^k on its diagonal. Then the transformed variable $\bar{\boldsymbol{q}}_t(\boldsymbol{X}) = \boldsymbol{\Gamma}_t \boldsymbol{q}_t(\boldsymbol{X})$ follows the ODE:

$$\bar{\boldsymbol{q}}_t(\boldsymbol{X}) = \bar{\boldsymbol{q}}_0(\boldsymbol{X}) + \int_0^t \boldsymbol{\Gamma}_s \boldsymbol{A} \boldsymbol{\Gamma}_s^{-1} \bar{\boldsymbol{q}}_s(\boldsymbol{X})ds$$

which can be solved numerically starting from point $\bar{\boldsymbol{q}}_0(\boldsymbol{X}) = \boldsymbol{q}_0(\boldsymbol{X}) = \hat{\boldsymbol{X}}_0$. Then

$$\hat{\boldsymbol{X}}_t = \frac{\boldsymbol{q}_t(\boldsymbol{X})}{\boldsymbol{q}_t(\boldsymbol{X})^\top \mathbf{1}} = \frac{\bar{\boldsymbol{q}}_t(\boldsymbol{X})}{\bar{\boldsymbol{q}}_t(\boldsymbol{X})^\top \mathbf{1}}$$

One issue with this representation is that it is not straightforward to compute the predictable quadratic variation (or covariation) of $\hat{\boldsymbol{X}}$ from this form.

C.1.1 Discretization and numerical solution

By discretizing the equation for $\bar{\boldsymbol{q}}$ over an equidistant grid $\{t_0, ..., t_N\}$ with timestep size Δt, we have

$$\boldsymbol{\Gamma}_{t_{k+1}}\boldsymbol{q}_{t_{k+1}}(\boldsymbol{X}) = \boldsymbol{\Gamma}_{t_k}\boldsymbol{q}_{t_k}(\boldsymbol{X}) + \int_{t_k}^{t_{k+1}} \boldsymbol{\Gamma}_s \boldsymbol{A}\boldsymbol{q}_s(\boldsymbol{X})ds$$
$$\approx \boldsymbol{\Gamma}_{t_k}\boldsymbol{q}_{t_k}(\boldsymbol{X}) + \boldsymbol{\Gamma}_{t_k}\boldsymbol{A}\boldsymbol{q}_{t_k}(\boldsymbol{X})\Delta t,$$

or equivalently,

$$\boldsymbol{q}_{t_{k+1}}(\boldsymbol{X}) \approx \boldsymbol{\Gamma}_{t_{k+1}}^{-1}\boldsymbol{\Gamma}_{t_k}(\boldsymbol{I}_M + \boldsymbol{A}\Delta t)\boldsymbol{q}_{t_k}(\boldsymbol{X}).$$

To get $\boldsymbol{\Gamma}_t$, we recall that $\boldsymbol{\Gamma}_t = \mathrm{diag}(\boldsymbol{\lambda}_t)$, where

$$\lambda_t^j = \exp\left(\int_0^t \frac{b^j}{c^j}dW_s - \frac{1}{2}\int_0^t \left(\frac{b^j}{c^j}\right)^2 ds + \int_0^t \int_{\mathbb{R}} (\frac{1}{L^j(y)} - 1)F^j(dy)ds + \int_0^t \log(L^j(y))J^Y(dy, ds)\right)$$
$$= \exp\left(\frac{b^j}{c^j}W_t + \frac{1}{2}\left(\frac{b^j}{c^j}\right)^2 t + \int_0^t \int_{\mathbb{R}} (\bar{F}(dy) - F^j(dy)) + \int_0^t \log(L^j(y))J^Y(dy, ds)\right)$$

Then $\boldsymbol{B} = \boldsymbol{\Gamma}_{t_{k+1}}^{-1}\boldsymbol{\Gamma}_{t_k}$ is going to be a diagonal matrix with diagonal elements

$$B_{jj} = \exp\left(\frac{b^j}{c^j}(W_{t_{k+1}} - W_{t_k}) + \frac{1}{2}\left(\frac{b^j}{c^j}\right)^2 \delta t\right.$$
$$\left. + \Delta t \int_{\mathbb{R}} (\bar{F}(dy) - F^j(dy)) + \int_{t_k}^{t_{k+1}} \log(L^j(y))J^Y(dy, ds)\right)$$

Here we note that the last term is simply going to be a weighted sum over all jumps observed between times t_k and t_{k+1}, where the weighting term L^j is the ratio of Lévy densities of the j-th Lévy process and the chosen reference process with Lévy measure $\bar{F}$.

C.2 Filtered dynamics for $\hat{\boldsymbol{X}}$

Lemma C.6. *Under assumptions 3.2 and 3.1, every $(\mathbb{P}, \hat{\mathcal{F}}_t)$-local martingale $\boldsymbol{m}$ can be decomposed as*

$$\boldsymbol{m}_t = \boldsymbol{m}_0 + \int_0^t \int_{\mathbb{R}} \boldsymbol{\omega}_s(y)(J^Y(ds, dy) - \hat{\boldsymbol{X}}_{t-}^{\top}\boldsymbol{F}(dy)ds) + \int_0^t \boldsymbol{\alpha}_s dI_s$$

where I_t is an $(\mathbb{P}, \hat{\mathcal{F}}_t)$-Wiener process, $\boldsymbol{\omega}_t(y)$ is an $\hat{\mathcal{F}}_t$-predictable process and $\boldsymbol{\alpha}_t$ an $\hat{\mathcal{F}}_t$-adapted process and both are bounded $\mathbb{P}$-a.s. When $\boldsymbol{m}$ is a M-dimensional vector, $\boldsymbol{\omega}_t(y) = [\omega_t^1(y), ..., \omega_t^M(y)]^\top$ and $\boldsymbol{\alpha}_t = [\alpha_t^1, ..., \alpha_t^M]^\top$ are vector processes.

Proof. See Ceci and Colaneri [2012, Proposition 2.4], or in different notation also in Schmidt and Frey [2012, Lemma 3.2]. $\qquad\square$

Proof of theorem 3.6. The proof of this theorem follows the ideas in Ceci and Colaneri [2012, Theorem 3.1], extended to a multidimensional signal X. (Note a similar theorem is provided in [Frey and Schmidt, 2012, Proposition 3.2]).

We use the innovation process previously defined:

$$I_t = W_t + \int_0^t \sum_j \frac{b^j}{\sqrt{\bar{c}}} (X_{s-}^j - \hat{X}_{s-}^j)\,ds = W_t + \frac{1}{\sqrt{\bar{c}}} \int_0^t \boldsymbol{b}^\top (\boldsymbol{X}_{s-} - \hat{\boldsymbol{X}}_{s-})ds$$

which is a $(\mathbb{P}, \hat{\mathcal{F}}_t)$-local martingale. We know $\boldsymbol{X}$ has the semimartingale decomposition:

$$\boldsymbol{X}_t = \boldsymbol{X}_0 + \int_0^t \boldsymbol{A}\boldsymbol{X}_{s-}ds + \boldsymbol{M}_t$$

The $\hat{\mathcal{F}}_t$-projection of $\boldsymbol{X}$ is

$$\hat{\boldsymbol{X}}_t = \hat{\boldsymbol{X}}_0 + \int_0^t \widehat{\boldsymbol{A}\boldsymbol{X}_{s-}}ds + \hat{\boldsymbol{M}}_t$$

$$= \hat{\boldsymbol{X}}_0 + \int_0^t \boldsymbol{A}\hat{\boldsymbol{X}}_{s-}ds - \int_0^t \boldsymbol{A}\hat{\boldsymbol{X}}_{s-}ds + \int_0^t \widehat{\boldsymbol{A}\boldsymbol{X}_{s-}}ds + \hat{\boldsymbol{M}}_t$$

From Lemma C.5 we have that $\int_0^t \boldsymbol{A}\hat{\boldsymbol{X}}_{s-}ds - \int_0^t \widehat{\boldsymbol{A}\boldsymbol{X}_{s-}}ds$ is a $(\hat{\mathcal{F}}_t, \mathbb{P})$-martingale. Furthermore, since we know $\boldsymbol{M}_t$ is a $(\mathcal{G}_t, \mathbb{P})$-martingale, it follows (see Lemma C.4) that $\hat{\boldsymbol{M}}_t$ is a $(\hat{\mathcal{F}}_t, \mathbb{P})$-martingale and hence $\hat{\boldsymbol{X}}_t - \hat{\boldsymbol{X}}_0 - \int_0^t \boldsymbol{A}\hat{\boldsymbol{X}}_{s-}ds$ is also an $\hat{\mathcal{F}}_t$-martingale. Lemma C.6 then ensures the existence of vector processes $\boldsymbol{\alpha}, \boldsymbol{\omega}$ such that

$$\hat{\boldsymbol{X}}_t - \hat{\boldsymbol{X}}_0 - \int_0^t \boldsymbol{A}\hat{\boldsymbol{X}}_{s-}ds = \int_0^t \int_{\mathbb{R}} \boldsymbol{\omega}_s(y)(J^Y(ds, dy) - \hat{\boldsymbol{X}}_{s-}^\top \boldsymbol{F}(dy)ds) + \int_0^t \boldsymbol{\alpha}_s dI_s$$

We will now proceed to find such processes $\boldsymbol{\alpha}, \boldsymbol{\omega}$ and find they can be written as functions of $\hat{X}_t$ (and jump size y for the case of ω).

168

For that, consider $\hat{\mathcal{F}}_t$-adapted process $\tilde{W}$:

$$\tilde{W}_t = I_t + \int_0^t \frac{\hat{\boldsymbol{X}}_s^\top \boldsymbol{b}}{\sqrt{\bar{c}}} ds = W_t + \int_0^t \frac{\boldsymbol{X}_s^\top \boldsymbol{b}}{\sqrt{\bar{c}}} ds$$

and a bounded $\hat{\mathcal{F}}_t$-adapted M-dimensional process $\boldsymbol{U}$ s.t.

$$\boldsymbol{U}_t = \int_0^t \int_{\mathbb{R}} \boldsymbol{\Gamma}(s, x) J^Y(ds, dx)$$

where $\boldsymbol{\Gamma}(s, x) = [\Gamma^1(s, x), ..., \Gamma^M(s, x)]^\top$.

We proceed following the steps in Ceci and Colaneri [2012]:

1. compute $\widehat{X^k \tilde{W}}_t$ and $\hat{X}_t^k \tilde{W}_t$ separately

2. compute $\widehat{X^k U^k}_t$ and $\hat{X}_t^k U_t^k$

Since in both cases, $\tilde{W}$ and U are $\hat{\mathcal{F}}_t$-adapted, it holds that $\widehat{X^k \tilde{W}}_t = \hat{X}_t^k \tilde{W}_t$ and $\widehat{X^k U^k}_t = \hat{X}_t^k U_t^k$. By comparing the drift on the left and right-hand side of each equation, we will find processes α_s^k and $\omega_s^k(y)$.

Step 1

$$\begin{aligned}
d(X_t^k \tilde{W}_t) &= X_{t-}^k d\tilde{W}_t + \tilde{W}_{t-} dX_t^k + d\langle X^k, \tilde{W} \rangle_t^G \\
&= X_{t-}^k dW_t + X_{t-}^k \frac{b^\top X_t}{\sqrt{\bar{c}}}) dt + \tilde{W}_{t-} \sum_j a_{kj} X_t^j dt + dm_t^1
\end{aligned}$$

where $m_t^1 = \int_0^t \tilde{W}_{s-} dM_s^k$ is a $(\mathbb{P}, \mathcal{G}_t)$-local martingale. Projecting onto $\hat{\mathcal{F}}_t$,

$$d(\widehat{X^k \tilde{W}}_t) = \left\{ \sum_j \frac{b^j}{\sqrt{\bar{c}}} \widehat{X_{t-}^k X_t^j} + \tilde{W}_{t-} \sum_j a_{kj} \hat{X}_t^j \right\} dt + \hat{X}^k dW_t + d\hat{m}_t^1 + d\tilde{m}_t^1$$

where $\tilde{m}_1$ is a $(\mathbb{P}, \hat{\mathcal{F}}_t)$-martingale and $\hat{m}^1$ has a sequence of stopping times on which it is a $(\mathbb{P}, \hat{\mathcal{F}}_t)$-martingale (see Ceci and Colaneri [2012]).

Computed separately,

$$d(\hat{X}_t^k \tilde{W}_t) = \left\{ \hat{X}_{t-}^k \frac{b^\top \hat{X}_t}{\sqrt{\bar{c}}} + \tilde{W}_{t-} \sum_j a_{kj} \hat{X}_t^j + \alpha_s^k \right\} dt + dm_t^2$$

where $m_t^2 = \int_0^t (\tilde{W}_s h^k(s) + \hat{X}_s^k) dI_s + \int_0^t \tilde{W}_s \int_{\mathbb{R}} \omega_s^k(x) (J(ds, dx) - \hat{\boldsymbol{X}}_{s-}^\top \boldsymbol{F}(ds, dx))$ is a $(\mathbb{P}, \hat{\mathcal{F}}_t)$-local martingale.

Since $\widehat{X^k \tilde{W}}_t = \hat{X}_t^k \tilde{W}_t$, the drift terms must equal,which gives

$$\alpha_t^k = \sum_j \frac{b^j}{\sqrt{\bar{c}}} \left(\widehat{X_{t-}^k X_{t-}^j} - \hat{X}_{t-}^k \hat{X}_{t-}^j \right).$$

This can be further simplified by noting that $\widehat{X_{t-}^k X_{t-}^j} = \hat{X}_{t-}^k$ when $k = j$, and otherwise equals 0:

$$\alpha_t^k = \hat{X}_{t-}^k \frac{1}{\sqrt{\bar{c}}} \left(b^k - \sum_j b^j \hat{X}_{t-}^j \right).$$

Thus the process $\boldsymbol{\alpha}_t = [\alpha_t^1, \cdot, \alpha_t^M]^\top$ can be written as $\boldsymbol{\alpha}_t = a(\hat{\boldsymbol{X}}_{t-})$ where function $a(\boldsymbol{x})$ is defined as follows:

$$a(\boldsymbol{x}) = \text{diag}(\boldsymbol{x}) \frac{\boldsymbol{b} - \boldsymbol{b}^\top \boldsymbol{x}}{\sqrt{\bar{c}}}$$

Defining $\tilde{a}(\boldsymbol{x}) := \frac{1}{\sqrt{\bar{c}}} (\boldsymbol{b} - \boldsymbol{b}^\top \boldsymbol{x})$, then the process can also be rewritten into vector form as follows:

$$\boldsymbol{\alpha}_t = \text{diag}(\hat{\boldsymbol{X}}_{t-}) \tilde{a}(\hat{\boldsymbol{X}}_{t-})$$

Step 2 As in step 1, we start under $\mathcal{G}_t$; if we assume that $\boldsymbol{X}$ and Y do not have any common jumps:

$$
\begin{aligned}
d(X_t^k U_t^k) &= X_{t-}^k dU_t^k + U_{t-}^k dX_t^k + d[X^k, U^k]_t \\
&= \left\{ U_{t-}^k \sum_j a_{kj} X_t^j + X_{t-}^k \int_{\mathbb{R}} \Gamma^k(t, x) \sum_j X_t^j F^j(dx) \right\} dt + dm_t^3
\end{aligned}
$$

where $m_t^3 = \int_0^t \int_{\mathbb{R}} X_{s-}^k \Gamma^k(s, x) \tilde{J}^Y(ds, dx) + \int_0^t U_{s-} dM_s^k$ is a $(\mathbb{P}, \mathcal{G}_t)$-martingale. Projecting onto $\hat{\mathcal{F}}_t$,

$$d(\widehat{X_t^k U_t^k}) = \left\{ U_{t-}^k \sum_j a_{kj} \hat{X}_t^j + \int_{\mathbb{R}} \Gamma^k(t, x) \sum_j \widehat{X_{t-}^k X_t^j} F^j(dx) \right\} dt + d\tilde{m}_t^3$$

where $\tilde{m}_t^3$ is a $(\mathbb{P}, \hat{\mathcal{F}}_t)$-martingale.

Independently computed,

$$d(\hat{X}_t^k U_t^k) = \hat{X}_{t-}^k dU_t^k + U_{t-}^k d\hat{X}_t^k + d[\hat{X}^k, U^k]_t$$

$$= \left\{ \int_{\mathbb{R}} (\hat{X}_{t-}^k + \omega_t(x)) \Gamma^k(t, x) \sum \hat{X}_{t-}^j F^j(dx) + U_{t-}^k \sum_j a_{kj} \hat{X}_{t-}^j \right\} dt + dm_t^4$$

where $m_t^4 = \int_0^t U_s^k h^k(s) dI_s + \int_0^t (\hat{X}_{s-}^k \Gamma^k(s, x) + U_{s-}^k \omega_s^k(x))(J^Y(ds, dx) - \sum_j \hat{X}_{s-}^j F^j(dx) ds)$ is a $\mathcal{F}_t^Y$-martingale.

Comparing drifts again, we find:

$$\sum_j \widehat{X_{t-}^k X_{t-}^j} \int_{\mathbb{R}} \Gamma^k(t, x) F^j(dx) = \int_R \Gamma^k(t, x)(\hat{X}_{t-}^k + \omega_t^k(x)) \sum_j \hat{X}_{t-}^j F^j(dx)$$

Rearranging, we can obtain the process $\omega_t^k(x)$:

$$\omega_t^k(x) = \frac{\sum_j (\widehat{X_{t-}^k X_{t-}^j} - \hat{X}_{t-}^k \hat{X}_{t-}^j) F^j(dx)}{\sum_j \hat{X}_{t-}^j F^j(dx)}$$

But $\widehat{X_{t-}^k X_t^j} = \hat{X}_{t-}^k$ when $k = j$ and is equal to 0 for all other cases, which means our result simplifies:

$$\omega_t^k(x) = \hat{X}_{t-}^k \left(\frac{F^k(dx)}{\sum_j \hat{X}_{t-}^j F^j(dx)} - 1 \right)$$

This can be written in vector form as $\boldsymbol{\omega}_t(y) = w(\hat{\boldsymbol{X}}_{t-}, y)$ where function $w(\boldsymbol{x}, y)$ is defined as follows:

$$w(\boldsymbol{x}, y) := \operatorname{diag}(\boldsymbol{x}) \left(\frac{\boldsymbol{F}(dy)}{\boldsymbol{x}^\top \boldsymbol{F}(dy)} - 1 \right)$$

We can define a separate function $\tilde{w}(\boldsymbol{x}, y) := (\frac{\boldsymbol{F}(dy)}{\boldsymbol{x}^\top \boldsymbol{F}(dx)} - 1)$, which means $w(\boldsymbol{x}, y) = \operatorname{diag}(\boldsymbol{x})\tilde{w}(\boldsymbol{x}, y)$, or alternatively $\boldsymbol{w}(\boldsymbol{x}, y) = \operatorname{diag}(\tilde{w}(\boldsymbol{x}, y))\boldsymbol{x}$. $\qquad \square$

C.3 Hedging error via PIDEs

Here, we will describe an alternative approach that gives us the hedging error

$$\varepsilon_0^2 = \mathbb{E}[\langle V \rangle_T - \xi^2 \cdot \langle S \rangle_T]$$

as the solution of a partial integral differential equation (PIDE). For that, we define the following function h:

$$h(t, Y_t, \hat{\boldsymbol{X}}_t) := \mathbb{E}[\int_t^T d\langle V - \xi \cdot S\rangle_u | \hat{\mathcal{F}}_t]$$

where $Y = \log(S/S_0)$. Our aim then is to compute

$$h(0, Y_0, \hat{\boldsymbol{X}}_0) = \varepsilon_0^2 = \mathbb{E}[\int_0^T d\langle V - \xi \cdot S\rangle_u | \hat{\mathcal{F}}_0]$$

Theorem C.7. *Assuming $h(t, y, x) \in \mathcal{C}^{1,2,2}$ and $V_t = f(t, \hat{\boldsymbol{X}}_t, Y_t) \in \mathcal{C}^{1,2,2}$, then*

$$\varepsilon_0^2 = h(0, Y_0, \hat{\boldsymbol{X}}_0)$$

is the solution to the following backwards PIDE

$$
\begin{aligned}
0 = &\left(\frac{\partial f}{\partial y} - S_{t-}\xi_{t-}\right)^2 \bar{c} \\
&+ \int_{\mathbb{R}} \Bigg[f(t, Y_{t-} + y, \hat{\boldsymbol{X}}_{t-} + w(\hat{\boldsymbol{X}}_{t-}, y)) \\
&\qquad - f(t_-, Y_{t-}, \hat{\boldsymbol{X}}_{t-}) - \xi_{t-}S_{t-}(e^y - 1) \Bigg]^2 \hat{\boldsymbol{X}}_{t-}^\top \boldsymbol{F}(dy) \\
&+ (\nabla_x f)^\top \left(\int_{\mathbb{R}} \hat{\boldsymbol{X}}_{t-} diag(\tilde{w}^2) \hat{\boldsymbol{X}}_{t-}^\top \hat{\boldsymbol{X}}_{t-}^\top \boldsymbol{F}(dy) + \hat{\boldsymbol{X}}_{t-} diag(\tilde{a}(\hat{\boldsymbol{X}}_{t-})^2) \hat{\boldsymbol{X}}_{t-}^\top \right) \nabla_x f \\
&+ \frac{\partial}{\partial t} h + \frac{1}{2} \frac{\partial^2}{\partial y^2} h \bar{c} + (\nabla_x h)^\top \boldsymbol{A} \hat{\boldsymbol{X}}_{t-} + \sum_{i,j} \frac{\partial^2}{\partial x_i \partial x_j} h a^i(\hat{\boldsymbol{X}}_{t-}) a^j(\hat{\boldsymbol{X}}_{t-}) \\
&+ \int_{\mathbb{R}} \left(h(t, Y_{t-} + y, \hat{\boldsymbol{X}}_{t-} + w(\hat{\boldsymbol{X}}_{t-}, y)) - h(t, Y_{t-}, \hat{\boldsymbol{X}}_{t-}) - \frac{\partial}{\partial y} h(t, Y_{t-}, \hat{\boldsymbol{X}}_{t-}) y \right) \hat{\boldsymbol{X}}_{t-}^\top \boldsymbol{F}(dy)
\end{aligned}
$$

with terminal condition

$$h(T, Y_T, \hat{\boldsymbol{X}}_T) = (f(T, Y_T, \hat{\boldsymbol{X}}_T) - H(T, S_0 e^{Y_T}))^2$$

Proof. To pin down the function h, we will use the fact that

$$\beta_t := \int_0^t d\langle V - \xi \cdot S\rangle_u + \mathbb{E}[\int_t^T d\langle V - \xi \cdot S\rangle_u | \hat{\mathcal{F}}_t] = \int_0^t d\langle V - \xi \cdot S\rangle_u + h(t, Y_t, \hat{\boldsymbol{X}}_t)$$

is a martingale:

$$\mathbb{E}[\beta_t|\hat{\mathcal{F}}_s] = \mathbb{E}[\int_0^t d\langle V - \xi \cdot S\rangle_u + \mathbb{E}[\int_t^T d\langle V - \xi \cdot S\rangle_u|\hat{\mathcal{F}}_t]|\hat{\mathcal{F}}_s]$$

$$= \mathbb{E}[\int_0^s d\langle V - \xi \cdot S\rangle_u + \mathbb{E}[\int_s^T d\langle V - \xi \cdot S\rangle_u|\hat{\mathcal{F}}_t]|\hat{\mathcal{F}}_s]$$

$$= \int_0^s d\langle V - \xi \cdot S\rangle_u + \mathbb{E}[\int_s^T d\langle V - \xi \cdot S\rangle_u|\hat{\mathcal{F}}_s] = \beta_s$$

As a consequence, it holds that

$$\mathbb{E}[d\langle V - \xi \cdot S\rangle_t + dh|\hat{\mathcal{F}}_t] = 0 \tag{C.1}$$

We will use this relation to obtain a PIDE that determines h.

Using Itô formula for jump processes and the Kushner-Stratonovich equation in theorem 3.6 we have

$$dh(t, Y_t, \hat{\boldsymbol{X}}_t) = \frac{\partial}{\partial t}h \, dt$$

$$+ \frac{\partial}{\partial y}h \, dY_t + \frac{1}{2}\frac{\partial^2}{\partial y^2}h \, d\langle Y^c\rangle_t$$

$$+ (\nabla_x h)^\top d\hat{\boldsymbol{X}}_t + \frac{1}{2}\sum_{i,j}\frac{\partial^2}{\partial x_i \partial x_j}h d\langle \hat{X}^i, \hat{X}^j\rangle_t^c$$

$$+ (h(t, Y_{t-} + y, \hat{\boldsymbol{X}}_{t-} + w(\hat{\boldsymbol{X}}_{t-}, y)) - h(t, Y_{t-}, \hat{\boldsymbol{X}}_{t-}) - \nabla_x h^\top \boldsymbol{x} - \frac{\partial h}{\partial y}y) * J^{\hat{\boldsymbol{X}}, Y}$$

Then the expectation in equation C.1 leads to the following PIDE for h (with jump measure truncation function $h(y) = y$):

$$0 = \frac{d\langle V - \xi \cdot S\rangle}{dt} + \frac{\partial}{\partial t}h + \frac{1}{2}\frac{\partial^2}{\partial y^2}h \, \bar{c}$$

$$+ (\nabla_x h)^\top \boldsymbol{A}\hat{\boldsymbol{X}}_{t-} + \frac{1}{2}\sum_{i,j}\frac{\partial^2 h}{\partial x_i \partial x_j}a^i(\hat{\boldsymbol{X}}_{t-})a^j(\hat{\boldsymbol{X}}_{t-})$$

$$+ \int_{\mathbb{R}}\Bigg(h(t, Y_{t-} + y, \hat{\boldsymbol{X}}_{t-} + w(\hat{\boldsymbol{X}}_{t-}, y)) - h(t, Y_{t-}, \hat{\boldsymbol{X}}_{t-})$$

$$- \frac{\partial}{\partial y}h(t, Y_{t-}, \hat{\boldsymbol{X}}_{t-})y - \nabla_x h^\top w(\hat{\boldsymbol{X}}_{t-}, y)\Bigg)\hat{\boldsymbol{X}}_{t-}^\top \boldsymbol{F}(dy)$$

To completely define this PIDE, we need to compute $\frac{d\langle V - \xi \cdot S\rangle}{dt}$.

To obtain it, we will apply Itô's lemma to V, compute $V - \xi \cdot S$, and then from that directly obtain its predictable quadratic variation.

Writing $V_t = f(t, \hat{\boldsymbol{X}}_t, Y_t)$ and using notation $h_y(t, y, x) := \frac{\partial h}{\partial y}(t, y, x)$, Itô's lemma says:

$$
\begin{aligned}
V_T &= f(T, \hat{\boldsymbol{X}}_T, Y_T) \\
&= V(0, Y_0, \hat{\boldsymbol{X}}) + \int_0^T f_t(t, \hat{\boldsymbol{X}}_{t-}, Y_{t-})dt + \int_0^T f_y(t, \hat{\boldsymbol{X}}_{t-}, Y_{t-})dY_t \\
&\quad + \frac{1}{2}\int_0^T f_{yy}(t, \hat{\boldsymbol{X}}_{t-}, Y_{t-})d\langle Y^c\rangle_t \\
&\quad + \int_0^T \int_{\mathbb{R}} (f(t, X_{t-} + w(\hat{\boldsymbol{X}}_{t-}, y), Y_{t-} + y) - f(t, \hat{\boldsymbol{X}}_{t-}, Y_{t-}) \\
&\qquad - f_y(t, \hat{\boldsymbol{X}}_{t-}, Y_{t-})y - \nabla_x f(t, \hat{\boldsymbol{X}}_{t-}, Y_{t-})w(\hat{\boldsymbol{X}}_{t-}, y))J(dy, dt) \\
&\quad + \int_0^T (\nabla_x f(t, \hat{\boldsymbol{X}}_{t-}, Y_{t-}))^\top d\hat{\boldsymbol{X}}_t + \frac{1}{2}\int_0^T \sum_{i,j} \frac{\partial^2}{\partial x_i \partial x_j} f(t, \hat{\boldsymbol{X}}_{t-}, Y_{t-})d\langle \hat{X}^i, \hat{X}^j\rangle_t^c
\end{aligned}
$$

We then write S in terms of Y (using truncation function $h(y) = y$):

$$
dS = S_0 d(e^Y) = S_-\left(dY + \frac{1}{2}d\langle Y^c\rangle + \int_{\mathbb{R}} (e^y - 1 - y)J(dy, dt)\right)
$$

to get $V - \xi \cdot S$:

$$
\begin{aligned}
V_T - \int_0^T \xi_t dS_t &= f(T, \hat{\boldsymbol{X}}_T, Y_T) \\
&= f(0, \hat{\boldsymbol{X}}, Y_0) + \int_0^T f_t(t, \hat{\boldsymbol{X}}_{t-}, Y_{t-})dt + \int_0^T (f_y(t, \hat{\boldsymbol{X}}_{t-}, Y_{t-}) - \xi_{t-} S_{t-})dY_t \\
&\quad + \frac{1}{2}\int_0^T (h_{yy}^f(t, \hat{\boldsymbol{X}}_{t-}, Y_{t-}) - \xi_{t-} S_{t-})d\langle Y^c\rangle_t \\
&\quad + \int_0^T \int_{\mathbb{R}} \Big[f(t, \hat{\boldsymbol{X}}_{t-} + w(\hat{\boldsymbol{X}}_{t-}, y), Y_{t-} + y) \\
&\qquad - f(t, \hat{\boldsymbol{X}}_{t-}, Y_{t-}) - f_y(t, \hat{\boldsymbol{X}}_{t-}, Y_{t-})y - \nabla_x f(t, \hat{\boldsymbol{X}}_{t-}, Y_{t-})w(\hat{\boldsymbol{X}}_{t-}, y) \\
&\qquad - \xi_{t-} S_{t-}(e^y - 1 - y)\Big]J(dy, dt) \\
&\quad + \int_0^T (\nabla_x f(t, \hat{\boldsymbol{X}}_{t-}, Y_{t-}))^\top d\hat{\boldsymbol{X}}_t + \frac{1}{2}\int_0^T \sum_{i,j} \frac{\partial^2}{\partial x_i \partial x_j} f(t, \hat{\boldsymbol{X}}_{t-}, Y_{t-})d\langle \hat{X}^i, \hat{X}^j\rangle_t
\end{aligned}
$$

Then we can see the predictable quadratic variation of this is:

$$\langle V - \xi \cdot S\rangle_T = \int_0^T (f_y(t_-, \hat{\boldsymbol{X}}_{t-}, Y_{t-}) - S_{t-}\xi_{t-})^2 d\langle Y^c\rangle_t$$

$$+ \int_0^T \int_{\mathbb{R}} \Big[f(t, \hat{\boldsymbol{X}}_{t-} + w(\hat{\boldsymbol{X}}_{t-}, y), Y_{t-} + y)$$

$$- f(t_-, \hat{\boldsymbol{X}}_{t-}, Y_{t-}) - \xi_{t-}S_{t-}(e^y - 1) \Big]^2 \hat{\boldsymbol{X}}_{t-}^\top \boldsymbol{F}(dy)dt$$

$$+ \int_0^T f(t_-, \hat{\boldsymbol{X}}_{t-}, Y_{t-})^\top d\langle \hat{\boldsymbol{X}}\rangle_t f(t_-, \hat{\boldsymbol{X}}_{t-}, Y_{t-})$$

From the Kushner-Stratonovich equation 3.6 for $\hat{\boldsymbol{X}}$, we know that

$$d\langle \hat{\boldsymbol{X}}\rangle = (\int_{\mathbb{R}} \hat{\boldsymbol{X}}_- \text{diag}(\tilde{w}^2)\hat{\boldsymbol{X}}_-^\top \hat{\boldsymbol{X}}_-^\top \boldsymbol{F}(dy) + \hat{\boldsymbol{X}}_- \text{diag}(\tilde{a}(\hat{\boldsymbol{X}}_-)^2)\hat{\boldsymbol{X}}_-^\top)dt$$

This leads to a final PIDE defining the backward evolution of the quadratic hedging error (omitting the inputs to h, V, $\tilde{w}$ where clear):

$$0 = (\frac{\partial f}{\partial y} - S_{t-}\xi_{t-})^2 \bar{c}$$

$$+ \int_{\mathbb{R}} \Big[f(t, \hat{\boldsymbol{X}}_{t-} + w(\hat{\boldsymbol{X}}_{t-}, y), Y_{t-} + y)$$

$$- f(t_-, \hat{\boldsymbol{X}}_{t-}, Y_{t-}) - \xi_{t-}S_{t-}(e^y - 1) \Big]^2 \hat{\boldsymbol{X}}_{t-}^\top \boldsymbol{F}(dy)$$

$$+ (\nabla_x f)^\top (\int_{\mathbb{R}} \hat{\boldsymbol{X}}_{t-} \text{diag}(\tilde{w}^2)\hat{\boldsymbol{X}}_{t-}^\top \hat{\boldsymbol{X}}_{t-}^\top \boldsymbol{F}(dy) + \hat{\boldsymbol{X}}_{t-} \text{diag}(\tilde{a}(\hat{\boldsymbol{X}}_-)^2)\hat{\boldsymbol{X}}_{t-}^\top)\nabla_x f$$

$$+ \frac{\partial}{\partial t}h + \frac{1}{2}\frac{\partial^2}{\partial y^2}h\,\bar{c} + (\nabla_x h)^\top \boldsymbol{A}\hat{\boldsymbol{X}}_{t-} + \sum_{i,j} \frac{\partial^2 h}{\partial x_i \partial x_j}a^i(\hat{\boldsymbol{X}}_{t-})a^j(\hat{\boldsymbol{X}}_{t-})$$

$$+ \int_{\mathbb{R}} \Big(h(t, Y_{t-} + y, \hat{\boldsymbol{X}}_{t-} + w(\hat{\boldsymbol{X}}_{t-}, y)) - h(t, Y_{t-}, \hat{\boldsymbol{X}}_{t-})$$

$$- \frac{\partial}{\partial y}h(t, Y_{t-}, \hat{\boldsymbol{X}}_{t-})y - \nabla_x h^\top w(\hat{\boldsymbol{X}}_{t-}, y) \Big) \hat{\boldsymbol{X}}_{t-}^\top \boldsymbol{F}(dy)$$

We remind the reader that in this equation, we aim to solve for function h as we want to obtain $\varepsilon_0^2 = h(0, \hat{\boldsymbol{X}}_0, Y_0)$, whereas $V = f(t, \hat{\boldsymbol{X}}_t, Y_t)$ is known. $\square$

It is worth pointing out that the dimensionality of the PIDE grows with the number of regimes. Already with 2 regimes, we have at best a 2-dimensional PIDE (with dimensions $\hat{X}_t^1, Y_t$, eliminating 1 dimension by noting that $\hat{X}_t^2 = 1 - \hat{X}_t^1$). Thus we are better served by using our recursive approximation scheme for this computation,

especially for more than 2-3 regimes. Alternatively we can resort to computing the quadratic hedging error via Monte Carlo simulations.

As there is no self-evident solution to the PIDE that gives us the hedging error, in the next subsection we outline a finite-difference scheme to solve it numerically. We do this for the case of 2 regimes, as this illustrates the principles behind higher-dimensional solutions but keeps the indexing and notation manageable.

C.3.1   Finite difference scheme for 2 regimes

We recall the PIDE:

$$
\begin{aligned}
0 = {} & (\frac{\partial f}{\partial y} - S_{t-}\xi_{t-})^2 \bar{c} \\
& + \int_{\mathbb{R}} \Big[f(t, \hat{\boldsymbol{X}}_{t-} + w(\hat{\boldsymbol{X}}_{t-}, y), Y_{t-} + y) \\
& \qquad - f(t_-, \hat{\boldsymbol{X}}_{t-}, Y_{t-}) - \xi_{t-}S_{t-}(e^y - 1) \Big]^2 \hat{\boldsymbol{X}}_{t-}^{\top} \boldsymbol{F}(dy) \\
& + (\nabla_x f)^{\top} (\int_{\mathbb{R}} \hat{\boldsymbol{X}}_{t-}\mathrm{diag}(\tilde{\boldsymbol{w}}^2)\hat{\boldsymbol{X}}_{t-}^{\top}\hat{\boldsymbol{X}}_{t-}^{\top}\boldsymbol{F}(dy) + \hat{\boldsymbol{X}}_{t-}\mathrm{diag}(\tilde{a}(\hat{\boldsymbol{X}}_{t-})^2)\hat{\boldsymbol{X}}_{t-}^{\top})\nabla_x f \\
& + \frac{\partial}{\partial t}h + \frac{1}{2}\frac{\partial^2}{\partial y^2}h\,\bar{c} + (\nabla_x h)^{\top}\boldsymbol{A}\hat{\boldsymbol{X}}_{t-} + \sum_{i,j} \frac{\partial^2}{\partial x_i \partial x_j}ha^i(\hat{\boldsymbol{X}}_{t-})a^j(\hat{\boldsymbol{X}}_{t-}) \\
& + \int_{\mathbb{R}} \Big(h(t, Y_{t-} + y, \hat{\boldsymbol{X}}_{t-} + w(\hat{\boldsymbol{X}}_{t-}, y)) - h(t, Y_{t-}, \hat{\boldsymbol{X}}_{t-}) - \frac{\partial}{\partial y}h(t, Y_{t-}, \hat{\boldsymbol{X}}_{t-})y \Big) \hat{\boldsymbol{X}}_{t-}^{\top}\boldsymbol{F}(dy).
\end{aligned}
$$

We notice that can define a drift term $d(t, \hat{\boldsymbol{X}}, Y)$ independent of the solution h:

$$
\begin{aligned}
d(t, \hat{\boldsymbol{X}}_{t-}, Y_{t-}) := {} & (\frac{\partial f}{\partial y} - S_{t-}\xi_{t-})^2 \bar{c} \\
& + \int_{\mathbb{R}} \Big[f(t, \hat{\boldsymbol{X}}_{t-} + w(\hat{\boldsymbol{X}}_{t-}, y), Y_{t-} + y) \\
& \qquad - f(t_-, \hat{\boldsymbol{X}}_{t-}, Y_{t-}) - \xi_{t-}S_{t-}(e^y - 1) \Big]^2 \hat{\boldsymbol{X}}_{t-}^{\top} \boldsymbol{F}(dy) \\
& + (\nabla_x f)^{\top} (\int_{\mathbb{R}} \hat{\boldsymbol{X}}_{t-}\mathrm{diag}(\tilde{\boldsymbol{w}}^2)\hat{\boldsymbol{X}}_{t-}^{\top}\hat{\boldsymbol{X}}_{t-}^{\top}\boldsymbol{F}(dy) + \hat{\boldsymbol{X}}_{t-}\mathrm{diag}(\tilde{a}(\hat{\boldsymbol{X}}_{t-})^2)\hat{\boldsymbol{X}}_{t-}^{\top})\nabla_x f
\end{aligned}
$$

This drift term can be pre-computed completely separately from any finite difference scheme solution.

Next, we simplify the solution for $h(t, Y, \hat{\boldsymbol{X}}) = h(t, Y, \hat{X}^1, \hat{X}^2)$ by noting that $\hat{X}^2 = 1 - \hat{X}^1$ and hence $h(t, Y, \hat{X}^1, \hat{X}^2) = h(t, Y, \hat{X}^1, (1 - \hat{X}^1)) =: \tilde{h}(t, Y, \hat{X}^1)$. We will continue this section by using this variant of $h(t, y, x)$ with 1-dimensional inputs. Using the simplifying relation above, and denoting the elements of vector $a(\hat{\boldsymbol{X}})$ as a^1, a^2, the PIDE can be written as follows:

$$
0 = d(t, x, y) + \frac{\partial h}{\partial t} + \frac{1}{2}\bar{c}\frac{\partial^2 h}{\partial y^2}
$$
$$
+ \frac{\partial h}{\partial x}((a_{11} - a_{21})x + (a_{12} - a_{22})(1 - x)) + \frac{\partial^2 h}{\partial x^2}(a^1 - a^2)^2
$$
$$
+ \int_{\mathbb{R}} (h(t, Y_- + y, \hat{X}^1_- + w(\hat{X}^1_-, y)) - h(t, Y_-, \hat{X}^1_-) - \frac{\partial h}{\partial y}(t, Y_-, \hat{X}^1_-)y - \nabla_x h^\top w(\hat{\boldsymbol{X}}_-, y))\hat{\boldsymbol{X}}^\top_- \boldsymbol{F}(dy)
$$

Now we define 3 grids on which we will solve the PIDE numerically:

- time grid: indexed $0, ..., t, t + 1, ..., T$,

- y grid: indexed $0, ..., n, n + 1, ..., N$,

- x grid: indexed $0, ..., m, m + 1, ..., M$

(note here M denotes how finely we can distinguish between different levels of $\hat{X}^1$, not the number of regimes) and denote $h_{t,n,m} := h(t, y_n, x_m)$. We then use an explicit finite difference implementation and central finite differences:

$$
\frac{\partial h}{\partial t} \approx \frac{h_{t,n,m} - h_{t-1,n,m}}{\Delta t}
$$
$$
\frac{\partial h}{\partial x} \approx \frac{h_{t,n,m+1} - h_{t,n,m-1}}{2\Delta x}
$$
$$
\frac{\partial^2 h}{\partial x^2} \approx \frac{h_{t,n,m+1} - 2h_{t,n,m} + h_{t,n,m-1}}{(\Delta x)^2}
$$
$$
\frac{\partial^2 h}{\partial y^2} \approx \frac{h_{t,n+1,m} - 2h_{t,n,m} + h_{t,n-1,m}}{(\Delta y)^2}.
$$

Using these approximations within our PIDE, we obtain an explicit recursive scheme

solving backward in time:

$$h_{t-1,n,m} = \Delta t \bigg(d_{t,n,m} + h_{t,n,m} + \frac{1}{2\Delta x}(h_{t,n,m+1} - h_{t,n,m-1})((a_{11} - a_{21})x_1 + (a_{12} - a_{22})(1 - x_1))$$

$$+ \frac{1}{2(\Delta y)^2}\bar{c}(h_{t,n+1,m} - 2h_{t,n,m} + h_{t,n-1,m})(a^1_{t,m} - a^2_{t,m})^2$$

$$+ \int_{\mathbb{R}}(h(t, y_n + z, x_m + w^1(x_m, z)) - h_{t,n,m}$$

$$- z\frac{h_{t,n+1,m} - h_{t,n-1,m}}{2\Delta y} - w^1(x_m, z)\frac{h_{t,n,m+1} - h_{t,n,m-1}}{2\Delta x})(x_m F^1 + (1 - x_m)F^2)(dz)$$

The terminal condition (ie. boundary on the time grid) is $h_{T,n,m} = (f_{T,n,m} - H_{T,n})^2$. For the boundaries on indices m, n we use the smooth-pasting condition $\frac{\partial^2 h}{\partial x^2} = 0, \frac{\partial^2 h}{\partial y^2} = 0$, which translates to the practical boundary conditions $h_{t,M,n} = 2h_{t,M-1,n} - h_{t,M-2,n}$ and $h_{t,m,N} = 2h_{t,m,N-1} - h_{t,m,N-2}$.

References

Milton Abramowitz and Irene A. Stegun, editors. *Handbook of Mathematical Function with Formulas, Graphs, and Mathematical Tables*, volume 55 of *Applied Mathematics Series*. National Bureau of Standards, 10th edition, 1972.

Aristotle. *Politics.* 1999. ISBN 9780486414249. URL `http://socserv2.socsci.mcmaster.ca/econ/ugcm/3ll3/aristotle/Politics.pdf`.

Marco Avellaneda, A Levy, and A Parás. Pricing and hedging derivative securities in markets with uncertain volatilities. *Applied Mathematical Finance*, 1995. URL `http://www.tandfonline.com/doi/abs/10.1080/13504869500000005`.

Elie Ayache. The irony in the variance swaps. *Wilmott Magazine*, pages 16–23, September 2006.

Louis Bachelier. *Théorie de la spéculation.* PhD thesis, L'École Normale Supérieure, 1900.

Alan Bain and Dan Crisan. *Fundamentals of Stochastic Filtering*. Springer Verlag, 2009.

Dirk Becherer. Bounded solutions to backward SDEs with jumps for utility optimization and indifference hedging. *The Annals of Applied Probability*, 16:2027–2054, 2006.

Aharon Ben-Tal and Marc Teboulle. An old-new concept of convex risk measures: The optimized certainty equivalent. *Mathematical Finance*, 17(3):449–476, 2007a. ISSN 09601627. doi: 10.1111/j.1467-9965.2007.00311.x.

Aharon Ben-Tal and Marc Teboulle. An old-new concept of convex risk measures: The optimized certainty equivalent. *Mathematical Finance*, 17(3):449–476, July 2007b.

Dimitris Bertsimas, Leonid Kogan, and Andrew W. Lo. When is time continuous? *Journal of Financial Economics*, 55:173–204, 2000.

Sara Biagini and Aleš Černý. Admissible strategies in semimartingale portfolio selection. *SIAM Journal on Control and Optimization*, 49(1):42–72, 2011.

Patrick Billingsley. *Probability and Measure.* John Wiley & Sons, third edition, 1995.

Tomas Björk and Irina Slinko. Towards a General Theory of Good-Deal Bounds. *Review of Finance*, 10(2):221–260, 2006. ISSN 1572-3097. doi: 10.1007/s10679-006-8279-1. URL `http://rof.oxfordjournals.org/cgi/doi/10.1007/s10679-006-8279-1`.

Fisher Black and Myron Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81:637–654, 1973.

Nina Boyarchenko, Mario Cerrato, John Crosby, and Stewart Hodges. No Good Deals - No Bad Models Federal Reserve Bank of New York. 2014.

Phelim P. Boyle and David Emanuel. Discretely adjusted option hedges. *Journal of Financial Economics*, 8(3):259 – 282, 1980.

Pierre Brémaud, editor. *Point Processes and Queues: Martingale Dynamics.* Springer Verlag, 1981.

Pierre Brémaud, editor. *Markov Chains: Gibbs fields, Monte Carlo simulation and Queues.* Springer Verlag, 1999.

Mats Broden and Peter Tankov. Tracking errors from discrete hedging in exponential Levy models. *International Journal of Theoretical and Applied Finance*, 14(6): 803–837, 2011. URL `http://arxiv.org/abs/1003.0709`.

Chris Brooks, Joelle Miffre, and Aleš Černý. Optimal hedging with higher moments. *Journal of Futures Markets*, 32(10):909–944, 2012. doi: 10.1002/fut. URL `http://onlinelibrary.wiley.com/doi/10.1002/fut.20542/full`.

Peter Carr and Roger Lee. Variation and share-weighted variation swaps on time-changed Lévy processes. *Finance and Stochastics*, 17(4):685–716, jul 2013. ISSN 0949-2984. doi: 10.1007/s00780-013-0212-9. URL `http://link.springer.com/10.1007/s00780-013-0212-9`.

Peter Carr and Dilip Madan. Towards a theory of volatility trading. In Robert Jarrow, editor, *Volatility: New Estimation Techniques for Pricing Derivatives*, pages 417–427. Risk Publications, 2002.

Peter Carr, Roger Lee, and Liuren Wu. Variance swaps on time-changed lévy processes. *Finance and Stochastics*, 16(2):335–355, 2012.

Álvaro Cartea and Sebastian Jaimungal. Modelling Asset Prices for Algorithmic and High-Frequency Trading. *Applied Mathematical Finance*, 20(6):512–547, dec 2013. ISSN 1350-486X. doi: 10.1080/1350486X.2013.771515. URL `http://www.tandfonline.com/doi/abs/10.1080/1350486X.2013.771515`.

CBOE. Vix white paper, 2009.

Claudia Ceci and Katia Colaneri. Nonlinear filtering for jump diffusion observations. *Advances in Applied Probability*, 44(3):678–701, 2012. ISSN 00018678. doi: 10. 1239/aap/1346955260.

Claudia Ceci and Katia Colaneri. The Zakai equation of nonlinear filtering for jump-diffusion observation: existence and uniqueness. *Applied Mathematics & Optimization*, 69(1):47–82, 2014. URL `http://arxiv.org/abs/1210.4279`.

Claudia Ceci, Katia Colaneri, and Alessandra Cretarola. Local risk-minimization under restricted information to asset prices. *Electronic Journal of Probability*, 20, 2015. URL `http://arxiv.org/abs/1312.4385`.

Aleš Černý and Stewart Hodges. The theory of good-deal pricing in financial markets. In Hélyette Geman, Dilip B. Madan, and Stanley R Pliska, editors, *Mathematical FinanceBachelier Congress 2000*, pages 175–202. 2002. URL `http://link.springer.com/chapter/10.1007/978-3-662-12429-1{_}9`.

Aleš Černý and Jan Kallsen. Hedging By Sequential Regressions Revisited. *Mathematical Finance*, 19(4):591–617, oct 2009. ISSN 09601627. doi: 10.1111/j.1467-9965.2009.00381.x. URL `http://doi.wiley.com/10.1111/j.1467-9965.2009.00381.x`.

Aleš Černý, Fabio Maccheroni, Massimo Marinacci, and Aldo Rustichini. On the computation of optimal monotone meanvariance portfolios via truncated quadratic utility. *Journal of Mathematical Economics*, 48(6):386–395, dec 2012. ISSN 03044068. doi: 10.1016/j.jmateco.2012.08.006. URL `http://linkinghub.elsevier.com/retrieve/pii/S0304406812000614`.

Aleš Černý. Generalized sharpe ratios and asset pricing in incomplete markets. *European Finance Review*, 7:161–233, 2003.

Aleš Černý. Optimal continuous-time hedging with leptokurtic returns. *Mathematical finance*, 17(2):175–203, 2007.

Aleš Černý. *Mathematical techniques in finance: Tools for incomplete markets.* Princeton University Press, 2nd edition, 2009.

Aleš Černý and Jan Kallsen. On the structure of general mean-variance hedging strategies. *The Annals of Probability*, 35(4):1479–1531, 2007.

Aleš Černý and Juraj Špilda. A note on "discrete time hedging errors for options with irregular payoffs". Electronic working paper published at http://ssrn.com/abstract=2042519, 2012.

Aleš Černý, Stephan Denkl, and Jan Kallsen. Hedging in lévy models and the time step equivalent of jumps. *preprint*, 2013.

Patrick Cheridito and Michael Kupper. Recursiveness of indifference prices and translation-invariant preferences. *Mathematics and Financial Economics*, 2(3): 173–188, 2009. ISSN 1862-9679. doi: 10.1007/s11579-009-0020-3. URL `http://link.springer.com/10.1007/s11579-009-0020-3`.

Patrick Cheridito, Ulrich Horst, Michael Kupper, and Traian A Pirvu. Equilibrium Pricing in Incomplete Markets under Translation Invariant Preferences . *Mathematics of Operations Research*, 41(1):174–195, 2015. ISSN 15265471. doi: 10.1287/moor.2015.0721.

Alexander Cherny and Michael Kupper. Divergence utilities. 2007. URL `https://papers.ssrn.com/sol3/papers.cfm?abstract{_}id=1023525`.

Julien Chevallier and Stéphane Goutte. Statistical method to estimate regime-switching Levy models. In *Stochastic Models, Statistics and Their Applications: Springer Proceedings in Mathematics and Statistics*. Springer, 2014. URL `https://halshs.archives-ouvertes.fr/halshs-01090825`.

Julien Chevallier and Stéphane Goutte. Detecting jumps and regime-switches in international stock markets returns. *Applied Economics Letters*, 22(13):1011–1019, 2015. URL `https://hal.archives-ouvertes.fr/hal-01090833`.

Kyriakos Chourdakis. Switching Levy Models in continuous time: Finite distributions and option pricing. 2005. URL `https://papers.ssrn.com/sol3/papers.cfm?abstract{_}id=838924`.

John H. Cochrane and Jesús Saá-Requejo. Beyond arbitrage: Good-deal asset price bounds in incomplete markets. *The Journal of Political Economy*, 108(1):79–119, 2000.

Samuel N. Cohen and Robert J. Elliott. Solutions of Backward Stochastic Differential Equations on Markov Chains. *Communications on Stochastic Analysis*, 2(2):251–262, 2008. URL `http://arxiv.org/abs/0809.5102`.

Rama Cont and Ekaterina Voltchkova. A finite difference scheme for option pricing in jump diffusion and exponential Lévy models. *SIAM Journal on Numerical Analysis*, 43(4):1596–1626, 2005. ISSN 0036-1429. doi: 10.1137/S0036142903436186. URL `http://epubs.siam.org/doi/abs/10.1137/S0036142903436186`.

Rama Cont, Peter Tankov, and Ekaterina Volotchkova. Hedging with options in models with jumps. In F. Benth, G. Di Nunno, T. Lindstrom, B.Øksendal, and T. Zhang, editors, *Stochastic Analysis and Applications - the Abel Symposium 2005 in honor of Kiyosi Itô*, pages 197–218. Springer, Berlin, 2007.

Jose Manuel Corcuera, David Nualart, and Wim Schoutens. Moment derivatives and levy-type market completion. In A.E. Kyprianou, Wim Schoutens, and Paul Wilmott, editors, *Exotic Option Pricing and Advanced Lévy models*, pages 67–97. Wiley, 2005.

John Crosby and Mark Davis. Variance derivatives: Pricing and convergence. http://www.john-crosby.co.uk/pdfs/VarianceDerivatives.pdf, 2011.

Mark Davis, Jan Obłój, and Vimal Raval. Arbitrage bounds for prices of weighted variance swaps. *Mathematical Finance*, 24(4):821–854, 2014.

Freddy Delbaen, Peter Grandits, Thorsten Rheinländer, Dominick Samperi, Martin Schweizer, and Christophe Stricker. Exponential Hedging and Entropic Penalties. *Mathematical Finance*, 12(2):99–123, apr 2002. ISSN 0960-1627. doi: 10.1111/1467-9965.02001. URL `http://doi.wiley.com/10.1111/1467-9965.02001`.

Kresimir Demeterfi, Emanuel Derman, Michael Kamal, and Joseph Zou. More than you ever wanted to know about volatility swaps. Technical report, Goldman Sachs, 1999.

Stephan Denkl, Martina Goy, Jan Kallsen, Johannes Muhle-Karbe, and Arnd Pauwels. On the performance of delta hedging strategies in exponential lévy models. *Quantitative Finance*, 13(8):1173–1184, 2013.

Giuseppe Di Graziano and L. C. G. Rogers. Equity with Markov-modulated dividends. *Quantitative Finance*, 9(1):19–26, 2009. ISSN 1469-7688. doi: 10.1080/14697680802036168.

Bruno Dupire. Arbitrage pricing with stochastic volatility. Technical report, Banque Paribas, 1992.

Robert J. Elliott and Guang-Hua Lian. Pricing variance and volatility swaps in a stochastic volatility model with regime switching: discrete observations case. *Quantitative Finance*, 13(5):687–698, may 2013. ISSN 1469-7688. doi: 10.1080/14697688.2012.676208. URL http://www.tandfonline.com/doi/abs/10.1080/14697688.2012.676208.

Robert J. Elliott and Carlton-James U. Osakwe. Option Pricing for Pure Jump Processes with Markov Switching Compensators. *Finance and Stochastics*, 10 (2):250–275, apr 2006. ISSN 0949-2984. doi: 10.1007/s00780-006-0004-6. URL http://link.springer.com/10.1007/s00780-006-0004-6.

Robert J. Elliott and Andrew J. Royal. Asset prices with regime-switching variance gamma dynamics. In P.G. Ciarlet, editor, *Handbook of Numerical Analysis*, volume XV, pages 685–711. Elsevier, 2008.

Robert J. Elliott and Tak Kuen Siu. On risk minimizing portfolios under a Markovian regime-switching Black-Scholes economy. *Annals of Operations Research*, 176 (1):271–291, oct 2008. ISSN 0254-5330. doi: 10.1007/s10479-008-0448-5. URL http://link.springer.com/10.1007/s10479-008-0448-5.

Robert J. Elliott and Tak Kuen Siu. Attainable Contingent Claims in a Markovian Regime-Switching Market. *International Journal of Theoretical and Applied Finance*, 15(08):1250055, dec 2012. ISSN 0219-0249. doi: 10.1142/S0219024912500550. URL http://www.worldscientific.com/doi/abs/10.1142/S0219024912500550.

Robert J. Elliott and Tak Kuen Siu. Option Pricing and Filtering with Hidden Markov-Modulated Pure-Jump Processes. *Applied Mathematical Finance*,

20(1):1–25, 2013. ISSN 1350-486X. doi: 10.1080/1350486X.2012.655929. URL http://www.tandfonline.com/doi/abs/10.1080/1350486X.2012.655929.

Robert J. Elliott, Lakhdar Aggoun, and John B. Moore. *Hidden Markov Models: Estimation and Control*. Springer, 1995. ISBN 9780387943640.

Robert J. Elliott, Leunglung Chan, and Tak Kuen Siu. Option pricing and Esscher transform under regime switching. *Annals of Finance*, 1(4): 423–432, 2005. ISSN 1614-2446. doi: 10.1007/s10436-005-0013-z. URL http://link.springer.com/10.1007/s10436-005-0013-z.

Robert J. Elliott, Tak Kuen Siu, Leunglung Chan, and John W. Lau. Pricing Options Under a Generalized Markov-Modulated Jump-Diffusion Model. *Stochastic Analysis and Applications*, 25(4):821–843, jun 2007. ISSN 0736-2994. doi: 10.1080/07362990701420118. URL http://www.tandfonline.com/doi/abs/10.1080/07362990701420118.

Robert J. Elliott, Tak Kuen Siu, and Alex Badescu. On mean-variance portfolio selection under a hidden Markovian regime-switching model. *Economic Modelling*, 27(3):678–686, 2010. ISSN 02649993. doi: 10.1016/j.econmod.2010.01.007. URL http://linkinghub.elsevier.com/retrieve/pii/S0264999310000088.

Robert J. Elliott, Leunglung Chan, and Tak Kuen Siu. Option valuation under a regime-switching constant elasticity of variance process. *Applied Mathematics and Computation*, 219(9):4434–4443, jan 2013. ISSN 00963003. doi: 10.1016/j.amc.2012.10.047. URL http://linkinghub.elsevier.com/retrieve/pii/S0096300312010636.

D Filipović and Michael Kupper. Monotone and cash-invariant convex functions and hulls. *Insurance: Mathematics and Economics*, 41(1):1–16, 2007. URL http://www.sciencedirect.com/science/article/pii/S0167668706001181.

D Filipović and Michael Kupper. Equilibrium prices for monetary utility functions. *International Journal of Theoretical and Applied Finance*, 11(3):1–20, 2008. URL http://www.worldscientific.com/doi/abs/10.1142/S0219024908004828.

Hans Föllmer and Philip E Protter. Local martingales and filtration shrinkage. *ESAIM: Probability and Statistics*, pages 1–21, 2010. ISSN 1292-8100. doi: 10.1051/ps/2010023. URL http://ganymedes.lib.unideb.hu:8080/udpeer/handle/2437.2/17408.

Hans Föllmer and Alexander Schied. Convex measures of risk and trading constraints. *Finance and stochastics*, 6(4):429–447, 2002. URL `http://link.springer.com/article/10.1007/s007800200072`.

Hans Föllmer and Martin Schweizer. Hedging by sequential regression: An introduction to the mathematics of option trading. *ASTIN Bulletin*, 18:147–160, 1989.

Hans Föllmer and Martin Schweizer. Hedging of contingent claims under incomplete information. In Mark H.A. Davis and Robert J. Elliott, editors, *Applied Stochastic Analysis*, volume 5, pages 389–414. Gordon and Breach, London/New York, 1990.

Hans Föllmer and Dieter Sondermann. *Hedging of nonredundant contingent claims*, chapter 12, pages 205–223. North Holland, Amsterdam, 1986.

Martin Forde. The Real P&L in Black-Scholes and Dupire Delta Hedging. *International Journal of Theoretical and Applied Finance*, 2003. URL `http://www.statistics.bristol.ac.uk/research/stats/reports/2003/0312.pdf`.

R Frey and T Schmidt. Pricing and Hedging of Credit Derivatives via the Innovations Approach to Nonlinear Filtering. *Finance and Stochastics*, 16(1):105–133, 2012. URL `http://www.tu-chemnitz.de/mathematik/fima/publikationen/frey-schmidt-pricingand`

Tsukasa Fujiwara. From the Minimal Entropy Martingale Measures to the Optimal Strategies for the Exponential Utility Maximization: the Case of Geometric Lévy Processes. *Asia-Pacific Financial Markets*, 11(4):367–391, jul 2006. ISSN 1387-2834. doi: 10.1007/s10690-006-9019-4. URL `http://www.springerlink.com/index/10.1007/s10690-006-9019-4`.

Tsukasa Fujiwara and Yoshio Miyahara. The minimal entropy martingale measures for geometric Levy processes. *Finance and Stochastics*, 7(4):509–531, oct 2003. ISSN 0949-2984. doi: 10.1007/s007800200097. URL `http://www.springerlink.com/Index/10.1007/s007800200097`.

Stefan Geiss. Quantitative approximation of certain stochastic integrals. *Stochastics and Stochastics Reports*, 73:241–270, 2002.

Emmanuel Gobet and Azmi Makhlouf. The tracking error rate of the delta-gamma hedging strategy. *Mathematical Finance*, 2(2):277–309, April 2012.

Emmanuel Gobet and Emmanuel Temam. Discrete time hedging errors for options with irregular payoffs. *Finance and Stochastics*, 5(3):357–367, jul 2001a. ISSN 0949-2984. doi: 10.1007/PL00013539. URL `http://www.springerlink.com/index/10.1007/PL00013539`.

Emmanuel Gobet and Emmanuel Temam. Discrete time hedging errors for options with irregular payoffs. *Finance and Stochastics*, 5:357–367, 2001b.

Stephane Goutte and Benteng Zou. Continuous time regime switching model applied to foreign exchange rate. *Mathematical Finance Letters*, 8:1–37, 2013. URL `http://hal.inria.fr/docs/00/64/39/00/PDF/FX{_}RS{_}ZG.pdf`.

Stéphane Goutte, Nadia Oudjane, and Francesco Russo. Variance Optimal Hedging for continuous time additive processes and applications. *Stochastics An International Journal of Probability and Stochastic Processes*, 86(1): 147–185, 2014. ISSN 1744-2508. doi: 10.1080/17442508.2013.774402. URL `http://arxiv.org/abs/0912.0372`.

M. R. Grasselli and T. R. Hurd. Indifference Pricing and Hedging for Volatility Derivatives. *Applied Mathematical Finance*, 14(4):303–317, 2007. ISSN 1350-486X. doi: 10.1080/13527260600963851. URL `http://www.tandfonline.com/doi/abs/10.1080/13527260600963851`.

Rainer Haidinger and Richard Warnung. Risk Measures in a Regime Switching Model Capturing Stylized Facts. 2012. URL `https://arxiv.org/abs/1212.4126`.

Donatien Hainaut. Financial modeling with switching Lévy processes. 2011. URL `http://www.crest.fr/ckfinder/userfiles/files/Pageperso/dhainaut/HainautSwitching`

Donatien Hainaut and David B. Colwell. A structural model for credit risk with switching processes and synchronous jumps. *The European Journal of Finance*, (June):1–23, 2014. ISSN 1351-847X. doi: 10.1080/1351847X.2014.924079. URL `http://www.tandfonline.com/doi/full/10.1080/1351847X.2014.924079`.

Donatien Hainaut and O Le Courtois. An intensity model for credit risk with switching Lévy processes. *Quantitative Finance*, 14(8):1453–1465, 2013. URL `http://www.tandfonline.com/doi/abs/10.1080/14697688.2012.756583`.

Donatien Hainaut and Christian Y Robert. Credit risk valuation with rating transitions and partial information. *International*

Journal of Theoretical and Applied Finance, 17(7), 2014. URL `http://www.crest.fr/ckfinder/userfiles/files/pageperso/dhainaut/PaperPartialInfo`

James Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, March 1989.

L.P. Hansen and T.J. Sargent. *Robustness*. Princeton University Press, 2008. ISBN 9780691114422.

Mary R Hardy. A Regime-Switching Model of Long-Term Stock Returns. *North American Actuarial Journal*, 5(2):41–53, 2001.

Peter Harremoës. An interpretation of Squared Skewness minus Excess Kurtosis. 2000.

Espen Gaarder Haug and Nassim Nicholas Taleb. Option traders use (very) sophisticated heuristics, never the black-scholes-merton formula. *Journal of Economic Behavior and Organization*, 77(2):97–106, 2011.

Takaki Hayashi and Per A. Mykland. Evaluating hedging errors: an asymptotic approach. *Mathematical finance*, 15(2):309–343, April 2005.

Vicky Henderson and David Hobson. Utility indifference pricing - an overview. In René Carmona, editor, *Indifference Pricing: Theory and Applications*. Princeton University Press, 2009.

Steven L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2): 327, 1993. doi: 10.1093/rfs/6.2.327.

David Hobson and Martin Klimmek. Model independent hedging strategies for variance swaps. To appear in Finance and Stochastics, http://arxiv.org/abs/1104.4010v2, 2011.

Stewart Hodges and Anthony Neuberger. Optimal replication of contingent claims under transaction costs. *Review of Futures Markets*, 8(2):222–242, 1989.

Friedrich Hubalek, Jan Kallsen, and Leszek Krawczyk. Variance-optimal hedging for processes with stationary independent increments. *The Annals of Applied Probability*, 16(2):853–885, may 2006. ISSN 1050-5164. doi: 10.1214/105051606000000178. URL `http://projecteuclid.org/Dienst/getRecord?id=euclid.aoap/1151592253/`.

Julien Hugonnier, Dmitry Kramkov, and Walter Schachermayer. On utility-based pricing of contingent claims in incomplete markets. *Mathematical Finance*, 15(2): 203–212, 2005. ISSN 09601627. doi: 10.1111/j.0960-1627.2005.00217.x.

Jean Jacod and Albert N. Shiryaev. *Limit Theorems for Stochastic Processes*. Springer-Verlag, 2 edition, 2003.

Victor Kac and Pokman Cheung. *Quantum Calculus*. Springer Verlag, 1st edition, 2002.

Jan Kallsen. Optimal portfolios for exponential Lévy processes. *Mathematical methods of operations research*, 51:357–374, 2000. URL http://link.springer.com/article/10.1007/s001860000048.

Jan Kallsen and Thorsten Rheinländer. Asymptotic utility-based pricing and hedging for exponential utility. *Statistics and Decisions*, 28:17–36, 2011.

Jan Kallsen and Albert Shiryaev. The cumulant process and Esscher's change of measure. *Finance and Stochastics*, 6(2):397–428, 2002. URL http://www.springerlink.com/index/57U0NYD9FWMFT6N6.pdf.

Jan Kallsen, Johannes Muhle-Karbe, Natalia Shenkman, and Richard Vierthauer. Discrete-time variance-optimal hedging in affine stochastic volatility models. *Alternative Investments and Strategies. World Scientific, Singapore*, 2009.

Constantinos Kardaras. No-free-lunch equivalences for exponential Lévy models under convex constraints on investment. *Mathematical Finance*, 19(2):161–187, 2009. URL http://onlinelibrary.wiley.com/doi/10.1111/j.1467-9965.2009.00363.x/full.

N Karoui, M Jeanblanc-Picquè, and S E Shreve. Robustness of the Black and Scholes Formula. *Mathematical Finance*, 8(2):93–126, 1998. ISSN 1467-9965. doi: 10.1111/1467-9965.00047.

Stefan Kassberger and Thomas Liebmann. q-optimal martingale measures for exponential lévy processes. *Finance and Stochastics*, 15(1):117–140, 2011.

Offer Kella and Marc Yor. Unifying the Dynkin and Lebesgue-Stieltjes formulae. *Journal of Applied Probability*, 52(1), 2017. URL http://arxiv.org/abs/1308.5795.

Joerg Kienitz and Daniel Wetterau. *Financial Modelling: Theory, Implementation and Practice (with Matlab source)*. Wiley Finance, 2012.

Young Shin Kim, Business Engineering, Frank J Fabozzi, Zuodong Lin, and Svetlozar T Rachev. Option pricing and hedging under a stochastic volatility Levy process model. *Review of Derivatives Research*, 15(1):81–97, 2011.

Susanne Klöppel and Martin Schweizer. Dynamic utility-based good-deal bounds. *Statistics and Decisions*, 25:285–309, 2008.

Roman Kozhan, Anthony Neuberger, and Paul Schneider. The skew risk premium in the equity index market. *Review of Financial Studies*, 26(9):2174–2203, 2013. ISSN 08939454. doi: 10.1093/rfs/hht039.

Dmitry Kramkov and Mihai Sîrbu. Asymptotic analysis of utility-based hedging strategies for small number of contingent claims. *Stochastic processes and their applications*, 117(11):1606–1620, 2007.

Hayne Leland. Option prices with transaction costs. *Journal of Finance*, 40(5): 1283–1301, December 1985.

Chuin Ching Liew and Tak Kuen Siu. A hidden Markov regime-switching model for option valuation. *Insurance: Mathematics and Economics*, 47(3):374–384, dec 2010. ISSN 01676687. doi: 10.1016/j.insmatheco.2010.08.003. URL `http://linkinghub.elsevier.com/retrieve/pii/S0167668710000946`.

Michael Mania and Martin Schweizer. Dynamic exponential utility indifference valuation. *The Annals of Applied Probability*, 15(3):2113 – 2143, 2005.

Harry M. Markowitz. Portfolio selection. *Journal of Finance*, 7:77–91, March 1952.

Robert C. Merton. Theory of rational option pricing. *Bell Journal of Economics and Management Science*, pages 141–183, 1973.

Yoshio Miyahara, Monique Jeanblanc, and Susanne Klöppel. Minimal f^q-martingale measures for exponential lévy processes. *Annals of Applied Probability*, 17:1615–1638, 2007.

Romuald Momeya and O Menoukeu Pamen. Local risk-minimization under a partially observed Markov-modulated exponential Lévy model. Technical Report 5, 2011. URL `https://www.duo.uio.no/handle/10852/10161`.

Sergey Nadtochiy and Jan Obłój. Robust Trading of Implied Skew. *International Journal of Theoretical and Applied Finance*, 20(2), 2017.

Anthony Neuberger. Volatility trading. Technical report, London Business School, 1990.

Anthony Neuberger. The log contract. *Journal of Portfolio Management*, 20(2): 74–80, 1994.

Anthony Neuberger. Realized Skewness. *Review of Financial Studies*, 25 (11):3423–3455, 2012. ISSN 0893-9454. doi: 10.1093/rfs/hhs101. URL `http://rfs.oxfordjournals.org/cgi/doi/10.1093/rfs/hhs101`.

Marcel Nutz. Power utility maximization in constrained exponential Lévy models. *Mathematical Finance*, 22(4):690–709, 2012. URL `http://onlinelibrary.wiley.com/doi/10.1111/j.1467-9965.2011.00480.x/full`.

Antoon Pelsser and Lukasz Delong. Instantaneous Mean-Variance Hedging and Sharpe Ratio Pricing in a Regime-Switching Financial Model. *Stochastic Models*, 31(1):67–97, 2015. URL `http://ideas.repec.org/p/arx/papers/1303.4082.html`.

Huyên Pham. Mean-variance hedging for partially observed drift processes. *International Journal of Theoretical and Applied Finance*, 4(2):263–284, 2001.

Phillip Protter. *Stochastic Integration and Differential Equations*. Springer-Verlag, 2 edition, 2004.

L C G Rogers and Liang Zhang. An asset return model capturing stylized facts. *Mathematics and Financial Economics*, 5(2):101–119, 2011.

L.C.G. Rogers and David Williams. *Diffusions, Markov Processes and Martingales, Volume 2: Ito Calculus*. Cambridge University Press, 2000.

Paul A. Samuelson. A rational theory of warrant pricing. *Industrial Management Review*, 6(2):13–32, 1965.

Ken-Iti Sato. *Lévy processes and infinitely divisible distributions*. Cambridge University Press, 1999.

Ulrike Schaede. Forwards and futures in tokugawa-period japan: A new perspective on the dojima rice market. *Journal of Banking and Finance*, 13:487–513, September 1989.

Thorsten Schmidt and Rüdiger Frey. Pricing and Hedging of Credit Derivatives via Nonlinear Filtering. *Finance and Stochastics*, 16(1):105–133, 2012.

Wim Schoutens. Moment swaps. *Quantitative Finance*, 5(6):525–530, dec 2005. ISSN 1469-7688. doi: 10.1080/14697680500401490. URL http://www.informaworld.com/openurl?genre=article{&}doi=10.1080/1469768050040149

Martin Schweizer. Variance-optimal hedging in discrete time. *Mathematics of Operations Research*, 20:1–32, 1995.

Steven E. Shreve. *Stochastic Calculus for Finance II: Continuous-Time Models.* Springer Verlag, 1st edition, 2004.

Tak-Kuen Siu. A hidden markov-modulated jump diffusion model for european option pricing. In Rogemar S. Mamon and Robert J. Elliott, editors, *Hidden Markov Models in Finance: Further Developments and Applications*, volume II. Springer Verlag, 2014.

Anatoliy Swishchuk, Maksym Tertychnyi, and Robert Elliott. Pricing Currency Derivatives with Markov-modulated Lévy Dynamics. *Insurance: Mathematics and Economics*, 57:67–76, 2014.

Peter Tankov and Ekaterina Voltchkova. Asymptotic analysis of hedging errors in models with jumps. *Stochastic processes and their applications*, 119:2004–2027, 2009.

Emmanuel Temam. *Couverture Approchée d'options Exotiques. Pricing des options asiatiques.* PhD thesis, Université Paris VI, 2001.

JP Temme. Power utility maximization in exponential Lévy models: convergence of discrete-time to continuous-time maximizers. *Mathematical Methods of Operations Research*, 76(1):21–41, 2012. URL http://link.springer.com/article/10.1007/s00186-012-0388-3.

Edward O. Thorp and Sheen T. Kassouf. *Beat The Market: A Scientific Stock Market System.* Random House New York, 1967.

Klaus B. Toft. On the mean-variance tradeoff in option replication with transaction costs. *The Journal of Financial and Quantitative Analysis*, 31(2):233–263, June 1996.

E.T. Whittaker and G.N. Watson. *A course of modern analysis.* Cambridge University Press, 4th edition, 1996.

Ruotao Zhang. *Couverture Approchée des Option Européennes.* PhD thesis, École Nationale des Ponts et Chaussées, 1999.