The Non-parametric Calibration of Jump-diffusion Lévy Models

Jaco van Zyl

29 December 2022

Abstract

This study investigates the effectiveness of relative-entropy-based regularised calibration procedures at addressing the ill-posedness encountered in the calibration of non-parametric jump-diffusion Lévy models. Calibrated models, which have been selected for their capability of simulating realistic price path evolutions, are typically employed to price path-dependent instruments or to perform dynamic hedging.

The financial risks associated with using pricing models in real-world transactions can be reduced by selecting an appropriate model together with a suitable calibration procedure. It has been well established that mispricing resulting from improper model-based pricing has resulted in significant financial losses.

Lévy processes aim at improving on the ability of continuous models to represent market-observed price evolutions by allowing highly flexible jump characteristics to be specified, enabling the inclusion of jump-discontinuities in pricing models. The capacity of parametric Lévy models to provide a general class of feature-rich models with jump-discontinuities is further enhanced by the extension to non-parametric Lévy models, where the jump characteristics can more freely be defined.

Model calibrations, like most other inverse problems, suffer from ill-posedness. Consequently, optimisation results from a calibration exercise generally do not converge to a unique solution and typically do not vary continuously with input data. The increase in dimension of the solution space associated with the non-parametric approach necessitates further measures to address the ill-posedness.

This study evaluates the performance of the relative-entropy-based regularised calibration procedures proposed by Cont and Tankov [CT04, CT06] at addressing the primary concern of ill-posedness encountered in the calibration of non-parametric jump-diffusion Lévy models. We will show that although the procedures provide some stability with respect to the input prices between subsequent calibrations, the procedures are of limited value at addressing the ill-posedness relating to the convergence to a unique solution. In our experiments, we expose the sensitivity of results to both the initial points as well as to the prior measure presented to the optimisation procedures.

Our study highlights some deficiencies in the process, showing that the regularised calibration procedure is unreliable, necessitating active user intervention to manage outcomes. We conclude that the observed difficulties are primarily the result of a persistent non-convexity of the regularised objective function at realistic levels of regularisation. Therefore ill-posedness continues to present a risk that needs to be managed by practitioners when applying these procedures to the recovery of non-parametric Lévy models.

Statement of own work

I, Jaco van Zyl, hereby declare that this dissertation is my own work.

I confirm that I appropriately acknowledge in the dissertation

- *the sources used, and*
- *assistance received*

in the execution of this work.

Signed by candidate

Jaco van Zyl

29 December 2022

Statement of own opinion

I, Jaco van Zyl, hereby declare that the opinions expressed in this dissertation are solely my own, and do not reflect the opinions or beliefs of the University of Cape Town.

Signed by candidate

Jaco van Zyl

29 December 2022

Acknowledgement

I wish to thank Professor Peter Ouwehand for agreeing to supervise this research project, and for his helpful mentorship and constructive guidance through this process.

Furthermore, I would like to thank wife, Christel, for her loving kindness, loyal support, and positive encouragement along this special journey.

Contents

List of Figures

List of Tables

Part I

Study outline

Chapter 1

Introduction

1.1 Problem statement

The worldwide over-the-counter (OTC) derivatives market was estimated at a notional amount of \$600 trillion[1] with a gross market value of \$12.4 trillion in a statistical survey[2] released by the *Bank of International Settlements* (BIS) in May 2022 [Ban22]. The OTC derivatives market has grown since the 2007 financial crisis and has overtaken the exchange-traded market, which in 2019 accounted for only 41% of the market share [Ban19]. Derivative instruments traded over the counter are usually illiquid, which means that market pricing information is typically insufficient for determining the prices of these instruments. Therefore OTC instruments largely depend on pricing models to determine their prices.

This study will evaluate a calibration technique which can be used in the calibration of such pricing models. Non-parametric Lévy models seek to improve the ability to capture the dynamics of the price evolution of an underlying asset by including jump-discontinuities which can be specified in a particularly general form through the Lévy measure.

In particular, this study is concerned with evaluating the performance of:

- the non-parametric calibration of
- jump-diffusion Lévy models, while
- specifically examining the capabilities of relative-entropy-based regularisation techniques employed to mitigate the ill-posed nature of these calibration problems.

We will concentrate specifically on examining and then evaluating the procedures proposed by Cont and Tankov in their literature on this topic [CT04, CT06, CT03, CT02].

[1] Trillion denotes 10^{12}.

[2] Based on data from the second half of 2021.

Model calibration is more than just an academic exercise to determine which models have superior abilities at capturing market dynamics for improved pricing performance. In practice, financial decision-makers rely on realistic pricing, including the pricing of illiquid or OTC financial instruments, to manage risk for large financial institutions such as banks. When pricing such instruments where no market prices are available, it is necessary to resort to a calibrated pricing model to replicate market dynamics. The calibration procedures considered in this study use options prices to acquire the implied dynamics of the underlying asset rather than directly observing historical time series data.

The calibrated pricing model is then used to extrapolate from the reference (calibration) set of liquid instruments to illiquid path-dependent instruments by simulating the pricing evolution of an underlying. In the absence of more accurate market information, practitioners use this extrapolation approach and are exposed to the risk of mispricing such illiquid path-dependent instruments, which could have significant real-world financial implications.

The reason for using *Lévy models,* which supplement traditional diffusion pricing processes with jump-discontinuities, is an effort to improve the reliability of pricing models. The addition of jump-discontinuities potentially allows one to capture a richer, more realistic set of market dynamics and thereby improve the ability of pricing models to reflect specific market dynamics. The objective is to reduce the risk of mispricing by the simulation of paths that more closely resemble market-observed dynamics in an effort to more effectively price path-dependent instruments.

The emphasis on *non-parametric* Lévy models is a further refinement aiming to reduce model-specific limitations on jump characteristics. Non-parametric models allow a more general "model-free" approach so that pricing dynamics are more explicitly determined by market data. The close relationship between the model characteristics and the market data without model-bound assumptions allows the model to more rigidly assume the form implied by the data. However, this close relationship between model and data also implies that inherent noise present in the market data will more readily pass through to this highly flexible model. This situation is analogous to fitting a high-order polynomial to noisy data. The over-fitting then leads to significantly different meaningless "solutions" primarily determined by the variations caused by the random noise.

Regularisation has been proposed as a strategy to address the ill-posed nature of this type of non-parametric model calibration. By applying regularisation, the aim is to reduce the aforementioned sensitivity to random market price variations in successive model calibrations and to assist with the convergence to a unique solution.

Next, we will provide an outline of the structure of this study, where our primary objective is to determine to what extent the relative-entropy-based regularisation techniques proposed by Cont and Tankov succeed in addressing the complexities associated with this ill-posed optimisation problem.

1.2 Outline of the dissertation

The chapters forming the subject matter of this dissertation are grouped into three parts. Part I, the *Study Outline*, consists of Chapter 1, the current chapter. In this chapter, we introduce the topic of our study, outline the structure of the material presented and then briefly discuss significant literature related to the study.

In Part II, the *Theoretical Framework*, we introduce a selection of fundamental theoretical concepts to allow the principles applied in the procedures of this study to be conveniently described.

The theoretical framework commences with Chapter 2, which introduces Lévy models with a discussion on the components and properties of Lévy processes with elementary results and terminology. We discuss the description of jump properties through the Lévy measure and the model presentation in the form of a characteristic function via the Lévy-Khinchin formula. Some parametric Lévy models are described before introducing the non-parametric form. We conclude this chapter by discussing pricing models and the simulation of Lévy price paths.

In Chapter 3, we discuss Fourier transform pricing techniques. This pricing method forms a core component of the pricing engine used in the calibration process.

Chapter 4 describes the model calibration process, where model parameters are recovered from options prices using optimisation algorithms. We introduce the problematic aspect of ill-posedness often encountered in this type of inversion process and then present regularisation as a potential remedy to this problem. We conclude the chapter by discussing calibration risk and model risk.

The next segment of chapters, titled *Examination of calibration procedures,* forms Part III of this study, where we look specifically at the procedures employed in the relative-entropy-based regularised optimisation process and perform an assessment through several experiments. Results are then analysed and discussed.

In Chapter 5, we break down the regularised calibration process into a sequence of step-wise numerical procedures and discuss the detail of each step. This constitutes a detailed explanation of the procedures proposed by Cont and Tankov in their literature on the topic of non-parametric calibration of Lévy models using relative-entropy-based regularisation [CT04, CT06].

Chapter 6 describes the experiments which we conducted to assess the performance of the procedures under examination in this study, along with the results obtained. In this chapter, we will illustrate through our experiments that the calibration process is not particularly effective. We will show that this procedure is difficult to apply as it requires that practitioners closely supervise the outcomes of calibration results, since it may often be necessary to make adjustments to compensate for the ineffectiveness of the overall process. We will expose some undesirable properties of the regularised calibration results, such as ill-posedness through non-uniqueness of solutions and observed instability with respect to small changes in market data.

In Chapter 7, we discuss the key findings from the experiments in the preceding chapter and express our views on the observations. We highlight the shortcomings of the

optimisation process illustrated by the results of our experiments and provide our interpretation of the underlying issues that contribute to these difficulties. Thereafter we discuss our findings in relation to published results from Cont and Tankov [CT04, CT06] as well as results published by other independent researchers. We will show that although our results correspond well with those of independent researchers, they do not correspond well with results published in the original articles [CT04, CT06]. We will closely examine selected results and explanations published in these articles by Cont and Tankov and expose their weakness.

Chapter 8 concludes this dissertation.

A summary of the mathematical notation used in this dissertation is available in Appendix A.

1.3 Significant literature

1.3.1 Principal articles

The primary focus of this study concerns the recovery of non-parametric Lévy models using the relative-entropy-based regularised calibration methods proposed by Rama Cont and Peter Tankov. Their procedures are described in a number of publications. The most significant[3] presentation of their methods can be found in the following articles, which we will refer to as the *principal articles*:

- *Nonparametric calibration of jump-diffusion option pricing models* [CT04] published in 2004, and
- *Retrieving Lévy processes from option prices: Regularization of an ill-posed inverse problem* [Con06b] published in 2006.

Supplementary information describing peripheral aspects of these procedures can be found in their working paper, *Calibration of jump-diffusion option pricing models: a robust non-parametric approach* [CT02] and in Chapter 13 of their acclaimed[4] book, *Financial Modelling with Jump Processes* [CT03].

Significant citations of the principal articles

Our literature review found that although the principal articles received a significant number of citations, there were hardly any citations contributing additional insights, evaluations, support or criticism of their work. The citations are mostly incidental[5] and they are frequently used to contextualise research in adjacent areas. Many of the

[3] The article *Nonparametric calibration of jump-diffusion option pricing models* [CT04] is the most cited article published by Peter Tankov to date according to *Google Scholar*, with a total of 279 citations recorded by August 2022 [Goo22b].

[4] Winner of the Riskbook.com *Best of 2004 Book Award* [Goo22a].

[5] Citations of these articles are typically used in student dissertations as a way to brush aside the severity of addressing problems associated with ill-posedness encountered in their own research.

articles mentioned in the following subsection cited the principal articles to show how their research relates to existing literature.

We did, however, find one article which presents clear evidence of efforts towards a practical implementation of the Cont and Tankov procedures. This important article by Huh [Huh19] describing his research in applying machine learning to the non-parametric recovery of Lévy properties also expresses some insightful criticism based on his evaluation of "existing" non-parametric procedures applied to Lévy jump-diffusions. The article will be discussed in more detail in the subsection below as an example of a different approach to the non-parametric recovery of Lévy jump-diffusions. Later, in Chapter 7, this article will be revisited when we discuss our results in relation to those of other independent researchers.

The article by Bondi et al. [BRR20], in which the authors claim that they successfully applied the relative-entropy-based calibration procedures developed by Cont and Tankov [CT04], unfortunately, does not contain any information of substance.

Guillaume and Schoutens briefly criticise the regularised entropy-based calibration procedures of the principal articles [CT04, CT06] in their article *A moment matching market implied calibration* [GS13] by pointing out that solutions depend on the starting parameters of the optimiser as well as the selected prior measure. They do not furnish any results or citations in support of their statement.

To the best of our knowledge, this dissertation is the first in-depth empirical analysis of the methods described in the above-mentioned principal articles by Cont and Tankov.

1.3.2 Parallel research

Although we found relatively little work directly evaluating the principal articles or directly expanding on them, there have been several parallel research paths with some elements in common.

For this discussion, we distinguish between the following broad categories of related calibration problems:

- Recovery of non-parametric Lévy properties from options prices (the core objective).
- Regularised calibration of *parametric* Lévy models from options prices.
- Recovery of Lévy properties by *direct observation* of the underlying process (time series).
- Other non-parametric regularisation techniques of interest (not applied to Lévy measures).

We selected the above-mentioned categorisation of articles according to the *primary objective* (task) of the calibration problems, stating "what" needs to be achieved or recovered in each problem class. Within each of these categories, we will mention articles describing different approaches to *how* the calibration objective could be achieved.

In the following subsections, we will discuss significant literature under a heading for each of the above-mentioned categories, with the first being the most important since it is directly related to the core objective of the principal articles.

1.3.2.1 Recovery of non-parametric Lévy properties from options prices

The key focus of this dissertation is on the retrieval of non-parametric Lévy properties by calibration to options prices. We are particularly interested in the extraction of the jump properties described by the Lévy measure and will now discuss literature describing other techniques to achieve this. The principal articles used relative entropy to assist with the non-parametric calibration, but here we consider literature describing other approaches.

In the article *Spectral calibration of exponential Lévy models* [BR06], Belomestny and Reiss describe a non-parametric calibration approach using "spectral regularisation". In the calibration process, they smooth their Lévy measure by disallowing large changes in jump activity for small changes in jump size. They achieve this by removing high-frequency components in the frequency-domain so that calibration, in essence, determines the optimal cut-off frequency for the best fit. This important article will be discussed later in Section 4.2.2. Minor variations in the implementation of this approach can be found in the article by Söhl and Trabs [ST12].

A more recent article by Huh [Huh19], which falls in the field of machine learning, is concerned with recovering non-parametric Lévy characteristics from options prices by using neural networks. In the article *Pricing options with exponential Lévy neural network* [Huh19], Huh suggests that the "existing"[6] non-parametric calibration procedures for exponential Lévy models fail to deliver useful results in practice and suggests a non-parametric approach based on neural networks. Unfortunately, the study compares the results of their method with those obtained from the *parametric* Kou and Merton models and not with truly non-parametric methods.

In their article, *Nonparametric implied Lévy densities* [QT19], Qin and Todorov recently developed a non-parametric Lévy density estimator applicable to short-maturity options where their procedure can conveniently separate volatility and jumps.

1.3.2.2 Regularised parametric Lévy calibration

Matsuda researched the effectiveness of calibrating with and without regularisation [Mat05], but his research did not explore non-parametric calibration and was limited to parametric Merton models.

Kim and Lee applied relative entropy to CGMY model calibration to choose the best "model preserving minimal entropy martingale measure" from the infinite number of available equivalent martingale measures [KL07].

Related work by Cont and Kokholm describes a procedure of dual calibration where calibration is performed on both VIX option prices as well as options prices on the

[6] Referring to the procedures of Cont and Tankov[CT04, CT06] and Belomestny and Reiss [BR06].

underlying index allowing for jumps in volatility and returns [CK13]. They chose to use *parametric* jump distributions.

Albani and Zubelli follow the approach of isolating the local volatility surface and parametric jump size distributions [Alb20]. They apply a "form of Tikhonov regularisation" to address the ill-posedness in their calibration procedure.

Although research in machine learning techniques is gaining popularity in solving financial modelling problems, there do not seem to be many articles addressing the calibration problems considered in this study. In a recent literature review of more than 150 articles on options pricing through neural networks [RW19], none of the articles was concerned with non-parametric Lévy models and only two articles related to parametric Lévy models. In their literature review [RW19], Ruf and Wang mentioned articles relating to parametric Lévy calibration by Yang et al. [YZH17] and Karatas et al. [KHO19].

Giebel and Rainer proposed a procedure for combining neural networks with the statistical estimation of stochastic process parameters recovered from time series data [GR13].

1.3.2.3 Recovery of Lévy properties from the underlying time series

The articles mentioned in this category are *not* concerned with the recovery of Lévy properties from options prices, but they rely instead on discrete observation of the Lévy process directly. When employed to recover a jump-diffusion, this generally involves disentangling jumps (required to recover the Lévy measure) from the diffusion component of the observed process.

Several articles have been published describing the recovery of Lévy properties by direct observation of the process. Examples of proposed techniques can be found in articles by Jongbloed et al. [JVDMVDV05], Van Es et al. [VEGS07], Neumann and Reiss [NR09], Comte and Genon-Catalot [CGC09], Gugushvili et al. [Gug09, GVdMS15, GvdMS18, GdMSS19, GMvdM20], Kappus and Reiss [KR10], Belomestny [Bel11], Duval and Mariucci [DM17, DM21] and Coca[Coc18].

In a different, but related setting, Todorov and Tauchen [TT11] propose that volatility jumps could be modelled through a pure jump Lévy process.

Belomestny proposes a method to estimate the fractional order of a Lévy process which he describes as a "semi-parametric approach" [Bel10]. It involves using both the options prices and the underlying data for the calibration and estimation.

Xu and Darve, in their article, *Calibrating multivariate Lévy processes with neural networks* [XD20], apply neural networks to the problem of non-parametric estimation of a two-dimensional Lévy density. They use time series data from the underlying process to perform their estimation.

1.3.2.4 Other non-parametric regularisation techniques of interest

In this section, we look specifically at influential non-parametric calibration techniques which are *not* concerned directly with the recovery of Lévy measures. We will commence with a discussion of the articles which preceded the primary articles, creating

a frame of reference for us to present the distinguishing properties of the methods examined in this study. Thereafter we will mention some significant research in subsequent articles.

Literature preceding the primary articles

In their discussion on "previous literature" in the primary articles [CT04, CT06], Cont and Tankov describe two significant methods: The first important article by Kallsen [Kal02] introduces the *minimal entropy consistent martingale measure* (MECMM), a pricing measure consistent with options prices while minimising the relative entropy with respect to a prior. He uses the MECMM in a theoretical sense with regard to exponential hedging and not as a pricing mechanism.

The non-parametric relative-entropy-based calibration methods proposed by Avellaneda [Ave98] and Avellaneda et al. [ABF+01] form the second group of significant articles described in the primary articles [CT04, CT06]. These procedures yield a pricing probability measure Q calibrated to market prices while maintaining a minimum entropy with respect to a prior pricing measure.

The following points describe features specific to Avellaneda's approach [Ave98, ABF+01] that are significantly different to the approach in the primary articles [CT04, CT06, Ave98]:

- Calibration is performed using Monte Carlo paths over a discretised *state space* rather than a discretised parameter space. Avellaneda's process yields a set of weights allowing pricing via discrete-time Monte Carlo methods, rather than a set of model parameters which can be used in continuous time processes.
- The pricing measures retrieved by Avellaneda are not guaranteed to be risk-neutral, whereas the recovered Lévy jump-diffusions in the primary articles are risk-neutral.
- Avellaneda's technique uses relative entropy as a selection criterion, rather than using it as a regularisation technique.

As mentioned by Cont and Tankov [CT04, CT06], there were a number of articles published prior to the principal articles where regularisation by relative entropy was applied to the non-parametric recovery of *probability measures* for asset pricing rather than the recovery of *Lévy measures*. The articles cited for their parallel research relationship in the non-parametric recovery of pricing measures are: Avellaneda [Ave98], Rouge and Karoui [REK00], Frittelli [Fri00], Kallsen [Kal02], Goll and Rüschendorf [GR01], Föllmer and Schied [FS02] and Fujiwara and Miyahara [FM03].

The article by Kallsen [Kal02] has the distinguishing feature that, unlike the procedures described in the other articles, the calibration is performed to options prices, much like the approach in the primary articles [CT04, Con06b].

Literature published since the primary articles

In their article, *Equity correlations implied by index options: estimation and model uncertainty analysis*, Cont and Deguest propose constructing a model based on the minimal entropy random mixture of a number of reference models [CD13]. They determine the distribution of the model mixture weights by applying relative-entropy-based optimisation, which they consider to be a well-posed problem. Here, the minimisation is over the space of model parameters and not market scenarios.

The article, *Nonparametric Bayesian volatility estimation for gamma-driven stochastic differential equations* [BGSS20], by Belomestny et al. describes a procedure using Bayesian estimation to obtain a non-parametric form of the volatility function. The volatility function forms a key component of Lévy-driven stochastic differential equations.

Todorov estimates tail jump variation measures non-parametrically from short-term options prices in his article, *Nonparametric jump variation measures from options* [Tod21].

Part II

Theoretical framework

Chapter 2

Lévy models

In mathematical finance, the drive to create models of price evolutions that more closely describe the observed characteristics of real-world markets, which at times exhibit sudden changes, has encouraged research into supplementing continuous models with jump-discontinuities. Lévy models provide a mathematically tractable means of extending continuous diffusion models to a class of models which allows discontinuities in the asset price process.

The procedures examined in this study focus specifically on the calibration of non-parametric *Lévy models*. This chapter will introduce the theoretical aspects of Lévy processes required to explain these procedures as follows. The fundamental elements and principal properties of Lévy processes will be introduced first. Then we will show how the Lévy measure can be employed to describe a particularly rich set of jump attributes. It will be shown how jumps described by the Lévy measure are suitable not only for extending a diffusion process to a jump-diffusion process, but also to represent the high jump-intensity in a pure-jump process.

We will show how Lévy models can be represented by a characteristic function through the Lévy-Khinchin formula, with the Lévy measure forming a core element of the characteristic function.

After introducing some important risk-neutral-pricing-theory applicable to Lévy models, we give examples of parametric Lévy models used for asset pricing and discuss the principles of non-parametric Lévy models.

Calibrated Lévy models are typically used to simulate price path evolutions which can be used for pricing path-dependent options or to perform dynamic hedging. We will illustrate the practical application of Lévy models in pricing path-dependent options by simulation techniques, discussing the generation of time series from the dynamics embedded in Lévy models.

Before proceeding with the theoretical aspects of Lévy models, we will examine some arguments concerning the inclusion of jump-discontinuities in financial models.

The actual observation of jump-discontinuities in the prices of financial instruments is somewhat subjective because it depends on the perspective of the observer as well as the magnitude of the time interval between successive price samples. Gennotte and

Leland argue that although some market crashes, such as the 1987 crash, could be considered evidence of a jump discontinuity by some, a closer inspection might reveal a different series of events [GL90]. They suggest that the 1987 crash was caused by several cascading hedging transactions combined with insufficient market liquidity [GL90].

When observed on a coarser time scale, the 1987 crash resembles a discontinuity in the price of the market index. Tankov and Voltchkova have a firm view in support of the observation of jump-discontinuities and state that "The last and probably the strongest argument for using discontinuous models is simply the presence of jumps in observed prices." [TV09]

Although the observation of discontinuities in market prices is arguable, the addition of jump-discontinuities in a pricing model can be used to improve certain models. In his book, *The volatility surface: a practitioner's guide* [Gat11], Gatheral states that diffusion models and diffusion-based stochastic volatility models cannot explain the steepness of the volatility skew at short expirations. He suggests that the inclusion of jumps might explain this steepness by allowing the possibility of larger moves over a short time period. He explains the observed near-term market behaviour by stating that "Large moves do sometimes occur and it makes economic sense to bid for out-of-the-money options—at the very least to cover existing risk." [Gat11]

In other words, the introduction of jump-discontinuities adds a modelling mechanism for creating large moves. This adds an additional flexible attribute which can be used to supplement continuous models to create more interesting models. We will explain in this chapter how Lévy models can be used to introduce jump-discontinuities which may be used to supplement a diffusion process or to form a pure jump process. Further, we will show how Lévy processes describe the jump attributes, allowing for a wide variety of jump characteristics to be specified, while remaining mathematically tractable.

In Section 2.1 of this chapter, we will start with a discussion of Lévy processes viewed as the sum of drift, a diffusion component and a jump process. This is followed by a definition for Lévy processes and a discussion of their distinguishing properties, including infinite divisibility.

In Section 2.2, we present an illustration of the specification of the jump component of Lévy processes with finite jump activity through the compound Poisson process. We show how this can be seen as a generalisation of the Poisson process, where uniform jump sizes are replaced with random jump sizes from a specified distribution.

In Section 2.3, we discuss how the Lévy measure is used to extend the properties of the jump-distribution encountered in the compound Poisson process. The extension allows the specification of jump components of a Lévy process to include a richer set of features, including infinite jump activity. Properties such as finite activity and finite variation are discussed in this section.

In Section 2.4, the characteristic function is defined, and examples are used to illustrate the progression of features from the Poisson processes to the compound Poisson process. We further extend this illustration of the progression towards the comprehensive Lévy-Khinchin formula.

Risk-neutral pricing and equivalent measures are dealt with in Section 2.5. In Section 2.6, we present a few examples of exponential Lévy pricing models and their

characteristic functions.

We conclude this chapter with a section discussing the simulation of Lévy processes. We will show how financial time series simulations can be generated by adding jump discontinuities to continuous components creating a jump-diffusion price path which can be used for pricing or hedging.

2.1 Lévy processes

Lévy processes provide a mechanism for the inclusion of jumps, diffusion and drift. Lévy processes, X_t, can typically be expressed as

$$X_t = \gamma t + \sigma W_t + Z_t, \tag{2.1}$$

with

- a deterministic drift component, γt,
- diffusion component, σW_t, and
- a random jump process Z_t, with a possibly infinite number of jumps per unit time described by a *Lévy measure*.

The first two terms form a Brownian motion with drift, a frequently used construct in continuous path models. One can therefore consider the class of Lévy processes to be a generalisation of these continuous path stochastic processes.

A key strength of Lévy processes lies in the flexibility of the jump process, Z_t, described through a Lévy measure, which allows for the description of a rich set of jump characteristics, including the possibility of infinitely many jumps. In essence, a Lévy measure describes the jump characteristics by specifying the expected number of jumps per unit time for each set of jump sizes. The Lévy measure is more formally defined and discussed in Section 2.3. Illustrations of the individual components forming Lévy paths can be seen in Figure 2.1 on page 19 and Figure 2.2 on page 20.

Definition 2.1.1. Lévy process (Cont and Tankov [CT03])

A càdlàg[1] *stochastic process* $(X_t)_{t\geq 0}$ *on* $(\Omega, \mathscr{F}, \mathbb{P})$ *with values in* $\mathbb{R}^d$ *such that* $X_0 = 0$ *is a* Lévy process *if it has the following properties:*

1. *Independent increments: The random variables*

   $$X_{t_o}, X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}}$$

   are independent for every increasing time sequence $t_0, \ldots, t_n$.

2. *Stationary increments: The probability distribution of* $X_{t+h} - X_t$ *does not depend on* t.

[1] A càdlàg process is right continuous with left limits.

3. *Stochastic continuity: For every* $\epsilon > 0$*, and any* $t \geq 0$ *it holds that*

$$\lim_{h \to 0} \mathbb{P}\left(|X_{t+h} - X_t| > \epsilon\right) = 0.$$

The first two properties are fundamental to both the Poisson process and Brownian motion.

The third property does not imply that the sample paths of the process are continuous, but rather that the discontinuities, or jumps, occur at discrete random times so that the probability of observing a jump at any particular instant is zero. Only jumps that occur at random times are considered, so no deterministic jumps are allowed.

Infinite divisibility is another distinguishing property of Lévy processes and will be discussed next.

2.1.1 Infinite divisibility

Definition 2.1.2. Infinite divisibility (Cont and Tankov [CT03])

A probability distribution F *on* $\mathbb{R}^d$ *is infinitely divisible if and only if for any integer,* $n \geq 2$*, there exists* n *i.i.d. random variables* $Y_1, \ldots Y_n$ *such that* $Y_1 + \ldots + Y_n$ *has distribution* F*.*

Infinite divisibility is a significant property of all Lévy processes and follows as a result of the above-mentioned definition of Lévy processes. In other words, if $(X_t)_{t \geq 0}$ is a Lévy process, then for every t, the distribution of X_t is infinitely divisible [CT03]. Furthermore, it can be shown that for an infinitely divisible distribution F, there exists a corresponding Lévy process (X_t) such that X_1 is given by F [CT03]. The significance of the infinite divisibility property of Lévy processes relates to its association with the Lévy-Khinchin formula, which represents the characteristic function of any infinitely divisible distribution [Sat99]. The Lévy-Khinchin formula, discussed in Section 2.4.2, provides an effective tool for expressing the properties of a Lévy process as a characteristic function for use in transform pricing, as discussed in Section 3.2.

2.2 Compound Poisson process

We will start this section by reviewing the Poisson distribution and then discuss the Poisson process before progressing to the compound Poisson process.

The Poisson distribution is defined as follows.

Definition 2.2.1. Poisson random variable (Ross [Ros10])

An integer-valued random variable N *is said to follow a Poisson distribution with parameter* $\lambda > 0$ *if*

$$\forall n \in \mathbb{N}_0, \quad \mathbb{P}\left(N = n\right) = e^{-\lambda} \frac{\lambda^n}{n!}, \tag{2.2}$$

where (2.2) defines the probability mass function, with $e^{-\lambda} \sum_0^\infty \frac{\lambda^n}{n!} = 1$*.*

One can define the Poisson process as a Lévy process in terms of the *Poisson distribution* as follows.

Definition 2.2.2. Poisson process as Lévy process (Sato [Sat99])

A stochastic process $(N_t)_{t\geq 0}$ *on* $\mathbb{R}^+$ *is a Poisson process with parameter* $\lambda > 0$ *if it is a Lévy process, and for* $t > 0$, N_t *has the Poisson distribution with parameter* λt.

From the preceding definitions, the distribution of the Poisson process $(N_t)_{t>0}$ for $t > 0$ can be written as

$$\forall n \in \mathbb{N}_0, \quad \mathbb{P}\left(N_t = n\right) = e^{-\lambda t}\frac{(\lambda t)^n}{n!}.$$

The Poisson process can also be viewed from the perspective of the *exponential distribution*. The following theorem expresses the Poisson process as an exponential counting process.

Theorem 2.2.3. The Poisson process as exponential counting process (Cont and Tankov [CT03])

Let $(\tau_j)_{j\geq 1}$ *be a sequence of independent exponential random variables with parameter* λ *and* $T_n = \sum_{j=1}^{n} \tau_j$. *The process* $(N_t)_{t\geq 0}$ *defined by*

$$N_t = \sum_{n\geq 1} 1_{t\geq T_n}$$

is a Poisson process with intensity λ.

The Poisson process is a pure jump process which can also be considered to be a counting process where jumps of unit size occur with i.i.d. inter-arrival times that follow the exponential distribution with parameter λ. The parameter λ specifies the average number of jumps per unit time which we call the jump intensity of the process.

The inter-arrival times of the events or jumps modelled by the Poisson process have the property of memorylessness associated with the exponential distribution[2], a feature for modelling random arrivals where the waiting time for the next event does not depend on the amount of time elapsed since the last event.

There is a close relationship between Lévy Processes, the Poisson process and Brownian motion. Both the Poisson process and Brownian motion share the key properties of independence of increments and stationary increments with Lévy processes. Not only are the Poisson process and the Wiener process fundamental examples of Lévy processes, they can also be considered as the basic building blocks of Lévy processes. Lévy processes can be considered to consist of a superposition of a (possibly infinite) number of independent Poisson processes and a Wiener process [CT03].

A compound Poisson process is a pure jump stochastic process which can be informally described as a Poisson process with random jump sizes. It forms a pure jump Lévy process and is an effective means of illustrating the principles applied in defining the jump component of a Lévy process. The compound Poisson process is an adequate

[2]The exponential distribution is the only continuous memoryless distribution [Ros14].

means of fully describing the jump component for the subset of Lévy processes with finite jump activity. Although the jumps in Lévy processes with infinite jump activity cannot be adequately described by the compound Poisson process, their description relies on an extension of the same principles discussed below. The extension requires the use of the Lévy measure, which will be discussed later in Section 2.3.

2.2.1 Definition of a compound Poisson process

The compound Poisson process is defined as follows.

Definition 2.2.4. Compound Poisson process (Ross [Ros10])

A stochastic process $(X_t)_{t\geq 0}$ *is said to be a compound Poisson process if it can be represented as*

$$X_t = \sum_{j=1}^{N_t} Y_j,$$

where N_t *is a Poisson process with intensity* $\lambda > 0$ *and* $(Y_j)_{j\geq 1}$ *is a family of independent and identically distributed random variables with law* f*. Furthermore,* Y_j *must be independent of* N_t*.*

When a compound Poisson process is used to model a jump process, the arrival times are modelled by the Poisson component, N_t, providing the memoryless exponential timing properties for the jumps. The sizes of the individual jumps are modelled by Y_j. The distribution of jump sizes, f, can be any arbitrary distribution with zero mass at the origin. An explicit density function for f is not required, as will be discussed in Section 2.4 on characteristic functions of Lévy processes. Essentially there are two independent sources of randomness in place: The first determines the memoryless *arrival* of jump "events" and the second determines the *size* of each jump.

Equipped with this means of describing the jumps, we can now combine the compound Poisson process with a diffusion process to form a jump-diffusion Lévy process written as

$$X_t = \gamma t + \sigma W_t + \sum_{j=1}^{N_t} Y_j. \tag{2.3}$$

This jump-diffusion Lévy process based on the compound Poisson process is sufficiently flexible for defining the sub-class of Lévy processes where the expected number of jumps per time unit is finite.

Figure 2.1 on page 19 shows the path components forming a Lévy jump-diffusion, where the jump distribution has been chosen as an illustration to generate a fairly narrow range of positive and negative jump sizes without small jumps. This jump distribution was defined by a bimodal normal distribution as shown in Figure D.1 on page 168. This can be compared with the path components shown in Figure 2.2 on page 20, which are based on the Merton jump-diffusion mode, where we expect to observe a larger number

of small jumps generated by a normal distribution. The Merton model is described in more detail in Section 2.6.1.1 on page 32.

This difference is further illustrated by multiple price-paths based on the exponential of these two processes presented in Section 2.7, where we discuss the simulation of Lévy processes. Figure 2.7 on page 38 presents multiple exponential paths based on the jump-diffusion with a bimodal jump distribution, and Figure 2.8 on page 38 shows exponential paths for the Merton jump-diffusion process.

In Section 2.4 on characteristic functions, one can see in Table 2.1 on page 36 the characteristic functions illustrating the progression of the key elements introduced above: from the Poisson distribution up to the compound Poisson jump-diffusion. A well-known example of a compound Poisson jump-diffusion model is the Merton model. In the Merton model, a stock price has the form $S_t = S_0 e^{rt+X_t}$, with r representing the interest rate, X_t, the jump-diffusion in equation (2.3) with the jumps, Y_j, following the normal distribution [TV09]. The Merton model will be discussed in more detail in Section 2.6.1.1 on page 32.

In the next section, we discuss the *Lévy measure*, an extension which allows the specification of the number of jumps per unit time to become infinite for a subset of small jump sizes.

2.3 The Lévy measure

In the following definition of the Lévy measure, the symbol $\mathcal{B}$ is used to represent the Borel-algebra.

Definition 2.3.1. Lévy Measure (Cont and Tankov [CT03])

Let $(X_t)_{t>0}$ *be a Lévy process on* $\mathbb{R}^d$. *The measure* ν *on* $\mathbb{R}^d$ defined *by:*

$$\nu(A) = E\left[\#\left\{t \in [0,1] : \Delta X_t \neq 0, \Delta X_t \in A\right\}\right], \qquad A \in \mathcal{B}\left(\mathbb{R}^d\right)$$

is called the Lévy measure of $X : \nu(A)$ *is the expected number, per unit time, of jumps with size belonging to* A.

A Lévy measure, ν, is a positive measure defined on $\mathbb{R}^d \setminus \{0\}$ satisfying the integrability constraint [CT03]

$$\int_{\mathbb{R}^d} \left(1 \wedge |x|^2\right) \nu(dx) < \infty. \tag{2.4}$$

The Lévy measure $\nu(dx)$ is the means for specifying the jump properties of a Lévy process, where the jumps with sizes in set A arrive according to a Poisson process with intensity $\int_A \nu(dx)$. This means that $\nu(A)$ is the expected number of jumps per time unit with *size* that belongs to set A. The value of $\nu(A)$ is positive, and possibly infinite, in particular $\nu(A) \in [0, \infty]$.

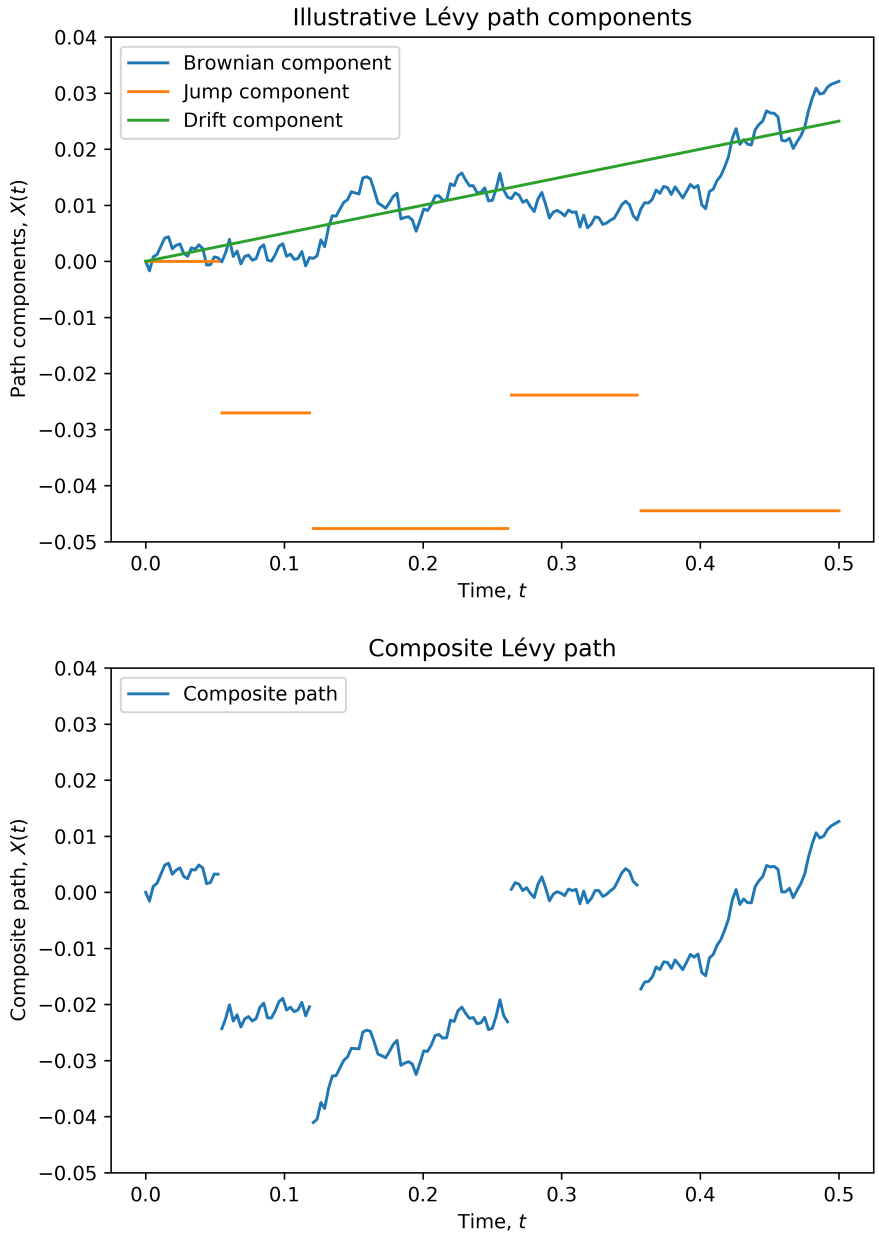

Figure 2.1: Example of a Lévy jump-diffusion path illustrating the individual components that form the path.

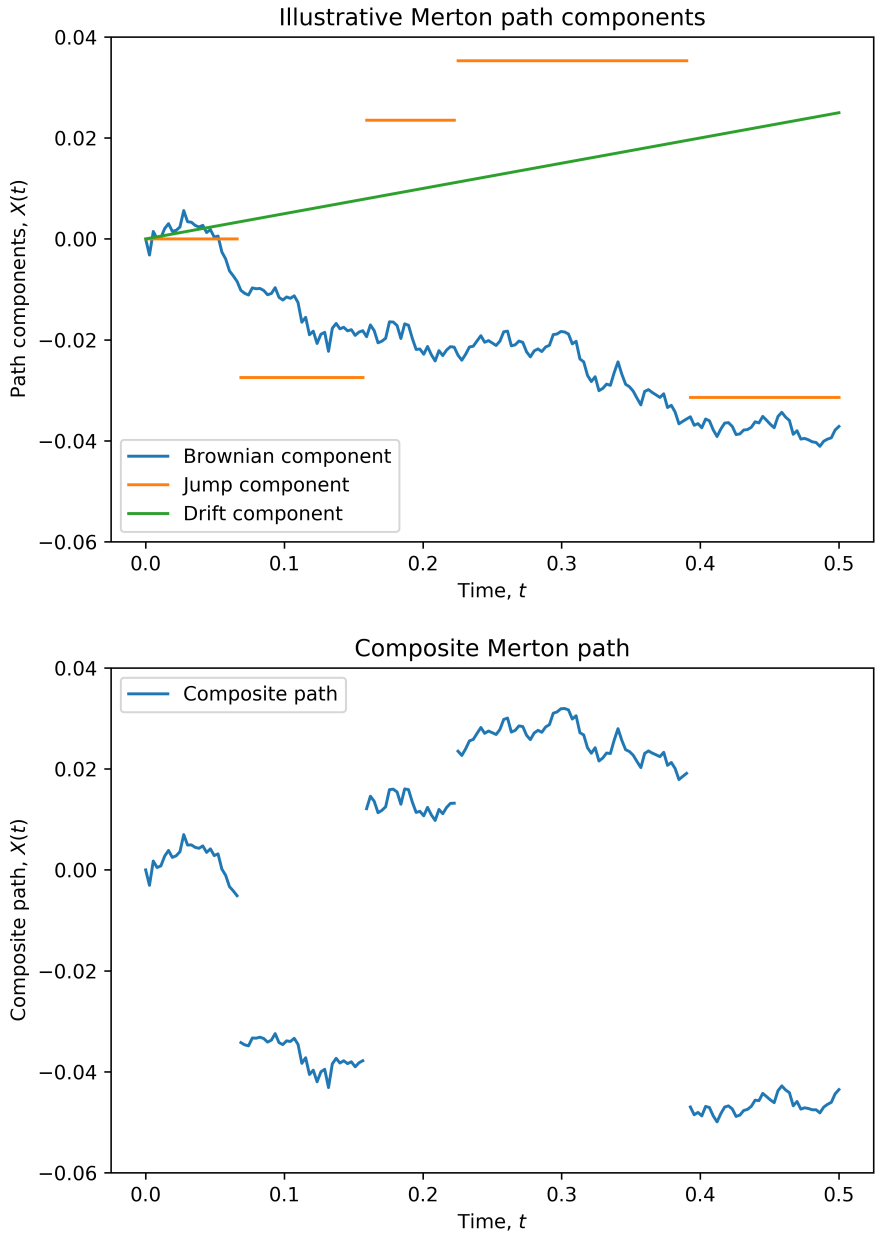

Figure 2.2: Example of a Merton jump-diffusion path illustrating the individual components that form the path.

From this definition, we note that using a Lévy measure to specify the jump properties rather than using a distribution function has the advantage of opening up a broader set of allowable jump profiles, including models with infinite activity.

When we use the Lévy measure in a one-dimensional process to describe the jumps of *finite activity*, with $\nu(\mathbb{R}) < \infty$, jumps follow a compound Poisson process, which can be written in the form

$$\nu(dx) = \lambda\mu(dx), \tag{2.5}$$

with jump intensity λ and the jump distribution described by the probability measure μ [CT03]. This implies that for finite activity models, the jumps described by its Lévy measure, $\nu(dx)$, can be separated into a jump intensity and a *distribution* of jump sizes, as can be seen on the right-hand side of (2.5). The Lévy density, $\nu(x)$, represents a combination of these two components, and is defined such that $\nu(x)dx = \nu(dx)$. From a different perspective, the Lévy density can be written in terms of a probability density, $f(x)$, as $\nu(x) = \lambda f(x)$. The Lévy densities of the finite activity Merton jump-diffusion and the Kou model are illustrated in Figure 2.3 on page 32 and Figure 2.4 on page 33, respectively.

The *path variation* of a Lévy process is determined by the presence of a Brownian motion component and the convergence of the sum of small jumps [Sch03].

Proposition 2.3.2. Finite variation of a Lévy process (Cont and Tankov [CT03])

A Lévy process is of finite variation *if and only if it has no Brownian motion component and a Lévy measure which satisfies the condition*

$$\int_{|x|\leq 1} |x|\,\nu(dx) < \infty.$$

The property of finite variation is important for integrability [Sch03].

As mentioned above, *finite activity* jump processes, with $\nu(\mathbb{R}) < \infty$, have a finite number of jumps in any finite interval and allow jumps to be represented by the compound Poisson process. On the other hand, infinite activity jump processes have jumps that arrive infinitely often with $\nu(\mathbb{R}) = \infty$ [CT03].

The concepts of *infinite activity* and *infinite variation* determine key characteristics in the formulation of Lévy measures describing *pure jump* models, where a diffusion component can be replaced by a high concentration of small jumps. A more detailed discussion about pure jump Lévy processes will follow in the next section after briefly introducing the characteristic function.

2.4 The characteristic function of a Lévy process

2.4.1 Characteristic functions

Definition 2.4.1. Fourier transform and inverse Fourier transform [BS94]

In mathematical statistics, the Fourier transform of a function f is defined as

$$\mathcal{F}f(u) = \int_{-\infty}^{\infty} e^{iux} f(x)\, dx.$$

The inverse transform is defined as

$$\mathcal{F}^{-1}f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-iux} f(u)\, du.$$

The integration is typically performed with respect to a real variable, but it is also possible to integrate along contours in the complex plane [Lew01].

The above definition, which will be applied in this dissertation, represents one of several formats for the Fourier and inverse-Fourier transforms. A more general definition that can easily be adapted to a chosen format is presented in Appendix B.1.1.

Definition 2.4.2. Characteristic function [CT03]

The characteristic function of a random variable X with distribution function F_X is defined as

$$\phi_X(u) = E\left[e^{iuX}\right] = \int_{\mathbb{R}} e^{iux} dF_X(x).$$

If the probability density function of X exists, we obtain the characteristic function by taking the Fourier transform of the probability density function f_X, giving

$$\phi_X(u) = E\left[e^{iuX}\right] = \int_{\mathbb{R}} e^{iux} f_X(x)\, dx. \tag{2.6}$$

The characteristic function uniquely defines a distribution, similar to the manner in which a probability density function describes a distribution [Sch03]. There is a one-to-one relationship between a distribution and its characteristic function. When a Lévy process represents the price model of an underlying asset, the characteristic function allows the use of efficient numerical Fourier transform pricing algorithms, discussed in Section 3.2.

When the characteristic function is of the form $\phi_X(u) = e^{t\psi(u)}$, we call $\psi(u)$ the *characteristic exponent* or the cumulant of the characteristic function.

Proposition 2.4.3. The characteristic function of the compound Poisson process (Cont and Tankov [CT03])

The characteristic function of a compound Poisson process $(X_t)_{t\geq 0}$ on $\mathbb{R}^d$ is represented by

$$E\left[e^{iu.X_T}\right] = \exp\left\{t\lambda \int_{\mathbb{R}^d} \left(e^{iu.x} - 1\right) f(x)\, dx\right\}, \quad \forall u \in \mathbb{R}^d,$$

where λ is the jump intensity and f represents the jump size distribution.

Proof. Conditioning $E\left[e^{iu.X_T}\right]$ on N_t , with $(Y_j)_{j\geq 0}$ representing a family of independent random variables from distribution f, and with ϕ_{Y_1} representing the characteristic function of f, we proceed as follows:

$$\begin{aligned}
E\left[e^{iu.X_t}\right] &= E\left[E\left[e^{iu.X_T} \mid N_t\right]\right] \\
&= \sum_{n\geq 0} \mathbb{P}\left(N_t = n\right) E\left[\exp\left(iu.\sum_{j=1}^{N_t} Y_j\right) \mid N_t = n\right] \\
&= \sum_{n\geq 0} \mathbb{P}\left(N_t = n\right) E\left[\exp\left(iu.\sum_{j=1}^{n} Y_j\right)\right] \\
&= \sum_{n=0}^{\infty} \frac{e^{-\lambda t}\left(\lambda t\right)^n \left(\phi_{Y_1}\left(u\right)\right)^n}{n!} \\
&= \exp\left\{\lambda t\left(\phi_{Y_1}\left(u\right) - 1\right)\right\} \\
&= \exp\left\{t\lambda \int_{\mathbb{R}^d} \left(e^{iu.x} - 1\right) f\left(x\right) dx\right\}
\end{aligned}$$

□

The distinct nature of the Lévy process consisting of the sum of individual independent components, as illustrated in equations (2.1) and (2.3), can be discerned when comparing the characteristic functions of the individual components with the terms in the characteristic exponent of the Lévy process. In Table 2.1 on page 36, one can identify the geometric Brownian motion and compound Poisson elements, which together form the Lévy jump-diffusion in item 3.

The progression from

- *Poisson distribution* with rate parameter λ to the
- *Poisson process* N_t, which follows the Poisson distribution with rate λt, and which is again extended to a
- *compound Poisson process* with jump distribution f,

can be seen in items 2a-c in the above-mentioned table.

Considering the one-dimensional form, the characteristic function of the pure-jump compound Poisson process,

$$E\left[e^{iuX_t}\right] = \exp\left\{t\lambda \int_{-\infty}^{\infty} \left(e^{iux} - 1\right) f\left(x\right) dx\right\},$$

which is of finite jump intensity, can be written in terms of the Lévy measure with $\nu\left(dx\right) = \lambda f\left(x\right)\ dx$ as

$$E\left[e^{iuX_t}\right] = \exp\left\{t \int_{-\infty}^{\infty} \left(e^{iux} - 1\right) \nu\left(dx\right)\right\}.$$

The corresponding characteristic function for a jump-diffusion based on the Poisson process can be written in terms of the Lévy measure as

$$E\left[e^{iuX_t}\right] = \exp\left\{t\left(iu\gamma - \frac{u^2\sigma^2}{2} + \int_{-\infty}^{\infty}\left(e^{iux} - 1\right)\nu\left(\mathrm{d}x\right)\right)\right\}, \tag{2.7}$$

where X_t follows the expression for the jump-diffusion in (2.3)

$$X_t = \gamma t + \sigma W_t + \sum_{j=1}^{N_t} Y_j.$$

The last term in the expression for X_t represents the compound Poisson jumps described in Definition 2.2.4.

This characteristic function of a jump-diffusion process is sufficient for describing the set of Lévy processes where the jump intensity is finite. The Lévy-Khinchin formula, which will be discussed next, generalises the characteristic function to include Lévy processes with infinitely many jumps in every time-interval.

2.4.2 Lévy-Khinchin formula

The one-dimensional representation of the Lévy-Khinchin[3] formula as a characteristic exponent is

$$\psi(u) = -\frac{u^2\sigma^2}{2} + iu\gamma + \int_{-\infty}^{\infty}\left(e^{iux} - 1 - iux1_{|x|\leq 1}\right)\nu\left(\mathrm{d}x\right), \tag{2.8}$$

where the components characterising the Lévy process are specified in a format known as the Lévy triplet, written as $\left(\sigma^2, \nu, \gamma\right)$ [CT03]. The Lévy-Khinchin formula is a general formulation of the characteristic function for Lévy processes [CT03]. Moreover, the Lévy-Khinchin formula represents the characteristic functions of all infinitely divisible distributions [Sat99].

The three components of a Lévy process can be distinguished in the Lévy triplet:

- Deterministic component, γ,
- Diffusion component, σ^2 and
- Jump component specified by the Lévy measure, ν.

Comparing the Lévy-Khinchin formula in equation (2.8) with the characteristic function of the compound Poisson jump-diffusion in equation (2.7), we note the addition of the *compensator term*, $iux1_{|x|\leq 1}$, in the Lévy-Khinchin formula re-written as

[3] French spelling: Lévy-Khintchine.

$$\psi(u) = -\frac{u^2\sigma^2}{2} + iu\gamma + \int_{|x|>1} \left(e^{iux} - 1\right) \nu\left(\mathrm{d}x\right) + \int_{|x|\leq 1} \left(e^{iux} - 1 - \boldsymbol{iux}\right) \nu\left(\mathrm{d}x\right)$$

for emphasis on the differences. With the generalisation from finite compound Poisson Lévy measures to a Lévy measure that could have an infinite number of small jumps, the deterministic compensator term, shown in bold, is subtracted from

$$\int_{|x|\leq 1} \left(e^{iux} - 1\right) \nu\left(dx\right),$$

which makes the expression square integrable. This technique illustrates the principles of the Lévy-Itô decomposition, which is discussed in detail in Chapter 4 of Sato [Sat99].

This compensation of small jumps allows for the creation of pure-jump Lévy processes using the Lévy-Khinchin formula and a triplet of the form $\left(\sigma^2 = 0, \gamma, \nu\right)$ where the infinite activity in small jumps is sufficient to eliminate the need for a diffusion component [CT03].

Simulating the infinite activity associated with a pure jump Lévy process can be problematic, but one can take advantage of the ambiguity between infinite activity in small jumps and a diffusion process to implement a numerical approximation of an infinite activity pure jump Lévy process. In the article, *Deterministic methods for option pricing in exponential Lévy models* [VT14], Voltchkova and Tankov suggest that one can substitute a Wiener process for jumps in a pure jump process that are smaller than some $\epsilon > 0$, by matching the variance in

$$\sigma^2\left(\epsilon\right) = \int_{-\epsilon}^{\epsilon} x^2 \nu\left(dx\right).$$

2.5 Risk-neutral pricing and equivalent measures

In the preceding subsection, we introduced the Lévy-Khinchin formula to represent Lévy processes, a mechanism which we intend to use for creating more realistic models for the price evolution of financial assets. In cases where such a Lévy process is used to represent an asset price process, it is essential that the characteristics of the process be appropriate for rational arbitrage-free pricing.

In mathematical finance, the risk-neutral equivalent martingale measure is used to maintain fair, arbitrage-free, pricing by ensuring that the expected return on a stock under this measure is equal to the risk-free rate of return. Furthermore, the existence of an equivalent martingale measure implies arbitrage-free pricing under that measure. This is expressed more formally in the following propositions. The notation follows the usual conventions used for describing martingale measures: a scenario space $(\Omega, \mathcal{F})$ with the flow of information modelled by the filtration $\left(\mathcal{F}_t\right)_{t\in[0,T]}$. This can be written as $(\Omega, \mathcal{F}, (\mathcal{F}_t))$. The maturity of the instrument being priced is T and n represents the number of assets. Furthermore, we represent the risk-free interest rate by r, the

terminal pay-off by H and $\hat{S}_t^i = e^{-rt} S_t^i$ represents the discounted price of S_t^i, which is the price of asset i at time t.

Definition 2.5.1. Equivalent probability measures (Cont and Tankov [CT03])

A probability measure $\mathbb{Q}$ on a scenario space $(\Omega, \mathcal{F})$ is equivalent to probability measure $\mathbb{P}$ if they have the same null sets. This can be expressed as

$$\mathbb{P} \sim \mathbb{Q}: \qquad \forall A \in \mathcal{F} \quad \mathbb{Q}(A) = 0 \Longleftrightarrow \mathbb{P}(A) = 0.$$

Definition 2.5.2. Equivalent martingale measures (Cont and Tankov [CT03])

A probability measure $\mathbb{Q}$ on a scenario space $(\Omega, \mathcal{F})$ is an equivalent martingale measure to $\mathbb{P}$ if they are equivalent and additionally satisfy the martingale requirement expressed as

$$\forall i = 1 \ldots n, \qquad E_{\mathbb{Q}}\left[\hat{S}_T^i \mid \mathcal{F}_t\right] = \hat{S}_T^i \qquad (\forall t \leq T).$$

Proposition 2.5.3. Risk-neutral pricing (Cont and Tankov [CT03])

In a market where the scenarios are described by a probability $\mathbb{P}$*, any arbitrage-free linear pricing rule* Π *can be represented as*

$$\Pi_t(H) = e^{-r(T-t)} E_{\mathbb{Q}}[H \mid \mathcal{F}_t],$$

where $\mathbb{Q}$ *is an* equivalent martingale measure*: a probability measure in the market scenarios such that*

$$\begin{aligned} &\mathbb{P} \sim \mathbb{Q}: \qquad \forall A \in \mathcal{F} \quad \mathbb{Q}(A) = 0 \Longleftrightarrow \mathbb{P}(A) = 0 \\ &\text{and} \qquad \forall i = 1 \ldots n, \qquad E_{\mathbb{Q}}\left[\hat{S}_T^i \mid \mathcal{F}_t\right] = \hat{S}_T^i \qquad (\forall t \leq T). \end{aligned}$$

We state the following simplified[4] form of the *fundamental theorem of asset pricing* as presented by Cont and Tankov in their book *Financial Modelling with Jump Processes* [CT03], which is adequate for illustrating the principles relating fair asset pricing and equivalent martingale measures.

Proposition 2.5.4. Fundamental theorem of asset pricing (Cont and Tankov [CT03])

The market model, represented by the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ *with asset prices* $(S_t)_{t \in [0,T]}$ *is arbitrage-free if and only if there exists a probability measure* $\mathbb{Q} \sim \mathbb{P}$ *such that the discounted asset prices* $\left(\hat{S}_t\right)_{t \in [0,T]}$ *are martingales with respect to* $\mathbb{Q}$*.*

The next proposition is included as the foundation for a significant result for asset-pricing with exponential Lévy models, which will be presented later in Proposition 2.5.6.

[4] A more rigorous treatment of this theorem is presented in the article by Delbaen and Schachermayer titled: *A general version of the fundamental theorem of asset pricing* [DS94].

Proposition 2.5.5. Martingale conditions for processes with independent increments (Cont and Tankov [CT03])

Let $(X_t)_{t\geq 0}$ be a real-valued process with independent increments. Then

1. $\left(\frac{e^{iuX_t}}{E\left[e^{iuX_t}\right]}\right)_{t\geq 0}$ *is a martingale* $\forall u \in \mathbb{R}$.
2. *If for some* $u \in \mathbb{R}$, $E\left[e^{uX_t}\right] < \infty \; \forall t \geq 0$ *then* $\left(\frac{e^{uX_t}}{E\left[e^{uX_t}\right]}\right)_{t\geq 0}$ *is a martingale.*
3. *If* $E\left[X_t\right] < \infty \; \forall t \geq 0$ *then* $M_t = X_t - E\left[X_t\right]$ *is a martingale (and also a process with independent increments).*
4. *If* $Var\left[X_t\right] < \infty \; \forall t \geq 0$ *then* $\left(M_t\right)^2 - E\left[\left(M_t\right)^2\right]$ *is a martingale, where* M *is the martingale defined above.*

If (X_t) is a Lévy process, then for all the processes mentioned in this proposition to be martingales, it suffices that the corresponding moments be finite for one value of t.

Proposition 2.5.6. Martingale conditions for Lévy processes and exponential Lévy processes (Cont and Tankov [CT03])

Let $(X_t)_{t\geq 0}$ be a Lévy process with characteristic triplet (A, ν, γ).

1. (X_t) *is a martingale if and only if* $\int_{|x|\geq 1} |x| \, \nu\left(dx\right) < \infty$ *and*
$$\gamma + \int_{|x|\geq 1} x\nu\left(dx\right) = 0.$$
2. $\exp\left(X_t\right)$ *is a martingale if and only if* $\int_{|x|\geq 1} e^x \nu\left(dx\right) < \infty$ *and*
$$\frac{A}{2} + \gamma + \int_{-\infty}^{\infty} \left(e^x - 1 - x1_{|x|\leq 1}\right) \nu\left(dx\right) = 0. \tag{2.9}$$

In Chapter 4, we will discuss techniques for the calibration of Lévy pricing models, where the problem can be expressed as the search for a risk-neutral Lévy process equivalent to a specific Lévy process, which represents our prior knowledge of the pricing process. When such a change of measure is required, and we wish to find Lévy processes with equivalent probability measures, we need to consider some requirements relating to the parameters that define the processes through their Lévy triplets.

The following proposition sets out the relationships between these characteristics associated with equivalent Lévy processes. These relationships allow one to identify the legitimate changes that can be made to the Lévy process characteristics while preserving equivalence of probability measures.

Proposition 2.5.7. Conditions for the equivalence of probability measures between Lévy processes (Sato [Sat99])

Let $(X_t, \mathbb{P})$ and $(X_t, \mathbb{P}')$ be two Lévy processes on $(\Omega, \mathcal{F})$ with characteristic triplets (σ, ν, γ) and (σ', ν', γ'). Then the restrictions $\mathbb{P}\mid_{\mathcal{F}_t}$ and $\mathbb{P}'\mid_{\mathcal{F}_t}$ are equivalent for all t (or equivalently for one $t > 0$) if and only if the following three conditions are satisfied:

1. $\sigma = \sigma'$.
2. *The Lévy measures are equivalent with*

$$\int_{-\infty}^{\infty} \left(e^{\phi(x)/2} - 1\right)^2 \nu\left(dx\right) < \infty$$

where $\phi\left(x\right)$ is the logarithm of the Radon-Nikodym density of ν' with respect to ν : $e^{\phi(x)} = \frac{d\nu'}{d\nu}$.

3. *If $\sigma = 0$, then in addition, γ' must satisfy*

$$\gamma' - \gamma = \int_{-1}^{1} x\left(\nu' - \nu\right)\left(dx\right).$$

The Radon-Nikodym derivative is given by

$$\frac{d\mathbb{P}'\mid_{\mathcal{F}_t}}{d\mathbb{P}\mid_{\mathcal{F}_t}} = e^{U_t}$$

where $\left(U_t\right)_{t\geq 0}$ is a Lévy process with characteristic triplet $\left(\sigma_U, \nu_U, \gamma_U\right)$ with

$$\sigma_U = \sigma\eta$$

$$\nu_U = \nu\phi^{-1}$$

$$\gamma_U = -\frac{1}{2}\sigma^2\eta^2 - \int_{-\infty}^{\infty}\left(e^y - 1 - y1_{|y|\leq 1}\right)\left(\nu\phi^{-1}\right)\left(dy\right)$$

and η is chosen so that

$$\gamma' - \gamma - \int_{-1}^{1} x\left(\nu' - \nu\right)\left(dx\right) = \sigma^2\eta$$

and U_t satisfies $E_{\mathbb{P}}\left[e^{U_t}\right] = 1$ for all t.

Looking at the first two conditions, we note that for Lévy models with a non-zero diffusion component, there is extensive freedom in changing the jump properties while still maintaining equivalence of the probability measures. Furthermore, the restrictions on the choice of γ', stated in the third condition above, cease to apply for non-zero σ. This allows the freedom to adjust γ' to satisfy the martingale requirements resulting from changes in the Lévy measure.

We mentioned in Section 2.3 that finite activity models, where $\lambda = \int \nu\left(dx\right) < \infty$, allow the jump size distribution, μ, to be obtained from the Lévy measure, ν, by $\mu\left(dx\right) = \frac{\nu(dx)}{\lambda}$. The jumps then follow the compound Poisson process with intensity λ. Compound Poisson processes do not require a truncation term, which conveniently simplifies the characteristic function so that the Lévy-Khinchin representation of the characteristic exponent from (2.8), which is repeated here

$$\psi(u) = iu\gamma - \frac{u^2\sigma^2}{2} + \int_{-\infty}^{\infty}\left(e^{iux} - 1 - iux1_{|x|\leq 1}\right)\nu\left(\mathrm{d}x\right),$$

becomes

$$\psi(u) = iub - \frac{u^2\sigma^2}{2} + \int_{-\infty}^{\infty} \left(e^{iux} - 1\right) \nu\left(\mathrm{d}x\right),$$

where b is the true drift without the truncation component $iux1_{|x|\leq 1}$ so that,

$$\gamma = b + \int_{-1}^{1} x\nu\left(dx\right).$$

The variable γ is sometimes loosely referred to as the "drift" of the Lévy process, but it is, in fact, dependent on the choice of truncation function [CT03].

The following corollary illustrates the simplification of conditions stated in Proposition 2.5.7 in the case of compound Poisson jumps.

Corollary 2.5.8. Conditions for the equivalence of probability measures for compound Poisson jumps (Cont and Tankov [CT04])

Let $(X_t, \mathbb{P})$ and $(X_t, \mathbb{P}')$ be two Lévy processes on $(\Omega, \mathcal{F})$ with characteristic triplets (σ, ν, γ) and (σ', ν', γ'). Suppose that the jump part of X_t is of compound Poisson type, then $\mathbb{P}\mid_{\mathcal{F}_t}$ and $\mathbb{P}'\mid_{\mathcal{F}_t}$ are mutually absolutely continuous for all t if and only if the following conditions are satisfied:

1. *$\sigma = \sigma'$.*
2. *The jump part of X_t is of compound Poisson type, and the two jump size distributions are mutually absolutely continuous.*
3. *If $\sigma = 0$, then in addition we must have $b' = b$.*

The definitions of the Radon-Nikodym derivative and U_t correspond to the definitions stated in Proposition 2.5.7 above.

Proof.

The second condition of Proposition 2.5.7 is fulfilled automatically, since

$$\int_{-\infty}^{\infty} \left(e^{\phi(x)/2} - 1\right)^2 \nu\left(dx\right) \leq 2\int_{-\infty}^{\infty} \left(\nu\left(dx\right) + \nu'\left(dx\right)\right) < \infty.$$

The Lévy process U_t also has a compound Poisson jump part, since

$$\int_{-1}^{1} \nu_U\left(dx\right) = \int_{-1}^{1} \left[\nu\phi^{-1}\right]\left(dx\right) = \int\limits_{-1\leq\phi(y)\leq 1} \nu\left(dy\right) < \infty.$$

□

We observe that for two jump-diffusion Lévy processes to have equivalent probability measures, we require

- diffusion coefficients that are equal and non-zero, and
- jump-size distributions with the same null sets.

Consider, for example, the case where X_t is a Merton process with a normally distributed jump size distribution, a model which will be discussed in Section 2.6.1.1. One can use the following approach to identify an equivalent martingale measure and still have wide scope for changing the jump characteristics: If one selects a compound Poisson process X'_t with the same diffusion component as the Merton model, so that $\sigma = \sigma'$, then one can select a jump distribution with any shape, provided that there are no jump-size sets that have probability zero. Since b' is not tied to b in the case of non-zero diffusion coefficients, one has the freedom to select b' to satisfy the martingale condition from (2.9) in

$$\psi(-i) = b' + \frac{\sigma'^2}{2} + \int_{-\infty}^{\infty} \left(e^x - 1\right) \nu'\left(\mathrm{d}x\right) = 0. \tag{2.10}$$

In the preceding propositions, we showed that one could obtain equivalent martingale measures with different jump distributions by adjusting the drift in cases where the diffusion component is non-zero. We will now show how the Esscher transform can be applied as an alternative method of obtaining equivalent martingale measures in Lévy processes without a diffusion component.

Definition 2.5.9. The Esscher transform (Hubalek and Sgarra [HS06])

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a random variable X, and a real valued parameter θ, The Esscher transform $\mathbb{P}^\theta$ is defined by

$$d\mathbb{P}^\theta = \frac{e^{\theta X} d\mathbb{P}}{E\left[e^{\theta X}\right]},$$

provided that the expectation exists.

An Esscher transform can be performed by using Proposition 2.5.7 with the transformation of measure function $\phi(x) = \theta x$ [CT03]. Cont and Tankov [CT03] illustrate that for a Lévy process X_t with characteristic triplet $\left(\sigma^2, \nu, \gamma\right)$, with ν satisfying $\int_{|x|\geq 1} e^{\theta x} \nu\left(dx\right) < \infty$, one obtains a Lévy process $\tilde{X}_t$ with

- an equivalent probability measure,
- zero diffusion component,
- Lévy measure $\tilde{\nu}\left(dx\right) = e^{\theta x} \nu\left(dx\right)$,
- drift $\gamma = \gamma + \int_{-1}^{1} x\left(e^{\theta x} - 1\right) \nu\left(dx\right)$ and

- Radon-Nikodym derivative

$$\frac{d\mathbb{Q}\mid_{\mathcal{F}_t}}{d\mathbb{P}\mid_{\mathcal{F}_t}} = \frac{e^{\theta X_t}}{E\left[e^{\theta X_t}\right]} = \exp\left(\theta X_t + \gamma\left(\theta\right) t\right),$$

  where $\gamma\left(\theta\right) = -\log\left(E\left[\exp\left(\theta X_1\right)\right]\right)$.

Proposition 2.5.10. Absence of arbitrage in exponential Lévy models (Cont and Tankov [CT03])

Let $(X, \mathbb{P})$ be a Lévy process. If the trajectories of X are neither almost surely increasing nor almost surely decreasing, then the exponential Lévy model given by $S_t = e^{rt+X_t}$ is arbitrage free: there exists a probability measure $\mathbb{Q}$ equivalent to $\mathbb{P}$ such that $\left(e^{-rt}S_t\right)_{t\in[0,T]}$ is a $\mathbb{Q}$-martingale, where r is the interest rate.

In the following subsection, we will discuss some examples of parametric exponential Lévy pricing models and present the principles for non-parametric exponential Lévy models.

2.6 Exponential Lévy pricing models

In this section, we discuss a number of exponential Lévy models used in asset pricing, with the price of an asset, S_t, represented as

$$S_t = S_0 e^{rt+X_t},$$

where r is the risk-free interest rate, S_0 is the price of the asset at $t = 0$ and X_t is a Lévy process.

We begin by presenting examples of parametric models in the jump-diffusion class and then discuss infinite-activity pure-jump models. We conclude this section by discussing the principles of non-parametric models.

2.6.1 Parametric exponential Lévy models

In parametric Lévy models, the Lévy measure is specified by a model-specific function with a number of parameters to determine its shape, describing the spread of jump activity over jump size.

In the following paragraphs, we illustrate how parameters can be used to specify the properties of exponential Lévy models by presenting two examples from each of the Jump-diffusion and Pure-jump categories. The examples follow the notation of Cont and Tankov [CT03].

Several parametric exponential Lévy models have been introduced in financial literature, and more detailed discussions on various models can be found in texts by Kienitz and Wetterau [KW12], Schoutens [Sch03], and Hirsa [Hir12].

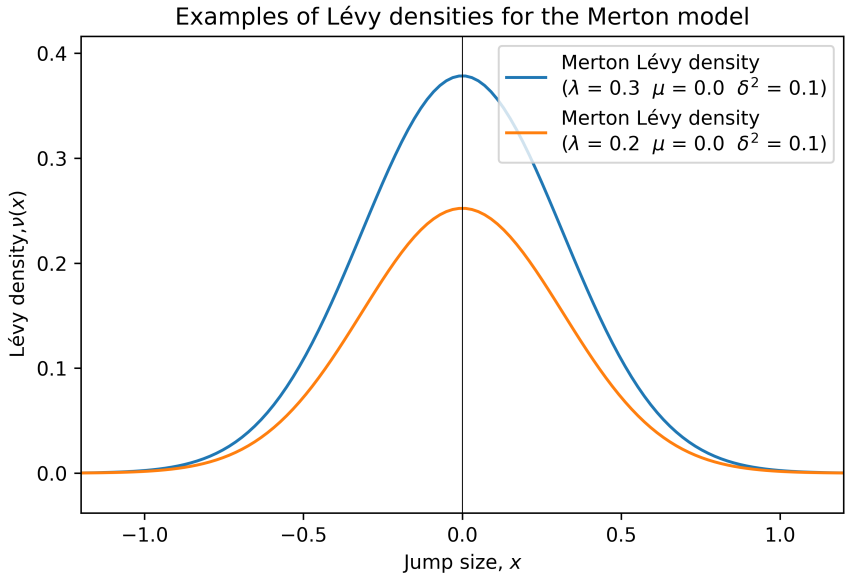

Figure 2.3: Lévy densities of the Merton model.

2.6.1.1 Jump-diffusion models

The Merton model and the Kou model are examples of parametric jump-diffusion models with jumps following the compound Poisson process, which can be expressed as a distribution of jump sizes arriving according to a Poisson process (N_t). The models can be written as

$$X_t = \gamma t + \sigma W_t + \sum_{j=1}^{N_t} Y_j.$$

The jump distribution of the Merton model follows a Gaussian distribution, $Y_i \sim N\left(\mu, \delta^2\right)$, with mean μ and variance δ^2. The characteristic exponent in the form of the Lévy-Khinchin formula is

$$\psi(u) = -\frac{u^2\sigma^2}{2} + iub + \int_{\mathbb{R}} \left(e^{iux} - 1\right) \frac{\lambda}{\sqrt{2\pi\delta^2}} \exp\left(-\frac{(x-\mu)^2}{2\delta^2}\right) (\mathrm{d}x)\,, \tag{2.11}$$

with jump intensity λ, drift parameter b, and diffusion volatility σ^2. Figure 2.3 illustrates some examples of the Lévy densities of the Merton jump-diffusion model.

In the Kou model the Lévy density, $\nu\left(x\right)$, follows a double exponential distribution which can be written as

$$\nu\left(x\right) = \lambda\left[p\lambda_1 e^{-\lambda_1 x} 1_{x>0} + (1-p)\,\lambda_2 e^{-\lambda_2|x|} 1_{x<0}\right],$$

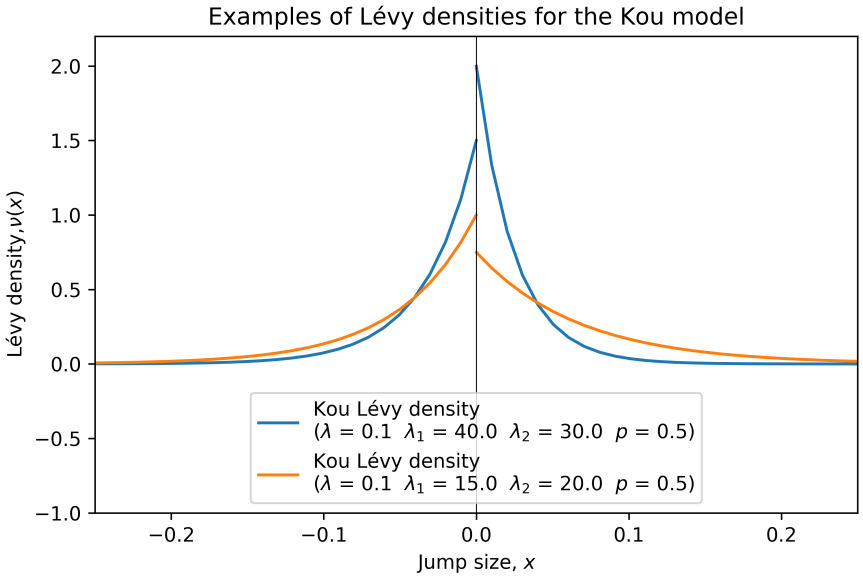

Figure 2.4: Lévy densities of the Kou model.

with λ_1 determining the exponential describing the distribution of positive jumps, λ_2 the negative jumps, and λ representing the overall jump intensity. The variable $p \in [0, 1]$ represents the probability of a positive jump. The corresponding characteristic exponent by the Lévy-Khinchin formula is

$$\psi(u) = -\frac{u^2\sigma^2}{2} + iub + iu\lambda\left(\frac{p}{\lambda_1 - iu} - \frac{1-p}{\lambda_2 + iu}\right),$$

with drift coefficient b, and diffusion volatility σ^2. Examples of Lévy densities from the Kou model are shown in Figure 2.4.

The resulting density functions have tails decaying exponentially in the Kou model, while the tails of the Merton model are heavier than the Normal distribution, but with finite exponential moments [CT03].

2.6.1.2 Pure-jump models

The *Variance gamma* (VG) model and the *Normal inverse Gaussian* (NIG) model are two examples of pure-jump Lévy models with infinitely many jumps in any time interval. Both models are based on Brownian subordination. A subordinator is an increasing Lévy process with positive jumps, positive drift and no diffusion component, which can be used in the construction of other Lévy processes by representing a random forward movement in time. The VG and NIG models are constructed using Brownian subordination with the volatility parameter σ, drift parameter θ, and subordinator variance κ. The Brownian subordination can be interpreted as a Brownian motion with drift on an independent stochastic time scale, S_t, so that the process can be expressed as

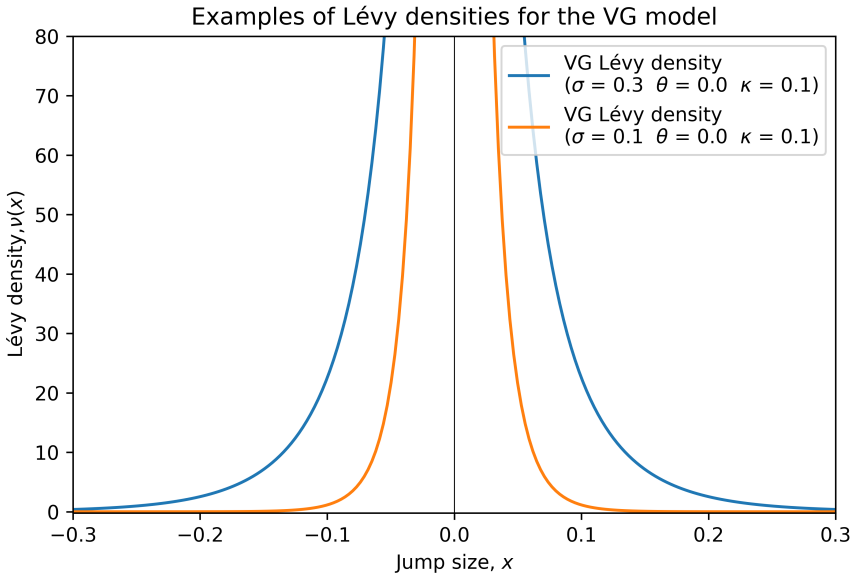

Figure 2.5: Lévy densities of the VG model.

$$X_t = \sigma W(S_t) + \mu S_t.$$

The statistical properties of S_t are determined by the respective subordinating processes. In these two models, the subordinators are the Gamma process and the Inverse Gaussian process. Here we will state key properties of the processes, illustrating how the shape of the Lévy measure is determined by parameters. Further details can be found in Chapter 4 of the book, *Financial modelling with jump processes,* by Cont and Tankov [CT03].

The Lévy measure of the Variance Gamma process is described by the Lévy density,

$$\nu(x) = \frac{1}{\kappa |x|} e^{Ax - B|x|},$$

with $A = \dfrac{\theta}{\sigma^2}$ and $B = \dfrac{\sqrt{\theta^2 + \frac{2\sigma^2}{\kappa}}}{\sigma^2}$.

The corresponding characteristic exponent is

$$\psi(u) = -\frac{1}{\kappa} \log\left(1 + \frac{u^2\sigma^2\kappa}{2} - i\theta\kappa u\right).$$

Figure 2.5 illustrates two examples of Lévy densities for the VG model.

The Normal inverse Gaussian process has Lévy density

$$\nu(x) \frac{C}{|x|} e^{Ax} K_1(B|x|),$$

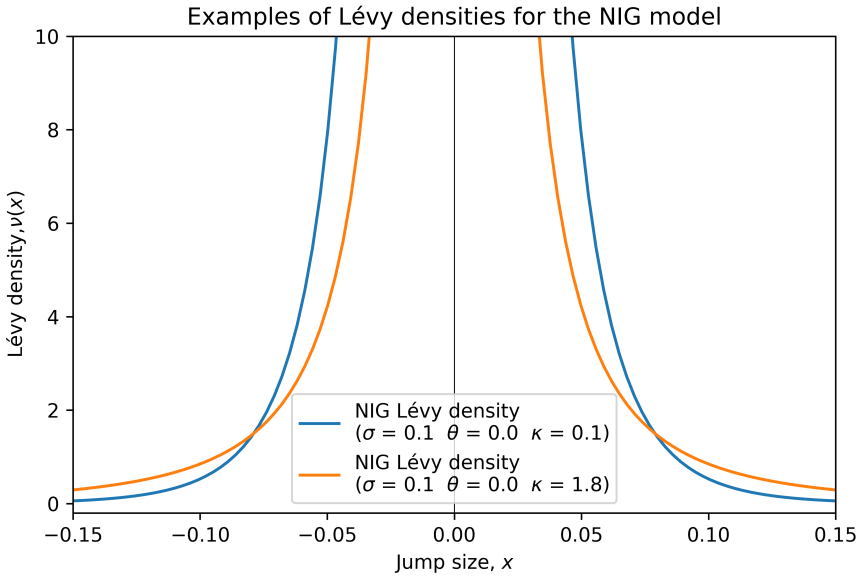

Figure 2.6: Lévy densities of the NIG model.

where K_1 is a type-1 Bessel function, $A = \frac{\theta}{\sigma^2}$, $B = \frac{\sqrt{\theta^2 + \frac{\sigma^2}{\kappa}}}{\sigma^2}$ and $C = \frac{\sqrt{\theta^2 + \frac{\sigma^2}{\kappa}}}{2\pi\sigma\sqrt{\kappa}}$. Examples of Lévy densities from the NIG model are shown in Figure 2.6.

Both models have infinite activity of small jumps, but the Normal inverse Gaussian model has infinite variation, whereas the Variance Gamma model has a lower level of activity corresponding with its finite variation. Both models have exponential tails in their probability densities. Their tail decay rates are given by $\lambda_{\text{left}} = B + A$ and $\lambda_{\text{right}} = B - A$.

A summary of significant features of the four parametric models discussed in this section is shown in Table 2.2 on the next page. We will progress to the discussion of *non-parametric* exponential Lévy models in the next section.

2.6.2 Non-parametric exponential Lévy models

The non-parametric specification of a Lévy density relies on numerical methods to construct a discrete form of the Lévy density composed of a finite number of possible jump sizes. The freedom to specify the shape of the Lévy density directly allows one to construct a remarkably general Lévy model without being constrained to identifying an appropriate parametric model, which may be less consistent with market data. This flexibility can assist in identifying an appropriate model, thereby reducing the risk of model-misspecification. Model risk, which is the uncertainty associated with the selection of a pricing model, is discussed in more detail in Section 4.3.1.

No.	Component	Characteristic function
1	Geometric Brownian motion	$E\left[e^{iuX_t}\right] = \exp\left\{t\left(iu\mu - \frac{u^2\sigma^2}{2}\right)\right\}$
2a	Poisson distribution	$E\left[e^{iuN}\right] = \exp\left\{\lambda\left(e^{iu} - 1\right)\right\}$
2b	Poisson process	$E\left[e^{iuN_t}\right] = \exp\left\{\lambda t\left(e^{iu} - 1\right)\right\}$
2c	Compound Poisson process	$E\left[e^{iuX_t}\right] = \exp\left\{t\lambda\int_{-\infty}^{\infty}\left(e^{iux} - 1\right) f\left(x\right) dx\right\}$
3	Compound Poisson jump-diffusion	$E\left[e^{iuX_t}\right] =$ $\exp\left\{t\left(iu\mu - \frac{u^2\sigma^2}{2} + \lambda\int_{-\infty}^{\infty}\left(e^{iux} - 1\right) f\left(x\right) dx\right)\right\}$

Table 2.1: Characteristic functions of Lévy components [CT03].

Model	Diffusion component	Jump activity	Variation from Jumps	Tail attributes of probability density	Number of parameters
Merton	Yes	Finite	Finite	Heavier than Gaussian	4
Kou	Yes	Finite	Finite	Exponential	5
VG	No	Infinite	Finite	Exponential	3
NIG	No	Infinite	Infinite	Exponential	3

Table 2.2: Key characteristics of selected parametric Lévy models [CT03].

The non-parametric approach applied in this study allows the jumps of a compound Poisson process to be specified as a uniformly spaced discrete Lévy density, forming the jump component of a jump-diffusion process. The jumps described by the compound Poisson process are rather general, and the basic assumptions are limited to:

- exponential (memoryless) inter-arrival times,
- zero mass at the origin and
- finite jump activity.

Within these broad boundaries, the Lévy density can assume virtually any shape. Section 5.3.3 discusses the particulars of the numerical implementation used in this study to represent non-parametric Lévy models in the form of a discrete Lévy density.

2.7 Simulation with Lévy models

The pricing of European options under Lévy models can usually be performed analytically. However, pricing path-dependent exotic options or performing dynamic hedging requires a more intensive reliance on numerical methods. In such cases, one typically resorts to Monte Carlo simulation or the solution of integro-differential equations by numerical procedures such as finite difference methods.

Monte Carlo simulations on a fixed time grid will be used in this study to demonstrate the principles of pricing path-dependent options from calibrated Lévy processes. We select this simple case of fixed-time observations since it is adequate for demonstrating the principles relevant to this study and easily provides exact results. This simplification avoids the non-essential technical detail associated with more complicated simulations, where for example, the entire trajectory may need to be considered rather than just the on-grid points.

A more comprehensive discussion on the Monte Carlo simulation techniques and integro-differential numerical procedures with a specific focus on Lévy processes can be found in Chapters 6 and 12 of the book, *Financial Modelling with Jump Processes* by Cont and Tankov [CT03]. A general discussion on Monte Carlo techniques can be found in *Monte Carlo Methods in Financial Engineering* by Glasserman [Gla04].

The jump component of a finite-activity Lévy process can be described by a compound Poisson process, as mentioned in Section 2.2. The compound Poisson process is well suited to path simulation since there are a finite number of jumps in a bounded interval, and the distribution of jump sizes is known. As a result, path simulation of compound Poisson processes in an interval can be performed without discretisation error as one is required to simulate a finite number of jump times with a correspondingly finite number of jump sizes.

Since the diffusion and jump components of a jump-diffusion process are independent, they can be simulated separately. This is illustrated in the simulation procedures described below, based on Algorithms 6.2 and 6.3 by Cont and Tankov [CT03].

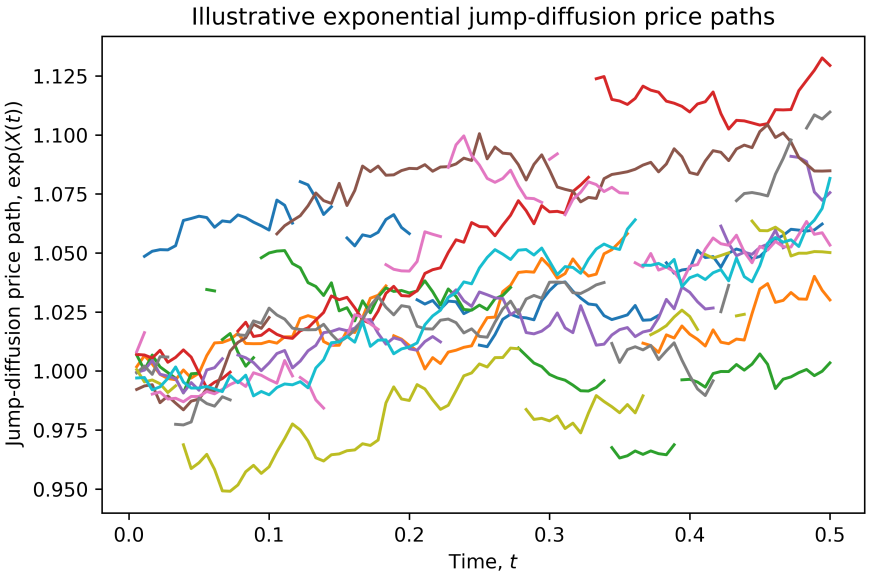

Figure 2.7: Price paths generated using an exponential jump-diffusion model with bimodal Lévy density avoiding small jump sizes.

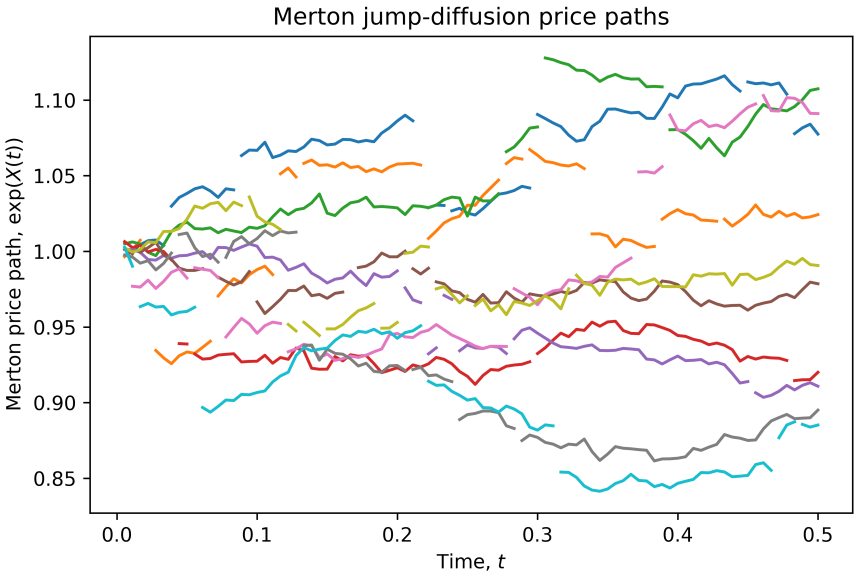

Figure 2.8: Price paths generated using an exponential Merton jump-diffusion model illustrating the inclusion of small jump sizes.

The steps to simulate a pure-jump compound Poisson process:

1. Determine the *number of jumps* N in the simulated trajectory on the interval $[0, T]$ by generating a Poisson random variable with parameter λT.
2. Determine the N independent *jump times* U_i for the simulation using the uniform distribution over $[0, T]$.
3. Determine the N independent *jump sizes* Y_i following the law given by $\frac{\nu(dx)}{\lambda}$.
4. The jump path is then given by

$$X(t) = \sum_{i=1}^{N} 1_{U_i<t} Y_i.$$

The above-mentioned pure jump compound Poisson process can be extended to a jump-diffusion process with characteristic triplet $\left(\sigma^2, \nu, b\right)$ on a discrete time-grid by adding terms for the diffusion and the drift at times $t_1, \ldots, t_n$ giving

$$X(t_j) = \sum_{i=1}^{N} 1_{U_i<t_j} Y_i + \sum_{k=1}^{j} G_k + bt_j,$$

where $t_0 = 0$ and G_j is a set of n independent Gaussian random variables with mean zero and variances $\mathrm{Var}(G_j) = (t_j - t_{j-1})\,\sigma^2$.

These discrete time jump-diffusion simulations are suitable for the simulation of paths for discretely observed path-dependent options. When pay-offs are based on discrete observations, they can be simulated exactly using this procedure without introducing *pay-off discretisation errors* [Gla04]. When pay-offs, however, depend on observing the entire trajectory, the diffusion component will need to be approximated using an approach such as an Euler approximation on a finer grid. In contrast to this, the jump times are simulated exactly, and therefore do not depend on the discrete time-grid. Simulations in this study will, for simplicity, be limited to a fixed uniform time grid suitable for discretely observed pay-offs.

Figure 2.7 on the preceding page shows price paths generated by an exponential jump-diffusion model with a bimodal jump distribution based on the density in Figure D.2 on page 168, illustrating a fairly narrow range of jump sizes. In contrast, Figure 2.8 on the preceding page illustrates exponential price paths from the Merton jump-diffusion model with a wider variety of jump sizes and a more significant incidence of small jumps.

Chapter 3

Options pricing by Fourier transform

In this chapter, we discuss the theoretical framework for options pricing by Fourier transform methods. The pricing procedures allow efficient pricing of the Lévy models explored in the previous chapter, which can then be used in calibration procedures as discussed in the next chapter.

We start by reviewing pricing by expectation and then advance to Fourier pricing techniques, with specific emphasis on the Discrete Fourier Transform (DFT) and its numerically efficient implementation in the Fast Fourier Transform (FFT).

Efficient Fourier transform pricing is an essential aspect of the calibration procedures described in this study since the procedures rely on pricing options over a whole range of strike prices repeatedly when model parameters are varied to search for an optimal fit to market prices. Furthermore, Fourier transform pricing is well suited to Lévy models as most Lévy models can be described by a characteristic function via the Lévy-Khinchin formula, discussed in Section 2.4.2.

3.1 Expectation pricing using a density function

The risk-neutral pricing of a pay-off function H under $\mathbb{Q}$ was presented in Proposition 2.5.3 on page 26 as

$$\Pi_t(H) = e^{-r(T-t)} E_{\mathbb{Q}}[H \mid \mathcal{F}_t].$$

The price of a vanilla call option with pay-off function $H = (S_T - K)^+$ can at time $t = 0$, under the risk-neural measure $\mathbb{Q}$, be expressed as

$$C_T(K) = E_{\mathbb{Q}}\left[e^{-rT}(S_T - K)^+\right]. \tag{3.1}$$

A known risk-neutral density function, $f_T(S)$, of the underlying at maturity, allows the call price to be obtained by evaluating the expected value [Sch03] by using

$$\begin{aligned} C_T(K) &= e^{-rT}\int_{-\infty}^{\infty} f_T(S)(S-K)^+ dS \\ &= e^{-rT}\int_{K}^{\infty} f_T(S)(S-K) dS. \end{aligned}$$

It is often convenient to express the pricing in exponential form. Consider a European call option with maturity T, log-strike $k = \log(K)$, where the risk-neutral density under measure $\mathbb{Q}$ is $q_T(s)$, and $s = \log(S)$ for underlying, S. Following the expectation pricing argument in (3.1), the option price, C_T is the discounted expected value under $\mathbb{Q}$ of the pay-off, written in exponential form as

$$C_T(k) = e^{-rT}E_{\mathbb{Q}}\left[\left(e^{s_T} - e^k\right)^+\right].$$

When q_T is available in closed form, the call price at a single strike can be calculated by direct integration as follows:

$$C_T(k) = e^{-rT}\int_{k}^{\infty} q_T(s)\left(e^s - e^k\right) ds. \tag{3.2}$$

However, for model classes where the density function is not available in closed form, one can, in several cases, rely on the characteristic function representation of a distribution and then use transform-techniques for pricing derivatives [Hir12].

3.2 Options pricing by Fourier inversion

The Fourier transform,

$$\mathcal{F}f(u) = \int_{-\infty}^{\infty} e^{iux} f(x)\, dx, \tag{3.3}$$

the inverse Fourier transform,

$$\mathcal{F}^{-1}f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-iux} f(u)\, du,$$

and the characteristic function were defined in Section 2.4, and then some examples of the characteristic functions of Lévy processes were discussed. This section explains how the characteristic function, introduced in Section 2.4.1, can be employed to perform Fourier transform pricing.

As explained in Section 2.4.1, in cases where the probability density function of a distribution is known, the characteristic function can be obtained by using

$$\phi_X(u) = E\left[e^{iuX}\right] = \int_{\mathbb{R}} e^{iux} f_X(x)\, dx,$$

which is the Fourier transform of the probability density function, f_X.

Similarly, if the characteristic function of a random variable is known, the inverse Fourier transform of the characteristic function can be used to calculate the expected value numerically. Therefore the characteristic function can be employed to determine the option price for the pay-off function by expectation pricing.

Pricing options via a characteristic function rather than directly from a probability density function has the advantage that one can price with underlying dynamics described by any distribution which has a characteristic function. Several distributions relevant to options pricing have a known characteristic function without a probability density function being available in closed form. Therefore, by using the characteristic function, we open up our study to a much richer set of models.

In their article, *Option valuation using the fast Fourier transform* [CM99], Carr and Madan present Fourier pricing formulations, which are:

- Generic pricing formulae, and allow pricing whenever a characteristic function is known. With this flexibility, they are providing access to a rich set of models by simply including the characteristic function in the pricing formula as a "parameter".
- Compatible with the FFT algorithm, allowing rapid calculation of options prices over a range of prices in a single execution.

Their procedure is limited to pricing path-independent options, and pricing formulae are derived for specific terminal pay-off functions.

After introducing the *Discrete Fourier Transform* and the *Fast Fourier Transform*, the following sections will briefly describe the principles used in the Carr-Madan article to find a Fourier pricing formula[1].

3.2.1 The Discrete Fourier Transform

When numerical solutions are required, the Fourier integral (3.3) can be approximated by taking the *discrete* Fourier transform, generally resulting in a more efficient computation.

Definition 3.2.1. *The discrete Fourier transform of a sequence of N points $\{x_k\}$, $k = 0, 1, \ldots, N-1$, is another sequence of N points $\{y_j\}$, $j = 0, 1, \ldots, N-1$ where*

$$y_j = \sum_{k=0}^{N-1} e^{\frac{2\pi i j k}{N}} x_k \quad j = 0, 1, \ldots, N-1. \tag{3.4}$$

[1] In the following subsections, we will follow the notation and forms of the pricing formulae introduced by Carr and Madan in their original article [CM99]. Some relevant variations in the formulations are mentioned in Appendix B.1.3.

Similarly, the inverse Fourier transform,

$$\mathcal{F}^{-1}f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-iux} f(u)\, du,$$

has a discrete representation

$$x_j = \frac{1}{N}\sum_{k=0}^{N-1} e^{\frac{-2\pi ijk}{N}} y_k \quad j = 0, 1, \ldots, N-1. \tag{3.5}$$

Various formats and conventions exist for the above formulations. This dissertation will use the above mentioned formats. Additional formulations are mentioned in Appendix B.1.1.

The discrete Fourier transform can be used to approximate the Fourier transform[2]. The approximation will result in truncation errors due to the finite summation span as well as errors introduced by the discretisation into a finite number of points. The following section will discuss the approximation and related errors of an inverse transform in more detail.

The fast Fourier transform is a computationally efficient algorithm for computing the discrete Fourier transform developed by Cooley and Tukey [CT65]. It can be most efficiently applied when the number of points, N, is an integer power of two and reduces the computational complexity from $O(N^2)$ for an ordinary DFT to $O(N \ln N)$ for the FFT.

Due to its computational efficiency, the FFT algorithm is well suited to cases where repeated calculations are required at several strike prices, such as model calibration. However, in cases where the emphasis is not on speed, but on the accuracy of the pricing of only a few options, Fourier transform pricing via numerical integration algorithms with variable step sizes yields superior results [CT03].

The term *Fourier-domain*[3] will be used in discussions referring to the vectors and values obtained from characteristic functions or functions of characteristic functions. On the other hand, the terms *pricing-domain* or *density-domain* will be used to refer to results obtained from an inverse-Fourier-transform performed on values in the Fourier-domain.

The discrete Fourier points can be transferred and inverse-transformed exactly between the two domains using (3.4) and (3.5). Similarly the FFT and the inverse fast Fourier transform (iFFT) can also transform the points between the domains exactly[4].

[2] The same applies to the inverse Fourier transform.

[3] In Signal Processing, transforms are typically performed between the time-domain and the frequency-domain while image processing usually refers to an image domain and a frequency-domain [BB09], but the transformed image could also be designated as belonging to the Fourier-domain.

[4] In the continuous case, for $f \in L^2(\mathbb{R})$, $\mathcal{F}^{-1}\mathcal{F}f = f$ [CT03].

3.2.2 Truncation and discretisation errors in discrete transforms

The finite number of points used in the discrete Fourier transform to approximate the continuous Fourier integral, which has an infinite integration interval, introduces truncation and discretisation errors.

When approximating options prices using numerical Fourier inversion of a pricing formula, $\zeta_T(u)$, we have

$$\begin{aligned}\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-iuk}\zeta_T(u)\ du &= \frac{1}{2\pi}\int_{-A/2}^{A/2} e^{-iuk}\zeta_T(u)\ du + \epsilon_T \\ &= \frac{A}{2\pi N}\sum_{m=0}^{N-1} w_m \zeta_T(u)\, e^{iku_m} + \epsilon_T + \epsilon_D,\end{aligned}$$

where ϵ_T is the truncation error, and ϵ_D is the discretisation error for the integration span, A. The variables in the summation are: the number of points, N, the weights for the selected integration rule[5], w_m, and discretisation points

$$u_m = -\frac{A}{2} + m\Delta,$$

where Δ is the discretisation step size,

$$\Delta = \frac{A}{N-1}.$$

Section 3.3.2 will discuss the practical considerations for managing the errors mentioned above.

3.2.3 European call pricing by Fourier inversion

In this section, we discuss the pricing procedure developed by Carr and Madan in their article *Option valuation using the fast Fourier transform* [CM99].

The task at hand is to express the Fourier transform of the call price function, C_T, for maturity T in terms of the *characteristic function* of the underlying, ϕ_T, but it is not possible to take the Fourier transform of the call price in equation (3.2) on page 41 directly. The expression is not integrable because it tends to S_0 as $k \to -\infty$. To work around this obstacle, Carr and Madan introduce a damped version of the call price, which they refer to as the *modified call price*, c_T defined as

$$c_T(k) = e^{\alpha k} C_T(k).$$

The damping factor, α, is chosen to allow convergence of the Fourier integral,

[5] An example of a common integration rule is the trapezoidal rule, which would have unity weights for all points, except the endpoints which would have $w_0 = w_{N-1} = 1/2$.

$$\begin{aligned}\hat{c}_T(u) &= \int_{-\infty}^{\infty} e^{iuk} c_T(k)\ dk \\ &= \int_{-\infty}^{\infty} e^{iuk} \int_k^{\infty} e^{-rT} e^{\alpha k} q_T(s) \left(e^s - e^k\right)\ ds\ dk,\end{aligned}$$

so that a Fourier pricing formula can be obtained for the modified call price in terms of ϕ_T, the characteristic function of the underlying. The pricing formula,

$$\hat{c}_T(u) = \frac{e^{-rT}\phi_T(u-(\alpha+1)i)}{\alpha^2+\alpha-u^2+iu(2\alpha+1)} \tag{3.6}$$

can generate modified call prices by applying Fourier inversion. Then, un-damping this result will yield normal call prices as follows:

$$C_T(k) = \frac{e^{-\alpha k}}{\pi} \int_0^{\infty} e^{-iuk} \hat{c}_T(u)\ du, \tag{3.7}$$

where the symmetry of a complex integrand with a real result allows the reduction of the limits of integration to the positive half of the real axis.

The estimation of the damping factor is dependent on the characteristics of the integrand. It can therefore vary according to the model and may even vary with the model parameters. This introduces an intricacy which needs to be considered for successful Fourier transform pricing and will be discussed in Section 3.3.1.

3.2.4 Out-of-the-money time-value pricing by Fourier inversion

Carr and Madan offer another variant of the pricing formula, which can be applied to avoid some numerical difficulties associated with the above-mentioned modified call pricing approach. Specifically, the integrand in (3.7) tends to become oscillatory at short maturities resulting in numerical inaccuracies. The alternative approach discussed below is not based on the modified call price, but operates on the time-value of out-of-the-money options.

Definition 3.2.2. Out of the money time-value prices (Carr and Madan [CM99])

The out-of-the-money time-value price, z_T is defined as:

$$z_T(k) = e^{-rT} E_Q\left[\left(e^k - e^s\right) 1_{s<k,k<0} + \left(e^s - e^k\right) 1_{s>k,k>0}\right]. \tag{3.8}$$

This can be expressed in integral form as

$$z_T(k) = e^{-rT} \int_{-\infty}^{\infty} \left[\left(e^k - e^s\right) 1_{s<k,k<0} + \left(e^s - e^k\right) 1_{s>k,k>0}\right] q_T(s)\ ds. \tag{3.9}$$

For ease of notation, prices have been scaled so that $S_0 = 1$.

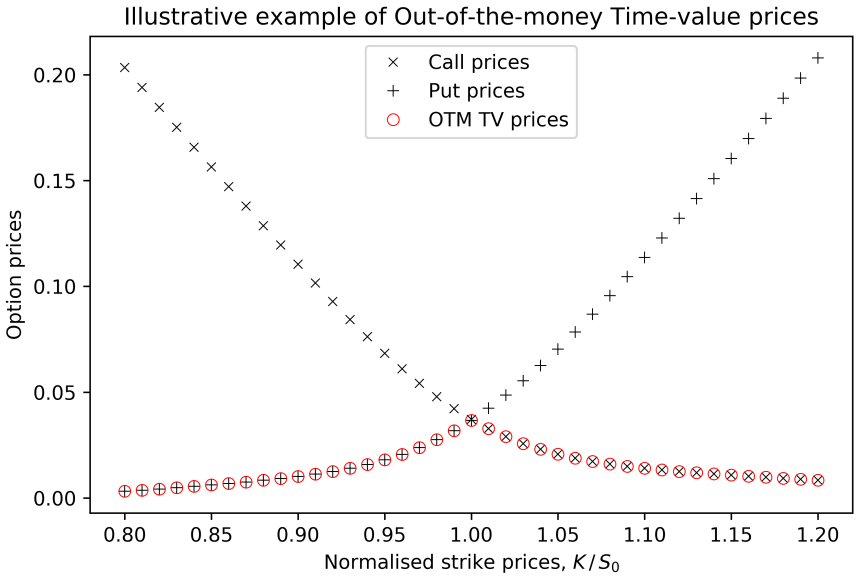

Figure 3.1: Out-of-the-money time-value (OTM TV) prices in relation to put prices and call prices.

In other words, for maturity T, z_T corresponds to the put price for out-of-the-money puts at strikes $K < S_0$, while representing out-of-the-money call prices at strikes $K > S_0$. One can express this as

$$z_T(K) = \begin{cases} e^{-rT} E_Q\left[(S_T - K)^+\right] & K < S_0 \\ e^{-rT} E_Q\left[(K - S_T)^+\right] & K \geq S_0. \end{cases}$$

An illustrative example of out-of-the-money time-value prices, along with corresponding put prices and call prices, is presented in Figure 3.1.

Since $z_T(k)$ decays fast enough as $k \to -\infty$ as well as when $k \to \infty$ without requiring a damping factor, one can perform a Fourier transform directly as follows

$$\zeta_T(v) = \int_{-\infty}^{\infty} e^{ivk} z_T(k)\ dk.$$

An inverse Fourier transform of ζ_T yields a means of pricing out-of-the-money options through z_T, expressed as

$$z_T(k) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ivk} \zeta_T(v)\ dv.$$

The out-of-the-money time-value options prices can be transformed to regular put or call prices by applying put-call parity.

Following the derivation of Carr and Madan [CM99], ζ_T can be expressed in terms of (3.9) as

$$\zeta_T(v) = \int_{-\infty}^{0} e^{ivk} e^{-rT} \int_{-\infty}^{k} \left(e^k - e^s\right) q_T(s)\ ds\, dk \quad + \int_{0}^{\infty} e^{ivk} e^{-rT} \int_{k}^{\infty} \left(e^s - e^k\right) q_T(s)\ ds\, dk.$$

Then, reversing the order of integration, one obtains

$$\zeta_T(v) = \int_{-\infty}^{0} e^{-rT} q_T(s) \int_{s}^{\infty} \left(e^{(1+iv)k} - e^s e^{ivk}\right)\ dk\, ds \quad + \int_{0}^{\infty} e^{-rT} q_T(s) \int_{0}^{s} \left(e^s e^{ivk} - e^{(1+iv)k}\right)\ dk\, ds.$$

Carr and Madan complete the derivation by carrying out the inner integrations, simplifying, and then rewriting the outer integrals in terms of the characteristic function [CM99]. This gives the Fourier pricing function in terms of the characteristic function of the underlying for out-of-the-money time-value prices as

$$\zeta_T(u) = e^{-rT}\left[\frac{1}{1+iu} - \frac{e^{rT}}{iu} + \frac{\phi_T(u-i)}{iu(1+iu)}\right].$$

The value of $z_T(k)$ has a cusp at $k = 0$, which becomes steeper as $T \to 0$. This cusp results in slower decaying tails in the Fourier-domain[6]. The authors suggest adding a damping function, $\sinh(\alpha k)$, so that the transform of $\sinh(\alpha k) z_T(k)$ is performed to reduce the steepness near $k = 0$. The options price obtained by Fourier inversion can then be written as

$$z_T(k) = \frac{1}{\pi \sinh(\alpha k)} \int_{0}^{\infty} e^{-iuk} \hat{z}_T(u)\ du, \tag{3.10}$$

where

$$\hat{z}_T(u) = \frac{\zeta_T(u - i\alpha) - \zeta_T(u + i\alpha)}{2}.$$

3.2.5 Black-Scholes adjusted transform pricing formula

In the numerical implementation of the above-mentioned out-of-the-money time-value approach, which will be discussed in Section C.2.2 on page 149, it was observed that the steepness of the price curve leads to inferior pricing near-the-money. In literature by Cont and Tankov [CT03, CT04], as well as the article by Tankov and Voltchkova [TV09], the authors propose a variation on the pricing formula, which proved to yield

[6] Smooth, continuous functions have faster decaying tails in the Fourier domain, while sharp transitions, such as the cusp in $z_T(k)$ have more "high frequency" components leading to more weight in the tails which will tend to increase truncation errors in the Fourier inversion.

a significant improvement in the pricing accuracy observed in this study. Instead of subtracting the intrinsic value from the options price as proposed by Carr and Madan above, the authors suggest subtracting the Black-Scholes call price as an alternative. This adjustment results in a smoother pricing function better suited to discrete transform pricing with a finite number of points. The specified Black-Scholes call price is calculated using the same interest rate, underlying and maturity as the primary characteristic function, but with an arbitrary volatility Σ. We state their proposition below without proof.

Proposition 3.2.3. Black-Scholes adjusted transform pricing formula (Tankov and Voltchkova [TV09])

Let $\{X_t\}_{t\geq 0}$ be a real valued Lévy process, with Lévy measure ν, such that $\left(e^{X_t}\right)$ is a martingale, and

$$\int_{x>1} e^{(1+\alpha)x}\nu\left(dx\right) < \infty$$

for some $\alpha > 0$. Then the Fourier transform of the modified option price

$$\tilde{z}_T\left(k\right) = C\left(k\right) - C_{BS}^{\Sigma}\left(k\right)$$

in log-strike k is given by

$$\tilde{\zeta}_T\left(u\right) = e^{ivrT}\frac{\Phi_T\left(u-i\right) - \Phi_T^{\Sigma}\left(u-i\right)}{iu\left(1+iu\right)}, \tag{3.11}$$

where Φ_T is the characteristic function of X_T and

$$\Phi_T^{\Sigma}\left(u\right) = \exp\left(-\frac{\Sigma^2 T}{2}\left(u^2+iu\right)\right)$$

is the characteristic function for the Black-Scholes model with a selected volatility of $\Sigma > 0$.

The authors state that the optimal value for Σ is obtained by finding the value for which $\tilde{\zeta}_T\left(0\right) = 0$ [TV09]. In our experiments, we not only observed a significant improvement in pricing accuracy obtained with this approach, but we also observed insensitivity of the level of pricing accuracy to the selected value for Σ. Varying the value of Σ did not cause any meaningful change in the pricing error in our pricing accuracy tests conducted in Section C.2.

3.3 Parameter selection for Fourier pricing

Although calibration techniques built on Fourier pricing methods benefit from the efficiency and flexibility offered by these pricing procedures, some configuration is required to achieve accurate pricing. The computationally efficient FFT is based on the discrete Fourier transform, which in this application serves as a discretised approximation of the continuous Fourier integral. Several parameters must be manually selected to manage

the magnitude of various numerical errors resulting from this discrete approximation. In this section, we consider the parameters that influence the numerical accuracy of the practical computation in Fourier pricing and need to be selected before commencing the calibration process.

The calibration techniques considered in this study rely on pricing through a characteristic function using the FFT. The authors of the articles describing the calibration methods do not specify details of the FFT parameters and treat the pricing process as an autonomous entity that generates acceptable prices without the need for any intervention. Additional verification experiments conducted in implementing these calibration techniques, discussed in Chapter 6 on page 77, suggest that careful selection of transform pricing parameters is essential. In his book, *Computational methods in finance*, Hirsa makes a similar observation after tests demonstrating the sensitivity of the pricing accuracy to parameters[7]:

> "Our analyses indicate when using FFT methods for option pricing, results are very sensitive to the choice of α, N, and η. FFT methods will work in theory for any damping variable α, but one has to be careful as to the choice of N and η. As shown, the most accurate results are achieved by using a very large N and very small interval η, but this choice of parameters is not only computationally expensive but also would not provide a useful range of strikes."

An appropriate damping factor and a good balance between input and output grid spacings will increase the accuracy of the transform pricing. However, the output grid spacing is in log-strike units, which requires interpolation to get options prices at the mostly off-grid market strike prices introducing another potential source of numerical errors.

3.3.1 Damping factor selection

When using the Carr-Madan Fourier pricing formulae, the convergence rate depends on an appropriate selection of the damping factor, α. This requirement applies to both numerical integration and FFT implementations.

In their implementation of the regularised non-parametric calibration, Cont and Tankov [CT04] chose to use their variant of the out-of-the-money time value Fourier pricing formula without any damping[8]. The standard time-value Fourier pricing formula described by Carr and Madan [CM99], shown in equation 3.10, can be used without the $\sinh(\alpha k)$ damping improvement.[9] Similarly, Belomestny and Reiss, in their description

[7] Hirsa uses η to represent the step size applied to the discretisation of the pricing function in the Fourier domain.

[8] Cont and Tankov opted to soften the discontinuity by using the Black-Scholes adjusted formula mentioned in Section 3.2.5 [CT04].

[9] When pricing by Fourier inversion of the European call pricing formula in Section 3.2.3 damping is required, but additional damping it is not essential for the out-of-the-money time-value Fourier pricing formula.

of the spectral approach to non-parametric calibration [BR06], use their variant of the out-of-the-money time-value pricing formula without any damping.

Both of these calibration implementations potentially have increased truncation errors when used with the standard Carr-Madan pricing formulae due to the discontinuity of the time value pricing at $k = 0$. The discontinuity causes a sudden (non-smooth) change, introducing "high frequency" components, which results in heavier tails in the Fourier pricing formula. In Appendix B, we will discuss some differences in the formulation of random variables and differences in pricing formulae specified in these articles. In Section C.2.2, we discuss the errors introduced by the discontinuity in more detail and provide the results of some numerical experiments illustrating the matter.

The call pricing transform formula (3.6) is more sensitive to the selection of an appropriate damping factor, which makes the out-of-the-money pricing formula (3.10) a better choice among the Carr-Madan formulae. The selection of the damping factor varies with the properties of the characteristic function [Lee04]. Lee gives examples illustrating how the choice of optimal α for a particular model, chosen to minimise error bounds, can vary with maturity [Lee04].

Although the damping factor is an important aspect when using the standard Carr-Madan formulae, the Black-Scholes adjusted formula mentioned in Section 3.2.5 uses a different approach to mitigate the truncation errors and therefore does not require a damping factor.

3.3.2 Discrete transform parameter selection

Using the discrete Fourier transform for numerical options pricing offers significant computational benefits, particularly when utilising the FFT algorithm. However, since the discrete Fourier transform is an approximation to the continuous case, one needs to consider certain restrictions and limitations associated with this approximation:

Manual selection of parameters: The grid parameters for the DFT need to be carefully selected to minimise truncation and discretisation errors. Appropriate parameter selection is dependent on the smoothness and rate of decay, which will vary according to the numerical properties of the specific function to be transformed. In order to ensure reasonable accuracy, an inspection of the function is required, and universal parameters are not available for blind application of the DFT in options pricing. In the article, *Option pricing by transform methods: extensions, unification and error control* [Lee04], Lee observes how a change in the maturation of an option can shift the magnitude of the errors between truncation and discretisation errors.

Uniform discretisation grid spacing: The discretisation grid has a uniform spacing, complicating discretisation error reduction in steep regions of the function to be transformed. Reduction of the discretisation error requires an increase in density for the entire grid. It is not possible to selectively increase the density only in the steep areas. An increase in the grid density can be achieved by increasing the number of points, which slows down the computation. The grid

density can also be increased by reducing the span, which generally increases the truncation error. This approach of reducing the discretisation grid spacing by reducing the span leads to increased spacing in the output (pricing) grid, which is discussed next.

Output grid step size: The output grid spacing is also uniform and has the added inconvenience that its step size is inversely related to the step size of the discretisation grid. If one, for example, chooses to decrease the step size of the discretisation grid to reduce discretisation errors, one will introduce a wider step size in the output grid.
The larger gaps between the strikes on the output grid generally increase the interpolation errors, which will be discussed in the next item below. This trade-off between the two step sizes implies that the only way to make both grids dense is to increase the number of grid points, which increases the computation time and the memory requirement.

Interpolation errors: When applying transform pricing, the output is usually on a grid of evenly spaced log-strike prices, which requires interpolation to obtain prices at market strikes which are usually not on the grid points. A sufficiently fine output grid will be required to manage the errors resulting from interpolation, particularly in the area that falls near the money, which is usually steeper.

Balancing errors: Truncation errors result from the inability of a finite integration interval to capture sufficient mass of the tails of the function being transformed due to an insufficient rate of decay of the tails for the selected interval. For a given number of points, reducing the truncation error by widening the discretisation grid step size will tend to increase the discretisation errors as the points are now further apart. This trade-off between truncation and discretisation errors requires a careful selection of parameters. The selection of these parameters is highly dependent on the characteristics of the function being transformed.

The number of points: Usually, the FFT (as a specific implementation of the DFT) has the additional restriction that the number of points must be an integer power of two[10]. Although the FFT is recognised for its efficiency, the number of points remains a key factor for the execution time of the algorithm and should be kept as low as possible within the error limit. For example, an increase in the number of points from 2^{10} to 2^{14} will slow down each iteration of the calibration process by a factor of more than of twenty.

Experience suggests that a degree of trial and error is necessary to "calibrate" the parameters of the FFT pricing against instruments with known model prices before proceeding with model calibration. Appendix C.2 presents the selection and verification processes applied to the transform pricing parameters used in this study.

Model calibration will be discussed in the next chapter.

[10]Some modern implementations of the FFT do not require the number of points to be an integer power of two, but the efficiency of the algorithm is optimal when the number of points is an integer power of two [VGO+19].

Chapter 4

Model calibration and regularisation

Model calibration is a central theme in the procedures evaluated in this study. The calibration procedures that will be examined in this dissertation seek to improve the performance of the calibration of non-parametric Lévy models used for options pricing. Non-parametric Lévy models will be calibrated to the market prices of liquid vanilla call options.

The calibration involves finding the set of model parameters that allows the model to generate prices that are closest to the reference market prices. These calibrated options models can then be used to extrapolate the market dynamics obtained from the liquid options prices to applications where we wish to model the price evolution of the underlying over time. This extrapolation of the dynamics into a price-path evolution can be used for pricing illiquid path-dependent options and for performing dynamic hedging.

In the first section below, we will show how the objective function is used to formulate the calibration as an optimisation process. Then, we will explain the problem of ill-posedness associated with this traditional approach to calibration. The core topic of *regularisation*, an approach to compensate for the ill-posedness by attempting to transform the optimisation procedure into a well-posed problem, will be discussed in the second section of this chapter. We conclude this chapter by discussing *model risk* and *calibration risk*, two sources of risk associated respectively with the appropriateness of the selected model and the properties of the calibration procedure used.

4.1 Objective function and ill-posedness

Once a particular model has been identified to be used for pricing derivatives, the model parameters need to be determined by *calibrating* the model against a set of liquid market prices. The calibration process aims to find the model parameters that will allow the model-generated prices to yield the closest match to the observed market prices. The "closeness" of the model prices to the market prices is measured using an

objective function that indicates the magnitude of the discrepancy between the price sets. The *quadratic pricing error,* calculated from the sum of the squares of errors and the closely related *root mean square error* (RMSE), are frequently used as the objective function in model calibration. For a set of M market prices, C_j, and corresponding model prices, C_j^θ, based on parameter vector, θ, the root mean square error is calculated using

$$\text{RMSE} = \sqrt{\sum_{j=1}^{M} \frac{\left(C_j - C_j^\theta\right)^2}{M}},$$

while the quadratic pricing error is calculated using

$$\text{SSE} = \sum_{j=1}^{M} \left(C_j - C_j^\theta\right)^2.$$

Guillaume and Schoutens suggest that although the RMSE is currently the most widely used objective function for calibration in industry, objective functions based on the average-absolute-error as a percentage of the mean options price (APE) written as

$$\text{APE} = \frac{1}{\underset{j}{\text{mean}}\, C_j^\theta} \sum_{j=1}^{M} \frac{\left|C_j - C_j^\theta\right|}{M},$$

and the average relative error, written as

$$\text{ARPE} = \frac{1}{M} \sum_{j=1}^{M} \frac{\left|C_j - C_j^\theta\right|}{C_j},$$

are equally appropriate measures for the goodness of fit [GS12]. The objective function can be further refined by assigning a weight, w_j to the error in each price pair.

After the selection of the objective function, the optimisation problem can then be formulated using the objective function. Then, an optimisation algorithm can be employed to find the optimal numerical solution, θ^*, which minimises the objective function. The formulation of an optimisation problem using a weighted least squares objective function can be written as

$$\theta^* = \arg\inf_{\theta} \sum_{j=1}^{M} w_j \left(C_j - C_j^\theta\right)^2. \tag{4.1}$$

The weights can be used to control the relative influence of price pairs in determining the optimal parameter set. Additional information, such as the bid-ask spread, can be used to balance the relative importance of the observations based on the liquidity reflected in the spread. This approach will give preference to the information embedded

in the more liquid prices, which will have narrower spreads. The weights in a bid-ask weighting scheme can be expressed as

$$w_j = \frac{1}{\left(C_j^{\text{bid}} - C_j^{\text{ask}}\right)^2}.$$

When the bid-ask spread data is not available, one can assign weights as

$$w_j = \frac{1}{\left[V\left(\tilde{\sigma}_j\right)\right]^2},$$

where $V\left(\tilde{\sigma}_j\right)$ is the Black-Scholes *Vega* calculated using the implied volatility, $\tilde{\sigma}_j$, from the options data [CT04]. This weighting provides an approximation that can be used to scale the options prices so that the objective will essentially minimise the difference in implied volatility between the data and the model.

Another practical approach would be to scale the errors to the relative magnitude of the options price. This scaling can be achieved by selecting weights as

$$w_j = \frac{1}{\left(C_j\right)^2}. \tag{4.2}$$

This weighting scheme levels the order of magnitude of each error term in the objective function relative to the price, similar to the relative scaling in ARPE above.

The Black-Scholes model is an example of a single-parameter model. When calibrating this model, the process aims to find the volatility (the single Black-Scholes parameter) that will minimise the error between the market prices and the formula-generated prices.

An example of a more difficult calibration is the four-parameter Merton model, where the non-convexity of a least-squares based calibration makes it difficult to find a global minimum. Different starting conditions could lead to different local minima returned by the optimisation algorithm. This difficulty is not specific to the Merton model, but is a general problem which illustrates *ill-posedness* in multi-parameter calibrations, a topic discussed further in the text below.

For non-parametric Lévy models, as described in Sections 2.6.2 and 5.3.2, there is even greater difficulty since the calibration aims to retrieve an entire vector of points representing a discrete Lévy measure. The effective increase in the number of unknowns widens the solution space, thereby complicating the search for an optimal solution.

A diagrammatic representation of the interaction between the key elements in a calibration process is shown in Figure 4.1 on page 56. The optimisation algorithm controls the iteration cycles and initiates a new cycle by "suggesting" an improvement on the parameter set. The new parameter set is used to generate a set of model prices, and the objective function measures the error with respect to the market prices. This error measurement serves as feedback to the optimiser on the outcome of the proposed improvement so that the algorithm can attempt to improve on previous attempts, usually with the support of a local gradient estimate.

Definition 4.1.1. Well-posed problem.

A problem is well-posed if a unique solution exists and the solution changes continuously with changes in the input variables to the problem.

There are three potential scenarios that could lead to an ill-posed problem [Cré10]:

- no solution exists,
- multiple solutions exist, or
- the solutions do not depend on the input variables continuously.

The calculation of an option price using a model with its parameter set is an example of a well-posed problem. However, the inverse of this problem, where we wish to find the parameters, given the options prices, is usually not well-posed [GS13]. The optimisation problem in equation (4.1) is an example of an inverse problem that generally does not satisfy the requirements for a *well-posed problem* and is, therefore, *ill-posed*. The non-convexity of the problem results in multiple local minima in the optimisation function. When applying an optimisation algorithm to such a non-convex function, the solutions returned by the algorithms are typically local extrema and will depend on the initial conditions supplied to the optimiser. This results in different "optimal" parameters returned by the optimiser depending on the starting point assigned to each parameter.

Furthermore, the small variations observed in market options prices from day to day may lead to large changes in the optimal parameters [GS13]. This instability with respect to the observed input prices illustrates the lack of continuity with respect to input variables associated with an ill-posed problem. These large changes in the parameters describing the dynamics of the price evolution can lead to unreasonably large daily fluctuations in the pricing of exotic options, even though the calibration precision may be quite similar. Apart from the undesirable fluctuation in pricing exotic options, the numerical instability has an adverse effect on the calibration algorithm [CT03].

In the context of calibrating a Lévy model to options data, one may find that there are several Lévy triplets capable of reproducing the market prices to similar precision and that these "solutions" do not vary continuously with respect to the market data.

Regularisation is a technique that can be used to address the above-mentioned issues associated with the ill-posed calibration problem and will be discussed next.

4.2 Regularisation

Engl et al. state that regularisation methods are numerical methods used to address ill-posed problems [EHN96]. Regularisation methods provide a means of obtaining approximate solutions to ill-posed problems which are stable with respect to small changes in the input variables [TA11]. Regularisation often relies on employing additional information to stabilise an ill-posed problem, thereby preventing overfitting, which is typically encountered in noisy least squares problems.

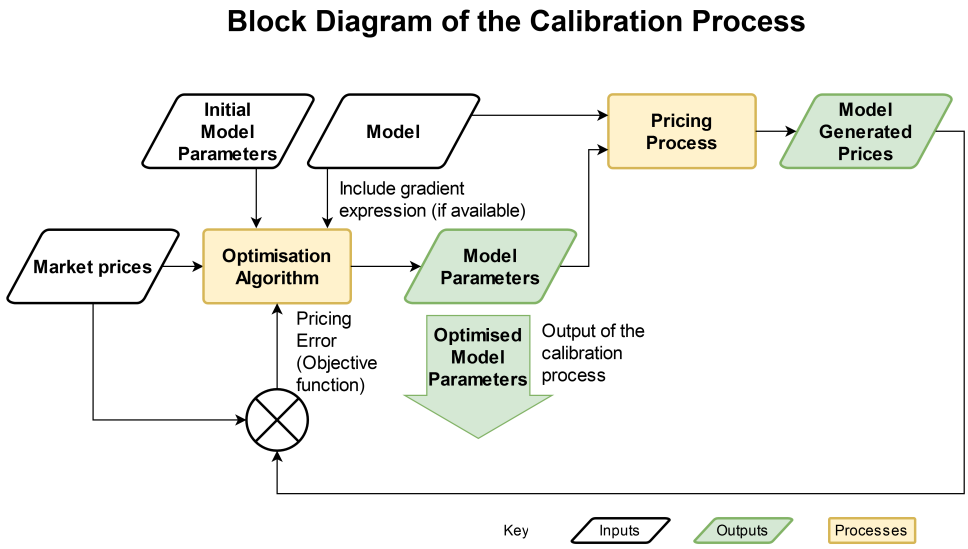

Figure 4.1: Block diagram representation of the calibration process. (Source: personal illustration).

Two forms of regularisation will be discussed in this chapter. Firstly we will discuss regularisation by a penalty term based on *relative entropy* as employed by Cont and Tankov in their research concerning the non-parametric calibration of Lévy models [CT04]. This will be followed by a discussion on *spectral regularisation*, which Belomestny and Reiss employed as regularisation technique in their approach to the non-parametric calibration of Lévy models [BR06].

4.2.1 Regularisation by relative entropy

Generalised Tikhonov regularisation techniques [Vog02] achieve regularisation by introducing a penalty term consisting of a penalty function, G, and a regularisation parameter, $\alpha > 0$, so that the objective function based on a weighted least squares would take the form

$$\theta^* = \arg\inf_{\theta} \sum_{j=1}^{M} w_j \left(C_j - C_j^{\theta}\right)^2 + \alpha G(\theta)\,. \tag{4.3}$$

The regularisation parameter is used to control the amount of regularisation in the objective function by adjusting the magnitude of the penalisation term.

When introducing regularisation as a penalty function in the recovery of a non-parametric Lévy measure, one selects a prior measure as a shaping-benchmark. In principle, one attempts to guide the optimiser to maintain a reasonable degree closeness to the prior measure, wishing to obtain some of the desirable shaping characteristics, while at the same time achieving a reasonable fit to the data. Different methods are employed to measure the "closeness" of the solution to the prior measure.

Conventional Tikhonov regularisation employs a penalty that based on the squared norm of an error vector calculated with respect to a prior parameter vector and is a widely used technique for solving ill-posed problems. Regularisation by relative entropy, another example of a generalised Tikhonov regularisation technique, is closely related to Tikhonov regularisation, but has the distinction that the norm of the error vector is not squared.

Definition 4.2.1. *The relative entropy* $\mathcal{E}(\mathbb{Q}, \mathbb{P})$ *of pricing measure* $\mathbb{Q}$ *with respect to* $\mathbb{P}$ *is defined as*

$$\mathcal{E}(\mathbb{Q}, \mathbb{P}) = E^{\mathbb{Q}}\left[\log\left(\frac{d\mathbb{Q}}{d\mathbb{P}}\right)\right] = E^{\mathbb{P}}\left[\frac{d\mathbb{Q}}{d\mathbb{P}}\log\left(\frac{d\mathbb{Q}}{d\mathbb{P}}\right)\right]. \tag{4.4}$$

The term *entropy* in this context refers to the *Shannon entropy* used in information theory[1] as a measure of information content. Relative entropy, also known as the Kullback-Leibler distance, provides a measure of the "deviation" of a probability distribution with respect to an equivalent reference distribution, often referred to as the prior distribution. Conceptually, relative entropy is analogous to a distance between two distributions but cannot be classified as a metric due to its distribution-wise asymmetry. Relative entropy can be extended from probability measures to Lévy measures. The following section provides a more detailed discussion on relative entropy for Lévy measures.

Relative entropy has some properties that are well suited to serve as a regularisation penalty function for probability measures and the extension to Lévy measures:

1. It is a strictly convex function, as can be seen in the $x \log x$ form of the right-most term in equation (4.4).
2. Relative entropy is non-negative with
   - $\mathcal{E}(\mathbb{Q}, \mathbb{P}) \geq 0$ and
   - $\mathcal{E}(\mathbb{Q}, \mathbb{P}) = 0$ if and only if $\frac{d\mathbb{P}}{d\mathbb{Q}} = 1 \quad a.s.$
3. When a risk-neutral measure is selected as the prior measure, it guarantees an equivalent martingale measure as a solution.

Furthermore, the penalty term based on relative entropy has the benefit of assisting the optimiser in guiding the solution vector points into positive territory in places where the prior is positive, preserving absolute continuity. This influence is caused by the gradient of the penalty term becoming steep near zero, steering the optimiser away from zero.

[1] Entropy is a probabilistic concept used in information theory; typically in situations where a probability distribution is to be estimated from incomplete data with a prior assumption.

4.2.1.1 Relative entropy in terms of Lévy measures

In Chapter 9 of their book, *Financial modelling with jump processes*, Cont and Tankov show how relative entropy can be generalised from the usual definition in terms of probability measures to apply to Lévy measures [CT03]. Their proposition is stated here without proof[2]:

Proposition 4.2.2. Relative entropy for exponential Lévy measures with equivalent probability measures

Let $\mathbb{P}$ and $\mathbb{Q}$ be equivalent measures on $(\Omega, \mathcal{F})$ generated by exponential Lévy models with Lévy triplets $\left(\sigma^2, \nu^P, \gamma^P\right)$ and $\left(\sigma^2, \nu^Q, \gamma^Q\right)$. Assume $\sigma > 0$. The relative entropy $\mathcal{E}(\mathbb{Q}, \mathbb{P})$ is then given by

$$\mathcal{E}(\mathbb{Q} \mid \mathbb{P}) = \frac{T}{2\sigma^2}\left\{\gamma^Q - \gamma^P - \int_{-1}^{1} x\left(\nu^Q - \nu^P\right)(dx)\right\}^2 + T\int_{-\infty}^{\infty}\left(\frac{d\nu^Q}{d\nu^P}\log\frac{d\nu^Q}{d\nu^P} + 1 - \frac{d\nu^Q}{d\nu^P}\right)\nu^P(dx). \quad (4.5)$$

In the case where $\mathbb{P}$ and $\mathbb{Q}$ are equivalent measures of risk-neutral exponential Lévy models, the relative entropy can be reduced to

$$\mathcal{E}(\mathbb{Q} \mid \mathbb{P}) = \frac{T}{2\sigma^2}\left\{\int_{-\infty}^{\infty}(e^x - 1)\left(\nu^Q - \nu^P\right)(dx)\right\}^2 + T\int_{-\infty}^{\infty}\left(\frac{d\nu^Q}{d\nu^P}\log\frac{d\nu^Q}{d\nu^P} + 1 - \frac{d\nu^Q}{d\nu^P}\right)\nu^P(dx). \quad (4.6)$$

It is important to note that there is only one diffusion component, σ^2, shared by both Lévy models. This single diffusion component follows from the requirements for equivalence of probability measures stated in Proposition 2.5.7.

Therefore, in a calibration process of exponential Lévy models relying on regularisation with relative entropy, the diffusion component cannot be optimised in the calibration process, but remains fixed at the value acquired from the prior model. This inability to calibrate the diffusion component is one of the disadvantages of using this calibration procedure based on relative entropy.

The prior measure needs to be carefully selected to ensure that the "prior information" used to regularise the calibration model is of value. A procedure for the selection of the prior measure is presented in Section 5.3.

[2] A more rigorous form of the proof with an extension allowing infinite activity Lévy measures can be found in Appendix A of the article *Retrieving Lévy processes from option prices: Regularization of an ill-posed inverse problem* [CT06].

4.2.1.2 The regularisation parameter

The regularisation parameter determines the trade-off between stability and precision in the objective function. The regularisation parameter is represented by α in (4.3) on page 56.

A regularisation parameter of zero at the one extreme will remove regularisation, causing the objective function to revert to the form used in a regular least squares calibration. On the other hand, a very large regularisation parameter will bias the calibration towards the prior measure with little regard for the market data.

Engl et al. provide a simple heuristic motivation [EHN96] describing the principles of choosing the regularisation parameter according to the method proposed by Morozov [Mor66]. Their motivation is based on the goal of choosing the regularisation parameter such that the regularisation is maximised while restricting the contribution of pricing error due to regularisation to the magnitude of the inherent noise in the data. Theoretically, applying this principle ensures maximum stability without introducing a significant increase in the error level [CT04].

If the options data has bid and ask prices available, then ϵ_0 can be determined using

$$\epsilon_0 = \sqrt{\sum_{j=1}^{M} w_j \left| C_j^{\text{bid}} - C_j^{\text{ask}} \right|^2}.$$

In cases where the bid-ask detail is unavailable, one can estimate the "model error" from a weighted least-squares calibration without regularisation. This standard calibration can be performed by setting the regularisation level, α, to zero in Equation (4.3) and determining the error using

$$\epsilon_0 = \sqrt{\sum_{j=1}^{M} w_j \left| C_j - C_j^{\theta^*} \right|^2}. \tag{4.7}$$

The value of ϵ_0 represents the error level at the best fit to the data without any contribution from the relative entropy term. Introducing a relative entropy penalty will sacrifice the precision of this fit for improved smoothness by "pulling" the optimal solution towards the prior measure. The optimal pricing error using regularisation applied with factor α, represented by $\epsilon(\alpha)$, should ideally be of a similar order of magnitude as the pricing error without regularisation, ϵ_0.

To achieve this similarity in error magnitude, Cont and Tankov suggest choosing an acceptable error inflation factor, δ, for example, $\delta = 1.1$, representing a magnitude increase of 10% [CT04]. Then solve for the regularisation factor, α, to satisfy $\delta\epsilon_0 = \epsilon(\alpha)$. One can find a value for α a with a few iterations of the bisection method.

Although this approach of selecting the regularisation factor seems reasonable from a theoretical point of view, applying this procedure in practice can be challenging. We will explain in Section 5.5 that the above procedure of choosing the regularisation parameter by first determining the "model error" can be problematic when dealing with a non-convex objective function.

4.2.1.3 The impact of regularisation by relative entropy

The beneficial impact of regularisation by relative entropy as a strategy to solve ill-posed non-parametric calibration problems seems to be somewhat uncertain and is a key point being explored in this dissertation. Published statements regarding the expected results of this technique vary. In Chapter 7, we will discuss some of these statements in more detail in conjunction with our experimental observations.

4.2.2 Spectral regularisation

Spectral regularisation relies on the signal processing principle of *filtering* to smooth out the "noise" resulting from the over-fitting of an ill-posed optimisation problem - typically in the presence of noisy data.

In the analysis of time-series data, the *moving average* is frequently used to smooth out short-term fluctuations resulting from noisy data so that underlying trends and seasonal cycles can be identified. One can then control the smoothing characteristics by selecting the number of points to include in the moving average as well as the weighting of the points.

When viewed from a signal processing perspective, the moving average is an elementary *low pass filter* which attenuates the high-frequency components corresponding to the short-term fluctuations while preserving (passing through) the lower frequency components representing the slower moving trends and cycles of the time series.

Time-domain signals can be transformed[3] and represented in their *frequency-domain* equivalent, revealing the magnitude of their constituent frequency components. Applying a low-pass filter in the frequency-domain corresponds to reducing the amplitude of the high-frequency components, while leaving the low-frequency components mostly unaltered. Figure 4.2 on the following page illustrates the filtering effect of an n-point moving average from a frequency-domain perspective. This diagram shows the corresponding attenuation in the higher frequencies from the time-domain operation of applying a moving average. Observe how both the yellow and the blue graphs have an amplitude of one at the low frequencies on the extreme left. This corresponds to low frequencies being "passed" at their full amplitude, without being attenuated. When we look at their amplitudes at a frequency near 0.4, we see that the blue graph is still passing a large portion of the signal, whereas the yellow graph, which represents a more sophisticated filter with a sharper rate of attenuation, will reduce the amplitude to less than 0.25 for all frequencies above 0.4.

In principle, one could have taken a Fourier transform of the original time-series data, applied the illustrated frequency-domain filter profile for the moving average, and then transformed it back to the time-domain and achieved the same result as a moving average.

Although useful for illustrative purposes, performing this particular operation in the frequency-domain would not have much benefit. It would be easier to just apply a

[3] In the context of this dissertation the Fourier transform will be used, but other transforms such as the Laplace transform and Z-transform are also used to switch between the time-domain and frequency-domain representations of a signal or function.

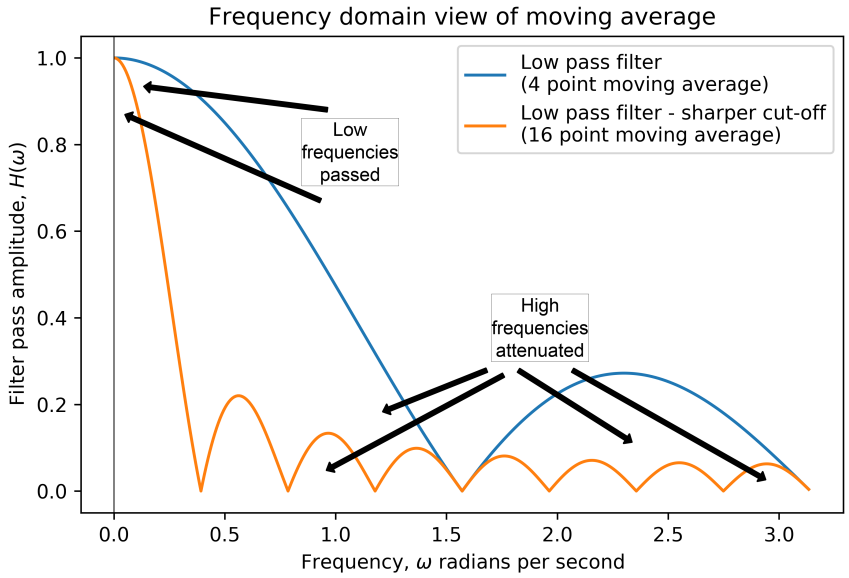

Figure 4.2: Frequency-domain representation of a moving average as a low pass filter.

moving average in the time-domain. Besides, from a signal processing point of view, this filter would be considered to be a rather poor low-pass filter with a weak rate of attenuation.

However, the true benefit of applying the filtering in the frequency-domain becomes evident when one considers that the filter properties can be precisely controlled in the frequency-domain. This allows the exact specification of the level of attenuation versus frequency for the signal, which could, for example, be set up to completely truncate the signals above a specified frequency.

In the article *Spectral calibration of exponential Lévy models* [BR06], Belomestny and Reiss apply spectral regularisation in their non-parametric calibration procedure by selecting the optimal cut-off frequency. All spectral information above the cut-off frequency is truncated, thereby completely removing the high-frequency components. Their procedure aims to select an optimal cut-off frequency in order to reduce the effect of the noise in the data on the recovered Lévy measure, while still preserving the relevant contribution of the data.

4.3 Model risk and calibration risk

When pricing instruments such as OTC path-dependent exotic options where market prices are not available, one needs to rely on the simulation of paths generated by calibrated models to obtain prices. This method of pricing is achieved by firstly extracting the dynamics of the underlying asset embedded in the market prices of vanilla options by calibrating a suitable model to the reference market prices. One then proceeds by generating a price-path simulation for pricing the instrument. This process requires

that one selects not only a *model* that is able to acquire the relevant fundamental characteristics from the market prices, but also a reliable *calibration procedure* to extract these characteristics from the market data. With these two components successfully selected, one would expect to replicate the paths with sufficient reliability to allow dependable pricing.

Mispricing of exotic derivatives resulting from such a calibration-and-extrapolation exercise can have severe real-world financial implications, and due consideration should be given when selecting the model as well as the calibration process [GS12]. In the context of pricing financial instruments, Morini defines *interpolation* as the use of a calibrated model to price "very similar products", where the term *extrapolation* is used when pricing a product that differs from the liquid product used for the calibration [Mor11]. He points out that there is a higher risk of mispricing when extrapolating compared to interpolating and that this is further increased as the complexity of the priced product increases relative to the calibration set.

In the immediately following section, we will discuss *model risk*, the risk associated with the uncertainty regarding the adequacy of the model used for the pricing. We conclude with a discussion on *calibration risk*, the consequence of pricing variations that occur when different calibration techniques are applied to the same model and identical data.

4.3.1 Model risk

When pricing options for which no market prices are available, such as illiquid or over the counter exotic options, it is *necessary* to use a pricing model to extrapolate or interpolate from market prices of more liquidly traded instruments. One can view this operation as an arbitrage-free pricing rule which extends from the liquid to the illiquid instruments [Con06b]. The appropriate pricing of these model-priced instruments is therefore dependent on the model to provide an effective interpolation and extrapolation tool based on the market prices of the liquid instruments. Financial decision-makers rely on realistic pricing delivered by these models. However, poor model selection has been identified as a significant factor contributing to financial losses resulting from mispriced instruments [Con06b].

In derivative pricing, *model risk* is typically used to indicate the risk that different models, which have been calibrated to the same data, could yield different prices for an exotic product [Reb02]. Model risk relates to the uncertainty of the choice of pricing model and could lead to the mispricing of derivative products [Con06b].

In the article, "*A perfect calibration! Now what?*", Schoutens et al. demonstrate that there can be significant differences in the pricing of exotic options across different models in spite of the fact that the models are well calibrated to vanilla options [SST03]. They conclude that models capable of matching European call prices do not necessarily lead to matching prices of path-dependent exotic options, even though the models selected for these tests exhibit a reasonable degree of sophistication, including features such as stochastic volatility.

It is evident that model risk is an emerging issue in quantitative finance and is receiving much attention from regulators and researchers. In the opening chapter of Morini's

book, "*Understanding and managing model risk*", he explains how the 2007 financial crisis led to changes in the Basel regulations aimed precisely at reducing model risk [Mor11].

Non-parametric Lévy models seek to contribute to the reduction of model risk in the following way:

- By allowing the inclusion of jump-discontinuities, the Lévy model creates a more realistic model of price evolution, which is a factor that influences the pricing of path-dependent exotics.
- The nature of the non-parametric Lévy measure means that there are no shape restrictions, and the model is not tightly coupled to the limitations of specifically selected parametric properties. This generality of the model takes a broader view and allows the calibration procedure to be guided by the data. The inherent noise in the data is then softened by the regularisation to prevent over-fitting.
- The non-parametric approach can be used to verify and test the validity of assumptions made in parametric models [CT04].

Although Lévy processes perform well in single-period calibrations to vanilla options, simultaneous multi-period calibrations appear to be inadequate [TV09, CT03]. In particular, Lévy processes are effective at generating an implied volatility skew for a single maturity but fail to generate a reasonable volatility surface in simultaneous multi-period calibrations [Tan07, CT04]. When calibrated separately, each maturity tends to have unique properties attached to that maturity for the same underlying and trading day. The independent and stationary increments of Lévy processes give rise to a time-homogeneous process where the distributional properties are fixed over time.

Tankov and Voltchkova in their article, *Jump-diffusion models: a practitioner's guide* [TV09], state the matter as follows:

> "If we have calibrated the model parameters for a single maturity T, this fixes completely the risk-neutral stock price distribution for all other maturities. A special kind of maturity dependence is therefore hard-wired into every Lévy jump-diffusion model, and (as shown) it does not always correspond to the term structures of market option prices."

One should therefore take account of the multi-period difficulties associated with Lévy processes and determine if it is appropriate for a particular application to reduce the possibility model risk

4.3.2 Calibration risk

In the article, "*Calibration risk: Illustrating the impact of calibration risk under the Heston model*" [GS12], Guillaume and Schoutens indicate that in addition to model risk, there is also a risk associated with the details of the calibration procedure. They refer to this risk as *calibration risk* and demonstrate how within the same model, different

calibration techniques deliver different prices for exotic instruments. Different formulations of the objective function, as discussed in Section 4.1, were shown to influence the outcome of the regularised pricing algorithms examined in this study.

In a lecture titled, *Inverse problems in options pricing*, presented by Rama Cont at a workshop on statistical methods for inverse problems, he states that different regularisation techniques tend to lead to different optimal parameter sets [Con06a]. Therefore, within a given model that is regularised using different methods, one may expect variations in the calibrated parameters.

Part III

Examination of calibration procedures

Chapter 5

Calibration process details

This chapter describes the numerical implementation of the relative-entropy-based calibration procedures. We will describe the numerical procedures required to calibrate non-parametric Lévy models using the procedures proposed by Cont and Tankov [CT04]. We will examine the practical steps of their procedure in detail. In addition, we will highlight some important decisions required for appropriate implementation.

The first section below will outline the process, breaking it down into individual steps. The subsequent sections will then examine the practical details of each of these steps.

5.1 Overview of the regularised calibration procedure

Cont and Tankov proposed a method of regularised least squares for the non-parametric calibration of jump-diffusion Lévy models [CT02, CT04, CT06]. They regularise the calibration by introducing a penalty term into the objective function. They selected the *relative entropy* of the discretised Lévy measure with respect to a pre-selected prior Lévy measure as their penalty function.

The final result of such a calibration process is the "optimal" discrete Lévy measure, resulting from the following key elements:

- a gradient-based optimisation algorithm,
- the market prices of vanilla options,
- a prior Lévy measure and
- a prior diffusion parameter for the Lévy process.

The diffusion parameter, representing the volatility of the Lévy diffusion component, is not calibrated but fixed at a value determined by the prior Lévy measure.

Cont and Tankov chose to simplify the numerical implementation by restricting the calibrated Lévy process and the prior measures to *jump-diffusion* models [CT04]. This simplification implies that models will be restricted to finite-activity models with a

non-zero diffusion component. Within the group of jump-diffusion models, there is still scope for approximating infinite activity behaviour by choosing a relatively small diffusion component and a sizeable jump-activity level for the small jumps.

In this dissertation, we intend to perform an unbiased assessment of these calibration procedures and therefore follow the original procedures as closely as possible.

We summarise the regularised calibration process as follows:

1. Configure the pricing error measurement term of the objective function.
2. Obtain a suitable *discretised Lévy measure* as the *prior* measure for calculating the relative entropy penalty term.
3. Determine the *diffusion component* that is compatible with the prior measure obtained in the preceding step.
4. Select a value for the *regularisation parameter* to control the amount of regularisation-penalty to apply in the objective function used by the optimisation algorithm.
5. Select an *initial value* to determine the starting location for the optimiser. This initial-value vector will be in the form of a discrete Lévy measure with the same grid spacing as the discretised prior measure.
6. Use the parameters of the preceding steps to determine the *optimal discrete Lévy measure* by running the optimiser: Calculating on each iteration:
   (a) the value of the *regularised* least squares objective function, along with
   (b) the corresponding gradient (with respect to the discrete Lévy measure).

In the subsequent sections of this chapter, we will discuss these steps in greater detail, emphasising practical implementation.

5.2 Configuration of the pricing error term

In the formulation proposed by Cont and Tankov, the pricing error in the objective function is determined by employing a *weighted* least-squares calculation [CT04]. This formulation allows the practitioner to specify the relative influence of the individual market prices in determining the optimal solution. Examples of weighting schemes are mentioned in Section 4.1. The first step of the optimisation procedure can be described in our notation[1] as follows:

Select appropriate weights, w_j for the option prices in the least squares calculation of

$$\sum_{j=1}^{M} w_j \left(C_j - C_j^{\theta}\right)^2,$$

[1] We follow the notation used for the objective functions introduced in Section 4.1. In addition, our notation is summarised in Appendix 4.1.

which is the first term of the objective function

$$J(\theta) = \sum_{j=1}^{M} w_j \left(C_j - C_j^{\theta} \right)^2 + \alpha H(\theta), \tag{5.1}$$

with relative entropy term $H(\theta) = \mathcal{E}(\mathbb{Q}_\theta \mid \mathbb{P}_0)$ as defined in (4.6) from Proposition 4.2.2, where $\mathbb{P}_0$ is the probability measure generated by the prior, and $\mathbb{Q}_\theta$ is the equivalent probability measure generated by parameter set θ. In Equation (5.1), α represents the regularisation parameter.

The weights are applied to a set of M market prices, C_j, and corresponding model prices, C_j^θ, based on parameter vector, θ. The values of θ and C_j^θ are calculated on each iteration of the optimiser.

The weights should be chosen to balance the relative importance of the market prices at different strike prices and is a reflection of our confidence in the integrity of each input price.

In our implementation, we tested optimisation using the weighted least squares with uniform weights and a weighting scheme to balance the relative magnitude of the options prices shown in (4.2) on page 54.

Furthermore, we tested the optimisation performance with the same weighting schemes applied to an RMSE error term. The RMSE formulation is mentioned in our discussion on objective functions in Section 4.1. In Section 6.2.6.5 we will explore the impact of varying the objective function formulation.

5.3 Preparing the discrete prior measure

In this section, we will start by explaining the process of acquiring a prior Lévy measure. This will be followed by a discussion on the non-parametric representation of a Lévy measure, which we introduced in Section 2.6.2. Finally, we will explain the discretisation process used to transform the acquired prior measure into the non-parametric form.

5.3.1 Selecting the prior measure

Upon starting a new relative-entropy-based calibration, we need to initialise the calibration process with a *prior* estimate of the Lévy measure, ν^p, and the associated diffusion parameter σ_p.

By introducing the prior Lévy measure, we are introducing additional "prior knowledge" about the market to supplement the noisy market data wishing to stabilise the non-parametric calibration problem. Cont and Tankov refer to the prior as the "user's view of the model" [CT04]. The magnitude of the relative-entropy-based penalty term increases with increased deviation from the prior measure, as explained in Section 4.2.1. Therefore, one should choose the prior to be a reasonable reflection of the views on

the market so that the regularisation can draw the Lévy measure towards a plausible reference.

The prior should be selected from the family of compound Poisson jump-diffusion models with a non-zero diffusion component. As described in Proposition 2.5.7, a non-zero diffusion component allows considerable freedom to change the shape of the Lévy measure while allowing easy adjustment of the drift to satisfy martingale conditions.

When performing the first calibration in a periodic series without the availability of historical calibration results, the "initial" prior needs to be estimated. Some approaches suggested by Cont and Tankov [CT04] for obtaining the initial prior measure are:

- Applying statistical estimation techniques to the time-series data of the underlying to obtain parameters for a Lévy model directly from the underlying.
- Choosing a model and parameters that "seem reasonable".
- Calibrating a parametric jump-diffusion model, such as the Merton model or the Kou model, to options data using ordinary least-squares optimisation methods.

Although the prior is pivotal in the regularisation and a key component in the relative entropy calculation, the authors claim that the importance of selecting the "correct" prior for this procedure is not critical [CT04]. They state that "exact values of the parameters of the prior model are not very important but that it is crucial to find the right shape of the prior" [CT04]. The relationship between the selected prior measure and the calibration result will be investigated more closely in the experiments discussed in Section 6.2.6.3. In our discussion in Section 7.3, we will furnish arguments challenging the above claims of the authors.

The stabilisation of calibrations is best achieved by keeping the prior reasonably static from day to day, considering that the market prices usually tend to fluctuate daily [CT03]. This stability in the prior can be achieved by obtaining a representative prior from observations of various parametric calibrations over a period of time and then keeping the prior fixed for future calibrations. One can also consider the effortless approach of using the Lévy measure from the most recent calibration for the subsequent calibration [CT04]. The process will then need to commence with one of the above-mentioned techniques to obtain the initial prior, yielding a Lévy triplet of the form $(\nu^p, \sigma_p, \gamma^p)$.

A prior measure can be easily obtained by calibrating the Merton model to the options data using a least squares approach. This will be the typical approach used in this study to obtain an initial prior measure. Although the non-convex nature of a least-squares-based calibration of the four-parameter Merton model is unlikely to result in the parameters corresponding to the global minimum, it generally yields an appropriately shaped prior Lévy measure [CT04]. The characteristic exponent of the Merton model is described in Equation (2.11) on page 32.

5.3.2 The discrete Lévy measure

The non-parametric representation of a Lévy model adopted by Cont and Tankov in their work on non-parametric calibration techniques [CT02, CT04, CT06, CT03]

involves representing the Lévy density as a finite set of evenly spaced points. We follow their approach closely, as described next.

In the numerical implementation of the non-parametric Lévy measure, we use an N_ν dimensional vector $(\nu_j,\ j = 0 \dots N_\nu - 1)$ to represent the Lévy measure, ν, as a set of discrete points with a corresponding finite number of N_ν discrete jump sizes $(x_j,\ j = 0 \dots N_\nu - 1)$. The values of the vector elements ν_j correspond to the mass of the jump intensity "around" each possible jump size x_j. The jump component of this Lévy model can be regarded as a weighted sum of N_ν independent Poisson processes, each with its own intensity.[2] The discrete form of the Lévy measure can be expressed as

$$\nu = \sum_{j=0}^{N_\nu - 1} \nu_j \delta\left(x - x_j\right),$$

where we use the sampling property of the the Dirac delta to extract the discrete values at each x_j.

The Lévy-Khinchin representation of the characteristic exponent of the non-parametric Lévy model can be written as

$$\psi(u) = -\frac{u^2\sigma^2}{2} + iub\left(\nu\right) + \sum_{j=0}^{N_\nu - 1} \left(e^{iux_j} - 1\right)\nu_j, \tag{5.2}$$

where $b\left(\nu\right)$ is the drift component dependent on ν. When the process is required to satisfy the martingale requirements mentioned in Proposition 2.5.6, the drift component can be chosen to satisfy $\psi\left(-i\right) = 0$ for a given ν as shown in (2.10) on page 30.

Choosing a uniformly spaced jump-size grid, x_j, allows one to employ the FFT to efficiently compute summation "integrals" used in the calibration process. In particular, one can calculate the gradient of the objective function (with respect to the discrete Lévy measure) by taking advantage of the FFT when the jump grid spacing is uniform. The computation of the gradient will be discussed in Section 5.6.4.

5.3.3 Discretisation of the prior measure

Given suitable model parameters from the parametric calibration of a jump-diffusion model, the next step is to determine a discrete from of the corresponding Lévy measure on an evenly spaced jump-size grid.

The prior measure ν^p can be discretised into the non-parametric representation

$$\nu^p = \sum_{j=0}^{N_\nu - 1} \nu_j^p \delta\left(x - x_j\right),$$

on an evenly spaced grid using

[2] This is an example of a superposition of independent Poisson processes as described in Section 2.2 on page 15.

$$\nu_j^p = \int_{x_j - \frac{\Delta x}{2}}^{x_j + \frac{\Delta x}{2}} \nu^p (dx) \quad j = 1, \ldots, N_\nu - 2,$$

where Δx is the uniform spacing between the points, x_j, representing the discrete jump sizes.

For the evaluation of the end-point, ν_0^p, integration is performed from minus infinity and similarly for $\nu_{N_\nu - 1}^p$ we integrate to infinity. This approach lumps the mass in the portion of the tails of ν^p that fall beyond the range of the discrete jump vector, into the end points of the vector.

Grid parameter selection for the discrete Lévy measure

It should be noted when selecting the grid parameters for the discretisation of the prior, that the identical grid parameters will apply to the solution vector, which is also a discrete Lévy measure. This consistency in the grid of jump sizes is required for the calculation of the relative entropy as well as the gradient with each iteration of the optimisation algorithm.

The non-parametric representation of the Lévy measure is, in essence, a discrete *approximation* of the Lévy measure on a vector, which represents the jump intensity at a finite number of evenly spaced jump sizes.

The *span* of the jump sizes in the discrete Lévy measure is determined by the *number of distinct jump sizes*, N_ν, and the *grid spacing*, Δx. These three interrelated parameters need to be balanced to achieve a sufficiently rich set of distinct jump sizes to resemble the pricing of the continuous equivalent, without introducing excessive computational burden.

In our preliminary verification tests, detailed in Appendix C, we performed a series of experiments to determine reasonable values for the grid parameters. We discretised the Lévy measure of a Merton model using various grid parameters and compared pricing accuracy against the Merton model.

Our findings suggest that a reasonable balance can be achieved with 256 distinct jump sizes over a span of [−2 to 2] for the jump size. This translates to a pricing error in the order of 10^{-6} for normalised strike prices which corresponds to an error of approximately 10^{-4} percent of the strike price. Detailed results are available in Appendix C.2.4.

5.4 The diffusion component

The volatility, σ_p, used to describe the diffusion component of the prior jump-diffusion is a further parameter of interest obtained from the initial prior Lévy triplet $(\nu^p, \sigma_p, \gamma^p)$. In other words, the diffusion component is established when determining the parameters of the prior by calibrating to a parametric model or by using one of the other techniques mentioned in Section 5.3.1.

Once obtained, the diffusion component remains static during the calibration process, not only as a static prior, but also retains the value to the optimal solution. Specifically, the prior volatility propagates directly to the optimal solution as it remains fixed at the prior value in the interim jump-diffusion candidate solutions of the optimisation.

Maintaining equivalence of measures and risk-neutrality

Fixing the volatility of the candidate calibration solutions to the prior volatility maintains equivalence with respect to the probability measure of the prior in accordance with Proposition 2.5.8. This equivalence of probability measures is required to satisfy the prerequisites for calculating the relative entropy of Lévy measures stated in Proposition 4.2.2 and further discussed in Section 4.2.1.1.

Selecting a finite activity jump-diffusion model with non-zero diffusion component as the prior offers a reasonably broad class of equivalent models: In essence, this opens as potential equivalent solutions, all jump-diffusion models, with significant freedom to determine the shape of the Lévy measure, provided that the same diffusion component as the prior is maintained. Therefore, when selecting a risk-neutral prior from the jump-diffusion class, we can guarantee that the equivalent solutions generated with respect to this prior will be a risk-neutral model as described in Section 2.5.

5.5 Determining the regularisation parameter

The regularisation parameter, introduced in Section 4.2.1, determines the weight assigned to the relative entropy penalty term in the modified objective function, regulating the balance between pricing precision and the stability with respect to input price fluctuations.

As mentioned in Section 4.2.1.2, we require an estimate of the intrinsic noise in the data, ϵ_0. The value of ϵ_0 serves as a lower bound reference level to the achievable calibration error when applying the Morozov discrepancy principle [Mor66]. Since bid-ask spreads were unavailable in the market data used to test the calibration procedures, we opted for the "model error" approach using Equation (4.7) on page 59.

Although it was beneficial to recognise the principles applied in choosing a regularisation parameter to keep the pricing error level roughly in the same magnitude as the model error, an exact execution of their procedure proved to be impossible. The non-convexity of the regularised optimisation problem with sensitivity to initial points prevented the full utilisation of this approach. The problem of non-convexity is discussed in detail in Chapter 7.

We circumvent this problem by avoiding the dependence on a single value for the regularisation parameter. Instead, we chose to consider a sequence of possible values for the regularisation parameter in determining the behaviour of the optimisation procedure.

Pricing errors from non-parametric calibrations (without regularisation) and the pricing errors observed in the parametric calibration of the prior measures served as rough

estimates in determining a feasible range for the magnitude of regularisation-induced errors.

Most of our optimisation experiments were performed by stepping through a range of potential regularisation parameters while observing the changes in optimisation results with changes in the level of regularisation. Our approach of running batches of optimisation tests while stepping through a parameter range is described in Section 6.2.3 on page 86.

5.6 Regularised Optimisation

The optimisation process follows after obtaining the parameters in the preceding steps. In particular, the user-specified-parameters required to proceed with the optimisation are: the prior measure, prior volatility, weights of option prices and the regularisation parameter.

Given a suitable gradient based optimisation algorithm, the optimisation process can be broken down into:

1. Selecting the initial points for the optimiser, more specifically: it is the selection of the initial discretised Lévy measure from which the optimiser will iterate towards the optimal (calibrated) Lévy measure by repeating the next two steps on each iteration.
2. Calculating the objective function with penalty term consisting of:
   (a) the weighted pricing error component and
   (b) the relative entropy of the Lévy measure with respect to the prior measure scaled by the regularisation parameter
3. Calculating the gradient of the objective function with respect to the vector representing the individual points of the discretised Lévy measure.

5.6.1 Choice of optimiser

Cont and Tankov suggest using the BFGS algorithm as optimiser when applying their calibration procedure [CT04]. BFGS is a gradient based quasi-Newton optimisation procedure written by Broyden, Fletcher, Goldfarb, and Shanno [NW06]. Quasi-Newton methods do not require a second derivative from the user [NW06], but instead the Hessian matrix of second derivatives is *approximated* in a computationally efficient manner and therefore not calculated on each iteration [NW06]. The computational requirements therefore reduce to calculating just the objective function and its gradient on each iteration of the optimiser.

In the numerical implementation used in this study we selected the L-BFGS-B adaptation by Zhu, Byrd and Nocedal, which is specifically tailored to allow for the optimisation of a large number of variables with bound constraints [ZBLN97, VGO+19].

In our discussing on key findings of this study in Section 7.1, we highlight the observation of ill-posedness and ill-conditioned circumstances in the procedures examined in this study. These observed difficulties have the potential to impact negatively on this gradient-based optimisation method.

In our evaluation we followed the procedures of Cont and Tankov as closely as possible. However, after observing some of the above-mentioned difficulties, we tested additional optimisation strategies in an attempt to improve our results. We wrapped the L-BFGS-B optimiser in a basin-hopping global optimiser. The increase in optimisation time was severe. We were forced to significantly reduce the number of points in our Lévy measure to obtain a bearable computation time. It was interesting to observe that the combination of a global optimiser and a significant reduction in the number of Lévy points led to improved convergence, not only for the regularised calibrations, but also in the un-regularised case. As shown in our discussion in Section 6.1.2, a significant reduction in the number of Lévy points will increase pricing errors. Therefore, reducing the number of Lévy points to improve convergence is not an appropriate strategy.

5.6.2 Selection of initial points

Influenced by our interpretation of the primary articles [CT04, CT06], we did not expect the selection of initial points to be critical in determining the outcome of the regularised optimisation results. However, we soon realised that optimisation results are indeed sensitive to the initial points supplied to the optimiser, as will be detailed in Section 6.2.5.

Our strategy for testing the impact of changes in the initial points follows the equivalent approach to the batch-testing approach used in testing the sensitivity to the regularisation parameter. The test method is explained in Section 6.2.3 on page 86.

5.6.3 Calculation of relative entropy

We introduced Equation (4.6) for calculating the relative entropy of equivalent risk-neural Lévy measures in Section 4.2.1. The corresponding approximation, applicable to discrete Lévy measures, can be expressed as

$$H\left(\nu^Q, \nu^P\right) = \mathcal{E}\left(\mathbb{Q} \mid \mathbb{P}\right) = \frac{T}{2\sigma^2}\left\{\sum_{j=1}^{N}\left(e^{x_j}-1\right)\left(\nu_j^Q - \nu_j^P\right)\right\}^2 + T\sum_{j=1}^{N}\left(\frac{\nu_j^Q}{\nu_j^P}\ln\frac{\nu_j^Q}{\nu_j^P} + 1 - \frac{\nu_j^Q}{\nu_j^P}\right)\nu_j^P.$$

5.6.4 Calculation of the gradient of the objective function

Cont and Tankov present an efficient method for calculating the objective function's gradient with respect to the discrete Lévy measure in the description of their calibration

procedure [CT04]. Their streamlined technique for calculating the gradient, which needs to be re-calculated at each iteration of the optimiser, significantly reduces the computation time for calibration. Our experiments suggest that their procedure offers at least a ten-fold reduction in calibration time compared to regular numerical gradient estimation. Below, we will reproduce the essence of their published derivation for the gradient formula [CT04].

We will follow the form of the time-value[3], $z_T(k)$, used by Cont and Tankov [CT03], where

$$z_T(k) = e^{-rT} E\left[\left(e^{rT+X_T} - e^k\right)^+\right] - \left(1 - e^{k-rT}\right)^+, \tag{5.3}$$

represents the time-value of the option at log-strike k for maturity T.

The second term, representing the option's intrinsic value, is not dependent on the Lévy measure. Consequently, the derivative of the time-value, $z_T(k)$, with respect to the Lévy measure is determined by the first term on the right, which is the expression for the call price, $C_T(k)$. Therefore, the derivative of the call price is the same as the derivative of the time-value.

We continue to follow the notation of Cont and Tankov [CT03] with discrete form of the Lévy-Khinchin formula,

$$\phi_T(u) = \exp\left(-\frac{u^2\sigma^2}{2} + iub(\nu) + \sum_{j=0}^{N_\nu - 1}\left(e^{iux_j} - 1\right)\nu_j\right), \tag{5.4}$$

based on (5.2), their time-value pricing formula,

$$\zeta_T(u) = \frac{e^{-rT}\phi_T(u-i) - e^{iurT}}{iu(1+iu)}, \tag{5.5}$$

and the related Fourier inversion formula,

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-iuk}\zeta_T(u)\, du.$$

The derivative of ζ_T from (5.5) is given by

$$\frac{\partial\zeta_T(u)}{\partial\nu_j} = \frac{Te^{-rT}\phi_T(u-i)}{iu(1+iu)}\left\{iu\left(1-e^{x_j}\right) + e^{x_j}\left(e^{iux_j} - 1\right)\right\},$$

which can be written as

[3] Appendix B.1.3 shows some applicable variations on the Carr-Madan pricing formulae [CM99].

$$\frac{\partial \zeta_T(u)}{\partial \nu_j} = T\left(1-e^{x_j}\right)e^{-rT}\frac{\phi_T(u-i)}{(1+iu)} + Te^{x_j}e^{iux_j}\zeta_T(u) - Te^{x_j}\zeta_T(u) + Te^{x_j}e^{iurT}\frac{e^{iux_j}-1}{iu(1+iu)}.$$

Applying a Fourier transform term-by-term leads to

$$\begin{aligned}\frac{\partial z_T(k)}{\partial \nu_j} &= T\left(1-e^{x_j}\right)e^{-rT}\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-iuk}\frac{\phi_T(u-i)}{1+iu}\,du \\ &\quad + Te^{x_j}\left\{z_T(k-x_j) + \left(1-e^{k-x_j-rT}\right)^+ - z_T(k) - \left(1-e^{k-rT}\right)^+\right\} \\ &= T\left(1-e^{x_j}\right)e^{-rT}\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-iuk}\frac{\phi_T(u-i)}{1+iu}\,du + Te^{x_j}\left\{C_T(k-x_j) - C_T(k)\right\}.\end{aligned} \tag{5.6}$$

In the discrete form, one can employ the FFT to compute the summation corresponding with the integral in (5.6).

The derivation of the gradient formula is a valuable result because it reduces the calibration time to an acceptable level, thereby allowing the opportunity for experimentation without excessive delays between results.

5.6.4.1 Gradient behaviour at small jump sizes

Observe that in Equation (5.6), the gradient for small jump sizes, where x is near zero, becomes smaller as we move closer to zero jump size. At zero jump size, where $x = 0$, the gradient is zero across all strikes. This quality implies that the objective function is not sensitive to changes in the jump intensity of the smaller jumps. In other words, from the perspective of a gradient-based optimiser, the small jump points will, in essence, stay near their initial values and will not be moved much as the low gradient value will "suggest" to the optimiser that the points are already near optimal.

This results in a "stickiness" with respect to the initial points encountered in the small jump portion of the Lévy measure. This aspect, which was also observed in our experiments, will be discussed in Section 6.2.2.

Additional discussions on the characteristics of the gradient are presented in Appendix C.3.1.

Chapter 6

Calibration tests and experiments

The primary objective of this study is to test the effectiveness of relative-entropy-based calibration procedures in addressing the ill-posedness encountered in the calibration of non-parametric jump-diffusion Lévy models. In this chapter, we set out the details and results of experiments conducted in our course of evaluating these procedures.

We commence with an abridged discussion of preliminary tests and model verification experiments, which we performed to determine suitable optimisation parameters and confirm our implementation's integrity. We will highlight significant observations, critical implementation decisions, and confirm that our implementation is correct with acceptable pricing error levels. In Appendix C, we present an extensive record of our model verification tests and results.

After discussing the preliminaries, we begin the performance assessment by conducting experiments to test the ability of the calibration procedure to converge to sensible solutions. Convergence tests form the core of our evaluation effort in this study.

We will present results showing that the non-uniqueness of solutions and sensitivity to initial points remain problematic despite generous levels of regularisation. Furthermore, we will illustrate difficulties in controlling the regularisation effect and the sensitivity of results to the properties of a prior measure.

After that, an investigation follows to test the calibration procedure's ability to recover a Lévy measure from its generated prices. We will explain that we could not retrieve meaningful results from our experiments for recovering a Lévy measure (with a known pre-defined shape) from its vanilla options prices.

In the final section of this chapter, we present the results of pricing path-dependent exotic options by simulation.

6.1 Preliminary tests and model verification

We performed verification tests to:

- Determine the appropriate configuration for a reliable transform pricing engine in terms of

  - pricing formula and
  - parameter selection.

- Determine the typical pricing error with the above configuration with respect to reference prices.
- Determine suitable parameters for the non-parametric representation of Lévy measures to achieve acceptable pricing relative to parametric equivalents.
- Verify the accurate operation of the gradient function for the objective.
- Verify proper operation of relative entropy.

Our results indicate an effective implementation with satisfactory pricing abilities. Details and results of these verification test are available in Appendix C.

6.1.1 Pricing mechanism

This section will highlight the essential transform pricing parameters and accuracy figures. A summary is presented in Table 6.1.

Our tests suggest that adequate pricing accuracy is achievable when using the *Black-Scholes adjusted pricing formula*, discussed in Section 3.2.5. Test details are discussed in Appendix C.2, where we show that with 2^{12} FFT grid points, one can achieve pricing errors[1] of less than $10^{-4}\%$ for Merton model call prices over a normalised strike range of $-0.6S_0$ to $1.4S_0$. Over the same range we observed a maximum absolute pricing error in the order of 10^{-7}. For 160 strike prices at the maximum error, this would limit the contribution from the pricing error to the sum of errors squared (SSE) to a maximum of 4.7×10^{-12}. This limit is well below the error levels that we will encounter in our calibration results later in this chapter. In other words, the contribution of inherent pricing errors to the error term in the objective function is insignificant and will, therefore, not impact on results.

Further details of our pricing tests are available in Appendix C.2. Pricing errors versus reference prices from various models across a range of normalised strikes can be seen in Figure C.2. Table C.1 shows the ranges of feasibility for various transform pricing parameters.

6.1.2 Non-parametric pricing errors

This section addresses the parameter selection for realistic pricing with a discrete Lévy measure. The last two entries in Table 6.1 show the values of interest.

We performed tests to determine suitable parameters for the non-parametric representation of Lévy measures to achieve pricing compatible with their continuous counterparts.

[1] We explain the presentation format for our measured pricing errors in Appendix C.2.1.

Parameter	**Value**	**Notes**
Best pricing formula	Black-Scholes adjusted formula	
FFT grid size	2^{12}	
FFT input grid spacing	0.1	
Maximum call pricing error (%)	$< 10^{-4}\%$	Tested to Merton model
Maximum absolute pricing error	$< 1.7 \times 10^{-7}$	(Prices normalised to $S_0 = 1$)
Worst case pricing error contribution to SSE	4.7×10^{-12}	For 160 strike prices at maximum error
Discrete Lévy measure number of jump sizes	2^{10}	
Discrete Lévy measure range of jump sizes	-2 to 2	

Table 6.1: Summary of significant pricing parameters.

Our results suggest that one can comfortably maintain a maximum absolute pricing error of less than 10^{-7} by selecting 2^{10} jump sizes, covering a jump size span of -2 to 2. In other words, errors introduced by the discretisation into the non-parametric calibration process are of the same order of magnitude as the inherent pricing errors. Therefore they will not negatively influence our calibration results.

Our non-parametric versus parametric pricing tests are available in Appendix C.2.4. Table C.3 shows pricing errors for a few combinations of the number of jump sizes and the range of jump sizes. Figure C.3 illustrates absolute pricing error over the range of normalised strike prices.

6.1.3 Gradient verification

We verified the gradient calculation discussed in Section 5.6.4 by applying a finite difference numerical approximation. Results are available in Appendix C.3.

In our tests, we observed that the gradient of the error term becomes small across all strikes as the jump size approaches zero. This aspect was mentioned in Section 5.6.4.1 when we discussed the gradient behaviour for small jumps. In our experiments, we observed this quality as a "stickiness" with respect to the initial points in the small jump portion of the Lévy measure. These observations will be discussed in Section 6.2.2.

6.1.4 Relative entropy verification

In the following section, where we investigate the convergence of the optimisation process, we will illustrate the performance of the relative entropy in isolation. The details of this experiment is presented in Section 6.2.2.3.

Additional verification tests for the operation of the relative entropy penalty function are presented in Appendix C.4.

In our tests, we observed that relative entropy in the discrete form appears to be slightly non-convex with small valleys (local-minima). The valleys tend to become shallower as the number of discrete points are increased, but the number of valleys seem to increase with the number of discrete points. Although most optimisation tests were unaffected by the local minima, some instances failed to converge. This aspect is discussed in Appendix C.4.3.

6.2 Convergence tests

In this section, we examine the robustness of the regularised calibration procedures by performing tests to assess the ability to neutralise the effects of ill-posedness. This examination forms a fundamental component of the research performed in this study.

Earlier, in Section on page 52, we discussed the issue of ill-posedness associated with non-parametric calibrations. This section presents the results of our detailed analysis of this important aspect. Later, in Chapter 7, we will compare the results of our

experiments performed in this section with specific statements in relevant literature about the abatement of ill-posedness when applying relative entropy to this type of calibration problem.

Next, we briefly discuss the key considerations for assessing the calibration results. Before observing the optimisation results at balanced regularisation levels, we demonstrate the undesirable properties at the extremes of regularisation.

These preliminary special case tests are then followed by a description of our experiments to test the convergence of regularised calibration under more feasible test conditions and then a summary of the resulting findings.

6.2.1 Considerations in the assessment of convergence results

The primary objective of the regularised calibration process is the retrieval of a Lévy measure that reflects the key characteristics of the jumps in the underlying, which is implied by the options prices. To achieve this, we need to strike a balance between a "perfect" fit to the (noisy) data and the absolute smoothness of our prior measure. The non-parametric calibration without regularisation generally leads to overfitting and extreme sensitivity to small changes in prices or sensitivity to initial points supplied to the optimiser. At the other extreme, optimising with an excessive amount of regularisation will cause the optimiser to converge to the prior measure, regardless of changes in the market data or the optimiser's initialisation points. Examples of calibrations at these extremes are shown in the next section. Striking a balance between these extremes may seem like a trivial exercise, but our observations suggest that simply terminating the search for valid parameters when a pleasing optimisation result is delivered would be ill-advised. Therefore, we extend our frame of assessment to include properties spanning across results from multiple optimisations and assess results using a number of criteria.

Ideally, we would like to see evidence of the following properties in the results of our non-parametric convergence tests:

1. An adequate level of control to balance calibration precision with the smoothing of regularisation to allow:
   (a) A user-selectable degree of smoothing, preferably with a fair degree of monotonicity. In other words, increasing the level of regularisation should ideally increase the smoothness.
   (b) Stability of the shape characteristics, avoiding extraordinary changes in optimal jump distributions for a small adjustment in regularisation weight.
2. Relief from ill-posedness by removing or reducing the high sensitivity of results to initial points and small changes in market data. In principle, we desire stability without extreme bias towards the prior, which could overpower the information contained in the market data.
3. Given suitable levels of regularisation, consistent convergence to a unique solution or possibly a set of solutions that are much alike.

4. A reasonably autonomous process capable of generating acceptable results without the need for considerable user intervention to compensate for low-grade results.

In essence, the ideal outcome amounts to balancing the goodness of fit and the smoothness of the retrieved Lévy measure while being able to extract consistent results largely independent of small price changes or changes in optimiser initialisation. When one obtains a reasonable degree of consistency in the delivered Lévy measure, it provides the opportunity to visualise the jump characteristics and to apply these stable results to generate price path simulations for pricing path-dependent options.

In the following sections, results will be assessed in terms of the above-mentioned characteristics. In some cases, the assessment may involve a degree of judgement, but we aim to present suitable examples to illustrate clear-cut cases allowing fair evidence-based evaluation.

In some cases, the assessment may involve a degree of subjective judgement. However, we aim to follow a fair evidence-based evaluation focused on showing clear-cut examples in support of our findings.

6.2.2 Calibration with extremes of regularisation

In this section, we examine non-parametric calibration at the two extremes of minimal and maximal regularisation. We firstly show some calibration results performed without regularisation while varying the initial points used to seed the optimisation process. We will also point out that the calibration process tends to leave the Lévy points representing the small-jump intensities close to their initial values.

Thereafter, we confirm that "optimisation" with regularisation only (disregarding the contribution of market data) will generally converge to the prior measure and is almost insensitive to the variation in initial points.

These "calibrations" were performed using the standard algorithms that will be used for convergence tests in the later sections of this chapter. Therefore, these tests serve as illustrative examples at the extremes and basic verification tests of algorithm functionality at the extremes.

The calibrations illustrated here were performed on artificially generated market prices based on the Merton model with discrete Lévy measures shown as dotted lines in Figure 6.1. In other words, there is a (known) perfect solution in the Lévy points labelled "Data".

6.2.2.1 Calibration without regularisation

The optimised discrete Lévy measures of the non-parametric calibration without regularisation for different initialisation points are shown in Figure 6.1, and the corresponding implied volatility curves are shown in Figure 6.2. As expected, without regularisation, the optimisation results differ significantly as the initial points are changed. However, judging by the goodness of fit in the implied volatility curves, they all present

a close fit to the source data. The closeness of the fit is also reflected in the mean squared error (MSE) of less than 6.0×10^{-9} for all three examples. This is an unusually low error when compared to the regularised examples shown in the next section, where we sacrifice some precision for smoothness.

Figure 6.1 shows the discrete Lévy measures from calibration without regularisation. The corresponding implied volatility curves are shown in Figure 6.2.

6.2.2.2 Insensitivity of small-jump intensities to pricing data

Note from the results shown in Figure 6.1 how the jump intensities of the small jumps (near zero) "stick" to their initial values. The low gradient values for small jumps mentioned earlier in Section 5.6.4.1 implies that the objective function is insensitive to changes in the intensity levels of small jumps. The contribution of small-jump intensities becomes insignificant as the jump sizes become smaller. In essence, the central parts of the discrete Lévy measure representing the small-jump intensities become somewhat superfluous in determining the behaviour in jump-diffusion models.

Moreover, the ambiguity between the dynamics of the diffusion component and the tiny jumps can further explain the unimportance of the intensity levels in the small-jump region. It is interesting to note that although the Kou and Merton models have entirely different shapes in their Lévy measures for the small-jump areas, the significance of changes in that central area is insubstantial.

In our experience from evaluating the quality of several optimisation results for Lévy measures, there seems to be a natural observer bias towards assigning too much importance to the shape of the central portions, which represent an area of lower significance. It is natural for an observer to prefer something like the aesthetically pleasing shape of a normal distribution, with its prominent peak, rather than a peculiar uneven shape. It requires a certain effort to override one's natural preference for order when evaluating the quality of the output of the calibration process.

We will return to the insignificance of the small-jump intensities when we take a closer look at our criticism of some arguments from the original articles presented in Section 7.3.

The results shown in Figure 6.1 clearly illustrate the calibration problem characterised by ill-posedness, with results exhibiting extreme sensitivity to initial points, particularly in the central regions of the retrieved discrete Lévy measures. These results also illustrate the worst-case results, providing a reference level which we aim to improve by applying the regularisation techniques.

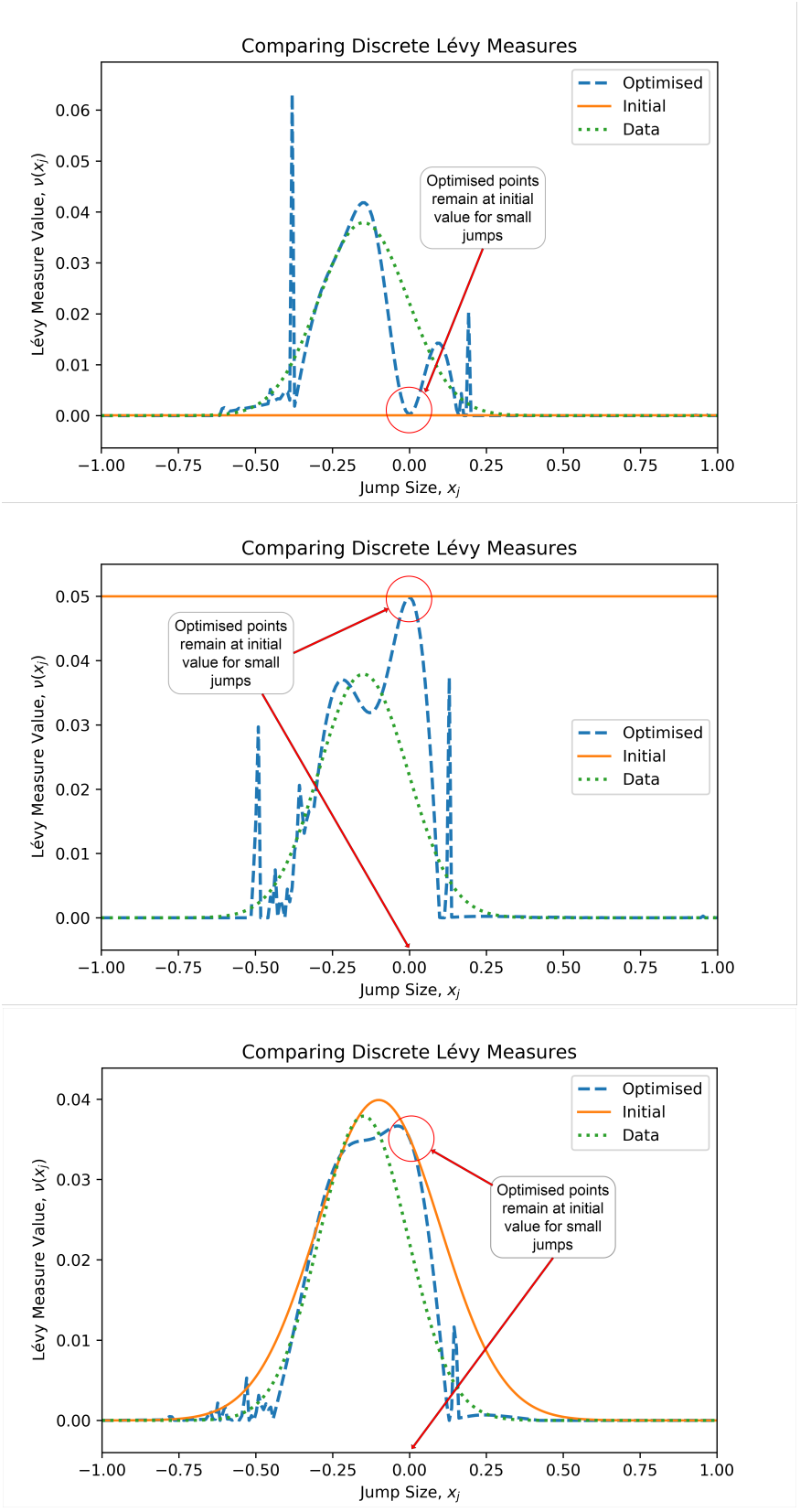

Figure 6.1: Non-parametric calibration results without regularisation.

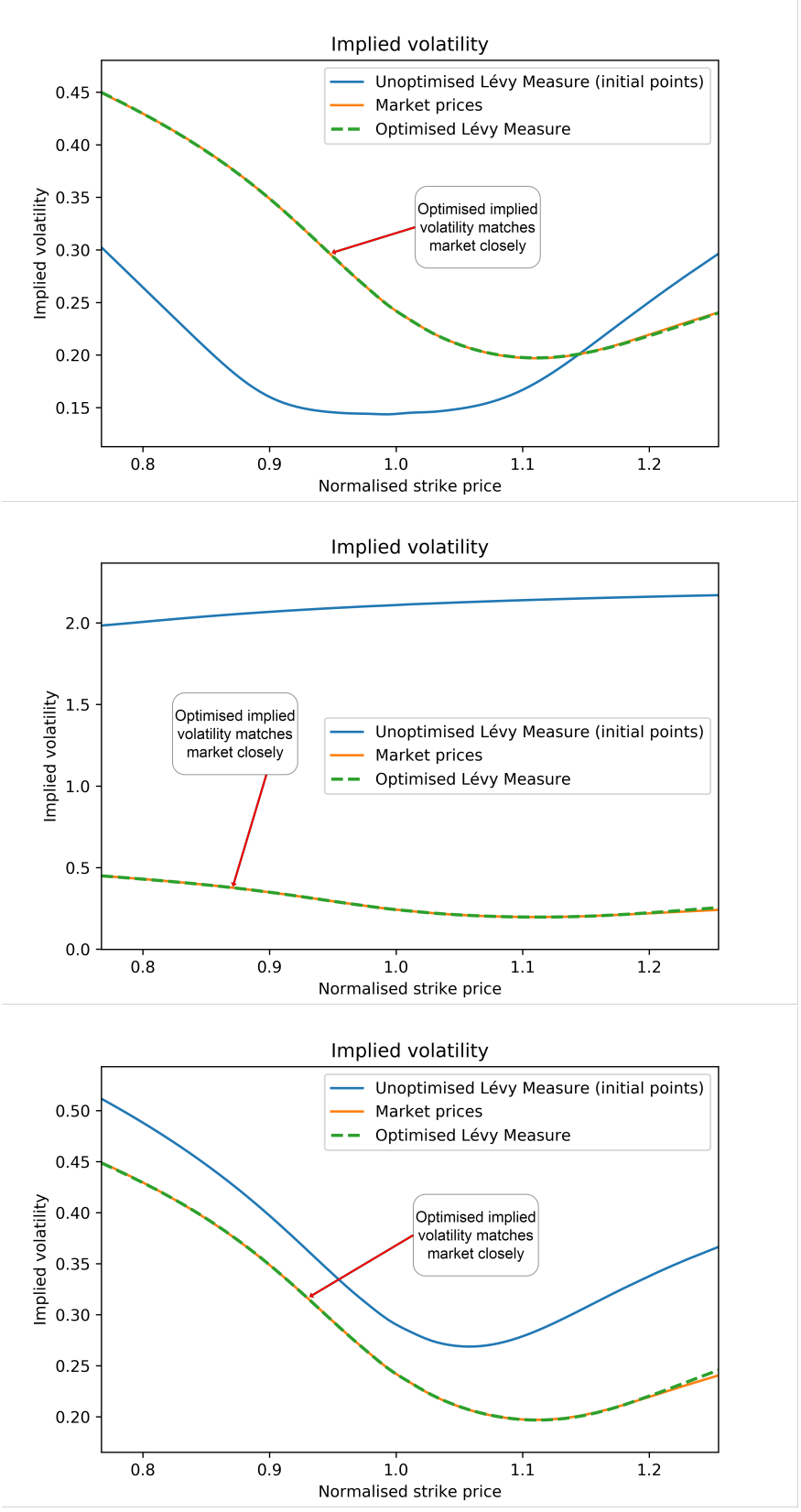

Figure 6.2: Implied volatility curves corresponding to the optimisation results without regularisation that were shown in Figure 6.1.

6.2.2.3 Calibration with relative entropy only

At the other extreme of the optimisation level, we tested the ability of the optimiser to converge to the prior discrete Lévy measure given an objective function with relative entropy only. We disabled the least squares component in the objective function along with its gradient and selected the generating measure (of the artificial Merton prices) as the prior. In other words, we allowed the optimiser to work without visibility to market pricing errors in the objective function. We provided the perfect prior measure with only the relative entropy component in the objective and its gradient. Figure 6.3 shows an example of the optimal discrete Lévy points of optimisation based solely on relative entropy.

After termination of the optimisation, we computed the final relative entropy and corresponding pricing error of the "optimal" Lévy points. This procedure was repeated for various starting points. Table 6.2 summarises the results of more than 20 optimisation tests where the optimiser returned a valid result. We can see from these results that for the "perfect prior", the optimisation results indicate an optimised relative entropy in the order of 10^{-3}, which corresponds to an MSE pricing error in the order of 10^{-7} for the optimal measure. This suggests that in the region close to the optimal discrete Lévy points, a regularisation constant in the order of 10^{-4} will provide a reasonable balance between the components of the regularisation penalty and the pricing error.

Our tests suggest that the relative entropy penalty term and its gradient function, when tested in isolation, perform well at steering the optimiser towards the prior measure[2]. In the above-mentioned experiment, where the prior coincides with the optimal solution, the optimiser achieved a reasonably close fit when judged by the mean pricing error of 10^{-7} obtained with the optimal points[3].

6.2.3 Calibration convergence test method

Several parameters influence the non-parametric calibration outcome, so it is unrealistic to expect a reasonable assessment by looking at parameters in isolation. Hence, in our non-parametric calibration convergence tests, we followed a manual "grid-search" strategy.

We would start by choosing a sequence of test values for each parameter of interest. Then, we would select two or three parameters to test together as a test group. We would then systematically step through the combinations of the selected parameter sets while running the calibration procedure for each combination.

[2] Approximately 15% of the tests over various initial conditions failed to deliver a successful optimisation outcome for the relative-entropy-only test, suggesting that there is an inconsistency with regard to the operation of the optimiser, the relative entropy and the gradient function for the relative entropy. A discussion on this observation can be found in Appendix C.4.2 on page 161.

[3] It is possible to achieve a closer fit by adjusting optimiser tolerances on the termination conditions with respect to the objective and gradient, but in this test we opted to use identical optimiser settings to those used in our convergence tests with realistic regularisation values in the following sections.

Parameter	Minimum	Maximum	Mean
Relative entropy value	3×10^{-4}	7×10^{-3}	2×10^{-3}
Pricing error (MSE)	7×10^{-10}	4×10^{-7}	1×10^{-7}

Table 6.2: Results from successful optimisation tests for convergence with relative entropy only.

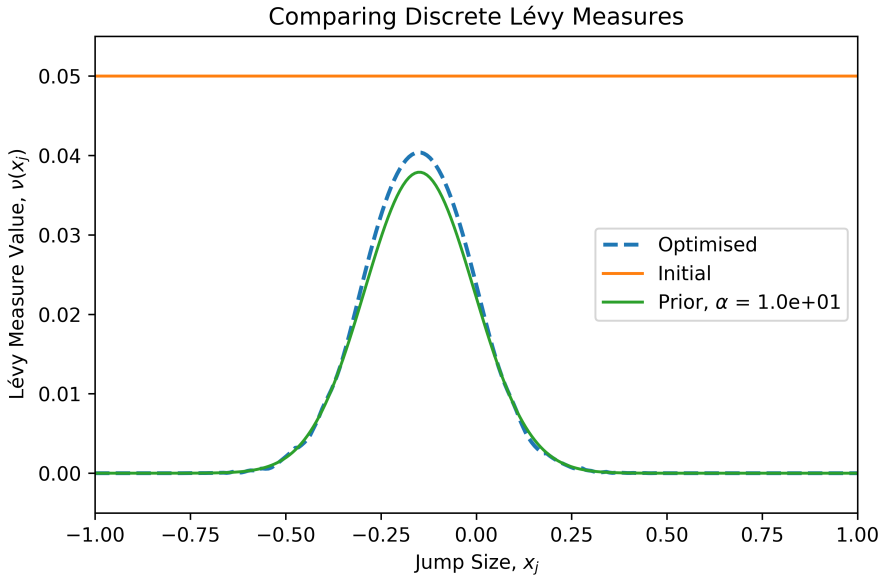

Figure 6.3: Optimisation test with using the relative entropy term only.

After observing inconsistent results in our initial convergence tests and experiencing difficulty isolating the cause, we adopted this meticulous strategy. In addition, it serves as a more thorough approach for analysing the behaviour of regularised calibration in terms of convergence.

Virtually all the test groups were based on the test values for two primary sets together with a test value set from one other secondary parameter.

The two primary parameter sets of interest are:

1. Sets of initial discrete Lévy point vectors to test optimisation performance from different starting points of the optimiser.
2. Sets of values for the regularisation constant, α, in our objective function as shown in Equation (5.1) on page 68 to select the balance between regularisation and pricing precision.

Examples of secondary parameters are:

1. The market data set, which also includes artificially generated prices.
2. The choice of prior measure.
3. The selected method of pricing-error measurement in the objective function.
4. Calibration precision and other optimiser-specific control parameters.
5. Pricing parameters and Lévy measure discretisation parameters (A subset of the relevant parameters are mentioned in Section 6.1 with further details available in Appendix C.2)

The parameter sets and other test configuration details for a group of tests were mostly defined in an optimisation test script. The script would systematically walk through the required parameter sets and perform each optimisation instance with its distinctive parameters, capturing results of interest. Each optimisation instance would yield an optimisation outcome with its own numerical results and graphs.

A combination of text files, graphs and packaged variables would be generated and stored for post-optimisation analysis of each optimisation instance. In addition, key numerical results were collated and written as a summary file to allow inspection of the entire set in a spreadsheet, with each row representing an optimisation outcome. Typical summary results include:

- Objective function values (initial and final)
- Pricing error as MSE and RMSE (initial and final)
- Relative entropy component (initial and final)
- Optimiser termination reason, number of iterations and optimisation time.

Several graphs are typically generated for each optimisation instance. The most important graphs are used to show the change from the initial to the final (optimised) state for elements such as:

- Discrete Lévy measures (some showing reference Lévy measures for source prices and prior measures).
- Gradients with a breakdown of gradient components between pricing error and relative entropy.
- Pricing errors or implied volatility against normalised strike prices.

Inspecting the graphs and numerical results of an optimisation outcome in isolation has value in understanding aspects of that particular result. However, we found it difficult to assess the performance of the regularised optimisation procedure based on analysing individual results in isolation. Instead, we found it beneficial to study the results using a comparative approach.

The test script was designed to collate the individual numerical results of each optimisation in the group into a summary file for further analysis in a spreadsheet. When viewed in a spreadsheet, comparing and interpreting results from different optimisation instances becomes more manageable. We found this approach useful for comparing the relative magnitudes of key results, obtaining summary statistics, and identifying outliers.

Similarly, the optimisation results presented as graphs are also more easily assessed when compared side-by-side to the corresponding graph from another optimisation instance. Presenting the graphs from an optimisation group in the form of a matrix of images has proven to be an effective approach for gaining insight into the impact of parameter changes on the results.

The graphs in the matrix can then be arranged for one parameter to change horizontally and for a second parameter to change vertically. An example is shown in Figure E.1. Typically, we arranged our image matrix so that each column would correspond to a particular initial point vector, and each row would correspond to a particular regularisation constant, with the magnitude increasing from top to bottom.

The matrix view can be created for any of the graphs delivered by the optimisation procedure. This allows one to gain insight into the performance of that particular aspect by comparing the graphs in the group as the parameters are varied.

6.2.4 Examining the control of smoothness versus precision

Our results for the control of the regularisation in terms of balancing smoothness and precision specifically looked for evidence of the following:

- Monotonicity of the regularisation. The results should exhibit an increase in smoothness with an increase in regularisation.

- The stability of shape characteristics. Optimisation results should exhibit a fair degree of insensitivity for slight variations in regularisation.

Our results indicate the following:

- Although there is evidence of general conformance to monotonicity, the monotonicity of the regularisation is not guaranteed.
  - Examples of a deterioration in the smoothness (or convergence) with increased regularisation can be seen in Figure 6.4 and Figure 6.5, compared to an example showing more consistent behaviour in Figure 6.6.
- The shape characteristics of regularised optimisation results are sometimes unpredictable, and stability is not definite.
  - Examples of major shape changes for a small change in regularisation are shown along with their implied volatility in Figure 6.7.
    Judging by the implied volatility graphs in this figure, this example also serves as another illustration of the above-mentioned monotonicity issue.

6.2.5 Examining the sensitivity to optimiser initialisation points

The regularisation techniques examined in this study aim to reduce the *ill-posedness* associated with retrieving non-parametric Lévy measures from options prices. Testing the functionality of this reduction of ill-posedness resulting from the relative-entropy-based regularised calibration is, therefore, a fundamental aspect of the research performed in this study.

We performed extensive testing on the sensitivity of the retrieved discrete Lévy measures to changes in the initial-point vector used to seed the optimiser with a discrete Lévy measure. Figure 6.8 on page 96 demonstrates how widely results can vary when the initial points are varied, but the regularisation and the prior remain unchanged.

We showed in Section 6.2.2 on page 82 that in the absence of regularisation, the retrieved discrete Lévy measure could vary significantly when the optimiser is presented with different initial points. We also demonstrated that in the cases of extremely high levels of regularisation, there is no sensitivity to the initial points and that, in general, we observed convergence to the prior measure. One can therefore argue that given sufficient regularisation, the ill-posedness can be entirely eliminated.

Results from our experiments based on stepping through sets of initial points with varying levels of regularisation confirm that for very high levels of regularisation, the resulting discrete Lévy measure can be forced to converge to a shape closely resembling the prior measure.

Our findings suggest that the sensitivity to initial points presents an issue that requires some user intervention to overcome. It is not advisable to merely rely on the optimisation process to automatically yield high-quality results. In our experience, obtaining

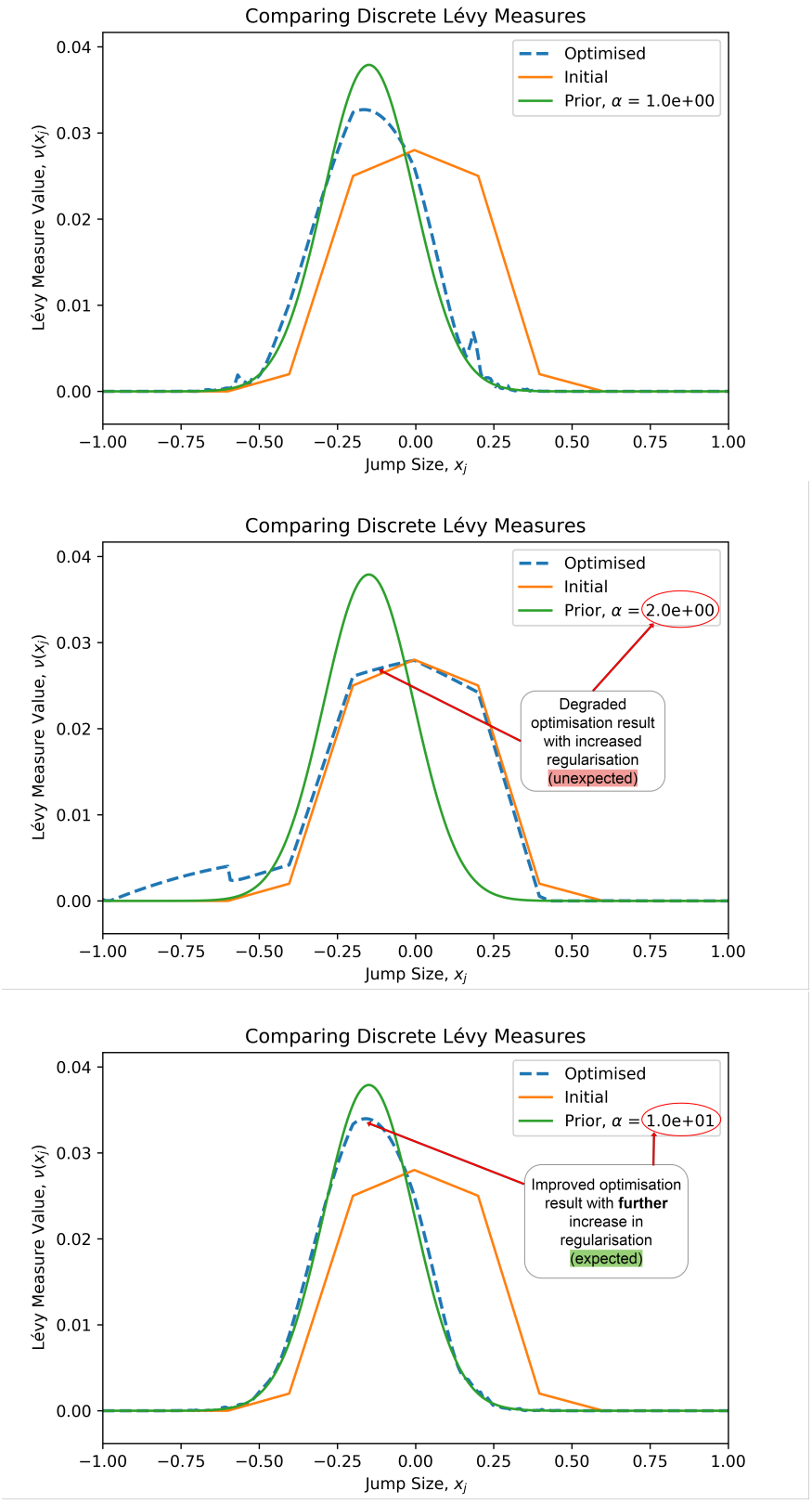

Figure 6.4: Example showing that increasing the regularisation (top to bottom) does not always improve convergence (true measure coincides with prior).

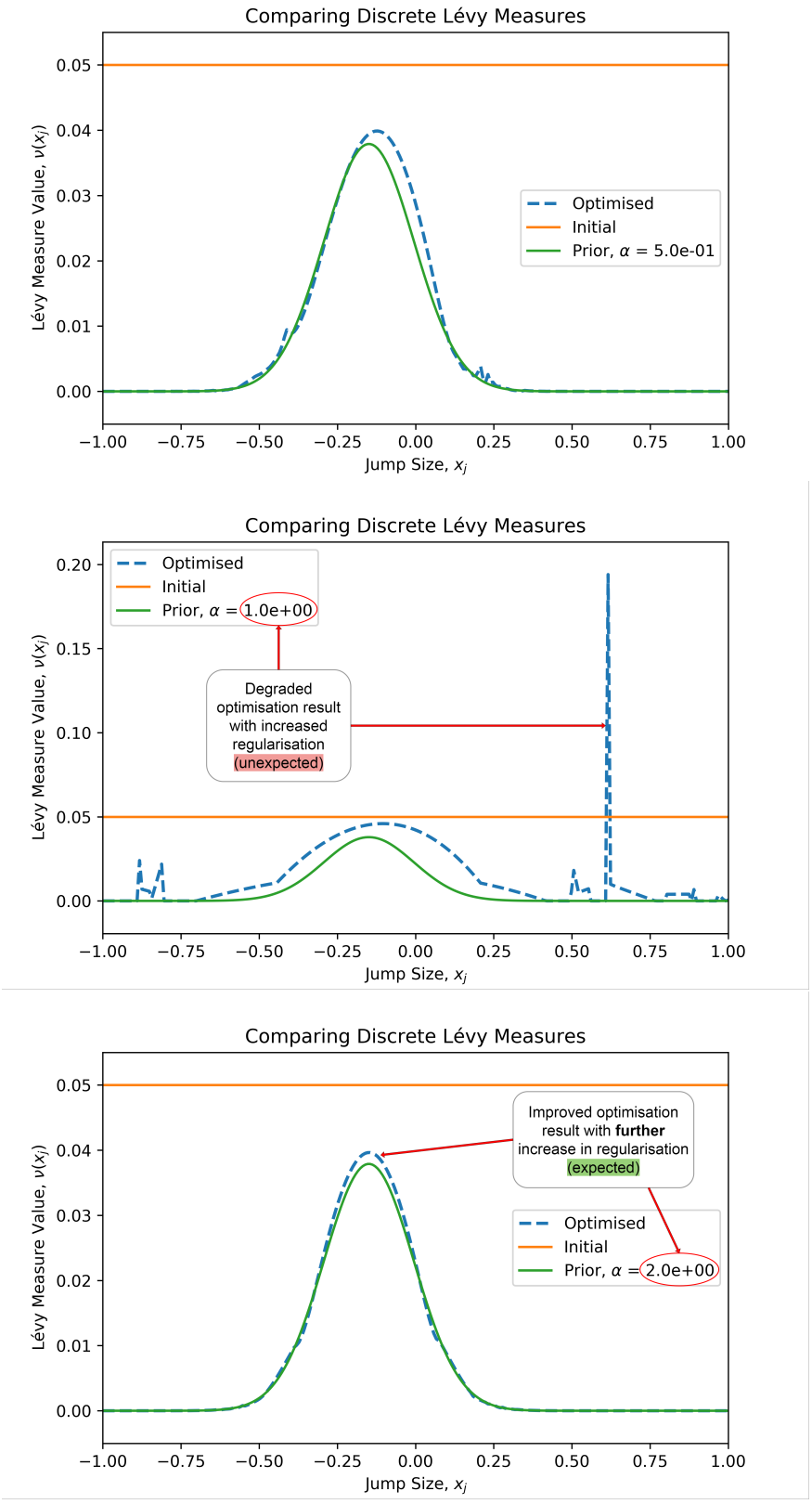

Figure 6.5: Example showing that increasing the regularisation (top to bottom) does not always improve convergence (true measure coincides with prior).

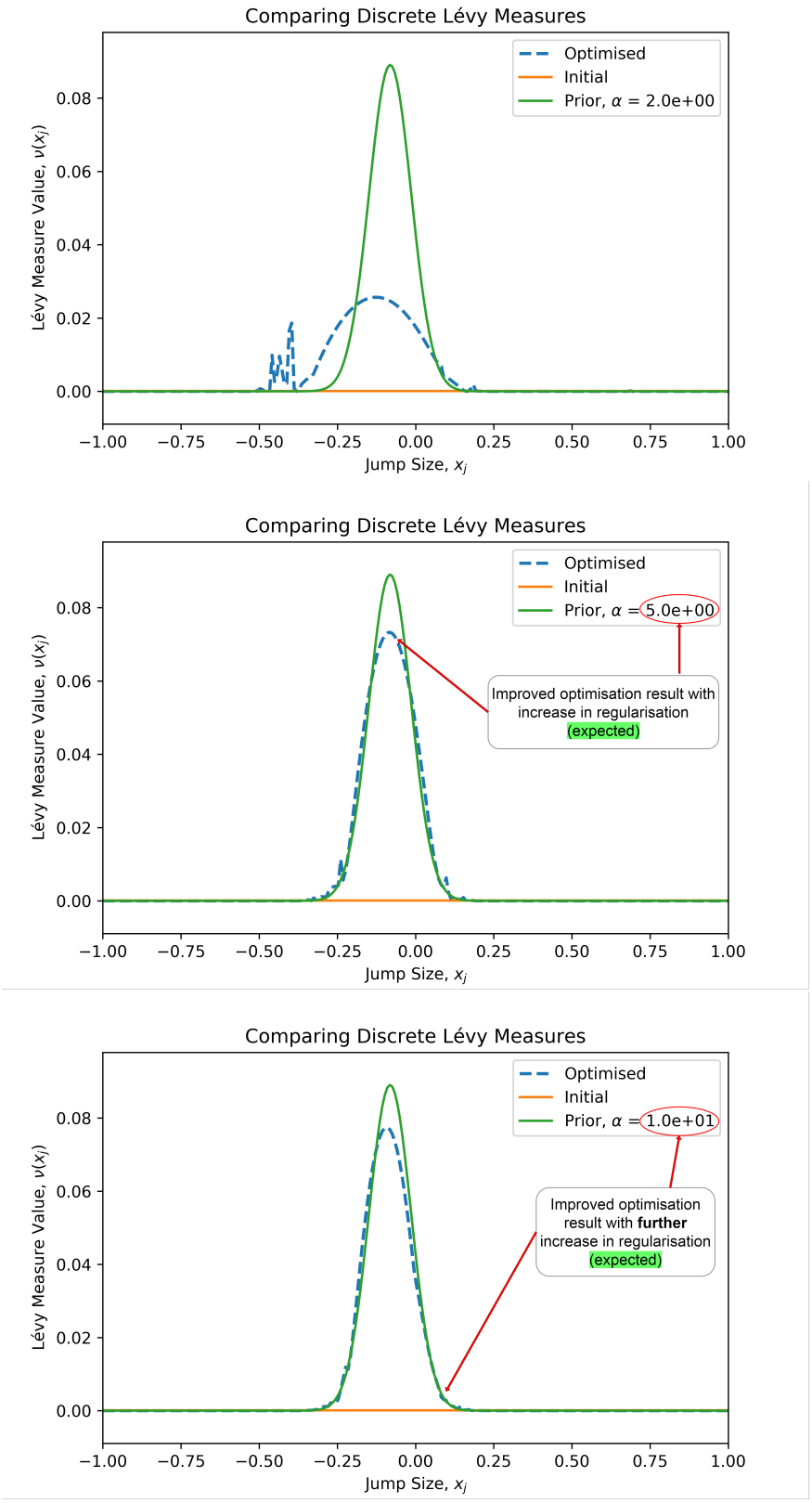

Figure 6.6: A typical example showing that increasing the regularisation (top to bottom) will in general improve convergence (real data, no known true measure).

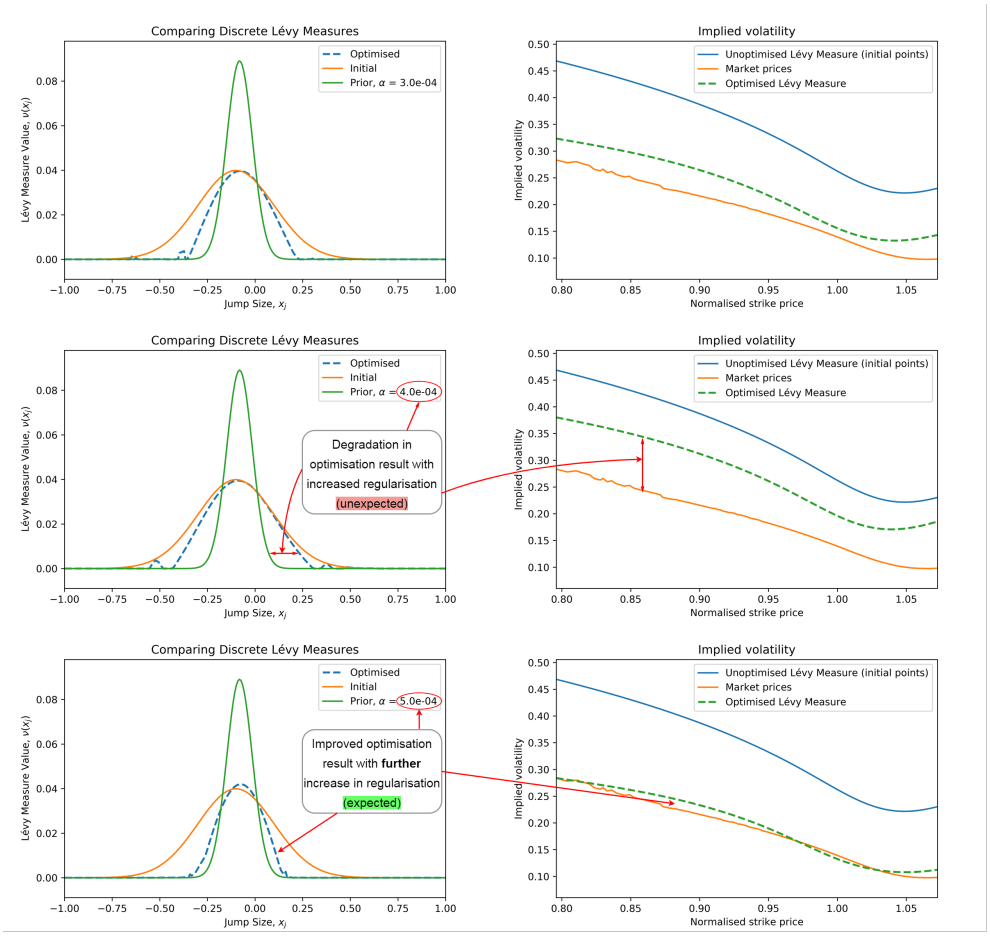

Figure 6.7: Example showing that shape stability with small increases in regularisation (top to bottom) does not always hold (real data, no known true measure).

the best balance between information contained in the data and smoothing requires an empirical approach.

It is advisable for the user to perform some trial-and-error in selecting initial points and to select the regularisation factor in conjunction with the initial point vector.

Our observations show that choosing the initial points close to the prior requires a smaller regularisation factor. This reduction in the regularisation is required to allow the optimiser to explore regions representative of the data and not become trapped by the influence of the prior. Without a reduction in the level of regularisation, the proximity of the initial points to the prior tends to prevent the optimiser from moving the solution away.

However, our experience suggests that the reduced regularisation levels required to explore the surrounding areas when starting too close to the prior tend to result in noisy optimal Lévy points with a spiky appearance.

In other words, when the initial points are chosen too close to the prior, we are presented with two undesirable options:

1. The optimiser seems to get "trapped" at the initial points unless the regularisation is reduced.
2. The much-reduced levels of regularisation required to break free from this holding position then result in under-regularisation with noisy results.

There does not seem to be an acceptable window for adjusting the regularisation to strike a good balance in cases where the initial points are too close to the prior. Figure 6.9 on page 97 shows optimisation results at different levels of regularisation for initial points coinciding with the prior measure. The regularisation levels in this figure are decreasing from top to bottom.

Starting the optimiser further from the prior with no particular proximity bias towards the prior and just relying on the gravitation towards the prior provided by the regularisation has the capacity to yield more balanced results. This, however, generally requires the manual manipulation of the regularisation factor for a given starting point vector.

In other words, the regularisation level needs to be adjusted to be "compatible" with particular initial points. There is no universal regularisation factor that will guarantee good results from any initial point vector for a given prior. Figure 6.10 on page 98 illustrates that reasonable levels of convergence can be achieved with vastly different levels of regularisation.

6.2.6 Examining the influence of secondary parameters

The group tests performed in the two preceding subsections were repeated while stepping through a series of values for each secondary parameter to observe the impact on the convergence and smoothness control tests. We summarise the influence of the secondary parameters on the optimisation outcome in the paragraphs below.

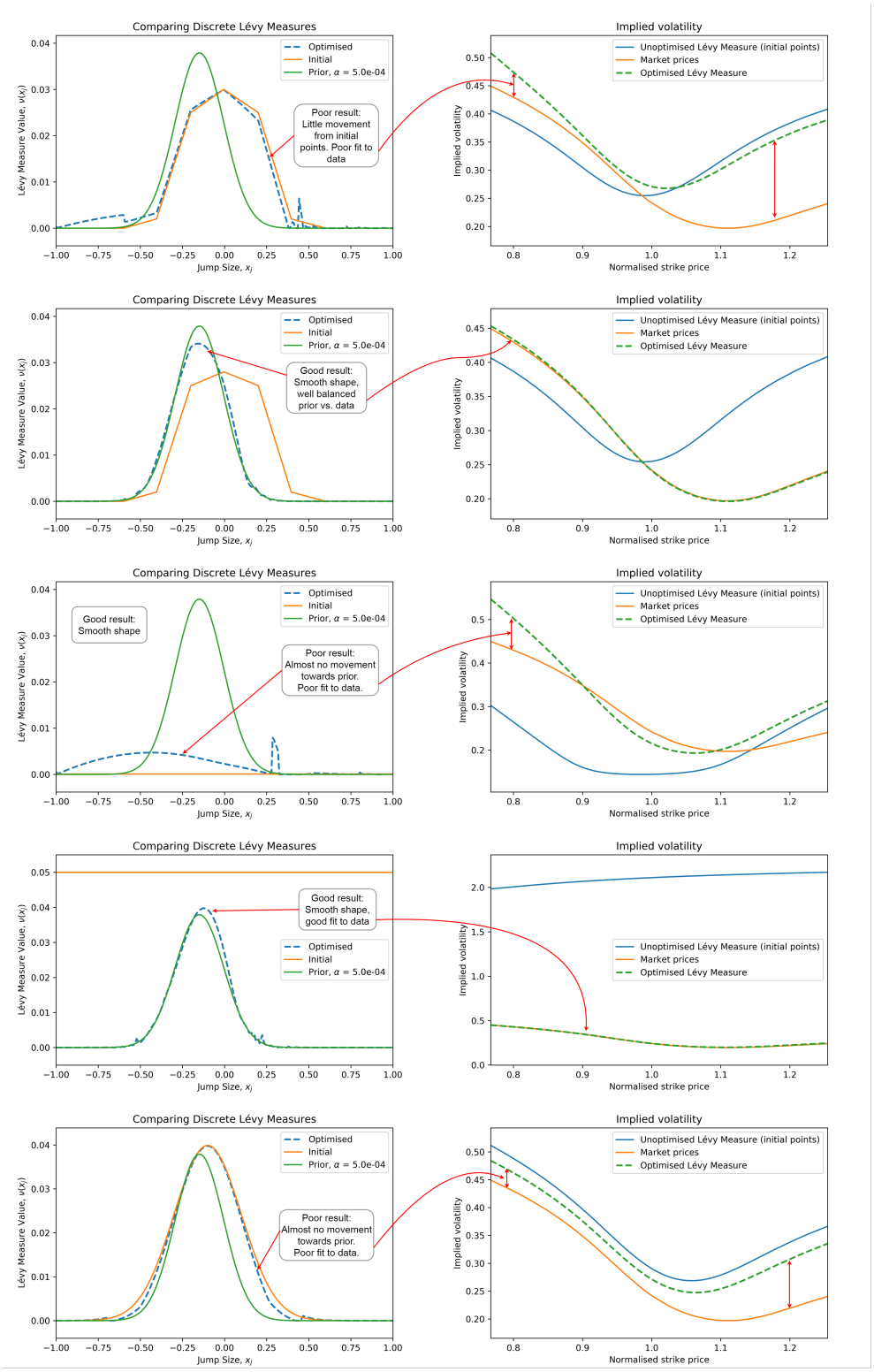

Figure 6.8: Results from various initial points (with the same regularisation factor).

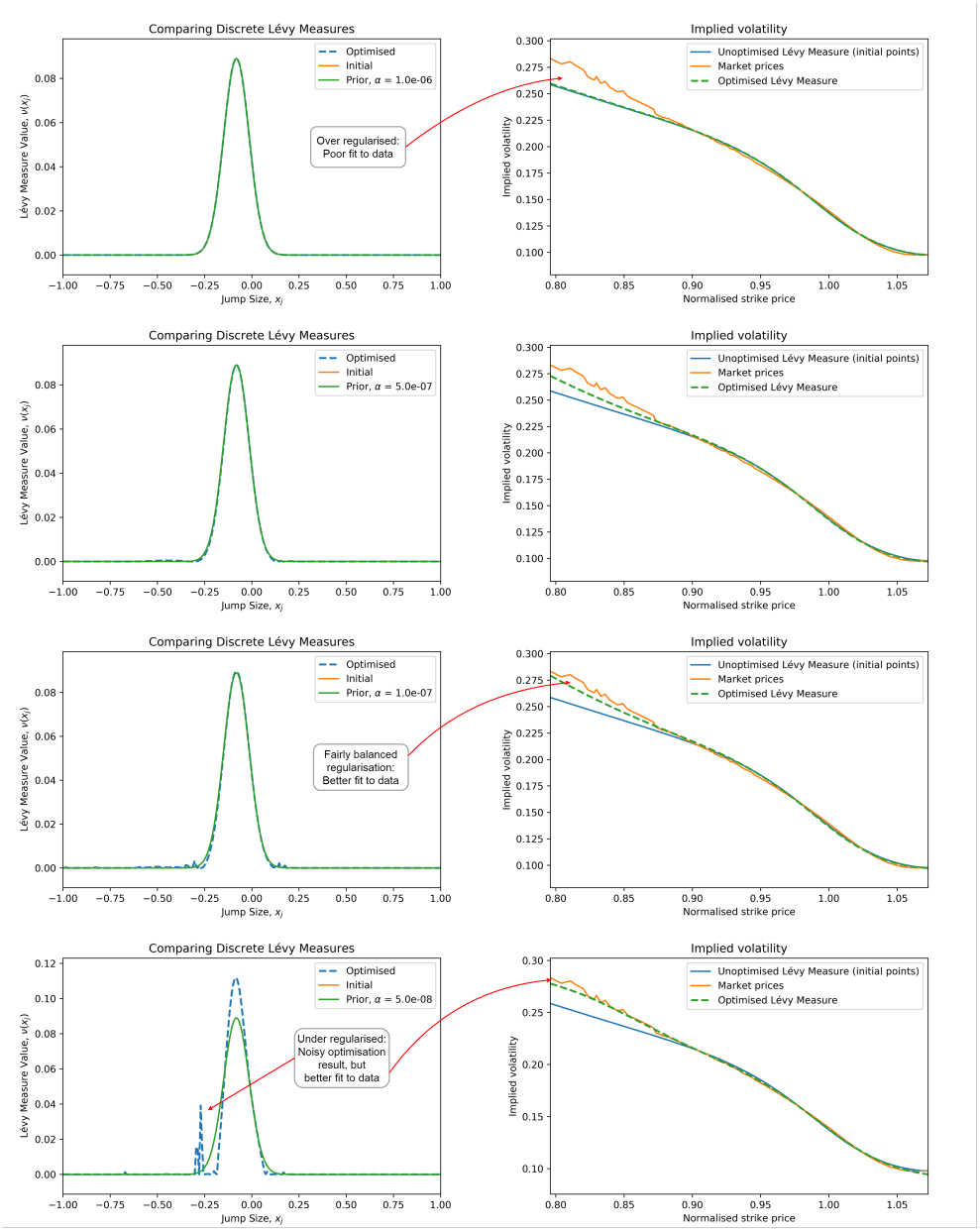

Figure 6.9: Initial points coinciding with the prior and regularisation decreasing top to bottom (real data, no known true measure).

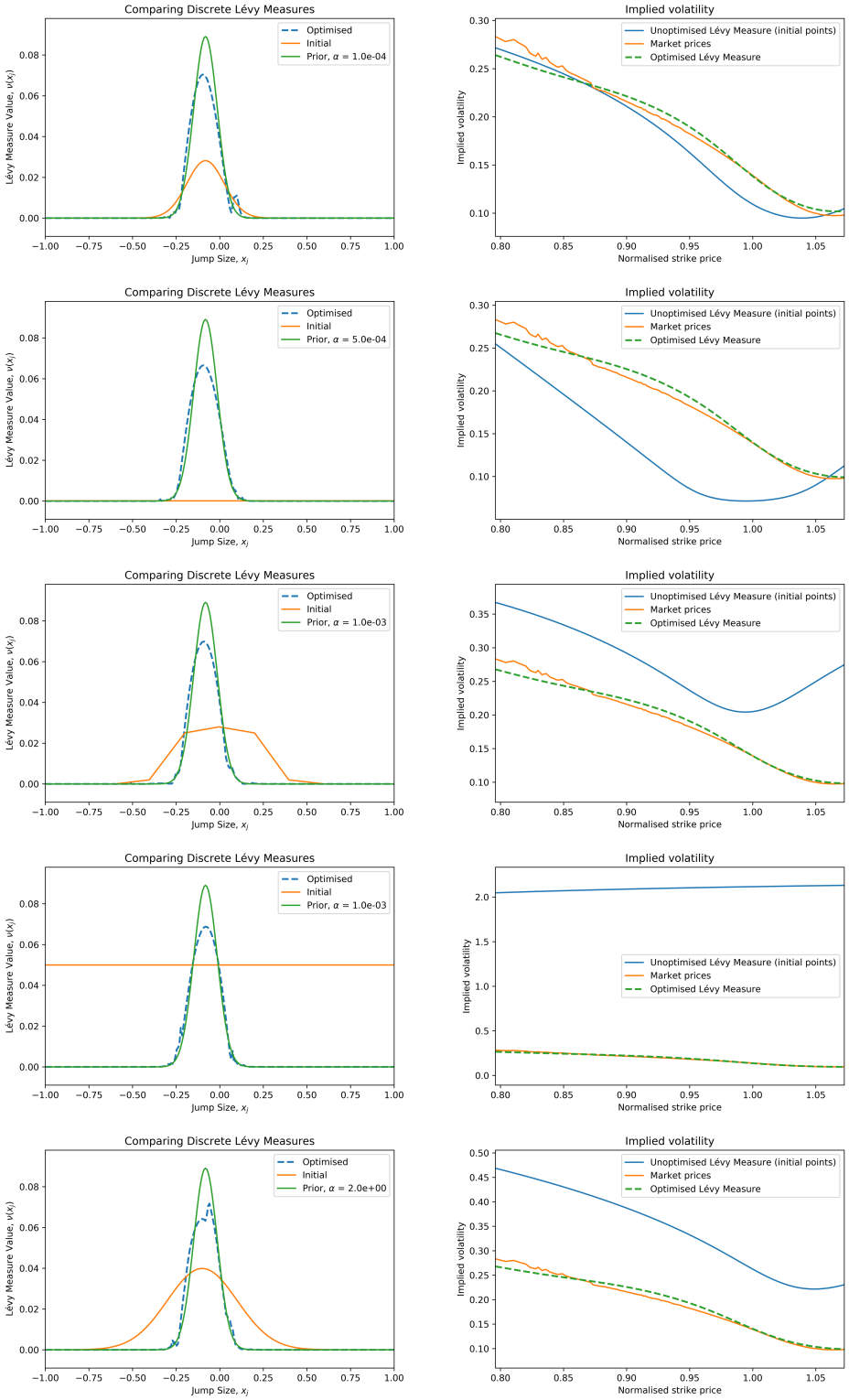

Figure 6.10: Reasonable convergence with diverse levels of regularisation (real data, no known true measure).

6.2.6.1 Data set variation

In general, changing the data set did not significantly alter the overall quality of the optimisation results. As expected, optimisation results from different data sets were by no means identical, but similar behavioural patterns were observed. In fact, the graphs selected to illustrate our findings in the preceding section were not selected from results based on a single data set. Calibration tests were performed using several data sets in the course of this study. We used various sets of market options prices based on the S&P 500 Index[4]. In addition, we generated our own artificial prices from known models and non-parametric models with known Lévy measures. In Section 6.3, we will discuss experiments where we specifically attempted to recover a known Lévy measure from its (artificial) options prices.

6.2.6.2 Sensitivity to small price changes

Cont and Tankov claim that their regularised calibration procedure is able to address ill-posedness and specifically emphasise the stability of their results in the presence of small price fluctuations between calibrations [CT04, CT06]. We will now discuss the results of our experiments to test the sensitivity of the calibration result to small price fluctuations.

Ideally, one would like to see a reasonable level of stability when applying regularised calibration in the presence of the price fluctuations observed in the market. Large changes in the characteristics of the retrieved discrete Lévy measure in response to small price fluctuations would be undesirable and would imply inadequate relief from ill-posedness.

We followed the following procedure to assess the sensitivity to small price fluctuations:

1. Obtain a discrete Lévy measure by parametric calibration of market data to the Merton model followed by discretisation.
2. Generate "artificial" market prices at a selection of strike prices using the reference discrete Lévy measure obtained in the preceding step.
3. Extract the implied volatility at each price.
4. Repeat for each inspection of sensitivity to price fluctuations:
   (a) Add Gaussian noise to the implied volatility vector.
   (b) Generate noisy artificial prices using the resulting noisy implied volatility.
   (c) Run an optimisation group test (as described in Section 6.2.3) for each generated noisy price set.

[4]We obtained SPX options prices (on the S&P 500 Index) from CBOE between June and December 2020. SPX product details are available at https://www.cboe.com/tradable_products/sp_500/spx_options/.

5. Inspect the graph matrices from the results of each test group for consistency with respect to each other. Particularly the discrete Lévy measures and the implied volatility of prices generated with these Lévy measures. In other words, we inspect optimisation results with identical parameters (including initial points and regularisation) to see if they yield similar-looking results in the presence of slight pricing variations from the Gaussian noise.

Our observations were mostly favourable. The regularised calibration is reasonably effective at reducing the sensitivity to noise in the market prices.

We assessed the capability of the process to maintain the stability of solutions in the presence of price fluctuations by looking at two essential properties: Firstly, we assessed the ability of the optimised Lévy measures to hold their shape. Secondly, we assessed pricing stability by comparing the implied volatility associated with the retrieved discrete Lévy measure to the known noisy implied volatility used to generate the "market" prices.

In Section 6.2.5, we observed the sensitivity of results to the variation in initial points. Many of those results were unsatisfactory. It should be noted that in this section, we considered a "good" result to be the *consistent replication* of all those results (good or bad) when exposed to price fluctuations. In other words, the consistency of results is the only criterion considered here.

The results were predominantly good, showing low sensitivity to minor price fluctuations. In contrast, we observed significant shape changes without the support of regularisation. An example of such shape changes is shown in Figure 6.11 on the next page, where the parameters of both graphs are identical, except for the minor change in market pricing.

Although most results were good, in a few exceptional cases, the price noise shifted the resulting discrete Lévy measures to a markedly different result. This shift resembles the observed variability of results with changes in initial points. In Figure 6.12 on the following page, we show an example of good optimisation results in the top row, with a poor optimisation result in the bottom row, once again with identical parameters except for the small market pricing differences. In other words, pricing stability is reasonably good, but definitely not guaranteed.

6.2.6.3 The influence of the prior measure

The selection of the prior discrete Lévy measure is typically performed with the best "prior knowledge" of the market dynamics, as discussed in Section 5.3 on page 68. In our experiments to characterise the convergence performance, we specifically looked at the significance of the prior measure in determining the overall quality of the results. We assessed results in terms of convergence to a well-balanced result with a plausible shape and a reasonable fit to the market-implied volatility. In essence, we were interested in establishing whether any valid prior measure would provide reasonable smoothing to oppose the irregular results observed in optimisation without regularisation or whether a realistic prior measure obtained from the data is actually required.

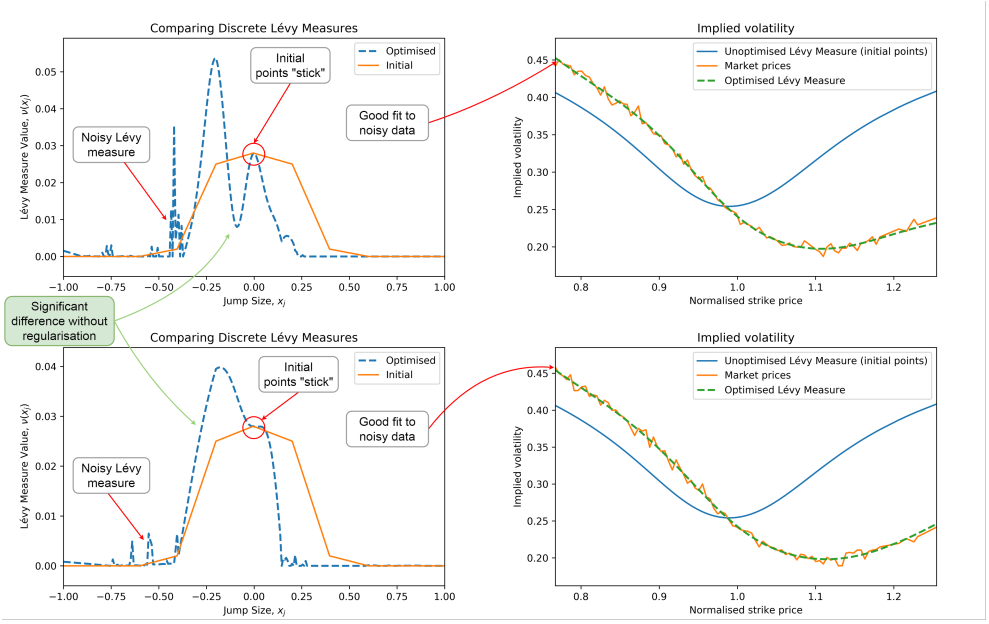

Figure 6.11: Without regularisation: significant shape changes from small price changes (identical parameters in both sets).

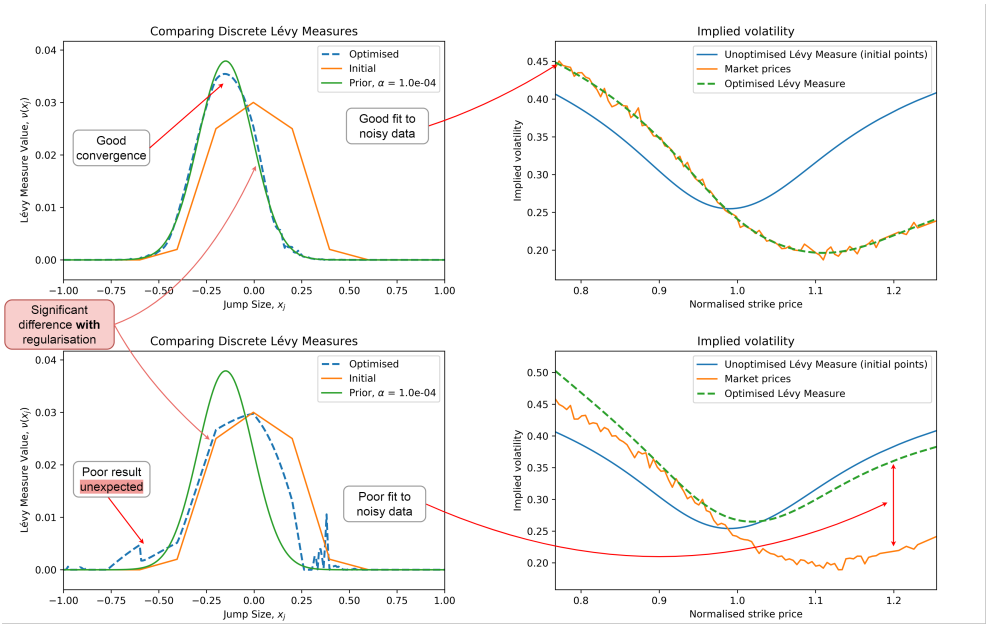

Figure 6.12: Non-convergence can result from small price changes (with all other parameters identical).

Tests with no shaping assistance We started our experiments by selecting an "infeasible" prior. We chose a prior representing equal (non-zero) activity across all jump sizes. With this experiment, we were interested in determining if the regularisation with such a prior would reduce the amplitude of the noisy spikes in the resulting discrete Lévy measure. In addition, we wished to establish whether it would be possible to extract a reasonably sensible Lévy measure using this prior, balanced with the directional contribution from the data component in the objective function. In other words, we were interested in assessing the results obtained with some smoothing from regularisation without any specific shaping capacity.

Most optimisation results with "infeasible" prior measures provided some smoothing to the retrieved Lévy measure but failed to converge to a sensible result. In many cases, the resulting implied volatility did not even roughly match the implied volatility of the market prices. The examples in Figure 6.13 on the following page show unsatisfactory pricing results, with hardly any smoothing visible in the first row and some discernable smoothing in the second row. Both have poorly fitting implied volatility curves.

There were a few isolated exceptions, where the optimisation results were sensible but not particularly impressive. Carefully selected, "helpful", initial points were applied in obtaining the results shown in Figure 6.14 on the next page. This demonstrates how initial points can contribute to good-looking results even when the prior is rather unreliable. An example of a more typical result with the same prior but without the assistance of the carefully placed initial points is shown in Figure 6.13.

These results show that carefully placed initial points are essential in providing a reasonable-looking result when the prior is inadequate. In their discussion on results with different prior measures, Cont and Tankov were silent about the impact of initial points on optimisation outcomes and showed well-shaped results from a similar "flat" prior [CT04, CT06]. We will explain later, in Section 7.3, that our interpretation of the significance of the prior measure differs from the statements made by Cont and Tankov in their articles based on a few selected results [CT04, CT06].

Tests with shaping assistance In this set of experiments, we selected a prior measure with a more appropriate shape. We chose a shape based on the Merton model, which has a normally distributed jump distribution. However, we selected parameters to provide a fairly poor match for the jump distribution implied by our market data. In this experiment, we wished to determine the quality of the results when applying regularisation with an imperfect prior that provides some stability and some support towards a plausible shape.

The shaping assistance generally provides improved results when compared to the results in the preceding paragraph, where tests were performed without any shaping support. We observed that for relatively small regularisation factors, one could reduce the magnitude of the noise while obtaining a good match to the implied-volatility from the market prices. However, this seems to require the initial points to be placed fairly close to the prior measure.

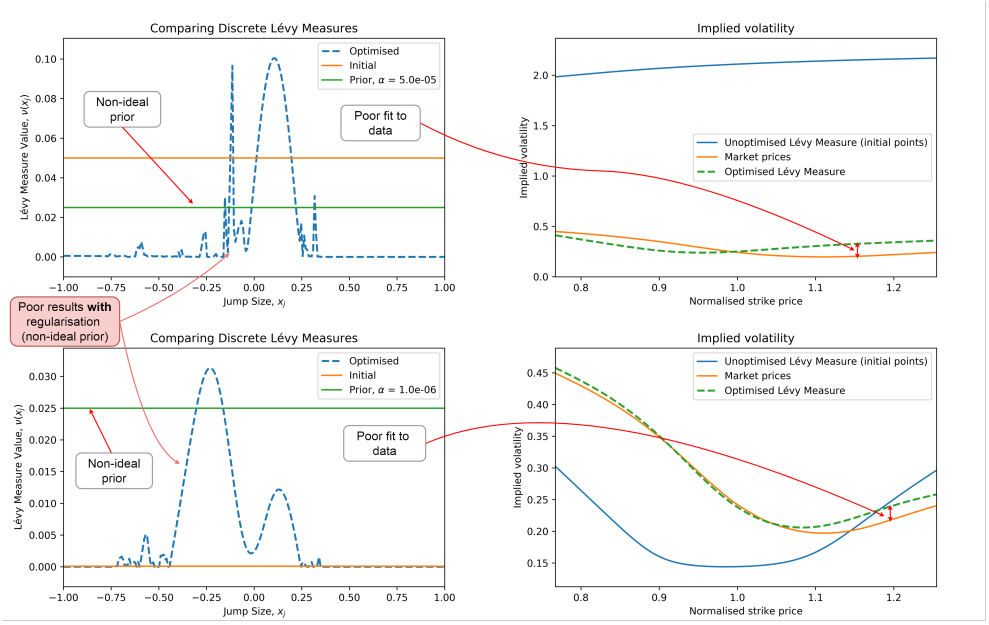

Figure 6.13: Example of unsatisfactory results obtained without an appropriately shaped prior measure.

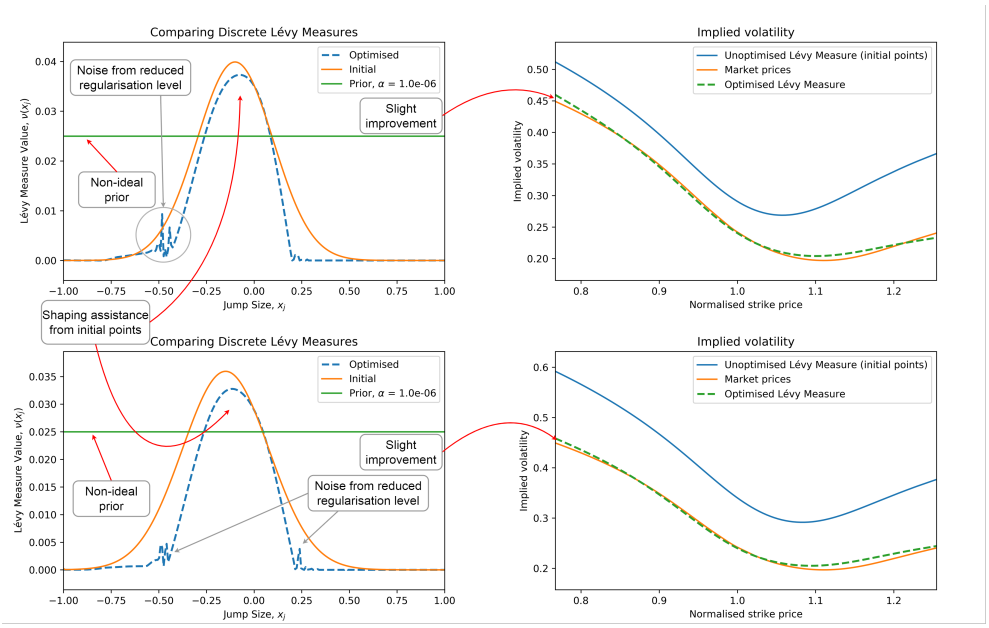

Figure 6.14: Examples of slightly improved results without an appropriately shaped prior measure.

When using these relatively small regularisation factors from more distant initial points, the results are significantly worse in terms of smoothness. As expected, increasing the regularisation yields a smoother-shaped discrete Lévy measure but results in an inferior fit to the market data. Figure 6.15 on the following page shows an example of inadequate smoothing with a reasonable pricing fit in the first row versus over-regularisation resulting in a poor pricing fit in the bottom row.

Once again, the "performance" of the prior in providing good regularisation is significantly influenced by the selected initial points. Making claims about the insensitivity of results to the prior measure without regard for the placement of the initial points is inappropriate, as will be argued in our discussion in Section 7.3.

As mentioned in our earlier discussion on initial points in Section 6.2.5, initial points close to the prior can be problematic. Starting the optimiser with initial points directly on the prior measure appears to degrade the quality of the results. When starting the optimiser with initial points coinciding with the prior, the resulting discrete Lévy measure tends to "stick" to the prior measure and a much lower regularisation factor is required to release this bond. This lower regularisation factor then results in a substantial increase in the noise levels once the optimiser is free to explore the surrounding areas. Figure 6.16 on the next page shows the results of starting the optimiser with initial points on the prior versus the slightly improved results of starting with a small offset from the prior.

The results indicate that the successful convergence of regularised optimisation results is tightly coupled with the selection of:

- Optimiser initialisation points, the
- Regularisation factor, and the
- Selected prior measure.

The results of our experiments suggest that the selection of a suitable regularisation factor is dependent on both the initial points as well as the selected prior measure. Therefore, contrary to our initial expectations, the selection of a universal regularisation factor that will provide good results across the board is not achievable. A practitioner will need to carefully select these three coupled parameters and make the required adjustments to obtain suitable results.

6.2.6.4 Adjusting optimiser parameters

Setting tighter tolerances on the termination criteria of the optimiser certainly had a positive impact on the quality of the results. Adjusting the termination criteria based on the gradient value and the minimum reduction threshold of successive optimisations proved a sensible approach to nudge results to a better solution. This approach managed to reduce optimisation failures, but only up to a point. In general, tighter tolerances lead to improved results. However, we observed diminishing returns at the cost of a significantly increased computational burden beyond a certain point.

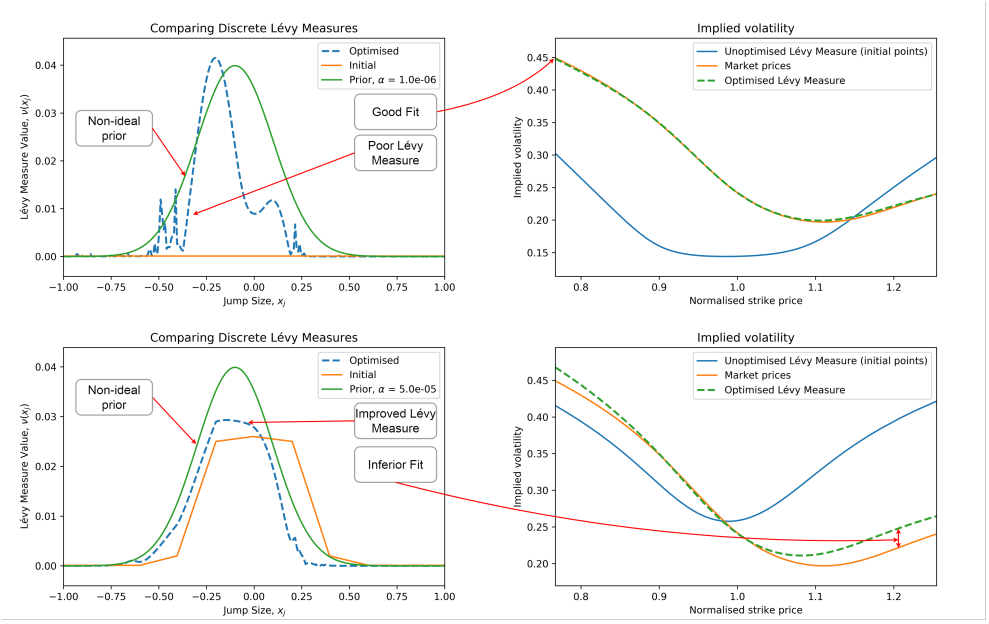

Figure 6.15: Examples of compromise between smoothness and price fit with a non-ideal prior measure.

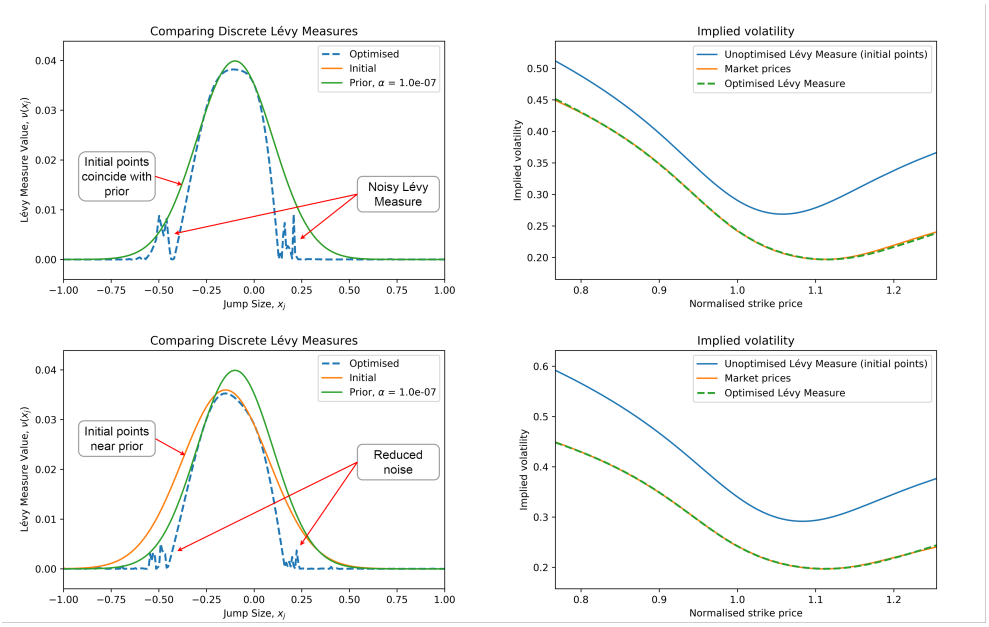

Figure 6.16: Example of slight improvement when initial points do not coincide with the prior (in second row).

Table 6.3 shows summary figures for an example drawn from the output data of three convergence test groups, showing a rough[5] trend of improvement in results with stricter tolerance on termination based on successive function values. The increase in the number of function calls illustrates the exponential increase in the optimisation time along with a reduced increase in convergence success rate.

6.2.6.5 Objective function variations

A single optimisation result could fail to converge with one objective function formulation while converging to a reasonable solution with another. We observed that changing the objective function's characteristics can significantly impact the results when viewed as individual results in isolation.

For example, a result that failed to converge using an objective function with a mean-squared error might converge when using a root-mean-squared error function for the same initial points and regularisation parameter. However, when evaluating a group of results with a different objective function, the overall convergence success rate did not improve when changing from a mean-squared error to a root-mean-squared error function.

6.2.6.6 Changes to pricing parameters

The impact of changes to the pricing parameters on optimisation results was observed to be similar to the above-mentioned observation for changes to the objective function.

On an individual basis, some impact was observed. Some results that failed to converge with one pricing parameter set would converge with a different set. However, when viewed as a group of results with a fixed pricing parameter set applied to the same group, there would not be a particular pricing parameter set yielding consistently superior results.

In other words, marginal changes in pricing were able to influence the convergence result of a particular optimisation instance. This could be similar to the observation of some sensitivity to small price changes in Section 6.2.6.2.

6.2.6.7 Overall impact of secondary parameters

In the preceding tests with variations in the secondary parameters, we examined the optimisation outcomes across a range of parameter values for a selection of parameters.

The variation of optimisation results with the prior measure and with small price variations could be considered regular performance tests since they test core characteristics

[5]These figures cannot be considered as significant in terms an absolute measurement of the convergence rate, but serve the purpose of showing the relative improvement versus optimiser termination tolerance. The figures were based on an arbitrary (but identical) set of initial conditions with an arbitrary error threshold as the convergence pass criteria, so the absolute convergence rates shown are unimportant.

Function value tolerance	Convergence pass rate	Average number of function calls
10^{-9}	58%	95
10^{-12}	67%	203
10^{-15}	72%	433

Table 6.3: Results illustrating improved convergence with stricter optimiser termination tolerance on the value of the objective function.

of the relative-entropy-based calibration technique. However, the other secondary tests developed in our search for improvement.

Our initial impressions of the calibration performance and usability of the regularised calibration procedures were rather negative. Since the low quality of the calibration results and concerns about the stability of the process suggested the possibility of an implementation error, we were compelled to significantly increase our test efforts.

The verification tests described in Appendix C were primarily the result of an intensified search for means to improve the overall performance. Similarly, most secondary parameters were introduced as part of an in-depth search for potential adjustments to improve the results.

We found that improvements from tweaking the secondary parameters were discernable, but unfortunately, they cannot be regarded as substantial improvements. Next, we will summarise our assessment of the tests.

6.2.7 Performance assessment

In terms of our assessment criteria, stated in Section 6.2.1, we summarise our evaluation based on the preceding results under the corresponding headings below.

6.2.7.1 The control of regularisation

Controlling the balance between smoothness and calibration precision has proven to be challenging. Our results revealed that a user could not rely on this regularised calibration procedure for consistent and predictable control over the fundamental qualities of the calibration outcome.

From a control perspective, the observed instances of a non-monotonic relationship between the regularisation level and the smoothness are unacceptable. Furthermore, the unexpected variations in shape from minor adjustments in regularisation are another undesirable peculiarity.

6.2.7.2 The relief from ill-posedness

Our experiments confirm that applying very high levels of regularisation will reduce the sensitivity of results to the initial points supplied to the optimiser. In these over-regularised cases, however, the retrieved discrete Lévy measures essentially converge to the prior measure, overriding the information in the data.

Our grid-search experiments suggest that for a given level of regularisation that yields acceptable results for a particular set of initial points, a different set of initial points would generally not yield good results at that level of regularisation.

In other words, the ill-posedness associated with the optimiser sensitivity to initial points can generally not be adequately removed when applying reasonable levels of regularisation.

We observed reasonable stability of solutions when subjected to minor variations in pricing data. Although this aspect of ill-posedness is significantly improved by regularisation, stability is not guaranteed.

6.2.7.3 The uniqueness of solutions

Solutions are generally not unique. Moreover, results from our experiments suggest that convergence to solutions with a consistent appearance is infrequent. However, when results are carefully inspected and parameters are manually adjusted with care, it seems possible to retrieve reasonable-looking optimal discrete Lévy measures from certain initial points. Figure 6.10 on page 98, for example, shows a number of comparable optimisation results for the same data and prior measure but with different initial points and regularisation parameters. Unfortunately, the heavy burden on the user to search for suitable combinations of parameters to yield appropriate results makes this approach impracticable.

6.2.7.4 Autonomous operation

As mentioned in the preceding paragraphs, a reasonable degree of user intervention is required to manage the optimisation process. This lack of autonomous operation, and the resulting need for meticulous user intervention is a significant limitation of this procedure.

Another drawback of the need for user intervention is that it opens up the opportunity for the manipulation of results. A selected prior combined with carefully placed initial points can provide a practitioner with the opportunity to shape results according to his/her personal views of the market. The objectivity of an autonomous process subsides with increasing levels of user intervention with increased potential of inadvertently introducing bias.

Section 7.1 will discuss the key findings from our experiments in this chapter and then compare our results with related results from existing literature.

6.3 Recovery of a known Lévy measure

In the preceding sections, we tested various individual aspects of the regularised calibration procedure with the objective of evaluating the performance against a set of suitable criteria. In this section, we test the performance of the overall system by testing the ability to recover a known Lévy measure from its prices.

We generated prices using the following Lévy measures:

1. Merton model
2. Kou model
3. Variance Gamma model
4. Uniformly distributed Lévy measure
5. Triangular distributed Lévy measure
6. A bimodal Lévy measure

The first three price sets were generated directly from the respective characteristic functions and verified against *Premia*[6]. The last three price sets were generated by defining custom discrete Lévy densities and then using the characteristic function based on the non-parametric form of the Lévy-Khinchin formula in (5.2). For each known model or Lévy measure, we generated prices at two sets of strike prices. For the first set, we generated 80 call prices at our standard strike range of $0.8 \times S_0$ to $1.2 \times S_0$. For the second set, we used a wider span of $0.6 \times S_0$ to $1.4 \times S_0$, yielding 160 prices at the same strike density.

For each of the ten price sets from Lévy measures 2 to 6 above, we calibrated (parametrically) to the Merton model to create ten Merton prior measures. Similarly, we calibrated ten sets to the Kou model for use as prior measures.

Using the 20 prior measures from the Merton and Kou models, we set up 20 experiments where we test the ability of the procedures to recover the generating Lévy measures from the price sets. For each experiment, we set up a search grid of parameters to vary the initial points and the regularisation constant to find feasible solutions. Our focus was on selecting reasonable values (aspiring to achieve success) without using unfairly obtained information about the characteristics of the generating measure.

Unfortunately, we were unable to recover meaningful results. Optimisation results failed to converge to anything resembling the generating measures. In our assessment of results, we allowed for leniency in the central regions where we know that the data has almost no impact, as mentioned in Section 6.2.2.2. However, in spite of our softened evaluation, we still could not claim any degree of success. Figure 6.17 shows an example of a typical result from our Lévy measure recovery experiments.

[6]We used the command line version of *Premia 16* option pricing software, written and maintained by the MathRisk project. The software and addition information can be obtained from https://www.rocq.inria.fr/mathfi/Premia/.

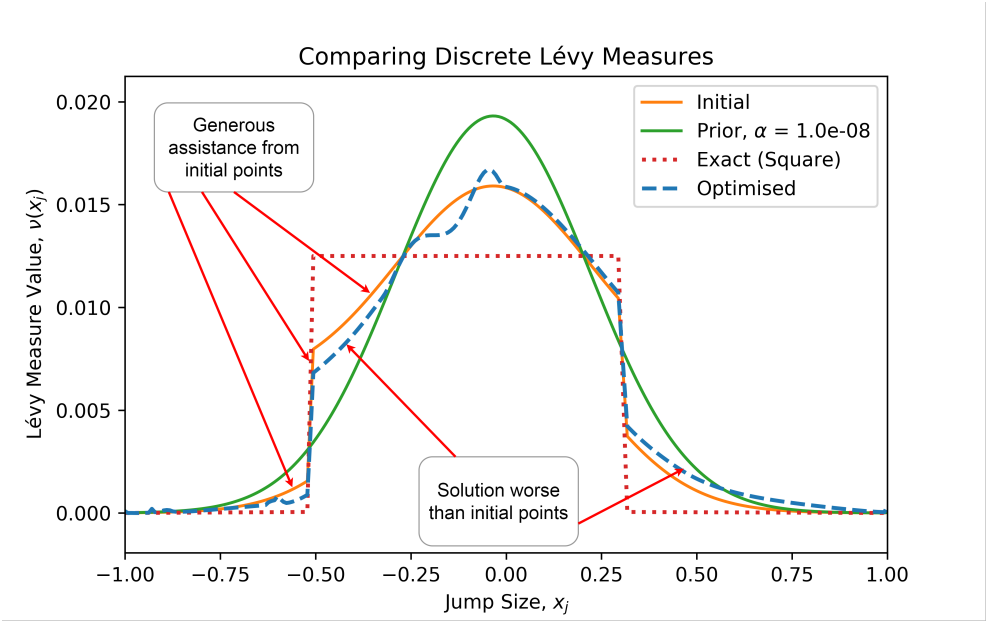

Figure 6.17: Unsuccessful attempt at the recovery of a known Lévy measure.

We conclude that at relatively objective initial points (not giving away any previous knowledge of the actual generating measures), the parametrically calibrated prior measures were unable to assist the regularised optimisation procedure in finding the generating Lévy measures. The regular-entropy-based calibration procedure fails to deliver meaningful results at recovering the Lévy measures. Even when supplied with additional assistance obtained from fairly generous initial points, results were relatively poor. This supports our earlier observation that the stability obtained from a prior measure can reduce the sharp fluctuations, but could result in an unreasonable result. In other words, the prior needs to be "appropriate" to guide the optimiser to a realistic solution. We will discuss this aspect in more detail in Section 7.3, where we will argue that the impact of the prior on the final result is significant. Our interpretation of the significance of the prior measure differs from the statements made by Cont and Tankov in the original articles [CT04, CT06].

6.4 Pricing through simulation

In this section, we demonstrate how path-dependent options can be priced by simulating paths generated by Lévy models. We started the experiment by calibrating a selection of Lévy models to the same vanilla options market data. We then simulated paths for these Lévy models and used simple Monte Carlo techniques, described earlier in Section 2.7, to price the options. We compared the path-dependent options prices from selected *parametric* Lévy models to results from pleasing regularised *non-parametric* calibrations. We found that prices from our sample of "reasonable-looking" non-parametric discrete Lévy measures were in a similar price range to the bulk of the parametric models.

Furthermore, we compared the same set of path-dependent prices from a poor non-parametric calibration result with "reasonable-looking" non-parametric results. Our results show reasonably close pricing for the group of good calibration results and a discernable pricing deviation for the prices of the poorly calibrated non-parametric Lévy measure.

In the next subsection, we describe the path-dependent barrier options used to perform the pricing tests. This chapter concludes with the results of our barrier option pricing experiments.

6.4.1 Selected barrier options

For our tests, we selected a set of path-dependent *Barrier options*. The options are briefly described in the following paragraphs, where we use the symbol B_{type} to represent the barrier level of the relevant barrier *type* along with S_T^* and S_{*T} to represent the respective maximum and minimum levels of the underlying during the period of the contract.

Knock-in options are activated by the underlying touching a specified barrier during the contract period [Hul12]. Once activated, they assume the usual pay-off at maturity. They remain worthless if the activation barrier is not reached.

Similarly, a *knock-out* option becomes worthless if a specified barrier is touched or otherwise assumes the usual pay-off at maturity [Hul12].

Up-and-in options

This is a knock-in option which is activated by the underlying touching the high barrier, B_{ui}, during the contract period. The up-and-in call option price can be expressed as

$$C_T\left(K, B_{\text{ui}}\right) = E_{\mathbb{Q}}\left[e^{-rT}\left(S_T - K\right)^+ 1_{S_T^* \geq B_{\text{ui}}}\right],$$

where S_T^* represents the highest value of S_t until maturity,

$$S_t^* = \max_{u<t} S_u.$$

Down-and-in options

This is a knock-in option which is activated by the underlying touching the low barrier, B_{di}, during the contract period. The down-and-in call option price can be expressed as

$$C_T\left(K, B_{\text{di}}\right) = E_{\mathbb{Q}}\left[e^{-rT}\left(S_T - K\right)^+ 1_{S_{*T} \leq B_{\text{di}}}\right],$$

where S_{*T} represents the lowest value of S_t until maturity,

$$S_{*t} = \min_{u<t} S_u.$$

Double-in options are activated when either the low or the high barrier is reached.

Up-and-out options

This is a knock-out option which becomes worthless if the underlying touches the high barrier, B_{uo}, during the contract period. The up-and-out call option price can be expressed as

$$C_T\left(K, B_{\text{uo}}\right) = E_{\mathbb{Q}}\left[e^{-rT}\left(S_T - K\right)^+ 1_{S_T^* < B_{\text{uo}}}\right],$$

where S_T^* represents the highest value of S_t until maturity,

$$S_t^* = \max_{u<t} S_u.$$

Down-and-out options

This is a knock-out option which becomes worthless if the underlying touches the low barrier, B_{do}, during the contract period. The down-and-out call option price can be expressed as

$$C_T\left(K, B_{\text{do}}\right) = E_{\mathbb{Q}}\left[e^{-rT}\left(S_T - K\right)^+ 1_{S_{*T} > B_{\text{do}}}\right],$$

where S_{*T} represents the lowest value of S_t until maturity,

$$S_{*t} = \min_{u<t} S_u.$$

Double-out options are knocked out when either the low or the high barrier is breached.

In the next subsection we will present results of pricing experiments for knock-out barrier options[7].

6.4.2 Barrier option pricing results

We calibrated the our parametric models to a sample of vanilla options. We used options prices based on the S&P index as described in earlier Section 6.2.6.1. We selected a sample of SPX options prices on 3 December 2019, corresponding with a value of 3109.3 for the underlying. We chose options with a reasonably short maturity of 59 days and calibrated the following parametric models to the data:

- VG,
- NIG,
- Kou, and

[7]Due to in-out parity, performing the experiments for knock-in options is unnecessary since no additional information will be revealed.

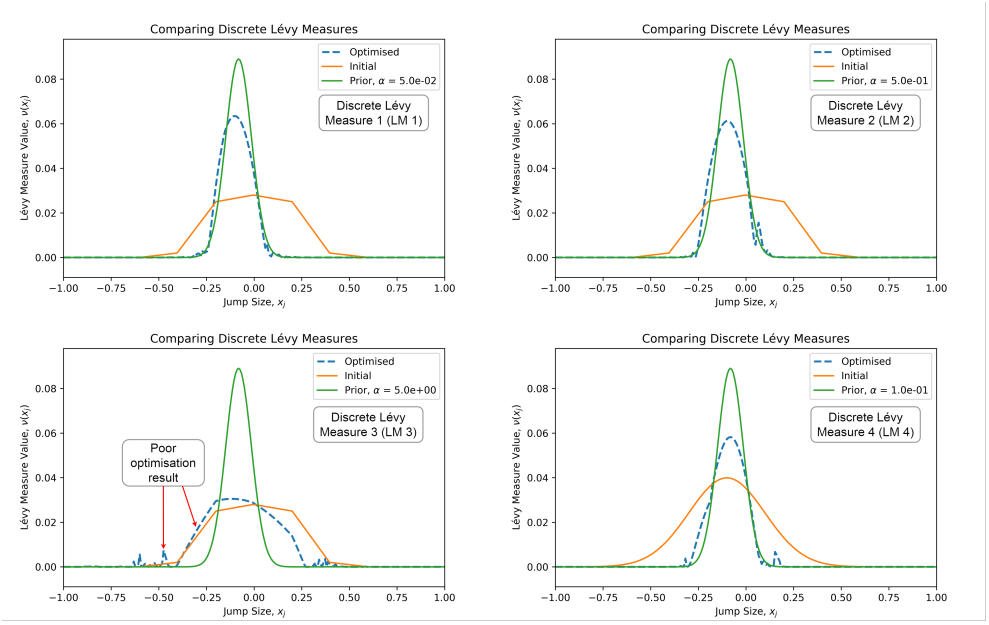

Figure 6.18: Discrete Lévy measures selected for illustrative path simulations.

- Merton.

We then performed regularised non-parametric calibrations using the Merton model as the prior. The different outcomes resulted from varying calibration parameters, such as initial points and the regularisation parameter. Figure 6.18 shows a sample of four optimisation results of our regularised non-parametric calibration. The result displayed in the bottom-left (LM3) is an example of a poor optimisation result.

In our barrier option pricing experiments, we first compare the prices of the parametric models together with the pricing obtained from one of our "reasonable" non-parametric results. We selected the result displayed in the top-left position in Figure 6.18 (labelled LM1) as our non-parametric example.

The second group barrier option experiments compare the prices generated by the four discrete Lévy measures from the optimisations shown in Figure 6.18.

In examining the comparative pricing performance, we focussed on the difference in the prices delivered by each Lévy measure. We chose to display the relative differences between the prices by subtracting a reference price set for each pricing experiment. We selected the barrier options prices generated by our calibrated Merton model (which also served as our prior) as the baseline price set for displaying the relative price differences. Selecting the Merton model as a reference has the additional advantage of illustrating the price deviation with respect to the prior's prices for the non-parametric pricing experiments.

Figures 6.19, 6.20, and 6.21 illustrate the results of our down-and-out pricing experiments at three different barrier levels. The graphs on the right-hand side illustrate the relative pricing differences with the Merton model as the reference offset price.

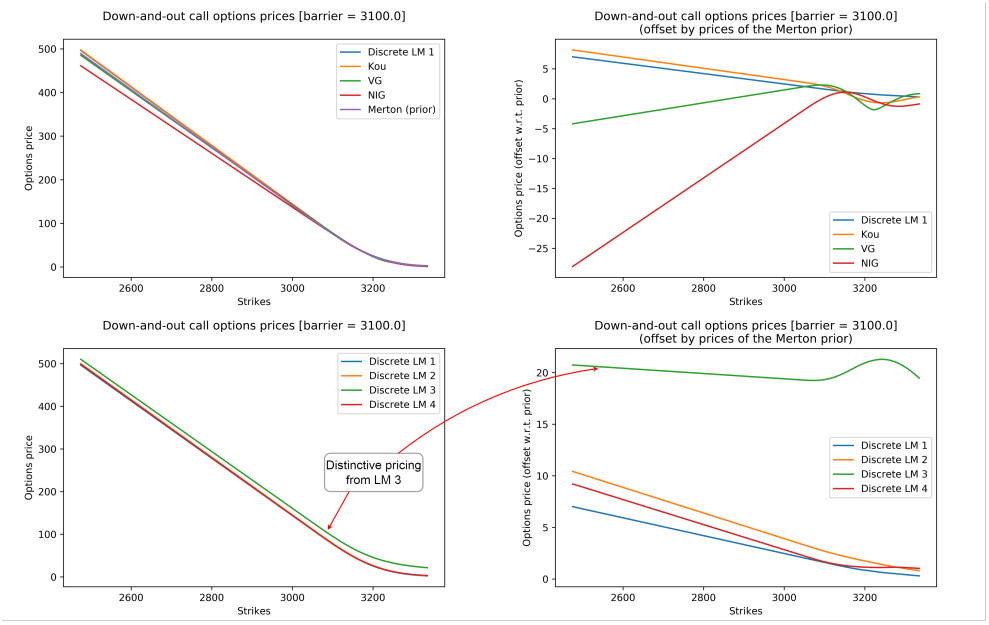

Figure 6.19: Down-and-out options prices with barrier set at 3100.

Similarly, Figures 6.22, 6.22, and 6.24 illustrate the results of the up-and-out pricing experiments at different barrier levels.

Our results suggest that we obtained reasonable path-dependent pricing from the recovered Lévy measures when compared with parametric models. We did not observe any abnormal behaviour and found that, in general, there was no significant pricing deviation from the Merton reference level. As expected, the pricing from the poor non-parametric result (LM3) is clearly discernable in most of our results.

Although we demonstrated pricing from an insignificant sample of recovered discrete Lévy measures, it is interesting to note the rather small deviation relative to the Merton pricing. One may wonder if this observation suggests a degree of over-regularisation with too much bias towards the prior.

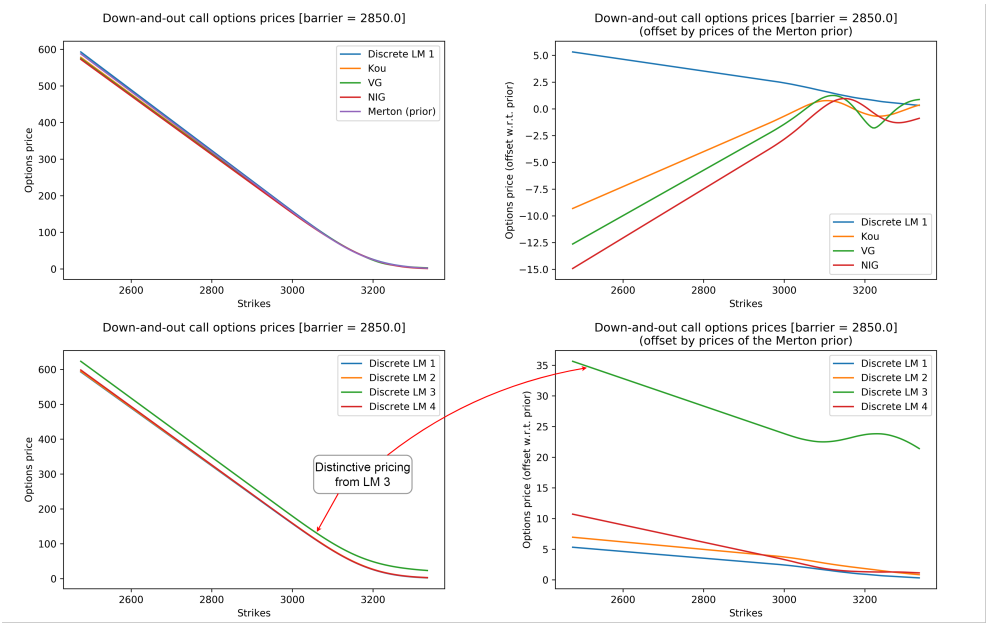

Figure 6.20: Down-and-out options prices with barrier set at 2850.

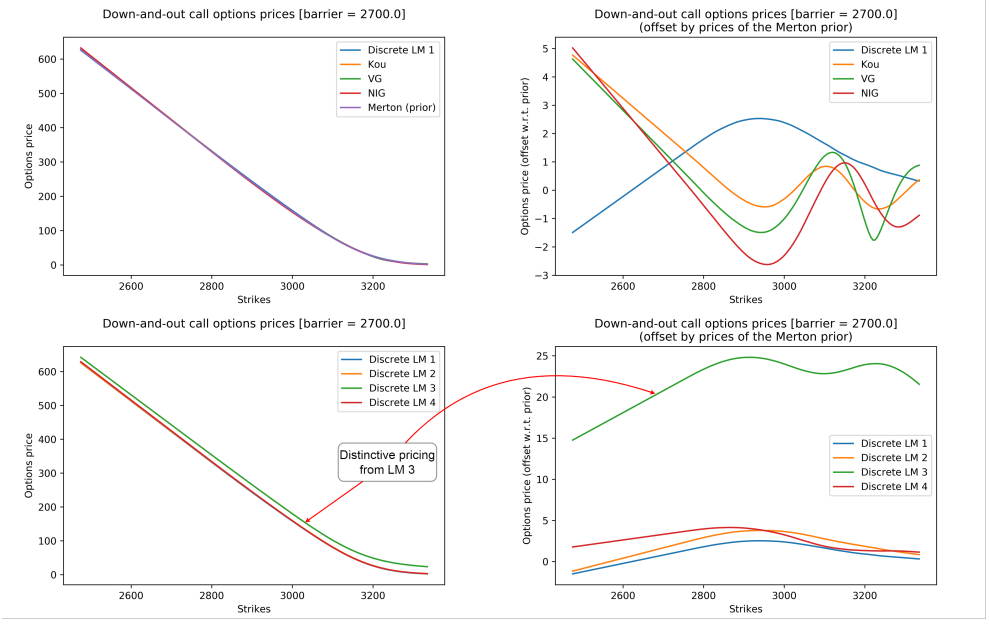

Figure 6.21: Down-and-out options prices with barrier set at 2700.

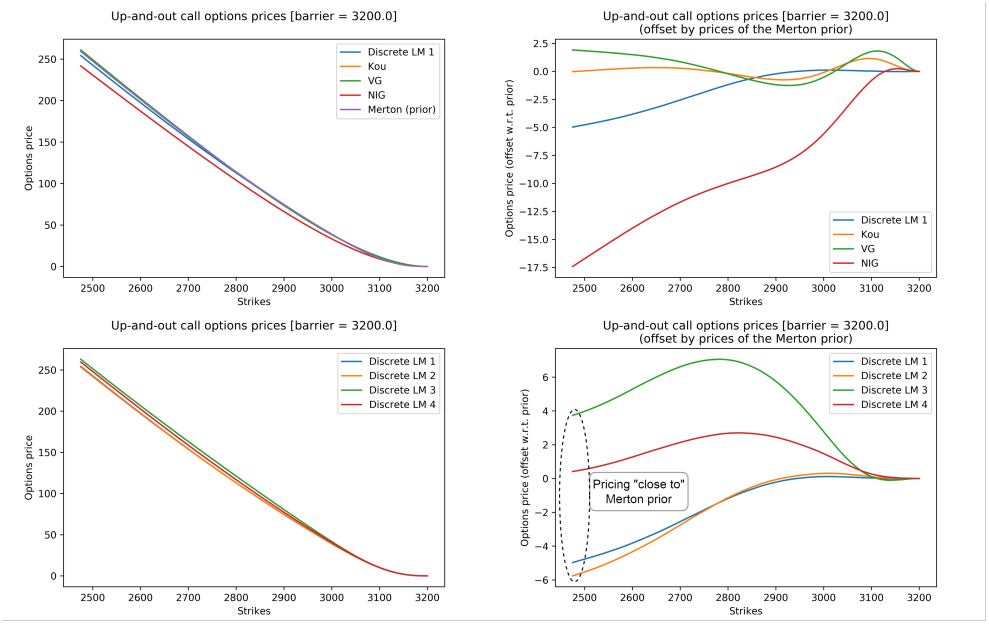

Figure 6.22: Up-and-out options prices with barrier set at 3200.

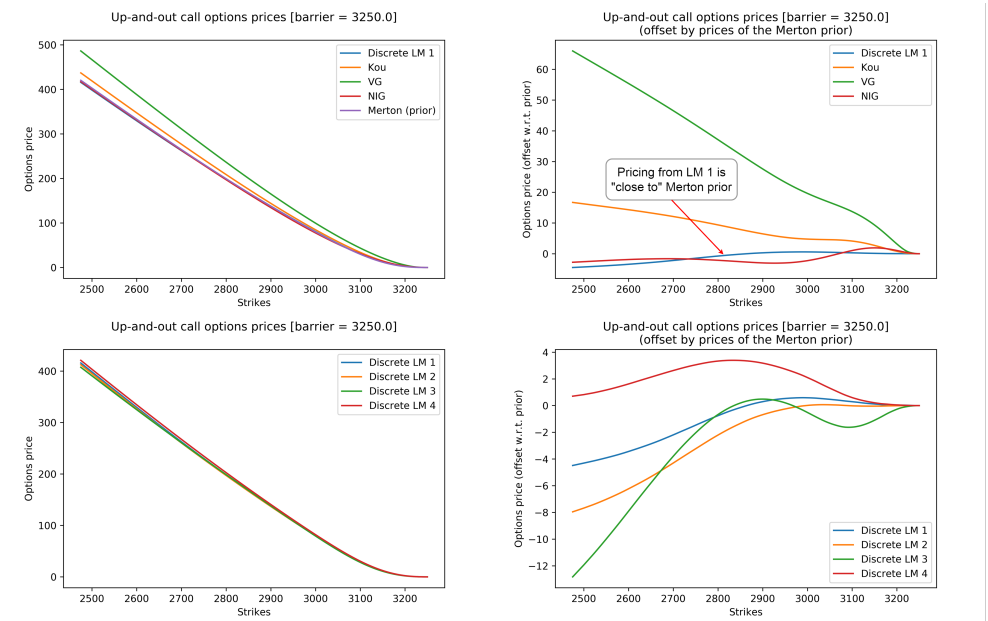

Figure 6.23: Up-and-out options prices with barrier set at 3250.

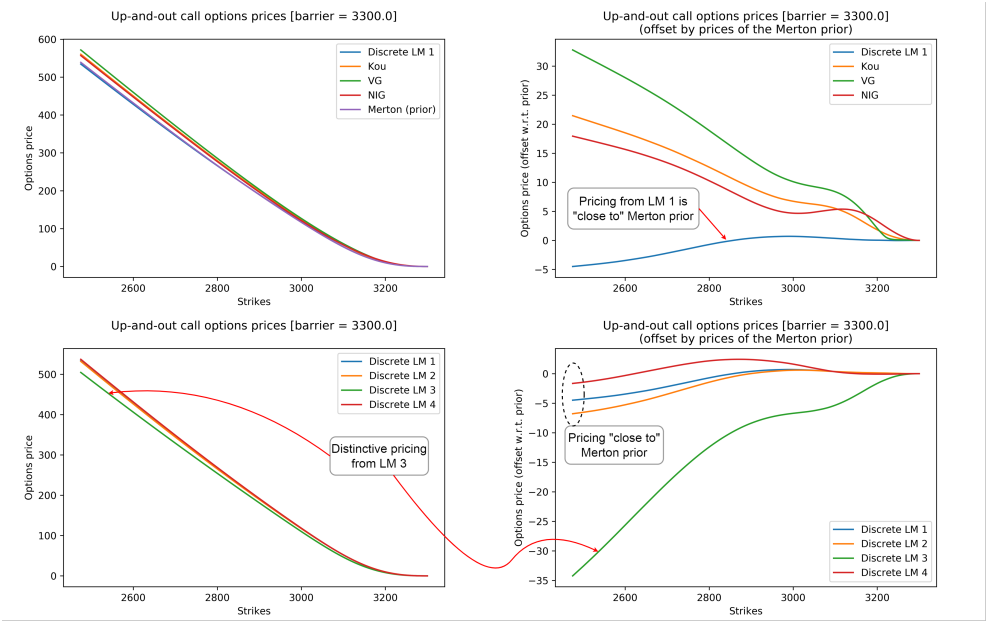

Figure 6.24: Up-and-out options prices with barrier set at 3300.

Chapter 7

Discussion

In this chapter, we take a closer look at the key observations from our experiments in the preceding chapter, compare our findings with those of other independent researchers, and show that others encountered similar problems. This is followed by a critical review of the articles by Cont and Tankov in relation to our findings, where we highlight some deficiencies in their work.

7.1 Key findings of this study

In Chapter 6, we performed a series of experiments to determine the qualities of the regularised optimisation procedures against a set of applicable assessment criteria. In Section 6.2.7, we presented the outcomes of the tests against the assessment criteria. We will now summarise the key findings arising from our assessment.

Although it is possible to squeeze good results from these regularised calibration procedures, the overall assessment is rather negative. We summarise our criticism as follows:

- Solutions are generally not unique and resulting Lévy measures can have vastly different shapes.
- Optimisation results are sensitive to the initial points supplied to the optimiser.
- The control over smoothing versus accuracy is troublesome, and small changes in regularisation can result in significant changes in the optimiser result. Monotonicity of the smoothing effect is also not guaranteed.
- Although stability with respect to small changes in market prices has been observed, stability is not guaranteed. Significant changes in the optimisation results following minor changes in market data have also been observed.
- Results depend not only on the amount of regularisation but also on the properties of the prior measure.

Overall the regularised calibration procedure is unreliable and requires active user intervention to manage outcomes. The process remains ill-posed with some sensitivity to market prices and significant sensitivity to initial points.

Interpretation of findings

In this research project, finding meaningful answers to the observed behaviour has been challenging. Several additional tests were performed and even repeated to verify our implementation, as our initial beliefs were strongly biased towards finding a fault in our own work. Our final interpretation is surprisingly simple, and we state our inference in the following paragraphs:

1. The optimisation problem remains non-convex and is therefore ill-posed with multiple local minima for any "reasonable" level of regularisation. In other words, the amount of regularisation required to force the result into convexity makes the overall procedure ineffective.

2. We believe that gradient-based optimisation routines may fail to deliver consistent results for this type of optimisation problem, which is inclined to be *poorly scaled*.[1] Poorly scaled optimisation problems may lead to ill-conditioned systems of equations, which tend to result in poor behaviour in gradient-based optimisation routines. This problem is usually associated with a vast range of magnitudes in the solution vector and a significant difference in the sensitivity of the objective function to changes in the individual components. A closer inspection of the range of magnitudes and sensitivities encountered in the Lévy measures of the Merton or Kou models, typically used as prior measures, will illustrate the point.

3. We harbour reservations about the convexity of the relative-entropy regularisation function in the case of a discrete Lévy measure. This aspect is discussed in Appendix C.4.2, and illustrated Figure C.12 on page 166.

The interplay between these three problematic items provides a troublesome platform for performing any worthwhile financial modelling.

As mentioned in Section 5.6.1, the use of a global optimiser was considered as an approach to combat the non-convexity, but it proved to be infeasible due to the steep computational burden.

Recovery of a price generating Lévy measure

More advanced experiments, such as recovering a Lévy measure with a known predefined shape from its vanilla options prices, proved disappointing.

We know that it is possible to achieve smooth-looking optimisation results by employing the combination of the attraction to a prior measure and carefully placed initial points.

[1] Nocedal and Wright describe a *poorly scaled optimisation problem* as an optimisation problem where changes to the vector x in a certain direction results in much larger variations in $f(x)$ than changes to x in another direction [NW06].

Take, for example, the case of initial points selected to coincide with the prior measure: Generally, this will strongly bias the result toward something closely resembling the prior measure, particularly in the small-jump central region. In the central (small jump) regions, the Lévy measure is insensitive to the data, as discussed in Section 6.2.2.2 on page 83.

In the unlikely event that the prior (from a calibrated parametric jump-diffusion) happens to be a particularly good fit for the embedded Lévy measure, this approach will generally yield good-looking, smooth results. However, our experiments suggest that it is unreasonable to expect meaningful results from this approach in the case of a truly unknown shape for the Lévy measure. Choosing a convenient prior such as calibrating a Merton Lévy measure to the data from a completely differently shaped Lévy measure and then initialising the optimiser to coincide with this prior will not result in anything resembling the price-generating Lévy measure.

The additional bias introduced by setting the initial points to coincide with the prior in this non-convex optimisation problem can therefore be used to manipulate the appearance of the results, especially when the generating measure happens to be closely related to the prior model.

7.2 Findings by other researchers

Guillaume and Schoutens

As mentioned earlier in the literature review, Guillaume and Schoutens expressed criticism regarding the value of the regularised calibration procedures of the primary articles [CT04, CT06] in finding a meaningful solution [GS13]. They objected to the dependence of the solution on initial points and on the selected prior and stated the following:

> "the solution of the regularized calibration problem depends on the choice of the prior and, more importantly, on the starting parameter set" [GS13].

This statement is in complete agreement with the results that we obtained in the previous chapter. We will also show in our discussion on the original articles below why we find the "empirical proof" of the insensitivity of results to the selection of the prior, furnished by Cont and Tankov [CT04], inadequate.

Jeonggyu Huh

In the article, *Pricing options with exponential Lévy neural network* [Huh19], Huh suggests that the "existing"[2] non-parametric calibration procedures for exponential Lévy models fail to deliver useful results in practice.

[2] Referring to the procedures of Cont and Tankov[CT04, CT06] and Belomestny and Reiss [BR06].

Huh states that he abandoned his attempts to get meaningful comparative results from existing non-parametric methods and was forced to settle for comparing his results with parametric Lévy models instead. He described it as being "too demanding" to obtain comparative results due to "vulnerabilities" encountered when applying the "existing" regularised non-parametric methods in practice [Huh19].

He stated that he focused his attention instead on highlighting the problems encountered with the regularised non-parametric methods. He claims that the methods of Cont and Tankov [CT04, CT06] are not only ill-posed with multiple local minima, but also *ill-conditioned* when applied in practice [Huh19].

Both the ill-posedness and the ill-conditioned nature of the problems that we observed agree with Huh's observations. [Huh19].

7.3 Critical assessment of articles by Cont and Tankov

Equipped with our results from the preceding chapter, we closely inspect selected results and statements published by the authors in the principal articles [CT04, CT06] and other related literature by the authors.

We will start by considering the most recent article, *Retrieving Lévy processes from option prices: Regularization of an ill-posed inverse problem* [CT06], and then systematically work back to address earlier work.

Retrieving Lévy processes from option prices: Regularization of an ill-posed inverse problem [CT06]

We wish to highlight two notable concerns about this article:

1. The article does not clearly state whether (or not) the regularisation procedure is successful at dealing with the non-convexity and ill-posedness mentioned in their problem statement.
2. The article suggests that the selection of the prior measure is insignificant, and that inaccuracy in the selection of the prior measure will not impact materially on the quality of solutions.

We will briefly motivate our concerns below.

Transparency relating to performance

A reader could easily be led to believe that this technique has the ability to fully address the problems of non-convexity and ill-posedness associated with the non-parametric calibration problem. A clear problem statement in the article describes the problem of ill-posedness associated with standard least squares optimisation. This is followed by a detailed description of the benefits associated with their regularisation procedure. This

could easily prompt belief that their technique solves the problems of non-convexity and ill-posedness. Moreover, the omission of a discussion about their performance limitations has the potential to obscure the fact that, in truth, solutions remain ill-posed and non-convex even after regularisation.

The authors tone down the problem of the persistent non-convexity of the regularised solutions by suggesting that it could be advantageous to have multiple solutions over a unique solution:

> "Using relative entropy for the selection of solutions removes, to some extent, the identification problem of least squares calibration. Whereas in the least squares case, this was an important nuisance, now, if two measures reproduce market option prices with the same precision and have the same entropy relative to the prior, this means that both measures are compatible with all the available information [CT06]."

Although the problem statement identified multiple solutions as a nuisance associated with the standard least squares optimisation, the authors omitted discussion on this problematic aspect of ill-posedness in evaluating their own regularised optimisation procedures.

They continue their justification for the "benefits" of multiple solutions by stating that:

> "Knowledge of many such probability measures instead of one may be seen as an advantage, because it allows us to estimate model risk and provide confidence intervals for the prices of exotic options [CT06]."

The dominant impact of the prior measure

Later, in their discussion on the importance of the selection of the prior measure, they omit discussing the issue of a potentially dominant impact of the prior on determining which solutions will be "selected" by the optimisation procedure. Instead, they lead the reader through a mathematical proof ([Con06b] Theorem 6.3), by which one could easily be lead to believe that the correctness of the prior is not important in the outcome of their procedure. Their narrative after the proof is as follows:

> "Another implication of the above theorem is that small changes in the prior Lévy process lead to small changes in the solution: the solution is not very sensitive to minor errors in the determination of the prior measure. This result confirms the empirical observations[3] made in [CT04]" [CT06].

The article prompts the belief that solutions are not tightly coupled with the selection of the prior measure used for regularisation. In the next section below, discussing the companion article [CT04], we will show that this claim about the prior measure is unsound. We demonstrated in our results that the prior measure has a significant impact

[3] These empirical observations will be challenged in our discussion of that article [CT04] later in this section.

on the solutions. This observation is in agreement with those of other independent researchers, as discussed in Section 7.2.

The authors focus attention on positive claims about their methods, but omit discussions regarding the limitations and weaknesses of these methods, even though the authors claim that the article is a "rigorous analysis" of the topic.

Nonparametric calibration of jump-diffusion option pricing models [CT04]

In the following paragraphs, we will discuss the following three selected issues from the article *Nonparametric calibration of jump-diffusion option pricing models* [CT04]:

1. The article omits details describing the capability of the regularisation procedure in addressing the non-convexity and ill-posedness that were mentioned in their problem statement. However, the authors prompt readers to believe that their procedure transforms the calibration operation into a well-posed problem.
2. The article omits discussions about the influence of initial points on solutions to the regularised calibration, even though the sensitivity to initial points formed part of their problem statement.
3. The article uses an unconvincing example in support of their suggestion that the correct selection of the prior measure is insignificant and that solutions are insensitive to the correct selection of the prior measure.

We will briefly discuss these concerns below. After addressing the three concerns, we will also present an observation that the shape of some centrally located Lévy points in solutions could possibly be the result of carefully placed initial points.

Well-posedness and uniqueness of solutions

In this article, the authors state that they propose a "well-posed formulation of the calibration problem" by introducing a regularisation term into the least-squares-based objective function [CT04]. They claim to enforce uniqueness and stability by introducing the regularisation term [CT04]. This statement could be interpreted as suggesting that their regularised solutions are well-posed and unique.

We know that uniqueness remains a problem and observe that this claim of well-posedness and uniqueness was omitted in their subsequent companion article [CT06], discussed above. As pointed out in our discussion above, the subsequent companion article contains an "admission" of non-uniqueness by stating that multiple "compatible" solutions exist. This article, however, does not mention this important issue.

The issue of initial points

The problem statement clearly mentions that the standard least-squares optimisation problem suffers from a sensitivity to the initial points.

One would expect an "empirical analysis" of the performance of the regularised optimisation procedures to include some tests to demonstrate the ability to converge to a unique solution from various initial points, particularly if the objective of the procedure is to address ill-posedness.

However, Cont and Tankov omitted this important aspect from their published results. Instead, the authors demonstrate the sensitivity to initial points in the un-regularised case (in their Figure 3 [CT04]). Then they proceed to state on page 15 in their text: "As can be seen in Figure 3 (left graph), the results of the minimization are totally different!" [CT04].

After proclaiming the flaws of the un-regularised optimisation in their problem statement, they omit discussing this matter when evaluating their own procedures.

We have already pointed out that our study suggests that the selection of initial points is a major factor in these procedures:

- The optimisation outcome is strongly influenced by initial points supplied to the optimiser.
- The optimisation problem remains ill-posed with sensitivity to initial points.
- The optimisation outcome can easily be manipulated to a desired appearance by supplying particularly favourable initial points fairly close to the preferred result.

Readers who have read the problem statement and the "suggestion" of well-posedness may be inclined to miss noticing the *absence* of:

- a clear statement about the performance of the procedures in addressing the ill-posedness and
- test results for the sensitivity of solutions to the initial points.

It is therefore conceivable that some readers may be led to believe that the regularised calibration delivers well-posed results with a unique solution and without any sensitivity to initial points.

In our opinion, a more open and complete empirical analysis by the authors would have been more valuable to the research community.

Insensitivity of solutions to the prior measure

The authors claim that this article provides "an empirical analysis of the effect of the choice of prior on the solution of the calibration problem. [CT06]". We will show that

this statement is based on a single trivial example and that it is unsuitable for the justification of such a claim.

In our discussion, we will focus our attention on the central portion of the Lévy measure representing the small jumps. We note that for small jumps, the data is largely irrelevant in determining the jump intensity of the optimised Lévy measure. The insensitivity can be explained by the ambiguity between small jumps and the diffusion component. This aspect was discussed earlier in Section 6.2.2.2, where we showed that initial points representing small jumps are not "motivated to move" away from their initial positions.

The insignificance of the small jumps is also reflected in the low values of the gradient associated with these points. Our experiments suggest that near-zero jump sizes are, in essence, meaningless in jump-diffusions. The attractive peaked shape of the Gaussian distribution from a Merton prior may be aesthetically pleasing, but in reality, it contains no useful information near its peak.

We argue that the selected prior measures that Cont and Tankov used as evidence of the "insensitivity to exact values of the prior" are somewhat inadequate. The two prior measures in their first experiment, shown in Figure 7.1 and labelled A and B, are essentially identical in terms of pricing substance.

The difference in the height of the peaks in the prior measures may seem impressive to the untrained eye, but unfortunately, this change is insignificant as it falls in the small-jump area. Simply changing the intensity of the somewhat irrelevant jumps near the centre of the prior Lévy measure is a rather unconvincing "parameter change" to the prior.

Any deduction based on these two prior measures with insignificant differences cannot be considered valid evidence for the insensitivity to the choice of prior parameters.

The potential role of initial points in shaping solutions

We observed a pattern in the optimisation results, that would seem consistent with the presence of some "assistance" from initial points to generate attractive results.

The choice of initial points for "Test A" and "Test B" in Figure 7.1 potentially influenced the above-mentioned result even further. The final result would likely have remained stationary at these initial points because the optimiser would be less inclined to move these insensitive points.

We once again point out the possibility that their selection of initial points may have contributed to the attractive shape of "Test B". Observe how the portion near the point marked "X", representing near-zero jump sizes, ended up quite close to the curve in "Test A". In the result of their second experiment, shown in Figure 7.2, they also present a rather pleasing shape in the central regions of their "Test B" (shown in yellow). If we consider the insensitivity of the central points to the data and the location of the prior well below the results, one could wonder: how did the section marked as "X" end up there?

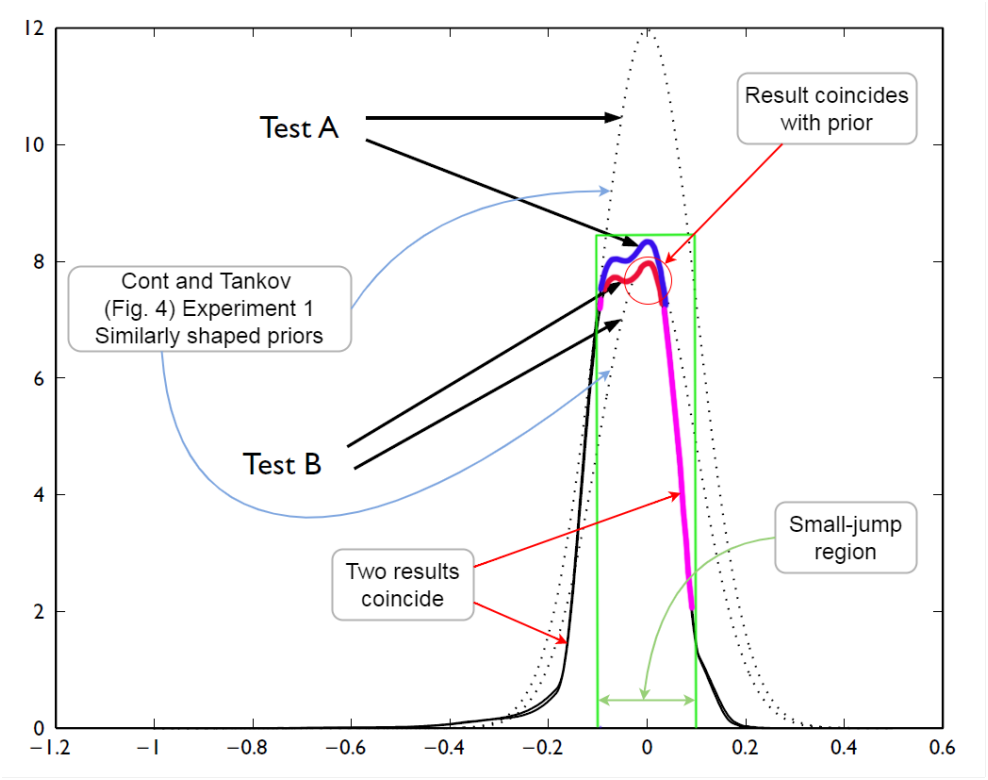

Figure 7.1: Similarly shaped prior measures (own markings in colour, source [CT04] Fig 4)

It is also interesting to note how the "Test A" reference results for both experiments coincide with the prior near the peak. This usually indicates either excessive regularisation or rather "accommodating" initial points selected to coincide with the prior and possibly the price-generating Lévy measure.

More transparency regarding the price-generating measure and initial points selected for their empirical analysis would have given their research more credibility.

Financial Modelling with Jump Processes [CT03]

In the book, *Financial Modelling with Jump Processes* [CT03], the authors state that their regularisation procedure "removes" the problem of multiple solutions associated with the ill-posed calibration problem.[4]

In other words, the authors claim that their regularisation procedure can completely eliminate the problem of ill-posedness, including the problem of non-uniqueness of

[4]See the text in the last paragraph at the bottom of page 433 [CT03].

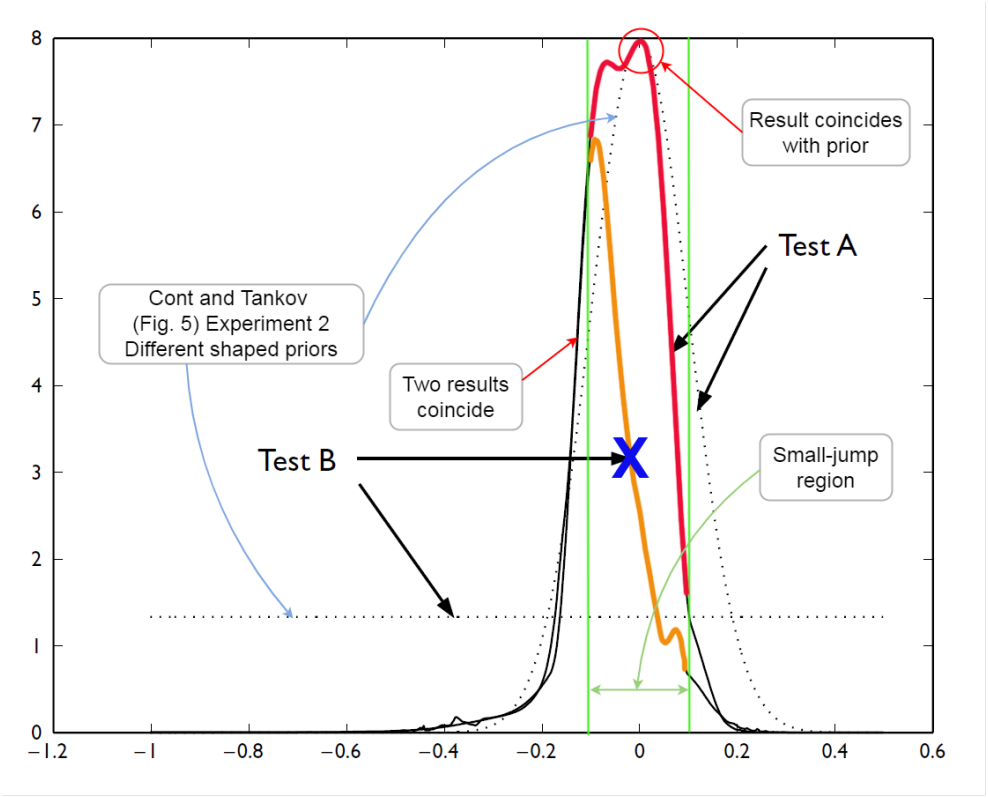

Figure 7.2: Differently shaped prior measures (own markings in colour, source [CT04] Fig 5)

solutions. We know from our results and the results of other researchers mentioned in Section 7.2 that this statement is inaccurate.

Moreover, in their own research in the two papers discussed above [CT04, CT06], Cont and Tankov were more conservative in their claims: They chose to avoid any explicit claims about eliminating ill-posedness, and toned down discussions on the issue of multiple solutions in their procedures.

Subsequent to the literature discussed above, Rama Cont made statements confirming that ill-posedness and non-convexity remain problematic, even after regularisation. These statements will be discussed next.

Constructive statements in later work

In a mini-workshop on solving inverse problems [HR06], Rama Cont provided a more unambiguous statement about the impact of their regularisation procedures, by stating: "Also, the regularized functional is still not convex so uniqueness of the regularized solution and convergence of gradient-based methods are not obvious." [Con06a].

In the overhead slides of the lecture titled, *Inverse problems in option pricing: a statistical approach using minimal entropy random mixtures* [CLD13], Rama Cont made a similar statement about their regularisation procedure. He stated that there are numerical difficulties even after regularisation due to the non-convex optimisation problem. He described the output of the process as a "point estimate", representing one of many solutions.

Chapter 8

Conclusion

8.1 Conclusion

Our study demonstrates that relative-entropy-based regularisation is not a particularly effective solution for managing the ill-posedness associated with the non-parametric recovery of a discrete Lévy measure from options prices. Ill-posedness is still a significant issue, and the process requires user intervention, so the problems encountered in other traditional (un-regularised) optimisation problems are not eliminated.

More specifically, we found problems relating to:

- The uniqueness of solutions.
- The sensitivity to initial points.
- A dependence of solutions on the selected prior measure.
- The stability with respect to the amount of regularisation.
- The stability with respect to changes in market prices.

These practical problems agree with observations reported by other independent researchers [Huh19, GS13].

We believe that these problems are related to the persistent non-convexity of the regularised objective function at reasonable levels of regularisation.

In addition, we believe that the problem is further exacerbated by

- a measure of non-convexity observed in the discrete form of the relative entropy function and by
- an inclination of the optimisation problem to be poorly scaled and, therefore, susceptible to ill-conditioning.

Gradient-based optimisation problems are prone to ill-conditioning when the points in the solution vector vary significantly in their sensitivity to the optimisation "step" size.

Although we observed some deficiencies in the procedures examined in this study, we are not questioning the mathematical validity of the work done by Cont and Tankov [CT04, CT06]. They furnish reasonable mathematical proofs substantiating the convexity of the relative entropy penalty term in the *continuous* case [CT04, CT06].

However, we question the practical application of their procedures at reasonable levels of regularisation and the validity of some of their empirically motivated conclusions about the capabilities of their procedures. The limitations of their procedures were also not clearly stated in their articles.

In particular, we found it rather difficult to understand from their articles exactly what the authors claim to have achieved through their procedures in terms of dealing with the problem of ill-posedness. Furthermore, we were not convinced by their demonstrations relating to the unimportance of the prior measure in determining the results. Moreover, their analysis and discussions omitted essential aspects of optimisation such as the sensitivity to initial points and the uniqueness of solutions.

We found that with the careful selection of a number of parameters one *can* harvest attractive-looking results. The intricate relationship of the parameters makes it impractical to specify a dependable strategy to automate the process of consistently extracting good results. An empirical approach seems to be the most effective strategy for obtaining acceptable results. Although brute force can be applied to an automated search for better candidate solutions by wrapping a global optimiser around the core optimiser, the additional computational burden becomes problematic. Even with such extreme measures, good results are not guaranteed.

On the other hand, the sensitivity of results to the selection of the initial points can be (unfairly) exploited to generate arbitrarily good-looking results. In other words, the optimisation outcome can be readily manipulated to a chosen appearance by selecting initial points reasonably close to the desired result, particularly in the central parts of the Lévy measure representing small jumps.

One such strategy could be to set the initial points to coincide with the prior measure. This strategy can result in nice-looking, smooth results in exceptional cases, such as when the prior happens to be close to the price-generating measure. However, as demonstrated in our experiments, initialising to a prior that poorly reflects the properties of the embedded Lévy measure is bound to generate meaningless results.

Similarly, excessive regularisation can create the impression of stability and smoothness, particularly in the exceptional case where the prior measure closely matches the price-generating measure. In practice, this approach amounts to achieving no more than a mere parametric calibration to the prior measure. In cases where the true shape of the generating measure is not well captured by the prior measure, this approach will fail to deliver useful results. Excessive pull towards the prior will significantly impair the influence of the data on the recovered Lévy measure.

We have also shown that reducing the regularisation to a lower level, allowing the optimiser some freedom to deviate from the prior, can yield results that “look reasonable”. However, the shape of such results can vary significantly depending on the initial

points used to seed the optimiser. Moreover, we have shown that for a given level of regularisation yielding reasonable-looking results, a slight change in initial points can lead to a completely different result. This small change in initial points could lead to a meaningless result with undesirable characteristics. In essence, the procedure yields many possible results due to its persistent ill-posedness, and some of these results are entirely absurd.

Consequently, it remains the practitioner's duty to filter through results from this ill-posed problem and to manually select "feasible-looking" results. This selection would then tend to promote the practitioner's desired view of the market. Therefore, the exercise becomes subjective because there are typically many potential jump profiles that could look plausible.

The most significant finding in our study indicates that it is generally not possible to recover a price-generating Lévy measure from its own options prices. Our tests found that we were unable to recover the Lévy measure from a set of its options prices when applying regularisation and using prior measures from calibrated parametric jump-diffusion models. Even when providing a rather wide and dense set of options prices, we failed to recover any Lévy measures that reasonably reflect the jump dynamics of the generating measures.

Our study suggests that there seems to be significant scope for improvement in the non-parametric recovery of Lévy properties. The available procedures all exhibit their own deficiencies [Huh19], thus creating an opportunity for further study to find improved solutions in this domain. For the time being, practitioners will need to treat non-parametric results with caution and consider supplementing their market views with findings from parametric models.

Although there is room for improvement in the quality of the non-parametric results, one should not expect the elimination of ill-posedness. In studies by Cont et al. several years since the primary articles, their remarks suggest that one should treat these regularised optimisation results as "point estimates" or "candidate solutions" where one solution is selected from many [CD13, CLD13]. They claim that it is unreasonable to expect this type of regularised calibration problem to uniquely define the pricing rule [CD13, CLD13]. They argue that the constraints of this optimisation problem are, in general, insufficient to pinpoint a unique model since we are dealing with "only a finite number of options prices" and we typically use European options which "constrain only the marginal laws of the process" [CD13, CLD13].

The non-parametric Lévy models studied in this dissertation constitute a sophisticated and flexible model class with the potential to capture many important aspects for modelling realistic price-paths. However, this study shows that the regularisation procedures employed were ineffective in addressing the problem of ill-posedness. The failure to address the ill-posedness is, in essence, a failure to address the risks associated with this model class, thereby increasing the risk of mispricing.

Part IV

Appendices

Appendix A

Notation

Symbol	Description
A	Multi dimensional representation of the diffusion parameter, σ, in a Lévy triplet (A, ν, γ).
(A, ν, γ)	Multi dimensional Lévy triplet.
α	Regularisation parameter for the penalty term weighting $\alpha H(\theta)$. Used in the context of regularisation in Chapters 4, 5 and 6.
α	Damping factor used in Fourier transform pricing for example $\sinh(\alpha k)$ or $e^{-\alpha k}$. Used in Chapter 3 in the context of Fourier pricing formulae..
$b(\nu)$	Drift component (without any lumped compensation, as opposed to γ which may include a deterministic compensation term for small jumps).
$\mathcal{B}(.)$	Borel-algebra
$C_{BS}^{\Sigma}(k)$	Black-Scholes call price with volatility parameter Σ. Used in the Black-Scholes adjusted transform pricing formula in Section 3.2.5.
C_j	Market price point j.
C_j^{θ}	Model price point j.

Symbol	Description
C_T	Call price at expiration.
c_T	"Modified" call price at expiration. (Carr and Madan pricing derivation [CM99].)
$\hat{c}_T(u)$	Pricing function for modified call prices in the Fourier domain. Fourier inversion gives c_T.
Δx	The uniform spacing between adjacent discrete jump sizes $(x_j - x_{j-1})$.
D_h	Directional derivative in direction h.
$\delta(x)$	Dirac delta function.
$E[.]$	Expected value.
$E_{\mathbb{P}}[.]$	Expected value under probability measure $\mathbb{P}$.
$\mathcal{E}(\mathbb{Q}_\theta \mid \mathbb{P}_0)$	Relative entropy of measure $\mathbb{Q}_\theta$ with respect to $\mathbb{P}_0$.
$\mathcal{F}f(u)$	Fourier transform of the function $f(u)$.
$\mathcal{F}^{-1}f(x)$	Inverse Fourier transform of the function $f(x)$.
$\mathcal{F}_t$	Information available at time t.
γ	The deterministic component of a Lévy process such as the drift in a jump-diffusion process.
$H(\theta)$	Regularisation penalty function for parameter vector θ.
H	Pay-off function, for example $H = (S_T - K)^+$.
1_A	Indicator function $1_A(x) := \begin{cases} 1 & \text{if} \quad x \in A, \\ 0 & \text{if} \quad x \notin A. \end{cases}$

Symbol	Description
$J(\theta)$	Objective function value for parameter set θ.
K	Strike price.
k	Log-strike price: $k = \log(K)$.
$\log(x)$	Natural logarithm.
$\log_{10}(x)$	Logarithm base 10.
λ	The intensity of a Poisson process and the jump-intensity of a compound Poisson process.
M	Number of market prices in the objective function.
N	Number of points in a discrete Fourier transform.
$\mathbb{N}, \mathbb{N}_0$	Non-negative integers.
N_ν	Number of points in a discrete Lévy measure.
N_t	Poisson process
$N\left(\mu, \delta^2\right)$	Normal distribution with mean μ and variance δ^2.
$(\Omega, \mathcal{F}, \mathbb{P})$	A probability space with scenario set Ω, equipped with σ-algebra $\mathcal{F}$ and $\mathbb{P}$ is a probability measure on the measurable space $(\Omega, \mathcal{F})$.
$\mathbb{P}\mid_{\mathcal{F}_t}$	Probability measure $\mathbb{P}$ restricted to filtration $\mathcal{F}_t$
$\nu(x)$	Lévy density.
ν, $\nu(dx)$ or $\nu(x)dx$	Lévy measure.
ν_j	Point j of the discrete Lévy density representing the "intensity" of jump size x_j.

Symbol	Description
ν^p	Prior measure.
$\phi_X(u)$	Characteristic function for random variable X.
$\psi(u)$	Characteristic exponent.
r	Risk free interest rate.
S_0	Price of the underlying asset at $t = 0$.
S_t	Price of the underlying asset at time t.
S_T	Price of the underlying asset at expiration of an option.
$\hat{S}_t$	Discounted price of the underlying asset at time t.
s	Log-price of the underlying: $s = \log(S)$
Σ	Black-Scholes volatility parameter Σ, for the Black-Scholes call price, $C_{BS}^{\Sigma}(k)$, in the adjusted transform pricing formula (Section 3.2.5).
$\tilde{\sigma}_j$	Implied volatility associated with price C_j.
σ	Diffusion parameter for the Lévy process with diffusion component σW_t.
σ_p	Diffusion parameter associated with the prior measure ν^p.
(σ, ν, γ)	Lévy triplet.
T	Time of expiration of an option.
t	Time.
θ	General representation of a parameter set for a pricing model.

Symbol	Description
θ^*	Optimal parameter set for a pricing model.
$V(\tilde{\sigma}_j)$	Black-Scholes *Vega* calculated using the implied volatility $\tilde{\sigma}_j$.
w_j	Weight associated with the market-model price pairs in the objective function.
W_t	A Wiener process which can be scaled by σ to form the Lévy process diffusion component, σW_t.
x_j	The jump size represented by point j of the discrete Lévy density element ν_j.
X_t	Random process.
Y_j	The size of the jth jump in a compound Poisson process.
z_T	Out-of-the-money time-value option price.
$\widetilde{z}_T$	Black-Scholes modified option price. (Section 3.2.5).
$\zeta_T(u)$	Pricing function for Out-of-the-money time-value prices in the Fourier domain. Fourier inversion gives z_T.
$\widetilde{\zeta}_T(u)$	Pricing function for Black-Scholes modified option prices in the Fourier domain. Fourier inversion gives $\widetilde{z}_T$. (Section 3.2.5)

Appendix B

Transform pricing

B.1 Transform pricing and characteristic function

B.1.1 General expression for the Fourier transform

In this dissertation we chose to use the format described in Definition 2.4.1. This is one of several formats used for the Fourier transform and matched inverse Fourier transform.

Weisstein [Wei22] presents a general form for the Fourier pair as

$$F(\omega) = \sqrt{\frac{|b|}{(2\pi)^{1-a}}} \int_{-\infty}^{\infty} f(t)\, e^{ib\omega t}\, dt$$

and

$$f(t) = \sqrt{\frac{|b|}{(2\pi)^{1+a}}} \int_{-\infty}^{\infty} F(\omega)\, e^{-ib\omega t}\, d\omega,$$

where a and b are arbitrary constants.

Typical examples of values for (a, b) [Wei22] encountered in various disciplines are:

- Probability theory $(1, 1)$ (used in this dissertation)
- Modern physics $(0, 1)$
- Classical physics $(-1, 1)$
- Mathematics $(1, -1)$
- Signal processing $(0, -2\pi)$

Similarly, different forms can be used for the discrete Fourier transform,

$$y_j = \sum_{k=0}^{N-1} e^{\frac{2\pi ijk}{N}} x_k \quad j = 0, 1, \ldots, N-1, \tag{B.1}$$

and discrete inverse transform,

$$x_j = \frac{1}{N} \sum_{k=0}^{N-1} e^{\frac{-2\pi ijk}{N}} y_k \quad j = 0, 1, \ldots, N-1. \tag{B.2}$$

Generally, the sign of the exponents can be swapped and the normalisation factor of $\frac{1}{N}$ in (B.2) can be changed, provided that the product of the normalisation factors of the DFT and iDFT is equal to $\frac{1}{N}$.

B.1.2 Complex conjugate symmetry of a characteristic function

Since the complex conjugate of a characteristic function has the property that $\overline{\Phi_X(u)} = \Phi_X(-u)$, the symmetry of a characteristic function about $u = 0$ can be utilised to simplify the numerical Fourier inversion by integrating over the positive half of the real axis.

B.1.3 Variations in transform variables and formulae

When referring to the characteristic functions and distributions encountered in statistics their one-to-one relationship allows one to unambiguously refer to them by name of the distribution. However, there is no uniform standard for formulating model-specific characteristic functions and pricing formulae as they could vary with the formulation of the specific pricing problem and authors adopt the approach that best suits their mathematical requirements.

Additional care is therefore required when one is dealing with a *model* name in options pricing, as there can be differences in the formulation of the random variables, their characteristic functions and the corresponding Fourier pricing formulae for the same model. The variation in the formulation across texts has the consequence that some inspection is required, and possibly some adjustment before a characteristic function from one text can be applied to the Fourier pricing formula of another.

The following three articles, each with their own variant of Fourier pricing formulae are of specific interest to this document:

Carr and Madan: In their article *Option valuation using the fast Fourier transform* [CM99], they present Fourier pricing formulations that are compatible with the FFT. These formulations have been widely adopted, sometimes in a slightly modified form, to solve various path-independent calibration problems.

Cont and Tankov: In their article *Non-parametric calibration of jump-diffusion option pricing models* [CT04], they apply the principles of the above-mentioned Carr-Madan article to a regularised non-parametric calibration problem.

Belomestny and Reiss: In the article *Spectral calibration of exponential Lévy models* [BR06], the authors again apply the Carr-Madan principles to their calibration procedure.

Examples in Table B.1 on the next page show different random variables selected by the authors to be characterised by the characteristic functions in a transform pricing procedure. The corresponding formulations of the call price by expectation for the same random variables are shown in Table B.2. The equivalent Fourier pricing formulae in time-value format are shown in Table B.3.

There are slight differences in the formulation for the time-value, $z_T(k)$, of the option prices used to determine the inversion formulae. Carr-Madan use the out-of-the-money time-value as defined in Definition 3.2.2 as

$$z_T(k) = e^{-rT} E_Q \left[\left(e^k - e^s \right) 1_{s<k, k<0} + \left(e^s - e^k \right) 1_{s>k, k>0} \right]. \tag{B.3}$$

Cont and Tankov define the time-value, $z_T(k)$, used in their inversion formula as

$$z_T(k) = e^{-rT} E \left[\left(e^{rT+X_T} - e^k \right)^+ \right] - \left(1 - e^{k-rT} \right)^+, \tag{B.4}$$

where the first term on the right is the expression for the call price of the option and the second term represents the (deterministic) intrinsic value $e^{-rT}\left(e^{s_0+rt} - e^k\right)^+$. Since an option price contains a time-value component and an intrinsic value component, the time value can be obtained by subtracting the intrinsic value as shown in (B.4).

B.2 Fourier transforms involving discontinuities

The inability of a Fourier series to adequately represent a function with discontinuities, even when increasing the number of Fourier terms, is known as the Gibbs phenomenon [Gib98]. This phenomenon is also applicable to Fourier transforms where the resulting function of the transform has a discontinuity as encountered in out-of-the-money time-value curve described in Section C.2.2 on page 149 and illustrated in Figure C.1 on page 150.

An analogous instance in Figure B.1 illustrates a simple example of a Fourier series with an increasing number of terms used to construct a saw tooth wave. One can see that the diagonal portion is reasonably well approximated as we increase the number of terms,

Author	**Random variable of the characteristic function**
Carr & Madan	The random variable is s_T, the log-price of S_t the terminal spot price. $s_T = \ln(S_T)$
Cont & Tankov	The stochastic process X_t, such that e^{X_t} is a martingale. Often a Lévy process described by the Lévy-Khinchin formula.
Belomestny & Reiss	The stochastic process X_t, such that e^{X_t} is a martingale. Often a Lévy process described by the Lévy-Khinchin formula.

Table B.1: Examples of formulation differences in the random variable described by the characteristic function.

Author	**Call price formulation as expectation**
Carr & Madan	$C_T(k) = e^{-rT} E\left[\left(e^{s_T} - e^k\right)^+\right]$, where $k = \ln(K)$.
Cont & Tankov	$C_T(k) = e^{-rT} E\left[\left(e^{X_T + rT} - e^k\right)^+\right]$, where $k = \ln(K)$.
Belomestny & Reiss	$C_T(x) = E\left[\left(e^{X_T} - e^x\right)^+\right]$, where $x = \ln(K) - rT$.

Table B.2: Examples of formulation differences in expectation pricing formula.

Author	**Out-of-the-money time-value Fourier pricing formula**
Carr & Madan	$\zeta_T(u) = e^{-rT}\left[\frac{1}{1+iu} - \frac{e^{rT}}{iu} + \frac{\phi_T(u-i)}{iu(1+iu)}\right]$
Cont & Tankov	$\zeta_T(u) = e^{iurT}\left[\frac{1}{1+iu} - \frac{1}{iu} + \frac{\phi_T(u-i)}{iu(1+iu)}\right]$
Belomestny & Reiss	$\zeta_T(u) = \left[\frac{1}{1+iu} - \frac{1}{iu} + \frac{\phi_T(u-i)}{iu(1+iu)}\right]$

Table B.3: Examples of formulation differences in Out-of-the-money time-value Fourier pricing formula.

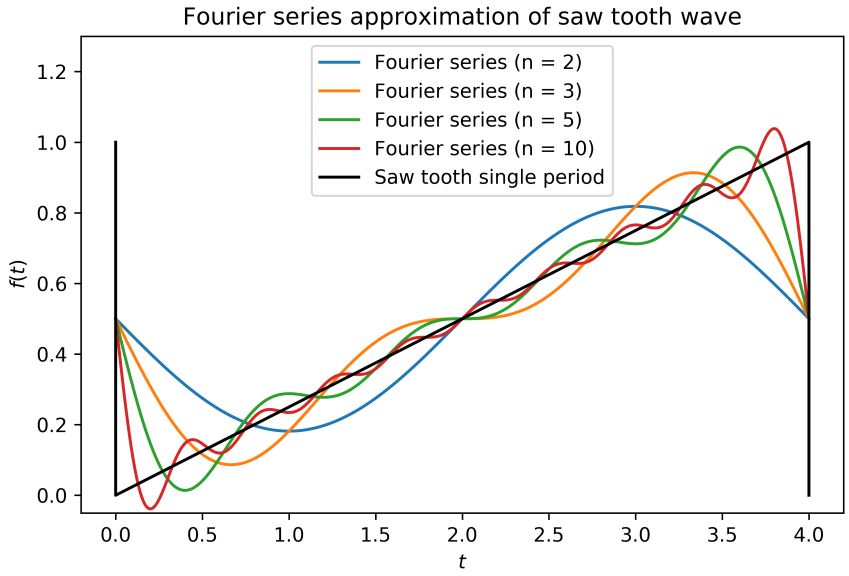

Figure B.1: Fourier series of a saw tooth wave with increasing terms over a single period.

whereas the corner portions (with sharp transition) are poorly rendered, needing a wider range of "frequencies", which in turn tend to increase the level of overshoot at the discontinuity. The following experiment involving the triangular distribution illustrates the Gibbs phenomenon for steep transitions in Fourier transforms.

The characteristic function of the triangular distribution can be used to generate its density function by Fourier transform. We demonstrate the recovery of the density functions by FFT of two triangular distributions. For this experiment we selected a reference triangular distribution and then chose a similar triangular distribution with a steeper slope and compared the performance of the FFT generated distributions against the known triangular distributions as shown in Figure B.2 on the following page. We used identical transform parameters and calculated the errors with respect to the benchmark distributions. The errors and log-errors are shown in Figure B.3 on page 144. From this experiment we can see that there is a noticeable increase in the error coinciding the non-smooth transitions and that this is significantly aggravated by the steepness of the transition.

Similarly, the steepness caused by the at-the-money discontinuity of our transform pricing curve in Figure C.1 is the reason for the significantly larger pricing errors observed near-the-money in our pricing tests in Section C.2.2. The increase in pricing errors near the money for various models can be seen in Figure on page 145, which represents the pricing errors resulting from the discontinuous Carr-Madan time-value curve. This can be compared with the much improved results from the Black-Scholes adjusted approach which eliminates the discontinuity shown in Figure on page 154.

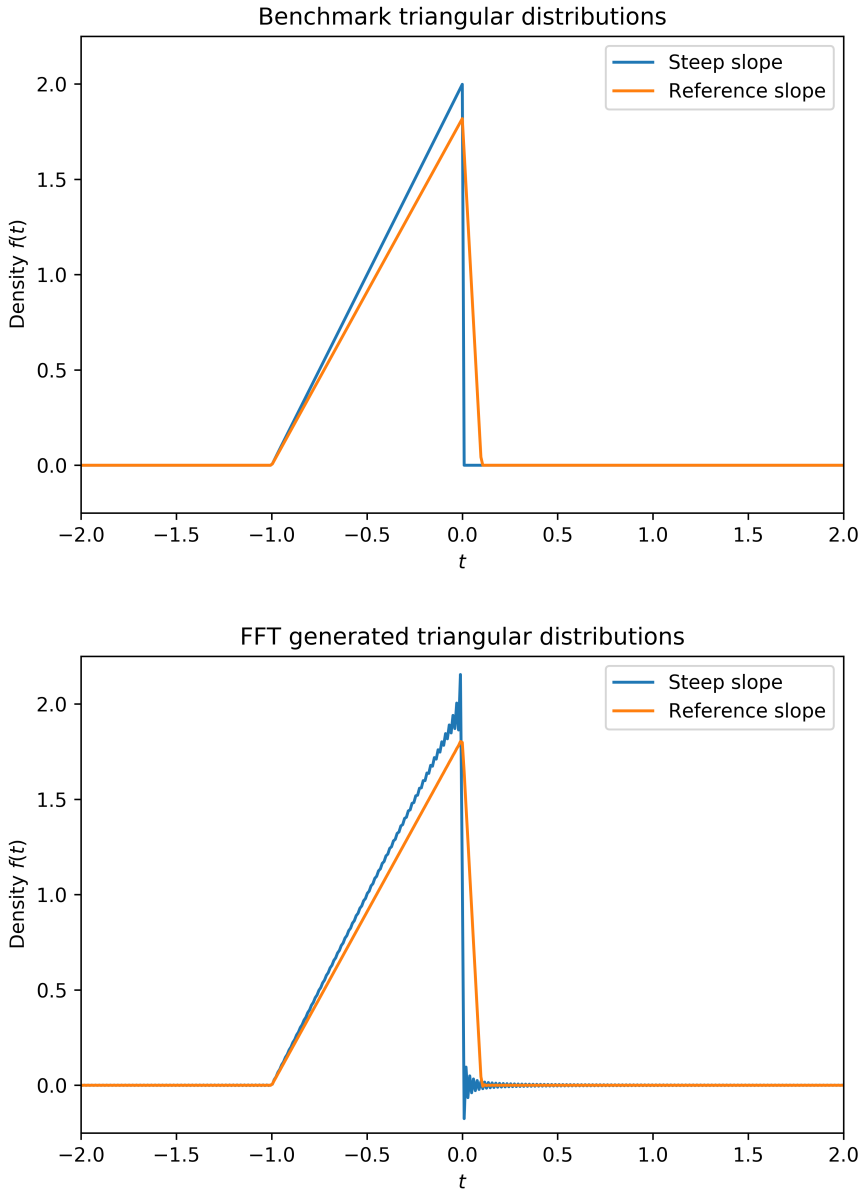

Figure B.2: Decline in FFT generated accuracy with a steeper slope.

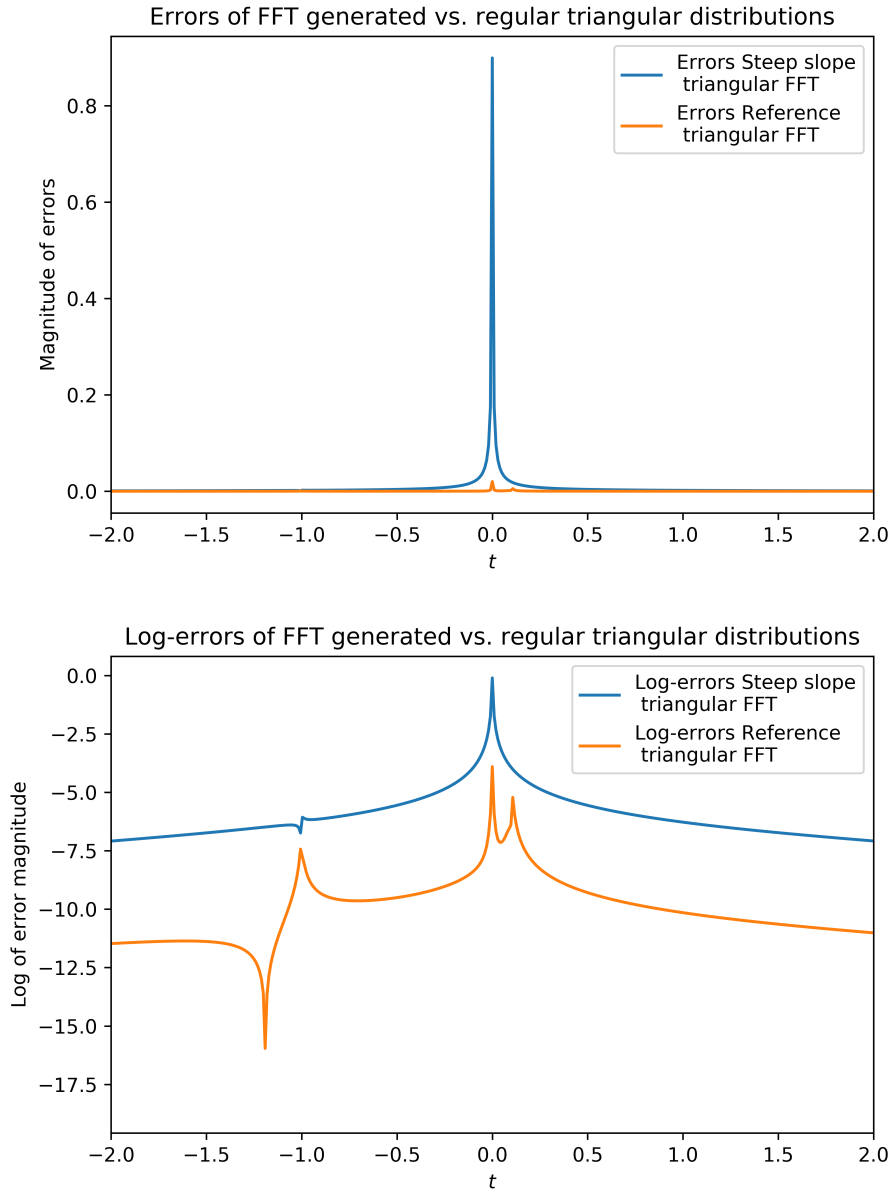

Figure B.3: Increase in the error of an FFT generated distribution with a steeper slope.

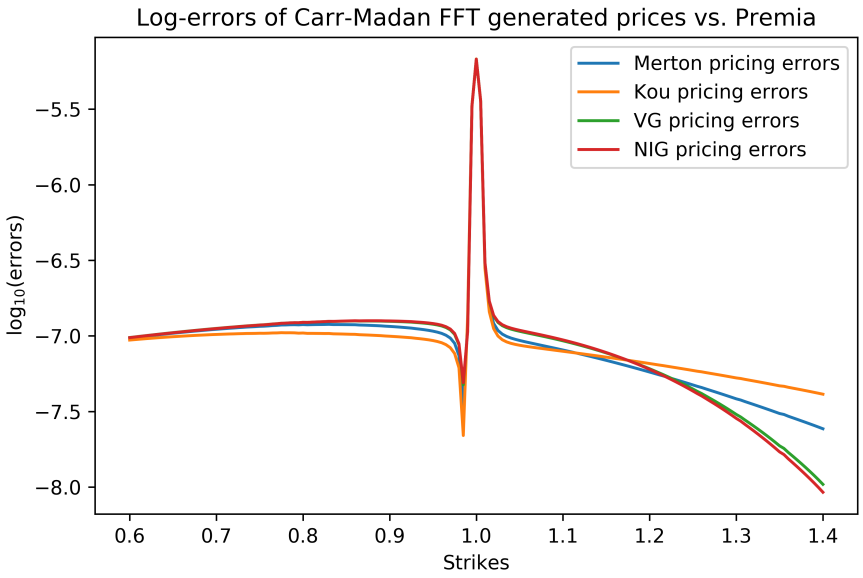

Figure B.4: Log-errors of transform pricing using the standard Carr-Madan time-value formula.

Appendix C

Pricing and model verification

C.1 Test strategy

In this appendix, we present a detailed examination of the performance of individual components used in the construction of a model to test the calibration procedures. This serves as more than a just simple verification of the integrity of the implemented algorithms, but demonstrates the importance of some behind-the-scenes parameters which can easily be overlooked. We will show that there are several detailed implementation decisions that a practitioner will need to make to achieve reliable performance.

Although it may seem reasonable to set up scripts cycling through various combinations of parameters for the entire model and observing the outcome of key results, this approach can easily get out of hand due to the large number of parameters which will multiply up creating an overwhelming volume of results to interpret.It is however possible to follow the principles of this approach when isolating individual components and looking at outcomes based on sweeping through a smaller set of parameters when examining the component of interest. In our experience the best evaluation strategy is to consider the influence of a maximum of three parameters simultaneously.

Our component level tests are arranged as follows:

- Pricing accuracy for differences in transform pricing variant and parameters
- Pricing performance with a discrete Lévy measure and related parameters
- Gradient computation performance and characteristics
- Regularisation performance and characteristics (in isolation)

This appendix deals with the component-level verification of the model implementation by performing specific tests designed for this purpose. The performance of the integrated set of components at carrying out non-parametric regularised optimisation is addressed in Chapter 6.

We begin our verification by testing and "calibrating" the transform pricing mechanism in the following section.

C.2 Pricing tests

A reliable pricing mechanism forms the cornerstone for proper model calibration. When implementing a Fourier pricing algorithm, it is important to remember that it is not a precise pricing mechanism, but instead an *approximation* which requires careful management of errors. In our discussion on parameter selection for Fourier pricing in Section 3.3 on page 48 we explained the relationship between the integration span, the number of discrete points and the log-strike spacing of the prices delivered. In essence, we need to "integrate" over a wide enough span of the characteristic function with an evenly spaced grid that is fine enough to control loss of precision resulting from discretisation errors. This grid spacing in turn has an inverse relationship with respect to the spacing of log-strike output prices. The log-strike price grid needs to be fine enough to control the interpolation errors encountered when translating options prices at the log-strike prices to options prices at (off-grid) market strikes.

Apart from grid parameter selection, it is also meaningful to consider other factors impacting on the pricing algorithm such as:

- the damping factor,
- the interpolation method used to obtain options prices at the market strikes, and
- the selection of the specific variant of the pricing algorithm.

Measuring the magnitude of pricing errors of the "calibrated" pricing algorithm against reference prices serves as a form of validation for the implementation of pricing mechanism and associated characteristic functions. Furthermore, these pricing errors provide a reference "best-fit" error magnitude for the measurement of pricing errors during the calibration process.

The non-parametric representation of Lévy models in the form of a discretised Lévy measure is another *approximation* that will be used extensively in following sections of this chapter. We will present the results of a number of pricing tests relating to this approximation.

In this section we present the results of a number of experiments selected to determine:

1. the range of acceptable values for the integration width, integration step size and log-strike step size by measuring the pricing errors as parameter values are changed in the form of a grid-search while pricing by FFT,
2. the maximum and mean pricing errors of our implementation of selected Fourier pricing algorithms applied to some parametric Lévy models versus reference prices for the same models,
3. the extent of pricing errors observed when pricing options using a non-parametric representation with of a Lévy measure acquired from a known parametric Lévy model versus the original parametric model.

Reference prices were generated using an installed command line version of *Premia 16* option pricing software, written and maintained by the MathRisk project. The software and addition information can be obtained from https://www.rocq.inria.fr/mathfi/Premia/.

C.2.1 Pricing-error presentation

We chose to measure our pricing precision in terms of the a *maximum absolute pricing error* across the range of normalised strike prices, as will be explained in the next paragraph. However, the percentage error with respect to the reference options prices provides the reader with some sense of the scale for the pricing precision and examples are provided for contextualisation purposes. Figure C.2 shows plots of the absolute error (top) and the corresponding percentage error plot at the bottom. One can see that for the Merton model, the highest percentage error occurs at a strike price of approximately 1.3 with an error level of about $10^{-4}\%$.

Model calibration typically relies on the sum of the square of the errors (SSE) to determine the closeness of model prices to the reference market prices). In other words, the price *difference* is important. For testing the pricing engine against reference prices, a *maximum absolute pricing error* across the (normalised) strike range with respect to (normalised) reference prices was found to be a sensible measure of pricing precision. In other words, normalising all prices with respect to $S_0 = 1$ and then choosing the maximum observed price deviation, proved to be a conservative approach to identifying the pricing capability.

Furthermore, observing the absolute pricing error against strikes provides insight which may be useful in tuning the pricing parameters. In the example is shown in Figure C.2, one can see that the maximum absolute pricing error for the Merton model occurs at a normalised strike price of approximately 0.9 with an error in the region of 10^{-7}.

In addition, we also considered the SSE when comparing repeated pricing (with various parameters) over the same (normalised) strike range against reference prices. In general, we found that in cases where the maximum absolute pricing errors were under control, there was not much additional benefit in analysing the SSE. The SSE does however provide an indication of the expected limit to calibration precision.

Since it is insightful to see the pricing error as a percentage of the option price, we included the percentage errors mainly for reference purposes. However, expressing the pricing errors as a percentage error with respect to the option price makes it difficult to identify the errors in regions where the intrinsic value is large. For example at low strikes, call prices have a large intrinsic value which conceals the true magnitude of the pricing error, as shown in the bottom graphs of Figure C.2. If the same call price at a low strike had been converted to the equivalent put price, the percentage error would be much larger.

One could remove the intrinsic value by using percentage prices of out-of-the-money time-value options, but we are satisfied that the above-mentioned maximum absolute error tests provided good error management across the strike range.

C.2.2 Pricing variants and parameters

In the early stages of the model development for this study, the initial pricing algorithm parameters were based on values found in readily available literature. The original transform pricing article by Carr and Madan [CM99] for example gives an indication of some of the parameters that they used. Similarly, in the text by Hirsa [Hir12] we found some values used in his assessments of transform pricing techniques. Using these as seeding values and applying some adjustments, we were able to verify a satisfactory level of functionality in the pricing engine. Verification was performed by testing against the Black-Scholes pricing formula and a few samples of parametric Lévy model prices with known parameters found in literature [GS13, Hir12, KW12, Sch03]. We further verified the pricing algorithms by testing the recovery of some parametric Lévy model parameters from a set of prices using a least squares optimisation[1].

The initial test results for the pricing mechanism proved adequate to allow us to proceed with further development of the model. The basic pricing mechanism was capable of using either of the two Carr and Madan methods [CM99] described in Sections 3.2.3 and 3.2.4 along with regular linear interpolation. Pricing errors with respect to a normalised strike set (with respect to S_0) appeared to be less than 1.0×10^{-5}.

Much later in the development of the modelling software and after performing several sets of regularised non-parametric calibrations, the pricing mechanism was re-visited with a more meticulous approach. The consistent observations of somewhat larger pricing errors near-the-money was interpreted as a need for an interpolation algorithm more sophisticated than the basic linear interpolation that was initially implemented. The abrupt change in slope near-the-money can be seen in our earlier example of out-of-the-money time-value prices in Figure 3.1 on page 46. An implementation using cubic splines to interpolate the prices led to a marginal increase in pricing accuracy near-the-money. We also tested an approach of performing the pricing interpolation on call prices instead. The much smoother curve obtained by converting the time-value grid points to call price points prior to interpolation still did not yield satisfactory pricing near-the-money.

A closer inspection revealed that the unsatisfactory prices near-the-money relates directly to the steepness of the out-of-the-money time-value curve in this area. As one can see from the above-mentioned figure, the out-of-the-money time-value graph has a cusp at $K = 1$. More importantly, there is usually as discontinuity at this point, since in general the call price and the put price are not equal at $K = S_0$. From put-call-parity,

$$c - p = S_0 e^{-qT} - K e^{-rT},$$

we see that for the put and call prices to be equal, we require that $r = q$. The additional steepness introduced by the discontinuity at the cusp, illustrated in Figure C.1 on the following page, significantly complicates the transform pricing. As mentioned

[1] The results in the article, *A moment matching market implied calibration,* by Guillaume and Schoutens [GS13] provided a useful selection of parameters for some parametric Lévy models. The authors kindly assisted by providing the market data used in their study which we applied to our early parametric calibration tests.

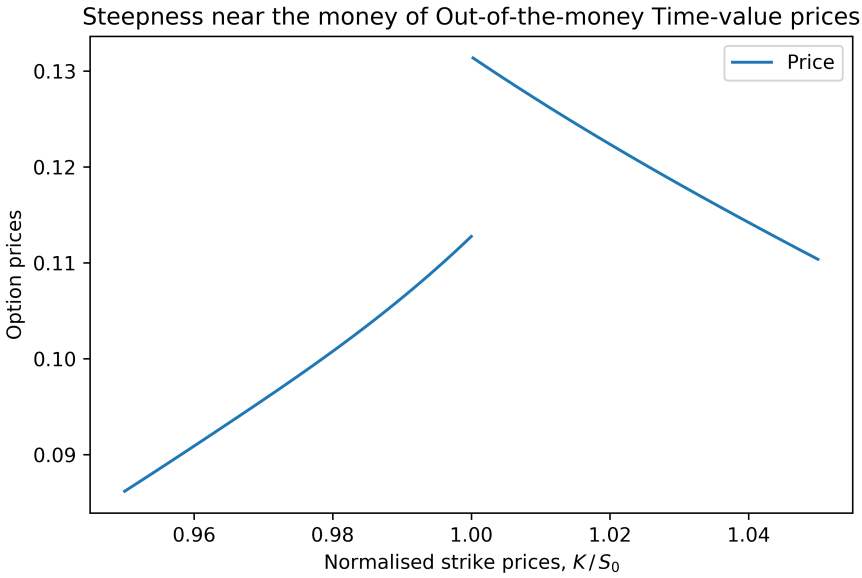

Figure C.1: OTM TV Discontinuity

in Section 3.3.1, the additional "high frequency" components introduced by any sudden transitions increase the truncation errors. This difficulty is known as the Gibbs phenomenon [Gib98] and is illustrated and discussed in Appendix B.2 on page 140.

The identification of the origin of the poor pricing led to some brute-force attempts at improving the pricing using the Carr-Madan time-value formula with various damping factors, larger FFT grids etcetera. Although a marginal improvement was observed as a result of adjusting the parameters, the magnitude of the improvement did not justify the additional computational burden required to achieve it. This led to the search for a a more appropriate improvement. The Black-Scholes adjusted pricing formula proposed by Tankov and Voltchkova [TV09] discussed in Section 3.2.5 completely resolved the issue. The results of this improvement will be presented in the following section.

C.2.3 Pricing relative to reference prices

In this section we summarise the results of selected experiments that were conducted to study the performance characteristics of transform pricing with a view to determine appropriate parameter sets for our pricing algorithms in preparation for their instrumental role in the calibration tests. This is an important exercise, because one or two poorly priced options prices have the potential to skew the calibration result clouding the evaluation of the calibration procedures. It is important to understand the capability of the pricing mechanism as it impacts directly on the effectiveness of the calibration.

Manual grid search for feasible FFT parameters

This rather extreme approach was undertaken after observing pricing errors while using an implementation of the Carr-Madan time-value approach which had previously demonstrated "acceptable" results. In essence we looked at pricing errors against a set of reference prices generated with *Premia* while sweeping a number of key parameters. In particular, we were interested in observing pricing errors with respect to changes in the following parameters:

- Damping factor
- Input grid step size
- Input integration width
- Output pricing grid step size (log-strike)
- Number of grid points

Since the last four items are related, as explained in Section on page 48, we varied only three parameters to perform our grid search.

We used prices generated by the closed form Merton formula over a reasonably wide normalised strike range of 0.6 to 1.4. The range of feasible parameter values together with the inspection ranges are shown in Table on the next page. The threshold of 5×10^{-5} was selected in determining the feasible region, based on observations from earlier work in this field. Transfer pricing using the Carr-Madan time-value formula achieved maximum pricing errors in the order of 1.1×10^{-5} using 2^{16} points. Adjusting the damping factor within a given set of grid parameters did not seem to deliver a noticeable impact on the magnitude of pricing errors.

Repeating the exercise using the Black-Scholes adjusted pricing formula achieved pricing within 1.7×10^{-7} using only 2^{12} points. This is a significant improvement when compared to the above-mentioned errors for the standard Carr-Madan time-value method. Table on the following page provides a summary of observed pricing errors for a number of variations on the grid size applied to these two pricing methods. We found that adjusting the Black-Scholes volatility parameter away from its optimal value (by a factor of 5 either way) did not have a material impact on observed pricing errors.

As discussed in Section on page 48, the characteristic function of a model and the model parameters will impact on the shape of the function to be transformed. Therefore the choice of transform parameters to balance the errors resulting from discretisation, truncation and interpolation will need to be examined routinely rather than blindly trusting rule-of-thumb values that had previously proven to be reliable.

Our verification against reference prices was performed by considering:

Parameter	**Inspection range**	**Feasible range**
Damping factor	1.1 to 1.8	Full inspection range
Input grid step size	0.05 to 10	≤ 1
Input integration width	40 to 10^6	≥ 3000
Output pricing grid step size (log-strike)	10^{-6} to 0.15	≤ 0.02
Number of grid points	$2^{10}$to 2^{20}	$\geq 2^{12}$

Table C.1: Investigation into feasible transform pricing parameters.

Pricing formula	**Grid size**	**Pricing error**
Carr-Madan time-value	2^{12}	5.0×10^{-5}
Carr-Madan time-value	2^{16}	1.1×10^{-5}
Carr-Madan time-value	2^{20}	5.3×10^{-6}
Black-Scholes adjusted	2^{12}	1.7×10^{-7}
Black-Scholes adjusted	2^{16}	1.7×10^{-7}
Black-Scholes adjusted	2^{20}	1.7×10^{-7}

Table C.2: Comparative pricing error magnitudes.

- Parametric Lévy models to include: Merton, Kou, VG and NIG
- A selection of interest rates
- A selection of maturities

A detailed discussion of the numerous verification tests performed for this study is beyond the scope of this dissertation, but noteworthy observations can be summarised as follows:

- Our implementation of the transform pricing algorithm applied to the Black-Scholes adjusted pricing formula produced good results across the spectrum of test cases. The largest pricing errors were less than 5.0×10^{-7}, which is in line with results in Table C.2.
- Pricing accuracy using the out-of-the-money time-value formula deteriorated with larger interest rates and longer maturities when compared to our results in the feasibility test, losing another two significant digits when compared to results in Table C.2.

C.2.4 Non-parametric relative to parametric prices

In this study the non-parametric representation of a Lévy model is based on a discrete approximation of the Lévy density by an evenly spaced vector representation of the jump sizes. Sections 5.3.2 and 5.3 describe the concepts as applied to the discretisation of a continuous Lévy density. This procedure is used to prepare a prior measure for regularisation and the optimised "solution" will also be presented in this non-parametric format.

In this section we investigate the selection of discretisation parameters and how they impact on pricing accuracy. The discretisation parameters present another example of an implementation decision by the practitioner. The parameters of interest are:

- The jumps size "bin" width, representing the range of jump sizes which will be represented by a single jump size vector point.
- The number of vector points representing the number of discrete jump sizes.
- Total width of the discretised Lévy density, which is determined by the preceding two parameters and determines the truncation points of the tails. The truncated mass of the tails are added to the mass of each end point respectively. Details are described in Section 5.3.3.

In our experiment to test the pricing performance of the non-parametric representation we selected a Merton model and observed the pricing errors of its discretised form relative to the original parametric model. We once again performed a systematic search for feasible values of the two key parameters by looking at the range of the parameters

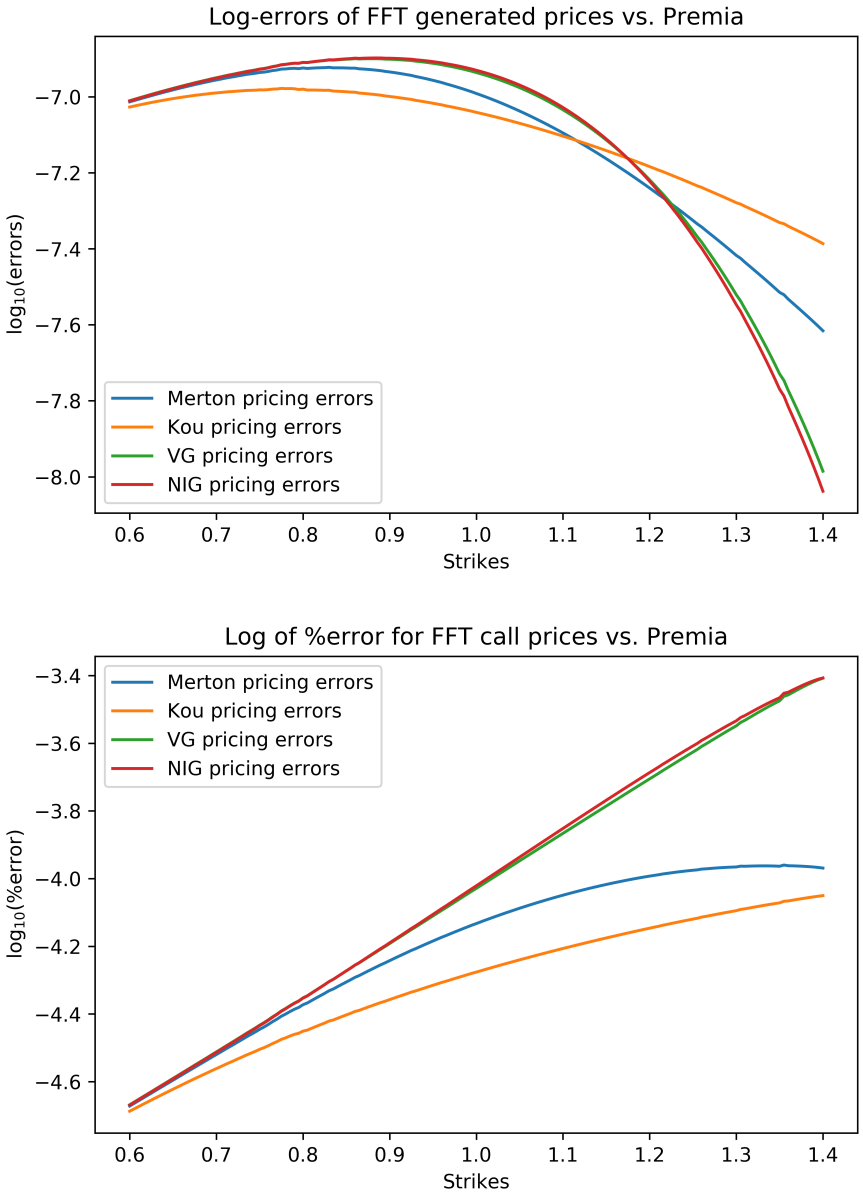

Figure C.2: Log-errors of transform pricing using the Black-Scholes adjusted method.

Maximum Pricing error threshold	Jump size vector size	Minimum jump-size span	Bin Width
1.0×10^{-4}	$\geq 2^6$	-1 to 1	0.03
1.0×10^{-5}	$\geq 2^8$	-1 to 1	0.008
1.0×10^{-7}	$\geq 2^{10}$	-2 to 2	0.004

Table C.3: Feasible discretisation parameters for a given error threshold.

which provide combinations that yield reasonable pricing errors. Our findings are summarised in Table C.3.

In general, the pricing errors observed when comparing a continuous model with its discrete equivalent are not of critical importance to the regularisation procedure of the calibration process. The role of the prior measure in the regularisation is primarily that of a "shaping" function to roughly nudge the optimiser towards a solution that more closely resembles our "view" of what it should be. It is, however, useful to have some benchmark of the parameter requirements to reproduce a continuous model to within a certain error threshold. The assurance that a discretisation parameter set provides a adequate approximation to a continuous model also provides an indication that it could be sufficiently rich to reasonably represent the jump dynamics of an arbitrary non-parametric Lévy density. The reasonable representation of the jump dynamics is essential when applying a retrieved Lévy measure to pricing path-dependent options.

In Figure on the next page we show that for a given jump size range of $[-2$ to $2]$ we can obtain a pricing error below 10^{-7} as we increase the number of points in our discrete Lévy density to 2^{10}. In Figure on the following page we show that the accuracy is limited by a too narrow range of $[-1$ to $1]$ and that a further increase in the number of points becomes ineffective.

C.2.5 A discrete Lévy density approximation applied to an infinite activity Lévy model

Although we limit our study to recovering non-parametric jump-diffusion models with *finite jump intensity* and non-zero diffusion component, we briefly examine the possibility of generating an analogous discrete approximation to a pure-jump model with infinite jump activity. A successful non-parametric approximation for an infinite activity model could give some indication of the feasibility of using such approximations to serve as a prior measure in our regularisation process.

We examine the approach of discretising the finite portions of the Lévy density, while constraining the high activity portions to a chosen maximum level. An appropriately determined diffusion parameter then serves to compensate for the resulting reduction in the small jump activity.

We selected a VG model, then chose an arbitrary maximum jump-activity level and expanded the parameter search procedure mentioned in Section on page 153 to search

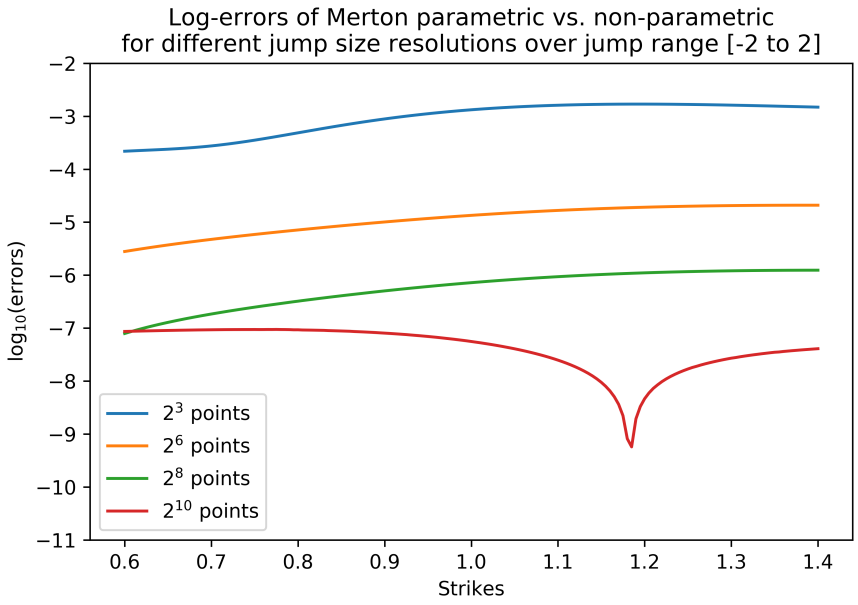

Figure C.3: Log of pricing error of various non-parametric resolutions with respect to the parametric Merton model.

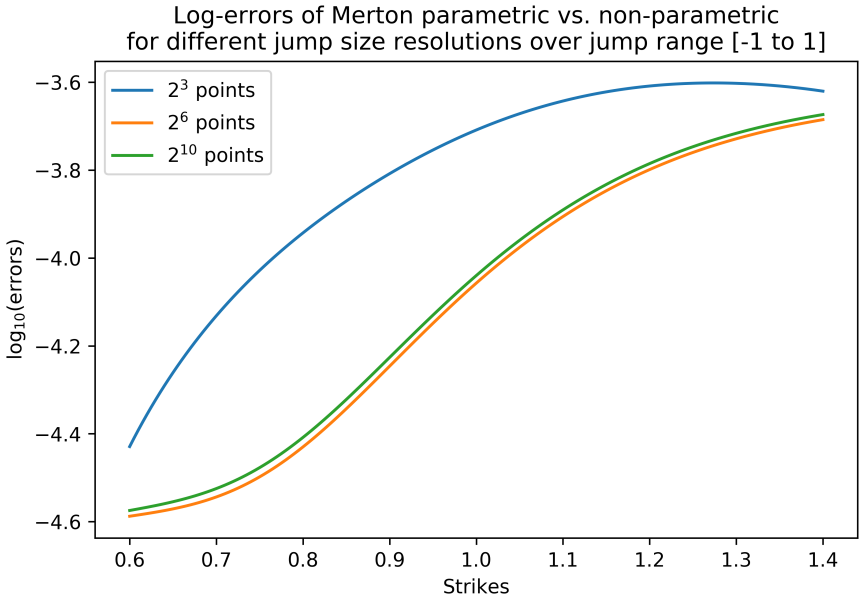

Figure C.4: Log of pricing error of various non-parametric resolutions with respect to the parametric Merton model for a narrower jump range.

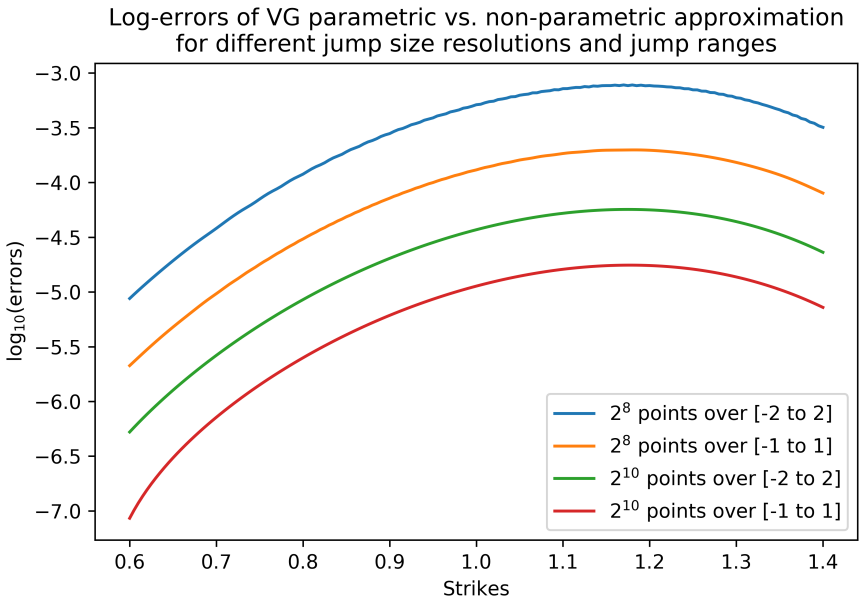

Figure C.5: Log of pricing error of various non-parametric resolutions with respect to the infinite activity parametric VG model.

for the discretisation parameters by adding a diffusion parameter to our parameter search list. Figure on this page shows the pricing errors of the discrete non-parametric jump-diffusion approximation against the parametric VG model for a few parameter sets. Based on a few superficial tests, we conclude that there is potential for generating prior measures from jump-diffusion approximations to infinite activity models.

C.3 Gradient tests

The gradient of the objective with respect to the discrete Lévy measure is an important component required for the efficient operation of the optimiser. The primary function of the gradient is that of directing the optimiser towards a solution. The gradient can also serve as an indicator that Lévy points have reached their optimal value. The optimiser will normally terminate either when the gradient is below its specified threshold or when the value of the objective is below its threshold. When the objective function includes a relative entropy regularisation term the gradient contains an additional component which directs the optimiser towards the prior measure. Calculating the gradient by using a formula has a significant improvement in the processing time over a numerical gradient approximation.

C.3.1 Gradient of the mean squared error component

We mentioned in Section 5.6.4 that the gradient is calculated with respect to each discrete Lévy point ν_i by first calculating the derivative, $\frac{\partial z(k_j)}{\partial \nu_i}$, of each price, $z\left(k_j\right)$, with

respect to that Lévy point. In this section we show examples of the formula calculated gradient components by applying Equation (5.6). We verify that the implementation of this calculation is reasonable by showing that this agrees with the finite difference numerical approximation obtained by applying a small change to the value of the selected Lévy point. Figure C.6 on the next page shows the derivatives and approximations with respect to three selected discrete Lévy points. For this illustration we selected points representing large positive jumps, large negative jumps and a point representing small jumps.

We see that the formula generated curves agree well with the corresponding curves generated by numerical differentiation.[2] It is further interesting to observe that:

1. Out of the money put option prices (at lower strikes) are more sensitive to changes in jump activity of large negative jumps.
2. Out of the money call option prices (at higher strikes) are more sensitive to changes in jump activity of large positive jumps.
3. The derivative curves are all non-negative.
4. Prices across the entire range of strikes are insensitive to changes in the activity level of small jumps.

The first observation can be explained by the fact that the put option will become more valuable in the event of a large negative move of the underlying, which becomes more likely with an increase in jump activity of large negative jumps. The option will be more likely to end up in the money or even further in the money if a large negative jump occurs.

Similarly, the second observation is explained by the increase in value of a call option as the event of a large increase in the underlying becomes more likely. The option will be more likely to end up in the money and possibly further in the money from the increased activity level of large positive jumps.

The non-negativity in our third observation is analogous to Black-Scholes pricing where an increase in both put and call prices occur with an increase in volatility. In other words, options prices increase with an increase in jump activity.

The final observation of price insensitivity to the intensity changes in small jumps is an obvious, but important property which impacts significantly on the non-parametric calibration results investigated in this study. We discussed the gradient behaviour of small jumps in Section 5.6.4.1. The consequence of this insensitivity will be illustrated and discussed further in Section 6.2.2 and will form part of our arguments in the critical assessment of the principal articles in Section 7.3.

Intuitively, it makes sense that the activity levels of larger jumps will have more impact on the price of the underlying than the activity levels of smaller jumps. Furthermore, the finite activity jump-diffusion model-class considered in this study already has a

[2] This was more carefully confirmed by the low error observed when examining the magnitude of difference between the two curves.

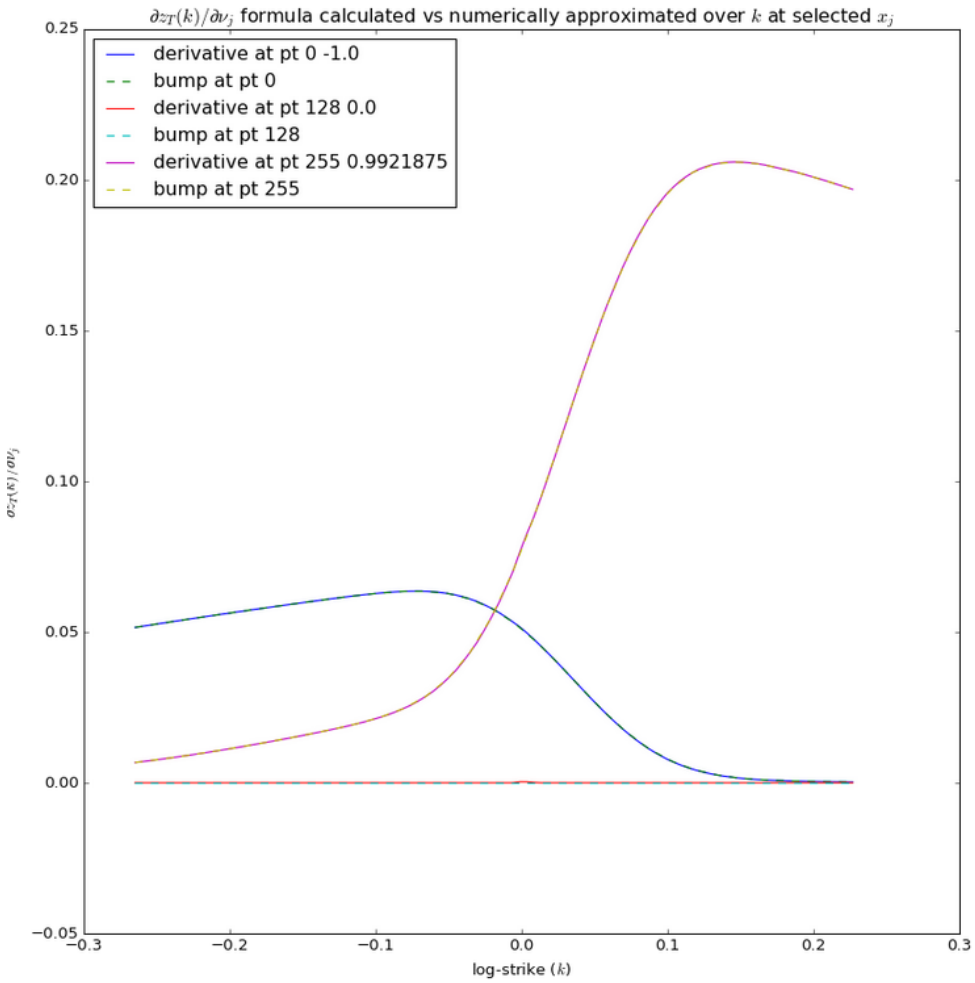

Figure C.6: Gradient plot for a few noteworthy jump sizes (without regularisation)

mechanism for representing the smaller movements in the form of a (non-zero) diffusion component. The diffusion component creates a form of ambiguity in the small movements when considered alongside the small jumps. This overlap of the diffusion and small jumps therefore causes a further reduction in the impact of the small-jump intensity on options pricing.

C.3.2 Gradient with regularisation

The regularisation term forms part of the regularised objective function by introducing a penalty associated with the deviation from the prior measure. Therefore the gradient of the regularised objective function will include a biasing contribution from this penalty term which will nudge the current iteration of the optimiser towards the prior.

We verified the relative entropy term of the gradient using a finite difference numerical approximation as we did with the error term described in the preceding subsection.

The next section contains additional information about the operational verification of the relative entropy component and its gradient.

C.4 Relative entropy verification

C.4.1 Discrete Lévy point sweep tests

After observing inconsistent results in some of the optimisation tests with varying initial Lévy points, we intensified the level of verification testing on our model as we searched for an explanation in a possible implementation error.

Among the verification tests performed to assess the implementation of our modelling software, was a test to investigate the relative proportion of influence between the error term and the regularisation term of the objective function. More specifically, we were interested to visualise the changes in value of the regularised objective and its gradient as the jump intensity of one point of the discrete Lévy density was swept across a range of values and through its optimal value. In addition we confirmed the observed composite gradient with a finite difference approximation to the gradient as the point was swept. We set up out experiment with the following steps:

1. Construct a simple discrete *reference pricing Lévy density* with a small number of points (shown in Figure on the following page).
2. Generate a set of reference prices from this *reference pricing Lévy density* (using an arbitrary diffusion component).
3. Choose a discrete *prior Lévy density* with its jump intensity points in reasonably close proximity to our reference-pricing density (shown in Figure on the next page).
4. Choose a discrete *test pricing Lévy density* with its jump intensity points near the prior density and the reference-pricing density
5. Select one point from the test pricing density and sweep its jump intensity across a range of values passing through the ideal value denoted by the reference density. For each position of this swept point record:
   (a) the pricing error component and regularisation component with the full objective function (shown in Figure on page 162).
   (b) the gradient of the pricing error component and the gradient of the regularisation term together with the composite gradient (shown in Figure on page 162 and Figure on page 163).
   (c) repeat the gradient calculations using the numerical approximation via finite difference.
6. Repeat for a different Lévy point and observe the changes relative to the jump size represented by the selected point.

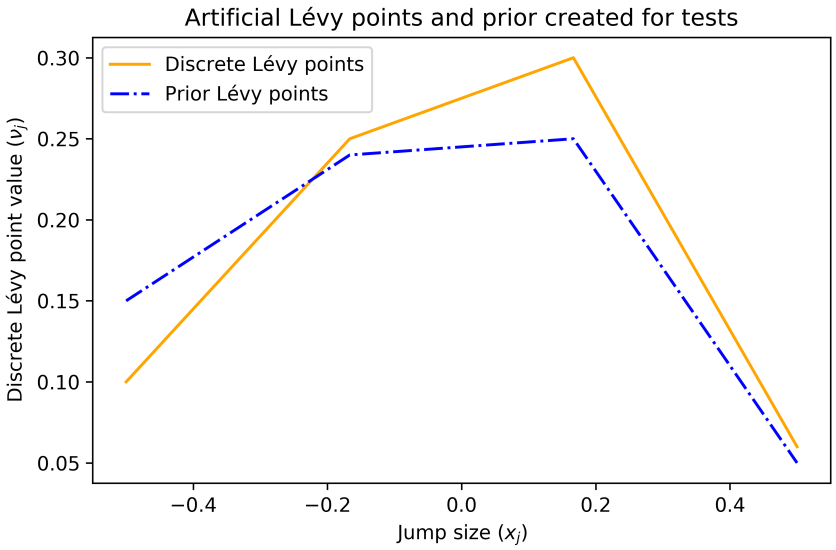

Figure C.7: An example of pricing Lévy points and prior Lévy points for a point sweep tests.

We found that this simple experiment provides an effective visualisation tool to observe the interrelationship between the components of the composite objective function and its gradient as the optimal solution is approached by a specific Lévy point. As expected, the relative entropy plays a significant role in directing the optimiser towards the prior measure, and conceivably towards a feasible solution. It is however significant to note that the relative entropy and its gradient provide the optimiser with a point-specific "suggestion" for improving the objective. In other words, when a particular point is in a non-optimal position, the gradient of the relative entropy term is more specific about which point requires adjustment than the error term. This aspect seems to become more prominent when the non-optimal Lévy point represents a relatively small jump size.

The above experiment was repeated to test the performance of a of a regularised root mean squared error (RMSE) objective and its gradient. An example is shown in Figure on page 164.

Our detailed investigation, which also included comparative finite difference gradient approximations, did not reveal any evidence of errors in our implementation.

C.4.2 Relative entropy convergence

We discussed the extreme test case for an optimisation of an objective consisting solely of the relative entropy penalty component in Section 6.2.2 on page 82. In this section of the appendix we provide additional observations for the mentioned non-convergence of some of these tests.

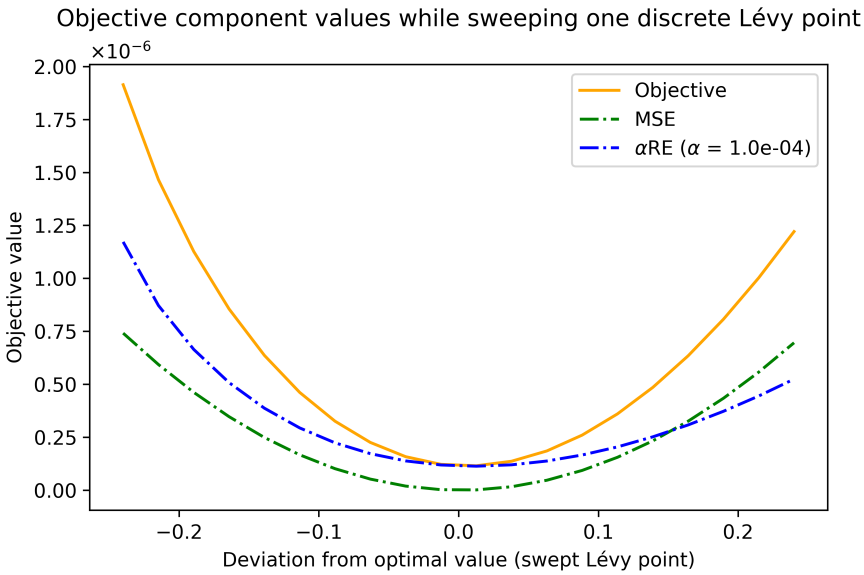

Figure C.8: Objective function with its components over the swept range of the selected Lévy point.

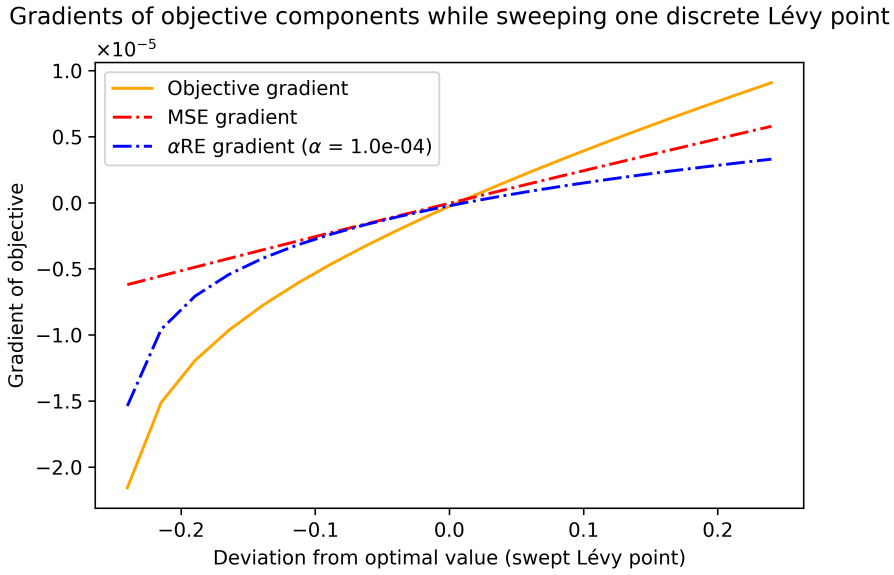

Figure C.9: Gradient of the objective function and its components with respect to the swept point.

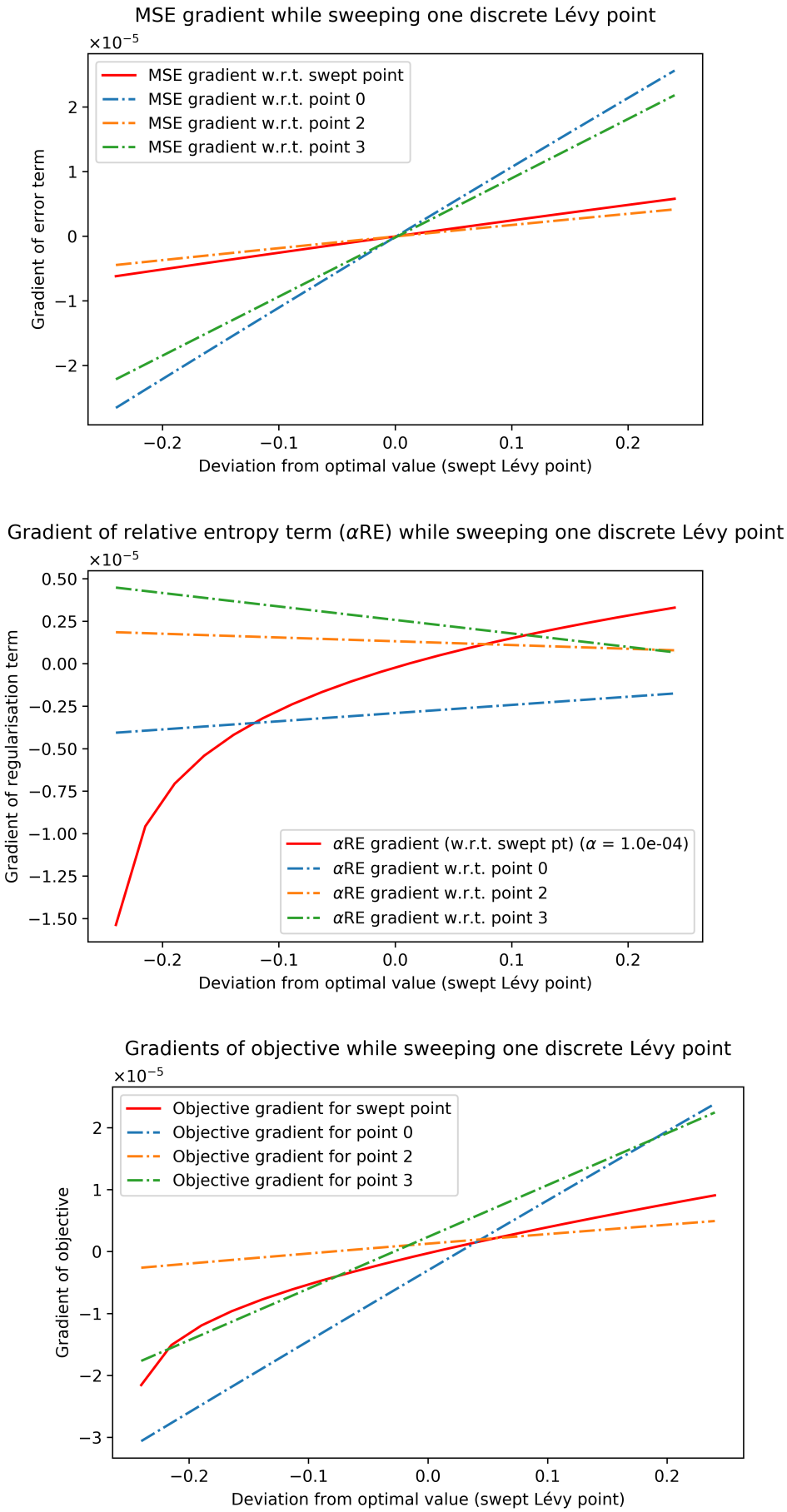

Figure C.10: Breakdown of gradient components for all four points Lévy points.

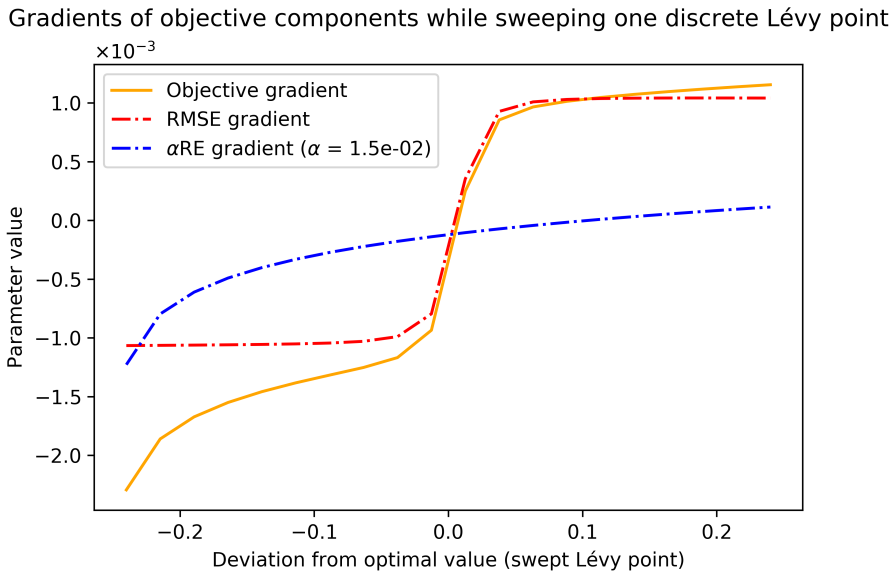

Figure C.11: Root mean squared gradient components with respect to the swept Lévy point.

We mentioned earlier that approximately 15% of the relative-entropy-only tests failed to converge. This might be an over-statement of the extent of the problem, since the failures coincided with one particular set of initial points. The optimiser terminated with a line-search error, suggesting that it was unable to determine a steepest descent line from the gradient function for a particular point. This problem is usually indicative of an incorrect implementation of the gradient function. Our results from several numerical differentiation tests, however, suggest that the gradient function was implemented correctly. These numerical approximation tests are discussed in Section C.3.2 and Appendix C.4.1. It is however possible that this particular set of initial points as shown in Figure 6.3 presented a problem. The initial points are rather far away from any sensible value in the tail areas of the discrete Lévy measure and may have exposed an issue relating to the wide dynamic range of the relative entropy function which becomes evident when the "distance" becomes excessively large. Perhaps the line search algorithm has some difficulty with such extreme non-linearity. Since the investigation of the internal operation of optimiser routines falls outside the scope of this study, the observation was noted, but not further investigated.

Surprisingly, we further observed that within the set of initial points that showed the above-mentioned phenomenon of non-convergence, there were some instances of successful convergence. In about 50% of the tests from this initial point vector we observed successful convergence. This was rather unexpected, since the parameter that was varied for the given initial point set, was simply the regularisation constant, which has an obsolete role when there is no pricing error term present[3]. In other words, the convergence was influenced by the multiplying of the objective function (in this

[3]The variation of regularisation constant was absolutely not required for this experiment. Its presence was an unintentional side effect of a least-effort modification to an existing test script which

case, the relative entropy term) with a constant. Furthermore it was observed that this behaviour was not monotonic with respect to the magnitude of the regularisation constant.

One could interpret this as an example of the failure of a gradient based optimiser to converge. Rama Cont stated the following in the context of regularised optimisation problems in the article *Inverse problems in options pricing:* "[The] convergence of gradient-based methods are not obvious" [Con06a].

C.4.3 Local minima in discrete relative entropy

Prompted by the above-mentioned observations of non-convergence of relative entropy functions, we conducted experiments to test the relative entropy function with a low number of Lévy points, to see if one could identify any signs of non-convexity. We started with a simple four-point discrete Lévy measure with matching prior measure and observed the levels of relative entropy as points were swept across a range of values. To our surprise, we observed the presence of local minima. The surface plot in Figure C.12 clearly shows the global minimum surrounded by a number of local minima in the relative entropy as two points were swept[4]. We found that increasing the number of Lévy points tends to reduce the depth of the individual valleys, but also leads to an increase in the number of valleys.

was originally intended for conventional optimisation tests with a varying regularisation constant for each initial point vector.

[4]We plotted the logarithm of the relative entropy to more clearly illustrate the magnitudes of the surface plot.

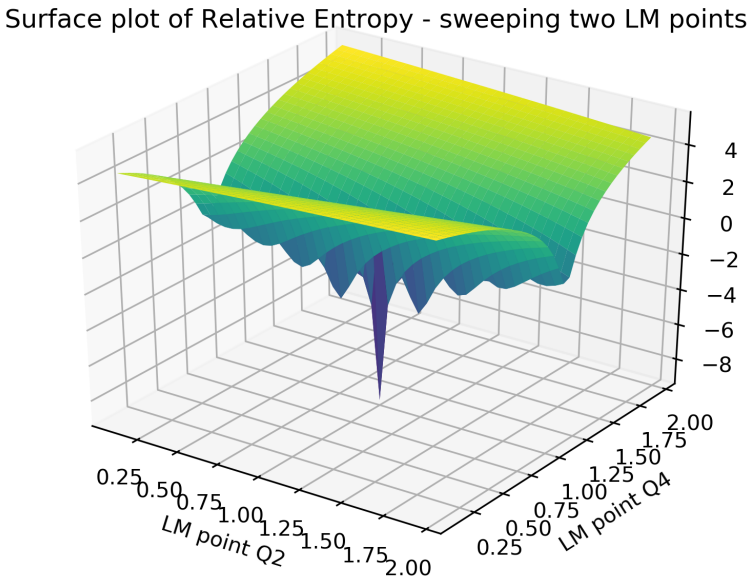

Figure C.12: Illustration of local minima in a relative entropy function with only four discrete Lévy points.

Appendix D

Additional experiments and results

D.1 Illustrative Lévy density from a probability density

In the development of the model the some experiments were performed with Lévy densities derived from arbitrary jump distributions. Below, we show an example representing an arbitrarily selected bimodal jump distribution. This particular shape was selected to generate a relatively narrow range of jump sizes, specifically excluding small jumps. Typical paths with these jump properties are shown in Figure 2.7 on page 38.

Figure D.1shows a probability density of the jump distributions, with the corresponding Lévy density with a jump intensity with $\lambda = 10$ shown in Figure D.2. The related discrete Lévy density can be seen in Figure D.3.

D.2 Original graphs illustrating sensitivity to prior measures

The unaltered graphs published in the article *Nonparametric calibration of jump-diffusion option pricing models* [CT04] to demonstrate are shown in Figure D.4 and Figure D.5 for reference purposes.

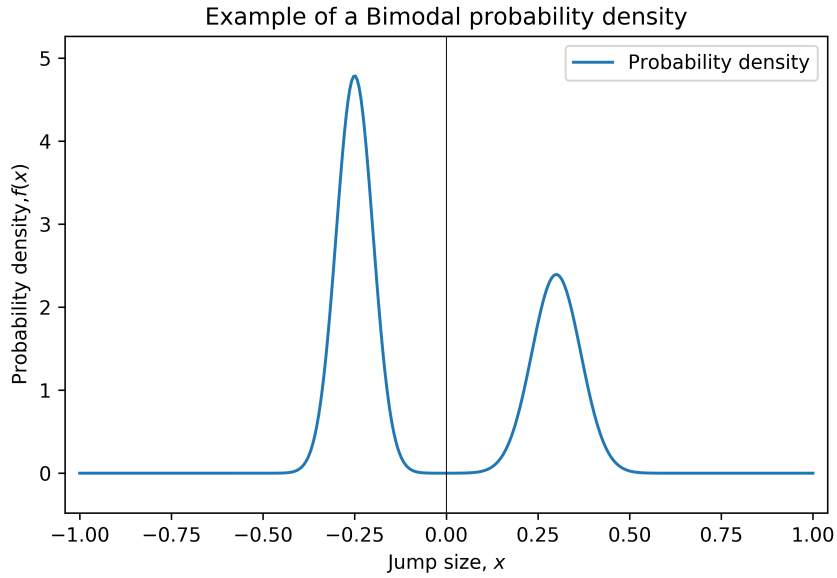

Figure D.1: An arbitrary "bimodal" probability density.

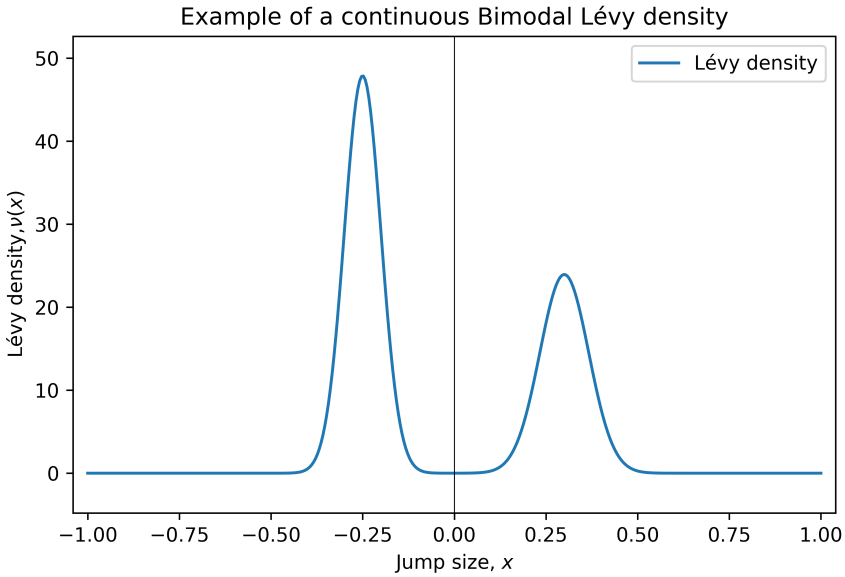

Figure D.2: Continuous "bimodal" Lévy density from a pdf with $\lambda = 10$.

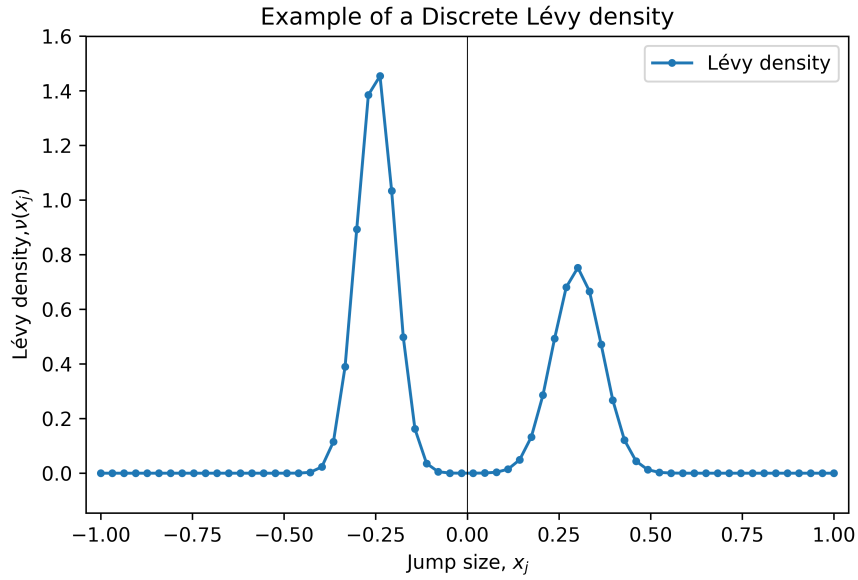

Figure D.3: Discrete "bimodal" Lévy density from a pdf with $\lambda = 10$.

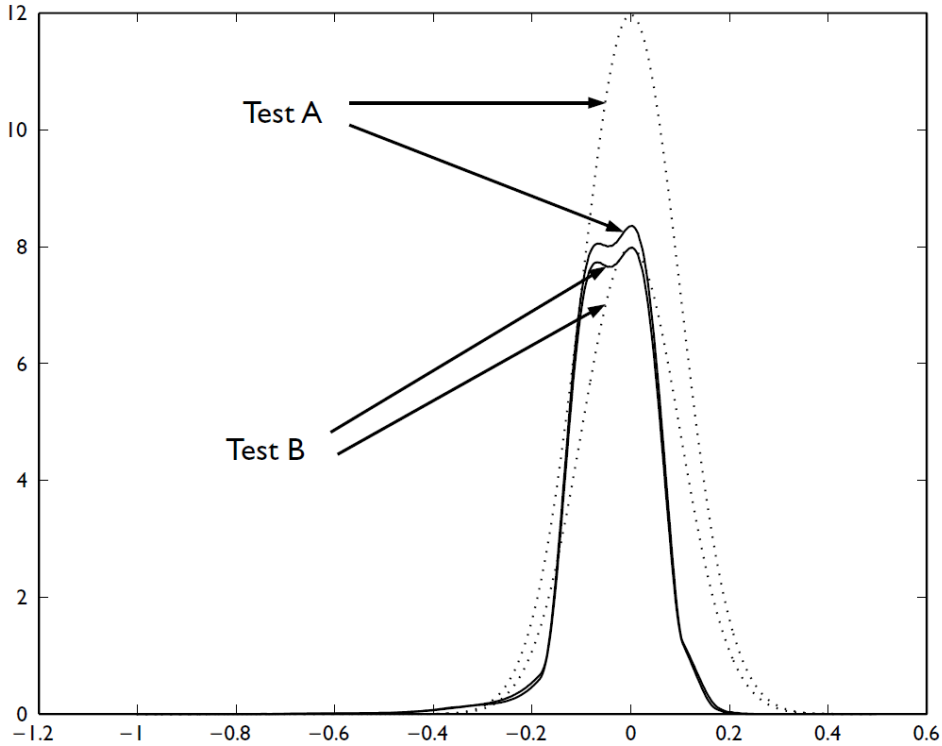

Figure D.4: Cont and Tankov "prior insensitivity test" with similarly shaped prior measures ([CT04] Fig 4)

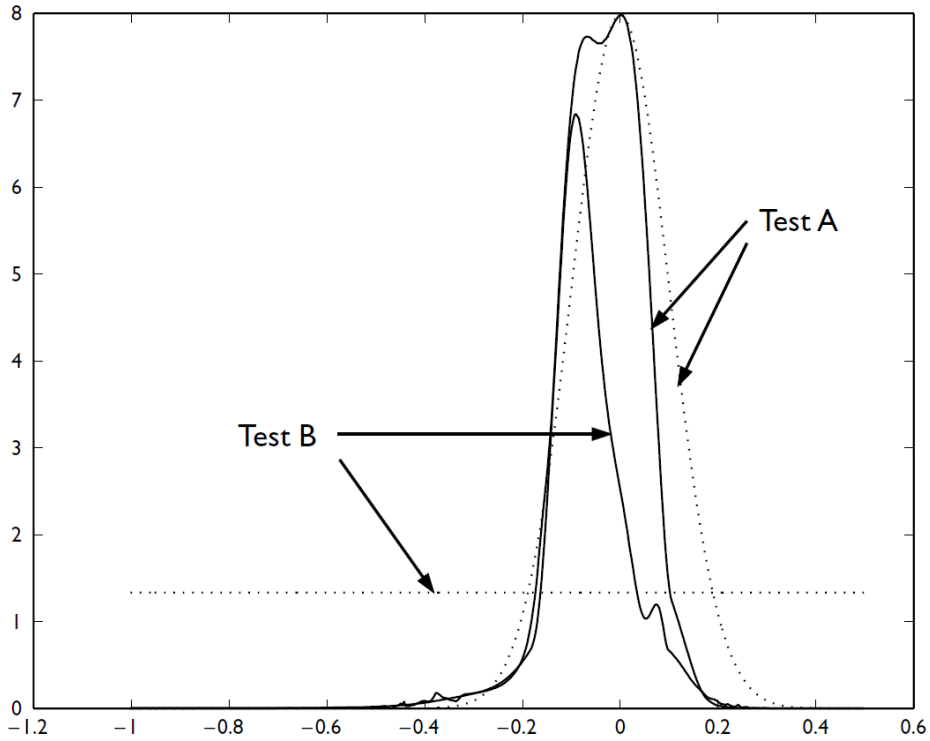

Figure D.5: Cont and Tankov "prior insensitivity test" with differently shaped prior measures ([CT04] Fig 5)

Appendix E

Illustrative examples of model generated results

The graph array below is a typical example of the format used to evaluate batches of optimisation results with parameters swept across a range of values.

Figure E.1: Sample graph array (low resolution)

Bibliography

[ABF+01] Marco Avellaneda, Robert Buff, Craig Friedman, Nicolas Grandechamp, Lukasz Kruk, and Joshua Newman. Weighted Monte Carlo: a new technique for calibrating asset-pricing models. *International Journal of Theoretical and Applied Finance*, 4(01):91–119, 2001.

[Alb20] Jorge Albani, Viniciusand Zubelli. A splitting strategy for the calibration of jump-diffusion models. *Finance and Stochastics*, 24(3):677–722, 2020.

[Ave98] Marco Avellaneda. Minimum-entropy calibration of asset-pricing models. *International Journal of theoretical and applied finance*, 1(04):447–472, 1998.

[Ban19] Bank for International Settlements (BIS). BIS Quarterly Review September 2019, September 2019.

[Ban22] Bank for International Settlements (BIS). OTC derivatives statistics at end December 2021, May 2022.

[BB09] S. Broughton and K. Bryan. *Discrete Fourier analysis and wavelets: application to signal and image processing.* John Wiley & Sons, 2009.

[Bel10] Denis Belomestny. Spectral estimation of the fractional order of a Lévy process. *Ann. Statist.*, 38(1):317–351, 02 2010. doi:10.1214/09-AOS715.

[Bel11] Denis Belomestny. Spectral estimation of the Lévy density in partially observed affine models. *Stochastic processes and their applications*, 121(6):1217–1244, 2011.

[BGSS20] Denis Belomestny, Shota Gugushvili, Moritz Schauer, and Peter Spreij. Nonparametric Bayesian volatility estimation for gamma-driven stochastic differential equations. *arXiv preprint arXiv:2011.08321*, 2020.

[BR06] Denis Belomestny and Markus Reiß. Spectral calibration of exponential Lévy models. *Finance and Stochastics*, 10(4):449–474, 2006.

[BRR20] Alessandro Bondi, Dragana Radojičić, and Thorsten Rheinländer. Comparing two different option pricing methods. *Risks*, 8(4):108, 2020.

[BS94] D.H. Bailey and P.N. Swarztrauber. A fast method for the numerical evaluation of continuous Fourier and Laplace transforms. *SIAM Journal on Scientific Computing*, 15(5):1105–1110, 1994.

[CD13] Rama Cont and Romain Deguest. Equity correlations implied by index options: estimation and model uncertainty analysis. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, 23(3):496–530, 2013.

[CGC09] Fabienne Comte and Valentine Genon-Catalot. Nonparametric estimation for pure jump Lévy processes based on high frequency data. *Stochastic Processes and their Applications*, 119(12):4088–4123, 2009.

[CK13] Rama Cont and Thomas Kokholm. A consistent pricing model for index options and volatility derivatives. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, 23(2):248–274, 2013.

[CLD13] Rama Cont, Christian Leonard, and Romain Deguest. Inverse problems in option pricing: a statistical approach using minimal entropy random mixtures (overhead slides), 2013.

[CM99] Peter Carr and Dilip Madan. Option valuation using the fast Fourier transform. *Journal of Computational Finance*, 2(4):61–73, 1999.

[Coc18] Alberto Coca. Efficient nonparametric inference for discretely observed compound Poisson processes. *Probability Theory and Related Fields*, 170(1):475–523, 2018.

[Con06a] Rama Cont. Inverse problems in options pricing. In Marc Hoffmann and Markus Reiß, editors, *Mini-Workshop: Statistical Methods for Inverse Problems*, volume 51 of *Oberwolfach Reports*, pages 3057–3086, 2006. URL: https://publications.mfo.de/bitstream/handle/mfo/2981/OWR_2006_51.pdf.

[Con06b] Rama Cont. Model uncertainty and its impact on the pricing of derivative instruments. *Mathematical Finance*, 16(3):519–547, 2006.

[Cré10] Stéphane Crépey. Tikhonov regularization. *Encyclopedia of Quantitative Finance*, 2010.

[CT65] James W. Cooley and John W. Tukey. An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, 19(90):297–301, 1965.

[CT02] Rama Cont and Peter Tankov. Calibration of jump-diffusion option pricing models: a robust non-parametric approach. *Working paper*, 2002.

[CT03] Rama Cont and Peter Tankov. *Financial Modelling with Jump Processes*. CRC press, 2003.

[CT04] Rama Cont and Peter Tankov. Nonparametric calibration of jump-diffusion option pricing models. *Journal of Computational Finance*, 7:1–49, 2004.

[CT06] Rama Cont and Peter Tankov. Retrieving Lévy processes from option prices: Regularization of an ill-posed inverse problem. *SIAM Journal on Control and Optimization*, 45(1):1–25, 2006.

[DM17] Céline Duval and Ester Mariucci. Compound Poisson approximation to estimate the Lévy density. 2017.

[DM21] Céline Duval and Ester Mariucci. Spectral-free estimation of Lévy densities in high-frequency regime. *Bernoulli*, 27(4):2649–2674, 2021.

[DS94] Freddy Delbaen and Walter Schachermayer. A general version of the fundamental theorem of asset pricing. *Mathematische annalen*, 300(1):463–520, 1994.

[EHN96] Heinz Werner Engl, Martin Hanke, and Andreas Neubauer. *Regularization of inverse problems*, volume 375. Springer Science & Business Media, 1996.

[FM03] Tsukasa Fujiwara and Yoshio Miyahara. The minimal entropy martingale measures for geometric Lévy processes. *Finance and Stochastics*, 7(4):509–531, 2003.

[Fri00] Marco Frittelli. The minimal entropy martingale measure and the valuation problem in incomplete markets. *Mathematical finance*, 10(1):39–52, 2000.

[FS02] Hans Föllmer and Alexander Schied. Stochastic finance. In *Stochastic Finance*. de Gruyter, 2002.

[Gat11] Jim Gatheral. *The volatility surface: a practitioner's guide*, volume 357. John Wiley & Sons, 2011.

[GdMSS19] Shota Gugushvili, Frank van der Meulen, Moritz Schauer, and Peter Spreij. Nonparametric Bayesian volatility estimation. In *2017 MATRIX annals*, pages 279–302. Springer, 2019.

[Gib98] J. W. Gibbs. Fourier's series. *Nature*, 59(1522):200–200, 1898.

[GL90] Gerard Gennotte and Hayne Leland. Market liquidity, hedging, and crashes. *The American Economic Review*, pages 999–1021, 1990.

[Gla04] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Stochastic Modelling and Applied Probability. Springer New York, 2004.

[GMvdM20] Shota Gugushvili, Ester Mariucci, and Frank van der Meulen. Decompounding discrete distributions: A nonparametric Bayesian approach. *Scandinavian journal of statistics*, 47(2):464–492, 2020.

[Goo22a] Google Books. Financial Modelling with Jump Processes. https://books.google.co.za/books/about/Financial_Modelling_with_Jump_Processes.html?id=bVlieV8GBrIC&redir_esc=y, 2022. [Online; accessed 4-August-2022].

[Goo22b] Google Scholar. Peter Tankov citations. https://scholar.google.com/citations?user=UoUi8tUAAAAJ&hl=en&oi=sra, 2022. [Online; accessed 4-August-2022].

[GR01] Thomas Goll and Ludger Rüschendorf. Minimax and minimal distance martingale measures and their relationship to portfolio optimization. *Finance and Stochastics*, 5(4):557–581, 2001.

[GR13] Stefan Giebel and Martin Rainer. Neural network calibrated stochastic processes: forecasting financial assets. *Central European Journal of Operations Research*, 21(2):277–293, 2013.

[GS12] Florence Guillaume and Wim Schoutens. Calibration risk: Illustrating the impact of calibration risk under the Heston model. *Review of Derivatives Research*, 15(1):57–79, 2012.

[GS13] Florence Guillaume and Wim Schoutens. A moment matching market implied calibration. *Quantitative Finance*, 13(9):1359–1373, 2013.

[Gug09] Shota Gugushvili. Nonparametric estimation of the characteristic triplet of a discretely observed Lévy process. *Journal of Nonparametric Statistics*, 21(3):321–343, 2009.

[GVdMS15] Shota Gugushvili, Frank Van der Meulen, and Peter Spreij. Nonparametric Bayesian inference for multidimensional compound Poisson processes. *Modern Stochastics: Theory and Applications*, 2(1):1–15, 2015.

[GvdMS18] Shota Gugushvili, Frank van der Meulen, and Peter Spreij. A nonparametric Bayesian approach to decompounding from high frequency data. *Statistical Inference for Stochastic Processes*, 21(1):53–79, 2018.

[Hir12] Ali Hirsa. *Computational methods in finance*. CRC Press, 2012.

[HR06] Marc Hoffmann and Markus Reiß. Mini-workshop: Statistical methods for inverse problems. *Oberwolfach Reports*, 3(4):3057–3086, 2006.

[HS06] Friedrich Hubalek and Carlo Sgarra. Esscher transforms and the minimal entropy martingale measure for exponential Lévy models. *Quantitative Finance*, 6(02):125–145, 2006.

[Huh19] Jeonggyu Huh. Pricing options with exponential Lévy neural network. *Expert Systems with Applications*, 127:128–140, 2019.

[Hul12] John C. Hull. *Options, Futures, and Other Derivatives*. Pearson, eighth edition, 2012.

[JVDMVDV05] G. Jongbloed, F. H. Van Der Meulen, and A. W. Van Der Vaart. Nonparametric inference for Lévy-driven Ornstein-Uhlenbeck processes. *Bernoulli*, 11(5):759–791, 2005.

[Kal02] Jan Kallsen. Utility-based derivative pricing in incomplete markets. In *Mathematical finance Bachelier congress 2000*, pages 313–338. Springer, 2002.

[KHO19] Tugce Karatas, Ali Hirsa, and Amir Oskoui. Supervised deep neural networks (DNNs) for pricing/calibration of vanilla/exotic options under various different processes. *arXiv preprint arXiv:1902.05810*, 2019.

[KL07] Young Shin Kim and Jeong Hyun Lee. The relative entropy in CGMY processes and its applications to finance. *Mathematical Methods of Operations Research*, 66(2):327–338, 2007.

[KR10] Johanna Kappus and Markus Reiß. Estimation of the characteristics of a Lévy process observed at arbitrary frequency. *Statistica Neerlandica*, 64(3):314–328, 2010.

[KW12] J. Kienitz and D. Wetterau. *Financial Modelling: Theory, Implementation and Practice with MATLAB Source*. The Wiley Finance Series. Wiley, 2012.

[Lee04] Roger W. Lee. Option pricing by transform methods: extensions, unification and error control. *Journal of Computational Finance*, 7(3):51–86, 2004.

[Lew01] Alan L. Lewis. A simple option formula for general jump-diffusion and other exponential Lévy processes. *Available at SSRN 282110*, 2001.

[Mat05] Kazuhisa Matsuda. Parametric regularized calibration of Merton jump-diffusion model with relative entropy: What difference does it make? 2005.

[Mor66] Vladimir Alekseevich Morozov. On the solution of functional equations by the method of regularization. In *Doklady Akademii Nauk*, volume 167, pages 510–512. Russian Academy of Sciences, 1966.

[Mor11] Massimo Morini. *Understanding and Managing Model Risk: A practical guide for quants, traders and validators*. John Wiley & Sons, 2011.

[NR09] Michael H Neumann and Markus Reiß. Nonparametric estimation for Lévy processes from low-frequency observations. *Bernoulli*, 15(1):223–248, 2009.

[NW06] Jorge Nocedal and Stephen Wright. *Numerical optimization*. Springer Science & Business Media, 2006.

[QT19] Likuan Qin and Viktor Todorov. Nonparametric implied Lévy densities. *The Annals of Statistics*, 47(2):1025–1060, 2019.

[Reb02] Riccardo Rebonato. Theory and practice of model risk management. *Modern Risk Management: A History, RiskWaters Group, London*, pages 223–248, 2002.

[REK00] Richard Rouge and Nicole El Karoui. Pricing via utility maximization and entropy. *Mathematical Finance*, 10(2):259–276, 2000.

[Ros10] Sheldon M. Ross. *Introduction to probability models*. Academic press, 2010.

[Ros14] M. Ross, Sheldon. *A first course in probability*. Pearson, 2014.

[RW19] Johannes Ruf and Weiguan Wang. Neural networks for option pricing and hedging: a literature review. *arXiv preprint arXiv:1911.05620*, 2019.

[Sat99] Ken-iti Sato. *Lévy processes and infinitely divisible distributions*. Cambridge University Press, 1999.

[Sch03] W. Schoutens. *Lévy Processes in Finance: Pricing Financial Derivatives*. Wiley Series in Probability and Statistics. Wiley, 2003.

[SST03] Wim Schoutens, Erwin Simons, and Jurgen Tistaert. A perfect calibration! Now what? *The best of Wilmott*, page 281, 2003.

[ST12] Jakob Söhl and Mathias Trabs. Option calibration of exponential Lévy models: implementation and empirical results. SFB 649 discussion paper 2012-017, Berlin, 2012. URL: http://hdl.handle.net/10419/56624.

[TA11] A. N. Tikhonov and V. Y. Arsenin. Regularization. In *Encyclopedia of Mathematics*. Springer, 2011. URL: http://www.encyclopediaofmath.org/index.php?title=Regularization&oldid=18991.

[Tan07] Peter Tankov. Lévy processes in finance and risk management. *Wilmott Magazine, September-October*, 2007.

[Tod21] Viktor Todorov. Nonparametric jump variation measures from options. *Journal of Econometrics*, 2021.

[TT11] Viktor Todorov and George Tauchen. Volatility jumps. *Journal of Business & Economic Statistics*, 29(3):356–371, 2011.

[TV09] Peter Tankov and Ekaterina Voltchkova. Jump-diffusion models: a practitioner's guide. *Banque et Marchés*, 99(1):24, 2009.

[VEGS07] Bert Van Es, Shota Gugushvili, and Peter Spreij. A kernel type nonparametric density estimator for decompounding. *Bernoulli*, 13(3):672–694, 2007.

[VGO+19] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, and Jonathan Bright. Scipy 1.0 - Fundamental algorithms for scientific computing in Python, 2019.

[Vog02] Curtis Vogel. *Computational methods for inverse problems*. SIAM, 2002.

[VT14] Ekaterina Voltchkova and Peter Tankov. Deterministic methods for option pricing in exponential Lévy models. *Premia 14 documentation*, 2014.

[Wei22] E. W. Weisstein. Fourier transform. *Mathworld - a Wolfram web resource*, 2022. URL: https://mathworld.wolfram.com/FourierTransform.html.

[XD20] Kailai Xu and Eric Darve. Calibrating multivariate Lévy processes with neural networks. In *Mathematical and Scientific Machine Learning*, pages 207–220. PMLR, 2020.

[YZH17] Yongxin Yang, Yu Zheng, and Timothy Hospedales. Gated neural networks for option pricing: Rationality by design. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, 2017.

[ZBLN97] Ciyou Zhu, Richard H. Byrd, Peihuang Lu, and Jorge Nocedal. Algorithm 778: L-BFGS-B: FORTRAN subroutines for large-scale bound-constrained optimization. *ACM Transactions on Mathematical Software (TOMS)*, 23(4):550–560, 1997.