Essays on Quantitative Finance

Ph.D. Thesis

Martin Jönsson

Department of Mathematical Sciences
Faculty of Science
University of Copenhagen

This thesis has been submitted to the Ph.D. School of the Faculty of Science,
University of Copenhagen.

Martin Jönsson
Department of Mathematical Sciences
University of Copenhagen
Universitetsparken 5
2100 København Ø
Denmark

Email: maj@math.ku.dk

Academic advisor:	Prof. Rolf Poulsen, University of Copenhagen
Assessment committee:	Prof. Mogens Steffensen, University of Copenhagen Prof. Antje Mahayni, University of Duisburg-Essen Prof. Claus Munk, Copenhagen Business School
Submission date:	June 3rd, 2016
ISBN:	978-87-7078-945-5

Preface

This thesis has been submitted in partial fulfilment of the requirements of the degree of Doctor of Philosophy at the Faculty of Science, University of Copenhagen. The work included here was carried out at the Department of Mathematical Sciences, University of Copenhagen, between March 2014 and December 2015, and at the Mathematical Institute, University of Oxford, during an academic visit between January 2016 and April 2016.

There is a lot of people who have offered me a great deal of support and cheering during my time as a Ph.D. student and I would like to express my sincere gratitude to all of them. Most of all, I would like to thank my supervisor Rolf Poulsen for his confident decision to admit me as his Ph.D. student; for his never-ending support, many ideas, academic advice and for making my Ph.D. journey feasible whilst giving all the academic liberty I could wish for. Likewise, my greatest gratitude to Samuel N. Cohen, my "supervisor abroad" who generously offered me the most interesting and enjoyable time in Oxford. Thanks to his hospitality and inspiring academic guidance this was a fantastic experience for me, both in terms of exploring new research and since I was introduced to the delightful social sphere of the department.

My very special thanks to my friend and fellow doctorate Simon Ellersgaard Nielsen: for our fruitful collaborations and for cheerful times at our office – it's been a true pleasure. The wit of Simon's pen as well as his sharp mathematical reasoning have greatly contributed to my work and I shall be deeply thankful. Many thanks also to my co-authors Rune Ramsdal Ernstsen and Anders Skajaa for the enjoyment of working together. For inspiring and challenging me when I was their student: thanks to Goran Peskir and Alexander Sokol. As well, special thanks to Mogens Steffensen for supporting me, to Jesper Lund Pedersen for answering my questions and to Susanne Ditlevsen for being the first to invite me to Copenhagen.

Further, I would like to thank Emil Jørgensen for inspiring discussions and for a great time at the department – I really hope we'll have an opportunity to work together in the future – and to Jannick Schreiner for happy evenings in Copenhagen and for a great time at our office. Thanks also to everyone at the Department of Mathematics including the administration for providing a friendly and stimulating environment. At the University of Oxford, special thanks to Laura Auger for all her kind help and administrative support, and to Gonçalo Simões and Victor Fedyashov for truly making me feel at home in the group.

Last but certainly not least: I would like to thank my dear parents and brothers for believing in me and for always supporting me regardless of what I have decided to do. Heartfelt thanks to my dear mother for helping me out in Lund during the last months of my studies. Thanks also to my parents-in-law for their encouragement, to my friends and to my high-school teacher Anders Blennar for telling me to study maths and physics in the very first place.

Finally, I would like to thank my wonderful wife for her never ending support and patience, for always standing by my side and for being the love of my life. Without you none of this would be possible. And to my daughter, thank you just for being You.

Lund, May 2016 *Martin Jönsson*

Summary

This thesis consists of five research papers written during the period March 2014 - April 2016. The papers can be read independently and their abstracts are:

1. **European Option Pricing with Stochastic Volatility Models under Parameter Uncertainty.** We consider stochastic volatility models under parameter uncertainty and investigate how model derived prices of European options are affected. We let the pricing parameters evolve dynamically in time within a specified region, and formalise the problem as a control problem where the control acts on the parameters to maximise/minimise the option value. Through a dual representation with backward stochastic differential equations, we obtain explicit equations for Heston's model and investigate several numerical solutions thereof. In an empirical study, we apply our results to market data from the S&P 500 index where the model is estimated to historical asset prices. We find that the conservative model-prices cover 98% of the considered market-prices for a set of European call options.

2. **The Fundamental Theorem of Derivative Trading.** When estimated volatilities are not in perfect agreement with reality, delta hedged option portfolios will incur a non-zero profit-and-loss over time. However, there is a surprisingly simple formula for the resulting hedge error, which has been known since the late '90s. We call this The Fundamental Theorem of Derivative Trading. This paper is a survey with twists of that result. We prove a more general version of it and discuss various extensions and applications, from incorporating a multi-dimensional jump framework to deriving the Dupire-Gyöngy-Derman-Kani formula. We also consider its practical consequences both in simulation experiments and on empirical data thus demonstrating the benefits of hedging with implied volatility.

3. **Risk Minimization in Electricity Markets.** This paper analyses risk management of fixed price, unspecified consumption contracts in energy markets. We model the joint dynamics of spot-price and consumption of electricity, study expected loss minimisation for different loss functions and derive optimal static hedge strategies based on forward contracts. These strategies are tested empirically on 2012-2014 Nordic market data and compared to a simpler hedge strategy which is widely employed by the industry. Results show that our suggested hedge outperforms the commonly used with a higher reward-to-risk ratio, which can be exploited to release

a premium from the contract. The realised cumulative profit-and-loss from our suggested hedge is greater for almost every single one-month period of the considered data, whilst the hourly realised payout results in a 66% out-performance probability.

4. **Stochastic Volatility for Utility Maximisers.** Using martingale methods we derive bequest optimising portfolio weights for a rational investor who trades in a bond-stock-derivative economy characterised by a generic stochastic volatility model. For illustrative purposes we then proceed to analyse the specific case of the Heston economy, which admits explicit expressions for plain vanilla Europeans options. By calibrating the model to market data, we find that the demand for derivatives is primarily driven by the myopic hedge component. Furthermore, upon deploying our optimal strategy on real market prices, we find only a very modest improvement in portfolio wealth over the corresponding strategy which only trades in bonds and stocks.

5. **Volatility is Log-Normal, But not for the Reason you Think.** Stochastic volatility models have increased enormously in popularity since their introduction in the late eighties. Not the least for hedging and option pricing purposes since they do well in fitting the implied volatility surface. In fact, their pricing ability is often the reason for advocating such a model, whilst their ability to capture the underlying dynamics is loosely motivated. In this paper we test for what is a good model of volatility based on the latter perspective: we briefly review three well-known stochastic volatility models, and concentrate on the instantaneous variance in Heston's model, a log-normal model and in the 3-over-2 model. Since volatility is a non-observable process, we employ the technique of realized volatility to obtain variance measurements and from these we form a goodness-of-fit analysis based on the concept of uniform residuals. To assess the model-classification ability of our analysis, we perform a Monte Carlo study. We then apply the methodology in an empirical study, where our results show that the log-normal model yields a much better goodness-of-fit than both Heston's and the 3-over-2 model.

The following papers were also completed during the course of my Ph.D studies but are not part of this thesis:

6. **Stochastic Volatility for Utility Maximisers - The Bond-Stock Economy**, with S. Ellersgaard, *unpublished*.

7. **Hedging Local Volume Risk using Forward Markets: Nordic Case**, with R. Ramsdal Ernstsen, T. Krogh Boomsma and A. Skajaa, *Energy Economics*, submitted.

8. **Robust feature representation for classification of bird song syllables**, with M. Sandsten and M. Große Ruse, *EURASIP Journal on Advances in Signal Processing*, 2016(1):1–16.

Contents

1 Introduction

The work included in this thesis – well, pretty much everything I did during my Ph.D. – has almost exclusively been initiated by interesting problems and my curiosity to learn more about the fascinating theory of financial mathematics. As such, there was no initial "big question" that underlies my work. Instead, appealing problems and pleasant collaborations have constantly found their way to me and as a result, I've had the great pleasure to work with a good variety of projects. Nevertheless, the papers of this thesis mainly concerns three classical math-finance themes, namely: *option pricing*, *hedging* and *optimal investment* in financial markets.

Interestingly enough, all three concepts lie at hear of the seminal work by Black, Scholes and Merton whose papers mark the beginning of contemporary mathematical finance. Black and Scholes (1973) and Merton (1973)[1] present the famous Black-Scholes financial market model[2] and several fundamental ideas are developed in their papers. Firstly, that of *option pricing*: with a method of continuous trading in the underlying stock and a risk-free bond, they show how an option payoff can be perfectly replicated. The fair price of the option, by no-arbitrage arguments, must then be equal to the value of the replicating portfolio and their approach give rise to the Black-Scholes pricing equation: If $\pi_t = F(t, S_t)$ is the price process of an option with terminal payoff $\Phi(S_T)$, where S_t is the price process of the underlying stock, the (deterministic) pricing function F must satisfy the partial differential equation

$$\frac{\partial F}{\partial t} + \mathcal{L}F - rF = 0 \tag{1.1}$$

with boundary condition $F(T, s) = \Phi(s)$. In particular, they derive an explicit expression for the price of a European call option – the Black-Scholes formula. Secondly, (and inseparable to the pricing problem) they provide a recipe of how exactly this replicating portfolio should be set-up in practice: the *delta-hedge*. With (ψ_t, φ_t) denoting a trading strategy in the bond and stock respectively, a (self-financing) portfolio with value dynamics $dV_t = \psi_t dB_t + \varphi_t dS_t$ replicates the option if and only if

$$\varphi_t = \Delta_t \equiv \frac{\partial F}{\partial s}(t, S_t) \tag{1.2}$$

[1] Modern finance saw yet another important development in 1973 when the world's first exchange for options opened in Chicago.

[2] The modelling framework of Black-Scholes-Merton has its roots in Louis Bachelier's thesis from 1900 and in the work by Samuelson (1965).

where F satisfies (1.1) and ψ_t is chosen such that $V_t = F(t, S_t)$ – the value of the replicating portfolio equals the option price at all times. Thirdly, Merton (1969, 1971) considers the problem of *optimal investment and consumption* under Black-Scholes model with the methods of continuous-time stochastic optimal control. If c_t denotes the consumption and w_t the stock-weight of an agent with initial wealth $\mathscr{W}_0$, the task at hand is to choose a consumption-investment plan (c_t, w_t) that maximises the agent's total utility over a time horizon $[0, T]$

$$\mathbb{E}\left[\int_0^T e^{-\delta s} u(c_s) ds + e^{-\delta T} u(\mathscr{W}_T)\right]. \tag{1.3}$$

Here, the controlled wealth has dynamics $d\mathscr{W}_t = (w_t(\mu - r) + r)\mathscr{W}_t dt - c_t dt + w_t \sigma \mathscr{W}_t dW_t$ while the agent's utility from consumption and a terminal endowment is given by the personal utility function u. The classical method for optimising (1.3) is that of dynamical programming which ties the problem to a partial differential equation – the Hamilton-Jacobi-Bellman equation – and Merton provides solutions thereof for a variety of utility functions and set-ups.

The ideas of Black, Scholes and Merton where soon extended and the modern martingale approach came to light in the papers by Harrison and Kreps (1979), and Harrison and Pliska (1981). Here, equivalent martingale measure(s) play an integral role in the pricing and hedging problem: The market model is arbitrage free if and only if a martingale measure exists (the first fundamental theorem) and further, an arbitrage-free market model is complete if and only if the measure is unique (the second fundamental theorem).[3] The connection to an operational risk-neutral valuation formula is almost immediate: Under such a martingale measure Q, the discounted price process of a contingent claim H is a martingale, $\pi_t B_t^{-1} = \mathbb{E}^Q[\pi_s B_s^{-1} | \mathcal{F}_t]$ for all $t \leq s \leq T$, and in particular

$$\pi_t = B_t \, \mathbb{E}^Q\left[\frac{H}{B_T}\middle| \mathcal{F}_t\right]. \tag{1.4}$$

If the set of martingale measures is non-singular, H is not necessarily replicable by a trading strategy and (1.4) generally yields different arbitrage-free prices for different choices of Q.[4] However, if H is an attainable claim, all replicating strategies for H must have the same portfolio value which equals the price of the risk-neutral valuation formula and hence, the right-hand side of (1.4) is independent of the choice of Q. In particular, if Q is unique, any contingent claim is attainable and (1.4) gives the unique fair price for every H. This is the case for the Black-Scholes market and due to the Markovian structure of the model,

[3] The results stated here are deliberately somewhat imprecise since the fundamental theorems have been subject to a great deal of research and the original versions have seen several extensions, most prominently by Delbaen and Schachermayer (1994), and Delbaen and Schachermayer (1998), where the latter even goes beyond semimartingale models.

[4] Even if H is not *perfectly replicable*: $\exists \psi, \varphi : V_T^{\psi,\varphi} = H$, a.s., one always has the existence of a (cheapest) *super-replicating* strategy with discounted value $\tilde{\pi}_t^{\sup} = \inf_{\psi,\varphi}\{\tilde{V}_t^{\psi,\varphi} : V_T^{\psi,\varphi} \geq H, \text{ a.s.}\}$. In particular, it generally holds that $\tilde{\pi}_t^{\sup} = \sup_Q \mathbb{E}^Q[\tilde{H}|\mathcal{F}_t]$ where the supremum is taken over the set of all equivalent martingale measures.

a connection between (1.1) and (1.4) is directly given by the Feynman-Kac representation formula.

The martingale approach came to play also for optimal investment problems in the papers by Karatzas et al. (1987), and Cox and Huang (1989). In the complete market case, it provides an alternative methodology to the dynamical programming principle that does not require a Markovian structure of the model. The basic idea is to maximise the agent's utility from investing in the set of contingent claims which have a price equal to the agent's initial wealth. Since the market is complete, this corresponds to maximising $\mathbb{E}[u(H)]$ over all integrable $H \in \mathcal{F}_T$ subject to $\mathscr{W}_0 = \mathbb{E}^Q[e^{-rT}H]$ and the key point is, since any claim is attainable, this is equivalent to finding an optimal portfolio strategy in the stock-bond market. Thus, the dynamic control problem is replaced by a static optimisation, conveniently expressed in terms of the Lagrangian

$$\mathscr{L} = \mathbb{E}\left[u(H) - \lambda(e^{-rT}\xi_T H - \mathscr{W}_0)\right] \tag{1.5}$$

where ξ_T is the Radon-Nikodym density of the martingale measure with respect to the objective measure. A point-wise optimisation of (1.5) immediately gives the optimal wealth $H^* = (u')^{-1}(\lambda e^{-rT}\xi_T)$ where the multiplier λ is obtained from the condition $\mathscr{W}_0 = \mathbb{E}[e^{-rT}\xi_T H^*]$. The associated optimal portfolio strategy may also be computed, in practice by matching the dynamics of the wealth process $\mathscr{W}_t^* = \mathbb{E}^Q[e^{-r(T-t)}H^*|\mathcal{F}_t]$ with the value dynamics of a general portfolio strategy in the stock-bond market.

The pricing and hedging theory has seen a rapid development since the early 1970s. One natural driver is the empirically based critique of the Black-Scholes-Merton framework on at least three points: (i) the constant-volatility model's inability to fit option prices observed on the market, (ii) the distributional mismatch of the stock-price model and market behaviour of asset returns, (iii) the assumption that price processes evolve continuously in time in contrast to sudden discontinuous moves by jumps. On the modelling front, stochastic volatility models addressing both (i) and (ii) appeared in the late 1980s with the model by Heston (1993) being the most influential. At the same time, Dupire (1994) and Derman and Kani (1994) showed how local volatility models could be used to perfectly fit a surface of market implied volatilities, across all observed strikes and maturities. Further, already Merton (1976) turned to (iii) with a compound Poisson process added to the continuous geometric Brownian motion of Black-Scholes model; a specification that was later adopted to Heston's stochastic volatility by Bates (1996). This to mention only a handful of the many developments since the beginning of the 1970s: As such, Heston, Dupire and Merton's jump are merely special cases among the great family of exponential Lévy models, general jump-diffusions, stochastic volatility models with jumps, Lévy driven stochastic volatility, local Lévy models and time-changed models, which all add to the model plethora of modern mathematical finance. For a comprehensive textbook-reading on the topic, see for instance Björk (2009), Cont and Tankov (2004), Bingham and Kiesel (2004), Shreve (2004) or Karatzas and Shreve (1998).

This thesis presents five unconnected papers each of which concerning one or several of the above themes. In the first paper, we study stochastic volatility models subject to

parameter uncertainty and how the option pricing is affected. With Heston's model as working example, our approach is to formulate a control problem similar to (1.3) based on the risk-neutral formula (1.4) where the parameters play the role of the controls. The aim is to maximise/minimise the option price over some uncertainty region for the parameters, to obtain a conservative interval for the value of the option. However, we use neither Hamilton-Jacobi-Bellman nor the martingale approach; instead we exploit a duality between optimal control problems and backward stochastic differential equations. Several numerical methods are investigated for their solution, and we analyse empirical implications based on market data from the S&P 500 index. This further brings us to a discussion of how parameter uncertainty can represent the incompleteness of stochastic volatility models by spanning the set of equivalent martingale measures. Ultimately, our approach gives us the cost for super-replication of the option, and we ask in the empirical section if market prices are covered by these inferred pricing bounds.

The next two papers are concerned with the concept of hedging in two fundamentally different contexts. The first of the two considers the classical problem of derivative hedging with a dynamic delta-strategy à la (1.2). Here we ask the question of what happens if an erroneous model is used to appoint the delta-hedge in a tripartite model set-up where the option market is also in discordance with the "veridical" evolution of the underlying. Specifically, with a general Itô process describing the multi-asset space of underlyings and local volatility models used for both hedging and by the option market, we prove a formula for the profit-and-loss incurred from holding the hedge portfolio. As special cases, we consider hedging with a volatility that matches (A) the real dynamics, (B) the option-market implied volatility, and investigate the practical consequences both with a simulation study and in an empirical experiment. Even if we do not find empirical support for the most conspicuous implication of the theorem – that arbitrage can be exploited if the underlying models are Black-Scholian – we see *some* evidence for the fact that hedging at implied volatility yields smoother, less erratic, P&L paths.

The third paper deals with a more practical problem where we wish to set up a hedge portfolio that replicates a so called *fixed-price-agreement* for electricity. For this purpose, we consider a set of forward contracts as given by the market, and aim to construct a static hedging strategy based on these instruments that efficiently replicates the issued claim. The key challenge is that the underlying processes – the spot-price of electricity and the consumption quantity – are non-tradable objects that also typically affects the outcome for the issuer in an adverse way. Hence, the situation is manifestly incomplete: There is no trading strategy in the underlyings that replicates the claim – there is not even a market for the underlyings. Instead, we resort to the method of risk-minimisation of the hedge-portfolio's financial payouts. For this purpose, we propose a joint model for the price-consumption processes that takes into account both spacial and temporal dependency and the seasonal behaviour of electricity. We consider the risk measure of expected positive loss and derive an expression for the hedge portfolio under our model. With the quadratic risk-measure as reference point – since it is commonly used by the industry – we then implement our strategies in an empirical experiment based on forward, spot-prices and consumption data from the Nordic energy market.

In the forth paper, we accede to the Mertonian tradition of control problems and present an optimal investment study where the financial market is driven by a stochastic volatility model. We consider an investment space consisting of a risky stock, a bond and a European derivative written on the stock, and investigate how an agent can maximise his utility from investing in such an economy. Since the fundamental market is completed – we adjoin an additional asset to the stock-bond market – the martingale approach proves useful and by solving (1.5) we arrive at expressions for the optimal wealth process and associated (bond, stock, derivative)-strategy. Once again, Heston's model receives our main attention and we provide explicit formulas for the optimal portfolio weights when the market derivative is a plain vanilla put or call option. Our theoretical study is finally demonstrated by an empirical experiment where we investigate if the trading of S&P 500 calls add any welfare gains to the rational investor.

In the last paper of the thesis we ask the simple question of what is a good model for volatility. However, in contrast to the usual pricing-oriented measure of model-to-market fit for option prices, we focus on the volatility process itself and how well stochastic volatility models match its distributional properties. For this purpose, we promote a goodness-of-fit analysis that builds on the conditional law of a model. Here we have to take into account that instantaneous variance – the object that a stochastic volatility model allege to describe – is a non-observable, latent process. Thus, we consider several choices of variance measures and how their precision affects our ability to distinguish between different models. We scrutinise on our approach in a simulation study before we apply the methodology to a large set of market quotes from both the S&P 500 index and ten individual stocks. Ultimately – as revealed already in the title of the paper – we aim to conclude which stochastic volatility model does the best job.

2

European Option Pricing with Stochastic Volatility Models under Parameter Uncertainty

Samuel N. Cohen[1] and Martin Jönsson[2]

Abstract. We consider stochastic volatility models under parameter uncertainty and investigate how model derived prices of European options are affected. We let the pricing parameters evolve dynamically in time within a specified region, and formalise the problem as a control problem where the control acts on the parameters to maximise/minimise the option value. Through a dual representation with backward stochastic differential equations, we obtain explicit equations for Heston's model and investigate several numerical solutions thereof. In an empirical study, we apply our results to market data from the S&P 500 index where the model is estimated to historical asset prices. We find that the conservative model-prices cover 98% of the considered market-prices for a set of European call options.

Keywords: Option Pricing, Stochastic Volatility, Model Uncertainty.

[1]Mathematical Institute, University of Oxford, United Kingdom.
[2]Department of Mathematical Sciences, University of Copenhagen, Denmark.

2.1 Introduction

In this paper, we consider the problem of European-option pricing when the underlying assets are assumed to follow a stochastic volatility model in a setting that accommodate for *parameter uncertainty*, and in particular, how this transfers to *conservative bounds* for derived option prices.

Stochastic volatility models feature an instantaneous variance of the asset price, the *volatility*, that evolves stochastically in time. It is a natural generalisation of the seminal constant-volatility model of Black and Scholes (1973), and examples include the models introduced by Hull and White (1987), Stein and Stein (1991), Heston (1993), Bates (1996) and Heston (1997) to mention a few. Evidence supporting this generalisation in terms of empirical asset-return behaviour goes back to Black (1976), while for instance Stein (1989) highlights the prediction mismatch of a constant-volatility models and option prices observed from the market. Stochastic volatility serves as an attractive alternative and numerous studies are available from the literature in their favour.

Being a parametric model immediately implies that the stochastic volatility model can be fitted with data. At least two approaches are conventional for this purpose: either estimation from historical asset-prices, or calibration from market option-prices by matching the model derived price (or a combination of the two, see e.g. Ait-Sahalia et al. (2007)). We will concentrate on the former which conforms with the model as postulated to represent the underlying financial market of risky assets: options are fundamentally *derived securities* thereof and their prices are not endogenously given by the model. The concept of parameter uncertainty then naturally arise from statistical inference of the model. Estimation from observed time-series of asset prices results in point-estimates and confidence intervals of the parameters associated with the *real-world* probability measure (as opposed to *risk-neutral* measure(s) used for no-arbitrage option pricing). The confidence interval thus defines the uncertainty set which contains the true value of the model parameters, at a given confidence level.

The question remains of how the parameter uncertainty affects option prices as outputted by the stochastic volatility model. First one needs to establish the relation between the parameters under the statistical measure and under a risk-neutral pricing measure. This is done by the market price of risk, and typically in such a way that the model remains form-invariant. The uncertainty may then be imposed on the risk-neutral parameters which are used for option pricing. Second, to avoid introducing arbitrage into the model, any pricing measure has to be equivalent to the original statistical measure with the consequence that diffusion parameters remain unchanged. Therefore, we assume values of diffusion parameters to be fixed (by their point-estimate) and consider uncertainty in drift- and jump parameters alone.[3] This further offers an interpretation of the parameter uncertainty as representative for the *incompleteness* of the stochastic volatility model: there exists a

[3] A supporting case for this assumption is the fact pointed out for instance by Rogers (2001): while volatilities may be estimated within reasonable confidence with a few years of data, drift estimation requires data from much longer time periods.

space of equivalent pricing measures as given by the span of risk-neutral parameter values in the uncertainty set (we elaborate on this in the introducing discussion of Section 2.4).

We immediately look at the model pricing from a best/worst -case point of view, and aim to obtain some *pricing bounds* inferred from the parameter uncertainty. Two approaches are fair: either optimising the pricing function over the parameters constrained by the uncertainty set, or treating the parameters as dynamical components of a *control process* which acts to optimise the option value. The former is thus a special case of the latter where the control process is restricted to take constant values only. We formalise the problem as a control problem and since all pricing measures are equivalent, this can be seen as change of measure problem. Following the results due to Quenez (1997), the optimal value function of the option price may then by expressed as a backward stochastic differential equation.

The postulation of parameter uncertainty, and more generally *model uncertainty*, as an inherent model feature is certainly not novel. Conceptually, parameter and model uncertainty draw on the principles due to Keynes (1921) and Knight (1921) of the *unknown unknown*, as distinguished from the *known unknown* where randomness emanate from a uniquely postulated probability distribution. In the case of parameter uncertainty, the distribution is unknown but within a specified family of parametrised distributions while model uncertainty more or less leaves the distribution completely unknown (which can be embedded into the former). The importance of model uncertainty in finance was early acknowledged by Derman (1996) and have since then received a great deal of attention. For a readable overview of model uncertainty in the context of option pricing see for instance Gupta et al. (2010), while the parameter uncertainty worst-case approach taken here follows in the same lines as El Karoui and Quenez (1995), Avellaneda et al. (1995), Lyons (1995) and Avellaneda and Paras (1996), among others.

Overview. The model proposed by Heston (1993) will be the working model of our study, and we present the risk-neutral pricing of European options in Section 2.2 along with the BSDE representation of the controlled value process. We show how to derive the optimal driver that generates the BSDE of the optimally controlled value processes, which gives us the pricing bounds for options under parameter uncertainty. To obtain actual values for the pricing bounds, we must resort to numerical solutions for the BSDE that governs the optimal value. In Section 2.3, we detail some simulation schemes for this purpose, and demonstrate the methods in a controlled setting to be able to compare and evaluate their performance. With a suggested numerical scheme in hand, we then proceed in Section 2.4 to illustrate our method empirically on real-world market data from the S&P 500 index. Again, we take the view that inferred statistical parameter-uncertainty represents the space of equivalent martingale measures postulated by the incompleteness of the model. For a set of market quotes of European call options on the S&P 500 index, we then investigate how well the (numerically calculated) model bounds actually cover the observed market prices. We also compare the results with the corresponding constant-parameter optimal price. For completion, we finally treat the general multi-asset case of a Markovian stochastic volatility model with jumps in Section 2.5. Section 2.6 concludes.

2.2 The Heston stochastic volatility model

To set the scene, we consider a financial market consisting of a risk-free money account and a risky asset over a fixed time period $[0, T]$. We assume the standard assumptions of a frictionless market: short selling is permitted and assets may be held in arbitrary amounts, there are no transaction costs and borrowing and lending are made at the same interest rate. The prices of the assets will be modelled as adapted stochastic processes on a filtered probability space, the notion of which will be formalised in the following section.

2.2.1 European option pricing

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P)$ be a filtered probability space where $\{\mathcal{F}_t\}_{t\geq 0}$ is the natural filtration generated by two independent Wiener processes W^1 and W^2, augmented to satisfy the usual conditions of P-completeness and right continuity. We assume that the asset price S and variance V follow the model by Heston (1993), with real-world dynamics (under the objective probability measure P) as given by

$$\begin{aligned} dS_t &= \mu(V_t)S_t dt + \sqrt{V_t}S_t(\rho dW_t^1 + \sqrt{1-\rho^2}dW_t^2), \\ dV_t &= \kappa(\theta - V_t)dt + \sigma\sqrt{V_t}dW_t^1, \end{aligned}$$

for nonnegative constants κ, θ, σ and instantaneous correlation $\rho \in (-1, 1)$. The variance process thus follows a square root process[4] and it is bounded below by zero. If Feller's condition is satisfied; $2\kappa\theta \geq \sigma^2$, then the boundary can not be achieved. Furthermore, the relative rate of return μ is taken to be a deterministic function of the variance. In addition to the risky asset, the market contains a risk-free money account which value processes is denoted B. The money account pays a constant rate of return r, which means that B obeys the deterministic dynamics $dB_t = rB_t dt$.

The market price of risk processes (γ^1, γ^2) associated with W^1 and W^2, are assumed to be specified such that

$$\frac{\mu(V) - r}{\sqrt{V}} = \left(\rho\gamma^1 + \sqrt{1-\rho^2}\gamma^2\right) \tag{2.1}$$

and as suggested by Heston, we let $\gamma^1 \equiv \lambda\sqrt{V}$ for some constant λ. We then have that the stochastic exponential of $-(\gamma^1, \gamma^2) \bullet (W^1, W^2)$ is given by[5]

$$\mathcal{E}(-\gamma \bullet W) = \exp\left(-\int_0^{\cdot} \lambda\sqrt{V_s}dW_s^1 - \int_0^{\cdot} \gamma_s^2 dW_s^2 - \frac{1}{2}\int_0^{\cdot}(\lambda^2 V_s + (\gamma_s^2)^2)ds\right)$$

and if we define the measure Q on $\mathcal{F}_T$ for a fixed deterministic time T by

$$\frac{dQ}{dP} = \mathcal{E}(-\gamma \bullet W)_T$$

[4] Also know as a CIR process, since its use as a model for short-term interest rates by Cox et al. (1985). The square root process goes back to Feller (1951).

[5] We use $\bullet$ to denote the stochastic integral of d-dimensional processes: $H \bullet M = \sum_{i=1}^d \int_0^{\cdot} H_t^i dM_t^i$ for H, M taking values in $\mathbb{R}^d$.

we have that Q is equivalent to P (provided the stochastic exponential is a martingale, i.e. $\mathbb{E}[\mathcal{E}(-\gamma \bullet W)_t] = 1$ for all $t \in [0,T]$, for which Novikov's and Kazamaki's conditions are sufficient. Wong and Heyde (2006) express this explicit in terms of the parameters). Further, by the Girsanov theorem, $\{\tilde{W}_t^1\}_{t\in[0,T]}$ and $\{\tilde{W}_t^2\}_{t\in[0,T]}$ defined by

$$\begin{aligned} d\tilde{W}_t^1 &= \lambda\sqrt{V_t}dt + dW_t^1, \\ d\tilde{W}_t^2 &= \gamma_t^2 dt + dW_t^2, \end{aligned}$$

are independent Wiener processes under Q. By virtue of equation (2.1), this gives the Q-dynamics of the model

$$\begin{aligned} dS_t &= rS_t dt + \sqrt{V_t}S_t(\rho d\tilde{W}_t^1 + \sqrt{1-\rho^2}d\tilde{W}_t^2), \\ dV_t &= (\kappa\theta - [\kappa + \sigma\lambda]\, V_t)\, dt + \sigma\sqrt{V_t}d\tilde{W}_t^1, \end{aligned} \tag{2.2}$$

for $t \in [0,T]$ and we note that the variance dynamics is form invariant under the measure change: V follows a square root process under Q as well with "risk-neutral" parameters $\tilde{\kappa}, \tilde{\theta}, \sigma$ where

$$\tilde{\kappa} = \kappa + \sigma\lambda \quad \text{and} \quad \tilde{\theta} = \frac{\kappa\theta}{\kappa + \sigma\lambda}.$$

We also see that the discounted asset price $B^{-1}S$ will be a Q-martingale (i.e. Q is an equivalent martingale measure) such that the financial market model (B,S) is arbitrage-free. However, as (γ^1, γ^2) may be arbitrarily chosen as long as (2.1) is satisfied, the model is incomplete. This means that λ could be determined by a single exogenously given asset (with a volatility dependent price) to complete the market, and γ^2 is uniquely determined by equation (2.1). Any other contingent claim will then be uniquely priced.

For a European option with payoff $g(S_T)$ at maturity T, we have that the $\mathcal{C}^{1,2}$ function $D(t,s,v)$ of the pricing rule $D_t = D(t,S_t,V_t)$, $t \in [0,T]$, for the option satisfies the following partial differential equation

$$\frac{\partial D}{\partial t} + rs\frac{\partial D}{\partial s} + \{\kappa\theta - v(\kappa+\sigma\lambda)\}\frac{\partial D}{\partial v} + \frac{1}{2}s^2 v\frac{\partial^2 D}{\partial s^2} + \rho\sigma vs\frac{\partial^2 D}{\partial v \partial s} + \frac{1}{2}\sigma^2 v\frac{\partial^2 D}{\partial v^2} = rD, \tag{2.3}$$

with terminal condition $D(T,s,v) = g(s)$. Notice that the expression in curly brackets can be equivalently written $\tilde{\kappa}(\tilde{\theta} - v)$ with the risk-neutral specification of the parameters. Equivalently, by Feynman-Kac, this is to say that we have the risk-neutral pricing formula

$$D(t,s,v) = \mathbb{E}^Q\left[e^{-r(T-t)}g(S_T)\middle|\,(S_t,V_t) = (s,v)\right]$$

where (S,V) follows the Q-dynamics with initial value $(S_t,V_t) = (s,v)$ at the initial time $t \in [0,T]$.

This pricing equation (2.3) is the same as in Heston's original paper if we let $\lambda_v = \sigma\lambda$ and $\lambda(t,S_t,V_t) = \lambda_v V_t$ being the price of volatility risk used in his exposition. The equation is solved by the Fourier transform of the price and the resulting "semi-closed" pricing formula

is obtained by the inverse transform. In practice, however, the inverse transform has to be calculated by numerical integration methods.

If the underlying asset pays a constant dividend yield δ (continuously compounded) we have the S-dynamics under Q

$$dS_t = (r-\delta)S_t dt + \sqrt{V_t}S_t(\rho d\tilde{W}_t^1 + \sqrt{1-\rho^2}d\tilde{W}_t^2)$$

since an investment in the asset should have an instantaneous return being the same as the risk-free rate under the equivalent martingale measure. The pricing function $D(t,s,v)$ for an option (without dividend payments) will then satisfy the pricing equation (2.3) with the second term $rs\partial D/\partial s$ substituted with $(r-\delta)s\partial D/\partial s$. One way of seeing this is to apply Itô's lemma to $B^{-1}D$ under Q and set the drift term to zero, as the discounted price process of any tradable asset must be a Q-martingale for the non-arbitrage condition to hold.

2.2.2 Pricing bounds under parameter uncertainty

Heston's model (and any other stochastic volatility model) is fundamentally a model for the underlying financial market even if it is predominantly used for option pricing purposes. The pricing measure is often taken as being fixed for convenience, for instance through model-to-market calibration of option prices, and the connection to the objective measure is not important for the analysis, and hence not necessarily made to be explicit.

Although we are dealing with option pricing as well, we will take a slightly converse approach where the pricing measure inherits its uncertainty from the objective measure. For this purpose, we assume a pricing measure Q to be given momentarily, just to be able to replace it with another pricing measure that is subject to uncertainty. Here we infer uncertainty of pricing parameters from statistical estimation of objective parameters and thus, the relation between the measures will play an integral role.

To this end, we introduce parameter uncertainty in our model by modifying our reference measure with the effect of a control that governs the parameter processes. Namely, we replace the risk-neutral measure Q with an equivalent measure Q^u under which we have the controlled dynamics as given by

$$\begin{aligned} dS_t &= r^u(u_t)S_t dt + \sqrt{V_t}S_t(\rho dW_t^{u1} + \sqrt{1-\rho^2}dW_t^{u2}), \\ dV_t &= \kappa^u(u_t)\left(\theta^u(u_t) - V_t\right)dt + \sigma\sqrt{V_t}dW_t^{u1}, \end{aligned} \tag{2.4}$$

for $t \in [0,T]$. The control process $\{u_t\}_{t\geq 0}$ is an $\mathcal{F}_t$-predictable process that takes values in a compact set $U \subset \mathbb{R}^3$, which we will call the parameter uncertainty set. We write $\mathcal{U}$ for the space of admissible control processes (that is, predictable processes taking values in U) and under Q^u, we have that W^{u1} and W^{u2} are independent Wiener processes, as will be explained in a moment. The control process realises its paths stochastically, and we simply do not know beforehand which $\{u_t\}_{t\geq 0} \in \mathcal{U}$ will be governing (2.4): the

uncertainty is tantamount to this choice. Furthermore, we denote the components of the control $\{u_t\}_{t\geq 0} = \{r_t, \kappa_t, \theta_t\}_{t\geq 0}$ and let the controlled drift-functions of (2.4), all $f : U \to \mathbb{R}^+$, be defined as

$$r^u(u_t) = r_t, \qquad \kappa^u(u_t) = \kappa_t + \sigma\lambda \quad \text{and} \quad \theta^u(u_t) = \frac{\kappa_t \theta_t}{\kappa_t + \sigma\lambda}. \tag{2.5}$$

Notice that this specification of the controlled drift relies on the premise that the Q-parameters $r, \tilde{\kappa}, \tilde{\theta}$ are subject to parameter uncertainty by their replacement with r^u, κ^u, θ^u. The uncertainty is in turn taken to be inferred from statistical estimation of the objective P-parameters, represented by $(r_t, \kappa_t, \theta_r) \in U$ where U is the statistical confidence interval, and transferred to the pricing parameters by the map

$$U \ni u_t \mapsto (r^u(u_t), \kappa^u(u_t), \theta^u(u_t)) \in U^\lambda$$

as given by (2.5). Here U^λ is the uncertainty set for the controlled parameters, induced by the same mapping. The parameter λ associated with Q thus plays an instrumental role in facilitating the uncertainty transfer and it determines the set U^λ where our uncertain price-parameters live. In practice, we forcefully set $\lambda = 0$ to obtain that the uncertainty in price-parameters is exactly that of the uncertainty in estimated real-world parameters, i.e. $U^\lambda \equiv U$. However, note that this does not imply $P \equiv Q$ nor $\mu = r$, cf. equation (2.1).

With the controlled dynamics of (S, V) representing the model under influence of parameter uncertainty, we proceed to define what we mean with the upper and lower boundary for the price of a European option. Namely, for an option written on S with terminal payoff at time T given by a square-integrable $\mathcal{F}_T$-measurable random variable G, we will take the most conservative prices from valuation under the controlled pricing measures (i.e. under parameter uncertainty) as given by the control problems

$$D_t^- = \operatorname*{ess\,inf}_{\{u_t\}\in\mathcal{U}} \mathbb{E}_u \left[e^{-\int_t^T r_s ds} G \middle| \mathcal{F}_t \right] \quad \text{and} \quad D_t^+ = \operatorname*{ess\,sup}_{\{u_t\}\in\mathcal{U}} \mathbb{E}_u \left[e^{-\int_t^T r_s ds} G \middle| \mathcal{F}_t \right] \tag{2.6}$$

for $t \in [0, T]$, where $\mathbb{E}_u(\cdot|\mathcal{F})$ denotes the conditional expectation under Q^u. In a sense, we thus consider the super-replication costs of selling a long/short position in the option when the uncertain parameters evolve stochastically in the uncertainty set, in an optimal way.[6]

In order to find a pricing PDE that corresponds to equation (2.3) of the previous section, we henceforth consider payoffs given by $G = g(S_T)$ for some non-negative function g and for simplicity, we assume g to be bounded. Due to the Markovian structure of the problem,

[6]This draws on the interpretation that $\{Q^u : u \in \mathcal{U}\}$ is the set of equivalent martingale measures of an incomplete market model, such that the most conservative risk-neutral price of an option equals the super-replication cost of a short position in the same: with $\Pi_t(G) = \inf_\phi\{\tilde{V}_t(\phi) : V_T(\phi) \geq G, \text{ a.s.}\}$ being the discounted portfolio value of the (cheapest) admissible strategy ϕ that super-replicates G, then $\Pi_t(G) = \operatorname{ess\,sup}_{u\in\mathcal{U}} \mathbb{E}_u[\tilde{G}|\mathcal{F}_t]$ and the supremum is attained. See for instance Cont and Tankov (2004), Section 10.2.

we then have that the optimal value processes will be functions of the current asset price and variance state

$$D_t^- = D^-(t, S_t, V_t),$$
$$D_t^+ = D^+(t, S_t, V_t),$$

for some continuous functions $D^{\pm} : [0,T] \times \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}$. As we will see later, these functions will satisfy a semilinear version of the standard pricing equation for European options. However, before we arrive at more precise expressions for the optimally controlled value processes (and their generating functions) we take one step backwards: we will first consider the value process for a fixed control and its link to a backward stochastic differential equation. Following the approach due to Quenez (1997) as outlined in Cohen and Elliott (2015), we then consider the optimally controlled value process as the solution to a closely related BSDE.

With the intention of finding the pricing-bound functions through a dual formulation with BSDEs, we begin by defining the effect of the control as the $\mathbb{R}^2$-valued process

$$\alpha(S_t, V_t, u_t) = \frac{1}{\sigma\sqrt{V_t}} \begin{bmatrix} \kappa^u(u_t)\theta^u(u_t) - \tilde{\kappa}\tilde{\theta} - (\kappa^u(u_t) - \tilde{\kappa})V_t \\ \frac{-\rho\left(\kappa^u(u_t)\theta^u(u_t) - \tilde{\kappa}\tilde{\theta} - (\kappa^u(u_t) - \tilde{\kappa})V_t\right) + \sigma(r_t - r)}{\sqrt{1-\rho^2}} \end{bmatrix}. \tag{2.7}$$

The stochastic exponential of the process $\alpha(S,V,u)^\top \bullet (\tilde{W}^1, \tilde{W}^2)$ is then what defines the measure change $Q \to Q^u$ on $\mathcal{F}_T$,

$$\frac{dQ^u}{dQ} = \mathcal{E}\left(\int_0^\cdot \alpha_1(S_t, V_t, u_t)d\tilde{W}_t^1 + \int_0^\cdot \alpha_2(S_t, V_t, u_t)d\tilde{W}_t^2\right)_T$$

where α_1 and α_2 are the two components of the effect process (provided $\mathcal{E}(\alpha^\top \bullet \tilde{W})$ is a martingale). By Girsanov's theorem

$$dW_t^{1u} = d\tilde{W}_t^1 - \alpha_1(t, S_t, V_t)dt,$$
$$dW_t^{2u} = d\tilde{W}_t^2 - \alpha_2(t, S_t, V_t)dt,$$

thus defines the two independent Wiener processes under Q^u. Further, we define the linear driver function $f : (0,\infty) \times (0,\infty) \times \mathbb{R} \times \mathbb{R}^{1\times 2} \times U \to \mathbb{R}$ as

$$f(s, v, y, z, u) = \varrho(s, v, u)y + z\alpha(s, v, u) \tag{2.8}$$

where $\varrho(s, v, u) \equiv -r$, that is, the (negative) first component of the control which represents the risk-free interest rate.

With the driver defined by (2.7)-(2.8), we now have the following representation of the option value subject to parameter uncertainty from Q^u: for a given, fixed control process $u = \{u_t\}_{t\geq 0} \in \mathcal{U}$, the controlled value process as given by

$$J_t(u) = \mathbb{E}_u\left[e^{-\int_t^T r_s ds} g(S_T)\middle|\, \mathcal{F}_t\right], \quad t \in [0,T],$$

is the unique solution to the linear Markovian backward stochastic differential equation

$$\begin{aligned} dJ_t(u) &= -f(S_t, V_t, J_t(u), Z_t, u_t)dt + Z_t d\tilde{W}_t, \\ J_T(u) &= g(S_T), \end{aligned} \tag{2.9}$$

where $Z = (Z^1, Z^2)^\top$ – the martingale representation part of Y – is a process taking values in $\mathbb{R}^{1\times 2}$ (being a part of the solution to the BSDE). To see this, consider the process $J(u)$ that solves (2.9) and let $\mathcal{E}(\Gamma)$ be the stochastic exponential of

$$\Gamma = \int_0^\cdot -r_t dt + \int_0^\cdot \alpha(S_t, V_t, u_t)^\top d\tilde{W}_t.$$

Apply Itô's product rule to $\mathcal{E}(\Gamma)J(u)$ to obtain

$$d\left(\mathcal{E}(\Gamma)_t J_t(u)\right) = \mathcal{E}(\Gamma)_t \left(Z_t + J_t(u)\alpha(S_t, V_t, u_t)^\top\right) d\tilde{W}_t$$

and thus, since $\mathcal{E}(\Gamma)J(u)$ is a martingale under Q, we have

$$\begin{aligned} J_t(u) &= \frac{1}{\mathcal{E}(\Gamma)_t}\mathbb{E}^Q\left[\mathcal{E}(\Gamma)_T\, g(S_T)\middle|\, \mathcal{F}_t\right] \\ &= e^{\int_0^t r_s ds}\frac{1}{\mathcal{E}(\alpha^\top \bullet \tilde{W})_t}\mathbb{E}^Q\left[e^{-\int_0^T r_s ds}\mathcal{E}(\alpha^\top \bullet \tilde{W})_T\, g(S_T)\middle|\, \mathcal{F}_t\right] \\ &= \mathbb{E}_u\left[e^{-\int_t^T r_s ds} g(S_T)\middle|\, \mathcal{F}_t\right] \end{aligned}$$

as $\mathcal{E}(\alpha^\top \bullet \tilde{W})$ is the density for the measure change $Q \to Q^u$.

The BSDE (2.9) governs the value process under the impact of a fixed parameter control process that evolves in the uncertainty set U. To obtain the lowest (highest) value scenario, the value process is to be minimised (maximised) over all admissible controls in $\mathcal{U}$ and as we retail in the following, this is done through pointwise optimisation with respect to $u \in U$ of the driver function for the value process. Hence, we define the following drivers optimised over the parameter uncertainty set

$$H^-(s,v,y,z) = \operatorname*{ess\,inf}_{u\in U} f(s,v,y,z,u) \quad \text{and} \quad H^+(s,v,y,z) \quad = \operatorname*{ess\,sup}_{u\in U} f(s,v,y,z,u),$$

where we note that as U is compact, the infimum and supremum are both attained. We then have the following main result which is due to the comparison principle for BSDEs: the lower/upper optimally controlled value processes $\{D_t^\pm\}_{t\in[0,T]} = \{\pm \operatorname{ess\,sup}_{u\in\mathcal{U}} \pm J_t(u)\}_{t\in[0,T]}$ have cadlag modifications that are the unique solutions of the BSDEs

$$\begin{aligned} dD_t^\pm &= -H^\pm(S_t, V_t, D_t^\pm, Z_t)dt + Z_t d\tilde{W}_t, \\ D_T^\pm &= g(S_T). \end{aligned} \tag{2.10}$$

In particular, the processes are equal to deterministic functions of (t, S_t, V_t), that is, $D_t^\pm = D^\pm(t, S_t, V_t)$ for some continuous functions $D^\pm : [0,T]\times\mathbb{R}^+\times\mathbb{R}^+ \to \mathbb{R}$. As the

infimum (supremum) of H is attained, we further have that there exists optimal controls $\{u_t^{\pm *}\}_{t\in[0,T]} \in \mathcal{U}$ which are optimal *feedback* controls. This means that the processes

$$u_t^{\pm *} = u^{\pm *}(t, S_t, V_t), \quad t \in [0, T],$$

are the optimal controls among all predictable controls for some deterministic functions $u^{\pm *} : [0, T] \times \mathbb{R}^+ \times \mathbb{R}^+ \to U$. Finally, by the semilinear Feynman-Kac formula (provided a solution exists), we have that $D^-(t, s, v)$ satisfies the following semilinear parabolic PDE

$$\frac{\partial D}{\partial t} + \frac{1}{2}s^2 v \frac{\partial^2 D}{\partial s^2} + \rho\sigma v s \frac{\partial^2 D}{\partial v \partial s} + \frac{1}{2}\sigma^2 v \frac{\partial^2 D}{\partial v^2} + \operatorname*{ess\,inf}_{(r,\kappa,\theta)\in U} \left\{ -rD + rs\frac{\partial D}{\partial s} + \kappa^u(\kappa)\left(\theta^u(\theta) - v\right)\frac{\partial D}{\partial v} \right\} = 0 \tag{2.11}$$

with terminal value $D^-(T, s, v) = g(s)$. In the corresponding equation for $D^+(t, s, v)$ we have a supremum substituted for the infimum.

Proof: For the first part of the result, since $H^-(s, v, y, z) \leq f(s, v, y, z, u)$ by definition, we have that the (unique) solution[7] Y to the BSDE with data $(g(S_T), H^-)$ satisfies $Y_t \leq J_t(u)$ for all controls $u \in \mathcal{U}$ (up to indistinguishability). This is a consequence of the comparison theorem for BSDEs (see e.g. Cohen and Elliott (2015), Theorem A.9.20). Further, by Filippov's implicit function theorem (Cohen and Elliott (2015), Theorem 21.3.4), for each $\epsilon > 0$ there exists a predictable control $u^\epsilon \in \mathcal{U}$ such that $f(s, v, y, z, u^\epsilon) \leq H^-(s, v, y, z) + \epsilon$. Since $Y_t + \epsilon(T - t)$ solves the BSDE with driver $H^-(s, v, y, z) + \epsilon$, the comparison theorem yields $J_t(u^\epsilon) \leq Y_t + \epsilon(T - t)$ (up to indistinguishability) and we have the inclusion

$$Y_t \leq J_t(u^\epsilon) \leq Y_t + \epsilon(T - t).$$

Letting $\epsilon \to 0$ we have that $Y_t = \operatorname{ess\,inf}_{u\in\mathcal{U}} J_t(u) = D_t^-$ for every t which is to say that Y is a version of the optimal value process. That D_t^- can be written as a continuous function of (t, S_t, V_t) is due to the fact that (2.10) is a Markovian BSDE. Further, as we have that the optimal control is attainable, Filippov's theorem gives that it is a function of $(t, S_t, V_t, D_t^-, Z_t)$ where $Z_t = z(t, S_t, V_t)$ – due to the Markovian BSDE[8] – and we have the result that $u_t^{-*} = u^{-*}(t, S_t, V_t)$ for a deterministic function. □

To obtain an expression for the optimised driver $H^\pm$, we note that the driver of the value function is conveniently expressed in terms of divergence of the control from the statistical

[7] As we assume the driver f to be sufficiently integrable for the $J(u)$-BSDE to admit a unique solution (i.e. it is a stochastic Lipschitz driver) the integrability carries over to H such that the Y-BSDE admits a unique solution as well.

[8] The function for the martingale representation Z is obtained explicitly by applying Itô's lemma to $D_t = D(t, S_t, V_t)$ and using the semilinear pricing PDE (2.11), which gives

$$dD(t, S_t, V_t) = -H(S_t, V_t, D_t, Z_t)dt + \partial_x D(t, S_t, V_t)\sigma(S_t, V_t)d\tilde{W}_t$$

where $\partial_x f \equiv (\partial_s f, \partial_v f)$ and $\sigma(s, v)$ should be understood as the diffusion matrix of (2.2). Hence, by uniqueness of the BSDE solution, $z(t, s, v) \equiv \partial_x D(t, s, v)\sigma(s, v)$ is the deterministic generating function for Z.

parameters. By rearrangement of (2.8)

$$\begin{aligned} f(S_t, V_t, Y_t, Z_t, u_t) &= (r_t - r)\left(\frac{Z_t^2}{\sqrt{1-\rho^2}\sqrt{V_t}} - Y_t\right) + (\kappa_t - \kappa)\left(\frac{-Z_t^1\sqrt{V_t}}{\sigma} + \frac{\rho Z_t^2\sqrt{V_t}}{\sigma\sqrt{1-\rho^2}}\right) \\ &+ (\kappa_t\theta_t - \kappa\theta)\left(\frac{Z_t^1}{\sigma\sqrt{V_t}} - \frac{\rho Z_t^2}{\sigma\sqrt{1-\rho^2}\sqrt{V_t}}\right) - rY_t \end{aligned} \tag{2.12}$$

since $\kappa^u(u_t) - \tilde{\kappa} = \kappa_t - \kappa$ and $\kappa^u(u_t)\theta^u(u_t) - \tilde{\kappa}\tilde{\theta} = \kappa_t\theta_t - \kappa\theta$, which is due to the linear form of the drift (and the simple form of the parameter change under $P \to Q$, regardless of the value of λ). If we let $\beta_t \equiv \kappa_t\theta_t$ and use the parametrisation $(r_t, \kappa_t, \beta_t) \mapsto (r_t, \kappa_t, \theta_t)$, we have that the driver is a linear function of the divergence $\tilde{u}_t = (r_t - r, \kappa_t - \kappa, \beta_t - \beta)$. Hence, the optimal drivers $H^{\pm}$ are obtained by minimising/maximising a linear objective subject to the constraint given by the compact uncertainty set U for the original P-parameters. In particular, from statistical inference of the estimation, we have that the $1-\alpha$ confidence ellipse

$$\tilde{u}^\top \Sigma_{r,\kappa,\beta}^{-1} \tilde{u} \leq \chi_3^2(1-\alpha) \tag{2.13}$$

represents $u \in U$ (for a significance level α) where $\Sigma_{r,\kappa,\beta}$ is the covariance matrix of the parameters and $\chi_3^2(1-\alpha)$ is the quantile of the chi-squared distribution with three degrees of freedom (see further Section 2.4.1). As $\tilde{u} \mapsto f(\tilde{u})$ is linear, it has no internal stationary points and the quadratic problems

$$\begin{aligned} &H^- = \inf f(\tilde{u}) \quad \text{and} \quad H^+ = \sup f(\tilde{u}) \\ &\text{subject to } \tilde{u}^\top \Sigma_{r,\kappa,\beta}^{-1} \tilde{u} = \chi_3^2(1-\alpha) \end{aligned}$$

give the optimised drivers. The solutions are (obtained e.g. by a Lagrange multiplier)

$$\begin{aligned} H^{\pm}(S_t, V_t, Z_t, Y_t) &= \pm\sqrt{\chi_3^2(1-\alpha) n_t^\top \Sigma_{r,\kappa,\beta}^\top n_t} - rY_t \\ \tilde{u}^{\pm}(S_t, V_t, Z_t, Y_t) &= \pm\sqrt{\frac{\chi_3^2(1-\alpha)}{n_t^\top \Sigma_{r,\kappa,\beta}^\top n_t}} \Sigma_{r,\kappa,\beta} n_t \end{aligned} \tag{2.14}$$

where n_t is the 3×1 vector of coefficients to the parameter deviances of equation (2.12) given by

$$n_t = \left[\left(\frac{Z_t^2}{\sqrt{1-\rho^2}\sqrt{V_t}} - Y_t\right), \left(\frac{-Z_t^1\sqrt{V_t}}{\sigma} + \frac{\rho Z_t^2\sqrt{V_t}}{\sigma\sqrt{1-\rho^2}}\right), \left(\frac{Z_t^1}{\sigma\sqrt{V_t}} - \frac{\rho Z_t^2}{\sigma\sqrt{1-\rho^2}\sqrt{V_t}}\right)\right]^\top.$$

The optimal drivers in (2.14) conclude our analysis since we now have an explicit form for the stochastic differential equation (2.10) that describes the evolution of the pricing boundaries. Before we proceed to obtaining approximative solutions of these equation by numerical methods, a few remarks are in order. Firstly, the approach applies unchanged to a portfolio of options with time-T terminal payoff $\sum_i g_i(S_T)$. Due to the non-linearity of the pricing boundaries (2.6), we further have $D^+(\sum w_i g_i(S_T)) \leq \sum w_i D^+(g_i(S_T))$ for

weights $\sum w_i = 1$, such that the super-replication cost for individual hedging might be lowered by hedging the portfolio as a whole. Secondly, for a general payoff represented by $G \in \mathcal{F}_T$, for instance a path-dependent European options on S, we have a value process equation corresponding to (2.10) with terminal condition $D_T^{\pm} = G$. However, this problem do no longer yield a Markovian structure, and we do not have $D^{\pm}$ (nor Z) being generated by deterministic functions, neither does the numerical methods of Section 2.3 apply. Thirdly, we deliberately impose parameter uncertainty by replacing $Q \to Q^u$ in contrast to replacing $P \to Q^u$ directly (which would yield the same form of the effect $\alpha(t, S_t, V_t)$ that governs the measure change, but with r replaced by μ in (2.7)). The reason is that the governing BSDEs (2.10) will have a terminal condition $g(S_T)$ where $S_T \sim Q$, for which we have accessible parameters (in particular, we may directly observe the Q-drift r instead of estimating the P-drift μ). Fourthly, the worst-case approach with pricing boundaries is not the only way of accounting for parameter uncertainty in option pricing. An interesting alternative that takes into account beliefs about the likelihood of the uncertainty is the Bayesian approach outlined in Gupta et al. (2010). We conclude this section with two more remarks.

Remark 2.1. So far, we have not expressed any integrability conditions on the process $\alpha(S_t, V_t, u_t)$ in order to guarantee that (i) the density dQ^u/dQ and (ii) the driver $f(S_t, V_t, Y_t, Z_t, u_t)$ are well defined, i.e. for the measure change $Q \to Q^u$ to be eligible and to certify that f (and hence $H^{\pm}$) yields a BSDE which admits a unique solution. For this purpose, Novikov's condition

$$\mathbb{E}^Q \left[e^{\frac{1}{2}\int_0^T ||\alpha(S_t,V_t,u_t)||^2 dt} \right] < \infty \tag{2.15}$$

is sufficient for both (i) and (ii) since then we have that the driver is stochastic Lipschitz in y and z (note that r_t is bounded in U), i.e.

$$|f(S_t, V_t, y, z, u_t) - f(S_t, V_t, y', z', u_t)| \leq ||\alpha(S_t, V_t, u_t)|| \left(|y - y'| + ||z - z'|| \right)$$

where $||\alpha(S_t, V_t, u_t)||$ is predictable and such that (2.15) holds. With a stochastic Lipschitz driver, the concerned BSDE admits a unique solution which is bounded if the terminal condition $g(S_T)$ is bounded (see Cohen and Elliott (2015), Chapter A.9.2).

For Novikov's condition in (2.15) we note that the integrand of the exponent can be written

$$||\alpha(S_t, V_t, u_t)||^2 = a(u_t)V_t + b(u_t)\frac{1}{V_t} + c(u_t)$$

with

$$\begin{aligned}
a(u_t) &= \frac{(\kappa_t - \kappa)^2}{\sigma^2(1-\rho^2)} \\
b(u_t) &= \frac{\sigma^2(r_t - r)^2 + (\kappa_t\theta_t - \kappa\theta)^2 + 2\rho\sigma(r - r_t)(\kappa_t\theta_t - \kappa\theta)}{\sigma^2(1-\rho^2)} \\
c(u_t) &= -2\frac{(\sigma\rho(r - r_t) + \kappa_t\theta_t - \kappa\theta)(\kappa_t - \kappa)}{\sigma^2(1-\rho^2)}
\end{aligned}$$

such that for the expectation in (2.15) we have

$$\mathbb{E}^Q\left[e^{\frac{1}{2}\int_0^T a(u_t)V_t dt} e^{\frac{1}{2}\int_0^T b(u_t)\frac{1}{V_t}dt} e^{\frac{1}{2}\int_0^T c(u_t)dt}\right] \leq k\sqrt{\mathbb{E}^Q\left[e^{\int_0^T a(u_t)V_t dt}\right]\mathbb{E}^Q\left[e^{\int_0^T b(u_t)\frac{1}{V_t}dt}\right]} \tag{2.16}$$

since $e^{\frac{1}{2}\int_0^T c(u_t)dt}$ is bounded by a constant k (for $u_t \in U$) and where we have used the Cauchy-Schwarz inequality. If we begin with the first expectation on the right hand side of (2.16) we have $a(u_t) \leq \bar{a}$ for a constant $\bar{a}$. As the Laplace transform of the integrated CIR process[9] is finite for $\bar{a} \leq \tilde{\kappa}^2/(2\sigma^2)$, we end up with the condition

$$|\kappa_t - \kappa| \leq \tilde{\kappa}\frac{\sqrt{1-\rho^2}}{\sqrt{2}}. \tag{2.17}$$

For the second expectation of (2.16), we use that $b(u_t) \leq \bar{b}$ for a constant $\bar{b}$ and that the Laplace transform of the integrated inverse-CIR process [10] is finite for

$$\bar{b} \leq \left(\frac{2\tilde{\kappa}\tilde{\theta} - \sigma^2}{2\sqrt{2}\sigma}\right)^2. \tag{2.18}$$

Rearranging this condition, we have that

$$\sigma^2(r_t - r)^2 + (\kappa_t\theta_t - \kappa\theta)^2 + 2\rho\sigma(r - r_t)(\kappa_t\theta_t - \kappa\theta) \leq \frac{1-\rho^2}{2}\left(\kappa\theta - \sigma^2/2\right)^2$$

together with (2.17) are sufficient conditions for (2.15) to hold.

[9] The Laplace transform of the integrated variance $\mathbb{E}[\exp(-\beta\int_0^T V_t dt)]$ goes back to Cox et al. (1985) and is well defined for $-\beta \leq \kappa^2/(2\sigma^2)$, see also Carr et al. (2003).

[10] Carr and Sun (2007) gives an expression for the joint transform of the log-price and integrated variance of a 3-over-2 process. Applying Itô's formula to $1/V_t$ we find that the inverse-CIR (κ, θ, σ) process is a 3-over-2 process with parameters $(\hat{\kappa} \equiv \kappa\theta - \sigma^2, \hat{\theta} \equiv \kappa/(\kappa\theta - \sigma^2), \hat{\sigma} \equiv -\sigma)$. Using their transform, provided $\hat{\kappa} > -\hat{\sigma}^2/2$,

$$\mathbb{E}\left[e^{-\lambda\int_0^T \frac{1}{V_t}dt}\right] = \frac{\Gamma(\gamma-\alpha)}{\Gamma(\gamma)}\left(\frac{2}{\hat{\sigma}^2 y(0,1/V_0)}\right)^{\alpha} M\left(\alpha, \gamma, -\frac{2}{\hat{\sigma}^2 y(0,1/V_0)}\right)$$

where

$$\begin{aligned} y(t,x) &\equiv x(e^{\hat{\kappa}\hat{\theta}(T-t)} - 1)/(\hat{\kappa}\hat{\theta}) = x(e^{\kappa(T-t)} - 1)/\kappa \\ \alpha &\equiv -(1/2 + \hat{\kappa}/\sigma^2) + \sqrt{(1/2 + \hat{\kappa}/\sigma^2)^2 + 2\lambda/\sigma^2} \\ \gamma &\equiv 2(\alpha + 1 + \hat{\kappa}/\sigma^2) = 1 + 2\sqrt{(1/2 + \hat{\kappa}/\sigma^2)^2 + 2\lambda/\sigma^2} \end{aligned}$$

and M is the confluent hypergeometric function. From this, we see that

$$\lambda \geq -\left(\frac{2\hat{\kappa} + \sigma^2}{2\sqrt{2}\sigma}\right)^2 = -\left(\frac{2\kappa\theta - \sigma^2}{2\sqrt{2}\sigma}\right)^2$$

is a sufficient condition for the transform to being well defined.

Remark 2.2. As a final point, we note that the above framework may be adapted to a situation where the underlying asset pays a continuously compounded dividend yield δ. A predictable dividend process in then added to the control process $\{u_t\}_{t\geq 0} = \{r_t, \delta_t, \kappa_t, \theta_t\}_{t\geq 0}$ taking values in a compact set $U \subset \mathbb{R}^4$. The change of measure $Q \to Q^u$ is given by the effect process

$$\alpha(S_t, V_t, u_t) = \frac{1}{\sigma\sqrt{V_t}} \begin{bmatrix} \kappa_t\theta_t - \kappa\theta - (\kappa_t - \kappa)V_t \\ \frac{-\rho(\kappa_t\theta_t - \kappa\theta - (\kappa_t - \kappa)V_t) + \sigma(r_t - r - \delta_t + \delta)}{\sqrt{1-\rho^2}} \end{bmatrix}$$

such that the asset price has a controlled drift coefficient $(r_t - \delta_t)S_t dt$ under Q^u. The pricing bounds are defined by the optimally controlled value processes and as in the zero-dividend case, they are given as solutions to the optimal value-process BSDE (2.10) with the driver H being (2.8) pointwise optimised over U. A before, this yields a linear optimisation problem in terms of the control's divergence from statistical parameters with a solution corresponding to the one given in (2.14):

$$H^{\pm}(S_t, V_t, Z_t, Y_t) = \pm\sqrt{\chi^2_4(1-\alpha) n_t^\top \Sigma^\top_{r,\delta,\kappa,\beta} n_t} - rY_t$$

where $\Sigma_{r,\delta,\kappa,\beta}$ is the covariance matrix of the parameters and n_t the 4×1 vector

$$n_t = \left[\left(\frac{Z_t^2}{\bar{\rho}\sqrt{V_t}} - Y_t \right), \left(\frac{Z_t^2}{\bar{\rho}\sqrt{V_t}} \right), \left(\frac{-Z_t^1\sqrt{V_t}}{\sigma} + \frac{\rho Z_t^2\sqrt{V_t}}{\sigma\bar{\rho}} \right), \left(\frac{Z_t^1}{\sigma\sqrt{V_t}} - \frac{\rho Z_t^2}{\sigma\bar{\rho}\sqrt{V_t}} \right) \right]^\top$$

where we denote $\bar{\rho} \equiv \sqrt{1-\rho^2}$. The optimal control/divergence $\tilde{u}^{\pm}(S_t, V_t, Z_t, Y_t) = (r_t - r, \delta_t - \delta, \kappa_t - \kappa, \beta_t - \beta)$ is given by the corresponding expression of (2.14).

2.3 Numerical methods for BSDEs

The optimally controlled value process (or the value process for a fixed feedback control, i.e. $u_t = u(t, S_t, V_t)$ for a deterministic function u) is given by the solution to the decoupled forward-backward stochastic differential equation (2.2)-(2.10). In general, there is not much hope of finding closed-form solutions to neither forward nor backward SDEs and one typically has to consider numerical methods. For our purposes, we consider the simulation technique by Bouchard and Touzi (2004).

2.3.1 The simulation scheme by Bouchard and Touzi

For a time gird $\pi : 0 = t_0 < \cdots < t_n = T$, Bouchard and Touzi (2004) propose a method to generate a discrete-time approximation (X^π, Y^π) of the solution to a decoupled equation with forward component X and backward component Y. In the first part of the scheme, the forward component X^π is simulated over the time grid π with a standard

Euler-Maruyama approximation to generate N paths of X^π (see e.g. Kloeden and Platen (1992)). The component Y^π is then generated by the backward induction

$$\begin{aligned} Y^\pi_{t_n} &= g(X^\pi_{t_n}) \\ Z^\pi_{t_{i-1}} &= \frac{1}{\Delta_i}\mathbb{E}\left[Y^\pi_{t_i}\Delta W_{t_i} \middle| X^\pi_{t_{i-1}} \right] \\ Y^\pi_{t_{i-1}} &= \mathbb{E}\left[Y^\pi_{t_i} + f(X^\pi_{t_{i-1}}, Y^\pi_{t_{i-1}}, Z^\pi_{t_{i-1}})\Delta_i \middle| X^\pi_{t_{i-1}} \right] \end{aligned} \tag{2.19}$$

where $\Delta_i \equiv t_i - t_{i-1}$ and $\Delta W_{t_i} \equiv W_{t_i} - W_{t_{i-1}}$ are the i^{th} time- and Wiener increments from the generation of X^π. The last equation in (2.19) is obtained by applying $\mathbb{E}[\cdot|\mathcal{F}_{t_{i-1}}]$ to the following simple discretization of the BSDE

$$Y^\pi_{t_i} - Y^\pi_{t_{i-1}} = -f(X^\pi_{t_{i-1}}, Y^\pi_{t_{i-1}}, Z^\pi_{t_{i-1}})\Delta_i + Z^\pi_{t_{i-1}}\Delta W_{t_i} \tag{2.20}$$

and using the Markov property of X and the fact that Y_t and Z_t are both deterministic functions of X_t for all $t \in [0,T]$. The second equation for Z is obtained similarly by multiplying (2.20) with ΔW_{t_i} and taking conditional expectations.

For the backward induction (2.19) one has to compute the conditional expectations and to make the scheme operational, this is made with an approximation $\hat{E}[\cdot|X^\pi_{t_{i-1}}]$ of the regression function $\mathbb{E}[\cdot|X^\pi_{t_{i-1}}]$ based on simulated training data. That is, the data

$$\{Y^{\pi(j)}_{t_i}, \Delta W^{(j)}_{t_i}, X^{\pi(j)}_{t_{i-1}}\}_{1\le j\le N}$$

is used for the first regression in (2.19) where $X^{\pi(j)}$ is the j^{th} simulated path of X^π and $Y^{\pi(j)}$ is the corresponding value from the induction of the previous time step. For the second regression, $\{Y^{\pi(j)}_{t_i}, Z^{(j)}_{t_{i-1}}, X^{\pi(j)}_{t_{i-1}}\}_{1\le j\le N}$ is used accordingly.

As an example of a non-parametric regression estimator, it is suggested to use the Nadaraya-Watson weighted average for a kernel estimator. We conveniently employ the k-nearest neighbour kernel for this purpose: for $X^\pi_{t_{i-1}}$ and $\xi \in \mathcal{F}_{t_i}$, each with N simulated outcomes, we approximate $\mathbb{E}[\xi|X^\pi_{t_{i-1}} = X^{\pi(j)}_{t_{i-1}}]$, $j = 1,\dots,N$, with

$$\hat{E}\left[\xi \middle| X^{\pi(j)}_{t_{i-1}}\right] = \frac{\sum_{l=1}^N \xi^{(l)}\mathbf{1}\left(||X^{\pi(l)}_{t_{i-1}} - X^{\pi(j)}_{t_{i-1}}|| \le d^{(j)}_k\right)}{k+1} \tag{2.21}$$

where $d^{(j)}_k$ is the distance between $X^{\pi(j)}_{t_{i-1}}$ and its k^{th} nearest neighbour, and $\mathbf{1}(\cdot)$ is the indicator function. The regression (2.21) together with (2.19) yields an implicit simulation method and as a last step of the scheme, it is suggested to truncate $Z^\pi_{t_{i-1}}$ and $Y^\pi_{t_{i-1}}$ if one has appropriate (possibly t and $X^\pi_{t_{i-1}}$ dependent) bounds for $\mathbb{E}[Y^\pi_{t_i}\Delta W_{t_i}|X^\pi_{t_{i-1}}]$, $\mathbb{E}[Y^\pi_{t_i}|X^\pi_{t_{i-1}}]$ and $Y_{t_{i-1}}$.

The k-nearest neighbours estimator (2.21) is a method that approximates the regression function in a local neighbourhood with a constant. As such, it has low bias but high variance (it is "wiggly" and unstable) and it suffers from the curse of dimensionality at

the boundaries. For an alternative regression estimator, we consider the MARS method[11] (multivariate adaptive regression spines) which uses piecewise linear basis functions in an adaptive manner to approximate the regression function. The model has the linear form

$$\hat{E}\left[\xi \left| X^{\pi}_{t_{i-1}} \right.\right] = \beta_0 + \sum_{m=1}^{M} \beta_m h_m(X^{\pi}_{t_{i-1}}) \tag{2.22}$$

where each basis function $h_m(X)$ is in the candidate set of paired piecewise linear splines: $(X^k - \eta)^+$, $(\eta - X^k)^+$ with $k = 1, 2$ referring to the components of X. The knots η are placed at any value in the set of X-observations: $\eta \in \{X^{k,\pi(j)}_{t_{i-1}}\}_{j=1}^{N}$ for $k = 1, 2$, and $h_m(X)$ are allowed to be d-times products of the splines as well, where d is limited by the model degree. The model (2.22) is built up in a sequential manner by adding a new spline from the candidate set in each stage based on the largest reduction of residual error. All coefficients $\beta_0, \beta_1, \ldots$ are estimated in each stage by least squares and the process continues until the model has a prescribed number of term. The model is then "pruned" with a backward deleting procedure where the optimal number of terms is estimated by cross-validation (for details, see Hastie et al. (2005)).

2.3.2 Modified simulation schemes

As a first modification of the Bouchard-Touzi method, we consider an explicit version of the implicit scheme by replacing the second regression in (2.19):

$$\begin{aligned} Y^{\pi}_{t_n} &= g(X^{\pi}_{t_n}) \\ Z^{\pi}_{t_{i-1}} &= \frac{1}{\Delta_i}\mathbb{E}\left[Y^{\pi}_{t_i} \Delta W_{t_i} \middle| X^{\pi}_{t_{i-1}} \right] \\ Y^{\pi}_{t_{i-1}} &= \mathbb{E}\left[Y^{\pi}_{t_i} + f(X^{\pi}_{t_{i-1}}, Y^{\pi}_{t_i}, Z^{\pi}_{t_{i-1}})\Delta_i \middle| X^{\pi}_{t_{i-1}} \right]. \end{aligned} \tag{2.23}$$

This comes from a discretization of the BSDE at the right time-point Y_{t_i} instead of $Y_{t_{i-1}}$ and since Y is a continuous process, the effect of using the value at the right time-point is vanishing as the time-grid becomes tighter. The same discretization is used by for instance Gobet and Lemor (2008), and the benefit is that this allows for an explicit calculation of $Y^{\pi}_{t_{i-1}}$ in the second regression step of each iteration.

As an additional step, to obtain an implicit method with a fixed point procedure, we may employ (2.23) to get a first candidate $\tilde{Y}^{\pi}_{t_{i-1}}$ and supplement each step in the backward induction with a small number of implicit iterations of

$$\tilde{Y}^{\pi}_{t_{i-1}} = \mathbb{E}\left[Y^{\pi}_{t_i} + f(X^{\pi}_{t_{i-1}}, \tilde{Y}^{\pi}_{t_{i-1}}, Z^{\pi}_{t_{i-1}})\Delta_i \middle| X^{\pi}_{t_{i-1}} \right], \tag{2.24}$$

and keeping $Y^{\pi}_{t_{i-1}} = \tilde{Y}^{\pi}_{t_{i-1}}$ as our final value for the next backward step.

[11] We are using the R package "earth" by Milborrow. Derived from mda:mars by T. Hastie and R. Tibshirani. (2011).

Secondly, to improve the stability of the scheme, we consider a modification of (2.23) based on the following recursion for the backward component

$$\begin{aligned} Y^\pi_{t_{i-1}} &= Y^\pi_{t_i} + f(X^\pi_{t_{i-1}}, Y^\pi_{t_i}, Z^\pi_{t_{i-1}})\Delta_i - Z^\pi_{t_{i-1}}\Delta W_{t_i} \\ &= Y^\pi_{t_{i+1}} + f(X^\pi_{t_{i-1}}, Y^\pi_{t_i}, Z^\pi_{t_{i-1}})\Delta_i + f(X^\pi_{t_i}, Y^\pi_{t_{i+1}}, Z^\pi_{t_i})\Delta_{i+1} - Z^\pi_{t_{i-1}}\Delta W_{t_i} - Z^\pi_{t_i}\Delta W_{t_{i+1}} \\ &= Y^\pi_{t_n} + \sum_{k=i}^{n} f(X^\pi_{t_{k-1}}, Y^\pi_{t_k}, Z^\pi_{t_{k-1}})\Delta_k - Z^\pi_{t_{k-1}}\Delta W_{t_k} \end{aligned}$$

such that we may write the explicit backward induction (2.23) as

$$\begin{aligned} Y^\pi_{t_n} &= g(X^\pi_{t_n}) \\ Z^\pi_{t_{i-1}} &= \frac{1}{\Delta_i}\mathbb{E}\left[\left(Y^\pi_{t_n} + \sum_{k=i+1}^{n} f(X^\pi_{t_{k-1}}, Y^\pi_{t_k}, Z^\pi_{t_{k-1}})\Delta_k\right)\Delta W_{t_i}\middle| X^\pi_{t_{i-1}}\right] \\ Y^\pi_{t_{i-1}} &= \mathbb{E}\left[Y^\pi_{t_n} + \sum_{k=i}^{n} f(X^\pi_{t_{k-1}}, Y^\pi_{t_k}, Z^\pi_{t_{k-1}})\Delta_k\middle| X^\pi_{t_{i-1}}\right]. \end{aligned} \tag{2.25}$$

The benefit of this is that errors due to approximating the conditional expectation do not accumulate at the same rate. As in the previous modification, we may complement (2.25) with a small number of iterations

$$\tilde{Y}^\pi_{t_{i-1}} = \mathbb{E}\left[Y^\pi_{t_n} + \sum_{k=i}^{n} f(X^\pi_{t_{k-1}}, \tilde{Y}^\pi_{t_{k-1}}, Z^\pi_{t_{k-1}})\Delta_k\middle| X^\pi_{t_{i-1}}\right] \tag{2.26}$$

for an implicit method.

For an alternative type of simulation schemes, recall that for Markovian forward-backward equations, both Y_t and Z_t may be written as functions of the current forward state (t, X_t). Hence, we use the regression estimator of (2.23) to write

$$\begin{aligned} Y^\pi_{t_{i-1}} &= \hat{E}\left[Y^\pi_{t_i} + f(X^\pi_{t_{i-1}}, Y^\pi_{t_i}, Z^\pi_{t_{i-1}})\Delta_i\middle| X^\pi_{t_{i-1}}\right] \\ &\equiv \hat{y}_{i-1}(X^\pi_{t_{i-1}}) \end{aligned} \tag{2.27}$$

that is, the function $y(t,x)$ that generates $Y_t = y(t, X_t)$ is approximated with $\hat{y}(\cdot)$. Further, if we use $Z^\pi_{t_i}$ in the driver of (2.27) ($Z^\pi_{t_i}$ from the previous time-step) to obtain $\hat{y}_{i-1}(\cdot)$, we get the following scheme

$$\begin{aligned} &Y^\pi_{t_n} = g(X^\pi_{t_n}), \quad Z^\pi_{t_n} = \partial_x g(X^\pi_{t_n})\sigma(X^\pi_{t_n}), \\ &Y^\pi_{t_{i-1}} = \hat{E}\left[Y^\pi_{t_i} + f(X^\pi_{t_{i-1}}, Y^\pi_{t_i}, Z^\pi_{t_i})\Delta_i\middle| X^\pi_{t_{i-1}}\right] \implies \hat{y}_{i-1}(\cdot), \\ &Z^\pi_{t_{i-1}} = \partial_x \hat{y}_{i-1}(X^\pi_{t_{i-1}})\sigma(X^\pi_{t_{i-1}}), \end{aligned} \tag{2.28}$$

since the function that generates Z_t is given by $Z_t = \partial_x y(t, X_t)\sigma(X_t)$, see footnote 8. In particular, if we employ the MARS regression, $\hat{y}(\cdot)$ will be a sum of piecewise linear

splines and products thereof up to the specified degree. Hence, the partial derivatives $(\partial_s \hat{y}(\cdot), \partial_v \hat{y}(\cdot))$ are easily calculated analytically. Further, the last two calculations of (2.28) may be iterated with $Y^\pi_{t_{i-1}}, Z^\pi_{t_{i-1}}$ for an implicit version of the scheme.

For a second type of modifications, we may include additional predictors for the regression functions. As an example, let $C_{\mathrm{He}}(t,x)$ denote the pricing function of an option with terminal payoff $g(X_T)$ calculated under Heston's model. As the pricing bound Y_t lies in a neighbourhood of the price $C_{\mathrm{He}}(t, X_t)$, we may add this as a predictor to our regression estimator

$$Y^\pi_{t_{i-1}} = \hat{E}\left[Y^\pi_{t_i} + f(X^\pi_{t_{i-1}}, Y^\pi_{t_i}, Z^\pi_{t_{i-1}})\Delta_i \middle| X^\pi_{t_{i-1}}, C_{\mathrm{He}}(t_{i-1}, X^\pi_{t_{i-1}}) \right]. \tag{2.29}$$

Finally, we mention a modification of the first regression in the standard scheme (2.19), as proposed by Alanko and Avellaneda (2013)

$$Z^\pi_{t_{i-1}} = \frac{1}{\Delta_i}\mathbb{E}\left[\left(Y^\pi_{t_i} - \mathbb{E}[Y^\pi_{t_i}|X^\pi_{t_{i-1}}] \right) \Delta W_{t_i} \middle| X^\pi_{t_{i-1}} \right] \tag{2.30}$$

with the purpose being a variance reduction of the regression estimate. The motivation is the following: since $Y^\pi_{t_i} = y(t_i, X^\pi_{t_i})$ for some continuous function $y(t,x)$, we have $Z^\pi_{t_{i-1}} = \mathbb{E}[y(t_{i-1}+\Delta_i, X^\pi_{t_{i-1}} + \Delta X^\pi_{t_i})\Delta W_{t_i}/\Delta_i | X^\pi_{t_{i-1}}]$ and the estimator thereof

$$\frac{1}{N}\sum_{j=1}^{N} y(t_{i-1}+\Delta_i, X^\pi_{t_{i-1}} + \Delta X^{\pi(j)}_{t_i}) \frac{\sqrt{\Delta_i} z^{(j)}}{\Delta_i} \tag{2.31}$$

where $z^{(j)}$ are independent standard normal random variables. As $\Delta X^{\pi(j)}_{t_i} = \text{drift} \times \Delta_i + \text{diff} \times \sqrt{\Delta_i} z^{(j)}$, we have that the variance of the estimate (2.31) is approximately $y(t_{i-1}, X^\pi_{t_{i-1}})^2/(N\Delta_i)$ for small Δ_i and hence, it blows up as $\Delta_i \to 0$. In return if we use

$$\frac{1}{N}\sum_{j=1}^{N} \left(y(t_i, X^\pi_{t_i}) - y(t_{i-1}, X^\pi_{t_{i-1}}) + f_{i-1}\Delta_i \right) \frac{\sqrt{\Delta_i} z^{(j)}}{\Delta_i} \tag{2.32}$$

where $f_{i-1} \equiv f(X^\pi_{t_{i-1}}, Y^\pi_{t_{i-1}}, Z^\pi_{t_{i-1}})$ and $y(t_{i-1}, X^\pi_{t_{i-1}}) - f_{i-1}\Delta_i = \mathbb{E}[Y^\pi_{t_i}|X^\pi_{t_{i-1}}]$ from (2.19), we have that the estimator (2.32) of (2.30) will have approximate variance $2y_x(t_{i-1}, X^\pi_{t_{i-1}})^2/N + \Delta_i f_{i-1}/N$ which do not depend on Δ_i as this goes to zero.

We end this section with a demonstration of the simulation schemes based on (2.23) and (2.25) in the following example.

Example 2.1. For the forward process, we simulate $N = 100,000$ paths of Heston's model (2.2) with parameters $(r, \kappa, \theta, \sigma, \rho) = (0, 5.07, 0.0457, 0.48, -0.767)$, initial value $(S^\pi_0, V^\pi_0) = (100, \theta)$ over an equidistant time grid with $n = 25$ points and terminal time $T = 1$. For the backward process, we consider the trivial driver $f(X_t, Y_t, Z_t) = 0$, i.e. $dY_t = Z_t d\tilde{W}_t$, together with the terminal condition $Y_T = S_T$. Hence, Y is a martingale and we have

$$Y_t = \mathbb{E}^Q\left[S_T|\mathcal{F}_t\right] = S_t$$

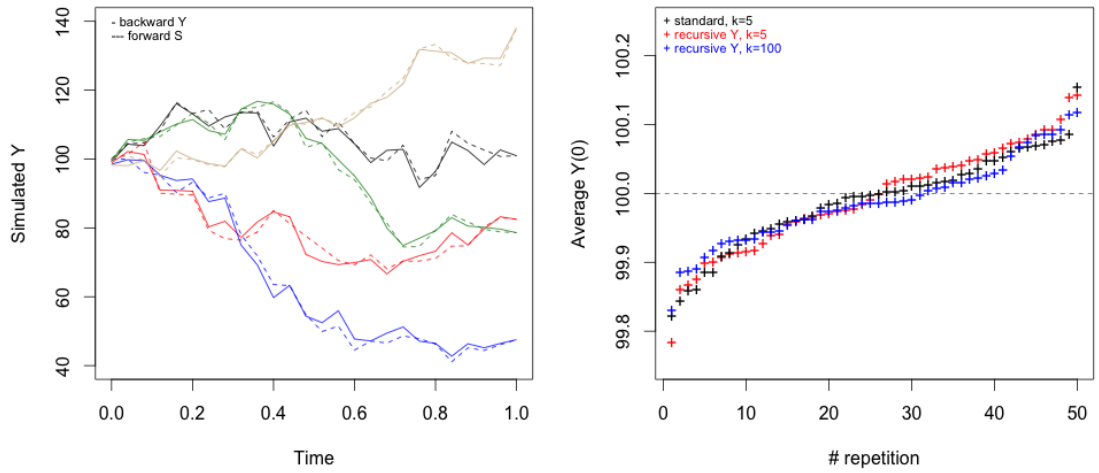

Figure 2.1: Left figure: five simulated paths of Y^π (solid lines) with the zero-driver of example 2.1. The explicit scheme (2.21)-(2.23) is employed with $k = 5$ nearest neighbours. The forward component $X^\pi = (S^\pi, V^\pi)$ is simulated from Heston's model and the dashed lines show the corresponding paths of S^π. **Right figure:** the N-sample average of Y_0^π (in increasing order) from 50 repetitions of the simulation with the $k = 5$ explicit scheme (2.23) (black crosses), the recursive-based scheme (2.25) with $k = 5$ (red crosses) and $k = 100$ (blue crosses).

since for a zero interest rate, S is a Q-martingale as well. As there is no dependency of Z in the driver, the backward induction simplifies to the regression $Y^\pi_{t_{i-1}} = \hat{E}[Y^\pi_{t_i}|X^\pi_{t_{i-1}}]$ repeated for $i = n, \dots, 1$ and a starting value $Y^\pi_{t_n} = S^\pi_{t_n}$. With $k = 5$ nearest neighbours of the regression estimator (2.21), the left pane of Figure 2.1 shows five simulated paths of the backward process Y^π with the explicit scheme (2.23) along with the corresponding paths of S^π and it can be seen that the components follow each other quite closely. Looking at the initial time value, the N-sample of Y_0^π has an average 98.532 to be compared with the true value $Y_0 = \mathbb{E}^Q[S_T] = S_0 = 100$, while the sample of $Y^\pi_{t_n} = S^\pi_{t_n}$ averages to 99.998.

If we repeat the simulation 50 times and calculate the average of Y_0^π for each repetition, we obtain the result in the right pane of Figure 2.1. The first explicit scheme based on (2.23) yields sample averages quite close to the true value and if we repeat the simulations with the Y^π-recursion scheme (2.25) instead, we obtain similar results. For comparison, we have included the recursive scheme with $k = 100$ nearest neighbours as well.

Finally, notice that this example corresponds to $g(x) = x$ and an effect $\alpha(S_t, V_t, u_t) = (0, 0)^\top$ such that $Q^u \equiv Q$. Hence, with $\varrho = 0$ for the driver (2.8), we have that the value process $J_t(u) = \mathbb{E}_u[g(S_T)|\mathcal{F}_t] = \mathbb{E}^Q[S_T|\mathcal{F}_t]$ is the solution to (2.9).

2.3.3 Simulation results for European options

For numerical calculation of the pricing bounds for European options, we consider the parameter setting given in Table 2.4 and a set of call options with strike-maturity structure as given in Table 2.2. The call prices are calculated from the so called semi-closed pricing formula of Heston's model, i.e. by numerical integration of the inverse Fourier transform of the price (see e.g. Gatheral (2011)). The corresponding implied volatilities are then obtained from Black-Scholes formula by numerical optimisation.

European call prices

Strike/Expiry	75	100	125
4m	26.0044 (0.2823)	4.8239 (0.2106)	0.0070 (0.1518)
1y	29.4915 (0.2482)	10.9174 (0.2124)	1.8403 (0.1832)
10y	57.4959 (0.2220)	46.4060 (0.2174)	37.1943 (0.2138)

Table 2.2: Prices and implied volatilities (in parenthesis) of European call options calculated by the semi-closed pricing formula of Heston's model with parameters from Table 2.4.

Prior to considering the pricing bounds as obtained from the optimally controlled value process, we take a look at the prices one achieves by minimising/maximising Heston's pricing formula $C_{\mathrm{He}}(\cdot)$ over the parameter uncertainty set U represented by the elliptic constraint in (2.13) with a 95% confidence level. That is

$$C^{\pm}_{\mathrm{He}} = \left| \min_{(r,\kappa,\theta)\in U} \pm C_{\mathrm{He}}(S, V; \tau, K, \Theta) \right| \tag{2.33}$$

where Θ is the vector of model parameters including (r, κ, θ) while K is the strike and τ the time to maturity. From numerical optimisation of (2.33) with parameters and elliptic uncertainty region based on Table 2.4, we get the results in Table 2.3. We will use these as a reference point for our forthcoming simulation study.

Simulation of the forward component

The forward component $X = (S, V)$ of the SDE (2.2) governing the asset price and variance is simulated in the first stage of the simulation scheme for the forward-backward equation. We employ the standard Euler-Maruyama scheme for the log-price and an implicit Milstein scheme to generate the variance

$$\log S^{\pi}_{t_i} = \log S^{\pi}_{t_{i-1}} + \left(\mu - \frac{1}{2} V^{\pi}_{t_{i-1}}\right) \Delta_i + \sqrt{V^{\pi}_{t_{i-1}}} (\rho \Delta W^1_{t_i} + \sqrt{1-\rho^2} \Delta W^2_{t_i})$$

$$V^{\pi}_{t_i} = \frac{V^{\pi}_{t_{i-1}} + \kappa\theta\Delta_i + \sigma\sqrt{V^{\pi}_{t_{i-1}}} \Delta W^2_{t_i} + \frac{1}{4}\sigma^2((\Delta W^2_{t_i})^2 - \Delta_i)}{1 + \tilde{\kappa}\Delta_i}$$

Optimised Heston pricing function

Strike/Expiry	75	100	125
4m	$[25.9316, 26.2591]$ $(0.0520, 0.3651)$	$[4.5758, 5.0572]$ $(0.1980, 0.2225)$	$[0.0040, 0.0124]$ $(0.1441, 0.1610)$
1y	$[28.6578, 30.4061]$ $(0.0303, 0.3060)$	$[9.9716, 11.8229]$ $(0.1872, 0.2364)$	$[1.3840, 2.4824]$ $(0.1659, 0.2053)$
10y	$[54.5102, 62.3675]$ $(0.0195, 0.3190)$	$[40.2004, 51.9955]$ $(0.1085, 0.2925)$	$[30.7291, 43.0811]$ $(0.1444, 0.2754)$

Table 2.3: Prices and implied volatilities of European call options, calculated by numerical minimisation/maximisation of the Heston pricing formula over the parameters (r, κ, θ) constrained by the parameter uncertainty region.

Model parameters

S_0	V_0	r	κ	θ	σ	ρ
100	0.0457	0.05	5.070	0.0457	0.4800	-0.767

	r	κ	β
r	2.5e-05	0	0
κ	0	0.25	0
β	0	0	1e-04

Table 2.4: Parameter setting and covariance matrix used for the numerical calculation of pricing bounds for European options.

where $\Delta W^1_{t_i}$, $\Delta W^2_{t_i}$ are independent variables generated from the zero-mean normal distribution with variance Δ_i. If the parameters satisfy $4\kappa\theta > \sigma^2$ this discretization scheme generates positive variance paths and we do not have to impose any truncation as in the case with the standard Euler-Maruyama scheme, see Andersen et al. (2010). We simulate $N = 100,000$ paths over an equidistant time gird with $n = 25$ knots.

The optimised Heston formula by backward simulation

As a first simulation example of the backward component Y, we consider the formula-optimal price of the at-the-money call with maturity one year (prices given in Table 2.3). Hence, we simulate the backward component with the non-optimised driver $f(X_t, Y_t, Z_t, u_t)$ of equation (2.8) with a constant u_t based on the resulting parameters from the price-optimisation (2.33) of the considered call option. This allows us to evaluate the accuracy of our simulations schemes in a situation where we know the true values we are aiming at calculating numerically. The reason for simulating the optimised price of (2.33) instead of the null-controlled plain price of the call (given in Table 2.2) is that the optimised-price

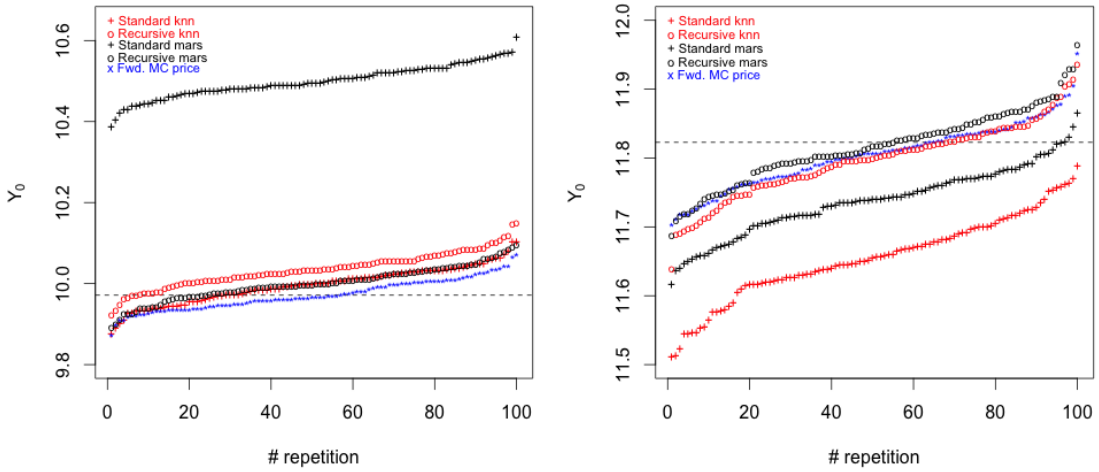

Figure 2.5: Numerical calculation of the formula-optimal price of the one-year at-the-money call (Table 2.3). **Left figure:** the N-sample average of the minimised price Y_0^π from 100 repetitions of the simulation (in increasing order). We use a equidistant time-grid with $n = 25$ time-points and generate $N = 100,000$ paths of Y^π in every simulation. **Right figure:** the corresponding maximised price. The figures show the results from four explicit schemes based on the $k = 5$ nearest neighbours estimator (red marks) and the MARS estimator of degree 2 (black marks). The dashed lines indicate the true call price as calculated by the (optimised) Heston's formula while the blue stars show the Monte-Carlo price as calculated from the N simulated paths of $X^\pi = (S^\pi, V^\pi)$ for each repetition.

simulation relies on a Z-dependent driver, while the Q-price has an effect being zero in (2.8) such that the Z-regression step of the simulation scheme expires for the plain price (cf. example 2.1).

For starters, we consider the following four variations of the simulation schemes from the previous section:

1. the explicit scheme (2.23) with $k = 5$ nearest neighbours regression (2.21)

2. the explicit scheme with MARS regression (2.22) of degree 2

3. the explicit-recursive scheme (2.25) with $k = 5$ nearest neighbours regression

4. the explicit-recursive scheme with MARS regression of degree 2.

For each of the schemes **1**–**4**, we repeatedly simulate the formula-optimal price 100 times and calculate sample- bias and root mean square errors. The results are given in Table 2.6, while Figure 2.5 shows the prices from all repetitions of the simulation.

Backward simulated optimised Heston price

Scheme	Ave. $\mathbb{E}(\hat{\pi})$	Bias: $\mathbb{E}(\hat{\pi}) - \pi$	RMSE: $\sqrt{\mathbb{E}[(\hat{\pi} - \pi)^2]}$
Explicit knn	11.6552	-0.1677	0.1783
Explicit MARS	11.7378	-0.0851	0.0987
Recursive knn	11.7968	-0.0261	0.0608
Recursive MARS	11.8164	-0.0065	0.0534
Forward MC	11.8041	-0.0188	0.0508
Explicit knn	9.9960	0.0244	0.0511
Explicit MARS	10.4993	0.5277	0.5292
Recursive knn	10.0351	0.0635	0.0766
Recursive MARS	10.0004	0.0288	0.0509
Forward MC	9.9719	0.0003	0.0380

Table 2.6: Accuracy of the simulated formula-optimised price of an at-the-money call option with maturity one year (true values in Table 2.3) for $N = 100,000$ and $n = 25$. Sample- average, bias and root mean square error calculated from 100 repetitions of each simulation.

From Table 2.6 we see that the explicit-recursive-MARS scheme performs best in terms of low bias and low RMSE although the simple explicit-knn scheme performs well for the lower price. Comparing the backward simulation with the Monte Carlo price calculated directly from forward simulation we have close to equal performance for the higher price. Since the backward simulation step is dependent of the forward step, we can not expect any improvement in accuracy beyond that of the forward simulation.

Next, we continue with the following modifications of the simulation schemes:

5. explicit-recursive-MARS with variance reduction (2.30)

6. explicit-recursive-MARS with two implicit iterations (2.26)

7. a combination of **5** and **6**

8. explicit-recursive-MARS with call-price predictor[12] (2.29).

The results are recorded in Table 2.7 and if we compare these with the result for the plain explicit-recursive-MARS scheme **4**, we observe similar accuracies for all of them. However,

[12]The calculation of the pricing-formula for the call relies on numerical integration and we need $N = 100,000$ such evaluations for each of $n = 25$ time-step which makes the scheme very computer intensive. For this reason, we calculate a subset of 500 call prices and use a polynomial regression to predict the remaining call prices. As the pricing formula is a "nice" function of S and V, this approximation only has a limited impact.

Backward simulated optimised Heston price II

Scheme	$\mathbb{E}(\hat{\pi})$	$\mathbb{E}(\hat{\pi}) - \pi$	$\sqrt{\mathbb{E}[(\hat{\pi} - \pi)^2]}$
Forward MC	11.8094	-0.0135	0.0411
Rec. MARS, var. reduction	11.8169	-0.0060	0.0433
Rec. MARS, two implicit	11.8196	-0.0033	0.0468
Rec. MARS, var. red. & two imp.	11.8164	-0.0065	0.0433
Rec. MARS, call-predictor	11.8166	-0.0063	0.0435
Rec. MARS, Z-function	11.6868	-0.1361	0.1489
Rec. MARS, Z-fun. & three imp.	11.6794	-0.1435	0.1558
Forward MC	9.9719	0.0003	0.0380
Rec. MARS, var. reduction	10.0075	0.0359	0.0501
Rec. MARS, two implicit	10.0027	0.0311	0.0495
Rec. MARS, var. red. & two imp.	10.0094	0.0378	0.0515
Rec. MARS, call-predictor	10.0082	0.0366	0.0507
Rec. MARS, Z-function	10.1096	0.1380	0.1502
Rec. MARS, Z-fun. & three imp.	10.1661	0.1945	0.2034

Table 2.7: Accuracy of the simulated formula-optimised price of an at-the-money call option with maturity one year (true values in Table 2.3) for $N = 100,000$ and $n = 25$. Sample- average, bias and root mean square error calculated from 100 repetitions of each simulation.

as both implicit schemes **6** and **7** add N regressions and N evaluations of the driver to the computational cost for each implicit iteration, we opt for the schemes **4** or **5**.

At last, we consider two schemes based on the MARS derivative: (**9**) explicit-recursive-MARS with Z-function (2.28), (**10**) explicit-recursive-MARS with Z-function and three implicit iterations. Both these modifications yield poor accuracy, see Table 2.7.

2.3.4 The optimally controlled value-process of a European call option

Here we simulate the backward component of equation (2.10) that governs the optimally controlled upper/lower pricing bound of the European call option with strike-maturity structure as in Table 2.2 based on parameters in Table 2.4. Hence, we simulate Y with an initial (terminal) condition $Y^{\pi}_{t_n} = (S^{\pi}_{t_n} - K)^+$ and an optimised driver $H(X_t, Y_t, Z_t)$ as of equation (2.14) with a confidence level of 95% for the parameter uncertainty region based on the covariance matrix in Table 2.4.

As before, we simulate the forward component $X^{\pi} = (S^{\pi}, V^{\pi})$ with the Euler-Maruyama implicit-Milstein scheme and for a start, we use $N = 100,000$ paths over an equidistant

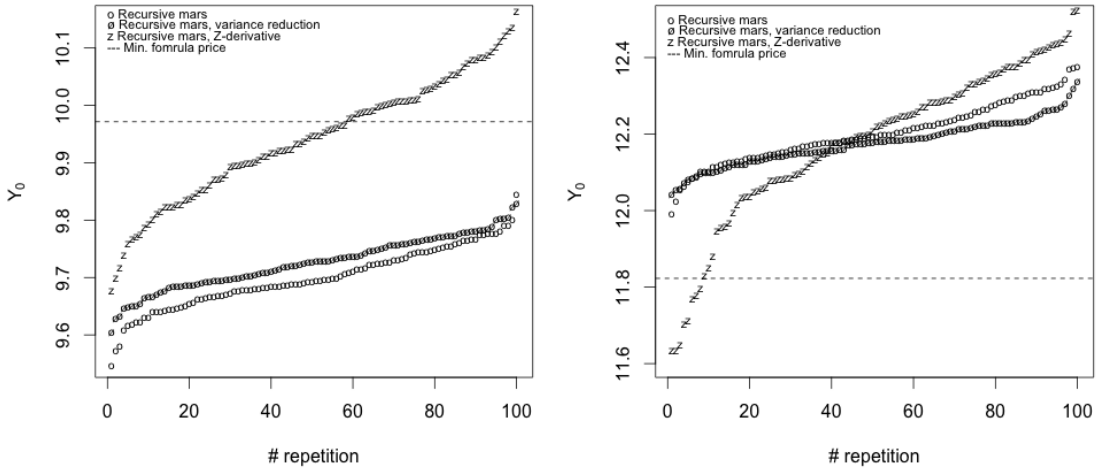

Figure 2.8: Numerical calculation for the pricing bounds of the one-year at-the-money call with $N = 100,000$ paths over $n = 25$ time-points. **Left figure:** the N-sample average Y_0^π of the lower bound for the call price (in increasing order) calculated from each of 100 repetitions of the simulation. **Right figure:** the corresponding upper bound. The dashed lines indicate the call price as calculated by the optimised Heston's formula.

time-gird with $n = 25$ points. Note that for each backwards time-step, we perform $3 \times N$ regressions to obtain the one-step recursion of $Z^{1\pi}$, $Z^{2\pi}$, Y^π and N evaluations of the matrix multiplication in (2.14) for the optimal driver. For the implicit versions of the schemes, we iterate two (or three) times which adds $2 \times N$ regressions and $2 \times N$ matrix multiplications to each time-step.

For a demonstrative example, we again consider the one-year the at-the-money call option and run 100 repetitions of the following simulation schemes:

1. explicit-recursive-MARS of degree 2

2. explicit-recursive-MARS of degree 2 with variance reduction

3. explicit-recursive-MARS of degree 2 with Z calculated from the MARS derivative

4. two implicit fixed-point iterations added to scheme number **1**

5. two implicit fixed-point iterations added to scheme number **2**

6. explicit-recursive-MARS of degree 2 with call-price predictor and variance reduction.

The resulting pricing bounds are shown in Figure 2.8 where, for clarity, we have plotted only the results from scheme number **1**, **2** and **3**.

From Figure 2.8 we see that if we add variance reduction to the explicit-recursive-MARS we obtain slightly higher (lower) prices for the lower (upper) boundary and a somewhat lower variance. Further, if we consider the two-step implicit versions of these schemes, we have that **1** and **4** coincide almost perfectly, and also **2** and **5**, for both the upper and lower bounds (these schemes are excluded from Figure 2.8 only for clarity). The same holds if we add the call-price predictor: **2** and **6** coincide for both the upper and lower bounds. As in the case for the formula-optimised price, the Z-function scheme yields a high lower bound (similar to the formula minimised price) and a upper bound similar to the other schemes. Both bounds also have a very high variance, and for these reasons we henceforth omit the Z-function schemes.

Recursive MARS degree 2 with variance reduction, $n = 25$

Strike/Expiry	75	100	125
4m	[25.7771, 26.2877] (0.0585,0.3714)	[4.5005,5.1597] (0.1942,0.2277)	[0.0016,0.0175] (0.1335,0.1672)
1y	[28.5910,30.5482] (0.0329,0.3138)	[9.7418,12.1603] (0.1811,0.2454)	[1.2374,2.6306] (0.1600,0.2101)
10y	[52.7314,64.9297] (0.0319,0.3645)	[40.5299,54.2037] (0.1176,0.3214)	[30.7684,45.1219] (0.1448,0.2970)

Table 2.9: Pricing bounds for the European call option and corresponding Black-Scholes implied volatilities. Calculated from numerical simulation schemes for the backward process with $N = 100,000$ simulated paths of the forward process following Heston's model over an equidistant time grid with $n = 25$ points.

Based on the previous results for the one-year ATM call, we choose to employ the explicit-recursive-MARS with variance reduction as our working scheme for pricing-boundary calculations. The simulation based results for the considered call options of Table 2.2 are given in Table 2.9 and if we compare these with the formula-optimal prices of Table 2.3, we generally see wider pricing intervals for the optimally controlled value process. This is what we should expect: the formula-optimal prices correspond to a controlled value-process with parameters held constant throughout the lifetime of the option, while in the former case, the parameters are allowed to vary in an optimal way. An illustration of this point is given in Figure 2.10 where it is shown how the parameters vary for the optimally controlled one-year at-the-money call option.

The previous pricing bounds where obtained from a simulation of $N = 100,000$ paths over a regular time-grid of $n = 25$ points. While N is chosen to be a high number for the gain of a low error of the simulation-based regression estimator, the discretization time-step $\Delta = T/n$ is relatively large (for the one-year option, $n = 25$ corresponds to having a time-step in the size of two weeks while for practical Monte Carlo pricing, one typically uses daily or even finer time-steps). For this reason we repeat the calculation of Table 2.9 with a finer time-step of $n = 100$. The results given in Table 2.11 show wider

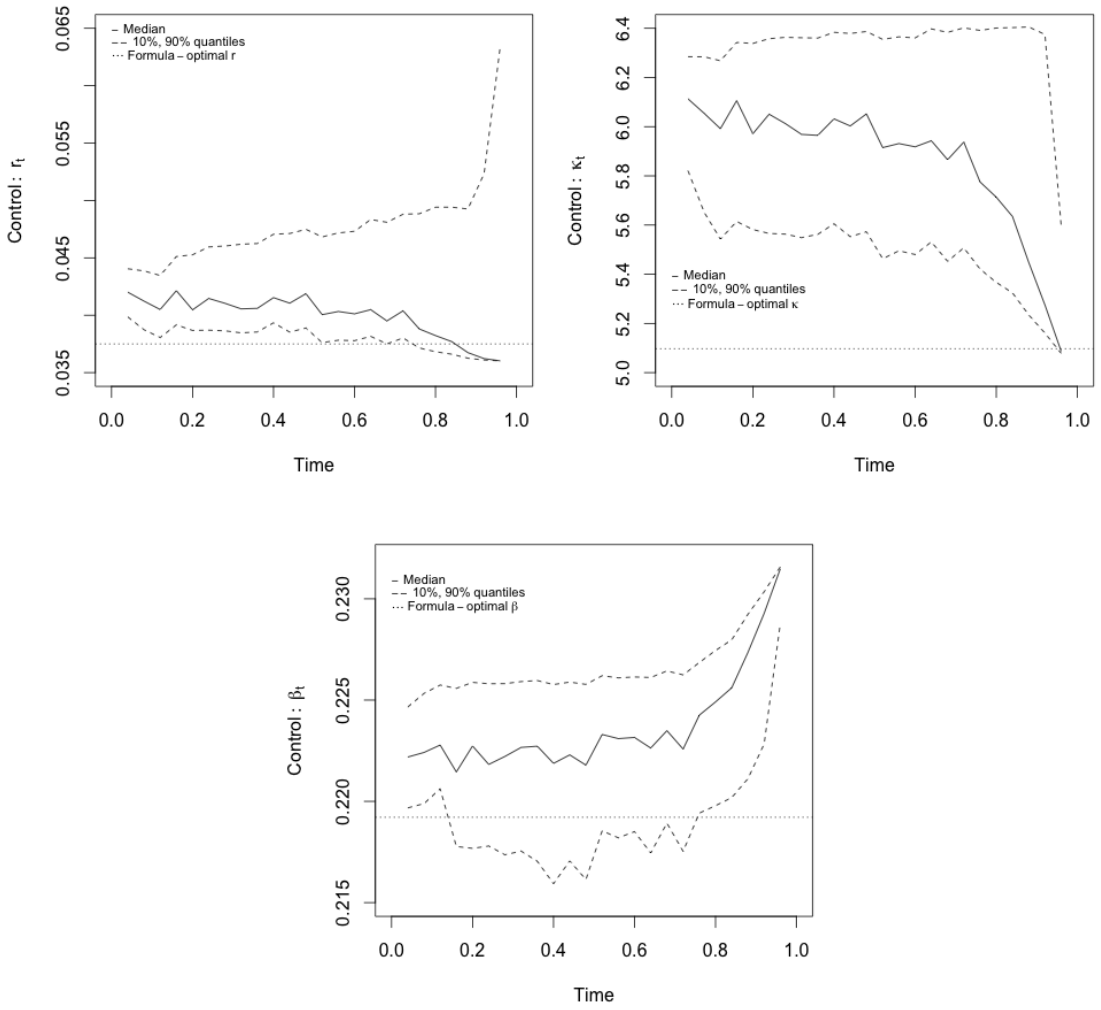

Figure 2.10: The optimal controls $u_t^* = (r_t^*, \kappa_t^*, \beta_t^*)$ as outputted from the optimisation of the driver H^- for the one-year ATM call option. Plotted median and quantiles of $N = 100,000$ simulation paths. The dotted lines show the corresponding constant parameter choice from the optimised Heston formula.

pricing bounds for all strikes/maturities when comparing to tabel 2.9 and the difference between the two step sizes increases with the maturity. A natural explanation for this is that with a higher number of n we also have a higher number of time-steps at which we optimise the driver $H^{\pm}$, and this should lead to a value process Y^{π} optimised to a higher degree. This effect is obvious for the long-maturity options while it is less apparent for the four-month option (the implied volatilities agrees down to 10^{-2}) which also indicates that the simulation error is not particularly affected by the finer time-discretization.

Recursive MARS degree 2 with variance reduction, $n = 100$

Strike/Expiry	75	100	125
4m	[25.7349, 26.3130] (0.05824, 0.3767)	[4.4748, 5.1885] (0.1929, 0.2292)	[0.0005054, 0.02127] (0.1225, 0.1710)
1y	[28.3640, 30.6353] (0.0330, 0.3184)	[9.6158, 12.2530] (0.1777, 0.2478)	[1.1033, 2.6726] (0.1543, 0.2115)
10y	[48.5895, 67.3999] (0.0217, 0.4076)	[36.9068, 57.1310] (0.0199, 0.3598)	[27.4504, 48.1860] (0.1053, 0.3298)

Table 2.11: Pricing bounds for the European call option and corresponding Black-Scholes implied volatilities. Calculated from numerical simulation schemes for the backward process with $N = 100,000$ simulated paths of the forward process following Heston's model over an equidistant time grid with $n = 100$ points.

2.4 The empirical perspective

Based on the numerical recipe for the calculation of pricing bounds of European options, we are now ready to take a look at how the method carries over for a set of real data with historical market prices of the S&P 500 index. The empirical study is carried out in a few steps and we motivate the analysis with the following rationale. We let the statistical parameters as estimated from historical observations of the (asset-) price and variance represent the financial market model under the objective measure P. Hence, (S, V) is assumed to evolve with P-dynamics according to Heston's model specified with the estimated parameters. We consider (B, S) exclusively as the traded assets in the financial market model driven by two random sources, W^1 and W^2, and refrain from the assumption that there exists an additional, exogenously given, (volatility dependent) asset which would complete the model. On the other hand, we dispense from arbitrage opportunities in the model and affirm the existence of a space of risk-neutral measures: Q exists (not necessarily unique) in a set $\mathcal{Q}$ of probability measures equivalent to P, such that the discounted asset price is a martingale under any measure in $\mathcal{Q}$. In our model context, this implies that (S, V) will have the same diffusion matrix under every $Q \in \mathcal{Q}$ as given by the diffusion matrix of the P-dynamics. This follows from the notion that the quadratic variation (continuous part) of a semimartingale is invariant under equivalent probability measures on a complete filtered space. Further, by Girsanov's theorem, we have that the law of the driving random sources is invariant: they will be (independent) Wiener process under all equivalent measures in $\mathcal{Q}$.

With this in mind, we fix the diffusion matrix of (S, V) to be as given by the estimated diffusion parameters from historical data. We then take the space of equivalent risk-neutral measures to be equal the space spanned by the controlled measure Q^u over all admissible controls: $\mathcal{Q} \equiv \{Q^u : u \in \mathcal{U}\}$ where $\mathcal{U}$ represents the space of predictable control processes $u = \{u_t\}_{t \geq 0}$ that lives in the compact uncertainty set $U \subset \mathbb{R}^d$. In particular, we deduce

the uncertainty set from inference of the statistical estimation problem and define U to be represented by the elliptical confidence region (for a given 95% level) as derived from the observed Fisher information. The question to ask is then, if market option prices are covered by the pricing rules in $\mathcal{Q}$ as given by the corresponding model pricing-boundaries.

Since the volatility process of an asset is latent by nature, it has to be measured with some method. In the following section we briefly present the realized volatility measure which gives a commonly used nonparametric estimator of the variance process. The measured quantity is the historical volatility, as opposed to implied volatility reversed from option market-prices, and it is estimated from historical prices of the asset. We then proceed to briefly retail some estimation methods that we employ for point estimation of the model parameters and for drawing inference thereof. The empirical study based on S&P 500 market data finally follows.

2.4.1 Measured variance and parameter estimation

The historical market data for the S&P 500 index contains of daily observations of the closing price and we use the variance as estimated from high-frequency observations (∼5min) of the asset returns with the realised volatility measure (see e.g. Andersen and Benzoni (2009)). Hence, if $s = (s_{t_1}, s_{t_2}, \dots)$ is a time-series with daily prices then the realized variance measure is calculated as

$$RV([t_{i-1}, t_i]) = \sum_{k: s_k, s_{k+1} \in [t_{i-1}, t_i]} (y_{s_{k+1}} - y_{s_k})^2,$$

where $[t_{i-1}, t_i]$ is the duration of the day and $s_k \in [t_{i-1}, t_i]$ are intra-day time points over which $y_{s_k} = \log(s_{s_k}) - \log(s_{s_{k-1}})$ are the log-returns. The realized variance approximates the integrated variance: $RV([t_{i-1}, t_i]) \xrightarrow{P} \int_{t_{i-1}}^{t_i} V_s ds$, (for a continuous return process, the quadratic variation for a general semimartingale), and the measured variance at time t_i is taken to be

$$v_{t_i} = \frac{1}{t_i - t_{i-1}} RV([t_{i-1}, t_i]).$$

such that one obtains a time-series of daily variances $v = (v_{t_1}, v_{t_2}, \dots)$. For convenience, we use precomputed variance estimates from the Oxford-Man Institute's realised library.[13] Figure 2.12 shows the daily variance and closing price of the S&P 500 index from the period January 3rd, 2000 to February 29th, 2016.

We consider the estimation of the parameters $\Theta = (\kappa, \theta, \sigma)$ from $n + 1$ observations $v = (v_0, \dots, v_n)$ of the variance. Here, v_i is treated as the observed value of V_{t_i} for a set of discrete time-points $(t_0, \dots, t_n)$ at which the observations are made, and we denote with $\Delta_i = t_{i+1} - t_i$ the length of the i^{th} time-interval between two consecutive observations.

[13] The Realised Library version 0.2 by Heber, Gerd, Lunde, Shephard and Sheppard (2009), `http://realized.oxford-man.ox.ac.uk`.

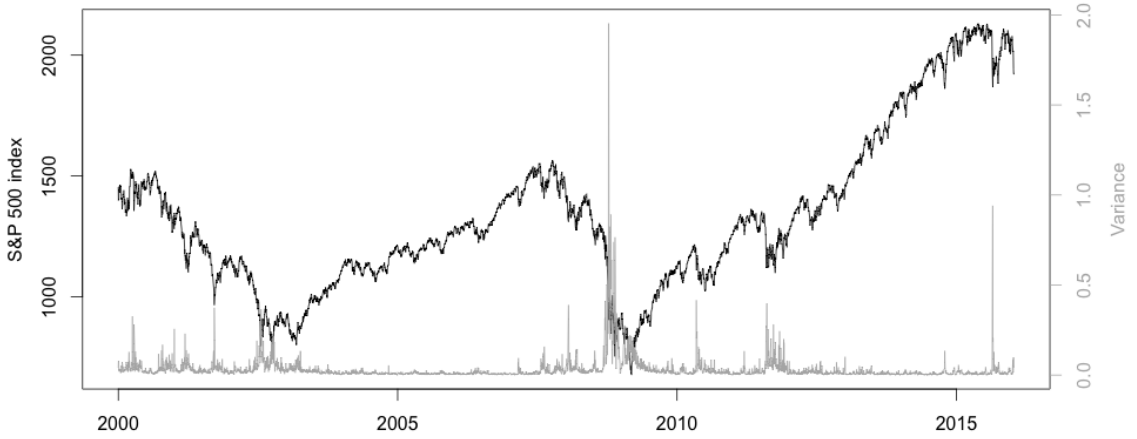

Figure 2.12: Historical closing prices and realized variances of the S&P 500 index, 4,035 daily observations from January 3rd, 2000 to February 29th, 2016. The data is sourced from the Oxford-Man Institute realized library.

In general, inference on the parameters Θ of a process with transition density $f(y; x, \delta, \Theta)$ for $V_{t+\delta}|V_t = x$ can be made with the likelihood function

$$L_n(\Theta) = \prod_{i=0}^{n-1} f(v_{i+1}; v_i, \Delta_i, \Theta)$$

and the maximising argument of the (log of) likelihood function is the maximum likelihood estimator $\hat{\Theta}$ of Θ. If the transition density is unknown in closed form, or, as in the case for the square root process, of a kind that is impenetrable for optimisation (both analytically and by numerical schemes), then one may consider alternatives based on suitable approximations of the likelihood. A direct way of doing so, is to consider the time-discrete approximation V^π of the process V as given by a Euler-Maruyama scheme; for the square root process

$$V^\pi_{t+\delta} = V^\pi_t + \kappa(\theta - V^\pi_t)\delta + \sigma\sqrt{V^\pi_t}\,(W_{t+\delta} - W_t)$$

which will give an approximative Gaussian log-likelihood function

$$l_n(\Theta) \equiv \log L_n(\Theta) = -\frac{1}{2}\sum_{i=0}^{n-1} \frac{(v_{i+1} - v_i - \kappa(\theta - v_i)\Delta_i)^2}{\sigma^2 v_i \Delta_i} + \log(2\pi\sigma^2 v_i \Delta_i) \tag{2.34}$$

suitable for optimisation. A function on the same form as above was considered for least-squares estimation of the drift parameters in Prakasa Rao (1983). For processes with ergodic property, Kessler (1997) considered the joint estimation of drift and diffusion parameters with a Gaussian approximation to the transition density of the form (2.34)

and showed that under general conditions, the estimator is asymptotically normal and efficient. Their approach addresses the case when the mean and variance of the transition density are unknown and uses approximations in their place. For the square root process, the explicit expressions

$$\begin{aligned}\mathbb{E}[V_{t+\delta}|V_t = v] &= \theta + (v-\theta)e^{-\kappa\delta} \equiv \mu(v,\delta),\\ \mathrm{Var}(V_{t+\delta}|V_t = v) &= v\frac{\sigma^2}{\kappa}(e^{-\kappa\delta} - e^{-2\kappa\delta}) + \theta\frac{\sigma^2}{2\kappa}(1-e^{-\kappa\delta})^2 \equiv s^2(v,\delta),\end{aligned}$$

in place of the approximations $\mu(v,\delta) \approx v + \kappa(\theta - v)\delta$ and $s^2(v,\delta) \approx \sigma^2 v^2 \delta$ in (2.34) give that

$$l_n(\Theta) = -\frac{1}{2}\sum_{i=0}^{n-1}\frac{(v_{i+1} - \mu(v_i,\Delta_i))^2}{s^2(v_i,\Delta_i)} + \log(2\pi s^2(v_i,\Delta_i))$$

forms an approximative Gaussian likelihood function with exact expressions for the conditional mean and variance.

An approximation for the variance of the maximum likelihood estimator $\hat{\Theta} = \arg\max l_n(\Theta)$ is given by the observed information matrix

$$I_o = -\frac{\partial^2 l_n(\Theta)}{\partial\Theta^\top \partial\Theta}\bigg|_{\Theta=\hat{\Theta}}$$

which may be calculated by numerical differentiation of the log-likelihood function at estimated values. The approximative covariance matrix of $\hat{\Theta}$ is then given by the inverse $\Sigma_{\hat{\Theta}} = I_o^{-1}$ and an estimated standard error of the j^{th} parameter by $\sqrt{(I_o^{-1})_{jj}}$. Hence, an approximate $1-\alpha$ confidence region for Θ is given by the ellipse

$$(\Theta - \hat{\Theta})\Sigma_{\hat{\Theta}}^{-1}(\Theta - \hat{\Theta})^\top \leq \chi_d^2(1-\alpha)$$

where d is the dimension of the row vector Θ and $\chi_d^2(1-\alpha)$ is the $1-\alpha$ quantile of the chi-square distribution with d degrees of freedom.

2.4.2 Empirical study

We base the empirical study on market data sourced from the Oxford-Man Institute of Quantitative Finance and from Wharton Research Data Services.[14] Data from Oxford-Man is mainly used for the *asset price* of the S&P 500 index, both historical closing prices observed with a daily frequency, and high-frequency ($\sim$ 5min) returns which are used for the calculation of historical volatility of the index price. Data from the Option Metrics database, sourced through Wharton Research, is used for the *option price* of European call options written on the S&P 500 index. We use historical quotes of bid and offer prices

[14]The former provides a publicly available data-source from `http://realized.oxford-man.ox.ac.uk` while the latter requires a user account, `https://wrds-web.wharton.upenn.edu/wrds/`.

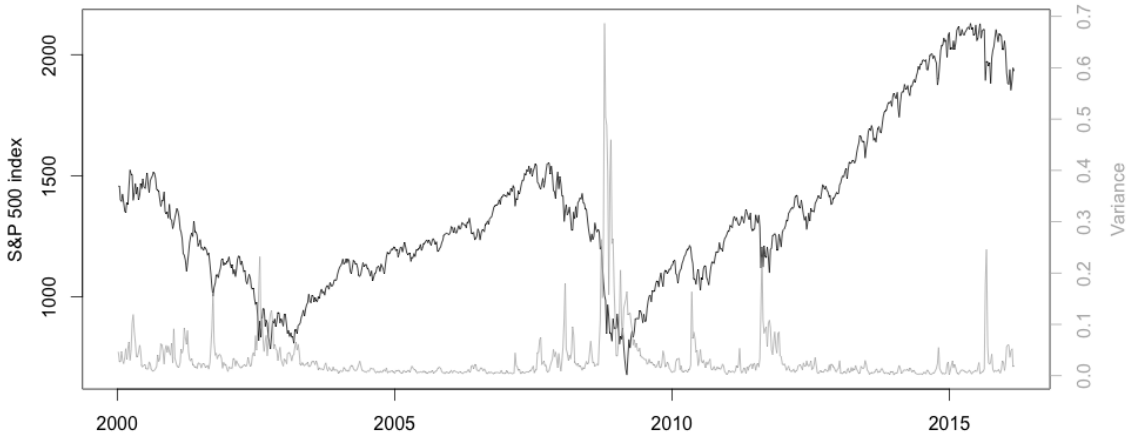

Figure 2.13: Historical closing prices and realized variances of the S&P 500 index, 843 weekly observations from January 3$^{\text{rd}}$, 2000 to February 29$^{\text{th}}$, 2016.

from options with different strike-prices and maturities, which we select from the available market quotes. In addition to price quotes, strike-level and maturity date, we also extract the relevant dividend yield paid by the underlying index, and the risk-free interest rate that corresponds to the maturity of each option in the data set.

Prior to the numerical calculation of pricing bounds for call options on the S&P 500 index, we estimate the parameters of Heston's model from the historical market price and variance according to the following steps:

1. First, we decimate the observation frequency of the variance to weekly observations by calculating the realized variance measure over week-long intervals, Figure 2.13. This operation smooths the measured variance process and in particular, it removes the extreme variance spikes (cf. Figure 2.12) which cause non-robust parameter estimates.

2. We estimate (κ, β, σ) from the weekly variance with the parametrisation $(\kappa, \beta, \sigma) \mapsto (\kappa, \theta, \sigma)$ of the model. We employ the approximative likelihood based on Euler conditional moments[15] and calculate the approximative covariance matrix accordingly by numerical differentiation. Results are given in Table 2.14 (with squared elements of the covariance matrix for a notion of standard errors) together with estimation results from the daily variance.

[15] Alternatively, we may employ the (approximative) likelihood with exact conditional moments. For daily observations, the numerical optimisation does not converge while for weekly data, this yields similar parameter estimates and standard errors as with approximative moments.

3. In addition to estimated parameters of the variance process, we estimate the correlation coefficient of the model with a realised covariation measure.[16] This gives an estimate $\rho = -0.274$ from the weekly variance and closing price of the S&P 500 index.

From the results in Table 2.14, note that the daily variance yields relatively high estimates of the mean-reversion speed to accommodate extreme observations, and also large standard errors of both drift parameters, which indicate that the square root process is a poorly fitting model for the daily variance data.

Estimates, daily data

	κ	θ	σ
	29.6	0.0315	2.58

Estimates, weekly data

	κ	θ	σ
	4.59	0.0307	0.775

Standard errors

	κ	β	σ
κ	4.314	0.544	2.37e-06
β	0.544	0.074	1.79e-06
σ	2.37e-06	1.79e-06	0.0288

Standard errors

	κ	β	σ
κ	1.395	0.621	1.91e-06
β	0.621	0.0269	4.84e-06
σ	1.91e-06	4.84e-06	0.0189

Table 2.14: Parameters and standard errors estimated from historical data of the S&P 500 index. **Left table:** results based on 4,035 daily observations. **Right table:** results based on 843 weekly observations as decimated from the original data. All estimates from numerical optimisation and differentiation of the approximative likelihood function based on Euler moments.

With estimated model parameters in hand, we continue to the calculation of upper/lower bounds for the price of European options. We proceed according to the following:

1. We consider call options on the S&P 500 index with historical market prices from the three-year period of August 31$^{\text{st}}$, 2012 to August 31$^{\text{st}}$, 2015. We select the dates from this period that coincides with the weekly index data (i.e. dates for which both the option and the S&P 500 closing price are quoted). This results in 157 dates and a total of 244,239 option quotes for different strikes and maturities.

2. For each of the 157 dates during the time period, we chose a single option with strike prices and times to maturity as shown in right Figure 2.15. The "initial" option of each period is selected with a medium-seized maturity and a strike-price as close as possible to being at-the-money. We then retain the same maturity and strike

[16] The quadratic covariation of logarithmic data gives $\frac{1}{t}[\log S, \frac{1}{\sigma}\log V]_t = \frac{1}{t}\int_0^t \sqrt{V_s}\frac{1}{\sqrt{V_s}} d[\rho W^1 + \sqrt{1-\rho^2}W^2, W^1]_s = \rho$ and we use a realized covariation estimate thereof.

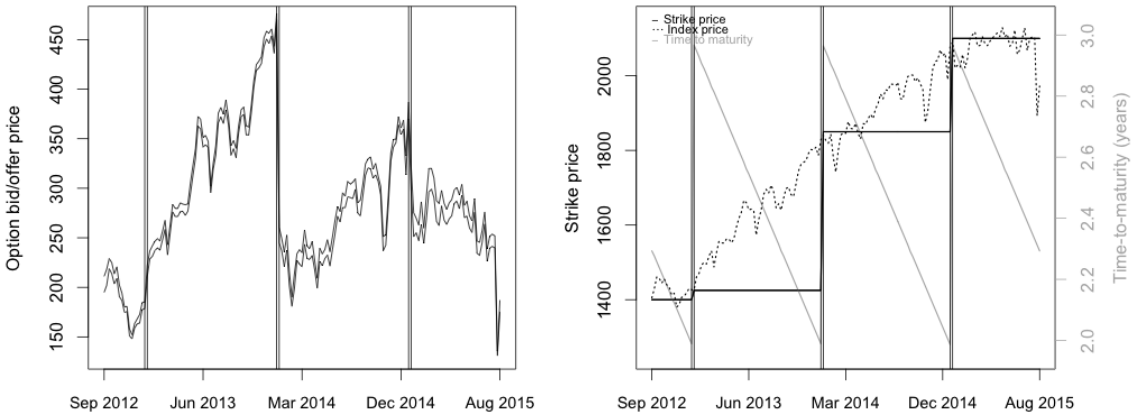

Figure 2.15: Left figure: historical market prices of call options on the S&P 500 index. The figure shows 157 bid/offer quotes (converted to zero-dividend prices) from the period August 31st, 2012 to August 31st, 2015. **Right figure:** the strike price and time-to-maturity of the call options.

price (the same option) as far as there is market quotes available from the OptionMetrics data. This gives us four options in total. We further record the relevant risk-free rate as given by the (continuously compounded) zero-coupon interest rate with corresponding maturity, and the current dividend yield, both retrieved from the OptionMetrics data. The left pane of Figure 2.15 shows the resulting bid/offer quotes after they have been converted to zero-dividend prices. This is done for each quote by first calculating the Black-Scholes implied volatility (with the effective dividend yield) and then recalculating the Black-Scholes price with zero dividend.

3. For the calculation of pricing bounds for the call option, we initially have the following numerical considerations. Firstly, for the parameter estimates in Table 2.14 (based on weekly data) we have $\sqrt{4\beta} = 0.751 < 0.775 = \sigma$ which implies that the implicit Milstein scheme may fail due to the generation of negative outcomes. To prevent this, we include a truncation step[17] to the simulation scheme according to the suggested method of Andersen et al. (2010). Further, we increase the number of time steps to $n = 1,000$ for the forward simulation to prevent the generation of negative variance values.[18] We then down-sample the simulated price and variance to the original time-grid of $n = 25$ steps for the backward simulation. Secondly, the

[17]The time-stepping of the scheme fails whenever $V_{t_i}^{\pi} < 0$ due to the computation of $\sqrt{V_{t_i}^{\pi}}$ and the truncation step is simply to replace with $\sqrt{(V_{t_i}^{\pi})^+}$. Although this prevents the scheme to fail, note that negative values may still be generated and in particular when the time-step Δ_i is large.

[18]Bookkeeping the sign of the generated variance values yields positive outcomes 99.3% of the time when using $n = 1,000$ time steps and 96.7% for $n = 25$.

driver of the backward process will explode for variance values approaching zero. As the forward simulation may output negative/zero values, we cancel the control: $u(X_t, Z_t, Y_t) \equiv 0$ giving $H(X_t, Z_t, Y_t) = -rY_t$, each time the variance is smaller than a threshold; $V_t < \varepsilon$, and we set $\varepsilon = 0.00041$ which is the minimum value of the S&P 500 variance.

4. We simulate the optimally controlled value-process by the explicit scheme with Y-recursion and the MARS method of degree 2 for the regressions. The control variate is included and we simulate $N = 100,000$ paths of the forward-backward process over a time grid with $n = 25$ time steps (with $n = 1,000$ down-sampled to $n = 25$ for the forward process). For each call option price, we run a separate simulation with the estimated model parameters and appropriate maturity/strike, risk-free rate and initial values for the forward process (S&P 500 index level and variance) from the market data. We simulate the backward process with the minimised driver (for the lower pricing bound) and the maximised driver (for the upper bound) based on the same forward simulation. Each evaluation of the optimal driver is calculated with the covariance matrix $\Sigma_{r,\kappa,\beta}$ where we use estimated standard errors and correlation for κ, β and a standard deviation of 0.00005 for the interest rate (r uncorrelated with κ, β). As before, we use a confidence level of 95% for the uncertainty region.

5. Finally, we calculate the corresponding minimum/maximum prices for each considered call option by numerical optimisation of Heston's pricing function over the same 95% uncertainty region.

The resulting upper and lower pricing bounds from the simulations are presented in Figure 2.16 together with the corresponding market quotes of bid and offer prices for the call options. The formula-optimal prices are depicted in Figure 2.18. The dates at which there is a common change in strike price and maturity (see right Figure 2.15) have been marked in the figures to separate the different options: we have calculated model bounds of four calls on the S&P 500 index with different strike/maturity structures on a weekly basis, the first during a time period of ~ 4 months and the remaining during periods of $\sim 8-12$ months. We label these options (I)-(IV) and give some additional results in Table 2.17.

As a first note on the results for the pricing bounds, we see that the market bid/offer quotes fall inside the model bounds for almost all considered call prices (154 out of 157) and in particular, for all prices when looking at the latest two options (III)-(IV), see Table 2.17. The lower bound is always covering the bid quote and the price interval of the bounds is fairly symmetrical around the mid-market price for options (III)-(IV). The offer prices of option (I) is close to the upper bound (occasionally above) and the same holds for (II). This option's moneyness is increasing with time (see right pane of Figure 2.15) while the bound-to-offer distance is shrinking. A possible explanation may be the model's inability to capture the slope and skew of market prices/implied volatilities (for the parameters we use, as estimated from historical data).

For an investigation of this point, we take a look at the option prices and model boundaries for a range of strikes at the first and last date of option (II) which we have depicted in

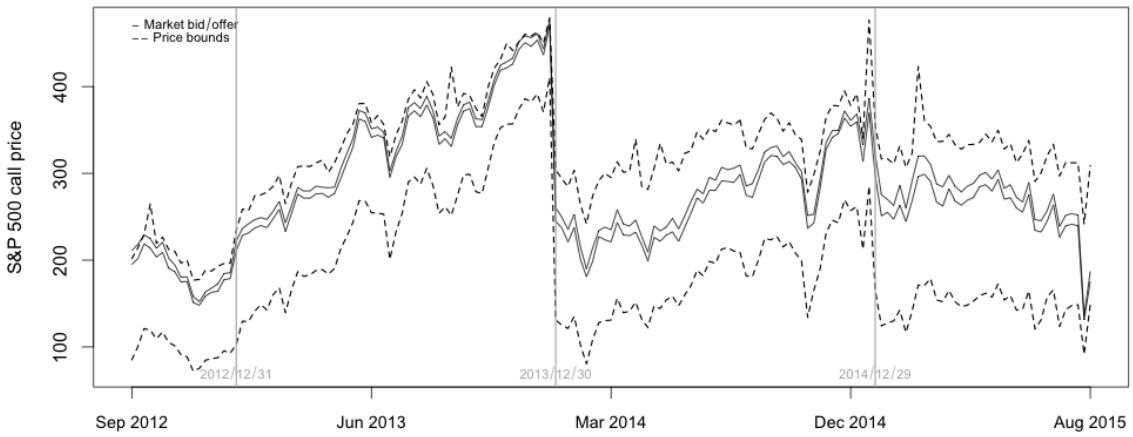

Figure 2.16: European call options on the S&P 500 index: upper and lower pricing bounds (dashed lines) as calculated by simulations of the optimally controlled value process. The graph shows model prices and historical market quotes of bid and offer prices (solid lines) on a weekly basis of options with four different strike/maturity structures (see right pane of Figure 2.15).

European call options on the S&P 500 index

Option:	(I)	(II)	(III)	(IV)
Duration	12/09/04-12/12/24	12/12/31-13/12/23	13/12/30-14/12/22	14/12/29-15/08/31
Maturity	14/12/20	15/12/19	16/12/16	17/12/15
Spread:price	5.6%	3.4%	5.6%	7.1%
Bounds:price	74.9%	42.7%	72.4%	85.7%
Bounds:spread	14.9	12.4	13.9	12.8
In bounds	88.2%	98.1%	100%	100%
Optim:price	39.2%	21.1%	37.5%	44.2%
Optim:spread	7.8	6.1	7.2	6.6
In interval	11.8%	0%	76.9%	72.2%

Table 2.17: Key figures for the model prices of S&P 500 call options from the optimally controlled value process (the pricing bounds) and the formula-optimal pricing interval. The spread-to-price ratio gives the average size of the market spread as a percentage of the average mid-market price of each option. Similarly, the bounds-to-price ratio gives the (average) size of model-bounds to the price, and to the market spread (bounds-to-spread ratio). The in-bounds figures give the proportion of market quotes that fall inside the model bounds. Corresponding figures are calculated for the intervals of the formula optimised model prices.

terms of implied volatilities in Figure 2.19. The left pane, which shows prices form the first considered date 2012-12-31 of option (II), shows a strongly skewed volatility curve from the

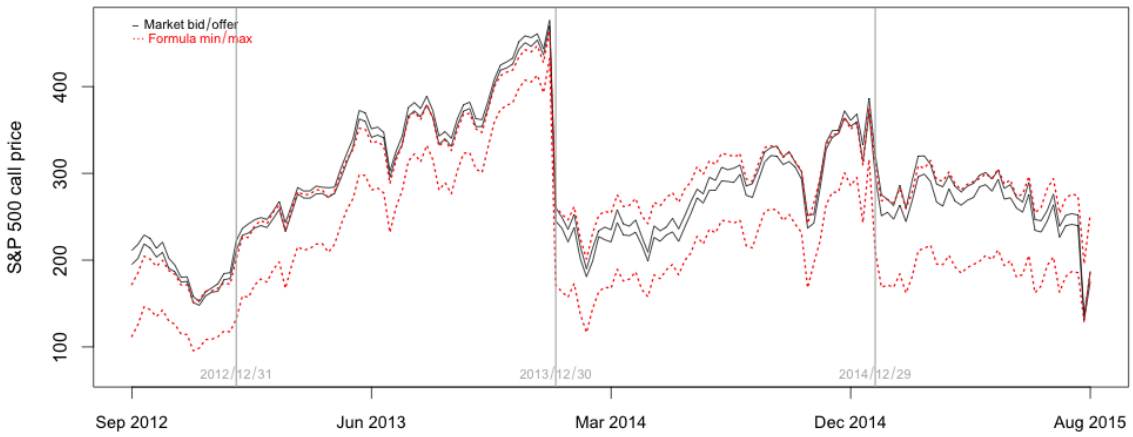

Figure 2.18: The minimum and maximum price (red dotted lines) as obtained from the optimisation of Heston's pricing formula for call options. The solid black lines show the market bid/offer quotes of the S&P 500 options.

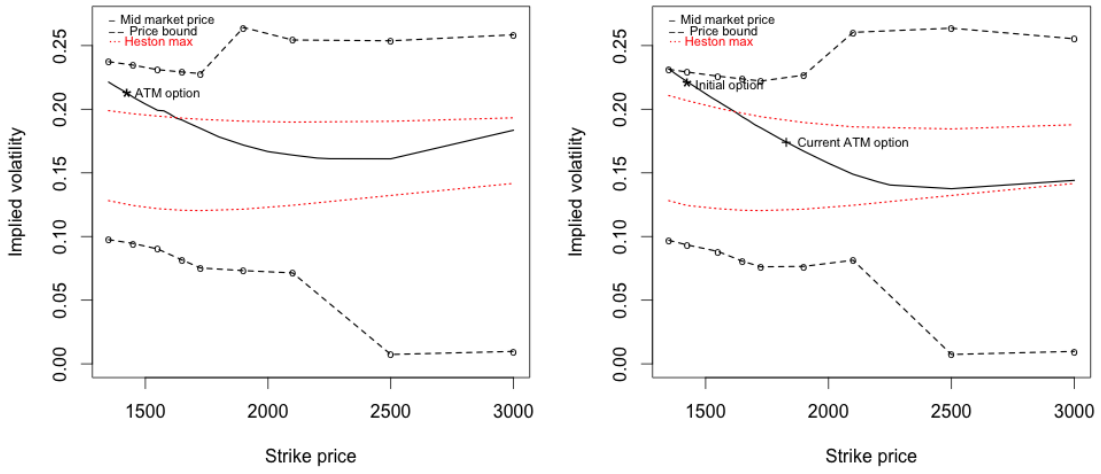

Figure 2.19: Left figure: mid-market implied volatility of the S&P 500 call option for different strikes as recorded on 2012-12-31. Corresponding model-boundaries (dashed lines) and formula-optimal prices (red dotted lines), both in terms of implied volatilities. The volatility of the ATM option (II) is marked with a star. **Right figure:** implied volatilities from the mid-market price, model-boundaries and formula-optimum, as recorded on 2013-12-23 (the last date of the considered period for option (II)).

market prices, while the prices from Heston's (optimised) formula yield much flatter, less sloped, curves. This indicates that we need a higher level of skewness to fit the curvature of market volatilities (roughly speaking, a stronger negative correlation to increase the slope and a higher level of "vol-of-vol" to increase the skew), and a higher volatility level overall for the ATM formula-price to fit with the market (a higher mean-reversion level and lower reversion speed; quite remarkably, this would require a negative market price of risk parameter λ, cf. equation (2.5)). For both dates, the pricing-bounds from the optimally controlled value process are wide enough to covers all strikes even if option (II) yields a boundary close to the market volatility at the later date where it is deeply in-the-money (see right pane of Figure 2.19). If we consider the ATM strike instead, we see model-bounds that are fairly symmetric around the market volatility. Furthermore, as in the case with formula-optimal prices, we note that the bounds do not exhibit any curvature in line with the market volatilities, supposedly because of the low level of negative correlation and vol-of-vol.

Returning to the prices of all options (I)-(IV), we see that average ranges of the model bounds are ~ 13 times the sizes of the market bid/offer spreads (Table 2.17). Comparing to the mid-market option price, we have that the average bound-range is $\sim 40 - 80$ percent of the option price while the spread is only $\sim 5\%$ of the price. The size of the model-bound interval is ultimately dependent on the uncertainty region for the drift parameters (effectively of the variance drift since the interest-rate uncertainty is negligible). So despite the fact that the price range of the model bounds is quite large, it's notable that it covers market-quotes almost everywhere since the model is completely specified from historical data of the underlying asset price, and not from the option-price data.

Furthermore, if we consider the option prices as obtained from optimising the Heston formula (Figure 2.18), we have that $\sim 40\%$ of the market quotes fall inside the model interval. The market quotes of option (II) are outside the model prediction throughout the period, and option (I) has only 11.8% of its quotes covered. The (average) ranges of the optimised prices are ~ 7 times the sizes of market spreads, and thus almost halved compared the ranges of the model bounds. The optimisation of Heston's formula based on statistical inference is clearly not sufficient to cover for the market quotes of the considered data. This method corresponds to an optimally controlled value process with parameter processes being constants, and we simply have to allow for these parameters to vary in order to cover the market pricing of options in a satisfactory way.

2.5 Stochastic volatility models with jumps

For the purpose of completion, we will generalise our modelling framework in this final section to a multi-asset setting under a Markovian stochastic volatility model with jumps. Our intension is to give a brief presentation of how the uncertainty pricing transfers to a general model and we deliberately avoid going too deep into details and technical assumptions.

2.5.1 A generic Markovian model

We consider a financial market model on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P)$ that consists of a money account B, paying a risk-free interest of deterministic rate r, and a $\mathbb{R}^d$-valued stochastic process $S = (S^1, \ldots, S^d)^\top$ representing the price processes of d risky assets. Furthermore, we have d' non-negative stochastic processes, V taking values in $\mathbb{R}^{d'}$, that represents the instantaneous variances. Typically, the two are of equal dimension such that each asset price is diffused by an individual volatility and we also assume that $d' = d$ is the case here. The statistical P-dynamics of the $m = 2d$ column-vector $X = (S; V)$ of state variables are assumed to be of the form

$$dX_t = \mu^P(X_t)dt + \sigma(X_t)dW_t + \int_{\mathcal{Z}} h(\xi, X_{t-})\tilde{\mu}(d\xi, dt)$$

where $\mu^p(\cdot)$ is the m-dimensional drift-function under the statistical measure, $\sigma(\cdot)$ the $m \times m$-valued diffusion matrix, and W is a $\mathbb{R}^m$-valued Wiener process. The jump part is driven by $\tilde{\mu}$, a compensated Poisson random measure on a Blackwell space $\mathcal{Z}$ with deterministic compensator $\mu_p(d\xi, dt) = \nu(d\xi)dt$, and $h(\cdot)$ is a state-dependent function valued in $\mathbb{R}^m$ that governs the jump sizes of X. Since we are working with a $\mathbb{R}^m$-dimensional state processes, we take $\mathcal{Z} = \mathbb{R}^m$. We assume that $\{\mathcal{F}_t\}_{t\geq 0}$ is generated by W and $\tilde{\mu}$ jointly, and augmented to satisfy the usual conditions. The functions μ^P, σ, h are assumed to be defined such that the SDE admits a unique solution up to a fixed deterministic time T (for instance of linear growth and locally Lipschitz continuous), and V being non-negative almost surely. Further, we assume sufficient integrability conditions such that the market model admits no arbitrage: there exists an equivalent martingale measure Q under which S and V follows

$$\begin{aligned} dS_t &= rS_t dt + \sigma_S(S_t, V_t)d\tilde{W}_t + \int_{\mathbb{R}^d} h(\xi, S_{t-}, V_t)\tilde{\mu}(d\xi, dt) \\ dV_t &= \mu_V(V_t, \Gamma)dt + \sigma_V(V_t)d\tilde{W}_t \end{aligned}$$

where $\sigma_S(\cdot)$ and $\sigma_V(\cdot)$, both with values in $\mathbb{R}^{d\times m}$, are the first and last d rows of σ. The $\mathbb{R}^d$-valued function $\mu_V(\cdot, \Gamma)$ is the Q-drift of the variance with parameters Γ, and $\tilde{W}$ is a m-dimensional Wiener process under Q. For our convenience, we have assumed that jumps affect the asset prices[19] only; $\mathcal{Z} = \mathbb{R}^d$ and $h(\cdot)$ is an $\mathbb{R}^d$-valued function while the compensator measure $\mu_p(d\xi, dt) = \nu(\gamma, d\xi)dt$ is dependent on Q-parameters γ (we use the same notation for μ_p here even if the compensator may be different under P and Q). Furthermore, we assume that the continuous variance has drift and diffusion functions dependent on the state of the variance alone. The risky assets are not assumed to carry any dividend payments, although the generalisation to non-zero dividends (as well as jumps

[19] As the jumps are generated by a Poisson random measure, S will have jumps given by

$$\Delta S_t = \int_{z\in\mathbb{R}^d} h(z, S_{t-}, V_t)\tilde{\mu}(dz, \{t\}) = h(z_t, S_{t-}, V_t)\mathbf{1}_{\{\Delta S_t \neq 0\}}$$

where $z_t \in \mathbb{R}^d$ is a (unique) point in the set where $\mu(\{z_t\}) = 1$.

and S-dependent coefficients for the variance) should be straightforward. As under P, we assume all coefficients under Q to sufficiently well behaved for an appropriate solution to exists.

The market model (S, B) is free of arbitrage but incomplete as it has more random sources $(2 \times d)$ than traded risky assets (d), and since the asset prices exhibit jumps (i.e. the risk-neutral measure Q is not unique). For a Markovian pricing rule

$$\begin{aligned} D_t &= D(t, S_t, V_t), \quad t \in [0, T], \\ D &: [0, T] \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}, \quad D \in \mathcal{C}^{1,2}, \end{aligned}$$

of a European option with terminal payoff $g(S_T)$, we have a pricing equation corresponding to (2.3) as given by

$$\begin{aligned} \frac{\partial D}{\partial t} + \mathcal{L}D - rD &= 0 \\ D(T, s, v) &= g(s) \end{aligned}$$

where $\mathcal{L}$ is the (time independent) inegro-differential operator that generates $(S; V)$ under Q. For a function $f(s, v) \in \mathcal{C}^2$ the operator is defined as

$$\begin{aligned} \mathcal{L}f(s,v) = \mu^Q(s,v)^\top \nabla_x f(s,v) + \frac{1}{2}\mathrm{tr}\left[\sigma^\top(s,v)\nabla^2_{xx} f(s,v)\sigma(s,v)\right] \\ + \int_{\xi \in \mathbb{R}^d} \Big(f(s + h(\xi, s, v), v) - f(s,v) - h(\xi, s, v)^\top \nabla_x f(s,v) \Big) \nu(d\xi) \end{aligned}$$

where ∇_x, ∇^2_{xx} with $x = (s, v)$ are the gradient and Hessian operators respectively, and $\mathrm{tr}[\cdot]$ the matrix trace. By the Feynman-Kac representation formula, this is equivalent to the risk-neutral valuation formula $D(t, s, v) = \mathbb{E}^Q \left[e^{-r(T-t)} g(S_T) \middle| (S_t, V_t) = (s, v) \right]$.

2.5.2 Pricing under parameter uncertainty

Here we introduce the controlled measure Q^u that represents the parameter uncertainty in our model. Hence, with $\{u_t\}_{t \geq 0}$ being a $\mathcal{F}_t$-predictable control process that takes its values in a compact uncertainty set $U \subset \mathbb{R}^k$, we let the controlled dynamics be

$$\begin{aligned} dS_t &= r_t S_t dt + \sigma_S(S_t, V_t) dW^u_t + \int_{\mathbb{R}^d} h(\xi, S_{t-}, V_t) \tilde{\mu}^u(d\xi, dt) \\ dV_t &= \mu_V(V_t, u_t) dt + \sigma_V(V_t) dW^u_t. \end{aligned}$$

where we assume that the controlled drift of the asset price and variance, $r_t S_t$ and $\mu_V(V_t, u_t) = \mu_V(V_t, \Gamma_t)$, to be of the same functional form under Q and Q^u. Hence, the control has components $u_t = (r_t, \Gamma_t, \gamma_t)$ where Γ are the parameters of μ_V, while γ are parameters to the controlled (form invariant) compensator measure

$$\mu^u_p(d\xi, dt) = \nu(\gamma_t, d\xi) dt$$

with Radon-Nikodym density $\beta(\xi,t,u_t) \equiv d\mu_p^u/d\mu_p$ with respect to μ_p. We let $\mu^{Q^u}(S_t,V_t,u_t) \equiv (r_tS_t;\mu_V(V_t,u_t))$ denote the common drift of $(S;V)$ under Q^u (and similarly for the common Q-drift). The effect of the control is then defined by the $\mathbb{R}^{1\times m}$-valued process

$$\alpha(S_t,V_t,u_t) = \sigma^{-1}(S_t,V_t)\left(\mu^{Q^u}(S_t,V_t,u_t) - \mu^Q(S_t,V_t) - \int_{\mathbb{R}^d} h(\xi,S_{t-},V_t)(\beta(\xi,t,u_t)-1)\nu(d\xi)\right)$$

to give the linear driver function $f(s,v,y,z,\theta,u) = -ry + z\alpha(s,v,u) + \int_\xi \theta(\xi)(\beta(\xi,t,u)-1)\nu(d\xi)$ where the second last argument is a function $\theta : \mathbb{R}^d \mapsto \mathbb{R}$. Modulo sufficient integrability and Lipschitz conditions, we have that the value function for a fixed admissible control $J_t(u) = \mathbb{E}_u[e^{-\int_t^T r_u du}g(S_T)|\mathcal{F}_t]$, $t \in [0,T]$, is given as part of the solution $(J(u),Z,\Theta)$ to the linear BSDE

$$\begin{aligned} dJ_t(u) &= -f(S_t,V_t,J_t(u),Z_t,\Theta_t,u_t)dt + Z_t d\tilde{W}_t + \int_{\xi\in\mathbb{R}^d}\Theta_t(\xi)\tilde{\mu}(d\xi,dt) \\ J_T(u) &= g(S_T) \end{aligned}$$

where Z is $\mathbb{R}^{1\times m}$-valued while Θ is a process taking its values in the space of functions $\theta : \mathbb{R}^d \mapsto \mathbb{R}$. The result follows similarly as in the case of Heston's model: apply Itô's product rule to $\mathcal{E}(\Lambda)J(u)$, where $\Lambda = -\int_0^\cdot r_t dt + \alpha(S_t,V_t,u_t)\bullet\tilde{W} + \int_0^\cdot\int(\beta(\xi,t,u_t)-1)\tilde{\mu}(d\xi,dt)$, to see that $\mathcal{E}(\Lambda)J(u)$ is a martingale, and use that $\mathcal{E}(\Lambda+\int_0^\cdot r_t dt)_T = dQ^u/dQ$ for the measure change of $\mathcal{E}(\Lambda)_t J_t(u) = \mathbb{E}[\mathcal{E}(\Lambda)_T J_T(u)|\mathcal{F}_t]$ to obtain the original expression for the value process after rearrangement. Further, defining the pointwise optimised driver functions over the compact uncertainty set

$$H^\pm(s,v,y,z,\theta) = \left|\operatorname*{ess\,sup}_{u\in U} \pm f(s,v,y,z,\theta,u)\right|,$$

we have by the comparison theorem that the the optimally controlled value processes (the upper/lower pricing boundaries) $\{D_t^\pm\}_{t\in[0,T]} = \{|\operatorname{ess\,sup}_{\{u_t\}} \pm J_t(u)|\}_{t\in[0,T]}$ are solutions to the BSDEs

$$\begin{aligned} dD_t^\pm &= -H^\pm(S_t,V_t,D_t^\pm,Z_t,\Theta_t)dt + Z_t d\tilde{W}_t + \int_{\xi\in\mathbb{R}^d}\Theta_t(\xi)\tilde{\mu}(d\xi,dt) \\ D_T^\pm &= g(S_T). \end{aligned}$$

Here as well, this is a consequence of the fact that we have a linear driver in y, z and θ, and from the comparison theorem for BSDEs. The proof, which we omit for brevity, follows in the same fashion as in the previous case with Heston's model (see Cohen and Elliott (2015), Chapter 21, for details). As well, since we work in a Markovian setting, we have that the solution can be written with a deterministic function $D_t = D(t,S_t,V_t)$, and the same holds for the optimal control: there exists a function $u^*(t,s,v)$ such that the feedback control $u_t^* = u^*(t,S_t,V_t)$ is the optimal control among all admissible controls.

Finally, as in the case with Heston's model, we have by the semilinear Feynman-Kac formula that $D(t,s,v)$ satisfies a semilinear partial differential equation

$$\frac{\partial D}{\partial t} + \frac{1}{2}\mathrm{tr}[\sigma^\top \nabla^2_{xx} D\sigma] + \operatorname*{ess\,inf}_{(r,\Gamma,\gamma)\in U} \left\{ -rD + (\mu^{Q^u})^\top \nabla_x D + \int_\xi (\Delta D - h^\top \nabla_x D)\nu(\gamma, d\xi) \right\} = 0$$

with terminal condition $D(T, s, v) = g(s)$, and where ΔD is shorthand notation for $D(t, s + h(\xi, s, v), v) - D(t, s, v)$. Although many numerical methods exist for a PIDE of this type, one may opt for simulating the BSDE solution instead (see e.g. Bouchard and Elie (2008)), especially when the dimensional of the problem is high.

2.6 Conclusion

Model uncertainty, here represented by parameter uncertainty, is an acknowledged concept formalised by Knight (1921) and its importance has been studied in the financial context at least since Derman (1996). The focus of this paper has been to investigate how parameter uncertainty could be incorporated into a stochastic volatility model, and how it affects derived prices of European option. The considered uncertainty was fairly general: interest rate and volatility drift parameters where allowed to change over time (constant-parameters being a special case) within a pre-described uncertainty region inferred from statistical estimation. The effect on pricing was then studied from a worst-case perspective with boundaries for the option price that could be embedded into a control problem, with the control playing a role of the uncertain parameters.

With Heston's model as a working example, the control problem–BSDE duality was then exploited and an explicit equation for the pricing boundary (the optimal value process) was derived in the form of a Markovian linear BSDE. A numerical scheme with several suggested modifications was considered for the solution of this BSDE, and an evaluation of the schemes was made in a known-outcome setting analogous to the dynamic-parameter setting. Based on bias/variance (and computational) considerations, a scheme was proposed for an empirical study of the methodology applied to real-world market data. Studying a set of bid/offer market quotes of European call options on the S&P 500 index and their corresponding model-price bounds, it was found that even if the model (and uncertainty set) was estimated from historical prices of the underlying, 98% of the market option prices was within the model-prescribed bounds. In contrast, $\sim 40\%$ of the market quotes was within the maximum/minimum model-price interval when constant parameters where used.

In both the dynamic and constant parameter setting, it was seen that the model implied volatilities did not follow the curvature of the market implied volatilities. A natural explanation for this observation is that the diffusion parameters, which effectively decide the slope and skew of the implied volatility curve, were estimated from asset-price data, and not from option-price data. An interesting empirical sequel would therefore be to study how the shape and coverage of model-price bounds change when parameters are calibrated from market option prices instead. We leave this for further investigation.

Finally, we note that prior beliefs and preferences about the uncertainty are not taken into consideration by the conservative approach with pricing boundaries. However, with

$L(u_t, u'_t)$ being some function that assigns a loss when u_t is used instead of the true parameters u'_t, we could incorporate beliefs with a value function of the form

$$J_t(u) = \mathbb{E}_u \left[e^{\int_t^T r_s ds} G + \int_t^T L(u_s, u'_s) ds \middle| \mathcal{F}_t \right]$$

that would lead to a similar linear BSDE. The loss could be based on the (approximative) normality of estimated parameters or some quantity related to an economic value, for instance a hedging error. In both cases, the value of the loss must be related to the value of the option payoff, an intricate task that we leave for further research along with this approach.

3

The Fundamental Theorem of Derivative Trading – Exposition, Extensions and Experiments

Simon Ellersgaard, Martin Jönsson and Rolf Poulsen[1]

Abstract. When estimated volatilities are not in perfect agreement with reality, delta hedged option portfolios will incur a non-zero profit-and-loss over time. However, there is a surprisingly simple formula for the resulting hedge error, which has been known since the late '90s. We call this The Fundamental Theorem of Derivative Trading. This paper is a survey with twists of that result. We prove a more general version of it and discuss various extensions and applications, from incorporating a multi-dimensional jump framework to deriving the Dupire-Gyöngy-Derman-Kani formula. We also consider its practical consequences both in simulation experiments and on empirical data thus demonstrating the benefits of hedging with implied volatility.

Keywords: Delta Hedging, Model Uncertainty, Volatility Arbitrage.

[1]All authors are with the Department of Mathematical Sciences, University of Copenhagen.

3.1 A meditation on the art of derivative hedging

Introduction. Of all possible concepts within the field mathematical finance, that of *continuous time derivative hedging* indubitably emerges as the central pillar. First used in the seminal work by Black and Scholes (1973),[2] it has become the cornerstone in the determination of no-arbitrage prices for new financial products. Yet a disconnect between this body of abstract mathematical theory and real world practice prevails. Specifically, successful hedging relies crucially on us having near perfect information about the model that drives the underlying asset. Even if we boldly adopt the standard stochastic differential equation paradigm of asset pricing, it remains to make exact specifications for the degree to which the price process reacts to market fluctuations (i.e. to specify the diffusion term, the volatility). Alas, volatility blatantly transcends direct human observation, being, as it were, a Kantian *Ding an sich*[3] of which we only have approximate knowledge.

One such source supervenes upon historical realisations of the underlying price process: for example, assuming that the governing model can at least *locally* be approximated as a geometric Brownian motion, one can proceed to measure the standard deviation of past log returns over time. Yet this procedure raises uncomfortable questions pertaining to statistical measurement: under ordinary circumstances, increasing the sample space should narrow the confidence interval around our sample parameter. Only here, there is no a priori way of telling if and when a model undergoes a structural change.[4] Inevitably, this implies that extending the time series of log returns too far into the past might lead to a less accurate estimator as we might end up sampling from a governing dynamics that is no longer valid. Of course, we may take *some* measures against this issue by trying our luck with ever more intricate time series analyses, until we stumble upon a model the parameters of which satisfy our arbitrary tolerance for statistical significance. Nevertheless, in practice this procedure invariably boils down to checking some finite basket of models and selecting the best one from the lot. Furthermore, unknown structural breaks continue to pose a problem no matter what.

Alternatively, we might try to extract an implied volatility from the market by fitting our model to observed option prices. Nevertheless the inadequacy of the methodology quickly becomes apparent: first, implied volatility might be ill-defined as it is the case for certain exotic products such as barrier options. Secondly, it is quite clear that the market hysteria which drives the prices of traded options need not capture the market hysteria

[2]That Black and Scholes along with Merton were the first is the general consensus, although the paper by Haug and Taleb (2011) shows that the view is not universal.

[3]Literally, *thing in itself* or the *noumenon*. Kant held that there is a distinction between the way things appear to observers (phenomena) and the way reality actually is construed (noumena).

[4]This scenario is not at all implausible. Unlike the physical sciences where the fundamental laws are assumed to have no *sufficient reason* to change (in the Leibnizian or Occamian sense), this philosophical principle would hardly withstand scrutiny in a social science context. Asset price processes are fundamentally governed by market agents and their reactions to various events (be they self-induced or exogenous). There is really no reason to assume that these market players will not drastically change their opinions at some point (for one reason or the other).

which drives the corresponding market for the underlying asset. Fair pricing ultimately boils down to understanding the true nature of the underlying product, not to mimic the collective madness of option traders.

Whilst volatility at its core remains elusive to us, the situation is perhaps not as dire as one might think. Specifically, we can develop a formal understanding of the profit-&-loss we incur upon hedging a portfolio with an erroneous volatility - at least insofar as we make some moderate assumptions of the dynamical form of the underlying assets. To give a concrete example of this, consider the simple interest rate free framework presented in Andreasen (2003) where the price process of a single non-dividend paying asset is assumed to follow the real dynamics

$$dX_t = X_t(\mu_{t,r}dt + \sigma_{t,r}dW_t).$$

Let V_t^i be the value of an option that trades in the market at a certain implied volatility σ_i - possibly quite different from the epistemically inaccessible $\sigma_{t,r}$. Now if we were to set up a hedge of a long position on such an option, using σ_i as our hedge volatility, an application of Itō's formula, coupled with the Black-Scholes equation, shows that the infinitesimal value change in the hedge portfolio $\Pi_t = V_t^i - \partial_x V_t^i \cdot X_t$, is of the form

$$d\Pi_t = \tfrac{1}{2}(\sigma_{t,r}^2 - \sigma_i^2)X_t^2\partial_{xx}^2 V_t dt, \tag{3.1}$$

which generally is non-zero unless $\sigma_i = \sigma_{t,r}$. For reasons that will become clearer below, the importance of this result is of such magnitude that Andreasen dubs it *The fundamental theorem of derivative trading*. Indeed, a more abstract variation of it will be the central object of study in this paper.

To the best of our knowledge, quantitative studies into the effect of hedging with an erroneous volatility first appeared in a paper on the robustness of the Black-Scholes formula by Karoui et al. (1998). They viewed the result as a largely negative one: unless volatility is bounded (which it is not in any stochastic volatility model) then there is no simple super-replication strategy. Easier derivations of the fundamental theorem are encountered in the papers by Gibson et al. (1999), Mahayni et al. (1999) and Rasmussen (2001). Indeed, there seems to be an added awareness of the result being a positive one: a decent forecast of volatility gives rise to a small hedge error. Jointly, what characterises these papers is their binary partition of volatility into a "wrong" and "right" category. Somewhat subtler treatments can be found in the unpublished works by Carr (2002) and Henrard (2001) in which the partition becomes tripartite: concretely, these papers explicitly differentiate between a true governing volatility of the underlying asset, an implied volatility characterising market consensus, and a hedge volatility which captures the personal belief of the option hedger. The benefits of this three-part structure are clearly enunciated in Ahmad and Wilmott (2005) in which the associated P&L paths under the various choices of volatility are exhibited, alongside mathematical (non-analytic) expressions for the mean implied P&L and its variance.

The fundamental theorem has in other words received extensive treatment in the academic literature, yet it has never quite reached the "textbook" status we reckon it deserves (in

particular, it rarely finds its way into the curriculum of aspiring financial engineers). To end this regrettable situation, we here provide a thorough exposition. Taking our vantage point in the tripartite philosophy of Henrard and Carr, we endeavour to provide a simple proof of the fundamental theorem, whilst simultaneously spicing up the result by considering a more general dynamics in a full-fledged multi-asset framework.

Overview. The structure of this paper is as follows: in Section 3.2 we state and prove a generalised version of the fundamental theorem of derivative trading and discuss its various implications for hedging strategies and applications. Key insights here include the benefit of hedging with the implied volatility with regards to attaining "smooth" P&L paths, and a somewhat surprising connection to Dupire's local volatility formula. In Section 3.3 we expose the implications of adding a multi-dimensional jump process to the dynamics, thus emphasising the relative ease with which the original proof can be adapted. In particular, we argue that long option positions with convex pay-out profiles profit from a discontinuous movement in one of the underlying stocks. Finally, Section 3.4 presents an empirical investigation into what actually happens to our portfolio when we hedge using various volatilities. Using actual quotes for stocks and call options, we demonstrate that self-financing portfolios hedged at the implied volatility indeed give rise to a smoother P&L over time, whilst allowing for some amount of volatility arbitrage to be picked up in the process.

3.2 The fundamental theorem of derivative trading

3.2.1 Model set-up

Consider a financial market comprised of a risk-free money account as well as n risky assets, each of which pays out a continuous dividend yield. We assume all assets to be infinitely divisible as to the amount which may be held, that trading takes place continuously in time, and that no trade is subject to financial friction. Formally, we imagine the information flow of this world to be captured by the stochastic basis $(\Omega, \mathscr{F}, \mathbb{F}, \mathbb{P})$, where Ω represents all possible states of the economy, $\mathbb{P}$ is the physical probability measure, and $\mathbb{F} = \{\mathscr{F}_t\}_{t\geq 0}$ is a filtration which satisfies the *usual conditions*.[5] The price processes of the risky assets, $\boldsymbol{X}_t = (X_{1t}, X_{2t}, ..., X_{nt})^\intercal$, are assumed to follow the real dynamics[6]

$$d\boldsymbol{X}_t = \mathbf{D}_{\boldsymbol{X}_t}[\boldsymbol{\mu}_r(t, \tilde{\boldsymbol{X}}_t)dt + \boldsymbol{\sigma}_r(t, \tilde{\boldsymbol{X}}_t)d\boldsymbol{W}_t], \tag{3.2}$$

[5]Specifically, it satisfies right-continuity, $\cap_{s\geq t}\mathscr{F}_s = \mathscr{F}_t, \forall t \geq 0$ (if we move incrementally forward in time there will be no jump in information), and completeness, i.e. $\mathscr{F}_0$ contains all $\mathbb{P}$ null sets.

[6]The nomenclature "real dynamics" is ripe with unfortunate connotations of Platonic realism (ontological significance of mathematical objects). Strictly speaking, this is not what we require, but rather the *expressive adequacy* of a model: i.e. it's ability to adequately capture the financial events unfolding.

where $\mathbf{D}_{\boldsymbol{X}}$ is the $n \times n$ diagonal matrix $\text{diag}(X_{1t}, X_{2t}, ..., X_{nt})$, and $\boldsymbol{W}_t = (W_{1t}, W_{2t}, ..., W_{nt})^\intercal$ is an n-dimensional standard Brownian motion adapted to $\mathbb{F}$. Furthermore, $\boldsymbol{\mu}_r : [0,\infty)\times\mathbb{R}^{n+m} \mapsto \mathbb{R}^n$ and $\boldsymbol{\sigma} : [0,\infty)\times\mathbb{R}^{n+m} \mapsto \mathbb{R}^{n\times n}$ are deterministic functions, sufficiently well-behaved for the SDE to have a unique strong solution (in particular, we assume the regularity conditions

$$\int_t^s |\mathbf{D}_{\boldsymbol{X}_u}\boldsymbol{\mu}_r(u, \widetilde{\boldsymbol{X}}_u)|du < \infty, \qquad \int_t^s |\mathbf{D}_{\boldsymbol{X}_u}\boldsymbol{\sigma}_r(u, \widetilde{\boldsymbol{X}}_u)|^2 du < \infty, \tag{3.3}$$

hold a.s. $\forall t \leq s$, where the first norm is to be understood in the Euclidian sense, whilst the latter should be construed in the matrical sense).[7] Finally, we define $\widetilde{\boldsymbol{X}}_t$ as the $n+m$ dimensional vector $(\boldsymbol{X}_t; \boldsymbol{\chi}_t)$ where $\boldsymbol{\chi}_t = (\chi_{1t}, \chi_{2t}, ..., \chi_{mt})^\intercal$ has the interpretation of an m-dimensional state variable, the exact dynamical nature of which is not integral to what follows.[8]

In what follows we consider the scenario of what happens when we hedge an option on $\boldsymbol{X}_t$, ignorant of the existence of the state variable $\boldsymbol{\chi}_t$, as well as the form of $\boldsymbol{\mu}_r(\cdot,\cdot)$ and $\boldsymbol{\sigma}_r(\cdot,\cdot)$. Specifically, we shall imagine that we are misguided to the extent that we would model the dynamics of $\boldsymbol{X}_t$ as a local volatility model with diffusion matrix $\boldsymbol{\sigma}_h(t, \boldsymbol{X}_t)$. Similar assumptions pertain to the market, although here we label the "implied" diffusion matrix $\boldsymbol{\sigma}_i(t, \boldsymbol{X}_t)$ to distinguish it from our personal belief. Irrespective of which dynamical specification is being made, we maintain that regularity conditions analogous to (3.3) remain satisfied. Finally, a cautionary remark: throughout these pages we use r and i to emphasise that the volatility is *real* and *implied* respectively, whilst h refers to an arbitrary *hedge* volatility. For a comprehensible reading, it is incumbent that the reader keeps these definitions in mind.

3.2.2 Theorem and derivation

Theorem 3.1. The fundamental theorem of derivative trading. Let $V_t = V(t, \boldsymbol{X}_t) \in \mathcal{C}^{1,2}([0,T]\times\mathbb{R}^n)$ be the price process of a European option with terminal pay-off $V_T = g(\boldsymbol{X}_T)$, the underlying of which follows the real dynamics (3.2). We assume the option trades in the market at the (not necessarily uniquely determined) implied volatility $\boldsymbol{\sigma}_i = \boldsymbol{\sigma}_i(t, \boldsymbol{X}_t)$. Now suppose we at time $t = 0$ acquire such an option for the implied price V_0^i and set out to Δ-hedge our position. Said hedge is performed under the notion that the volatility function ought, in fact, to be of the form $\boldsymbol{\sigma}_h = \boldsymbol{\sigma}_h(t, \boldsymbol{X}_t)$, leading to the fair price process V_0^h. Then the present value of the **profit-&-loss** we incur from holding such

[7] Specifically, if $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{A} \in \mathbb{R}^{n\times d}$, the the Euclidian norm is defined as $|\boldsymbol{x}| \equiv (\sum_{i=1}^n x_i^2)^{1/2}$, while the matrical norm is $|\boldsymbol{A}| \equiv (\sum_{i=1}^n \sum_{j=1}^d A_{ij}^2)^{1/2}$.

[8] Nonetheless, a common assumption in the stochastic volatility literature is obviously to let $\boldsymbol{\chi}$ be driven by a stochastic differential equation of the form $d\boldsymbol{\chi}_t = \boldsymbol{m}(\boldsymbol{\chi}_t)dt + \boldsymbol{v}(\boldsymbol{\chi}_t)d\boldsymbol{W}_t + \bar{\boldsymbol{v}}(\boldsymbol{\chi}_t)d\bar{\boldsymbol{W}}_t$, where $\bar{\boldsymbol{W}}_t$ is second standard Brownian motion (independent of the first), and $\boldsymbol{m}$, $\boldsymbol{v}$ and $\bar{\boldsymbol{v}}$ are dimensionally consistent, regularity conforming vectors and matrices.

a portfolio over the interval $\mathbb{T} = [0, T]$ is

$$\mathrm{P\&L}^h_{\mathbb{T}} = V^h_0 - V^i_0 + \tfrac{1}{2}\int_0^T e^{-\int_0^t r_u du}\ \mathrm{tr}[\mathbf{D}_{\boldsymbol{X}_t}\boldsymbol{\Sigma}_{rh}(t, \widetilde{\boldsymbol{X}}_t)\mathbf{D}_{\boldsymbol{X}_t}\nabla^2_{\boldsymbol{xx}}V^h_t]dt, \tag{3.4}$$

where $r_u = r(u, \boldsymbol{X}_u)$ is the locally risk free rate, $\nabla^2_{\boldsymbol{xx}}$ is the Hessian operator, and

$$\boldsymbol{\Sigma}_{rh}(t, \widetilde{\boldsymbol{X}}_t) \equiv \boldsymbol{\sigma}_r(t, \widetilde{\boldsymbol{X}}_t)\boldsymbol{\sigma}^\intercal_r(t, \widetilde{\boldsymbol{X}}_t) - \boldsymbol{\sigma}_h(t, \boldsymbol{X}_t)\boldsymbol{\sigma}^\intercal_h(t, \boldsymbol{X}_t), \tag{3.5}$$

is a matrix which takes values in $\mathbb{R}^{n\times n}$.

Proof: Let $\{\Pi^h_t\}_{t\in[0,T]}$ be the value process of the hedge portfolio long one option valued according to the implied market conception, $\{V^i_t\}_{t\in[0,T]}$, and short $\{\Delta^h_t = \nabla_{\boldsymbol{x}}V^h_t\}_{t\in[0,T]}$ units of the underlying with value process $\{\boldsymbol{X}_t\}_{t\in[0,T]}$, where $\nabla_{\boldsymbol{x}}$ is the gradient operator. We suppose the money account B is chosen such that the net value of the position is zero:

$$\Pi^h_t = V^i_t + B_t - \nabla_{\boldsymbol{x}}V^h_t \bullet \boldsymbol{X}_t = 0,$$

where $\bullet$ is the dot product. Now consider the infinitesimal change to the value of this portfolio over the interval $[t, t+dt]$, where $t \in [0, T)$. From the *self-financing condition* we have that

$$d\Pi^h_t = dV^i_t + r_t B_t dt - \nabla_{\boldsymbol{x}}V^h_t \bullet (d\boldsymbol{X}_t + \boldsymbol{q}_t \circ \boldsymbol{X}_t dt),$$

where $\boldsymbol{q}_t = (q_1(t, X_{1t}), q_2(t, X_{2t}), ..., q_n(t, X_{nt}))^\intercal$ codifies the continuous dividend yields and $\circ$ is the Hadamard product.[9] Jointly, the two previous equations entail that

$$d\Pi^h_t = dV^i_t - \nabla_{\boldsymbol{x}}V^h_t \bullet (d\boldsymbol{X}_t - (r_t\boldsymbol{\iota} - \boldsymbol{q}_t) \circ \boldsymbol{X}_t dt) - r_t V^i_t dt, \tag{3.6}$$

where $\boldsymbol{\iota} = (1, 1, ..., 1)^\intercal \in \mathbb{R}^n$.

Now consider the option valued under $\boldsymbol{\sigma}_h(t, \boldsymbol{X}_t)$; from the multi-dimensional Itō formula[10] we have that

$$dV^h_t = \{\partial_t V^h_t + \tfrac{1}{2}\mathrm{tr}[\boldsymbol{\sigma}^\intercal_r(t, \widetilde{\boldsymbol{X}}_t)\mathbf{D}_{\boldsymbol{X}_t}\nabla^2_{\boldsymbol{xx}}V^h_t\mathbf{D}_{\boldsymbol{X}_t}\boldsymbol{\sigma}_r(t, \widetilde{\boldsymbol{X}}_t)]\}dt + \nabla_{\boldsymbol{x}}V^h_t \bullet d\boldsymbol{X}_t, \tag{3.7}$$

where we have used the fact that $\boldsymbol{X}_t$ is governed by (3.2). Meanwhile, V^h_t satisfies the multi-dimensional Black Scholes equation for dividend paying underlyings,[11]

$$r_t V^h_t = \partial_t V^h_t + \nabla_{\boldsymbol{x}}V^h_t \bullet ((r_t\boldsymbol{\iota} - \boldsymbol{q}_t) \circ \boldsymbol{X}_t) + \tfrac{1}{2}\mathrm{tr}[\boldsymbol{\sigma}^\intercal_h(t, \boldsymbol{X}_t)\mathbf{D}_{\boldsymbol{X}_t}\nabla^2_{\boldsymbol{xx}}V^h_t\mathbf{D}_{\boldsymbol{X}_t}\boldsymbol{\sigma}_h(t, \boldsymbol{X}_t)]. \tag{3.8}$$

Combining this expression with the Itō expansion we obtain,

$$\begin{aligned} 0 = &- dV^h_t + r_t V^h_t dt + \nabla_{\boldsymbol{x}}V^h_t \bullet (d\boldsymbol{X}_t - (r_t\boldsymbol{\iota} - \boldsymbol{q}_t) \circ \boldsymbol{X}_t dt) \\ &+ \tfrac{1}{2}\mathrm{tr}[\mathbf{D}_{\boldsymbol{X}_t}\big(\boldsymbol{\sigma}_r(t, \widetilde{\boldsymbol{X}}_t)\boldsymbol{\sigma}^\intercal_r(t, \widetilde{\boldsymbol{X}}_t) - \boldsymbol{\sigma}_h(t, \boldsymbol{X}_t)\boldsymbol{\sigma}^\intercal_h(t, \boldsymbol{X}_t)\big)\mathbf{D}_{\boldsymbol{X}_t}\nabla^2_{\boldsymbol{xx}}V^h_t]dt \end{aligned} \tag{3.9}$$

[9] Otherwise known as the entry-wise product. Per definition, if A and B are matrices of equal dimensions, then $(A \circ B)_{ij} = A_{ij}B_{ij}$.

[10] See for instance Björk (2009), p. 65.

[11] See for instance Björk, Theorem 13.1 and Proposition 16.7.

where we have used the fact that the trace is invariant under cyclic permutations of its constituent matrices. Finally, defining $\mathbf{\Sigma}_{rh}(t, \widetilde{\boldsymbol{X}}_t)$ as in (3.5), and adding (3.9) to (3.6) we obtain

$$\begin{aligned} d\Pi_t^h = & \; dV_t^i - dV_t^h - r_t(V_t^i - V_t^h)dt + \tfrac{1}{2}\mathrm{tr}[\mathbf{D}_{\boldsymbol{X}_t}\mathbf{\Sigma}_{rh}(t, \widetilde{\boldsymbol{X}}_t)\mathbf{D}_{\boldsymbol{X}_t}\nabla^2_{\boldsymbol{xx}}V_t^h]dt \\ = & \; e^{\int_0^t r_u du}d(e^{-\int_0^t r_u du}(V_t^i - V_t^h)) + \tfrac{1}{2}\mathrm{tr}[\mathbf{D}_{\boldsymbol{X}_t}\mathbf{\Sigma}_{rh}(t, \widetilde{\boldsymbol{X}}_t)\mathbf{D}_{\boldsymbol{X}_t}\nabla^2_{\boldsymbol{xx}}V_t^h]dt. \end{aligned} \tag{3.10}$$

Whilst a perfect hedge would render this infinitesimal value-change in the portfolio *zero*, this is clearly not the case here. In fact, upon discounting (3.10) back to the present ($t = 0$) and integrating up the infinitesimal components, we find that net profit-&-loss incurred over the life-time of the portfolio is

$$\begin{aligned} P\&L_{\mathbb{T}}^h = & \int_0^T d(e^{-\int_0^t r_u du}(V_t^i - V_t^h)) + \int_0^T e^{-\int_0^t r_u du}\tfrac{1}{2}\mathrm{tr}[\mathbf{D}_{\boldsymbol{X}_t}\mathbf{\Sigma}_{rh}(t, \widetilde{\boldsymbol{X}}_t)\mathbf{D}_{\boldsymbol{X}_t}\nabla^2_{\boldsymbol{xx}}V_t^h]dt \\ = & \; V_0^h - V_0^i + \tfrac{1}{2}\int_0^T e^{-\int_0^t r_u du}\;\mathrm{tr}[\mathbf{D}_{\boldsymbol{X}_t}\mathbf{\Sigma}_{rh}(t, \widetilde{\boldsymbol{X}}_t)\mathbf{D}_{\boldsymbol{X}_t}\nabla^2_{\boldsymbol{xx}}V_t^h]dt. \end{aligned}$$

where $P\&L_{\mathbb{T}}^h \equiv \int_0^T e^{-\int_0^t r_u du}d\Pi_t^h$, and the last line makes use of the fact that $V_T^i = V_T^h = g(\boldsymbol{X}_T)$. This is the desired result. □

Remark 3.1. A few observations on this proof are in order: first, the relative simplicity of (3.4) clearly boils down to the assumption that the market is perceived to be driven by a local volatility model. If this assumption is dropped equation (3.8) no longer holds. Secondly, it should be clear that the value of the $P\&L$ changes sign if we are short on the derivative and long on the underlying. Thirdly, the market price of the derivative enters only though the initial price V_0. That is because we look at the profit-&-loss accrued over the entire life-time of the portfolio. The case of marking-to-market requires further analysis and/or assumption. We will elaborate on this in the following subsection.

Remark 3.2. From a generalist's perspective, theorem 1 suffers from a number of glaring limitations: for instance, the governing asset price dynamics only considers Brownian stochasticity, the hedge is assumed to be a workaday Δ-hedge, and the option type is vanilla European in the sense that the terminal pay-off is determined by the instantaneous price of the underlying assets. Fortunately, the fundamental theorem can readily be extended in various directions: e.g. it can be shown that if $V_t = V(t, X_t, A_t)$ is an Asian option written on the continuous average A_t of the underlying process X_t, then the fundamental theorem remains form invariant. In Section 1.3 we consider one particularly topical dynamical modification viz. the incorporation of possible market crashes through jump diffusion.

3.2.3 The implications for Δ-hedging.

From a first inspection, the fundamental theorem quite clearly demonstrates that reasonably successful hedging is possible *even* under significant model uncertainty. Indeed, as

Davis (2010) puts it "without some robustness property of this kind, it is hard to imagine that the derivatives industry could exist at all". In this section, we dive further into the implications of what happens to our portfolio, by considering the case where we hedge with (a) the real volatility, and (b) the implied volatility. It is a standing assumption in this subsection that whatever is used as hedge volatility is of the form that allows the use of Theorem 1 (see the discussion in Remark 1 above). The reader can think of the cases of constant volatility.

Hedging with the real volatility. Suppose we happen to be bang-on our estimate of the real volatility matrix in our Δ-hedge, i.e. let $\boldsymbol{\sigma}_h(t, \boldsymbol{X}_t) = \boldsymbol{\sigma}_r(t, \widetilde{\boldsymbol{X}}_t)$ a.s. $\forall t \in [0, T]$, then $\boldsymbol{\Sigma}_{rr}(t, \widetilde{\boldsymbol{X}}_t) = \mathbf{0}$ and the present valued profit-&-loss amounts to

$$P\&L^r_{\mathbb{T}} = V^r_0 - V^i_0,$$

which is manifestly deterministic.[12] However, we observe that this relies crucially on us holding the portfolio until expiry of the option. Day-to-day fluctuations of the profit-&-loss still vary stochastically (erratically) as it is vividly demonstrated by combining equation (3.9) (where $h = i$) with equation (3.6) (where $h = r$):

$$\begin{aligned} d\Pi^r_t = {} & \tfrac{1}{2}\ \mathrm{tr}[\mathbf{D}_{\boldsymbol{X}_t}\boldsymbol{\Sigma}_{ri}(t, \widetilde{\boldsymbol{X}}_t)\mathbf{D}_{\boldsymbol{X}_t}\nabla^2_{\boldsymbol{xx}}V^i_t]dt \\ & + \nabla_{\boldsymbol{x}}\left(V^i_t - V^r_t\right) \bullet \left\{(\boldsymbol{\mu}^r_t - r_t\boldsymbol{\iota} + \boldsymbol{q}_t) \circ \boldsymbol{X}_t dt + \mathbf{D}_{\boldsymbol{X}_t}\boldsymbol{\sigma}_r(t, \widetilde{\boldsymbol{X}}_t)d\boldsymbol{W}_t\right\}, \end{aligned}$$

cf. the explicit dependence of the Brownian increment. As for the *profitability* of the Δ-hedging strategy, this is a complex issue which ultimately must be studied on a case-by-case basis. However, for options with positive *vega*,[13] it suffices to require that the real volatility everywhere exceeds the implied volatility.

Hedging with the implied volatility. Suppose instead we hedge the portfolio using the implied volatility matrix $\boldsymbol{\sigma}_i(t, \boldsymbol{X}_t)$, $\forall t \in [0, T]$, then the associated present-valued profit-&-loss is of the form

$$P\&L^i_{\mathbb{T}} = \tfrac{1}{2}\int_0^T e^{-\int_0^t r_u du}\ \mathrm{tr}[\mathbf{D}_{\boldsymbol{X}_t}\boldsymbol{\Sigma}_{ri}(t, \widetilde{\boldsymbol{X}}_t)\mathbf{D}_{\boldsymbol{X}_t}\nabla^2_{\boldsymbol{xx}}V^i_t]dt.$$

As we find ourselves integrating over the stochastic process $\boldsymbol{X}_t$, this profit-&-loss is manifestly stochastic. Notice though that $d\Pi^i_t$ here does *not* depend explicitly on the Brownian increment (the daily profit-and-loss is $\mathcal{O}(dt)$) which gives rise to point that "bad models cause bleeding - not blow-ups". As for the *profitability* of the strategy, again this is a complex issue: however, insofar as $\boldsymbol{\Sigma}_{ri}(t, \widetilde{\boldsymbol{X}}_t) \circ \nabla^2_{\boldsymbol{xx}}V^i_t$ is *positive definite* a.s. for all $t \in [0, T]$,

[12] Obviously, this can only be the case if there is no underlying state variable.

[13] A clear example of vega being positive would be European calls and puts, which satisfy the assumptions needed to derive the Black-Scholes formula. Explicitly, $\nu \equiv \frac{\partial V}{\partial \sigma} = S_t e^{-\delta(T-t)}\phi(d_1)\sqrt{T-t} > 0$ where ϕ is the standard normal pdf and d_1 has the usual definition.

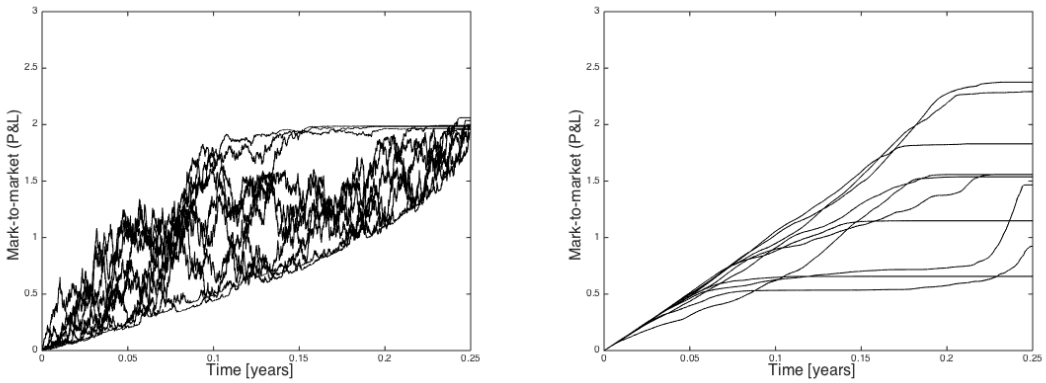

Figure 3.1: Left: Delta hedging a portfolio assuming that $\sigma_h = \sigma_r$. The parameter specifications are: $r = 0.05$, $\mu = 0.1$, $\sigma_i = 0.2$, $\sigma_r = 0.3$, $S_0 = 100$, $K = 100$, $q = 0$ and $T = 0.25$. The portfolio is rebalanced 5000 times during the lifetime of the option. Observe that while the P&L fluctuates randomly along the path of S_t due to the presence of dW_t, the accumulated P&L at the maturity of the option is the deterministic quantity $\Pi_T = e^{rT}(V_0^r - V_0^i)$. From the Black-Scholes formula it follows that $V_0^r = 6.583$ and $V_0^i = 4.615$ so $\Pi_{T=1} = 1.993$. The fact that our ten paths only approximately hit this terminal value is attributable to the discretisation of the hedging which should be done in continuous time. **Right:** Delta hedging a portfolio assuming that $\sigma_h = \sigma_i$.The parameter specifications are as before. Evidently, the accumulated P&L stays highly path dependent for the *entire* duration of the option. However, the curves per se are smooth, which highlights that $d\Pi_t^i$ does not depend explicitly on the Brownian increment.

then we're making a profit with probability one. To see this, recall that the trace can be written as[14]

$$\mathrm{tr}[\mathbf{D}_{\boldsymbol{X}_t}\boldsymbol{\Sigma}_{ri}(t,\widetilde{\boldsymbol{X}}_t)\mathbf{D}_{\boldsymbol{X}_t}\nabla^2_{\boldsymbol{xx}}V_t^i] = \boldsymbol{X}_t^\intercal(\boldsymbol{\Sigma}_{ri}(t,\widetilde{\boldsymbol{X}}_t)\circ\nabla^2_{\boldsymbol{xx}}V_t^i)\boldsymbol{X}_t,$$

In particular, if $\boldsymbol{\Sigma}_{ri}(t,\widetilde{\boldsymbol{X}}_t)\circ\nabla^2_{\boldsymbol{xx}}V_t^i$ is positive definite at all times, i.e.

$$\forall t\in[0,T],\quad \forall \boldsymbol{X}_t\in\mathbb{R}^n:\quad \boldsymbol{X}_t^\intercal(\boldsymbol{\Sigma}_{ri}(t,\widetilde{\boldsymbol{X}}_t)\circ\nabla^2_{\boldsymbol{xx}}V_t^i)\boldsymbol{X}_t>0,$$

then $P\&L^i_{\mathbb{T}} > 0$. A sufficient condition for this to be the case is that $\boldsymbol{\Sigma}_{ri}(t,\widetilde{\boldsymbol{X}}_t)$ and $\nabla^2_{\boldsymbol{xx}}V_t^i$ individually are positive definite $\forall t$, as demonstrated by the *Schur Product Theorem*.

Wilmott's hedge experiment. The points imbued in the previous two paragraphs are forcefully demonstrated in the event that there is only one risky asset in existence, the derivative is a European call option and all volatilities are assumed constant. Based on Wilmott and Ahmad, Figure 3.1 clearly illustrates the behaviour of the profit-&-loss paths insofar as we hedge with (a) the real volatility, and (b) the implied volatility. Again, the main insights are as follows: hedging V_t^i with the real volatility causes the $P\&L$ of the portfolio to fluctuate erratically over time, only to land at a deterministic value

[14]This follows from the general identity for matrices $\mathbf{A}$ and $\mathbf{B}$ of corresponding dimensions: $\mathbf{x}^\intercal(\mathbf{A}\circ\mathbf{B})\mathbf{y} = \mathrm{tr}[\mathbf{D}_{\mathbf{x}}\mathbf{A}\mathbf{D}_{\mathbf{y}}\mathbf{B}^\intercal]$ where $\mathbf{x}$ and $\mathbf{y}$ are vectors.

at maturity. On the other hand, hedging V_t^i with the implied volatility yields smoother (albeit still stochastic) P&L curves. Nonetheless, here there is no way of telling what the P&L actually amounts to at maturity.

Rather perturbingly, both strategies blatantly suggest the relative ease with which we can make *volatility arbitrage*. Specifically, assuming that the historical volatility is a reasonable proxy for the real volatility, $\sigma_{\text{hist}} \approx \sigma_r$, and that $\sigma_{\text{hist}} > \sigma_i$ ($\sigma_{\text{hist}} < \sigma_i$), it would suffice to go long (short) on the hedge portfolio for $\mathbb{P}(P\&L_{\mathbb{T}} \geq 0) = 1$ and $\mathbb{P}(P\&L_{\mathbb{T}} > 0) > 0$. Reality, of course, is not always as simple as our abstract idealisations, wherefore we dedicate Section 3.4 to an empirical investigation of Wilmott's experiment.

Remark 3.3. Wilmott's experiment reflects the concerting approach of traders: hedge with market implied parameters (see e.g. Hull and White (2016)). But for the experiment's strong smoothness property to hold for implied-volatility hedging, it is important that the implied volatility is assumed constant, or at least that it does not diffuse randomly as it would in e.g. Heston's model. Whenever the latter is the case, anything might happen. However, since implied volatility is intricately linked to (conditional expectation of) time-integrated real volatility,[15] it is likely to be quantitatively smoother than real volatility, so there is hope that the conclusion of the idealized experiment carries over to empirical analysis.

Remark 3.4. It is tempting to think of an option trade as a zero-sum game (up to a risk-premium) between the buyer and the seller; if one wins the other loses. That, however, it not true. Imagine a stock dynamics which is described by geometric Brownian motion with 15% volatility. A buyer is strongly bullish and makes a directional bet by buying an at-the-money call-option, which he is willing to pay 20% implied volatility on. The buyer does not hedge his position, the market rallies, and he makes a nice profit. Meanwhile, the seller is right in her 15%-volatility forecast, Δ-hedges and makes a nice (arbitrage) profit too. Needless to say, option writing can also imply losses for both the buyer and the seller as we saw it during the financial crisis.[16]

3.2.4 Applications

Due to the presence of the real volatility, the exact nature of which transcends our epistemic domain, one might reasonably ponder whether the fundamental theorem conveys any practical points besides those of the preceding subsection. Using two poignant (even if somewhat eccentric) examples, we will argue that the gravity of the fundamental theorem propagates well into risk management and volatility surface calibration. Zero rates and dividends will be assumed throughout.

Example 3.1. Let $V_t(T, K)$ be the price process of a European strike K maturity T call or put option, written on an underlying which obeys geometric Brownian motion,

[15] Think of the Black-Scholes model with time-dependent volatility or see Romano and Touzi (1997), Proposition 4.1.

[16] This example is inspired by Antoine Savine.

$dX_t = X_t[\mu_r dt + \sigma_r dW_t]$, where μ_r, σ_r are constants. Suppose we Δ-hedge a long position on V_t at the *implied* volatility, $\sigma_h = \sigma_i$, then the fundamental theorem implies that

$$P\&L^i_{\mathbb{T}} = \tfrac{1}{2}\int_0^T (\sigma_r^2 - \sigma_i^2) X_t^2 \Gamma_t^i dt,$$

where

$$\Gamma_t^i \equiv \frac{\phi(d_1^i)}{X_t \sigma_i \sqrt{T-t}},$$

is the option's gamma, $\phi : \mathbb{R} \mapsto \mathbb{R}_+$ is the standard normal pdf and

$$d_1^i \equiv \tfrac{1}{\sigma_i\sqrt{T-t}}\left\{\ln(X_t/K) + \tfrac{1}{2}\sigma_i^2(T-t)\right\}.$$

Since $\Gamma_t^i > 0$, $\forall t$, the strategy is profitable if and only if $\sigma_r^2 > \sigma_i^2$. Furthermore, by maximising the integrand with respect to X_t we find that the $P\&L^i_{\mathbb{T}}$ is maximal when

$$X_t^* = K e^{\frac{1}{2}\sigma_i^2(T-t)},$$

Specifically, upon evaluating the integral explicitly it can be shown that[17]

$$\max_{X_t} P\&L^i_{\mathbb{T}} = \sqrt{\frac{T}{2\pi}}\frac{K}{\sigma_i}(\sigma_r^2 - \sigma_i^2).$$

From a risk management point of view, the important point is that we can compute a confidence interval for the real volatility based on historical observations. Hence, we can compute a confidence interval for the maximal profit-&-loss we might face upon holding the hedge portfolio till expiry.

Example 3.2. Let $V_t = C_t(T,K)$ be the price process of a European strike K maturity T call option written on an underlying price process X. As in (3.2) we assume the fundamental dynamics to be of the form $dX_t = X_t[\mu_r(t,\widetilde{X}_t)dt + \sigma_r(t,\widetilde{X}_t)dW_t]$, where $\widetilde{X}_t$ is defined as the $(1+m)$-dimensional vector $(X_t;\boldsymbol{\chi}_t)$ and $\boldsymbol{\chi}$ is a state variable. Also, we suppose the integrability condition $\mathbb{E}[\int_0^T \sigma_r^2(t,\widetilde{X}_t)X_t^2 dt] < \infty$, and that there exists an equivalent martingale measure, $\mathbb{Q}$, which renders X_t a martingale (recall the risk free rate is assumed zero):[18]

$$dX_t = \sigma_r(t,\widetilde{X}_t)X_t dW_t^{\mathbb{Q}}.$$

[17] We notice that a similar result can be found in Derman (2008).

[18] Obviously, such an existence claim is not altogether innocuous. Indeed, the measure change is here further complicated by the fact that we have not made formal specifications for the dynamical form of the state variable $\boldsymbol{\chi}_t$. However, insofar as we adopt the standard dynamical assumption $d\boldsymbol{\chi}_t = \boldsymbol{m}(\boldsymbol{\chi}_t)dt + \boldsymbol{v}(\boldsymbol{\chi}_t)dW_t + \bar{\boldsymbol{v}}(\boldsymbol{\chi}_t)d\bar{\boldsymbol{W}}_t$, our existence claim is tantamount to positing the existence of a market price of risk vector $\boldsymbol{\theta} \in \mathbb{R}^m$ which renders $L(T) = L_X(T)L_{\boldsymbol{\chi}}(T)$ a true martingale, where

$$L_X(T) \equiv \exp\Big\{-\int_0^T \frac{\mu_r(t,\widetilde{X}_t)}{\sigma_r(t,\widetilde{X}_t)}dW_t - \tfrac{1}{2}\int_0^T \frac{\mu_r^2(t,\widetilde{X}_t)}{\sigma_r^2(t,\widetilde{X}_t)}dt\Big\},$$

and

$$L_{\boldsymbol{\chi}}(T) \equiv \exp\left\{-\int_0^T \boldsymbol{\theta}_t^{\intercal} d\bar{\boldsymbol{W}}_t - \tfrac{1}{2}\int_0^T |\boldsymbol{\theta}_t|^2 dt\right\}.$$

Now consider the admittedly somewhat contrived scenario of a Δ-hedged portfolio, long one unit of the call, for which σ_h and σ_i are both zero.[19] The associated value process is

$$\begin{aligned}\Pi_t^i &= C_t^i(T,K) + B_t - \partial_x C_t^h(T,K)\cdot X_t \\ &= (X_t - K)^+ + B_t - \mathbf{1}_{\{X_t>K\}} X_t,\end{aligned} \tag{3.11}$$

where $\mathbf{1}_{\{X_t>K\}}$ is the indicator function. The important point here is that $(X_t-K)^+$ may be reinterpreted as the terminal pay-off of a strike K maturity t call option (obviously, the specification $\sigma_h = \sigma_i = 0$ is paramount here). Substituting (3.11) into the infinitesimal form of the fundamental theorem,

$$d\Pi_t^i = \tfrac{1}{2}(\sigma_r^2(t,\widetilde{X}_t) - \sigma_i^2) X_t^2 \partial_{xx}^2 C_t^i(T,K)dt,$$

we find that

$$d((X_t-K)^+ + B_t - \mathbf{1}_{\{X_t>K\}}X_t) = \tfrac{1}{2}\sigma_r^2(t,\widetilde{X}_t)X_t^2\delta(X_t-K)dt, \tag{3.12}$$

where we once again have made use of $\sigma_i = 0$, alongside the fact that $\partial_x \mathbf{1}_{\{X_t>K\}}$ is the Dirac delta-function $\delta(X_t-K)$. Taking the risk neutral expectation of (3.12), conditional on $\mathscr{F}_0$, the left-hand side reduces to

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}}[LHS] &= \mathbb{E}^{\mathbb{Q}}[d(X_t-K)^+] + \mathbb{E}^{\mathbb{Q}}[dB_t - \mathbf{1}_{\{X_t>K\}}dX_t] \\ &= d\mathbb{E}^{\mathbb{Q}}[(X_t-K)^+] - \mathbb{E}^{\mathbb{Q}}[\mathbf{1}_{\{X_t>K\}}dX_t] \\ &= dC_0^r(t,K) - \mathbb{E}^{\mathbb{Q}}[\mathbb{E}^{\mathbb{Q}}[\mathbf{1}_{\{X_t>K\}}dX_t|\mathscr{F}_t]] \\ &= dC_0^r(t,K) - \mathbb{E}^{\mathbb{Q}}[\mathbf{1}_{\{X_t>K\}}\mathbb{E}^{\mathbb{Q}}[dX_t|\mathscr{F}_t]] \\ &= dC_0^r(t,K),\end{aligned} \tag{3.13}$$

where the second line uses $dB_t = 0$, whilst the third line uses the law of iterated expectations and the fact that $\mathbb{E}^{\mathbb{Q}}[(X_t-K)^+]$ is the time zero price of a strike K maturity t call option. Finally, the fourth line follows from the $\mathscr{F}_t$-measurability of $\mathbf{1}_{\{X_t>K\}}$, whilst the fifth line exploits the martingale property $\mathbb{E}^{\mathbb{Q}}[dX_t] = 0$.

As for the right-hand side, define the joint density

$$f^{\mathbb{Q}}_{\sigma_r^2,X_t}(\sigma^2,x)d\sigma^2 dx \equiv \mathbb{Q}(\{\sigma^2 \le \sigma_r^2 \le \sigma^2 + d\sigma^2\} \cap \{x \le X_t \le x + dx\}),$$

then

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}}[RHS] &= \tfrac{1}{2}\iint_{\mathbb{R}_+^2} \sigma^2 x^2 \delta(x-K) f^{\mathbb{Q}}_{\sigma_r^2,X_t}(\sigma^2,x)d\sigma^2 dx dt \\ &= \tfrac{1}{2}\iint_{\mathbb{R}_+^2} \sigma^2 x^2 \delta(x-K) f^{\mathbb{Q}}_{\sigma_r^2}(\sigma^2|X_t=x) f^{\mathbb{Q}}_{X_t}(x)d\sigma^2 dx dt\end{aligned}$$

[19] To be precise, the contrived part is the assumption that the call trades at zero volatility; less so that we hedge it at zero volatility. The latter corresponds to a so-called *stop-loss strategy*, see Carr and Jarrow (1990).

$$
\begin{aligned}
&= \tfrac{1}{2}\int_{\mathbb{R}_+} x^2\delta(x-K)f^{\mathbb{Q}}_{X_t}(x)\Big\{\int_{\mathbb{R}_+}\sigma^2 f^{\mathbb{Q}}_{\sigma_r^2}(\sigma^2|X_t=x)d\sigma^2\Big\}dxdt \qquad (3.14)\\
&\equiv \tfrac{1}{2}\int_{\mathbb{R}_+} x^2\delta(x-K)f^{\mathbb{Q}}_{X_t}(x)\mathbb{E}^{\mathbb{Q}}[\sigma_r^2(t,\widetilde{X}_t)|X_t=x]dxdt\\
&= \tfrac{1}{2}K^2 f^{\mathbb{Q}}_{X_t}(K)\mathbb{E}^{\mathbb{Q}}[\sigma_r^2(t,\widetilde{X}_t)|X_t=K]dt.
\end{aligned}
$$

Recalling that $\partial_K\mathbb{E}^{\mathbb{Q}}[(X_t-K)\mathbf{1}_{\{X_t>K\}}] = -\mathbb{E}^{\mathbb{Q}}[\mathbf{1}_{\{X_t>K\}}]$, and $-\partial_K\mathbb{E}^{\mathbb{Q}}[\mathbf{1}_{\{X_t>K\}}] = \mathbb{E}^{\mathbb{Q}}[\delta(X_t-K)]$ we arrive at the Breeden-Litzenberger formula

$$f^{\mathbb{Q}}_{X_t}(K) = \partial^2_{KK}C_0^r(t,K). \qquad (3.15)$$

Combining equations (3.13), (3.14), and (3.15) we thus have that

$$\frac{dC_0^r}{dt}(t,K) = \tfrac{1}{2}\partial^2_{KK}C_0^r(t,K)K^2\mathbb{E}^{\mathbb{Q}}[\sigma_r^2(t,\widetilde{X}_t)|X_t=K],$$

which using the change of notation[20] $t=T$ amounts to the celebrated Dupire-Gyöngy-Derman-Kani formula

$$\mathbb{E}^{\mathbb{Q}}[\sigma_r^2(T,\widetilde{X}_T)|X_T=K] = \frac{\partial_T C_0^r(T,K)}{\frac{1}{2}K^2\partial^2_{KK}C_0^r(T,K)}, \qquad (3.16)$$

- see Derman and Kani (1998), Dupire (1994), Gyöngy (1986). Conceptually, the important point here is that the righthand side, modulo some amount of interpolation,[21] is empirically measurable, whence (3.16) provides a way of calibrating the volatility surface to observed call option prices in the market. Specifically, the formula enables us to square a local diffusion model with the infamous skew/smile effect of implied volatilities across different values for the strike and time to maturity, whilst delicately sidestepping the deeper issue as to why this phenomenon prevails.

Remark 3.5. In Wittgensteinian terms we must "throw away the ladder" to arrive at this final conclusion, Wittgenstein (1922), Proposition 6.54. Hitherto, we have assumed that the real parameters (r) are fundamentally unobservable, whilst the implied parameters (i) are those we are exposed to in the market. Yet, no such distinction exists in the works of Dupire et al., whence the r superscript in (3.16) really ought to be dropped.

Remark 3.6. The above derivation is arguably unconventional and neither rigorous nor the quickest way to demonstrate (3.16). In fact, the entire point of setting $\sigma_i = 0$ is essentially to extract the Itō-(Tanaka) formula applied to $(X_t-K)^+$, from which Derman et al.'s derivation takes its starting point. We keep the derivation here, as it provides a curious glimpse into how two philosophically quite distinct theorems can be interconnected.

[20] We do this to emphasise that t is the *maturity* of the option (not its value at time t).

[21] Exactly how to do this extrapolation has turned out to be sufficiently non-trivial to spurn numerous papers and successive quant-of-the-year awards a-decade-and-a-half later, see Andreasen and Huge (2010) (pure local volatility), Guyon and Henry-Labordère (2012) (decorated stochastic volatility models).

3.3 The gospel of the jump

Following remark 3.2, it is worthwhile exploring how the fundamental theorem can be adapted to new terrain. For instance, it is well known that Brownian motion in itself does not adequately capture the sporadic discontinuities that emerge in stock price processes. Hence, it is opportune to scrutinise the effect of a jump diffusion process, which in turn will give rise to another valuable lesson on the profitability of imperfect hedging.

Already, it is a well-known fact that exact hedges generally do not exist in a jump economy where the true dynamics of the underlying is perfectly disseminated (see Merton (1976) or the expositions by Privault (2013) and Shreve (2004)). It is thus of some theoretical interest to see how this preexisting hedge error is further complicated under the model error framework of the fundamental theorem. Stepping-stones towards answering this question are found in Andreasen (2003) and Davis (2010) both of whom consider a mono-dimensional implied-hedge scenario with perfect information about the jump diffusion component. Our contribution is to generalise their results to a multi-dimensional framework with arbitrary (mis)-specifications for the volatility and jump distribution. For an overview of multi-dimensional jump-diffusion theory we refer the reader to the appendix.

Suppose the real dynamics of the underlying price process obeys

$$d\boldsymbol{X}_t = \mathbf{D}_{\boldsymbol{X}_t}[\boldsymbol{\mu}_r(t,\widetilde{\boldsymbol{X}}_t)dt + \boldsymbol{\sigma}_r(t,\widetilde{\boldsymbol{X}}_t)d\boldsymbol{W}_t] + \mathbf{D}_{\boldsymbol{X}_{t^-}}d\boldsymbol{Y}_t, \tag{3.17}$$

where $\{\boldsymbol{Y}_t\}_{t\geq 0}$ is an n-dimensional vector of independent compound Poisson processes. Specifically, the j^{th} component is given by

$$Y_t^j = \sum_{k=1}^{N_t^j} Z_k^j,$$

where $\{N_t^j\}_{t\geq 0}$ is an intensity-λ_j Poisson process, and $\{Z_k^j\}_{k\geq 1}$ is a sequence of relative jump-sizes, assumed to be i.i.d. square-integrable random variables with cumulative distribution function (cdf) $\nu_j : \mathbb{R} \mapsto [0,1]$. For shorthand, we shall refer to the vectors $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, ..., \lambda_n)^\intercal$ and $\boldsymbol{\nu} = (\nu_1, \nu_2, ..., \nu_n)^\intercal$ as the intensity and cdf of $\boldsymbol{Y}_t$.

Oblivious to the true nature of (3.17), we imagine that pricing and hedging should be performed (with obvious notation) under the tuple $\langle \boldsymbol{\phi}, \boldsymbol{\lambda}_h^{\mathbb{Q}}, \boldsymbol{\nu}_h^{\mathbb{Q}}, \boldsymbol{\sigma}_h(t,\boldsymbol{X}_t), \mathbb{Q}\rangle$, where $\mathbb{Q}$ is the risk neutral measure

$$\begin{aligned} d\mathbb{Q}_{\boldsymbol{\phi},\boldsymbol{\lambda}^{\mathbb{Q}},\boldsymbol{\nu}^{\mathbb{Q}}} = \exp&\left\{\int_0^T \boldsymbol{\phi}_s \bullet d\boldsymbol{W}_s - \tfrac{1}{2}\int_0^T |\boldsymbol{\phi}_s|^2 ds - \sum_{j=1}^n (\lambda_{h,j}^{\mathbb{Q}} - \lambda_{h,j})T\right\} \\ &\cdot \prod_{j=1}^n \prod_{k=1}^{N_t^j} \frac{\lambda_{h,j}^{\mathbb{Q}} d\nu_{h,j}^{\mathbb{Q}}(Z_k^j)}{\lambda_{h,j} d\nu_{h,j}(Z_k^j)} d\mathbb{P}^h, \end{aligned} \tag{3.18}$$

such that $\{\boldsymbol{\phi}_t\}_{t\geq 0}$ is a bounded adapted n-dimensional process (the so-called Girsanov kernel), and $\boldsymbol{\lambda}_h^{\mathbb{Q}}, \boldsymbol{\nu}_h^{\mathbb{Q}}$ respectively represent the jump intensity and jump-size distribution under $\mathbb{Q}$. Specifically, the price of an option with terminal pay-off $g(\boldsymbol{X}_T)$ is determined as

$$V_t^h = \mathbb{E}^{\mathbb{Q}}[e^{-\int_t^T r_u du} g(\boldsymbol{X}_T)|\mathscr{F}_t^{\boldsymbol{X}}],$$

with the underlying supposedly driven by

$$d\boldsymbol{X}_t = \mathbf{D}_{\boldsymbol{X}_t}[r_t\boldsymbol{\iota}dt + \boldsymbol{\sigma}_h(t,\boldsymbol{X}_t)d\boldsymbol{W}_t^{\mathbb{Q}}] + \mathbf{D}_{\boldsymbol{X}_{t-}}[d\boldsymbol{Y}_t - \boldsymbol{\lambda}_h^{\mathbb{Q}} \circ \mathbb{E}_{\boldsymbol{\nu}^{\mathbb{Q}}}[\boldsymbol{Z}_1]],$$

with $\boldsymbol{Z}_1 = (Z_1^1, Z_1^2, ..., Z_1^n)^\intercal$, and $\mathbb{Q}$ has been specified such that

$$\boldsymbol{\mu}_h(t,\boldsymbol{X}_t) + \boldsymbol{\lambda}_h^{\mathbb{Q}} \circ \mathbb{E}_{\boldsymbol{\nu}^{\mathbb{Q}}}[\boldsymbol{Z}_1] + \boldsymbol{\sigma}_h(t,\boldsymbol{X}_t)\boldsymbol{\phi}_t = r_t\boldsymbol{\iota}, \tag{3.19}$$

is satisfied almost everywhere.[22]

Remark 3.7. We emphasise that (3.18) is a risk neutral measure transformation of the *hedge* dynamics with the associated measure $\mathbb{P}^h$. This is to be contrasted with example 2 in Subsection 3.2.4 in which $\mathbb{Q}$ is the risk neutral measure of the *real* dynamics.

Theorem 3.2. The fundamental theorem of derivative trading with jumps. Let $V_t = V(t,\boldsymbol{X}_t) \in \mathcal{C}^{1,2}([0,T]\times\mathbb{R}^n)$ be the price process of a European option with terminal pay-off $V_T = g(\boldsymbol{X}_T)$, the underlying of which follows the real dynamics (3.17). We assume the option trades in the market at the (not necessarily uniquely determined) implied volatility $\boldsymbol{\sigma}_i = \boldsymbol{\sigma}_i(t,\boldsymbol{X}_t)$. Now suppose we at time $t=0$ acquire such an option for the implied price V_0^i and set out to Δ-hedge our position. Said hedge is performed under the notion $\langle \boldsymbol{\phi}, \boldsymbol{\lambda}_h^{\mathbb{Q}}, \boldsymbol{\nu}_h^{\mathbb{Q}}, \boldsymbol{\sigma}_h(t,\boldsymbol{X}_t), \mathbb{Q}\rangle$, leading to the fair price process V_t^h. Then the present value of the **profit-&-loss** we incur from holding such a portfolio over the interval $\mathbb{T} = [0,T]$ is

$$\begin{aligned}
\mathrm{P\&L}_{\mathbb{T}}^h = V_0^h - V_0^i &+ \tfrac{1}{2}\int_0^T e^{-\int_0^t r_u du}\ \mathrm{tr}[\mathbf{D}_{\boldsymbol{X}_t}\boldsymbol{\Sigma}_{rh}(t,\widetilde{\boldsymbol{X}}_t)\mathbf{D}_{\boldsymbol{X}_t}\nabla^2_{\boldsymbol{xx}}V_t^h]dt, \\
&+ \int_0^T \sum_{j=1}^n e^{-\int_0^t r_u du}\Big\{\Big(\Delta_j V_t^h(t,\boldsymbol{X}_{t-}) - X_{j,t-}Z_{N_t^j}\partial_{x_j}V_t^h\Big)dN_t^j \\
&- \lambda_{h,j}^{\mathbb{Q}}\Big(\mathbb{E}^{\mathbb{Q}}[\Delta_j V_t^h(t,\boldsymbol{x})]|_{\boldsymbol{x}=\boldsymbol{X}_{t-}} - X_{j,t-}\mathbb{E}^{\mathbb{Q}}[Z_1^i]\partial_{x_j}V_t^h\Big)dt\Big\},
\end{aligned} \tag{3.20}$$

where

$$\Delta_j V_t^h(t,\boldsymbol{X}_{t-}) \equiv V^h(t,\boldsymbol{X}_{t-}\circ(\boldsymbol{\iota} + \hat{\mathbf{e}}_j Z^j_{N_t^j})) - V^h(t,\boldsymbol{X}_{t-}),$$

represents the change in value of the option when the underlying jumps in the j^{th} component, and $[\hat{\mathbf{e}}_j]_k = \delta_{j,k}$ is a unit vector in $\mathbb{R}^n$.

[22] It should be clear the $\mathbb{Q}$ is not uniquely determined. In fact, for (3.19) to admit only one solution, we would require that either (i) $\boldsymbol{\lambda}_h = \boldsymbol{\lambda}_h^{\mathbb{Q}} = 0$ (there are *no* jumps), in which case we recover the standard Girsanov theorem with $\boldsymbol{\phi}_t = \boldsymbol{\sigma}_h^{-1}(r\boldsymbol{\iota} - \boldsymbol{\mu}_h)$, or (ii) when $\boldsymbol{\sigma}_h = 0$ (there are *only* jumps) and $\boldsymbol{\nu}^{\mathbb{Q}} = \boldsymbol{\nu} = \boldsymbol{\delta}_1$ (the simple Poisson process case) in which case $\boldsymbol{\lambda}_h^{\mathbb{Q}} = r_t\boldsymbol{\iota} - \boldsymbol{\mu}_h$.

Sketch of proof: The proof runs in parallel with that of theorem 1. Specifically, the analogue of expression (3.6) is

$$d\Pi_t^h = dV_t^i - \nabla_{\boldsymbol{x}} V_t^h \bullet (d\boldsymbol{X}_t^{\text{cont.}} - (r_t \boldsymbol{\iota} - \boldsymbol{q}_t) \circ \boldsymbol{X}_t dt) - \nabla_{\boldsymbol{x}} V_t^h \bullet d\boldsymbol{Y}_t - r_t V_t^i dt,$$

where $d\boldsymbol{X}_t^{\text{cont.}}$ is the continuous part of (3.17) i.e.

$$d\boldsymbol{X}_t^{\text{cont.}} = \mathbf{D}_{\boldsymbol{X}_t}[\boldsymbol{\mu}_r(t, \widetilde{\boldsymbol{X}}_t)dt + \boldsymbol{\sigma}_r(t, \widetilde{\boldsymbol{X}}_t)d\boldsymbol{W}_t].$$

Furthermore, in analogy with (3.7) and (3.8) we have the Itō formula

$$\begin{aligned} dV_t^h = &\{\partial_t V_t^h + \tfrac{1}{2}\text{tr}[\boldsymbol{\sigma}_r^\intercal(t, \widetilde{\boldsymbol{X}}_t)\mathbf{D}_{\boldsymbol{X}_t}\nabla^2_{\boldsymbol{xx}} V_t^h \mathbf{D}_{\boldsymbol{X}_t}\boldsymbol{\sigma}_r(t, \widetilde{\boldsymbol{X}}_t)]\}dt + \nabla_{\boldsymbol{x}} V_t^h \bullet d\boldsymbol{X}_t^{\text{cont.}}, \\ &+ \sum_{j=1}^n [V^h(t, \boldsymbol{X}_{t-} \circ (\boldsymbol{\iota} + \hat{\mathbf{e}}_j Z^j_{N_t^j})) - V^h(t, \boldsymbol{X}_{t-})] dN_t^j, \end{aligned}$$

and the partial integro-differential equation for pricing purposes

$$\begin{aligned} r_t V_t^h = &\partial_t V_t^h + \nabla_{\boldsymbol{x}} V_t^h \bullet ((r_t \boldsymbol{\iota} - \boldsymbol{q}_t) \circ \boldsymbol{X}_t) + \tfrac{1}{2}\text{tr}[\boldsymbol{\sigma}_h^\intercal(t, \boldsymbol{x})\mathbf{D}_{\boldsymbol{X}_t}\nabla^2_{\boldsymbol{xx}} V_t \mathbf{D}_{\boldsymbol{X}_t}\boldsymbol{\sigma}_h(t, \boldsymbol{X})] \\ &+ \textstyle\sum_{j=1}^n \lambda_{h,j}^{\mathbb{Q}} \mathbb{E}_{\boldsymbol{\nu}^{\mathbb{Q}}}[V(t, \boldsymbol{x} \circ (\boldsymbol{\iota} + \hat{\mathbf{e}}_j Z_1^j)) - V(t, \boldsymbol{x}) - x_j Z_1^j \partial_{x_j} V(t, \boldsymbol{x})]_{\boldsymbol{x}=\boldsymbol{X}_{t-}}. \end{aligned}$$

Combining these three expressions as above yields the desired result. □

Remark 3.8. The last two lines in (3.20), which we denote by $P\&L_J$, represent the present-valued profit-&-loss brought about by our inability to hedge the jump risk completely. A more compact way of writing this result is attained by considering the associated Poisson random measure $J_j(dt \times dz^j)$ with intensity measure $\mathbb{E}[J_j(dt \times dz^j)] = \lambda_j dt d\nu_j(z^j)$ for $j = 1, 2, ..., n$. Specifically, upon defining the pseudo-compensated random measure

$$\tilde{J}_{h,j}(dt \times dz^j) \equiv J_j(dt \times dz^j) - \lambda_{h,j}^{\mathbb{Q}} dt d\nu_{h,j}^{\mathbb{Q}}(z^j), \tag{3.21}$$

for $j = 1, 2, ..., n$, we see that the jump contribution to the profit-and-loss may be written as

$$P\&L_J = \int_0^T \int_{\mathbb{R}} \sum_{j=1}^n e^{-\int_0^t r_u du} \{\Delta_j V_t^h(t, \boldsymbol{X}_{t-}) - z^j X_{j,t-} \partial_{x_j} V_t^h\} \tilde{J}_{h,j}(dt \times dz_j), \tag{3.22}$$

where "pseudo" is used to emphasise that (3.21) is a real-world Poisson random measure compensated by a mis-specified intensity measure term: it is neither a martingale measure under $\mathbb{P}$ nor under the listed $\mathbb{Q}$. Only if we remove parameter uncertainty on the jump parameters in the sense $(\boldsymbol{\lambda}_h^{\mathbb{Q}}, \boldsymbol{\nu}_h^{\mathbb{Q}}) = (\boldsymbol{\lambda}^{\mathbb{Q}}, \boldsymbol{\nu}^{\mathbb{Q}})$ do we recover the $\mathbb{Q}$ martingale measure property. In particular, for the Merton specification $(\boldsymbol{\lambda}^{\mathbb{Q}}, \boldsymbol{\nu}^{\mathbb{Q}}) = (\boldsymbol{\lambda}, \boldsymbol{\nu})$ the compensated random measures become martingale measures under $\mathbb{P}$, which in turn implies $\mathbb{E}[P\&L_J] = 0$ insofar as the integrand is square integrable.

Conceptually, the takeaway message from the formulation (3.22) is that if V is convex in all of its components (a property it will inherit from the payoff function under mild

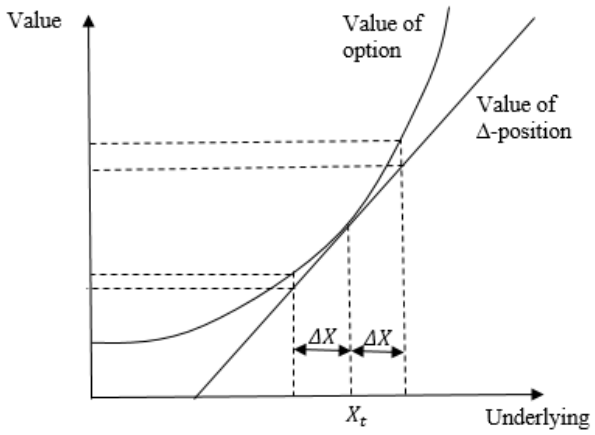

Figure 3.2: Suppose we Δ-hedge a long position in an option with a convex pricing function. Insofar as a jump in the underlying occurs, $X_t \mapsto X_t \pm \Delta X_t$, it follows that the value of the option will exceed the value of the Δ-position. Hence, our net *P&L* benefits from such an occurrence. Obviously, the converse will be true if we hold a short position in the option.

conditions) then $\forall j : \Delta V > \partial_{x_j} V \Delta X_j$ whence the integrand in $P\&L_J$ is positive. Thus, our hedge portfolio actually benefits from jumps in either direction of any of the underlying price process. Conversely, if we had shorted the option, the hedge profit would obviously take a hit in the event of a jump (in Talebian terms, holding a hedge portfolio with a short option position corresponds to "picking pennies in front of a steam roller").[23] A vivid illustration of this point is provided in Figure 3.2 for an option written on a single underlying.

3.4 Insights from empirics: on arbitrage and erraticism

Inspired by Wilmott's theoretical hedge experiment, we now look into the empirical performance of Δ-hedging strategies based on (I) forecasted implied volatilities and (II) forecasted actual (i.e. historical) volatilities. Specifically, we are interested in the properties of the accumulated P&L, insofar as we Δ-hedge, till expiry, a three-month call-option on the S&P500 index, initially purchased at-the-money. We investigate a totality of 36 such portfolios over disjoint intervals between July 2004 and July 2013. This involves market data on both the underlying index and on options. Daily data on the S&P500 index is readily and freely available. For option data, we combine a 2004-2009 data set from a

[23] Nassim (2007), p.19, "'*Most traders were just "picking pennies in front of a steam roller," exposing themselves to the high impact rare event yet sleeping like babies, unaware of it.*"

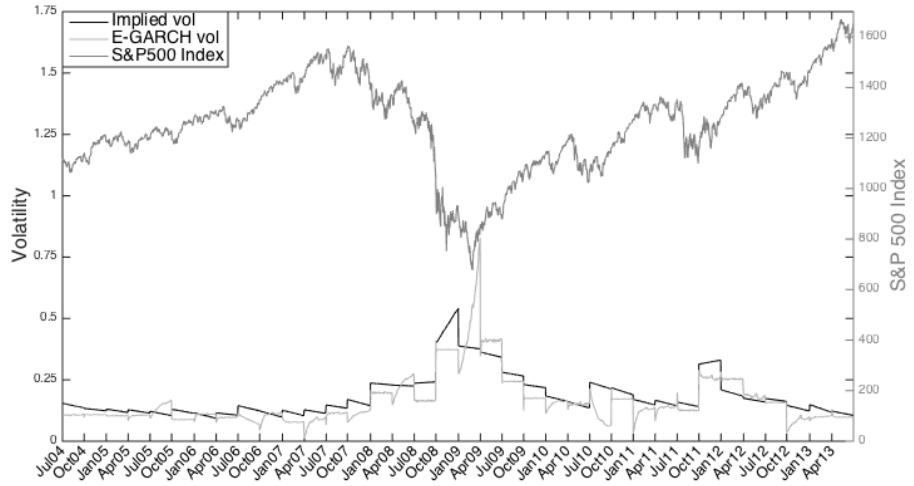

Figure 3.3: The top **grey curve** is the S&P500 Index plotted from July 2004 to July 2013 [units on right hand axis]. The tic-dates on the time axis have deliberately been chosen to match the purchasing dates $\{t_i\}_{i=1}^{36}$ of the 36 delta-hedged portfolios under investigation (each of which is of three months' duration). The **light grey curve** is the actual (stochastic) volatility estimated from a lognormal volatility model. Specifically, every time segment between purchasing dates $[t_i, t_{i+1})$ reflects a Monte Carlo simulated forecast based upon an EGARCH(1,1) fitted to market data from the previous time segment $[t_{i-1}, t_i)$. Finally, the **black curve** is the three-month ATM implied volatility. Specifically, every time segment between purchasing dates $[t_i, t_{i+1})$ is a static forecast based upon ATM implied volatility data from the purchasing date t_i. Both volatility curves have their units on the left hand axis.

major commercial bank[24] with more recent prices from OptionMetrics obtained via the Wharton Financial Database.

Whilst ATM call option prices straightforwardly are obtained from the data set, the (forecasted) implied and actual volatilities require a bit of manipulation. In case of the former, we define the daily implied volatility, over the life-time of the portfolio, as the ATM implied volatility of corresponding tenor obtained at the portfolio purchasing date (the resulting volatility process is illustrated by the black curve in Figure 3.3). In case of the latter, we require a suitable volatility model fitted to historical data in order to predict the "actual" volatility process. Specifically, we define the daily actual volatility, over the life-time of the portfolio, as the conditional expectation of a volatility model which has been fitted to market data from the previous portfolio period. In this context, we observe that models with lognormal volatility dynamics generally have more empirical support than, say, Heston's model.[25] The exponential general autoregressive conditional heteroskedasticity

[24]The bank shall remain nameless, but the data can be downloaded from `http://www.math.ku.dk/~rolf/Svend/`

[25]See Gatheral et al. (2014) and their references.

model (EGARCH(1,1)) has proven particularly felicitous in the context of S&P 500 forecasting[26] - a result we assume applies universally for each of the 36 portfolios investigated. Thus, we hold it to be the case that daily log returns, r_t, can be modelled as $r_t = \mu + \varepsilon_t$, where μ is the mean return, and ε_t has the interpretation of a hetereoskedastic error. In particular, ε_t is construed to be the product between a white noise process, $z_t \sim \mathcal{N}(0,1)$, and a daily standard deviation, σ_t, which obeys the relation

$$\log \sigma_t^2 = \alpha_0 + \alpha_1 \log \sigma_{t-1}^2 + \alpha_2 \left[\frac{|\varepsilon_{t-1}|}{\sigma_{t-1}} - \sqrt{\frac{2}{\pi}} \right] + \alpha_3 \frac{\varepsilon_{t-1}}{\sigma_{t-1}}, \tag{3.23}$$

where $\alpha_0, \alpha_1, \alpha_2$ and α_3 are constants. The resulting volatility process is illustrated by the light grey curve in Figure 3.3.

A few remarks on the estimated volatility processes are in order. First, we clearly see that volatility can change dramatically during the life-time of a portfolio. We also see that implied volatility typically is higher than actual volatility. This oft-reported result can be explained theoretically by the stochastic volatility having a market price of risk attached, see for instance Henderson et al. (2005). Finally, there is a clear negative correlation between stock returns and volatility during the financial turmoil which followed the Lehman default in September 2008. All in all, reality unsurprisingly turns out to be a bit more complicated than the set-up in Wilmott's experiment. Still and all, does its main messages carry over? To test this, we perform a hedge experiment with the following design:

- For any given portfolio, we compute the daily implied volatilities $\{\sigma_t^{\text{imp}}\}_{t=1}^{63}$ and the daily actual volatilities $\{\sigma_t^{\text{act}}\}_{t=1}^{63}$ as outlined above. We assume there are 63 trading days over a three months period (labelled by $t = 1, 2, ..., 63$) and let S_t, r_t and q_t denote the time t value of the index, interest rate and dividend yield.

- For each of the two hedging strategies $x \in \{\sigma^{\text{imp}}, \sigma^{\text{act}}\}$ we do the following: If $\sigma_1^{\text{act}} < \sigma_1^{\text{imp}}$ we short the call ($\gamma = -1$); otherwise, we go long the call ($\gamma = +1$) in accordance with the remark made in the section on Wilmott's hedge experiment. Then, we set up the delta neutral portfolio $\Pi_1 = B_1 - \gamma \Delta_1^{\text{BS}}(x_1) S_1 + \gamma C_1^{\text{BS}}(\sigma_1^{\text{imp}})$ s.t. $\Pi_1 = 0$, where $\Delta_1^{\text{BS}}(x_1)$ is the well-known Black-Scholes delta.

- For $t = 2, 3, ..., 63$ we do the following: compute the time t value of the portfolio set up the previous day: $\tilde{\Pi}_t = B_{t-1} e^{r_{t-1}\Delta t} - \gamma \Delta_t^{\text{BS}}(x_t) S_t e^{q_{t-1}\Delta t} + \gamma C_t^{\text{BS}}(\sigma_t^{\text{imp}})$. The quantity $dP\&L_t = \tilde{\Pi}_t - \Pi_{t-1}$ defines the profit-&-loss accrued over the interval $[t-1, t]$. Next, we rebalance the portfolio such that it, once again, is delta-neutral, $\Pi_t = B_t - \gamma \Delta_t^{\text{BS}}(x_t) S_t + \gamma C_t^{\text{BS}}(\sigma_t^{\text{imp}})$, where B_t is chosen in accordance with the self-financing condition: $\tilde{\Pi}_t = \Pi_t$.

- Finally, at the option expiry, we compute the terminal P&L, as well as its lifetime quadratic variation, $\sum_{t=1}^{63} |dP\&L_t|^2/63$.

[26] See Awartani and Corradi (2005).

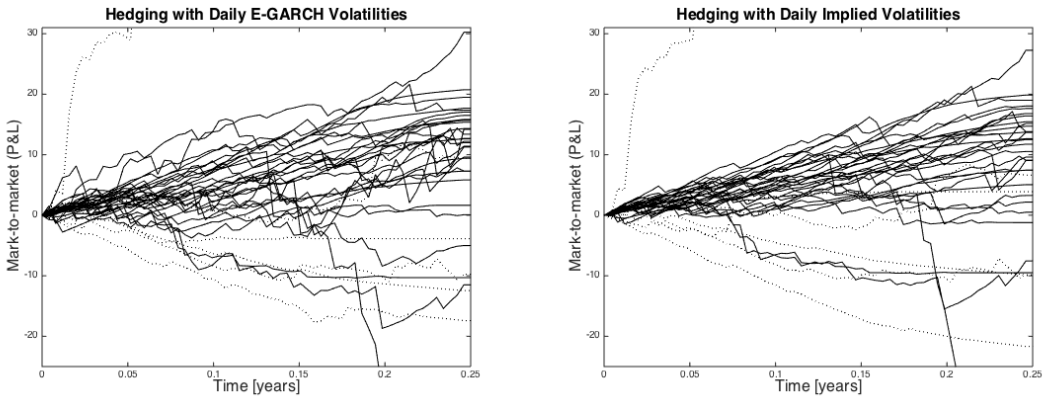

Figure 3.4: Panels (a) (actual) and (b) (implied) show the path-for-path hedge error behaviour for the 36 non-overlapping three-month hedges. Dotted paths correspond to cases where we initially take a long position in the option.

The 36 hedge error (or P&L) paths and the distributions of the quadratic variation of the two methods are shown in Figure 3.4. Table 3.5 reports descriptive statistics and statistical tests of various hypotheses.

Quantity	**Mean**	**SD.**	**Notes (Hypotheses Tests)**
Hedge error, actual volatility	7.7	17.3	The mean hedge error is statistically greater than zero (p-value = 1%) when we hedge with the actual vol. forecast.
Hedge error, implied volatility	7.7	15.6	The mean hedge error is statistically greater than zero (p-value = 1%) when we hedge with the implied vol. forecast. We cannot reject the hypothesis that the standard deviations $sd_{act} = sd_{imp}$ are equal (p-value = 55%).
Quadratic var., actual volatility	1.2	2.1	
Quadratic var., implied volatility	0.81	2.0	The mean quadratic variation of the hedge error when we hedge with the actual vol. forecast is statistically less than the quadratic variation when we hedge with the implied vol. forecast (p-value = 1.4%).

Table 3.5: Summary statistics and hypothesis tests for different hedge strategies.

First, we note (Figure 3.4) that even though implied volatility typically is above actual volatility, this far from creates arbitrage. Hedge errors for the two methods readily become negative. A primary explanation for this is the randomness of volatility. Our Δ-hedged strategy only makes us a profit if realised volatility ends up "on the right side" of initial implied volatility. And that we don't know for sure until after the hedging period is over; we have to base our decisions on forecasts; initial forecasts even, for the fundamental theorem to apply. Notice though that the averages for both hedge errors are significantly positive. This shows that there is a risk premium that can be picked up, most often by selling options and Δ-hedging them. Because the hedge is not perfect, this compensation is anticipated.

The question is, is it financially significant? In theory the hedged portfolio has an initial cost of zero, so it is not obvious how to define a rate of return, but the initial option price would seem a reasonable (possibly conservative) benchmark for the collateral that would need to be posted on a hedged short call option position. From column three in Table 3.6 the average option price is \$49.2. Comparing this to the means (~ 7.7; remember this is over a three-month horizon) and standard deviations (~ 15.5; ditto) of the hedge errors in Table 3.5 shows that the gains are also significant in economic terms. Put differently, the crude calculation $(4 \cdot \frac{7.7}{49.2} - 0.02)/(\sqrt{4} \cdot \frac{15.5}{49.2})$ gives annualised Sharpe-ratios around 1.

If we look just at the terminal hedge errors, then the difference in riskiness (as measured by standard deviation) between hedging with actual and hedging with implied volatility is in no way statistically significant (the p-value for equality of variances is 55%). Also, the correlation between the terminal hedge error from the two approaches is 0.97. However, if we consider the quadratic variations as the measure of riskiness, then the picture changes. The average quadratic variation of the implied hedge error (0.81) is only two-thirds of the average quadratic variation of the actual hedge error (1.2) (a paired t-test for equality yields a p-value of 1.4%).

All in all this shows that volatility arbitrage is difficult, but the following insight from Wilmott's experiment stands: if you are in the business of hedging, then the use of implied volatility should make you sleep better at night.

3.5 Conclusion

In the world of finance, no issue is more pressing than that of hedging our risks, yet remarkably little attention has been paid to the risk brought about by the possibility that our models might be wrong. To remedy this deplorable situation, we have in this paper derived a meta-theorem that quantifies the P&L of a Δ-hedged portfolio with an erroneous volatility specification. *Meta-* to the extent that one of the constituent parameters (the real volatility) is transcendental; yet, also a theorem with some very concrete "real world" corollaries. For instance, a specific case was investigated in which the implied volatility gives rise to smooth (i.e. $\mathcal{O}(dt)$) P&L-paths, whilst any other hedge volatility yields erratic (i.e. $\mathcal{O}(dW_t)$) P&L paths. In a somewhat quirkier context, the Dupire-Gyöngy-Derman-Kani formula for volatility surface calibration was shown to be a corollary.

Whilst the theorem proved in Section 3.2 is more general than the versions typically found in the literature, it does not go *far enough*. Extensive empirical support has been added to the case of discontinuities in the stock price process: thus, in the gospel of the jump we extended the fundamental theorem to include compound Poisson processes, which came with the revelation that jumps unambiguously hurt you when you try to hedge short put and call option positions.

One of the most conspicuous implications of the fundamental theorem is undoubtedly the apparent ease with which arbitrage can be made: e.g. in the constant parameter framework of Wilmott's experiment, a free lunch is guaranteed insofar as we can establish $\max\{\sigma_r, \sigma_i\}$

(in case of the former, we go long on the option - otherwise, we short it). Studying this strategy empirically, we find that the mean P&L indeed is in the positive; nonetheless, qua a significant dispersion the profit readily turns negative: the statistical arbitrage accordingly relies on us being willing to take so some significant hits along the way. Indeed, this is without even factoring in the non-negligible role of transaction costs. On the other hand, there is strong evidence that hedging at the implied volatility does yield smoother P&L paths.

One final remark: This paper can be seen as an exhaustive exposition of which volatility to use when delta hedging. To keep the length manageable and the presentation self-contained we have ignored an aspect that is of both theoretical and practical importance. It can be posed thus: which delta should I use? In models where the underlying and the volatility are correlated a strong case can be made for using the so-called risk-minimizing delta, which is in broad terms is the usual delta plus the (underlying, volatility)-correlation times the volatility of volatility times the Vega of the target option, see for instance Poulsen et al. (2009), Andreasen (2013), or Hull and White (2016). We leave the connection of this theory to the fundamental theorem of derivative trading, theoretically as well empirically, to future research.

Appendix A: Multi-dimensional jumps

In this section we establish Girsanov's Theorem and a pricing PDE for multi-dimensional jump-diffusion models. The equivalent results for 1-dimensional models are ubiquitous - see for instance Cont and Tankov (2004), Privault (2013) or Runggaldier (2003).

The Radon-Nikodym derivative

Theorem 3.3. A generalised Girsanov theorem for jump-diffusion processes. Let $\{\boldsymbol{W}_t\}_{t\in[0,T]}$ be a d_w-dimensional vector of independent Wiener processes on the filtered probability space $(\Omega, \mathscr{F}, \mathbb{P}, \{\mathscr{F}_t\}_{t\in[0,T]})$. One the same space, let $\{\boldsymbol{Y}_t\}_{t\in[0,T]}$ be a d_y-dimensional vector of independent compound Poisson processes, the i^{th} component of which is

$$Y_t^i = \sum_{k=1}^{N_t^i} Z_k^i,$$

where $N_t^i \sim \text{Pois}(\lambda_i t)$, $\lambda_i > 0$ is an intensity parameter, and $\{Z_k^i\}_{k\in\mathbb{N}}$ is a sequence of i.i.d. random variables with jump distribution $d\nu_i(z)$. Finally, let $\{\boldsymbol{\phi}_t\}_{t\in[0,T]}$ be a d_w-dimensional Girsanov kernel (some bounded, adapted process), then the processes

$$\left\{\boldsymbol{W}_t^{\mathbb{Q}} := \boldsymbol{W}_t - \int_0^t \boldsymbol{\phi}_s ds\right\}_{t\in[0,T]}, \quad \text{and} \quad \left\{\tilde{\boldsymbol{Y}}_t^{\mathbb{Q}} := \boldsymbol{Y}_t - \boldsymbol{\lambda}^{\mathbb{Q}} \circ \mathbb{E}_{\boldsymbol{\nu}^{\mathbb{Q}}}[\boldsymbol{Z}_1]t\right\}_{t\in[0,T]}$$

are martingales under the probability measure $\mathbb{Q}$ defined as

$$d\mathbb{Q}_{\boldsymbol{\phi},\boldsymbol{\lambda}^{\mathbb{Q}},\boldsymbol{\nu}^{\mathbb{Q}}} = \mathscr{E}(\boldsymbol{\phi}\star\boldsymbol{W})(T)\cdot e^{-\sum_{i=1}^{d_y}(\lambda_i^{\mathbb{Q}}-\lambda_i)T}\prod_{i=1}^{d_y}\prod_{k=1}^{N_t^i}\frac{\lambda_i^{\mathbb{Q}} d\nu_i^{\mathbb{Q}}(Z_k^i)}{\lambda_i d\nu_i(Z_k^i)} d\mathbb{P}$$

where $\mathscr{E}(\boldsymbol{\phi}\star\boldsymbol{W})(T) = e^{\int_0^T \boldsymbol{\phi}_s \bullet d\boldsymbol{W}_s - \frac{1}{2}\int_0^T |\boldsymbol{\phi}_s|^2 ds}$ is the Doleans exponential with respect to $\boldsymbol{W}_t$, and we have defined $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, ..., \lambda_{d_y})^\intercal$, and $\boldsymbol{\nu} = (\nu_1, \nu_2, ..., \nu_{d_y})^\intercal$.

Proof. The diffusion part is well known from Girsanov's theorem and will not be treated here. Instead we will show that for any bounded measurable function $f : \mathbb{R}^{d_y} \mapsto \mathbb{R}$, the following equivalence obtains

$$\mathbb{E}_{\boldsymbol{\lambda}^{\mathbb{Q}},\boldsymbol{\nu}^{\mathbb{Q}}}[f(\boldsymbol{Y}_T)] = \mathbb{E}_{\boldsymbol{\lambda},\boldsymbol{\nu}}\left[f(\boldsymbol{Y}_T)\tfrac{d\mathbb{Q}}{d\mathbb{P}}\right],$$

for the defined measure $\mathbb{Q}$. To this end, define $\boldsymbol{Y}_t^h$ as the vector $\boldsymbol{Y}_t$ with the upper limit of the summation, N_t^i, replaced by some fixed number $h^i \in \mathbb{N}_0$ for all i. Then the RHS can be written as

$$e^{-\sum_{i=1}^{d_y}(\lambda_i^{\mathbb{Q}}-\lambda_i)T}\mathbb{E}_{\boldsymbol{\lambda},\boldsymbol{\nu}}\left[f(\boldsymbol{Y}_t)\prod_{i=1}^{d_y}\prod_{k=1}^{N_t^i}\frac{\lambda_i^{\mathbb{Q}} d\nu_i^{\mathbb{Q}}(Z_k^i)}{\lambda_i d\nu_i(Z_k^i)}\right]$$

$$
\begin{aligned}
&= e^{-\sum_{i=1}^{d_y}(\lambda_i^{\mathbb{Q}}-\lambda_i)T} \sum_{h^1=0}^{\infty} \cdots \sum_{h^{d_y}=0}^{\infty} \mathbb{P}\left(\bigcap_{i=1}^{d_y}\{N_T^i = h^i\}\right) \cdot \\
&\quad \mathbb{E}_{\boldsymbol{\lambda},\boldsymbol{\nu}}\left[f(\boldsymbol{Y}_t)\prod_{i=1}^{d_y}\prod_{k=1}^{N_t^i}\frac{\lambda_i^{\mathbb{Q}} d\nu_i^{\mathbb{Q}}(Z_k^i)}{\lambda_i d\nu_i(Z_k^i)}\middle|\bigcap_{i=1}^{d_y}\{N_T^i = h^i\}\right] \\
&= e^{-\sum_{i=1}^{d_y}(\lambda_i^{\mathbb{Q}}-\lambda_i)T} \sum_{h^1=0}^{\infty} \cdots \sum_{h^{d_y}=0}^{\infty} \prod_{i=1}^{d_y} \mathbb{P}\left(N_T^i = h^i\right) \mathbb{E}_{\boldsymbol{\lambda},\boldsymbol{\nu}}\left[f(\boldsymbol{Y}_t^h)\prod_{i=1}^{d_y}\prod_{k=1}^{h^i}\frac{\lambda_i^{\mathbb{Q}} d\nu_i^{\mathbb{Q}}(Z_k^i)}{\lambda_i d\nu_i(Z_k^i)}\right] \\
&= e^{-\sum_{i=1}^{d_y}(\lambda_i^{\mathbb{Q}}-\lambda_i)T} \sum_{h^1=0}^{\infty} \cdots \sum_{h^{d_y}=0}^{\infty} \prod_{i=1}^{d_y} \frac{e^{-\lambda_i T}(\lambda_i T)^{h_i}}{k^i!} \mathbb{E}_{\boldsymbol{\lambda},\boldsymbol{\nu}}\left[f(\boldsymbol{Y}_t^h)\prod_{i=1}^{d_y}\prod_{k=1}^{h^i}\frac{\lambda_i^{\mathbb{Q}} d\nu_i^{\mathbb{Q}}(Z_k^i)}{\lambda_i d\nu_i(Z_k^i)}\right] \\
&= \sum_{h^1=0}^{\infty} \cdots \sum_{h^n=0}^{\infty} \prod_{i=1}^{d_y} \frac{e^{-\lambda_i^{\mathbb{Q}} T}(\lambda_i^{\mathbb{Q}} T)^{h_i}}{k^i!} \mathbb{E}_{\boldsymbol{\lambda},\boldsymbol{\nu}}\left[f(\boldsymbol{Y}_t^h)\prod_{i=1}^{d_y}\prod_{k=1}^{h^i}\frac{d\nu_i^{\mathbb{Q}}(Z_k^i)}{d\nu_i(Z_k^i)}\right] \\
&= \sum_{h^1=0}^{\infty} \cdots \sum_{h^{d_y}=0}^{\infty} \prod_{i=1}^{d_y} \frac{e^{-\lambda_i^{\mathbb{Q}} T}(\lambda_i^{\mathbb{Q}} T)^{h_i}}{k^i!} \int_{\mathbb{R}^{h^1}} \cdots \int_{\mathbb{R}^{h^{d_y}}} f(\boldsymbol{y}_t^h)\prod_{i=1}^{d_y}\prod_{k=1}^{h^i}\frac{d\nu_i^{\mathbb{Q}}(z_k^i)}{d\nu_i(z_k^i)} d\nu_i(z_k^i) \\
&= \sum_{h^1=0}^{\infty} \cdots \sum_{h^{d_y}=0}^{\infty} \prod_{i=1}^{d_y} \frac{e^{-\lambda_i^{\mathbb{Q}} T}(\lambda_i^{\mathbb{Q}} T)^{h_i}}{k^i!} \int_{\mathbb{R}^{h^1}} \cdots \int_{\mathbb{R}^{h^{d_y}}} f(\boldsymbol{y}_t^h)\prod_{i=1}^{d_y}\prod_{k=1}^{h^i} d\nu_i^{\mathbb{Q}}(z_k^i) \\
&= \sum_{h^1=0}^{\infty} \cdots \sum_{h^n=0}^{\infty} \prod_{i=1}^{d_y} \mathbb{Q}\left(N_T^i = h^i\right) \mathbb{E}_{\boldsymbol{\lambda}^{\mathbb{Q}},\boldsymbol{\nu}^{\mathbb{Q}}}\left[f(\boldsymbol{Y}_t^h)\right] \\
&= \sum_{h^1=0}^{\infty} \cdots \sum_{h^{d_y}=0}^{\infty} \mathbb{Q}\left(\bigcap_{i=1}^{d_y}\{N_T^i = h^i\}\right) \mathbb{E}_{\boldsymbol{\lambda}^{\mathbb{Q}},\boldsymbol{\nu}^{\mathbb{Q}}}\left[f(\boldsymbol{Y}_t)\middle|\bigcap_{i=1}^{d_y}\{N_T^i = h^i\}\right]
\end{aligned}
$$

which by the law of total probability corresponds to the LHS, $\mathbb{E}_{\boldsymbol{\lambda}^{\mathbb{Q}},\boldsymbol{\nu}^{\mathbb{Q}}}[f(\boldsymbol{Y}_T)]$ as desired. To show independence of increments under $\mathbb{Q}$, let $\xi_s = d\mathbb{Q}/d\mathbb{P}(s)$, and let f and g be two bounded measurable functions. Suppose $s < t \leq T$ then

$$
\begin{aligned}
\mathbb{E}_{\boldsymbol{\lambda}^{\mathbb{Q}},\boldsymbol{\nu}^{\mathbb{Q}}}[f(\boldsymbol{Y}_s)g(\boldsymbol{Y}_t - \boldsymbol{Y}_s)] &= \mathbb{E}_{\boldsymbol{\lambda},\boldsymbol{\nu}}[f(\boldsymbol{Y}_s)g(\boldsymbol{Y}_t - \boldsymbol{Y}_s)\xi_t] \\
&= \mathbb{E}_{\boldsymbol{\lambda},\boldsymbol{\nu}}[f(\boldsymbol{Y}_s)\xi_s]\mathbb{E}_{\boldsymbol{\lambda},\boldsymbol{\nu}}[g(\boldsymbol{Y}_t - \boldsymbol{Y}_s)\xi_t/\xi_s] \\
&= \mathbb{E}_{\boldsymbol{\lambda},\boldsymbol{\nu}}[f(\boldsymbol{Y}_s)\xi_s]\mathbb{E}_{\boldsymbol{\lambda},\boldsymbol{\nu}}[g(\boldsymbol{Y}_t - \boldsymbol{Y}_s)\xi_t] \\
&= \mathbb{E}_{\boldsymbol{\lambda}^{\mathbb{Q}},\boldsymbol{\nu}^{\mathbb{Q}}}[f(\boldsymbol{Y}_s)] \cdot \mathbb{E}_{\boldsymbol{\lambda}^{\mathbb{Q}},\boldsymbol{\nu}^{\mathbb{Q}}}[g(\boldsymbol{Y}_t - \boldsymbol{Y}_s)].
\end{aligned}
$$

□

Measure changes in jump-diffusion dynamics

To appreciate the gravity of this result, consider the jump-diffusion dynamics of the n-dimensional stock price process

$$d\boldsymbol{X}_t = \mathbf{D}_{\boldsymbol{X}_{t-}}[\boldsymbol{\mu}(t,\boldsymbol{X}_t)dt + \boldsymbol{\sigma}(t,\boldsymbol{X}_t)d\boldsymbol{W}_t] + \mathbf{D}_{\boldsymbol{X}_t}\boldsymbol{\theta}(t,\boldsymbol{X}_{t-})d\boldsymbol{Y}_t \tag{3.24}$$

where $\mathbf{D}_{\boldsymbol{X}_{t-}} = \text{diag}(X_{1,t-}, X_{2,t-}, ..., X_{n,t-})$, $\boldsymbol{\mu} : [0,\infty) \times \mathbb{R}^n \mapsto \mathbb{R}^n$, $\boldsymbol{\sigma} : [0,\infty) \times \mathbb{R}^n \mapsto \mathbb{R}^{n\times d_w}$ and $\boldsymbol{\theta} : [0,\infty) \times \mathbb{R}^n \mapsto \mathbb{R}^{n\times d_y}$. For the purposes of no-arbitrage pricing we want to determine the measure $\mathbb{Q}$ such that the discounted process

$$\boldsymbol{X}_t^* := e^{-\int_0^t r_u du}\boldsymbol{X}_t,$$

is a martingale. From Ito's lemma, it can readily be deduced that

$$\begin{aligned} d\boldsymbol{X}_t^* = \mathbf{D}^*_{\boldsymbol{X}_t}[(\boldsymbol{\mu}(t,\boldsymbol{X}_t) - r_t\boldsymbol{\iota} + \boldsymbol{\theta}(t,\boldsymbol{X}_t)\boldsymbol{\lambda}\circ\mathbb{E}_{\boldsymbol{\nu}}[\boldsymbol{Z}_1])dt + \boldsymbol{\sigma}(t,\boldsymbol{X}_t)d\boldsymbol{W}_t] \\ + \mathbf{D}^*_{\boldsymbol{X}_{t-}}\boldsymbol{\theta}(t,\boldsymbol{X}_{t-})(d\boldsymbol{Y}_t - \boldsymbol{\lambda}\circ\mathbb{E}_{\boldsymbol{\nu}}[\boldsymbol{Z}_1]dt), \end{aligned}$$

where we have added and subtracted $\boldsymbol{\theta}(t,\boldsymbol{X}_t)\boldsymbol{\lambda}\circ\mathbb{E}_{\boldsymbol{\nu}}[\boldsymbol{Z}_1]dt$, where $\boldsymbol{Z}_1 := (Z_1^1, Z_1^2, ..., Z_1^{d_y})^\intercal$. Using the measure transformation in the theorem above this transforms to

$$\begin{aligned} d\boldsymbol{X}_t^* = \mathbf{D}^*_{\boldsymbol{X}_t}[(\boldsymbol{\mu}(t,\boldsymbol{X}_t) - r_t\boldsymbol{\iota} + \boldsymbol{\theta}(t,\boldsymbol{X}_t)\boldsymbol{\lambda}^{\mathbb{Q}}\circ\mathbb{E}_{\boldsymbol{\nu}^{\mathbb{Q}}}[\boldsymbol{Z}_1] + \boldsymbol{\sigma}(t,\boldsymbol{X}_t)\boldsymbol{\phi}_t)dt + \boldsymbol{\sigma}(t,\boldsymbol{X}_t)d\boldsymbol{W}_t^{\mathbb{Q}}] \\ + \mathbf{D}^*_{\boldsymbol{X}_{t-}}\boldsymbol{\theta}(t,\boldsymbol{X}_{t-})d\tilde{\boldsymbol{Y}}_t^{\mathbb{Q}}. \end{aligned}$$

Hence, $\boldsymbol{X}_t^*$ is a $\mathbb{Q}$-martingale iff the tuple $\langle\boldsymbol{\phi}_t, \boldsymbol{\lambda}^{\mathbb{Q}}, \boldsymbol{\nu}^{\mathbb{Q}}, \mathbb{Q}\rangle$ is chosen such that

$$\boldsymbol{\mu}(t,\boldsymbol{X}_t) + \boldsymbol{\theta}(t,\boldsymbol{X}_t)\boldsymbol{\lambda}^{\mathbb{Q}}\circ\mathbb{E}_{\boldsymbol{\nu}^{\mathbb{Q}}}[\boldsymbol{Z}_1] + \boldsymbol{\sigma}(t,\boldsymbol{X}_t)\boldsymbol{\phi}_t = r_t\boldsymbol{\iota}, \tag{3.25}$$

almost everywhere. Needless to say, the infinite number of tuples which satisfies the no-arbitrage condition (3.25) ruins our chances of establishing *unique* prices for financial derivatives depending on the underlying jump diffusion dynamics. This is obvious by recalling that the discounted price process of V also should be a martingale, whence:

$$V(t,\boldsymbol{X}_t) = \mathbb{E}_{t,\boldsymbol{\lambda}^{\mathbb{Q}},\boldsymbol{\nu}^{\mathbb{Q}}}\left[e^{-\int_t^T r_u du}g(\boldsymbol{X}_T)\right], \tag{3.26}$$

where $g(\boldsymbol{X}_T)$ is the terminal pay-off.

PIDE methods

Theorem 3.4. The partial integro-differential pricing equation (PIDE) Consider a jump diffusion dynamics of the form (3.24), and let $V_t = V(t,\boldsymbol{X}_t)$ be a derivative the value of which is contingent upon it. Let $\langle\boldsymbol{\phi}_t, \boldsymbol{\lambda}^{\mathbb{Q}}, \boldsymbol{\nu}^{\mathbb{Q}}, \mathbb{Q}\rangle$ be a tuple such that the no-arbitrage condition (3.25) is satisfied. Then

$$d\boldsymbol{X}_t = \mathbf{D}_{\boldsymbol{X}_t}[r_t\boldsymbol{\iota}dt + \boldsymbol{\sigma}(t,\boldsymbol{X}_t)d\boldsymbol{W}_t^{\mathbb{Q}}] + \mathbf{D}_{\boldsymbol{X}_{t-}}\boldsymbol{\theta}(t,\boldsymbol{X}_{t-})[d\boldsymbol{Y}_t - \boldsymbol{\lambda}^{\mathbb{Q}}\circ\mathbb{E}_{\boldsymbol{\nu}^{\mathbb{Q}}}[\boldsymbol{Z}_1]dt]],$$

and $V_t = V(t, \boldsymbol{x})$ satisfies the PIDE

$$\begin{aligned} r_t V_t = \partial_t V_t + r_t \boldsymbol{x} \bullet \nabla_{\boldsymbol{x}} V_t + \tfrac{1}{2}\mathrm{tr}[\boldsymbol{\sigma}^\intercal(t,\boldsymbol{x})\mathbf{D}_{\boldsymbol{x}}\nabla^2_{\boldsymbol{xx}}V_t\mathbf{D}_{\boldsymbol{x}}\boldsymbol{\sigma}(t,\boldsymbol{x})] \\ + \mathbb{E}_{\boldsymbol{\nu}^{\mathbb{Q}}}[\textstyle\sum_{i=1}^{d_y}\lambda_i^{\mathbb{Q}}\{V(t,\boldsymbol{x}\circ(\boldsymbol{\iota}+\boldsymbol{\theta}_{:,i}(t,\boldsymbol{x})Z_1^i)) - V(t,\boldsymbol{x})\} \\ - (\mathbf{D}_{\boldsymbol{x}}\boldsymbol{\theta}(t,\boldsymbol{x})\boldsymbol{\lambda}^{\mathbb{Q}}\circ\boldsymbol{Z}_1)\bullet\nabla_{\boldsymbol{x}}V(t,\boldsymbol{x})], \end{aligned}$$

where $\nabla_{\boldsymbol{x}}$ is the gradient operator, $\nabla^2_{\boldsymbol{xx}}$ is the Hessian operator, $\boldsymbol{\iota} := (1,1,...,1)^\intercal \in \mathbb{R}^n$, and $\boldsymbol{\theta}_{:,i}(t,\boldsymbol{x})$ denotes the i^{th} column of the matrix $\boldsymbol{\theta}(t,\boldsymbol{x})$. Particularly, when $n = d_y$ and $\boldsymbol{\theta} = \mathbb{I}$ is the identity matrix then

$$\begin{aligned} r_t V_t = \partial_t V_t + r_t \boldsymbol{x} \bullet \nabla_{\boldsymbol{x}} V_t + \tfrac{1}{2}\mathrm{tr}[\boldsymbol{\sigma}^\intercal(t,\boldsymbol{x})\mathbf{D}_{\boldsymbol{x}}\nabla^2_{\boldsymbol{xx}}V_t\mathbf{D}_{\boldsymbol{x}}\boldsymbol{\sigma}(t,\boldsymbol{x})] \\ + \textstyle\sum_{i=1}^{n}\lambda_i^{\mathbb{Q}}\mathbb{E}_{\boldsymbol{\nu}^{\mathbb{Q}}}[V(t,\boldsymbol{x}\circ(\boldsymbol{\iota}+\hat{\mathbf{e}}_i Z_1^i)) - V(t,\boldsymbol{x}) - x_i Z_1^i \partial_{x_i}V(t,\boldsymbol{x})], \end{aligned}$$

where $\hat{\mathbf{e}}_i$ is a unit vector in the i^{th} direction.

Proof. Suppose a jump occurs in the i^{th} component of the compound Poisson process $\boldsymbol{Y}_t$: $Y_t^i = Y_{t-}^i + Z_t^i$. From the governing dynamics (3.24), this means that the stock price process jumps by $\Delta\boldsymbol{X}_t = \boldsymbol{X}_{t-} \circ (\boldsymbol{\iota} + \boldsymbol{\theta}_{:,i}(t,\boldsymbol{X}_{t-})Z_t^i)$. Defining the continuous SDE

$$d\boldsymbol{X}_t^{\text{cont.}} := \mathbf{D}_{\boldsymbol{X}_t}[r_t\boldsymbol{\iota}dt + \boldsymbol{\sigma}(t,\boldsymbol{X}_t)d\boldsymbol{W}_t^{\mathbb{Q}}] - \mathbf{D}_{\boldsymbol{X}_{t-}}\boldsymbol{\theta}(t,\boldsymbol{X}_{t-})\boldsymbol{\lambda}^{\mathbb{Q}}\circ\mathbb{E}_{\boldsymbol{\nu}^{\mathbb{Q}}}[\boldsymbol{Z}_1]dt],$$

we find by Ito's lemma,

$$\begin{aligned} dV(t,\boldsymbol{X}_t) &= \partial_t V(t,\boldsymbol{X}_t)dt + \nabla_{\boldsymbol{x}}V(t,\boldsymbol{X}_t)\bullet d\boldsymbol{X}_t^{\text{cont.}} \\ &\quad + \tfrac{1}{2}\mathrm{tr}[\boldsymbol{\sigma}^\intercal(t,\boldsymbol{X}_t)\mathbf{D}_{\boldsymbol{X}_t}\nabla^2_{\boldsymbol{xx}}V(t,\boldsymbol{X}_t)\mathbf{D}_{\boldsymbol{X}_t}\boldsymbol{\sigma}(t,\boldsymbol{X}_t)]dt \\ &\quad + \textstyle\sum_{i=1}^{d_y}(V(t,\boldsymbol{X}_{t-}+\Delta\boldsymbol{X}_t) - V(t,\boldsymbol{X}_{t-}))dN_t^i \\ &= \partial_t V(t,\boldsymbol{X}_t)dt + \nabla_{\boldsymbol{x}}V(t,\boldsymbol{X}_t)\bullet(\mathbf{D}_{\boldsymbol{X}_t}r_t\boldsymbol{\iota})dt \\ &\quad + \nabla_{\boldsymbol{x}}V(t,\boldsymbol{X}_t)\bullet(\mathbf{D}_{\boldsymbol{X}_t}\boldsymbol{\sigma}(t,\boldsymbol{X}_t)d\boldsymbol{W}_t^{\mathbb{Q}}) \\ &\quad - \nabla_{\boldsymbol{x}}V(t,\boldsymbol{X}_t)\bullet(\mathbf{D}_{\boldsymbol{X}_{t-}}\boldsymbol{\theta}(t,\boldsymbol{X}_{t-})\boldsymbol{\lambda}^{\mathbb{Q}}\circ\mathbb{E}_{\boldsymbol{\nu}^{\mathbb{Q}}}[\boldsymbol{Z}_1])dt \\ &\quad + \tfrac{1}{2}\mathrm{tr}[\boldsymbol{\sigma}^\intercal(t,\boldsymbol{X}_t)\mathbf{D}_{\boldsymbol{X}_t}\nabla^2_{\boldsymbol{xx}}V(t,\boldsymbol{X}_t)\mathbf{D}_{\boldsymbol{X}_t}\boldsymbol{\sigma}(t,\boldsymbol{X}_t)]dt \\ &\quad + \textstyle\sum_{i=1}^{d_y}(V(t,\boldsymbol{X}_{t-}\circ(\boldsymbol{\iota}+\boldsymbol{\theta}_{:,i}(t,\boldsymbol{X}_{t-})Z_{N_t^i}^i)) - V(t,\boldsymbol{X}_{t-}))dN_t^i \\ &= \partial_t V(t,\boldsymbol{X}_t)dt + \nabla_{\boldsymbol{x}}V(t,\boldsymbol{X}_t)\bullet(\mathbf{D}_{\boldsymbol{X}_t}r_t\boldsymbol{\iota})dt \\ &\quad - \nabla_{\boldsymbol{x}}V(t,\boldsymbol{X}_t)\bullet(\mathbf{D}_{\boldsymbol{X}_{t-}}\boldsymbol{\theta}(t,\boldsymbol{X}_{t-})\boldsymbol{\lambda}^{\mathbb{Q}}\circ\mathbb{E}_{\boldsymbol{\nu}^{\mathbb{Q}}}[\boldsymbol{Z}_1])dt \\ &\quad + \tfrac{1}{2}\mathrm{tr}[\boldsymbol{\sigma}^\intercal(t,\boldsymbol{X}_t)\mathbf{D}_{\boldsymbol{X}_t}\nabla^2_{\boldsymbol{xx}}V(t,\boldsymbol{X}_t)\mathbf{D}_{\boldsymbol{X}_t}\boldsymbol{\sigma}(t,\boldsymbol{X}_t)]dt \\ &\quad + \textstyle\sum_{i=1}^{d_y}\lambda_i^{\mathbb{Q}}\mathbb{E}_{\boldsymbol{\nu}^{\mathbb{Q}}}\big[(V(t,\boldsymbol{x}\circ(\boldsymbol{\iota}+\boldsymbol{\theta}_{:,i}(t,\boldsymbol{x})Z_1^i)) - V(t,\boldsymbol{x}))\big]_{\boldsymbol{x}=\boldsymbol{X}_{t-}}dt \\ &\quad + \nabla_{\boldsymbol{x}}V(t,\boldsymbol{X}_t)\bullet(\mathbf{D}_{\boldsymbol{X}_t}\boldsymbol{\sigma}(t,\boldsymbol{X}_t)d\boldsymbol{W}_t^{\mathbb{Q}}) \\ &\quad + \textstyle\sum_{i=1}^{d_y}\big\{(V(t,\boldsymbol{X}_{t-}\circ(\boldsymbol{\iota}+\boldsymbol{\theta}_{:,i}(t,\boldsymbol{X}_{t-})Z_{N_t^i}^i)) - V(t,\boldsymbol{X}_{t-}))dN_t^i \\ &\quad - \lambda_i^{\mathbb{Q}}\mathbb{E}_{\boldsymbol{\nu}^{\mathbb{Q}}}\big[(V(t,\boldsymbol{x}\circ(\boldsymbol{\iota}+\boldsymbol{\theta}_{:,i}(t,\boldsymbol{x})Z_1^i)) - V(t,\boldsymbol{x}))\big]_{\boldsymbol{x}=\boldsymbol{X}_{t-}}dt\big\}. \end{aligned}$$

Under $\mathbb{Q}$, the expectations of the diffusion term and the *compensated* jump terms (the last three lines) vanish. Furthermore, since

$$V_t^* := e^{-\int_0^t r_u du} V(t, \boldsymbol{X}_t),$$

is a $\mathbb{Q}$ martingale; dV_t^* should be driftless. These facts jointly imply that

$$\begin{aligned}
&- r_t V(t, \boldsymbol{X}_t) + \partial_t V(t, \boldsymbol{X}_t) + r_t \mathbf{D}_{\boldsymbol{X}_t} \boldsymbol{\iota} \bullet \nabla_{\boldsymbol{x}} V(t, \boldsymbol{X}_t) \\
&- \nabla_{\boldsymbol{x}} V(t, \boldsymbol{X}_t) \bullet (\mathbf{D}_{\boldsymbol{X}_{t-}} \boldsymbol{\theta}(t, \boldsymbol{X}_{t-}) \boldsymbol{\lambda}^{\mathbb{Q}} \circ \mathbb{E}_{\boldsymbol{\nu}^{\mathbb{Q}}}[\boldsymbol{Z}_1]) \\
&+ \tfrac{1}{2} \mathrm{tr}[\boldsymbol{\sigma}^{\intercal}(t, \boldsymbol{X}_t) \mathbf{D}_{\boldsymbol{X}_t} \nabla^2_{\boldsymbol{x}\boldsymbol{x}} V(t, \boldsymbol{X}_t) \mathbf{D}_{\boldsymbol{X}_t} \boldsymbol{\sigma}(t, \boldsymbol{X}_t)] \\
&+ \textstyle\sum_{i=1}^{d_y} \lambda_i^{\mathbb{Q}} \mathbb{E}_{\boldsymbol{\nu}^{\mathbb{Q}}} \big[(V(t, \boldsymbol{x} \circ (\boldsymbol{\iota} + \boldsymbol{\theta}_{:,i}(t, \boldsymbol{x}) Z_1^i)) - V(t, \boldsymbol{x})) \big]_{\boldsymbol{x} = \boldsymbol{X}_{t-}} = 0,
\end{aligned}$$

which essentially is what we wanted to show.

□

Appendix B: Data

Contract	ATM strike	Option price	$P\&L_T^{\text{actual}}$	$P\&L_T^{\text{implied}}$	$Q.V.^{\text{actual}}$	$Q.V.^{\text{implied}}$
07-Jul-2004	1118.3	36.5852	12.2045	15.1591	0.5269	0.2615
05-Oct-2004	1134.5	33.0392	5.8372	5.0520	0.1683	0.1386
05-Jan-2005	1183.7	34.7050	11.4080	13.6705	0.1975	0.1759
06-Apr-2005	1184.1	34.9985	7.3072	9.0917	0.3162	0.1693
06-Jul-2005	1194.9	34.4864	11.9818	10.5282	0.2974	0.0894
04-Oct-2005	1214.5	37.8141	7.2779	7.4261	0.5384	0.1670
06-Jan-2006	1285.4	37.1621	12.6952	12.4934	0.1539	0.1406
07-Apr-2006	1295.5	38.2703	0.0765	0.5022	0.2827	0.2444
07-Jul-2006	1265.5	45.5356	15.3714	13.7452	0.3655	0.1974
05-Oct-2006	1353.2	42.7682	12.6179	12.6400	0.0904	0.0945
08-Jan-2007	1412.8	45.4682	-5.0096	2.1741	2.4476	1.0569
09-Apr-2007	1444.6	47.0689	19.4885	7.4564	0.7699	0.0865
09-Jul-2007	1531.8	55.8378	-11.4976	-7.5524	1.8603	1.1396
05-Oct-2007	1557.6	63.1625	1.6451	-1.2115	1.2330	0.4542
09-Jan-2008	1409.1	74.2874	9.6117	9.6158	1.1975	0.6555
09-Apr-2008	1354.5	66.2276	17.3617	19.0049	0.8019	0.6270
09-Jul-2008	1244.7	62.8179	-56.9636	-47.0345	8.0872	10.4193
07-Oct-2008	996.2	83.8510	55.3847	51.8900	9.7721	6.4129
09-Jan-2009	890.3	69.9489	14.1892	3.2637	3.0947	0.4083
10-Apr-2009	856.6	62.9702	30.2400	27.2551	0.5701	0.4336
09-Jul-2009	882.7	49.8464	-12.4499	-9.8467	0.1245	0.1039
07-Oct-2009	1057.6	49.0640	17.0496	18.0507	0.2944	0.2135
07-Jan-2010	1141.7	42.2410	16.4989	16.4106	0.2595	0.1990
09-Apr-2010	1194.4	36.6784	-10.3121	-9.5031	0.5463	0.5578
09-Jul-2010	1078.0	52.2001	15.6833	17.6455	3.0501	0.3326
07-Oct-2010	1158.1	50.6050	20.7394	19.8607	0.1926	0.2166
06-Jan-2011	1273.8	43.6970	9.4015	11.7384	0.3762	0.2400
07-Apr-2011	1333.5	44.7866	13.3942	13.8116	0.3490	0.3055
08-Jul-2011	1343.8	43.0900	-3.8722	3.8883	0.1692	0.3196
06-Oct-2011	1165.0	73.4417	14.3245	16.8015	0.8601	0.7112
06-Jan-2012	1277.8	53.9770	-17.4158	-21.6739	0.3472	0.1853
09-Apr-2012	1382.2	48.9735	-9.7517	-9.9641	0.4760	0.4018
09-Jul-2012	1352.5	47.5814	15.8417	15.4475	0.3181	0.3184
05-Oct-2012	1460.9	42.9608	11.2648	9.0422	3.0925	0.8156
09-Jan-2013	1461.0	43.7355	17.6935	14.7094	0.2747	0.1360
11-Apr-2013	1593.4	39.2535	9.4000	6.6261	1.0037	0.7546

Table 3.6: The first column lists the purchasing dates of the 36 contracts. Column two shows the ATM strikes at which the contracts are purchased and column three show the prices at which this happens. The fourth column gives the terminal P&L for each contract, when the hedge is performed with an "actual" (EGARCH(1,1)) volatility forecast. Column five likewise, but when the hedge is with the implied volatilities. Finally, columns six and seven give the quadratic variation, defined as $\sum_{i=1}^{N} |dP\&L_i|^2/N$, where $N = 63$ is the number of trading days, for the entire actual and implied paths respectively.

4

Risk Minimization in Electricity Markets: Fixed Price, Unknown Consumption

Martin Jönsson, Rolf Poulsen, Rune Ramsdal-Ernstsen and Anders Skajaa[1]

Abstract. This paper analyses risk management of fixed price, unspecified consumption contracts in energy markets. We model the joint dynamics of spot-price and consumption of electricity, study expected loss minimisation for different loss functions and derive optimal static hedge strategies based on forward contracts. These strategies are tested empirically on 2012-2014 Nordic market data and compared to a simpler hedge strategy which is widely employed by the industry. Results show that our suggested hedge outperforms the commonly used with a higher reward-to-risk ratio, which can be exploited to release a premium from the contract. The realised cumulative profit-and-loss from our suggested hedge is greater for almost every single one-month period of the considered data, whilst the hourly realised payout results in a 66% out-performance probability.

Keywords: Quantity Risk, Electricity Markets, Hedging, Fixed Price Contracts.

[1]Jönsson, Poulsen and Ramsdal are from the Department of Mathematical Sciences, University of Copenhagen. Skajaa is from DONG Energy.

4.1 Introduction

A popular product sold by retail power-suppliers (in broad terms: energy companies) to medium-to-large costumers is a contract where an unspecified amount of power can be bought at a fixed price during a certain time-interval.[2] We call this a fixed price, unspecified consumption contract or, where it causes no confusion, simply a *fixed price agreement* or just *the contract*.

Having sold[3] the fixed price agreement, the energy company must asses and manage the risks associated with the contract. This problem is difficult for several reasons: If the seller faced only the price risk, he would simply buy forward contracts to set up a perfect hedge. But with a criterion function that takes into account both (expected) gains and losses, the fact that forward prices are typically above expected future spot-prices[4] will make it tempting to under-hedge. Indeed; the main complicating factor from a theoretical point of view is the quantity risk. If the consumer's quantity choice was all-or-nothing, the contract would have the payoff of a regular call option. Handling that would require modelling the price randomness of the electricity market (which has several characteristics that are different from standard financial markets, such as stocks and exchange rates). However, the consumer has more flexibility and because his energy demand is created by physical needs, his choice will not be all-or-nothing. Further, the demand will be positively correlated with the price (high demand drives up the price), which for the contract seller creates what in other areas of finance would be known as wrong-way risk: He will have to deliver a lot of energy for a fixed price when buying it in the spot market is expensive. In this paper, we will deal with this risk with a joint model of spot-price and consumption load.

The asymmetric structure faced by retail suppliers of inelastic demand, an obligation to deliver electricity at fixed prices (over fixed periods) procured from a spot market and a spot-price/demand correlation affecting adversely are all typical characteristics for the electricity market, see Stoft (2002). The non-storability of electricity (in contrary to other commodities) is at heart of the problem not only since it makes physical hedging inoperable, it also causes highly volatile spot-markets with structural price jumps (Benth et al. (2008)). As a consequence, the problem of joint quantity and price risk on electricity markets is commonly studied in the literature, and it stresses the importance of incorporating price–quantity correlation into the models. For an example, Bessembinder and Lemmon (2002) consider an equilibrium market model for which the correlation has a significant impact on optimal hedging strategies in forward markets. Their equilibrium perspective has several extensions, for instance Willems and Morbee (2010) who extend their market model to include additional derivatives and study the effects of offering an increasing number of options on the market. From a retailers perspective, with a exogenously given spot-

[2] The entire business-to-business electricity market in Denmark in 2014 was 23.1 TWh, or about 800 million euros in monetary terms. 60% was sold as fixed price agreements in some form.

[3] *Entered into* might be a better description than *sold*, as the contract typically requires no up-front payment.

[4] This so-called *contango* is the norm in Nordic power markets; see e.g. Botterud et al. (2002).

and forward market, Oum et al. (2006) employ utility maximisation to derive optimal hedging strategies where they exploit the correlation to manage price-quantity risk in a single period setting. As another example, Boroumand et al. (2015) address the problem of price–quantity risk on an intra-day scale and demonstrate with a simulation approach that intra-day hedging outperforms daily and longer-term portfolios.

From the retailer's perspective, we aim to capture the price-demand correlation with a bivariate model where both the spot-price and the load have temporal dependency. Based on our model, we set out to hedge the price-quantity risk of a fixed price agreement with forward contracts in a multi-period setting. In relation to our approach, Coulon et al. (2013) develop a more intricate three-factor model and consider hedging strategies based on options that are priced with closed-form expressions. Of course, numerous variations of spot-price and derivative pricing models exist in the literature (see Carmona and Coulon (2014) for a survey) but in contrast to the latter, we do not consider the problem of pricing derivatives under our model. Instead we concentrate on a hedging approach based on exogenously given market data.

Overview. The rest of the paper is organised as follows. In Section 4.2 we formulate a dynamic model for electricity price and load – a two-dimensional Ornstein-Uhlenbeck model with a seasonality component. In Section 4.3 we analyse expected loss minimization. Looking specifically at static hedging strategies with forward contracts, we derive explicit formulas for optimal hedge portfolios based on two different loss functions (the square and the "hockey stick"). Section 4.4 is an empirical study of the hedge performance based on publicly available consumption, spot and forward prices in the Nordic region. We find substantial benefits for our approach. Most notably, when we look at realised hourly payoffs from the hedge portfolio, we find that our suggested strategy has a greater tendency to generate positive outcomes (an average of 821 EUR), while the widely employed strategy we use for comparison often ends up with losses (-1,107 EUR). In other terms, we have a relatively larger hourly payoff from the model hedge with a 66% out-performance probability. This comes to the cost of a higher standard deviation (5,419 versus 3,325) and thus, with a higher risk in classical terms. But if we look at the reward-to-risk ratios of the two (0.15 versus -0.33), we find that our strategy outperforms to the extend that we may turn over the excess reward-to-risk into a contract premium. Viewed in different terms, we see that our strategy yields a greater accumulated profit-and-loss at month-end, for all but two one-month periods of the considered market data. At last, we repeat the hedging experiment for a proprietary data set from DONG Energy and we obtain even stronger results in this case.

4.2 Electricity price and consumption model

Our approach will be to model jointly the electricity spot price and the consumption load with a continuous stochastic process. Both the price and the load exhibit seasonal patterns – time of the day, day of the week and climatological season naturally effect the

spot price and consumption, see Benth et al. (2008) and Haugom (2011) – and we will fit deterministic periodic functions to capture this seasonality. The residual price and load will then be modelled with a stationary process and we conveniently choose the continuous Ornstein-Uhlenbeck process for this purpose (see for instance Weron (2007) and Benth et al. (2008) for general treatments of electricity-price modelling).

4.2.1 The bivariate Ornstein-Uhlenbeck process

The bivariate, mean-reverting Ornstein-Uhlenbeck process $(X_t, Y_t)_{t\geq 0}$ is given by the solution to the stochastic differential equation

$$\begin{aligned} dX_t &= \kappa_x(\theta_x - X_t)dt + \sigma_x dW_t^{(1)}, \\ dY_t &= \kappa_y(\theta_y - Y_t)dt + \sigma_y dW_t^{(2)}, \end{aligned}$$

where $\kappa_x, \kappa_y, \sigma_x, \sigma_y$ are positive parameters, $\theta_x, \theta_y \in \mathbb{R}$ and the Brownian motions $W^{(1)}$, $W^{(2)}$ are correlated with $\rho_w \in [-1, 1]$. The solution, obtained by integrating factors, is given by

$$\begin{aligned} X_t &= X_0 e^{-\kappa_x t} + \sigma_x \int_0^t e^{-\kappa_x(t-u)} dW_u^{(1)} + \theta_x, \\ Y_t &= Y_0 e^{-\kappa_y t} + \sigma_y \int_0^t e^{-\kappa_y(t-u)} dW_u^{(2)} + \theta_y, \end{aligned} \tag{4.1}$$

for $t \geq 0$ (see for instance Gardiner et al. (1985) for details). From (4.1) it is straightforward to calculate the (univariate) covariance functions

$$\mathrm{Cov}(X_{t+\Delta}, X_t) = \Sigma_x \, e^{-\kappa_x|\Delta|}, \;\; \mathrm{Cov}(Y_{t+\Delta}, Y_t) = \Sigma_y \, e^{-\kappa_y|\Delta|},$$

and the cross-covariance function

$$\begin{aligned} \mathrm{Cov}(X_{t+\Delta}, Y_t) &= \Sigma_{xy} \, e^{-\kappa_x|\Delta|}, \;\; \Delta \geq 0, \\ \mathrm{Cov}(X_{t+\Delta}, Y_t) &= \Sigma_{xy} \, e^{-\kappa_y|\Delta|}, \;\; \Delta < 0, \end{aligned}$$

where we have defined the parameters

$$\Sigma_x = \frac{\sigma_x^2}{2\kappa_x}, \;\; \Sigma_y = \frac{\sigma_y^2}{2\kappa_y} \text{ and } \Sigma_{xy} = \rho_w \frac{\sigma_x \sigma_y}{\kappa_x + \kappa_y}.$$

Since from (4.1) we see that X and Y are Gaussian processes, the bivariate process (X, Y) has a stationary distribution which is bivariate Normal

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} \sim BVN\left(\begin{bmatrix} \theta_x \\ \theta_y \end{bmatrix}, \begin{bmatrix} \Sigma_x & \Sigma_{xy} \\ \Sigma_{xy} & \Sigma_y \end{bmatrix} \right),$$

while the conditional distribution is given by

$$\begin{pmatrix} X_{t+\Delta} \\ Y_{t+\Delta} \end{pmatrix} | \begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} \sim BVN\left(\boldsymbol{\mu}_{t+\Delta|t}, \boldsymbol{\Sigma}_{t+\Delta|t} \right), \tag{4.2}$$

with mean vector and covariance matrix

$$\begin{aligned}\boldsymbol{\mu}_{t+\Delta|t} &= \begin{bmatrix}\theta_x + (x-\theta_x)e^{-\kappa_x|\Delta|} \\ \theta_y + (y-\theta_y)e^{-\kappa_y|\Delta|}\end{bmatrix}, \\ \boldsymbol{\Sigma}_{t+\Delta|t} &= \begin{bmatrix}\Sigma_x(1-e^{-2\kappa_x|\Delta|}) & \Sigma_{xy}(1-e^{-(\kappa_x+\kappa_y)|\Delta|}) \\ \Sigma_{xy}(1-e^{-(\kappa_x+\kappa_y)|\Delta|}) & \Sigma_y(1-e^{-2\kappa_y|\Delta|})\end{bmatrix}.\end{aligned}$$

With conditional distributions in hand for both the univariate and bivariate process, we proceed in the next section to the maximum likelihood estimator of the parameters.

4.2.2 Parameter estimation

With observed data we refer to a set of observations made at $n+1$ time-points $t_0, \ldots, t_n$ with $\Delta_i = t_i - t_{i-1}$ denoting the spacings of the time grid. An observed path $x = (x_0, \ldots, x_n)$ of the process $(X_t)_{t\geq 0}$ is thus a set of observations where x_i is the observed value of X_{t_i} for $i = 0, \ldots, n$. We write the log-likelihood for the observed path x of the univariate Ornstein-Uhlenbeck process as a function of the parameters $\psi = (\kappa, \theta, \sigma)$

$$l(\psi; x) = \sum_{i=0}^{n-1} \log \phi(x_{i+1}; \mu_{i+1|i}, \sigma^2_{i+1|i}) + \log \phi(x_0; \theta, \sigma^2/(2\kappa))$$

where ϕ is the normal density function, here with (conditional) mean $\mu_{i+1|i} = x_i e^{-\kappa\Delta_{i+1}} + \theta(1-e^{-\kappa\Delta_{i+1}})$ and variance $\sigma^2_{t+1|i} = \sigma^2(1-e^{-2\kappa\Delta_{i+1}})/(2\kappa)$. The decomposition of the density comes from the fact that X is a Markov process: with $f(x_{i+1}|x_i, \ldots, x_0) = f(x_{i+1}|x_i)$ for the conditional density, this allows us to write

$$f(x_0, \ldots, x_n) = f(x_n|x_{n-1}) \ldots f(x_1|x_0) f(x_0).$$

The maximum likelihood estimate $\hat{\psi}$ is the argument that maximizes the log-likelihood

$$\hat{\psi} = \arg\max_{\psi} l(\psi, x)$$

which may be obtained by a numerical optimizer. Calculating the observed information matrix

$$I_o = -\frac{\partial^2 l}{\partial \psi^T \partial \psi}\bigg|_{\psi=\hat{\psi}}$$

by numerical differentiation at estimated values gives an estimated standard error of the jth parameter as $\sqrt{(I_o^{-1})_{j,j}}$ where $(A^{-1})_{i,j}$ denotes element i, j of the inverse matrix of A. Further, if $\hat{\psi}_x$ and $\hat{\psi}_y$ are univariate estimates from two observed paths x and y, then

$$\hat{\rho}_w = \frac{\hat{\kappa}_x + \hat{\kappa}_y}{\hat{\sigma}_x \hat{\sigma}_y} \cdot \hat{\rho}(x, y)$$

gives an estimator of ρ_w where $\hat{\rho}(x, y)$ is the sample correlation coefficient between x and y.

A bivariate observation (x, y) from the bivariate Ornstein-Uhlenbeck process with parameters $\psi_{xy} = (\kappa_x, \theta_x, \sigma_x, \kappa_y, \theta_y, \sigma_y, \rho_w)$ yields the log-likelihood

$$(\psi_{xy}; x, y) = \sum_{i=0}^{n-1} \log \phi(x_{i+1}, y_{i+1}; \boldsymbol{\mu}_{i+1|i}, \boldsymbol{\Sigma}_{i+1|i}) + \log \phi(x_0, y_0; \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$$

where $\boldsymbol{\mu}_{i+1|i}$ and $\boldsymbol{\Sigma}_{i+1|i}$ are the conditional mean vector and covariance matrix given in the previous section. Here ϕ denotes the bivariate normal density and the maximum likelihood estimate $\hat{\psi}_{xy}$ may be obtained by numerical optimization along with estimated standard errors from the numerically calculated information matrix.

4.2.3 The seasonal Ornstein-Uhlenbeck model

Since the consumption and price of electricity exhibit seasonal behaviours, we continue to work with a model for the price S and load L as given by

$$\begin{aligned} S_t &= \tilde{S}_t + \theta_S(t) \\ L_t &= \tilde{L}_t + \theta_L(t) \end{aligned}$$

where $(\tilde{S}, \tilde{L})$ follows a bivariate Ornstein-Uhlenbeck process with constant mean while $\theta_S(t)$ and $\theta_L(t)$ are deterministic seasonality functions. For this purpose, we let $\theta(t)$ be a periodic function of the form

$$\theta(t) = \alpha_0 + \sum_{i=1}^{p} \alpha_i \cdot \sin(\frac{2\pi}{\tau_i} t + \phi_i) \tag{4.3}$$

with p periods $\tau_1, \ldots \tau_p$ and phases $\phi = (\phi_1, \ldots, \phi_p)$. For estimation of the parameters, notice that we may write (4.3) as a regression

$$\mathbf{y} = \mathbf{A}(\phi)\mathbf{a}$$

where $\mathbf{y} = (y_1, \ldots, y_n)^T$, $\mathbf{a} = (\alpha_0, \ldots, \alpha_p)^T$ and the regression matrix $\mathbf{A}(\phi)$ of dimension $[n \times p + 1]$ is a function of the phase vector ϕ

$$\mathbf{A}(\phi) = \begin{pmatrix} 1 & \sin(\frac{2\pi}{\tau_1} t_1 + \phi_1) & \ldots & \sin(\frac{2\pi}{\tau_p} t_1 + \phi_p) \\ 1 & \sin(\frac{2\pi}{\tau_1} t_2 + \phi_1) & \ldots & \sin(\frac{2\pi}{\tau_p} t_2 + \phi_p) \\ \vdots & \vdots & & \vdots \\ 1 & \sin(\frac{2\pi}{\tau_1} t_n + \phi_1) & \ldots & \sin(\frac{2\pi}{\tau_p} t_n + \phi_p) \end{pmatrix}.$$

If, for a moment, we consider the phases to be known, then the regression is a standard least-squares optimization problem

$$\mathbf{a}^* = \arg\min_{\mathbf{a}} ||\mathbf{y} - \mathbf{A}(\phi)\mathbf{a}||^2$$

which has the solution

$$\mathbf{a}^* = (\mathbf{A}(\phi)^T\mathbf{A}(\phi))^{-1}\mathbf{A}(\phi)^T\mathbf{y}$$

i.e. a function of the phase vector $\mathbf{a}^* = \mathbf{a}^*(\phi)$. This lead us to perform the parameter estimation in two steps: First we minimize

$$\hat{\phi} = \arg\min_{\phi} ||\mathbf{y} - \mathbf{A}(\phi)\mathbf{a}^*(\phi)||^2$$

by numerical optimization, then we perform the regression

$$\hat{\mathbf{a}} = \mathbf{a}^*(\hat{\phi}) = (\mathbf{A}(\hat{\phi})^T\mathbf{A}(\hat{\phi}))^{-1}\mathbf{A}(\hat{\phi})^T\mathbf{y}$$

to obtain parameter estimates $(\hat{\mathbf{a}}, \hat{\phi})$ of the periodic function. Hence, our approach will be to fit seasonal functions in a first step to the price and load respectively, and to fit an Ornstein-Uhlenbeck process to the de-seasonalized price and load in the second step. Henceforth, we refer to this model as the seasonal Ornstein-Uhlenbeck model.

4.3 Risk-minimizing static hedging

The hedging problem we are facing is to replicate a future uncertain financial obligation. That is, we will try to replicate the future payment with other financial products, typically liquid contracts that are traded in the market, and in our case it will be forward contracts. We will design a static hedging strategy: a buy-and-hold position in the forward contract that aims at hedging our financial obligation. Our approach to determine the forward position will be the method of expected loss minimization.

To this end, let π_T denote the payoff at time T of the hedged position, that is, the financial obligation we want to hedge together with a portfolio of hedging instruments. Then $-\pi_T$ gives the loss of our position and

$$\mathbb{E}[u(-\pi_T)] \tag{4.4}$$

gives the risk measure for some specified loss function u (see Artzner et al. (1999) for details of risk measures). The problem at hand is to choose the hedging portfolio such that this risk is minimized. Hence, the task is to choose both type of hedging instruments, and positions to buy/sell of these instruments at initiation. Notice that if we were to set up a dynamic hedging strategy instead, then we would have to choose the time-points at which to rebalance the hedging portfolio as well[5].

[5]For an overview of hedging in incomplete markets, and the case of dynamics hedging based on quadratic risk minimization, see e.g. Bingham and Kiesel (2004), Chapter 7.2 and references therein.

4.3.1 Hedging a fixed price agreement

Assume that we want to hedge a fixed price agreement that expires at a specified future time $T > t_0$ where t_0 is the initiation time. This is a financial contract with payoff $(S_T - F^{\text{fpa}})L_T$ at time T for the long position and its features are explained by the following hypothetical situation: At time t_0, we agree a price F^{fpa} with our counterparty and obligate ourselves to deliver an unspecified quantity L_T of electricity at T, for the price F^{fpa}. The quantity will be the unknown consumption of the counterparty. At the expiry time T, we will have to cover our contracted delivery of electricity by purchasing the same quantity from the spot market at a price S_T. Our cash-flow for the naked position at time T amounts to

$$F^{\text{fpa}} L_T - S_T L_T$$

which is the short position of the fixed price agreement.

To set up a static, buy-and-hold, hedge at time t_0 for the naked position (notice that S_T and L_T are unknown at t_0), we enter a forward contract with expiry T and forward price F, that is, a contract with payoff $S_T - F$ at time T, and we enter the forward with a position V to be determined. If we assume that the fixed price agreement has a price relative to the forward $F^{\text{fpa}} = F + m$, where m is some non-negative margin, then the payoff of our hedged position is

$$\pi_T = (S_T - F)(V - L_T) + mL_T.$$

The hedging problem is to specify, at time t_0, the forward position V in an optimal way. For this purpose, we employ the approach of minimizing the expected positive loss, i.e. with $u(x) = \max(x, 0)$ as the loss function of equation (4.4). Thus, we aim to minimize

$$f(V) = \mathbb{E}[\max\left(-(S_T - F)(V - L_T) + mL_T, 0\right)] \tag{4.5}$$

where the minimizing argument V^* forms the optimal hedge of our hedging strategy.

Financial contracts traded at the exchange for power derivatives are typically not specified with a single expiry-time as in the previous example. A power forward obligate the counterparties to settle the difference between spot and fixed price at a set of time-points during a specified period, for example a month, quarter or a year. There is typically a *base-load* and a *peak-load* version of the contract as well, where the latter has settlements on peak-hours only, that is weekdays 08:00 to 20:00 for the Nordic market, while the former has settlements on every full hour during the expiry period.

We consider fixed price agreements and forwards that are specified with an expiry month M and we let $M(p)$ denote peak-hours during this month and $M(op)$ the remaining off-peak hours. The payoff at time T_i for the hedged position during off-peak is thus $(S_{T_i} - F^{\text{b}})(V^{\text{b}} - L_{T_i})$ for all $T_i \in M(op)$, where V^{b} is the position in the base-load forward with price F^{b}. Notice that we assume the pricing $F^{\text{b-fpa}} = F^{\text{b}} + m$ with margin $m \geq 0$ for delivery during off-peak of the fixed price agreement; in particular, we let $m = 0$. For a time-point during peak periods we have the choice to include in our hedge portfolio

a position V^{p} in the peak-load forward with price $F^{\mathrm{p}} \geq F^{\mathrm{b}}$. This gives a payoff of the hedged position $(S_{T_i} - F^{\mathrm{b}})V^{\mathrm{b}} + (S_{T_i} - F^{\mathrm{p}})V^{\mathrm{p}} + (F^{\text{p-fpa}} - S_{T_i})L_{T_i}$ for $T_i \in M(p)$. If we define $\tilde{F} = F^{\mathrm{p}} - \frac{V^{\mathrm{b}}}{(V^{\mathrm{b}}+V^{\mathrm{p}})}(F^{\mathrm{p}} - F^{\mathrm{b}})$ such that $F^{\mathrm{b}} \leq \tilde{F} \leq F^{\mathrm{p}}$, we obtain

$$\pi_{T_i} = (S_{T_i} - \tilde{F})(V^{\mathrm{b}} + V^{\mathrm{p}} - L_{T_i}), \quad T_i \in M(p)$$

where we assume a pricing $F^{\text{p-fpa}} \geq \tilde{F} + m$ with margin $m \geq 0$ and we set $m = 0$ here as well. To determine the optimal hedge for the entire month M, we minimize the function

$$\begin{aligned} f(V^{\mathrm{b}}, V^{\mathrm{p}}) &= \sum_{T_i \in M(op)} \mathbb{E}\left[\max\left(-(S_{T_i} - F^{\mathrm{b}})(V^{\mathrm{b}} - L_{T_i}), 0\right)\right] \\ &+ \sum_{T_i \in M(p)} \mathbb{E}\left[\max\left(-(S_{T_i} - \tilde{F})(V^{\mathrm{b}} + V^{\mathrm{p}} - L_{T_i}), 0\right)\right] \end{aligned}$$

such that the minimizing argument $(V^{\mathrm{b}*}, V^{\mathrm{p}*})$ is the base- and peak-load forward positions of our hedge.

Remark 4.1. Firstly, note that our objective function $f(V^{\mathrm{b}}, V^{\mathrm{p}})$ is a sum over the individual risks of each hourly payoff π_{T_i}. An alternative to this is to view the collection of payoffs as a portfolio and minimise the portfolio risk instead, i.e. $\mathbb{E}\left[\max\left(-\sum_i \pi_{T_i}, 0\right)\right]$. Due to the sub-additivity property of the risk measure (indeed: of the function $\max(x, 0)$) we have that $\sum_i \mathbb{E}\left[\max\left(-\pi_{T_i}, 0\right)\right] \geq \mathbb{E}\left[\max\left(-\sum_i \pi_{T_i}, 0\right)\right]$ so we can potentially be more efficient, but this would come at a cost of higher complexity when calculating the objective function. Secondly, as $\max(x, 0)$ is convex, and convex functions are invariant under affine maps, we have that $V \mapsto \max(-(S - F)(V - L), 0)$ is convex, as well as $V \mapsto \sum_i \max(-(S_i - F)(V - L_i), 0)$. Thus we have that any local minimum is indeed a global minimum and the same holds for minimising $f(V^{\mathrm{b}}, V^{\mathrm{p}})$ jointly with respect to $(V^{\mathrm{b}}, V^{\mathrm{p}})$. Finally, the same will hold for the squared risk measure $\mathbb{E}\left[(\pi_T)^2\right]$, as the loss function x^2 is convex as well.

4.3.2 Hedging with the seasonal Ornstein-Uhlenbeck model

To be able to minimize the total expected loss of a monthly contract and find the optimal hedge, we will derive an expression for the expectation in equation (4.5) with the margin set to zero. To simplify the notation, assume a current time $t_0 = 0$ at which $S_0 = s$ and $L_0 = l$ are given for the spot and load, that is $\tilde{S}_0 = \tilde{s} = s - \theta_S(0)$ and $\tilde{L}_0 = \tilde{l} = l - \theta_L(0)$ are the initial values of the deseasonalized spot and load. Let $(\kappa_S, \kappa_L, \theta_S, \theta_S, \sigma_S, \sigma_L, \rho_W)$ be the parameters of the Ornstein-Uhlenbeck process $(\tilde{S}, \tilde{L})$. Then

$$\begin{pmatrix} S_T \\ L_T \end{pmatrix} | \begin{pmatrix} S_0 \\ L_0 \end{pmatrix} \overset{d}{=} \begin{pmatrix} \mu_S(T,T) + \sigma_S(T)Z \\ \mu_L(T,T) + \sigma_L(T)\left(\rho_{SL}(T)Z + \sqrt{1 - \rho_{SL}^2(T)}Z^{\perp}\right) \end{pmatrix}$$

where Z, $Z^{\perp}$ are independent standard normal variables and where we have defined

$$\mu_S(t, \tau) = \theta_S(t) + \theta_S + (\tilde{s} - \theta_S)e^{-\kappa_S \tau}$$

$$\begin{aligned}
\mu_L(t,\tau) &= \theta_L(t) + \theta_L + (\tilde{l} - \theta_L)e^{-\kappa_L \tau} \\
\sigma_S^2(\tau) &= \frac{\sigma_S^2}{2\kappa_S}(1 - e^{-2\kappa_S \tau}) \\
\sigma_L^2(\tau) &= \frac{\sigma_L^2}{2\kappa_L}(1 - e^{-2\kappa_L \tau}) \\
\rho_{SL}(\tau) &= 2\rho_W \frac{\sqrt{\kappa_S \kappa_L}}{(\kappa_S + \kappa_L)} \frac{(1 - e^{-(\kappa_S + \kappa_L)\tau})}{\sqrt{(1 - e^{-2\kappa_S \tau})(1 - e^{-2\kappa_L \tau})}}
\end{aligned}$$

with a slightly ambiguous notation. Thus, we may write the payoff at time T of the hedged position as the product of two linear functions of Z and $Z^{\perp}$

$$\begin{aligned}
&(S_T - F)(V - L_T) \\
&\stackrel{d}{=} \underbrace{(\mu_S(T,T) - F + \sigma_S(T)Z)}_{=f(Z)} \underbrace{\left(V - \mu_L(T,T) - \sigma_L(T)(\rho_{SL}(T)Z + \sqrt{1 - \rho_{SL}^2(T)}Z^{\perp})\right)}_{=g(Z,Z^{\perp})}
\end{aligned}$$

that is, a product of two dependent normal variables where F, V will play the role of $F^{\mathrm{b}}, V^{\mathrm{b}}$ and $\tilde{F}, V^{\mathrm{b}} + V^{\mathrm{p}}$. The expected positive loss may then be written as

$$\begin{aligned}
&\mathbb{E}[-\min(f(Z)g(Z,Z^{\perp}), 0)] \\
&= -\mathbb{E}[f(Z)g(Z,Z^{\perp})(\mathbf{1}_{f(Z)<0}\mathbf{1}_{g(Z,Z^{\perp})>0} + \mathbf{1}_{f(Z)>0}\mathbf{1}_{g(Z,Z^{\perp})<0})].
\end{aligned}$$

To calculate the terms $\mathbb{E}[fg\mathbf{1}_{f<0}\mathbf{1}_{g>0}]$ and $\mathbb{E}[fg\mathbf{1}_{f>0}\mathbf{1}_{g<0}]$, we use the conditional expectations

$$\begin{aligned}
h_1(z) &= \mathbb{E}\left(g(Z,Z^{\perp})\mathbf{1}_{g(Z,Z^{\perp})>0} | Z = z\right), \\
h_2(z) &= \mathbb{E}\left(g(Z,Z^{\perp})\mathbf{1}_{g(Z,Z^{\perp})<0} | Z = z\right),
\end{aligned}$$

and notice that we may write the expected loss as

$$\begin{aligned}
&-\mathbb{E}[fg\mathbf{1}_{f<0}\mathbf{1}_{g>0}] - \mathbb{E}[fg\mathbf{1}_{f>0}\mathbf{1}_{g<0}] \\
&= -\mathbb{E}\left[f(Z)h_1(Z)\mathbf{1}_{f<0}\right] - \mathbb{E}\left[f(Z)h_2(Z)\mathbf{1}_{f>0}\right] \\
&= -\mathbb{E}\left[f(Z)\left(h_1(Z)\mathbf{1}_{f<0} + h_2(Z)\mathbf{1}_{f>0}\right)\right].
\end{aligned}$$

For a general Gaussian variable, $X \sim N(\mu, \sigma^2)$,

$$\begin{aligned}
\mathbb{E}[X\mathbf{1}_{X>0}] &= \mu\Phi\left(\frac{\mu}{\sigma}\right) + \sigma\phi\left(\frac{\mu}{\sigma}\right), \\
\mathbb{E}[X\mathbf{1}_{X<0}] &= \mu\Phi\left(-\frac{\mu}{\sigma}\right) - \sigma\phi\left(\frac{\mu}{\sigma}\right),
\end{aligned}$$

where Φ and ϕ are the standard normal distribution and density respectively. These expressions may readily be used for h_1 and h_2 since

$$g(Z,Z^{\perp})|Z = z \sim N(\underbrace{V - \mu_L(T,T) - \sigma_L(T)\rho_{SL}(T)z}_{=\mu_g(z)}, \underbrace{\sigma_L^2(T)(1 - \rho_{SL}^2(T))}_{=\sigma_g^2}).$$

This leads us to define

$$\psi(z) = -f(z)\left(\left\{\mu_g(z)\Phi\left(\frac{\mu_g(z)}{\sigma_g}\right) + \sigma_g\phi\left(\frac{\mu_g(z)}{\sigma_g}\right)\right\}\mathbf{1}_{f(z)<0} + \right.$$
$$\left. + \left\{\mu_g(z)\Phi\left(-\frac{\mu_g(z)}{\sigma_g}\right) - \sigma_g\phi\left(\frac{\mu_g(z)}{\sigma_g}\right)\right\}\mathbf{1}_{f(z)>0}\right)$$

which gives us a final expression for the expected positive loss

$$\mathbb{E}[\max(-(S_T - F)(V - L_T), 0)] = \mathbb{E}[\psi(Z)]$$

where Z is a standard normal variable.

For the special case when $\tilde{S}$ and $\tilde{L}$ are uncorrelated, and thus independent, i.e. $\rho_W = 0$ and $\rho_{SL}(\tau) = 0$, we obtain

$$\begin{aligned}&(S_T - F)(V - L_T)\\ &\overset{d}{=} \underbrace{(\mu_S(T,T) - F + \sigma_S(T)Z)}_{=f(Z)}\underbrace{\left(V - \mu_L(T,T) - \sigma_L(T)Z^{\perp}\right)}_{=g(Z^{\perp})}\end{aligned}$$

this simplifies the expected positive loss to a closed-form expression

$$\begin{aligned}&\mathbb{E}[-\min(f(Z)g(Z^{\perp}), 0)]\\ &= -\mathbb{E}[f(Z)g(Z^{\perp})(\mathbf{1}_{f(Z)<0}\mathbf{1}_{g(Z^{\perp})>0} + \mathbf{1}_{f(Z)>0}\mathbf{1}_{g(Z^{\perp})<0})]\\ &= -\mathbb{E}[f(Z)\mathbf{1}_{f(Z)<0}]\cdot\mathbb{E}[g(Z^{\perp})\mathbf{1}_{g(Z^{\perp})>0}] - \mathbb{E}[f(Z)\mathbf{1}_{f(Z)>0}]\cdot\mathbb{E}[g(Z^{\perp})\mathbf{1}_{g(Z^{\perp})<0}]\\ &= \left\{(F - \mu_S(T,T))\Phi\left(\tfrac{F-\mu_S(T,T)}{\sigma_S(T)}\right) + \sigma_S(T)\phi\left(\tfrac{\mu_S(T,T)-F}{\sigma_S(T)}\right)\right\}\\ &\cdot\left\{(V - \mu_L(T,T))\Phi\left(\tfrac{V-\mu_L(T,T)}{\sigma_L(T)}\right) + \sigma_L(T)\phi\left(\tfrac{V-\mu_L(T,T)}{\sigma_L(T)}\right)\right\}\\ &+ \left\{(\mu_S(T,T) - F)\Phi\left(\tfrac{\mu_S(T,T)-F}{\sigma_S(T)}\right) + \sigma_S(T)\phi\left(\tfrac{\mu_S(T,T)-F}{\sigma_S(T)}\right)\right\}\\ &\cdot\left\{(\mu_L(T,T) - V)\Phi\left(\tfrac{\mu_L(T,T)-V}{\sigma_L(T)}\right) + \sigma_L(T)\phi\left(\tfrac{V-\mu_L(T,T)}{\sigma_L(T)}\right)\right\}.\end{aligned}$$

The expected loss calculated with dependency between the deseasonalized spot and load[6] is shown in Figure 4.1 along with the expected loss calculated from the independent specification of the model. The difference between the two expected loss functions is small. In particular, the minimum expected loss from the two models is obtained at the same argument, that is, the zero-correlation case yields the same optimal hedge position as the dependent case.

To end the section with a second choice of risk measure, notice that a quadratic loss function $u(x) = x^2$ gives the expected squared loss

$$\mathbb{E}[(S_T - F)^2(V - L_T)^2]$$

[6]The is calculated by approximating $\mathbb{E}[\psi(Z)]$ with $\frac{1}{n}\sum\psi(z_i)$ where z_i are 1,000 generated outcomes from the standard normal distribution.

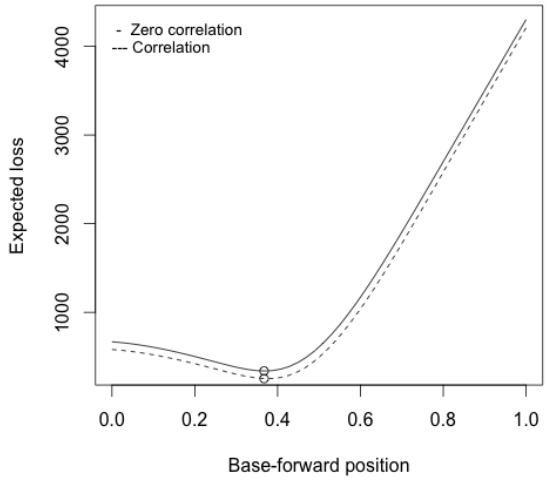

Figure 4.1: Expected positive loss calculated from the dependent (solid line) and independent (dashed line) model for the deseasonalized data with parameters in Table 4.5 and the January-13 contract in Table 4.3. Notice that both models yield the same minimizing argument (circles), i.e. the same optimal hedging position in the forward contract.

$$\begin{aligned} &= \mathbb{E}[f(Z)^2 g(Z, Z^\perp)^2] \\ &= \mathbb{E}[f(Z)^2 \cdot \mathbb{E}(g(Z, Z^\perp)^2 | Z = z)] \end{aligned} \tag{4.6}$$

where

$$\begin{aligned} \mathbb{E}(g(Z, Z^\perp)^2 | Z = z) = \sigma_L^2(T)(1 - \rho_{SL}^2(T)) + (V - \mu_L(T,T))^2 \\ +\sigma_L^2(T)\rho_{SL}^2 z^2 - 2(V - \mu_L(T,T))\sigma_L(T)\rho_{SL}(T)z. \end{aligned}$$

Substitute the conditional expectation into equation (4.6) to obtain, after a few simplifications, the quadratic function

$$\begin{aligned} &\mathbb{E}[(S_T - F)^2 (V - L_T)^2] \\ &= \left((\mu_S(T,T) - F)^2 + \sigma_S^2(T)\right)\left((V - \mu_L(T,T))^2 + \sigma_L^2(T)\right) \\ &+2\sigma_S^2(T)\sigma_L^2(T)\rho_{SL}^2(T) - 4(\mu_S(T,T) - F)(V - \mu_L(T,T))\sigma_S(T)\sigma_L(T)\rho_{SL}(T) \end{aligned}$$

where we have employed the fact that $\mathbb{E}[Z] = \mathbb{E}[Z^3] = 0$ while $\mathbb{E}[Z^2] = 1$ and $\mathbb{E}[Z^4] = 3$ for a standard normal variable. Differentiation with respect to V and equating to zero yields

$$V^* = \mu_L(T,T) + 2\frac{(\mu_S(T,T) - F)\sigma_S(T)\sigma_L(T)\rho_{SL}(T)}{(\mu_S(T,T) - F)^2 + \sigma_S^2(T)} \tag{4.7}$$

which is the forward position of the optimal hedge obtained from minimizing the expected quadratic loss. Notice here that quadratic hedging suggests to hedge with a forward position equal to the expectation of the load in the case when the deseasonalized spot and load are uncorrelated.

4.4 Empirical study

In this section, we present an empirical study of the performance of our hedging approach. We base the study partly on publicly available market data, partly on a set of proprietary load data obtained from DONG Energy, who is a major player in the Danish power market. The data is presented in the next section along with fitted seasonal functions and estimated Ornstein-Uhlenbeck parameters from the residual processes. In the forthcoming section, we present the results from our hedging approach based on expected loss minimization and we compare these with an "average load" hedging strategy, which is a strategy widely employed by the industry.

4.4.1 Data and fitted model

We will test the empirical performance of our model's loss-minimizing hedge using spot prices, forward prices and consumption data from the price area DK1 of the Danish power market.[7] The price and load data are recorded with an hourly observation frequency during the period 1 January 2012 to 31 December 2014 and it is shown in Figure 4.2. The spot price (top pane) is measured in EUR/MWh (euros per megawatt hours) while the load data (bottom pane) is measured in MWh. In addition to the publicly available consumption data in Figure 4.2, we have access to a set of proprietary load data that has been made available to us from DONG Energy. This data has been anonymized by scaling with the maximum load to show a percentage between zero and one and it is shown in Figure 4.17 in the appendix. We return to the data from DONG Energy in Section 4.4.3.

We use market quotes of peak- and base forwards with monthly expiries for the DK1 price area in addition to the spot-price and load data. We will initiate our hedge approximately two weeks prior to the expiry month of the froward contract and we have listed the corresponding forward prices in Table 4.3.

Remark 4.2. The data from the Nordic power market contains quotes of base- and peak forwards noted on the Nord-Pool system sport-price and *certificate of difference* forwards (CDF) for the different Nordic price areas (see Table 4.15 in the appendix). As the CDF forward for the DK1–system difference is quoted for a base version only, we are, in fact, subject to a bias risk that we do not hedge against. However, since the bias affects peak-hours only with the DK1–system spot-price difference (typically an order of magnitude smaller than the actual price) and with the peak-position (typically half the size of the base-position), we chose to ignore the bias and perform our hedging experiment as if the peak-forward was noted on the DK1 spot price as well. We ignore the same bias for the average-load hedge to be able to fairly compare the two strategies.

[7] Spot- and forward prices for the Nordic market are publicly available at `http://www.nordpoolspot.com` and `http://www.nasdaqomx.com` while load data for Denmark is available at `http://energinet.dk`.

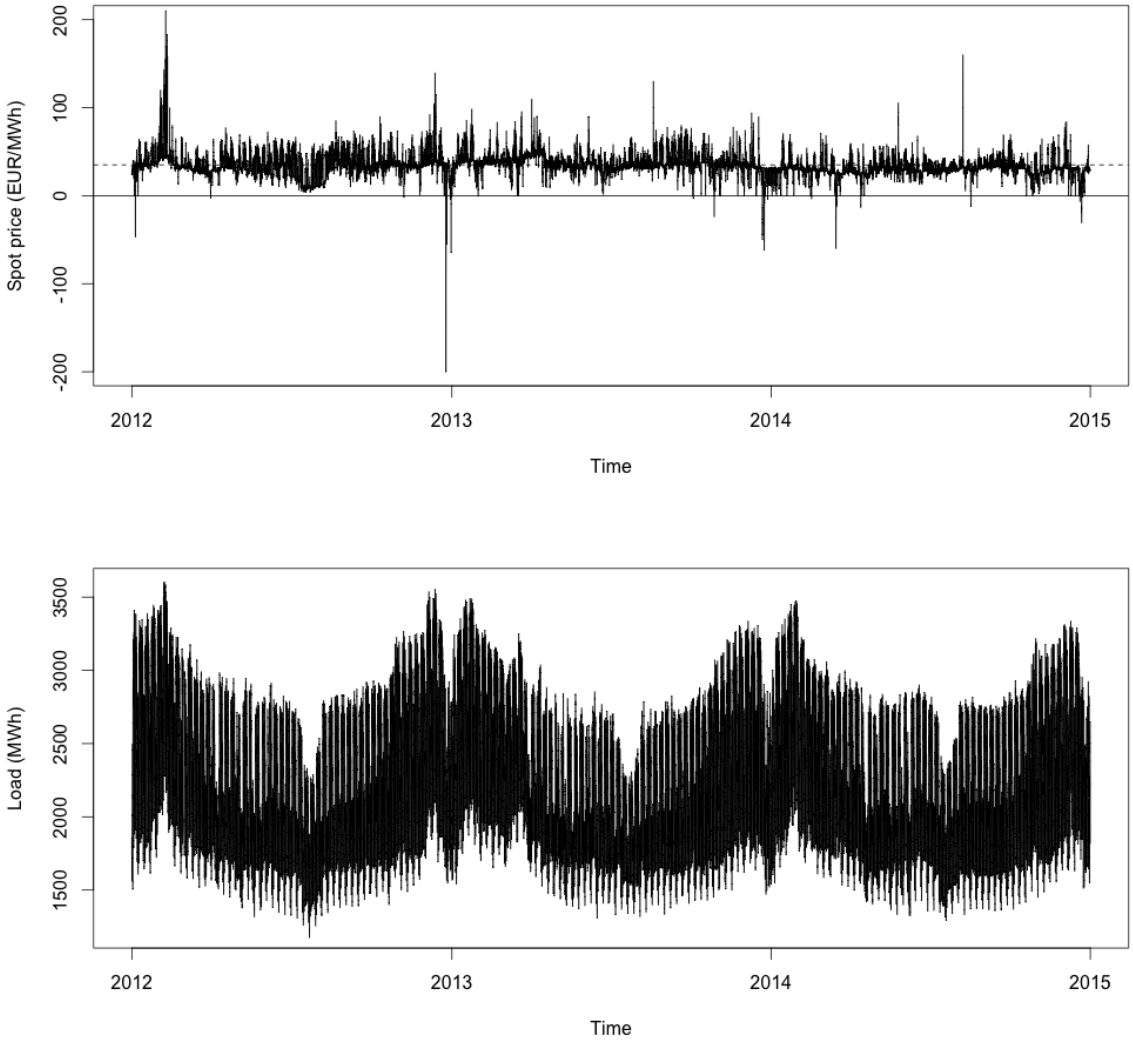

Figure 4.2: Spot price (top) and consumption (bottom) from the DK1 price area of the electricity market in Denmark. The dashed line marks the average spot price 34.96 EUR/MWh. The data is recorded with an hourly frequency and covers the period from 1 January 2012 to 31 December 2014.

Upon inspection of Figure 4.2, we see that both the price and load series exhibit seasonal patters which are particularly evident for the load: Power consumption inherently varies with the time of the day and day of the week. Due to electrical heating during winter (and air cooling during summer) there is a variety with the climatological season as well. The

Initiation t_0	Expiry month M	Base forward F^{b}	Peak forward F^{p}
2012-12-17	January-13	44.55	53.25
2013-01-17	February-13	45.90	52.95
2013-02-14	March-13	37.90	38.80
2013-03-18	April-13	37.15	39.65
2013-04-16	May-13	36.60	36.05
2013-05-16	June-13	33.55	34.85
2013-06-17	July-13	32.90	35.30
2013-07-17	August-13	38.90	41.95
2013-08-16	September-13	40.20	43.10
2013-09-16	October-13	40.20	45.30
2013-10-17	November-13	40.30	46.30
2013-11-18	December-13	37.40	39.45
2013-12-17	January-14	36.00	41.10
2014-01-17	February-14	36.02	41.02
2014-02-17	March-14	29.90	33.95
2014-03-17	April-14	31.15	35.18
2014-04-16	May-14	30.35	34.35
2014-05-16	June-14	30.10	34.15
2014-06-17	July-14	30.79	33.60
2014-07-17	August-14	34.25	37.15
2014-08-18	September-14	35.50	38.50
2014-09-17	October-14	34.60	37.15
2014-10-17	November-14	32.60	34.65
2014-11-17	December-14	30.25	32.05

Table 4.3: Forward prices (in EUR) at time t_0 for contracts with expiry months during 2013-2014.

seasonality of consumption is transferred to the spot price as demand for electricity affects its price. Further, as climatological season affects the production of hydroelectricity, there will also be seasonal dependent factors that affects supply, and thus the price. Finally, notice that the spot price may turn negative with some remarkably large negative values recorded around the new years of[8] 2012, 2013 and 2014.

To achieve a better fit for the seasonality functions, we separate weekends (including holidays) from weekdays (Monday to Friday) and fit two periodic functions separately to the two data sets.[9] The combined seasonality functions for load and spot price from the period 2012-01-01 to 2012-12-31 (Figure 4.4) clearly captures a seasonal behaviour where the fitted function for the load is particularly reminiscent of the original data. We used

[8] An extreme price entry of 2000 EUR/MWh recorded on 2013-07-08 is treated as an outlier and removed from the data.

[9] The market data is subject to wintertime changes (one "missing" hour) and summertime changes (one "additional" hour) that will incur an one-hour time shift of all time points during the summertime. To align the data with the 24-hour periodicity of the seasonal functions, we duplicate the last data-point of the wintertimes and remove the last of summertimes.

five periods of 12, 24, 168 (one week), 4380 (half year) and 8760 (one year) hours for this purpose. Notice that the axes' scales of Figure 4.4 are kept unchanged from Figure 4.2 to highlight how the amplitude of fitted periodic functions compare to the amplitude of unfitted data. The residual processes (the *deseasonalized data*, middle Figure 4.7) are then

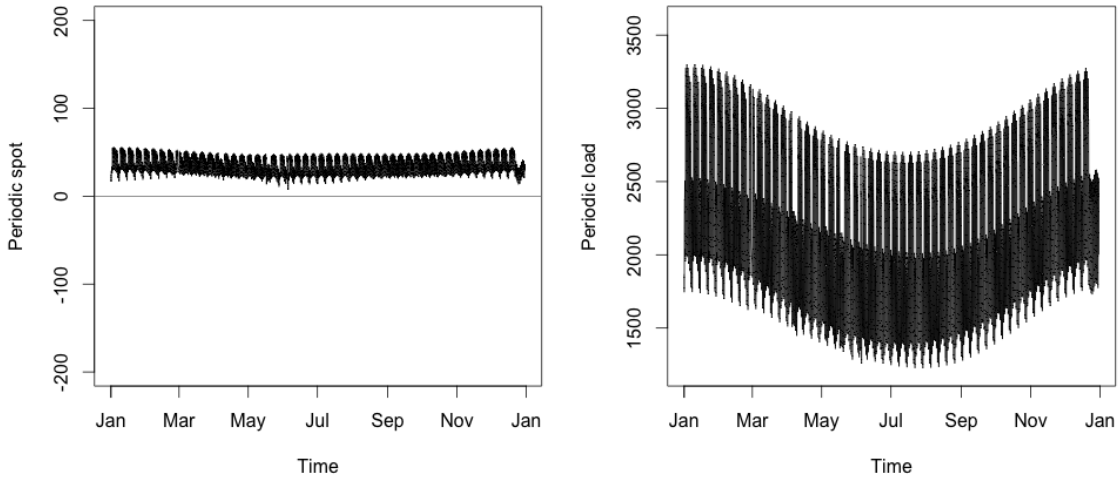

Figure 4.4: The figures show seasonality functions fitted to the spot price (left) and consumption (right) of the period 2012-01-01 to 2012-12-31.

fitted to a bivariate Ornstein-Uhlenbeck process with resulting estimated parameters and standard errors recorded in Table 4.5.

	κ	θ	σ	ρ_w
Spot	0.12 (0.005)	0.00 (0.59)	6.36 (0.051)	0.32 (0.010)
Load	0.19 (0.007)	0.00 (5.45)	94.2 (0.78)	

Table 4.5: Estimated parameters of the Ornstein-Uhlenbeck process from deseasonalized data 2012-01-01 to 2012-12-31 with standard errors in parenthesis. The parameters are estimated by numerical optimization of the bivariate maximum likelihood function, where estimates from the univariate maximum likelihood are employed as starting values for the numerical optimization.

We intend to do a rough investigation of the goodness-of-fit of our model before we continue with the hedging problem. The histograms of deseasonalized data (Figure 4.6) show a fairly good fit to the normal distribution, (the marginal distribution of the Ornstein-Uhlenbeck process is normal) for the spot and in particular, for the load. Next, if we compare sim-

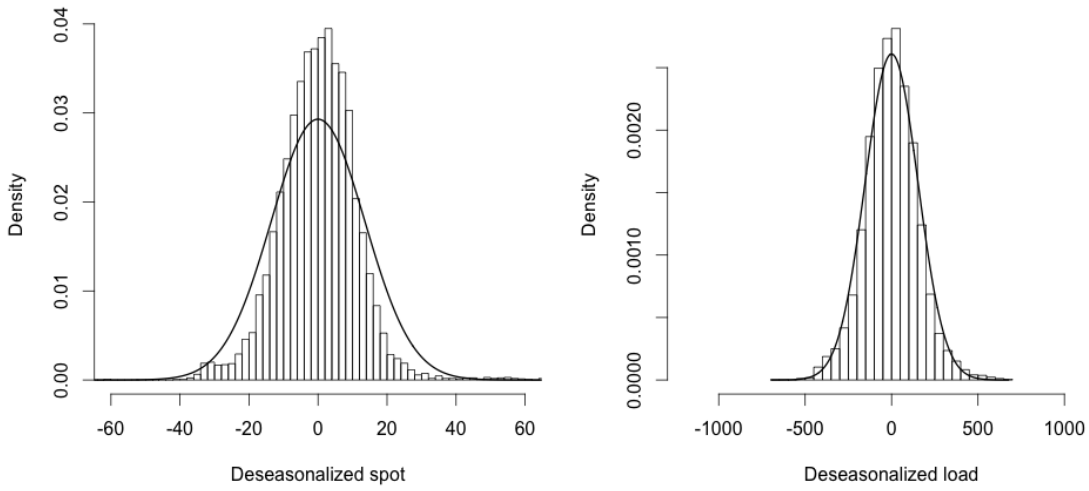

Figure 4.6: Histograms of the deseasonalized spot (left) and load (right) from the period 2012-01-01 to 2012-12-31. The solid lines are the fitted normal densities of the spot and load respectively.

ulated paths from the bivariate Ornstein-Uhlenbeck process[10] (left Figure 4.7) with the deseasonalised data (middle Figure 4.7), we find that the simulated processes are fairly reminiscent of the data we intend to model. Obviously, our model can not generate any

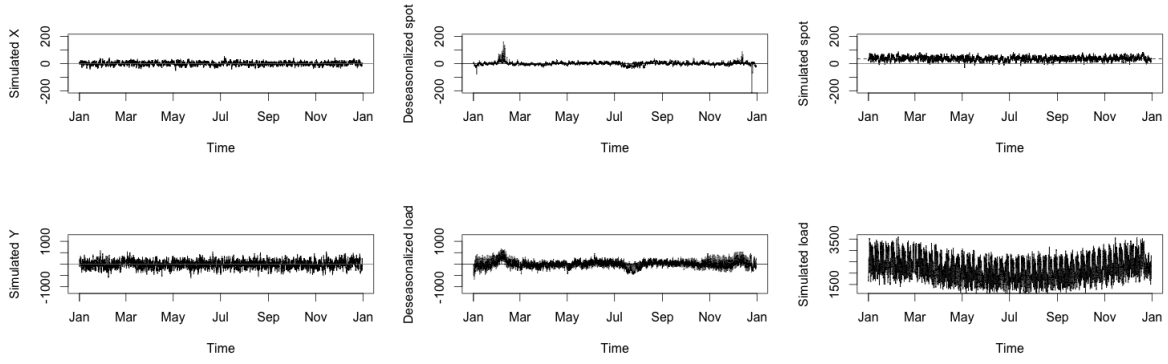

Figure 4.7: Left figures show simulated sample paths from the bivariate Ornstein-Uhlenbeck processes which are used to model the deseasonalized spot and load (middle figures). Right figures show simulated spot (top) and load (bottom) obtained by adding fitted seasonality functions to the simulated paths in the left figures.

jump discontinuities (the Ornstein-Uhlenbeck process and the periodic function are con-

[10]We employ an exact simulation method for the bivariate Ornstein-Uhlenbeck process, see Platen and Bruti-Liberati (2010).

tinuous functions) or jump-alike extremes (the Ornstein-Uhlenbeck process has normal distributed increments) and thus we fail to capture the jumps of our data. This is particularly evident when we look at the simulated spot (i.e. the simulated Ornstein-Uhlenbeck process added to the fitted seasonality function, top right Figure 4.7) compared to the original data (top Figure 4.2) whilst it is not as evident for the load. However, as our ultimate objective is to find an efficient hedging strategy and not to find the best fitting model, we are confident with our modelling approach whilst aware of some of its shortcomings.

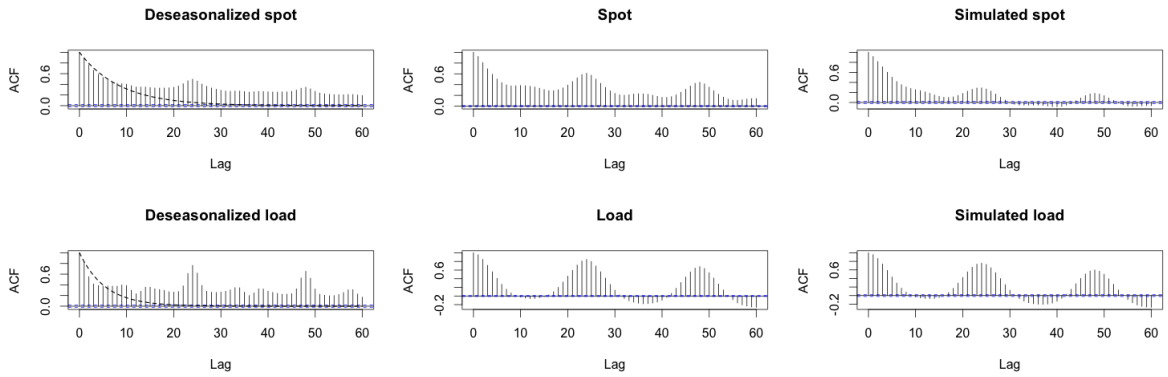

Figure 4.8: Sample autocorrelation of deseasonalized data (left figures). The dashed lines show the autocorrelation function of the Ornstein-Uhlenbeck process calculated with estimated parameters. Sample autocorrelation functions of spot and load (middle figures) and corresponding functions of the simulated spot and load (right figures).

If we look at the sample autocorrelation functions of the deseasonalized data and compare to the correlation function $e^{-\kappa\Delta}$ of the Ornstein-Uhlenbeck process, we find a mediocre match (left Figure 4.8). However, notice that the sample autocorrelation of the spot and load (middle Figure 4.8) are quite similar to the sample autocorrelation functions of the simulated processes (right Figure 4.8). This is due to fact that the temporal dependency is partly captured by the seasonality functions.

Remark 4.3. When we derived an expression for the expected positive loss under the seasonal Ornstein-Uhlenbeck model (Section 4.3.2), we obtained a simpler, closed-form expression in the case when there is no cross-correlation between the deseasonalized processes. We then compared to the dependent case with an example, and saw that the two loss functions where quit close (in particular, their minimum was obtained at the same argument). An explanation for this is that the cross dependency is partly captured by the seasonality functions, partly by the Ornstein-Uhlenbeck process (just as for the temporal dependency). Indeed: for the 2012 data, the empirical cross-correlation is 0.37 of the deseasonalized price and load, and 0.93 of the seasonality functions. This means that we might do good enough with the closed-form expression of the uncorrelated model specification, since much of the cross dependency is captured in the seasonality functions.

4.4.2 Empirical hedging experiment

We set up our empirical experiment for the DK1 market to hedge 24 fixed-price agreements with monthly expires during 2013 and 2014. At initiation time t_0, set to be approx 15 days prior to the first day of the expiry month, we enter a fixed price agreement with reference load according to the consumption data presented in previous section and a fixed price according to the corresponding t_0 market forward price[11] of Table 4.3. To hedge the contract, we enter forwards (on the same expiry month) with hedge positions decided from either the minimum expected loss function (the *model hedge*) or from the average-load strategy.

For the experiment, we rescale the load data with the maximum load to improve numerical stability and to obtain a time-series that corresponds to the rescaled load data from DONG Energy. This has the only consequence that our hedging positions $(V^{\mathrm{b}}, V^{\mathrm{p}})$ will be expressed in units of percentage of maximum load. To begin with the January-13 contract, we employ six months of spot and (rescaled) load data up until t_0, that is 2012-06-18 to 2012-12-17, to estimate the seasonal Ornstein-Uhlenbeck model in the first step (parameters recorded in Table 4.16 in the appendix). We use three periods of 12, 24 and 168 hours. Including periods of a half/one year will improve the fit within the estimation set, whilst extrapolation from long periods will impair the out-of-sample fit (i.e. the prediction of the periodic function throughout the expiry month).

Based on estimated parameters, we calculate the expected positive loss as a function of peak- and base forward positions in the hedge portfolio (Figure 4.9) and find the argument $(V^{\mathrm{b}*}, V^{\mathrm{p}*}) = (0.361, 0.340)$ that achieves the minimum-loss to obtain our optimal hedge. Note that, as the load data is rescaled with the maximum load, $V^{\mathrm{b}*}$, $V^{\mathrm{p}*}$, are percentages of the maximum load as well. The realised payoff from the optimally hedged position is then calculated for each hour t_i of the expiry month M as

$$\begin{aligned}\pi_{t_i} &= (S_{t_i} - F^{\mathrm{b}})(V^{\mathrm{b}*} - L_{t_i}), \quad t_i \in M(op),\\ \pi_{t_i} &= (S_{t_i} - \tilde{F})(V^{\mathrm{b}*} + V^{\mathrm{p}*} - L_{t_i}), \quad t_i \in M(p),\end{aligned}$$

where S_{t_i} and L_{t_i} are the realised spot-price and (unscaled) consumption while the optimal forward positions have been scaled back with the maximum load. The results for the January-13 contract are shown in left Figure 4.10.

In order to verify our results for the model hedge, we calculate the average-load, $(\bar{V}^{\mathrm{b}}, \bar{V}^{\mathrm{p}}) = (0.632, 0.222)$, (percentage of maximum load) and show the realised hourly payoff in the right-hand panel of Figure 4.10. The average-load strategy means holding forward positions corresponding to the expected load as obtained by minimizing the quadratic loss function under zero correlation.

[11] The Nordic power market supplies forwards noted on the Nord-Pool system spot-price and *Contracts for difference* (CFD forwards) for the different Nordic price-areas whereas Table 4.3 shows a combination of the two for DK1 prices, see Remark 4.2. The original data is listed in Table 4.15 in the appendix.

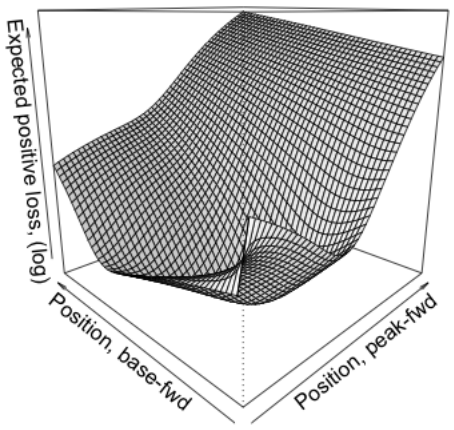

Figure 4.9: Expected positive loss of the hedged position for the January-13 contract calculated from the seasonal Ornstein-Uhlenbeck model. The minimizing argument $(V^{\mathrm{b}*}, V^{\mathrm{p}*}) = (0.361, 0.340)$ gives the optimal hedge (measured as percentages of maximum load).

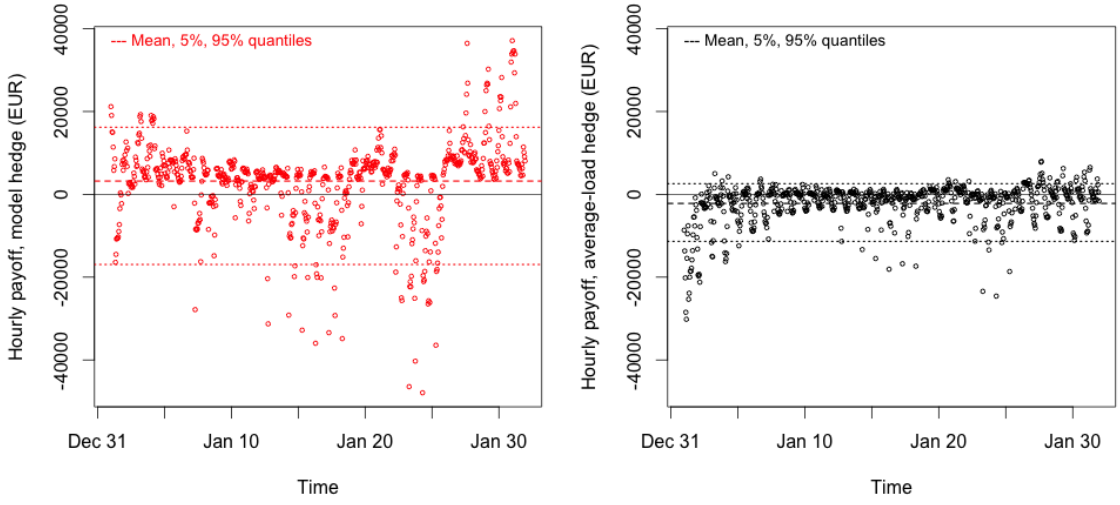

Figure 4.10: Realised hourly payoff from the model hedge (left-hand panel) calculated for the January-13 contract. The realised payoff from the average-load strategy (right-hand panel) has a lower (negative) average, while the variance of the average-load strategy's payoff is lower than the variance of the payoff from the model hedge.

The most notable feature from the hourly payoffs of the model hedge and average-load hedge (Figure 4.10) is that the model hedge yields an average payoff per hour that is positive at 3,208 EUR, while the average-load hedge yields a negative average hourly payoff of -2,192 EUR. The negative average payoff inexorably leads to a negative accumulative payoff at expiry for the January-13 contract (Figure 4.11), while the model hedge dispenses a postitive net profit-and-loss at the expiry. On the other hand, the standard deviation of hourly model-hedge payoff is 10,224 EUR, while the average-load hedge yields a lower standard deviation of 4,808 EUR. This means reducing the risk in classical terms with a lower variation of the outcome.

Another measure of efficiency for the hedging strategies is the probability of loss: 22.6% for the model hedge, while it is considerable higher at 63.2% for the average-load hedge. In concrete terms this means that we make a loss in 63.2% of the hourly payoff outcomes with the average-load strategy (more than every second outcome) while the model-hedge yields losses in 22.6% of the outcomes.

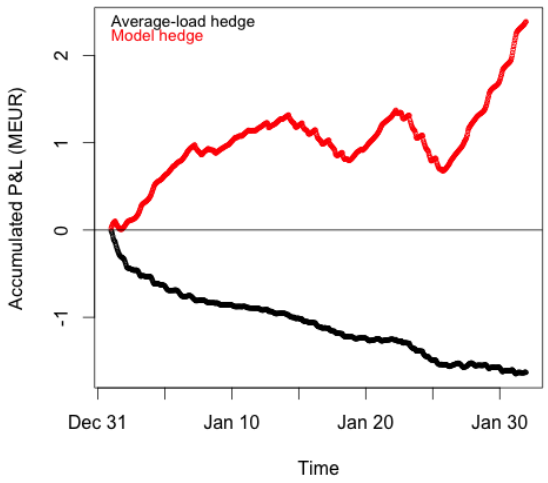

Figure 4.11: Accumulated profit-&-loss in million euros of the hedged position for the January-13 contract. The model hedge (red) yields a month-end profit of 2.39 MEUR while the average load-hedge (black) yields a loss of -1.63 MEUR.

4.4.3 Results

If we repeat the above procedure and determine the static hedges for all 24 contracts, we obtain the optimal peak/load forward positions reported in Table 4.13 for the model hedge and average-load hedge respectively. As for the January-13 contract, we see in Table 4.13 that the average hourly payoffs of the model hedge are positive for a majority of the

contracts, and always higher (except for two contracts) than the average hourly payoff for the corresponding average-load hedge. This is natural as the purpose of the model hedge is to minimise the (expected) positive loss. We also observe that the standard deviation of hourly payoffs is higher for the model hedge for all but one contract (where the mean is still higher) when we compare to the average-load strategy. As we basically minimise the variance of the payoff (or the loss) with the average-load strategy, this should be expected as well.

The effect of a relatively high hourly payoff with a higher standard deviation for the model hedge can be seen for the entire set of contracts in the top and middle panes of Figure 4.14. The model hedge has a slightly positive average hourly payoff of 821 EUR (when looking at the entire 24-month period) while the average-load has a negative average payoff per hour of -1,107 EUR. In return, we observe from the quantiles that the model hedge yields a greater variation (a standard deviation of 5,419 EUR while the average-load yields 3,325 EUR). From the figures, it is notable that the negative part of the hourly payoff outcomes are fairly similar for the two strategies (with a few more large negative payoffs for the model hedge) while the set of positive outcomes is considerably reduced for the average-load hedge. If we compare the relative size of the hourly payoff from the strategies, we find that the model hedge yields the largest payout 66% of the time. This agrees with the probability of loss, which is lower for the model hedge for almost all contracts as recorded in Table 4.13. We can also observe from the hourly payoff distribution in Figure 4.12 that the model hedge yields a positive skew, while the average-load has a negativity skewed distribution.

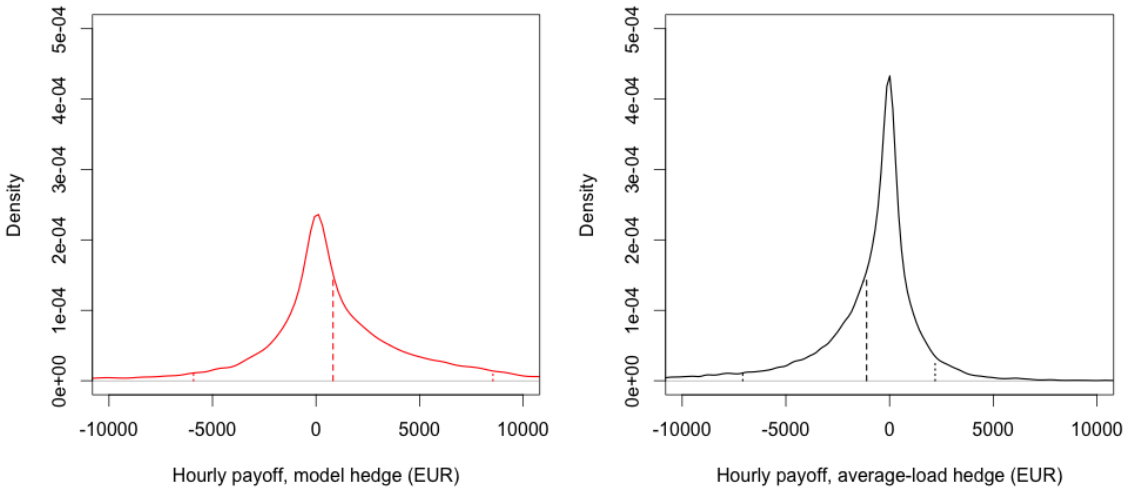

Figure 4.12: Left figure shows (estimated) density of the hourly payoff of the model hedge from all 24 contract. The respective density for the mean-load is shown in the right figure. The dashed vertical lines show means, 5% and 95% quantiles.

In all, it thus seems that we have reduced the opportunity of large, positive payoff outcomes with the average-load strategy, for the benefit of a smaller variance. But this comes at a large cost in terms of profit. If we look at the month-end, accumulated profit-&-loss for all 24 contracts plotted in the bottom pane of Figure 4.14, we find negative P&Ls for all contracts hedged with the average-load strategy, while the model hedge yields a profit in the end of the month for most of the contracts.

To make this qualitative observation somewhat more quantitative, we calculate a reward-to-variability ratio of the realised monthly returns (the accumulated P&L at month-end). For the model hedge, this gives 0.60 while the average-load hedge yields a negative ratio of -2.07. Inevitably, this emanates from the fact that the average-load strategy yields negative realised monthly returns.

One way of dealing with this problem for the average-load strategy is to charge a safety margin m to the contract price, such that the total realised payoff adds to

$$\sum_{t_i \in M(op)} \pi_{t_i} + \sum_{t_i \in M(p)} \pi_{t_i} + \sum_{t_i \in M} mL_{t_i}.$$

To make the comparison between the two strategies more concrete, we may then calculate a margin "certainty equivalent" that makes the two strategies have the same realised reward-risk ratio. Namely, we have to add a margin of $m = 0.62$ EUR/MWh to the average-load strategy to obtain a reward-to-risk of 0.60. In economic terms, this gives roughly $0.62 \times 30 \cdot 24 \cdot 2240 = 1.00$ MEUR per month, where 2,240 MWh is the average load and $30 \cdot 24$ is the number of hours per month. Considering that the month-end P&Ls are in the range of -1 – 3 MEUR, (see bottom pane of Figure 4.14), this is quite a hefty premium. In terms of the forward prices, a margin of 0.62 will add about 1.5% to the price.

If we look at realised hourly returns instead, we have a reward-risk-ratio of 0.15 for the model hedge and -0.33 for the average-load. To achieve the same reward-to-risk for the latter, we have to add a margin of 0.72 EUR/MWh to the contract price.

Remark 4.4. In Remark 4.3 we noted that a specification with zero cross-dependency between the deseasonalized processes could potentially perform as good as the original specification. If we repeat our hedging experiment with the independent specification (employing the closed-form expression for the expected positive loss), we obtain an average hourly payoff 742 EUR and a standard deviation 5,259. This gives a reward-to-risk ratio of 0.14, to be compared with 0.15 for the dependent specification, while the out-performance probability is 65.5%.

At last, if we repeat our empirical experiment with the proprietary set of consumption data from DONG Energy (Figure 4.17 in the appendix) we obtain comparable results, as can be seen from Table 4.19 in the appendix. For realised hourly payouts, the model hedge outperforms with a 70% probability and if we look at reward-to-risk ratios, we have 0.14 for the model hedge and -0.30 for the average load which corresponds to a margin equivalent of 0.65 EUR/MWh. Finally, for every single one-month period, we have that the

model based strategy outperforms the average-load strategy in terms of a larger realised cumulative profit-and-loss at month-end (see Figure 4.18 in the appendix).

To conclude: if a minimal variance makes you sleep better at night, chose the average-load strategy. But if you don't want to charge a solid safety premium, you might as well loosing it all in the end.

Contract	$V^{\mathrm{b}*}$	$V^{\mathrm{p}*}$	Payoff	P(loss)	$\bar{V}^{\mathrm{b}}$	$\bar{V}^{\mathrm{p}}$	Payoff	P(loss)
Jan -13	0.361	0.34	0.89(2.84)	0.23	0.632	0.222	-0.608(1.33)	0.63
Feb -13	0.388	0.354	1.296(1.33)	0.08	0.618	0.21	-0.204(0.74)	0.56
Mar -13	0.512	0.363	0.146(1.63)	0.41	0.607	0.169	-0.265(1.11)	0.57
Apr -13	0.535	0.335	0.14(1.6)	0.48	0.541	0.167	-0.403(1.23)	0.61
May -13	0.585	0.301	0.05(1.3)	0.57	0.504	0.161	-0.269(0.51)	0.74
Jun -13	0.637	0.271	0.097(2.08)	0.48	0.505	0.181	-0.271(0.74)	0.67
Jul -13	0.641	0.26	0.729(2.06)	0.39	0.484	0.149	-0.088(0.47)	0.6
Aug -13	0.486	0.337	0.093(1.23)	0.51	0.505	0.182	-0.338(0.69)	0.7
Sep -13	0.438	0.31	-0.165(1.57)	0.37	0.52	0.188	-0.488(1.04)	0.72
Oct -13	0.443	0.294	0.209(1.21)	0.34	0.544	0.201	-0.29(0.95)	0.61
Nov -13	0.418	0.294	0.512(1.33)	0.18	0.579	0.227	-0.365(0.96)	0.63
Dec -13	0.441	0.3	0.584(1.78)	0.25	0.587	0.202	-0.642(1.51)	0.6
Jan -14	0.485	0.321	0.37(1.15)	0.25	0.62	0.231	-0.487(1.21)	0.56
Feb -14	0.486	0.345	0.553(1.02)	0.27	0.605	0.201	-0.233(0.87)	0.52
Mar -14	0.608	0.299	-0.394(1.09)	0.76	0.567	0.199	-0.203(0.76)	0.61
Apr -14	0.528	0.324	-0.352(1.24)	0.66	0.526	0.179	-0.305(0.79)	0.64
May -14	0.506	0.335	0.133(1.1)	0.53	0.507	0.187	-0.278(0.67)	0.68
Jun -14	0.493	0.338	0.023(0.83)	0.51	0.519	0.199	-0.273(0.62)	0.68
Jul -14	0.48	0.353	0.056(0.56)	0.55	0.485	0.174	-0.108(0.32)	0.61
Aug -14	0.407	0.351	0.24(0.95)	0.3	0.51	0.202	-0.127(0.61)	0.6
Sep -14	0.401	0.308	-0.135(0.8)	0.44	0.519	0.201	-0.217(0.49)	0.71
Oct -14	0.418	0.307	0.419(1.32)	0.27	0.545	0.213	-0.151(0.87)	0.59
Nov -14	0.453	0.319	0.168(1.19)	0.29	0.582	0.213	-0.301(0.93)	0.6
Dec -14	0.489	0.275	-0.108(1.75)	0.61	0.599	0.187	-0.445(1.27)	0.69

Table 4.13: Positions for the peak- and base forwards for optimal model hedging of the fixed price agreements (measured in percentage of maximum load). The monthly average (standard deviation) of the hourly payoff (measured in EUR/maximum load) and the probability of loss. The average-load hedge positions, $(\bar{V}^{\mathrm{b}}, \bar{V}^{\mathrm{p}})$, with corresponding average hourly payoff and loss probability are also included in the table.

4.5 Conclusion

The focus of this paper was to formulate a hedging strategy for a fixed price, unspecified consumption contract. For this purpose, we considered the expected positive loss as our risk measure under a joint model for the spot-price and consumption load. We derived a semi-closed formula for this model which we minimised numerically to find the optimal

hedge position. In an empirical experiment we compared our model-hedge strategy to an average-load strategy that is commonly used by the industry. For our set of market data, we showed that the suggested model hedge outperformed the average-load strategy to the extent that we have to add a significant premium to the latter in order for the two to be financially equal. If we looked at the realised hourly payouts, we found that the model hedge yields a relatively larger outcome with a 66% probability. Further, for all but two single one-month periods of the considered market data, we showed that the model based strategy outperformed the average-load in terms of a greater realised cumulative profit-and-loss at month-end. At the very end, we repeated the hedging experiment on a proprietary set of load data from DONG Energy where we obtained even stronger results than for the publicly available data.

Appendix

Notation time	Expiry	Base forward	Peak forward	Base CFD
2012-12-17	January-13	43.30	52.00	1.25
2013-01-17	February-13	45.30	52.35	0.60
2013-02-14	March-13	38.10	39.00	-0.20
2013-03-18	April-13	40.55	43.05	-3.40
2013-04-16	May-13	38.10	37.55	-1.50
2013-05-16	June-13	34.70	36.00	-1.15
2013-06-17	July-13	28.30	30.70	4.60
2013-07-17	August-13	36.95	40.00	1.95
2013-08-16	September-13	36.50	39.40	3.70
2013-09-16	October-13	38.90	44.00	1.30
2013-10-17	November-13	41.00	47.00	-0.70
2013-11-18	December-13	41.45	43.50	-4.05
2013-12-17	January-14	35.90	41.00	0.10
2014-01-17	February-14	36.30	41.30	-0.28
2014-02-17	March-14	29.00	33.05	0.90
2014-03-17	April-14	25.60	29.63	5.55
2014-04-16	May-14	24.50	28.50	5.85
2014-05-16	June-14	25.30	29.35	4.80
2014-06-17	July-14	22.99	25.80	7.80
2014-07-17	August-14	31.25	34.15	3.00
2014-08-18	September-14	32.70	35.70	2.80
2014-09-17	October-14	34.75	37.30	-0.15
2014-10-17	November-14	31.95	34.00	0.65
2014-11-17	December-14	33.00	34.80	-2.75

Table 4.15: Nord-Pool system- and CFD forward prices (in EUR) for base- and peak contracts.

Contract	κ_S	θ_S	σ_S	κ_L	θ_L	σ_L	ρ_W
Jan-13	0.103 (0.007)	0.000 (0.792)	5.42 (0.061)	0.065 (0.005)	0.000 (0.006)	0.025 (0.00027)	0.368 (0.013)
Feb-13	0.114 (0.007)	0.001 (0.896)	6.77 (0.076)	0.065 (0.005)	0.000 (0.006)	0.026 (0.00028)	0.291 (0.014)
Mar-13	0.121 (0.008)	0.000 (0.814)	6.53 (0.074)	0.093 (0.007)	0.000 (0.004)	0.027 (0.00004)	0.303 (0.014)
Apr-13	0.125 (0.008)	0.000 (0.79)	6.51 (0.073)	0.132 (0.008)	0.000 (0.003)	0.028 (0.00032)	0.31 (0.014)
May-13	0.107 (0.007)	0.000 (0.907)	6.42 (0.071)	0.133 (0.008)	0.000 (0.003)	0.028 (0.00032)	0.303 (0.014)
Jun-13	0.103 (0.007)	0.000 (0.917)	6.23 (0.069)	0.08 (0.006)	0.000 (0.005)	0.027 (0.00029)	0.311 (0.014)
Jul-13	0.099 (0.007)	0.000 (0.892)	5.82 (0.065)	0.06 (0.005)	-0.001 (0.006)	0.025 (0.00027)	0.259 (0.014)
Aug-13	0.086 (0.006)	0.000 (0.67)	3.82 (0.042)	0.053 (0.005)	-0.001 (0.007)	0.024 (0.00026)	0.348 (0.013)
Sep-13	0.085 (0.006)	0.000 (0.633)	3.56 (0.039)	0.055 (0.005)	0.000 (0.006)	0.021 (0.00022)	0.327 (0.014)
Oct-13	0.089 (0.006)	0.084 (0.608)	3.58 (0.04)	0.085 (0.006)	0.000 (0.003)	0.019 (0.00021)	0.285 (0.014)
Nov-13	0.091 (0.007)	0.000 (0.611)	3.69 (0.041)	0.148 (0.009)	0.000 (0.002)	0.02 (0.00023)	0.33 (0.014)
Dec-13	0.096 (0.007)	0.000 (0.638)	4.03 (0.045)	0.109 (0.007)	0.000 (0.003)	0.023 (0.00025)	0.324 (0.014)
Jan-14	0.110 (0.007)	-0.019 (0.616)	4.5 (0.051)	0.068 (0.006)	0.000 (0.005)	0.024 (0.00027)	0.32 (0.014)
Feb-14	0.103 (0.007)	-0.001 (0.716)	4.86 (0.055)	0.065 (0.006)	0.000 (0.006)	0.025 (0.00028)	0.28 (0.014)
Mar-14	0.096 (0.007)	-0.001 (0.774)	4.93 (0.056)	0.079 (0.007)	0.000 (0.005)	0.027 (0.00029)	0.262 (0.014)
Apr-14	0.106 (0.007)	0.000 (0.712)	4.98 (0.056)	0.113 (0.008)	0.000 (0.004)	0.028 (0.00031)	0.255 (0.014)
May-14	0.123 (0.008)	-0.001 (0.578)	4.69 (0.053)	0.143 (0.009)	0.000 (0.003)	0.028 (0.00031)	0.255 (0.014)
Jun-14	0.124 (0.008)	-0.001 (0.548)	4.49 (0.051)	0.098 (0.007)	0.000 (0.004)	0.026 (0.00029)	0.256 (0.014)
Jul-14	0.109 (0.007)	0.000 (0.558)	4.04 (0.046)	0.078 (0.006)	-0.001 (0.005)	0.024 (0.00027)	0.222 (0.015)
Aug-14	0.108 (0.007)	0.000 (0.492)	3.51 (0.04)	0.062 (0.006)	0.000 (0.006)	0.023 (0.00025)	0.255 (0.014)
Sep-14	0.109 (0.008)	0.000 (0.521)	3.75 (0.042)	0.091 (0.007)	0.000 (0.003)	0.02 (0.00022)	0.219 (0.015)
Oct-14	0.105 (0.007)	0.018 (0.475)	3.31 (0.037)	0.127 (0.008)	0.000 (0.002)	0.019 (0.00021)	0.207 (0.015)
Nov-14	0.123 (0.008)	0.000 (0.412)	3.36 (0.038)	0.16 (0.009)	0.000 (0.002)	0.019 (0.00022)	0.259 (0.014)
Dec-14	0.100 (0.007)	-0.001 (0.535)	3.54 (0.04)	0.12 (0.008)	0.000 (0.003)	0.023 (0.00025)	0.214 (0.015)

Table 4.16: Estimated parameters and standard errors of the Ornstein-Uhlenbeck process from rescaled deseasonalized spot- and (scaled) load data.

Contract	$V^{\mathrm{b}*}$	$V^{\mathrm{p}*}$	Payoff	P(loss)	$\bar{V}^{\mathrm{b}}$	$\bar{V}^{\mathrm{p}}$	Payoff	P(loss)
Jan -13	0.32	0.274	0.623(1.98)	0.25	0.499	0.167	-0.371(1.07)	0.6
Feb -13	0.349	0.279	0.949(1.07)	0.1	0.512	0.169	-0.1(0.56)	0.53
Mar -13	0.431	0.305	0.134(1.32)	0.4	0.509	0.145	-0.206(0.96)	0.55
Apr -13	0.452	0.281	0.015(1.35)	0.46	0.492	0.142	-0.252(1.04)	0.56
May -13	0.481	0.266	0.074(0.84)	0.52	0.454	0.137	-0.208(0.47)	0.66
Jun -13	0.509	0.249	0.094(1.3)	0.51	0.458	0.156	-0.202(0.68)	0.65
Jul -13	0.534	0.216	0.52(1.45)	0.4	0.436	0.128	-0.044(0.47)	0.51
Aug -13	0.436	0.266	-0.016(0.92)	0.43	0.471	0.158	-0.301(0.69)	0.65
Sep -13	0.402	0.259	-0.144(1.39)	0.4	0.48	0.163	-0.355(0.94)	0.65
Oct -13	0.411	0.256	0.234(1)	0.32	0.493	0.163	-0.171(0.76)	0.59
Nov -13	0.391	0.26	0.449(0.94)	0.21	0.517	0.181	-0.232(0.73)	0.59
Dec -13	0.412	0.259	-0.042(1.38)	0.38	0.492	0.143	-0.666(1.49)	0.57
Jan -14	0.45	0.272	0.25(0.99)	0.26	0.555	0.18	-0.406(1.06)	0.54
Feb -14	0.447	0.284	0.442(0.71)	0.26	0.551	0.171	-0.258(0.81)	0.57
Mar -14	0.526	0.267	-0.17(0.8)	0.65	0.531	0.169	-0.177(0.72)	0.60
Apr -14	0.475	0.273	-0.249(1.21)	0.51	0.527	0.131	-0.398(0.96)	0.65
May -14	0.472	0.279	0.172(1.02)	0.51	0.48	0.131	-0.229(0.57)	0.66
Jun -14	0.463	0.278	0.066(0.73)	0.4	0.508	0.137	-0.218(0.54)	0.65
Jul -14	0.466	0.286	0.081(0.49)	0.52	0.471	0.118	-0.08(0.26)	0.62
Aug -14	0.407	0.277	0.235(0.68)	0.26	0.5	0.153	-0.092(0.37)	0.61
Sep -14	0.406	0.239	-0.093(0.73)	0.42	0.514	0.156	-0.145(0.44)	0.65
Oct -14	0.435	0.236	0.338(1.02)	0.29	0.523	0.158	-0.054(0.7)	0.58
Nov -14	0.453	0.251	0.048(0.91)	0.32	0.543	0.184	-0.265(0.71)	0.65
Dec -14	0.485	0.214	-0.211(1.31)	0.65	0.546	0.146	-0.435(1.16)	0.70

Table 4.19: Results from the hedging experiment based on the consumption data from DONG Energy. Forward positions in percentage of maximum load while the monthly average (standard deviation) of the hourly payoff is measured in EUR/maximum load.

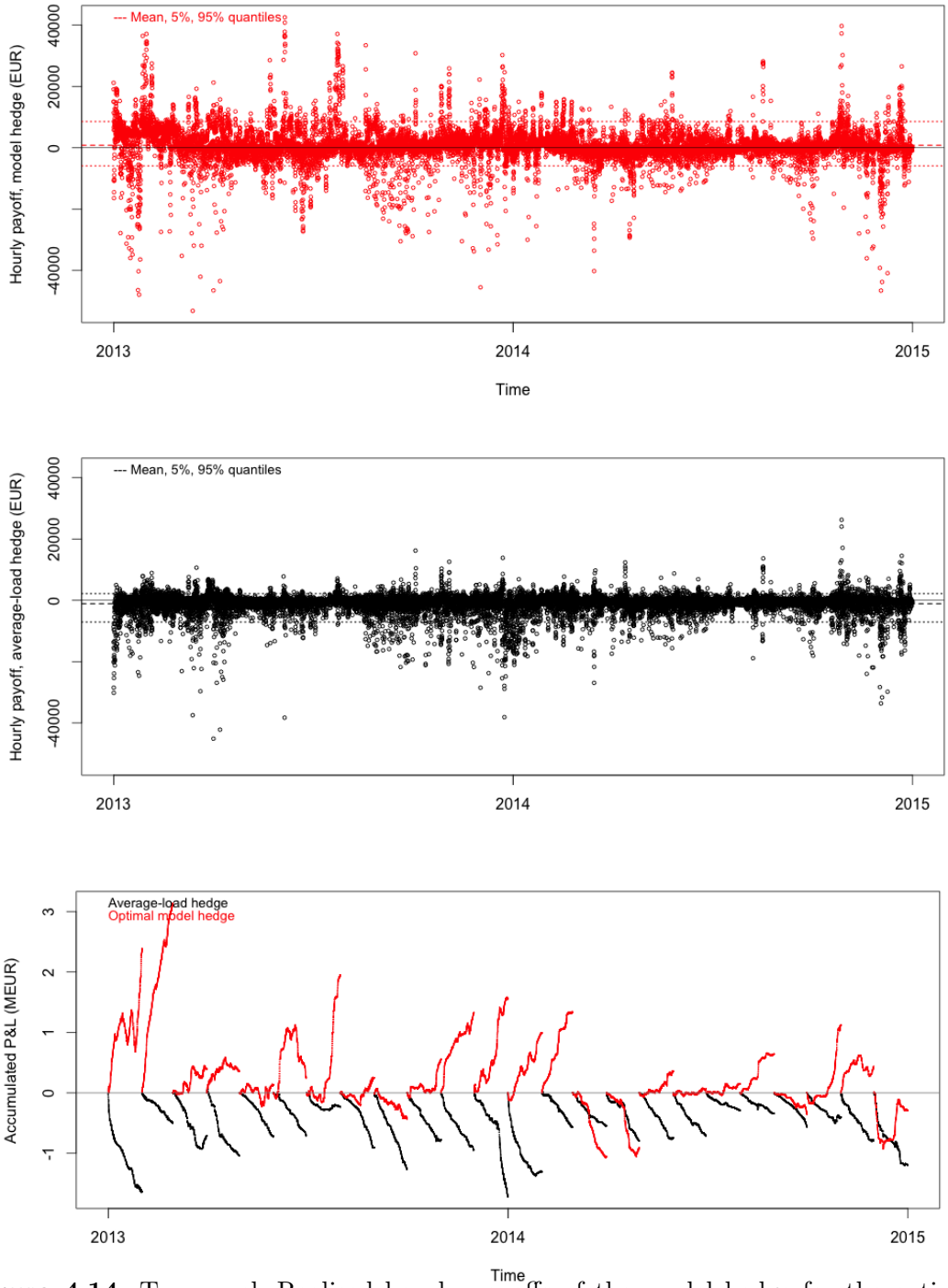

Figure 4.14: Top panel: Realised hourly payoffs of the model hedge for the entire set of contracts. Middle panel: Realised hourly payoffs for the average-load strategy. Bottom panel: accumulated monthly profit-&-loss in million euros of the strategies for all 24 contract. The model hedge (red) yields a higher month-end P&L for all but two contracts.

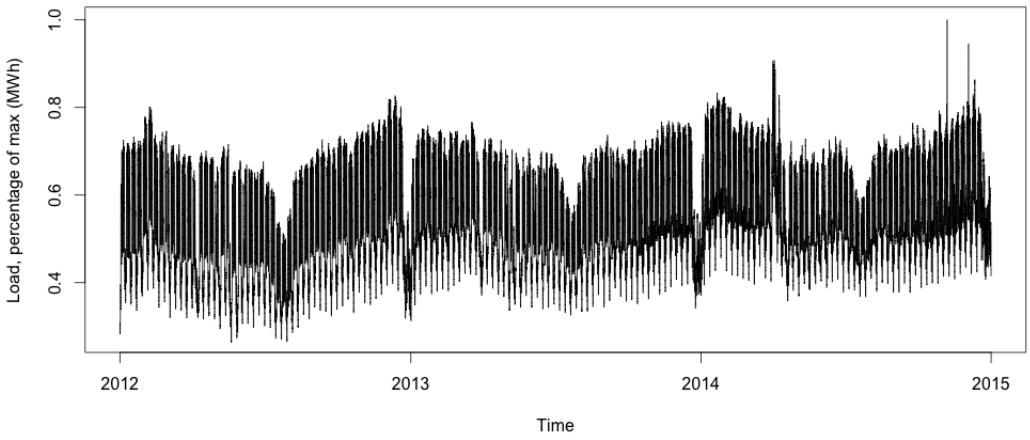

Figure 4.17: The consumption data from DONG Energy. Note that the consumption has been anonymized and rescaled with the maximum load to lie between zero and one. The data is recorded with an hourly frequency and covers the period from 1 January 2012 to 31 December 2014.

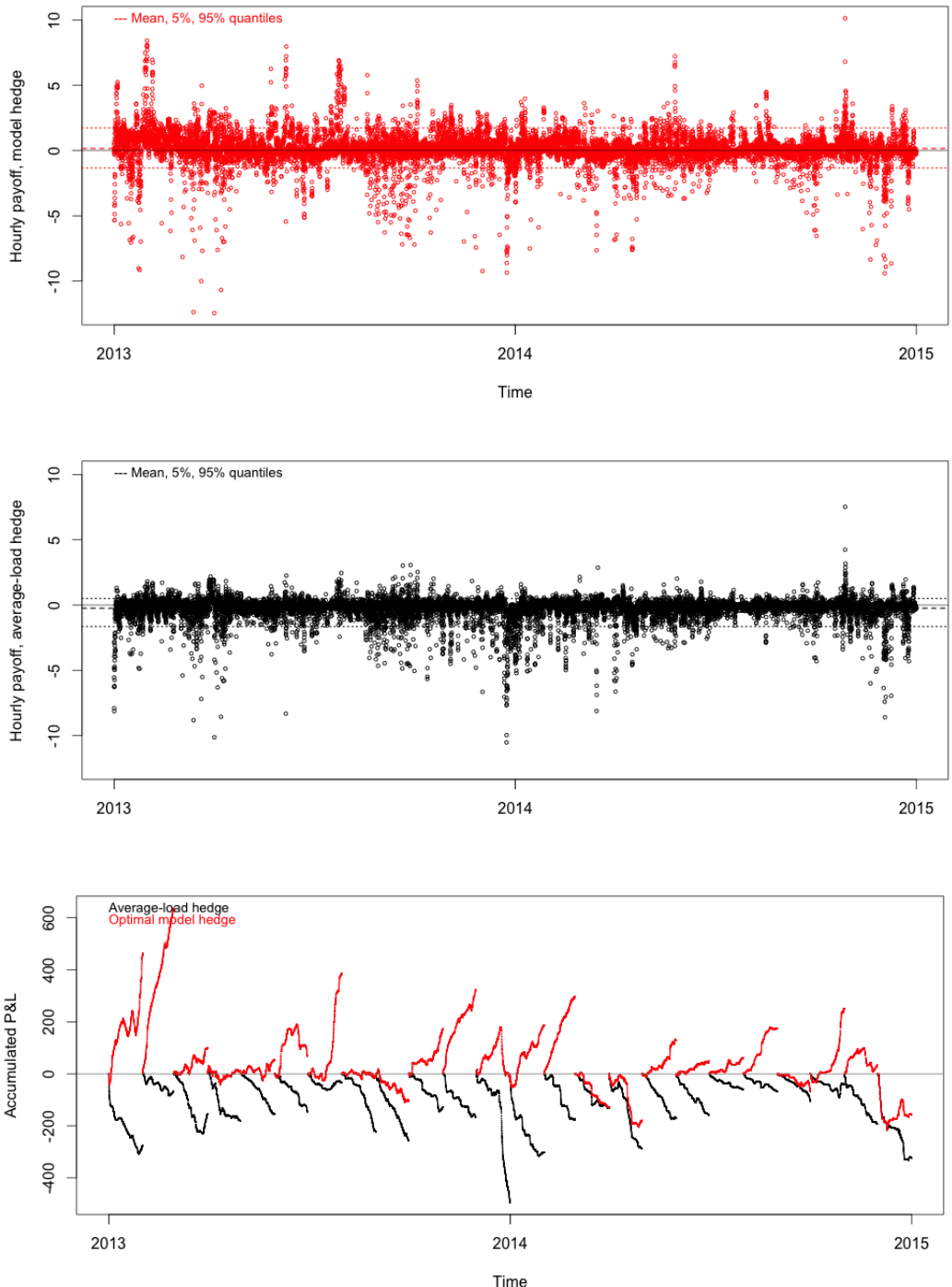

Figure 4.18: Results for the consumption data from DONG Energy. Top panel: Realised hourly payoffs of the model hedge. Middle panel: Realised hourly payoffs for the average-load strategy. Bottom panel: accumulated monthly profit-&-loss of the strategies for all 24 contract. The model hedge (red) yields a higher month-end P&L for every contract. All figures measured in EUR/maximum load.

5

Stochastic Volatility for Utility Maximisers – The Case of Derivatives

Simon Ellersgaard and Martin Jönsson[1]

Abstract. Using martingale methods we derive bequest optimising portfolio weights for a rational investor who trades in a bond-stock-derivative economy characterised by a generic stochastic volatility model. For illustrative purposes we then proceed to analyse the specific case of the Heston economy, which admits explicit expressions for plain vanilla Europeans options. By calibrating the model to market data, we find that the demand for derivatives is primarily driven by the myopic hedge component. Furthermore, upon deploying our optimal strategy on real market prices, we find only a very modest improvement in portfolio wealth over the corresponding strategy which only trades in bonds and stocks.

Keywords: Merton's Portfolio Problem, Stochastic Volatility, HJB Equation, Martingale Approach.

[1]Both authors are with the Department of Mathematical Sciences, University of Copenhagen.

5.1 Introduction

Regardless of whether we are dealing with the running variance associated with a financial time series, or the implied volatility surface extracted from traded option prices, one thing is abundantly clear: contrary to the assumption of the Black-Scholes-Merton model, volatility is far from constant. In fact, universally accepted stylised facts of the economy include the highly erratic nature of the variance process through time (see figure 5.3), or the skew/smile effect characteristic of the implied volatility surface - as reported in Cont and Tankov (2004).[2] Naturally, a plethora of possible resolutions to these effects have been proposed on the modelling front, most prominently *local volatility models* in which σ is a deterministic function of the random stock price, and *diffusion-based stochastic volatility models* in which σ is modelled directly as a stochastic differential equation. Both approaches must be considered significant steps towards designing calibratable models to observed market phenomena although neither can be said to be void of imperfections. However, the latter is arguably the more sophisticated of the two, being as it were more readily susceptible to theoretical augmentation. Derivatives pricing likewise becomes a matter of some interest: whilst local volatility models will have us believe that options are perfectly hedgeable using bonds and the underlying stocks (thereby making them formally redundant), this is not so for valuation under stochastic volatility models. Here, incompleteness (Björk (2009)) forces us to make further exogenous assumptions about the behaviour of the market in order to pin down our risk neutral measure, $\mathbb{Q}$. Specifically, to value one option, enough similar traded options must already exist on the market, in order for us to say anything concrete (in somewhat more abstract terms: a supply-and-demand induced market price of risk must prevail).

Surprisingly, while derivative pricing and calibration in connexion with stochastic volatility constitute major research areas in the quant-finance community, relatively few papers deal with the impact of stochastic volatility on portfolio optimisation. In fact, to the best of our knowledge, the first authors to deal explicitly with the issue are Jun Liu and Jun Pan, a little more than a decade ago. The more pedestrian of their analyses is found in Liu (2007), in which bequest optimisation in a Heston-driven[3] bond-stock economy is used to illustrate a grander theoretical point about solutions to HJB equations. Briefly, under the assumption that the market price of risk is proportional to volatility, $\lambda_1 = \bar{\lambda}_1\sigma$, Liu shows that the optimal portfolio weight to be placed on the stock by a rational CRRA investor is

$$\pi_{S,t}^{Liu} = \frac{\bar{\lambda}_1}{\gamma} - \rho\sigma_v\frac{\gamma-1}{\gamma}L(T-t), \tag{5.1}$$

where $\bar{\lambda}_1/\gamma$ is the Merton (1969) ratio, and the second term is a stoch-vol *hedge correction*

[2] In the words of Cont and Tankov, Chapter 1: "For equity and foreign exchange options, implied volatilities $\sigma_t(T,K)$ display a strong dependence with respect to the strike price: this dependence may be decreasing ("skew") or U-shaped ("smile") and has greatly increased since the 1987 crash".

[3] See Heston (1993) and Section 5.4 for an exposition of this model.

(Munk (2013)) in which L is the deterministic function

$$L(\tau) = \frac{\bar{\lambda}_1^2}{\gamma} \frac{(e^{\eta\tau} - 1)}{(\epsilon + \eta)(e^{\eta\tau} - 1) + 2\eta},$$

where we have defined the parameters $\epsilon \equiv \kappa + \frac{\gamma-1}{\gamma}\rho\sigma_v\bar{\lambda}_1$ and

$$\eta \equiv \sqrt{\epsilon^2 + \frac{\gamma - 1}{\gamma^2}\sigma_v^2 \left(\rho^2 + \gamma[1 - \rho^2]\right) \bar{\lambda}_1^2}.$$

Throughout this paper, we refer to this result as Liu's strategy. Little is said by Liu on the empirical implications, yet it is well-known that the correction hedge is negligible. For example, our own investigation Ellersgaard and Jönsson (2013) reveals that the hedge correction is multiple orders of magnitude smaller than the Merton weight for realistic parameter specifications, thus leading to non-measurable improvements in the investor's welfare.

A far richer theoretical account of the role of stochastic volatility in portfolio maximisation is provided in Liu and Pan (2003), who extend the above framework to include jumps in the underlying price process, and complete the market by including tradeable derivative securities (specifically, a straddle, chosen for its sensitivity to volatility risk). By solving the relevant HJB equation optimal portfolio weights are provided (in terms of certain partial derivatives); moreover, through the employ of market calibrated parameters Liu and Pan estimate that the primary demand for derivatives is nested in the myopic component of the portfolio weight (rather than the volatility hedge correction). Based on the same parameters they also establish significant improvements in *certainty equivalent wealth* through the act of including derivatives in a utility maximised portfolio.

Overview. This paper might be read as a quasi-exposition of the work above (in the sense that we obtain similar formulae albeit using different techniques) with elements of novelty. Specifically, through martingale considerations we establish the optimal investment strategy for a fairly generic stochastic volatility model. These formulas are instantiations of more general state-variable expressions found in e.g. Munk (2013), but we expose them as (i) there is some pedagogical value in seeing how they can be derived in a martingale framework, (ii) they readily can be adapted to more exotic volatility models. Upon specialising to the Heston model, we then proceed to find explicit expressions in the event the derivative is a plain vanilla call or put option. Vis-a-vis the straddle strategy mentioned above this is a minor variation, yet the value of our analysis lies in its attention to detail, both from a conceptual and computational point of view. In either case, the change to the optimal investment plan is considerable with respect to Liu's strategy (5.1) for calibrated market parameters. Furthermore, in corroboration of the findings by Liu and Pan we find through Monte Carlo simulation that the hedge component specific to stochastic volatility is irrelevant. Finally, in a novel portfolio rebalancing experiment using real market prices, we show that access to plain vanillas for utility optimisers create only a very modest (dubious) improvement expressed in terms of a certainty equivalent in wealth, in discordance with the quasi-empirical musings of Liu and Pan.

5.2 Problem set-up

5.2.1 Market assumptions

Following the path betrodden by Black, Scholes, and Merton we start out by considering a financial landscape which is frictionless, arbitrage free, and allows for continuous trading. Three assets which jointly complete the market are assumed to prevail, viz. a risk-free money account (a *bond*), one fundamental risky security (a *stock*), and one derivative security with a European exercise feature at time $t = T'$. As it is commonplace, we define the dynamical equations of these securities by first introducing the filtered probability space $(\Omega, \mathscr{F}, \mathbb{F} = \{\mathscr{F}_t\}_{t\in[0,T']}, \mathbb{P})$, where Ω represents all possible states of the economy, $\mathbb{P}$ is the real-world probability measure, and $\mathscr{F}_t$ is the augmented natural filtration of two independent Wiener processes W_1 and W_2:

$$\mathscr{F}_t = \sigma(\mathscr{F}_t^{\boldsymbol{W}} \cup \mathcal{N}),$$

where $\mathscr{F}_t^{\boldsymbol{W}} = \sigma(\{W_{1s}, W_{2s}\}_{s\in[0,t]})$ and $\mathcal{N}$ represents the sets of $\mathbb{P}$-null events. With this in mind, we specify the price process dynamics of the money account $\{B_t\}_{t\in[0,T']}$ as the deterministic equation,

$$dB_t = rB_t dt, \tag{5.2}$$

where $B_0 = b_0 \in \mathbb{R}^+$ and r is the constant rate of interest. As for the fundamental risky security $\{S_t\}_{t\in[0,T']}$ we posit an SDE model of a rather generic stochastic volatility form, viz.

$$\begin{aligned} dS_t &= \mu_S(t, V_t)S_t dt + \sqrt{V_t}S_t dW_{1t}, \\ dV_t &= \alpha(t, V_t)dt + \beta(t, V_t)(\rho dW_{1t} + \sqrt{1-\rho^2}dW_{2t}), \end{aligned} \tag{5.3}$$

where $(S_0, V_0) = (s_0, v_0) \in \mathbb{R}^{2+}$, and $\{V_t\}_{t\in[0,T']}$ is the variance process which we assume strictly positive. As for the dynamical constituents: μ_S, α, and β are taken to be real valued deterministic functions $[0, T'] \times \mathbb{R}^+ \mapsto \mathbb{R}$, whilst $\rho = \mathbb{C}\text{orr}[dS_t, dV_t] \in (-1, 1)$, is a Pearson correlation coefficient between the stock and variance processes. Finally, as for the European derivative, we envision a one-time pay-off $\Phi(S_{T'})$ based on the magnitude of the contemporaneous stock value at expiry time T'. Letting $\{D_t\}_{t\in[0,T']} = \{D(t, S_t, V_t)\}_{t\in[0,T']}$ represent the price process of the derivative, it follows from Itô's lemma that

$$dD_t = \mu_D(t, S_t, V_t)D_t dt + \sigma_{1D}(t, S_t, V_t)D_t dW_{1t} + \sigma_{2D}(t, S_t, V_t)D_t dW_{2t}, \tag{5.4}$$

where $D_{T'} = \Phi_{T'}$, and $\mu_D, \sigma_{1D}, \sigma_{2D} : [0, T'] \times \mathbb{R}^+ \times \mathbb{R}^+ \mapsto \mathbb{R}$ are the functions

$$\begin{aligned} \mu_D(t, s, v) &\equiv D^{-1}\Big[\partial_t D + \mu_S(t, v)s\partial_s D + \alpha(t, v)\partial_v D + \tfrac{1}{2}vs^2\partial_{ss}^2 D \\ &\qquad + \tfrac{1}{2}\beta^2(t, v)\partial_{vv}^2 D + \rho\beta(t, v)\sqrt{v}s\partial_{sv}^2 D\Big], \\ \sigma_{1D}(t, s, v) &\equiv D^{-1}\Big[\rho\beta(t, v)\partial_v D + \sqrt{v}s\partial_s D\Big], \\ \sigma_{2D}(t, s, v) &\equiv D^{-1}\Big[\sqrt{1-\rho^2}\beta(t, v)\partial_v D\Big], \end{aligned} \tag{5.5}$$

assuming, of course, that $D \in \mathcal{C}^{1,2,2}$.

Crucial to our derivations in the subsequent sections, we now enforce the following minimal structure upon the *aggregate* risk preferences of agents trading in our tripartite economy:

Assumption 5.1. The market prices of risk λ_1 and λ_2 associated with W_1 and W_2 are functions of v **only**. In concrete terms this means that

$$\lambda_1(v) = \frac{\mu_S(t,v) - r}{\sqrt{v}}, \tag{5.6}$$

and

$$\lambda_2(v) = \frac{\mu_D(t,s,v) - r}{\sigma_{2D}(t,s,v)} - \frac{\sigma_{1D}(t,s,v)}{\sigma_{2D}(t,s,v)}\lambda_1(v). \tag{5.7}$$

We call this the **weak Heston assumption** for reasons which will become clearer below.[4]

Now, from (5.6) we may define the risk-neutral measure $\mathbb{Q}$ on $\mathcal{F}_{T'}$ through the Radon-Nikodym derivative

$$\left.\frac{d\mathbb{Q}}{d\mathbb{P}}\right|_{\mathcal{F}_{T'}} \equiv \xi_{T'} \equiv \exp\left\{ -\tfrac{1}{2}\int_0^{T'} \sum_{i=1}^{2} \lambda_i^2(V_t)dt - \int_0^{T'} \sum_{i=1}^{2} \lambda_i(V_t)dW_{it} \right\}. \tag{5.8}$$

Assuming the Novikov condition, $\mathbb{E}[\exp\{\frac{1}{2}\int_0^{T'}(\lambda_1^2(V_t) + \lambda_1^2(V_t))dt\}] < \infty$, then $\mathbb{E}[\xi_{T'}] = 1$, whence $\mathbb{Q}$ is an equivalent local martingale measure (ELMM). All discounted asset prices under $\mathbb{Q}$ are therefore local martingales, which can be verified by combining Girsanov's transformation

$$dW_{it} = -\lambda_i(V_t)dt + dW_{it}^{\mathbb{Q}},$$

for $i = 1, 2$ with the price dynamics (5.3) and (5.4). Finally, upon combining the market price of risk (5.7) with the Itô expressions (5.5) we readily find that the partial differential equation governing the price of the derivative is of the form

$$\begin{aligned} 0 = {} & \partial_t D + rs\partial_s D + \{\alpha(t,v) - \beta(t,v)[\rho\lambda_1(v) + \sqrt{1-\rho^2}\lambda_2(v)]\}\partial_v D \\ & + \tfrac{1}{2}vs^2\partial_{ss}^2 D + \tfrac{1}{2}\beta^2(t,v)\partial_{vv}^2 D + \rho\beta(t,v)\sqrt{v}s\partial_{sv}^2 D - rD, \end{aligned} \tag{5.9}$$

subject to the terminal condition $D(T', s, v) = \Phi(s)$.

Remark 5.1. We assume the Novikov condition in order to establish the existence of the ELMM. Rather remarkably, existence is something all too often glossed over in the stochastic volatility literature. Indeed, there are somewhat spectacular examples of stochastic volatility models where the no-arbitrage condition generally breaks down cf. e.g. the Stein

[4] We emphasise that this is not a vacuous statement: specifically, the weak Heston assumption is not a *gauge freedom*, as it invariably does say *something* about supply and demand in the market. On the other hand, it is not an approximation either: clearly, we have the mathematical freedom to suppose whatever we want here.

& Stein model. Furthermore, it is reasonable to show formally that discounted asset prices are *true martingales* as opposed to strictly local ones. Although failure of the true martingale property does not entail arbitrage, it can lead to peculiarities ("bubble problems") such as the breakdown of put-call parity. For examples and general theory pertaining to these fascinating issues we refer the reader to Wong and Heyde (2006).

5.2.2 Investor assumptions

We consider the case of an investor who trades in the three asset classes in a self-financing manner over the temporal horizon $[0,T] \subseteq [0,T']$, with the intention of maximising the expected discounted utility from her terminal wealth, $\mathscr{W}_T$. Specifically, we are interested in determining the optimal portfolio weights $\pi^*_{S,t}$ and $\pi^*_{D,t}$ which the investor should place on the stock and the derivative[5] such that

$$\{\pi^*_{S,t}, \pi^*_{D,t}\}_{t\in[0,T]} = \underset{\{\pi_S,\pi_D\}\in\mathscr{L}^2[0,T]}{\operatorname{argmax}} \mathbb{E}[e^{-\delta T} u(\mathscr{W}_T)], \tag{5.10}$$

where $\delta \in \mathbb{R}_+$ is a subjective discounting factor, and $u : \mathbb{R}^+ \mapsto \mathbb{R}$ is a utility function which we assume isoelastic, i.e.

$$u(x) = \frac{x^{1-\gamma}}{1-\gamma},$$

where $\gamma > 1$ codifies the investor's level of risk aversion. No restrictions on short-selling and leveraging are enforced upon the portfolio weights. However, to rule out arbitrage through doubling-strategies we assume that the weights belong to the space of square-integrable processes, which we have denoted by $\mathscr{L}^2$.

Note that from the self-financing condition it follows that the optimal wealth process $\{\mathscr{W}^*_t\}_{t\in[0,T]}$ obeys the dynamics

$$\begin{aligned} d\mathscr{W}^*_t = &[r + \pi^*_{S,t}\sqrt{V_t}\lambda_1(V_t) + \pi^*_{D,t}(\sigma_{1D}\lambda_1(V_t) + \sigma_{2D}\lambda_2(V_t))]\mathscr{W}^*_t dt \\ &+ [\pi^*_{S,t}\sqrt{V_t} + \pi^*_{D,t}\sigma_{1D}]\mathscr{W}^*_t dW_{1t} + \pi^*_{D,t}\sigma_{2D}\mathscr{W}^*_t dW_{2t}, \end{aligned} \tag{5.11}$$

which can be used to set up a Hamilton-Jacobi-Bellman equation for the value function associated with (5.10). This, traditional approach is nonetheless not the route by which we shall be proceeding: rather, we opt for a martingale theoretic approach, which is well-equipped to handle bequest-optimisation problems in complete financial markets.

[5]We assume that the remaining fraction of the wealth $1 - \pi_{S,t} - \pi_{D,t}$ is allocated to the risk free money account.

5.3 The martingale solution

5.3.1 The optimal wealth process

It is well-known that the dynamic programming problem highlighted above may be reformulated as a static optimisation problem by solving for the optimal wealth process, whence the optimal portfolio weights can be deduced, Björk (2009). Specifically, we are scrutinising the optimisation problem

$$\mathscr{W}_T^* = \underset{\mathscr{W}_T \in \mathscr{K}_T}{\operatorname{argmax}} \, \mathbb{E}[e^{-\delta T} u(\mathscr{W}_T)], \tag{5.12}$$

over the class of adapted self-financing portfolios, $\mathscr{K}_T$, which is to say subject to the static *budget constraint*,

$$w_0 = \mathbb{E}^{\mathbb{Q}}[e^{-rT}\mathscr{W}_T],$$

where $\mathscr{W}_0 = w_0$.[6] In Lagrangian terms, we are accordingly dealing with

$$\mathfrak{L} = \mathbb{E}[e^{-\delta T} u(\mathscr{W}_T) - \eta \xi_T e^{-rT}\mathscr{W}_T], \tag{5.13}$$

where η is the Lagrange multiplier, and $\xi_T = \mathbb{E}[\xi_{T'}|\mathcal{F}_T]$ is the Radon-Nikodym derivative defined in (5.8), here introduced to write the entire expectation under the $\mathbb{P}$-measure. By differentiating partially with respect to $\mathscr{W}_T$ and equating to zero, we may extract the optimal terminal wealth

$$\mathscr{W}_T^* = (u')^{-1}(\eta e^{(\delta - r)T}\xi_T),$$

where $(u')^{-1}(\cdot) = (\cdot)^{-1/\gamma}$ is the inverse marginal utility, i.e.

$$\mathscr{W}_T^* = \eta^{-1/\gamma} e^{-qT} \xi_T^{-1/\gamma},$$

where $q \equiv (\delta - r)/\gamma$. To determine the multiplier η we combine this expression with the $\mathbb{P}$-budget constraint

$$w_0 = \mathbb{E}[e^{-rT}\xi_T \mathscr{W}_T^*],$$

to get

$$\eta^{-1/\gamma} = \frac{w_0}{\mathbb{E}[e^{-(r+q)T}\xi_T^{1-1/\gamma}]}.$$

Thus,

$$\mathscr{W}_T^* = \frac{w_0 e^{rT}\xi_T^{-1/\gamma}}{\mathbb{E}[\xi_T^{1-1/\gamma}]}. \tag{5.14}$$

Now, consider the denominator $\mathbb{E}[\xi_T^{1-1/\gamma}]$. From (5.8) it is not hard to see that this *almost* looks like a $\mathbb{P}$-expectation of a Radon-Nikodym ξ^0 defined as:

$$\left.\frac{d\mathbb{Q}_0}{d\mathbb{P}}\right|_T \equiv \xi_T^0 \equiv \exp\left\{ -\tfrac{1}{2}(1-1/\gamma)^2 \int_0^T \sum_{i=1}^2 \lambda_i^2(V_t)dt - (1-1/\gamma)\int_0^T \sum_{i=1}^2 \lambda_i(V_t)dW_{it} \right\}. \tag{5.15}$$

[6] For a formal proof of this result, see Munk (2013).

In fact, one may readily check that $\xi_T^{1-1/\gamma}$ and ξ_T^0 are related through

$$\xi_T^{1-1/\gamma} = \xi_T^0 \exp\left\{\frac{1-\gamma}{2\gamma^2}\int_0^T \sum_{i=1}^2 \lambda_i^2(V_t)dt\right\}, \tag{5.16}$$

whence

$$\mathbb{E}[\xi_T^{1-1/\gamma}] = \mathbb{E}^{\mathbb{Q}_0}\left[\exp\left\{\frac{1-\gamma}{2\gamma^2}\int_0^T \sum_{i=1}^2 \lambda_i^2(V_t)dt\right\}\right]. \tag{5.17}$$

The explicit dependence on the Wiener increments has thus been suppressed through a second change of measure. This expectation is sufficiently important to what follows that we are prompted to introduce the function $H : [0,T] \times \mathbb{R}^+ \mapsto \mathbb{R}$:

$$H_t = H(t,v) = \mathbb{E}_{t,v}^{\mathbb{Q}_0}\left[\exp\left\{\frac{1-\gamma}{2\gamma^2}\int_t^T \sum_{i=1}^2 \lambda_i^2(V_s)ds\right\}\right]. \tag{5.18}$$

To see how this comes in handy, let us determine the optimal wealth process $\mathscr{W}_t^*$ for all times $t \in [0,T]$. From the budget constraint

$$\begin{aligned}
\mathscr{W}_t^* &= \mathbb{E}_{t,v}^{\mathbb{Q}}[e^{-r(T-t)}\mathscr{W}_T^*] \\
&= e^{-r(T-t)}\tfrac{1}{\xi_t}\mathbb{E}_{t,v}[\xi_T\mathscr{W}_T^*] \\
&= e^{-r(T-t)}\tfrac{1}{\xi_t}\tfrac{w_0 e^{rT}}{H_0}\mathbb{E}_{t,v}[\xi_T^{1-1/\gamma}] \\
&= \tfrac{e^{rt}w_0}{\xi_t H_0}\mathbb{E}_{t,v}[\xi_T^0 \exp\{\tfrac{1-\gamma}{2\gamma^2}\textstyle\int_0^T(\lambda_1^2(V_s)+\lambda_2^2(V_s))ds\}] \\
&= \tfrac{e^{rt}w_0}{\xi_t H_0}\xi_t^0\mathbb{E}_{t,v}^{\mathbb{Q}_0}[\exp\{\tfrac{1-\gamma}{2\gamma^2}\textstyle\int_0^T(\lambda_1^2(V_s)+\lambda_2^2(V_s))ds\}] \\
&= \tfrac{e^{rt}w_0}{\xi_t H_0}\xi_t^0 \exp\{\tfrac{1-\gamma}{2\gamma^2}\textstyle\int_0^t(\lambda_1^2(V_s)+\lambda_2^2(V_s))ds\}\mathbb{E}_{t,v}^{\mathbb{Q}_0}[\exp\{\tfrac{1-\gamma}{2\gamma^2}\textstyle\int_t^T(\lambda_1^2(V_s)+\lambda_2^2(V_s))ds\}] \\
&= \tfrac{e^{rt}w_0}{\xi_t H_0}\xi_t^{1-1/\gamma}\mathbb{E}_{t,v}^{\mathbb{Q}_0}[\exp\{\tfrac{1-\gamma}{2\gamma^2}\textstyle\int_t^T(\lambda_1^2(V_s)+\lambda_2^2(V_s))ds\}],
\end{aligned}$$

where the second line uses the abstract Bayes' formula (Björk (2009), proposition B.41), the third line uses the optimal wealth expression (5.14), the fourth line uses the identity (5.16), the fifth line uses (5.17), the sixth line splits the integral at the point of measurability $\mathscr{F}_t$, and the final line uses (5.16) again. Hence, from the definition of the H function (5.18) we find that the optimal wealth process can be written as

$$\mathscr{W}_t^* = e^{rt}w_0\frac{H_t}{H_0}\xi_t^{-1/\gamma}. \tag{5.19}$$

5.3.2 Notes on the H-function

Since the H-function (5.18) is assumed to be a function of t and V_t it follows from Itô's lemma and the dynamics (5.3) that

$$dH_t = \mu_H(t,V_t)H_t dt + \sigma_{1H}(t,V_t)H_t dW_{1t} + \sigma_{2H}(t,V_t)H_t dW_{2t}, \tag{5.20}$$

where

$$\begin{aligned}\mu_H(t,v) &\equiv H^{-1}\Big[\partial_t H + \alpha(t,v)\partial_v H + \tfrac{1}{2}\beta^2(t,v)\partial^2_{vv}H\Big],\\ \sigma_{1H}(t,v) &\equiv H^{-1}\rho\beta(t,v)\partial_v H,\\ \sigma_{2H}(t,v) &\equiv H^{-1}\sqrt{1-\rho^2}\beta(t,v)\partial_v H.\end{aligned}$$

Now, from Girsanov's theorem it follows that the $\mathbb{Q}_0$-Brownian increments are related to the $\mathbb{P}$-Brownian increments through

$$dW_{it} = -(1-1/\gamma)\lambda_i(V_t)dt + dW^{\mathbb{Q}_0}_{it},$$

for $i = 1, 2$. Substituting these into the dynamics (5.3) for the variance V_t, we find that its drift changes as $\alpha(t,v) \to \alpha^{\mathbb{Q}_0}(t,v)$ where

$$\alpha^{\mathbb{Q}_0}(t,v) \equiv \alpha(t,v) - (1-1/\gamma)\beta(t,v)\Big[\rho\lambda_1(v) + \sqrt{1-\rho^2}\lambda_2(v)\Big].$$

Thus, from Feynman-Kac we may deduce that H solves the linear PDE

$$0 = \partial_t H + \alpha^{\mathbb{Q}_0}(t,v)\partial_v H + \tfrac{1}{2}\beta^2(t,v)\partial^2_{vv}H + \frac{1-\gamma}{2\gamma^2}\sum_{i=1}^{2}\lambda_i^2(V_t)H, \tag{5.21}$$

subject to the terminal condition $H(T,v) = 1$. The point here is well-worth appreciating: rather than enduring the non-linearity inherent to the HJB formalism, we have transformed the optimisation problem into something as comparatively pedestrian as having to solve (5.21). Whether we aim for an analytic or a numerical solution, it is clear which approach imbues the greatest allure.

5.3.3 The optimal portfolio weights

Finally, we are in a position to determine the optimal portfolio weights π^*_{St} and π^*_{Dt}. Applying Itô's lemma to (5.19) we find

$$\begin{aligned}d\mathscr{W}^*_t &= re^{rt}w_0\tfrac{H_t}{H_0}\xi_t^{-1/\gamma}dt + e^{rt}w_0\tfrac{1}{H_0}\xi_t^{-1/\gamma}dH_t + e^{rt}w_0\tfrac{H_t}{H_0}d(\xi_t^{-1/\gamma}) + e^{rt}w_0\tfrac{1}{H_0}dH_t d(\xi_t^{-1/\gamma})\\ &= \text{drift} + e^{rt}w_0\tfrac{1}{H_0}\xi_t^{-1/\gamma}[\sigma_{1H}H_t dW_{1t} + \sigma_{2H}H_t dW_{2t}] - e^{rt}w_0\tfrac{H_t}{H_0}\tfrac{1}{\gamma}\xi_t^{-1/\gamma-1}d\xi_t\\ &= \text{drift} + \mathscr{W}^*_t[\sigma_{1H}dW_{1t} + \sigma_{2H}dW_{2t}] + e^{rt}w_0\tfrac{H_t}{H_0}\tfrac{1}{\gamma}\xi_t^{-1/\gamma-1}\xi_t[\lambda_1 dW_{1t} + \lambda_2 dW_{2t}]\\ &= \text{drift} + \mathscr{W}^*_t[\sigma_{1H}dW_{1t} + \sigma_{2H}dW_{2t}] + \mathscr{W}^*_t\tfrac{1}{\gamma}[\lambda_1 dW_{1t} + \lambda_2 dW_{2t}]\\ &= \text{drift} + \big[\sigma_{1H} + \tfrac{1}{\gamma}\lambda_1\big]\mathscr{W}^*_t dW_{1t} + \big[\sigma_{2H} + \tfrac{1}{\gamma}\lambda_2\big]\mathscr{W}^*_t dW_{2t},\end{aligned} \tag{5.22}$$

where the first line uses the product rule, the second line makes use of (5.20) and the chain rule, the third line makes use of (5.19) and the definition of the Radon-Nikodym derivative (5.8), and the fourth line makes use of (5.19) again. Comparing (5.22) with our expression for the self-financing condition (5.11) we see that we have established two simultaneous equations from which π^*_{St} and π^*_{Dt} can be determined

$$\pi^*_{S,t}\sqrt{v} + \pi^*_{D,t}\sigma_{1D} = \sigma_{1H} + \tfrac{1}{\gamma}\lambda_1(v), \quad \text{and} \quad \pi^*_{D,t}\sigma_{2D} = \sigma_{2H} + \tfrac{1}{\gamma}\lambda_2(v).$$

Solving these we ultimately arrive at

$$\pi^*_{St} = \frac{1}{\sqrt{v}}\left\{\frac{\lambda_1(v)}{\gamma} - \frac{\sigma_{1D}}{\sigma_{2D}}\frac{\lambda_2(v)}{\gamma} + \sigma_{1H} - \frac{\sigma_{1D}\sigma_{2H}}{\sigma_{2D}}\right\}, \tag{5.23a}$$

$$\pi^*_{Dt} = \frac{1}{\sigma_{2D}}\left\{\frac{\lambda_2(v)}{\gamma} + \sigma_{2H}\right\}. \tag{5.23b}$$

Theorem 5.1. It is well-worth summarising our findings in this section. Consider the control problem stated in (5.12). Defining the function

$$H_t = H(t,v) = \mathbb{E}^{\mathbb{Q}_0}_{t,v}\left[\exp\left\{\frac{1-\gamma}{2\gamma^2}\int_t^T (\lambda_1^2(V_s) + \lambda_2^2(V_s))ds\right\}\right], \tag{5.24}$$

where $\mathbb{Q}_0$ is the measure on $\mathcal{F}_T$ defined through $\xi^0 \equiv d\mathbb{Q}_0/d\mathbb{P}$ where

$$d\xi^0_t = -(1-1/\gamma)\xi^0_t(\lambda_1(V_t)dW_{1t} + \lambda_2(V_t)dW_{2t}), \tag{5.25}$$

the optimal wealth process can be written on the form

$$\mathscr{W}^*_t = e^{rt}w_0\frac{H_t}{H_0}\xi_t^{-1/\gamma}. \tag{5.26}$$

Furthermore, the optimal controls which give rise to this maximal wealth process are of the form

$$\pi^*_{St} = \frac{1}{\sqrt{V_t}}\left\{\frac{\lambda_1(V_t)}{\gamma} - \frac{\sigma_{1D}}{\sigma_{2D}}\frac{\lambda_2(V_t)}{\gamma} + \sigma_{1H} - \frac{\sigma_{1D}\sigma_{2H}}{\sigma_{2D}}\right\}, \tag{5.27a}$$

$$\pi^*_{Dt} = \frac{1}{\sigma_{2D}}\left\{\frac{\lambda_2(V_t)}{\gamma} + \sigma_{2H}\right\}, \tag{5.27b}$$

where

$$\sigma_{1D} \equiv D^{-1}[\rho\beta\partial_v D + \sqrt{v}s\partial_s D], \qquad \sigma_{2D} \equiv D^{-1}[\sqrt{1-\rho^2}\beta\partial_v D], \tag{5.28}$$

$$\sigma_{1H} \equiv H^{-1}\rho\beta\partial_v H, \qquad \sigma_{2H} \equiv H^{-1}\sqrt{1-\rho^2}\beta\partial_v H. \tag{5.29}$$

5.4 Example: the Heston model

Undoubtedly, the most well-known of all stochastic volatility models is that proposed by Heston (1993). The Heston model stands out for a number of reasons: first, the variance process is non-negative and mean-reverting, which harmonises with market data; secondly, the model is sufficiently parsimonious to allow for swift calibrations (Ribeiro and Poulsen (2013), Weron and Wystup (2011)); thirdly, as exposed below, it famously admits comparatively simple expressions for plain vanilla options; finally, said expressions yield implied

volatilities which are found to fit the empirically observed volatility smile closely for a broad range[7] of medium-seized times to maturity Weron and Wystup (2011).

Formally, the Heston model is a Cox-Ingersoll-Ross model for the variance process:

$$dV_t = \kappa(\theta - V_t)dt + \sigma_v\sqrt{V_t}(\rho dW_{1t} + \sqrt{1-\rho^2}dW_{2t}), \tag{5.30}$$

where κ, θ, and σ_v are nonnegative parameters which signify the speed of mean reversion, the long term variance, and the volatility of variance respectively. Insofar as the Feller condition is satisfied, it can be shown that the variance process stays strictly positive at all times.[8] Moreover, the distribution of V_t under (5.30) is non-central χ^2, which in the asymptotic limit $t \to \infty$ tends towards a gamma distribution. This effectively disposes of one of the key shortfalls of classical GBM valuation as the resulting density function of log returns will be fatter (exponential) than the bell curve.

5.4.1 Vanilla valuation

What really propelled the Heston model into the academic limelight is largely its ability to price European calls (and *ipso facto* European puts). For the reader's convenience we here briefly review the valuation formula, and tie it to the theory of pricing in incomplete markets alluded to in Subsection 5.2.1. Specifically, the relevant pricing PDE (5.9) is of the form

$$\begin{aligned} 0 = {} & \partial_t D + rs\partial_s D + \{\kappa(\theta - v) - \sigma_v\sqrt{v}[\rho\lambda_1(v) + \sqrt{1-\rho^2}\lambda_2(v)]\}\partial_v D \\ & + \tfrac{1}{2}vs^2\partial^2_{ss}D + \tfrac{1}{2}\sigma_v^2 v\partial^2_{vv}D + \rho\sigma_v vs\partial^2_{sv}D - rD, \end{aligned} \tag{5.31}$$

subject to $D(T', s) = [\phi(s-K)]^+$, where ϕ is a binary variable which takes on the value +1 if the option is a call, and -1 if the option is a put. Upon solving this equation, Heston crucially makes the assumption that the *market price of volatility risk*, λ_v, here defined as[9]

$$\lambda_v \equiv \sigma_v[\rho\lambda_1 + \sqrt{1-\rho^2}\lambda_2], \tag{5.32}$$

[7]Matching the smile for very short or very long times to maturity proves more difficult. In particular, with regards to the former, the so-called volatility of variance, σ_v, tends to explode, which indicates that there's a jump effect neglected by the dynamics.

[8]It is dubious that calibrated parameters actually satisfy this condition; Ribeiro and Poulsen (2013).

[9]The market price of volatilty risk (5.32) is concept which arises naturally insofar as the dynamical equations (5.3) have *not* had their random components decorellated through a Cholesky decomposition. Specifically, for the market price of risk vector

$$\boldsymbol{\lambda} = \boldsymbol{\sigma}^{-1}(\text{excess return vector}),$$

we would set $\boldsymbol{\sigma} = [\sqrt{V_t}, 0; D^{-1}s\sqrt{v}\partial_s D + D^{-1}\rho\sigma_v\sqrt{v}\partial_v D, D^{-1}\sqrt{1-\rho^2}\sigma_v\sqrt{v}\partial_v D]$, whilst Heston sets $\boldsymbol{\sigma} = \boldsymbol{\sigma}' := [\sqrt{V_t}, 0; D^{-1}\sqrt{v}\partial_s D, D^{-1}\sigma_v\sqrt{v}\partial_v D]$ (the latter is related to the former through the multiplication of the lower triangular matrix $\mathbf{L} = [1, 0; \rho, \sqrt{1-\rho^2}]$: $\boldsymbol{\sigma} = \boldsymbol{\sigma}'\mathbf{L}$). For convenience, Heston also absorbs the constant σ_v in his definition.

is proportional to $\sqrt{v}$, i.e.

$$\exists \bar{\lambda}_v \in \mathbb{R} \text{ s.t. } \lambda_v(v) = \bar{\lambda}_v \sqrt{v}. \tag{5.33}$$

We call this the **Heston assumption** and note that it constitutes a concrete instantiation of the weak Heston assumption explicated above. Nonetheless, based on our desire to solve the PDE for the H-function, it will in fact be convenient to assume something slightly stronger, viz.

Assumption 5.2. There exist constants $\bar{\lambda}_1$ and $\bar{\lambda}_2$ such that $\lambda_1(v) = \bar{\lambda}_1\sqrt{v}$ and $\lambda_2(v) = \bar{\lambda}_2\sqrt{v}$. We call this the **strong Heston assumption**.

Remark 5.2. A partial motivation for (5.33) is provided through Breeden's consumption based model, $\lambda_v(V_t)dt = \gamma \mathbb{C}\text{ov}[dV_t, dc_t/c_t]$, when the consumption process is chosen as in the (general equilibrium) Cox, Ingersoll, and Ross (1985) model (see Heston (1993)). Less generously, we might view it as a postulate detached from empirical evidence, which purposefully has been engineered in order to allow (5.31) to be solved. Be that as it may, under the proportionality assumption it can be shown that the ELMM, $\mathbb{Q}$, exists and that discounted asset prices are true martingales insofar as certain inequalities on the parameters are satisfied, Wong and Heyde (2006). This may be taken as a formal justification for Heston's well-known valuation formula:

Theorem 5.2. (Heston's valuation formula for European vanillas) The no-arbitrage price of a European vanilla option with maturity time T' is given by

$$\begin{aligned} D(t,s,v) &= \text{HestonVanilla}(\kappa, \theta, \sigma_v, \rho, \bar{\lambda}_1, \bar{\lambda}_2, r, v, s, K, \tau', \phi) \\ &= \phi\{sQ_1(\phi) - e^{-r\tau'}KQ_2(\phi)\}, \end{aligned} \tag{5.34}$$

where $\phi = +1$ if D is a call, and $\phi = -1$ if D is a put, $\tau' \equiv T' - t$,

$$Q_j(\phi) \equiv \frac{1-\phi}{2} + \phi P_j(\ln s, v, \tau', \ln K), \tag{5.35}$$

for $j = 1, 2$, and we have defined

$$P_j(\ln s, v, \tau', \ln K) \equiv \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \Re\left\{\frac{e^{-i\varphi \ln K} f_j(\ln s, v, \tau', \varphi)}{i\varphi}\right\} d\varphi \tag{5.36a}$$

$$f_j(\ln s, v, \tau', \varphi) \equiv \exp\{C_j(\tau', \varphi) + D_j(\tau', \varphi)v + i\varphi \ln s\}, \tag{5.36b}$$

$$D_j(\tau', \varphi) \equiv \frac{b_j - \rho\sigma_v\varphi i + d_j}{\sigma_v^2}\left(\frac{1 - e^{d_j\tau'}}{1 - g_j e^{d_j\tau'}}\right), \tag{5.36c}$$

$$C_j(\tau', \varphi) \equiv r\varphi i\tau' + \frac{a}{\sigma_v^2}\left\{(b_j - \rho\sigma_v\varphi i + d_j)\tau' - 2\ln\left(\frac{1 - g_j e^{d_j\tau'}}{1 - g_j}\right)\right\}, \tag{5.36d}$$

where $d_j \equiv ((\rho\sigma_v\varphi i - b_j)^2 - \sigma_v^2(2u_j\varphi i - \varphi^2))^{1/2}$, $g_j \equiv (b_j - \rho\sigma_v\varphi i + d_j)/(b_j - \rho\sigma_v\varphi i - d_j)$, $a \equiv \kappa\theta$, $u_1 \equiv \frac{1}{2}$, $u_2 \equiv -\frac{1}{2}$, $b_1 \equiv \kappa + \bar{\lambda}_v - \rho\sigma_v$, and $b_2 \equiv \kappa + \bar{\lambda}_v$.

Remark 5.3. From the generalised option pricing formula (Björk (2009), proposition 26.11)

$$D_t = \phi\{s\mathbb{S}(\phi S_{T'} \geq \phi K|\mathscr{F}_t) - e^{-r\tau'}K\mathbb{Q}(\phi S_{T'} \geq \phi K|\mathscr{F}_t)\}, \tag{5.37}$$

where $\mathbb{S}$ is the stock measure defined through the Radon-Nikodym derivative

$$\xi_u^s \equiv \frac{d\mathbb{S}}{d\mathbb{Q}} = e^{ru}\frac{S_u}{S_0},$$

it follows that Q_1 and Q_2 in (5.34) are risk-adjusted probabilities that the option expires in the money. Specifically,

$$Q_1 = \mathbb{S}_t(\phi S_{T'} \geq \phi K|S_t = s; V_t = v), \tag{5.38a}$$
$$Q_2 = \mathbb{Q}_t(\phi S_{T'} \geq \phi K|S_t = s; V_t = v). \tag{5.38b}$$

.

Remark 5.4. The formula stated in (5.34) is rather unconventionally expressed in terms of the market price of risk constants $\bar{\lambda}_i$, $i = 1, 2$ along with the $\mathbb{P}$-parameters of the variance process. More commonly, the valuation formula is specified directly in terms of the risk-neutral $\mathbb{Q}$-parameters: $\kappa^{\mathbb{Q}} \equiv \kappa + \sigma_v(\rho\bar{\lambda}_1 + \sqrt{1-\rho^2}\bar{\lambda}_2)$ and $\theta^{\mathbb{Q}} \equiv \theta\kappa/\kappa^{\mathbb{Q}}$ (diffusion parameters invariant).

5.4.2 The optimal Heston controls

We assume the investor trades in a risk free money account, a stock and a European vanilla derivative in a market characterised by Hestonian stochastic volatility. From the generic optimal control functions (5.27) it follows that we must determine $\sigma_{1D}, \sigma_{2D}, \sigma_{1H}$, and σ_{2H} and thence the quantities $\partial_s D, \partial_v D, H$, and $\partial_v H$. We do this over the two subsequent lemmas.

Lemma 5.1. The option **delta** is given by

$$\Delta_t \equiv \partial_s D = \phi Q_1(\phi), \tag{5.39}$$

while the option **vega** is given by

$$\nu_t \equiv \partial_v D = s\nu_1(\ln s, v, \tau', \ln K) - Ke^{-r\tau'}\nu_2(\ln s, v, \tau', \ln K), \tag{5.40}$$

where

$$\nu_j(\ln s, v, \tau', \ln K) \equiv \frac{1}{\pi}\int_0^\infty \Re\left\{\frac{D_j(\tau', \varphi)e^{-i\varphi \ln K} f_j(\ln s, v, \tau', \varphi)}{i\varphi}\right\} d\varphi. \tag{5.41}$$

Proof. A hands-on differentiation of (5.34) with respect to the underlying, s, and subsequent algebraic simplification is a tedious exercise better avoided. A somewhat subtler but

arguably simpler argument may be presented by noting that the valuation formula is first order homogenous in the variable pair (s, K), i.e.

$$\begin{aligned}\text{HestonVanilla}(\kappa, \theta, \sigma_v, \rho, \bar{\lambda}_1, \bar{\lambda}_2, r, v, as, aK, \tau', \phi) = \\ a \cdot \text{HestonVanilla}(\kappa, \theta, \sigma_v, \rho, \bar{\lambda}_1, \bar{\lambda}_2, r, v, s, K, \tau', \phi),\end{aligned}$$

for any $a \in \mathbb{R}$, whence *Euler's Homogenous Function Theorem*[10] entails

$$D = s\partial_s D + K\partial_K D. \tag{5.42}$$

Comparing (5.34) with (5.42) it is tempting to deduce that $\partial_s D = \phi Q_1(\phi)$, yet some care must be taken here. Specifically, it is not immediately obvious that (5.34) is the so-called *natural form* of D, Reiß and Wystup (2001): with two terms present in the equation we can add any arbitrary component to one term, as long as we cancel it through a corresponding subtraction to the other term.[11] To establish that (5.34) is the natural one, we employ a well-known result from Breeden and Litzenberger (1978), viz.

$$\begin{aligned}\partial_K D &= e^{-r\tau'}\partial_K \mathbb{E}^{\mathbb{Q}}_{t,s,v}[[\phi(S_{T'} - K)]^+] \\ &= e^{-r\tau'}\phi\partial_K \mathbb{E}^{\mathbb{Q}}_{t,s,v}[(S_{T'} - K)\mathbf{1}\{\phi S_{T'} \geq \phi K\}] \\ &= -e^{-r\tau'}\phi\mathbb{E}^{\mathbb{Q}}_{t,s,v}[\mathbf{1}\{\phi S_{T'} \geq \phi K\}] \\ &= -e^{-r\tau'}\phi\mathbb{Q}_{t,s,v}(\phi S_{T'} \geq \phi K).\end{aligned}$$

Comparing this with (5.37) the result follows.

Equation (5.40) follows immediately from differentiating (5.34) with respect to v. Note that the result is independent of ϕ. □

Lemma 5.2. Under the strong Heston assumption the H function has the exponential affine form

$$H(t, v) = \exp\{A(\tau) + B(\tau)v\}, \tag{5.43}$$

where $\tau \equiv T - t$. Here $B : [0, T] \mapsto \mathbb{R}$ is the function

$$B(\tau) = \frac{1 - \gamma}{\gamma^2}(\bar{\lambda}_1^2 + \bar{\lambda}_2^2) \cdot \frac{e^{\omega\tau} - 1}{(\omega + \alpha)(e^{\omega\tau} - 1) + 2\omega}, \tag{5.44}$$

[10]Recall that the function $g : \mathbb{R}^2 \mapsto \mathbb{R}$ is said to be homogenous of degree n if

$$g(ax_1, ax_2) = a^n g(x_1, x_2).$$

Let $x_1' = nx_1$ and $x_2' = nx_2$ then we find upon differentiating g with respect to a that $na^{n-1}g = \partial_{x_1'}g\partial_a x_1' + \partial_{x_2'}g\partial_a x_2' = x_1\partial_{ax_1}g + x_2\partial_{ax_2}g$. In particular, upon setting $a = 1$ we get Euler's result for homogenous functions:

$$ng = x_1\partial_{x_1}g + x_2\partial_{x_2}g.$$

[11]Let $g = x_1\partial_{x_1}g + x_2\partial_{x_2}g$ and $g = x_1h_1(x_1, x_2) + x_2h_2(x_1, x_2)$ then a necessary and sufficient condition for $\partial_{x_1}g = h_1(x_1, x_2)$ and $\partial_{x_2}g = h_2(x_1, x_2)$ is that $x_1^2\partial_{x_1}h_1 = x_2^2\partial_{x_2}h_2$ - see Reiß and Wystup (2001).

where $\alpha \equiv \kappa + (1 - 1/\gamma)\bar{\lambda}_v$ and

$$\omega \equiv \sqrt{\alpha^2 + \sigma_v^2 \frac{\gamma - 1}{\gamma^2}(\bar{\lambda}_1^2 + \bar{\lambda}_2^2)},$$

while $A : [0, T] \mapsto \mathbb{R}$ is the function

$$A(\tau) = \frac{\kappa\theta}{\alpha^2 - \omega^2} \left\{ (\alpha + \omega)\tau + 2 \ln \left| \frac{2\omega}{(\alpha + \omega)(e^{\omega\tau} - 1) + 2\omega} \right| \right\}. \tag{5.45}$$

Proof. Substituting in the relevant parametric specifications (5.30), (5.33), into the governing PDE (5.21) we find

$$0 = -\partial_\tau H + [\kappa\theta - \{\kappa + (1 - 1/\gamma)\bar{\lambda}_v\}v]\partial_v H + \tfrac{1}{2}\sigma_v^2 v \partial_{vv}^2 H + \frac{1 - \gamma}{2\gamma^2}(\bar{\lambda}_1^2 + \bar{\lambda}_2^2)vH, \tag{5.46}$$

subject to the initial condition $H(0, v) = 1$, where we have invoked the time transformation $t \mapsto \tau$. Since the coefficients are linear functions of v we form the ansatz that the solution is of an exponential affine form. Thus, upon substituting (5.43) into (5.46) and using the fact that the expression should hold for any value of v we find the coupled ODEs:

$$B'(\tau) = \tfrac{1}{2}\sigma_v^2 B^2(\tau) - \{\kappa + (1 - 1/\gamma)\bar{\lambda}_v\}B(\tau) + \frac{1 - \gamma}{2\gamma^2}(\bar{\lambda}_1^2 + \bar{\lambda}_2^2), \tag{5.47a}$$

$$A'(\tau) = \kappa\theta B(\tau), \tag{5.47b}$$

subject to the boundary conditions $A(0) = B(0) = 0$, where $'$ denotes the derivative with respect to τ. The first equation is Riccatian, which readily allows us to extract the solution (5.44).[12] Note that ω is a real number insofar as $\gamma > 1$ which we assume to be the case. As for the function A we observe that (5.47b) can we written as $A(\tau) = \kappa\theta \int_0^\tau B(t)dt$. Performing this tedious integration we get the desired result. □

Putting these results together we can finally state our theorem on the optimal (B, S, D)-portfolio weights in a Heston driven economy:

Theorem 5.3. The optimal stock weight is given by

$$\pi_{S,t}^* = \frac{\bar{\lambda}_1}{\gamma} - \frac{1}{\sqrt{1 - \rho^2}} \left[\rho + \frac{s}{\sigma_v} \frac{\Delta_t}{\nu_t} \right] \frac{\bar{\lambda}_2}{\gamma} - s \frac{\Delta_t}{\nu_t} B(\tau), \tag{5.48}$$

while the optimal vanilla option weight is

$$\pi_{D,t}^* = \frac{D_t \bar{\lambda}_2}{\gamma \sigma_v \sqrt{1 - \rho^2} \nu_t} + \frac{D_t}{\nu_t} B(\tau), \tag{5.49}$$

where B is defined in (5.44), D_t is the option price given by (5.34), Δ_t is the option delta given in (5.39), and ν_t is the option vega given in (5.40). Note that the time parameter in B is τ (the investment horizon), while it for option quantities $\{D, \Delta, \nu\}$ is τ' (the maturity of the option).

[12] Recall that the generic Riccati equation $y'(x) = ay^2(x) + by(x) + c$ with $y(0) = 0$ has the solution $y(x) = [2c(e^{\delta x} - 1)]/[(\delta - b)(e^{\delta x} - 1) + 2\delta]$ where $\delta \equiv \sqrt{b^2 - 4ac}$ assuming $b^2 > 4ac$.

We note that the first term in (5.48) is Merton's optimal stock weight in a simple (B, S)-economy with constant volatility. More generally, referencing standard results in the literature[13], we see that the first two terms in (5.48) and the first term in (5.49) constitute the optimal portfolio weights in a (B, S, D)-economy for a utility maximising investor who disregards stochastic fluctuations in the state variable v (otherwise known as the *myopic* or *1-period* strategy). Thus, the hedge against stochastic volatility is nested in the time-price-volatility dependent term $-s\frac{\Delta_t}{\nu_t}B(\tau)$ in (5.48) and $\frac{D_t}{\nu_t}B(\tau)$ in (5.49). In this connection we note that Δ is a function bounded by the interval $[0, 1]$ for a call option ($[-1, 0]$ for a put option), whilst $B(\tau)$ is a monotonically decreasing function bounded by the interval $((1-\gamma)(\bar{\lambda}_1^2 + \bar{\lambda}_2^2)/(\gamma^2[\omega + \alpha]), 0]$. s, D, and ν are all positive quantities unbounded from above. The signs of the volatility hedge corrections on the stock and the derivative are thus respectively positive and negative if D is a call option, and negative and negative if D is a put option. To appreciate the implications of this, figure 5.1 plots the optimal (bank,stock,ATM call option)-weights for different times to maturity, when the tuple (s, v) is held constant at $(100, \theta)$. We assume that the risk free rate is 0.02, that the option expires at the end of the investment horizon ($\tau = \tau'$), and that the investor's level of risk aversion γ is 2. Other parameters are estimated from the S&P 500 index and are exhibited in table 5.5.

Upon examining figure 5.1 we make the following observations: relative to the optimal Merton weight [**dash-dotted grey line**], Liu's volatility correction [**full grey line**] is barely noticeable, perturbing π_S^* at the order of magnitude 10^{-3}. By comparison, access to derivative trading prompts the investor to drastically increase her holding in the stock [**full red line**], by decreasing her long position in the money account [**full black line**], and shorting the call option [**full blue line**] at a rather modest level. This makes good sense: by shorting the call, the investor has a *negative* exposure to the risk endemic to the variance process thereby collecting positive risk premium, Munk (2013). Out of interest, we have also plotted the effect of including the volatility hedge terms in the optimal investment ratios [**dash-dotted lines**]. As argued above, this respectively underestimates and overestimates the weights on stock and the derivative: here by as much as six percentage points for the stock and a single percentage point for the call (see the RHS figure for greater clarity). Volatility hedge corrections thus *seemingly* have the magnitude to perturb the terminal wealth of a rational investor by a measurable amount. Yet, this is in fact *not* the case when we Monte Carlo simulate the wealth process (5.11) of an investor trading in a Hestonian economy: although the expected return is higher for someone who hedges volatility vis-à-vis one who does not, so is the associated variance. A Welch's t-test therefore cannot reject the null hypothesis that the two trading strategies have equal returns (p-value ≈ 0.65). This suggests that the real capital gains (if any) are to be garnered from access to the derivative security, and not the hedge corrections to volatility per se, in accordance with the findings by Liu and Pan (2003).

[13] See for example Munk (2013) theorems 6.2 and 7.5

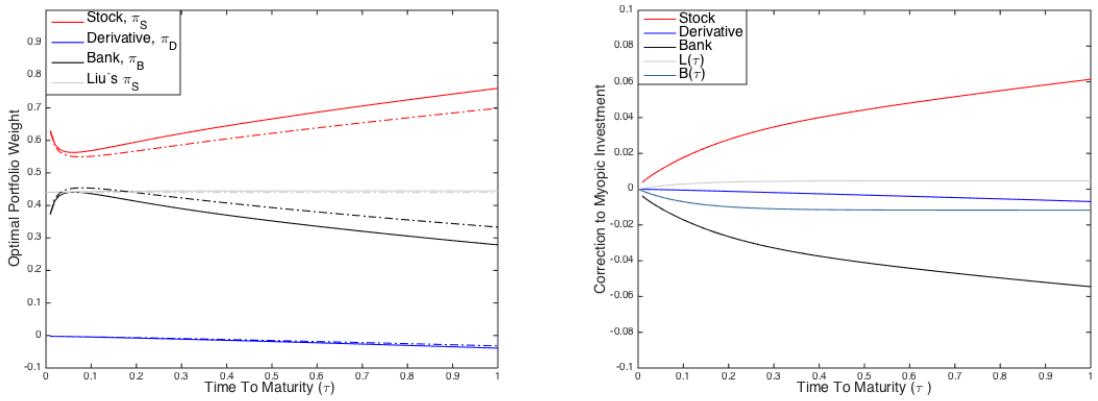

Figure 5.1: **Left:** Optimal investment strategies in a (bank,stock,call option)-economy for different times to maturity with s and v held constant. Note that the investor shorts the derivative in order to enter into a significant long position in the underlying stock and deposit money in the bank. The grey line shows Liu's optimal portfolio weight on the stock when the investor disregards derivatives. The dash-dotted lines represent the corresponding strategies when we do not hedge stochastic variations in the state variable (volatility). For Liu's model this corresponds to the Merton weight λ_1/γ. **Right:** The size of the correction to the various portfolio weights brought about by hedging volatility. The figure also exhibits the magnitude of the deterministic functions $L(\tau)$ and $B(\tau)$. **NB:** as TTM approaches zero, so does ν which creates problems with numerical instability in this region.

5.4.3 The value of derivative trading

Whilst the inclusion of a derivative to the investment portfolio has a significant impact on the optimal trading strategy (as demonstrated in the previous section), one should perhaps be more vigilant of a potential *value-gain* from such a position. In particular, since both $(\pi_B^*, \pi_S^*, \pi_D^*)$ and Liu's strategy $(\pi_B^{Liu}, \pi_S^{Liu})$ are optimal in the sense that they maximise the expected utility of terminal wealth, their respective utilities are the quantities to compare. In fact, Liu's optimum is a constrained strategy of the former, in that no derivative trading is allowed ($\pi_D \equiv 0$) and thus suboptimal: with $J(w_0, \pi) = \mathbb{E}[e^{-\delta T} u(\mathscr{W}_T)]$ denoting the expected utility from investing an initial wealth w_0 according to a strategy π, we readily have from (5.10)

$$J(w_0, \pi^{Liu}) \leq J(w_0, \pi^*) = \sup_{\pi \in \mathscr{L}^2} J(w_0, \pi).$$

A common measure for the comparison of two such strategies in monetary terms is the *certainty equivalent in wealth*

$$c^* = \sup \left\{ c \geq 0 : J(w_0, \pi^{Liu}) \leq J(w_0 - c, \pi^*) \right\}$$

which gives the reduction in initial wealth the investor is willing to sacrifice in order to trade in the derivative market (see Munk (2013), Chapter 5.4). For the optimal strategy,

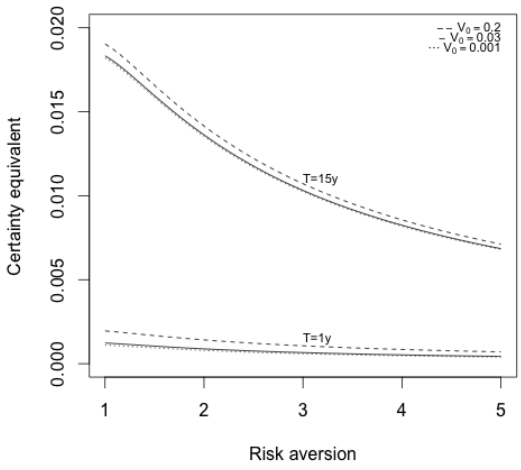

Figure 5.2: Certainty equivalents in initial wealth: the fraction of wealth an investor is willing to sacrifice to be able to trade in the derivative market, as a function of investor risk-aversion. The upper curves show wealth equivalents for a trading horizon of 15 years while the lower curves show a one-year trading horizon. Three different current volatility levels are included in the figure.

we have an optimal terminal wealth W_T^* from (5.14) that yields

$$J(w_0, \pi^*) = e^{(r(1-\gamma)-\delta)T} \frac{w_0^{1-\gamma}}{1-\gamma} H(0,v)^\gamma$$

where we have used (5.18) for the expectation (5.17). Similarly, following Ellersgaard and Jönsson (2013), the optimal expected utility from investing in a stock-bond economy is given by $J(w_0, \pi^{Liu}) = e^{(r(1-\gamma)-\delta)T} \frac{w_0^{1-\gamma}}{1-\gamma} \hat{H}(0,v)^\gamma$ where the function $\hat{H}$ is of the same exponential form as H given in (5.43). Denoting its corresponding exponential coefficients with $\hat{A}(\cdot)$ and $\hat{B}(\cdot)$ respectively, a straightforward calculation gives the certainty equivalent

$$c^* = \left(1 - \left[e^{\hat{A}(T)-A(T)-(\hat{B}(T)-B(T))v}\right]^{\frac{\gamma}{1-\gamma}}\right) w_0.$$

The functions $\hat{A}$, $\hat{B}$ are identical to (5.45)-(5.44) with slightly modified coefficients. For their exact expressions, we refer to Ellersgaard and Jönsson (2013).

Figure 5.2 depicts certainty equivalents in wealth as a fraction of w_0 for different levels of risk aversion γ. The certainty equivalent is based on estimated parameters from the S&P 500 data (table 5.5) and two different investment horizons (one year and 15 years) corresponding to the horizons that will be used for the empirical trading experiment. We have included three different levels of current volatility: $v = 0.03$ is equal the estimated long-term variance θ, $v = 0.001$ is the same order of magnitude as the minimum historical S&P 500 variance while $v = 0.20$ could be considered to be a relatively high (and rare) level, cf. figure 5.3.

Clearly, figure 5.2 demonstrates that the investment horizon has a high impact on the certainty equivalent whilst the current variance level plays a rather minor role: the investor is prepared to refrain from a few percentage points from her initial wealth in return of being able to trade in the derivative market, and the longer the period, the more to gain from trading. The current variance is ephemeral, and thus yields a minor effect. We further note that the wealth equivalent decreases with level of risk aversion γ: more risk-averse investors will involve less in the risky trading of derivatives and thus have less to gain from this business.

For our empirical study we will assume $\gamma = 2$ which yields certainty equivalents $c^* \approx 0.014 \cdot w_0$ (for initial variance levels between 0.001 and 0.20) for the first experiment with $T = 15$ years and $c^* \approx 0.0009 \cdot w_0$ for the second experiment with $T = 1$ year. For both horizons, only a very modest fraction of initial wealth has to be sacrificed for the access to derivative trading. Whether we will actually see any difference at all in *realised utility* (expressed in monetary terms with the realised certainty equivalent) is yet for the empirical trading experiment to reveal.

5.4.4 Towards higher generality

A natural generalisation of the Heston model is to allow the model parameters to be time-dependent functions. This will not only allow for a financial landscape that changes dynamically over time, but also provide a more realistic pricing model the implied volatility surface of which calibrates much closer to market data. A tractable yet flexible functional form for this purpose is to specify $\left\{\kappa(t), \theta(t), \sigma_v(t), \rho(t), \bar{\lambda}_1(t), \bar{\lambda}_2(t)\right\}$ to be piecewise constant functions over some finite partition $t_0 < t_1 < \cdots < t_n$ of the market horizon $[0, T']$. That is, $\kappa(t) = \bar{\kappa}_1$ for $t \in (t_{n-1}, t_n]$, $\kappa(t) = \bar{\kappa}_2$ for $t \in (t_{n-2}, t_{n-1}]$ etc. with corresponding structures for the remaining parameters.[14] The key observation is then that we may solve all PDEs sequentially backwards in time. Firstly, we have that the optimal weights (5.48)-(5.49) depends on the function B which solves the Riccati equation (5.47a). For the first subinterval $(t_{n-1}, t_n]$ expressed in backwards-time, $\tau \in [\tau_0, \tau_1) = [0, T - t_{n-1})$ with $\tau_k \equiv T - t_{n-k}$, we have the familiar boundary condition $B_1(0) = 0$ and solution B_1 as in equation (5.44) with $\{\bar{\kappa}_1, \bar{\theta}_1, \bar{\sigma}_{v1}, \bar{\rho}, \bar{\lambda}_{11}, \bar{\lambda}_{21}\}$. This gives us a value for $B_1(\tau_1)$ which will act as a (non-zero) boundary condition for B_2 on the next interval $[\tau_1, \tau_2)$. We may then proceed sequentially backwards in time[15] to obtain $B_2, \ldots B_n$ over $[\tau_2, \tau_3), \ldots, [\tau_{n-1}, \tau_n)$ to cover the whole of $[0, T] \subseteq [0, T']$. Secondly, we need to calculate Δ and ν over the same sequence of subintervals (extended to cover $[0, T']$) to complete the expression for our market weights. It turns out that this calculation is virtually identical to the one just considered: D_j and C_j of (5.36c)-(5.36d) solve a system of ODEs which is the same as

[14] We shall still assume that the parameter-constants are specified within reasonable limits, i.e. that the governing dynamics (5.30) allows for a positive solution. Moreover, the Feller condition might be desired to be satisfied locally for each time interval of the partition.

[15] To this end we need the Ricatti equation with a *non-zero* initial condition i.e. $y'(x) = ay^2(x) + by(x) + c$, $y(0) = y_0$, which has the solution $y(x) = y_0 + [2(ay_0^2 + by_0 + c)(e^{\delta x} - 1)]/[(\delta - b - 2ay_0)(e^{\delta x} - 1) + 2\delta]$ where $\delta \equiv \sqrt{b^2 - 4ac}$ assuming $b^2 > 4ac$.

(5.47a)-(5.47b) (see Mikhailov and Nögel (2004) for details). Hence, we may compute the optimal weights for a time-dependent parameter specification of the Heston model as well (for reasons of brevity we exclude the technical details from this paper).

5.5 The empirical perspective

Based on our optimal portfolio weights in a Heston driven (B, S, D)-economy, we proceed to perform an empirical experiment which aims to measure the degree to which the inclusion of plain vanillas impacts the financial wealth of a utility maximising investor. In particular, we set out to perform a simple automated trading experiment where we let historical market prices from the S&P 500 index play the role of the fundamental risky security and where market prices of call options written on the same index constitute the derivative available in the economy. We use market interest rates for the money account.

5.5.1 Market data

For the tradable stock, we use 3,909 daily prices of the S&P 500 index from the period 2000-01-03 to 2015-08-31. The market price (sourced from Wharton Research Data Services[16]) is plotted in figure 5.3 together with the daily variance. The variance process is measured from high-frequency data with the realised volatility measure and we use precomputed estimates from the Oxford-Man Institute's realised library.[17].

For the tradable derivative, we use daily mid-market prices of European call options on the S&P 500 index from the same time period, as shown in figure 5.4. The time period covers prices of 23 call options with medium-sized times to maturity (in the range 18 to 36 months subject to data availability) and the strike-price of each option is selected to be ATM at initiation (or as close as possible thereto - subject to availability) thereby sowing the seed for high exposure to volatility risk. The strike-price and time-to-maturity structure is shown in figure 5.6. Note the varying TTMs for the options, which again are symptomatic of the data set available (we use the Option Metrics database sourced through Wharton Research Data Services).

For our last asset in the (B, S, D) economy, we use the daily short-term LIBOR rate for an interest to the risk-free money account. The LIBOR market-data is from the Option Metrics database as well.

[16] `https://wrds-web.wharton.upenn.edu/wrds/`.

[17] The Realised Library version 0.2 by Heber, Gerd, Lunde, Shephard and Sheppard (2009) - see `http://realized.oxford-man.ox.ac.uk` For details on the realised volatility measure, see e.g Andersen and Benzoni (2009).

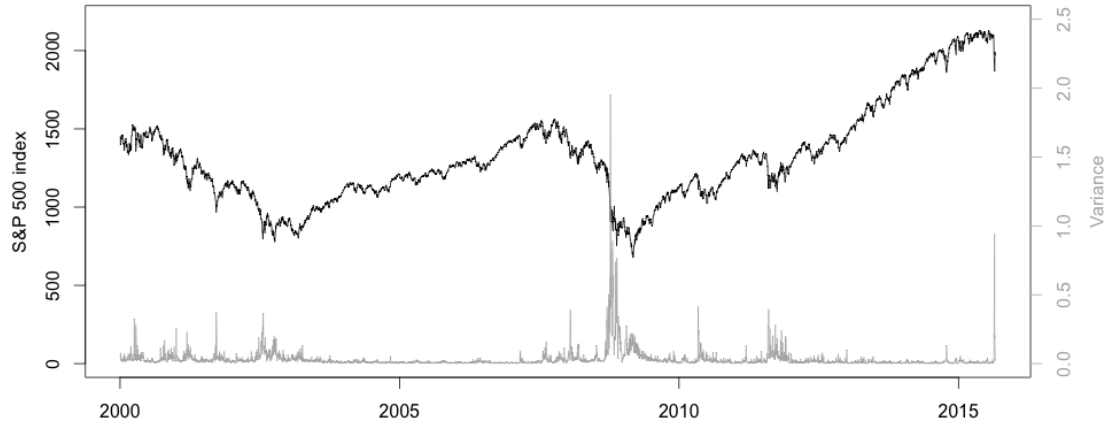

Figure 5.3: Daily market prices and measured variance of the S&P 500 index from the period 2000-01-03 to 2015-08-31. The price data is sourced from Wharton Research Data Services while the variance data is sourced from the Oxford-Man Institute's realised library.

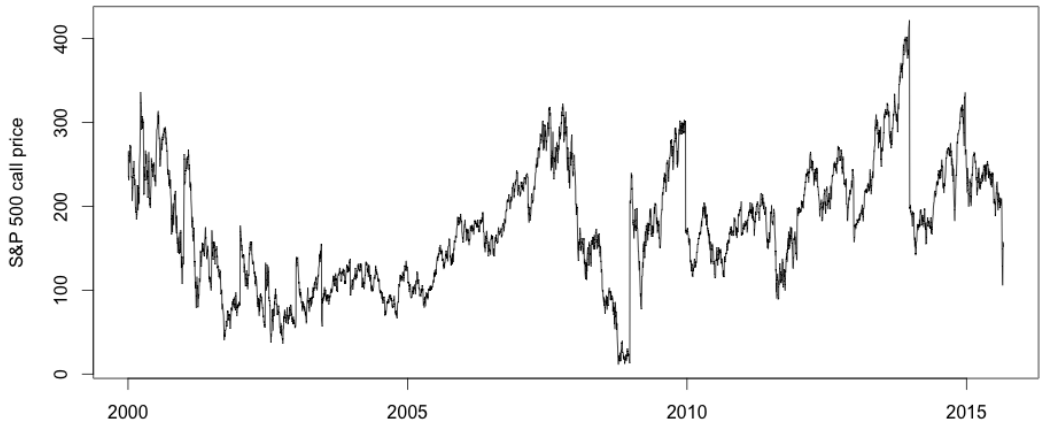

Figure 5.4: Daily mid-market prices of European call options on the S&P 500 index from the period 2000-01-03 to 2015-08-31, sourced from Wharton Research Data Services (Option Metrics data). The corresponding strike prices and maturities are shown in figure 5.6.

5.5.2 Parameter estimation

For our empirical experiment, we will trade according to the optimal portfolio weights $(\pi^*_{Bt}, \pi^*_{St}, \pi^*_{Dt})$ which are functions of the model parameters, the current stock price, call price and variance level at time t. To this end, we estimate the model parameters with the

following approach. First, we estimate $(\kappa, \theta, \sigma_v, \rho)$ of the Cox-Ingersoll-Ross process from the daily variance data with a maximum likelihood method.[18] Secondly, based on these parameters, to determine the market prices of risk, $(\bar{\lambda}_1, \bar{\lambda}_2)$, we minimise the squares error between daily observed S&P 500 call option prices, $\{\hat{C}_t\}$, and corresponding theoretical Heston prices $\{C_t^{He} = C_t^{He}(\bar{\lambda}_1, \bar{\lambda}_2)\}$, where the strike-maturity structure is chosen as in figure 5.6. Specifically, we solve the minimisation problem

$$(\bar{\lambda}_1, \bar{\lambda}_2) = \underset{(\bar{\lambda}_1, \bar{\lambda}_2)}{\operatorname{argmin}} \sum_{t \in \mathbb{T}} \left(\hat{C}_t - C_t^{He}(\bar{\lambda}_1, \bar{\lambda}_2) \right)^2,$$

where $\mathbb{T} = \{\text{trading days between 2000-01-03 and 2015-08-31}\}$. The resulting parameters estimated from the market data are given in table 5.5. Note here in particular that the market price of risk $\bar{\lambda}_2$ associated with W_2 is negative, corroborating standard empirical findings a la Bakshi and Kapadia (2003).

	κ	θ	σ_v	ρ	$\bar{\lambda}_1$	$\bar{\lambda}_2$
Estimate	9.71	0.030	2.72	-0.17	0.88	-0.29

Table 5.5: Estimated model parameters with the two-step approach. The estimates are based on 3,909 daily observations of the stock price, the variance and the call price. Note that $\bar{\lambda}_1 > 0$ whilst $\bar{\lambda}_2 < 0$.

Remark 5.5. A few remarks on the above estimation procedure are in order here. First, note that when we will use the parameter estimates for the forthcoming trading experiment, we employ ex-ante estimates based on actual "future" market data. An alternative is to estimate parameters from historical data prior to the trading period. However, due to the amount of available data, this would impair the accuracy our estimates. Since we are primarily interested in the the efficiency of the trading strategy, (and not in the parameter estimation problem per se) we require as robust estimates as possible. Thus, we opt for the former alternative. Secondly, we use a rather unconventional estimation procedure whereby we separately estimate the variance parameters under the statistical measure $\mathbb{P}$, followed by a mean square optimisation to back out the market prices of risk. A more commonplace approach is to formulate the pricing model under $\mathbb{Q}$ directly (see remark 5.4), and to estimate the risk-neutral parameters from option data alone, again through mean square principles . This estimation approach is commonly referred to as model-to-market calibration, and typically a whole surface of option prices is employed for day-to-day estimation of the parameters. Since we require statistical CIR-parameters along with the market price of risks separately for calculation of the optimal portfolio weights, we use the two-step approach in place of the calibration method.

[18] We use numerical optimisation of a Gaussian likelihood, from the method of Sørensen (1999) based on estimating functions.

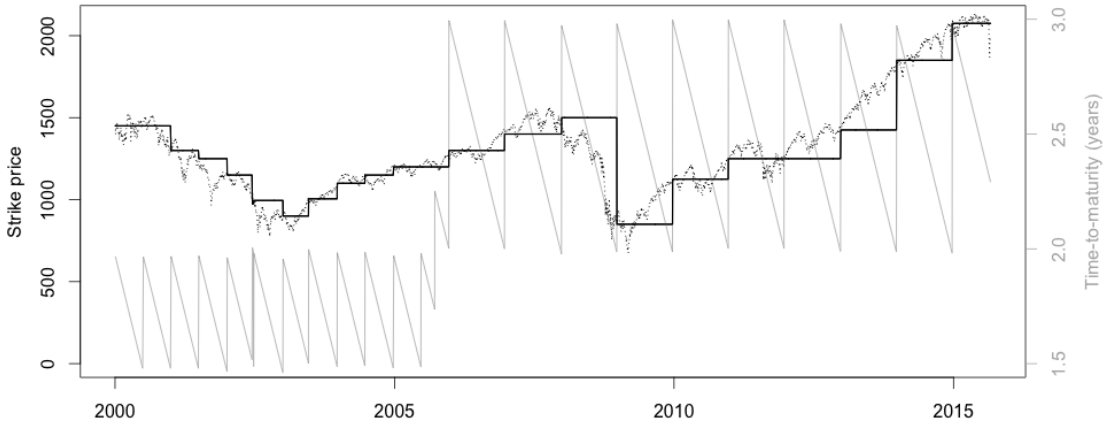

Figure 5.6: The daily strike-price (black line) and time-to-maturity (grey line) for the call price in figure 5.4. The dotted line shows the S&P 500 index level.

5.5.3 Empirical trading experiment

With market data for the (B, S, D) economy from the S&P 500 index and the LIBOR rate, we set out to perform an empirical trading experiment. We intend to invest in a portfolio according to the optimal weights $(\pi^*_{Bt}, \pi^*_{St}, \pi^*_{Dt})$ and we trade dynamically with daily rebalancing as the time evolves during the period 2000-01-03 to 2015-08-31. Hence, if ΔX_{t_i} denotes the daily price-change from t_i to t_{i+1} of an asset in the economy, this means that we realise a daily change in the portfolio value

$$\begin{aligned} \Delta \mathscr{W}_{t_i} &= \mathscr{W}_{t_i} \left(\pi^*_{Bt_i} \frac{\Delta B_{t_i}}{B_{t_i}} + \pi^*_{St_i} \frac{\Delta S_{t_i}}{S_{t_i}} + \pi^*_{Dt_i} \frac{\Delta D_{t_i}}{D_{t_i}} \right), \\ \mathscr{W}_{t_0} &= w_0, \end{aligned}$$

for all dates $t_0, t_1, \ldots$ in the investment period and the realised daily wealth amounts to $\mathscr{W}_{t_{i+1}} = \mathscr{W}_{t_i} + \Delta \mathscr{W}_{t_i}$. Note that the portfolio is self-financing: an initial amount w_0 is invested at the initial time and there is no infusion or withdraw of capital from the portfolio during the investment period.

In addition to the parameters in table 5.5, we fix the risk aversion parameter to the arbitrary value $\gamma = 2$. For comparison purposes, we include a "naive" trading strategy with a constant equal weight invested in each asset, $(\pi_{Bt}, \pi_{St}, \pi_{Dt}) = (1/3, 1/3, 1/3)$. We also trade according to Liu's optimal investment strategy in the limited economy, that is, we invest in (B, S) with portfolio weights $(\pi^{Liu}_{Bt}, \pi^{Liu}_{St})$ (and $\pi_{Dt} = 0$ for the call option).

With this in mind, we conduct the following two experiments for each of the three trading strategies:

1. **Trading throughout 2000-2015.** We set the investment period to be 2000-01-03 to 2015-08-31 and initialise the portfolios with a wealth $w_0 = 1,000$. We do not trade over the dates when there is a "new" option, i.e. every time there is a new expiry date, since this would give false price moves of the call option due to changes in the strike price (cf. figure 5.4 and 5.6). The same rule pertains to Liu's strategy, even though there is no trading in the option. The realised wealth processes from trading according to the three strategies are shown in figure 5.8 while the optimal portfolio weights are shown in figure 5.11.

   The realised terminal wealth $\mathscr{W}_T$ is 1,220 with the naive strategy, 1,497 with Liu's strategy and 1,524 with the optimal strategy respectively. For the two latter strategies to be equivalent, we have to initiate the optimal (B, S, D)-portfolio with at least $w_0 = 982.5$ which yields a realised wealth of 1,497. Since our experiment results in a single wealth-path here, this yields equivalent realised utilities as well. Hence, in terms of a relative certainty equivalent, a reduction of 1.75% from the initial wealth of the optimal strategy makes it comparable to the strategy being restricted to bond and stock trading. As predicted by the certainty equivalent from expected utility, $c^* = 1.4\%$, this is quite a modest reduction which suggests that the two strategies are close in monetary performance.

   To further compare the strategies (all three having $w_0 = 1,000$) with a measure standard for financial investments, we calculate the realised Sharpe-ratios of daily returns as

$$\mathcal{S} = \frac{\text{Mean}(R_{t_i} - r_{t_i})}{\text{SD}(R_{t_i})},$$

   where $R_{t_i} = \log(\mathscr{W}_{t_i}) - \log(\mathscr{W}_{t_{i-1}})$, $t_1, t_2, \ldots$ are the daily returns of the investment portfolio and r_{t_i} is the daily returns of the money account (the daily return from the LIBOR rate). The results are shown in table 5.7.

Strategy	Mean return	Std. Dev.	Sharpe-ratio	Sharpe-R. annual
$(\pi^*_{Bt}, \pi^*_{St}, \pi^*_{Dt})$	0.011%	0.65%	0.39%	7.46%
$(\pi^{Liu}_{Bt}, \pi^{Liu}_{St}, 0)$	0.010%	0.56%	0.37%	7.16%
$(1/3, 1/3, 1/3)$	0.005%	2.27%	-0.14 %	-2.64%

Table 5.7: The Sharpe-ratio, mean and standard deviation of daily portfolio returns from the three strategies when trading throughout the whole period 2000-01-03 to 2015-08-31. The last column shows the annualised Sharpe-ratio. The daily mean-return of the money account is 0.0082%, which corresponds to an annualised return of 3.0%.

2. **Investment periods according to the option-expiry structure.** For our second empirical experiment, we reset our investment portfolio with an initial $w_0 = 1,000$ every time there is a "new" option, i.e. every time there is a new expiry date and strike-price (see figure 5.6). We set the investment period accordingly, i.e. to start when we reset the portfolio and to end at the date on which we will reset

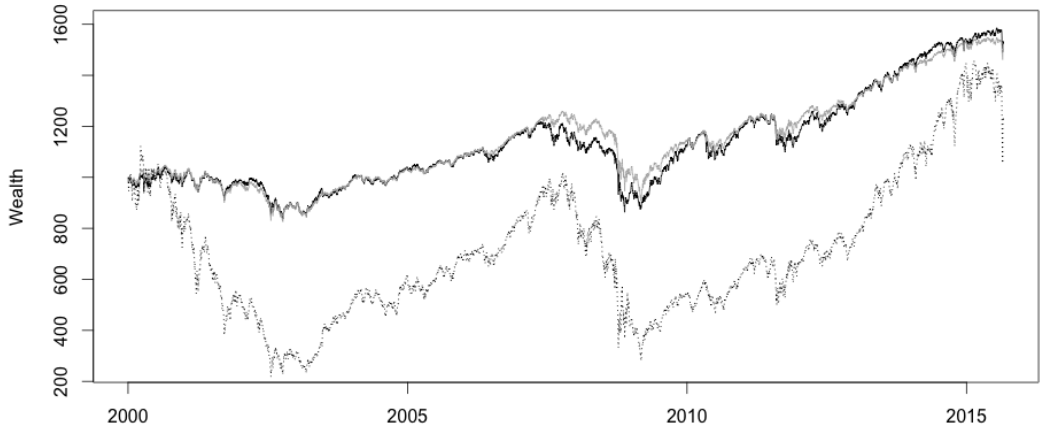

Figure 5.8: Trading throughout 2000-01-03 to 2015-08-31: wealth processes from trading in {LIBOR,S&P500,Call} according to the naive strategy (dotted line) and optimal investment strategy (black line), and from trading in {LIBOR,S&P500} with Liu's optimal strategy (grey line).

the portfolio the next time. The resulting realised portfolio value-processes from the optimal (B, S, D)- and (B, S) strategies are shown in figure 5.10 and the portfolio weights of the two strategies are plotted in figure 5.12.

For this experiment we have a total of 25 investment periods with horizons in the range of six months to one year. Calculating the average realised utility as

$$\bar{J}(\pi) = \sum_{j=1}^{25} \frac{(\mathscr{W}_{T_j})^{1-\gamma}}{1-\gamma},$$

where $\mathscr{W}_{T_j}$ is the realised terminal wealth of period j, we obtain $-10.26 \cdot 10^{-4}$ from the naive strategy, $-9.858 \cdot 10^{-4}$ from Liu's strategy and $-9.860 \cdot 10^{-4}$ from the optimal strategy. Thus, Liu's strategy actually achieves a slightly greater utility. In terms of a certainty equivalent, we have to initiate the optimal (B, S, D) portfolio (for each of the 25 periods) with $w_0 = 1,000.50$ in order to obtain the same average realised utility. This corresponds to a negligible reduction (increase) of -0.05% of initial wealth, to be compared with the prediction in certainty equivalent for expected utility being 0.9%.

The Sharpe-ratios based on realised daily returns of the investment portfolios are collected in table 5.9, where the results for the naive strategy are included as well.

A note on the interpretation of these results is in order here. First, whilst trading strategy (1) picks up a slightly higher mean return along the way for the (B, S, D) portfolio, the associated variance of returns is also higher (the risk averse investor might choke on this).

Strategy	Mean return	Std. Dev.	Sharpe-ratio	Sharpe-R. annual
$(\pi^*_{Bt}, \pi^*_{St}, \pi^*_{Dt})$	0.010%	0.55%	0.37%	7.13%
$(\pi^{Liu}_{Bt}, \pi^{Liu}_{St}, 0)$	0.011%	0.65%	0.36%	6.93%
$(1/3, 1/3, 1/3)$	0.005%	2.28%	-0.14 %	-2.65%

Table 5.9: The Sharpe-ratio, mean and standard deviation of daily portfolio returns from the three strategies when trading according to the option-expiry structure. The last column shows the annualised Sharpe-ratio.

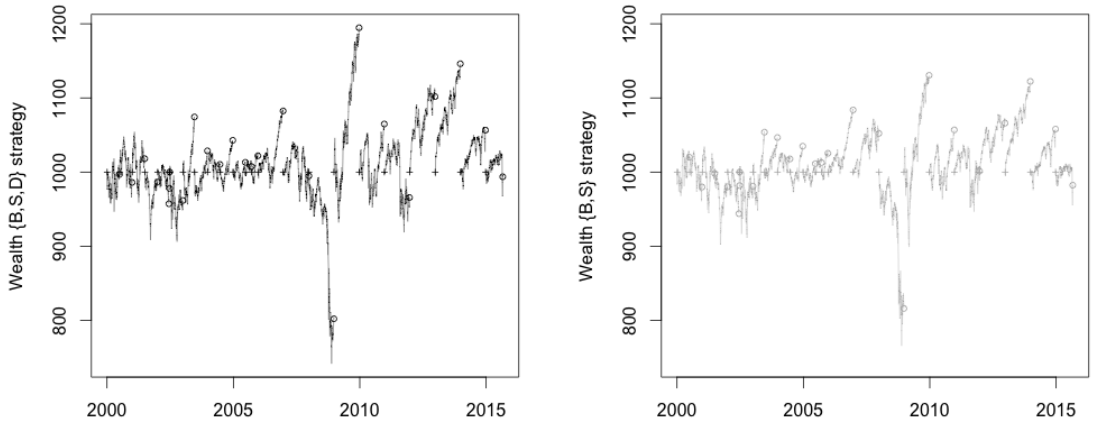

Figure 5.10: Trading according to the option-expiry structure. **Left:** wealth process from trading in {LIBOR,S&P500,Call} according to the optimal investment strategy during investment periods that matches the expiry/strike structure of the call options. A cross indicates the beginning of an investment period (with initial wealth $w_0 = 1,000$) while a circle shows the terminal wealth at the end of the period. **Right:** wealth process from trading in {LIBOR,S&P500} according to Liu's optimal investment strategy during the same investment periods.

Granted: the (B, S, D) strategy does come out victorious in the end, but this is largely happenstance: had we terminated our algorithm during the financial brouhaha, our conclusion would have been different. Indeed, the realised Sharpe ratio obtained vis-à-vis Liu's derivative-free trading strategy is of a very modest nature (0.3 percentage points in annualised Sharpe ratio) and so is the certainty equivalent relative initial wealth.

As for strategy (2) the conclusion is largely invariant, only here the mean return and standard deviation for the (B, S, D) strategy are actually lower than Liu's strategy, jointly leading to a Sharpe ratio of (modest) superiority and negative realised certainty equivalent. We stress that this is about as far as we can go in our analysis here: while strategy (2) on first sight seems to warrant Welchian hypothesis testing (after all, we are seemingly performing the same experiment 25 times), this would be profoundly statistically flawed.

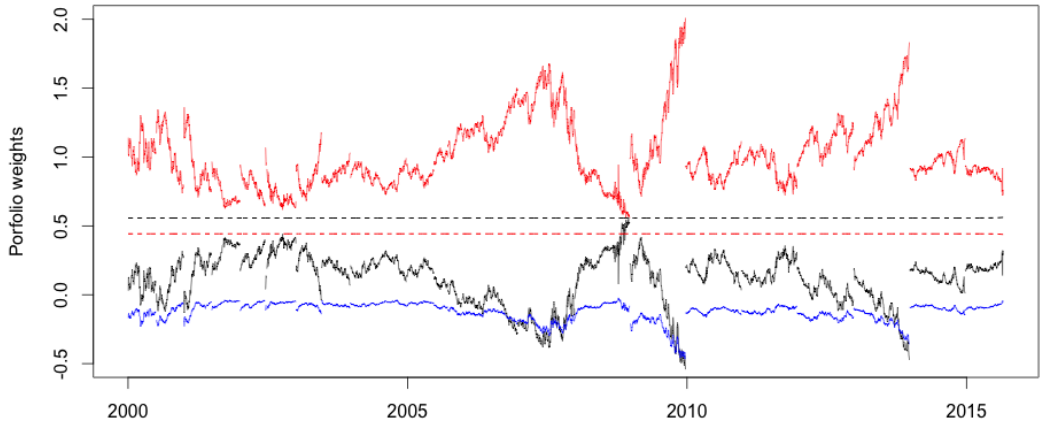

Figure 5.11: Trading throughout 2000-01-03 to 2015-08-31: portfolio weights from the optimal {B,S,D}-strategy (solid lines) and Liu's {B,S}-strategy (dashed lines). Black lines show the weights in B, red lines the weights in S and blue line the weight in D. Notice that the weights of the {B,S,D}-strategy makes sudden jumps at the time-points where the expiry/strike of the option changes, and that we do not trade during these dates. Furthermore, observe that the derivative position is everywhere negative.

Essentially, whenever we sample a wealth path we do so from a different space, where virtually every underlying parameter (bar the risk aversion) is different. Our conclusion is thus of a purely observational nature: we see that the utility maximising trading strategy including derivatives outperforms the one without; yet the financial benefit is *hardly* worth talking about. Furthermore, this statement is obviously contingent upon the overall setting of our experiments: from when we chose to terminate our algorithm, to which derivative product we decided to consider. Indeed, our analysis has shamelessly disregarded transactions costs, which will have a significant impact on our wealth paths. Nevertheless, as a first analysis, the results are certainly noteworthy.

5.6 Conclusion

In the first part of this paper we derived optimal portfolio weights for a utility maximiser who trades in a (B, S, D)-economy in a *generic* stochastic volatility framework, thus extending the work by Liu and Pan. In the second part, we derived explicit expressions for the Heston model, which benefits by admitting closed form expressions for plain vanilla European options. Here, empirically based Monte Carlo simulations suggest that there is no tangible welfare benefit associated with hedging volatility per se: in other words, if our portfolio benefits from the inclusion of derivatives, it does so through shear myopic diversification. Liu and Pan are optimistic on this account: through quasi-empirical con-

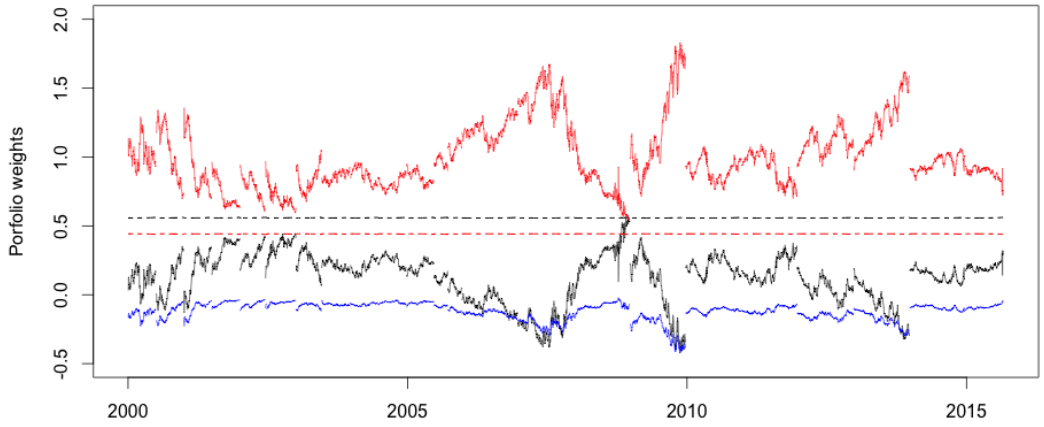

Figure 5.12: Trading according to the option-expiry structure: portfolio weights from the optimal {B,S,D}-strategy (solid lines) and Liu's {B,S}-strategy (dashed lines). Black lines show the weights in B, red lines the weights in S and blue line the weight in D. Notice that the weights of the {B,S,D}-strategy makes sudden jumps at the time-points where the expiry/strike of the option changes, and that the investment portfolio is reset at these dates. Again, the derivative position is everywhere negative.

siderations they find considerable improvements in the certainty equivalent in wealth for investors who trade in derivatives (the reported number is 14.2% for an investor who trades in a straddle position (long call, long put, same strike)). Our own findings, which arguably have a much firmer grounding in empirical data, are considerably more pessimistic: whilst we can extract a certainty equivalent of 1.75% for our long-term trading experiment (and a higher Sharpe ratio than for Liu's (B, S)-strategy) this is of a very modest nature, and arguably a "fluke" brought about by the overall circumstances of our experiment. Obviously, we cannot rule out that more "cognisant" derivative strategies will have a greater impact upon the investor's wealth level: we will leave this for future research.

6

Volatility is Log-Normal, But not for the Reason you Think

Martin Jönsson and Rolf Poulsen[1]

Abstract. Stochastic volatility models have increased enormously in popularity since their introduction in the late eighties. Not the least for hedging and option pricing purposes since they do well in fitting the implied volatility surface. In fact, their pricing ability is often the reason for advocating such a model, whilst their ability to capture the underlying dynamics is loosely motivated. In this paper we test for what is a good model of volatility based on the latter perspective: we briefly review three well-known stochastic volatility models, and concentrate on the instantaneous variance in Heston's model, a log-normal model and in the 3-over-2 model. Since volatility is a non-observable process, we employ the technique of realized volatility to obtain variance measurements and from these we form a goodness-of-fit analysis based on the concept of uniform residuals. To assess the model-classification ability of our analysis, we perform a Monte Carlo study. We then apply the methodology in an empirical study, where our results show that the log-normal model yields a much better goodness-of-fit than both Heston's and the 3-over-2 model.

Keywords: Stochastic Volatility, High-Frequency Data, Goodness-of-fit Analysis.

[1]Both authors are with the Department of Mathematical Sciences, University of Copenhagen.

6.1 Prelude: Any signs of log-normality?

A widespread method for measuring the volatility[2] of some price process S is through an exponentially weighted average of past squared rates of returns

$$\sigma_t^2 = \frac{1-\lambda}{1-\lambda^{N+1}} \sum_{j=0}^{N} \left[\ln\left(\frac{S_{t-j}}{S_{t-(j+1)}} \right) \right]^2 \lambda^j \tag{6.1}$$

see for instance RiskMetrics (1996), Table 5.1, Wilmott (1998), Section 45, or Hull (2009), Chapter 17. Applying this to the last 20 years of daily returns on the S&P 500 index[3] (with the suggested values $\lambda = 0.94$ and $N = 20$) gives the volatility shown in Figure 6.1 and if we estimate the empirical density (i.e. a non-parametrically smoothed histogram) of the volatility, we obtain what is shown in Figure 6.2. In that graph we have also plotted the

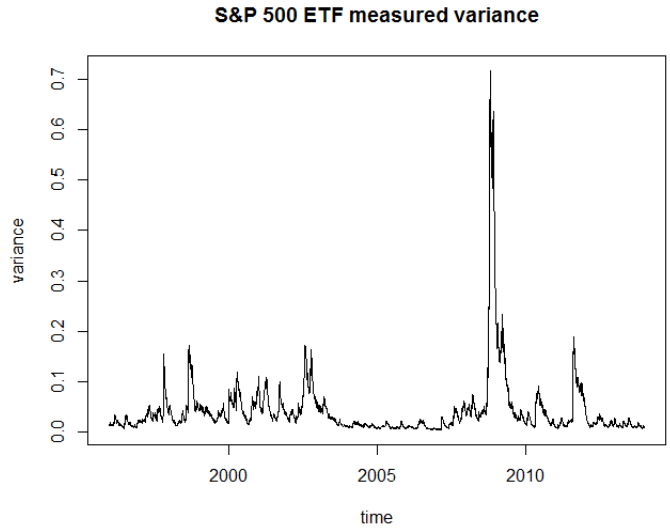

Figure 6.1: Daily variance of the S&P 500 index measured with the EWMA model of equation (6.1) with $\lambda = 0.94$ and $N = 20$.

best-fitting log-normal and gamma densities: the log-normal distribution gives a decent fit, the gamma distribution does not. The most commonly used continuous-time stochastic

[2]We should refer to the quantity on the right hand side of equation (6.1) as "instantaneous variance", multiply it by 252 to annualize, and then define volatility as its square root. For the reason of readability we will drop this distinction in the introduction. Note that log-normal models are stable to roots and squares, so it is not misleading to say "volatility is log-normal" even though we model variance. Heston and the 3-over-2 models do not posses this stability property.

[3]We will use data from the SPDR S&P 500 ETF (SPY) as substitute for the S&P 500 index. The reason for not looking at the index directly is simply a lack of accessible data. The S&P 500 ETF data is retrieved from Wharton Research Data Services, see Section 6.4.

volatility model is that of Heston (1993), in which the variance has a stationary distribution of gamma-type. It thus seems that there is clear empirical evidence against the Heston model and in favour of a log-normal model.

There are, however, a few problems with this reasoning. First, the observations used to estimate the densities in Figure 6.2 are far from independent: their first order auto-correlation is about 0.99. This means that standard distributional tests (such as the Kolmogorov-Smirnov test) are grossly invalid since they assume independent observations and things may look a lot more significant than they actually are. Second, equation (6.1) is not the true continuous-time object that Heston's and other volatility models describe, it is a measured quantity and as such contaminated by noise.

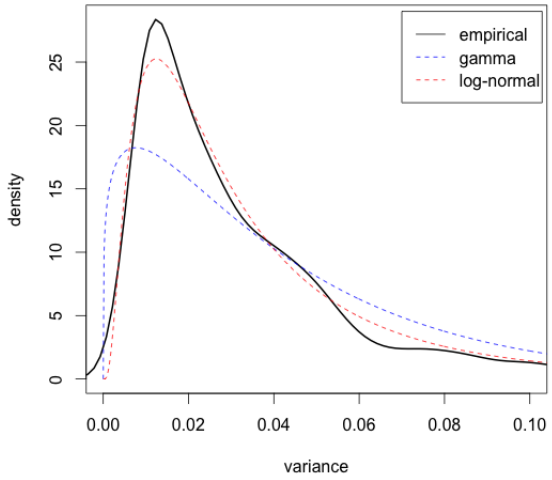

Figure 6.2: Estimated empirical density of the measured S&P 500 variance together with fitted log-normal and gamma densities.

In this paper we attack these problems in order to discover, what we can and cannot say about volatility with statistical certainty. The answers are in the title. More specifically, we look at three continuous-time models – Heston, log-normal, and 3-over-2 – and show that with daily measurements, equation (6.1) (and Figure 6.2 that relies on it) cannot be used to distinguish the models. For all models, unconditional distributions look log-normal and one-day-ahead conditional distributions are too noisy. However, shifting to a 5-minute observation frequency for the volatility measurement, which is quite feasible for data from liquid markets, it is possible to discriminate between the models both in controlled simulation studies and on market data. We find overwhelming support for the log- normal model, both for the S&P 500 index and at an individual stock level, as well as when we take jump corrections into consideration.

Overview. We review three stochastic volatility models in Section 6.2.1 with focus on their

process laws as given by conditional distributions. This to form statements about the fit of a model that take distributional properties and temporal dependence into consideration. We revisit some methods for parameter estimation in Section 6.2.2 and variance measurement in Section 6.2.3 that will lie at the base of our goodness-of-fit analysis, outlined in Section 6.2.4. To address the question of credibility of our study, we perform a controlled numerical assessment in Section 6.3 before we apply the methodology to real market data in the empirical study of Section 6.4. Section 6.5 concludes.

6.2 Models, estimation and test

6.2.1 Continuous time stochastic volatility models

To fix some notation, we assume that the price of an asset can be represented by a stochastic process $S = (S(t))_{t\geq 0}$ that satisfies the stochastic differential equation

$$dS(t) = \mu S(t)dt + \sqrt{V(t)}S(t)dZ(t) \tag{6.2}$$

where μ is the drift coefficient, $Z = (Z(t))_{t\geq 0}$ is a Wiener process and $V = (V(t))_{t\geq 0}$ is the instantaneous variance process which will be our main modelling target. Specifically, we consider the following model specifications:

Log-normal model. For this model, let $V(t) = \exp X(t)$ where $X = (X(t))_{t\geq 0}$ is a mean-reverting Gaussian process (also known as an Ornstein-Uhlenbeck or Vasicek process) satisfying

$$dX(t) = \kappa(\theta - X(t))dt + \varepsilon dW(t)$$

with mean-reversion speed κ, long-term mean θ and diffusion parameter ε, while $W = (W(t))_{t\geq 0}$ is another Wiener process correlated to Z with correlation coefficient ρ (see for instance Scott (1987) for an early account of this model). By applying Ito's formula to $X(t)e^{\kappa t}$, one obtains after rearrangement

$$X(t) = X(s)e^{-\kappa(t-s)} + \theta(1 - e^{-\kappa(t-s)}) + \varepsilon \int_s^t e^{-\kappa(t-u)}dW(u)$$

such that X is a Gaussian process (making V log-normal) with conditional distribution

$$X(t+h)|X(t) \sim \mathcal{N}\left(X(t)e^{-\kappa h} + \theta(1 - e^{-\kappa h}), \frac{\varepsilon^2(1 - e^{-2\kappa h})}{2\kappa}\right) \tag{6.3}$$

and with a stationary distribution that is normal with mean θ and variance $\varepsilon^2/(2\kappa)$.

Heston's model. As proposed by Heston (1993), the instantaneous variance V follows a square-root process (also known as Feller or Cox-Ingersoll-Ross process)

$$dV(t) = \kappa(\theta - V(t))dt + \varepsilon\sqrt{V(t)}dW(t) \tag{6.4}$$

with mean reversion κ, long-term mean θ and diffusion parameter ε. Taking conditional expectation of the integrated version of (6.4) leads to a first order differential equation for the conditional mean. The solution is

$$\mathbb{E}(V(t+h)|V(t)) = V(t)e^{-\kappa h} + \theta(1-e^{-\kappa h}) \tag{6.5}$$

while the conditional variance

$$\mathrm{Var}(V(t+h)|V(t)) = V(t)\frac{\varepsilon^2}{\kappa}(e^{-\kappa h} - e^{-2\kappa h}) + \theta\frac{\varepsilon^2}{2\kappa}(1-e^{-\kappa h})^2 \tag{6.6}$$

is derived by taking conditional expectation of $dX^2(t)$ (obtained by Ito's formula) and solving the corresponding equation (see for instance Cairns (2004) for details). The conditional distribution of V in Heston's model is a non-central chi-squared distribution with $df = 4\kappa\theta/\varepsilon^2$ degrees of freedom, non-centrality parameter $\nu(t) = V(t)\cdot 4\kappa e^{-\kappa h}/(\varepsilon^2(1-e^{-\kappa h}))$ and scaling parameter $c = 4\kappa/(\varepsilon^2(1-e^{-\kappa h}))$, that is

$$c\cdot V(t+h)|V(t) \sim \chi^2\left(df, \nu(t)\right)$$

see Cox et al. (1985). If the parameters satisfy the Feller condition $2\kappa\theta > \epsilon^2$, then V will have a stationary gamma distribution with shape parameter $df/2$ and scale parameter $\varepsilon^2/(2\kappa)$.

The 3-over-2 model. The dynamics of the instantaneous variance in the 3-over-2 model is given by

$$dV(t) = V(t)\kappa(\theta - V(t))dt + \varepsilon V(t)^{3/2}dW(t)$$

where κ, θ and ε are parameters. By applying Ito's formula to $Y(t) = 1/V(t)$ we see that

$$dY(t) = \kappa\theta\left(\frac{\kappa+\varepsilon^2}{\kappa\theta} - Y(t)\right) - \varepsilon\sqrt{Y(t)}dW(t) \tag{6.7}$$

and thus $1/V$ follows a square root process with reversion speed $\tilde{\kappa} = \kappa\theta$ and long-term mean $\tilde{\theta} = (\kappa+\varepsilon^2)/(\kappa\theta)$. We will refer to the model specified by (6.7) as the *reciprocal 3-over-2 model* with reciprocal parameters $(\tilde{\kappa}, \tilde{\theta}, \tilde{\varepsilon})$.[4] This implies that V will have a stationary reciprocal-gamma distribution with shape parameter $2\tilde{\kappa}\tilde{\theta}/\varepsilon^2$ and rate parameter $2\tilde{\kappa}/\varepsilon^2$. For a treatment of the 3-over-2 model including the derivation of option-pricing formulas, see Heston (1997) and Lewis (2000).

6.2.2 Estimation of discretely observed diffusions

Suppose for now that we have observed the instantaneous variance process, or some non-parameter-dependent transformation of it, at discrete points in time $t_0, \ldots, t_N$ where the

[4]Note that if the Feller condition holds for Y then $\mathbb{P}(Y(t) > 0) = 1$ for all $t \geq 0$ which implies that the variance process $V = 1/Y$ will take finite values almost surely. Thus, we should require that this is the case. Indeed: $2\tilde{\kappa}\tilde{\theta} > \varepsilon^2$ is equivalent to $2\kappa > -\varepsilon^2$ for the parameters of V and this condition is assumed to hold by definition.

spacings $h_i = t_{i+1} - t_i$ are one day or a few minutes for high-frequency data. An observed path $x = (x_0, \ldots, x_N)$ of the variance/transformed process, generically denoted X, is thus a set of $N+1$ observations where x_i is the observed value of $X(t_i)$ for $i = 0, \ldots, N$.

For the log-normal model with parameters $\psi = (\kappa, \theta, \varepsilon)$, the role of X is played by $\ln(V)$, and we can write the log-likelihood as

$$l(\psi; x) = \sum_{i=0}^{N-1} \log \phi(x_{i+1}; m_{i+1|i}, \sigma^2_{i+1|i}) \tag{6.8}$$

where ϕ denotes the normal density, here with (conditional) mean $m_{i+1|i}$ and variance $\sigma^2_{i+1|i}$ as given in (6.3). The maximum likelihood estimate $\hat{\psi} = (\hat{\kappa}, \hat{\theta}, \hat{\varepsilon})$ is the argument that maximizes the log-likelihood (6.8). It may be found in closed form if the observations are equidistant, and easily by a numerical optimizer if they are not. Calculating the observed information matrix

$$I_o = -\frac{\partial^2 l}{\partial \psi^T \partial \psi}\bigg|_{\psi=\hat{\psi}}$$

by numerical differentiation at estimated values gives an estimated standard error of the j$^{\text{th}}$ parameter as $\sqrt{(I_o^{-1})_{j,j}}$ where $(A^{-1})_{i,j}$ denotes element i, j of the inverse of a matrix A.

For the Heston model, the role of X is played by V itself. Optimization on the non-central chi-squared distribution that enters the model's log-likelihood is not possible in closed form, and numerically it is delicate and slow. A simple way to obtain a consistent estimator of $\psi = (\kappa, \theta, \varepsilon)$ is by plugging the conditional moments into a Gaussian likelihood and maximizing. In other words, this means using the approximate log-likelihood

$$l^a(\psi; v) = \sum_{i=0}^{N-1} \log \phi(v_{i+1}; m_{i+1|i}, \sigma^2_{i+1|i}), \tag{6.9}$$

where the conditional mean and variance are given by equations (6.5)-(6.6). The reason to why this gives consistent estimators is that the first order conditions for optimization of (6.9) are a set of so-called martingale estimating equations, see Sørensen (1999). As for the previous model, we approximate standard errors of the estimated parameters from the observed information matrix of the log-likelihood by numerical differentiation.

In a similar fashion for the (reciprocal) 3-over-2 model, the approximate log-likelihood (6.9) with $1/V$ in the role of X can be optimized to obtain maximum likelihood estimates of the reciprocal parameters $(\tilde{\kappa}, \tilde{\theta}, \tilde{\varepsilon})$. These estimates may then be transformed to estimates of the original parameters through $\kappa = \tilde{\kappa}\tilde{\theta} - \tilde{\varepsilon}^2$ and $\theta = \tilde{\kappa}/(\tilde{\kappa}\tilde{\theta} - \tilde{\varepsilon}^2)$ while $\varepsilon = \tilde{\varepsilon}$. Standard errors of the reciprocal parameters may be calculated from the observed information matrix such that standard errors of the original parameters are obtained by the delta method.

Alternatively, we may re-parametrize (6.9) to express a likelihood function with respect to the original parameters of V

$$\hat{l}^a(\kappa, \theta, \varepsilon; v) = l^a(\tilde{\psi}(\kappa, \theta, \varepsilon); v) \tag{6.10}$$

where $\tilde{\psi}(\kappa, \theta, \varepsilon)$ is short-hand notation for the inverse transformations $\tilde{\kappa} = \kappa\theta$ and $\tilde{\theta} = (\kappa+\varepsilon^2)/(\kappa\theta)$. Numerical optimization of the approximate log-likelihood (6.10) with respect to $(\kappa, \theta, \varepsilon)$ may then be performed to obtain maximum likelihood estimates of the original parameters directly.

6.2.3 Measuring volatility

Contrary to the simplifying assumption in the previous subsection, the instantaneous variance process V is not directly observable and in practice it must be measured from observations of the asset price process. One method for this, as already mentioned in the introduction, is given by the EWMA measure (6.1). Another method that also provides quantitative statements about the accuracy of the measurement comes by following Andersen and Benzoni (2009) in a study of so-called realized volatility.

To this end, denote the logarithmic asset price by $Y(t) = \log S(t)$ and the continuously compounded return over a measurement interval $[t-k, t]$ by $r(t, k) = Y(t) - Y(t-k)$. Applying Ito's formula with S following equation (6.2) then gives

$$dY(t) = \alpha(t)dt + \sqrt{V(t)}dZ(t)$$

where $\alpha(t) = \mu - \frac{V(t)}{2}$ such that the continuously compounded return over $[t-k, t]$ amounts to

$$r(t, k) = \int_{t-k}^{t} \alpha(u)du + \int_{t-k}^{t} \sqrt{V(u)}dZ(u).$$

From this, we see that the quadratic variation of the log price over the same time interval is given by

$$QV(t, k) = \int_{t-k}^{t} V(u)du. \tag{6.11}$$

For a partition of $[t-k, t]$, for instance $\{t-k+\frac{j}{n}, j = 0, \ldots, nk\}$, the realized volatility is defined by

$$RV(t, k; n) = \sum_{j=1}^{nk} r\left(t-k+\frac{j}{n}, \frac{1}{n}\right)^2$$

and semimartingale theory ensures uniform convergence in probability

$$RV(t, k; n) \longrightarrow QV(t, k) \text{ as } n \to \infty. \tag{6.12}$$

The point is, for high-frequency returns (a large n) we have that RV gives a good approximation of QV and we may use that $\int_{t-k}^{t} V(u)du \approx V(\bar{t}) \cdot k$ for some $\bar{t} \in [t-k, t]$ to obtain a measured variance path. Thus, to adopt our notation, for an asset log-price Y observed at time points $t_0, \ldots t_N$ and a measurement interval $[t_{i-n}, t_i]$ (say one day) such that the observation frequency is set by n (say every 5 minutes), we calculate

$$v_i = \frac{1}{t_i - t_{i-n}} \sum_{j=1}^{n} \left(y_{i-n+j} - y_{i-n+(j-1)}\right)^2 \tag{6.13}$$

for $i = n, 2n, \dots, N$ as the measured variance path $v = (v_n, v_{2n}, \dots, v_N)$ of the non-observable variance $V(t_n), V(t_{2n}), \dots$. Note that we obtain daily variance measurements while we require asset price observations of a much higher frequency. Of course, we may reduce n to get more frequent variance measurements but for a cost of poorer approximation with respect to the convergence in (6.12).

Remark 6.1. Based on the same reasoning as for realized volatility, we may employ the measurement technique for estimating the vol-of-vol parameter ε. For Heston's model, let $Y(t) = \sqrt{V(t)}$ and apply Ito's formula to obtain

$$dY(t) = \left(\frac{\kappa}{2Y(t)}(\theta - Y^2(t)) - \frac{\varepsilon^2}{8Y(t)} \right) dt + \frac{\varepsilon}{2} dW(t)$$

with quadratic variation $QV([0,T]) = \varepsilon^2 T/4$ over the interval $[0,T]$. This leads us to the vol-of-vol estimator

$$\hat{\varepsilon}^2 = \frac{4}{T} \sum_{t_i \in [0,T]} (Y(t_{i+1}) - Y(t_i))^2 \tag{6.14}$$

based on the "realized volatility" of the transformed variance process over $[0,T]$. In a similar fashion for the 3-over-2 model, $Y(t) = 1/\sqrt{V(t)}$ gives by Ito's formula

$$dY(t) = (\dots)dt - \varepsilon/2 \cdot dW(t)$$

such that equation (6.14) can be reused for an estimator. For the log-normal model, $Y(t) = \log V(t)$ yields a diffusion term $\varepsilon dW(t)$ and quadratic variation $QV([0,T]) = \varepsilon^2 T$ so that equation (6.14) divided by 4 gives the desired vol-of-vol estimator.

The realized volatility measure relies on the fact that S is described by a stochastic process following (6.2) that does not have any jumps[5] such that the convergence (6.12) provides an approximation of the integrated variance due to (6.11). However, one may argue that this is a somewhat simplified description of reality when looking at market data, and call for a model that allows for jumps in the asset price path as well.[6] One way of doing so, as described by Andersen and Benzoni (2009), is to add a jump process to the price dynamics

$$dS(t) = \mu S(t)dt + \sqrt{V(t)}S(t)dZ(t) + (e^{\xi(t)} - 1)S(t^-)dN(t)$$

where $N = (N(t))_{t \geq 0}$ is a Poisson process uncorrelated to Z with finite intensity, while $\xi = (\xi(t))_{t \geq 0}$ are random variables that determine the magnitudes of the jumps: $\Delta S(t) = S(t) - S(t^-) = S(t^-)e^{\xi(t)}$. Ito's formula for jump-diffusions applied to $Y(t) = \log S(t)$ then gives

$$dY(t) = \alpha(t)dt + \sqrt{V(t)}dZ(t) + \xi(t)dN(t) \tag{6.15}$$

[5]Note that S being the solution of (6.2) has continuous sample paths also in the case when V displays jumps.

[6]Even if the question of discontinuities for asset prices has been under scrutiny for a long time, the assumption of continuous price paths is perhaps not as restrictive as previously believed, see Christensen et al. (2014).

where the jump term $\xi(t)dN(t)$ is zero except when a jump occurs at time t. It should be understood in integral form

$$\int_0^t \xi(u)dN(u) = \sum_{T_i \in [0,t]} \xi(T_i)$$

where $T_1, \ldots, T_{N(t)}$ are the jump times of N during the time interval $[0, t]$. Under these dynamics, the quadratic variation of the log-price Y following (6.15) includes both the integrated variance and a cumulative squared jump part

$$QV(t,k) = \int_{t-k}^t V(u)du + \sum_{T_i \in [t-k,t]} \xi^2(T_i)$$

while the convergence of the realized volatility measure

$$RV(t,k;n) \longrightarrow QV(t,k) \text{ as } n \to \infty.$$

still holds (ucp). Thus, an estimator is needed for the integrated variance part separately and here we employ the *realized bipower variation* as introduced by Barndorff-Nielsen and Shephard (2004)

$$BV(t,k;n) = \frac{\pi}{2} \sum_{j=2}^{nk} |r\left(t-k+\frac{j}{n}, \frac{1}{n}\right)||r\left(t-k+\frac{j-1}{n}, \frac{1}{n}\right)|.$$

This measure provides a consistent estimate of the variance part that is robust to jumps. This also gives an estimator of the cumulative squared jump part since we have

$$RV(t,k;n) - BV(t,k;n) \longrightarrow \sum_{T_i \in [t-k,t]} \xi^2(T_i) \text{ as } n \to \infty$$

from which the quantity on the left hand side can be employed as an estimator.

6.2.4 Goodness of fit and uniform residuals

A goodness-of-fit analysis that takes into consideration the full conditional structure of a diffusion process X uses the so-called *uniform residuals*, see Pedersen (1994). With $F_{i+1|i}(x;\psi)$ denoting the conditional distribution function of $X(t_{i+1})|X(t_i)$, the uniform residuals are defined by the probability transformation

$$U_{i+1} = F_{i+1|i}(X(t_{i+1});\psi), \;\; i = 0, \ldots, N-1.$$

The random variables $U_1, \ldots, U_N$ will be independent and identically $U(0,1)$ distributed if the true distribution of X is that of the family $\{F_{i+1|i}, i = 0, \ldots, N-1\}$, a result that goes back at least to Rosenblatt (1952). The proof is straightforward: if the random variable X has a strictly increasing, continuous distribution function F, then the random variable

$F(X)$ will be uniformly distributed over the unit interval. This fact, combined with the Markov property of diffusion processes and an application of iterated expectations, then give that $\{U_i\}$ is a family of independent random variables.

For the log-normal model, we calculate uniform residuals from a variance path v as

$$u_{i+1} = F_N\left(x_{i+1}; x_i e^{-\kappa h_i} + \theta(1 - e^{-\kappa h_i}), \frac{\varepsilon^2}{2\kappa}(1 - e^{-2\kappa h_i})\right) \tag{6.16}$$

where $x_i = \log v_i$ and with F_N denoting the normal distribution function. For Heston's model, we use the true conditional non-central chi-squared distribution function F_{χ^2} and calculate residuals as

$$u_{i+1} = F_{\chi^2}\left(c_i v_{i+1}; \mathrm{df} = \frac{4\kappa\theta}{\varepsilon^2}, \mathrm{ncp}_i = c_i v_i e^{-\kappa h_i}\right) \tag{6.17}$$

where $c_i = 4\kappa/(\varepsilon^2(1-e^{-\kappa h_i}))$. Similarly, for the 3-over-2 model (6.17) gives the corresponding uniform residuals when v is substituted with $1/v$ and the parameters are substituted with $\tilde{\kappa} = \kappa\theta$ and $\tilde{\theta} = (\kappa + \varepsilon^2)/(\kappa\theta)$ from $(\kappa, \theta, \varepsilon)$ being the 3-over-2 estimates (or directly from the reciprocal parameters).

To test if a family of observed residuals $(u_1, \ldots, u_N)$ – calculated with the probability transformation corresponding to one of the models – are independently $U(0,1)$-distributed, we employ a quantile-quantile plot, which is a standard diagnostic tool. We also look at a standard chi-square test according to the following: (i) divide the interval $[0,1]$ into n_b subintervals of equal sizes, "bins", (ii) calculate $E_j = n_b/n$, the theoretical frequency of observations in bin j, (iii) calculate O_j, the observed number of observations respectively. Under the null hypothesis

$$H_0 : (u_1, \ldots, u_N) \text{ are independent observations from the uniform distribution}$$

we then have that the test statistic

$$\sum_{j=1}^{n_b} \frac{(O_j - E_j)^2}{E_j} \tag{6.18}$$

is chi-square distributed with $n_b - 1$ degrees of freedom. As an alternative, the Kolmogorov-Smirnov test yields that the test statistic

$$\sqrt{n} \sup_x |\hat{F}_n(x) - F_U(x)|$$

converges to the Kolmogorov-Smirnov distribution under the same null hypothesis, where $\hat{F}_n$ denotes the estimated empirical distribution function while F_U is the uniform distribution over $[0,1]$.

To summarise, we perform the following steps for a goodness-of-fit analysis:

1. Measure variance $v = (v_1, \ldots v_N)$ by the realized volatility measure (6.13).

2. For each model: obtain $(\hat{\kappa}, \hat{\theta}, \hat{\varepsilon})$ from the maximum likelihood estimator (6.8) or (6.9).

3. For each model: calculate residuals $u = (u_1, \ldots, u_N)$ from (6.16) or (6.17) with $(\hat{\kappa}, \hat{\theta}, \hat{\varepsilon})$.

4. Perform quantile diagnostics and/or a test under H_0 : residuals are iid $U(0, 1)$.

In the next section, we will apply these steps in a controlled simulation study to see if we can tell with certainty, based on quantile diagnostics and test results, which model is the correct one.

6.3 Simulation study

In what follows, we apply our goodness-of-fit test described above to simulated data. We generate asset price and variance paths from each of the three models with the simulation methods described in the appendix, Section 6.5.1. With properties of the market data in mind we use the following specification: we simulate each path with 362,881 observations to mimic a time grid of length $T = 15$ years with equidistant spacing $\Delta = 5$min. This corresponds to 252 trading days per year, 8h per day and 12 observations per hour. We set $S_0 = 100$ and $\mu = 0.05$ for the asset price, a correlation $\rho = -0.20$ for the driving Wiener process and $V_0 = 0.04$ for the initial variance along with the parameters given in Table 6.3.

	κ	θ	ε
Heston	7.00	0.035	0.92
log-normal	71.9	-3.89	12.9
3-over-2	3.65	17.1	111.5

Table 6.3: Model parameters for the variance processes used for the simulation study.

Figure 6.4 shows simulated asset price and variance from Heston's model[7] along with realized volatility (RV) with $n = 192$ price observations for each variance measurement. This yields 1,890 measured observations in total. Parameter estimates obtained from optimizing the likelihood based on simulated variance data[8] yields $(\hat{\kappa}, \hat{\theta}, \hat{\varepsilon}) = (10.0, 0.030, 0.94)$

[7]We employ the implicit Milstein scheme for Heston's model (note that $4\kappa\theta > \varepsilon^2$ holds), the exact variance scheme for the log-normal model and the implicit Milstein of the reciprocal variance for the 3-over-2 model ($4\tilde{\kappa}\tilde{\theta} > \varepsilon^2$ hold as well for reciprocal parameters) – see Section 6.5.1 in the appendix for details.

[8]Here, we use a subset of data taken at every 192$^{\text{nd}}$ index of the original variance path. This to estimate parameters from a data set that corresponds to the measured variance at the measurement time-points.

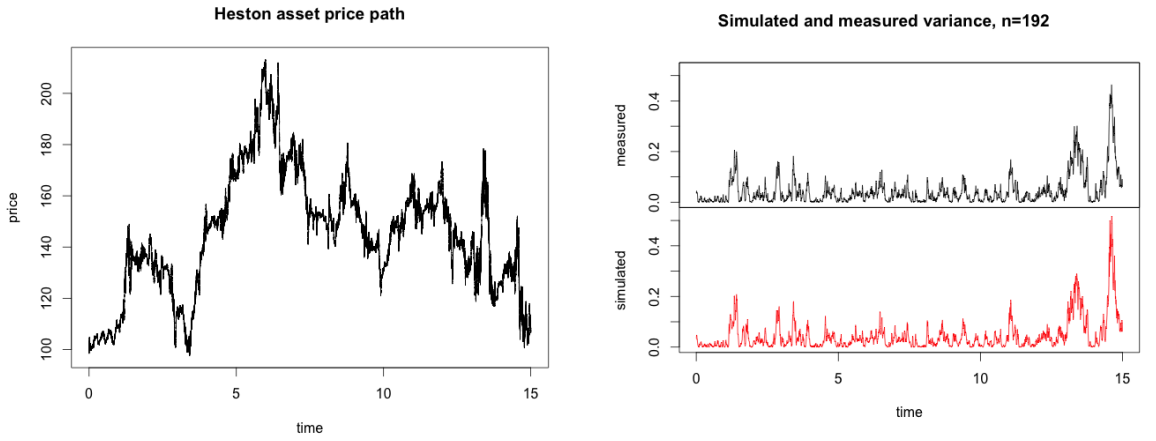

Figure 6.4: Left figure: A simulated price path from Heston's model with the implicit Milstein scheme. $T = 15$ years with 252 days per year and $n = 96$ points per day (every 5 minutes) giving 362,881 observations. **Right figure:** simulated (black) and RV-measured variance (red) with $n = 192$ prices per variance measurements, giving 1,890 observation.

while the RV-measured variance results in the estimates $(\hat{\kappa}, \hat{\theta}, \hat{\varepsilon}) = (5.78, 0.030, 0.82)$. Estimated parameters are then used in the next step to calculate uniform residuals with resulting uniform quantile plots shown in Figure 6.5. The quantile plots exhibit good fits for both sets of residuals which is also confirmed by a chi-square test based on $n_b = 250$ bins: the p-value is 0.79 for the simulated variance path and 0.57 for the measured data. This means that we can not reject the null-hypothesis, residuals are uniformly distributed, which implies that Heston's model shows a strong goodness-of-fit, for both the simulated and measured data.

If we redo the procedure but with calculated residuals (and estimated parameters) corresponding to the log-normal and 3-over-2 model, we obtain the results in Table 6.6. The

P-value:	Simulated variance	Realized volatility	EWMA measure
Heston	0.79	0.57	0
log-normal	0	0	0
3-over-2	0	0	0

Table 6.6: Results from the chi-square test of uniform residuals calculated from the three different model assumptions on underlying data simulated from Heston's model.

table is supplemented with the EWMA measure ($\lambda = 0.94$) based on daily asset prices which yields zero p-values for all models. In all, it is clear that the log-normal and 3-over-2 model have no support in terms of our goodness-of-fit analysis. Further, if we use the

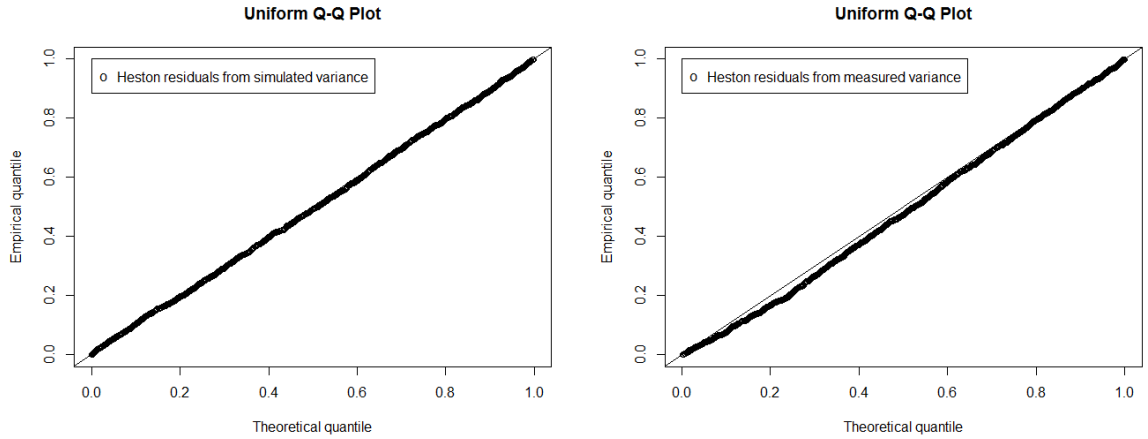

Figure 6.5: Quantile plots of Heston uniform residuals calculated from a simulated variance path (**left**) and from the corresponding measured variance (**right**).

largest p-value as a model discrimination criterion based on RV-measured data we would be able to distinguish between the three models.

In a similar fashion as above we run the test procedure on generated data from the log-normal model. Estimation of log-normal parameters from simulated variance yields $(\hat{\kappa}, \hat{\theta}, \hat{\varepsilon}) = (71.7, -3.88, 13.0)$ while measured variance gives $(\hat{\kappa}, \hat{\theta}, \hat{\varepsilon}) = (46.0, -3.79, 9.57)$. The p-values of calculated uniform residuals are reported in Table 6.7 along with the Heston and 3-over-2 models. The result supports the hypothesis that variance data, both

P-value:	Simulated variance	Realized volatility	EWMA measure
Heston	0	0	0
log-normal	0.72	0.78	0
3-over-2	0	0	0

Table 6.7: Results from the chi-square test of uniform residuals calculated from the three different model assumptions on underlying data simulated from the log-normal model.

simulated and measured, fits well with being log-normal, but not with Heston or the 3-over-2 model. Based on p-values we can discriminate between the models for the log-normal case as well.

Data generated from the 3-over-2 model yields parameter estimates $(\hat{\kappa}, \hat{\theta}, \hat{\varepsilon}) = (15.6, 3.9, 110.3)$ and $(\hat{\kappa}, \hat{\theta}, \hat{\varepsilon}) = (16.8, 2.52, 89.2)$ respectively, which results in the p-values given in Table 6.8. The results show support for the goodness-of-fit of the 3-over-2 model (both for simulated and measured variance) but not for Heston or the log-normal model and once again

P-value:	Simulated variance	Realized volatility	EWMA measure
Heston	0	0	0
log-normal	0	0	0
3-over-2	0.90	0.26	0

Table 6.8: Results from the chi-square test of uniform residuals calculated from the three different model assumptions on underlying data simulated from the 3-over-2 model.

is it possible to predict the correct model based on the largest p-value from residuals calculated with the realized volatility.

We conclude that in all three cases we get support for the correct underlying simulation model while the alternatives do not get any support in terms of model fit. This might not be as surprising for simulated data as for the realized volatility measure – here we have potential sources of errors in several steps that may affect the distributional properties in a discommoding way – and we can in fact discriminate between the three models. We also note that variance measured with the EWMA model yields zero p-values in all considered cases, most likely since this is a measure being too crude for the preservation of tractable distributional properties.

To examine the *robustness* of this goodness-of-fit analysis, and in particular its model discrimination power, we run 100 repetitions of the above procedure with each of the models as underlying. In each repetition, we calculate p-values based on uniform residuals corresponding to the three alternatives and record which model achieves the highest p-value. As a result, we obtain a robust 100% prediction hit-rate (that is, the correct model is identified based on the highest p-value obtained among the three models) in all three cases: when uniform residuals are calculated from RV-measured variance of the log-normal, Heston's and the 3-over-2 model, our goodness-of-fit analysis always points us to the correct model. Graphs of recorded p-values from all 100 repetitions can be found in the appendix, Figure 6.24. Similarly as in the one-case study above, we also observe zero (or negligible) p-values for the alternative models, for every repetition[9].

Remark 6.2. If we look at corresponding p-values from the Kolmogorov-Smirnov test instead, the results remain unchanged with correct model predictions for all 100 repetitions. A minor difference between the results is that Heston simulations obtain relatively lower

[9]We observe quite some variation in realized p-values of the correct model across different repetitions, regardless of underlying model (the alternatives obtain stable values ∼zero). To put this p-value variation into perspective, we perform 100 additional simulations with draws directly from the uniform distribution. The resulting p-values, shown in the appendix, exhibit a comparable degree of variation as well, even though we have drawn the residuals directly from the true distribution. We therefore remind that we should not rely to hard on the actual magnitude of the p-values, and that diagnostics from quantile plots make an important supplement.

p-values for the true model when compared to the chi-squared test, while log-normal simulations lead to relatively higher p-values for the correct underlying.

Remark 6.3. The above Monte Carlo study was performed with simulations schemes that introduce bias to the discrete-time approximations. First, the simulation of all three models employ the Euler-Maruyama scheme of equation (6.19) for the log-asset price which may disrupt the distributional properties of the simulated price paths. Secondly, the implicit Milstein scheme introduces bias to the simulated variance process of the Heston and the 3-over-2 model. As a result we present simulated data to our goodness-of-fit analysis that might affect several steps of the test procedure in an adverse way. Regardless, our test procedure preforms with a 100% success-rate on the simulated data. The almost exact simulation schemes of Section 6.5.1 do not introduce bias to the asset price although they include approximations for the integrated variance processes. With this in mind we repeat the full Monte Carlo study with 100 repetitions based on each model with the almost exact simulation schemes. The results are practically identical to the previous results with Euler/Milstein based schemes: the correct model achieves the highest p-value in all repetitions, for all three models, whilst the alternative models achieve zero support throughout.

Finally, we investigate how robust the goodness-of-fit analysis is with respect to (i) the accuracy of the realized volatility measurement and (ii) the total number of asset price observations. For the first point, we see that a resolution of $n > 50$ gives 100% correct model predictions (based on 100 repetitions) for all three model while for the second point, we require at least three years of high-frequency data to predict the correct model. See Section 6.5.2 in the appendix for details.

6.3.1 Daily EWMA measured variance revisited

To return to the question of creditability of the introductory analysis in Section 6.1, we perform a mimicking experiment and apply the exponentially weighted average measure to 4,532 daily prices simulated from Heston model. Figure 6.9 shows fitted empirical distributions based on EWMA measured variance and most likely these curves will lead us to the same conclusion as for the S&P 500 data: the distribution looks log-normal! This is the case even though the underlying is Heston and it should indeed encourage us to question the preluding conclusion about the log-normality of S&P 500 variance.

Immediately, as we have a large number of observations (4,532 compared to 1,890 used in the previous study) the failing link is the accuracy of the EWMA measure, not an insufficient number of observations: the root mean square error of EWMA measured variance is 0.037 compared to 0.009 for the realized volatility of Figure 6.4. Further, if we employ our goodness-of-fit analysis with the EWMA measure along with the corresponding true variance, we get the results in Table 6.10. The EWMA measure credits Heston's model with zero p-value while the simulated variance points to the correct model. Note that the log-normal model, which has our support from a visual diagnose of Figure 6.9, (at least

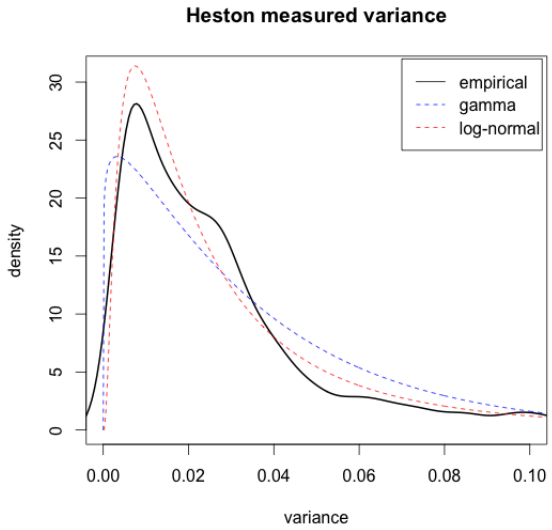

Figure 6.9: Fitted empirical, log-normal and gamma densities to EWMA measured variance with $\lambda = 0.94$ of 4,532 simulated prices with Heston's model.

P-value:	Simulated variance	EWMA measure
Heston	0.39	0
log-normal	0	0
3-over-2	0	0

Table 6.10: Heston's model with 4,532 observations: results from the chi-square test of uniform residuals calculated from the three different model assumptions for the underlying data.

for a fitted log-normal *distribution*) obtains no support based on the EWMA measure. Similar results apply if we use simulated data from the log-normal and 3-over-2 model, no support to any of the models from the EWMA measure with 4,532 observations while the true variance points us to the correct model in both cases.

This simple experiment shows that there is not much hope of predicting the correct underlying when using the exponentially weighted average measure, nor can we tell which model is true from a diagnose of empirical distributions – this will even lead to false conclusions.[10] On the other hand, when we measure variance from $\sim 360,000$ price observations and apply our goodness-of-fit analysis, we have good reasons to believe that we can indeed

[10]Indeed, if we repeat a 1000 times the steps leading to Figure 6.9 and record the closest distribution in L^2-distance, we have that the log-normal function is closest to the empirical in 73% of the cases.

predict a model that nicely fits the data with a significant p-value. Even more importantly, we have good reasons to believe that a poorly fitting model will not get any support at all from the test. In the next section, we apply these findings and reconsider the S&P 500 data with our sharper analysis based on high-frequency price quotes.

6.4 Empirical study: horserace between models

Our empirical analysis is based on market data from the S&P 500 index and ten different securities listed on the New York Stock Exchange (NYSE). The data is retrieved from Wharton Research Data Services[11] and it consists of intraday *tick-by-tick* trade quotes. This raw data is, however, not directly suited for analysis and we apply a series of preprocessing methods before our analysis. The first method is a cleaning procedure as proposed by Barndorff-Nielsen et al. (2009). First, entries with a zero-price are deleted along with entries with an abnormal sale condition as indicated by the code of the trade. The dataset is then restricted to contain trades from the exchange where the security is listed. Next, entries with the same time stamp are replaced by their median price.

The cleaned data will be on an irregular tick-by-tick time scale. In the second preprocessing step, we aggregates prices to an equidistant 5 minute time-grid by taking the last realized price before each grid point. Aggregation to a 5 minute frequency is motived by the fact that price processes are increasingly perturbed by micro-structure noise for higher frequencies. As a final step, the data is restricted to exchange trading-hours, being 9:30 am to 4:00 pm from Monday to Friday.

With this data in hand, we apply our goodness-of-fit analysis to the SPDR S&P 500 ETF, which we also analysed in the introduction. The preprocessed data contains 368,724 price quotes observed with a 5 minute frequency from the period 1996-01-02 to 2013-12-31 and a graph of the historical prices is depicted in the left pane of Figure 6.11. In the first step of our analysis, we apply the realized volatility measure with resolution $n = 192$ and a plot of the resulting 1,920 variance observations is shown in the right pane of Figure 6.11 (cf. Figure 6.1 for the daily exponentially weighted average variance). Numerical optimization is then performed of the likelihood function from each model with resulting parameter estimates and standard errors recorded in Table 6.12. In the next step, estimated parameters are used to calculate uniform residuals from the probability transform of respective model with resulting quantile-quantile plots depicted in Figure 6.13.

From the plots, we see a perfect fit of the log-normal residuals to the standard uniform distribution while the Heston and 3-over-2 model exhibit very poor fits. From a visual diagnose we thus have good reasons to prefer the log-normal model for the S&P 500

[11] `https://wrds-web.wharton.upenn.edu/wrds`

[12] These estimates and errors are obtained from the reciprocal specification of the 3-over-2 model (6.7) with the log-likelihood function (6.9) optimised and differentiated with respect to reciprocal parameters.

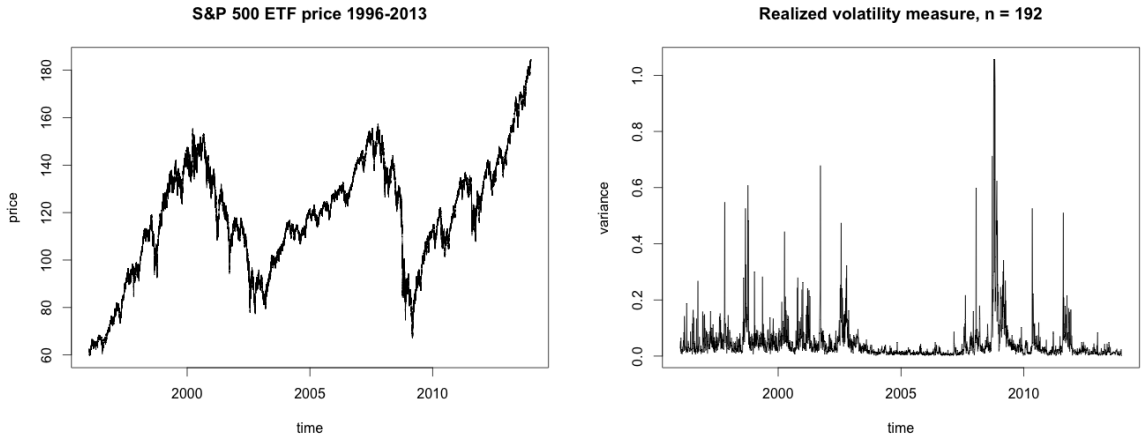

Figure 6.11: The left figure shows the S&P 500 ETF price from 1996-01-02 to 2013-12-31 with a 5 minute frequency and a total of 368,724 observations. Right figure: variance of the S&P 500 index measured by realized volatility with frequency $n = 192$ and a total of 1,920 observations.

	$\hat{\kappa}$	$\hat{\theta}$	$\hat{\varepsilon}$
Heston	21.7 (3.63)	0.022 (0.0049)	2.98 (0.0791)
log-normal	56.1 (3.36)	-3.65 (0.049)	11.5 (0.24)
3-over-2[12]	24.6 (4.01)	9.18 (6.98)	128.2 (3.95)

Table 6.12: S&P 500 ETF data: maximum likelihood estimated parameters and standard errors from the realized volatility measure.

index, in favour of dismissing the other two. The diagnose is supported by the last step of our goodness-of-fit analysis with results from a chi-square test with $n_b = 250$ bins recorded in Table 6.14: we can not reject the null hypothesis that residuals are standard uniform when they are calculated from the log-normal model. This agrees with p-values from the Kolmogorov-Smirnoff test as well. On the other hand, we reject with strong confidence the hypothesis when residuals are calculated from both Heston's and the 3-over-2 model.

Remark 6.4. A remark on estimated Heston parameters is in order here. Comparing our estimates to those reported elsewhere in the literature, Ait-Sahalia et al. (2007), Hurn et al. (2015) and Guillaume and Schoutens (2012), as well as to market practice,[13] we get a combination of relatively high volatility-of-volatility ε and high speed of mean reversion κ. This is needed to match the spiky time series behaviour of instantaneous variance, which is basically what our likelihood approach aims at capturing, and what we believe must lie at the heart of any good volatility model.

[13]See `http://www.wilmott.com/messageview.cfm?catid=8\&threadid=36596`

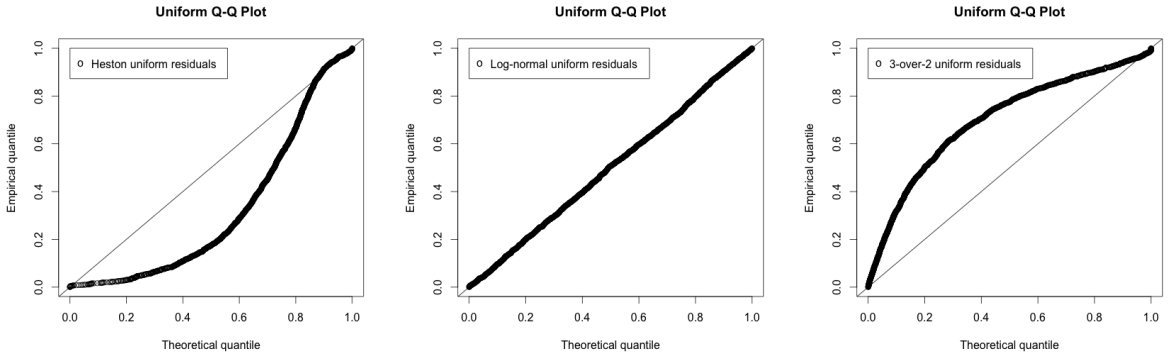

Figure 6.13: Quantile plots of the uniform distribution from uniform residuals calculated with measured S&P 500 variance.

P-value:	Chi-square	Kolmogorov-Smirnov
Heston	0.00	0.00
log-normal	0.79	0.68
3-over-2	0.00	0.00

Table 6.14: S&P 500 data: p-values of chi-square and Komogorov-Smirnov distributional tests of uniform residuals calculated from each model.

Remark 6.5. For the 3-over-2 model, parameter estimates of Table 6.12 and the resulting fit of the model residuals are based on the reciprocal specification. If we instead optimize the likelihood (6.10) we obtain what is reported in Table 6.15 and here we observe standard errors that are quite large for kappa and theta. The estimation is performed by optimizing (6.10) with respect to original parameters $(\kappa, \theta, \varepsilon)$ such that the square-root process parameters $(\tilde{\kappa}, \tilde{\theta}, \tilde{\varepsilon})$ is a transformation of these. This transformation might cause the standard error calculation to be unstable as it is based on numerical differentiation of the log-likelihood function and a matrix inversion of the information matrix, both involving the transformation. Howbeit, if we continue our goodness-of-fit analysis from here on, we obtain zero p-values from both the chi-square and Kolmogorov-Smirnov test as well – the support for the 3-over-2 model remains grossly absent.

	$\hat{\kappa}$	$\hat{\theta}$	$\hat{\varepsilon}$
3-over-2	2.7 (5.64)	30.0 (62.7)	98.2 (1.95)

Table 6.15: S&P 500 ETF data: maximum likelihood estimated parameters and standard errors for the 3-over-2 model.

# obs:	IBM	MCD	CAT	MMM	MCO
price	246,308	328,111	328,107	328,108	271,580
variance	1,282	1,708	1,708	1,708	1,414

# obs:	XOM	AZN	GS	HPQ	FDX
price	288,846	302,350	300,759	239,356	328,104
variance	1,504	1,574	1,566	1,246	1,708

Table 6.16: Number of asset price and measured variance observations for the NYSE stock data.

Based on the results from all models, a visual diagnose of quantile plots and p-values from two distributional tests, we see strong support for the log-normal model as a good volatility model for the S&P 500 data. But what if this is only a simple aggregation effect? After all, the S&P 500 index represents a portfolio of ~ 500 individual stocks: a critic may argue that realized volatility measured from this portfolio will be approximately (log)normal distributed, due to the central limit theorem, regardless of the stocks' individual variance distributions.[14]

With the above criticism in mind – and since the analysis is interesting in its own right – we next look at market data of ten individual stocks randomly chosen from the New York Stock Exchange. The stocks' tick codes along with their respective number of observations are reported in Table 6.16. The time lengths of the price series are in the range of 12 to 16 years, subject to availability from Wharton Research Data Services.

Estimated parameters and standard errors from the measured realized volatility are listed in the appendix, Table 6.25. Based on these, we perform our goodness-of-fit analysis for each individual stock with resulting p-values from the chi-square and Kolmogorov-Smirnov tests reported in Table 6.17, where we have included both the original and reciprocal specification of the 3-over-2 model. The results are thoroughgoing: for each of the ten stocks, we obtain a striking support for the log-normal model while the null hypotheses, suggesting either of the alternative models, are rejected with strong confidence. The only figure that stands out is the p-value of 1% from the chi-square test of log-normal residuals calculated from Hawlett-Packard (HPQ). Based on a confidence $< 99\%$ we thus reject all three models for the HPQ data. On the other hand, we observe a p-value of 5% from the Kolmogorov-Smirnov test which indicates that we should prefer, if any, the log-normal among the three models. This is also supported by a visual diagnose of the quantile-quantile plots: the log-normal residuals provide a convincing fit to the uniform distribution, while

[14] Even if the realized volatility measure is non-linear, a rough calculation gives (with illustrative notation) $RV(S\&P500) = \sum_t pf\,return(t)^2 \approx \sum_t \left(\sum_k stock_k\,return(t)\right)^2 = \sum_t \sum_k stock_k\,r(t)^2 + \sum_{k\neq j} 2 stock_k\,r(t) stock_j\,r(t) = \sum_k RV(stock_k) + \sum_{k\neq j} 2 \sum_t stock_k\,r(t) stock_j\,r(t)$. The first term is the sum of individual variances, while the second term sums up *realized covariations* across all stocks – a quantity that converges to the sum of individual quadratic variations. In particular, if returns are mutually independent across all stocks, we have $RV(S\&P500) \approx \sum_k^{500} RV(stock_k)$.

chi-square:	IBM	MCD	CAT	MMM	MCO	XOM	AZN	GS	HPQ	FDX
Heston	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
log-normal	0.48	0.67	0.36	0.84	0.15	0.61	0.91	0.12	0.01	0.90
3-over-2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3-over-2 inv	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Kolm.-Smir.:	IBM	MCD	CAT	MMM	MCO	XOM	AZN	GS	HPQ	FDX
Heston	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
log-normal	0.21	0.53	0.83	0.79	0.12	0.07	0.56	0.12	0.05	0.47
3-over-2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3-over-2 inv	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 6.17: P-values from uniform distributional tests of uniform residuals calculated from the realized volatility measure with $n = 192$.

the other two do not, see Figure 6.27 in the appendix.

6.4.1 What about jumps in the asset prices?

The previous analyses of price data was based on realized volatility since this measure gives a good approximation of quadratic variation. For a general semimartingale, the quadratic variation consists of an integrated variance part and a jump term which is non-zero if and only if the price process exhibits jumps. Therefore, realized volatility is a suitable measure when the price process is continuous, or at least has negligible jumps of low frequency and small magnitudes. Otherwise, it will provide a biased measurement of the integrated variance and the bias will always be positive as the squared jumps accumulate to the quadratic variation. This could give an explanation to the variance "spikiness" observed for instance in Figure 6.13 which may in turn affect the fit of a particular volatility model.

For this reason, we repeat the goodness-of-fit analysis based on the realized bipower measure with $n = 192$ prices for each variance observation. For the S&P 500 data, Figure 6.18 shows variance obtained from both measures and we can indeed observe some variance peaks that seem to be due jumps of the asset price, when looking at the difference between the two graphs. If we calculate $(RV - BV)/RV$ as the estimated quota of quadratic variation that can be attributed jumps, we obtain 20%. Nevertheless, for estimated parameters from the bipower variation, as reported in Table 6.19, we observe similar estimates compared to realized volatility estimates, cf. Table 6.12. Furthermore, quantile plots from the bipower based uniform residuals are almost identical to the plots in Figure 6.13 which lead us to the same diagnose: the log-normal model fits well with the S&P 500 data while Heston's and the 3-over-2 model present poor fits (p-values supporting this conclusion are presented in Table 6.20).

If we continue the analysis with bipower measured variance from the individual stocks we obtain the test results as reported in Table 6.21 (estimated parameters and standard

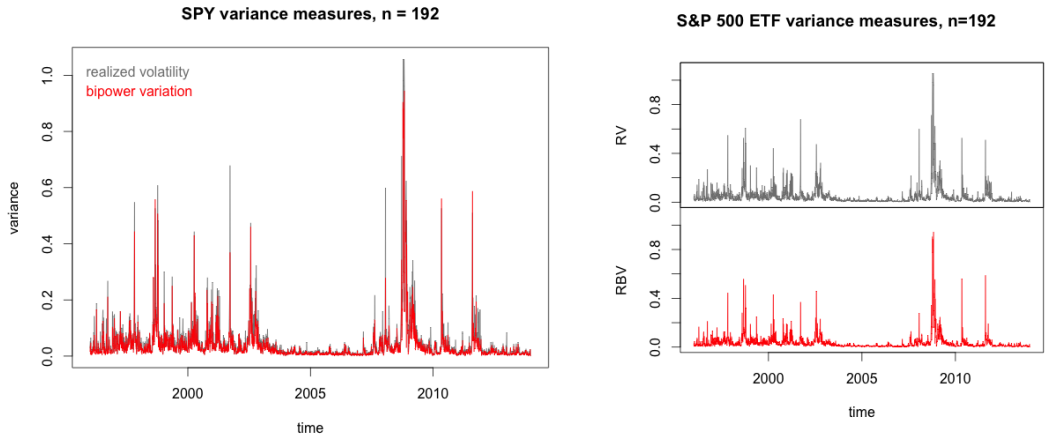

Figure 6.18: The figures show S&P 500 variance measured from the realized variance (gray) and realized bipower variation (red).

	$\hat{\kappa}$	$\hat{\theta}$	$\hat{\varepsilon}$
Heston	21.4 (3.77)	0.011 (0.0037)	2.78 (0.0786)
log-normal	55.3 (3.31)	-3.92 (0.05)	11.5 (0.24)
3-over-2	25.4 (3.98)	17.2 (8.82)	142.4 (4.26)

Table 6.19: S&P 500 data: maximum likelihood estimated parameters and standard errors from the realized bipower measure.

errors are given in Table 6.26 in the appendix). The p-values are consistently pointing us to reject all but the log-normal model for each of the individual stock with statistical certainty. Only the data from Caterpillar Inc. obtains a non-significant p-value when applying the chi-square test. On the other hand, again when we diagnose the quantile plots, and look at the Kolmogorov-Smirnov results, we can with certainty say that the log-normal model fits well. Our goodness-of-fit analysis stands robust to jump-corrections in the asset price and for all considered data sets – index as well as individual stocks – we have a winner: the stochastic volatility model of log-normal type.

P-values	Chi-square	Kolmogorov-Smirnov
Heston	0.00	0.00
log-normal	0.04	0.17
3-over-2	0.00	0.00
3-over-2 inv.	0.00	0.00

Table 6.20: S&P 500 data: p-values of chi-square and Komogorov-Smirnov distributional tests of uniform residuals calculated from each model with the realized bipower measure.

chi-square:	IBM	MCD	CAT	MMM	MCO	XOM	AZN	GS	HPQ	FDX
Heston	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
log-normal	0.44	0.15	0.006	0.42	0.73	0.66	0.22	0.21	0.94	0.53
3-over-2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3-over-2 inv.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Kolm.-Smir.:	IBM	MCD	CAT	MMM	MCO	XOM	AZN	GS	HPQ	FDX
Heston	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
log-normal	0.41	0.43	0.055	0.15	0.30	0.08	0.98	0.05	0.83	0.04
3-over-2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3-over-2 inv	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 6.21: Stock data: p-values from uniform distributional tests of uniform residuals calculated from the realized bipower variation measure with $n = 192$.

6.5 Epilogue: A taxonomy of log-normal volatility models

In this paper we have demonstrated that log-normal models provide a significantly better description of the empirical behaviour of instantaneous variance than the Heston and the 3-over-2 models do. However, at coarser time-resolutions it can be hard to distinguish the models, in particular with too crude variance measures and a false model might even be preferred in certain situations.

We are by no means the first to suggest log-normality. Similarly to us, Andersen et al. (2001) (see for instance Figures 1-3) and Wilmott (2006) base their argumentation solely on the empirical behaviour of measured volatility of the underlying, that is, on fundamentals. Papers such as Hagan et al. (2002) and Sepp (2014) suggest log-normal('ish) models with option pricing in mind – although while abandoning empirics. In one recent work Gatheral et al. (2014) even makes the point that we should question the diffusive nature of volatility (which we don't in this paper) before we question its log-normality.

Appendix

6.5.1 Scenario simulation schemes

Ito-Taylor approximations. A classical and simple simulation method for solutions to stochastic differential equations is the Euler-Maruyama scheme which involves the application of Ito-Taylor expansions (see Kloeden and Platen (1992)). With $\hat{Y}$ denoting a discrete-time approximation of the log asset-price $Y = \log S$, the Euler-Maruyama scheme reads

$$\hat{Y}(t+\Delta) = \hat{Y}(t) + (\mu - \frac{1}{2}V(t))\Delta + \sqrt{V(t)\Delta} \cdot Z_Y \tag{6.19}$$

where Z_Y is a $\mathcal{N}(0,1)$ Gaussian variable. This one-step scheme gives a sample of $\hat{Y}(t+\Delta)$ given $\hat{Y}(t)$ and $V(t)$ for an arbitrary time-increment Δ. Repeated use will give a full path of $\hat{S} = \exp \hat{Y}$ at any chosen set of discrete time-points.

The Euler-Maruyama scheme for the variance process of Heston's model is given by

$$\hat{V}(t+\Delta) = \hat{V}(t) + \kappa(\theta - \hat{V}(t))\Delta + \varepsilon\sqrt{\hat{V}(t)\Delta} \cdot Z_V$$

where Z_V is a $\mathcal{N}(0,1)$ Gaussian variable correlated to Z_Y by ρ. The correlation is typically handled by a Cholesky decomposition

$$Z_Y = \rho Z_V + \sqrt{1-\rho^2} Z^{\perp}$$

where Z_V and $Z^{\perp}$ are independent standard Gaussian variables. The Euler-Maruyama scheme suffers from several drawbacks, it is a biased scheme and it may produce negative values of the variance in this case. Andersen et al. (2010) discuss these issues and review several alternative solutions, one of which is the *truncated Euler* scheme

$$\hat{V}(t+\Delta) = \hat{V}(t) + \kappa(\theta - \hat{V}(t)^+)\Delta + \varepsilon\sqrt{\hat{V}(t)^+\Delta} \cdot Z_V$$

where $\hat{V}(t)$ is replaced by $\hat{V}(t)^+$ in equation (6.19). The *truncated explicit Milstein* reads

$$\hat{V}(t+\Delta) = \hat{V}(t) + \kappa(\theta - \hat{V}(t)^+)\Delta + \varepsilon\sqrt{\hat{V}(t)^+\Delta} \cdot Z_V + \frac{1}{4}\varepsilon^2\Delta(Z_V^2 - 1)$$

where the Milstein term should reduce the bias, while the *implicit Milstein scheme*

$$\hat{V}(t+\Delta) = \frac{\hat{V}(t) + \kappa\theta\Delta + \varepsilon\sqrt{\hat{V}(t)\Delta} \cdot Z_V + \frac{1}{4}\varepsilon^2\Delta(Z_V^2 - 1)}{1 + \kappa\Delta}$$

produces positive variance paths if the condition $4\kappa\theta > \varepsilon^2$ holds for the parameters.

The conditional distribution of the logarithmic variance is Gaussian in the log-normal model and this allows for a simple and exact simulation scheme for the variance process

$$V(t+\Delta) = \exp\left[\log\left(V(t)\right) \cdot e^{-\kappa\Delta} + \theta(1 - e^{-\kappa\Delta}) + \varepsilon\sqrt{\frac{1-e^{-2\kappa\Delta}}{2\kappa}} Z_V\right] \tag{6.20}$$

where Z_V is a $\mathcal{N}(0,1)$ Gaussian variable.

The simulation of the variance process in the 3-over-2 model is unfortunately not as straight forward. The Euler-Maruyama scheme

$$\hat{V}(t+\Delta) = \hat{V}(t) + \kappa\hat{V}(t)(\theta - \hat{V}(t))\Delta + \varepsilon\hat{V}(t)^{3/2}\sqrt{\Delta}\cdot Z_V$$

obviously allows for negative outcomes. A truncation is however not a good solution due to the extra $\hat{V}(t)$ term in the diffusion part: as soon as $\hat{V}(t+\Delta)$ reaches below zero it stays on the same negative level as both the drift and diffusion is set zero due to the truncation.

In stead of simulating the variance in the 3-over-2 process we simulate the reciprocal of the variance since there exists schemes for the square-root process. With $\tilde{\kappa} = \kappa\theta$ and $\tilde{\theta} = (\kappa+\varepsilon^2)/(\kappa\theta)$, the Euler-Maruyama scheme reads

$$\begin{aligned}\hat{Y}(t+\Delta) &= \hat{Y}(t) + \tilde{\kappa}(\tilde{\theta} - \hat{Y}(t))\Delta - \varepsilon\sqrt{\hat{Y}(t)\Delta}\cdot Z_V \\ \hat{V}(t+\Delta) &= 1/\hat{Y}(t+\Delta)\end{aligned}$$

for the (reciprocal) variance. To further rule out the possibility of negative values of Y; given that $4\tilde{\kappa}\tilde{\theta} > \varepsilon^2$, we may employ the implicit Milstein scheme

$$\hat{Y}(t+\Delta) = \frac{\hat{Y}(t) + \tilde{\kappa}\tilde{\theta}\Delta - \varepsilon\sqrt{\hat{Y}(t)\Delta}\cdot Z_V + \frac{1}{4}\varepsilon^2\Delta(Z_V^2 - 1)}{1+\tilde{\kappa}\Delta}$$

for the reciprocal variance.

Almost exact scenario simulation schemes. The Euler-Maruyama and Milstein schemes of the previous section are biased schemes which means that the distributional properties of $(\hat{Y}, \hat{V})$ will be approximate and converge, in some sense, to (Y, V) as the time increment goes to zero. Broadie and Kaya (2006) suggest an exact simulation scheme for the dynamics of Heston's model that we modify in the following way to form an *almost* exact scheme.

First, an application of Ito's formula to the logarithmic asset price gives

$$\begin{aligned}Y(t+\Delta) = Y(t) + \int_t^{t+\Delta}\left(\mu - \frac{1}{2}V(u)\right)du + \rho\int_t^{t+\Delta}\sqrt{V(u)}dW(u) \\ +\sqrt{1-\rho^2}\int_t^{t+\Delta}\sqrt{V(u)}dW^{\perp}(u)\end{aligned} \tag{6.21}$$

where the Wiener process correlation between Z in the asset price of equation (6.2) and W of the variance in equation (6.4) is handled by the Cholesky decomposition

$$dZ(t) = \rho dW(t) + \sqrt{1-\rho^2}dW^{\perp}(t)$$

with $W^{\perp}$ denoting a Wiener process independent of W. The variance dynamics

$$V(t+\Delta) = V(t) + \int_t^{t+\Delta}\kappa(\theta - V(u))du + \varepsilon\int_t^{t+\Delta}\sqrt{V(u)}dW(u)$$

is rewritten to get an expression for the stochastic integral

$$\int_t^{t+\Delta} \sqrt{V(u)} dW(u) = \frac{1}{\varepsilon} \left(V(t+\Delta) - V(t) - \int_t^{t+\Delta} \kappa(\theta - V(u)) du \right).$$

This equation is then substituted into equation (6.21) to arrive at

$$\begin{aligned} Y(t+\Delta) = Y(t) + \mu\Delta + \frac{\rho}{\varepsilon}(V(t+\Delta) - V(t) - \kappa\theta\Delta) \\ + \left(\frac{\rho\kappa}{\varepsilon} - \frac{1}{2} \right) \int_t^{t+\Delta} V(u) du + \sqrt{1-\rho^2} \int_t^{t+\Delta} \sqrt{V(u)} dW^{\perp}(u). \end{aligned} \tag{6.22}$$

The exact generation of $Y(t+\Delta)$ given $Y(t), V(t)$ from the one-step scheme (6.22) requires sampling of three random variables. First, $V(t+\Delta)$ given $V(t)$ is non-central chi-square distributed and this can be sampled by draws from the Poisson and chi-square distribution according to the following. With $n(\Delta) = 4\kappa e^{-\kappa\Delta}/(\varepsilon^2(1-e^{-\kappa\Delta}))$ and $d = 4\kappa\theta/\varepsilon^2$:

1. Draw N from a Poisson distribution with mean $\frac{1}{2}V(t)n(\Delta)$.
2. Given N, draw X from a chi-square distribution with $d + 2N$ degrees of freedom.
3. Set $V(t+\Delta) = Xe^{-\kappa\Delta}/n(\Delta)$.

Secondly, if we have a generated value of the integrated variance $\int V(u)du$, the last term of equation (6.22) is sampled from the normal distribution

$$\int_t^{t+\Delta} \sqrt{V(u)} dW^{\perp}(u) \sim \mathcal{N}\left(0, \int_t^{t+\Delta} V(u) du\right)$$

which is due to the fact that $W^{\perp}$ and V are independent. Broadie and Kaya (2006) recount a method for the generation of $\int V(u)du$ given $V(t+\Delta), V(t)$ that is based on the Laplace transform of the integrated variance. In return, the above steps result in their exact simulation scheme although the last generation of the integrated variance involves a series of complicated steps.

A simple alternative, as suggested in Platen and Bruti-Liberati (2010), is to replace the integrated variance with an approximation as given by the trapezoidal rule

$$\int_t^{t+\Delta} V(u) du \approx \frac{V(t+\Delta) + V(t)}{2} \Delta \tag{6.23}$$

and this is employed for the almost exact simulation scheme of Heston's model.

Baldeaux (2012) retails a method for the exact simulation of the 3-over-2 model that is similar to the Broadie-Kaya scheme. Their method may be adapted for an almost exact scheme for the 3-over-2 model. First, $1/V(t+\Delta)$ given $1/V(t)$ is non-central chi-square distributed and it may be sampled by the above procedure where κ is substituted with

$\tilde{\kappa} = \kappa\theta$ and θ with $\tilde{\theta} = (\kappa + \varepsilon^2)/(\kappa\theta)$. Next, Ito's formula applied to the logarithmic variance $X(t) = \log V(t)$ gives

$$X(t+\Delta) = X(t) + \kappa\theta\Delta - \bar{\kappa}\int_t^{t+\Delta} V(u)du + \varepsilon\int_t^{t+\Delta}\sqrt{V(u)}du$$

where $\bar{\kappa} = \kappa + \varepsilon^2/2$ and hence for the stochastic integral

$$\int_t^{t+\Delta}\sqrt{V(u)}dW(u) = \frac{1}{\varepsilon}\left(\log V(t+\Delta) - \log V(t) - \kappa\theta\Delta + \bar{\kappa}\int_t^{t+\Delta} V(u)du\right).$$

This is substituted into equation (6.21) to arrive at

$$Y(t+\Delta) = Y(t) + \mu\Delta + \frac{\rho}{\varepsilon}\left(\log\frac{V(t+\Delta)}{V(t)} - \kappa\theta\Delta\right)$$
$$+\left(\frac{\rho\bar{\kappa}}{\varepsilon} - \frac{1}{2}\right)\int_t^{t+\Delta} V(u)du + \sqrt{1-\rho^2}\int_t^{t+\Delta}\sqrt{V(u)}dW^{\perp}(u)$$

which again relies on the generation of $\int V(u)du$ given $V(t+\Delta), V(t)$ where V follows the 3-over-2 model. As for Heston's model, we resort to approximating the integrated variance by the trapezoidal rule (6.23) to complete the almost exact simulation scheme for the 3-over-2 model.

For the log-normal model the generation of $V(t+\Delta)|V(t)$ is made by a draw from the normal distribution according to equation (6.20). For the next step, Ito's formula applied to $\sqrt{V(t)}$ gives

$$d\sqrt{V(t)} = \frac{1}{2}\kappa\sqrt{V(t)}\ (\bar{\theta} - \log V(t))dt + \frac{1}{2}\varepsilon\sqrt{V(t)}dW(t)$$

where $\bar{\theta} = \theta + \varepsilon^2/(4\kappa)$. The integrated version may then be substituted into equation (6.21) to obtain the logarithmic asset price

$$Y(t+\Delta) = Y(t) + \mu\Delta - \frac{1}{2}\int_t^{t+\Delta} V(u)du + \sqrt{1-\rho^2}\int_t^{t+\Delta}\sqrt{V(u)}dW^{\perp}(u)$$
$$+\frac{2\rho}{\varepsilon}\left(\sqrt{V(t+\Delta)} - \sqrt{V(t)} - \frac{\kappa\bar{\theta}}{2}\int_t^{t+\Delta}\sqrt{V(u)}du + \frac{\kappa}{2}\int_t^{t+\Delta}\sqrt{V(u)}\log V(u)du\right).$$

This is an expression that has several integrals of the variance process and we use the trapezoidal rule for their generation

$$\begin{aligned}
\int_t^{t+\Delta} V(u)du &\approx \frac{V(t+\Delta)+V(t)}{2}\,\Delta \\
\int_t^{t+\Delta}\sqrt{V(u)}du &\approx \frac{\sqrt{V(t+\Delta)}+\sqrt{V(t)}}{2}\,\Delta \\
\int_t^{t+\Delta}\sqrt{V(u)}\log V(u)du &\approx \frac{\sqrt{V(t+\Delta)}\log V(t+\Delta)+\sqrt{V(t)}\log V(t)}{2}\,\Delta
\end{aligned}$$

to end up with an almost exact simulation scheme for the log-normal model.

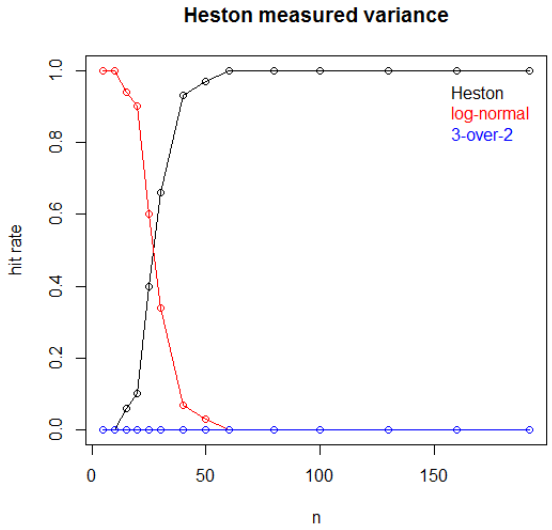

Figure 6.22: The figure shows prediction hit-rates as a function of resolution n: the percentage out of 100 simulations for which our test selects the correct model from the largest p-value.

6.5.2 Robustness of the model selection test

To address the first question of how accurate variance measure we need to be able to predict the correct model, we focus on the resolution n of the realized volatility measure (see equation 6.13). For a simulated path and a selected value of n (starting low with $n = 5$), we use $1,890 \times n$ prices for the variance measurement in order to keep the total number of measured observations constant at $1,890$. We then apply our goodness-of-fit analysis and record the model with strongest support according to the p-value. We repeat this for 100 simulated paths and record the hit-rate for each model as the percentage of all simulations for which the correct model is chosen. The result is shown in Figure 6.22 with Heston as the underlying simulation model: the hit-rates show that the test points us to the correct model for $n > 50$ while a low resolution $n \sim 10$ points to the log-normal model in 100% of the simulations. The accuracy of the realized volatility measure is too poor in these cases.

The hit-rate from 100 simulations with the log-normal model as underlying is plotted in left Figure 6.23. The figure shows that the analysis points us to the correct model for all simulations, also for poor accuracies. Similarly, with the 3-over-2 as underlying model, we obtain the hit-rates plotted in right Figure 6.23 which shows that the test points to the correct model for $n > 60$. As for the case with Heston simulations, a low resolution $n \sim 10$ falsely points to the log-normal model. Thus, we conclude that we need a resolution that gives a sufficient accuracy in order for our test to be robust. As we saw in the Monte Carlo study as well, we have a robust test for $n \sim 200$ and we might go down to $n \sim 60$

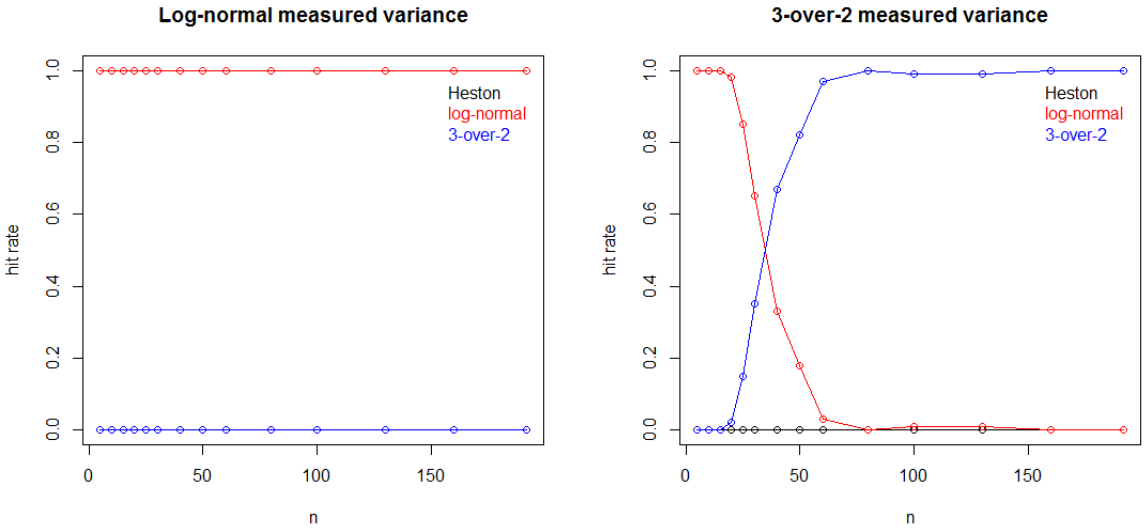

Figure 6.23: The figures show hit-rates as a function of resolution n when the underlying simulation model is the log-normal (left) and the 3-over-2 model (right).

before our procedure gives unreliable results, in particular when the underlying model is the 3-over-2 model.

For the second question of how long price series we require to get a robust test, we proceed in a similar manner. We alter the length of the asset price path and apply our analysis for each simulated path with $n = 192$ prices per variance measurement. The hit-rate is calculate from 100 simulated paths and we repeat the procedure starting from 1 year of data (24,193 prices, 126 variances) up to 15 years of data. The result for simulated Heston's model shows that the test works well even for short price-paths, and the same holds for simulated log-normal data. The 1 year data points us to the correct model in 75 (60) simulations out of 100 when we generate from Heston's (log-normal) model. For paths longer than 5 years (120,961 prices, 630 variances) we have a robust test with a hit-rate of 100% for both models. The result for the 3-over-2 model is similar although we need a little more data: we have a robust test for data longer than 6 years (145,153 prices, 756 variances) while short price paths with 100–400 variance observations might point us to the wrong model with hit-rates in the range 55–85%.

6.5.3 Figures and tables

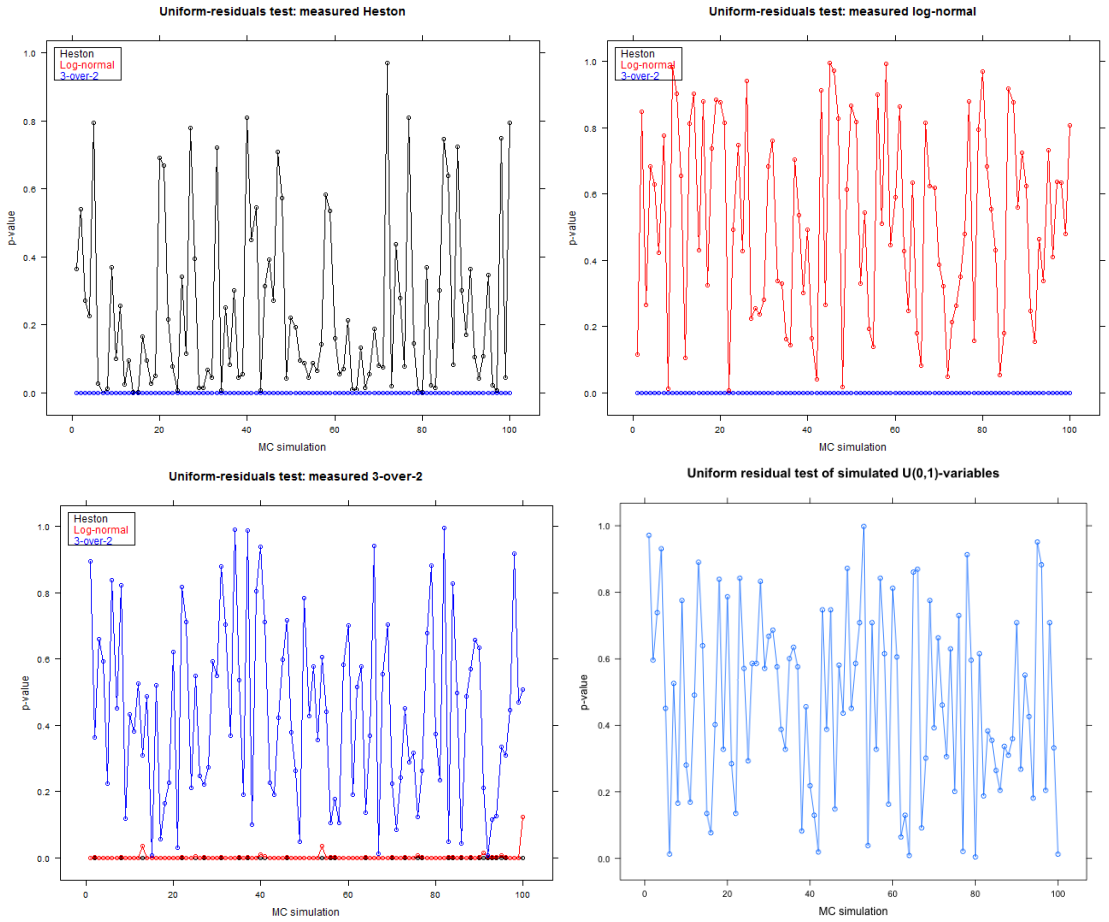

Figure 6.24: The figures show p-values from the goodness-of-fit analysis based on 100 simulations of Heston's model (**top left figure**) the log-normal model (**top right figure**) and the 3-over-2 model respectively (**bottom left figure**). In all three cases, the highest recorded p-value points us to the correct underlying model, while the alternative models obtain close to zero p-values. To put the variation of the p-values into perspective, the **bottom right figure** shows results when the procedure is applied directly to data from the uniform distribution: here we see the same degree of randomness. The important message is that the actual magnitude of the p-value is less relevant: the correct underlying model always obtains the largest, albeit varying, p-value, whilst the alternative models yield poor goodness-of-fits.

Heston:	**IBM**	**MCD**	**CAT**	**MMM**	**MCO**
$\hat{\kappa}$	75.1 (7.19)	18.2 (3.99)	71.6 (5.34)	37.0 (4.52)	100.0 (6.85)
$\hat{\theta}$	0.024 (0.0043)	0.052 (0.0123)	0.112 (0.0087)	0.061 (0.0061)	0.096 (0.0093)
$\hat{\varepsilon}$	6.89 (0.351)	4.14 (0.115)	7.21 (0.215)	3.89 (0.105)	10.1 (0.39)
log-normal:					
$\hat{\kappa}$	86.6 (6.09)	61.1 (3.82)	83.9 (5.12)	76.9 (4.69)	71.6 (4.82)
$\hat{\theta}$	-3.31 (0.047)	-2.96 (0.048)	-2.45 (0.039)	-3.08 (0.040)	-2.60 (0.051)
$\hat{\varepsilon}$	13.5 (0.38)	11.4 (0.25)	12.7 (0.31)	12.1 (0.29)	43.8 (0.33)
3-over-2:					
$\hat{\kappa}$	10.4 (22.8)	2.21 (3.61)	3.21 (1.37)	2.79 (5.45)	5.98 (2.38)
$\hat{\theta}$	14.1 (30.9)	30.6 (49.8)	36.6 (15.5)	44.7 (87.1)	15.8 (6.26)
$\hat{\varepsilon}$	86.5 (2.87)	57.8 (1.18)	52.6 (1.28)	74.0 (1.92)	54.0 (1.39)
reciprocal 3-2:					
$\hat{\kappa}$	105.0 (9.96)	27.3 (4.15)	77.7 (6.47)	77.6 (7.49)	48.8 (5.65)
$\hat{\theta}$	40.0 (1.70)	27.2 (3.04)	16.3 (0.79)	32.0 (1.49)	19.2 (1.48)
$\hat{\varepsilon}$	95.2 (3.39)	67.8 (1.72)	60.7 (1.71)	82.8 (2.47)	61.9 (1.81)

Heston:	**XOM**	**AZN**	**GS**	**HPQ**	**FDX**
$\hat{\kappa}$	45.4 (4.55)	113.0 (7.92)	41.6 (5.14)	116.9 (12.1)	65.2 (6.79)
$\hat{\theta}$	0.052 (0.0048)	0.069 (0.0053)	0.027 (0.0094)	0.14 (0.012)	0.082 (0.008)
$\hat{\varepsilon}$	3.61 (0.102)	7.76 (0.281)	7.97 (0.322)	12.6 (0.57)	6.62 (0.211)
log-normal:					
$\hat{\kappa}$	72.3 (4.73)	126.2 (8.58)	58.3 (3.82)	102.2 (7.39)	88.0 (5.39)
$\hat{\theta}$	-3.10 (0.043)	-2.97 (0.037)	-2.49 (0.053)	-2.54 (0.045)	-2.63 (0.038)
$\hat{\varepsilon}$	11.3 (0.28)	16.6 (0.50)	11.6 (0.27)	14.9 (0.46)	12.8 (0.32)
3-over-2:					
$\hat{\kappa}$	3.04 (5.34)	7.53 (8.29)	3.20 (2.90)	6.21 (2.94)	4.72 (5.78)
$\hat{\theta}$	29.9 (52.6)	29.4 (32.3)	32.1 (29.0)	23.0 (10.9)	27.9 (34.1)
$\hat{\varepsilon}$	60.9 (1.47)	101.1 (4.03)	51.4 (1.32)	61.1 (1.87)	59.5 (1.52)
reciprocal 3-2:					
$\hat{\kappa}$	60.9 (5.77)	152.0 (12.5)	45.6 (6.02)	105.8 (9.65)	91.6 (7.52)
$\hat{\theta}$	31.1 (1.61)	30.6 (1.24)	15.7 (1.29)	18.7 (0.88)	19.4 (0.83)
$\hat{\varepsilon}$	67.2 (1.87)	122.5 (4.50)	56.6 (1.62)	71.9 (2.58)	67.9 (2.01)

Table 6.25: Ten individual stocks from NYSE: maximum likelihood parameters and standard errors estimated from the realized volatility measure.

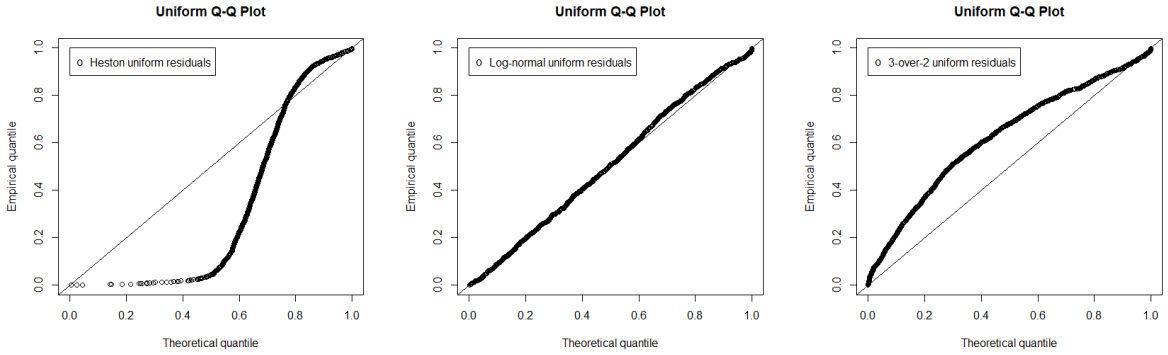

Figure 6.27: Quantile plots from uniform residuals calculated with realized volatility measured Hewlett-Packard variance.

Heston:	**IBM**	**MCD**	**CAT**	**MMM**	**MCO**
$\hat{\kappa}$	25.3 (4.13)	14.6 (3.59)	31.9 (3.84)	24.8 (3.70)	46.8 (4.82)
$\hat{\theta}$	0.031 (0.0053)	0.051 (0.0115)	0.083 (0.0084)	0.049 (0.0060)	0.096 (0.0087)
$\hat{\varepsilon}$	2.91 (0.090)	3.30 (0.085)	3.91 (0.099)	2.94 (0.074)	5.23 (0.16)
log-normal:					
$\hat{\kappa}$	73.5 (5.18)	56.5 (3.57)	78.9 (4.82)	71.4 (4.38)	69.1 (4.67)
$\hat{\theta}$	-3.57 (0.047)	-3.18 (0.049)	-2.72 (0.038)	-3.32 (0.041)	-2.86 (0.050)
$\hat{\varepsilon}$	11.7 (0.32)	11.0 (0.24)	11.7 (0.28)	11.4 (0.26)	12.3 (0.31)
3-over-2:					
$\hat{\kappa}$	6.69 (9.53)	2.17 (2.82)	3.34 (1.45)	6.16 (15.5)	4.76 (15.3)
$\hat{\theta}$	20.9 (29.7)	28.3 (36.8)	31.5 (13.6)	16.1 (40.6)	22.3 (71.4)
$\hat{\varepsilon}$	93.5 (3.22)	61.6 (1.24)	56.0 (1.34)	72.9 (1.78)	63.5 (1.78)
reciprocal 3-2:					
$\hat{\kappa}$	89.1 (10.06)	20.9 (3.8)	60.4 (6.00)	56.5 (6.11)	49.7 (6.28)
$\hat{\theta}$	49.5 (2.28)	31.0 (4.48)	20.2 (1.16)	40.5 (2.17)	24.5 (1.87)
$\hat{\varepsilon}$	98.7 (3.60)	72.1 (1.80)	62.7 (1.74)	79.2 (2.18)	70.2 (2.12)

Heston:	**XOM**	**AZN**	**GS**	**HPQ**	**FDX**
$\hat{\kappa}$	44.6 (4.31)	37.1 (3.99)	36.4 (5.01)	43.0 (5.71)	41.4 (4.42)
$\hat{\theta}$	0.043 (0.0037)	0.030 (0.0028)	0.002 (0.0062)	0.087 (0.0091)	0.073 (0.0058)
$\hat{\varepsilon}$	3.06 (0.084)	2.37 (0.063)	6.77 (0.27)	4.83 (0.16)	3.73 (0.097)
log-normal:					
$\hat{\kappa}$	74.7 (4.88)	86.7 (5.51)	60.3 (3.93)	82.6 (5.87)	84.8 (5.18)
$\hat{\theta}$	-3.34 (0.042)	-3.76 (0.038)	-2.73 (0.051)	-2.81 (0.045)	-2.89 (0.037)
$\hat{\varepsilon}$	11.5 (0.29)	12.1 (0.31)	11.5 (0.27)	12.4 (0.35)	12.0 (0.29)
3-over-2:					
$\hat{\kappa}$	5.15 (8.22)	5.45 (16.3)	3.77 (3.28)	5.06 (3.20)	3.97 (4.45)
$\hat{\theta}$	19.8 (31.5)	21.0 (62.7)	30.4 (26.4)	24.7 (15.5)	29.9 (33.5)
$\hat{\varepsilon}$	71.8 (1.80)	102.0 (2.61)	60.0 (1.64)	64.4 (1.97)	62.8 (1.60)
reciprocal 3-2:					
$\hat{\kappa}$	68.2 (6.32)	70.7 (6.79)	53.0 (6.86)	77.4 (8.30)	78.3 (6.99)
$\hat{\theta}$	38.0 (1.90)	60.6 (3.33)	20.8 (1.46)	23.1 (1.33)	24.7 (1.11)
$\hat{\varepsilon}$	79.2 (2.28)	117.0 (3.5)	64.6 (1.92)	72.9 (2.51)	69.4 (2.00)

Table 6.26: Maximum likelihood estimated parameters and standard errors from the realized bipower variation measure.

Bibliography

Ahmad, R. and Wilmott, P. (2005). Which free lunch would you like today, sir?: Delta hedging, volatility arbitrage and optimal portfolios. *Wilmott*, pages 64–79.

Ait-Sahalia, Y., Kimmel, R., et al. (2007). Maximum likelihood estimation of stochastic volatility models. *Journal of Financial Economics*, 83(2):413–452.

Alanko, S. and Avellaneda, M. (2013). Reducing variance in the numerical solution of bsdes. *Comptes Rendus Mathematique*, 351(3):135–138.

Andersen, L. B., Jackel, P., and Kahl, C. (2010). *Simulation of Square-Root Processes*. John Wiley and Sons, Ltd.

Andersen, T. and Benzoni, L. (2009). Realized volatility. *Handbook of Financial Time Series*, pages 555–575.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Ebens, K. (2001). The distribution of realized stock return volatility. *Journal of Financial Economics*, 61:43–76.

Andreasen, J. (2003). Derivatives - the view from the trenches. Presentation at the University of Copenhagen, available at http://www.math.ku.dk/rolf/jandreasen.pdf.

Andreasen, J. (2013). Option anatomy. Inagural lectures at Department of Mathematical Sciences, University of Copenhagen. http://www.math.ku.dk/rolf/AndreasenInagural.pdf.

Andreasen, J. and Huge, B. N. (2010). Volatility interpolation. *Available at SSRN 1694972*.

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9:203–228.

Avellaneda, M., Levy, A., and Paras, A. (1995). Pricing and hedging derivative securities in markets with uncertain volatilities. *Applied Mathematical Finance*, 2(2):73–88.

Avellaneda, M. and Paras, A. (1996). Managing the volatility risk of portfolios of derivative securities: the lagrangian uncertain volatility model. *Applied Mathematical Finance*, 3(1):21–52.

Awartani, B. M. and Corradi, V. (2005). Predicting the volatility of the s&p-500 stock index via garch models: the role of asymmetries. *International Journal of Forecasting*, 21(1):167–183.

Bakshi, G. and Kapadia, N. (2003). Delta-hedged gains and the negative market volatility risk premium. *Review of Financial Studies*, 16(2):527–566.

Baldeaux, J. (2012). Exact simulation of the 3/2 model. *International Journal of Theoretical and Applied Finance*, 15(05).

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2009). Realized kernels in practice: Trades and quotes. *The Econometrics Journal*, 12(3):C1–C32.

Barndorff-Nielsen, O. E. and Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics*, 2(1):1–37.

Bates, D. S. (1996). Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options. *Review of financial studies*, 9(1):69–107.

Benth, F. E., Benth, J. S., and Koekebakker, S. (2008). *Stochastic modelling of electricity and related markets*, volume Advanced Series on Statistical Science & Applied Probability, vol. 11. World Scientific.

Bessembinder, H. and Lemmon, M. L. (2002). Equilibrium pricing and optimal hedging in electricity forward markets. *Journal of Finance*, 57(3):1347–1382.

Bingham, N. H. and Kiesel, R. (2004). *Risk-Neutral Valuation: Pricing and Hedging of Financial Derivatives*. Springer Science & Business Media.

Björk, T. (2009). *Arbitrage Theory in Continuous Time*. Oxford University Press.

Black, F. (1976). Studies of stock price volatility changes. *In: Proceedings of the 1976 Meetings of the American Statistical Association*, pages 171–181.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *The journal of political economy*, pages 637–654.

Boroumand, R. H., Goutte, S., Porcher, S., and Porcher, T. (2015). Hedging strategies in energy markets: The case of electricity retailers. *Energy Economics*, 51:503–509.

Botterud, A., Bhattacharyya, A. K., and Ilic, M. (2002). Futures and spot prices âĂŞ an analysis of the scandinavian electricity market. In *Proceedings of the 34th North American Power Symposium*.

Bouchard, B. and Elie, R. (2008). Discrete-time approximation of decoupled forward–backward sde with jumps. *Stochastic Processes and their Applications*, 118(1):53–75.

Bouchard, B. and Touzi, N. (2004). Discrete-time approximation and monte-carlo simulation of backward stochastic differential equations. *Stochastic Processes and their applications*, 111(2):175–206.

Breeden, D. T. and Litzenberger, R. H. (1978). Prices of state-contingent claims implicit in option prices. *Journal of business*, pages 621–651.

Broadie, M. and Kaya, Ö. (2006). Exact simulation of stochastic volatility and other affine jump diffusion processes. *Operations Research*, 54(2):217–231.

Cairns, A. J. G. (2004). *Intereast Rate Models: an introduction*. Princeton.

Carmona, R. and Coulon, M. (2014). A survey of commodity markets and structural models for electricity prices. In *Quantitative Energy Finance*, pages 41–83. Springer.

Carr, P. (2002). Faq's in option pricing theory. Working paper, available at http://www.math.nyu.edu/research/carrp/papers/pdf/faq2.pdf.

Carr, P., Geman, H., Madan, D. B., and Yor, M. (2003). Stochastic volatility for lévy processes. *Mathematical Finance*, 13(3):345–382.

Carr, P. and Jarrow, R. (1990). The stop-loss start-gain paradox and option valuation: A new decomposition into intrinsic and time value. *Review of Financial Studies*, 3(3):469–492.

Carr, P. and Sun, J. (2007). A new approach for option pricing under stochastic volatility. *Review of Derivatives Research*, 10(2):87–150.

Christensen, K., Oomen, R. C., and Podolskij, M. (2014). Fact or friction: Jumps at ultra high frequency. *Journal of Financial Economics*, 114(3):576–599.

Cohen, S. N. and Elliott, R. J. (2015). *Stochastic Calculus and Applications*. Springer.

Cont, R. and Tankov, P. (2004). *Financial modelling with jump processes*. Chapman & Hall/CRC.

Coulon, M., Powell, W. B., and Sircar, R. (2013). A model for hedging load and price risk in the texas electricity market. *Energy Economics*, 40:976–988.

Cox, J. C. and Huang, C.-f. (1989). Optimal consumption and portfolio policies when asset prices follow a diffusion process. *Journal of economic theory*, 49(1):33–83.

Cox, J. C., Ingersoll Jr, J. E., and Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica: Journal of the Econometric Society*, pages 385–407.

Davis, M. H. (2010). Black–scholes formula. *Encyclopedia of Quantitative Finance*.

Delbaen, F. and Schachermayer, W. (1994). A general version of the fundamental theorem of asset pricing. *Mathematische annalen*, 300(1):463–520.

Delbaen, F. and Schachermayer, W. (1998). The fundamental theorem of asset pricing for unbounded stochastic processes. *Mathematische annalen*, 312(2):215–250.

Derman, E. (1996). Model risk. *Risk Magazine*, 9(5):34–37.

Derman, E. (2008). Lectures on the smile. Columbia University, http://www.emanuelderman.com/writing/entry/lectures-on-the-smile-2008.

Derman, E. and Kani, I. (1994). Riding on the smile. *Risk*, 7:32–39.

Derman, E. and Kani, I. (1998). Stochastic implied trees: Arbitrage pricing with stochastic term and strike structure of volatility. *International journal of theoretical and applied finance*, 1(01):61–110.

Dupire, B. (1994). Pricing with a smile. *Risk*, 7(1):18–20.

El Karoui, N. and Quenez, M.-C. (1995). Dynamic programming and pricing of contingent claims in an incomplete market. *SIAM journal on Control and Optimization*, 33(1):29–66.

Ellersgaard, S. and Jönsson, M. (2013). Stochastic volatility for utility maximisers - the bond-stock economy. Unpublished manuscript.

Feller, W. (1951). Two singular diffusion problems. *Annals of mathematics*, pages 173–182.

Gardiner, C. W. et al. (1985). *Handbook of stochastic methods*, volume 4. Springer.

Gatheral, J. (2011). *The volatility surface: a practitioner's guide*, volume 357. John Wiley & Sons.

Gatheral, J., Jaisson, T., and Rosenbaum, M. (2014). Volatility is rough. Workning paper available at SSRN.

Gibson, R., Lhabitant, F., Pistre, N., and Talay, D. (1999). Interest rate model risk: What are we talking about? *Journal of Risk*, 1:37–62.

Gobet, E. and Lemor, J.-P. (2008). Numerical simulation of bsdes using empirical regression methods: theory and practice. *arXiv preprint arXiv:0806.4447*.

Guillaume, F. and Schoutens, W. (2012). Use a Reduced Heston or Reduce the Use of Heston? *Wilmott Journal*, 2:171–192.

Gupta, A., Reisinger, C., and Whitley, A. (2010). Model uncertainty and its impact on derivative pricing. *Rethinking risk measurement and reporting*, 122.

Guyon, J. and Henry-Labordère, P. (2012). Being particular about calibration. *Risk*, 25(1):88.

Gyöngy, I. (1986). Mimicking the one-dimensional marginal distributions of processes having an itô differential. *Probability theory and related fields*, 71(4):501–516.

Hagan, P. S., Kumar, D., Lesniewski, A. S., and Woodward, D. E. (2002). Managing smile risk. *Wilmott Magazine*, July:84–108.

Harrison, J. M. and Kreps, D. M. (1979). Martingales and arbitrage in multiperiod securities markets. *Journal of Economic theory*, 20(3):381–408.

Harrison, J. M. and Pliska, S. R. (1981). Martingales and stochastic integrals in the theory of continuous trading. *Stochastic processes and their applications*, 11(3):215–260.

Hastie, T., Tibshirani, R., Friedman, J., and Franklin, J. (2005). The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer*, 27(2):83–85.

Haug, E. G. and Taleb, N. N. (2011). Option traders use (very) sophisticated heuristics, never the black–scholes–merton formula. *Journal of Economic Behavior & Organization*, 77(2):97–106.

Haugom, E. (2011). Some stylized facts about high-frequency Nord Pool forward electricity prices. *Journal of Energy Markets*, 4(1):21–49.

Henderson, V., Hobson, D., Howison, S., and Kluge, T. (2005). A comparison of option prices under different pricing measures in a stochastic volatility model with correlation. *Review of Derivatives Research*, 8(1):5–25.

Henrard, M. (2001). Parameter risk in the black and scholes model. Working paper, http://128.118.178.162/eps/ri/papers/0310/0310002.pdf.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of financial studies*, 6(2):327–343.

Heston, S. L. (1997). A simple new formula for options with stochastic volatility.

Hull, J. (2009). *Options, Futures and Other Derivatives*. Options, Futures and Other Derivatives. Pearson/Prentice Hall.

Hull, J. and White, A. (1987). The pricing of options on assets with stochastic volatilities. *The journal of finance*, 42(2):281–300.

Hull, J. C. and White, A. (2016). Optimal delta hedging. *Available at SSRN 2658343*.

Hurn, A. S., Lindsay, K. A., and McClelland, A. J. (2015). Estimating the parameters of stochastic volatility models using option price data. *Journal of Business & Economic Statistics*, 33:579–594.

Karatzas, I., Lehoczky, J. P., and Shreve, S. E. (1987). Optimal portfolio and consumption decisions for a "small investor" on a finite horizon. *SIAM journal on control and optimization*, 25(6):1557–1586.

Karatzas, I. and Shreve, S. E. (1998). *Methods of mathematical finance*, volume 39. Springer Science & Business Media.

Karoui, N. E., Jeanblanc-Picquè, M., and Shreve, S. E. (1998). Robustness of the black and scholes formula. *Mathematical finance*, 8(2):93–126.

Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scandinavian Journal of Statistics*, 24(2):211–229.

Keynes, J. M. (1921). A treatise on probability.

Kloeden, P. E. and Platen, E. (1992). *Numerical solution of stochastic differential equations*, volume 23. Springer.

Knight, F. H. (1921). Risk, uncertainty and profit. *Boston: Houghton Mifflin*.

Lewis, A. L. (2000). *Option valuation under stochastic volatility*. Finance press.

Liu, J. (2007). Portfolio selection in stochastic environments. *Review of Financial Studies*, 20(1):1–39.

Liu, J. and Pan, J. (2003). Dynamic derivative strategies. *Journal of Financial Economics*, 69(3):401–430.

Lyons, T. J. (1995). Uncertain volatility and the risk-free synthesis of derivatives. *Applied mathematical finance*, 2(2):117–133.

Mahayni, A. B., Schlogl, E., and Schlögl, L. (1999). Robustness of gaussian hedges and the hedging of fixed income derivatives. *Available at SSRN 159668*.

Merton, R. C. (1969). Lifetime portfolio selection under uncertainty: The continuous-time case. *The review of Economics and Statistics*, pages 247–257.

Merton, R. C. (1971). Optimum consumption and portfolio rules in a continuous-time model. *Journal of economic theory*, 3(4):373–413.

Merton, R. C. (1973). Theory of rational option pricing. *The Bell Journal of economics and management science*, pages 141–183.

Merton, R. C. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of financial economics*, 3(1-2):125–144.

Mikhailov, S. and Nögel, U. (2004). *Hestons stochastic volatility model: Implementation, calibration and some extensions*. John Wiley and Sons.

Milborrow. Derived from mda:mars by T. Hastie and R. Tibshirani., S. (2011). *earth: Multivariate Adaptive Regression Splines*. R package.

Munk, C. (2013). Dynamic asset allocation. Unpublished lecture notes.

Nassim, T. (2007). The black swan: the impact of the highly improbable.

Oum, Y., Oren, S., and Deng, S. (2006). Hedging quantity risks with standard power options in a competitive wholesale electricity market. *Naval Research Logistics (NRL)*, 53(7):697–712.

Pedersen, A. R. (1994). Uniform residuals for discretely observed diffusion processes. Technical report, Department of Theoretical Statistics, University of Aarhus. Research Report No. 295.

Platen, E. and Bruti-Liberati, N. (2010). *Numerical solution of stochastic differential equations with jumps in finance*, volume 64. Springer.

Poulsen, R., Schenk-Hoppé, K. R., and Ewald, C.-O. (2009). Risk minimization in stochastic volatility models: Model risk and empirical performance. *Quantitative Finance*, 9(6):693–704.

Prakasa Rao, B. (1983). Asymptotic theory for non-linear least squares estimator for diffusion processes. *Statistics: A Journal of Theoretical and Applied Statistics*, 14(2):195–209.

Privault, N. (2013). Notes on stochastic finance. Lecture Notes FE6516, http://www.ntu.edu.sg/home/nprivault/indext.html.

Quenez, M.-C. (1997). *Stochastic control and BSDEs*. Addison Wesley Longman, Harlow.

Rasmussen, N. (2001). Hedging with a misspecified model. *Available at SSRN 260754*.

Reiß, O. and Wystup, U. (2001). Computing option price sensitivities using homogeneity and other tricks. *The Journal of Derivatives*, 9(2):41–53.

Ribeiro, A. and Poulsen, R. (2013). Approximation behoves calibration. *Quantitative Finance Letters*, 1(1):36–40.

RiskMetrics (1996). Riskmetrics technical document. Technical report, JP Morgan/Reuters. https://www.msci.com/documents/10199/5915b101-4206-4ba0-aee2-3449d5c7e95a.

Rogers, L. C. G. (2001). The relaxed investor and parameter uncertainty. *Finance and Stochastics*, 5(2):131–154.

Romano, M. and Touzi, N. (1997). Contingent claims and market completeness in a stochastic volatility model. *Mathematical Finance*, 7(4):399–412.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *The Annals of Mathematical Statistics*, 23(3):pp. 470–472.

Runggaldier, W. J. (2003). Jump-diffusion models. *Handbook of heavy tailed distributions in finance*, 1:169–209.

Samuelson, P. A. (1965). Proof that properly anticipated prices fluctuate randomly. *IMR; Industrial Management Review (pre-1986)*, 6(2):41.

Scott, L. O. (1987). Option pricing when the variance changes randomly: Theory, estimation, and an application. *Journal of Financial and Quantitative analysis*, 22(04):419–438.

Sepp, A. (2014). Log-normal stochastic volatility model: New insight and closed-form solution for vanilla options. Workning paper available at SSRN.

Shreve, S. E. (2004). *Stochastic calculus for finance II: Continuous-time models*, volume 11. Springer Science & Business Media.

Sørensen, M. (1999). On the asymptotics of estimating functions. *Brazilian Journal of Probability and Statistics*, 13:111–136.

Stein, E. M. and Stein, J. C. (1991). Stock price distributions with stochastic volatility: an analytic approach. *Review of financial Studies*, 4(4):727–752.

Stein, J. (1989). Overreactions in the options market. *The Journal of Finance*, 44(4):1011–1023.

Stoft, S. (2002). Power system economics. *Journal of Energy Literature*, 8:94–99.

Weron, R. (2007). *Modeling and forecasting electricity loads and prices: A statistical approach*, volume 403. John Wiley & Sons.

Weron, R. and Wystup, U. (2011). Heston's model and the smile. In *Statistical tools for finance and insurance*, pages 161–181. Springer.

Willems, B. and Morbee, J. (2010). Market completeness: How options affect hedging and investments in the electricity sector. *Energy Economics*, 32(4):786–795.

Wilmott, P. (1998). *Derivatives*. Wiley.

Wilmott, P. (2006). Volatility forecasting, option trading and crashmetrics. Talk available http://www.wilmott.com/images/246/VolForecastingOpTradingCM.wmv.

Wittgenstein, L. (1922). *Tractatus logico-philosophicus*.

Wong, B. and Heyde, C. (2006). On changes of measure in stochastic volatility models. *International Journal of Stochastic Analysis*, 2006.