René Rosenberger, BSc

ON AUTOCALLABLE STRUCTURED PRODUCTS IN THE HYBRID HESTON-HULL-WHITE MODEL WITH CORRELATION

MASTER'S THESIS

to achieve the university degree of

Diplom-Ingenieur

Master's degree programme: Financial and Actuarial Mathematics

submitted to

Graz University of Technology

Supervisor

O. Univ.-Prof. Dr. phil. Robert TICHY
Institute of Analysis and Number Theory

Co-Supervisor

Dipl.-Ing. Dr. techn. Stefan THONHAUSER
Institute of Analysis and Number Theory

Graz, July 2016

AFFIDAVIT

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present master‘s thesis.

Date

Signature

Abstract

This thesis examines the valuation of autocallable structured products and the calibration of the associated model to market data. There is a big variety of such products in the market. The ones this thesis examines are path dependent options in a hybrid interest rate and equity market. An appropriate formulation of such products requires a model which takes correlated stochastic interest rates into account, i.e., we combine the very sophisticated stochastic volatility model from Heston with the Hull-White interest rate process. Due to their path dependence, pricing is usually done using a PDE or MC approach. For the calibration to standard European call prices from the market, however, highly efficient Fourier techniques can be applied.

The contribution of this thesis is the following: We will implement a PDE method and also an efficient MC simulation scheme, which compute prices for the desired products. Both methods will be compared using a set of challenging test cases from practice. Calibration of the Heston-Hull-White model is due to the correlated stochastic interest rates a quite challenging task. To simplify the procedure we will perform a switch to the forward measure. A variety of Fourier based valuation formulas can be found in the literature. We will adapt our preferred one to our needs and use it in a standard optimization routine to find the desired parameters of the model. Additionally the obtained parameters will be compared to the ones which were calibrated to the pure Heston model. In the end, we will discuss the results and give some thoughts on how the calibration can be simplified and the pricing more efficient.

Acknowledgements

This thesis denotes the end of my studies. Therefore, I would like to take the opportunity to express my gratitude to a number of people who supported me on my way.

I want to thank Dr. Stefan Thonhauser, who guided me through this thesis and also facilitated my visit to the University of Lausanne. I want to thank my supervisor Professor Robert Tichy, for his support and cooperativeness. I want to thank Professor Joël Wagner from the University of Lausanne, for his useful input during my stay; he and his whole team made me feel very welcome and I can look back to an awesome time I had in Switzerland. I also want to thank Dr. Ingo Schneider from DEKA bank, who introduced me to this interesting subject, gave useful comments and provided me with some market data. I am also thankful that I have received a funding from the local insurance company GRAWE, who honor research topics in the area of financial and insurance mathematics.

On a personal level, I want to thank my parents; not only for their support and for making it possible to live my student life without any sorrows, but rather for teaching me important values in life which made me the person I am today.

Finally, and in particular, I want to say that I am very glad that I have always had friends, who shared the pleasures of life outside of the university with me. I was lucky to get to know, and become friends with, some incredible characters who I can now share countless unforgettable stories with. Thanks to you, I will always have a smile on my face looking back to my student years. I am sure we will share a lot more of those memories in the future.

Contents

Chapter 1

Introduction

Over the last couple of decades, the study of finance transformed from an almost exclusively descriptive, non-mathematical business, to a highly analytical one. Sophisticated mathematics was deployed due to the emerge and rapid growth of a global market in financial derivatives. This market emerged into a hundreds of trillions dollar business. To trade this huge amount of derivatives without a guide on how to value them and how to hedge the risk they involve would be very problematically. In recent years financial engineers created more and more complex products. These exotic and hybrid derivatives are based on a combination of underlyings from different asset classes. One kind of them are autocallable structured products. The so-called autocallables have become increasingly popular among private investors in the last years. They are structured products linked to one (or more) underlying assets and belong to the class of exotic options. The name autocallable stems from the feature that, if certain predefined market conditions regarding its underlying are met, they will be called automatically, paying off the investor. This work focuses on autocallables linked "only" to one underlying. We also presume to have discrete call dates - where, in contrast to the continuous ones - the product will only be called at certain observation dates prior to its scheduled maturity. These observation dates are typically annually based. An early kick out will occur, for example, if the observed level of the underlying is at or above a fixed trigger level.

There is a big variety of autocallables on the market. A very common structure would have payoffs of the following form: If the price of the linked asset is greater or equal to the trigger level at one of the observation dates, it is called, and pays a pre-specified fixed-rate return. The trigger level is allowed to vary and the return is accumulated if the product has not been called previously. If the product is never called, meaning it is below the trigger levels at every call date, the investor will be exposed to the downside of

the reference asset and receives the same negative percentage return. Figure 1.1 visualizes the just described product. These contracts are exclusively traded over-the-counter (OTC). This implies a great interest on efficient and accurate pricing methods for these products, as well as appropriate market modeling.

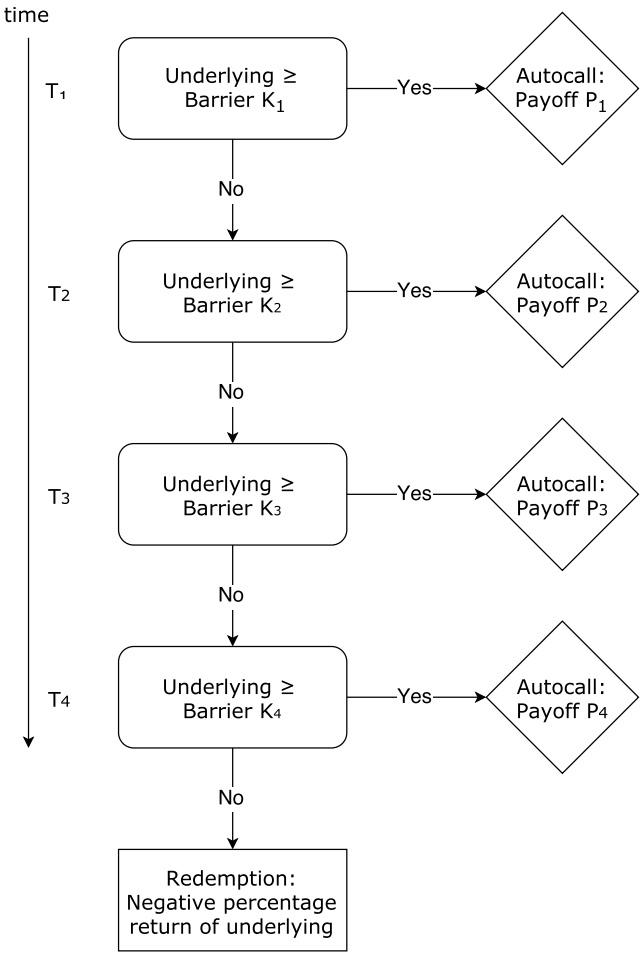

Figure 1.1: Payoff structure of our autocallable

The valuation of financial products is, due to the *fundamental Theorem of Asset Pricing*, a task of calculating their expected discounted payoffs at maturity with respect to a risk neutral pricing measure. For exotic options like our autocallables, this is mainly limited to the use of Monte-Carlo methods in order to compute this expectation numerically. The essence of those methods is to directly simulate the underlying dynamics and then calculating

the average result. There is a big variety of literature on MC applications in finance available, as for instance the book from Glassermann [14]. Since a very high number of paths need to be simulated, Monte-Carlo methods require tremendous computational effort, making them not applicable in many situations.

However, there is a link between a stochastic process driving the dynamics of the underlying and partial differential equations (PDEs). Those PDEs are then solved numerically, i.e. via finite difference discretization schemes. Such schemes are widely spread in applied mathematics and hence a lot of literature is available, see for example the books of Seydel [42] or Duffy [13] for financial applications. Of special interest for us is the work from Haentjens & in 't Hout [18] who use a so-called alternating direction implicite (ADI) time discretization scheme. Those are superior over standard methods such, as the Crank-Nicolson scheme, when dealing with multi dimensional PDEs with mixed derivatives. In general, their implementation demands more effort than for MC simulation, nevertheless, they are advantageous in almost every other aspect, like computational costs. Additionally, one gets option values for every knot of the grid on which the PDE is solved. This means, while having to run several MC simulations to calculate an options sensitivities one gets the, so called, "greeks" for free using PDE methods. Bottom line, these two methods are very powerful when pricing exotic options, for which no closed form solution can be obtained. Whereas they are of no use if valuation needs to be done in real-time or for numerous contracts at once. Evaluating OTC traded products like autocallables in real-time is of no big concern, however, when calibrating the parameters of a model to market data, speed is crucial. Conveniently, calibration only makes sense using highly liquid products such as vanilla options, which are, after all the simplest ones.

When implementing an algorithm to calibrate an option pricing model, one has to choose a method for pricing such vanilla options. The Fourier framework offers a highly efficient way of doing so, in any case where a closed-form expression of the underlying models characteristic function (ChF) is available. Evaluating an options price is left to a numerical integration of that ChF. Through Duffie's transform analysis [12], a characteristic function is easily derived, for model dynamics in the class of affine jump diffusions. Also for a high number of Lévy processes the ChF is available in closed form, however, in this thesis we will abandon jumps and focus on an affine diffusion. Carr and Madan, in their famous work [7], introduced the Fast Fourier transform (FFT) for pricing plain vanilla options through Fourier inversion. They criticize the direct integration method to be unable to use the computational power of FFT and point out numerical instabilities in case of using a decomposition of an option price into probability elements,

like Heston [20] first did when he introduced his famous stochastic volatility model in 1993. However, Lewis [30] and Attari [4] came up with modifications of the direct integration method which are free from this instability. Further, Kilin [27] points out two crucial restrictions using the FFT method. The first is that the grid spacings must satisfy certain conditions and the number of grid points must be a power of two. Additionally, the grid points for the numerical integration must be chosen equidistantly, eliminating the use of effective integration rules, i.e., the Gaussian quadrature. His point is, that if one implements a smart caching technique, also the direct integration has the advantage of simultaneously computing the ChF for different strikes because the integral is independent of K. Consequently, the number of evaluations of the ChF is the driving factor for the speed of the calibration. Depending on the desired accuracy the FFT method is outperformed with a factor of seven and even higher.

Developing an appropriate model which describes the market well is inevitable. Yet it has to be mathematically managable, i.e., allowing for fast and accurate pricing. This has been the moving cause for a great amount of literature. A milestone was definitely the work of Black and Scholes [6] and Merton [34], with their model based on the geometric Brownian motion. It is still the benchmark in practice, particularly due to the availability of explicit results. Nevertheless, some downsides are also very well known. It is not able to deal with many features that are empirically observed in asset return data, such as fat tails, volatility smiles (and smirks), volatility clustering and leverage effects. Consequently, more sophisticated models have been developed. In the so called *Stochastic Volatility* (SV) models, volatility is assumed to evolve according to a stochastic diffusion process; contrary to the Black-Scholes case, where it is a constant. Among the most famous works are the ones by Hull and White [21], Stein and Stein [43] and, above all, Heston [20]. In his famous work Heston added a square root diffusion to the price process as instantaneous volatility, and also allowed for a correlation between the two driving Brownian motions. Alternatively to the SV models, authors like Merton [35] and Kou [29] introduced jump-diffusions. They tried to describe the observed features of the market by adding jumps, from an independent compound Poisson process, to the assets price dynamics. Furthermore, Bates [5] combines both approaches. Subsequently, modeling with jumps gained a lot of popularity; in particular Lévy processes, which are stochastic processes with independent and stationary increments. See the book of Cont and Tankov [9] for a vast overview on this type of models. Even though a lot of (successfull) research has been done for this very general class of processes, this thesis will focus on continuous diffusions. Models have been developed to particularly meet the hybrid features of autocallables.

Therefore, stochastic volatility models were extended to include the term structure of interest rates. The aim is to combine them, with the Hull-White [22] or Cox-Ingersoll-Ross [10] model, in which the parameters are consistent with market prices of caps and swaptions. In particular, we chose the Heston stochastic volatility model and combine it with the Hull-White interest rate process. Correlations will be imposed for the assets price and its volatility, and between price and interest rate. Regarding the pricing of the autocallables, this will be no problem when using PDE or MC methods. However, when trying to calibrate the model to market data, extra work has to be done for the later correlation. We shall see under which conditions the Heston-Hull-White (HHW) model, with two correlations, fits into the class of affine (jump-) diffusions (AJD) from Duffie [12], in order to establish highly efficient vanilla option valuation.

This thesis is organized as follows: In Chapter 2, we will introduce the models in use and explain some basic theory in order to put market modeling on a solid mathematical base. Additionally some fundamental results on stochastic differential equations, market modeling and complex analysis are referred to the Appendix. The market modeling is followed by discussing the three main pricing methods for financial products in Chapter 3. Where the MC simulation and PDE technique are straight forward and the advanced reader should be very familiar with those. For the third, namely Fourier method, a lot of variations can be found in the literature. Due to our special demand on correlated interest rates a little extra work has to be done in the last section of this chapter. After that, Chapter 4 will show in detail how these three pricing methods were implemented within the Heston-Hull-White model. All the computational studies are contained in Chapter 5, where we verify the correctness of our implementations and interpret the results. Finally, Chapter 6 will conclude.

Chapter 2

Market modeling

2.1 Continuous time security markets

Before discussing the used models, a short review of general theory for a continuous time security market is given here. We follow the lines of Musiela & Rutkowski [37, Ch. 10].

Definition 1 (continuous time market model)**.** *A continuous time market model consists of the following:*

(i) a probability space $(\Omega, \mathcal{A}, \mathbb{P})$*,*

(ii) the filtration $\mathcal{F} = (\mathcal{F}_t)_{0 \le t \le T}$ *of* $\mathcal{A}$*, describing the progress of information (the filtration is assumed to satisfy the "usual conditions", meaning it is right continuous, i.e.,* $\mathcal{F}_t = \cap_{u>t} \mathcal{F}_u$ *for every* $t < T$*; and* $\mathcal{F}_0$ *contains all null sets, i.e., if* $B \subset A \in \mathcal{F}_0$ *and* $\mathbb{P}\{A\} = 0$*, then* $B \in \mathcal{F}_0$*)*

(iii) and a $(d+1)$*-dimensional price process* $S = (S_t)_{0 \le t \le T}$ *following a semimartingale with* $S_t = (S_t^0, S_t^1, \ldots, S_t^d)$*.*

We also assume that S^0 is a strictly positive benchmark, thus the numeraire asset. Further $\tilde{S} = (\tilde{S}_t)_{0 \le t \le T}$ with $\tilde{S}_t = S_t / S_t^0 = (1, \tilde{S}_t^1, \ldots, \tilde{S}_t^d)$ is called the discounted (relative) price process.

Definition 2 (wealth process, self financing trading strategy and gains process)**.** *A* $\mathbb{R}^{(d+1)}$ *valued predictable stochastic process* $\phi_t = (\phi_t^0, \ldots, \phi_t^d)$ *is called a self financing trading strategy if the wealth process* $V(\phi)$*, which is*

$$V_t(\phi) = \phi_t \cdot S_t = \sum_{i=0}^{d} \phi_t^i S_t^i, \quad \forall t \in [0, T],$$

satisfies $V_t(\phi) = V_0(\phi) + G_t(\phi)$ *for* $t \leq T$*, with the gains process*

$$G_t(\phi) = \int_0^t \phi_u \cdot dS_u = \sum_{i=0}^{d} \int_0^t \phi_u^i dS_u^i, \quad \forall t \in [0, T].$$

This definition reflects, that there is no infusion of new funds to, nor withdrawal of money from, the portfolio; the only changes are due to capital gains. It also means that there are no transaction costs or restrictions on short selling in the market.

Definition 3 (arbitrage)**.** *[37, Def. 10.1.1] A trading strategy* ϕ *is called an arbitrage, if it is admissible and the wealth process satisfies*

$$V_0(\phi) = 0, \quad \mathbb{P}(V_T(\phi) \geq 0) = 1, \quad \mathbb{P}(V_T(\phi) > 0) > 0.$$

Definition 4 (claim, attainable claim and replicating trading strategy)**.** *A trading strategy* ϕ *replicates a European contingent claim* X*, which is modeled by means of a* $\mathcal{F}_T$*-measurable random variable, if* $V_T(\phi) = X$ $I\!P - a.s.$*. The claim* X *is said to be attainable if it admits at least one replicating trading strategy.* $V(\phi)$ *is called replicating process of* X*.*

Definition 5 (equivalent martingale measure)**.** *A probability measure* $\mathbb{Q}$*, equivalent to* $\mathbb{P}$*, is called a martingale measure for* $\tilde{S}$ *if the relative price process follows a local martingale under* $\mathbb{Q}$*.*

Definition 6 (admissible trading strategy)**.** *If the relative gains process*

$$\tilde{G}_t(\phi) = \int_0^t \phi_u \cdot d\tilde{S}_u, \quad \forall t \in [0, T],$$

follows a martingale under $\mathbb{Q}$*, then the trading strategy* ϕ *is called* $\mathbb{Q}$*-admissible. A trading strategy is said to be tame if it is bounded from below, i.e. there exist* $m \in \mathbb{R}$ *such that* $\mathbb{P}$*-a.s.:* $\tilde{V}_t(\phi) \geq m$ *for all* t*.*

For an attainable claim, the wealth process $V(\phi)$ of any $\mathbb{Q}$-admissible strategy replicating X is called the arbitrage price of X and denoted by $\pi(X|\mathbb{Q})$.

Definition 7 (expected value process of a claim)**.** *Let* X *be a (not necessarily attainable) claim such that the random variable* X/S_T^0 *is* $\mathbb{Q}$*-integrable. Define the expected value process of* X *relative to* S^0 *under* $\mathbb{Q}$ *by*

$$\nu_t(X|\mathbb{Q}) = S_t^0 \mathbb{E}_{\mathbb{Q}}(X/S_T^0|\mathcal{F}_t), \quad \forall t \in [0, T].$$

The following Proposition shows that the arbitrage price of any contingent claim agrees with the associated expected value process.

Proposition 1 (risk neutral valuation formula)**.** *[37, Prop. 10.1.2 and 10.1.4] For any attainable claim X we have*

$$\pi_t(X|\mathbb{Q}) = \nu_t(X|\mathbb{Q}), \quad \forall t \in [0,T].$$

Further, the no-arbitrage price of X satisfies

$$\pi_0(X) = \sup_{\mathbb{Q}} S_0^0 \mathbb{E}_{\mathbb{Q}}(X/S_T^0) = \inf_{\phi \in \Theta(X)} V_0(\phi),$$

where $\Theta(X)$ is the class of tame trading strategies which replicate X. Such a trading strategy is also called a hedge.

Theorem 1. *[37, Sec. 10.1.5][fundamental theorem of asset pricing] If there exists a to $\mathbb{P}$ equivalent measure $\mathbb{Q}$ such that the discounted price process $\tilde{S}$ is a local martingale, then the model is arbitrage free.*

Remark: This is the practically more relevant direction of the theorem. In the case of positive locally bounded price processes Delbaen & Schachermayer [11] showed that the existence of $\mathbb{Q}$ is equivalent to the weaker condition "no free lunch with vanishing risk".

Theorem 2 (complete market)**.** *If every claim X has a hedge, then the equivalent martingale measure is unique. An arbitrage free market model is complete, if and only if the equivalent martingale measure is unique.*

2.1.1 The Black-Scholes model

In 1973, by far the most famous market model for pricing derivatives was introduced by Black and Scholes [6] and Merton [34]. Their associated Black-Scholes formula for the pricing of European Call and Put Options was revolutionizing the financial world and the surviving inventors Merton and Scholes received the Nobel Price for Economic Sciences in 1997. The model consists of a risk-free bond $S_t^0 = e^{rt}$ as the numéraire with $r > 0$ being the constant non-risky interest rate, and a risky asset $S^1 = (S_t^1)_{0 \le t \le T}$ whose price dynamics are modeled as

$$dS_t = \mu S_t dt + \sigma S_t dB_t, \tag{2.1.1}$$

with constant and deterministic drift $\mu \in \mathbb{R}$, the asset prices volatility $\sigma > 0$, a given initial price $S_0 > 0$ and a standard Brownian motion $B = (B_t)_{0 \le t \le T}$

on $\{F_t^B\}_{t\in[0,T]}$. Solving this stochastic differential equation gives the following explicit solution

$$S_t = S_0 \exp\left\{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma B_t\right\}, \quad 0 \le t \le T.$$

The Brownian motion B is with respect to the "real life" measure $\mathbb{P}$. For pricing purposes one needs to switch to an equivalent Martingale measure $\mathbb{Q}$, which is done with Girsanov's theorem. A multidimensional version of this theorem can be found in the Appendix A.3. Now let us derive the, as we shall see, unique equivalent martingale measure $\mathbb{Q}$ on $\mathcal{F}_T$: If the discounted price process $\tilde{S}^1$ admits the representation

$$d\tilde{S}_t^1 = Y_t dW_t$$

where W is a Brownian motion with respect to the measure $\mathbb{Q}^1$, then $\tilde{S}^1$ is a local Martingale with respect to $\mathbb{Q}$. Calculating $d\tilde{S}_t^1$ yields

$$\begin{aligned} d\tilde{S}_t^1 = d(e^{-rt}S_t^1) &= d(e^{-rt})S_t^1 + e^{-rt}dS_t^1 + d(e^{-rt})d(S_t^1) \\ &= -re^{-rt}S_t^1 dt + e^{-rt}S_t^1(\mu dt + \sigma dB_t) \\ &= \tilde{S}_t^1\left((\mu - r)dt + \sigma dB_t\right) \\ &= \sigma\tilde{S}_t^1 dW_t, \end{aligned}$$

therefore, granted by Girsanov's Theorem, W needs to be a Brownian motion with respect to $\mathbb{Q}$ with

$$W_t = B_t - \frac{r-\mu}{\sigma}t. \tag{2.1.2}$$

Due to $d\tilde{S}_t^1 = \sigma S_t^1 dW_t$ we have a local Martingale $\tilde{S}_t^1$ with respect to $\mathbb{Q}$; it is even a Martingale. The quantity $\lambda := \frac{\mu-r}{\sigma}$ is called *market price of risk*. In this case one can easily see that this choice of λ is the only one making $\tilde{S}_t^1$ a Martingale, hence the Black-Scholes model is complete.

Apparently these very restrictive assumptions deliver some downsides which motivated the development of more general models. For example, the Black-Scholes model claims for the logarithmic returns $\ln S_t$ to be normally distributed. The distribution observed from real market data, however, shows that the tails are significantly heavier than those induced by the Gaussian distribution. There is also a certain skewness noticeable and moreover it doesn't describe the so called *leverage effect*. This effect refers to the

[1] In this thesis we will always denote the Brownian motions with respect to $\mathbb{P}$ and $\mathbb{Q}$ with B and W respectively.

phenomenon that large down-moves in the assets price correlate with an increasing volatility. Also the implied volatility function, which is obtained if one plugs in the current market value of an option into the option pricing formula and solve for the volatility, can be considered as a measure of weakness of the model. Doing so, the implied volatility as a function in strike, is non constant and one observes *volatility smiles* or *smirks*, concluding that the Black-Scholes model does not describe the market correctly.

To address the above mentioned shortcomings the model has been altered in various ways. Merton [35] and Kou [29] for example added a jump process to the asset price dynamics. However, jump diffusions still cannot describe the *leverage effect*; therefore, other than adding jumps to the asset price process, the most successful variations introduced a (positive) stochastic process $v = (v_t)_{t\geq 0}$ to model the diffusion coefficient. This process describes the instantaneous volatility of the assets returns over the infinitesimal interval $(t, t + dt)$. The most common diffusion based stochastic volatility model is the following Heston model.

2.1.2 The Heston model

This model was introduced by Heston [20] in 1993. Here the constant volatility parameter σ from (2.1.1), is chosen to be replaced by a Cox-Ingersoll-Ross (CIR) process which follows the dynamics given by the SDE

$$dY_t = \kappa(\theta - Y_t)dt + \sigma\sqrt{Y_t}dB_t, \tag{2.1.3}$$

where $B = (B_t)_{t\geq 0}$ is again a standard Brownian motion and $\sigma > 0, \kappa > 0, \theta > 0$. Under these assumptions, with a positive starting value $Y_0 = y_0$, the process stays non-negative. Furthermore the mean reversion feature ensures for finite volatility. This square root process was introduced by Cox et al. [10], trying to model the dynamics of short rates. It can be shown that Y almost sure never reaches zero, if the parameters are chosen as to grant the restriction $2\kappa\theta \geq \sigma^2$, also known as the *Feller condition*. Although this condition is very rarely satisfied for financial applications, the process remains a good choice due to its main characteristic of mean reversion. If $Y_t \geq \theta$ the negative drift pulls the process downwards and vice versa, making it oscillate around θ. Therefore θ is interpreted as the long term volatility. The parameter κ determines the rate of this mean reversion.

Contrary to the Black-Scholes SDE it is not obvious that (2.1.3) has a solution on $[0, T]$ since the square root is not Lipschitz continuous in zero. However, in the case of a one dimensional Brownian motion it can be shown, e.g. in Müller [36], that the condition on the diffusion term can be weakened

to be Hölder continuous with order $1/2$, that is:

$$|\sigma(x) - \sigma(y)| \leq K \max(|x-y|^{1/2}, |x-y|),$$

hence, the SDE (2.1.3) has a unique strong solution[2]. Finally, the Heston model describes the price process as the dynamics of the two dimensional SDE

$$\begin{aligned} dS_t/S_t &= \mu dt + \sqrt{v_t} dB_t^S, \\ dv_t &= \kappa(\theta - v_t)dt + \sigma\sqrt{v_t} dB_t^v, \end{aligned} \tag{2.1.4}$$

with a constant correlation $dB_t^S dB_t^v = \rho dt$ between the two driving Brownian motions B_t^S and B_t^v. A Cholesky decomposition of the correlation matrix gives their relation as

$$\begin{pmatrix} B_t^v \\ B_t^S \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix} \begin{pmatrix} Z_t^1 \\ Z_t^2 \end{pmatrix}$$

where Z^1 and Z^2 are two independent Brownian motions. The coefficient $\rho \in [-1, 1]$ is indeed the instantaneous correlation; as when calculating the differential of the quadratic covariation

$$\begin{aligned} d[B^v, B^S]_t &= d[Z^1, \rho Z^1 + \sqrt{1-\rho^2} Z^2]_t \\ &= d([Z^1, \rho Z^1]_t + [Z^1, \sqrt{1-\rho^2} Z^2]_t) \\ &= d(\rho[Z^1, Z^1]_t + \sqrt{1-\rho^2}[Z^1, Z^2]_t) = \rho dt. \end{aligned}$$

Contrary to the Black-Scholes model now we have two sources of uncertainty, B^S and B^v, yet only one risky asset available for trade. Applying Girsanov's theorem from A.3 we get infinitely many equivalent martingale measures. Similar to (2.1.2) in the Black-Scholes case we have

$$W_t^v = B_t^v + \frac{\lambda}{\sigma}\sqrt{v_t} t,$$

where $(\lambda/\sigma)\sqrt{v}$ with $\lambda \in \mathbb{R}$ is the market price of volatility risk. The model under the new martingale measure therefore is

$$\begin{aligned} dS_t/S_t &= r dt + \sqrt{v_t} dW_t^S, \\ dv_t &= (\kappa(\theta - v_t) - \lambda v_t)dt + \sigma\sqrt{v_t} dW_t^v. \end{aligned}$$

[2]For completeness reasons, the Appendix defines the terms *weak* and *strong* solution of a SDE.

Thus the Heston model is incomplete but still free of arbitrage; hence a particular martingale measure (or equivalently market price of volatility risk) has to be chosen to determine option prices.

The correlation introduced in (2.1.4) provides no difficulties since the system is still of the affine form[3], and hence we can model the above mentioned *leverage effect* by choosing a negative value for ρ. Moreover, it accounts for long term implied volatility smiles or skews and volatility clustering. Unfortunately it does not model the short term implied volatility in a realistic way. To address this problem one can add jumps to the dynamics of the price process like Bates [5] did. However, to model short term volatilities does not seem that important when pricing autocallables. Therefore we leave this possible extension aside and focus on another weakness. Autocallables are hybrid products with a typical maturity of several years, a more realistic model therefore would account for a stochastic interest rate. A good candidate for a short rate process is the one proposed from Hull and White [22], which is widely used in practice.

2.2 Fixed income markets

A fixed income market is one on which various interest rate sensitive instruments are traded, such as bonds, swaps, swaptions, caps, etc. Before introducing the model we need to adapt the definition of arbitrage from the previous security market. So let $B(t,T)$ be the price of a zero coupon bond (ZCB) at time t with face value 1; this means that it pays 1 at maturity T. The price process $(B(t,T))_{t\in[0,T]}$ follows a strictly positive adapted process on the filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathcal{F}$ is the completed filtration of an underlying d-dimensional standard Brownian motion. Suppose the adapted process r, given on the same probability space, models the short term interest rate (=spot rate), with $B_t = \exp(\int_0^t r_u du)$ the savings account process, is defined.

Definition 8 (arbitrage free family of bond prices)**.** *[37, Def. 12.1.1] We call a family $B(t,T)$ with $t \le T \le T^*$ of adapted processes an arbitrage free family of bond prices relative to r, if the following conditions hold:*

(i) $B(T,T) = 1$ *and*

(ii) there exists a probability measure $\mathbb{Q}$ equivalent to $\mathbb{P}$ such that the relative bond price

$$\tilde{B}(t,T) = B(t,T)/B_t, \quad \forall t \in [0,T] \text{ and } \forall T \in [0,T^*],$$

[3]It will be explained in section 4.3.1 what that means and why it is important for us.

follows a martingale under $\mathbb{Q}$.

The reason why it is assumed that the relative bond price follows a martingale, and not a local martingale, is that then we have trivially $\tilde{B}(t,T) = \mathbb{E}_{\mathbb{Q}}(\tilde{B}(T,T)|\mathcal{F}_t)$, such that the bond price is

$$B(t,T) = B_t \mathbb{E}_{\mathbb{Q}}\left(B_T^{-1}|\mathcal{F}_t\right) = \mathbb{E}_{\mathbb{Q}}\left(e^{-\int_t^T r_u du}|\mathcal{F}_t\right), \quad \forall t \in [0,T].$$

Conversely, given any non negative short rate r on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and any equivalent probability measure $\mathbb{Q}$, the family $(B(t,T))_{t\leq T\leq T^*}$ is easily seen to be an arbitrage free family of bond prices.

If one now wants to match the initial yield curve, given a spot rate process r and probability measure $\mathbb{Q}$, the term structure $B(0,T)$ is uniquely determined via

$$B(0,T) = \mathbb{E}_{\mathbb{Q}}\left(e^{\int_0^T r_u du}\right),$$

for all maturities $T \in [0,T^*]$. Matching the yield curve based on a specific spot rate process a lot more effort is needed than in the case where one incorporates the initial term structure as an input of the model.

2.2.1 The Hull-White model

Extending Vasicek's model by a time dependent mean reverting drift term, Hull and White [22] introduced their model for pricing interest rate derivative securities in 1990. It is given through the following SDE

$$dr_t = a(b(t) - r_t)dt + \sigma dB_t, \tag{2.2.1}$$

where $b(t) > 0, t \in \mathbb{R}^+$, is a time dependent drift term, which reflects the observed term structure from the market. The parameters σ and a, as usual, determine the overall level of volatility and the reversion rate, respectively. Note the similarity to the CIR type process used for the instantaneous volatility in the Heston model. In contrary to CIR the diffusion term is kept constant whereas the mean reversion $b(t)$ is time dependent. Applying Itô's formula to $\mathrm{d}(\exp(at)r_t)$ and integration yields, for $t \geq 0$,

$$r_t = r_0 e^{-at} + a\int_0^t e^{-a(t-s)}b(s)ds + \sigma\int_0^t e^{-a(t-s)}dB_s.$$

Hence, r_t is normally distributed with

$$\mathbb{E}^{\mathbb{Q}}(r_t|\mathcal{F}_0) = r_0 e^{-at} + a\int_0^t e^{-a(t-s)}b(s)ds,$$

and

$$\mathrm{Var}^{\mathbb{Q}}(r_t|\mathcal{F}_0) = \frac{\sigma^2}{2a}\left(1 - e^{-2at}\right).$$

A very useful simplification is achieved by decomposing it into a simpler stochastic process plus a deterministic one.

Proposition 2. *[16, Prop. 2.1][Hull-White decomposition] The Hull-White process can be decomposed into $r_t = \tilde{r}_t + \psi_t$, where*

$$\psi_t = r_0 e^{-at} + a\int_0^t e^{-a(t-s)} b(s) ds,$$

and

$$d\tilde{r}_t = -a\tilde{r}_t dt + \sigma dB_t, \qquad \textit{with } \tilde{r}_0 = 0.$$

The process $\tilde{r}_t$ is now a classical Ornstein Uhlenbeck mean reverting process, determined by a and σ only, and independent of the term structure described by the deterministic ψ_t. Consequently, it is easier to analyze than the original one.

The value of a zero coupon bond (ZCB) at time t which pays 1 at maturity T, given the initial short rate r_0, is given by the very well known equation

$$\begin{aligned} B(r_0,t) =& e^{c(r_0,t)} \\ c(r_0,t) =& -\frac{r_0}{a}\left(1 - e^{-a(T-t)}\right) - \int_t^T b(s)\left(1 - e^{-a(T-s)}\right) ds \\ &+ \frac{\sigma_2^2}{2a^2}\left(T - t + \frac{2}{a}e^{-a(T-t)} - \frac{1}{2a}e^{-2a(T-t)} - \frac{3}{2a}\right). \end{aligned} \tag{2.2.2}$$

This model is widely used and analytically tractable, making it a good choice for our applications.

2.3 Hybrid markets

2.3.1 The hybrid Heston-Hull-White model

We have already discussed the importance of consideration for every single asset class, and a correlation between them, when modeling hybrid products in the Introduction. The autocallable structured products that are treated in this thesis are equity-interest-rate based derivatives. After introducing widely established models for each of those asset classes in the preceding Sections,

we will now combine them in the following hybrid Heston-Hull-White model, which is given through the three dimensional SDE

$$\begin{aligned} dS_t/S_t &= r_t dt + \sqrt{v_t} dW_t^S, \\ dv_t &= \kappa(\theta - v_t)dt + \sigma_1 \sqrt{v_t} dW_t^v, \\ dr_t &= a(b(t) - r_t)dt + \sigma_2 dW_t^r, \end{aligned} \tag{2.3.1}$$

where the parameters $\kappa, \theta, \sigma_1, a, b(t)$ and σ_2 are as introduced in Chapters (2.1.2) and (2.2.1), respectively. The SDEs are correlated in the sense of

$$\begin{aligned} dW_t^S dW_t^v &= \rho_{Sv} dt, \\ dW_t^S dW_t^r &= \rho_{Sr} dt, \end{aligned}$$

whereby the first one of them is predefined by the Heston model. The latter one describes the influence, movements in stock markets have to the behavior of interest rates. We leave the third possible correlation between volatility and interest rates aside[4].

A switch from the real life measure $\mathbb{P}$ to $\mathbb{Q}$ can again be achieved with Girsanov's theorem and is straight forward as in the preceding examples. To simplify those calculations, here the market prices of volatility and interest rate risk are set to zero and the model (2.3.1) is already stated under $\mathbb{Q}$, the risk neutral martingale measure. It goes without saying that the Heston-Hull-White model is incomplete, having three sources of uncertainty and only two asset classes available for trading.

With the given correlation structure the Brownian motions W^r, W^v, W^S can be written in terms of the independent ones, Z^1, Z^2, Z^3, as

$$\begin{pmatrix} W_t^r \\ W_t^v \\ W_t^S \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \rho_{Sr} & \rho_{Sv} & \sqrt{1 - \rho_{Sv}^2 - \rho_{Sr}^2} \end{pmatrix} \begin{pmatrix} Z_t^1 \\ Z_t^2 \\ Z_t^3 \end{pmatrix}, \tag{2.3.2}$$

where the above matrix is again the Cholesky decomposition of the correlation matrix. The correlation between stock prices and interest rates violates affinity properties, as will be discussed in Section 4.3.1; therefore, an approximation has to be found which makes the model affine. Duffie's famous work [12] then provides transform analysis to derive the characteristic function. Having knowledge of the ChF, highly efficient Fourier techniques can be applied for calibration. All of that will be discussed in section 4.3.

[4]Grzelak and Oosterlee [15] also discuss a model with a full correlation structure. However, they conclude that the benefits of deploying this $v - r-$correlation are not worth the mathematical effort of a more complicated model.

Various alternations of hybrid models can be found in the literature. For example, Grzelak and Oosterlee [15] use for the interest rate process the one from Cox-Ingersol-Ross. Also different stochastic volatility models can be used, like in [16], where the Schöbel-Zhu SV model [41] is used instead of the Heston model. The authors also compare these alternations with each other. In this thesis only the hybrid Heston-Hull-White model will be used.

Chapter 3

Pricing techniques

Valuing derivatives is due to the *fundamental Theorem of asset pricing* 1 and Proposition 1 a matter of computing the discounted expected payoff of the product under a risk neutral pricing measure. Depending on the problem probabilistic or numerical methods can be applied. Probabilistic methods demand knowledge of the (joint) distribution of the underlying model. When valuing exotic options this is rarely the case, hence numerical approaches have to be used. Two very powerful tools are Monte Carlo simulation and PDE techniques. Both will be introduced with their mathematical background in the following two sections. Further we discuss a Fourier transform based method which allows for pricing of plain vanilla options. With that technique highly efficient (semi) closed form solutions can be obtained if the characteristic function of the model is known. Consequently, we will use this method to calibrate the parameters of the model to the given market data.

3.1 Monte Carlo methods

A huge advantage of the Monte Carlo simulation is, that it is straight forward to implement and can be applied easily to every imaginable underlying structure. However, this requires huge computational power to produce decent values. A lot of research has been done to reduce the variance of MC estimators. The book of Glassermann [14] provides a comprehensive collection on this topic. In this section we will just introduce the mathematical foundations following the lines of this book and implement an efficient MC scheme, only as a reference for the comparison with other, for our application more efficient, methods in the following chapter.

The basis of the Monte Carlo methods is the analogy between volume and probability. This means that one samples randomly from a universe of

possible outcomes and then takes the fraction of all the random draws that fall in a given set as an estimate of the set's volume. The estimate is of the form

$$\hat{C}_n = \frac{1}{n}\sum_{i=1}^{n} C_i,$$

where the $\{C_i\}$ are i.i.d. random variables, $\mathbb{E}(C_i) = C$ and $\mathrm{Var}(C_i) = \sigma_C^2 < \infty$. With an increasing number of draws the law of large numbers provides convergence of the estimate to the correct value almost surely. Further, for a finite number of draws, the central limit theorem gives information about the error made in the estimate. The standardized estimator converges for $n \to \infty$ in distribution to the standard normal distribution

$$\frac{\hat{C}_n - C}{\sigma_C/\sqrt{n}} \xrightarrow{d} N(0,1).$$

Put in another way, this means that the distribution of the error estimate is

$$\hat{C}_n - C \approx N(0, \sigma_C^2/n),$$

and hence by replacing σ_C by the sample standard deviation s_C of $C_1, \ldots, C_n$ one can state the asymptotically valid confidence interval

$$\hat{C}_n \pm z_{\delta/2}\frac{s_C}{\sqrt{n}}, \tag{3.1.1}$$

where z_δ is the $1-\delta$ quantile of the standard normal distribution, meaning $\Phi(z_\delta) = 1 - \delta$.

Remark: Strictly speaking, for a finite number n of draws, the expression

$$\frac{\hat{C}_n - C}{s_C/\sqrt{n}}$$

is t-distributed with n degrees of freedom. For the confidence interval in (3.1.1) one therefore has to use the quantile t_δ of the t-distribution, which would make it slightly wider. However, for significant sample sizes ($> 10^5$) the difference is negligible.

Convergence order

In addition to the above described MC error, there is another error category based on the discretization of the SDE. To compare different discretization schemes with respect to their convergence order two categories of error are

commonly used to measure the quality of a discretization scheme. For the SDE

$$\mathrm{d}X_t = b(X_t, t)\mathrm{d}t + \sigma(X_t, t)\mathrm{d}W_t,$$

where b and σ satisfy the conditions for existence and uniqueness of a strong solution (see Appendix A); take $t_0 < t_1 < \cdots < t_n = T$ and let $\hat{X}(t_i)$ be any discrete time approximation to the continuous time process X. Then *strong* and *weak* error criteria are

$$\mathbb{E}(||\hat{X}(T) - X(T)||) \quad \text{or} \quad \mathbb{E}(\sup_{t_i} ||\hat{X}(t_i) - X(t_i)||),$$

and

$$|\mathbb{E}(f(\hat{X}(T))) - \mathbb{E}(f(X(T)))|,$$

with $f : \mathbb{R}^d \to \mathbb{R}$ satisfying some smoothness conditions, respectively. In derivative pricing naturally weak errors are most relevant. A discretization scheme is said to have weak order of convergence $p > 0$ if

$$|\mathbb{E}(f(\hat{X}(T))) - \mathbb{E}(f(X(T)))| \leq ch^p,$$

for some constant c and stepwidth h. Thus, a larger value of p implies a faster convergence and the same scheme will often have a smaller strong order of convergence than weak one. For example, under the conditions discussed in [14, Ch. 6.1.2], the Euler-Maruyama scheme given through

$$\hat{X}(t_{i+1}) = \hat{X}(t_i) + b(\hat{X}(t_i))(t_{i+1} - t_i) + \sigma(\hat{X}(t_i))\sqrt{t_{i+1} - t_i}Z_{i+1},$$

has strong order of $1/2$, but often achieves weak order of 1. However, for more complex models it is less likely that such basic schemes like the Euler yield satisfactory results. In such a case it is therefore often wise to aim for a more efficient scheme with higher order of convergence.

Andersen [3] proposed several efficient extensions for the Heston model. To sample from the Heston-Hull-White model we will use one of his schemes and extend it with a very efficient simulation scheme for the interest rate process based on a Hull-White decomposition. Details are referred to Section 4.1 where we will explain explicitly how to derive and implement it.

3.2 PDE approach

The second very powerful tool when valuing complex structured products are partial differential equations (PDEs). Under mild conditions the characteristics of a SDE can be described by a corresponding PDE reflecting

its dynamics. The specific structure of the derivative can then be modeled via the PDEs final and boundary conditions. Except for some simple cases, like the Black-Scholes PDE, they cannot be solved explicitly and therefore a numerical discretization, e.g. finite differences, has to be applied. In the following steps we will show how to use the famous Feynman-Kac theorem to derive the corresponding HHW PDE. Further we will state a result by Heath and Schweizer [19] which gives sufficient conditions for the existence of a unique classical solution to this PDE. The mathematical theory on the solution, existence and uniqueness of an SDE and its relation to a corresponding PDE is taken from Karatzas & Shreve [26] and Oksendal [38] and summarized in the Appendix A.

3.2.1 The Feynman-Kac theorem

The point of departure is of course the d-dimensional SDE

$$X_s^{t,x} = x + \int_t^s b(u, X_u^{t,x})du + \int_t^s \sigma(u, X_u^{t,x})dW_u \quad \text{for } t \leq s < \infty,$$

for which we assume that,

(i) the drift and diffusion coefficient functions $b_i(t,x), \sigma_{ij}(t,x) : [0,\infty) \times \mathbb{R}^d \to \mathbb{R}$ for $1 \leq i \leq d$, $1 \leq j \leq k$ are continuous and fulfill the linear growth condition

$$\|b(t,x)\|^2 + \|\sigma(t,x)\|^2 \leq K^2 \left(1 + \|x\|^2\right),$$

(ii) further the SDE has a weak solution $(X^{t,x}, W), (\Omega, \mathcal{F}, \mathbb{P}), (\mathcal{F}_s)$ for every starting point (t,x) and

(iii) the solution is unique in law.

The generator of an Itô diffusion (see equation (A.1.3) from the Appendix), in the more general case where the coefficients also depend on the time, is

$$\mathcal{A}_t u(x) = \sum_{i=1}^{d} b_i(t,x)\frac{\partial}{\partial x_i}u(x) + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{k} a_{ij}(t,x)\frac{\partial^2}{\partial x_i \partial x_j}u(x), \qquad (3.2.1)$$

with diffusion $a = \sigma\sigma^T$.

Definition 9 (ellipticity)**.** *We say $\mathcal{A}_t$ is elliptic in $x \in \mathbb{R}^d$ if $\exists \delta > 0$ such that*

$$\sum_{i=1}^{d}\sum_{j=1}^{k} a_{ij}(x)\xi_i\xi_j > \delta \|\xi\|^2 \ \ \forall \xi \in \mathbb{R}^d \backslash \{0\} \ \text{and} \ \forall t \geq 0.$$

Let $D \subseteq \mathbb{R}^d$ be open, then $\mathcal{A}_t$ is elliptic in D if it is elliptic in all $x \in D$ and it is called uniformly elliptic, if there $\exists \delta > 0$ such that $\mathcal{A}_t$ is elliptic in D with the same $\delta > 0, \forall x \in D$.

Cauchy problem and Feynman-Kac representation

Let $T > 0$ arbitrary but fixed, constants $L > 0, \lambda \geq 1$ and $f : \mathbb{R}^d \to \mathbb{R}, g : [0,T] \times \mathbb{R}^d \to \mathbb{R}, k : [0,T] \times \mathbb{R}^d \to \mathbb{R}$. The functions f, g and k are continuous and satisfy

$$\begin{aligned} |f(x)| &\leq L(1 + \|x\|^{2\lambda}) \quad \text{or} \quad f(x) \geq 0 \ \forall x \in \mathbb{R}^d, \\ |g(t,x)| &\leq L(1 + \|x\|^{2\lambda}) \quad \text{or} \quad g(t,x) \geq 0 \ \forall x \in \mathbb{R}^d, 0 \leq t \leq T, \end{aligned}$$

and the operator $\mathcal{A}_t$ as in 3.2.1.

Theorem 3 (Feynman-Kac representation)**.** *[26, Th. 7.6] Under the preceding conditions on b, σ and X, suppose that $v : [0,T] \times \mathbb{R}^d \to \mathbb{R}$ is continuous, $v \in C^{1,2}((0,T) \times \mathbb{R}^d)$ and satisfies the following Cauchy problem*

$$\begin{aligned} -\frac{\partial v}{\partial t} + kv &= \mathcal{A}_t v + g \quad \text{in } [0,T) \times \mathbb{R}^d, \\ v(T,x) &= f(x) \quad \forall x \in \mathbb{R}^d, \end{aligned}$$

as well as a polynomial growth condition

$$\max_{0 \leq t \leq T} |v(t,x)| \leq M(1 + \|x\|^{2\mu}) \quad \forall x \in \mathbb{R}^d, M > 0, \mu \geq 1.$$

Then v has the form

$$\begin{aligned} v(t,x) = \mathbb{E}^{t,x} \Bigg[& f(X_T) \exp\left(- \int_t^T k(u, X_u^{t,x}) du \right) \\ & + \int_t^T g(s, X_s^{t,x}) \exp\left(- \int_t^s k(u, X_u^{t,x}) du \right) ds \Bigg] \end{aligned}$$

on $[0,T] \times \mathbb{R}^d$, in particular v is unique.

Remark:

- Sufficient conditions for the existence of a classical solution to the Cauchy problem are:
  - (i) $\mathcal{A}_t$ is uniformly elliptic,
  - (ii) a_{ij}, b_i, k are bounded,

(iii) a_{ij}, b_i, k, g are Hölder continuous and

(iv) f, g are continuous and satisfy a polynomial growth condition

- Applied to our financial environment this means, that by departing from the risk neutral price process of a risky asset, which is a SDE $(X_t)_{t\leq T}$, with a risk-less interest rate r and a payoff function $f(x)$ we now know, by setting $g(x) = 0$ and $k = r$, that: The expected discounted payoff,

$$v(t,x) = \mathbb{E}^{t,x}\left(\exp\left(-\int_t^T r(u, X_u^{t,x})du\right) f(X_T^{t,x})\right),$$

is linked to the PDE

$$\frac{\partial}{\partial t}v + \mathcal{A}_t v - rv = 0$$

on $(0,T)\times\mathbb{R}^d$ with

$$v(T,x) = f(x).$$

Clearly this is a very powerful result. Whenever it is not possible to obtain a (semi) closed form solution to compute such expectations, the Feynman-Kac Theorem provides a way out via PDEs. Using well known finite difference schemes one can then numerically solve the PDE to get a solution. This is way more efficient[1] than applying standard Monte-Carlo techniques, which require a lot of computational power.

3.2.2 The HHW PDE

In the following lines we will derive the HHW PDE using the Feynman-Kac theorem from the previous section. After changing the order of the state variables to $X_t^* = (r_t, v_t, S_t)^T$ we have a symmetric correlation matrix

$$C = \begin{pmatrix} 1 & 0 & \rho_{Sr} \\ * & 1 & \rho_{Sv} \\ * & * & 1 \end{pmatrix}, \tag{3.2.2}$$

and its Cholesky decomposition $C = LL^T$, where L is a lower triangular matrix with strictly positive entries:

$$L = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \rho_{Sr} & \rho_{Sv} & \sqrt{1-\rho_{Sr}^2-\rho_{Sv}^2} \end{pmatrix}, \tag{3.2.3}$$

[1] As the dimension of the SDE increases, MC simulation becomes more efficient at some point, due to its independence. We are dealing here with three dimensions where PDE methods clearly outperform MC.

With the Cholesky decomposition, X_t^* can now be rewritten in terms of independent Brownian motions dZ_t^1, dZ_t^2, dZ_t^3. The system (2.3.1) reads now

$$\begin{pmatrix} dr_t \\ dv_t \\ dS_t/S_t \end{pmatrix} = \begin{pmatrix} a(b(t)-r_t) \\ \kappa(\eta - v_t) \\ r_t \end{pmatrix} dt + \begin{pmatrix} \sigma_2 & 0 & 0 \\ 0 & \sigma_1\sqrt{v_t} & 0 \\ L_{3,1}\sqrt{v_t} & L_{3,2}\sqrt{v_t} & L_{3,3}\sqrt{v_t} \end{pmatrix} \begin{pmatrix} dZ_t^1 \\ dZ_t^2 \\ dZ_t^3 \end{pmatrix}. \tag{3.2.4}$$

Squaring the diffusion matrix yields the symmetric instant covariance matrix:

$$\Sigma(X_t^*) = \begin{pmatrix} \sigma_2^2 & 0 & \rho_{Sr}\sigma_2\sqrt{v_t} \\ * & \sigma_1^2 v_t & \rho_{Sv}\sigma_1 v_t \\ * & * & v_t \end{pmatrix} \triangleq a(X_t^*).$$

Applying the Feynman-Kac Theorem (3) the operator from (3.2.1) in the HHW model is:

$$\begin{aligned} \mathcal{A}_t = \quad & a(b(t)-r_t)\frac{\partial}{\partial r} + \kappa(\eta - v_t)\frac{\partial}{\partial v} + r_t S_t \frac{\partial}{\partial S} \\ & + \frac{1}{2}\left[\sigma_2^2\frac{\partial^2}{\partial r^2} + \sigma_1^2 v_t \frac{\partial^2}{\partial v^2} + v_t S_t^2 \frac{\partial^2}{\partial S^2}\right. \\ & \qquad \left. + 2\rho_{Sr}\sigma_2\sqrt{v_t}S_t\frac{\partial^2}{\partial S \partial r} + 2\rho_{Sv}\sigma_1 v_t S_t \frac{\partial^2}{\partial S \partial v}\right]. \end{aligned}$$

Consequently, this yields the HHW PDE:

$$\begin{aligned} 0 = \quad & \frac{\partial u}{\partial t} + \frac{1}{2}s^2 v\frac{\partial^2 u}{\partial s^2} + \frac{1}{2}\sigma_1^2 v\frac{\partial^2 u}{\partial v^2} + \frac{1}{2}\sigma_2^2\frac{\partial^2 u}{\partial r^2} \\ & + \rho_{Sv}\sigma_1 sv\frac{\partial^2 u}{\partial s \partial v} + \rho_{Sr}\sigma_2 s\sqrt{v}\frac{\partial^2 u}{\partial s \partial r} \\ & + rs\frac{\partial u}{\partial s} + \kappa(\eta - v)\frac{\partial u}{\partial v} + a(b(t)-r)\frac{\partial u}{\partial r} - ru. \end{aligned} \tag{3.2.5}$$

The Feynman-Kac theorem tells us nothing about the existence of a solution, indeed we have to assume prior that there is one. Heath and Schweizer [19] provide an extension of Theorem 3, as a set of very general yet sufficient conditions under which both approaches are equivalent, without assuming existence of $u(t,x)$ beforehand.

3.2.3 Martingales vs PDEs: An equivalence result

To establish their equivalence result, Heath & Schweizer's [19] key assumptions are that the process X does not exit from a given domain $D \subseteq \mathbb{R}^d$ and that the drift and diffusion coefficient functions are smooth enough in

the interior of D. Therefore, typical degeneracies at the boundaries can be handled, e.g. for $v = 0$ in the Heston model. There is no general solution to this problem, but the following analytic and probabilistic assumptions are weak enough to be satisfied in some typical examples from finance. After establishing those assumptions and stating Theorem 4 it will also be shown how to verify them in practice.

Let $T \in (0, \infty)$ be the maturity and take an open and connected domain D in $\mathbb{R}^d$. Consider again the SDE

$$dX_s^{t,x} = b(s, X_s^{t,x})ds + \sum_{j=1}^{m} \sigma_j(s, X_s^{t,x})dW_s^j, \quad X_t^{t,x} = x \in D, \tag{3.2.6}$$

with $b : [0, T] \times D \to \mathbb{R}^d$ and $\sigma_j : [0, T] \times D \to \mathbb{R}^d, j = 1, \ldots, m$ continuous functions and $W = (W^1 \ldots W^m)^T$ a m-dimensional Brownian motion. Let $h : D \to [0, \infty), g : [0, T] \times D \to (-\infty, 0]$ and $c : [0, T] \times D \to \mathbb{R}$ be given measurable functions. Then define $u : [0, T] \times D \to [0, \infty]$ by

$$\begin{aligned} u(t, x) := \mathbb{E}_{t,x} \Bigg[& h(X_T^{t,x}) \exp\left(\int_t^T c(s, X_s^{t,x})ds \right) \\ & - \int_t^T g(s, X_s^{t,x}) \exp\left(\int_t^s c(u, X_u^{t,x})du \right) ds \Bigg]. \end{aligned} \tag{3.2.7}$$

The equation above is well-defined in $[0, \infty]$, if $X^{t,x}$ does not explode or leave D before time T. They then define for sufficiently smooth functions $u : [0, T] \times D \to \mathbb{R}$ an operator $\mathcal{L}$, which is very similar to the generator $\mathcal{A}_t$ in (3.2.1), by

$$(\mathcal{L}u)(t, x) := \sum_{i=1}^{d} b_i(t, x) \frac{\partial f}{\partial x_i}(t, x) + \frac{1}{2} \sum_{i,j=1}^{d} a_{ij}(t, x) \frac{\partial^2 f}{\partial x_i \partial x_j}(t, x) + c(t, x)u(t, x),$$

where $a = \sigma\sigma^T$ is again the diffusion matrix. The goal is now to give sufficient conditions on X, D, b, σ, h, c, g to make sure that u satisfies the PDE

$$\frac{\partial u}{\partial t} + \mathcal{L}u = g \quad \text{on } (0, T) \times D \tag{3.2.8}$$

with boundary conditions

$$u(T, x) = h(x) \text{ for } x \in D. \tag{3.2.9}$$

Theorem 4 (Heath and Schweizer)**.** *[19, Th. 1] Suppose that the following conditions hold:*

($A1$) *The coefficients b and σ_j, $j = 1, \ldots, m$ are on $[0,T]\times D$ locally Lipschitz-continuous in x, uniformly in t, i.e., for each compact subset F of D, there is a constant $K_F < \infty$ such that*

$$|G(t,x) - G(t,y)| \leq K_F|x-y|$$

for all $t \in [0,T], x, y \in F$ and $G \in \{b, \sigma_1, \ldots, \sigma_m\}$.

($A2$) *For all $(t,x) \in [0,T) \times D$, the solution $X^{t,x}$ of (3.2.6) neither explodes nor leaves D before T, i.e., $\mathbb{P}(sup_{t\leq s\leq T}|X_s^{t,x}| < \infty) = 1$ and $\mathbb{P}(X_s^{t,x} \in D$ for all $s \in [t,T]) = 1$.*

($A3$) *There exists a sequence $(D_n)_{n\in\mathbb{N}}$ of bounded domains contained in D such that $\cup_{n=1}^{\infty} D_n = D$ and for each n, the PDE*

$$\frac{\partial w}{\partial t} + \mathcal{L}w = g \quad \text{on } (0,T)\times D_n$$

with boundary condition $w(t,x) = h(t,x)$ on $(0,T)\times\partial D_n \cup \{T\}\times D_n$ has a classical solution $w_n(t,x)$.

Then u satisfies the PDE (3.2.8) with boundary condition (3.2.9). In particular, u is in $C^{1,2}$ and there exists a unique classical solution to this PDE.

Whereby ($A1$) can be easily verified - it is, for instance, satisfied if b and a are differentiable in x on the open set $(0,T)\times D$ with derivatives that are continuous on $[0,T]\times D$ - condition ($A2$) involves a more careful study of the process X. Finally, ($A3$) is easier to verify as the implication of the combination of:

(A3') there exists a sequence $(D_n)_{n\in\mathbb{N}}$ of bounded domains with $\overline{D}_n \subseteq D$ such that $\cup_{n=1}^{\infty} D_n = D$, each D_n has a C^2-boundary and for each n,

- ($A3a'$) the functions b and $a = \sigma\sigma^T$ are uniformly Lipschitz-continuous on $[0,T]\times\overline{D}_n$,
- ($A3b'$) $a(t,x)$ is uniformly elliptic on $\mathbb{R}^d$ for $(t,x) \in [0,T)\times D_n$, i.e., there is $\delta_n > 0$ such that $y^T a(t,x) y \geq \delta_n |y|^2$ for all $y \in \mathbb{R}^d$,
- ($A3c'$) c is uniformly Hölder-continuous on $[0,T]\times\overline{D}_n$,
- ($A3d'$) g is uniformly Hölder-continuous on $[0,T]\times\overline{D}_n$ and
- ($A3e'$) u is finite and continuous on $[0,T]\times\partial D_n \cup \{T\}\times\overline{D}_n$.

Note that the restrictive uniformity assumptions on a, b, g, c are imposed only for the local bounded domains D_n, and not globally for all $x \in D$. This is an essential difference compared to standard results from the literature, which allows to handle certain degeneracies on the boundary of D. Also one can often readily verify $(A3')$.

Continuity of u, at first, seems hard to verify. Due to $(A1)$ and $(A2)$ ensure that X is well-behaved in the interior of D, condition $(A3e')$ can be simplified by the following Lemma.

Lemma 1 (easier verification of $(A3e')$)**.** *[19, Lem. 2] Assume that $(A1)$ and $(A2)$ hold. If h, g and c are continuous, h and g are bounded and c is bounded from above, then u is continuous on $[0,T] \times D$.*

Also the uniform ellipticity condition $(A3b')$ on a can be verified easier via the next Lemma. A matrix a can't be uniformly elliptic unless its determinant is non zero.

Lemma 2 (easier verification of $(A3b')$)**.** *[19, Lem. 3] Assume that σ is continuous in (t,x). Fix any bounded domain $D' \subseteq D$ and let $\det a(t,x) \neq 0$ for all $(t,x) \in [0,T] \times \overline{D'}$. Then $a(t,x)$ is uniformly elliptic on $\mathbb{R}^d$ for $(t,x) \in [0,T) \times D'$.*

Finally, Theorem 4 with Lemmata 1 and 2 give a fairly nice set of assumptions under which the martingale and the PDE approaches are equivalent.

3.2.4 Equivalence of the Martingale and PDE approach for the HHW model

In this section Theorem 4 is applied to the HHW model to show equivalence between the martingale and PDE approaches. The price at time t of an European put option with maturity T and strike K is

$$u(t, S_t, v_t, r_t) = \mathbb{E}_{\mathbb{Q}} \left(\exp\left(- \int_t^T r_s ds \right) (K - S_T)^+ \right),$$

which is of the form as in (3.2.7) with $g \equiv 0, c = -r$ and $h(S_T) = (K - S_T)^+$. The reason why we argue here with the put option is the boundedness of its payoff function. Equivalence for the call payoff function is then a simple implication of the put-call parity, $C_t - P_t = S_t - KB(t,T)$. Now recall the system of SDEs (2.3.1) in the HHW model with coordinates $x := (r, v, S)$ and write the functions $b(t,x)$ and $\sigma(t,x)$ again as in (3.2.4):

$$b(t,x) = \begin{pmatrix} a(b(t) - r) \\ \kappa(\eta - v) \\ r \end{pmatrix},$$

$$\sigma(t,x) = \begin{pmatrix} \sigma_2 & 0 & 0 \\ 0 & \sigma_1\sqrt{v} & 0 \\ \rho_{Sr}\sqrt{v} & \rho_{Sv}\sqrt{v} & \sqrt{1-\rho_{Sr}^2-\rho_{Sv}^2}\sqrt{v} \end{pmatrix},$$

with $x \in D := \mathbb{R} \times (0,\infty)^2$. Now we check if the conditions from Theorem 4 hold: Condition $(A1)$ is clearly satisfied since b and σ are in $\mathcal{C}^1$ for x in $[0,T] \times D$. The Hull-White part does not explode and if the Feller condition $2\kappa\eta \geq \sigma_1^2$ is presumed neither v leaves $(0,\infty)$. Subsequently the same is true for $S = S_0 \exp(\int (r_s - \frac{1}{2}v_s)ds + \int \sqrt{v_s}dW_s)$, and thus $(A2)$ is satisfied. To verify $(A3)$ we take for D_n the cuboids $(-n,n) \times (1/n,n)^2$ with smoothed corners so $(A3')$ is satisfied. The coefficients b and σ are in $\mathcal{C}^1$, hence $(A3a')$ is obvious; so are $(A3c')$ and $(A3d')$. The matrix $a(t,x)$ has full rank and therefore Lemma 2 yields $(A3b')$. Finally, $(A3e')$ follows by Lemma 1.

3.3 Fourier methods

From Martingale pricing it is well known that option prices are the expected discounted payoff:

$$\mathbb{E}\left(e^{-\int_0^T r_s ds}\, w(\ln S_T)\right). \tag{3.3.1}$$

For a deterministic r, and if we write the expectation as its integral representation, we have

$$V(S_0) = \int_{\mathbb{R}} e^{-\int_0^T r_s ds}\, w(\ln S_T) p_T(S_T) dS_T,$$

which is the integral of the transition density $p_T(x)$ times the payoff function $w(\ln S_T)$. Already for simple models, those transition densities are usually very complicated and analytically hard to handle. Instead, it is much easier to compute the option value in Fourier space and inverse it. Many authors have worked with this Fourier technique, such as Lewis [30]. He proposes a simple option valuation formula for very general stock price processes; yet only considering a constant interest rate. We account for the essential feature of correlated stochastic interest rates in the HHW model by using the so called forward measure approach. This way, the discounting will be decoupled from formula (3.3.1).

In the next sections, we will describe the forward measure approach, introduce the Fourier technique used by Lewis [30] and finally propose our slightly adapted option valuation formula.

3.3.1 Forward measure approach

Following the lines of [37, Sec. 13.2], this section describes the arbitrage valuation of contingent claims under stochastic interest rates. Let's again denote by $B(t,T)$ the value of a ZCB at time t maturing at $T \leq T^*$ (with fixed horizon $T^* > 0$), which follows an Itô process

$$dB(t,T) = B(t,T)\left(r_t dt + b(t,T)dW_t^*\right), \tag{3.3.2}$$

under the martingale measure $\mathbb{Q}$ with $B(T,T) = 1$. The d-dimensional standard Brownian motion W^* is defined on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{Q})$, and r_t is the instantaneous and continuously compounded interest rate. This means, that the existence of an arbitrage free family of bond prices associated with a certain short term interest rate process[2] r is taken for granted. Further, we do also assume that the construction of this market, with primary tradeable securities being a family of bonds and a certain number of other assets (stocks), is already arbitrage free. Let again be

$$\pi_t(X) = B_t \mathbb{E}_{\mathbb{Q}}\left(X B_T^{-1} | \mathcal{F}_t\right), \quad \forall t \in [0,T],$$

the no-arbitrage price of an attainable contingent claim X, with savings account B_t as in Section 2.2.

Definition 10 (forward contract)**.** *[37, Def. 13.2.1] For fixed $0 \leq t \leq T \leq T^*$ and a time T contingent claim X, a forward contract written at t is represented by the time T contingent claim $G_T = X - F_X(t,T)$ which satisfies:*

(i) $F_X(t,T)$ is a $\mathcal{F}_t$-measurable random variable;

(ii) the arbitrage price at time t of G_T equals zero, i.e., $\pi_t(G_T) = 0$.

The random variable $F_X(t,T)$ is called the forward price of a contingent claim X at t which settles at T. In particular, X may be defined as a preassigned amount of the underlying asset to be delivered at settlement, i.e. one share of a stock $X = S_T$ or a ZCB $X = B(T,U)$ of maturity $U \geq T$.

To express the forward price of a claim X in terms of its arbitrage price $\pi_t(X)$ and $B(t,T)$ we observe, by simply using the above definition, that

$$\begin{aligned}\pi_t(G_T) &= B_t \mathbb{E}_{\mathbb{Q}}\left(G_T B_T^{-1} | \mathcal{F}_t\right) \\ &= B_t\left[\mathbb{E}_{\mathbb{Q}}\left(X B_T^{-1} | \mathcal{F}_t\right) - F_X(t,T)\mathbb{E}_{\mathbb{Q}}\left(B_T^{-1} | \mathcal{F}_t\right)\right] = 0,\end{aligned}$$

[2] In our case this is the Hull-White process.

and further

$$F_X(t,T) = \frac{\mathbb{E}_{\mathbb{Q}}\left(XB_T^{-1}|\mathcal{F}_t\right)}{\mathbb{E}_{\mathbb{Q}}\left(B_T^{-1}|\mathcal{F}_t\right)} = \frac{\pi_t(X)}{B(t,T)}.$$

Consequently, if the asset to be delivered is a stock the forward price equals

$$F_{S_T}(t,T) = \frac{S_t}{B(t,T)}, \quad \forall t \in [0,T].$$

Definition 11 (forward martingale measure)**.** *[37, Def. 13.2.2] An equivalent martingale measure $\mathbb{Q}_T$ on $(\Omega, \mathcal{F}_T)$ to $\mathbb{Q}$, with Radon-Nikodým derivative*

$$\frac{d\mathbb{Q}_T}{d\mathbb{Q}} = \frac{B_T^{-1}}{\mathbb{E}_{\mathbb{Q}}\left(B_T^{-1}\right)} = \frac{1}{B_T B(0,T)}, \quad \mathbb{Q}-a.s.,$$

is called forward martingale measure for the settlement date T.

Notice, if the Radon-Nikodým derivative is restricted to the σ-field $\mathcal{F}_t$, it satisfies

$$\eta_t = \frac{d\mathbb{Q}_T}{d\mathbb{Q}|\mathcal{F}_t} = \mathbb{E}_{\mathbb{Q}}\left(\frac{1}{B_T B(0,T)}|\mathcal{F}_t\right) = \frac{B(t,T)}{B_t B(0,T)}, \quad \forall t \in [0,T].$$

Furthermore, if the bond price is driven by 3.3.2, we have

$$\eta_t = \exp\left(\int_0^t b(u,T)\cdot dW_u^* - \frac{1}{2}\int_0^t |b(u,T)|^2 du\right),$$

and the process

$$W_t^T = W_t^* - \int_0^t b(u,T)du, \quad \forall t \in [0,T],$$

follows a standard Brownian motion under $\mathbb{Q}_T$. Now the forward price can be expressed in terms of the conditional expectation under the forward measure.

Lemma 3. *[37, Lem. 13.2.2] For an $\mathbb{Q}_T$-integrable attainable contingent claim X settling at T, the forward price at t for the time T equals*

$$F_X(t,T) = \mathbb{E}_{\mathbb{Q}_T}\left(X|\mathcal{F}_t\right), \quad \forall t \in [0,T].$$

Particularly, the forward price process $F_X(t,T)$, $t \in [0,T]$ follows a martingale under $\mathbb{Q}_T$.

Clearly, if the asset is again a stock, we have $F_S(t,T) = \mathbb{E}_{\mathbb{Q}_T}(S_T|\mathcal{F}_t)$. More generally, the relative price follows a local martingale under $\mathbb{Q}_T$, provided that the price of a bond maturing at T is taken as a numeraire. Finally, we can establish a version of the risk neutral valuation formula from Proposition 1 which fits to stochastic interest rates.

Lemma 4 (forward risk neutral valuation formula)**.** *[37, Lem. 13.2.3] The arbitrage price of an attainable contingent claim X settling at T is given by*

$$\pi_t(X) = B(t,T)\mathbb{E}_{\mathbb{Q}_T}\left(X|\mathcal{F}_t\right), \quad \forall t \in [0,T].$$

With this result we can use Fourier techniques to derive a simple option valuation formula. Let us first define the Fourier transform of an integrable function and see how it is linked to the characteristic function of a process.

3.3.2 Fourier transform and characteristic function

Definition 12 (Fourier transform and inversion)**.** *For $u \in \mathbb{R}$ the Fourier transform of a piecewise continuous function $f(x)$ is defined as*

$$\hat{f}(u) = \mathcal{F}[f(x)] = \int_{\mathbb{R}} e^{iux} f(x) dx,$$

where $f(x)$ has to be integrable, i.e. $\int_{\mathbb{R}} |f(x)| dx < \infty$. Conversely, given the transform $\hat{f}(u)$, the original function can be recovered using the Fourier inversion given through:

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-iux} \hat{f}(u) du.$$

Remark: We will later see that it is necessary to generalize this definition to the whole complex plane. For $z \in \mathbb{C}$, $\operatorname{Im} z \neq 0$ the function $\hat{f}(z)$ is called the generalized Fourier transform of f. The generalized Fourier transform then is inverted by integrating along a straight line parallel to the real axis within the strip of regularity

$$f(x) = \frac{1}{2\pi} \int_{iv_1-\infty}^{iv_1+\infty} e^{-izx} \hat{f}(z) dz, \quad z = u + iv \in \mathcal{S}_f,$$

with constant $\operatorname{Im} \hat{z} = v_1$ where $\hat{z} \in \mathcal{S}_f$ and $\mathcal{S}_f$ is the strip in which $\hat{f}$ is regular.

Definition 13 (characteristic function (ChF))**.** *For $c \in \mathbb{C}$ with $a < Im\, z < b$, the characteristic function of the process X_t is defined as*

$$\phi_t(z) = \mathbb{E}(\exp(izX_t)).$$

Now say $p_t(x)dx = \mathbb{P}(X_t \in \{dx\})$ is the transition density for a process to reach $X_t = x$ after time t. For $a < \operatorname{Im} z < b$, the ChF of the process X_t is

identical to the ChF of its transition density $p_t(x)$, which is the generalized Fourier transform of it:

$$\phi_t(z) = \mathcal{F}[p_t(x)] = \int_{\mathbb{R}} \exp(izx) p_t(x) dx, \quad a < \operatorname{Im} z < b. \tag{3.3.3}$$

Conversely the transition density can be restored using the inverse Fourier transform

$$p_t(x) = \mathcal{F}^{-1}[\phi_t(z)] = \frac{1}{2\pi} \int_{iv-\infty}^{iv+\infty} \exp(-izx) \phi_t(z) dz, \quad v \in (a, b).$$

The ChF has to be regular[3] in a strip parallel to the real axis to compute option prices.

Theorem 5 (Lukacs)**.** *[33, Th. 7.1.1] A ChF $\phi(z)$ that is regular in the neighborhood of $z = 0$, is also regular in a strip parallel to the real axis and can be represented as a Fourier integral in it. This strip is either the whole z-plane, or it has one or two boundary lines. The purely imaginary points on that boundary of the strip, if its not the whole z-plane, are singular points of $\phi(z)$.*

For a good stock market model the ChF $\phi_t(z)$ must exist at $z = 0$ and $z = -i$, hence regular in the strip $\mathcal{S}_X = \{z : a < \operatorname{Im} z < b\}$, where $a \leq -1$ and $b \geq 0$. The first one is trivial. For the second one, with $\tilde{S}_t = \tilde{S}_0 \exp\{X_t\}$ and $X_0 = 1$, known as martingale identity, one observes that $\phi_t(-i) = \mathbb{E}(\exp(-i^2 X_t)) = \mathbb{E}(\exp(X_t)) = 1$[4] for a martingale $\widetilde{S}_t$. We consider only good models in that sense.

3.3.3 Fourier transform of payoff functions

We have seen that ChFs typically have a z-plane strip in which they are regular. The key to Lewis' [30] approach is, that this is also true for typical payoff functions. Lets take, for example, a call option and set $x = \ln S_T$. The payoff function reads $w(x) = (e^x - K)^+$, with strike K, and so its generalized Fourier transform $\hat{w}(z) = \mathcal{F}[w(x)]$, $z \in \mathbb{C}$ is by a simple integration

$$\hat{w}(z) = \left(\frac{\exp((iz+1)x)}{iz+1} - K \frac{\exp(izx)}{iz} \right) \Bigg|_{x=\ln K}^{\infty}.$$

[3] A complex function of a complex variable is analytic in a region if it has a derivative at every point in that region. It is called single valued if it has the same value at every point. If it's both analytic and single valued in a region it is called regular.

[4] For processes with $\mathbb{E}(\exp X_t) < \infty$ this normalization can be achieved with a drift adjustment.

The upper limit $x = \infty$ does not exist for all $z \in \mathbb{C}$. Applying the restriction $\operatorname{Im} z > 1$, however, the transform $\hat{w}(z)$ is well defined and regular in that strip. Thus we have

$$\hat{w}(z) = -\frac{K^{1+iz}}{z^2 - iz}, \quad \operatorname{Im} z > 1, \tag{3.3.4}$$

for a call option. The same calculations can be done, e.g., for put options or various other standard payoff functions.

Remark: Now we see why it is important to admit a complex valued transform variable. Typical option payoffs have generalized Fourier transforms and exist in a strip, i.e., $z \in \mathcal{S}_w$, just like a ChF. Note also, that so far both regions of regularity (from the ChF and the payoff function) do not necessarily overlap. The next Section clarifies this necessity for deriving a option valuation formula.

Inverting the generalized Fourier transform of the payoff function, is again done by integrating along a straight line parallel to the real axis within the strip of regularity.[5] Fix $v = \operatorname{Im} z$, then

$$w(x) = \frac{1}{2\pi} \int_{iv-\infty}^{iv+\infty} e^{-izx} \hat{w}(z) dz, \quad z \in \mathcal{S}_w.$$

3.3.4 A simple option valuation formula

We have mentioned already in the beginning of this chapter that there is need for some adaptions towards Lewis [30] formula. The non correlated, even constant, interest rate he uses makes a lot of things easier. When having a correlation between S and r one can not simply move the discounting in (3.3.1) in front of the expectation. As discussed in Section 3.3.1, we perform a switch to the forward stock price given through

$$F_t = \frac{S_t}{B(t,T)}, \tag{3.3.5}$$

where $B(t,T)$ is the price of a ZCB at time t paying 1 at maturity T. With this switch from the spot measure $\mathbb{Q}$, to the T-forward measure $\mathbb{Q}^T$, we decouple the discounting from (3.3.1), and hence the value of the option becomes:

$$V(S_t) := \pi_t(w(F_T)) = B(t,T)\, \mathbb{E}_{\mathbb{Q}^T}(w(F_T)).$$

[5] Later we will vary the integration contour to derive different variations of option valuation formulas. By Cauchy's theorem this can be done if the contour extends to $\pm\infty$ as long as we remain within the strip of regularity.

Assume that under the forward pricing measure the forward stock price evolves as $F_T = F_0 \exp(X_T)$, where $\exp(X_T)$ is a martingale under $\mathbb{Q}^T$ with $\mathbb{E}(\exp(X_T)) = 1$. Now let $f_0 = \ln F_0$ and $f_T = \ln F_T$ and define the distribution function for the log forward price to reach a final value after time T as

$$Q_T(x, f_0) = \mathbb{P}(f_T \leq x | f_0) = \int_{-\infty}^{x} q_T(\xi, f_0) d\xi.$$

Remember from (3.3.3) that the ChF $\phi_T(z)$ of the process X_T is the generalized Fourier transform of the transition density $p_T(x)$. Its distribution is:

$$P_T(x) = \mathbb{P}(X_T \leq x) = \int_{-\infty}^{x} p_T(\xi) d\xi.$$

Now

$$\mathbb{P}(f_T \leq x | f_0) = \mathbb{P}(f_0 + X_T \leq x | f_0) = \mathbb{P}(X_T \leq x - f_0),$$

which means that $Q_T(x, f_0) = P_T(x - f_0)$ and by differentiating with respect to x, the densities are $q_T(x, f_0) = p_T(x - f_0)$. Taking the Fourier transform of $q_T(x, f_0)$ yields

$$\begin{aligned}
\hat{q}_T(z, f_0) &= \int_{-\infty}^{\infty} \exp(izx) q_T(x, f_0) dx = \\
&= \int_{-\infty}^{\infty} \exp(izx) p_T(x - f_0) dx = \\
&= \int_{-\infty}^{\infty} \exp(iz(x' + f_0)) p_T(x') dx' = \hat{p}_T(z) \exp(izf_0),
\end{aligned}$$

with $x' = x - f_0$. Recalling that $\phi_T(z) \equiv \hat{p}_T(z)$ we find, for z within the strip $\mathcal{S}_X$ with $a < \mathrm{Im} z < b$ and $a \leq -1, b \geq 0$, that $\phi_T(z) \exp(izf_0) = \hat{q}_T(z)$.

We have seen that both, the ChF $\phi_t(z)$ and the transform of a typical payoff function $\hat{w}(z)$ have their own strips of regularity. Denote by $\mathcal{S}^*$ the conjugate strip consisting of all $z = u - iv$, if $\mathcal{S}$ contains all $z = u + iv$.

Theorem 6 (option valuation)**.** *Let F_t be the forward price as in (3.3.5). Assume that $w(e^{f_T}) \geq 0$, with $f_T = \ln F_T$, is a European style payoff function which is Fourier integrable in a strip and bounded for $|f_T| < \infty$, with transform $\hat{w}(z)$, $z \in \mathcal{S}_w$. Further, let the dynamics of the model be given as $F_t = F_0 \exp(X_t)$, where $\exp(X_t)$ is a martingale. The process X_T has an analytic characteristic function $\phi_T(z)$, regular in the strip $\mathcal{S}_X = \{z = u + iv : v \in (a, b),\ a < -1$ and $b > 0\}$. Then, if we assume that $\mathcal{S}_V = \mathcal{S}_w \cap \mathcal{S}_X^*$ is not empty, the option value under the forward price is given through:*

$$V(F_0) = \frac{1}{2\pi} \int_{iv_1 - \infty}^{iv_1 + \infty} e^{-izf_0} \phi_{X_T}(-z) \hat{w}(z) dz, \tag{3.3.6}$$

with $z \in \mathcal{S}_V$ and a fixed $v_1 = Im\,\hat{z}$, $\hat{z} \in \mathcal{S}_V$. Consequently, the spot option value is

$$V(S_0) = B(0,T)\,V(F_0). \tag{3.3.7}$$

This formula provides a starting point with many variations.

Proof. The option value under the forward measure is given as the expectation of the payoff function under the forward price. Straightforward we obtain:

$$\begin{aligned}
V(F_0) &= \mathbb{E}(w(\ln F_T)) = \\
&= \int_{-\infty}^{\infty} w(f_T) q_T(f_T) df_T = \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi} \int_{iv-\infty}^{iv+\infty} e^{-izf_T} \hat{w}(z) dz q_T(f_T) df_T = \\
&= \frac{1}{2\pi} \int_{iv-\infty}^{iv+\infty} \underbrace{\int_{-\infty}^{\infty} e^{-izf_T} q_T(f_T) df_T}_{\hat{q}_T(-z) = \phi_{X_T}(-z)\exp(-izf_0)} \hat{w}(z) dz = \\
&= \frac{1}{2\pi} \int_{iv-\infty}^{iv+\infty} e^{-izf_0} \phi_{X_T}(-z) \hat{w}(z) dz,
\end{aligned}$$

as long as $z \in \mathcal{S}_V$. By simply switching back to the spot price we find $V(S_0) = B(0,T)V(F_0)$, which completes the proof. □

Remark: If we take the result in (3.3.6) and compare it with the integral representation of the expectation we see that,

$$\int_{-\infty}^{\infty} w(f_T) q_T(f_T) df_T = \frac{1}{2\pi} \int_{iv_1-\infty}^{iv_1+\infty} \hat{w}(z) \hat{q}_T(-z, f_0) dz, \tag{3.3.8}$$

can be interpreted as Parseval style identity as found in Lewis [30, Th. 3.1 and Lem. 3.3].

Corollary 1 (call option formula)**.** *Under the assumptions of Theorem (6), for a European call with payoff $w(S_T) = (S_T - K)^+$ the option value is given as:*

$$V(S_0) = -\frac{KB(0,T)}{2\pi} \int_{iv_1-\infty}^{iv_1+\infty} e^{-iz\ln(F_0/K)} \phi_{X_T}(-z) \frac{dz}{z^2 - iz}. \tag{3.3.9}$$

Proof. Recall the Fourier transform for the payoff function of a call from (3.3.4). Since it is regular for $\mathrm{Im}\,\hat{z} > 1$ with $\hat{z} = u + iv_1$, the strip $\mathcal{S}_V$ is non empty and simply plugging the transform into (3.3.6) and (3.3.7) yields the result. □

Contour variations

With help of the residue calculus from Appendix B (Theorem 16 and Corollary 3), variations of the derived formulas can be obtained. Since the integrand in (3.3.9) is regular throughout $\mathcal{S}_X^*$, except for simple poles at $z = 0$ and $z = i$, we can move the integration contour to, let's say, $v_2 \in (0, 1)$. By doing so we pick up the residue at $z = i$.

Now say $f(z)$ is the integrand in formula (3.3.9). Consequently, from the residue Theorem 16, we know that for sufficiently large $r > 0$ we have:

$$\int_{iv_1-r}^{iv_1+r} f(z)dz + \int_{iv_1+r}^{iv_2+r} f(z)dz + \int_{iv_2+r}^{iv_2-r} f(z)dz + \int_{iv_2-r}^{iv_1-r} f(z)dz = 2\pi i \operatorname{Res}_i f.$$

Since $f(z)$ is zero at ∞, the second and fourth integral vanish if we let $r \to \infty$, which yields

$$\int_{iv_1-\infty}^{iv_1+\infty} f(z)dz - \int_{iv_2-\infty}^{iv_2+\infty} f(z)dz = 2\pi i \operatorname{Res}_i f,$$

where we flipped the integration path in the second integral. According to Corollary 3, the residue at $z = i$, for our function f with simple zero in the denominator, is

$$\begin{aligned} \operatorname{Res}_i f &= -\frac{KB(t,T)}{2\pi} \frac{e^{-i^2 \ln(F_0/K)}\phi(-i)}{2i - i} \\ &= -\frac{B(t,T)F_0}{2\pi i}, \end{aligned}$$

and hence, after bringing the second integral to the right side, we obtain

$$V(S_0) = B(0,T)F_0 - \frac{KB(0,T)}{2\pi} \int_{iv_1-\infty}^{iv_1+\infty} e^{-izf_0}\phi_{X_T}(-z)\frac{dz}{z^2 - iz}, \quad v_2 \in (0,1).$$

Finally by choosing, the symmetrically between the poles lying $v_2 = 1/2$, the formula becomes

$$V(S_0) = B(0,T)\left[F_0 - \frac{1}{\pi}\sqrt{F_0 K}\int_0^\infty \operatorname{Re}\left\{e^{iu\ln(F_0/K)}\phi_{X_T}(u - \frac{i}{2})\right\}\frac{du}{u^2 + \frac{1}{4}}\right]. \tag{3.3.10}$$

Chapter 4

Autocallables in the HHW model with correlated stochastic interest rates

In the previous chapters we have discussed market models which suit the needs of an equity-interest-rate hybrid product, like the in Chapter 1 presented autocallables. Subsequently we have introduced pricing techniques, applicable to this challenging product and model combination. Further, despite the tricky yet desired price-interest-rate correlation, also a method to calibrate the parameters of the model to market data was derived in Section 3.3.4 from the previous Chapter. The current chapter shows how these concepts will be implemented within our framework. Firstly, an efficient MC approach, used only as a comparison to the following, much more efficient, PDE solution is presented. Secondly, we will show how to use Fourier techniques to calibrate the models' parameters to market data.

4.1 An efficient MC simulation scheme

Our scheme is based on a simple and efficient simulation scheme of the Heston stochastic volatility model from Andersen [3], combined with an exact way of sampling from the Hull-White model. We will use his QE (quadratic exponential) algorithm for the variance process and extend the formula for the price discretization with the stochastic interest rate component, which will be simulated based on the Hull-White decomposition.

4.1.1 Simulation scheme for the volatility process

For a mean reverting square-root diffusion process like in (2.1.4),

$$dv_t = \kappa(\theta - v_t)dt + \sigma_1\sqrt{v_t}dW_t,$$

we know the following analytical results as in [3] Proposition 1 and Corollary 1.

Proposition 3. *[3, Prop. 1] Take the cumulative distribution function of the non-central chi-square distribution with ν degrees of freedom and non-centrality parameter λ,*

$$F_{\chi'^2}(z;\nu,\lambda) = e^{-\lambda/2}\sum_{j=0}^{\infty}\frac{(\lambda/2)^j}{j!2^{\nu/2+j}\Gamma(\nu/2+j)}\int_0^z z^{\nu/2+j-1}e^{-x/2}dx,$$

and set

$$d = \frac{4\kappa\theta}{\sigma_1^2},$$
$$n(t,T) = \frac{4\kappa e^{-\kappa(T-t)}}{\epsilon^2(1-e^{-\kappa(T-t)})},\ T > t.$$

The distribution of $v(T)$, conditional on $v(t)$, is then,

$$I\!P(v(T) < x|v(t)) = F_{\chi'^2}\left(\frac{x\,n(t,T)}{e^{-\kappa(T-t)}}; d, v(t)n(t,T)\right).$$

Corollary 2. *[3, Cor. 1] For $T > t$, $v(T)$ has, conditional on $v(t)$, the first two moments*

$$\mathbb{E}(v(T)|v(t)) = \theta + (v(t)-\theta)e^{-\kappa(T-t)},$$
$$\mathrm{Var}(v(T)|v(t)) = \frac{v(t)\sigma_1^2e^{-\kappa(T-t)}}{\kappa}\left(1-e^{-\kappa(T-t)}\right) + \frac{\theta\sigma_1^2}{2\kappa}\left(1-e^{-\kappa(T-t)}\right)^2.$$

In Figure (4.1), the exact distribution of v is compared to a Gaussian and lognormal who were matched to the first two moments of v. Clearly, neither of them is a real good approximation for low values of the variance process. Unfortunately v has a strong affinity for the area around zero volatility and in many practical applications can even reach the origin, since often the Feller condition is not satisfied, i.e., $2\kappa\theta << \sigma_1^2$.

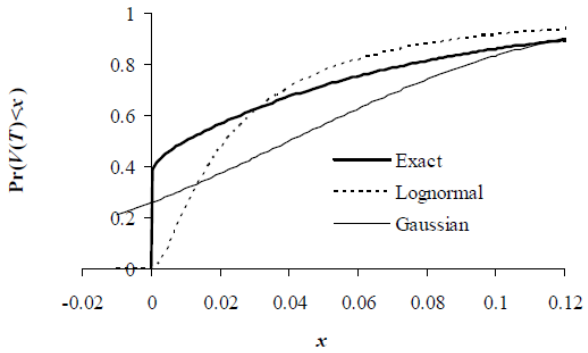

Figure 4.1: [3, Fig. 1] Example of exact distribution of v compared to a Gaussian and lognormal matched to the first two moments of v. The parameters are $T = 0.1, v(0) = \eta = 0.04, \kappa = 0.5$, and $\epsilon = 1$.

The non-central chi-square distribution approaches a Gaussian one for a non-centrality parameter approaching infinity. With a non-centrality parameter of $v(t)n(t, t+\Delta)$, Proposition 3 above yields that a Gaussian variable with its first two moments like in the last Corollary would be a good proxy for large $v(t)$. For small $v(t)$, the non-centrality vanishes and its distribution becomes proportional to that of a central chi-square distribution. Recalling the density of a central chi-square distribution with ν degrees of freedom,

$$f_{\chi^2}(x;\nu) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} e^{-x/2} x^{\nu/2-1}. \tag{4.1.1}$$

With the Feller condition not being satisfied the term $x^{\nu/2-1}$ implies a large density of $v(t+\Delta)$ around zero, for small $v(t)$. Approximating it with a Gaussian variable is clearly not accurate in this case.

The QE scheme is designed to address this weakness by dividing the problem into two schemes that work well for each ends. At first we concentrate on the high values of $v(t)$. Andersen [3] uses an observation that a non-central Chi-square distribution with moderate or high non-centrality parameter can be well represented by a power function applied to a Gaussian variable. For sufficiently large values of $\hat{v}(t)$ we have

$$\hat{v}(t+\Delta) \overset{d}{\approx} a(b + Z_v)^2, \tag{4.1.2}$$

with Z_v standard Gaussian, and a and b, who depend on $\hat{v}(t)$, determined by moment matching. We note that a cubic transformation of Z_v would be preferable. However, it allows negative values for the v-process, and therefore

the above quadratic representation is preferred. Unfortunately, the scheme does not work well for low values of $\hat{v}(t)$; it even fails to work for sufficiently low values (see Remark after Proposition 4). Now take a look at the asymptotic density of 4.1.1 if one lets $x \to 0$. Inspired by this, Andersen [3] uses an approximated density for $\hat{v}(t+\Delta)$, conditional on small $\hat{v}(t)$, of the form

$$\mathbb{P}\left(\hat{v}(t+\Delta) \in [x, x+dx] \,|\hat{v}(t)\right) \overset{d}{\approx} \left(p\delta_0 + \beta(1-p)e^{-\beta x}\right) dx, \quad x \geq 0, \quad (4.1.3)$$

where δ is the Dirac delta. With $p \in [0,1]$ and $\beta \geq 0$ one can easily verify that this is a valid density function. The strength of the probability mass at the origin is defined by p. This mass has an exponential tail similar to that of the central chi-square density. To sample from (4.1.3) one simply needs to integrate this density to get its cumulative distribution function and calculate the inverse.

$$\Psi(x) = \mathbb{P}\left(\hat{v}(t+\Delta) \leq x | \hat{v}(t)\right) = p + (1-p)\left(1 - e^{-\beta x}\right), x \geq 0,$$

with its inverse

$$\Psi^{-1}(u) = \Psi^{-1}(u; p, \beta) = \begin{cases} 0, & 0 \leq u \leq p, \\ \beta^{-1} \ln\left(\frac{1-p}{1-u}\right), & p < u \leq 1. \end{cases}$$

By applying the inverse distribution function method we get the sampling scheme

$$\hat{v}(t+\Delta) = \Psi^{-1}(U_v; p, \beta) \quad (4.1.4)$$

where U_v is uniformly distributed and p and β again depend on $\hat{v}(t)$. This scheme is simple and fast to execute. Figure (4.2) shows the quality of the two estimates, (4.1.2) for high and (4.1.4) low values respectively. The only thing left to do now is determine the constants a, b, p, β, which is done by moment matching techniques, and a rule for when to switch between both schemes.

Computing a and b:

Proposition 4. *Set*

$$\begin{aligned} m &= \theta + (\hat{v}(t) - \theta)e^{-\kappa\Delta}, \\ s^2 &= \frac{\hat{v}(t)\sigma_1^2 e^{-\kappa\Delta}}{\kappa}\left(1 - e^{-\kappa\Delta}\right) + \frac{\theta\sigma_1^2}{2\kappa}\left(1 - e^{-\kappa\Delta}\right)^2, \\ \psi &= \frac{s^2}{m^2}. \end{aligned}$$

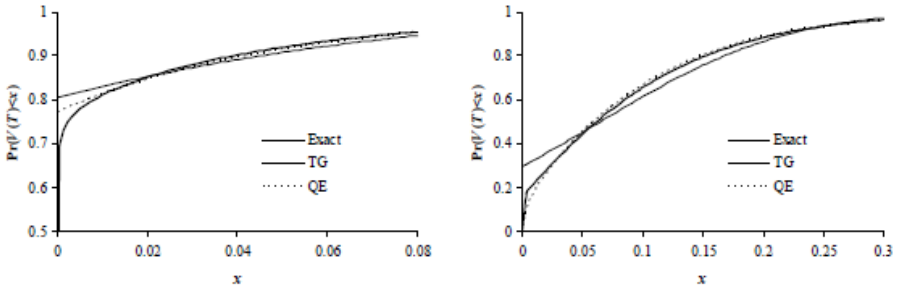

Figure 4.2: [3, Fig. 4] Quality of estimators - the parameters are as in figure (4.1), except for $v(0) = 0.01$ in the left and 0.09 in the right picture.

Let $\hat{v}(t+\Delta)$ be as in (4.1.2). With Corollary 2, $\mathbb{E}(\hat{v}(t+\Delta)|\hat{v}(t)) = m$ and $\mathrm{Var}(\hat{v}(t+\Delta)|\hat{v}(t)) = s^2$. Then

$$b^2 = 2\psi^{-1} - 1 + \sqrt{2\psi^{-1}}\sqrt{2\psi^{-1} - 1} \geq 0,$$
$$a = \frac{m}{1+b^2},$$

for $\psi \leq 2$.

Proof. For a proof see [3, Propositions 4, 5 and Corollary 1]. □

Remark: It can easily be seen that this only works for $\psi \leq 2$. With higher values of ψ, (meaning low values of $\hat{V}(t)$), this scheme is failing.

Computing p and β:

Proposition 5. *Set m, s, ψ as in proposition (4) and assuming $\psi \geq 1$. Then*

$$p = \frac{\psi - 1}{\psi + 1} \in [0, 1),$$
$$\beta = \frac{1-p}{m} = \frac{2}{m(\psi+1)} > 0.$$

For $\hat{v}(t+\Delta)$ as in (4.1.4), then again $\mathbb{E}(\hat{v}(t+\Delta)|\hat{v}(t)) = m$ and $\mathrm{Var}(\hat{v}(t+\Delta)|\hat{v}(t)) = s^2$ as in Proposition 4 before.

Proof. For a proof see [3, Propositions 4, 6 and Corollary 1]. □

Switching rule:

The quadratic and exponential sampling schemes can only be moment matched for $\psi \leq 2$ respectively $\psi \geq 1$. Luckily these domains overlap, meaning at least one of both can always be used. The most natural approach now is to introduce some level $\psi_c \in [1, 2]$ where both domains overlap. Since m and s^2 depend on $v(t)$, this level shows us when to use which scheme at every time step.

QE-algorithm:

The following algorithm summarizes how to perform one simulation step.

1: Choose an arbitrary level $\psi_c \in [1, 2]$ for the switching rule.
2: Given $\hat{v}(t)$, compute m, s^2 and ψ from Proposition 4.
3: Draw a standard normal random variable Z_v and compute its corresponding uniform $U_v = \Phi(Z_v)$[1].
4: **if** $\psi \leq \psi_c$ **then**
5: Compute a and b from Proposition 4.
6: Compute $\hat{v}(t + \Delta)$ with (4.1.2) and Z_v.
7: **else**
8: Compute β and p from Proposition 5.
9: Compute $\hat{v}(t + \Delta)$ with (4.1.4) and U_v.
10: **end if**

Note that the computations in this algorithm involve a lot of time independent calculations. To speed up the procedure they should be pre-cached before running a loop over all time steps. Also, due to the fact that the exact choice of ψ_c has very little effects on the quality of the scheme, it is set to $\psi_c = 1.5$, as in [3, Ch. 3.2.3].

4.1.2 Simulation scheme for the short rate process

The Hull-White interest rate process,

$$dr_t = a(b(t) - r_t)dt + \sigma_2 dW_t,$$

can be simulated very easily and efficiently. In order to do this we will first decompose it as in Proposition 2, into

$$r_t = \tilde{r}_t + \psi_t, \tag{4.1.5}$$

[1] Φ is the cumulative standard normal distribution function (CDF), which simply transforms a $N(0, 1)$ random variable into a uniformly $U(0, 1)$ one.

with

$$d\tilde{r}_t = -a\tilde{r}_t dt + \sigma dW_t, \qquad \text{with } \tilde{r}_0 = 0, \tag{4.1.6}$$

and

$$\psi_t = r_0 e^{-at} + a\int_0^t e^{-a(t-s)} b(s) ds. \tag{4.1.7}$$

For the remaining stochastic Ornstein-Uhlenbeck part (4.1.6) the following is very well known:

Theorem 7. *[39, Th. 1] The Ornstein-Uhlenbeck stochastic process as in (4.1.6) is normally distributed with expectation and variance*

$$\mathbb{E}\left(\tilde{r}(t)|\tilde{r}(s)\right) = \tilde{r}(s)e^{-a(t-s)},$$

$$\mathrm{Var}\left(\tilde{r}(t)|\tilde{r}(s)\right) = \frac{\sigma_2^2}{2a}\left(1 - e^{-2a(t-s)}\right).$$

Hence, it can be simulated exactly. The function $b(t)$ is supposed to reflect the market's term structure, which is usually given as a discount curve $P(0,T_i)$ with maturities T_i. One can easily calculate the spot rates $b(t)$ from the discount curve via the formula

$$b(T) = f(0,T) + \frac{1}{a}\frac{\partial}{\partial T} f(0,T) + \frac{\sigma_2^2}{2a^2}\left(1 - e^{-2aT}\right), \tag{4.1.8}$$

as in Musiela [37, p. 294], where $f(0,T)$ is the instantaneous forward rate, which is defined as

$$f(0,T) = -\frac{\partial \ln P(0,T)}{\partial T}.$$

We will approximate the forward rates $f(0,T)$ with the left sided difference quotient, whereas the derivative $\frac{\partial f(0,T_i)}{\partial T_i}$ is approximated with the central difference quotient, which gives a smoother result. For the spot rates $b(t)$ with $t \in (T_i, T_{i+1})$ a cubic interpolation should be sufficient.

Finally, the Hull-White decomposition (4.1.5) combined with Theorem 7 and equation (4.1.8) give a fairly handy set of formulas to efficiently simulate the interest rate process, as described by the following algorithm:

Exact simulation of r_t:

1: Calculate the spot rates from the given market discount curve with formula (4.1.8).
2: Draw a standard normal random variable Z_r.
3: Compute $\tilde{r}(s+\Delta)$ from $\tilde{r}(s)$ with expectation and variance as given in Theorem 7.
4: Compute the deterministic part ψ_t with formula (4.1.7).
5: Compute $r(t) = \tilde{r}(t) + \psi_t$.

4.1.3 Simulation scheme for the price process

Recall from Section 2.3.1 that we can write the three correlated Brownian motions W_t^S, W_t^v, W_t^r in terms of independent ones with the Cholesky decomposition of the correlation matrix as

$$\begin{pmatrix} W_t^r \\ W_t^v \\ W_t^S \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \rho_{Sr} & \rho_{Sv} & \sqrt{1-\rho_{Sv}^2-\rho_{Sr}^2} \end{pmatrix} \begin{pmatrix} Z_t^1 \\ Z_t^2 \\ Z_t^3 \end{pmatrix}.$$

Subsequently, we will propose a discretization scheme as in Andersen [3, Ch. 4.2] joined by the interest rate dynamics. After performing a log transformation ($x_t = \ln S_t$) to the price process, using Itô's formula, and plugging in the correlation structure from before, we obtain for the log price process the following exact representation:

$$\begin{aligned} \mathrm{d}x(t) &= (r(t) - \frac{1}{2}v(t))dt + \sqrt{v(t)}dW_t^S \\ &= (r(t) - \frac{1}{2}v(t))dt + \sqrt{v(t)}\left[\rho_{Sr}dZ_1(t) + \rho_{Sv}dZ_2(t) + \sqrt{1-\rho_{Sv}^2-\rho_{Sr}^2}Z_3(t)\right] \end{aligned}$$

Integrating the independent HHW SDEs yields:

$$\begin{aligned} v(t+\Delta) &= v(t) + \int_t^{t+\Delta} \kappa(\theta - v(u))du + \sigma_1 \int_t^{t+\Delta} \sqrt{v(u)}dW_u^v \\ r(t+\Delta) &= r(t) + \int_t^{t+\Delta} a(b(T-u) - r(u))du + \sigma_2 \int_t^{t+\Delta} dW_u^r \\ x(t+\Delta) &= x(t) + \int_t^{t+\Delta} (r(u) - \frac{1}{2}v(u))du \\ &\quad + \rho_{Sr} \int_t^{t+\Delta} \sqrt{v(u)}dW_u^r + \rho_{Sv} \int_t^{t+\Delta} \sqrt{v(u)}dW_u^v \\ &\quad + \sqrt{1-\rho_{Sv}^2-\rho_{Sr}^2} \int_t^{t+\Delta} \sqrt{v(u)}dZ_3(u) \end{aligned}$$

Isolating $\int_t^{t+\Delta} \sqrt{v(u)} dW_v(u)$ from the first equation and substituting it into the third we find

$$\begin{aligned}
x(t+\Delta) &= x(t) + \int_t^{t+\Delta} (r(u) - \frac{1}{2} v(u)) du \\
&\quad + \rho_{Sv} \frac{1}{\sigma_1} \left(v(t+\Delta) - v(t) - \kappa\theta\Delta + \kappa \int_t^{t+\Delta} v(u)) du \right) \\
&\quad + \rho_{Sr} \int_t^{t+\Delta} \sqrt{v(u)} dW_u^r + \sqrt{1 - \rho_{Sv}^2 - \rho_{Sr}^2} \int_t^{t+\Delta} \sqrt{v(u)} dZ_u^3 \\
&= x(t) + \frac{\rho_{Sv}}{\sigma_1} \left(v(t+\Delta) - v(t) - \kappa\theta\Delta \right) \\
&\quad + \int_t^{t+\Delta} r(u) du + \left(\frac{\kappa\rho_{Sv}}{\sigma_1} - \frac{1}{2} \right) \int_t^{t+\Delta} v(u) du \\
&\quad + \rho_{Sr} \int_t^{t+\Delta} \sqrt{v(u)} dW_u^r + \sqrt{1 - \rho_{Sv}^2 - \rho_{Sr}^2} \int_t^{t+\Delta} \sqrt{v(u)} dZ_u^3.
\end{aligned}$$

The time integrals for $v(u)$ and $r(u)$ are approximated by the simple quadrature

$$\int_t^{t+\Delta} v(u) du \approx \Delta \left[\gamma_1 v(t) + \gamma_2 v(t+\Delta) \right]$$

for constants γ_1 and γ_2. Since W^r is independent of v, conditional on $v(t)$ and $\int_t^{t+\Delta} v(u) du$, the two Itô integrals are Gaussian with mean zero and variance $\int_t^{t+\Delta} v(u) du$, hence they become

$$\int_t^{t+\Delta} \sqrt{v(u)} dW_r(u) \overset{d}{\approx} \sqrt{\Delta} \sqrt{\gamma_1 v(t) + \gamma_2 v(t+\Delta)} Z_r,$$

where Z_r is a standard normal random variable independent of v. These approximations yield

$$\begin{aligned}
\hat{x}(t+\Delta) &= \hat{x}(t) + \frac{\rho_{Sv}}{\sigma_1} \left(v(t+\Delta) - v(t) - \kappa\theta\Delta \right) + \Delta \left(\gamma_1 r(t) + \gamma_2 r(t+\Delta) \right) \\
&\quad + \Delta \left(\frac{\kappa\rho_{Sv}}{\sigma_1} - \frac{1}{2} \right) \left(\gamma_1 v(t) + \gamma_2 v(t+\Delta) \right) \\
&\quad + \sqrt{\Delta} \rho_{Sr} \sqrt{\gamma_1 v(t) + \gamma_2 v(t+\Delta)} Z_1 \\
&\quad + \sqrt{\Delta} \sqrt{1 - \rho_{Sv}^2 - \rho_{Sr}^2} \sqrt{\gamma_1 v(t) + \gamma_2 v(t+\Delta)} Z_S.
\end{aligned}$$

Finally, we use the central approximation $\gamma_1 = \gamma_2 = \frac{1}{2}$ for the integrals and replace $v(t)$ and $r(t)$ by their approximations $\hat{v}(t)$ and $\hat{r}(t)$. After rearranging

and collecting terms we get the following prefered expression for the log price

$$\begin{aligned}\hat{x}(t+\Delta) = \hat{x}(t) + K_0 + K_1\left(\hat{r}(t) + \hat{r}(t+\Delta)\right) + K_2\hat{v}(t) + K_3\hat{v}(t+\Delta) \\ + K_4\sqrt{\hat{v}(t) + \hat{v}(t+\Delta)}Z_r + K_5\sqrt{\hat{v}(t) + \hat{v}(t+\Delta)}Z_S,\end{aligned} \tag{4.1.9}$$

with

$$K_0 = -\frac{\rho_{Sv}\kappa\theta\Delta}{\sigma_1}, \quad K_2 = \left(\frac{1}{2}\Delta\left(\frac{\kappa\rho_{Sv}}{\sigma_1} - \frac{1}{2}\right) - \frac{\rho_{Sv}}{\sigma_1}\right), \quad K_4 = \sqrt{\frac{1}{2}\Delta}\rho_{Sr},$$

$$K_1 = \frac{1}{2}\Delta, \qquad K_3 = \left(\frac{1}{2}\Delta\left(\frac{\kappa\rho_{Sv}}{\sigma_1} - \frac{1}{2}\right) + \frac{\rho_{Sv}}{\sigma_1}\right), \quad K_5 = \sqrt{\frac{1}{2}\Delta}\sqrt{1-\rho_{Sv}^2-\rho_{Sr}^2}.$$

Note that the K_i depend also on the time step. By combining the above scheme for the log price with the two schemes for the spot rate and volatility we have a efficient way of sampling from the HHW process.

Algorithm for the log price x_t:

1: Given $r(t), v(t)$ and $x(t)$,
2: Compute $r(t+\Delta)$ from $r(t)$.
3: Compute $v(t+\Delta)$ from $v(t)$.
4: Compute $x(t+\Delta)$ from equation (4.1.9).

4.1.4 Convergence and consistency considerations

To analyse the convergence of the x process is complicated due to the fact that high order moments may not exist. Andersen [2, Prop. 3.1] shows for which maturity and model parameters the k-th moment $\mathbb{E}(x_T^k)$, for $k > 1$, is finite. Such analysis is for example performed in Lord [32], yet other authors turn to a simpler concept, namely the one of weak consistency. There is a strong connection between the two, as shown in [28, p. 328]. However, for a proof on the weak consistency we refer the interested reader to [3, Prop. 10].

4.2 An ADI finite difference approach for the HHW PDE

We have derived the HHW PDE already in Section 3.2.2, which is given as

$$\begin{aligned}\frac{\partial u}{\partial t} = &\frac{1}{2}s^2 v\frac{\partial^2 u}{\partial s^2} + \frac{1}{2}\sigma_1^2 v\frac{\partial^2 u}{\partial v^2} + \frac{1}{2}\sigma_2^2\frac{\partial^2 u}{\partial r^2} \\ &+ \rho_{Sv}\sigma_1 sv\frac{\partial^2 u}{\partial s\partial v} + \rho_{Sr}\sigma_2 s\sqrt{v}\frac{\partial^2 u}{\partial s\partial r} \\ &+ rs\frac{\partial u}{\partial s} + \kappa(\eta - v)\frac{\partial u}{\partial v} + a(b(T-t) - r)\frac{\partial u}{\partial r} - ru,\end{aligned} \tag{4.2.1}$$

with $u = u(s, v, r, T-t), s > 0, v > 0, -\infty < r < \infty$ and $0 < t \leq T$. The HHW PDE is a time dependent convection diffusion equation on a three-dimensional spatial unbounded domain. Initial and boundary conditions complete the PDE. Those are given by the specific option under consideration. If the volatility $v \to 0$ we see that all the terms in front of derivatives of the second order, except the one from $\partial^2 u/\partial r^2$, disappear. The correlations between the three Brownian motions yield the mixed spatial derivative terms. Furthermore, we notice that the coefficient of $\partial u/\partial r$ depends on the time.

Solutions in (semi) closed form only exist for plain vanilla style options. For the numerical solution we therefore consider the following approach based on finite differences. First the PDE is discretized in the spatial variables, (s, v, r), leading a system of stiff ODEs (ordinary differential equations). The resulting semi-discrete system is then solved by application of a suitable time discretization method. Three dimensions yield a very large semi-discrete system, with a high bandwidth also. Simple (and pure) explicit or implicit time discretization schemes are not suitable to tackle those challenges. Therefore, splitting schemes of the ADI (alternate direction implicit) type have to be used. Those are known to perform very well for PDEs of the above type.

In the following subsection we state the specific initial and boundary conditions for our option. The finite difference scheme and its spatial discretization on a non uniform grid will be presented subsequently, followed by the chosen ADI scheme, which will be explained in the corresponding Section 4.2.4.

4.2.1 Initial and boundary conditions

For the introduced product with payoff function as in Figure 1.1, the initial condition is given as

$$u(s,v,r,T) = \begin{cases} P_4, & s \geq K_4, \\ s, & s < K_4, \end{cases} \tag{4.2.2}$$

where the last observation date $T_4 = T$. The payoffs at the observation dates T_i for $i \in \{1,2,3\}$ serve as boundary conditions in the space dimensions. Before stating them we first restrict the spatial domain to a bounded set $[0, S_{max}] \times [0, V_{max}] \times [-R_{max}, R_{max}]$ with $S_{max}, V_{max}, R_{max}$ sufficiently large. Hence, we have
for $s = 0$

$$u(s,v,r,t) = 0, \tag{4.2.3}$$

for $s = S_{max}$

$$u(s,v,r,t) = \begin{cases} B(r,t,T_1)P_1, & t \leq T_1 \\ B(r,t,T_2)P_2, & T_1 < t \leq T_2 \\ B(r,t,T_3)P_3, & T_2 < t \leq T_3 \\ B(r,t,T_4)P_4, & T_3 < t \leq T_4, \end{cases} \tag{4.2.4}$$

for $v = V_{max}$

$$\begin{aligned} &\frac{\partial u}{\partial v}(s,v,r,t) = 0, \\ &\frac{\partial^2 u}{\partial v^2}(s,v,r,t) = 0, \\ &\frac{\partial^2 u}{\partial s \partial v}(s,v,r,t) = \frac{\partial^2 u}{\partial s \partial r}(s,v,r,t) = \frac{\partial^2 u}{\partial v \partial r}(s,v,r,t) = 0, \end{aligned} \tag{4.2.5}$$

for $r = \pm R_{max}$

$$\frac{\partial u}{\partial r}(s,v,r,t) = 0, \tag{4.2.6}$$

and finally for $t \in \{T_1, T_2, T_3\}$

$$u(s,v,r,t) = \begin{cases} P_1, & t = T_1 \text{ and } s \geq K_1 \\ P_2, & t = T_2 \text{ and } s \geq K_2 \\ P_3, & t = T_3 \text{ and } s \geq K_3. \end{cases} \tag{4.2.7}$$

The function $B(r,t,T_i)$ again denotes the value of a ZCB, given that the short rate at t equals r. Concerning the boundary condition at $r = \pm R_{max}$, it is

easy to show that the 'rho' of an European call option (sensitivity of option price for small changes in the spot interest rate) vanishes for extreme values of the spot interest rate in the Black Scholes model, and it is therefore plausible that this holds for our model (2.3.1) as well. Further, at $v = V_{max}$, the first order derivative has been set to zero, as done by Haentjens and 't Hout [18] for up-and-out call options. All second order derivatives at this boundary were also set to zero for convenience reasons. The other, Dirichlet boundary conditions are given through the structure of the autocallable. Finally, the boundary at $v = 0$ is simply treated by inserting $v = 0$ into the PDE (4.2.1).

4.2.2 Generating the spatial grid

For the spatial dimensions we use a nonuniform grid like in 'tHout & Haentjens [18] or 'tHout & Foulon [23]. The goal is, of course, to have relatively many grid points near the strikes K_i in the S-direction. This is done so as to reduce numerical difficulties, where the payoff function w has discontinuous derivatives, and because these are the regions of interest. Applying a non uniform grid therefore improves the accuracy of the discretization over a uniform one. We strictly follow the steps in [18, Ch. 2.2] except for splitting the domain $[0, S_{\max}]$ into four parts for every strike K_i.

Let m_1 be the number of grid points in the S-direction. Since only $\frac{m_1}{4}$ points are created for every part of the domain we firstly need to take care that $m_1 \mod 4 = 0$. The split subsets of the domain are given as $[B_1,\ B_2] \in \left\{\left[0,\ \frac{K_1+K_2}{2}\right], \left[\frac{K_1+K_2}{2},\ \frac{K_2+K_3}{2}\right], \left[\frac{K_2+K_3}{2},\ \frac{K_3+K_4}{2}\right], \left[\frac{K_3+K_4}{2},\ S_{\max}\right]\right\}$. Now let, for every one of them, be $d_1 > 0$ and equidistant points $\xi_0 < \cdots < \xi_{m_1/4}$ given by

$$\xi_i = \sinh^{-1}(-K_i/d_1) + i\Delta\xi, \quad (0 \le i \le \frac{m_1}{4}),$$

where

$$\Delta\xi = \frac{4}{m_1}\left[\sinh^{-1}((B_2 - B_1 - (K_i - B_1))/d_1) - \sinh^{-1}(-(K_i - B_1)/d_1)\right].$$

Transforming the equidistant points ξ_i to

$$s_i = K_i + d_1\ \sinh(\xi_i), \quad (0 \le i \le \frac{m_1}{4}),$$

gives a non uniform grid for each subset. Merging them then yields the grid for the whole S-domain $0 = s_0 < s_1 < \cdots < s_{m_1} = S_{\max}$ which concentrates around the four barriers. The parameter d_1 controls the fraction of points that lie 'close' to these barriers. We have chosen $d_1 = S_0/100$.

In the v- and r-direction we also define non uniform grids according to Haentjens and 't Hout [18] which concentrate around 0 and the initial value

of the short rate process R_0 respectively. Let again be m_2 and m_3 the number of grid points in the spatial domains, $d_2, d_3 > 0$ and equidistant points $\eta_0 < \cdots < \eta_{m_2}$ and $\zeta_0 < \cdots < \zeta_{m_3}$ be given through

$$\eta_j = j\,\Delta\eta \quad (0 \le j \le m_2),$$
$$\Delta\eta = \frac{1}{m_2}\sinh^{-1}(V_{\max}/d_2),$$

and

$$\zeta_k = \sinh^{-1}((-R_{\max} - R_0)/d_3) + k\dot{\Delta}\zeta \quad (0 \le k \le m_3),$$
$$\Delta\zeta = \frac{1}{m_3}\left[\sinh^{-1}((R_{\max} - R_0)/d_3) - \sinh^{-1}((-R_{\max} - R_0)/d_3)\right].$$

Then the grid points $0 = v_0 < v_1 < \cdots < v_{m_2} = V_{\max}$, respectively $-R_{\max} = r_0 < r_1 < \cdots < r_{m_3} = R_{\max}$ are given by

$$v_j = d_2\,\sinh(\eta_j) \quad (0 \le j \le m_2)$$

and

$$r_k = R_0 + d_3\,\sinh(\zeta_k) \quad (0 \le k \le m_3).$$

Like before the parameters d_2, d_3 control the fraction of points in the neighborhood of $v = 0$ and $r = R_0$.

To neglect the error made by restricting the spatial domain we set our boundaries far enough from the regions of interest. In our implementation we set $S_{\max} = 10S_0, V_{\max} = 5$ and $R_{\max} = 1$. Figure 4.3 shows an example of grid points in all three spatial dimensions.

4.2.3 Finite difference discretization

Given an increasing sequence of grid points $\{x_i\}_{i\in\mathbb{Z}}$ and $\Delta x_i = x_i - x_{i-1}$ for all i, first and second order derivatives for any given function $f : \mathbb{R} \to \mathbb{R}$ are approximated by the following finite difference formulas:

$$f'(x_i) \approx \alpha_{-2}f(x_{i-2}) + \alpha_{-1}f(x_{i-1}) + \alpha_0 f(x_0), \tag{4.2.8}$$
$$f'(x_i) \approx \beta_{-1}f(x_{i-1}) + \beta_0 f(x_i) + \beta_1 f(x_{i+1}), \tag{4.2.9}$$
$$f'(x_i) \approx \gamma_0 f(x_0) + \gamma_1 f(x_{i+1}) + \gamma_2 f(x_{i+2}), \tag{4.2.10}$$
$$f''(x_i) \approx \delta_{-1}f(x_{i-1}) + \delta_0 f(x_i) + \delta_1 f(x_{i+1}). \tag{4.2.11}$$

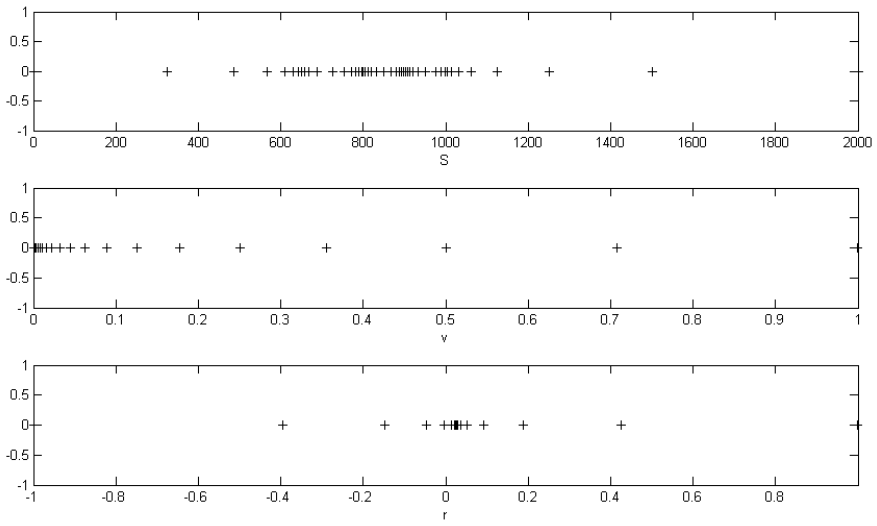

Figure 4.3: Example of the distribution of grid points - they concentrate around the strike prices K_i, zero volatility and the initial short rate r_0 respectively.

Approximations (4.2.8) and (4.2.10) are backward and forward schemes, whereas (4.2.9) and (4.2.11) are central first and second order schemes respectively. For the weights $\alpha_{\{-2,-1,0\}}, \beta_{\{-1,0,1\}}, \gamma_{\{0,1,2\}}, \delta_{\{-1,0,1\}}$ we have

$$\alpha_{-2} = \frac{\Delta x_i}{\Delta x_{i-1}(\Delta x_{i-1} + \Delta x_i)}, \quad \alpha_{-1} = \frac{-\Delta x_{i-1} - \Delta x_i}{\Delta x_{i-1}\Delta x_i}, \quad \alpha_0 = \frac{\Delta x_{i-1} + 2\Delta x_i}{\Delta x_i(\Delta x_{i-1} + \Delta x_i)},$$
$$\beta_{-1} = \frac{-\Delta x_{i+1}}{\Delta x_i(\Delta x_i + \Delta x_{i+1})}, \quad \beta_0 = \frac{\Delta x_{i+1} - \Delta x_i}{\Delta x_i \Delta x_{i+1}}, \quad \beta_1 = \frac{\Delta x_i}{\Delta x_{i+1}(\Delta x_i + \Delta x_{i+1})},$$
$$\gamma_0 = \frac{-2\Delta x_{i+1} - \Delta x_{i+2}}{\Delta x_{i+1}(\Delta x_{i+1} + \Delta x_{i+2})}, \quad \gamma_1 = \frac{\Delta x_{i+1} + \Delta x_{i+2}}{\Delta x_{i+1}\Delta x_{i+2}}, \quad \gamma_2 = \frac{-\Delta x_{i+1}}{\Delta x_{i+2}(\Delta x_{i+1} + \Delta x_{i+2})},$$
$$\delta_{-1} = \frac{2}{\Delta x_i(\Delta x_i + \Delta x_{i+1})}, \quad \delta_0 = \frac{-2}{\Delta x_i \Delta x_{i+1}}, \quad \delta_1 = \frac{2}{\Delta x_{i+1}(\Delta x_i + \Delta x_{i+1})}.$$

Note that the PDE (4.2.1) also has mixed derivative terms $f_{xy}(x_i, y_j)$ who are treated by successively applying the central scheme (4.2.9) in x- and y-directions. Those schemes possess a truncation error of second order on smooth grids[2] whenever the discretized function is often enough continuously differentiable.

Considering the two Dirichlet boundary conditions (4.2.3) and (4.2.4) at

[2]The term smooth refers to the function used to generate the non uniform grid.

the boarders $s = 0$ and S_{max}, our relevant set of grid points is

$$\mathcal{G} = \{(s_i, v_j, r_k) : 1 \leq i \leq m_1 - 1, 0 \leq j \leq m_2, 0 \leq k \leq m_3\}.$$

We replace every spatial derivative by its corresponding central scheme, except for the following regions:

- At $v = 0$ we simply set this value in the HHW PDE. Except for the first spatial derivative $\partial u/\partial v$ all other terms in the v-direction vanish. For $\partial u/\partial v$ we then use the forward scheme (4.2.10), independently of whether or not the Feller condition holds.

- In the region $v > \eta$ we apply the backward scheme (4.2.8) for the derivative $\partial v/\partial v$. It reduces spurious oscillations in the FD solution when σ_1 is small. Central schemes often produce oscillations making one sided schemes preferable.

- At the boundaries $r = \pm R_{max}$ the Neumann conditions (4.2.6) give $\partial u/\partial r$. It follows that the mixed derivative terms $\partial^2 u/\partial s \partial r$ and $\partial^2 u/\partial v \partial r$ vanish there. The second order derivative $\partial^2 u/\partial r^2$ is approximated using the central scheme (4.2.11) with virtual points $r_{m_3} + \Delta r_{m_3} > R_{max}$ and $r_0 - \Delta r_0 < -R_{max}$. Due to the vanishing derivative we can simply copy the value from the boarder $\pm R_{max}$.

- In the $r-$direction the backward (4.2.8) and (4.2.10) schemes are used if $r < 0$ respectively $r > 0$. This is, again, done to reduce spurious oscillations.

After the discretization, of the three space dimensions, the initial-boundary value problem of the HHW PDE transforms into an initial value problem for a system of stiff ordinary differential equations (ODEs):

$$U'(t) = A(t)U(t) + g(t), \quad 0 \leq t \leq T \text{ and } U(0) = U_0,$$

with $A(t)$ a given matrix and $g(t)$ a given vector determined by the boundary conditions. Ordered in a convenient way $U(t)$ contains the approximations of the option values $u(s, v, r, t)$ for every $(s, v, r, t) \in \mathcal{G} \times [0, T]$. Obviously U_0 is directly obtained from the payoff function at $t = T$. It is important here to mention that the size of this ODE system, which equals $m = (m_1 - 1)(m_2 + 1)(m_3 + 1)$, is usually very large making it numerically challenging. In our numerical tests we had to deal with sizes up to one million. Due to this fact alternating direction implicit (ADI) schemes have been introduced.

4.2.4 Time discretization (ADI schemes)

Those huge sizes of linear equation systems reveal the downsides of standard methods like the Crank-Nicolson scheme, it is therefore very important to choose an appropriate time discretization scheme. In every new time step the system of linear equations needs to be solved. The large bandwidth of $A(t)$ makes this computationally very demanding and therefore Crank-Nicolson numerically inefficient. We have to watch out for a more effective solution method. Dividing the matrix $A(t)$ into simpler ones, containing only derivatives in one spatial dimension, solves the problem of large bandwidths. This is actually the key approach leading us to the ADI splitting schemes. Here, we will decompose $A(t)$ into four simpler matrices,

$$A(t) = A_0 + A_1 + A_2 + A_3(t),$$

containing the part coming from all the mixed derivative terms and the ones from the $s-, v-$, and $r-$directions respectively. The remaining term ru from equation 4.2.1 is distributed evenly over the latter three. In analogy to the matrices we will split the vector into $g(t) = g_0(t) + g_1(t) + g_2(t) + g_3(t)$. Note, that for the matrices the time dependency of $A(t)$ is only passed on to $A_3(t)$, which makes A_0, A_1, A_2 time independent. In contrary to standard European call options, like in 'tHout & Haentjens [18], the time dependency of $g(t)$ is passed on to $g_0(t), g_1(t), g_2(t)$ and $g_3(t)$. This is due to the more complex boundary conditions for autocallables. This requires a bit of adjustment in the time discretization schemes which are in this paper. Finally, depending on how one arranges the sequence of the solution vector $U(t)$, the matrices $A_1, A_2, A_3(t)$ are essentially tridiagonal or pentadiagonal.

Now we discretize time by setting $t_n = n \cdot \Delta t$ with $\Delta t = T/N$, a fixed integer $N \geq 1$ and $\theta > 0$ a given real parameter. These schemes generate successive approximations U_n to the solution $U(t_n)$ for $n = 1, 2, \ldots, N$. Adapting the schemes to our case where every spatial boundary vector is time dependent we end up with the following ADI schemes:

Douglas (Do) scheme:

$$\begin{cases} Y_0 &= U_{n-1} + \Delta t(A(t_{n-1})U_{n-1} + g(t_{n-1})), \\ Y_j &= Y_{j-1} + \theta\Delta t(A_j Y_j - A_j U_{n-1} + g_j(t_n) - g_j(t_{n-1})), \qquad (j = 1, 2) \\ Y_3 &= Y_2 + \theta\Delta t(A_3(t_n)Y_3 - A_3(t_{n-1})U_{n-1} + g_3(t_n) - g_3(t_{n-1})), \\ U_n &= Y_3. \end{cases} \tag{4.2.12}$$

Craig-Sneyd (CS) scheme:

$$\begin{cases} Y_0 &= U_{n-1} + \Delta t(A(t_{n-1})U_{n-1} + g(t_{n-1})), \\ Y_j &= Y_{j-1} + \theta\Delta t(A_jY_j - A_jU_{n-1} + g_j(t_n) - g_j(t_{n-1})), \qquad (j = 1, 2) \\ Y_3 &= Y_2 + \theta\Delta t(A_3(t_n)Y_3 - A_3(t_{n-1})U_{n-1} + g_3(t_n) - g_3(t_{n-1})), \\ \tilde{Y}_0 &= Y_0 + \frac{1}{2}\Delta t(A_0Y_3 - A_0U_{n-1} + g_0(t_n) - g_0(t_{n-1})), \\ \tilde{Y}_j &= \tilde{Y}_{j-1} + \theta\Delta t(A_j\tilde{Y}_j - A_jU_{n-1} + g_j(t_n) - g_j(t_{n-1})), \qquad (j = 1, 2) \\ \tilde{Y}_3 &= \tilde{Y}_2 + \theta\Delta t(A_3(t_n)\tilde{Y}_3 - A_3(t_{n-1})U_{n-1} + g_3(t_n) - g_3(t_{n-1})), \\ U_n &= \tilde{Y}_3. \end{cases} \tag{4.2.13}$$

Modified Craig-Sneyd (MCS) scheme:

$$\begin{cases} Y_0 &= U_{n-1} + \Delta t(A(t_{n-1})U_{n-1} + g(t_{n-1})), \\ Y_j &= Y_{j-1} + \theta\Delta t(A_jY_j - A_jU_{n-1} + g_j(t_n) - g_j(t_{n-1})), \qquad (j = 1, 2) \\ Y_3 &= Y_2 + \theta\Delta t(A_3(t_n)Y_3 - A_3(t_{n-1})U_{n-1} + g_3(t_n) - g_3(t_{n-1})), \\ \hat{Y}_0 &= Y_0 + \theta\Delta t(A_0Y_3 - A_0U_{n-1} + g_0(t_n) - g_0(t_{n-1})), \\ \tilde{Y}_0 &= \hat{Y}_0 + (\frac{1}{2} - \theta)\Delta t(A(t_n)Y_3 - A(t_{n-1})U_{n-1} + g(t_n) - g(t_{n-1})), \\ \tilde{Y}_j &= \tilde{Y}_{j-1} + \theta\Delta t(A_j\tilde{Y}_j - A_jU_{n-1} + g_j(t_n) - g_j(t_{n-1})), \qquad (j = 1, 2) \\ \tilde{Y}_3 &= \tilde{Y}_2 + \theta\Delta t(A_3(t_n)\tilde{Y}_3 - A_3(t_{n-1})U_{n-1} + g_3(t_n) - g_3(t_{n-1})), \\ U_n &= \tilde{Y}_3. \end{cases} \tag{4.2.14}$$

It is easy to see that the A_0 part is always treated in an *explicit* way. This is due to its large bandwidth. All the other parts, however, are dealt with in an *implicit* fashion. The main task, in every time step, is to solve systems of linear equations, with coefficient matrices of the form $(I-\theta\Delta tA_j)$ for $j = 1, 2$ and $(I - \theta\Delta tA_3(t_n)$, with I being the corresponding identity matrix. The matrices $A_1, A_2, A_3(t)$ have a small bandwidth so we can do this efficiently by LU factorization. Due to their time independence the matrices A_1, A_2 can be created once beforehand. If the time step is constant we can do the same with their LU factorizations L_1, U_1, L_2, U_2. For every new time step Δt we have to create a new factorization. Unfortunately, the time dependency of $A_3(t)$ requires the whole procedure of creation and factorization to be done in every time step.

4.2.5 Comments on stability and error analysis

The CS and MCS schemes are simply extensions of the Do scheme, whereas the former are the same for $\theta = \frac{1}{2}$. One can obtain the order of consistency (in

the non stiff sense) of these schemes via Taylor expansion and some tedious calculations. The order of the Do scheme is one for any given θ if the matrix A_0 is non zero. The CS scheme has order two for a θ of $\frac{1}{2}$, whereas the MCS has it for any given parameter value. With the latter scheme one can therefore choose θ to meet some additional requirements.

Stability is also very important when applying such numerical schemes. For the above ADI schemes such results are discussed in detail by 'tHout & Welfert [24]. To be precise, they are speaking of *unconditional* stability, meaning without a restriction on the time step Δt, and in the l_2-norm. Lets state the semidiscretized convection diffusion equation as

$$\frac{\partial u}{\partial t} = c \cdot \nabla u + \nabla \cdot (D \nabla u),$$

with c and $D = (d_{ij})$ being a constant real vector and a constant positive semidefinite real matrix. Due to the occurrence of mixed derivatives the matrix D will be non diagonal.

Unfortunately, sufficient conditions on θ regarding stability, for three dimensional convection diffusion equations with mixed derivatives do not exist in the corresponding literature. Thus, in this implementation, the parameter θ will be selected on the base of results for two dimensional convection diffusion and three dimensional pure diffusion equations, both with mixed derivative terms. To express the results lets state the following condition on D

$$|d_{ij}| \leq \gamma \sqrt{d_{ii} d_{jj}} \qquad \text{for all } i \neq j, \tag{4.2.15}$$

where $\gamma \in [0, 1]$ can be interpreted as the relative size of the mixed derivative coefficients. For $\gamma = 1$ the above condition is always fulfilled because D is positive semidefinite. With the HHW PDE we have more information, expressed in a $\gamma < 1$.

The Do and CS schemes are both stable if $\theta \geq \frac{1}{2}$ for two dimensional convection diffusion equations with mixed derivatives, whereas for the MCS scheme we have $\frac{1}{2} \leq \theta \leq 1$. For a $\gamma \leq 0.96$ the latter scheme is also stable for $\theta = \frac{1}{3}$. In the case of three dimensional pure diffusion equations, with mixed derivatives, the Do, CS and MCS schemes are stable whenever $\theta \geq \frac{2}{3}, \theta \geq \frac{1}{2}$ and $\theta \geq \max\{\frac{1}{4}, \frac{2}{13}(2\gamma + 1)\}$ respectively. Also, a smaller value for θ turns out to produce a smaller error.

Looking at the diffusion matrix for the HHW PDE

$$D(s, v) = \frac{1}{2} \begin{pmatrix} s^2 v & \rho_{Sv} \sigma_1 s v & \rho_{Sr} \sigma_2 s \sqrt{v} \\ * & \sigma_1^2 v & \rho_{23} \sigma_1 \sigma_2 \sqrt{v} \\ * & * & \sigma_2^2 \end{pmatrix},$$

it's easy to see that the condition (4.2.15) holds for $\gamma = \max\{|\rho_{Sv}|, |\rho_{Sr}|, |\rho_{23}|\}$. In the implementation we therefore select for the Do, CS and MCS schemes the values $\theta = \frac{2}{3}, \theta = \frac{1}{2}$ and $\theta = \max\{\frac{1}{3}, \frac{2}{13}(2\gamma + 1)\}$ respectively. As one would naturally guess, the largest discretization errors occur around the barrier levels and observation dates, due to the non smoothness of the payoff function. For such regions, in the literature, often a damping procedure is applied to reduce the errors. At the observation dates $T_1, \ldots, T_4$, instead of performing one step with length Δt, one could make two substeps of length $\Delta t/2$ with the Do scheme and a parameter value of $\theta = 1$. Although, numerical tests have shown, that in our case the benefits are not significant.

In 'tHout & Haentjens [18] temporal errors, for numerically challenging up and out call options, were inspected. They found that for sufficiently small Δt the errors behave as $C(\Delta t)^p$, where $p = 1.0$ for the Do and $1.6 \leq p \leq 2.0$ for the CS and MCS scheme, with a constant C. Whereas p and C are only weakly dependent on the number of grid points.

4.3 Calibrating the HHW model to market data

In the last chapter we have seen that knowledge of the characteristic function of a model leads to highly efficient pricing formulas within the Fourier framework. Following Duffie [12], the class of affine diffusion processes will be introduced and further, some transform analysis which leads us to their characteristic function. To perform this analysis, the non affine HHW model additionally requires approximation of its non affine terms. We will then plug the ChF into a Lewis [30] style formula which was derived in Chapter 3.3.4. Finally this formula will be implemented in a way which makes the calibration routine even faster than the celebrated FFT.

4.3.1 Affine diffusion processes

Lets say we have a process X, which is Markovian in some state space $D \subset \mathbb{R}^n$, and solves the following system of SDEs,

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t,$$

where $\mu : D \to \mathbb{R}^n$, $\sigma : D \to \mathbb{R}^{n\times m}$ and W a m-dimensional standard Brownian motion.

Definition 14 (processes of affine form)**.** *The above system is said to be of the affine form if:*

$$\begin{aligned}
\mu(X_t) &= a_0 + a_1 X_t, && \text{for some } (a_0, a_1) \in \mathbb{R}^n \times \mathbb{R}^{n\times n},\\
[\sigma(X_t)\sigma(X_t)^T]_{ij} &= (c_0)_{ij} + (c_1)_{ij}^T X_t, && \text{for } (c_0, c_1) \in \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n\times n},\\
r(X_t) &= r_0 + r_1^T X_t, && \text{for arbitrary } (r_0, r_1) \in \mathbb{R} \times \mathbb{R}^n,
\end{aligned}$$

for $i, j = 1, \ldots, n$, where $r(X_t) : D \to \mathbb{R}$ is a fixed affine discount rate function (interest rate component).

Proposition 6. *Under the assumptions of Definition 14 before, a transform $\phi(z, X_t, t, T) : \mathbb{C}^n \times D \times \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{C}$ of X_t given by*

$$\phi(z, X_t, t, T) = \mathbb{E}\left(\exp\left(-\int_t^T r(X_s)ds\right) e^{izX_T} | \mathcal{F}_t\right).$$

is

$$\phi(z, X_t, t, T) = e^{A(z,t)+B(z,t)X_t},$$

where the coefficients $A(z,t)$ and $B(z,t)$ have to satisfy the complex valued ODEs

$$\begin{aligned}
\frac{d}{dt}B(z,t) &= r_1 - a_1^T B(z,t) - \frac{1}{2}B(z,t)^T c_1 B(z,t),\\
\frac{d}{dt}A(z,t) &= r_0 - a_0^T B(z,t) - \frac{1}{2}B(z,t)^T c_0 B(z,t),
\end{aligned}$$

with boundary conditions $B(z,T) = z$ and $A(z,T) = 0$.

Remark: Note that $\phi(z, X_t, t, T)$ differs from the familiar characteristic function of X_T due to the discounting $r(X_t)$. Conversely to Grzelak and Oosterlee [15] we also use the generalized characteristic function where the argument is complex valued, to be able to apply the Fourier methods from Lewis [30] as in chapter 3.3.4 to obtain a semi closed formula used for calibration. Duffie [12] also handles jumps in his transform analysis, which we dropped here since we do not have any in the HHW model.

4.3.2 Change to the forward measure

As explained in Chapter 3.3.4 already, we need to switch to the forward price, defined by $F(t) = \frac{S(t)}{B(t,T)}$, to decouple discounting from the expectation and being able to apply formula (3.3.10). Following Grzelak and Oosterlee [17] the price for an European call option is then,

$$V(t, S, v) = B(t,T)\mathbb{E}^T\left(\max(F(T) - K, 0) \,|\, \mathcal{F}_t\right),$$

where the interest rate r disappeared from the argument. Applying Itô we find the dynamics of the forward price as,

$$\mathrm{d}F(t) = \frac{1}{B(t,T)}\mathrm{d}S(t) - \frac{S(t)}{B^2(t,T)}\mathrm{d}B(t,T), \tag{4.3.1}$$

which is a Martingale under the T-forward measure and hence we don't find any drift in it anymore. Since the short rate is modeled as Hull-White process, the dynamics for the ZCB $B(t,T)$ in (4.3.1) are:

$$\mathrm{d}B(t,T)/B(t,T) = r(t)dt + \sigma_2 B_r(t,T)\mathrm{d}W_r^{\mathbb{T}}(t), \tag{4.3.2}$$

with $B_r(t,T) = \frac{1}{a}\left(e^{-a(T-t)} - 1\right)$. Hence, the model under the T-forward measure is:

$$\begin{aligned} \mathrm{d}F(t)/F(t) &= \sqrt{v(t)}\mathrm{d}W_F^T(t) - \sigma_2 B_r(t,T)\mathrm{d}W_r^T(t), \\ \mathrm{d}v(t) &= \kappa(\eta - v(t))\mathrm{d}t + \sigma_1\sqrt{v(t)}\mathrm{d}W_v^T(t), \end{aligned} \tag{4.3.3}$$

with correlations $\mathrm{d}W_F^T(t)\mathrm{d}W_v^T(t) = \rho_{Fv}\mathrm{d}t$ and $\mathrm{d}W_F^T(t)\mathrm{d}W_r^T(t) = \rho_{Fr}\mathrm{d}t$ as before.

4.3.3 The forward HHW PDE

In order to retrieve the characteristic function one first has to determine its corresponding pricing PDE. The spot HHW PDE is useless here since we need to change to the forward measure, like discussed before, to apply Fourier techniques. We simply proceed, with the forward HHW model (4.3.3), as in chapter (3.2.2) to derive the corresponding PDE.

Recall the HHW forward model from equation (4.3.3) above with state vector $X_t^* = (v_t, F_t)^T$. With the correlation structure for the three BM from (2.3.2), the SDEs for the HHW forward model can be written as

$$\begin{pmatrix} dv_t \\ dF_t/F_t \end{pmatrix} = b(X_t^*)dt + \sigma(X_t^*)\begin{pmatrix} dZ_t^1 \\ dZ_t^2 \\ dZ_t^3 \end{pmatrix} \tag{4.3.4}$$

with

$$b(X_t^*) = \begin{pmatrix} \kappa(\eta - v_t) \\ 0 \end{pmatrix},$$

$$\sigma(X_t^*) = \begin{pmatrix} \sigma_1\sqrt{v_t} & 0 & 0 \\ \rho_{Fv}\sqrt{v_t} & \rho_{Fr}\sqrt{v_t} - \sigma_2 B_r(t,T) & \sqrt{1-\rho_{Fr}^2-\rho_{Fv}^2}\sqrt{v_t} \end{pmatrix},$$

with independent Brownian motions Z_t^1, Z_t^2 and Z_t^3. Hence, the instantaneous covariance matrix is

$$\Sigma(X_t^*) = \begin{pmatrix} \sigma_1^2 v_t & \rho_{Fv}\sigma_1 v_t \\ * & v_t + \sigma_2^2 B_r^2(t,T) - 2\rho_{Fr}\sigma_2 B_r(t,T)\sqrt{v_t} \end{pmatrix}. \tag{4.3.5}$$

Applying Feynman-Kac yields the HHW forward PDE

$$\begin{aligned} 0 = \frac{\partial u}{\partial t} + \kappa(\eta - v)\frac{\partial u}{\partial v} + F^2\left(\frac{1}{2}v + \frac{1}{2}\sigma_2^2 B_r^2(t,T) - \rho_{Fr}\sigma_2 B_r(t,T)\sqrt{v}\right)\frac{\partial^2 u}{\partial F^2} \\ + \frac{1}{2}\sigma_1^2 v\frac{\partial^2 u}{\partial v^2} + \rho_{Fv}\sigma_1 v F\frac{\partial^2 u}{\partial F \partial v}. \end{aligned} \tag{4.3.6}$$

4.3.4 A deterministic approach to an affine version of the forward HHW model

In the pure Heston case a simple switch to the log-price x_t made the model affine. Applying Itô's formula with $x_t = \log F_t$ to (4.3.6) yields

$$\begin{aligned} 0 = \frac{\partial u}{\partial t} + \left(\frac{1}{2}v + \frac{1}{2}\sigma_2^2 B_r^2(t,T) - \rho_{Fr}\sigma_2 B_r(t,T)\sqrt{v}\right)\left(\frac{\partial^2 u}{\partial x^2} - \frac{\partial u}{\partial x}\right) \\ + \kappa(\eta - v)\frac{\partial u}{\partial v} + \frac{1}{2}\sigma_1^2 v\frac{\partial^2 u}{\partial v^2} + \rho_{Fv}\sigma_1 v\frac{\partial^2 u}{\partial x \partial v}. \end{aligned}$$

From that, or equivalently from the instantaneous covariance matrix in (4.3.5), we obtain that the system is still not of the affine form due to $\Sigma_{(2,2)}$ containing $\sqrt{v(t)}$. Applying Itô's formula again to $\sqrt{v(t)}$ is not possible because the second derivative of the square root at $v = 0$ does not exist. A trivial step to make the model affine would be to set ρ_{Fr} to zero, which clearly is no meaningful solution. Hence another way has to be found.

Detailed thoughts on how to reformulate the model, under the spot martingale measure, to impose an indirect correlation between stock price and interest rate can be found in Grzelak and Oosterlee [15, Ch. 2 ff]. They conclude that it is sufficient to approximate the non affine term in the covariance matrix. Also the alternative extension where the short rate is modeled as a CIR type process is handled there. Note that the HW process is just a special case of CIR where the exponent of the state variable in the diffusion is set to zero.

Deterministic approach

The easiest way for an approximation for the term $\Sigma_{(2,2)}$ is to replace $\sqrt{v_t}$ by its expectation:

$$\Sigma_{(2,2)} \approx v(t) + \sigma_2^2 B_r^2(t,T) - 2\rho_{Fr}\sigma_2 B_r(t,T)\mathbb{E}\left(\sqrt{v_t}\right). \tag{4.3.7}$$

Now the model is of the affine form as in chapter 4.3.1 discussed and we can derive its characteristic function. To calculate $\mathbb{E}(\sqrt{v_t})$, Lemma 5 gives a closed form expression for the expectation and variance of a CIR type process.

Lemma 5 (expectation and variance of a CIR type process)**.** *[15, Lem. 3.1] Suppose v_t is of CIR type as in 4.3.3. Then, for a given time $t > 0$, its square root's expectation and variance are given by*

$$\mathbb{E}\left(\sqrt{v_t}\right) = \sqrt{2c(t)}e^{-\lambda(t)/2}\sum_{k=0}^{\infty}\frac{1}{k!}(\lambda(t)/2)^k\frac{\Gamma(\frac{1+d}{2}+k)}{\Gamma(\frac{d}{2}+k)}, \tag{4.3.8}$$

and

$$\mathrm{Var}\left(\sqrt{v_t}\right) = c(t)(d+\lambda(t)) - 2c(t)e^{-\lambda(t)}\left(\sum_{k=0}^{\infty}\frac{1}{k!}(\lambda(t)/2)^k\frac{\Gamma(\frac{1+d}{2}+k)}{\Gamma(\frac{d}{2}+k)}\right)^2,$$

where

$$c(t) = \frac{1}{4\kappa}\sigma_1^2(1-e^{-\kappa t}), \quad d = \frac{4\kappa\eta}{\sigma_1^2}, \quad \lambda(t) = \frac{4\kappa v(0)e^{-\kappa t}}{\sigma_1^2(1-e^{-\kappa t})},$$

with $\Gamma(k)$ being the gamma function:

$$\Gamma(k) = \int_0^{\infty} t^{k-1}e^{-t}dt.$$

However, evaluating this analytic expression is quite expensive and for that reason [15] derived a much cheaper approximation based on the *delta method* for a random variable X. This method states, that for a random variable X, with expectation $\mathbb{E}(X)$ and variance $\mathrm{Var}(X)$, one can approximate a function $\phi(X)$ by a first order Taylor expansion at $\mathbb{E}(X)$ provided that its first derivative with respect to X exists and is sufficiently smooth.

Lemma 6 (delta method for the expectation of the variance process)**.** *[15, Lem. 3.2] The expectation of $\sqrt{v_t}$ can be approximated as:*

$$\mathbb{E}\left(\sqrt{v_t}\right) \approx \sqrt{c(t)(\lambda(t)-1) + c(t)d + \frac{c(t)d}{2(d+\lambda(t)}}, \tag{4.3.9}$$

with

$$c(t) = \frac{1}{4\kappa}\sigma_1^2(1 - e^{-\kappa t}), \quad d = \frac{4\kappa\eta}{\sigma_1^2}, \quad \lambda(t) = \frac{4\kappa v(0)e^{-\kappa t}}{\sigma_1^2(1 - e^{-\kappa t})}.$$

Note that in order for the expression under the square root to stay positive we need $8\kappa\eta/\sigma_1^2 \geq 1$ as discussed in the Appendix of [15]. Clearly this is the case if the Feller condition is satisfied. The good quality of this approximation is shown in [15, Fig: 3.1]. This means, that whenever formula (4.3.9) is applicable, we will use it. And for the parameter sets where the square root would become negative, we use the exact representation from Lemma 5 which is always save to do so.

4.3.5 Characteristic function for the deterministic approximation of the HHW model

We can now use Duffie's transform analysis and apply it to the HHW forward PDE. Recall Proposition 6 which states, applied to the HHW forward model with its two state variables $X_t = (x_t, v_t)$, that the discounted characteristic function is of the following form[3]:

$$\phi(u, X_t, \tau) = \exp\{A(u,\tau) + B(u,\tau)x_t + C(u,\tau)v_t\} \tag{4.3.10}$$

with boundary conditions $A(u,0) = 0, B(u,0) = iu$ and $C(u,0) = 0$. Where $\tau := T - t$ is the time lag.

Lemma 7. *For $u \in \mathbb{C}$ and $\tau \geq 0$ the functions $A(u,\tau), B(u,\tau)$ and $C(u,\tau)$ from (4.3.10) are given by:*

$$\begin{aligned}
B(u,\tau) &= iu, \\
C(u,\tau) &= \frac{1 - e^{-D_1\tau}}{\sigma_1^2(1 - ge^{-D_1\tau})}(\kappa - \sigma_1\rho_{xv}iu - D_1), \\
A(u,\tau) &= \kappa\eta I_2(\tau) + \frac{1}{2}\sigma_2^2 I_3(\tau) + \rho_{xr}\sigma_2 I_4(\tau),
\end{aligned}$$

[3]Note that we only put the summand $B(u,\tau)x_t$ in it here for completeness reasons. In the derivation of the Fourier formula we assumed the dynamics of the forward price as $F_t = F_0 e^{X_t}$ in order for the ChF to fulfill the martingale identity $\phi_{X_t}(-i) = 1$.

with $D_1 = \sqrt{(\sigma_1\rho_{xv}iu-\kappa)^2 - \sigma_1^2 iu(iu-1)}$ *and* $g = \frac{\kappa-\sigma_1\rho_{xv}iu-D_1}{\kappa-\sigma_1\rho_{xv}iu+D_1}$, *where*

$$I_2(\tau) = \frac{\tau}{\sigma_1^2}(\kappa - \sigma_1\rho_{xv}iu - D_1) - \frac{2}{\sigma_1^2}\log\left(\frac{1-ge^{-D_1\tau}}{1-g}\right),$$
$$I_3(\tau) = iu(iu-1)\frac{1}{2a^3}(-3 - e^{-2a\tau} + 4e^{-a\tau} + 2a\tau),$$
$$I_4(\tau) = iu(iu-1)\frac{1}{a}\int_0^\tau \mathbb{E}\left(\sqrt{v(T-s)}\right)(1-e^{-as})ds.$$

Proof. With the state vector $X_t = (x_t, v_t)^T$ the form $\phi := \phi(u, X_t, t, T)$ has to satisfy the pricing PDE

$$0 = \frac{\partial\phi}{\partial t} + \kappa(\eta - v)\frac{\partial\phi}{\partial v} + \frac{1}{2}\sigma_1^2 v\frac{\partial^2\phi}{\partial v^2} + \rho_{Fv}\sigma_1 v\frac{\partial^2\phi}{\partial x\partial v}$$
$$+ \left(\frac{1}{2}v + \frac{1}{2}\sigma_2^2 B_r^2(t,T) - \rho_{Fr}\sigma_2 B_r(t,T)\mathbb{E}\left(\sqrt{v}\right)\right)\left(\frac{\partial^2\phi}{\partial x^2} - \frac{\partial\phi}{\partial x}\right),$$

subject to the terminal condition $\phi(u, X_T, T, T) = \exp(iux_T)$. Since this PDE is affine, its solution is of the form as in equation (4.3.10). With the notation $A := A(u,t,T), B := B(u,t,T)$ and $C := C(u,t,T)$, the partial derivatives are given by:

$$\frac{\partial\phi}{\partial t} = \phi\left(\frac{\partial A}{\partial t} + x\frac{\partial B}{\partial t} + v\frac{\partial C}{\partial t}\right),$$
$$\frac{\partial\phi}{\partial x} = B\phi, \frac{\partial^2\phi}{\partial x^2} = B^2\phi,$$
$$\frac{\partial\phi}{\partial v} = C\phi, \frac{\partial^2\phi}{\partial v^2} = C^2\phi,$$
$$\frac{\partial^2\phi}{\partial x\partial v} = BC\phi.$$

Substituting these into the PDE:

$$0 = \frac{\partial A}{\partial t} + x\frac{\partial B}{\partial t} + v\frac{\partial C}{\partial t} + \kappa(\eta - v)C + \frac{1}{2}\sigma_1^2 vC^2 + \rho_{Fv}\sigma_1 vBC$$
$$+ \left(\frac{1}{2}v + \frac{1}{2}\sigma_2^2 B_r^2(t,T) - \rho_{Fr}\sigma_2 B_r(t,T)\mathbb{E}(\sqrt{v})\right)\left(B^2 - B\right).$$

By collecting terms for x and v we get the following set of ODEs:

$$\frac{\partial B}{\partial t} = 0,$$
$$\frac{\partial C}{\partial t} = \kappa C - \frac{1}{2}\sigma_1^2 C^2 - \frac{1}{2}(B^2 - B) - \rho_{Fv}\sigma_1 BC,$$
$$\frac{\partial A}{\partial t} = -\kappa\eta C - \left(\frac{1}{2}\sigma_2^2 B_r^2(t,T) - \rho_{Fr}\sigma_2 B_r(t,T)\mathbb{E}\left(\sqrt{v}\right)\right)\left(B^2 - B\right),$$

subject to the boundary conditions $A(u,0) = 0, B(u,0) = iu$ and $C(u,0) = 0$. Now set $\tau = T - t$. Due to the boundary condition for B, we obviously have, $B(u,\tau) = iu$. The ODE for $C(u,\tau)$ is given as a Riccati type equation

$$\frac{d}{d\tau}C(u,\tau) = q_2 C^2(u,\tau) + q_1 C(u,\tau) + q_0,$$

with corresponding q_0, q_1, q_2. Its solution is given by:

$$C(u,\tau) = \frac{1 - e^{-D_1\tau}}{\sigma_1^2(1 - ge^{-D_1\tau})}(\kappa - \sigma_1\rho_{xv}iu - D_1),$$

with D_1 and g as above. Integrating A yields

$$\begin{aligned} A(\tau) - A(0) = \kappa\eta \int_0^\tau C(s)ds \\ + iu(iu-1)\left(\frac{1}{2}\sigma_2^2 \int_0^\tau B_r^2(s)ds - \rho_{Fr}\sigma_2 \int_0^\tau B_r(s)\mathbb{E}\left(\sqrt{v(T-s)}\right)ds\right) \end{aligned}$$

and a straight forward evaluation of the integrals

$$A(\tau) = \kappa\eta I_2(\tau) + \frac{1}{2}\sigma_2^2 I_3(\tau) + \rho_{Fr}\sigma_2 I_4(\tau)$$

with $I_2(\tau), I_3(\tau)$ and $I_4(\tau)$ as stated above. □

Remark: Note that the equations for $\frac{d}{d\tau}C(u,\tau)$ and $I_2(\tau)$ are of the same form as in Heston [20]. The ODE for the variance $C(u,\tau)$ includes the complex root D_1, which has, of course, two possible values. Albrecher et al. [1] point out that, when choosing this value as Heston did, instabilities occur due to the branch cut[4] of the logarithm. Choosing the second value[5], as we did, results in a stable procedure.

Finally, plugging this ChF into our preferred option valuation formula (3.3.10), we can implement highly efficient calibration routines. The next chapter discusses how to do this in a way to be as fast as possible.

[4] A branch cut is a curve, with possibly open, closed, or half-open ends, in the complex plane across which an analytic multivalued function is discontinuous.

[5] This results in a minus in front of D_1; in physics a term of the form $e^{-D_1\tau}$ is often referred to as damping term.

4.3.6 Accelerating the calibration

The calibration routine we are going to implement is based on the direct integration of (3.3.10). This method is often criticized in the literature. Carr and Madan [7] fault the inability of it to keep up with their celebrated FFT method in terms of speed. They also point out numerical instabilities in case of decomposing the option price into probabilistic elements as in Heston [20]. However, for the latter modifications have been proposed, as for example the one from Lewis [30] we derived in Section 3.3.4 or another one as in Attari [4], which are free from this instability. Also the speed argument is only valid if one implements the direct integration method in a very needlessly unoptimized way. The purpose of this chapter is to show, as Kilin did in [27], a naturally seeming, optimized way of implementing the direct integration method.

Fast Fourier Transform (FFT)

Let us just quickly review the basic idea of FFT to better understand its major downside, namely the harsh restrictions on the grid spacings, which blow up computational speed. The value of a call option $V(S_0)$ is given as

$$V(S_0) = \frac{e^{-\gamma k}}{\pi} \int_0^\infty e^{-iku} \psi(u) du,$$

with

$$\psi(u) = \frac{e^{-rT} \phi(u - (\gamma + 1)i)}{\gamma^2 + \gamma - u^2 + (2\gamma + 1)ui},$$

where γ is a damping parameter, k the log strike price and $\phi(u)$ the ChF of the log price. The above integral is then approximated using an integration rule

$$\begin{gathered} \int_0^\infty e^{-iku} \psi(u) du \approx \sum_{j=0}^{N_{FFT}-1} e^{-iku_j} \psi(u_j) w_j \delta, \\ u_j = j\delta, \end{gathered} \tag{4.3.11}$$

with weights w_j and N_{FFT} the number of grid points. The crucial restriction here is, that the grid points u_j must be chosen equidistantly. Effective integration rules, i.e. the Gaussian quadrature, cannot be used therefore. With a set of log strikes $\{k_m = -(\frac{N_{FFT}\lambda}{2}) + m\lambda,\ m = 0, \ldots, N_{FFT} - 1\}$, the FFT simultaneously computes the above integral approximation. If the second restriction for the grid spacings,

$$\lambda\delta = \frac{2\pi}{N_{FFT}},$$

is satisfied, the sums (4.3.11) can be written as

$$\sum_{j=0}^{N_{FFT}-1} e^{-iku_j}\psi(u_j)w_j\delta = \sum_{j=0}^{N_{FFT}-1} e^{-i\left(\frac{2\pi}{N_{FFT}}\right)jm}h_j,$$

which allows the application of the FFT with vector $h = \{h_j = e^{i\left(\frac{N\lambda}{2}j\delta\right)}\psi(u_j)w_j\delta, j = 0, \ldots, N_{FFT} - 1\}$. Chourdakis [8] showed how to accelerate this method with his fractional FFT algorithm. Within this method the latter condition on the grid spacings can be relaxed. Further the accuracy of the FFT strongly depends on the choice of the dampening parameter γ. Using a practicable grid size of $N_{FFT} < 4096$ there is no value for γ which yields acceptable pricing errors for all possible parameter values as Lord and Kahl [31] provide examples for that. Whereas a refinement of the grid would solve this problem, it also slows down the calibration.

Caching technique for the direct integration method

However, the simultaneous calculation of all strikes is not only an exclusive advantage of FFT, since a smart implementation of direct integration yields the same. This is due to the fact that the value of the ChF does not depend on the strike. Hence the values for the various maturities can be stored in the cache and used in every loop of the strike. A basic algorithm to avoid these recalculations would look like:

1: Loop over maturities.
2: Loop over strikes.
3: Loop over grid points for the integration rule.
4: If at the first step of the strike loop: Evaluate the ChF and save the value to the cache.
5: If not at the first step of the strike loop: Read the value of the ChF from the cache.
6: Evaluate the whole integrand.
7: Calculate the price of the option.

The number of evaluations of the ChF is the main driving factor slowing down the calibration and therefore should be as low as possible. There are even more terms in the valuation formula which do not depend on the strike or maturity. Hence, further acceleration could be done also pre-calculating and storing them in the cache. Applying the above caching technique and being able to use effective integration rules, which do not require any restrictions on the grid spacings, such as the Gaussian quadrature drastically lower this number.

To verify all that Kilin has conducted two experiments[6]: At first, by taking 100 option values with 100 different (reasonable) parameter sets, he evaluated the minimum grid size for the numerical integration rule to reach a certain accuracy. The results can be seen in [27, p. 12, Table 1]. For all the tested models the grid for the FFT based methods must be at least seven times finer than for the direct integration method. In the second experiment he compared calibration times directly where the grid size has been set as to have an accuracy of 1.0 and 0.02 basis points[7]. For that, the calibration time of the direct integration method, is approximately 16, respectively, seven times faster than for the others. This corresponds to the grid sizes from the first experiment.

Bottom line, when comparing the calibration speed, one has to take into account every practicably applicable optimization opportunity for that method. The simultaneous calculation of option prices for several strikes is not an exclusive advantage of FFT based methods. Hence, it is not a valid criteria for comparing the calibration methods. We have seen that the number of evaluations of the ChF is a much more appropriate measure for that.

[6]Even though he did not experiment with the HHW model directly, but with the Heston and several other Levy models, we can safely conclude similar results in our case.

[7]A basis point is one hundredth of a percent.

Chapter 5

Computational studies

In this final chapter the results from our numerical studies will be presented. Additionally to the, in this thesis focused HHW model, we will also calculate prices for the well known and approved Heston model mentioned in Section 2.1.2. We do that with all three pricing techniques, as far as they are applicable. This refers, of course, to the Fourier technique, which excludes the possibility to price path dependent options, like mentioned a couple of times.

We will start by doing some convergence and consistency checks on the two models and all three pricing techniques for simple European options. This is due to two reasons: The first, like just stated, is applicability of our semi closed formula within the Fourier framework. Hence, we can compare it to the other methods and see if it works, which we have to do in order to calibrate later with a peaceful conscience. The second reason is, that naturally the payoff structure is much simpler for the European call option making it a less challenging test to see if the methods were implemented correctly.

Followed by that we will apply the MC and PDE methods to the in the beginning introduced autocallable structured products. This is again done for the HHW and as well the Heston model. The last Section then will concentrate on the calibration to market data within the HHW model. Calibration of such a hybrid model is usually done in two steps: Firstly, the parameters of the Hull-White model are calibrated to market prices from caps[1].Knowing them they can be plugged into the HHW model to calibrate the remaining ones. Finally, we will also compare these calibrated parameters to those one would obtain from the pure Heston model.

Concerning the coding environment it shall be mentioned that all programs were written in Matlab Version 7.5.0 (R2007b)©. Table 5.1 shows

[1]Caps are the structural equivalent to call options in the interest rate market.

	Case A	Case B	Case C	Case D	Case E	Case F
κ	3	0.6067	2.5	0.5	0.3	1
η	0.12	0.0707	0.06	0.04	0.04	0.09
σ_1	0.04	0.2928	0.5	1	0.9	1
a	0.2	0.05	0.15	0.08	0.16	0.22
c_1	0.05	0.055	0.101	0.103	0.055	0.074
c_2	0.01	0.005	0.001	0.003	0.025	0.014
c_3	1	4	2.3	1	1.6	2.1
σ_2	0.03	0.06	0.1	0.09	0.03	0.07
ρ_{sv}	0.6	-0.7571	-0.1	-0.9	-0.5	-0.3
ρ_{sr}	0.2	0.6	-0.3	0.4	0.2	-0.5
T	1	3	0.25	10	15	5
K	100	100	100	100	100	100

Table 5.1: Parameters used in the tests. Note that for cases D,E and F the Feller condition is not satisfied. They are referred to as very challenging real life examples.

our six main test cases for the computational studies. They were already used in Haentjens & 't Hout [18], Oosterlee et al. [17] and Andersen[2] [3]. In the cases D,E and F the Feller condition is not satisfied and hence they are referred to as very challenging real life examples from the market. For the time dependent mean reverting function $b(t)$ in the HHW model (2.3.1) we will, if not explicitly stated, use the form:

$$b(t) = c_1 - c_2 e^{-c_3 t}.$$

5.1 Consistency and convergence studies for European calls

As mentioned in the first lines of this Chapter, we will not only concentrate on the autocallables for our pricing methods, but also basic European calls due to the applicability of the Fourier method. This way we see if our derived formula works, which is important when calibrating later.

[2]Andersen worked with the Heston model and the additional parameters were just set that the HHW model converges to it.

5.1.1 Convergence of the HHW to the Heston model for European calls

When discussing the different market models in Chapter 2 we've introduced the HHW model as the extension of the Heston stochastic volatility model with the Hull-White interest rate process. So at first we wanted to check if our derived methods and formulas converge to the very sophisticated Heston model when choosing the parameters appropriately. We have implemented the Fourier method also for the Heston model. Due to the constant interest rate, where no switch to the forward measure is needed, this was straight forward. The ChF[3] and option valuation formula are given in Albrecher et al. [1, p. 4] and Lewis [30, p. 14, formula 3.11].

To reduce the HHW model to the Heston model we simply set c_2, σ_2 and ρ_{sr} equal to zero for a constant and non correlated interest rate. Option prices, which were calculated for various strikes $K = \{70, 100, 140\}$ with initial values $S_0 = 100, v_0 = \eta$ and $r_0 = b(0)$, can be seen in Table 5.2.

The first line for each strike displays the prices obtained from the Fourier method in the Heston model. Whereas the following three show the absolute error compared to the prices in the HHW model. They are negligibly low for the Fourier and PDE methods and for the MC simulation they are all within a confidence interval of three standard deviations.

5.1.2 Consistency for European calls

This section investigates the consistency between the three methods in the HHW model. We will use again our standard test cases from Table 5.1 and compute European call prices for *in-the-money, at-the-money* and *out-of-the-money* options, e.g. $K = \{70, 100, 140\}$, respectively with start values for S_0, v_0 and r_0 as before. It shall also be mentioned that the grid generation for the PDE is done as in Section 4.2.2 suggested, except for the price dimension where we need to concentrate grid lines around one strike K only. Further, the time steps are chosen equidistantly.

For the *at-the-money* and *out-of-the-money* options the first three cases in Table 5.3 show that the bias between the two methods is very low, whereas it is higher for the latter three. This is due to the long maturities of those challenging examples where one can see that the deterministic approximation from Section 4.3.4 is not optimal. The *in-the-money* options show a better result, yet the negative effects of this approximation can be seen.

[3]Note that we used the second version of the ChF which has the minus in the complex exponent. As Albrecher pointed out in his paper, this so called damping term avoids numerical instabilities in the form of the branch cut of the complex logarithm.

		Case A	Case B	Case C	Case D	Case E	Case F
		$K = 100$					
Heston	Fourier	15.994	25.110	6.064	66.024	59.178	39.412
HHW	Fourier	0.000	0.008	0.000	0.040	0.012	0.002
	PDE	0.011	0.010	0.004	0.007	0.008	0.014
	MC	0.038	0.019	0.005	0.063	0.031	0.032
		$K = 70$					
Heston	Fourier	35.025	43.829	31.761	75.970	70.988	55.227
HHW	Fourier	0.001	0.008	0.003	0.012	0.002	0.003
	PDE	0.013	0.008	0.000	0.004	0.007	0.014
	MC	0.129	0.184	0.110	0.335	0.109	0.219
		$K = 140$					
Heston	Fourier	4.628	8.788	0.043	53.110	44.165	23.291
HHW	Fourier	0.001	0.004	0.000	0.081	0.034	0.003
	PDE	0.007	0.015	0.000	0.014	0.007	0.012
	MC	0.038	0.170	0.003	0.062	0.051	0.160

Table 5.2: Results for the test cases A-F and different strikes K with $c_2 = \sigma_2 = \rho_{sr} = 0$. The grid for the PDE method was chosen to be 100/50/50 in the (S, v, r) directions with $\Delta t = 1/16$ and for the MC simulation we had $\Delta t = 1/32$ with $5 \cdot 10^5$ paths.

As a consequence of the error induced by this approximation within the Fourier method, we are using the PDE results as the comparison to the MC prices. For all six cases Table 5.4 below lists MC estimates of the bias to the PDE prices, for a time step Δ ranging from 1 year to 1/32 year. Again, due to computational limitations the number of paths was set to $5 \cdot 10^5$. For a step size of four steps per year the bias is moderately low. Using a grid of eight steps a year we obtain already accurate results. Refining it further to 16 and 32 the bias becomes negligible with respect to the sampling error induced by the MC simulation. The starred results indicate that the PDE price lies within the MC confidence interval of three standard deviations. As in Andersen [3] it is hard to establish an empirical convergence order[4] since the biases are too low to be statistically significant. It however seems to be higher than one, which is the one of a basic Euler scheme. Andersen conducts further tests in [3, Ch. 5.1, Fig. 5] to show that this assumption is true. Since we implemented an exact simulation scheme for the interest

[4]Recall from Section 3.1 that an order n means that the absolute value of the bias decreases as $c \cdot \Delta^n$.

	Case A	Case B	Case C	Case D	Case E	Case F
	$K = 100$					
Fourier	15.662	27.019	5.983	57.048	50.266	34.281
PDE	15.651	26.941	5.982	55.984	50.067	34.627
Bias	0.011	0.078	-0.002	1.064	0.199	-0.347
	$K = 70$					
Fourier	34.524	43.839	31.752	65.317	63.264	51.109
PDE	34.512	43.929	31.741	64.660	63.226	50.982
Bias	0.012	-0.090	0.012	0.657	0.039	0.127
	$K = 140$					
Fourier	4.521	12.993	0.040	48.747	36.493	18.392
PDE	4.516	12.615	0.040	47.186	36.012	19.436
Bias	0.005	0.377	0.000	1.561	0.481	-1.402

Table 5.3: Comparison of Fourier and PDE prices for a set of three different strikes with $S_0 = 100, v_0 = \eta$ and $r_0 = b(0)$. The grid for the PDE method was chosen to be 100/40/30 points in the S, v and r dimension respectively, with 100 time steps a year.

rate process, as in Section 4.1.2 described, and used the same QE-scheme for the volatility, only extending his discretization of the price process we can savely assume the same here. All that results in a highly efficient simulation scheme when it comes to speed and accuracy. Additionally we conducted one further test to validate our Fourier formula and compared it to the PDE solution by calculating Fourier prices for every grid point on the restricted (s, v, r, t) domain $[70\ 140] \times v_0 \times r_0 \times 0$, for two different v_0 and r_0. Figure 5.1 shows the bias for three combinations of initial volatility and interest rate.

These two cases were chosen because A satisfies the Feller condition whereas E does not. They also have very diverse maturities of one and 15 years respectively. From Figure 5.1 we can see that the Fourier approximation formula is accurate for different (even extreme) initial values of v_0 and r_0 and maturities. In Table 5.5 we located the point S_i with the maximum error for each setting. Especially for case A we find very accurate results. For the challenging example E the maximum error is higher, yet it occurs for very far *out-of-the-money* or *in-the-money* initial prices.

	Case A	Case B	Case C	Case D	Case E	Case F
Δ	$K = 100$					
1	0.516	2.124		6.206	2.392	2.185
1/2	0.115*	1.248		3.228	1.008	1.099
1/4	0.095*	0.535	0.093	1.635	0.439	0.686
1/8	0.064*	0.270	0.071	0.511	0.261	0.392
1/16	0.014*	0.100*	0.053*	0.051*	0.073*	0.243*
1/32	0.087*	0.007*	-0.003*	-0.050*	-0.015*	-0.013*
Δ	$K = 70$					
1	1.258	3.188		7.065	2.872	2.676
1/2	0.511	1.790		3.776	1.274	1.334
1/4	0.285	0.796	0.769	2.017	0.569	0.805
1/8	0.177	0.418	0.406	0.822	0.339	0.426
1/16	0.042*	0.181	0.226	0.316	0.115*	0.229*
1/32	0.134*	0.050*	-0.030*	0.185*	0.001*	-0.046*
Δ	$K = 140$					
1	0.132*	1.050		5.219	1.882	1.886
1/2	-0.027*	0.656		2.545	0.717	0.894
1/4	0.005*	0.258	0.008	1.138	0.263	0.570
1/8	0.009*	0.118	0.003*	0.082*	0.139*	0.339
1/16	-0.007*	0.022*	0.002*	-0.109*	-0.006*	0.213*
1/32	0.054*	-0.036*	0.000*	-0.188*	-0.057*	0.007*

Table 5.4: Convergence of the MC simulation to the PDE prices with same initial values as in Table 5.3. The star behind a result indicates that the PDE price is within the confidence interval of three standard deviations.

Case	v_0	r_0	max error	S_i	price at S_i
A	0.12	0.04	0.019	133.4	40.954
	0.12	0.13	0.019	129.0	43.237
	0.32	0.04	0.024	138.5	47.716
	0.32	0.13	0.024	138.5	53.646
E	0.04	0.03	0.324	70.3	24.839
	0.04	-0.01	0.418	71.5	19.719
	0.22	0.03	0.528	70.3	28.672
	0.22	-0.01	0.684	71.5	23.984

Table 5.5: Location of the maximum error for different initial values v_0 and r_0.

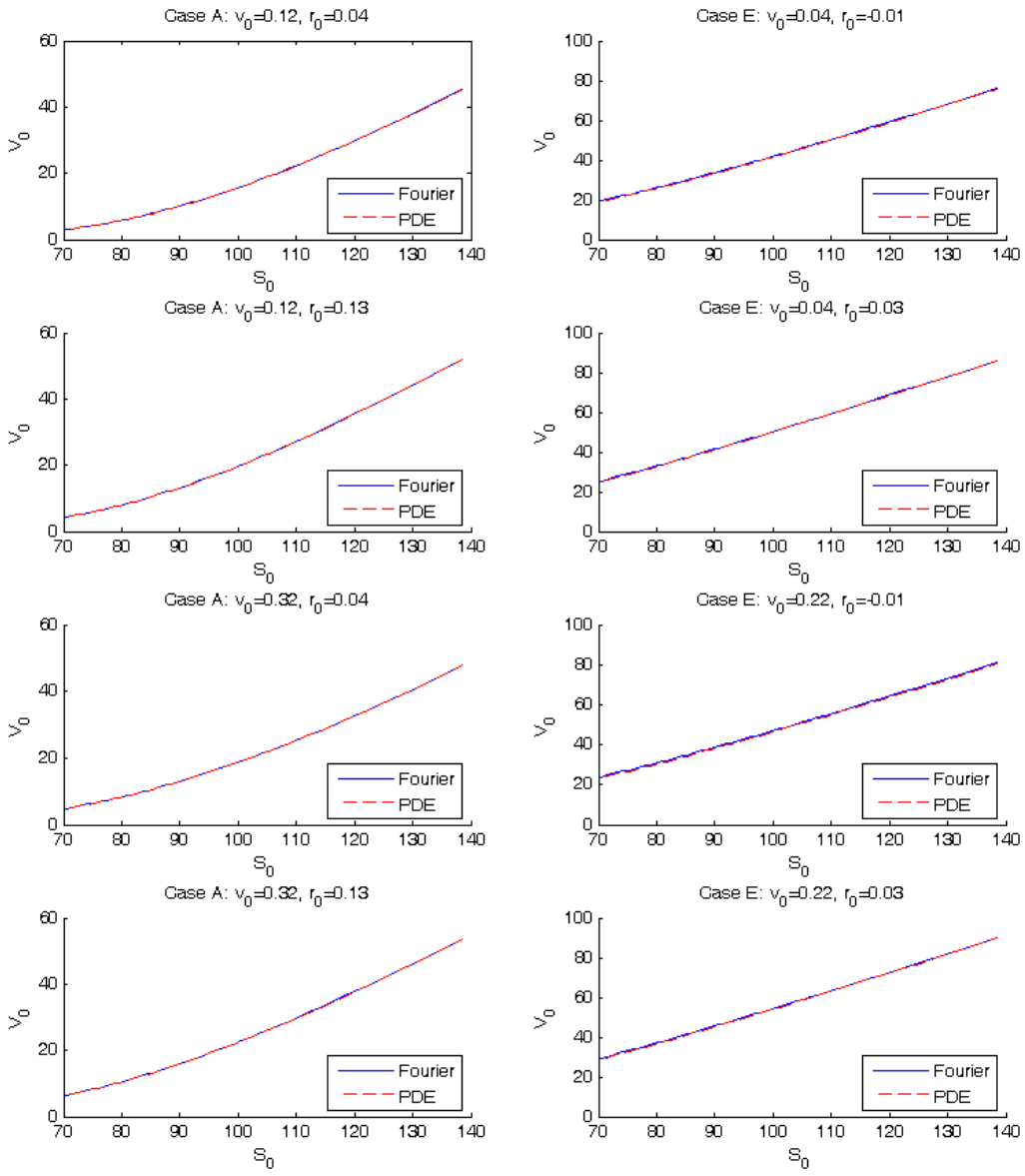

Figure 5.1: Comparison between Fourier and PDE method. The plots on the left side show results for case A (Feller condition satisfied) and the plots on the right side show case E (Feller condition not satisfied). We used again a grid of 100/40/40 in the (s, v, r) direction with 100 time steps a year.

5.2 Consistency studies for autocallables

From the above numerical studies we saw that the prices for European call options are consistent within all three pricing methods. Hence, we can savely say that their implementations are correct. Lets now take a look on the more complex autocallable structured products. Of course the Fourier method drops out here and the PDE scheme and MC simulation are the remaining pricing methods under consideration. We will use again the six parameter sets from Table 5.1, except for the parameters defining the payoff structure, i.e. the strike K and the maturity T. Instead, recall the structure for the autocallables introduced in Figure 1.1 from Chapter 1. For the numerical tests we will use the following set of observation days T_i, barriers K_i and payoffs P_i (in EUR), for $i = 1, \ldots, 4$, respectively,

16-Feb-2015,	15-Feb-2016,	14-Feb-2017,	14-Feb-2018,
100%,	90%,	80%,	65%,
1053.5,	1107.0,	1160.5,	1214.0,

and as valuation day t the 14-Feb-2014. Further, the issue price of the product is 1000EUR per unit (share).

In light of the already implemented routines for call options only minor changes need to be done for the autocallables, namely initial and boundary conditions. The grid in the price dimension S is generated according to Section 4.2.2 in order to concentrate around the barrier levels. Additionally the time step is reduced substantially around the observation days. During the course of the year we put grid points only quarterly, whereas all other lie symmetrically in a 1/10-th a year interval around those days. The intention of that non uniform grid is of course again, as for the call options before, to reduce numerical difficulties. To visualize the grid spacings, Figure 5.2 shows the surface of autocallable prices for all S and t, with initial $v_0 = \eta$ and $r_0 = b(0)$. It's easy to see that the grid lines concentrate around the discontinuities in the payoff structure.

We also tried some alternations on these grid spacings in the S and t dimensions but none of them seemed very beneficial. Especially in the time direction one might think that this rough grid spacing is not optimal and tend to a smoother one as for the price dimension. But doing so would cause heavy losses in efficiency because for every change in the time step all the LU factorizations for the ADI time discretization would have to be recalculated again and again, slowing down the whole routine. This way we only need to do this twice a year and tests showed that numerical accuracy does not suffer from that. Haentjens and 't Hout [18] also suggest implementing a damping

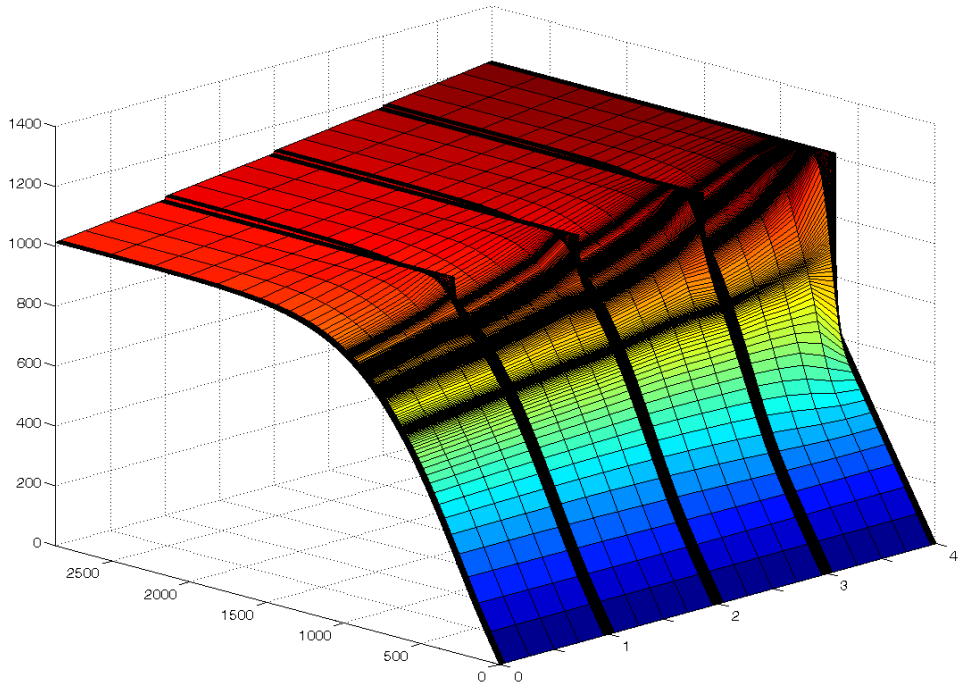

Figure 5.2: Surface of autocallable prices for all initial prices S and valuation days t with fixed initial volatility and interest rate.

procedure right after the points of discontinuity. They use it for the numerically very tough up-and-out call options, where the payoff vanishes above a predefined barrier level. The idea behind such a damping procedure is that instead of performing one full time step, from this point of discontinuity, they do two sub steps of size $\Delta t/2$ using the Douglas scheme (4.2.12) from Section 4.2.4 before continuing with the more sophisticated Craig-Sneyd (4.2.13) or Modified Craig-Sneyd (4.2.14) schemes. This flattens out possible oscillations faster. We also did some experiments with damping after every observation day, however the improvements were not really significant.

To test the consistency (and also convergence), we calculated MC prices for a refining number of time steps a year, each with a sufficient[5] amount of $5 \cdot 10^5$ paths. They are compared to the PDE prices for which the number of grid points is 400/20/20/100 in the S, v, r and t dimension respectively. These numbers were chosen irrespective of the run time but with focus to a highest possible accuracy. The PDE and MC results for all six test cases can

[5]In the sense that the resulting MC confidence interval is sufficient small.

		Case A	Case B	Case C	Case D	Case E	Case F
PDE		886.396	925.421	903.783	928.327	999.445	928.521
	1	14.059	4.203	-3.840	-6.664	-3.697	0.735*
	1/2	6.054	2.782	-1.532	-2.412	-0.914	-0.067*
	1/4	2.156	1.789	-0.965	-1.557	-0.359*	0.376*
MC	1/8	1.834	1.011	0.284*	-0.903	-0.562*	0.266*
	1/16	1.274	1.412	0.145*	-0.674	-0.273*	0.648*
	1/32	1.106*	-0.056*	0.350*	-0.630*	-0.061*	0.804*
	1/64	0.467*	0.823*	0.656*	-0.164*	0.246*	0.261*

Table 5.6: Convergence of the MC simulation to the PDE prices with same initial values as in Table 5.3. The star behind a result indicates that the PDE price is within the MC confidence interval of three standard deviations.

be found in Table 5.6 below. Using 32 steps a year we already obtain highly accurate prices with a bias less than 10^{-3}. They all lie within the confidence interval of three standard deviations. For the cases C, E and F even a much lower step size gives satisfying results. Also the convergence rate of the bias obtained by these, and other tested scenarios, seems to be higher than one, and hence consistent with our observations for the call options before and those of Andersen [3] in the simpler Heston model.

Altogether we can say that both methods work very well. As mentioned in the preceding Chapters already, the PDE method is our preferred pricing tool, whereas the MC simulation is used rather as a backtest.

5.3 Calibration to market data

We have mentioned already in the beginning of this Chapter that calibrating a hybrid market model like the HHW one is performed in two steps. At first the Hull-White model is calibrated to caps, yielding a and σ_2. The mean reversion function of the short rate $b(t)$ is also given through the discount curve from the market used to calibrate the Hull-White model. Secondly those parameters are plugged into the HHW model which then is calibrated to European call (or put) option prices. The focus of this thesis lies in the second part. For the Hull-White parameters we were lucky to receive calibrated values from DEKA Bank, i.e. a mean reversion intensity of $a = 0.0874$ and a interest rate volatility of $\sigma_2 = 0.0102$. The discount curve including the cap price, to which the Hull-White model was calibrated, as well as the

implied volatility surface[6] for the equity market are given in Appendix D, also provided by DEKA. From the discount curve we can calculate the function of the short rates $b(t)$ via formula (4.1.8) as described in Section 4.1.2, which completes the Hull-White part. The results can be seen in Figure 5.3. Only the remaining parameters need to be calibrated anymore.

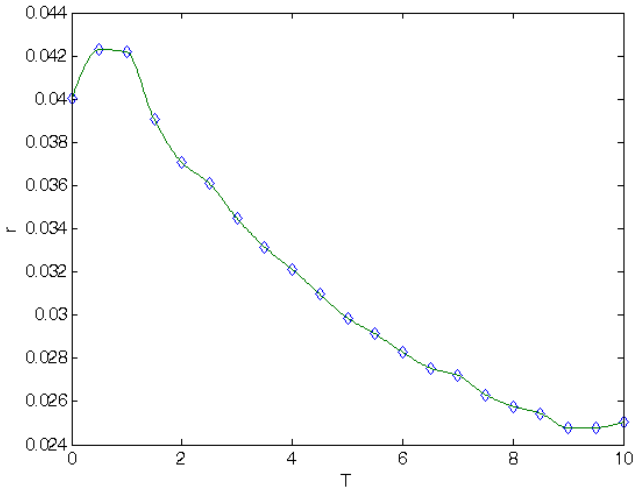

Figure 5.3: Short rates from discount curve

Calibration of course is a matter of optimizing an objective function with respect to some constraints on the parameters. When calibrating our model to given market data we want to find the parameter values which minimize the error made by our formula 3.3.10 compared to the markets given prices. The constraints on the parameters $\kappa, \eta, v_0, \sigma_1, a, \sigma_2, \rho_{Sv}$ and ρ_{Sr} are given as in Section 2.3.1, where we introduced the HHW model. For those only restricted (bounded) from one side we will set another bound on the other side, as can be seen in Table 5.7 below, to avoid restrictions that would make no sense in real life.

	κ	η	v_0	σ_1	ρ_{Sv}	ρ_{Sr}
Lower bound	e	e	e	0.1	-0.99	-0.99
Upper bound	20	2	2	3	0.99	0.99

Table 5.7: Lower and upper bounds for the parameters from the HHW model. The value 'e' was set to 10^{-5}.

[6]The corresponding option prices are of course the ones in the Black-Scholes sense.

5.3.1 Calibration of the pure Heston model

To compare the parameter estimates to a sophisticated model we also calibrated market data to the pure Heston model. Since we already implemented the Fourier method for previous tests to calculate Heston prices for European calls, this was pretty much no additional effort. For the calibration routine we used the caching technique from Section 4.3.6 in combination with the Gauss quadrature from the Appendix C to keep the number of evaluations of the ChF as low as possible. Further, as optimization solver we chose the Matlab function *fmincon*[7], which is an interior point algorithm. This one is supposed to work well for objective functions of the least squares type with smooth constraints. With a constant interest rate $r = 0.04$ and initial values $\{2.0, 0.05, 0.05, 0.3, -0.5\}$ for $\kappa, \eta, v_0, \sigma_1$ and ρ_{Sv} respectively, we got the following parameter estimates:

κ	η	σ_1	v_0	ρ_{Sv}	Est time	MSE
1.9750	0.0758	1.0514	0.0347	-0.7266	0.4733s	0.002413

Table 5.8: Parameter estimates for the pure Heston model.

Plugging those values, and the ones for the Hull-White part from before, into our PDE and MC pricing routines for the autocallable from Section 5.2 we get the prices 962.59 and 961.94, with three standard deviation confidence interval [961.14, 962.75], respectively.

5.3.2 Calibration of the Heston-Hull-White model

Like described before, now we are going to take the Hull-White parameters and plug them into the option valuation formula. It shall also be mentioned, that, in light of Section 4.3.6 we did some additional precaching to speed up the calibration, i.e. the expectation $\mathbb{E}\left(\sqrt{v_t}\right)$ from Lemmata 5 and 6 is independent of the strike K and hence can be evaluated beforehand. There are some other values which were precached, e.g. the forward price, but the before mentioned expectation is the most time saving.

The calibration was performed in two different ways. For the first, we fixed ρ_{Sr} and calibrated the model for a whole range $[-0.9, 0.9]$, with step size 0.1, of correlations to see how the mean squared error (MSE) behaves. The initial values for the optimization algorithm were $\{2.0, 0.05, 0.05, 0.3, -0.5\}$ for $\kappa, \eta, v_0, \sigma_1$ and ρ_{Sv} as before.

[7]Check the Matlab page http://de.mathworks.com/help/optim/ug/choosing-a-solver.html for further informations on the solver and other possible choices.

κ	η	σ_1	v_0	ρ_{Sv}	ρ_{Sr}	Est time	MSE
1.8527	0.0797	1.0358	0.0345	-0.7023	-0.90	4.9940s	0.002313
1.8536	0.0789	1.0125	0.0343	-0.7134	-0.80	3.1547s	0.002248
1.8210	0.0791	1.0055	0.0343	-0.7160	-0.70	3.0607s	0.002236
2.3292	0.0746	1.1793	0.0350	-0.6925	-0.60	6.3726s	0.002578
1.7601	0.0798	0.9979	0.0344	-0.7191	-0.50	2.7659s	0.002227
1.7301	0.0802	0.9943	0.0344	-0.7207	-0.40	2.2816s	0.002225
1.7005	0.0806	0.9907	0.0344	-0.7223	-0.30	2.1839s	0.002223
1.6712	0.0809	0.9872	0.0344	-0.7239	-0.20	2.7661s	0.002223
1.6421	0.0813	0.9837	0.0344	-0.7255	-0.10	3.5440s	0.002224
1.6134	0.0818	0.9803	0.0344	-0.7270	0.00	2.0958s	0.002225
1.5850	0.0822	0.9770	0.0344	-0.7286	0.10	2.8213s	0.002228
1.5569	0.0826	0.9738	0.0344	-0.7301	0.20	2.7206s	0.002232
1.5292	0.0831	0.9706	0.0344	-0.7317	0.30	2.4610s	0.002236
1.5019	0.0836	0.9677	0.0344	-0.7332	0.40	3.3126s	0.002242
1.4747	0.0840	0.9647	0.0344	-0.7347	0.50	2.9297s	0.002248
1.4481	0.0846	0.9620	0.0344	-0.7362	0.60	3.3604s	0.002255
2.0194	0.0749	1.1003	0.0350	-0.7327	0.70	3.2139s	0.002765
2.0208	0.0747	1.0991	0.0350	-0.7360	0.80	3.4318s	0.002807
2.0196	0.0745	1.0971	0.0350	-0.7389	0.90	3.5347s	0.002851

Table 5.9: Parameter estimates for a given range -0.9:0.1:0.9 of correlations ρ_{Sr} in the HHW model.

From Table 5.9 we can see that the MSE is slightly better when we assume a negative correlation ρ_{Sr}. The error gets worse with increasing correlation, whereas one outlier is encountered at $\rho_{Sr} = -0.6$. Note that wherever we obtain a relatively high MSE, the condition $8\kappa\eta/\sigma_1^2 \geq 1$ is not fulfilled. This might be due to numerical difficulties that arise when we are forced to use formula 4.3.8 instead of the simpler approximation 4.3.9 for the expectation of the volatility. Also the much higher runtime in those cases suggests that the optimization algorithm more often jumps to parameter constellations where the above condition is violated.

Now, the other way we calibrated the remaining parameters is to also let the algorithm optimize the second correlation factor. We can clearly see that the minimum of the MSE in the above results must be somewhere between -0.2 and -0.3. That is why, for the next calibration run, we used an initial ρ_{Sr} of -0.25.

As expected, the algorithm converged to the region where the MSE was the lowest in the previous experiment, yielding a correlation of $\rho_{Sr} =$

κ	η	σ_1	v_0	ρ_{Sv}	ρ_{Sr}	Est time	MSE
1.6759	0.0809	0.9877	0.0344	-0.7236	-0.2153	4.5947s	0.002223

Table 5.10: Parameter estimates in the HHW model.

-0.2153. Naturally the runtime is higher due to this additional degree of freedom for the optimization routine. The results can be seen in Table 5.10. In the following Table 5.11 we compared the prices for the autocallable when using those different sets of parameter estimates to see how big the gap between them is.

	PDE	MC	MCl	MCu
Heston params	962.73	962.18	961.37	962.99
HHW params	961.47	960.96	960.14	961.77

Table 5.11: Comparison of prices for the autocallables with Heston and HHW parameter estimates.

The fact that the parameter estimates from Tables 5.8 and 5.10, for the Heston and HHW model respectively, are very similar, also reflects in the autocallable's prices. Their MC confidence intervalls even overlap. Unfortunately we didn't have more sets of market data to analyse this matter further. Finally, we also observe, that it is save for a Bank to issue such a product. The contract we received states that the product is issued for 1000EUR plus additional charges of 30EUR. This results in an expected profit margin of about 7%.

Chapter 6

Conclusion

Within this thesis, we have discussed the pricing of autocallable structured products in the hybrid Heston-Hull-White model with correlated stochastic interest rates. Due to the complexity of these products we used PDE and MC methods to do so. We also calibrated the model to market data using highly efficient Fourier techniques.

Regarding the pricing techniques, we implemented a PDE method based on finite differences and alternate direction implicit time stepping as used in Haentjens and 't Hout [18]. Even in case of a three dimensional convection diffusion equation, which we had to solve, those methods work very well and therefore are superior to MC simulation. Nonetheless, we have also implemented an efficient MC method based on the algorithms from Andersen [3] and an exact simulation of the interest rate process. Due to the huge computational demand, the use of the MC method was more as a reference, whereas the PDE method is the primary pricing tool. For the calibration we used a, to our needs adapted, Fourier style formula introduced in Lewis [30] to calculate European call prices. The correlated stochastic interest rates demanded a change of measure, to the T-forward measure, to be able to use that technique.

During our computational studies we found that all implemented methods work very well and give consistent results. To verify the correctness of our Fourier formula we additionally implemented the PDE and MC methods to price European call options. All three methods were compared using a set of challenging parameters which cover different market scenarios. Consequently, we also calculated prices for autocallables and compared the PDE solutions to the MC confidence interval. Finally, to calibrate our model, we received a set of market data from DEKA bank, i.e., an implied volatility surface for a various number of strike prices and maturities and a discount curve. Calibration of a hybrid model usually works in two steps. The first one,

which is calibrating the interest rate process to caps, was not part of our work. Hence, we were also lucky to receive some calibrated parameter values. In the second step, the rest of the parameters is calibrated. On one side, we plugged the already known parameters into the Heston-Hull-White model to get the remaining ones. On the other side, also the pure Heston model was calibrated, to compare both parameter estimates. We found that they were only slightly different, which expectedly lead to very similar autocallable prices. To analyse the effects, of correlated stochastic interest rates, on the price better, various sets of market data and more tests are essential.

Finally, and because it would have gone beyond the constraints of this thesis, we want to present some thoughts for interesting future work. Naturally, having various sets of market data would be an immediate extension of this work in order to analyse the effects of using the more general hybrid Heston-Hull-White model compared to, e.g., the Heston model. Oosterlee and Grzelak [15] present more ways of approximating the model in order to derive semi closed solutions. Even though they find that the deterministic approximation is not inferior to more advanced (stochastic) ones. Contrary to that, one could also try to abandon the stochastic interest rate and imply merely a non constant but deterministic one, which reflects the term structure. Such a model would be much easier to handle and maybe only a negligible drawback compared to the other one. Concerning the efficiency of the PDE and MC methods, it would be definitely worth considering to implement the forward dynamics to reduce the number of dimensions to two. Especially for the PDE method this should bring substantial performance.

Appendix A

Fundamental results

For completeness reasons we will state some fundamental results concerning stochastic differential equations and probability theory, following the lines of Karatzas & Shreve [26, Ch. 5] and Oksendal [38, Ch. 5,7 and 8].

A.1 Existence and uniqueness of strong and weak solutions to SDEs

There are two concepts of solutions to stochastic differential equations with respect to Brownian motion, namely weak and strong ones. Also the existence and uniqueness of such solutions will be discussed.

Let $b_i(t,x)$ and $\sigma_{ij}(t,x)$ for $1 \leq i \leq d, 1 \leq j \leq r$, from $[0,\infty) \times \mathbb{R}^d \to \mathbb{R}$ be Borel-measurable functions. Define the d-dimensional drift vector $b(t,x) = \{b_i(t,x)\}_{1\leq i\leq d}$ and $(d \times r)$-dimensional dispersion matrix $\sigma(t,x) = \{\sigma_{ij}(t,x)\}_{1\leq i\leq d, 1\leq j\leq r}$. The intention is to give the stochastic differential equation

$$\mathrm{d}X_t^i = b_i(t,X_t)\mathrm{d}t + \sum_{j=1}^{r} \sigma_{ij}(t,X_t)\mathrm{d}W_t^j,\ 1 \leq i \leq d,$$

or written compact as

$$\mathrm{d}X_t = b(t,X_t)\mathrm{d}t + \sigma(t,X_t)\mathrm{d}W_t, \tag{A.1.1}$$

where $W = \{W_t; 0 \leq t < \infty\}$ is an $r-$dimensional Brownian motion, a meaning. A solution $X = \{X_t; 0 \leq t < \infty\}$, of this equation, is a suitable stochastic process with continuous sample paths and values in $\mathbb{R}^d$ and coefficients $b(t,x)$ and $\sigma(t,x)$; the matrix $a(t,x) = \sigma(t,x)\sigma^T(t,x)$ is called diffusion matrix.

For the concept of a strong solution we fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a r-dimensional Brownian motion $W = \{W_t, \mathcal{F}_t^W; 0 \leq t < \infty\}$ on it with its natural filtration $\{\mathcal{F}_t^W\}_{t\geq 0}$ and $\mathcal{F}_t^W = \sigma\{W_s, 0 \leq s \leq t\}$. An initial value X_0 is a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ independent of the Brownian motion with distribution μ, $\mu(A) = \mathbb{P}(X_0 \in A)$ for $A \in \mathcal{B}(\mathbb{R}^d)$. The natural filtration $\{\mathcal{F}_t^W\}_{t\geq 0}$ is extended by

$$\mathcal{G}_t \triangleq \sigma(x_0) \vee \mathcal{F}_t^W = \sigma(x_0, W_s, 0 \leq s \leq t),$$

and completed with the $\mathbb{P}$-Null sets (since there are Null sets which aren't in the σ-algebra)

$$\mathcal{N} \triangleq \{N \subseteq \Omega : \exists G \in \mathcal{G}_\infty \text{ with } N \subseteq G \text{ and } \mathbb{P}(G) = 0\},$$

to the augmented filtration $(\mathcal{F}_t)_{t\geq 0}$ given by

$$\mathcal{F}_t \triangleq \sigma(\mathcal{G}_t \cup \mathcal{N}),\ 0 \leq t < \infty; \quad \mathcal{F}_\infty \triangleq \sigma\left(\bigcup_{t\geq 0} \mathcal{F}_t\right). \tag{A.1.2}$$

It is obvious that $\{W_t, \mathcal{G}_t; 0 \leq t < \infty\}$ is a r-dimensional Brownian motion and hence, $\{W_t, \mathcal{F}_t; 0 \leq t < \infty\}$ is one with the filtration $\{\mathcal{F}_t\}$ satisfying the usual conditions.

Definition 15 (strong solution)**.** *[26, Def. 2.1] A strong solution to the SDE (A.1.1), on the given probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with respect to the fixed Brownian motion $(W_t)_{t\geq 0}$ and initial value x_0, is a process $X = (X_t)_{t\geq 0}$ with continuous sample paths and*

(i) *X is adapted to the filtration $(\mathcal{F}_t)_{t\geq 0}$,*

(ii) *$X_0 = x_0$ $\mathbb{P}$-a.s.,*

(iii) *$\mathbb{P}\left(\int_0^t |b_i(s, X_s)| + \sigma_{ij}^2(s, X_s)ds < \infty\right) = 1$ holds for all $1 \leq i \leq d$, $1 \leq j \leq r$, $0 \leq t < \infty$, and*

(iv) *$X_t = X_0 + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s$ $\forall 0 \leq t < \infty$ almost surely.*

Remark: The crucial point here is that the version of the Brownian motion W_t is given in advance and the solution constructed from it is $(\mathcal{F}_t)_{t\geq 0}$ adapted. This is expressed in (i) and meets an intuitive understanding of X as the output of a dynamical system described by b and σ, whose input are the Brownian motion W and initial value x_0.

Definition 16 (strong uniqueness)**.** *[26, Def. 2.3] Let the coefficients (b, σ), a r-dimensional Brownian motion W on some probability space $(\Omega, \mathcal{F}, I\!P)$ and independent d-dimensional start vector X_0 be given. Further the filtration is given as in (A.1.2). Suppose that X and $\tilde{X}$ are two strong solutions of (A.1.1) with respect to W and initial value X_0, then $I\!P(X_t = \tilde{X}_t; 0 \leq t < \infty) = 1$ and we say that strong uniqueness holds for (b, σ).*

Theorem 8 (strong uniqueness)**.** *[26, Th. 2.5] Let the coefficients $b(t, x), \sigma(t, x)$ be locally Lipschitz continuous in the space variable, i.e., for every $n \geq 1$ there exists a constant $K_n > 0$ such that $\forall t \geq 0$, $\|x\| \leq n$ and $\|y\| \leq n$. Then if*

$$\|b(t, x) - b(t, y)\| + \|\sigma(t, x) - \sigma(t, y)\| \leq K_n \|x - y\|,$$

strong uniqueness holds for (A.1.1).

Theorem 9 (existence of a strong solution)**.** *[26, Th. 2.9] Suppose that, for every $t \geq 0$, $x, y \in \mathbb{R}^d$ and positive constant K, the coefficients satisfy the global Lipschitz and linear growth conditions*

$$\begin{aligned} \|b(t, x) - b(t, y)\| + \|\sigma(t, x) - \sigma(t, y)\| &\leq K \|x - y\|, \\ \|b(t, x)\|^2 + \|\sigma(t, x)\|^2 &\leq K^2(1 + \|x\|^2), \end{aligned}$$

on some probability space $(\Omega, \mathcal{F}, I\!P)$ with $\{\mathcal{F}_t\}$ as in (A.1.2). Further, let x_0 be an $\mathbb{R}^d$ valued random vector, independent of the r-dimensional Brownian motion $W = \{W_t, \mathcal{F}_t^W; 0 \leq t < \infty\}$, with finite second moment

$$\mathbb{E} \|x_0\|^2 < \infty.$$

Then there exists a strong solution of equation (A.1.1) with respect to W and X_0 which is a continuous, adapted process $X = \{X_t, \mathcal{F}_t; 0 \leq t < \infty\}$.

A weaker notion of solvability, under which existence can be shown for much milder assumptions on the drift term, is presented in the following definition. Given only the coefficients $b(t, x)$ and $\sigma(t, x)$, and asking for a pair (X_t, W_t) on some filtration $\mathcal{F}_t$ and probability space $(\Omega, \mathcal{F}_t, \mathbb{P})$ leads to the so called weak solution. The uniqueness attached to it will lead naturally to the strong Markov property of the solution process.

Definition 17 (weak solution)**.** *[26, Def. 3.1] A weak solution to (A.1.1) is a triple (X, W), $(\Omega, \mathcal{F}, I\!P)$, $\{\mathcal{F}_t\}$, where*

(i) $(\Omega, \mathcal{F}, I\!P)$ is a probability space, $\{\mathcal{F}_t\}$ a filtration of sub-σ-fields of $\mathcal{F}$ ($\mathcal{F}_t \subset \mathcal{F}\, \forall t \geq 0$) satisfying the usual conditions,

(ii) $X = \{X_t, \mathcal{F}_t; 0 \leq t < \infty\}$ *is a continuous, adapted* $\mathbb{R}^d$*-valued process,* $W = \{W_t, \mathcal{F}_t; 0 \leq t < \infty\}$ *a* r*-dimensional Brownian motion and*

(iii) the integrability conditions (iii) *and* (iv) *from Definition 15 are satisfied.*

The filtration $\{\mathcal{F}_t\}$ is not necessarily the augmentation of $\mathcal{G}_t = \sigma(X_0) \vee \mathcal{F}_t^W$, $0 \leq t < \infty$, thus, the value of the solution $X_t(\omega)$ is not necessarily given by a measurable functional of the Brownian path $\{W_s(\omega); 0 \leq s \leq t\}$ and initial value $X_0(\omega)$. Consequently, the existence of a weak solution (X, W), $(\Omega, \mathcal{F}, \mathbb{P})$, $\{\mathcal{F}_t\}$ is no guarantee, given another Brownian motion $\{\widetilde{W}_t, \widetilde{\mathcal{F}}_t; 0 \leq t < \infty\}$ on an eventually different probability space $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}})$, for the existence of another weak solution $(\widetilde{X}, \widetilde{W})$, $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}})$, $\{\mathcal{F}_t\}$.

Definition 18 (pathwise uniqueness)**.** *[26, Def. 3.2] Suppose that both* (X, W), $(\Omega, \mathcal{F}, I\!P)$, $\{\mathcal{F}_t\}$ *and* $(\widetilde{X}, W)$, $(\Omega, \mathcal{F}, I\!P)$, $\{\widetilde{\mathcal{F}_t}\}$ *are weak solutions to (A.1.1) with common probability space and Brownian motion relative to a possibly different filtration and same initial value, i.e.* $I\!P(X_0 = \widetilde{X}_0) = 1$*. Whenever* X *and* $\widetilde{X}$ *are indistinguishable, i.e.* $I\!P(X_t = \widetilde{X}_t; \forall 0 \leq t < \infty) = 1$*, we say that pathwise uniqueness holds for (A.1.1).*

Remark: The strong uniqueness results from Theorems 8 and 9 also hold for the pathwise uniqueness since their proofs do not take advantage of the special form of the filtration for a strong solution.

Definition 19 (uniqueness in probability law)**.** *[26, Def. 3.4] If for two solutions* $(X, W), (\Omega, \mathcal{F}, I\!P), (\mathcal{F}_t)$ *and* $(\widetilde{X}, \widetilde{W}), (\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{I\!P}), (\widetilde{\mathcal{F}}_t)$*, with same initial distribution,* X *and* $\widetilde{X}$ *have the same law, we say that uniqueness in the sense of probability law holds for (A.1.1), i.e. they have the same finite dimensional distributions.*

Proposition 7 (Yamada & Watanabe)**.** *[26, Prop. 3.20] Uniqueness in probability law is implied by pathwise uniqueness.*

This means that under the assumptions of Theorem 8 a solution, weak or strong, is unique in probability law (weakly unique). From a modeling point of view, this result is very useful, because the weak solution concept, which does not specify the explicit representation of the Brownian motion beforehand, is more natural. Solving a SDE in the weak sense is equivalent to finding the law of a diffusion process in terms of the martingale problem developed by Stroock & Varadhan [45]; and further leads to the important Markov property.

Definition 20 (Itô diffusion). *[38, Def. 7.1.1] A stochastic process satisfying*

$$dX_t = b(X_t)\,dt + \sigma(X_t)\,dW_t, \quad t \geq s,\ X_s = x;$$

and the conditions in Theorem 8 is called a (time-homogeneous) Itô diffusion. (Note that the general case where the coefficients also depend on the time t can be reduced to this situation [38, Ch. 10,11])

Theorem 10 (generator of an Itô diffusion). *[38, Def. 7.3.1 and Th. 7.3.3] The generator of an Itô diffusion, defined as*

$$Af(x) = \lim_{t \to 0} \frac{\mathbb{E}^x(f(X_t)) - f(x)}{t}, \quad x \in \mathbb{R}^n,$$

for $f \in C_0^2(\mathbb{R}^n)$ is

$$\mathcal{A}f(x) = \sum_{i=1}^{d} b_i(x) \frac{\partial}{\partial x_i} f(x) + \sum_{i=1}^{d} \sum_{j=1}^{k} a_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j} f(x), \tag{A.1.3}$$

with $a = \sigma\sigma^T$.

A.2 Martingale problem and Markov property

Theorem 11 (Martingale problem). *[38, Th. 8.3.1] Suppose X_t is an Itô diffusion with generator $\mathcal{A}$. Then for every $f \in C_0^2(\mathbb{R}^n)$ the process*

$$M_t = f(X_t) - \int_0^t \mathcal{A}f(X_s)ds$$

is a Martingale w.r.t $\mathcal{M}$ generated by M_t.

Theorem 12 (Markov property for Itô diffusions). *[38, Th. 7.1.2] For a bounded Borel function f from $\mathbb{R}^n$ to $\mathbb{R}$ and $t, h \geq 0$ we have:*

$$\mathbb{E}^x \left(f(X_{t+h})|\mathcal{F}_t\right)_{(\omega)} = \mathbb{E}^{X_t(\omega)}(f(X_h)). \tag{A.2.1}$$

A relation of the form (A.2.1) continues to hold if t is replaced by a stopping time.

Theorem 13 (strong Markov property for Itô diffusions). *[38, Th. 7.2.4] Let f be as before, a stopping time $\tau < \infty$ a.s. and $h \geq 0$. Then*

$$\mathbb{E}^x(f(X_{\tau+h})|\mathcal{F}_\tau) = \mathbb{E}^{X_\tau}(f(X_h)). \tag{A.2.2}$$

A.3 Girsanov change of measure

Girsanov's theorem has an outstanding meaning in financial mathematics. It provides the base for switching from the "real life" measure $\mathbb{P}$ to the risk neutral pricing measure $\mathbb{Q}$.

Theorem 14 (Girsanov change of measure)**.** *Let $B = (B^1, \dots, B^d)$ be a $\mathbb{P}$ d-dimensional Brownian motion and $Y^i \in L^1(B^i)$. Let*

$$Z_s = \exp\left(\sum_{i=1}^{d} \int_0^s Y_u^i dB_u^i - \frac{1}{2}\sum_{i=1}^{d} \int_0^s (Y_u^i)^2 du\right)$$

be a $\mathbb{P}$-martingale. Then $\frac{d\mathbb{Q}_t}{d\mathbb{P}}(B) = Z_t$ defines an equivalent measure $\mathbb{Q}$ and the process

$$W_s = B_s - \int_0^s Y_u du,$$

is a $\mathbb{Q}$-Brownian motion.

Remark: The condition $Y \in L^1(B)$ is equivalent with Y being predictable and $\int_0^t Y_s^2 ds < \infty$ $\mathbb{P}$-a.s. for all $t \geq 0$. Under the Novikov condition

$$\mathbb{E}\left(\exp\left(\frac{1}{2}\sum_{i=1}^{d} \int_0^t (Y_u^i)^2 du\right)\right) < \infty$$

$(Z_s)_{s \leq t}$ is a martingale.

Appendix B

Residue calculus

In order to retrieve the preferred formula 3.3.10 via contour variations we have to apply residue calculus. The residue Theorem and some basic definitions are recalled, e.g. from Rudin [40] or Jaenich [25], in the following lines.

Definition 21 (holomorphic functions)**.** *[40, Def. 10.2] Let f be a complex function defined in Ω. If, for $z_0 \in \Omega$, the limit*

$$\lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0},$$

exists, we denote it by $f'(z_0)$ and call it the derivative of f at z_0. A function f is called holomorphic (or analytic) in Ω, if the derivative exists for every $z_0 \in \Omega$. This class of functions will be denoted by $H(\Omega)$.

Remark: $H(\Omega)$ is a ring, i.e. if $f, g \in H(\Omega)$, then also $f + g \in H(\Omega)$ and $fg \in H(\Omega)$. The usual differentiation rules from real analysis also apply here. Further, a function which is analytic and single valued in a region is called regular.

Definition 22 (regular and singular points)**.** *[40, Def. 16.1] Let D be an open circular disc, suppose $f \in H(D)$, and let β be a boundary point of D. We call β a regular point of f if there exists a disc D_1 with center β and a function $g \in H(D_1)$ such that $g(z) = f(z)$ for all $z \in D \cap D_1$. Any boundary point of D which is not a regular point is called a singular point of f.*

Definition 23 (isolated singularity)**.** *[40, Def. 10.19] A function f is said to have an isolated singularity at $a \in \Omega$ if $f \in H(\Omega - \{a\})$. The singularity is called removable if f can be so defined at a that the extended function is holomorphic in Ω.*

Definition 24 (meromorphic function)**.** *[40, Def. 10.41] Let Ω be an open set and $A \subset \Omega$. A function f is called meromorphic if*

(i) A has no limit point in Ω,

(ii) $f \in H(\Omega - A)$,

(iii) f has a pole at each point of A.

Note that every $f \in H(\Omega)$ is also meromorphic in Ω since $A = \emptyset$ is not excluded in this definition and by (i) A is at most countable.

A holomorphic function is not developable into a power series around an isolated singularity in general, although it is in a so called Laurent series.

Definition 25 (Laurent series)**.** *[25, Sec. 4.3] A Laurent series around a is a series of the form*

$$\sum_{n=-\infty}^{\infty} c_n (z-a)^n, \tag{B.0.1}$$

or more specific speaking the pair

$$\sum_{n=1}^{\infty} c_{-n} (z-a)^{-n} \quad \textit{and} \quad \sum_{n=0}^{\infty} c_n (z-a)^n,$$

where the first one is called principal part.

Theorem 15 (index)**.** *[40, Th. 10.10] Let γ be a closed path and Ω the compliment of γ^* (relative to the plane). Define*

$$\mathrm{Ind}_\gamma(z) = \frac{1}{2\pi i} \int_\gamma \frac{d\xi}{\xi - z} \quad z \in \Omega,$$

the index of z with respect to γ. Then Ind_γ is an integer valued function on Ω which is constant in each component of Ω and 0 in the unbounded component of Ω.

Definition 26 (residue)**.** *Lets say the holomorphic function f has an isolated singularity at a. For sufficient small $\epsilon > 0$ the residue of f at a is*

$$\mathrm{Res}_a f(z) := \frac{1}{2\pi i} \int_{|z-a|=\epsilon} f(z) dz. \tag{B.0.2}$$

Developing f around a into a Laurent series, only the coefficient c_{-1} contributes to its value. The residue is therefore also the -1-th coefficient of the Laurent series of f in a circular disc around a.

Theorem 16 (residue theorem)**.** *[40, Th. 10.42] Let f be meromorphic in Ω and A the set of points in Ω at which f has poles. If Γ is a cycle in $\Omega - A$ such that* $\mathrm{Ind}_\Gamma(\alpha) = 0$ for all $\alpha \neq \Omega$, then

$$\frac{1}{2\pi i}\int_\Gamma f(z)dz = \sum_{a\in A} \mathrm{Ind}_\Gamma(a)\,\mathrm{Res}_a f.$$

Residue calculus

If $f(z)$ has a pole of order k at z_0, then

$$\mathrm{Res}_{z_0} f(z) = \frac{1}{(k-1)!}\frac{\mathrm{d}^{(k-1)}}{\mathrm{d}z^{(k-1)}}\,|_{z_0}(\,z - z_0)^k f(z).$$

Corollary 3 (residue in the case of simple denominator zeros)**.** *[25, p. 78] If $g(z)$ and $h(z)$ are holomorphic at z_0 and $h(z)$ has a simple zero there, then*

$$\mathrm{Res}_{z_0}\frac{g(z)}{h(z)} = \frac{g(z_0)}{h'(z_0)}.$$

With the residue Theorem 16 and the last Corollary 3 we can obtain our preferred option valuation formula in the end of Section 3.3.4.

Appendix C

Numerical integration

The Gauss quadrature is a very efficient way of approximating integrals of the following form:

$$\int_a^b w(x)f(x)dx \approx \sum_{i=1}^{n} w_i f(x_i),$$

where $w(x)$ is a non-negative weight function and the x_i and w_i are the abscissae and weights. For our option valuation formula this means that we have a weight function

$$w(x) = e^{-x},$$

and the interval $[a, b] = [0, \infty]$. From Stoer [44, Sec. 3.6] we know that the x_i and w_i are therefore given through the Laguerre polynomials

$$L_n(x) = \sum_{j=0}^{n} (-1)^j \frac{(n!)^2}{(j!)^2(n-j)!} x^j,$$

with x_i being the i-th root of $L_n(x)$ and the weights are

$$w_i = \frac{x_i}{(n+1)^2[L_{n+1}(x_i)]^2}.$$

From Theorem [44, Th. 3.6.24] we have, for $f \in C^{2n}[a, b]$, that the error of the above integration rule is

$$\int_a^b w(x)f(x)dx - \sum_{i=1}^{n} w_i f(x_i) = \frac{f^{(2n)}(\xi)}{(2n)!} < L_n, L_n >,$$

with $\xi \in (a, b)$ and the scalar product for a real function g is given through

$$< g, g >= \int_a^b w(x)g(x)^2 dx.$$

Appendix D

Market data

SP500 Volatility Market Data

```
MktIV = [ ...
          27.8000   26.3800   25.3200   25.1800
          24.7700   24.0200   23.6400   23.6900
          21.8600   21.5800   22.0300   22.3900
          18.7800   19.3000   20.4700   20.9800
          15.7200   17.1200   18.9400   19.7000
          13.3400   15.1700   17.4800   18.4900
          13.2300   13.7300   16.1800   17.3600 ] ./ 100;
T = [45 98 261 348] ./ 365;
K = [120   125   130   135   140   145   150];
```

EUR discount curve and market price for cap

```
PV_disc = [ ...
            0.99850    0.98495   0.96299   0.93648   0.90790
            0.99620    0.98007   0.95667   0.92946   0.90062
            0.99310    0.97474   0.95012   0.92234   0.89333
            0.98931    0.96903   0.94338   0.91514   0.88604 ];
dt = 0.5;

MarketPR = 650;  % in bps
```

Bibliography

[1] Hansjörg Albrecher, Philipp Mayer, Wim Schoutens, and Jurgen Tistaert. The little heston trap. *WILMOTT*, 6:83–92, 2007.

[2] Leif B. G. Andersen and Vladimir V. Piterbarg. Moment explosions in stochastic volatility models. *Finance Stoch.*, 11(1):29–50, 2007.

[3] Leif BG Andersen. Efficient simulation of the heston stochastic volatility model. *Available at SSRN 946405*, 2007.

[4] Mukarram Attari. Option pricing using fourier transforms: A numerically efficient simplification. *Available at SSRN 520042*, 2004.

[5] David S Bates. Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options. *Review of financial studies*, 9(1):69–107, 1996.

[6] Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *The journal of political economy*, pages 637–654, 1973.

[7] Peter Carr and Dilip Madan. Option valuation using the fast fourier transform. *Journal of computational finance*, 2(4):61–73, 1999.

[8] Kyriakos Chourdakis. Option pricing using the fractional fft. *Journal of Computational Finance*, 8(2):1–18, 2004.

[9] Rama Cont and Peter Tankov. *Financial modelling with jump processes*. Chapman & Hall/CRC Financial Mathematics Series. Chapman & Hall/CRC, Boca Raton, FL, 2004.

[10] John C Cox, Jonathan E Ingersoll Jr, and Stephen A Ross. A theory of the term structure of interest rates. *Econometrica: Journal of the Econometric Society*, pages 385–407, 1985.

[11] Freddy Delbaen and Walter Schachermayer. A general version of the fundamental theorem of asset pricing. *Mathematische annalen*, 300(1):463–520, 1994.

[12] Darrell Duffie, Jun Pan, and Kenneth Singleton. Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 68(6):1343–1376, 2000.

[13] Daniel J Duffy. *Finite Difference methods in financial engineering: a Partial Differential Equation approach*. John Wiley & Sons, 2013.

[14] Paul Glasserman. *Monte Carlo methods in financial engineering*, volume 53. Springer Science & Business Media, 2003.

[15] Lech A. Grzelak and Cornelis W. Oosterlee. On the Heston model with stochastic interest rates. *SIAM J. Financial Math.*, 2(1):255–286, 2011.

[16] Lech A Grzelak, Cornelis W Oosterlee, and Sacha Van Weeren. Extension of stochastic volatility equity models with the hull–white interest rate process. *Quantitative Finance*, 12(1):89–105, 2012.

[17] Shimin Guo, Lech A. Grzelak, and Cornelis W. Oosterlee. Analysis of an affine version of the Heston-Hull-White option pricing partial differential equation. *Appl. Numer. Math.*, 72:143–159, 2013.

[18] Tinne Haentjens and Karel J Hout. Adi finite difference schemes for the heston-hull-white pde. *arXiv preprint arXiv:1111.4087*, 2011.

[19] David Heath and Martin Schweizer. Martingales versus PDEs in finance: an equivalence result with examples. *J. Appl. Probab.*, 37(4):947–957, 2000.

[20] Steven L Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of financial studies*, 6(2):327–343, 1993.

[21] John Hull and Alan White. The pricing of options on assets with stochastic volatilities. *The journal of finance*, 42(2):281–300, 1987.

[22] John Hull and Alan White. Pricing interest-rate-derivative securities. *Review of financial studies*, 3(4):573–592, 1990.

[23] K. J. In 't Hout and S. Foulon. ADI finite difference schemes for option pricing in the Heston model with correlation. *Int. J. Numer. Anal. Model.*, 7(2):303–320, 2010.

[24] K. J. in 't Hout and B. D. Welfert. Unconditional stability of second-order ADI schemes applied to multi-dimensional diffusion equations with mixed derivative terms. *Appl. Numer. Math.*, 59(3-4):677–692, 2009.

[25] Klaus Jänich. *Funktionentheorie*. Springer-Lehrbuch. [Springer Textbook]. Springer-Verlag, Berlin, third edition, 1993. Eine Einführung. [An introduction].

[26] Ioannis Karatzas and Steven Shreve. *Brownian motion and stochastic calculus*, volume 113. Springer Science & Business Media, 2012.

[27] Fiodar Kilin. Accelerating the calibration of stochastic volatility models. *MPRA*, pages 1–20, 2006.

[28] Peter E. Kloeden and Eckhard Platen. *Numerical solution of stochastic differential equations*, volume 23 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1992.

[29] Steven G Kou. A jump-diffusion model for option pricing. *Management science*, 48(8):1086–1101, 2002.

[30] Alan L Lewis. A simple option formula for general jump-diffusion and other exponential lévy processes. *Available at SSRN 282110*, 2001.

[31] Roger Lord and Christian Kahl. Optimal fourier inversion in semi-analytical option pricing. *Tinbergen Institute Discussion Paper*, pages 1–21, 2007.

[32] Roger Lord, Remmert Koekkoek, and Dick Van Dijk. A comparison of biased simulation schemes for stochastic volatility models. *Quant. Finance*, 10(2):177–194, 2010.

[33] Eugene Lukacs. *Characteristic functions*. Hafner Publishing Co., New York, 1970. Second edition, revised and enlarged.

[34] Robert C Merton. Theory of rational option pricing. *The Bell Journal of economics and management science*, pages 141–183, 1973.

[35] Robert C Merton. Option pricing when underlying stock returns are discontinuous. *Journal of financial economics*, 3(1-2):125–144, 1976.

[36] Wolfgang Mueller. *Stochastische Analysis*. TU Graz, 2012.

[37] Marek Musiela and Marek Rutkowski. *Martingale methods in financial modelling*, volume 36 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1997.

[38] Bernt Oksendal. *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media, 2013.

[39] Vladimir Ostrovski. Efficient and exact simulation of the hull-white model. *Available at SSRN 2304848*, 2013.

[40] Walter Rudin. *Real and complex analysis*. McGraw-Hill Book Co., New York, third edition, 1987.

[41] Rainer Schöbel and Jianwei Zhu. Stochastic volatility with an ornstein–uhlenbeck process: an extension. *European Finance Review*, 3(1):23–46, 1999.

[42] Rüdiger Seydel. *Tools for computational finance*. Universitext. Springer-Verlag, Berlin, second edition, 2004.

[43] Elias M Stein and Jeremy C Stein. Stock price distributions with stochastic volatility: an analytic approach. *Review of financial Studies*, 4(4):727–752, 1991.

[44] Josef Stoer. *Numerische Mathematik. 1.* Springer-Lehrbuch. [Springer Textbook]. Springer-Verlag, Berlin, sixth edition, 1993. Eine Einführung—unter Berücksichtigung von Vorlesungen von F. L. Bauer. [An introduction—based on the lectures of F. L. Bauer].

[45] Daniel W Stroock and SRS Varadhan. Diffusion processes with continuous coefficients, i. *Communications on Pure and Applied Mathematics*, 22(3):345–400, 1969.