Universität Ulm
Fakultät für Mathematik und Wirtschaftswissenschaften
Graduiertenkolleg 1100
Modellierung, Analyse und Simulation
in der Wirtschaftsmathematik

Option pricing under time-varying risk aversion with applications to risk forecasting

DISSERTATION

zur Erlangung des Doktorgrades Dr. rer. pol.
der Fakultät für Mathematik und Wirtschaftswissenschaften
der Universität Ulm

vorgelegt von
Florentin Rahe
aus Stuttgart

2012

Amtierender Dekan: Professor Dr. Dieter Rautenbach
1. Gutachter: Professor Dr. Rüdiger Kiesel
2. Gutachter: Professor Dr. Gunter Löffler
Tag der Promotion: 2. Mai 2013

Contents

List of Figures	V
List of Tables	VII
List of Abbreviations	IX
List of Symbols	XI
1. Introduction	1
2. Asset pricing	5
2.1. Introduction	5
2.2. Economic perspective	6
2.2.1. Consumption-based model	6
2.2.2. Pricing kernel	7
2.3. Financial mathematical perspective	9
2.3.1. Risk-neutral pricing	10
2.3.2. Pricing kernel	13
2.4. Parametric pricing kernels	16
2.4.1. Utility transformation	16
2.4.2. Calibration transformation	17
2.4.3. Estimation	18
2.5. Empirical pricing kernels	20
2.5.1. Stylized facts	20
2.5.2. The pricing kernel puzzle	24
2.5.3. Applications in forecasting – a literature review	25
2.5.4. An empirical example	28
3. Affine models	35
3.1. Introduction	35
3.2. Affine processes	35
3.3. Fourier inversion methods	37
3.3.1. Black-Scholes style formula	37
3.3.2. Carr-Madan formula	38
3.3.3. Lewis formula	42
3.4. Heston model	43
3.4.1. Statistical asset price dynamics	43
3.4.2. Risk-neutral asset price dynamics	44
3.4.3. Pricing kernel	48

3.4.4.	Characteristic function	49
3.4.5.	Variance risk premium	52
3.5.	A 2-Factor stochastic volatility model	56
3.5.1.	Statistical asset price dynamics	56
3.5.2.	Risk-neutral asset price dynamics	57
3.5.3.	Pricing kernel	61
3.5.4.	Characteristic function	61
3.5.5.	Variance risk premium	63
4.	Estimation of option pricing models	65
4.1.	Introduction	65
4.2.	The classical Kalman filter	66
4.2.1.	State-space representation	66
4.2.2.	Derivation of the Kalman filter	67
4.2.3.	Maximum likelihood estimation	69
4.2.4.	Statistical inference	71
4.3.	The unscented Kalman filter	72
4.3.1.	Motivation	72
4.3.2.	Unscented transformation	74
4.3.3.	Derivation of the unscented Kalman filter	79
4.3.4.	Quasi-maximum likelihood estimation	82
4.3.5.	Statistical inference	83
4.4.	Estimation of option pricing models via the unscented Kalman filter	84
4.4.1.	Information content of option prices	84
4.4.2.	State-space representation for option pricing models	86
4.4.3.	Vega-scaling	87
4.4.4.	Estimation	90
5.	Monte-Carlo study	91
5.1.	Introduction	91
5.2.	Numerical difficulties	91
5.3.	A caching technique for the 2-Factor SV model	94
5.4.	Data	96
5.4.1.	Data characteristics	96
5.4.2.	Data simulation	98
5.5.	Results of the Monte Carlo study	101
5.5.1.	Heston model	101
5.5.2.	2-Factor SV model	104
6.	Empirical study	109
6.1.	Introduction	109
6.2.	Literature review	109
6.3.	Data	112
6.4.	Estimation results	118
6.4.1.	Heston model	118
6.4.2.	2-Factor SV model	123

6.5. Forecasting performance	128
6.5.1. Setup	128
6.5.2. Value-at-Risk	129
6.5.3. Forecasting models	131
6.5.4. Results	133
7. Conclusion	141
8. Zusammenfassung	147
A. Supplementary theory and notation	151
A.1. Properties of the multivariate normal distribution	151
A.2. Multi-dimensional Taylor series expansion	152
A.3. Runge-Kutta method	153
B. Additional calculations	157
B.1. Supplement to the proof of Lemma 3.1	157
B.2. Discretization of the Cox-Ingersoll-Ross process	158
B.3. Return distribution	159
C. Additional figures and tables	161
C.1. Value-at-Risk forecasts for CDS spreads	161
C.2. Data analysis	163
C.3. Estimation results for the Heston model	165
C.4. Estimation results for the 2-Factor SV model	168
D. A modified Heston model	171
Bibliography	173
Acknowledgment	183

List of Figures

2.1. Pricing kernel based on the power utility function.	9
2.2. Pricing kernel of the Black-Scholes model.	15
2.3. Various pricing kernel estimates in the literature.	21
2.4. Firm value, volatility and transformation estimates for Valeo.	31
3.1. Integrand of the Carr-Madan formula.	41
3.2. Statistical properties of the Heston model.	44
3.3. Volatility surface of the Heston model.	48
3.4. Pricing kernel of the Heston model.	49
3.5. Risk aversion index estimate of Bollerslev et al. [36].	55
3.6. Volatility term-structure of the 2-Factor SV model.	59
3.7. Volatility surface of the 2-Factor SV model.	59
3.8. Volatility term-structure of S&P 500 index options.	60
3.9. Pricing kernel of the 2-Factor SV model.	61
4.1. Mean and variance of a transformed random variable.	74
4.2. Time series of regular and vega-scaled option prices.	88
5.1. Log-likelihood functions based on the adaptive Simpson rule.	93
5.2. Log-likelihood functions based on the adaptive Gauss-Kronrod rule.	95
5.3. Strike choice function.	97
5.4. Volatility sensitivity of option prices.	98
5.5. Estimation accuracy based on deep in-the-money and out-of-the-money options.	99
5.6. Filtering accuracy of the state equation for the Heston model.	102
5.7. Log-likelihood functions for the parameters of the Heston model.	103
5.8. Filtering accuracy of the state equation for the 2-Factor SV model.	105
5.9. Log-likelihood functions for the parameters of the 2-Factor SV model.	106
6.1. Statistical properties of the S&P 500 index.	112
6.2. Maturities of the S&P 500 index options.	113
6.3. Moneyness of the S&P 500 index options.	114
6.4. Black-Scholes implied volatilities of the S&P 500 index options.	116
6.5. Volatility term-structure of the S&P 500 index options.	117
6.6. Volatility and variance risk premium estimates for the S&P 500 index based on the Heston model.	120
6.7. Volatility term-structure of the S&P 500 index options and corresponding Heston model estimates.	121

List of Figures

6.8. Evolution of the Heston model parameters before and during the subprime mortgage crisis.	122
6.9. Volatility, risk aversion and variance risk premium estimates for the S&P 500 index based on the 2-Factor SV model.	125
6.10. Volatility term structure of the S&P 500 index options and corresponding 2-Factor SV model estimates.	126
6.11. Evolution of the 2-Factor SV model parameters before and during the subprime mortgage crisis.	127
6.12. The Value-at-Risk for a continuous profit and loss distribution.	130
6.13. Evolution of the GARCH parameters.	137
6.14. Evolution of the 99% one week ahead VaR forecast for different models. . .	139
6.15. Comparison of the 99% one week ahead VaR forecasts.	140
C.1. Dividends for the S&P 500 index and riskfree rate for the U.S. market. . .	163
C.2. Vega-scaled S&P 500 index options.	164
C.3. Measurement errors for the Heston model.	165
C.4. Measurement errors for the 2-Factor SV model.	168

List of Tables

2.1. Value-at-Risk forecasts for 2-year CDS spreads.	32
4.1. Sigma points and weights.	79
4.2. Standard deviation of regular and vega-scaled option prices.	89
5.1. Simulation parameters for the Heston model.	101
5.2. Estimation results of the Monte-Carlo study for the Heston model.	103
5.3. Simulation parameters for the 2-Factor SV model.	104
5.4. Estimation results of the Monte-Carlo study for the 2-Factor SV model.	107
6.1. Previous estimation results for the parameters of the Heston model.	110
6.2. Descriptive statistics for the S&P 500 index returns.	113
6.3. Regular vs. vega-scaled option prices.	118
6.4. Estimation results for the parameters of the Heston model.	119
6.5. Estimation results for the parameters of the 2-Factor SV model.	123
6.6. Value-at-Risk forecasting performance: Precrash period.	134
6.7. Value-at-Risk forecasting performance: Crash period.	135
6.8. Value-at-Risk forecasting performance: Total period.	136
C.1. Value-at-Risk forecasts for 3-year CDS spreads.	161
C.2. Value-at-Risk forecasts for 4-year CDS spreads.	162
C.3. Value-at-Risk forecasts for 5-year CDS spreads.	162
C.4. Measurement errors for the Heston model.	165
C.5. Evolution of the Heston model parameters (Part 1/2).	166
C.6. Evolution of the Heston model parameters (Part 2/2).	167
C.7. Measurement errors for the 2-Factor SV model.	168
C.8. Evolution of the 2-Factor SV model parameters (Part 1/2).	169
C.9. Evolution of the 2-Factor SV model parameters (Part 2/2).	170

List of Abbreviations

ATM	at-the-money
CBOE	Chicago Board Options Exchange
CCAPM	consumption-based capital asset pricing model
cdf	cumulative distribution function
CDS	credit default swap
CIR	Cox-Ingersoll-Ross
EKF	Extended Kalman filter
FFT	Fast Fourier transform
GARCH	generalized autoregressive conditional heteroscedasticity
GMM	generalized method of moments
GPS	global positioning system
GS	grid size
ITM	in-the-money
ML	maximum likelihood
ODE	ordinary differential equation
OTM	out-of-the-money
pdf	probability density function
PK	pricing kernel
QML	quasi-maximum likelihood
RND	risk-neutral density
RWD	real-world density
SDE	stochastic differential equation
SV	stochastic volatility
UKF	unscented Kalman filter
VaR	Value-at-Risk

List of Symbols

In the following, we provide a summary of selected important symbols.

$\mathcal{F}_t$	σ -algebra representing the information at time t
$f(\cdot)$	probability density function
$f(\cdot \mathcal{F}_t)$	conditional probability density function
$\mathbb{E}[\cdot]$	expectation
$\mathbb{E}[\cdot \mathcal{F}_t]$	conditional expectation
$Var[\cdot]$	variance
$Var[\cdot \mathcal{F}_t]$	conditional variance
$\mathcal{N}(\mu, \Sigma)$	normal distribution with mean μ and variance Σ
φ_X	characteristic function of the random variable X
$\mathbb{P}$	statistical probability measure
$\mathbb{Q}$	risk-neutral probability measure
W_t	standard Brownian motion under the statistical measure
$\widetilde{W}_t$	standard Brownian motion under the risk-neutral measure
S_t	asset price process/underlying
σ_t	variance process
θ_t	risk aversion process
r	riskfree rate
K	strike price
τ	maturity
$\mathcal{C}$	theoretical call option price
$\bar{\mathcal{C}}$	observed call option price
σ_{BS}	Black-Scholes implied volatility
v_{BS}	Black-Scholes vega
U	utility function
x_t	vector of state variables
y_t	vector of measurements
η_t	disturbance term of the state equation
ϵ_t	disturbance term of the measurement equation
$\mathcal{I}_t$	information set, which comprises all observations y_t up to time t
Δt	time spacing of an equidistant data set
$\mathcal{L}$	log-likelihood function
$A^\top$	transposed of the matrix A
I_N	identity matrix of dimension N
$\odot$	Hadamard product

1. Introduction

A precise forecast of the future development of asset prices, such as stock or bonds, is important in many areas of finance. Besides the valuation of options and the portfolio theory this includes, in particular, risk management. Indeed, it is usually difficult to make a precise statement which price an asset will reach in the future. However, it is possible to assess the expected price development, volatility or the probability of single scenarios by means of the corresponding asset price distribution. In risk management the scenario of a large loss and its probability of occurrence are of particular interest. The Value-at-Risk is the probably most popular risk measure and quantifies the maximum loss of a position, which is not exceeded for a given confidence level and forecasting horizon (see e.g. Tasche [147]).

In order to estimate the distribution of future asset prices or returns, one usually applies an econometric or financial mathematical model, which is estimated based on historical data. Accordingly, the estimation results reflect the characteristics of the past. However, if there is a structural interruption at the end of the data period, the forecasting model does not or only insufficiently capture the current market situation. Therefore, it is not surprising that banks' risk models had big problems to measure their risk during the subprime mortgage crises. Exemplarily, one can quote Charles Prince, the former CEO of the Citigroup, who commented billions in capital write-offs with the words "our risk models simply did not work" [32].

An alternative to backward-looking forecasts based on historical data offer option prices, which aggregate all currently available market information about the future development of the underlying. This forward-looking information, which in particular reflects the current market conditions, can be extracted by means of option pricing models. Empirical studies suggest that the forward-looking information embedded in option prices, as for example Black-Scholes implied volatilities, do have a better forecasting performance than forecasts based on historical data (see Christoffersen et al. [57] for a literature review).

An option pricing model consists of two components, the statistical and the risk-neutral asset price dynamics. While the statistical dynamics describe the characteristics of the actual price evolution of the underlying asset, the risk-neutral dynamics are determined by both the actual price evolution and the risk preferences of the market participants. In particular, the risk-neutral dynamics are relevant for the purpose of option pricing. The ratio of risk-neutral to statistical dynamics is also referred to as "pricing kernel" and illustrates the supposed risk preferences of the option pricing model. The name "pricing kernel" is mainly used in the context of so called "consumption-based" models, which de-

1. Introduction

scribe the intertemporal consumption and investment decision problem of a representative investor. Material to the structure of the pricing kernel in the consumption-based model is the supposed utility function. Common utility functions imply pricing kernels, which are monotonically decreasing with the consumption level. Accordingly, a representative investor is willing to pay more for payoffs in bad states of the world (low consumption level) than for payoffs in good states of the world (high consumption level). An example for assets, which payoff in bad states of the world, are insurances [59, p. 13 f.].

In order to extract the forward-looking information embedded in option prices by means of option pricing models both the supposed risk-neutral dynamics and the pricing kernel need to be modeled in an appropriate manner. The empirical properties of the risk-neutral dynamics are well known and typically described in the form of Black-Scholes implied volatilities (see e.g. Cont et al. [61]). Modern option pricing models, such as e.g. the stochastic volatility model of Heston [95], are well able to capture these properties. In contrast, the empirical properties of pricing kernels are less understood. This can mainly be attributed to the fact that the risk preferences of market participants are not directly observable and thus difficult to estimate. However, some empirical studies suggest that pricing kernels are typically S-shaped or rather exhibit a monotonically increasing area (see e.g. Aït-Sahalia and Lo [1], Jackwerth [99] and Rosenberg and Engle [138]). Since the pricing kernels of option pricing and consumption-based models are usually not able to capture this typical characteristic, this phenomenon is also called “pricing kernel puzzle” (see e.g. Brown and Jackwerth [44]). In addition, the shape of pricing kernels is usually not constant but changes over time, which indicates that risk preferences are time-varying (see e.g. Rosenberg and Engle [138] and Grith et al. [89]).

Besides an adequate option pricing model, one also needs an appropriate estimation method in order to extract the forward-looking information embedded in option prices. Moreover, for the purpose of forecasting it is not sufficient to estimate only the risk-neutral dynamics, since they are biased by the risk preferences of the market participants. One possibility to jointly estimate the statistical and risk-neutral dynamics offers the Kalman filter. The Kalman filter is an algorithm to estimate the state of a time-varying system based on noisy measurements. Originally, the Kalman filter was applied in engineering applications, such as e.g. GPS (Global Positioning System) navigation. Meanwhile, the Kalman filter is also applied in the context of econometric issues. In contrast to the classical Kalman filter, the unscented Kalman filter developed by Julier und Uhlmann [106] can also be applied to nonlinear system, whereby its capabilities are considerably more versatile. Thus, the unscented Kalman filter can be applied in order to estimate option pricing models, which was first proposed by Bakshi et al. [11] and Bakshi and Wu [16].

The first contribution of this thesis is the development of a novel option pricing model, which incorporates the stylized facts of asset returns as well as the empirical properties of pricing kernels. The basis of this new 2-Factor stochastic volatility (SV) model is the Heston model, which is extended by an additional “risk aversion process”. However, this risk aversion process only enters the risk-neutral dynamics and not the statistical dynamics. Thereby, we are able to reproduce changes in option prices, which are only due to changes in risk preferences and not caused by changes in the statistical dynamics. In

particular, this model is able to produce S-shaped pricing kernels and time-varying risk preferences. Last but not least, the 2-Factor SV model has a more flexible volatility term structure than the Heston model, which makes it easier to explain time series of option prices.

The second contribution of this thesis is a thorough Monte-Carlo study, where we check the estimation accuracy of the unscented Kalman filter both for the Heston model and for the 2-Factor SV model. Thereby, we find that the parameters of the Heston model can be estimated very accurately. In contrast, it is more challenging to estimate the parameters of the 2-Factor SV model, which is due to the additional risk aversion process. Nevertheless, most parameters of the 2-Factor SV model can be estimated sufficiently well. Furthermore, we highlight some numerical difficulties within this Monte-Carlo study, which can arise when estimating option pricing models together with the unscented Kalman filter. More precisely, numerical difficulties in the form of discontinuous log-likelihood functions can arise if the adaptive Simpson rule, which can be used as numerical integration technique for the calculation of option prices, is combined with an inappropriate parameter choice for the unscented transformation. Moreover, we propose a caching technique, which offers a great speed advantage when estimating the 2-Factor SV model together with the unscented Kalman filter.

The third contribution of this thesis is a bipartite empirical study. In the first part, we estimate the Heston model and the 2-Factor SV model via the unscented Kalman filter based on S&P 500 index options. The underlying data period ranges from November 3, 1999 to October 28, 2009 and thus includes several market shocks like the burst of the dot-com bubble, the terror attacks 9/11 and the subprime mortgage crisis. For both models, we obtain highly significant estimation results, which also take on plausible values. Special mention must be made for the estimation results of the variance risk premium, which allow drawing conclusions about the risk preferences of the market participants. Similarly as in the previous literature (see e.g. Chernov and Ghysels [54], Pan [130] and Jones [104]), we find a negative variance risk premium for the Heston model, which indicates that market participants are risk averse. Also the variance risk premium of the 2-Factor SV model is negative during turbulent market conditions. However, during the calm market period from mid-2004 to mid-2007 the variance risk premium of the 2-Factor SV model is slightly positive, which indicates a risk seeking market behavior during that time. From mid-2007 the variance risk premium turns negative and the risk aversion process continuously increases, which reflects the growing fear of the market participants during the subprime mortgage crises. Moreover, we find that the risk aversion process has a lower variability compared to the volatility process, which makes it easier to identify changing market conditions. In addition, the 2-Factor SV model fits the empirical volatility term structure much better than the Heston model and also features a higher parameter stability.

In the second part of this empirical study, we analyze in which quality the Heston model and the 2-Factor SV model are able to extract the forward-looking information embedded in option prices. We therefore perform out-of-sample Value-at-Risk forecasts for the S&P 500 index during the period of the subprime mortgage crises based on these two models. In order to better classify the corresponding forecasting results, we also perform Value-

1. Introduction

at-Risk forecasts based on two commonly used Value-at-Risk models, namely the Black-Scholes model, which also relies on option-implied information, and the GARCH model, which relies on historical information. As a result, the 2-Factor SV model has the best forecasting performance, followed by the Black-Scholes, the Heston and the GARCH model. In particular, the 2-Factor SV model is the only one, which is able to perform highly accurate Value-at-Risk forecasts for all confidence levels (95%, 99% and 99.9%) and forecasting horizons (1, 2, 3 and 4 weeks) despite the challenging forecasting period.

The following thesis is structured as follows. First, Chapter 2 introduces the concept of pricing kernels and describes their empirical properties. Moreover, we review the literature about the performance of forecasting models based on option-implied information. Finally, we highlight the difficulty in using common pricing kernels for the purpose of forecasting within the framework of a small empirical study. Chapter 3 presents the class of affine processes and discusses different Fourier inversion techniques, which allow determining the price of an option based on the characteristic function of the underlying. Moreover, Chapter 3 derives the Heston model and the 2-Factor SV model, which both belong to the class of affine models. Then Chapter 4 explains how option pricing models can be estimated by means of the unscented Kalman filter. The corresponding estimation accuracy for the Heston model and the 2-Factor SV model is analyzed in Chapter 5. In Chapter 6, we present the empirical study where the Heston model and the 2-Factor SV model are estimated based on S&P 500 index options. Moreover, we analyze the application of these models for the purpose of forecasting by performing a Value-at-Risk study. Finally, Chapter 7 summarizes the results of this thesis and gives an outlook for further research opportunities. In addition, all MATLAB implementations and data sets used for this thesis can be found on the attached CD.

2. Asset pricing

2.1. Introduction

One of the fundamental issues in financial economics is asset pricing, which deals with the question of how to determine the price of a claim with uncertain future payments. Such risky assets are rather the norm than the exception and include investment opportunities, such as real estate, stock or options. The consumption-based model offers a very general framework, which allows to price any kind of asset. Within this approach the price of an asset corresponds to the expectation of the “preference-weighted” payoff profile. More precisely, the payoff profile of an asset is weighted by a pricing kernel, which is based on the utility function of the representative investor and thus reflects his risk preferences over different states of the world. In contrast to the consumption-based model, option pricing models can only be used to price derivatives, such as e.g. a European call option. Thereby, the price of an option can be determined by computing the expectation of the discounted payoff profile under the so called “risk-neutral measure”. Besides the risk-neutral asset price dynamics of the underlying, which are relevant for the purpose of option pricing, an option pricing model also needs to specify the statistical asset price dynamics of the underlying, which should reflect its empirical characteristics. In particular, the ratio of risk-neutral to statistical asset price dynamics approximately corresponds to the pricing kernel in the consumption-based model and thus reflects the supposed risk preferences of an option pricing model.

Recently, pricing kernels are applied in the context of forecasting. Thereby, the risk-neutral density of the underlying is first estimated based on current option prices and then transformed to a real-world density via some pricing kernel. In contrast to alternative forecasting methods based on historical data, this approach allows to make use of the forward-looking information contained in option prices. The crucial point in using transformed risk-neutral densities for the purpose of forecasting is that the underlying theoretical pricing kernel needs to capture the true empirical properties of pricing kernels. Otherwise, also the resulting real-world density does not adequately describe the characteristics of the statistical asset price dynamics. However, empirical studies suggest that the empirical properties of pricing kernels can only partly be captured by the commonly used theoretical pricing kernels. Nevertheless, there is some empirical evidence that transformed risk-neutral densities outperform risk-neutral densities in forecasting.

In the following, Chapter 2.2 first describes how asset prices can be determined within the framework of the consumption-based model. Then Chapter 2.3 explains the concept

2. Asset pricing

of risk-neutral pricing. Some commonly used parametric pricing kernels are presented in Chapter 2.4, where we also introduce an estimation method for parametric pricing kernels. Finally, Chapter 2.5 describes the empirical properties of pricing kernels and their application for forecasting. We conclude this section with an empirical example.

2.2. Economic perspective

In the following, we will describe how asset prices can be determined from an economic perspective. More precisely, we will show how asset prices can be determined within the framework of a *consumption-based model*. A good introduction to this topic is given by Cochrane [59], which is also the basis for the next two sections.¹

2.2.1. Consumption-based model

The starting point of the consumption-based model is the consideration of an investor about his composition of current and future consumption given some income stream. Following Cochrane [59, p. 4f], we consider a one period model and assume that the investor has a deterministic income Y_t at times t and $t + 1$.² This income can either be used for immediate consumption or for an investment opportunity. The only investment opportunity in this setting is an asset with time- t price P_t and random payoff Π_{t+1} . Furthermore, we assume that the investor can buy an arbitrary (possibly real-valued) number α of this asset. Hence, the consumption streams at times t and $t + 1$ are as follows [59, p. 5]:

$$\begin{aligned} t : \quad C_t &= Y_t - \alpha P_t , \\ t + 1 : \quad C_{t+1} &= Y_{t+1} + \alpha \Pi_{t+1} . \end{aligned} \tag{2.1}$$

The investor evaluates the consumption stream (2.1) via the following *intertemporal utility function* [59, p. 4]:

$$\mathcal{U}(C_t, C_{t+1}) = U(C_t) + \mathbb{E}[\beta U(C_{t+1}) \mid \mathcal{F}_t] , \tag{2.2}$$

where U denotes the *utility function* of the investor. A utility function is typically increasing, which reflects the appetite for more consumption, and concave, which captures the risk aversion of the investor [59, p. 5]. The aversion with regard to uncertain future consumption compared to certain immediate consumption increases with the degree of the concavity of the utility function. The *subjective discount factor* β , which is usually between zero and one, captures the impatience of the investor [59, p. 5]. In particular, the smaller β , the smaller gets the expectation in (2.2), which in turn increases the appreciation of immediate consumption relative to future consumption. In the following,

¹Alternatively, one can also consult Duffie [71] or Singleton [146].

²The one period consumption-based model can also be extended to a multi-period or continuous-time model. For more details see e.g. Cochrane [59, p. 25 ff.].

we will assume that β is deterministic. Finally, $\mathbb{E}[\cdot | \mathcal{F}_t]$ denotes the time- t conditional expectation operator evaluated based on the information of the investor at time t .

Now, given a subjective discount factor and a utility function, the investor will maximize his intertemporal utility function (2.2) by appropriately choosing α , which is [59, p. 5]

$$\arg \max_{\alpha \in \mathbb{R}} U(C_t) + \mathbb{E}[\beta U(C_{t+1}) | \mathcal{F}_t]$$

subject to

$$\begin{aligned} C_t &= Y_t - \alpha P_t , \\ C_{t+1} &= Y_{t+1} + \alpha \Pi_{t+1} . \end{aligned}$$

Following Cochrane [59, p. 5], this optimization problem can be solved by substituting the constraints (2.1) into the objective function (2.2) and then setting the derivative with respect to α equal to zero, which results in

$$P_t U'(C_t) = \mathbb{E}[\beta U'(C_{t+1}) \Pi_{t+1} | \mathcal{F}_t] . \quad (2.3)$$

The left hand side of (2.3) represents the current loss in utility if the investor buys one more unit of the asset. In contrast, the right hand side of (2.3) describes the gain in discounted expected utility, which results from the future payoff of that additional unit asset. Hence, the investor buys exactly as much of the asset such that the marginal utility loss equals the marginal utility gain. In particular, the first-order condition (2.3) determines the price of an asset with random payoff Π_{t+1} via the following equation [59, p. 5]:

$$P_t = \mathbb{E} [M_{t+1} \Pi_{t+1} | \mathcal{F}_t] , \quad (2.4)$$

where

$$M_{t+1} = \beta \frac{U'(C_{t+1})}{U'(C_t)} .$$

The term M_{t+1} is called *stochastic discount factor* or *pricing kernel* and captures the impatience as well as the risk aversion of the investor.³ It should be noted that the pricing kernel of an investor does not depend on the payoff profile of an asset and therefore can be used to price any asset [59, p. 7].

2.2.2. Pricing kernel

The asset specific risk adjustment in the consumption-based model depends on the co-variation between the pricing kernel and the future payoff. In order to see this, first note that the covariance between the pricing kernel and the future payoff can be written as [59, p. 13]

$$Cov [M_{t+1}, \Pi_{t+1} | \mathcal{F}_t] = \mathbb{E} [M_{t+1} \Pi_{t+1} | \mathcal{F}_t] - \mathbb{E} [M_{t+1} | \mathcal{F}_t] \mathbb{E} [\Pi_{t+1} | \mathcal{F}_t] . \quad (2.5)$$

³Sometimes M_{t+1} is also referred to as *state-price density* or *change of measure*.

2. Asset pricing

Furthermore, the time- t price of a certain future payoff in the amount of $\Pi_{t+1} = 1$ can be determined by discounting with the risk-free rate r , which is

$$P_t = \frac{1}{1+r}. \quad (2.6)$$

Alternatively, the price of this certain cash-flow can be determined via the basic pricing equation (2.4), which results in

$$P_t = \mathbb{E}[M_{t+1} | \mathcal{F}_t]. \quad (2.7)$$

Combining equation (2.6) and (2.7) yields

$$\frac{1}{1+r} = \mathbb{E}[M_{t+1} | \mathcal{F}_t]. \quad (2.8)$$

Now, the basic pricing equation can be written as [59, p. 13]:

$$\begin{aligned} P_t &= \mathbb{E}[M_{t+1}\Pi_{t+1} | \mathcal{F}_t] \\ &\stackrel{(2.5)}{=} \mathbb{E}[M_{t+1} | \mathcal{F}_t] \mathbb{E}[\Pi_{t+1} | \mathcal{F}_t] + \text{Cov}[M_{t+1}, \Pi_{t+1} | \mathcal{F}_t] \\ &\stackrel{(2.8)}{=} \frac{\mathbb{E}[\Pi_{t+1} | \mathcal{F}_t]}{1+r} + \text{Cov}[M_{t+1}, \Pi_{t+1} | \mathcal{F}_t]. \end{aligned} \quad (2.9)$$

Hence, the price of an asset with random payoff Π_{t+1} can be represented as the sum of the discounted expected cash flow and a risk adjustment term. The latter depends on the covariation of the pricing kernel and the payoff. In particular, a positive covariation of M_{t+1} and Π_{t+1} increases the corresponding asset price and thus lowers the corresponding return [59, p. 13]. The intuition behind this mechanism becomes clearer if we write (2.9) as follows

$$P_t = \frac{\mathbb{E}[\Pi_{t+1} | \mathcal{F}_t]}{1+r} + \frac{\text{Cov}[\beta U'(C_{t+1}), \Pi_{t+1} | \mathcal{F}_t]}{U'(C_t)}.$$

First note that the marginal utility U' is decreasing in consumption due to the concavity of the utility function. Hence, a positive covariation between C_{t+1} and Π_{t+1} results in a negative risk adjustment term and thus lowers the corresponding asset price. Conversely, a negative covariation between C_{t+1} and Π_{t+1} results in a positive risk adjustment term and thus raises the corresponding asset price. The sign of the risk adjustment term can be explained by the desire of the investor for a smooth consumption stream in any state of the world [59, p. 13]. Thus, investors prefer assets, which are negatively correlated with consumption and thus tend to have high payoffs in bad states of the world, where consumption would otherwise be low. An example for such an asset is fire insurance, which exactly pays off when your house burns down [59, p. 13 f.].

To further illustrate the concept of risk adjustments via pricing kernels, we consider an investor with a *power utility function*

$$U(x) = \frac{x^{1-\gamma}}{1-\gamma},$$

Figure 2.1.: Pricing kernel based on the power utility function.

where γ denotes the *relative risk aversion* [148, p. 452]. Moreover, we suppose that the subjective discount factor is constant with $\beta = 1$, so that the pricing kernel is given by

$$\begin{aligned}
 M_{t+1} &= \beta \frac{U'(C_{t+1})}{U'(C_t)} \\
 &= \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} \\
 &= e^{-\gamma \ln(C_{t+1}/C_t)} .
 \end{aligned} \tag{2.10}$$

The term $\ln(C_{t+1}/C_t)$ corresponds to the *growth rate of consumption*, which can further be specified by some distribution.⁴ For $\gamma > 0$, which is typically the case, the pricing kernel (2.10) is decreasing in the consumption growth rate, which is also illustrated in Figure 2.1. A decreasing pricing kernel indicates that an investor is willing to pay more for payoffs in rather bad than good states, where negative consumption growth rates are referred to as “bad states” and positive consumption growth rates as “good states”. In particular, the larger the relative risk aversion γ , the more is an investor willing to pay for payoffs in “bad states”.

2.3. Financial mathematical perspective

In the following, we will describe how asset prices can be determined from a financial mathematical perspective. More precisely, we will show how option prices can be computed via *risk-neutral pricing*. A good introduction into this topic is given by Shreve [144], which is also the basis for the following section.⁵

⁴See Cochrane [59, p. 11 f.] for an example.

⁵Alternatively, one can also consult Bingham and Kiesel [27], who describe the concept of risk-neutral valuation in a more general framework.

2.3.1. Risk-neutral pricing

Following Shreve [144, p. 214 ff.], the framework for the concept of risk-neutral pricing is a frictionless financial market, where market participants can continuously buy or sell all traded assets up to a fixed final time T . For simplicity, we consider a financial market model with only two assets, namely the stock and the money market account.⁶ The uncertainty in this financial market is modeled with a Brownian motion W_t , which is defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The filtration $(\mathcal{F}_t)_{0 \leq t \leq T}$, with $\mathcal{F} = \mathcal{F}_T$, is generated by W_t and represents all available information of the investors. Moreover, the stock price S_t is assumed to evolve as a *generalized geometric Brownian motion*, which is

$$dS_t = \mu_t S_t dt + \sigma_t S_t dW_t .$$

The adapted processes μ_t and σ_t denote the *mean rate of return* and *volatility*, respectively [144, p. 214]. In addition, we assume that the volatility process is almost surely not zero. Finally, we suppose a money market account with variable interest rate r_t , which is also $\mathcal{F}_t$ -adapted [144, p. 215].

The starting point for the concept of risk-neutral valuation is the idea to replicate the price evolution of a *contingent claim* via some hedging portfolio X_t , so that the final payoff of this contingent claim is attained almost surely [144, p. 218]. In our setup the contingent claim only depends on the stock price S_t , so that the final payoff can be expressed as $\Pi(S_T)$. An example for such a contingent claim is a European call option with strike price K , where the final payoff is given by

$$\Pi(S_T) = \max\{S_T - K, 0\} .$$

Following Shreve [144, p. 217 f.], the hedging portfolio within this setting can be composed out of the stock and the money market account. Let X_0 be the required initial capital for the hedging portfolio and α_t the number of shares invested in the stock. Here α_t is an adapted process, which is also referred to as *trading strategy*. In particular, the trading strategy needs to be *self-financing*, which essentially means that the price evolution of the hedging portfolio is only due to capital gains or losses and not a result of withdrawals or cash injections [27, p.231].⁷ Based on this trading strategy, the amount $\alpha_t S_t$ is invested in the stock and the remainder $X_t - \alpha_t S_t$ in the money market account. Hence, the evolution of the hedging portfolio is given by [144, p. 217]

$$dX_t = \alpha_t dS_t + r_t(X_t - \alpha_t S_t)dt . \quad (2.11)$$

Following Shreve [144, p. 217], the dynamics of the hedging portfolio can also be expressed as

$$dX_t = r_t X_t dt + \alpha_t \sigma_t S_t [\gamma_t dt + dW_t] , \quad (2.12)$$

⁶For a more general financial market model with d assets see e.g. Bingham and Kiesel [27].

⁷For a more formal definition of a self-financing trading strategy see Bingham and Kiesel [27, p.230, Definition 6.1.2].

where

$$\gamma_t = \frac{\mu_t - r_t}{\sigma_t} , \quad (2.13)$$

denotes the *market price of risk*.

Before we continue to derive the concept of risk-neutral valuation, we first need to state the following theorem [144, p. 224 f.]:⁸

Theorem 2.1 (Girsanov). *Let $W_t = [W_{1,t} \dots W_{d,t}]^\top$ be a d -dimensional Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $(\mathcal{F}_t)_{0 \leq t \leq T}$ be a filtration for this Brownian motion. Moreover, let $\gamma_t = [\gamma_{1,t} \dots \gamma_{d,t}]^\top$ be a d -dimensional adapted process. Define*

$$Z_t = \exp\left\{-\int_0^t \sum_{i=1}^d \gamma_{i,s} dW_{i,s} - \frac{1}{2} \int_0^t \sum_{i=1}^d \gamma_{i,s}^2 ds\right\} , \quad (2.14)$$

$$\widetilde{W}_{i,t} = W_{i,t} + \int_0^t \gamma_{i,s} ds \quad \text{for } i = 1, \dots, d , \quad (2.15)$$

with $\widetilde{W}_t = [\widetilde{W}_{1,t} \dots \widetilde{W}_{d,t}]^\top$ and suppose that

$$\mathbb{E}\left[\int_0^T \left(\sum_{i=1}^d \gamma_{i,s}^2\right) Z_s^2 ds\right] < \infty .$$

Then under the probability measure $\mathbb{Q}$, which is defined as

$$\mathbb{Q}[A] = \int_A Z_T(\omega) d\mathbb{P}(\omega) \quad \text{for all } A \in \mathcal{F} , \quad (2.16)$$

$\widetilde{W}_t$ is a d -dimensional Brownian motion.

The measure $\mathbb{Q}$ in Theorem 2.1 is also called *risk-neutral* or *equivalent martingale measure*. Moreover, Z_T is referred to as *change of measure* or *Radon-Nikodým derivative* of $\mathbb{Q}$ with respect to $\mathbb{P}$.⁹

If we apply Girsanov's theorem to the hedging portfolio dynamics (2.12), where (2.14) and (2.15) are specified by the market price of risk (2.13), then [144, p. 217]

$$dX_t = r_t X_t dt + \alpha_t \sigma_t S_t d\widetilde{W}_t .$$

Hence, under the risk-neutral measure, the mean rate of return of the hedging portfolio corresponds to the risk-free rate r_t .

⁸See Shreve [144, p. 212 ff.] for a proof of *Girsanov's theorem* in one dimension.

⁹For more details on Girsanov's theorem in the context of financial mathematics see Bingham and Kiesel [27]. Moreover, for a general treatment of the Radon-Nikodým derivative, we refer to Folland [80].

2. Asset pricing

Following Shreve [144, p. 215], we next define the *discount process*

$$D_t = e^{-\int_0^t r_s ds} .$$

Then the discounted hedging portfolio evolves as

$$d(D_t X_t) = \alpha_t \sigma_t D_t S_t \widetilde{W}_t ,$$

which is a result of *Itô's product rule* [144, p. 217].¹⁰ Therefore, under the risk-neutral measure, the discounted hedging portfolio has no drift term and thus is a *martingale*.¹¹ The martingale property of the risk-neutral discounted hedging portfolio dynamics allows to write

$$D_t X_t = \mathbb{E}_{\mathbb{Q}}[D_T X_T | \mathcal{F}_t] ,$$

where the subscript of the conditional expectation operator indicates the underlying probability measure. Following Shreve [144, p. 218], we assume that there exists a trading strategy α_t , such that

$$X_T = \Pi(S_T) ,$$

holds almost surely.¹² Then

$$\begin{aligned} D_t X_t &= \mathbb{E}_{\mathbb{Q}}[D_T X_T | \mathcal{F}_t] \\ &= \mathbb{E}_{\mathbb{Q}}[D_T \Pi(S_T) | \mathcal{F}_t] . \end{aligned}$$

Note that X_t corresponds to the capital, which is required to perfectly hedge a contingent claim with payoff $\Pi(S_T)$ via the trading strategy α_t [144, p. 218]. Hence, X_t also corresponds to the time- t price P_t of this contingent claim, so that

$$D_t P_t = \mathbb{E}_{\mathbb{Q}}[D_T \Pi(S_T) | \mathcal{F}_t] .$$

Since D_t is $\mathcal{F}_t$ -measurable, the time- t price of a contingent claim with payoff $\Pi(S_T)$ can be determined by

$$P_t = \mathbb{E}_{\mathbb{Q}}[e^{-\int_t^T r_s ds} \Pi(S_T) | \mathcal{F}_t] , \tag{2.17}$$

which is also referred to as *risk-neutral pricing formula* [144, p. 218]. One possibility to solve the risk-neutral pricing formula is the *Feynman-Kac approach*, which relates the conditional expectation in (2.17) to a partial differential equation, which can either be solved in closed-form or numerically.¹³

¹⁰See Protter [136] for more details about Itô calculus.

¹¹A stochastic process M_t is a martingale if $\mathbb{E}[M_t | \mathcal{F}_s] = M_s$ for all $0 \leq s \leq t \leq T$. Hence, a martingale has no tendency to rise or fall [144, p. 74].

¹²Note that under certain conditions such a trading strategy α_t indeed exists, which can be shown based on the *martingale representation theorem*. For more details see e.g. Shreve [144, p. 221 ff.]

¹³For more details about the Feynman-Kac approach see e.g. Shreve [144, p. 268 ff.].

2.3.2. Pricing kernel

The risk adjustment within the risk-neutral valuation approach works very similar to the one in the consumption-based model. Following Taylor [148, p. 251 f.], the close relation between these two concepts can be seen by reformulating the risk-neutral pricing formula as follows:

$$\begin{aligned}
P_t &= \mathbb{E}_{\mathbb{Q}}[e^{-\int_t^T r_s ds} \Pi(S_T) \mid \mathcal{F}_t] \\
&= \int_0^\infty e^{-\int_t^T r_s ds} \Pi(x) f_{S_T|\mathcal{F}_t}^{\mathbb{Q}}(x) dx \\
&= \int_0^\infty e^{-\int_t^T r_s ds} \Pi(x) \frac{f_{S_T|\mathcal{F}_t}^{\mathbb{Q}}(x)}{f_{S_T|\mathcal{F}_t}^{\mathbb{P}}(x)} f_{S_T|\mathcal{F}_t}^{\mathbb{P}}(x) dx \\
&= \mathbb{E}_{\mathbb{P}} \left[e^{-\int_t^T r_s ds} \frac{f_{S_T|\mathcal{F}_t}^{\mathbb{Q}}(S_T)}{f_{S_T|\mathcal{F}_t}^{\mathbb{P}}(S_T)} \Pi(S_T) \mid \mathcal{F}_t \right]. \tag{2.18}
\end{aligned}$$

Here, $f_{S_T|\mathcal{F}_t}^{\mathbb{P}}$ and $f_{S_T|\mathcal{F}_t}^{\mathbb{Q}}$ denote the statistical, respectively, the risk-neutral probability density function of the future asset price S_T conditional on the information $\mathcal{F}_t$. In the following, we will refer to $f_{S_T|\mathcal{F}_t}^{\mathbb{P}}$ simply as *real-world density* and to $f_{S_T|\mathcal{F}_t}^{\mathbb{Q}}$ as *risk-neutral density*. The term “real-world” emphasizes the fact that the real-world density reflects the actually observed asset price dynamics.¹⁴

If we compare (2.18) with (2.4), it is apparent that the term

$$e^{-\int_t^T r_s ds} \frac{f_{S_T|\mathcal{F}_t}^{\mathbb{Q}}(S_T)}{f_{S_T|\mathcal{F}_t}^{\mathbb{P}}(S_T)}, \tag{2.19}$$

in the risk-neutral pricing formula corresponds to

$$M_{t+1} = \beta \frac{U'(C_{t+1})}{U'(C_t)},$$

in the consumption-based model.¹⁵ Hence, the subjective discount factor β is replaced by a discount factor, which captures the term-structure of the risk-free rate. Moreover, the marginal rate of substitution $U'(C_{t+1})/U'(C_t)$ is replaced by the ratio of risk-neutral to real-world density, which reflects the risk preferences of the investor. Therefore, in analogy to the consumption-based model, the term (2.19) is also referred to as pricing kernel [148, p. 452]. However, in order to characterize the properties of pricing kernels, one usually considers the ratio of risk-neutral to statistical asset return density (see e.g. Brown and Jackwerth [44]). Further note, that a deterministic (subjective) discount factor only changes the scale but not the general shape of the pricing kernel. Therefore, we will subsequently loosely define the ratio of risk-neutral to statistical asset return density as

¹⁴Note that the real-world density is sometimes also called empirical, historical, risk-adjusted, subjective, statistical or true distribution [143, p. 2679].

¹⁵Note that the expectation in (2.4) is defined with respect to the statistical measure $\mathbb{P}$ even though not stated explicitly.

2. Asset pricing

pricing kernel and thus neglect the (subjective) discount factor.

To further illustrate the risk adjustment within the risk-neutral valuation approach, we consider the option pricing model of Black and Scholes [29], who suppose that the asset price under the statistical measure $\mathbb{P}$ evolves as

$$dS_t = \mu S_t dt + \sigma S_t dW_t .$$

This *stochastic differential equation* (SDE) has the following solution¹⁶

$$S_t = S_0 \exp\{(\mu - 1/2\sigma^2)t + \sigma W_t\} .$$

Next, we define the asset return for the period $[0, t]$ as follows:

$$ret_t = \ln(S_t/S_0) . \quad (2.20)$$

Since W_t is a normally distributed random variable with mean zero and variance t , also the asset return (2.20) under the statistical measure $\mathbb{P}$ is normally distributed with

$$\mathbb{P} : ret_t \sim \mathcal{N}(\mu_{\mathbb{P}}, \sigma^2) ,$$

where $\mu_{\mathbb{P}} = (\mu - 1/2\sigma^2)t$.

Furthermore, the asset price dynamics under the risk-neutral measure $\mathbb{Q}$ are defined as

$$dS_t = r S_t dt + \sigma S_t d\widetilde{W}_t ,$$

which can be obtained by an application of Girsanov's theorem.¹⁷ Accordingly, also the asset return (2.20) under the risk-neutral measure $\mathbb{Q}$ is normally distributed with

$$\mathbb{Q} : ret_t \sim \mathcal{N}(\mu_{\mathbb{Q}}, \sigma^2) ,$$

where $\mu_{\mathbb{Q}} = (r - 1/2\sigma^2)t$.

Now, the pricing kernel of the Black-Scholes model can be computed by building the ratio between the risk-neutral and the statistical asset return density, which is

$$\begin{aligned} \frac{f_{ret_t}^{\mathbb{Q}}(x)}{f_{ret_t}^{\mathbb{P}}(x)} &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x - \mu_{\mathbb{Q}})^2}{2\sigma^2}\right\} \Bigg/ \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x - \mu_{\mathbb{P}})^2}{2\sigma^2}\right\} \\ &= \exp\left\{-\frac{1}{2\sigma^2}(x - \mu_{\mathbb{Q}})^2 + \frac{1}{2\sigma^2}(x - \mu_{\mathbb{P}})^2\right\} \\ &= \exp\left\{-\frac{\mu_{\mathbb{P}} - \mu_{\mathbb{Q}}}{\sigma^2}x + \frac{\mu_{\mathbb{P}}^2 - \mu_{\mathbb{Q}}^2}{2\sigma^2}\right\} . \end{aligned} \quad (2.21)$$

First note that the pricing kernel (2.21) of the Black-Scholes model has the same exponential form as the pricing kernel (2.10) of the consumption-based model together with

¹⁶See Shreve [144, p. 148] for a proof.

¹⁷See Shreve [144, p. 214] for a proof.

Figure 2.2.: Pricing kernel of the Black-Scholes model. The underlying parameters for this illustration are $r = 0$, $\sigma = 0.1$ and $t = 1/12$.

the power utility function. Furthermore, the term $(\mu_{\mathbb{P}} - \mu_{\mathbb{Q}})/\sigma^2$ is positive if $\mu > r$, which is typically the case. Therefore, the pricing kernel (2.21) is decreasing in the asset return, which is illustrated in Figure 2.2. The interpretation of a decreasing pricing kernel is the same as before: investors are willing to pay more for payoffs in “bad states” (negative returns) than for payoffs in “good states” (positive returns), which reflects the risk aversion of the investor. In particular, the higher the ratio of the excess return $\mu - r$ compared to the variance σ^2 , the steeper is the pricing kernel. Hence, also the appreciation of an investor for payments in “bad states” increases compared to payments in “good states”. Conversely, one can also say that the higher the fear of having nothing in “bad states”, the steeper is the corresponding pricing kernel and thus the higher is the required return of an investor in order to agree to invest in an asset, which tends to pay off in “good states” and not in “bad states”.

Of course, the shape of the pricing kernel depends on the particular asset pricing model, so that Figures 2.1 and 2.2 are only illustrative examples. The pricing kernel of the consumption-based model mainly depends on the supposed utility function. In contrast, the pricing kernel of the risk-neutral valuation approach depends on the supposed asset price dynamics. In *incomplete markets* there exists no unique risk-neutral measure, so that the modeler has some freedom to specify the pricing kernel in order to capture the empirical properties of pricing kernels.¹⁸ In Chapter 2.5, we will describe the empirical characteristics of pricing kernels in more detail.

¹⁸A market is called incomplete if there exist more risk factors than traded assets, so that it is not possible to perfectly hedge all risk. An example for such a market model is the stochastic volatility model of Heston [95], which will be introduced in Chapter 3.4. For more details about risk-neutral pricing in incomplete markets see e.g. Bingham and Kiesel [27].

2.4. Parametric pricing kernels

In the following section, we will introduce two parametric pricing kernels, which are commonly used in the empirical literature (see e.g. Bliss and Panigirtzoglou [31], Liu et al. [122] and Shackleton et al. [143]). Moreover, we will describe an estimation method in order to estimate parametric pricing kernels.

2.4.1. Utility transformation

The utility transformation is a parametric pricing kernel and was first proposed by Rosenberg and Engle [138]. In order to derive the utility transformation, first recall the pricing kernel of the consumption-based model, which is defined as

$$M_T = \beta \frac{U'(C_T)}{U'(C_t)} . \quad (2.22)$$

Following Bliss and Panigirtzoglou [31, p. 409], we replace consumption in (2.22) by an asset S_t , which could represent a broad stock index like the S&P 500 index, so that

$$M_T = \beta \frac{U'(S_T)}{U'(S_t)} . \quad (2.23)$$

According to Brown and Jackwerth [44, p. 4], this modified pricing kernel is a projection of the consumption-based pricing kernel (2.22) onto the space of payoffs, which depend on the asset S_t . As a consequence, the pricing kernel (2.23) cannot be used to price all securities but a large and important subset composed of all financial securities, which depend on the asset S_t [44, p. 4].¹⁹

Next, we equalize the modified consumption-based pricing kernel (2.23) and the financial mathematical pricing kernel (2.19), so that

$$\begin{aligned} e^{-\int_t^T r_s ds} \frac{f_{S_T|\mathcal{F}_t}^{\mathbb{Q}}(S_T)}{f_{S_T|\mathcal{F}_t}^{\mathbb{P}}(S_T)} &= \beta \frac{U'(S_T)}{U'(S_t)} , \\ \iff f_{S_T|\mathcal{F}_t}^{\mathbb{P}}(S_T) &= \frac{1}{\xi} \frac{U'(S_T)}{U'(S_t)} f_{S_T|\mathcal{F}_t}^{\mathbb{Q}}(S_T) , \end{aligned} \quad (2.24)$$

where $\xi = e^{\int_t^T r_s ds} \beta$. Hence, the real-world density can be obtained by multiplying the risk-neutral density with an adjustment term, which depends on the (subjective) discount factor and the supposed utility function. Since the right hand side of (2.24) must integrate to one, Bliss and Panigirtzoglou [31, p. 419] propose the following normalized version of

¹⁹For more details about pricing kernel projections see Cochrane [59].

(2.24):

$$\begin{aligned}
f_{S_T|\mathcal{F}_t}^{\mathbb{P}}(S_T) &= \frac{1}{\xi} \frac{U'(S_T)}{U'(S_T)} f_{S_T|\mathcal{F}_t}^{\mathbb{Q}}(S_T) \Big/ \int_0^\infty \frac{1}{\xi} \frac{U'(x)}{U'(x)} f_{S_T|\mathcal{F}_t}^{\mathbb{Q}}(x) dx \\
&= \frac{f_{S_T|\mathcal{F}_t}^{\mathbb{Q}}(S_T)}{U'(S_T)} \Big/ \int_0^\infty \frac{f_{S_T|\mathcal{F}_t}^{\mathbb{Q}}(x)}{U'(x)} dx ,
\end{aligned} \tag{2.25}$$

where ξ is assumed to be deterministic. In particular, due to the normalization, the *utility transformation* (2.25) does not depend on ξ , which alleviates its application in the context of empirical studies.

2.4.2. Calibration transformation

The calibration transformation was first proposed by Fackler and King [75], who correct a misspecified density by virtue of a calibration function. To understand the theoretical background of this approach, we follow Taylor [148, p. 455 f.] and first consider a random variable S_T with *cumulative distribution function* (cdf) F . Then the random variable

$$U = F(S_T) \tag{2.26}$$

is uniformly distributed on the interval $[0, 1]$, which will be denoted by $U \sim \mathcal{U}(0, 1)$. Recall that a random variable U is uniformly distributed on the interval $[0, 1]$ if

$$\mathbb{P}[U \leq u] = u .$$

Hence, the proposition that (2.26) is uniformly distributed can be verified by noting that

$$\mathbb{P}[U \leq u] = \mathbb{P}[F(S_T) \leq u] = \mathbb{P}[S_T \leq F^{-1}(u)] = F(F^{-1}(u)) = u ,$$

where F^{-1} denotes the inverse of the cdf F [148, p. 456].

Unfortunately, the true cdf of S_T is not always known, so that a possibly misspecified cdf $\tilde{F}$ for S_T is used instead. In order to correct a misspecified cdf, Fackler and King [75] propose to use a calibration function, which captures the actual distribution of the transformed random variable

$$\tilde{U} = \tilde{F}(S_T) .$$

Assume C is the true cdf of the transformed random variable $\tilde{U}$, so that $C(\tilde{U}) \sim \mathcal{U}(0, 1)$. Then

$$C(\tilde{F}(x)) = \mathbb{P}[\tilde{U} \leq \tilde{F}(x)] = \mathbb{P}[\tilde{F}(S_T) \leq \tilde{F}(x)] = \mathbb{P}[S_T \leq x] = F(x) , \tag{2.27}$$

where F again denotes the correct cdf of S_T [122, p. 1504]. Thus, the calibration function C allows to correct a possibly misspecified cdf $\tilde{F}$. However, the calibration function is usually unknown and needs to be estimated.

2. Asset pricing

In our context, the “misspecified” cdf is the risk-neutral cdf of the future asset price S_T , which needs to be adjusted for risk preferences in order to correctly explain real-world observations. Therefore, in accordance to (2.27), we can write

$$F_{\mathbb{P}}(x) = C(F_{\mathbb{Q}}(x)) , \quad (2.28)$$

where $F_{\mathbb{Q}}$ and $F_{\mathbb{P}}$ denote the risk-neutral and real-world cdf, respectively. The relation between risk-neutral and real-world density can be obtained by deriving (2.28) with respect to x , which is [148, p. 456]

$$f_{\mathbb{P}}(x) = \frac{dC(F_{\mathbb{Q}}(x))}{dx} = \frac{dC(F_{\mathbb{Q}}(x))}{dF_{\mathbb{Q}}} \frac{dF_{\mathbb{Q}}(x)}{dx} = \frac{dC(F_{\mathbb{Q}}(x))}{dF_{\mathbb{Q}}} f_{\mathbb{Q}}(x) . \quad (2.29)$$

Fackler and King [75] suggest to use the *beta distribution* as calibration function, which is defined as [148, p. 439 f.]

$$F_{\beta}(x | j, k) = \frac{1}{B(j, k)} \int_0^x u^{j-1} (1-u)^{k-1} du , \quad (2.30)$$

where

$$B(j, k) = \frac{\Gamma(j)\Gamma(k)}{\Gamma(j+k)} .$$

Here Γ denotes the *gamma function*. Hence, the *calibration transformation* (2.29) together with the beta distribution (2.30) results in [148, p. 456]

$$f_{\mathbb{P}}(x) = \frac{F_{\mathbb{Q}}(x)^{j-1} (1 - F_{\mathbb{Q}}(x))^{k-1}}{B(j, k)} f_{\mathbb{Q}}(x) .$$

The beta distribution has flexible shape, which can be adjusted by the parameters j and k . In particular, for $j = k = 1$, the beta distribution corresponds to the cdf of a uniformly distributed random variable, which would be adequate if market participants are risk-neutral, so that the risk-neutral density correctly explains real-world observations.

2.4.3. Estimation

Following Liu et al. [122], the utility and calibration transformation can be estimated by virtue of a two-step *maximum likelihood* (ML) method. We therefore suppose that a time series of call options $\bar{\mathcal{C}}_0, \dots, \bar{\mathcal{C}}_N$ can be observed at times $t_0, \dots, t_N$.²⁰ More precisely, we suppose that at each time t_i a cross-section of call options $\bar{\mathcal{C}}_i = [\bar{\mathcal{C}}_{i,1} \dots \bar{\mathcal{C}}_{i,n}]$ with strike prices $K_i = [K_{i,1} \dots K_{i,n}]$ can be observed. Furthermore, all call options are supposed to have the same underlying S_i as well as the same maturity τ_i , such that $\tau_i = t_{i+1} - t_i$.

²⁰Of course, this estimation method can also be applied with put options or a combination of call and put options.

Recall from Chapter 2.3 that the price of a European call option can be determined by

$$\begin{aligned}\mathcal{C}(S_t, K, \tau, \theta_1) &= \mathbb{E}_{\mathbb{Q}}[e^{-r\tau} \max\{S_T - K, 0\} | \mathcal{F}_t] \\ &= e^{-r\tau} \int_0^\infty \max\{x - K, 0\} f_{S_T | \mathcal{F}_t}^{\mathbb{Q}}(x | \theta_1) dx ,\end{aligned}$$

where $\tau = T - t$. Here, $f_{S_T | \mathcal{F}_t}^{\mathbb{Q}}$ denotes some parametric risk-neutral density for the asset price S_T conditional on the information at time t . Moreover, the vector θ_1 summarizes the parameters of this risk-neutral density.²¹

Following Shackleton et al. [143, p. 2681], the risk-neutral densities embedded in the observed call option prices can be estimated by minimizing the squared differences between observed and theoretical call option prices, which is

$$\hat{\theta}_{1,i} = \arg \min_{\theta_{1,i}} \sum_{j=1}^n [\bar{\mathcal{C}}_{i,j} - \mathcal{C}(S_i, K_{i,j}, \tau_i, \theta_{1,i})]^2 \quad \text{for } i = 0, \dots, N . \quad (2.31)$$

It is worth mentioning that (2.31) is a ML estimator for the parameter vector $\theta_{1,i}$ if the pricing errors are independent and identically normally distributed [143, p. 2681]. In particular, the risk-neutral density estimate for the asset price S_i conditional on the information at time t_{i-1} is given by $f_{S_i | \mathcal{F}_{i-1}}^{\mathbb{Q}}(x_i | \hat{\theta}_{1,i-1})$, where $\mathcal{F}_{i-1}$ denotes the information set generated by $S_0, \dots, S_{i-1}$. These risk-neutral densities can be transformed to real-world densities via the utility or calibration transformation. For ease of notation, we will write these transformations more generally as

$$f_{S_i | \mathcal{F}_{i-1}}^{\mathbb{P}}(x_i) = g \left(f_{S_i | \mathcal{F}_{i-1}}^{\mathbb{Q}}(x_i | \hat{\theta}_{1,i-1}) \mid \theta_2 \right) ,$$

where g denotes the parametric transformation function specified by either (2.25) or (2.29). Moreover, the vector θ_2 summarizes the parameters of the transformation function g . As a result the joint density for observing $S_1, \dots, S_N$ can be written as

$$\begin{aligned}f_{S_1, \dots, S_N}^{\mathbb{P}}(x_1, \dots, x_N) &= \prod_{i=1}^N f_{S_i | \mathcal{F}_{i-1}}^{\mathbb{P}}(x_i) \\ &= \prod_{i=1}^N g \left(f_{S_i | \mathcal{F}_{i-1}}^{\mathbb{Q}}(x_i | \hat{\theta}_{1,i-1}) \mid \theta_2 \right) .\end{aligned}$$

Consequently, the log-likelihood function for the parameters of the transformation function g is given by [122, p. 1508]

$$\mathcal{L}(\theta_2 | x_1, \dots, x_N) = \sum_{i=1}^N \ln \left[g \left(f_{S_i | \mathcal{F}_{i-1}}^{\mathbb{Q}}(x_i | \hat{\theta}_{1,i-1}) \mid \theta_2 \right) \right] . \quad (2.32)$$

²¹Liu et al. [122] suppose to use a mixture of two log-normal densities and a generalized beta distribution as parametric risk-neutral density. In contrast, Shackleton et al. [143] suppose to model the risk-neutral density via some diffusion or jump-diffusion process. The latter approach has the advantage that a stochastic process specifies the risk-neutral density for all time horizons.

2. Asset pricing

The parameters of the transformation function g can now be estimated by maximizing (2.32) with respect to θ_2 , which is

$$\hat{\theta}_2 = \arg \max_{\theta_2} \mathcal{L}(\theta_2 | x_1, \dots, x_N) .$$

Note that the parameter vector of the transformation function θ_2 is time-invariant, so that the market participants are supposed to have constant risk preferences, irrespective of the market conditions.

2.5. Empirical pricing kernels

2.5.1. Stylized facts

In the following, we will describe the empirical properties of pricing kernels based on different pricing kernel estimates drawn from the literature. Similarly as in the previous section, the pricing kernel will now be defined as the ratio of the risk-neutral density to the real-world density, i.e. we neglect the (subjective) discount factor.²² Hence, the *risk-neutral density* (RND), *real-world density* (RWD) and *pricing kernel* (PK) are related as follows:

$$\text{RND} = \text{RWD} \times \text{PK} . \quad (2.33)$$

In particular, if we know any two of the three components in (2.33), we can recover the third. Before we discuss the empirical properties of pricing kernels found in the literature, we first categorize the therein employed estimation methods:

1. Independently estimate the risk-neutral and real-world density, which are respectively specified by some (non)parametric model. Then the pricing kernel can be inferred by dividing the risk-neutral density through the real-world density.
2. Jointly estimate the risk-neutral and real-world density, which are both specified within the framework of an option pricing model. Then the risk premia of the supposed option pricing model can be determined, which also specify the corresponding pricing kernel.
3. First estimate the risk-neutral density, which is specified by some (non)parametric model. Then choose a parametric pricing kernel based on which the risk-neutral density can be related to the real-world density. Finally, the parameters of the pricing kernel can be estimated based on asset price observations since the risk-neutral density together with the pricing kernel results in a parametric real-world density.²³

²²This pricing kernel specification is also commonly used in the literature (see e.g. Jackwerth [99]).

²³Of course, the order of this estimation approach can be reversed, i.e. the real-world density is first estimated based on some (non)parametric method and then related to the risk-neutral density via

(b)

(c)

Figure 2.3.: Various pricing kernel estimates in the literature. Figure (a) is a copy from Jackwerth [100, p. 57, Fig. 11], Figure (b) from Aït-Sahalia and Lo [1, p. 36, Fig. 3] and Figure (c) from Rosenberg and Engle [138, p. 361, Fig. 5].

2. Asset pricing

The *first approach* has the advantage that no particular assumptions about the pricing kernel are required. Moreover, there already exists a large number of (non)parametric models, which are well able to capture the properties of the risk-neutral and statistical asset price dynamics, respectively.²⁴ Hence, if the risk-neutral and real-world density are adequately estimated, their ratio should also reflect the most important properties of pricing kernels. Jackwerth [99] follows this approach and estimates the risk-neutral density by fitting a smooth function to the volatility smile implied by a cross-section of S&P 500 index options.²⁵ This implied volatility function allows to compute Black-Scholes option prices for the whole continuum of strike prices, which can further be used to compute the corresponding risk-neutral density.²⁶ Furthermore, Jackwerth [99] recovers the real-world density via kernel density estimation and finds that pricing kernels “dramatically change shapes around the 1987 crash [99, p. 435]”. More precisely, pricing kernels are monotonically decreasing before the crash and exhibit increasing regions thereafter. Similarly, Jackwerth [100] estimates pricing kernels for different markets, which are illustrated in Figure 2.3(a). Note that, except for the Japanese market, all pricing kernels feature increasing regions for moderate returns around -10% to $+10\%$. In the following, this particular shape of the pricing kernel will be referred to as *S-shape*. Alternatively, Aït-Sahalia and Lo [1] estimate the risk-neutral and real-world density for the S&P 500 index via kernel density estimation. The resulting pricing kernel estimate is illustrated in Figure 2.3(b), which also exhibits the S-shape. Finally, the German market is inspected by Grith et al. [89], who also recover S-shaped pricing kernels for the DAX. More precisely, Grith et al. [89] estimate time series of pricing kernels by virtue of shape invariant modeling. Thereby, the shape of some starting pricing kernel is only changed in scale and location. Again, the starting pricing kernel is calculated as the ratio of risk-neutral to real-world density, which are both estimated nonparametrically.

The *second approach* is an integrated approach, where the risk-neutral density, real-world density and pricing kernel are all specified within the framework of an option pricing model. The joint estimation of the risk-neutral and statistical asset price dynamics based on time series of asset prices and corresponding option prices allows to explicitly account for time variation in the pricing kernel. The properties of the pricing kernel and, in particular, the possibility for time variation, depend on the underlying option pricing model. For example, the pricing kernel of the Black-Scholes model is time-invariant and monotonically decreasing. However, if we impose too much structure on the pricing kernel, the true empirical properties might not be captured. Among others, Bakshi and Wu [16], Eraker [74], Jones [104] and Pan [130] jointly estimate the risk-neutral and statistical asset price dynamics, which are either specified by a pure diffusion or jump-diffusion process. Moreover, the applied statistical techniques for jointly estimating the risk-neutral and real-world density vary and include filtering techniques, the Bayesian

some parametric pricing kernel, which in turn is estimated based on option price data.

²⁴See Taylor [148] for an overview over different approaches to model the asset price dynamics.

²⁵The term “volatility smile” vividly describes the relation between strike price and Black-Scholes implied volatility. The latter refers to the volatility parameter of the Black-Scholes model, which is chosen such that the Black-Scholes model exactly matches the observed option price.

²⁶Breeden and Litzenberg [41] show that the risk-neutral density of the underlying can be recovered in a model-free fashion by twice differentiating the call option price with respect to the strike price. More precisely, they show that $f_Q(S_T) = e^{r\tau} \partial^2 \mathcal{C}(S_t, K, \tau) / \partial K^2 |_{K=S_T}$, where $\tau = T - t$.

approach and the generalized method of moments. The main interest of these authors was to determine the sign and magnitude of the risk premia of the inspected option pricing models. However, none of these authors explicitly calculated the corresponding pricing kernels. This question is partly worked out by Brown and Jackwerth [44], who calculate the pricing kernel implied by the risk premia estimates of Pan [130] and find that the resulting pricing kernel is monotonically decreasing. Moreover, Brown and Jackwerth [44] point out that if the jump intensity of the supposed asset price process is proportional to the jump risk premium, which is a common assumption in the literature, then the pricing kernel is monotonically decreasing and thus cannot feature the typical S-shape found in the first estimation approach [44, p. 15 f.].

The *third approach* is partly similar to the first approach since the risk-neutral density is estimated independently from the real-world density based on some (non)parametric method. But in contrast to the first approach, the risk-neutral density is then related to the real-world density via some parametric pricing kernel, which also allows to account for time variation in the pricing kernel. However, if we relate the risk-neutral density to the real-world density via some parametric pricing kernel, it must be ensured that this pricing kernel is flexible enough to capture the true empirical properties. Similarly as Jackwerth [99], Bliss and Panigirtzoglou [31] estimate the risk-neutral density by fitting a cubic spline to the volatility smile implied by cross-sections of FTSE 100 index options as well as S&P 500 index options. These risk-neutral density estimates are then transformed to real-world densities via the utility transformation (2.25), which is either specified by a power utility or exponential utility function.²⁷ Bliss and Panigirtzoglou [31] find that their risk aversion estimates take on plausible values and, in addition, depend on the current level of volatility. The latter indicates that risk preferences are time-varying. Moreover, their pricing kernel estimates are monotonically decreasing, which is due to the supposed utility functions. Alternatively, Rosenberg and Engle [138] first estimate the real-world density based on a *generalized autoregressive conditional heteroscedasticity* (GARCH) model and then translate this real-world density estimate to a risk-neutral density via two different parametric pricing kernels.²⁸ In the following, any member of the GARCH family will simply be referred to as GARCH-type model. The first pricing kernel is based on the power utility function and the second pricing kernel based on a more flexible four-parameter orthogonal polynomial. Rosenberg and Engle [138] find that the polynomial specification fits option prices much better than the power utility specification. Moreover, as can be seen in Figure 2.3(c), the pricing kernel estimates based on the polynomial specification are S-shaped, irrespective of the estimation date. Further note, that the polynomial pricing kernel estimates differ from each other, which indicates that the shape of pricing kernels, and hence also risk aversion, is time-varying. In particular, Rosenberg and Engle [138] find that risk aversion is autocorrelated, mean-reverting and counter cyclical. Finally, Alonso et al. [3] suppose a mixture of two log-normal densities as risk-neutral density, which they estimate based on Spanish IBEX 35 index options. Similarly as Bliss and Panigirtzoglou

²⁷In contrast to Chapter 2.4.3, Bliss and Panigirtzoglou [31] estimate the utility transformation by minimizing the test statistic of Berkowitz [26], which checks if the cumulative probabilities of the observed asset prices are uniformly and independently distributed.

²⁸The GARCH model was first proposed by Bollerslev [33]. For an introduction to GARCH models see Bollerslev et al. [35]. Moreover, Bollerslev [34] gives an overview over the different GARCH models.

2. Asset pricing

[31], these risk-neutral density estimates are then transformed to real-world densities via power and exponential utility functions. In addition, Alonso et al. [3] suppose a modified version of the power utility function, where the coefficient of relative risk aversion is modeled as a time-varying process.²⁹ In particular, Alonso et al. [3] also find evidence that risk aversion is time-varying.

All together, we conclude that there is strong evidence that empirical pricing kernels are on the one hand S-shaped and on the other hand time-varying. However, the knowledge about the time series properties of pricing kernels is limited and there are only few researchers, who address this issue.³⁰

2.5.2. The pricing kernel puzzle

In the previous section, we described the empirical properties of pricing kernels based on the results in the literature. In particular, we found that pricing kernels tend to be S-shaped. The increasing region of the pricing kernel however suggests that also marginal utility is increasing, which corresponds to a locally convex utility function. Market participants are thus risk seeking over a specific range of wealth, which is in contrast to economic theory [44, p. 4 f.]. Moreover, S-shaped pricing kernels cannot be captured by common utility functions, such as exponential or power utility, wherefore this phenomenon is also called *pricing kernel puzzle*. This puzzle was first recognized by Aït-Sahalia and Lo [1], Jackwerth [99] and Rosenberg and Engle [138]. Since then researchers work on possible explanations and models, which are able to capture S-shaped pricing kernels. Ziegler [159] investigates within the framework of a consumption-based model if i.) misestimation of the risk-neutral or statistical asset price dynamics, ii.) misestimation of investors beliefs or iii.) heterogeneous investor beliefs can explain the pricing kernel puzzle. Ziegler [159] finds that neither the first nor the second hypothesis can resolve the puzzle. In contrast, he could show that heterogeneous investor beliefs can result in S-shaped pricing kernels. However, Ziegler [159] already noted that the necessary degree of heterogeneity is implausibly large. In addition, Brown and Jackwerth [44] criticize Zieglers [159] “polar setting of having only pessimists and optimists” [44, p. 17] and point out that the pricing kernel puzzle recurs if there also exists a considerable group of moderate investors. Alternatively, Brown and Jackwerth [44] propose a consumption-based model, where the volatility dynamics of the wealth process are extended by an additional state variable. In particular, their model allows to produce S-shaped pricing kernels. Similarly, Benzoni et al. [25] and Chabi-Yo et al. [52] consider consumption-based models, where the representative investor has state-dependent utility, so that his risk aversion depends on the wealth level, which in turn allows to produce S-shaped pricing kernels. Most notably in this context is the contribution of Hens and Reichlin [94], who show that by relaxing one of the three standard assumptions – complete markets, averse investors and correct beliefs – pricing

²⁹More precisely, Alonso et al. [3] model the coefficient of relative risk aversion by a (non)linear model, where the relative risk aversion is either explained by past returns or implied volatility.

³⁰One example is Grith et al. [89], who propose to model time-varying pricing kernels via shape invariant modeling.

kernels can exhibit increasing regions. In particular, if the assumption of complete markets is not fulfilled, i.e. not all risk faced by the investors can completely be hedged, such as, for example, labor income risk, pricing kernels can have a very flexible shape [94, p. 4].

2.5.3. Applications in forecasting – a literature review

Since decades the forecasting of future return volatility or, more generally, the forecasting of future return distributions is an active and important field of research with applications in option pricing, portfolio optimization and risk management. Whereas the risk-neutral asset price dynamics are relevant for the purpose of option pricing, portfolio optimization and risk management issues require the knowledge of the statistical asset price dynamics. As mentioned earlier, risk-neutral densities can be transformed to real-world densities by virtue of pricing kernels. But why should we take the detour from the risk-neutral density over the pricing kernel to the real-world density if the real-world density can be estimated directly? Risk-neutral density estimates are based on cross-sections of option prices, which aggregate the *forward-looking* views of many market participants who make their trading decisions based on all currently available information [7, p. 1]. This information is based on different sources, which include the historical time series of the underlying among many others. In contrast, direct real-world density estimates are usually “only” based on the historical time series of the underlying and thus rely solely on *backward-looking* information. Moreover, the backward-looking approach reacts more slowly to changing market conditions than the forward-looking approach since the current market situation is only reflected by a single observation out of the complete historical time series. Unfortunately, the risk-neutral density also contains risk preferences and thus is a biased estimator of future asset prices unless market participants are risk-neutral. However, the risk-neutral density can be adjusted for risk preferences by some pricing kernel, so that the resulting real-world density is also based on forward-looking information. The crucial point of this approach is that the supposed pricing kernel must adequately describe the current risk preferences. Otherwise, also the forward-looking real-world density is a biased estimator of future asset prices. To answer the question if the forward-looking information embedded in option prices is indeed valuable for the purpose of forecasting, we will review the volatility and density forecasting literature with a special focus on the forecasting performance. Before we will outline the recent developments in density forecasting, where risk-neutral densities are transformed to real-world densities by virtue of pricing kernels, we will first discuss if option-implied information is valuable for the purpose of volatility forecasting.

Volatility forecasting

First, there is a large agreement in the literature that volatility forecasts based on option-implied information, as e.g. *Black-Scholes implied volatilities*, outperform volatility forecasts based on historical information, as e.g. via GARCH-type models. Worth mentioning in this context is the survey of Poon and Granger [135], who review 93 articles with regard

2. Asset pricing

to the forecasting performance of the therein analyzed volatility models. In particular, Poon and Granger [135] find that in 17 out of 18 empirical studies Black-Scholes implied volatilities outperform GARCH-type models and in 26 out of 34 empirical studies Black-Scholes implied volatilities outperform historical volatility models, such as historical averages, moving averages or exponential weights of squared returns.

The increasing availability of high frequency data allows to forecast the future volatility by virtue of *realized volatility* models, which refer to volatility estimators based on intra-day returns, such as e.g. the sum of the squared intra-day returns.³¹ Many empirical studies (see e.g. Andersen et al. [8] and Koopman et al. [116]) find that realized volatility models outperform GARCH-type models in forecasting future volatility. However, there is also strong evidence (see e.g. Blair et al. [30], Giot and Laurent [86], Jiang and Tian [102] and Busch et al. [47]) that volatility forecasts based on option-implied information are more accurate than volatility forecasts based on high frequency data. Particularly notable in this context is the recent empirical study of Busch et al. [47], who compare the forecasting performance of Black-Scholes implied volatilities with the one of a highly sophisticated realized volatility model. More precisely, Busch et al. [47] analyze three different markets, namely the bond, the foreign exchange and the stock market, and find that Black-Scholes implied volatilities always outperform their realized volatility model.

Recently, scholars estimate the option-implied volatility in a model-free fashion, i.e. without using any particular option pricing model.³² The probably most prominent example of this model-free approach is the VIX index of the *Chicago Board Options Exchange* (CBOE), which measures the implied volatility of S&P 500 index options.³³ However, Andersen and Bondarenko [7] find that Black-Scholes implied volatilities do have a better forecasting performance than the more complex *model-free implied volatilities*. They explain the inferior forecasting performance of the VIX compared to Black-Scholes implied volatilities by the lack of liquid far in-the-money and out-of-the money options, which are required to calculate model-free implied volatilities. Furthermore, Taylor et al. [149] compare the forecasting performance of Black-Scholes implied volatilities, model-free implied volatilities and GARCH-type volatility estimators for 149 individual US firms as well as for the S&P 100 index. They find that volatility forecasts for the S&P 100 index based on Black-Scholes implied volatilities and model-free implied volatilities outperform volatility forecasts based on GARCH-type models. Moreover, the same results are obtained for 126 out of 149 individual firms if the forecasting horizon corresponds to the expiry date of the option. Then, in addition, Black-Scholes implied volatilities outperform model-free implied volatilities for two third of all firms. Furthermore, Taylor et al. [149] find that the forecasting performance of Black-Scholes implied volatilities and model-free implied volatilities increases with the option trading activity. Interestingly, for over one third of all analyzed firms, one day ahead volatility forecasts based on GARCH-type models perform better than corresponding Black-Scholes implied volatilities or model-free implied

³¹For an introduction to realized volatility see e.g. Andersen and Benzoni [5]. Moreover, a comprehensive survey about the literature on realized volatility is given by Andersen et al. [6].

³²See Carr and Madan [49], Britten-Jones and Neuberger [42], Jiang and Tian [102, 103] and Figlewski [77] for more details about the model-free estimation of option-implied volatility.

³³The VIX index, which is also referred to as “investor fear gauge” (see [154]), is an important indicator of future market conditions. For more details about the VIX index see e.g. Whaley [154, 155].

volatilities.

Even though Black-Scholes implied volatilities are extremely valuable in forecasting future volatility, there is also a broad consensus in the literature (see e.g. Bollerslev and Zhou [39] and the literature therein) that Black-Scholes implied volatility is a biased estimator.³⁴ The bias can be attributed to the variance risk premium, which is usually found to be negative.³⁵ In particular, a negative variance risk premium increases the expected future volatility under the risk-neutral measure compared to the statistical measure, so that Black-Scholes implied volatility is an upwards biased estimator of future volatility. Interestingly, within the framework of a linear regression model, Chernov [53] finds that unbiased estimates of the future volatility can be obtained if Black-Scholes implied volatility is combined with an estimate of the current realized volatility.

Altogether, we conclude that the forward-looking information contained in option prices allows to forecast future volatility more accurate than volatility forecasts based on historical information. However, these risk-neutral volatility forecasts are biased due to the risk preferences embedded in option prices.

Density forecasting

Cross-sections of option prices can also be used to estimate the complete risk-neutral density of the underlying. This risk-neutral density can then be transformed to a forward-looking real-world density via some pricing kernel. Liu et al. [122] examine the forecasting performance of different risk-neutral density specifications, which are estimated based on FTSE 100 index options and then transformed to real-world densities via a utility and calibration transformation, respectively.³⁶ More precisely, Liu et al. [122] suppose three different risk-neutral densities: i) a mixture of two log-normal distributions (see e.g. Ritchey [137]) ii) a generalized beta distribution (see e.g. Anagnou-Basioudis et al. [4]) and iii) a more flexible risk-neutral density derived from a cubic spline, which is fitted to the implied volatility smile (see e.g. Bliss and Panigirtzoglou [31]). Moreover, Liu et al. [122] estimate the real-world density of the underlying based on a GARCH-type model. In the following, real-world densities, which are estimated based on historical data, such as a time series of stock returns, will be referred to as *historical real-world density*. Liu et al. [122] find that the information content of the three different risk-neutral densities is lower than their risk-adjusted counterpart and that real-world densities outperform the historical real-world density.³⁷ Moreover, they find that a combination of real-world densities

³⁴To test whether Black-Scholes implied volatility is a biased estimator of future volatility, one can regress the ex-post realized volatility $RV_{t,t+\Delta t}$ for the period $[t, t + \Delta t]$ on a constant c_1 and the ex-ante Black-Scholes implied volatility $\sigma_{t,t+\Delta t}$ for the same period, which is $RV_{t,t+\Delta t} = c_1 + c_2\sigma_{t,t+\Delta t} + \epsilon_t$, where ϵ_t is an independent and identically distributed random variable with zero mean. In particular, if the constant c_1 is equal to zero and the coefficient c_2 equal to one, Black-Scholes implied volatility is an unbiased estimator of future volatility [39, p. 131].

³⁵For more details about the variance risk premium see Chapter 3.4.5.

³⁶Liu et al. [122] suppose a power utility function for the utility transformation and a beta distribution for the calibration transformation.

³⁷The information content of the risk-neutral densities, real-world densities and historical real-world density is determined by the ex-post likelihood of the FTSE 100 index realizations.

and the historical real-world density has the highest information content. Interestingly, the more flexible risk-neutral density based on the spline method has an inferior forecasting performance compared to the risk-neutral densities based on parametric distributions. Also Kang and Kim [114] estimate risk-neutral densities via a cubic spline, which is fitted to the implied volatility smile of FTSE 100 index options. They find that more flexible pricing kernels derived from utility functions generally improve the forecasting performance of the resulting real-world densities.³⁸ Along the same lines, Alonso et al. [3] find that pricing kernels, which account for time-varying risk-aversion, can improve the forecasting performance. Most notably in this context is the empirical study of Shackleton et al. [143], who compare out-of-sample density forecasts for the S&P 500 index based on option-implied and historical return information. More precisely, Shackleton et al. [143] suppose three different risk-neutral densities, which are specified by the Black-Scholes, the Heston and the Bates model.³⁹ These risk-neutral densities are then transformed to real-world densities via a parametric and non-parametric calibration function as well as a risk-premium adjustment.⁴⁰ Alternatively, Shackleton et al. [143] directly estimate the historical real-world density via a GARCH-type model based on daily and intra-daily returns. In particular, Shackleton et al. [143] find that risk-neutral densities as well as real-world densities outperform historical real-world densities for all forecasting horizons, which encompass 1 day as well as 1, 2, 4, 6, 8 and 12 weeks.⁴¹ Moreover, real-world densities always outperform risk-neutral densities except for a forecasting horizon of 12 weeks. However, depending on the forecasting horizon, different risk-neutral densities and pricing kernels are optimal. Interestingly, the risk-neutral density specified by the Bates model does not improve the forecasting performance of the corresponding real-world density compared to the simpler risk-neutral density specification of the Heston model. A similar empirical study for the FTSE 100 index is conducted by Wang [152], who's results are in line with Shackleton et al. [143].

2.5.4. An empirical example

In the following, we will illustrate the estimation of parametric pricing kernels based on *credit default swap* (CDS) data, which were provided by the company *CDO*². More precisely, we will use weekly CDS spreads of the companies *Adecco*, *BMW*, *Deutsche Telekom*, *Linde*, *Lufthansa*, *Thyssen*, *Valeo* and *VW* with maturities of 2, 3, 4 and 5 years. The data period ranges from March 7, 2005 to May 25, 2009 and encompasses 221 observations. Before we discuss the estimation results in more detail, we shortly explain the concept of a CDS and its valuation via risk-neutral pricing.

³⁸Kang and Kim [114] suppose the following utility functions: exponential utility, hyperbolic absolute risk aversion, power utility and log plus power utility.

³⁹The Heston model is a stochastic volatility model, which will be described in Chapter 3.4 in more detail. The option pricing model of Bates [21] is an extension of the Heston model, which also accounts for jumps in the asset price process.

⁴⁰The risk-premium adjustment is the “initially intended” pricing kernel of an option pricing model, which is specified by Girsanov's theorem.

⁴¹Shackleton et al. [143] evaluate the forecasting performance of each forecasting model by the out-of-sample likelihood of the S&P 500 index realizations.

Following Lando [119, p. 198], a CDS is the most common credit derivative, which allows to hedge the default risk of a bond or credit. Usually, within a CDS contract, the *protection buyer* pays an annual premium until the reference entity either defaults or the contract expires. The premium of the CDS is calculated as the product of the CDS spread and the face value of debt. In the case of a *credit event*, such as bankruptcy, failure to pay or restructuring, the *protection seller* has to compensate the protection buyer for the suffered loss either by paying the face value of the reference security in exchange for the defaulted bond/credit (*physical settlement*) or by paying the difference between the face value and the market value of the defaulted bond/credit (*cash settlement*).

The fair value of a CDS can be determined with the Merton model.⁴² Following Lando [119, p. 8 ff.], the Merton model assumes that a limited company issues equity S_t and debt B_t at time $t = 0$. Moreover, the evolution of the firm value V_t , which corresponds to the sum of market value of equity S_t and debt B_t , is modeled as a geometric Brownian motion, which is

$$dV_t = \mu V_t dt + \sigma V_t dW_t .$$

Finally, the debt is assumed to be a *zero-coupon bond* with face value D and maturity T .⁴³ Hence, the cash flows at time T are as follows [119, p. 9]:

$$B_T = \min\{D, V_T\} \quad \text{and} \quad S_T = \max\{V_T - D, 0\} .$$

By virtue of risk-neutral valuation, the price of a zero-coupon bond can now be determined as follows [119, p. 9]:

$$\begin{aligned} B_t &= \mathbb{E}_{\mathbb{Q}}[e^{-r\tau} \min\{D, V_T\} \mid \mathcal{F}_t] \\ &= e^{-r\tau} \int_0^\infty \min\{D, x\} f_{V_T|\mathcal{F}_t}^{\mathbb{Q}}(x \mid V_t, \sigma) dx , \end{aligned}$$

where $\tau = T - t$. Since the firm value is assumed to follow a geometric Brownian motion, the risk-neutral density of the future firm value V_T conditional on the information at time t is log-normal with⁴⁴

$$f_{V_T|\mathcal{F}_t}^{\mathbb{Q}}(x \mid V_t, \sigma) = \frac{1}{\sqrt{2\pi}bx} \exp\left\{-\frac{(\ln(x) - a)^2}{2b^2}\right\} , \quad (2.34)$$

where

$$a = \ln(V_t) + (r - 1/2\sigma^2)\tau \quad \text{and} \quad b = \sigma\sqrt{\tau} .$$

Further note, that the *credit spread* of a bond is defined as the difference between the continuously compounded yield of a bond and the risk-free rate, which is [119, p. 11 f.]

$$s(t, \tau) = \frac{1}{\tau} \ln(D/B_t) - r .$$

Due to arbitrage considerations, the CDS spread of a bond should equal the corresponding

⁴²See Lando [119] for an overview of alternative credit risk models.

⁴³A zero-coupon bond pays no coupon until the end of maturity.

⁴⁴See Shreve [144, p. 147 f.] for a proof.

2. Asset pricing

credit spread, which is⁴⁵

$$cds(t, \tau) = s(t, \tau) . \quad (2.35)$$

In accordance to Chapter 2.4.3, the parameters of the risk-neutral density of the firm value can now be estimated via the following

$$\hat{V}_t, \hat{\sigma} = \min_{V_t, \sigma} \sum_{i=1}^n [\overline{cds}(t, \tau_i) - cds(t, \tau_i, V_t, \sigma)]^2 , \quad (2.36)$$

where $\overline{cds}(t, \tau_i)$ denotes the time t price of a CDS with τ_i years to maturity. Moreover, $cds(t, \tau_i, V_t, \sigma)$ refers to the corresponding theoretical CDS spread based on the Merton model. In particular, the risk-neutral density of the firm value (2.34) is fully specified by the parameters V_t and σ .

Now, we turn towards the empirical study, where we first computed the firm value and volatility via (2.36) for each observation date and company, respectively. Figures 2.4(a) and 2.4(b) show this exemplarily for the company Valeo. Based on these estimation results, which fully specify the risk-neutral density of the firm value (2.34), the utility and calibration transformation can be estimated as described in Chapter 2.4.3. The only difference in this setting is that the firm value, as opposed to the stock price, cannot be observed directly and thus needs to be replaced by the corresponding estimate $\hat{V}_t$. Moreover, in contrast to Chapter 2.4.3, the maturities of the CDS do not correspond to the sampling frequency of the data. However, since the risk-neutral firm value dynamics are specified by a stochastic process, the risk-neutral density of the future firm value can be calculated for all time horizons.

Next, we estimate the parameters of the utility transformation together with the power utility function and the calibration transformation together with the beta distribution based on a rolling window of 50 observation days. This allows us to check how sensitive these estimates react to changing market conditions. Again, Figures 2.4(c) and 2.4(d) illustrate this exemplarily for the company Valeo. First note, that neither the parameters of the utility transformation nor the parameters of the calibration transformation are constant, which indicates that the risk preferences of market participants are time-varying. Furthermore, since the parameter γ represents the relative risk aversion, Figure 2.4(c) can be interpreted in a way that market participants are getting more and more risk seeking until mid-2007 and more and more risk averse thereafter. In particular, this example shows that the assumption of a time-invariant pricing kernel is crucial.

The previous estimation results can be used for the purpose of *Value-at-Risk* (VaR) forecasting. Usually, the VaR refers to the maximum loss of a position with a holding period of Δt years, which will not be exceeded with a probability of (at least) α .⁴⁶ Within the framework of this section, the VaR of a τ -year CDS with a confidence level of α and a

⁴⁵Note that equation (2.35) only holds approximately since the CDS premium payments are usually discrete.

⁴⁶The risk measure Value-at-Risk will be explained in Chapter 6.5 in more detail.

Figure 2.4.: Firm value, volatility and transformation estimates for the company Valeo based on CDS data. Figures (a) and (b) show the evolution of the firm value and volatility, respectively. Moreover, the parameters of the utility and calibration transformation are illustrated in Figures (c) and (d), respectively. For simplicity, we set the face value of debt to $D = 100$ and the riskfree rate to $r = 0.03$.

holding period of Δt years will be defined as

$$cds(t + \Delta t, \tau, q_{1-\alpha}(V_{t+\Delta t} | \mathcal{F}_t), \sigma) . \quad (2.37)$$

Here $q_{1-\alpha}(V_{t+\Delta t} | \mathcal{F}_t)$ denotes the $1 - \alpha$ quantile of the firm value $V_{t+\Delta t}$ conditional on the information at time t , which can be calculated based on the previous estimates of the risk-neutral and real-world firm value density, respectively. For this empirical study, we suppose a confidence level of 95% and a forecasting horizon of one week.

Table 2.1 summarizes the results of the VaR forecasts for the 2-year CDS spreads of eight different companies.⁴⁷ More precisely, Table 2.1 lists the number of VaR exceptions as well as the corresponding p -values of the test of Kupiec [118], which checks if the proportion of VaR exceptions is consistent with the chosen confidence level.⁴⁸ Note that,

⁴⁷The VaR forecasts for the CDS with maturities of 3, 4 and 5 years are listed in Tables C.1 – C.3 in the Appendix.

⁴⁸For more details about the test of Kupiec [118] see Chapter 6.5.

2. Asset pricing

Table 2.1.: Value-at-Risk forecasts for 2-year CDS spreads.

Company	Utility transformation	Calibration transformation	No transformation
<i>Adecco</i>	8***	7***	9***
<i>p-value</i>	(0.85)	(0.58)	(0.88)
<i>BMW</i>	16*	10***	17*
<i>p-value</i>	(0.02)	(0.62)	(0.01)
<i>Deutsche Telekom</i>	7***	48	9***
<i>p-value</i>	(0.58)	(0.00)	(0.88)
<i>Linde</i>	12***	18	13***
<i>p-value</i>	(0.25)	(0.00)	(0.15)
<i>Lufthansa</i>	11***	5***	11***
<i>p-value</i>	(0.41)	(0.18)	(0.41)
<i>Thyssen</i>	15*	15*	16*
<i>p-value</i>	(0.04)	(0.04)	(0.02)
<i>Valeo</i>	15*	23	12***
<i>p-value</i>	(0.04)	(0.00)	(0.25)
<i>VW</i>	10***	27	12***
<i>p-value</i>	(0.62)	(0.00)	(0.25)

Notes: This table summarizes the number of VaR exceptions out of 171 VaR forecasts for the 2-year CDS spreads of various companies, where we supposed a forecasting horizon of one week and a confidence level of 95%. In order to test if the VaR with a confidence level of α is violated more or less than $(1 - \alpha) \cdot 100\%$ of the time, we apply the LR test of Kupiec [118]. The reported p -values correspond to the test statistic LR_K (6.5), where $0.01 \leq p < 0.05$ is indicated by *, $0.05 \leq p < 0.1$ by ** and $p \geq 0.1$ by ***.

given the null hypothesis is true, the p -value denotes the probability of obtaining a test statistic, which is at least as “extreme” as the realized test statistic. Hence, large p -values indicate that the null hypothesis is true.⁴⁹ In the following, we call a VaR model *slightly accurate* indicated by one asterisk if $0.01 \leq p < 0.05$, *accurate* indicated by two asterisk if $0.05 \leq p < 0.1$ and *highly accurate* indicated by three asterisk if $p \geq 0.1$. Furthermore, VaR models with p -values smaller than 0.01 will be referred to as *inaccurate*. The p -value classifications of the different VaR models summarized in Table 2.1 suggest that the utility transformation does not improve the VaR forecasting performance compared to VaR forecasts based on the risk-neutral density (no transformation). Moreover, the calibration transformation has the worst forecasting performance of all VaR models since, except for the company BMW, the p -value classifications of the other VaR models are either equally good (Adecco, Lufthansa, Thyssen) or better (Deutsche Telekom, Linde, Valeo, VW). Further note that VaR forecasts based on the risk-neutral density and the utility transformation, respectively, are at least slightly accurate.

Altogether, we can conclude that neither the utility transformation nor the calibration transformation is able to improve the forecasting performance compared to VaR-forecasts based on the risk-neutral density. This result is in contrast to the findings of the previous

⁴⁹For more details about hypothesis testing see e.g. Hayashi [93, p. 33 ff.].

empirical studies presented in Chapter 2.5.3. A possible explanation for this “unpleasant result” is the apparant parameter instability of the two transformation methods, which might result in biased real-world density estimates. Moreover, in contrast to the previous literature, we evaluate the VaR forecasting performance, which only requires the left tail of the distribution and not the whole distribution.

3. Affine models

3.1. Introduction

In econometrics as well as in financial mathematics it is important to model asset prices, so that their real empirical characteristics are captured. Moreover, as described in the previous Chapter, it is also crucial to capture the stylized facts of pricing kernels. A class of processes, which offers great flexibility in modeling asset prices while keeping valuation problems, as e.g. option pricing, analytically tractable are *affine processes*. Duffie et al. [73] develop a relatively general framework for affine processes, which can be applied for a wide variety of valuation problems. A prominent example for an affine model is the stochastic volatility model of Heston [95], which is well able to capture the statistical properties of asset prices. Due to the close relation between the variance risk premium and risk aversion, the Heston model also allows to account for risk aversion. However, the Heston model is too restrictive to capture the real empirical characteristics of pricing kernels, as e.g. the commonly found S-shape or time-varying risk preferences. Therefore, we propose an extension of the Heston model, which also allows to model the risk preferences in a realistic fashion by introducing an additional risk aversion process.

In the following, Chapter 3.2 defines the notion of affine processes and shows that the characteristic function of an affine process is known in semi closed-form. Chapter 3.3 describes how option prices can be calculated via Fourier inversion technique based on the characteristic function of the underlying. A detailed description of the Heston model is given in Chapter 3.4. Finally, a new 2-Factor stochastic volatility model is introduced in Chapter 3.5.

3.2. Affine processes

Following Kahl [108, pp. 13 ff.], let $W_t = [W_{1,t} \dots W_{d,t}]^\top$ be a d -dimensional standard Brownian motion defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, which satisfies the usual conditions.¹ Next, define the d -dimensional stochastic process $X_t = [X_{1,t} \dots X_{d,t}]^\top$ via the following SDE

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

¹For technical details, see for example Protter [136].

3. Affine models

The drift vector $\mu : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is assumed to be continuous differentiable and the diffusion matrix $\sigma : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$ is assumed to be twice continuous differentiable. Now, an *affine process* can be defined as follows [108, p. 15]:²

Definition 3.1 (Affine process). *A process X_t is called affine if the covariance matrix $\sigma(x)\sigma(x)^\top$ and the drift $\mu(x)$ are affine in x . That is,*

$$\begin{aligned}\sigma(x)\sigma(x)^\top &= a + \sum_{i=1}^d x_i \alpha_i, \\ \mu(x) &= b + \sum_{i=1}^d x_i \beta_i,\end{aligned}\tag{3.1}$$

for some real-valued $d \times d$ -matrices a and α_i and real-valued d -vectors b and β_i .

As first shown by Duffie et al. [73], the characteristic function of an affine process X_t is known up to the solution of a system of *Riccati equations* and can be specified via the following theorem [78, p. 127]:³

Theorem 3.1. *Suppose X_t is an affine process with drift and covariance matrix specified by (3.1). Then the $\mathcal{F}_t$ -conditional characteristic function of X_T is exponential affine in X_t for all $t \leq T$. That is, there exist $\mathbb{C}$ - and $\mathbb{C}^d$ -valued functions $\phi(t, u)$ and $\psi(t, u)$ with jointly continuous t -derivatives, such that*

$$\mathbb{E}[e^{iu^\top X_T} | \mathcal{F}_t] = e^{\phi(T-t, u) + \psi(T-t, u)^\top X_t},$$

for all $u \in \mathbb{R}^d$ and $t \leq T$. Moreover, ϕ and $\psi = [\psi_1 \dots \psi_d]^\top$ solve the system of Riccati equations

$$\begin{aligned}\partial_t \phi(t, u) &= \frac{1}{2} \psi(t, u)^\top a \psi(t, u) + b^\top \psi(t, u), \\ \partial_t \psi_i(t, u) &= \frac{1}{2} \psi(t, u)^\top \alpha_i \psi(t, u) + \beta_i^\top \psi(t, u), \quad 1 \leq i \leq d,\end{aligned}\tag{3.2}$$

subject to the boundary conditions $\phi(0, u) = 0$ and $\psi(0, u) = iu$. In particular, ϕ is determined by ψ via simple integration:

$$\phi(t, u) = \int_0^t \left(\frac{1}{2} \psi(s, u)^\top a \psi(s, u) + b^\top \psi(s, u) \right) ds.$$

The functions ϕ and ψ will also be called *characteristic exponents*. For some special cases, as e.g. the Heston model, the *ordinary differential equations* (ODE)s (3.2) can be solved explicitly. Unfortunately, in many cases this is not possible, so that numerical methods need to be applied in order to approximate the solution. Duffie et al. [73] propose to

²In the literature there exist alternative definitions for affine processes. Filipović and Mayerhofer [78] e.g. call a process affine if its characteristic function is exponential affine.

³See Filipović and Mayerhofer [78] for a proof.

use the *Runge-Kutta method* in this context, which will be described in Appendix A.3 in more detail.

3.3. Fourier inversion methods

As shown in the previous section, the characteristic function of an affine process is known up to the solution of a system of Riccati equations, which can either be solved in closed-form or via numerical methods. In the following, we will present some Fourier inversion methods, which allow to calculate option prices based on the characteristic function of the underlying. A good survey about Fourier inversion techniques in the context of option pricing is given by Schmelzle [141].

3.3.1. Black-Scholes style formula

Recall from Chapter 2.3 that the time $t = 0$ price of a European call option with underlying S_t , strike price K and maturity T can be determined by

$$\mathcal{C}(K, T) = \mathbb{E}_{\mathbb{Q}_1} [e^{-rT} \max \{S_T - K, 0\}] , \quad (3.3)$$

where $\mathbb{Q}_1$ denotes an equivalent martingale measure with the bank account as numéraire.⁴ In contrast to the risk-neutral pricing formula (2.17), we now supposed a constant risk-free rate r . In accordance to Schmelzle [141, p. 17], equation (3.3) can be reformulated as follows

$$\begin{aligned} \mathcal{C}(K, T) &= \mathbb{E}_{\mathbb{Q}_1} [e^{-rT} \max \{S_T - K, 0\}] \\ &= \mathbb{E}_{\mathbb{Q}_1} [e^{-rT} (S_T - K) \mathbf{1}_{\{S_T \geq K\}}] \\ &= \mathbb{E}_{\mathbb{Q}_2} \left[e^{-rT} S_T \mathbf{1}_{\{S_T \geq K\}} \frac{d\mathbb{Q}_1}{d\mathbb{Q}_2} \right] - e^{-rT} K \mathbb{E}_{\mathbb{Q}_1} [\mathbf{1}_{\{S_T \geq K\}}] \\ &= S_0 \mathbb{Q}_2 [S_T \geq K] - e^{-rT} K \mathbb{Q}_1 [S_T \geq K] . \end{aligned} \quad (3.4)$$

Here $\mathbb{Q}_2$ denotes an alternative equivalent martingale measures with the stock price as numéraire.⁵ Since equation (3.4) has the form of the classical Black-Scholes formula, Schmelzle [141] calls it *Black-Scholes style formula*. The probabilities $\mathbb{Q}_1 [S_T \geq K]$ and $\mathbb{Q}_2 [S_T \geq K]$ can be calculated by integrating the *probability density function* (pdf) of the underlying. Unfortunately, for more complex asset price processes the pdf of the underlying can become very complicated, which aggravates the (numerical) integration.⁶ For

⁴The bank account in this setup is defined as $B_t = e^{rt}$, which is also the classical choice as numéraire.

⁵A *change of numéraire* from the bank account to the stock price can be achieved via the Radon-Nikodým derivative $d\mathbb{Q}_2/d\mathbb{Q}_1 = B_0 S_T / B_T S_0$. For more details on the change of numéraire technique see Geman et al. [83].

⁶An example for such a complex model is the jump-diffusion model of Kou [117], where the computation of the probabilities $\mathbb{Q}_1 [S_T \geq K]$ and $\mathbb{Q}_2 [S_T \geq K]$ involves infinite series.

3. Affine models

some models, as e.g. the stochastic volatility model of Heston [95], the pdf of the underlying is even not known, so that this classical approach cannot be applied. To circumvent this problem, Heston [95] pioneered the use of Fourier analysis to price options by means of the characteristic function. Thereby he utilizes that the probabilities $\mathbb{Q}_1[S_T \geq K]$ and $\mathbb{Q}_2[S_T \geq K]$ can be obtained by Fourier inversion of the characteristic function of the underlying. This can be achieved with the following equation [141, p. 11]:

$$\mathbb{Q}[X \leq x] = \frac{1}{2} + \frac{1}{2\pi} \int_0^\infty \frac{e^{iux} \varphi_X(-u) - e^{-iux} \varphi_X(u)}{iu} du \quad (3.5)$$

$$= \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \Re \left(\frac{e^{-iux} \varphi_X(u)}{iu} \right) du, \quad (3.6)$$

where X is some random variable with characteristic function φ_X and probability measure $\mathbb{Q}$.⁷ Moreover, $\Re(z)$ denotes the real part of a complex variable z . In contrast to the pdf of the underlying, the corresponding characteristic function is often much simpler or even available at all. Even though the Black-Scholes style formula is theoretically appealing, this approach has the drawback of high computational costs due to two numerical integrations together with slow decaying integrands. Hence, we will not further follow this approach and refer to Zhu [158] for more details.⁸

To overcome the drawbacks of the Black-Scholes style formula, Attari [10] proposes a modification of the Black-Scholes style formula, where only one integration is needed. Furthermore, the resulting integrand decays at quadratic rate, whereas the integrands of the Black-Scholes style formula only decay at linear rate. Depending on the option pricing model, Attari [10] shows that his new formulation is up to 50% faster than the Black-Scholes style formula. Moreover, since each numerical integration suffers from an approximation and truncation error, he could also show that his new formulation is more accurate.

3.3.2. Carr-Madan formula

Another disadvantage of the Black-Scholes style formula is its inability to harness the computational power of the *Fast Fourier transform* (FFT). In contrast, Carr and Madan [50] propose a formulation, which allows to price options by using the FFT. This is, in particular, attractive if one needs to calculate option prices for a given maturity but many strike prices, as e.g. encountered in calibration problems. Furthermore, the approach of Carr and Madan [50] involves only one integral, where the integrand decays at quadratic rate. Hence, this formulation has the same advantages as the method proposed by Attari [10] but, as will be shown later, with an additional degree of freedom. Following Carr and

⁷Equation (3.5) is a variant of *Lévy's inversion theorem* and was shown by Gil-Pelaez [84]. Furthermore, the second equation (3.6) can be obtained by using that fact that $\overline{\varphi_X}(u) = \varphi_X(-u)$, where the overlining bar denotes the complex conjugate. Further recall that the real part of a complex number z is defined as $\Re(z) = (z + \bar{z})/2$.

⁸In particular, Zhu [158] describes how the probabilities $\mathbb{Q}_1[S_T \geq K]$ and $\mathbb{Q}_2[S_T \geq K]$ can be obtained via Fourier inversion.

Madan [50, p. 63], we first introduce a slightly different notation for the time $t = 0$ price of a call option by defining

$$\begin{aligned}\mathcal{C}(k, T) &= \mathbb{E}[e^{-rT} (e^{s_T} - e^k) \mathbf{1}_{\{s_T \geq k\}}] \\ &= e^{-rT} \int_{-\infty}^{\infty} (e^u - e^k) \mathbf{1}_{\{u \geq k\}} f_{s_T}^{\mathbb{Q}}(u) du ,\end{aligned}$$

where $k = \ln(K)$ and $s_T = \ln(S_T)$. Furthermore, $f_{s_T}^{\mathbb{Q}}$ denotes the risk-neutral density of the log-asset price s_T . Hence, the characteristic function of s_T can be defined as [50, p. 63]

$$\varphi_{s_T}(x) = \int_{-\infty}^{\infty} e^{ixu} f_{s_T}^{\mathbb{Q}}(u) du .$$

In contrast to Heston [95] who directly works with the characteristic function of the underlying, Carr and Madan [50] follow a different approach and consider the Fourier transformation of the call option price. Unfortunately, the call option price is not square-integrable with respect to k since $\mathcal{C}(k, T)$ converges to the spot price of the underlying for $k \rightarrow -\infty$. To support the integrability of the call option price over the negative domain of integration, Carr and Madan [50, p. 63] introduce the *damping coefficient* $e^{\alpha k}$, with $\alpha > 0$, and consider the *modified call option price*

$$c(k, T) = e^{\alpha k} \mathcal{C}(k, T) ,$$

instead. However, the damping coefficient deteriorates the integrability of $c(k, T)$ over the positive domain of integration. A sufficient condition for the modified call option price to be integrable over the positive domain of integration, and hence also to be square-integrable, is the requirement that the Fourier transformation of the modified call option price

$$\psi(x) = \int_{-\infty}^{\infty} e^{ixk} c(k, T) dk ,$$

is finite for $x = 0$ [50, p. 64]. As can be seen from equation (3.8), $\psi(0)$ is finite if $\varphi_{s_T}(-(\alpha + 1)i)$ is finite. Following Carr and Madan [50, p. 64], the latter requirement is fulfilled if the $(1 + \alpha)^{th}$ moment of S_T exists, i.e.

$$\mathbb{E}[S_T^{1+\alpha}] < \infty , \tag{3.7}$$

which results from the definition of the characteristic function. If the condition (3.7) holds, the Fourier transform of the modified call option price is well defined and can be computed as follows [50, p. 63]:

$$\begin{aligned}\psi(x) &= \int_{-\infty}^{\infty} e^{ixk} c(k, T) dk \\ &= \int_{-\infty}^{\infty} e^{ixk} e^{\alpha k} e^{-rT} \int_{-\infty}^{\infty} (e^u - e^k) \mathbf{1}_{\{u \geq k\}} f_{s_T}^{\mathbb{Q}}(u) du dk \\ &= e^{-rT} \int_{-\infty}^{\infty} f_{s_T}^{\mathbb{Q}}(u) \int_{-\infty}^{\infty} (e^u e^{k(\alpha+ix)} - e^{k(1+\alpha+ix)}) \mathbf{1}_{\{u \geq k\}} dk du\end{aligned}$$

3. Affine models

$$\begin{aligned}
&= e^{-rT} \int_{-\infty}^{\infty} f_{s_T}^{\mathbb{Q}}(u) \int_{-\infty}^u (e^u e^{k(\alpha+ix)} - e^{k(1+\alpha+ix)}) dk du \\
&= e^{-rT} \int_{-\infty}^{\infty} \left[e^u \frac{e^{k(\alpha+ix)}}{\alpha+ix} - \frac{e^{k(1+\alpha+ix)}}{1+\alpha+ix} \right]_{-\infty}^u f_{s_T}^{\mathbb{Q}}(u) du \\
&= e^{-rT} \int_{-\infty}^{\infty} \left(\frac{e^{u(1+\alpha+ix)}}{\alpha+ix} - \frac{e^{u(1+\alpha+ix)}}{1+\alpha+ix} \right) f_{s_T}^{\mathbb{Q}}(u) du \\
&= e^{-rT} \frac{\varphi_{s_T}(x - (\alpha+1)i)}{(\alpha+ix)(1+\alpha+ix)} . \tag{3.8}
\end{aligned}$$

Following Carr and Madan [50, p. 64], the price of a call option can be recovered by a subsequent Fourier inversion of $\psi(x)$ multiplied by the inverse of the damping coefficient, which is

$$\mathcal{C}(k, T) = \frac{e^{-\alpha k}}{2\pi} \int_{-\infty}^{\infty} e^{-ixk} \psi(x) dx . \tag{3.9}$$

Since $\mathcal{C}(k, T)$ is real, it follows that $\psi(x)$ is odd in its imaginary part and even in its real part, so that (3.9) can also be written as

$$C(k, T) = \frac{e^{-\alpha k}}{\pi} \int_0^{\infty} e^{-ixk} \psi(x) dx , \tag{3.10}$$

which will be referred to as *Carr-Madan formula* [50, p. 64].

Kahl and Lord [110] show that for short maturities and, in particular, for far from at-the-money options the integrand in (3.10) gets highly oscillating, which aggravates the numerical integration. We illustrate this behavior in Figure 3.1 for two different strikes prices and maturities based on the Heston model.⁹ To circumvent this problem, Carr and Madan [50] propose a different formulation, which is based on time values and works much better than (3.10) in this specific setting.¹⁰

The actual intention of Carr and Madan [50] was to find an option pricing formula, so that the FFT can be applied. The FFT is an algorithm to efficiently compute the discrete Fourier transform

$$\hat{f}(k) = \sum_{j=0}^{N-1} e^{-2\pi i k \frac{j}{N}} f(x_j) , \quad k = 1, \dots, N , \tag{3.11}$$

where N is a power of 2 [50, p. 67]. Thereby the computational costs are reduced from order N^2 to $N \log_2(N)$ [50, p. 67]. Carr and Madan [50] show that the original formulation (3.10) can be brought into the required discrete form (3.11). This can essentially be achieved by approximating (3.10) via N trapezoids of width η and by defining an equidistant log-strike grid with spacing λ .¹¹ Furthermore, we need to impose the restriction [50, p. 68]

$$\lambda \eta = \frac{2\pi}{N} .$$

⁹The Heston model will be explained in Chapter 3.4, where also the parameters of Figure 3.1 will be defined.

¹⁰For more details on the *time value approach* see Carr and Madan [50, pp. 65-67].

¹¹For more details see Carr and Madan [50, p. 67 f.].

Figure 3.1.: Integrand of the Carr-Madan formula based on the Heston model with parameters $S_0 = 100$, $\sigma_0 = 0.25^2$, $\phi = 0.25^2$, $\kappa = 3$, $\lambda = 0.25$ and $\rho = 0$.

Hence, for a given N , there is a trade off between the accuracy of the numerical integration, which can be adjusted via η , and the refinement of the log-strike grid controlled by λ . In order to get accurate integration results and a fine log-strike grid, we need to choose N large. Carr and Madan [50] e.g. choose $N = 4096$ and $\eta = 0.25$, so that $\lambda = 0.00613$. However, for applications the actual number of required strike prices is by far less than a typical choice of N , so that the calculation of most option prices is “unnecessary”. Furthermore, the imposed strike-grid does not always match the required strike prices, so that option prices need to be interpolated, which is an additional source of inaccuracy.

Zhu [158] compares the FFT approach of Carr and Madan [50] with a direct numerical integration of (3.10). In contrast to the FFT, where due to the restriction of an equidistant integration grid only simple integration methods, as e.g. the Trapezoid rule or the Simpson rule, can be applied, the direct integration allows for more sophisticated methods, as e.g. the Gaussian quadrature [158, p. 92]. Zhu [158] shows that a direct integration of (3.10) with the Gaussian quadrature is 27 times faster than the FFT if one needs to calculate 15 strike prices per maturity, which is a reasonable number from a practitioners point of view [158, p. 93]. Also Kilin [115] confirms the speed advantage of the direct integration compared to the FFT. Furthermore, he proposes a caching technique, which exploits that the characteristic function of the log-asset only depends on the maturity and not on the strike price. Hence, options with the same maturity but different strike prices can be calculated based on the same evaluations of the corresponding characteristic function. We conclude that for most applications the direct integration will be much faster than the FFT.

3.3.3. Lewis formula

Lewis [121] shows that the approach of Carr and Madan [50] can be embedded in a more general framework. Lewis [121] thereby pursues a modular approach, where the *generalized Fourier transformation* of the payoff profile as well as of the underlying risk-neutral asset price dynamics are calculated separately.¹² Option prices can then be obtained by a version of *Parseval's identity*. More precisely, the price of a European option with payoff profile Π and underlying $S_t = S_0 e^{rt+X_t}$, where X_t is some *Lévy process*, is given by [121, p. 2]

$$V(\Pi, T) = \frac{e^{-rT}}{2\pi} \int_{iy-\infty}^{iy+\infty} e^{-iz(\ln(S_0)+rT)} \varphi_{X_T}(-z) \widehat{\Pi}(z) dz, \quad z = x + iy \in \mathcal{S}. \quad (3.12)$$

Here φ_{X_T} denotes the characteristic function of the Lévy process X_T and $\widehat{\Pi}$ the Fourier transformation of the payoff profile Π . Furthermore, the domain of the integrand in (3.12) is restricted to the particular strip in the complex plan $\mathcal{S} = \mathcal{S}_X \cap \mathcal{S}_\Pi$, where $\mathcal{S}_X \subset \mathbb{C}$ is the domain of the characteristic function reflected across the real z -axis and $\mathcal{S}_\Pi \subset \mathbb{C}$ is the domain of the payoff profile.¹³ In order to compute option prices via (3.12), we need to specify a particular contour of integration, which can be achieved by fixing some $\bar{y}$ such that for every $x \in \mathbb{R}$, $z = x + i\bar{y} \in \mathcal{S}$. Lewis [121] shows that the Black-Scholes style formula and the Carr-Madan formula can be obtained as special cases of (3.12) by restricting $\mathcal{S}$ to a particular subset and by applying *residue calculus* [121, pp. 15-16]. Hence, the damping coefficient α of the Carr-Madan formula can be interpreted as a free parameter, which specifies the contour of integration. In particular, a smart choice of α can alleviate the numerical integration. More precisely, the undesirable oscillating behavior of the integrand in (3.10) for options with short maturities can be mitigated by choosing α appropriately (see Figure 3.1(d)). Lee [120] and Kahl and Lord [110] propose some procedures in order to determine the optimal choice of α . Furthermore, Lee [120] extends the approach of Lewis [121] to allow for more general pay-off profiles and stochastic interest rates. The latter can essentially be achieved by considering the characteristic function of the discounted underlying, which was already noted by Bakshi and Madan [14] and Duffie et al. [73].

¹²A generalized Fourier transformation has a domain in $\mathbb{C}$ and not, as usual, in $\mathbb{R}$. For an introduction to Fourier analysis see e.g. Grafakos [87].

¹³Lewis [121] specifies $\mathcal{S}_X$ and $\mathcal{S}_\Pi$ for various asset price processes and payoff profiles.

3.4. Heston model

3.4.1. Statistical asset price dynamics

The Heston model is an extension of the classical Black-Scholes model, which accounts for stochastic volatility. Following Heston [95], the asset price dynamics under the statistical measure $\mathbb{P}$ can be defined as follows:

$$dS_t = \mu S_t dt + \sqrt{\sigma_t} S_t dW_{1,t} , \quad (3.13)$$

$$d\sigma_t = \kappa(\phi - \sigma_t)dt + \lambda\sqrt{\sigma_t}dW_{2,t} . \quad (3.14)$$

Here the Brownian motions are correlated, so that $dW_{1,t}dW_{2,t} = \rho dt$. In contrast to the classical Black-Scholes model, where the variance of the asset price process is assumed to be constant, the variance σ_t is now modeled as a *Cox-Ingersoll-Ross* (CIR) process.¹⁴ The parameter κ controls how fast the current variance returns to its long term average ϕ . The smaller κ , the more persistent is the variance and consequently the higher is the level of *volatility clustering*. The phenomenon of volatility clustering is a well known empirical feature of stock returns and was first noted by Mandelbrot [125]. More precisely, Mandelbrot [125] discovered by observing time series of asset prices that “(...) *large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes* (...) [125, p. 418]”. Furthermore, the variance of σ_t can be adjusted by the parameter λ .¹⁵ The higher the variance of σ_t , the more probability mass is shifted from the center to the tails of the return distribution, which are therefore also called *fat tails*. The property of fat tails is another well known empirical feature of stock returns and can be characterized by the *kurtosis* of a distribution.¹⁶ If the kurtosis of a distribution is larger than three, which is the kurtosis of a normally distributed random variable, a distribution is said to have fat tails [60, p. 226]. Figure 3.2(a) illustrates the return distribution of the Heston model, which apparently has fat tails compared to the normally distributed asset returns of the Black-Scholes model.

Another well known empirical feature of stock returns is the *leverage effect*. As first noted by Black [28], volatility tends to rise when stock prices fall. This phenomenon can be explained by the leverage effect, which is a term of corporate finance and describes the relation between return on equity and leverage. The latter is usually defined as the ratio of debt to equity. If equity decreases, the leverage increases, which in turn results in higher business risk and volatility, respectively.¹⁷ To accommodate the leverage effect, the

¹⁴Even though the SDE for the variance has no explicit solution, Cox et al. [62] show that σ_t conditional on σ_s , with $s < t$, is noncentral chi-square distributed up to a scaling factor. Moreover, the variance process is always positive if the *Feller condition* $2\kappa\phi > \lambda^2$ is satisfied.

¹⁵This can be seen from the expression for the conditional variance, which is given by $\text{Var}[\sigma_t | \sigma_s] = \sigma_s \frac{\lambda^2}{\kappa} (e^{-\kappa(t-s)} - e^{-2\kappa(t-s)}) + \phi \frac{\lambda^2}{2\kappa} (1 - e^{-\kappa(t-s)})^2$ [62, p. 392].

¹⁶The kurtosis is defined as the ratio of the fourth central moment to the squared variance of a random variable.

¹⁷The relation between leverage and business risk and volatility, respectively, becomes clearer if we define the return on equity as $r_E = r_A + \frac{D}{E}(r_A - r_D)$, where r_A denotes the weighted average cost of capital, D and E the market value of debt and equity, respectively, and r_D the cost of debt. If we assume

3. Affine models

Figure 3.2.: Statistical properties of the Heston model, which accounts for (a) fat tails and (b) skewness in the return distribution.

Brownian motions $W_{1,t}$ and $W_{2,t}$ are correlated. In particular, if ρ is negative, volatility tends to rise if stock prices fall. As can be seen in Figure 3.2(b), the leverage effect results in a negatively skewed return distribution. Altogether, the Heston model accounts for the most important statistical features of stock returns.¹⁸

3.4.2. Risk-neutral asset price dynamics

In accordance to Shreve [144, p. 214 ff.] and van der Ploeg [134, p. 247 ff.], we will now derive the risk-neutral asset price dynamics of the Heston model, which are necessary for the purpose of option pricing. Subsequently, we will then discuss the properties of the risk-neutral asset price dynamics.

First recall the asset price dynamics of the Heston model under the statistical measure $\mathbb{P}$, which are defined as

$$\begin{aligned} dS_t &= \mu S_t dt + \sqrt{\sigma_t} S_t dW_{1,t} , \\ d\sigma_t &= \kappa(\phi - \sigma_t)dt + \lambda \sqrt{\sigma_t} dW_{2,t} , \end{aligned}$$

where $dW_{1,t}dW_{2,t} = \rho dt$. The two correlated Brownian motions $W_{1,t}$ and $W_{2,t}$ can also be

that r_D is deterministic, the variance of r_E increases quadratically with the leverage $L = D/E$. For more details on the leverage effect see e.g. Perridon and Steiner [131, p. 498 ff.].

¹⁸For a survey on the empirical properties of stock returns see e.g. Cont [60].

constructed out of two independent Brownian motions $\bar{W}_{1,t}$ and $\bar{W}_{2,t}$ via [134, p. 247]

$$\begin{aligned} W_{1,t} &= \sqrt{1 - \rho^2} \bar{W}_{1,t} + \rho \bar{W}_{2,t} , \\ W_{2,t} &= \bar{W}_{2,t} . \end{aligned}$$

Next, we extend the two standard Brownian motions $\bar{W}_{1,t}$ and $\bar{W}_{2,t}$ by a drift term via

$$\begin{aligned} d\widehat{W}_{1,t} &= d\bar{W}_{1,t} + \gamma_{1,t} dt , \\ d\widehat{W}_{2,t} &= d\bar{W}_{2,t} + \gamma_{2,t} dt , \end{aligned}$$

where $\gamma_t = [\gamma_{1,t} \ \gamma_{2,t}]^\top$ is adapted to the filtration generated by $\bar{W}_t = [\bar{W}_{1,t} \ \bar{W}_{2,t}]^\top$ [134, p. 248]. The drift term $\gamma_{1,t}$ will be referred to as *market price of asset risk* and the drift term $\gamma_{2,t}$ as *market price of variance risk*.¹⁹ Now, the risk-neutral asset price dynamics of the Heston model can be obtained by an application of Girsanov's theorem (see Theorem 2.1), which tells us how to construct an equivalent probability measure $\mathbb{Q}$, such that $\widehat{W}_{1,t}$ and $\widehat{W}_{2,t}$ are two standard Brownian motions without drift. According to equation (2.14), we define the process Z_t by

$$Z_t = \exp \left\{ - \int_0^t \sum_{i=1}^2 \gamma_{i,u} d\bar{W}_{1,u} - \frac{1}{2} \int_0^t \sum_{i=1}^2 \gamma_{i,u}^2 du \right\} , \quad (3.15)$$

and assume that

$$\mathbb{E} \left[\int_0^T \left(\sum_{i=1}^2 \gamma_{i,u}^2 \right) Z_u^2 du \right] < \infty .$$

Now, under the risk-neutral measure $\mathbb{Q}$ specified via (2.16), $\widehat{W}_{1,t}$ and $\widehat{W}_{2,t}$ are two standard Brownian motions without drift. Next, we consider the differential of the discounted asset price process²⁰

$$\begin{aligned} d(e^{-rt} S_t) &= -re^{-rt} S_t dt + e^{-rt} dS_t \\ &= e^{-rt} \sqrt{\sigma_t} S_t \left[\frac{\mu - r}{\sqrt{\sigma_t}} dt + dW_{1,t} \right] \\ &= e^{-rt} \sqrt{\sigma_t} S_t \left[\frac{\mu - r}{\sqrt{\sigma_t}} dt + \sqrt{1 - \rho^2} d\bar{W}_{1,t} + \rho d\bar{W}_{2,t} \right] \\ &= e^{-rt} \sqrt{\sigma_t} S_t \left[\frac{\mu - r}{\sqrt{\sigma_t}} dt + \sqrt{1 - \rho^2} (d\widehat{W}_{1,t} - \gamma_{1,t} dt) + \rho (d\widehat{W}_{2,t} - \gamma_{2,t} dt) \right] \\ &= e^{-rt} \sqrt{\sigma_t} S_t \left[\left(\frac{\mu - r}{\sqrt{\sigma_t}} - \sqrt{1 - \rho^2} \gamma_{1,t} - \rho \gamma_{2,t} \right) dt + \sqrt{1 - \rho^2} d\widehat{W}_{1,t} + \rho d\widehat{W}_{2,t} \right] \\ &= e^{-rt} \sqrt{\sigma_t} S_t \left[\left(\frac{\mu - r}{\sqrt{\sigma_t}} - \sqrt{1 - \rho^2} \gamma_{1,t} - \rho \gamma_{2,t} \right) dt + d\widehat{W}_{1,t} \right] , \end{aligned}$$

¹⁹If the asset price is the only source of risk, the market price of asset risk is usually just called the *market price of risk*.

²⁰For a derivation of the discounted stock price process in the case of a generalized geometric Brownian motion see Shreve [144, p. 216].

3. Affine models

where we set $\widetilde{W}_{1,t} = \sqrt{1 - \rho^2} \widehat{W}_{1,t} + \rho \widehat{W}_{2,t}$. Note that also $\widetilde{W}_{1,t}$ is a standard Brownian motion under the risk-neutral measure $\mathbb{Q}$. The discounted asset price process will be a martingale under $\mathbb{Q}$ if we choose $\gamma_{1,t}$ and $\gamma_{2,t}$ such that the drift term equals zero, i.e.

$$\frac{\mu - r}{\sqrt{\sigma_t}} - \sqrt{1 - \rho^2} \gamma_{1,t} - \rho \gamma_{2,t} = 0 . \quad (3.16)$$

Hence, the market price of asset risk $\gamma_{1,t}$ is defined as

$$\gamma_{1,t} = \frac{1}{\sqrt{1 - \rho^2}} \left(\frac{\mu - r}{\sqrt{\sigma_t}} - \rho \gamma_{2,t} \right) ,$$

which depends on the so far unspecified market price of variance risk.

Finally, we can proceed to determine the risk-neutral dynamics of S_t and σ_t , which can be calculated as follows:

$$\begin{aligned} dS_t &= \mu S_t dt + \sqrt{\sigma_t} S_t dW_{1,t} \\ &= \mu S_t dt + \sqrt{\sigma_t} S_t \left[\sqrt{1 - \rho^2} d\widehat{W}_{1,t} + \rho d\widehat{W}_{2,t} \right] \\ &= \mu S_t dt + \sqrt{\sigma_t} S_t \left[\sqrt{1 - \rho^2} \left(d\widehat{W}_{1,t} - \gamma_{1,t} dt \right) + \rho \left(d\widehat{W}_{2,t} - \gamma_{2,t} dt \right) \right] \\ &= \mu S_t dt + \sqrt{\sigma_t} S_t \left[-\frac{\mu - r}{\sqrt{\sigma_t}} dt + \sqrt{1 - \rho^2} d\widehat{W}_{1,t} + \rho d\widehat{W}_{2,t} \right] \\ &= r S_t dt + \sqrt{\sigma_t} S_t \widetilde{W}_{1,t} , \end{aligned}$$

and

$$\begin{aligned} d\sigma_t &= \kappa(\phi - \sigma_t) dt + \lambda \sqrt{\sigma_t} dW_{2,t} \\ &= \kappa(\phi - \sigma_t) dt + \lambda \sqrt{\sigma_t} \left(d\widehat{W}_{2,t} - \gamma_{2,t} dt \right) \\ &= [\kappa(\phi - \sigma_t) - \lambda \sqrt{\sigma_t} \gamma_{2,t}] dt + \lambda \sqrt{\sigma_t} d\widetilde{W}_{2,t} , \end{aligned}$$

where we set $\widetilde{W}_{2,t} = \widehat{W}_{2,t}$. The term $\lambda \sqrt{\sigma_t} \gamma_{2,t}$ represents the *risk premium* or *expected excess return* of the “asset” variance. Following Heston [95], we assume that this term is proportional to σ_t , which is

$$\lambda \sqrt{\sigma_t} \gamma_{2,t} = \gamma \sigma_t , \quad (3.17)$$

where $\gamma \in \mathbb{R}$.²¹ Then the *market price of variance risk* is given by

$$\gamma_{2,t} = \frac{\gamma}{\lambda} \sqrt{\sigma_t} , \quad (3.18)$$

so that the risk-neutral variance dynamics can be expressed as

$$d\sigma_t = \kappa^*(\phi^* - \sigma_t) dt + \lambda \sqrt{\sigma_t} d\widetilde{W}_{2,t} ,$$

²¹An economic justification for this choice of the variance risk premium will be given in Chapter 3.4.5.

where

$$\kappa^* = \kappa + \gamma \quad \text{and} \quad \phi^* = \kappa\phi/(\kappa + \gamma) .$$

Altogether, the risk-neutral asset price dynamics are given by

$$\begin{aligned} dS_t &= rS_t dt + \sqrt{\sigma_t} S_t d\widetilde{W}_{1,t} , \\ d\sigma_t &= \kappa^*(\phi^* - \sigma_t) dt + \lambda \sqrt{\sigma_t} d\widetilde{W}_{2,t} . \end{aligned} \tag{3.19}$$

Note that the structure of the asset price dynamics are the same under both measures $\mathbb{P}$ and $\mathbb{Q}$. In particular, if we assume that the market price of variance risk is zero, the variance processes are equal under $\mathbb{P}$ and $\mathbb{Q}$.

In praxis, options are often quoted in terms of Black-Scholes implied volatilities. Recall that the term Black-Scholes implied volatility refers to the volatility parameter of the Black-Scholes model, which is chosen such that the Black-Scholes price and the market price of an option agree. More formally, if $\bar{\mathcal{C}}(K, \tau)$ denotes the market price of a call option with strike price K and maturity τ , then the Black-Scholes implied volatility $\sigma_{BS}(K, \tau)$ of $\bar{\mathcal{C}}(K, \tau)$ has to fulfill the relation

$$\bar{\mathcal{C}}(K, \tau) = \mathcal{C}_{BS}(K, \tau, \sigma_{BS}(K, \tau)) ,$$

where $\mathcal{C}_{BS}$ denotes the corresponding Black-Scholes call option price [81, p. 11]. Usually, options for a specific underlying and date can be observed for different strike prices K_i , $i = 1, \dots, n$, and maturities τ_j , $j = 1, \dots, m$, respectively. The collection of all Black-Scholes implied volatilities $\sigma_{BS}(K_i, \tau_j)$ is called *volatility surface*.

The volatility surface can also be characterized by some empirical features.²² The most prominent feature is the so called *volatility smile*, which vividly describes a plot of the Black-Scholes implied volatility against the strike price. The volatility smile can be observed, in particular, for foreign currency options [98, p. 382 ff.]. For equity options, the Black-Scholes implied volatility tends to decrease as the strike price increases, which is sometimes referred to as *volatility skew* [98, p. 385 ff.]. As can be seen in Figure 3.3(a), the Heston model can generate both smile and skew shapes. Moreover, the smile shape can be attributed to the property of fat tails and the skew shape to the leverage effect [98, p. 383 ff.]. Black-Scholes implied volatilities do not only depend on the strike price but also on the maturity. If we regard the Black-Scholes implied volatility solely as a function of the maturity, this is also called the *volatility term structure*. The volatility term structure of the Heston model is determined by the variance process under the risk-neutral measure. In particular, the volatility term structure is upward sloping for $\sigma_0 < \phi^*$ and downward sloping for $\sigma_0 > \phi^*$. Mixon [128] finds that the volatility term structure of the S&P 500 index is often upward sloping and sometimes downward sloping. Figure 3.3(b) illustrates the volatility surface of the Heston model with a smile shape and an upward sloping volatility term structure.

²²For an overview over the empirical properties of volatility surfaces see Cont et al. [61] and the references therein.

3. Affine models

Figure 3.3.: Volatility surface of the Heston model, which accounts for (a) smile and skew shapes and (b) a flexible volatility term structure. The underlying parameters for these illustrations are $\kappa^* = 1$, $\phi = 0.25^2$ and $T = 0.3$ for (a), $\kappa^* = 5$, $\phi = 0.30^2$ and $\rho = 0$ for (b) and $\sigma_0 = 0.25^2$ and $\lambda = 0.5$ for both illustrations. For comparison reasons, the red and green curve in (a) are shifted upwards, such that all curves do have the same minimum.

3.4.3. Pricing kernel

As shown by Christoffersen et al. [56], the asset price dynamics of the Heston model together with a negative variance risk premium parameter γ translate into a U-shaped pricing kernel, which is illustrated in Figure 3.4. Recall from Chapter 2, that pricing kernels incorporate the preferences of the representative investor. Hence, their analysis allow us to draw conclusion about the implied risk aversion. A U-shaped pricing kernel indicates that the representative investor is risk averse and thus willing to pay a premium to receive payments in rather good or bad states of the economy. Bakshi and Madan [15] show that a U-shaped pricing kernel can be explained by heterogeneous beliefs about future equity returns. More precisely, the decreasing region of the pricing kernel is due to a group of risk averse investors, who are long in the equity market and want to hedge their position with corresponding put options. Conversely, the increasing region of the pricing kernel can be explained by a group of risk averse investors, who are short in the equity market and want to hedge their position with corresponding call options. However, the Heston model cannot capture S-shaped pricing kernels, which are commonly found in the empirical literature.²³

²³For more details on the empirical properties of pricing kernels see Chapter 2.5.

Figure 3.4.: Pricing kernel of the Heston model. This figure illustrates the pricing kernel of the Heston model for different values of the variance risk premium parameter γ . The pricing kernels are computed as the ratio of risk-neutral to statistical return distribution based on the following parameterization: $\sigma_t = 0.15^2$, $\phi = 0.25^2$, $\kappa = 10$, $\lambda = 0.5$, $\rho = -0.8$ and $\tau = 1$.

3.4.4. Characteristic function

As described in Chapter 3.3, option prices can be calculated based on the characteristic function of the log-asset price. Furthermore, if the evolution of the log-asset price is specified by an affine process, the corresponding characteristic function can be derived by an application of Theorem 3.1. In accordance to Kahl [108, p. 17 ff.], we will now first show that the Heston model is specified by an affine process and then derive the corresponding characteristic function.

Lemma 3.1. *Let S_t and σ_t be the risk-neutral asset price dynamics of the Heston model specified by (3.19). Then $X_t = [\ln(S_t) \ \sigma_t]^\top$ is an affine process with drift and covariance specified by (3.1), where*

$$b = \begin{bmatrix} r \\ \kappa^* \phi^* \end{bmatrix}, \quad \beta_1 = 0, \quad \beta_2 = \begin{bmatrix} -1/2 \\ -\kappa^* \end{bmatrix}, \quad (3.20)$$

$$a = 0, \quad \alpha_1 = 0, \quad \alpha_2 = \begin{bmatrix} 1 & \lambda \rho \\ \lambda \rho & \lambda^2 \end{bmatrix}. \quad (3.21)$$

Proof. A process with drift and covariance specified by (3.1) is by Definition 3.1 affine. Hence, we only need to check if the resulting dynamics coincide with $X_t = [\ln(S_t) \ \sigma_t]^\top$, which is equivalent to

$$\begin{aligned} d \ln(S_t) &= (r - 1/2 \sigma_t) dt + \sqrt{\sigma_t} d\widetilde{W}_{1,t}, \\ d\sigma_t &= \kappa^* (\phi^* - \sigma_t) dt + \lambda \sqrt{\sigma_t} d\widetilde{W}_{2,t}. \end{aligned} \quad (3.22)$$

3. Affine models

Next, we plug (3.20) in $\mu(x) = b + \sum_{i=1}^2 x_i \beta_i$, it is apparent that the resulting drift term is the same as in (3.22). Moreover, the covariance matrix of (3.22) is given by

$$dX_t dX_t = \begin{bmatrix} \sigma_t & \lambda \rho \sigma_t \\ \lambda \rho \sigma_t & \lambda^2 \sigma_t \end{bmatrix} dt . \quad (3.23)$$

Again, if we plug (3.21) in $\sigma(x)\sigma(x)^\top = a + \sum_{i=1}^2 x_i \alpha_i$, we can see that the resulting covariance matrix coincides with (3.23). $\square$

For the purpose of option pricing, we are only interested in the characteristic function of the log-asset price $\ln(S_T)$, which will be stated in the following proposition.

Proposition 3.1. *Let $\ln(S_t)$ be the risk-neutral log-asset price dynamics of the Heston model specified by (3.22). Then the characteristic function of $\ln(S_T)$ is given by*

$$\mathbb{E}[e^{iu \ln(S_T)} | \mathcal{F}_t] = e^{\phi(T-t, u) + \psi_1(T-t, u) \ln(S_t) + \psi_2(T-t, u) \sigma_t} , \quad (3.24)$$

where

$$\begin{aligned} \phi(t, u) &= iurt + \kappa^* \phi^* \frac{1}{\lambda^2} \left[t(c - \sqrt{d}) - 2 \ln \left(\frac{1 - ge^{-\sqrt{d}t}}{1 - g} \right) \right] , \\ \psi_1(t, u) &= iu , \\ \psi_2(t, u) &= \frac{1}{\lambda^2} (c - \sqrt{d}) \frac{1 - e^{-\sqrt{d}t}}{1 - ge^{-\sqrt{d}t}} , \end{aligned}$$

with $c = \kappa^* - iu\lambda\rho$, $d = c^2 - \lambda^2(iu + u^2)$ and $g = (c - \sqrt{d})/(c + \sqrt{d})$.

Proof. By Lemma 3.1, the Heston model is specified by an affine process. Hence, we can apply Theorem 3.1 together with the particular choice of $u = [u \ 0]^\top$, so that the characteristic exponents of (3.24) have to solve the following ODEs:

$$\partial_t \phi(t) = r\psi_1(t) + \kappa^* \phi^* \psi_2(t) , \quad (3.25)$$

$$\partial_t \psi_1(t) = 0 , \quad (3.26)$$

$$\partial_t \psi_2(t) = -\frac{1}{2}\psi_1(t) - \kappa^* \psi_2(t) + \frac{1}{2}\psi_1(t)^2 + \lambda\rho\psi_1(t)\psi_2(t) + \frac{1}{2}\lambda^2\psi_2(t)^2 , \quad (3.27)$$

subject to the boundary conditions $\phi(0) = 0$, $\psi_1(0) = iu$ and $\psi_2(0) = 0$. For better clarity, we omit the dependence of the characteristic exponents on u . Due to the boundary condition $\psi_1(0) = iu$, the solution of the second ODE (3.26) is simply $\psi_1(t) = iu$. Following Kahl [108, p. 17], the third ODE (3.27) can be written as

$$\frac{dy}{dt} = c_1 y^2 + c_2 y + c_3 , \quad (3.28)$$

where we set $y = \psi_2(t)$, $c_1 = \frac{1}{2}\lambda^2$, $c_2 = iu\lambda\rho - \kappa^*$ and $c_3 = \frac{1}{2}(iu + u^2)$. The quadratic

form of this ODE allows us to write (3.28) as [108, p. 18]

$$\frac{dy}{dt} = c_1 (y - \tilde{c}_2^\mp) (y - \tilde{c}_3^\mp) , \quad (3.29)$$

with

$$\tilde{c}_2^+ = \frac{-c_2 + \sqrt{d}}{2c_1} \quad \text{and} \quad \tilde{c}_3^- = \frac{-c_2 - \sqrt{d}}{2c_1} , \quad (3.30)$$

or

$$\tilde{c}_2^- = \frac{-c_2 - \sqrt{d}}{2c_1} \quad \text{and} \quad \tilde{c}_3^+ = \frac{-c_2 + \sqrt{d}}{2c_1} . \quad (3.31)$$

The coefficients $\tilde{c}_2^\pm$ and $\tilde{c}_3^\mp$, where the superscript indicates the choice of $+$ or $-$ in the nominator, can be obtained by comparison of the coefficients of the equations (3.28) and (3.29). This results in a quadratic equation, wherefore we have two possible solutions. As shown by Albrecher et al. [2] and Lord and Kahl [109, 111], the choice (3.31) is numerically more stable.²⁴ Therefore, we will choose the coefficients (3.31), which will subsequently simply be denoted by $\tilde{c}_2$ and $\tilde{c}_3$.

Next, the ODE (3.29) can be solved by *seperation of variables*, which is

$$\begin{aligned} \frac{dy}{dt} &= c_1 (y - \tilde{c}_2) (y - \tilde{c}_3) , \\ \iff \frac{1}{(y - \tilde{c}_2) (y - \tilde{c}_3)} dy &= c_1 dt , \\ \iff \int \frac{1}{(y - \tilde{c}_2) (y - \tilde{c}_3)} dy &= \int c_1 dt , \\ \iff \frac{1}{\tilde{c}_2 - \tilde{c}_3} \ln \left(\frac{y - \tilde{c}_2}{y - \tilde{c}_3} \right) &= c_1 t + \text{const}. \end{aligned}$$

If we reset $y = \psi_2(t)$ and solve for $\psi_2(t)$, we get

$$\psi_2(t) = \frac{\tilde{c}_2 - \tilde{c}_3 e^{-\sqrt{d}t + (\tilde{c}_2 - \tilde{c}_3)\text{const}}}{1 - e^{-\sqrt{d}t + (\tilde{c}_2 - \tilde{c}_3)\text{const}}} .$$

Due to the boundary condition $\psi_2(0) = 0$, it follows that

$$\text{const} = \frac{1}{\tilde{c}_2 - \tilde{c}_3} \ln \left(\frac{\tilde{c}_2}{\tilde{c}_3} \right) ,$$

²⁴The reason for the numerical instability of the characteristic function (3.24) together with the coefficients (3.30) is that this characteristic function has a branch cut – i.e. discontinuities in the complex plane – if we restrict the complex logarithm to its principal branch. In particular, if we need to calculate option prices with long maturities, the branch cut can lead to pricing errors. In contrast, the characteristic function based on the coefficients (3.31) has no branch cut and thus is numerically more stable.

3. Affine models

and hence

$$\begin{aligned}\psi_2(t) &= \tilde{c}_2 \frac{1 - e^{\sqrt{d}t}}{1 - \frac{\tilde{c}_2}{\tilde{c}_3} e^{-\sqrt{d}t}} \\ &= \frac{1}{\lambda^2} (c - \sqrt{d}) \frac{1 - e^{-\sqrt{d}t}}{1 - g e^{-\sqrt{d}t}} .\end{aligned}$$

Finally, the first ODE (3.25) can simply be solved by integrating the differential with respect to t , which is

$$\begin{aligned}\phi(t) &= \int_0^t [r\psi_1(s) + \kappa^* \phi^* \psi_2(s)] ds , \\ &= iurt + \kappa^* \phi^* \frac{1}{\lambda^2} (c - \sqrt{d}) \int_0^t \frac{1 - e^{-\sqrt{d}s}}{1 - g e^{-\sqrt{d}s}} ds , \\ &= iurt + \kappa^* \phi^* \frac{1}{\lambda^2} (c - \sqrt{d}) \left[t + \frac{g-1}{\sqrt{d}g} \ln \left(\frac{1 - g e^{-\sqrt{d}t}}{1 - g} \right) \right] \\ &= iurt + \kappa^* \phi^* \frac{1}{\lambda^2} \left[t (c - \sqrt{d}) - 2 \ln \left(\frac{1 - g e^{-\sqrt{d}t}}{1 - g} \right) \right] .\end{aligned}\tag{3.32}$$

The integral in the second equation (3.32) can be solved via tedious calculations, which are carried out in Appendix B.1. $\square$

3.4.5. Variance risk premium

In the following section, we will first derive the variance risk premium specification of the Heston model. Furthermore, we will give a possible economic justification for a negative variance risk premium. Subsequently, we will provide some empirical evidence for a negative variance risk premium.

Derivation of the variance risk premium

Heston [95] motivated his choice for the market price of variance risk (3.17) by Breeden's [40] *consumption-based capital asset pricing model* (CCAPM). In Breeden's [40] CCAPM, the instantaneous *expected excess return* or *risk premium* of an asset S_t is given by

$$\mathbb{E}[dS_t] - rdt = rra_t \text{Cov}[dS_t, \frac{dC_t}{C_t}] ,\tag{3.33}$$

where

$$rra_t = -C_t U''(C_t) / U'(C_t) ,$$

denotes the *relative risk aversion*, which is usually positive. Moreover, C_t denotes the consumption rate and U the utility function of the representative investor [40, p. 274 f.].

Accordingly, the risk premium of an asset S_t depends on its covariation to the consumption rate and on the attitude of the representative investor towards risk. In particular, the magnitude of the risk premium increases with the degree of relative risk aversion.

Heston [95] applies this risk premium definition to the “asset” variance and assumes that the relative risk aversion is constant with $rra_t = \tilde{\gamma}$. The latter is fulfilled if the representative investor has power utility.²⁵ Hence, in accordance to equation (3.33), the risk premium of the asset variance is given by

$$\mathbb{E}[d\sigma_t] - rdt = \tilde{\gamma} \text{Cov}[d\sigma_t, \frac{dC_t}{C_t}] .$$

Furthermore, Heston [95] assumes that the consumption rate process is specified as in the general equilibrium model of Cox et al. [62], so that

$$dC_t = \tilde{\mu}\sigma_t C_t dt + \tilde{\lambda}\sqrt{\sigma_t} C_t dW_t .$$

Finally, Heston [95] assumes that the consumption rate process C_t is correlated with the asset price process S_t , which can be achieved if we suppose that the Brownian motions $W_{1,t}$ and W_t are correlated, so that

$$dW_{1,t}dW_t = \tilde{\rho}dt .$$

Now, under these assumptions, the risk premium of the variance process σ_t is given by

$$\mathbb{E}[d\sigma_t] - rdt = \tilde{\gamma}\xi\sigma_t dt ,$$

where $\xi = \tilde{\lambda}\rho\tilde{\rho}$. Hence, the market price of variance risk essentially reflects the risk aversion of the representative investor $\tilde{\gamma}$ scaled by the covariance between the variance and the consumption rate process ξ . In particular, if we set $\gamma = \tilde{\gamma}\xi$, we obtain the risk premium, which we supposed in Chapter 3.4.2 for the derivation of the risk-neutral asset price dynamics.

A reasonable assumption would be that C_t and S_t are positively correlated, i.e. $\tilde{\rho} > 0$. Hence, if the representative investor is invested in the asset S_t , the rate of consumption gets higher if asset prices go up and lower if asset prices go down. Furthermore, due to the leverage effect, S_t and σ_t are negatively correlated, i.e. $\rho < 0$, so that also σ_t and C_t are negatively correlated. Altogether, this results in a negative variance risk premium, provided that the representative investor is risk averse, i.e. $\tilde{\gamma} > 0$. A negative variance risk premium implies that if we are invested in the “asset” σ_t , we earn, on average, less than the risk-free rate. Further note that if we are invested in the asset S_t , a negative variance risk premium increases the compensation you require for holding S_t . This is due to the leverage effect and can be seen from the following equation:

$$\mu - r = \sqrt{1 - \rho^2}\gamma_{1,t}\sqrt{\sigma_t} + \rho\frac{\gamma}{\lambda}\sigma_t ,$$

²⁵Recall that the power utility function is defined as $U(x) = (x^{1-\tilde{\gamma}} - 1) / (1 - \tilde{\gamma})$ and thus has a constant relative risk aversion in the amount of $\tilde{\gamma}$ [148, p. 452].

3. Affine models

which results from (3.16) together with (3.18). Hence, a negative variance risk premium parameter γ increases the compensation you require for holding S_t , provided that $\rho < 0$.

At first glance it might not be intuitive why an investor should buy an asset with a negative risk premium. But a closer look at equation (3.33) can resolve the confusion. We therefore assume that the representative investor is risk averse, i.e. $rra_t > 0$. Hence, the risk premium of an asset S_t is negative if it is negatively correlated to the consumption rate C_t . Now, recall from Chapter 2.2.2 that such an asset tends to pay off in states where consumption would otherwise be low. Since investors dislike volatile consumption streams, they are willing to pay a premium to smooth consumption with such an asset. Next, we will provide some empirical evidence that the variance risk premium is indeed negative.

Evidence for a negative variance risk premium

There is a vast literature on the sign and magnitude of the variance risk premium. The first strand of literature compares variance estimates under the statistical and risk-neutral measure, where a negative difference indicates that the variance risk premium is negative. Among others, Fleming et al. [79], Jackwerth and Rubinstein [101] and Christensen and Prabhala [55] find that Black-Scholes implied volatilities are consistently higher than realized volatilities, which indicates that the variance risk premium is negative. However, Bollerslev et al. [36] show that variance risk premium estimates are biased and inefficient if one uses Black-Scholes implied volatilities as risk-neutral volatility estimates and realized volatilities from daily returns as statistical volatility estimates. In contrast, they suggest to use model-free implied volatilities, which can be computed without using any parametric option pricing model. Furthermore, they make use of high-frequency data in order to get a more accurate estimate of the realized volatility. Bollerslev et al. [36] estimate the variance risk premium via the *generalized method of moments* (GMM) subject to moment conditions derived from a general stochastic volatility model. In particular, they find that the variance risk premium is significantly negative. Moreover, Bollerslev et al. [36] extend the Heston model by specifying a linear but time-varying model for the variance risk premium parameter γ_t , which is explained by a lagged value of γ_t and macro-finance state variables, such as e.g. the *price-earnings* ratio or industrial production.²⁶ Due to the connection between risk aversion and the variance risk premium, Bollerslev et al. [36] call γ_t also *risk aversion index*, which allows for a more flexible modeling of risk preferences through time.²⁷ In particular, they find that the variance risk premium parameter, and hence also risk aversion, is time-varying and responding to the state of the economy, which is illustrated in Figure 3.5. Interestingly, Bollerslev et al. [36] find that their estimates of the variance risk premium parameter predict future stock returns much better than

²⁶More precisely, Bollerslev et al. [36] propose to model the dynamics of the variance risk premium parameter via $\gamma_t = a + b\gamma_{t-1} + \sum_{k=1}^K c_k \text{state}_{k,t-1}$, where $\text{state}_{k,t}$ denotes some macro-finance state variable.

²⁷Recall that the variance risk premium parameter γ represents the risk aversion of the representative investor scaled by the covariance between σ_t and C_t . In particular, if we assume that this covariance is constant, time variation in the variance risk premium parameter is solely due to time variation in risk aversion.

Figure 3.5.: Risk aversion index estimate of Bollerslev et al. [36, p. 242 Fig. 6]

several established predictor variables, as e.g. price-earnings ratios or dividend yields.²⁸ Alternatively, Carr and Wu [51] use the notion of *variance swaps*, which are over-the-counter derivatives that pay the difference between a standard estimate of the realized variance and an (ex ante) fixed variance swap rate. Since a variance swap has zero net market value at entry, the variance swap rate equals the risk-neutral expectation of the future realized variance [51, p. 1311]. Carr and Wu [51] show that the variance swap rate can be approximated by a portfolio composed of specific options, which is closely related to the model-free implied volatility estimates. In particular, they propose to use the difference between the (ex post) realized variance and the variance swap rate in order to quantify the variance risk premium. Carr and Wu [51] find that the variance risk premium is strongly negative and time-varying. Moreover, they show that the variance risk premium cannot be explained by other commonly used risk factors, such as size, book-to-market or momentum. Hence, they conclude that the majority of the variance risk premium is generated by an independent risk factor. The results of the aforementioned studies are mostly based on index options. Interestingly, Driessen, Maenhout and Vilkov [69] could only find variance risk premia embedded in index options but not in options of individual firms.

The second strand of literature estimates option pricing models, where the volatility is modeled as a stochastic process, as e.g. the Heston model. Hence, these models comprise a variance risk premium, which can be determined by jointly estimating the supposed statistical and risk-neutral asset price dynamics. Among others, Bates [21, 22, 23, 24], Chernov and Ghysels [54], Eraker [74], Jones [104] and Pan [130] find that the variance risk premium is negative. Most notably in this context is the contribution of Bakshi and Wu [16], who model the risk premia of a jump-diffusion option pricing model via an autoregressive process. In order to keep the model computationally tractable, they assume that the risk premia evolve deterministically. In particular, Bakshi and Wu [16]

²⁸Similarly, Nyberg and Wilhelmsson [129] find that a two factor model, where the factors are the market factor and estimates of the variance risk premium specification of Bollerslev et al. [36], has larger explanatory power than the three-factor benchmark model of Fama and French [76]. Moreover, Bollerslev et al. [37] show that the variance risk premium has high predictability for stock returns, in particular, at the quarterly horizon.

3. Affine models

find that the average variance risk premium is negative and, in addition, further decreases during the NASDAQ bubble, which also indicates that risk aversion is time-varying.

The third strand of literature researches if variance risk is a systematic and thus a priced risk factor, which helps to explain asset returns. First, there are several studies, which estimate the expected return of a portfolio composed of a long position in a call option and a corresponding delta-hedge with the underlying stock.²⁹ Hence, this portfolio is immune against all risk except variance risk and should earn the risk-free rate if variance risk is not priced [12, p. 528]. Bakshi and Kapadia [12] find that delta hedged index options earn much less than the risk-free rate, which is in line with a negative variance risk premium. In another study, Bakshi and Kapadia [13] show that also individual equity options embed a negative variance risk premium, even though the magnitude of the risk premium is much smaller than for index options. Other researchers follow an econometric approach and regress option returns on variance and other risk factors in order to check if variance is a priced risk factor. Jones [104], Duarte and Jones [70] and Driessen and Maenhout [69] find that a negative variance risk premium helps to explain option returns. Moreover, Ang et al. [9] show that variance risk is priced negatively in the cross-section of stock returns.

3.5. A 2-Factor stochastic volatility model

3.5.1. Statistical asset price dynamics

In the following, we will introduce a new 2-Factor *stochastic volatility* (SV) model. A preliminary version of this model is described in Appendix D, which was the basis for the 2-Factor SV model. The motivation to formulate yet another option pricing model is based on two shortcomings of the Heston model. First, the Heston model is not able to capture S-shaped pricing kernels as commonly observed in the empirical literature (see Chapter 2.5.1). Hence, the risk preferences of market participants cannot be modeled in a realistic fashion. In particular, if the pricing kernel of the Heston model is used to adjust risk-neutral densities for risk preferences, the resulting real-world density might be biased. Second, the dynamics of the variance risk premium and hence also the risk aversion dynamics are too restrictive in the Heston model. This is mainly due to the constant variance risk premium parameter γ . As described in the previous section, Bollerslev et al. [36] find that γ is indeed time-varying and, in addition, depends on the state of the economy. In order to account for these shortcomings, we will extend the Heston model by an additional risk aversion process, which is unrelated to the statistical asset price dynamics but part of the risk-neutral asset price dynamics. More precisely, we assume

²⁹The delta of an option denotes the first partial derivative of the option price with respect to the underlying and also quantifies the number of shares necessary to hedge a short option position [144, p. 157]. The latter is also called delta-hedging.

that the asset price dynamics under the statistical measure are given by

$$dS_t = \mu S_t dt + \sqrt{\sigma_t} S_t dW_{1,t} , \quad (3.34)$$

$$d\sigma_t = \kappa_1(\phi_1 - \sigma_t)dt + \lambda_1 \sqrt{\sigma_t} dW_{2,t} , \quad (3.35)$$

$$d\theta_t = \kappa_2(\phi_2 - \theta_t)dt + \lambda_2 \sqrt{\theta_t} dW_{3,t} , \quad (3.36)$$

where $dW_{1,t} dW_{2,t} = \rho dt$. Furthermore, $W_{3,t}$ is uncorrelated to $W_{1,t}$ and $W_{2,t}$, so that also the process θ_t is uncorrelated to S_t and σ_t . The process θ_t will serve as a proxy variable for risk aversion, which enters the pricing kernel but is not directly observable. Just like the variance process, also the risk aversion process θ_t is a CIR process. In particular, θ_t is a mean-reverting process with long-term average ϕ_2 and mean-speed of reversion κ_2 . Since the risk aversion process θ_t is independent of S_t and σ_t , the 2-Factor SV model has the same statistical asset price properties as the Heston model, wherefore we refer to Chapter 3.4.1 for more details.

3.5.2. Risk-neutral asset price dynamics

Before we describe the properties of the 2-Factor SV model under the risk-neutral measure, we first derive the corresponding asset price dynamics. Similarly as for the Heston model, the risk-neutral asset price dynamics of the 2-Factor SV model can be derived by virtue of Girsanov's theorem. First recall the asset price dynamics under the statistical measure $\mathbb{P}$, which are given by

$$\begin{aligned} dS_t &= \mu S_t dt + \sqrt{\sigma_t} S_t dW_{1,t} , \\ d\sigma_t &= \kappa_1(\phi_1 - \sigma_t)dt + \lambda_1 \sqrt{\sigma_t} dW_{2,t} , \\ d\theta_t &= \kappa_2(\phi_2 - \theta_t)dt + \lambda_2 \sqrt{\theta_t} dW_{3,t} . \end{aligned}$$

Again, the two correlated Brownian motions $W_{1,t}$ and $W_{2,t}$ can be constructed out of two independent Brownian motions $\bar{W}_{1,t}$ and $\bar{W}_{2,t}$ via

$$\begin{aligned} W_{1,t} &= \sqrt{1 - \rho^2} \bar{W}_{1,t} + \rho \bar{W}_{2,t} , \\ W_{2,t} &= \bar{W}_{2,t} . \end{aligned}$$

Next, we define the following Brownian motions with drift

$$\begin{aligned} d\widehat{W}_{1,t} &= d\bar{W}_{1,t} + \gamma_{1,t} dt , \\ d\widehat{W}_{2,t} &= d\bar{W}_{2,t} + \gamma_{2,t} dt , \\ d\widehat{W}_{3,t} &= d\bar{W}_{3,t} + \gamma_{3,t} dt , \end{aligned}$$

where we set $\bar{W}_{3,t} = W_{3,t}$. The process $\gamma_t = [\gamma_{1,t} \ \gamma_{2,t} \ \gamma_{3,t}]^\top$ is assumed to be adapted to the filtration generated by $\bar{W}_t = [\bar{W}_{1,t} \ \bar{W}_{2,t} \ \bar{W}_{3,t}]^\top$. In the following, we will assume that the risk aversion process θ_t does not carry any market price of risk, so that $\gamma_{3,t} = 0$. Hence, the change from the statistical measure $\mathbb{P}$ to an equivalent martingale measure $\mathbb{Q}$

3. Affine models

can be performed via the Radon-Nikodým derivative (3.15) of the Heston model. If we impose the following condition:

$$\frac{\mu - r}{\sqrt{\sigma_t}} - \sqrt{1 - \rho^2} \gamma_{1,t} - \rho \gamma_{2,t} = 0 ,$$

the risk-neutral asset price dynamics are given by

$$\begin{aligned} dS_t &= rS_t dt + \sqrt{\sigma_t} S_t \widetilde{W}_{1,t} , \\ d\sigma_t &= [\kappa_1(\phi_1 - \sigma_t) - \lambda_1 \sqrt{\sigma_t} \gamma_{2,t}] dt + \lambda_1 \sqrt{\sigma_t} d\widetilde{W}_{2,t} , \\ d\theta_t &= \kappa_2(\phi_2 - \theta_t) dt + \lambda_2 \sqrt{\theta_t} d\widetilde{W}_{3,t} , \end{aligned}$$

where $\widetilde{W}_{1,t} = \sqrt{1 - \rho^2} \widehat{W}_{1,t} + \rho \widehat{W}_{2,t}$, $\widetilde{W}_{2,t} = \widehat{W}_{2,t}$ and $\widetilde{W}_{3,t} = \widehat{W}_{3,t}$.³⁰ Finally, we assume that the *market price of variance risk* is defined as³¹

$$\gamma_{2,t} = \frac{\kappa_1(\phi_1 - \sigma_t) - \kappa_1^*(\theta_t - \sigma_t)}{\lambda_1 \sqrt{\sigma_t}} ,$$

so that the risk-neutral variance dynamics are given by

$$d\sigma_t = \kappa_1^*(\theta_t - \sigma_t) dt + \lambda_1 \sqrt{\sigma_t} d\widetilde{W}_{2,t} . \quad (3.37)$$

Altogether, the risk-neutral asset price dynamics of the 2-Factor SV model are given by

$$\begin{aligned} dS_t &= rS_t dt + \sqrt{\sigma_t} S_t d\widetilde{W}_{1,t} , \\ d\sigma_t &= \kappa_1^*(\theta_t - \sigma_t) dt + \lambda_1 \sqrt{\sigma_t} d\widetilde{W}_{2,t} , \\ d\theta_t &= \kappa_2(\phi - \theta_t) dt + \lambda_2 \sqrt{\theta_t} d\widetilde{W}_{3,t} . \end{aligned} \quad (3.38)$$

The risk aversion process θ_t is now part of the variance process and replaces the long-term average ϕ_1 by the central tendency θ_t . Note that Duffie et al. [73] propose to use a similar specification to model asset prices. Furthermore, also Bates [24] supposes a 2-Factor SV model with jumps but assumes that θ_t stays constant over intra-daily intervals in order to keep the model computationally tractable. However, none of these authors proposes to use the second volatility factor θ_t only under the risk-neutral measure. Further note that 2-Factor models of this form are also applied in the interest rate literature, where one factor reflects the central tendency of the short rate (see e.g. Balduzzi et al. [17] and Dai and Singleton [63]).

The risk-neutral asset price dynamics of the 2-Factor SV model allow for a more flexible and realistic volatility term-structure than the Heston model. This is due to the additional risk aversion process, which drives the variance at first into the direction of θ_t . In particular, if the current level of risk aversion differs from its long-term average, i.e. $\theta_t \neq \phi_2$, the variance is meanwhile pushed into a direction contrary to its long-term average. Moreover, the smaller κ_2 the longer it takes θ_t to revert back to its long-term

³⁰For a detailed derivation of the risk-neutral dynamics of S_t and σ_t see Chapter 3.4.2.

³¹See Chapter 3.5.5 for an economic motivation of this market price of variance risk specification.

Figure 3.6.: Volatility term-structure of the 2-Factor SV model for different values of θ_t together with low ($\kappa_2 = 0.5$) and high ($\kappa_2 = 3$) levels of the mean-speed of reversion, respectively. Moreover, for comparison reasons we also show the volatility term-structure of the Heston model. The remaining parameters of the 2-Factor SV model are as follows: $\sigma_0 = 0.25^2$, $\kappa_1^* = 7$, $\phi_2 = 0.35^2$, $\lambda_1 = 0.3$, $\lambda_2 = 0.1$, $\rho = -0.5$. The Heston model is parameterized by $\sigma_0 = 0.25^2$, $\kappa^* = 7$, $\phi = 0.35^2$, $\lambda = 0.3$, $\rho = -0.5$.

Figure 3.7.: Volatility surface of the 2-Factor SV model with parameters $\sigma_0 = 0.25^2$, $\theta_0 = 0.15^2$, $\kappa_1^* = 7$, $\kappa_2 = 0.5$, $\phi_2 = 0.35^2$, $\lambda_1 = 0.3$, $\lambda_2 = 0.1$ and $\rho = 0$.

3. Affine models

average ϕ_2 , which is also the long-term average of the variance process σ_t .

Figure 3.6 illustrates the volatility term-structure of the 2-Factor SV model. In addition, for the purpose of comparison, Figure 3.6 also shows the volatility term-structure of the Heston model. For both models we assume that the underlying variance processes do have the same long-term average, i.e. $\phi = \phi_2$. Furthermore, we suppose three different starting values for the risk aversion process θ_0 . In the first case, we assume that $\theta_0 < \sigma_0 < \phi_2$, so that the variance is first pushed into the opposite direction of its long-term average (magenta line). In the second case, we suppose that $\sigma_0 < \theta_0 = \phi_2$, so that the long-term average of θ_t equals its starting value (blue line). Hence, the 2-Factor SV model resembles the setting of the Heston model (red line). If we choose κ_2 to be large, the volatility term-structure of both models is quite similar (see Figure 3.6(b)). At last, we consider the case $\theta_0 > \phi_2 > \sigma_0$, so that the variance is first pushed to a higher level than its long-term average (green line). Of course, these properties of the 2-Factor SV model also form the shape of the whole volatility surface, which is illustrated in Figure 3.7. However, the general volatility smile/skew structure of the 2-Factor SV model is not different to the one of the Heston model.

The different volatility term-structure shapes of the 2-Factor SV model can also be observed in reality, as illustrated in Figure 3.8 for S&P 500 index options.³² In particular, this kind of volatility term-structure can only be captured by the 2-Factor SV model and not by the Heston model.

Figure 3.8.: Volatility term-structure of S&P 500 index options at July 28, 2011 (a) and September 8, 2011 (b) (Source: <http://www.cboe.com>).

³²The volatility term-structure of S&P 500 index options were calculated by the CBOE and can freely be downloaded at their homepage (<http://www.cboe.com>).

3.5.3. Pricing kernel

In contrast to the Heston model, the 2-Factor SV model is well able to capture S-shaped pricing kernels. Figure 3.9 illustrates this feature of the 2-Factor SV model for different risk aversion levels θ_t while keeping the remaining parameters fix. Note that the supposed parameterization of the 2-Factor SV model reflects a calm market period, where the current level of the variance σ_t is lower than its long-term average ϕ_1 . Moreover, ϕ_1 is supposed to be lower than the long-term average of the risk aversion ϕ_2 , which is in line with the estimation results of the empirical study of Chapter 6. The lower the current level of risk aversion, the more pronounced is the S-shape of the pricing kernel. In contrast, for high levels of current risk aversion, the pricing kernel gets U-shaped. Further note that the pricing kernel with $\theta_t = 0.20^2$ (blue line) looks pretty similar to the pricing kernel estimates of Aït-Sahalia and Lo [1] and Rosenberg and Engle [138], which are illustrated in Figure 2.3.

Figure 3.9.: Pricing kernel of the 2-Factor SV model. This figure illustrates the pricing kernel of the 2-Factor SV model for different values of the risk aversion θ_t . The pricing kernels are computed as the ratio of risk-neutral to statistical return distribution based on the following parameterization: $\sigma_t = 0.15^2$, $\phi_1 = 0.25^2$, $\phi_2 = 0.35^2$, $\kappa_1 = 10$, $\kappa_1^* = 5$, $\kappa_2 = 1$, $\lambda_1 = 0.5$, $\lambda_2 = 0.5$, $\rho = -0.8$ and $\tau = 1$.

3.5.4. Characteristic function

For the purpose of option pricing, we are interested in the characteristic function of the log-asset price $\ln(S_T)$. If the log-asset price follows an affine process, the corresponding characteristic function can be computed via Theorem 3.1.

Lemma 3.2. *Let S_t , σ_t and θ_t be the risk-neutral asset price dynamics of the 2-Factor SV model specified by (3.38). Then $X_t = [\ln(S_t) \ \sigma_t \ \theta_t]^\top$ is an affine process with drift and*

3. Affine models

covariance specified by (3.1), where

$$b = \begin{bmatrix} r \\ 0 \\ \kappa_2 \phi_2 \end{bmatrix}, \quad \beta_1 = 0, \quad \beta_2 = \begin{bmatrix} -1/2 \\ -\kappa_1^* \\ 0 \end{bmatrix}, \quad \beta_3 = \begin{bmatrix} 0 \\ \kappa_1^* \\ -\kappa_2 \end{bmatrix}, \quad (3.39)$$

$$a = 0, \quad \alpha_1 = 0, \quad \alpha_2 = \begin{bmatrix} 1 & \lambda_1 \rho & 0 \\ \lambda_1 \rho & \lambda_1^2 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \alpha_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \lambda_2^2 \end{bmatrix}. \quad (3.40)$$

Proof. A process with drift and covariance specified by (3.1) is by Definition 3.1 affine. Hence, we only need to check if the resulting dynamics coincide with $X_t = [\ln(S_t) \ \sigma_t \ \theta_t]^\top$, which is equivalent to

$$\begin{aligned} d \ln(S_t) &= (r - 1/2 \sigma_t) dt + \sqrt{\sigma_t} d\widetilde{W}_{1,t}, \\ d\sigma_t &= \kappa_1^* (\theta_t - \sigma_t) dt + \lambda_1 \sqrt{\sigma_t} d\widetilde{W}_{2,t}, \\ d\theta_t &= \kappa_2 (\phi_2 - \theta_t) dt + \lambda_2 \sqrt{\theta_t} d\widetilde{W}_{3,t}. \end{aligned} \quad (3.41)$$

If we plug (3.39) in $\mu(x) = b + \sum_{i=1}^3 x_i \beta_i$, it is apparent that the resulting drift term is the same as in (3.41). Moreover, the covariance matrix of (3.41) is given by

$$dX_t dX_t = \begin{bmatrix} \sigma_t & \lambda_1 \rho \sigma_t & 0 \\ \lambda_1 \rho \sigma_t & \lambda_1^2 \sigma_t & 0 \\ 0 & 0 & \lambda_2^2 \theta_t \end{bmatrix} dt. \quad (3.42)$$

Again, if we plug (3.40) in $\sigma(x)\sigma(x)^\top = a + \sum_{i=1}^3 x_i \alpha_i$, we can see that the resulting covariance matrix coincides with (3.42). $\square$

Hence, the characteristic function of the log-asset price $\ln(S_T)$ can be calculated by virtue of Theorem 3.1.

Proposition 3.2. *Let $\ln(S_t)$, σ_t and θ_t be the risk-neutral log-asset price dynamics of the 2-Factor SV model specified by (3.41). Then the characteristic function of $\ln(S_T)$ is given by*

$$\mathbb{E}[e^{iu \ln(S_T)} | \mathcal{F}_t] = e^{\phi(T-t,u) + \psi_1(T-t,u) \ln(S_t) + \psi_2(T-t,u) \sigma_t + \psi_3(T-t,u) \theta_t}, \quad (3.43)$$

where

$$\begin{aligned} \phi(t, u) &= iurt + \kappa_2 \phi_2 \int_0^t \psi_3(s, u) ds, \\ \psi_1(t, u) &= iu, \\ \psi_2(t, u) &= \frac{1}{\lambda^2} (c - \sqrt{d}) \frac{1 - e^{-\sqrt{d}t}}{1 - ge^{-\sqrt{d}t}}, \\ \partial_t \psi_3(t, u) &= \kappa_1^* \psi_2(t, u) - \kappa_2 \psi_3(t, u) + 1/2 \lambda_2^2 \psi_3(t, u)^2, \quad \psi_3(0, u) = 0, \end{aligned} \quad (3.44)$$

with $c = \kappa_1^* - iu \lambda_1 \rho$, $d = c^2 - \lambda_1^2 (iu + u^2)$ and $g = (c - \sqrt{d}) / (c + \sqrt{d})$.

Proof. According to Lemma 3.2, X_t is an affine process, so that the characteristic function of $\ln(S_T)$ can be calculated by virtue of Theorem 3.1. In particular, the ODEs of the characteristic exponents in (3.43) are simply functions of the affine drift and covariance specification (3.39) and (3.40), respectively. Since we are interested in the characteristic function of $\ln(S_T)$, we make the particular choice of $u = [u \ 0 \ 0]^\top$. Hence, the ODEs for the characteristic exponents are given by

$$\partial_t \phi(t) = r\psi_1(t) + \kappa_2 \phi_2 \psi_3(t) , \quad (3.45)$$

$$\partial_t \psi_1(t) = 0 , \quad (3.46)$$

$$\partial_t \psi_2(t) = -1/2\psi_1(t) - \kappa_1^* \psi_2(t) + 1/2\psi_1(t)^2 + \lambda_1 \rho \psi_1(t) \psi_2(t) + 1/2\lambda_1^2 \psi_2(t)^2 , \quad (3.47)$$

$$\partial_t \psi_3(t) = \kappa_1^* \psi_2(t) - \kappa_2 \psi_3(t) + 1/2\lambda_2^2 \psi_3(t)^2 , \quad (3.48)$$

subject to the boundary conditions $\phi(0) = 0$, $\psi_1(0) = iu$, $\psi_2(0) = 0$ and $\psi_3(0) = 0$. Again, for better clarity, we omitted the dependence of the characteristic exponents on u . The ODEs (3.46) and (3.47) are the same as in the Heston model, wherefore we refer to the proof of Proposition 3.1 for a detailed derivation. Unfortunately, $\psi_3(t, u)$ cannot be calculated in closed-form, wherefore we only state the corresponding ODE at this point. Based on the numerical solution of (3.48), the ODE (3.45) can then be solved via numerical integration. $\square$

For applications, the ODE (3.44) has to be solved numerically. A commonly used approach for this kind of problem is the *Runge-Kutta method*, which is described in Appendix A.3 in more detail.

3.5.5. Variance risk premium

The variance risk premium of the 2-Factor SV model is defined as

$$\begin{aligned} \mathbb{E}[d\sigma_t] - rdt &= [\kappa_1 (\phi_1 - \sigma_t) - \kappa_1^* (\theta_t - \sigma_t)] dt , \\ &= [\kappa_1 (\phi_1 - \theta_t) + \gamma (\sigma_t - \theta_t)] dt , \end{aligned} \quad (3.49)$$

where $\gamma = \kappa_1^* - \kappa_1$. Hence, the sign of the variance risk premium depends on the parameters κ_1 and γ as well as on the current level of risk aversion θ_t compared to the long-term average of the variance ϕ_1 and the current level of the variance σ_t , respectively. The sign of the variance risk premium can easily be determined via the following boundary:

$$b(\sigma_t) = \phi_1 + \frac{\gamma}{\kappa_1^*} (\sigma_t - \phi_1) ,$$

which can be obtained by setting (3.49) equal to zero and then solving for θ_t . Since the variance risk premium specification (3.49) is decreasing in θ_t , the variance risk premium is positive if $\theta_t < b(\sigma_t)$ and negative if $\theta_t > b(\sigma_t)$. Furthermore, if $\gamma < 0$ and $\kappa_1^* > 0$,

3. Affine models

the boundary $b(\sigma_t)$ is decreasing in the current level of the variance σ_t .³³ Hence, during normal market conditions (low levels of σ_t) a negative variance risk premium is less likely since the boundary for θ_t to cross is high. Conversely, during turbulent market conditions (high levels of σ_t) a negative variance risk premium is more likely since the boundary for θ_t to cross is low. Due to the connection between the variance risk premium and risk aversion, a positive/negative variance risk premium indicates risk seeking/risk averse market participants.³⁴ Hence, the mechanism of the 2-Factor SV model for the sign of the variance risk premium supposes that market sentiment is rather relaxed during normal market conditions and rather stressed during turbulent market conditions, which is economically intuitive. Further note that the 2-Factor SV model can capture both, positive and negative variance risk premia over time. In contrast, the Heston model can only capture either positive or negative variance risk premia.

³³Note that empirical studies suggest that the mean speed of reversion of the variance is usually smaller under the risk-neutral measure than under the statistical measure (see e.g. Bakshi and Wu [16], Bates [24] and Pan [130]). Hence, it is reasonable to assume that $\gamma < 0$.

³⁴Recall from Chapter 3.4.5 that the variance risk premium can be specified by $rra_t \text{Cov}[d\sigma_t, \frac{dC_t}{C_t}]$. Due to the leverage effect, the covariance term is typically negative. Hence, a negative variance risk premium indicates that $rra_t > 0$, i.e. market participants are risk averse, and a positive variance risk premium indicates that $rra_t < 0$, i.e. market participants are risk seeking.

4. Estimation of option pricing models

4.1. Introduction

This chapter describes how to jointly estimate the statistical and risk-neutral asset price dynamics of an option pricing model. The proposed estimation method is based on the *unscented Kalman filter* (UKF), which is an extension of the classical Kalman filter. The Kalman filter is an algorithm in order to estimate the state of a time-varying system, which is indirectly observed through noisy measurements. Originally, the Kalman filter was applied in engineering applications, as for example in the *global positioning system* (GPS).¹ Recently, the Kalman filter is also applied in the context of financial econometrics, where the state of the system refers to unobservable economic variables, which are estimated based on related financial data.² Furthermore, the Kalman filter can be used to estimate the parameters of the underlying model, which in our context can be utilized to estimate option pricing models.

In the following, Chapter 4.2 derives the classical Kalman filter together with the underlying state-space representation. Moreover, this section describes how to estimate the parameters of the state-space representation. Then Chapter 4.3 introduces the unscented transformation and derives the UKF, which allows to handle nonlinearities in the state-space representation. Finally, Chapter 4.4 develops a state-space representation for option pricing models and describes how to apply the UKF in order to jointly estimate the statistical and risk-neutral asset price dynamics of an option pricing model. Moreover, the concept vega-scaling is introduced, which facilitates a sparse parametrization of the measurement error.

¹The GPS is a global satellite navigation system for positioning and time measurement, where the state of the system refers to the position and the velocity of an object, which is estimated based on GPS satellite signals.

²For example, the volatility of a stock is not directly observable but related to corresponding option prices, which in turn can be used to estimate the volatility.

4.2. The classical Kalman filter

4.2.1. State-space representation

In order to apply the Kalman filter, we first need to express the underlying model in a particular form called the *state-space representation*. The state-space representation consists of the *state equation* and the *measurement equation*. The state equation describes the state of a system, which is only indirectly observed through noisy measurements.

Following Hamilton [90, p. 372 ff.] and Harvey [92, p. 100 ff.], we denote by x_i the state of a system at time i , which is a $m \times 1$ vector and assumed to be generated by a first-order Markov process

$$x_i = Ax_{i-1} + \eta_i \quad \text{for } i = 1, \dots, N ,$$

where A is a $m \times m$ parameter matrix and η_i is a $m \times 1$ vector of serially uncorrelated disturbance terms with

$$\mathbb{E}[\eta_i] = 0 \quad \text{and} \quad \text{Var}[\eta_i] = P \quad \text{for } i = 1, \dots, N . \quad (4.1)$$

The state x_i is linked to a $n \times 1$ vector of observations y_i via the measurement equation

$$y_i = Bx_i + \epsilon_i \quad \text{for } i = 1, \dots, N ,$$

where B is an $n \times m$ parameter matrix and ϵ_i is a $n \times 1$ vector of serially uncorrelated disturbance terms with

$$\mathbb{E}[\epsilon_i] = 0 \quad \text{and} \quad \text{Var}[\epsilon_i] = Q \quad \text{for } i = 1, \dots, N .$$

Moreover, we assume that the mean and variance of the initial state x_0 are given by

$$\mathbb{E}[x_0] = m_0 \quad \text{and} \quad \text{Var}[x_0] = R_0 .$$

Finally, the disturbance terms η_i and ϵ_i are assumed to be uncorrelated at all times as well as uncorrelated with the initial state, which is

$$\mathbb{E}[\eta_i \epsilon_j^\top] = 0 \quad \text{for } i, j = 1, \dots, N , \quad (4.2)$$

as well as

$$\mathbb{E}[\eta_i x_0^\top] = 0 \quad \text{and} \quad \mathbb{E}[\epsilon_i x_0^\top] = 0 \quad \text{for } i = 1, \dots, N . \quad (4.3)$$

This is the basic state-space representation, which can be further extended in two directions. On the one hand, we can incorporate some exogenous input, both in the state equation and in the measurement equation. On the other hand, we can allow the parameter matrices A , B , P and Q to be functions of time.³

In the next section, we will describe how to estimate the state x_i via the Kalman filter.

³See Harvey [92, p. 100 ff.] for more details.

Subsequently, Chapter 4.2.3 explains how to estimate the parameter matrices A , B , P and Q , which are usually unknown.

4.2.2. Derivation of the Kalman filter

The Kalman filter, named after its inventor Rudolf E. Kálmán [112, 113], is an algorithm in order to recursively calculate the conditional mean and variance of the state x_i based on current and past observations $\mathcal{I}_i = \{y_1, \dots, y_i\}$. The conditional mean of the state is an optimal estimator for x_i in the sense that it minimizes the corresponding *mean squared error* [90, p. 73]. The classical way to derive the Kalman filter is via the *least square approach*, where the conditional mean and variance of the state are calculated by sequentially updating the linear projection of the system in order to account for new measurements.⁴ Alternatively, the Kalman filter can also be derived via the *Bayesian approach*, which further requires to specify the probability distribution of the state-space representation. In the following, we will show how to derive the Kalman filter via the *Bayesian approach* and thereby mainly follow Särkkä [140, p. 31 ff.].

First recall the state-space representation from the previous section, which is given by

$$\begin{aligned} x_i &= Ax_{i-1} + \eta_i, \\ y_i &= Bx_i + \epsilon_i, \end{aligned} \tag{4.4}$$

where $i = 1, \dots, N$. So far, we only specified the first two moments of the disturbance terms η_i and ϵ_i as well as of the initial state x_0 . Moreover, we assumed that η_i , ϵ_i and x_0 are uncorrelated to each other at all times. Now, we make the additional assumption that the disturbance terms η_i and ϵ_i as well as the initial state x_0 are normally distributed, so that

$$\eta_i \sim \mathcal{N}(0, P) \quad \text{for } i = 1, \dots, N, \tag{4.5}$$

$$\epsilon_i \sim \mathcal{N}(0, Q) \quad \text{for } i = 1, \dots, N, \tag{4.6}$$

$$x_0 \sim \mathcal{N}(m_0, R_0). \tag{4.7}$$

Furthermore, we assume that the parameter matrices A , B , P and Q as well as the parameterization of the initial state m_0 and R_0 are known. Under these assumptions, we can formulate the Kalman filter algorithm in accordance with Särkkä [140, p. 36 f.] as follows:

Algorithm 4.1 (Kalman filter). *Let (4.4) be the underlying linear state-space representation and assume that the disturbance terms η_i and ϵ_i as well as the initial state x_0 are normally distributed as specified in (4.5)–(4.7). Moreover, η_i , ϵ_i and x_0 are assumed to be uncorrelated to each other at all times as specified in (4.2)–(4.3). Then the state x_i given all observations up to time i is normally distributed with*

$$x_i | \mathcal{I}_i \sim \mathcal{N}(m_i, R_i).$$

⁴See Hamilton [90, p. 377 ff.] for a derivation of the Kalman filter based on the least square approach.

4. Estimation of option pricing models

Then the conditional mean m_i and the conditional variance R_i can be calculated recursively as follows:

1. Prediction step:

$$\begin{aligned} m_i^- &= Am_{i-1} , \\ R_i^- &= AR_{i-1}A^\top + Q . \end{aligned}$$

2. Update step:

$$\begin{aligned} m_i &= m_i^- + C_i^\top S_i^{-1}(y_i - n_i^-) , \\ R_i &= R_i^- - C_i^\top S_i^{-1}C_i , \end{aligned}$$

where

$$\begin{aligned} n_i^- &= Bm_i^- , \\ C_i &= BR_i^- , \\ S_i &= BR_i^- B^\top + Q . \end{aligned}$$

The recursion is started with the unconditional mean m_0 and the unconditional variance R_0 of the initial state x_0 .

Proof. The following proof is based on the Kalman filter derivation of Särkkä [140, p. 37]. Starting the recursion at time $i = 1$ with the unconditional distribution (4.7), the distribution of the state at time $i - 1$ conditional on the information $\mathcal{I}_{i-1}$ is given by

$$x_{i-1} | \mathcal{I}_{i-1} \sim \mathcal{N}(m_{i-1}, R_{i-1}) ,$$

where

$$\begin{aligned} m_{i-1} &= \mathbb{E}[x_{i-1} | \mathcal{I}_{i-1}] , \\ R_{i-1} &= \text{Var}[x_{i-1} | \mathcal{I}_{i-1}] . \end{aligned}$$

Due to the linearity of the state equation and the independence of the random variables $x_{i-1} | \mathcal{I}_{i-1}$ and η_i , the state x_i conditional on the information $\mathcal{I}_{i-1}$ is normally distributed with

$$x_i | \mathcal{I}_{i-1} \sim \mathcal{N}(m_i^-, R_i^-) ,$$

where

$$\begin{aligned} m_i^- &= \mathbb{E}[x_i | \mathcal{I}_{i-1}] = Am_{i-1} , \\ R_i^- &= \text{Var}[x_i | \mathcal{I}_{i-1}] = AR_{i-1}A^\top + P . \end{aligned}$$

Again, due to the linearity of the measurement equation and the independence of the random variables $x_i | \mathcal{I}_{i-1}$ and ϵ_i , we can apply Lemma A.1 to compute the joint distribution

of x_i and y_i conditional on the information set $\mathcal{I}_{i-1}$, which is

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} \Big| \mathcal{I}_{i-1} \sim \mathcal{N} \left(\begin{bmatrix} m_i^- \\ n_i^- \end{bmatrix}, \begin{bmatrix} R_i^- & C_i^\top \\ C_i & S_i \end{bmatrix} \right), \quad (4.8)$$

where

$$\begin{aligned} n_i^- &= \mathbb{E}[y_i | \mathcal{I}_{i-1}] = Bm_i^-, \\ C_i &= \text{Cov}[x_i, y_i | \mathcal{I}_{i-1}] = BR_i^-, \\ S_i &= \text{Var}[y_i | \mathcal{I}_{i-1}] = BR_i^- B^\top + Q. \end{aligned}$$

By Lemma A.2, the conditional distribution of x_i given the information $\mathcal{I}_i$ is given by

$$x_i | \mathcal{I}_i \sim \mathcal{N}(m_i, R_i),$$

where

$$\begin{aligned} m_i &= \mathbb{E}[x_i | \mathcal{I}_i] = m_i^- + C_i^\top S_i^{-1} (y_i - n_i^-), \\ R_i &= \text{Var}[x_i | \mathcal{I}_i] = R_i^- - C_i^\top S_i^{-1} C_i. \end{aligned}$$

□

The matrix product $C_i^\top S_i^{-1}$ is sometimes referred to as *Kalman filter gain* and weights the difference between the predicted and actual observations in order to correct the predicted state m_i^- based on the newly available information.

4.2.3. Maximum likelihood estimation

In this section, we show how to estimate the parameter matrices A , B , P and Q of the state-space representation (4.4). Furthermore, we discuss some approaches in order to determine the mean and variance of the initial state x_0 .

At first, we turn towards the issue, that the mean and variance of x_0 are usually not known. To overcome this problem, we first need to distinguish if the state has a stationary or non-stationary distribution. If the state is stationary, the Kalman filter can always be initialized with the unconditional mean and variance of x_i [92, p. 127]. In contrast, if the state is non-stationary, the unconditional distribution of x_i is not defined. In such a case, the distribution of the initial state can be specified via a *diffuse prior*, where the variance of the initial state is set to a large but finite number.⁵ Alternatively, the initial state can be treated as a fixed parameter, so that x_0 has a degenerated distribution with zero variance. Hence, x_0 is an additional parameter, which needs to be estimated.⁶

Next, we discuss how the unknown parameters of the state-space representation can be

⁵For more details see Harvey [92, p. 133 ff.].

⁶For more details see Harvey [92, p. 137 ff.].

4. Estimation of option pricing models

estimated by virtue of the *maximum likelihood* (ML) method.⁷ First recall from equation (4.8) that the joint distribution of x_i and y_i conditional on the information set $\mathcal{I}_{i-1}$ is given by

$$\begin{pmatrix} x_i \\ y_i \end{pmatrix} \Big| \mathcal{I}_{i-1} \sim \mathcal{N} \left(\begin{bmatrix} m_i^- \\ n_i^- \end{bmatrix}, \begin{bmatrix} R_i^- & C_i^\top \\ C_i & S_i \end{bmatrix} \right), \quad (4.9)$$

where the mean and variance of (4.9) are calculated as a by-product of the Kalman filter. Hence, by Lemma A.2, the distribution of y_i conditional on the information set $\mathcal{I}_{i-1}$ is given by

$$y_i \mid \mathcal{I}_{i-1} \sim \mathcal{N}(n_i^-, S_i). \quad (4.10)$$

In particular, the mean and variance of $y_i \mid \mathcal{I}_{i-1}$ depend on the parameter matrices A , B , P and Q and, if the initial state is treated as a fixed parameter, also on x_0 .⁸ To highlight this dependence, we denote the mean and variance of $y_i \mid \mathcal{I}_{i-1}$ by $n_i^-(\theta)$ and $S_i(\theta)$, where the vector θ summarizes the parameters of the state-space representation. Moreover, we will denote by θ_0 the corresponding true parameter vector. Now, we can write (4.10) also as

$$f_{y_i \mid \mathcal{I}_{i-1}}(\bar{y}_i \mid \theta) = \frac{1}{(2\pi)^{n/2} |S_i(\theta)|^{1/2}} \exp \left\{ -\frac{1}{2} [\bar{y}_i - n_i^-(\theta)]^\top S_i(\theta)^{-1} [\bar{y}_i - n_i^-(\theta)] \right\}, \quad (4.11)$$

where $\bar{y}_i$ denotes a possible realization of y_i . The joint distribution of all observations $y_1, \dots, y_N$ is given by

$$f_{y_1, \dots, y_N}(\bar{y}_1, \dots, \bar{y}_N \mid \theta) = \prod_{i=1}^N f_{y_i \mid \mathcal{I}_{i-1}}(\bar{y}_i \mid \theta), \quad (4.12)$$

where $\mathcal{I}_0 = \emptyset$ denotes the empty set.

We are now able to apply the ML method in order to estimate the unknown parameter vector θ_0 . The basic idea of the ML method is to find an estimator for θ_0 , which maximizes the joint distribution (4.12) for the given observations $\bar{y}_1, \dots, \bar{y}_N$. However, it is often easier to maximize the logarithm of the likelihood function (4.12), which is a monotone transformation and thus has the same extrema. In our setting, the log-likelihood function is given by

$$\mathcal{L}(\theta) = -\frac{nN}{2} \ln(2\pi) - \frac{1}{2} \sum_{i=1}^N \ln(|S_i(\theta)|) - \frac{1}{2} \sum_{i=1}^N \xi_i^\top S_i(\theta)^{-1} \xi_i, \quad (4.13)$$

where

$$\xi_i = \bar{y}_i - n_i^-(\theta),$$

is called the *prediction error*. This name is due to the fact that the conditional mean $n_i^- = \mathbb{E}[y_i \mid \mathcal{I}_{i-1}]$ is the optimal forecast for y_i in a mean squared error sense. Hence, equation (4.13) is sometimes also called the *prediction error decomposition* form of the

⁷For a detailed treatment of the ML method see e.g. Davidson and MacKinnon [65].

⁸Of course, also the remaining moments of (4.9) depend on the parameterization of the state-space representation.

log-likelihood function [92, p. 126]. The ML estimator of the parameter vector θ_0 is given by

$$\hat{\theta}_{ML} = \arg \max_{\theta \in \Theta} \mathcal{L}(\theta) , \quad (4.14)$$

where Θ denotes the admissible parameter space. The optimization in (4.14) is usually performed by a numerical optimization procedure.⁹

Under certain regularity conditions, the ML estimator $\hat{\theta}_{ML}$ has the following desirable properties [88, p. 473 f.]:

1. *Consistency*: $\text{plim}_{N \rightarrow \infty} \hat{\theta}_{ML} = \theta_0$.¹⁰
2. *Asymptotic normality*: $\sqrt{N} (\hat{\theta}_{ML} - \theta_0) \xrightarrow{d} \mathcal{N}(0, \mathfrak{I}_0^{-1})$, where $\mathfrak{I}_0$ denotes the information matrix of the true parameter vector.¹¹
3. *Asymptotic efficiency*: $\hat{\theta}_{ML}$ has minimal variance for $N \rightarrow \infty$.¹²

One important requirement for these properties to hold is that the assumed parametric distribution of the observations, which in our context is specified by the state-space representation, includes the true distribution of the observations. Otherwise, the estimator is not efficient anymore, as will be discussed in Chapter 4.3.4.

4.2.4. Statistical inference

In the previous section, we showed how to find the ML estimator for the parameters of the state-space representation. In this section, we will discuss how to perform statistical inference for the ML estimator.

First recall from the previous section that under certain regularity conditions, the ML estimator $\hat{\theta}_{ML}$ is asymptotically normally distributed, which is

$$\sqrt{N} (\hat{\theta}_{ML} - \theta_0) \xrightarrow{d} \mathcal{N}(0, \mathfrak{I}_0^{-1}) ,$$

⁹For numerical optimization method in this context see e.g. Hamilton [90, p. 133 ff.].

¹⁰The term “plim” refers to *convergence in probability*. A sequence of random variables $\{X_n\}$ is said to converge in probability to X if and only if for every $\epsilon > 0$ we have $\lim_{n \rightarrow \infty} \mathbb{P}[|X_n - X| > \epsilon] = 0$ [58, p. 70].

¹¹See Chapter 4.2.4, for a definition of the information matrix.

¹²Note that the variance of a consistent estimator is bounded from below by the *Cramér-Rao bound*, which is given by the inverse of the information matrix [93, p. 50]. Hence, the ML estimator achieves the Cramér-Rao bound.

4. Estimation of option pricing models

where the *information matrix* $\mathfrak{I}_0$ is defined as

$$\mathfrak{I}_0 = -\mathbb{E} \left[\frac{\partial^2 \mathcal{L}(\theta)}{\partial \theta \partial \theta^\top} \Big|_{\theta=\theta_0} \right] .$$

Recall that $\mathcal{L}(\theta)$ denotes the log-likelihood function for the observations $y_1, \dots, y_N$, which is given by

$$\mathcal{L}(\theta) = \sum_{i=1}^N \ln (f_{y_i | \mathcal{I}_{i-1}}(\bar{y}_i | \theta)) .$$

If the sample size is sufficiently large, it is reasonable to approximate the distribution of $\hat{\theta}_{ML}$ by

$$\hat{\theta}_{ML} \approx \mathcal{N}(\theta_0, N^{-1} \mathfrak{I}_0^{-1}) .$$

Following Hamilton [90, p. 143], the information matrix can either be estimated via the *second-derivative estimate*

$$\hat{\mathfrak{I}}_{2D} = -\frac{1}{N} \frac{\partial^2 \mathcal{L}(\theta)}{\partial \theta \partial \theta^\top} \Big|_{\theta=\hat{\theta}} , \quad (4.15)$$

or by the *outer-product estimate*

$$\hat{\mathfrak{I}}_{OP} = \frac{1}{N} \sum_{i=1}^N \left(\frac{\partial \ln (f_{y_i | \mathcal{I}_{i-1}}(\bar{y}_i | \theta))}{\partial \theta} \right) \cdot \left(\frac{\partial \ln (f_{y_i | \mathcal{I}_{i-1}}(\bar{y}_i | \theta))}{\partial \theta} \right)^\top \Big|_{\theta=\hat{\theta}} . \quad (4.16)$$

Hence, the distribution of the ML estimator can be approximated by

$$\hat{\theta}_{ML} \approx \mathcal{N}(\theta_0, N^{-1} \hat{\mathfrak{I}}_{2D/OP}^{-1}) ,$$

which can further be used to perform hypothesis tests or to build confidence intervals for $\hat{\theta}_{ML}$.¹³ The derivatives involved in the information matrix estimates usually cannot be calculated in closed-form, so that numerical methods need to be applied. Since the outer-product estimate only involves first-order derivatives, it is often easier to calculate $\hat{\mathfrak{I}}_{OP}$ than $\hat{\mathfrak{I}}_{2D}$.

4.3. The unscented Kalman filter

4.3.1. Motivation

The classical Kalman filter can only be applied if the underlying model can be represented via the linear state-space representation (4.4). In practice, most issues are nonlinear, so that the classical Kalman filter cannot be applied. One possibility to overcome this problem is to linearize the underlying model via a first order *Taylor approximation*, so that the resulting filtering equations are linear and thus the Kalman filter can be applied.

¹³For more details on hypothesis testing under normality see e.g. Hayashi [93, p. 33 ff.].

This method is also known as *Extend Kalman filter* (EKF) and according to Julier and Uhlmann [106] the most common state estimation algorithm for nonlinear systems.¹⁴ However, the use of the EKF has two well-known drawbacks [106]:

1. The linearization of the state-space representation via the Taylor approximation can produce highly unstable filters if the assumption of local linearity is violated.
2. The derivation of the Jacobian matrices, which are necessary to perform the Taylor approximation, are nontrivial in most applications and often lead to significant implementation difficulties.

The first issue is due to the fact that the prediction and updating equations for a nonlinear state-space representation involve the computation of the mean and variance for the transformed random variables. But if this transformation is strongly nonlinear, the true mean and variance differ substantially from their corresponding first order Taylor approximation, which in turn causes the EKF to be unstable. This problem is illustrated in the following example.

Example 4.1. Consider the random variable $X \sim \mathcal{N}(0, 1)$, which is transformed by the nonlinear function $f(x) = x^2$, so that

$$Y = f(X) .$$

The mean and variance of Y are given by

$$\mathbb{E}[Y] = 1 \quad \text{and} \quad \text{Var}[Y] = 2 ,$$

respectively.¹⁵ In contrast, the first order Taylor approximation of f expanded around $x_0 = 0$ is given by $f_{TA}(x) = 0$, so that the mean and variance of Y can be approximated by

$$\mathbb{E}[f_{TA}(X)] = 0 \quad \text{and} \quad \text{Var}[f_{TA}(X)] = 0 ,$$

respectively.

Instead of approximating the transformation function, we can also approximate the untransformed random variable via some sample points from the underlying distribution. In particular, if we propagate these sample points through the transformation, we explicitly account for any nonlinearity. The mean and variance of the transformed random variable can be approximated via the sample mean and variance of the transformed sample points. This approach is called *Monte-Carlo simulation* and illustrated in Figure 4.1 based on the setting of the previous Example 4.1. In particular, we can see that the approximation of the mean and variance for the transformed random variable via Monte-Carlo simulation is much better than the corresponding first order Taylor approximation.

¹⁴For more details on the EKF see e.g. SIMON [145].

¹⁵The mean of the transformed random variable Y is simply the variance of a standard normal random variable, which is equal to 1. The variance of the transformed random variable can be computed as follows: $\text{Var}[Y] = \mathbb{E}[(X^2 - 1)^2] = \mathbb{E}[X^4] - 2\mathbb{E}[X^2] + 1$. Since the fourth raw moment of a standard normal random variable is equal to $\mathbb{E}[X^4] = 3$, it follows that the variance of Y is equal to 2.

4. Estimation of option pricing models

Figure 4.1.: Mean and variance of a transformed random variable. This figure compares the sample mean (a) and the sample variance (b), both calculated via Monte-Carlo simulation, of the transformed random variable $Y = X^2$, where $X \sim \mathcal{N}(0, 1)$, with the corresponding true moments.

The Kalman filter based on the Monte-Carlo simulation is called *particle filter*, where the conditional moments involved in the computation of the prediction and updating equations are approximated via Monte-Carlo simulation.¹⁶ Unfortunately, the particle filter is computationally very demanding, especially if the desired accuracy is high and the underlying data set is large. This can also be seen in Figure 4.1, where at least 20,000 sample points are needed in order to accurately approximate the mean and variance of the transformed random variable. But instead to arbitrarily draw random numbers from the underlying distribution, we could also carefully choose some representative sample points, such that the mean and variance of the transformed random variable can be computed with high accuracy but low computational complexity. This is exactly the aim of the unscented transformation, which will be presented in the next section.

4.3.2. Unscented transformation

The *unscented transformation* is an approximation method in order to calculate the mean and variance of a transformed random variable. The unscented transformation was first developed by Julier et al. [105, 106, 107] and further extended by Wan and van der Merwe [150, 151]. The following explanation of the unscented transformation is based on Wan and van der Merwe [151].

Let X be a random variable of dimension $m \times 1$ with mean μ_X and variance Σ_X , which

¹⁶For more details on the particle filter see e.g. Simon [145].

is propagated through a nonlinear function $f : \mathbb{R}^m \rightarrow \mathbb{R}^n$, so that

$$Y = f(X) .$$

The mean and variance of Y can be approximated via the unscented transformation. Therefore, we first build a matrix $\mathcal{X}$ of $2m + 1$ *sigma vectors* $\mathbf{x}_i$

$$\mathcal{X} = [\mathbf{x}_0 \ \mathbf{x}_1 \ \dots \ \mathbf{x}_{2m}] . \quad (4.17)$$

The sigma vectors in (4.17) are defined as follows:

$$\begin{aligned} \mathbf{x}_0 &= \mu_X , \\ \mathbf{x}_i &= \mu_X + \sqrt{m + \lambda} \sigma_i \quad \text{for } i = 1, \dots, m , \\ \mathbf{x}_i &= \mu_X - \sqrt{m + \lambda} \sigma_{i-m} \quad \text{for } i = m + 1, \dots, 2m , \end{aligned} \quad (4.18)$$

where σ_i is the i^{th} row of the matrix square root $\sqrt{\Sigma_X}$.¹⁷ The scaling parameter

$$\lambda = \alpha^2 (m + \kappa) - m ,$$

where $0 < \alpha \leq 1$, determines the spread of the sigma points around the mean of X . The parameter α is usually set to a small positive value, e.g. 10^{-3} , and the secondary scaling parameter κ is usually set to zero [150, p. 155].

Next, the weights for the sigma vectors $\mathbf{x}_i$, which are used to approximate the mean and variance of the transformed random variable Y , are defined as follows:

$$\mathbf{w}_0^\mu = \lambda / (m + \lambda) , \quad (4.19)$$

$$\mathbf{w}_0^\Sigma = \lambda / (m + \lambda) + (1 - \alpha^2 + \beta) , \quad (4.20)$$

$$\mathbf{w}_i^\mu = \mathbf{w}_i^\Sigma = 1 / (2m + 2\lambda) \quad \text{for } i = 1, \dots, 2m , \quad (4.21)$$

where $\mathbf{w}_i^\mu$ and $\mathbf{w}_i^\Sigma$ denote the weights for the mean and variance approximation, respectively. The parameter β allows to incorporate prior knowledge about the distribution of X . For example, if X is normally distributed, $\beta = 2$ would be the optimal choice [105].¹⁸

The sigma vectors and corresponding weights are chosen such that the mean and variance approximation of X via the unscented transformation exactly match the true mean and variance of X , which is

$$\begin{aligned} \mu_X &= \sum_{i=0}^{2m} \mathbf{w}_i^\mu \mathbf{x}_i , \\ \Sigma_X &= \sum_{i=0}^{2m} \mathbf{w}_i^\Sigma (\mathbf{x}_i - \mu_X)(\mathbf{x}_i - \mu_X)^\top . \end{aligned}$$

In particular, the mean and variance of the random variable Y can now be approximated.

¹⁷The matrix square root can be obtained e.g. via Cholesky decomposition.

¹⁸See Julie [105] for more details about the parameterization of the unscented transformation.

4. Estimation of option pricing models

Therefore we first propagate the sigma vectors x_i through the nonlinear function f , which is

$$y_i = f(x_i) \quad \text{for } i = 0, \dots, 2m ,$$

and then calculate the sample mean and variance of the transformed sigma vectors y_i , which is

$$\hat{\mu}_Y = \sum_{i=0}^{2m} w_i^\mu y_i , \quad (4.22)$$

$$\hat{\Sigma}_Y = \sum_{i=0}^{2m} w_i^\Sigma (y_i - \hat{\mu}_Y)(y_i - \hat{\mu}_Y)^\top . \quad (4.23)$$

The accuracy of the mean and variance approximation via the unscented transformation will be summarized in the following proposition.

Proposition 4.1. *Let X be a m -dimensional random variable with mean μ_X and variance Σ_X and define*

$$Y = f(X) ,$$

where $f : \mathbb{R}^m \rightarrow \mathbb{R}^n$ is some nonlinear transformation. Then the mean and variance approximation of Y via the unscented transformation, which is specified by (4.22) and (4.23), is accurate up to the second order Taylor approximation. Moreover, if X is normally distributed, then the mean and variance approximation of Y via the unscented transformation is accurate up to the third order Taylor approximation.

Proof. The following proof is based on Simon [145, p. 441 ff.]. Consider the Taylor series expansion of Y around the mean of X , which is

$$Y = f(X) = f(\mu_X) + D_\epsilon^1 f + \frac{1}{2!} D_\epsilon^2 f + \frac{1}{3!} D_\epsilon^3 f + \dots , \quad (4.24)$$

where

$$D_\epsilon^k f = \left(\sum_{i=1}^m \epsilon_i \frac{\partial}{\partial x_i} \right)^k f(x) \Big|_{x=\mu_X} ,$$

and

$$\epsilon = X - \mu_X ,$$

with $\epsilon = [\epsilon_1 \dots \epsilon_m]^\top$.¹⁹ Hence, the mean of Y is equal to

$$\mathbb{E}[Y] = f(\mu_X) + \mathbb{E}[D_\epsilon f] + \frac{1}{2!} \mathbb{E}[D_\epsilon^2 f] + \frac{1}{3!} \mathbb{E}[D_\epsilon^3 f] + \dots .$$

The mean of the first order term is zero, since

$$\mathbb{E}[D_\epsilon f] = \mathbb{E} \left[\left(\sum_{i=1}^m \epsilon_i \frac{\partial}{\partial x_i} \right) f(x) \Big|_{x=\mu_X} \right]$$

¹⁹The operator D^k is explained in Appendix A.2 in more detail.

$$\begin{aligned}
&= \left(\sum_{i=1}^m \underbrace{\mathbb{E}[\epsilon_i]}_{=0} \frac{\partial}{\partial x_i} \right) f(x) \Big|_{x=\mu_X} \\
&= 0 .
\end{aligned}$$

Furthermore, if we assume that X is multivariate normally distributed with mean μ_X , then ϵ is multivariate normally distributed with zero mean, so that all odd-order moments of ϵ are zero, i.e. $\mathbb{E}[\epsilon^k] = 0$ for k odd.²⁰ Moreover, since $\mathbb{E}[D_\epsilon^{2k+1}f]$ with $k \geq 1$ consists only of odd-order moments of ϵ , the mean of all odd-order terms in (4.24) is zero.

The mean of the second order term equals

$$\begin{aligned}
\mathbb{E}[D_\epsilon^2 f] &= \mathbb{E} \left[\left(\sum_{i=1}^m \epsilon_i \frac{\partial}{\partial x_i} \right)^2 f(x) \Big|_{x=\mu_X} \right] \\
&= \mathbb{E} \left[\left(\sum_{i,j=1}^m \epsilon_i \epsilon_j \frac{\partial^2}{\partial x_i \partial x_j} \right) f(x) \Big|_{x=\mu_X} \right] \\
&= \left(\sum_{i,j=1}^m \mathbb{E}[\epsilon_i \epsilon_j] \frac{\partial^2}{\partial x_i \partial x_j} \right) f(x) \Big|_{x=\mu_X} \\
&= \left(\sum_{i,j=1}^m \Sigma_X(i,j) \frac{\partial^2}{\partial x_i \partial x_j} \right) f(x) \Big|_{x=\mu_X} ,
\end{aligned}$$

where $\Sigma_X(i,j)$ denotes the element in the i^{th} row and j^{th} column of Σ .

Next, we look at the Taylor series expansion of the transformed sigma vectors y_j around μ_X , which is given by

$$y_j = f(x_j) = f(\mu_X) + D_{\tilde{\epsilon}_j} f + \frac{1}{2!} D_{\tilde{\epsilon}_j}^2 f + \frac{1}{3!} D_{\tilde{\epsilon}_j}^3 f + \dots \quad \text{for } j = 0, \dots, 2m , \quad (4.25)$$

where

$$\tilde{\epsilon}_j = x_j - \mu_X , \quad (4.26)$$

with $\tilde{\epsilon}_j = [\tilde{\epsilon}_{1,j} \dots \tilde{\epsilon}_{m,j}]^\top$. The mean approximation of Y via the unscented transformation can be computed by weighting the transformed sigma vectors with their corresponding weights, which is

$$\hat{\mu}_Y = \sum_{j=0}^{2m} w_j^\mu y_j$$

²⁰To get an intuition why this holds true, consider the k^{th} moment $\mathbb{E}[X^k]$ of a univariate normally distributed random variable X with zero mean. Due to the symmetry of X , $\mathbb{E}[X^k] = \int_{-\infty}^{\infty} x^k f_X(x) dx = \int_{-\infty}^0 x^k f_X(x) dx + \int_0^{\infty} x^k f_X(x) dx = -\int_0^{\infty} x^k f_X(x) dx + \int_0^{\infty} x^k f_X(x) dx = \int_0^{\infty} (-x)^k f_X(x) dx + \int_0^{\infty} x^k f_X(x) dx$, which is zero if k is odd. For a proof of the multivariate case see e.g. Holmquist [96].

4. Estimation of option pricing models

$$\begin{aligned}
&= \frac{\lambda}{m + \lambda} f(\mu_X) + \frac{1}{2(m + \lambda)} \sum_{j=1}^{2m} \left(f(\mu_X) + D_{\tilde{\epsilon}_j}^1 f + \frac{1}{2!} D_{\tilde{\epsilon}_j}^2 f + \frac{1}{3!} D_{\tilde{\epsilon}_j}^3 f + \dots \right) \\
&= f(\mu_X) + \frac{1}{2(m + \lambda)} \sum_{j=1}^{2m} \left(D_{\tilde{\epsilon}_j}^1 f + \frac{1}{2!} D_{\tilde{\epsilon}_j}^2 f + \frac{1}{3!} D_{\tilde{\epsilon}_j}^3 f + \dots \right) . \tag{4.27}
\end{aligned}$$

Again, all odd order terms in (4.27) are zero since for any integer $k \geq 1$

$$\begin{aligned}
\sum_{j=1}^{2m} D_{\tilde{\epsilon}_j}^{2k+1} f &= \sum_{j=1}^{2m} \left[\left(\sum_{i=1}^m \tilde{\epsilon}_{i,j} \frac{\partial}{\partial x_i} \right)^{2k+1} f(x) \right] \Big|_{x=\mu_X} \\
&= \sum_{j=1}^{2m} \left[\sum_{i=1}^m \tilde{\epsilon}_{i,j}^{2k+1} \frac{\partial^{2k+1}}{\partial x_i^{2k+1}} f(x) \right] \Big|_{x=\mu_X} \\
&= \sum_{i=1}^m \left[\sum_{j=1}^{2m} \tilde{\epsilon}_{i,j}^{2k+1} \frac{\partial^{2k+1}}{\partial x_i^{2k+1}} f(x) \right] \Big|_{x=\mu_X} \\
&= 0 .
\end{aligned}$$

The last equality holds since $\tilde{\epsilon}_{i,j} = -\tilde{\epsilon}_{i,m+j}$ for $j = 1, \dots, m$ due to equation (4.18). Next, the mean of the second order term is given by

$$\begin{aligned}
\frac{1}{2(m + \lambda)} \sum_{k=1}^{2m} D_{\tilde{\epsilon}_k}^2 f &= \frac{1}{2(m + \lambda)} \sum_{k=1}^{2m} \left(\sum_{i=1}^m \tilde{\epsilon}_{i,k} \frac{\partial}{\partial x_i} \right)^2 f(x) \Big|_{x=\mu_X} \\
&= \frac{1}{2(m + \lambda)} \left(\sum_{k=1}^{2m} \sum_{i,j=1}^m \tilde{\epsilon}_{i,k} \tilde{\epsilon}_{j,k} \frac{\partial^2}{\partial x_i \partial x_j} \right) f(x) \Big|_{x=\mu_X} \\
&= \frac{1}{2(m + \lambda)} \left(\sum_{k=1}^{2m} \sum_{i,j=1}^m (m + \lambda) \sigma_{i,k} \sigma_{j,k} \frac{\partial^2}{\partial x_i \partial x_j} \right) f(x) \Big|_{x=\mu_X} \\
&= \left(\sum_{i,j=1}^m \sum_{k=1}^m \sigma_{i,k} \sigma_{j,k} \frac{\partial^2}{\partial x_i \partial x_j} \right) f(x) \Big|_{x=\mu_X} \\
&= \left(\sum_{i,j=1}^m \Sigma_X(i, j) \frac{\partial^2}{\partial x_i \partial x_j} \right) f(x) \Big|_{x=\mu_X} ,
\end{aligned}$$

where $\sigma_{i,k}$ denotes the i^{th} element of the sigma vector σ_k .

Hence, the mean of the first and second order terms in the Taylor series expansion of Y and y_i , respectively, are equal, so that the mean approximation of Y via the unscented transformation is accurate up to the second order Taylor approximation. Moreover, if X is multivariate normally distributed, also the mean of the third order term is zero, so that the mean approximation of Y via the unscented transformation is accurate up to the third order Taylor approximation. The proof for the accuracy of the variance approximation of

Y via the unscented transformation is very similar to the proof for the mean, wherefore we refer to Simon [145, p. 444 f.] or Wan and van der Merwe [151, p. 271 ff.] for more details. $\square$

The following example illustrates how the unscented transformation works.

Example 4.2. Consider the transformed random variable $Y = f(X)$ from Example 4.1, where $f(x) = x^2$ and $X \sim \mathcal{N}(0, 1)$, so that $\mu_X = 0$, $\Sigma_X = 1$ and $m = 1$. If we set $\alpha = 1/2$ and $\kappa = 0$, we can compute the (transformed) sigma points as well as the corresponding weights, which are summarized in the following table:

Table 4.1.: Sigma points and weights.

i	w_i^μ	w_i^Σ	x_i	y_i
0	-3	$\beta - 2.25$	0	0
1	2	2	$1/2$	$1/4$
2	2	2	$-1/2$	$1/4$

Thus, the sample mean and variance of the transformed sigma points are given by

$$\hat{\mu}_Y = \sum_{i=0}^2 w_i^\mu y_i = 1 ,$$

$$\hat{\Sigma}_Y = \sum_{i=0}^2 w_i^\Sigma (y_i - \hat{\mu}_Y)^2 = \beta .$$

Since X is normally distributed, we set $\beta = 2$, so that $\hat{\mu}_Y$ and $\hat{\Sigma}_Y$ exactly match the true mean and variance of Y .²¹ In particular, the mean and variance approximation via the unscented transformation is better than the corresponding Taylor approximation (see Example 4.1) and Monte-Carlo simulation (see Figure 4.1).

4.3.3. Derivation of the unscented Kalman filter

In the following, we derive the *unscented Kalman filter* (UKF) in accordance to Särkkä [140, p. 55 ff.]. The UKF is a Gaussian approximation to the classical Kalman filter based on the unscented transformation. In contrast to the classical Kalman filter, the state-space representation of the UKF no longer needs to be linear and can more generally be written as

$$x_i = f(x_{i-1}, \eta_i | \theta_1) , \quad (4.28)$$

$$y_i = g(x_i, \epsilon_i | \theta_2) , \quad (4.29)$$

²¹Recall that $\beta = 2$ is the optimal choice if X is normally distributed.

4. Estimation of option pricing models

where $f : \mathbb{R}^m \rightarrow \mathbb{R}^m$ and $g : \mathbb{R}^m \rightarrow \mathbb{R}^n$ are possibly nonlinear functions. Moreover, θ_1 and θ_2 denote parameter vectors, which summarize the parameters of the state- and measurement equation, respectively. Again, the state equation (4.28) describes the dynamics of the possibly unobservable state x_i , which is related to observable variables y_i via the measurement equation (4.29). Furthermore, the disturbance terms η_i and ϵ_i as well as the initial state x_0 are assumed to be normally distributed with

$$\eta_i \sim \mathcal{N}(0, P) \quad \text{for } i = 1, \dots, N, \quad (4.30)$$

$$\epsilon_i \sim \mathcal{N}(0, Q) \quad \text{for } i = 1, \dots, N, \quad (4.31)$$

$$x_0 \sim \mathcal{N}(\hat{m}_0, \hat{R}_0) . \quad (4.32)$$

Since f and g are possibly nonlinear, we face two difficulties if we would like to derive the optimal filtering equations in this setup:

1. The conditional mean and variance of the system can usually not be calculated in closed-form.
2. The system is in general not Gaussian.

To circumvent these problems, we first approximate the conditional mean and variance of the system by the unscented transformation, which will be denoted by

$$\mathbb{E} \begin{bmatrix} x_i \\ y_i \end{bmatrix} \Big| \mathcal{I}_{i-1} \Big] \overset{UT}{\approx} \begin{bmatrix} \hat{m}_i^- \\ \hat{n}_i^- \end{bmatrix} \quad \text{and} \quad \text{Var} \begin{bmatrix} x_i \\ y_i \end{bmatrix} \Big| \mathcal{I}_{i-1} \Big] \overset{UT}{\approx} \begin{bmatrix} \hat{R}_i^- & \hat{C}_i^\top \\ \hat{C}_i & \hat{S}_i \end{bmatrix} .$$

Then we use these moment approximations in order to form a Gaussian approximation to the conditional distribution of the system, which is

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} \Big| \mathcal{I}_{i-1} \approx \mathcal{N} \left(\begin{bmatrix} \hat{m}_i^- \\ \hat{n}_i^- \end{bmatrix}, \begin{bmatrix} \hat{R}_i^- & \hat{C}_i^\top \\ \hat{C}_i & \hat{S}_i \end{bmatrix} \right) . \quad (4.33)$$

In particular, the distributional approximation (4.33) together with Lemma A.2 also specifies the distributional approximation of $x_i | \mathcal{I}_i$, which is

$$x_i | \mathcal{I}_i \approx \mathcal{N}(\hat{m}_i, \hat{R}_i)$$

where

$$\begin{aligned} \hat{m}_i &= \hat{m}_i^- + \hat{C}_i^\top \hat{S}_i^{-1} (y_i - \hat{n}_i^-) , \\ \hat{R}_i &= \hat{R}_i^- - \hat{C}_i^\top \hat{S}_i^{-1} \hat{C}_i . \end{aligned}$$

Consequently, these approximations result in the prediction and updating equations of the classical Kalman filter, which will be summarized in the following algorithm in accordance to Särkkä [140, p. 55 ff.].

Algorithm 4.2 (Unscented Kalman filter). *Let (4.28) – (4.29) be the underlying, possibly nonlinear state-space representation and assume that the disturbance terms η_i and ϵ_i as*

well as the initial state x_0 are normally distributed as specified in (4.30) – (4.32). Then the state x_i given all observations up to time i is approximately normally distributed with

$$x_i | \mathcal{I}_i \approx \mathcal{N}(\hat{m}_i, \hat{R}_i) .$$

Furthermore, the conditional mean $\hat{m}_i$ and variance $\hat{R}_i$ can be calculated recursively as follows. First, define the augmented state vector $\tilde{x}_{i-1} = [x_{i-1}^\top \ \eta_i^\top \ \epsilon_i^\top]^\top$, which is of dimension $d = 2m + n$, and approximate the conditional mean $\mathbb{E}[\tilde{x}_{i-1} | \mathcal{I}_{i-1}]$ and variance $\text{Var}[\tilde{x}_{i-1} | \mathcal{I}_{i-1}]$ by

$$\tilde{m}_{i-1} = \begin{bmatrix} \hat{m}_{i-1} \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \tilde{R}_{i-1} = \begin{bmatrix} \hat{R}_{i-1} & 0 & 0 \\ 0 & P & 0 \\ 0 & 0 & Q \end{bmatrix} ,$$

respectively. Moreover, implicitly define the matrices A , B and C , such that

$$x_{i-1} = A\tilde{x}_{i-1} , \quad \eta_i = B\tilde{x}_{i-1} \quad \text{and} \quad \epsilon_i = C\tilde{x}_{i-1} .$$

Next, build the sigma vectors $\mathcal{X}_{i-1} = [\tilde{x}_{i-1,0} \dots \tilde{x}_{i-1,2d}]$ with mean $\tilde{m}_{i-1}$ and variance $\tilde{R}_{i-1}$ according to (4.18). Finally, define the transformed sigma vectors

$$\begin{aligned} x_{i,j}^- &= f(A\tilde{x}_{i-1,j}, B\tilde{x}_{i-1,j}) \\ y_{i,j}^- &= g(x_{i,j}^-, C\tilde{x}_{i-1,j}) \end{aligned}$$

and proceed with the following steps:

1. Prediction step:

$$\begin{aligned} \hat{m}_i^- &= \sum_{j=0}^{2d} w_j^\mu x_{i,j}^- , \\ \hat{R}_i^- &= \sum_{j=0}^{2d} w_j^\Sigma (x_{i,j}^- - \hat{m}_i^-)(x_{i,j}^- - \hat{m}_i^-)^\top . \end{aligned}$$

2. Update step:

$$\begin{aligned} \hat{m}_i &= \hat{m}_i^- + \hat{C}_i^\top \hat{S}_i^{-1} (y_i - \hat{n}_i^-) , \\ \hat{R}_i &= \hat{R}_i^- - \hat{C}_i^\top \hat{S}_i^{-1} \hat{C}_i , \end{aligned}$$

where

$$\begin{aligned} \hat{n}_i^- &= \sum_{j=0}^{2d} w_j^\mu y_{i,j}^- , \\ \hat{C}_i &= \sum_{j=0}^{2d} w_j^\Sigma (x_{i,j}^- - \hat{m}_i^-)(y_{i,j}^- - \hat{n}_i^-)^\top , \end{aligned}$$

4. Estimation of option pricing models

$$\hat{S}_i = \sum_{j=0}^{2d} w_j^\Sigma (y_{i,j}^- - \hat{n}_i^-)(y_{i,j}^- - \hat{n}_i^-)^\top .$$

The weights w_j^μ and w_j^Σ are specified in (4.19). The recursion is started with the unconditional mean $\hat{m}_0$ and the unconditional variance $\hat{R}_0$ of the initial state x_0 .

The computational complexity of the UKF can further be reduced if the disturbance terms η_i and ϵ_i are additive, so that the state-space representation is given by

$$\begin{aligned} x_i &= f(x_{i-1} | \theta_1) + \eta_i , \\ y_i &= g(x_i | \theta_1) + \epsilon_i . \end{aligned}$$

In this special case, we do not need to augment the state vector x_i with the disturbance terms η_i and ϵ_i in order to calculate the moments involved in the prediction and update step, so that the number of necessary sigma points can be reduced.²² Note that if f and g are linear, the UKF coincides with the classical Kalman filter.

4.3.4. Quasi-maximum likelihood estimation

Analog to Chapter 4.2.3, this section shows how to estimate the parameter vectors θ_1 and θ_2 of the general state-space representation

$$\begin{aligned} x_i &= f(x_{i-1}, \eta_i | \theta_1) , \\ y_i &= g(x_i, \epsilon_i | \theta_2) , \end{aligned} \tag{4.34}$$

where $i = 1, \dots, N$. The main difference to the linear state-space representation (4.4) is that the joint distribution of the observations $y_1, \dots, y_N$ specified via (4.34) is in general not Gaussian and in most cases even not known. Nevertheless, the UKF performs a Gaussian approximation to the joint distribution of x_i and y_i conditional on the information set $\mathcal{I}_{i-1}$ via

$$\begin{pmatrix} x_i \\ y_i \end{pmatrix} \Big| \mathcal{I}_{i-1} \approx \mathcal{N} \left(\begin{bmatrix} \hat{m}_i^- \\ \hat{n}_i^- \end{bmatrix} , \begin{bmatrix} \hat{R}_i^- & \hat{C}_i^\top \\ \hat{C}_i & \hat{S}_i \end{bmatrix} \right) , \tag{4.35}$$

where the mean and variance of (4.35) are calculated via the unscented transformation as described in Algorithm 4.2. Now, by Lemma A.2, the distribution of y_i conditional on the information set $\mathcal{I}_{i-1}$ can be approximated by

$$y_i | \mathcal{I}_{i-1} \approx \mathcal{N}(\hat{n}_i^-(\theta), \hat{S}_i(\theta)) ,$$

where $\theta = [\theta_1 \ \theta_2]$ indicates the dependence of $\hat{n}_i^-$ and $\hat{S}_i$ on the parameterization of (4.34). Hence, in accordance to Chapter 4.2.3, the log-likelihood function can be approximated

²²See Wan and van der Merwe [151, p. 233 ff.] for an treatment of the UKF with additive disturbance terms.

by

$$\widehat{\mathcal{L}}(\theta) = -\frac{nN}{2} \ln(2\pi) - \frac{1}{2} \sum_{i=1}^N \ln \left(|\widehat{S}_i(\theta)| \right) - \frac{1}{2} \sum_{i=1}^N \xi_i^\top \widehat{S}_i(\theta)^{-1} \xi_i, \quad (4.36)$$

where

$$\xi_i = \bar{y}_i - \widehat{n}_i^-(\theta).$$

In particular, the unknown parameter vector θ_0 can be estimated by maximizing (4.36) with respect to θ , which is

$$\widehat{\theta}_{QML} = \arg \max_{\theta \in \Theta} \widehat{\mathcal{L}}(\theta).$$

The ML method applied to a misspecified likelihood function is called *quasi-maximum likelihood* (QML) method. Under certain regularity conditions, Bollerslev and Wooldridge [38] show that the QML estimator is still weakly consistent and asymptotically normally distributed. In particular, this requires that the conditional mean and variance of $y_i | \mathcal{I}_{i-1}$ are correctly specified [38, p. 148].²³ The UKF approximates these moments up to second order accuracy, and in some special cases even up to third order accuracy, which should be sufficiently accurate in most cases in order to justify the usage of the QML method.

4.3.5. Statistical inference

Also the QML method allows to perform statistical inference. But due to the possible misspecification of the supposed likelihood function, the asymptotic distribution of the QML estimator is slightly different to the one of the ML estimator. Again, under certain regularity conditions, Bollerslev and Wooldridge [38] show that

$$\sqrt{N} \left(\widehat{\theta}_{QML} - \theta_0 \right) \xrightarrow{d} \mathcal{N}(0, A_0^{-1} B_0 A_0^{-1}),$$

where

$$A_0 = -\mathbb{E} \left[\frac{\partial^2 \mathcal{L}(\theta)}{\partial \theta \partial \theta^\top} \Big|_{\theta=\theta_0} \right],$$

$$B_0 = \mathbb{E} [s(\theta_0) s(\theta_0)^\top].$$

Here $s(\theta_0)$ denotes the *score vector*, which is defined as

$$s(\theta_0) = \frac{\partial \mathcal{L}(\theta)}{\partial \theta} \Big|_{\theta=\theta_0}.$$

Moreover, White [156] shows that $\widehat{\mathcal{J}}_{2D} \widehat{\mathcal{J}}_{OP}^{-1} \widehat{\mathcal{J}}_{2D}$ is a consistent estimator for $A_0^{-1} B_0 A_0^{-1}$, so that the distribution of the QML estimator can be approximated by

$$\widehat{\theta}_{QML} \approx \mathcal{N}(\theta_0, N^{-1} \widehat{\mathcal{J}}_{2D} \widehat{\mathcal{J}}_{OP}^{-1} \widehat{\mathcal{J}}_{2D}),$$

²³See Bollerslev and Wooldridge [38, p. 167] for the remaining regularity conditions.

which can be used to perform hypothesis test or to build confidence intervals.²⁴

4.4. Estimation of option pricing models via the unscented Kalman filter

4.4.1. Information content of option prices

Option prices comprise information about the underlying asset price dynamics, both under the statistical and risk-neutral measure. Whereas the risk-neutral information is contained in cross-sections of option prices, we need to consider time series of option prices to recover the statistical information. In the following, we will illustrate these two different kinds of information in a general setting and assume that the asset price process under the statistical measure $\mathbb{P}$ is given by a d -dimensional affine diffusion process

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t , \quad (4.37)$$

where W_t is a d -dimensional standard Brownian motion. Moreover, the affine drift and diffusion term $\mu(x)$ and $\sigma(x)$ are specified by (3.1).²⁵ For ease of notation, we summarize the parameters, which specify $\mu(x)$ and $\sigma(x)$, and thus also the statistical asset price dynamics, in the parameter vector θ_1 . Moreover, the first element of $X_t = [X_{1,t} \dots X_{d,t}]^\top$ describes solely the dynamics of the asset price itself, which is usually the underlying for options. The remaining elements of X_t describe other risk factors, such as the volatility of the underlying, which affect the probability distribution of $X_{1,t}$. For better clarity, we set $S_t = X_{1,t}$ in the following.

The asset price dynamics under an equivalent martingale measure $\mathbb{Q}$ will be denoted by

$$dX_t = \tilde{\mu}(X_t)dt + \tilde{\sigma}(X_t)d\tilde{W}_t , \quad (4.38)$$

which are relevant for the purpose of option pricing.²⁶ In particular, the time t price of a call option with underlying S_t , strike price K and maturity τ can be obtained in accordance to the risk-neutral pricing formula (2.17) as follows:

$$\mathcal{C}(X_t, K, \tau | \theta_2) = e^{-r\tau} \int_0^\infty \max\{x - K, 0\} f_{S_T|\mathcal{F}_t}^\mathbb{Q}(x | \theta_2) dx , \quad (4.39)$$

where $T = t + \tau$. Moreover, $f_{S_T|\mathcal{F}_t}^\mathbb{Q}(x | \theta_2)$ denotes the parametric risk-neutral distribution of S_T conditional on the information set $\mathcal{F}_t$. Similarly as before, the vector θ_2 summarizes the parameters of $\tilde{\mu}(x)$ and $\tilde{\sigma}(x)$.

²⁴Recall that $\mathcal{I}_{2D}$ and $\mathcal{I}_{OP}$ are different estimates for the information matrix, which are defined in (4.15) and (4.16), respectively.

²⁵For more details on affine processes see Chapter 3.2.

²⁶The asset price dynamics under an equivalent martingale measure $\mathbb{Q}$ can be derived by virtue of Girsanov's theorem, which is described in Chapter 2.3 in more details.

Since option prices can be determined by integrating the discounted pay-off profile with respect to the risk-neutral density of the underlying, each option price comprises risk-neutral information. Following Breeden and Litzenberg [41], the discounted risk-neutral density of the underlying can be recovered by twice differentiating the call option price with respect to the strike price, which is²⁷

$$e^{-r\tau} f_{S_T|\mathcal{F}_t}^{\mathbb{Q}}(x) = \left. \frac{\partial^2 \mathcal{C}(X_t, K, \tau)}{\partial K^2} \right|_{K=x}.$$

Hence, if call option prices for a given underlying and maturity can be observed for the whole continuum of strike prices, the implied risk-neutral density of the underlying can be recovered. This highlights the cross-sectional information content of option prices.

Since (4.38) implies a parametric risk-neutral density for the underlying, it is sufficient to fit the option pricing model (4.39) to a relatively small cross-section of option prices in order to recover the complete risk-neutral asset price dynamic. We therefore assume that a cross-section of option prices $\bar{\mathcal{C}}_{1,t}, \dots, \bar{\mathcal{C}}_{n,t}$ with strike prices $K_1, \dots, K_n$ and maturities $\tau_1, \dots, \tau_n$ can be observed. Then the quality of the fit for the option pricing model (4.39) to the cross-section of option prices can be evaluated by some loss function. A commonly used loss function in this context is the sum of the squared pricing errors (see e.g. Shackleton et al. [143, p. 2681] or Taylor [148, p. 434]) so that

$$\hat{\theta}_2 = \arg \min_{\theta_2} \sum_{i=1}^n [\bar{\mathcal{C}}_{i,t} - \mathcal{C}(X_t, K_i, \tau_i | \theta_2)]^2,$$

is a possible estimator for θ_2 .²⁸ The advantage of this parametric approach compared to the non-parametric approach of Breeden and Litzenberg [41] is that θ_2 determines the risk-neutral distribution of the underlying for all time horizons and not only for some specific time horizon.

Option prices also contain information about the statistical dynamics of the underlying asset price process X_t . Of course, if X_t is directly observable, we can also use observations of the asset price process in order to estimate its statistical dynamics. Unfortunately, X_t is not always (completely) observable, which aggravates the estimation. For example, the firm-value process of a company, which is relevant for the pricing of credit derivatives, is usually not observable. But since the time series behavior of option prices depends on the statistical dynamics of the underlying asset price process, option prices can be used to infer the dynamics of (4.37). To illustrate this, assume that a time series of call option prices $\bar{\mathcal{C}}_{t_1}, \dots, \bar{\mathcal{C}}_{t_N}$ with strike prices $K_{t_1}, \dots, K_{t_N}$ and maturities $\tau_{t_1}, \dots, \tau_{t_N}$ can be observed. Furthermore, assume that the time series $\bar{\mathcal{C}}_{t_1}, \dots, \bar{\mathcal{C}}_{t_N}$ can completely be explained by the option pricing model (4.39), so that $\bar{\mathcal{C}}_{t_1} = \mathcal{C}(X_{t_1}), \dots, \bar{\mathcal{C}}_{t_N} = \mathcal{C}(X_{t_N})$.²⁹ Now, if we could invert the option prices for the underlying asset prices, we could also recover the asset price time series $X_{t_1} = \mathcal{C}^{-1}(\bar{\mathcal{C}}_{t_1}), \dots, X_{t_N} = \mathcal{C}^{-1}(\bar{\mathcal{C}}_{t_N})$, which in turn can

²⁷This result holds for any risk-neutral density and is not restricted to some parametric density.

²⁸Alternatively, we can express observed and theoretical option prices in terms of Black-Scholes implied volatilities and then use the sum of the squared difference of the implied volatilities as a loss function.

²⁹For better clarity, we omitted the dependence of the theoretical option prices on K_t , τ_t and θ_2 .

4. Estimation of option pricing models

be used to estimate the statistical asset price dynamics. Unfortunately, observed option prices cannot fully be explained by theoretical option pricing models due to measurement errors and the partial inadequacy of the underlying model. Moreover, the inversion of option pricing models is generally an ill-posed problem. Despite these difficulties, one possibility to recover the asset price dynamics from time series of option prices is an application of the Kalman filter, which also allows to account for measurement errors. In particular, the Kalman filter allows to jointly estimate the statistical and risk-neutral asset price dynamics.

4.4.2. State-space representation for option pricing models

In the context of option pricing, the state-space representation describes the statistical dynamics of the underlying asset price process, which is linked to observed option prices by a theoretical option pricing model. In the following, we assume that the asset price process is specified by (4.37) and the theoretical option pricing model is given by (4.39). Moreover, we suppose an equidistant time grid $0 = t_1 < \dots < t_N = T$, where $\Delta t = t_i - t_{i-1}$. Furthermore, we will set $i = t_i$ if not stated otherwise. In accordance to Bakshi and Wu [16], the state equation can now be obtained by an Euler discretization of the statistical asset price dynamics (4.37), which is given by³⁰

$$X_i = X_{i-1} + \mu(X_{i-1})\Delta t + \sigma(X_{i-1})\sqrt{\Delta t}\eta_i \quad \text{for } i = 1, \dots, N,$$

where

$$\eta_i \sim \mathcal{N}(0, I_d).$$

Here I_d denotes the d -dimensional identity matrix. Of course, this is not the only possible state equation in order to describe the asset price dynamics X_t . In particular, some SDEs do have a closed-form solution, which can be used to formulate the state equation.³¹ Unfortunately, for many other SDEs there exists no closed-form solution, so that the state equation can only be obtained via an approximation.³² Further note that the classical Kalman filter cannot be applied in this setting since the disturbance term η_i is not a single additive term.

Next, the measurement equation relates the state vector X_i to observable variables. As measurements, we will use a time series of observed call option prices $\bar{\mathcal{C}}_1, \dots, \bar{\mathcal{C}}_N$, where $\bar{\mathcal{C}}_i = [\bar{\mathcal{C}}_{1,i} \dots \bar{\mathcal{C}}_{n,i}]^\top$ is a $n \times 1$ vector of call option prices with strike prices $K_i = [K_{1,i} \dots K_{n,i}]^\top$ and maturities $\tau_i = [\tau_{1,i} \dots \tau_{n,i}]^\top$. Following Bakshi and Wu [16], these observed call option prices can be explained by the option pricing model (4.39), so that the measurement equation can be expressed as

$$\bar{\mathcal{C}}_i = \mathcal{C}(X_i, K_i, \tau_i | \theta_2) + \epsilon_i \quad \text{for } i = 1, \dots, N.$$

³⁰See Seydel [142, p. 101 ff.] for a general treatment of discretizing stochastic differential equations.

³¹An example for a SDE with closed-form solution is the *geometric Brownian motion*, which is used, for example, to model the asset price process in the Black-Scholes model.

³²An example for a SDE with no closed-form solution is the *Cox-Ingersoll-Ross* process, which is used, for example, to model the variance in the Heston model.

4.4. Estimation of option pricing models via the unscented Kalman filter

The measurement error $\epsilon_i = [\epsilon_{1,i}, \dots, \epsilon_{n,i}]^\top$ offsets the difference between observed and theoretical option prices and is assumed to be normally distributed with

$$\epsilon_i \sim \mathcal{N}(0, \Sigma) \quad \text{for } i = 1, \dots, N .$$

Moreover, ϵ_i is assumed to be uncorrelated to η_i at all times, so that

$$\mathbb{E}[\eta_i \epsilon_j^\top] = 0 \quad \text{for } i, j = 1, \dots, N .$$

In accordance to Bakshi et al. [11], we assume that the variance matrix of the measurement error is defined as

$$\Sigma = \sigma_\epsilon^2 I_n , \tag{4.40}$$

where σ_ϵ is a scalar. Hence, the single measurement errors $\epsilon_{1,i}, \dots, \epsilon_{n,i}$ are uncorrelated with each other and do all have the same variance. In particular, the variance matrix of ϵ_i is parameterized with only one parameter. This is essentially the most simple assumption that can be made for the specification of a measurement error. Of course, other parameterizations for the variance matrix of the measurement error might be more realistic but do have the drawback of high complexity, which aggravates the parameter estimation. An example for a slightly more complex parameterization of Σ is to allow the single measurement errors to have distinctive variances but still suppose that they are uncorrelated to each other. But with $n = 10$ option prices per measurement, the number of parameters of Σ already exceeds the number of parameters of most option pricing models. Moreover, the more complex the specification of the measurement error, the more structure of the data is captured by the measurement error and not by the option pricing model. This in turn aggravates the estimation of the statistical and risk-neutral asset price dynamics parameterized by θ_1 and θ_2 , respectively.

4.4.3. Vega-scaling

Instead of modifying the measurement error in order to better capture the characteristics of the data, one can alternatively transform the data, such that the assumptions of the measurement error are fulfilled as good as possible. A helpful transformation in this context is to divide option prices by their corresponding Black-Scholes vega. Bakshi et al. [11] propose this data transformation for currency options and Bakshi and Wu [16] for index options, both in order to estimate the underlying option pricing model with the UKF, where the variance of the measurement error is specified as in (4.40).

Black-Scholes vega refers to the derivative of the Black-Scholes option price with respect to the volatility parameter σ and is given by [98, p. 365]

$$v_{BS}(S_t, K, \tau, \sigma) = S_t \phi(d_1) \sqrt{\tau} , \tag{4.41}$$

where

$$d_1 = \frac{\ln(S_t/K) + (r + \sigma^2/2)\tau}{\sigma \sqrt{\tau}} .$$

4. Estimation of option pricing models

Figure 4.2.: Time series of regular and vega-scaled option prices. These figures illustrate time series of regular (a) and vega-scaled (b) call option prices with strike price $K_t = 1,05 \cdot S_t$ and constant maturity $\tau = 1/12$. The data are simulated option prices based on the Heston model with parameters $\phi = 0.25^2$, $\kappa = 5$, $\lambda = 0.25$, $\rho = -0.5$ and $\gamma = 0$. Furthermore, we suppose weekly observations, i.e. $\Delta t = 1/52$, and $N = 1000$ observation days.

Here ϕ denotes the density of a standard normal random variable. Hence, Black-Scholes vega is a measure for the sensitivity of option prices to changes in volatility and depends on the maturity and moneyness of an option as well as on the volatility level. The latter can be determined for each observed call option price $\bar{\mathcal{C}}_{k,i}$ by the corresponding Black-Scholes implied volatility $\sigma_{BS}(K_{k,i}, \tau_{k,i})$.³³ Following Bakshi et al. [11] and Bakshi and Wu [16], vega-scaled option prices can now be obtained if we divide observed option prices by their corresponding Black-Scholes vega, which is

$$y_{k,i} = \frac{\bar{\mathcal{C}}_{k,i}}{v_{BS}(S_i, K_{k,i}, \tau_{k,i}, \sigma_{BS}(K_{k,i}, \tau_{k,i}))}.$$

Usually, a whole cross-section $\bar{\mathcal{C}}_i = [\bar{\mathcal{C}}_{1,i} \dots \bar{\mathcal{C}}_{n,i}]^\top$ of call options with strike prices $K_i = [K_{1,i} \dots K_{n,i}]^\top$ and maturities $\tau_i = [\tau_{1,i} \dots \tau_{n,i}]^\top$ can be observed, which is used as data basis for the estimation. For ease of notation, we will refer to $v_{BS}(\bar{\mathcal{C}}_i)$ as the $n \times 1$ vector of Black-Scholes vegas for the observed option prices $\bar{\mathcal{C}}_i$. Hence, the corresponding vector of vega-scaled option prices can be denoted by

$$y_i = v_{BS}^{-1}(\bar{\mathcal{C}}_i) \odot \bar{\mathcal{C}}_i, \quad (4.42)$$

where $\odot$ denotes the *Hadamard product*.³⁴ Moreover, the minus one in the exponent of

³³The Black-Scholes implied volatility is defined in Chapter 3.4.1.

³⁴The Hadamard product of two matrices $A, B \in \mathbb{R}^{n \times m}$ is defined as the entrywise product of the two

v_{BS}^{-1} indicates that we consider the inverse of the elements of $v_{BS}(\bar{\mathcal{C}}_i)$.

Table 4.2.: Standard deviation of regular and vega-scaled option prices.

(a)						
	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6
$K_{1,t}$	10.28	19.10	26.28	32.49	38.05	43.15
$K_{2,t}$	3.43	9.12	14.60	19.75	24.60	29.17
$K_{3,t}$	1.00	4.01	7.63	11.44	15.28	19.08

(b)						
	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6
$K_{1,t}$	2.34	2.15	1.89	1.65	1.45	1.28
$K_{2,t}$	1.64	1.70	1.57	1.42	1.27	1.14
$K_{3,t}$	1.15	1.32	1.28	1.19	1.09	1.00

Notes: These tables summarize the standard deviations for time series of regular (a) and vega-scaled (b) option prices with different strike prices and maturities. In each table, we divided all standard deviations by the respective minimum standard deviation in order to make the figures comparable. The calculations are based on time series of simulated option prices via the Heston model with parameters $\phi = 0.25^2$, $\kappa = 5$, $\lambda = 0.25$, $\rho = -0.5$ and $\gamma = 0$, $K_{i,t} = (1 + 0.05i)S_t$ and $\tau_i = i/12$. Furthermore, we suppose weekly observations, i.e. $\Delta t = 1/52$, and $N = 1000$ observation days.

Figure 4.2 illustrates time series of regular and vega-scaled simulated option prices with constant maturity and moneyness.³⁵ In particular, we can see that the time series of vega-scaled option prices is stationary. Moreover, the variance of time series of vega-scaled option prices is much more comparable across different strike prices and maturities. Hence, also the assumption of an identical variance for all measurement errors is much more realistic [16, p. 2257]. Table 4.2 compares the standardized standard deviation for time series of regular and vega-scaled simulated option prices with different strike prices and maturities. Note that the standard deviations of the regular option price time series summarized in Table 4.2(a) depend strongly on the moneyness and maturity and differ up to a factor of over 40. In contrast, the standard deviations of the vega-scaled option price time series summarized in Table 4.2(b) differ only up to a factor of two.

matrices, i.e. $(A \odot B)_{i,j} = (A)_{i,j} \cdot (B)_{i,j}$.

³⁵Even though time series of observed option prices do usually not fulfill these stylized characteristics, we can select option prices with maturities and moneyness within a specific range. For more details see also the empirical study in Chapter 6.3.

4.4.4. Estimation

In the previous two sections, we formulated a general state-space representation for option pricing models and discussed how vega-scaling facilitates a sparse parameterization of the measurement error. In accordance to Bakshi et al. [11] and Bakshi and Wu [16], we will now show how to jointly estimate the statistical and risk-neutral asset price dynamics by means of the UKF.

First note that vega-scaling can be incorporated into the state-space representation described in Chapter 4.4.2 if we divide both observed and theoretical option prices by their corresponding Black-Scholes vega. Following Bakshi et al. [11, p. 144 f.], an alternative state-space representation is then given by

$$\begin{aligned} X_i &= X_{i-1} + \mu(X_{i-1})\Delta t + \sigma(X_{i-1})\sqrt{\Delta t}\eta_i, \\ y_i &= v_{BS}^{-1}(\bar{\mathcal{C}}_i) \odot \mathcal{C}(X_i, K_i, \tau_i | \theta_2) + \epsilon_i, \end{aligned} \quad (4.43)$$

where $\eta_i \sim \mathcal{N}(0, I_m)$ and $\epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon^2 I_n)$ for $i = 1, \dots, N$. Moreover, y_i is defined as in (4.42). In particular, the distributional assumption of the measurement error is now much more realistic than without vega-scaling. Now, we are able to apply the UKF together with the state-space representation (4.43), so that the log-likelihood function for the observations $y_1, \dots, y_N$ can be approximated by

$$\hat{\mathcal{L}}(\theta) = -\frac{nN}{2} \ln(2\pi) - \frac{1}{2} \sum_{i=1}^N \ln \left(|\hat{S}_i(\theta)| \right) - \frac{1}{2} \sum_{i=1}^N \xi_i^\top \hat{S}_i(\theta)^{-1} \xi_i, \quad (4.44)$$

where $\xi_i = y_i - \hat{n}_i^-(\theta)$. Again, the vector θ summarizes the parameters of the statistical and risk-neutral asset price dynamics, which in turn are summarized in θ_1 and θ_2 , respectively. In particular, the statistical and risk-neutral asset price dynamics can be estimated by maximizing (4.44) with respect to θ , which requires a repeated application of the UKF.

5. Monte-Carlo study

5.1. Introduction

In this Chapter, we will assess how accurate the parameters of the Heston model and the 2-Factor SV model can be estimated by virtue of the UKF. In order to determine the estimation accuracy of the UKF, we will perform a Monte-Carlo study. The setup of this Monte-Carlo study is orientated towards the empirical study of Chapter 6, so that these empirical results can better be classified. Before we discuss the Monte-Carlo study in more details, Chapter 5.2 first examines some numerical difficulties within this estimation setup. Numerical difficulties in the form of discontinuous log-likelihood functions can arise if the adaptive Simpson rule, which can be used as numerical integration technique for the calculation of option prices, is combined with an inappropriate parameter choice for the unscented transformation. Moreover, Chapter 5.3 proposes a caching technique, which offers a great speed advantage for the estimation of the 2-Factor SV model together with the unscented Kalman filter. Then Chapter 5.4 describes the data characteristics as well as the underlying methodology for this Monte-Carlo study. Finally, in Chapter 5.5, we will present the results of this Monte-Carlo study.

5.2. Numerical difficulties

Numerical difficulties within the estimation framework presented in Chapter 4.4 can arise from different aspects. The reasons for these numerical difficulties are related to the measurement equation

$$y_i = v_{BS}^{-1}(\overline{\mathcal{C}}_i) \odot \mathcal{C}(S_i, x_i, K_i, \tau_i | \theta_2) + \epsilon_i , \quad (5.1)$$

which is part of the state-space representation (4.43). In contrast to the state equation of (4.43), the measurement equation (5.1) is no closed-form expression since option prices are calculated via Fourier inversion based on some numerical quadrature. This issue can cause problems in the prediction step of the UKF, where a set of sigma points is propagated

5. Monte-Carlo study

through the measurement equation. More precisely, at iteration step i the sigma points¹

$$\begin{aligned} x_{0,i} &= m_i^-(\theta) \\ x_{j,i} &= m_i^-(\theta) + \alpha\sqrt{p}\sigma_{j,i}^-(\theta) \quad \text{for } j = 1, \dots, p, \\ x_{j,i} &= m_i^-(\theta) - \alpha\sqrt{p}\sigma_{j,i}^-(\theta) \quad \text{for } j = p+1, \dots, 2p, \end{aligned} \quad (5.2)$$

are propagated through the measurement equation (5.1). Here $\sigma_{j,i}^-(\theta)$ refers to the j^{th} row of the matrix square root $\sqrt{R_i^-(\theta)}$, where θ indicates the dependence of m_i^- and R_i^- on the parameterization of the state-space representation.² Moreover, the parameter α determines the spread of the sigma points around $m_i^-(\theta)$. Within this Monte-Carlo study, the elements of $\sigma_{j,i}^-(\theta)$ are approximately of order 10^{-3} and $p \leq 2$, so that the spread of the sigma points around $m_i^-(\theta)$ is approximately of order $\alpha \cdot 10^{-3}$. Hence, the smaller α , the higher is the required degree of accuracy of the measurement equation in order to precisely map the dispersion of the sigma points. According to Wan and van der Merwe [151, p. 229], α is set to some small positive value between 10^{-4} and 1. In our setup, these bounds approximately translate into a spread of the sigma points of order 10^{-7} to 10^{-3} . This issue is not so much a problem for a single application of the UKF as for the QML estimation, where the UKF is applied for different values of θ . The QML estimation, in particular, requires that the log-likelihood function is smooth with respect to all parameters of the state-space representation. Otherwise, a frictionless optimization is not possible. Hence, the measurement equation also needs to precisely map small changes in θ , which usually results in even smaller changes in $m_i^-(\theta)$ and $\sigma_{j,i}^-(\theta)$.

Heart of the measurement equation (5.1) is the calculation of theoretical option prices, which is carried out by virtue of the Carr-Madan formula (3.10). The integral in (3.10) cannot be calculated in closed-form, so that numerical integration techniques are required in order to determine the theoretical option prices. A commonly used quadrature for this purpose is the *adaptive Simpson rule*, as for example proposed by Dobránszky [67] or Mikhailov and Nögel [127]. However, the adaptive Simpson rule together with a reasonable value for the absolute error tolerance is not able to calculate theoretical option prices accurate enough in order to precisely propagate sigma points with low dispersion through the measurement equation.³ As a consequence, the resulting log-likelihood function might exhibit discontinuities of which degree also depends on the sensitivity of $m_i^-(\theta)$ and $\sigma_{j,i}^-(\theta)$ to changes in θ . In particular, the smaller the sensitivity of $m_i^-(\theta)$ and $\sigma_{j,i}^-(\theta)$ to changes in θ , the flatter is the corresponding log-likelihood function and the higher is the degree of discontinuity.

Figure 5.1 illustrates this issue exemplarily for the parameters κ^* and ρ of the Heston model. In addition to the log-likelihood functions for these parameters, we also plot the corresponding first and second order derivatives, each with $\alpha = 0.001$ (see Figure 5.1(a)) and $\alpha = 0.1$ (see Figure 5.1(b)). While the first order derivative is important for the

¹Note that we set the secondary scaling parameter κ to zero in order to build the sigma points (5.2).

²Recall that m_i^- and R_i^- denote the mean and variance of the state x_i conditional on the information set $\mathcal{I}_{i-1}$.

³We used the MATLAB implementation `quad` for the adaptive Simpson rule and supposed the default value of 10^{-6} for the absolute error tolerance.

(a)

(b)

Figure 5.1.: Log-likelihood functions based on the adaptive Simpson rule. These figures illustrate the log-likelihood functions and the corresponding first and second order derivatives for the parameters κ^* and ρ of the Heston model, where option prices are calculated based on the MATLAB implementation `quad` for the adaptive Simpson rule with the default value of 10^{-6} for the absolute error tolerance. The upper figures (a) are computed with a low spread of $\alpha = 0.001$ for the sigma points and the lower figures (b) with a high spread of $\alpha = 0.1$ for the sigma points.

5. Monte-Carlo study

numerical optimization, the second order derivative is necessary to compute the Hessian of the log-likelihood function, which in turn is the basis to compute the standard errors for the estimated parameters. On the one hand, it is apparent that the degree of discontinuity is higher for small values of α than for large values. Furthermore, even though the log-likelihood function of ρ is smooth for $\alpha = 0.1$, the corresponding first and second order derivatives still exhibit discontinuities. On the other hand, the degree of discontinuity of the log-likelihood function for κ^* is smaller than for ρ . A possible explanation for the latter observation might be the fact that $m_i^-(\theta)$ and $\sigma_{j,i}^-(\theta)$ are more sensitive to changes in κ^* than in ρ .

To overcome the problem of discontinuous log-likelihood functions, we recommend to choose rather large values for α when applying the UKF in this context. Furthermore, in contrast to the adaptive Simpson rule, the *adaptive Gauss-Kronrod rule* together with reasonable values for the absolute error tolerance is well able to calculate theoretical option prices with the required accuracy, so that the resulting log-likelihood functions are smooth with respect to all parameters.⁴ Figure 5.2 illustrates this issue exemplarily for the parameters κ^* and ρ of the Heston model. In particular, the log-likelihood functions for κ^* and ρ as well as their first and second order derivatives are now smooth, irrespective of the choice for α . In addition, the adaptive Gauss-Kronrod rule is much faster than the adaptive Simpson rule. For example, the generation of one data sample for the Monte-Carlo study of the Heston model, wherefore 4000 option prices need to be calculated, requires 56 seconds based on the adaptive Gauss-Kronrod rule and 469 seconds based on the adaptive Simpson rule.⁵ This result is in line with Schmelzle [141, p. 75], who reports that the adaptive Gauss-Kronrod rule is approximately 5 times faster than the adaptive Simpson rule.

5.3. A caching technique for the 2-Factor SV model

The calculation of option prices based on the 2-Factor SV model is numerically even more challenging compared to the Heston model. This is due to the fact that the characteristic exponents $\phi(\tau, u)$ and $\psi_3(\tau, u)$, which partly specify the characteristic function of the 2-Factor SV model, need to be calculated numerically. This can be performed e.g. via the 4th-order Runge-Kutta method, which iteratively calculates $\psi_3(t, u)$ for $t \in \{t_1, \dots, t_M\}$, where $t_1 = 0$ and $t_M = \tau$.⁶ The calculation of $\phi(\tau, u)$ essentially involves to integrate $\psi_3(t, u)$ with respect to time, which can be performed e.g. via the *trapezoid rule* based on the numerical approximation of $\psi_3(t, u)$. In this Monte-Carlo study, we therefore evaluate $\psi_3(t, u)$ based on an equidistant time grid with $t_i - t_{i-1} = 0.01$. The computationally

⁴We used the MATLAB implementation `quadgk` for the adaptive Gauss-Kronrod rule and supposed an absolute error tolerance of 10^{-6} (10^{-10} by default).

⁵These calculations were performed on a Windows 7 computer with an Intel Core i3 CPU 550 with 3.2GHz and 4 GB RAM. We therefore supposed an absolute error tolerance of 10^{-6} and an upper integration bound of 100.

⁶We used the MATLAB implementation `ode45` for the 4th-order *Runge-Kutta method* with a relative error tolerance of 10^{-6} (10^{-3} by default).

5.3. A caching technique for the 2-Factor SV model

(a)

(b)

Figure 5.2.: Log-likelihood functions based on the adaptive Gauss-Kronrod rule. These figures illustrate the log-likelihood functions and the corresponding first and second order derivatives for the parameters κ^* and ρ of the Heston model, where option prices are calculated based on the MATLAB implementation `quadgk` for the adaptive Gauss-Kronrod rule with an absolute error tolerance of 10^{-6} (10^{-10} by default). The upper figures (a) are computed with a low spread of $\alpha = 0.001$ for the sigma points and the lower figures (b) with a high spread of $\alpha = 0.1$ for the sigma points.

most expensive part in calculating option prices based on the 2-Factor SV model is the calculation of the characteristic exponent $\psi_3(\tau, u)$ via the Runge-Kutta method. For example, the calculation of $\psi_3(\tau, u)$ for $\tau = 1/2$ and $u = 0, 0.1, \dots, 100$, which is a reasonable number of evaluations of $\psi_3(\tau, u)$ in order to calculate one specific option prices with high accuracy, takes approximately 25 seconds.⁷ In contrast, the calculation of 4000 different option prices based on the Heston model together with the adaptive Gauss-Kronrod rule takes “only” 56 seconds. Of course, we can also apply an adaptive integration method, as e.g. the Gauss-Kronrod rule, together with the 2-Factor SV model. But in our setup this would require to apply the Runge-Kutta method for each option price calculation at each iteration step of the UKF anew. This is due to the fact that the integrand of the Carr-Madan formula (3.10) changes if either the strike price K , the maturity τ or the state variables S_t , σ_t and θ_t change. Hence, also the optimal abscissas for the adaptive integration change, which makes it necessary to evaluate $\psi_3(\tau, u)$ always at different locations. Alternatively, we propose to calculate the characteristic exponents of the 2-Factor SV model in advance for the supposed parameterization θ_2 , the required maturities $\bar{\tau} = \bigcup_{i=1}^N \tau_i$ and a prespecified integration grid $u_1, \dots, u_N$. Based on these pre-calculations of the characteristic exponents, option prices can be calculated for all levels of S_t , σ_t , θ_t as well as all K and $\tau \in \bar{\tau}$. This can be achieved e.g. by the trapezoid rule, where the accuracy of the option price calculation depends on the refinement of the integration grid and the upper integration bound u_N .⁸ In particular, this approach is sufficient to evaluate the measurement equation at each iteration step of the UKF and thus is much faster than to recalculate the characteristic exponents for each option price calculation anew. Note that this approach is similar to the caching technique of Kilin [115], who proposes to store the evaluations of the characteristic function for a given maturity. But in contrast to Kilin [115], we are also able to calculate option prices for different maturities and state variables and not only for different strike prices. Of course, we can also apply this caching technique for the Heston model, but the speed advantage compared to the 2-Factor SV model is relatively low, since the characteristic exponents for the Heston model are known in closed-form.

5.4. Data

5.4.1. Data characteristics

The data characteristics for this Monte-Carlo study are chosen in accordance to the data characteristics of our empirical study, which will be presented in Chapter 6. Hence, we will simulate weekly data, i.e. $\Delta t = 1/52$, over a period of approximately 10 years, i.e. $N = 500$ observation days. Moreover, we will simulate not far *out of the money*

⁷Again, these calculations were performed on a Windows 7 computer with an Intel Core i3 CPU 550 with 3.2GHz and 4 GB RAM. We therefore parameterized ψ_3 by $\kappa_1^* = 3$, $\kappa_2 = 7$, $\phi_2 = 5/3 \cdot 0.25^2$, $\lambda_1 = 0.25$, $\lambda_2 = 0.25$ and $\rho = -0.5$.

⁸In this Monte-Carlo study, we assume that the integration grid is equidistant with $u_i - u_{i-1} = 0.1$ and $u_N = 100$.

(OTM) call option prices with constant maturities of 1, 2, 3 and 5 months.⁹ These maturities correspond approximately to the average maturities of the call options used in our empirical study. Furthermore, the strike prices of traded options are usually on a certain strike price grid. For example, S&P 500 index options, which are also used for our empirical study, are usually traded with strike prices on an equidistant strike price grid with a *grid size* (GS) of 25 points.¹⁰ In accordance to these market conditions for S&P 500 index options, we simulate OTM call options by choosing strike prices via the following:

$$K_{j,i} = \lfloor S_t / GS \rfloor \cdot GS + j \cdot GS \quad \text{for } j = 1, \dots, l. \quad (5.3)$$

Here $\lfloor \cdot \rfloor$ denotes the floor function and l determines the number of different strike prices per measurement.¹¹ The strike choice function (5.3) is illustrated in Figure 5.3. For this Monte-Carlo study, we will set $GS = 25$ and $l = 2$, so that the strike price and maturity vectors can be defined as

$$K_i = [K_{1,i} \ K_{1,i} \ K_{1,i} \ K_{1,i} \ K_{2,i} \ K_{2,i} \ K_{2,i} \ K_{2,i}] , \quad (5.4)$$

and

$$\tau_i = [1/12 \ 2/12 \ 3/12 \ 5/12 \ 1/12 \ 2/12 \ 3/12 \ 5/12] , \quad (5.5)$$

respectively.

Figure 5.3.: Strike choice function. This figure illustrates the strike choice function (5.3), where we set $GS = 25$ and $l = 2$.

Besides the advantage of high liquidity, the sensitivity of near the money options to changes in volatility is higher than for far OTM options. In particular, not far OTM call options are highly sensitive to changes in volatility. This is stylized in Figure 5.4, where the dotted line indicates the sensitivity of an *at the money* (ATM) call option, so that left/right of the dotted line are OTM/ITM (*in the money*) call options. Hence,

⁹The option prices are calculated based on the Heston model and the 2-Factor SV model. The specific parameterization for these option pricing models will be discussed in Chapter 5.5.

¹⁰The strike price grid for near the money S&P 500 index options is usually finer with a grid size of 5 points.

¹¹The floor function is defined as $\lfloor x \rfloor = \max\{m \in \mathbb{Z} \mid m \leq x\}$.

Figure 5.4.: Volatility sensitivity of option prices. This figure illustrates the sensitivity of the Black-Scholes option prices to changes in volatility by virtue of Black-Scholes vega. The underlying parameters are $\sigma = 0.3$ and $T = 5$.

not far OTM call options do have the highest sensitivity to changes in volatility. The sensitivity of option prices to changes in volatility is important from an estimation point of view. This is due to the fact that the Kalman filter exploits the relation between the state x_i and the measurement y_i . The closer these variables are related to each other, the more information about the state can be inferred from the measurement and in turn the “easier” is the estimation procedure for the Kalman filter. This reasoning becomes clearer if we consider the updating equation for the conditional mean of the state, which is given by

$$m_i = m_i^- + C_i^\top S_i^{-1} (y_i - n_i^-) ,$$

where $C_i = \text{Cov}[x_i, y_i | \mathcal{I}_{i-1}]$ and $S_i = \text{Var}[y_i | \mathcal{I}_{i-1}]$. Hence, if the covariance between the state x_i and the measurement y_i approaches zero, also the Kalman filter gain $C_i^\top S_i^{-1}$ approaches zero, so that the measurement y_i has no influence in predicting x_i . The superior estimation accuracy of the Kalman filter based on ATM options compared to deep ITM options is illustrated in Figure 5.5.

5.4.2. Data simulation

In order to perform this Monte-Carlo study, we first need to simulate data based on the Heston model and the 2-Factor SV model. Both models suppose that the statistical asset price dynamics are given by

$$dS_t = \mu S_t dt + \sqrt{\sigma_t} S_t dW_{1,t} , \quad (5.6)$$

$$d\sigma_t = \kappa_1 (\phi_1 - \sigma_t) dt + \lambda_1 \sqrt{\sigma_t} dW_{2,t} , \quad (5.7)$$

Figure 5.5.: Estimation accuracy based on deep in-the-money and out-of-the-money options. This figure illustrates the estimation accuracy of the state equation for the Heston model, which describes the evolution of the volatility. The estimation is performed with the UKF based on a time series of simulated Heston call option prices with a constant maturity of one year. Moreover, the left figure (a) is based on deep ITM call options with $S_t/K = 2$ and the right figure (b) based on ATM call options with $S_t/K = 1$.

where we set $\kappa_1 = \kappa$, $\phi_1 = \phi$ and $\lambda_1 = \lambda$ in the Heston model. In addition, the 2-Factor SV model supposes that the risk-aversion process evolves as

$$d\theta_t = \kappa_2(\phi_2 - \theta_t)dt + \lambda_2\sqrt{\theta_t}dW_{3,t} . \quad (5.8)$$

In the following, we will suppose an equidistant time grid $t_0, \dots, t_N$ with spacing Δt . Moreover, we assume that σ_t and θ_t stay constant over the time intervals $[t_{i-1}, t_i]$ for $i = 1, \dots, N$, so that the underlying S_t evolves piecewise as a geometric Brownian motion. Under these simplifying assumptions, the processes (5.6) – (5.8) can be approximated via the following:

$$S_i = S_{i-1} \exp\left\{\left(\mu - \frac{1}{2}\sigma_{i-1}\right)\Delta t + \sqrt{\sigma_{i-1}}\xi_i\right\} , \quad (5.9)$$

$$\sigma_i = e^{-\kappa_1\Delta t}\sigma_{i-1} + \phi_1(1 - e^{-\kappa_1\Delta t}) + e^{-\kappa_1\Delta t}\lambda_1\sqrt{\sigma_{i-1}}(\rho\xi_i + \sqrt{1-\rho^2}\eta_{1,i}) , \quad (5.10)$$

$$\theta_i = e^{-\kappa_2\Delta t}\theta_{i-1} + \phi_2(1 - e^{-\kappa_2\Delta t}) + e^{-\kappa_2\Delta t}\lambda_2\sqrt{\theta_{i-1}}\eta_{2,i} , \quad (5.11)$$

where ξ_i , $\eta_{1,i}$ and $\eta_{2,i}$ are independent and identically normally distributed random variables with zero mean and variance Δt .¹² For ease of notation, we now abstract from a specific option pricing model and set $x_i = \sigma_i$ and $\eta_i = \eta_{1,i}$ for the Heston model and $x_i = [\sigma_i \ \theta_i]^\top$ and $\eta_i = [\eta_{1,i} \ \eta_{2,i}]^\top$ for the 2-Factor SV model, so that the evolution of the

¹²The supposed discretization of the processes σ_t and θ_t , which both follow a CIR process, are derived in Appendix B.2.

5. Monte-Carlo study

state x_i can more generally be written as

$$x_i = f(x_{i-1}, \eta_i | \theta_1) \quad \text{for } i = 1, \dots, N.$$

Here the function f is indirectly specified by (5.10) – (5.11), where θ_1 denotes the corresponding parameters of the Heston model and the 2-Factor SV model, respectively.

Now, the data for this Monte-Carlo study can be generated by the following algorithm:

Algorithm 5.1 (Data simulation). *Initialize the algorithm with some starting values for the underlying S_0 and the state x_0 . Moreover, parameterize the state dynamics and the corresponding option pricing model via θ_1 and θ_2 , respectively. Then perform the following steps recursively for $i = 1, \dots, N$:*

1. *Simulate the disturbance terms ξ_i (1×1), η_i ($m \times 1$) and ϵ_i ($n \times 1$), where*

$$\begin{aligned} \xi_i &\sim \mathcal{N}(0, \Delta t) , \\ \eta_i &\sim \mathcal{N}(0, \Delta t I_m) , \\ \epsilon_i &\sim \mathcal{N}(0, \sigma_\epsilon^2 I_n) . \end{aligned}$$

2. *Determine the time i values for the underlying, the state and the theoretical call option prices via the following:*

$$\begin{aligned} S_i &= S_{i-1} \exp\left\{\left(\mu - \frac{1}{2}\sigma_{i-1}^2\right)\Delta t + \sqrt{\sigma_{i-1}^2} \xi_i\right\} , \\ x_i &= f(x_{i-1}, \eta_i | \theta_1) , \\ \bar{\mathcal{C}}_i &= \mathcal{C}(S_i, x_i, K_i, \tau_i | \theta_2) , \end{aligned}$$

where $\bar{\mathcal{C}}_i$ is a $n \times 1$ call option price vector with strike prices K_i ($n \times 1$) and maturities τ_i ($n \times 1$), which are specified in (5.4) and (5.5), respectively.

3. *Calculate the Black-Scholes implied volatilities $\sigma_{BS}(\bar{\mathcal{C}}_i)$ and then the Black-Scholes vegas $v_{BS}(\bar{\mathcal{C}}_i)$ for the theoretical call option prices $\bar{\mathcal{C}}_i$.*

4. *Generate vega-scaled and distorted call option prices via the following:*

$$y_i = v_{BS}^{-1}(\bar{\mathcal{C}}_i) \odot \bar{\mathcal{C}}_i + \epsilon_i .$$

For this Monte-Carlo study, we set the starting value for the underlying and the drift term to $S_0 = 1000$ and $\mu = 0.09$, respectively. The state-space representation within this Monte-Carlo study setup is given by

$$\begin{aligned} x_i &= f(x_{i-1}, \eta_i | \theta_1) , \\ y_i &= v_{BS}^{-1}(\bar{\mathcal{C}}_i) \odot \mathcal{C}(S_i, x_i, K_i, \tau_i | \theta_2) + \epsilon_i . \end{aligned} \tag{5.12}$$

Note that this formulation slightly differs from the state-space representation (4.43), which is applied together with empirical data, in the way that the vega-scaling factors are

calculated based on theoretical option prices and not based on observed option prices.

5.5. Results of the Monte Carlo study

5.5.1. Heston model

In order to estimate the Heston model by virtue of the UKF, we need to put the model into a suitable state-space representation. The state equation in this framework can be obtained by discretizing the SDE of the variance process (3.14), which is derived in Appendix B.2. Then, in accordance with the general state-space representation (5.12) for simulated data, the state-space representation for the Heston model is given by

$$\begin{aligned}\sigma_i &= e^{-\kappa\Delta t}\sigma_{i-1} + \phi(1 - e^{-\kappa\Delta t}) + e^{-\kappa\Delta t}\lambda\sqrt{\sigma_{i-1}}\eta_i, \\ y_i &= v_{BS}^{-1}(\bar{\mathcal{C}}_i) \odot \mathcal{C}_{Heston}(S_i, \sigma_i, K_i, \tau_i | \theta_2) + \epsilon_i,\end{aligned}\tag{5.13}$$

where $\eta_i \sim \mathcal{N}(0, \Delta t)$ and $\epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon I_8)$ for $i = 1, \dots, N$. Moreover, $\mathcal{C}_{Heston}$ denotes the call option price formula for the Heston model with risk-neutral parameters $\theta_2 = [\kappa^* \ \phi^* \ \lambda \ \rho]$.¹³ Recall that the risk-neutral parameters κ^* and ϕ^* are defined as $\kappa^* = \kappa + \gamma$ and $\phi^* = \kappa\phi/(\kappa + \gamma)$, where γ specifies the variance risk premium. Hence, the state-space representation for the Heston model can fully be parameterized by $\theta = [\kappa^* \ \phi^* \ \lambda \ \rho \ \gamma \ \sigma_\epsilon]$.

Table 5.1.: Simulation parameters for the Heston model.

	κ^*	ϕ^*	λ	ρ	γ	σ_ϵ
true	3	$5/3 \cdot 0.25^2$	0.25	-0.5	-2	0.01
guess	2	0.25^2	0.35	0	-1	0.02
min	10^{-6}	10^{-6}	10^{-6}	-1	$-\infty$	10^{-6}
max	∞	1	1	1	∞	1

The supposed parameterization of the Heston model within this Monte-Carlo study is listed in Table 5.1. Among the true parameters, which are used for the data simulation, Table 5.1 also states the starting values and the supposed bounds for the optimization. Moreover, we suppose that the starting value for the variance $\sigma_0 = 0.25^2$ is known.¹⁴ Before we discuss the results of the Monte-Carlo study for the Heston model, we illustrate some properties of this specific estimation setup based on a single simulated data set. Hence, these results should be interpreted with caution since they are rather indicative than representative.

¹³For more details about the Heston model see Chapter 3.4.

¹⁴For applications with real market data, the starting value for the variance can be initialized by the squared Black-Scholes implied volatility of an ATM option traded at the first day of the sample period.

Figure 5.6.: Filtering accuracy of the state equation for the Heston model. The solid line indicates the true simulated sample path and the dotted line the corresponding estimation result of the UKF. Moreover, the state-space representation (5.13) is parameterized by the QML estimator of true parameters summarized in Table 5.1, which was estimated based on a simulated data set.

At first, Figure 5.6 illustrates the filtering accuracy of the state equation for the Heston model, which is apparently very good. In addition, Figure 5.7 compares the log-likelihood functions for the single parameters of the state-space representation of the Heston model. The green lines indicate the true parameter values and the dotted lines the corresponding QML estimators. First note that the log-likelihood functions are smooth with respect to all parameters, which is necessary for a frictionless optimization.¹⁵ Furthermore, the numerical optimization routine has found all optima. The parameters ϕ^* , κ^* and σ_ϵ can be estimated very accurately since the deviations between the true and the estimated parameter values are very small. In contrast, the deviations between the true and the estimated parameter values for λ , ρ and γ are apparent, which also manifests itself in a flat log-likelihood function.¹⁶ The poor estimation accuracy of the parameter γ is due to the rare information content of the data about the statistical dynamics of the variance process.¹⁷ Also the poor estimation accuracy of the parameter ρ can be attributed to the data. The information about the correlation is contained in the smile/skew structure of the implied volatility surface. Hence, a more accurate estimation of ρ would require to use option price data with many and widely spread strike prices.

Next, Table 5.2 summarizes the estimation results of the Monte-Carlo study for the Heston model, which was performed based on 500 simulated data sets with characteristics as described in Chapter 5.4.2. The operators $\mu(\cdot)$ and $\sigma(\cdot)$ denote the sample mean and standard deviation, respectively. Furthermore, $\hat{\theta}$ denotes the QML estimator of the Heston parameters and $SE(\hat{\theta})$ the corresponding standard error estimates.¹⁸ At first, all QML estimators are unbiased since their sample means are approximately equal to their

¹⁵We used the MATLAB implementation `fmincon` for the optimization together with the following options: `TolFun = 10-12`, `algorithm = interior-point` and `FinDiffType = central`.

¹⁶Note that the scales of the y-axes in Figure 5.7 differ strongly. In particular, the range of the y-axis scale of the log-likelihood functions for the parameters λ , ρ and γ is very small, which indicates that the corresponding log-likelihood functions are flat.

¹⁷Recall from Chapter 4.4.1 that the statistical information about the variance process can only be extracted from the time series behavior of option prices.

¹⁸The standard errors of the QML estimators are calculated based on the second-order estimate of the information matrix.

Figure 5.7.: Log-likelihood functions for the parameters of the Heston model. These figures compare the log-likelihood functions for the single parameters of the state-space representation of the Heston model. The green lines indicate the true parameter values and the dotted lines the corresponding QML estimators.

Table 5.2.: Estimation results of the Monte-Carlo study for the Heston model.

	ϕ^*	κ^*	λ	ρ	γ	σ_ϵ
$\mu(\hat{\theta})$	0.104	3.012	0.250	-0.487	-2.030	0.010
$\sigma(\hat{\theta})$	(0.002)	(0.141)	(0.010)	(0.093)	(0.315)	(0.000)
$\mu(SE(\hat{\theta}))$	0.009	0.192	0.040	0.112	0.352	0.001
$\sigma(SE(\hat{\theta}))$	(0.148)	(0.383)	(0.573)	(0.137)	(0.418)	(0.012)

5. Monte-Carlo study

corresponding true values. In line with the previous discussion about the properties of this specific estimation setup, the standard errors of the parameters λ , ρ and γ are relatively high compared to the remaining parameters.¹⁹ Therefore, the estimation of these parameters is more difficult compared to the remaining parameters. Finally, the sample mean of the standard error estimates for the QML estimators are more or less similar to the standard deviations of the corresponding QML estimators, i.e. $\mu(SE(\hat{\theta})) \approx \sigma(\hat{\theta})$. However, the real size of the standard errors is overestimated for some parameters. Hence, our standard error estimates can be referred to as conservative estimates.

5.5.2. 2-Factor SV model

Analog to the Heston model, the state-space representation for the 2-Factor SV model can be written as

$$\begin{bmatrix} \sigma_i \\ \theta_i \end{bmatrix} = \begin{bmatrix} e^{-\kappa_1 \Delta t} \\ e^{-\kappa_2 \Delta t} \end{bmatrix} \odot \begin{bmatrix} \sigma_{i-1} \\ \theta_{i-1} \end{bmatrix} + \begin{bmatrix} \phi_1 (1 - e^{-\kappa_1 \Delta t}) \\ \phi_2 (1 - e^{-\kappa_2 \Delta t}) \end{bmatrix} + \begin{bmatrix} e^{-\kappa_1 \Delta t} \lambda_1 \\ e^{-\kappa_2 \Delta t} \lambda_2 \end{bmatrix} \odot \sqrt{\begin{bmatrix} \sigma_{i-1} \\ \theta_{i-1} \end{bmatrix}} \odot \eta_i, \quad (5.14)$$

$$y_i = v_{BS}^{-1}(\bar{\mathcal{C}}_i) \odot \mathcal{C}_{2FSV}(S_i, x_i, K_i, \tau_i | \theta_2) + \epsilon_i,$$

where $x_i = [\sigma_i \ \theta_i]^\top$, $\eta_i \sim \mathcal{N}(0, \Delta t I_2)$ and $\epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon^2 I_8)$ for $i = 1, \dots, N$. Moreover, $\mathcal{C}_{2FSV}$ denotes the call option price formula for the 2-Factor SV model, which is parameterized by $\theta_2 = [\kappa_1^* \ \kappa_2 \ \phi_2 \ \lambda_1 \ \lambda_2 \ \rho]$. Hence, the state-space representation for the 2-Factor SV model can fully be parameterized by $\theta = [\kappa_1 \ \kappa_1^* \ \kappa_2 \ \phi_1 \ \phi_2 \ \lambda_1 \ \lambda_2 \ \rho \ \sigma_\epsilon]$.

Table 5.3.: Simulation parameters for the 2-Factor SV model.

	κ_1	κ_1^*	κ_2	ϕ_1	ϕ_2	λ_1	λ_2	ρ	σ_ϵ
true	5	3	7	0.25^2	$5/3 \cdot 0.25^2$	0.25	0.25	-0.5	0.01
guess	4	2	6	0.20^2	0.25^2	0.35	0.35	0	0.02
min	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	-1	10^{-6}
max	∞	∞	∞	1	1	1	1	1	1

The supposed parameterization of the Monte-Carlo study for the 2-Factor SV model is listed in Table 5.3. In accordance to the previous setting of the Monte-Carlo study for the Heston model (see Table 5.1), we choose the same values for the corresponding parameters. Among the true parameterization, which is used for the data simulation, Table 5.3 also states the starting values and the bounds for the optimization. Moreover, we suppose that the starting values for the variance $\sigma_0 = 0.25^2$ and the risk aversion $\theta_0 = 5/3 \cdot 0.25^2$ are known.²⁰ Again, before we discuss the results of the Monte-Carlo study for the 2-Factor

¹⁹Note that the standard errors need to be related to the absolute value of the corresponding parameter estimates in order to judge if the standard error is relatively high or low.

²⁰For applications with real market data, the starting values for the variance and the risk aversion can both be initialized by the squared Black-Scholes implied volatility of an ATM option traded at the first day of the sample period.

SV model, we illustrate some properties of this specific estimation setup based on a single simulated data set.

Figure 5.8.: Filtering accuracy of the state equation for the 2-Factor SV model. The solid line indicates the true simulated sample path and the dotted line the corresponding estimation result of the UKF. Figure (a) shows the filtering accuracy of the variance σ_t and figure (b) the filtering accuracy of the risk aversion θ_t . Moreover, the state-space representation (5.14) is parameterized by the QML estimator of true parameters summarized in Table 5.3, which was estimated based on a simulated data set.

At first, Figure 5.8 illustrates how accurate the state equation of the 2-Factor SV model can be estimated by virtue of the UKF. Apparently, the filtering accuracy of the variance σ_t is very good whereas the filtering accuracy of the risk aversion θ_t is only satisfactory. Hence, the estimation setup of the 2-Factor SV model is more challenging than the one of the Heston model. This can also be seen from Figure 5.9, which compares the log-likelihood functions of the single parameters of the state-space representation for the 2-Factor SV model. First note that the log-likelihood functions are smooth with respect to all parameters, which is necessary for a frictionless optimization. Moreover, the numerical optimization routine has found all optima. The log-likelihood functions for the parameters κ_1 , κ_2 , ϕ_1 , λ_2 and ρ are extremely flat, which indicates that the estimation of these parameters is quite difficult. This reasoning is also confirmed by the estimation results for these parameters (dotted line), which partly deviate strongly from the corresponding true parameter values (green line). By chance the parameter κ_1 could be estimated very accurately, which is in contrast to the results of the following Monte-Carlo study (see Table 5.4). Again, the difficulty in estimating the parameters κ_1 and ϕ_1 can be explained by the rare information content of the data about the statistical dynamics of the variance process. In contrast, the curvature of the log-likelihood functions for the parameters κ_1^* , ϕ_2 , λ_1 and σ_ϵ is much stronger, which indicates that the estimation of these parameters is easier. This reasoning is confirmed by the estimation results of these parameters (dotted line), which more or less agree with the corresponding true parameter values (green line).

5. Monte-Carlo study

Figure 5.9.: Log-likelihood functions for the parameters of the 2-Factor SV model. These figures compare the log-likelihood functions for the individual parameters of the state-space representation of the 2-Factor SV model. The green lines indicate the true parameter values and the dotted lines the corresponding QML estimators.

Next, Table 5.4 summarizes the estimation results of the Monte-Carlo study for the 2-Factor SV model, which again was performed based on 500 simulated data sets with characteristics as described in Chapter 5.4.2. First note that the sample mean of the QML estimators for the parameters κ_1 , κ_2 and λ_2 slightly deviate from the corresponding true values. This poor result can be explained by the general difficulty to estimate these parameters within this estimation setup, which is also evident from the corresponding high standard errors $\sigma(\hat{\theta})$.²¹ In contrast, the remaining QML estimators are unbiased since their sample means are approximately equal to their corresponding true values. Finally, the sample mean of the standard error estimates for the QML estimators are more or less similar to the standard deviations of the corresponding QML estimators, i.e. $\mu(SE(\hat{\theta})) \approx \sigma(\hat{\theta})$. However, the real size of the standard errors is overestimated for some parameters. Hence, our standard error estimates can again be referred to as conservative estimates.

Table 5.4.: Estimation results of the Monte-Carlo study for the 2-Factor SV model.

	κ_1	κ_1^*	κ_2	ϕ_1	ϕ_2	λ_1	λ_2	ρ	σ_ϵ
$\mu(\hat{\theta})$	5.436	3.034	7.708	0.062	0.104	0.250	0.287	-0.486	0.010
$\sigma(\hat{\theta})$	(1.166)	(0.291)	(2.279)	(0.004)	(0.003)	(0.013)	(0.122)	(0.092)	(0.000)
$\mu(SE(\hat{\theta}))$	1.694	0.453	2.552	0.014	0.020	0.065	0.155	0.155	0.003
$\sigma(SE(\hat{\theta}))$	(7.311)	(1.300)	(4.313)	(0.138)	(0.105)	(0.820)	(0.406)	(0.381)	(0.023)

²¹Maybe the results of this Monte-Carlo study with respect to the unbiasedness of the estimated parameters could be further improved for a larger number of Monte Carlo simulations, which was not realized due to the already high computational complexity of this Monte-Carlo study.

6. Empirical study

6.1. Introduction

In the following Chapter we will jointly estimate the statistical and risk-neutral asset price dynamics of the Heston and 2-Factor SV model. The estimation is performed via the UKF based on time series of at-the-money and out-of-the-money S&P 500 index options. The data period ranges from November 3, 1999 to October 28, 2009 and thus includes several market shocks like the burst of the *dot-com bubble*, the *terror attacks 9/11* and the *subprime mortgage crisis*. Besides the analysis of the estimated parameters, we also examine the evolution of the filtered volatility and variance risk premium for both option pricing models. In addition, the 2-Factor SV model allows to infer the risk aversion of the market participants during this period. Finally, we perform Value-at-Risk forecasts for the S&P 500 index based on the estimation results of the Heston model and the 2-Factor SV model. More precisely, we perform Value-at-Risk forecasts during the subprime mortgage crisis, where we assume forecasting horizons of 1, 2, 3 and 4 weeks together with confidence levels of 95%, 99% and 99.9%. In order to better classify these results, we also perform Value-at-Risk forecasts based on two commonly used Value-at-Risk models, namely the Black-Scholes model, which also relies on option-implied information, and the GARCH model, which relies on historical information.

Before we discuss the results of our empirical study, Chapter 6.2 first gives an overview over the previous estimation results in the literature for the parameters of the Heston model. Furthermore, we review the literature with respect to the forecasting performance of various Value-at-Risk models. Then Chapter 6.3 analyzes the data set based on which our empirical study is performed. The estimation results for the Heston and 2-Factor SV model are presented in Chapter 6.4. Finally, the results of our Value-at-Risk study are discussed in Chapter 6.5.

6.2. Literature review

The following literature review consists of two parts. First, we give an overview over the previous estimation results in the literature for the parameters of the Heston model. Since the 2-Factor SV model is an extension of the Heston model, these results can also be used to better classify the estimation results for our new model. Subsequently, we

6. Empirical study

will summarize the previous findings in the literature about the forecasting performance of various Value-at-Risk models. Since most Value-at-Risk models assume a parametric profit or loss distribution, the topics volatility/density forecasting and Value-at-Risk forecasting are closely related. Hence, the literature review about the forecasting performance of various volatility and density forecasting methods, which we already gave in Chapter 2.5.3, should give an indication for promising and less promising Value-at-Risk models.

Table 6.1.: Previous estimation results for the parameters of the Heston model.

	Chernov and Ghysels [54]	Pan [130]	Jones [104]	Van der Ploeg [133]
Data	Daily S&P 500 returns + 1 ATM call option series	Weekly S&P 500 returns + 1 ATM call option series	Daily S&P 100 returns + VIX index	Daily FTSE 100 returns + 1 ATM call option series
Period	1986 – 1993	1989 – 1996	1988 – 2000	1997 – 2001
Estimation method	Efficient Method of Moments	Implied state GMM	Bayesian approach	Extended Kalman filter QML
ϕ^*	0.01	-0.20	-0.06	0.11
κ^*	0.69	-0.50	-2.98	3.10
λ	0.06	0.32	0.38	0.69
ρ	-0.02	-0.53	-0.68	-0.67
γ	-0.24	-7.60	-3.46	-4.76
$\oslash$ variance risk premium	-0.37%	-10.4%	-12.15%	-20.99%

Source: Van der Ploeg [134, p. 263]

Table 6.1 summarizes the previous estimation results in the literature for the parameters of the Heston model. First note that all listed empirical studies found negative variance risk premium estimates but with partly significant differences in the magnitude. Interestingly, Pan [130] and Jones [104] estimates for the mean-speed of reversion κ^* and the long-term average of the variance ϕ^* are negative. Pan [130] explains this phenomenon by the large difference between the statistical and risk-neutral volatility estimates in their sample, which can only be compensated by a large negative value for the variance risk premium parameter γ . Furthermore, the previous estimates in the literature for the parameters λ and ρ are plausible, even though the corresponding estimation results of Chernov and Ghysels [54] seem to be relatively small. In particular, the negative correlation estimates are in line with the leverage effect and a negatively skewed return distribution as commonly found in the literature. Further note that all previous empirical studies only used univariate option price time series, which makes it more difficult to estimate the risk-neutral asset price dynamics.

Next, we turn to the issue of Value-at-Risk forecasting, which is an important application for volatility and density forecasting methods. The Value-at-Risk is a commonly used

risk measure in the financial industry and quantifies the maximum loss, which is not exceeded for a given confidence level and forecasting horizon.¹ In general, the results of the forecasting performance for the different volatility and density forecasting methods presented in Chapter 2.5.3 are in line with the forecasting performance of the corresponding Value-at-Risk models. In particular, Value-at-Risk models based on option-implied information are found to be at least as good (see e.g. Giot [85]) or better (see e.g. Louzis et al. [123]) than Value-at-Risk models based on historical information. More precisely, Giot [85] finds that Value-at-Risk forecasts based on Black-Scholes implied volatilities and GARCH-type models do have similar forecasting performance, even during challenging market conditions.² Louzis et al. [123] examine the forecasting performance of various Value-at-Risk models based on well established volatility forecasting methods, as for example option-implied volatilities or realized volatilities. Moreover, they make use of the recent developments in volatility forecasting, like the realized GARCH model of Hensen et al. [91], which combines the GARCH model with realized volatility. Louzis et al. [123] find that no Value-at-Risk model based on historical data, neither daily nor intra-daily, can outperform Value-at-Risk models based on option-implied information, irrespective of the confidence level, forecasting horizon and evaluation metric.³ Furthermore, Watanabe [153] finds that the realized GARCH model of Hensen et al. [91] outperforms classical GARCH-type models for the purpose of Value-at-Risk forecasting.⁴ Interestingly, there is a disagreement in the literature whether Value-at-Risk models based on GARCH-type volatility or realized volatility estimators deliver the best forecasting performance.⁵ On the one hand, Giot and Laurent [86] find that Value-at-Risk models based on realized volatility do not outperform GARCH-type Value-at-Risk models. On the other hand, Brownless and Gallo [45] and Louzis et al. [123] find that Value-at-Risk models based on realized volatility improve the forecasting performance considerably relative to GARCH-type Value-at-Risk models.

¹For more details about the Value-at-Risk see Chapter 6.5.2.

²The Value-at-Risk study of Giot [85] is based on the S&P 100 index as well as the NASDAQ 100 index. Furthermore, Giot [85] assumes forecasting horizons of 5 and 10 days, confidence levels of 90%, 95% and 99% as well as a skewed student's t distribution for the return innovations.

³The Value-at-Risk study of Louzis et al. [123] is based on the S&P 500 index. Furthermore, Louzis et al. [123] assume forecasting horizons of 1 day, 1, 2, 3, and 4 weeks, confidence levels of 95% and 99% as well as a skewed student's t distribution for the return innovations.

⁴Watanabe [153] performs Value-at-Risk forecasts based on the S&P 500 index together with a forecasting horizon of one day and confidence levels of 90%, 95% and 99%. Furthermore, Watanabe [153] assumes three different distributions for the return innovations, namely the normal, student's t and the skewed student's t distribution.

⁵For a literature summary of Value-at-Risk studies based on realized volatility see Louzis et al. [123, Tab. 1].

6.3. Data

The following empirical study is based on Wednesday's weekly closing prices of the S&P 500 index and corresponding call options. The data are drawn from *OptionMetrics Ivy DB US* and were provided by the *Collaborative Research Center 649 – Economic Risk* at the *Humboldt-Universität zu Berlin*. We will analyze the time period from November 3, 1999 to October 28, 2009, which includes several market shocks like the burst of the *dot-com bubble* in 2000, the *terror attacks 9/11* in 2001 or the *subprime mortgage crisis* starting in 2008.

Figure 6.1.: Statistical properties of the S&P 500 index. These figures illustrate the evolution of the S&P 500 index (a) and the corresponding index returns (b). Moreover, figure (c) shows a kernel density estimation of the index returns (solid line), which is contrasted to the normal distribution (dotted line). The latter is specified by the sample mean and variance of the index returns. In addition, figure (d) shows the Quantile-Quantile-plot of the index returns.

First, Figure 6.1(a) gives an overview over the evolution of the S&P 500 index for the underlying time period, where the several market shocks are reflected by sudden falls in the index prices. Figure 6.1(b) illustrates the corresponding index returns, which exhibit low levels of volatility during the calm period 2004 – 2007 and high levels of volatility before and after this calm period. This feature is also called *volatility clustering*. Moreover, Figures 6.1(c) and 6.1(d) show a kernel density estimation and a Quantile-Quantile-plot of the index returns, which both indicate that the distribution of the index returns is

fat tailed and slightly negatively skewed. These observations are in line with the sample skewness and kurtosis of the S&P 500 index returns, which are summarized in Table 6.2.

Table 6.2.: Descriptive statistics for the S&P 500 index returns.

N	min	max	mean	SD	skewness	kurtosis
521	-16.4508	10.1824	-0.0503	2.6125	-0.4910	7.6359

Notes: This table reports the summary statistics of the weekly S&P 500 index returns. All figures are calculated based on percentage returns.

Further note that the S&P 500 index is a *price return index*, which does not account for dividend payments. Hence, for the purpose of option pricing, we also need to know the annualized dividend payments of the S&P 500 index, which are provided by *OptionMetrics Ivy DB US* and are illustrated in Figure C.1(a). Furthermore, *OptionMetrics Ivy DB US* also provides the “riskfree rate” for the US market, which is illustrated in Figure C.1(b).

In accordance with the option price characteristics assumed in the previous Monte-Carlo study (see Chapter 5.4.1), we choose at-the-money and out-of-the-money call options each with four different maturities. In contrast to the simulated option prices, the maturities of S&P 500 index options are not constant but time-varying. Hence, we will select call options within four different maturity classes τ_1 to τ_4 , which are illustrated in Figure 6.2. The apparent change of the maturity pattern since 2007 is due to a change in the offered maturities for S&P 500 index options. The average maturities of the maturity classes τ_1 to τ_4 are 22, 45, 89 and 148 days, respectively.

Figure 6.2.: Maturities of the S&P 500 index options. These figures illustrate the evolution of the maturity classes τ_1 (a), τ_2 (b), τ_3 (c) and τ_4 (d).

We select two different strike prices for each maturity class, which in total results in eight different call option prices time series. The strike prices are chosen according to the

6. Empirical study

Figure 6.3.: Moneyness of the S&P 500 index options. These figures illustrate the evolution of the moneyness of the different call option price time series. More precisely, figures (a) – (d) illustrate the moneyness of the at-the-money call options with moneyness $M_{1,t}$ to $M_{4,t}$ and figures (e) – (h) the moneyness of the out-of-the-money call options with moneyness $M_{5,t}$ to $M_{8,t}$, both with maturities $\tau_{1,t}$ to $\tau_{4,t}$.

following strike choice function:

$$K_{j,t} = \lfloor S_t/25 \rfloor \cdot 25 + j \cdot 25 \quad \text{for } j = 1, 2 ,$$

where S_t denotes the S&P 500 index. Unfortunately, option prices with the required strike prices are not always traded for all maturities, so that the corresponding closest available strike price is chosen instead. Figure 6.3 illustrates the resulting time series for the moneyness of the eight different call option price time series. Recall that the moneyness M_t of a call option with spot price S_t and strike price K_t is defined as $M_t = S_t/K_t$. The time series $M_{1,t}$ to $M_{4,t}$ represent the moneyness of the at-the-money call options ($j = 1$) and the time series $M_{5,t}$ to $M_{8,t}$ the moneyness of the out-of-the-money call options ($j = 2$), both with maturities $\tau_{1,t}$ to $\tau_{4,t}$, respectively. In particular, we can see that at some dates and for some maturities only far in-the-money or far out-of-the-money options are available.

Altogether, this particular maturity and strike price choice results in the call option price time series

$$\bar{C}_t = [\bar{C}_{1,t} \ \bar{C}_{2,t} \ \bar{C}_{3,t} \ \bar{C}_{4,t} \ \bar{C}_{5,t} \ \bar{C}_{6,t} \ \bar{C}_{7,t} \ \bar{C}_{8,t}]^\top ,$$

with maturities

$$\tau_t = [\tau_{1,t} \ \tau_{2,t} \ \tau_{3,t} \ \tau_{4,t} \ \tau_{1,t} \ \tau_{2,t} \ \tau_{3,t} \ \tau_{4,t}]^\top ,$$

and moneyness

$$M_t = [M_{1,t} \ M_{2,t} \ M_{3,t} \ M_{4,t} \ M_{5,t} \ M_{6,t} \ M_{7,t} \ M_{8,t}]^\top .$$

The call option price time series $\bar{C}_t$ is illustrated in Figure 6.4 in terms of Black-Scholes implied volatilities, which clearly reflect the different market conditions of this specific time period. Apart from some outliers, the Black-Scholes implied volatilities are very similar within the same maturity class. However, at-the-money call options tend to have higher Black-Scholes implied volatilities than out-of-the-money call options. This can also be seen from the positive difference between the Black-Scholes implied volatilities of at-the-money and out-of-the-money call options, which is illustrated in the additional small figures. This positive difference indicates a skew shaped volatility smile for all maturity classes.⁶ Further note that the differences between the Black-Scholes implied volatilities of at-the-money and out-of-the-money call options are larger for short maturities than for long maturities.

Figure 6.5 shows the differences between the Black-Scholes implied volatilities for at-the-money (left figures) and out-of-the-money (right figures) call options. Apparently, there are differences between the Black-Scholes implied volatilities for call options with different maturities, both for at-the-money and out-of-the-money call options. In particular, call options with long maturities tend to have higher Black-Scholes implied volatilities than call options with short maturities, which is often referred to as increasing volatility term structure. This positive relation between the magnitude of the Black-Scholes implied volatility and the maturity of the underlying option breaks down during turbulent market conditions as can be seen from 2000 to 2003 and from 2008 to 2009. In particular, in times

⁶This feature is illustrated in Figure 3.3(a) based on the Heston model.

6. Empirical study

Figure 6.4.: Black-Scholes implied volatilities of the S&P 500 index options. Figures (a) – (d) compare the Black-Scholes implied volatilities of at-the-money (blue line) and out-of-the-money (green line) call options for the maturity classes τ_1 to τ_4 . Moreover, the difference between the blue and the green line, which indicates the slope of the volatility smile, are illustrated by the red line.

Figure 6.5.: Volatility term structure of the S&P 500 index options. These figures illustrate the differences between the Black-Scholes implied volatilities for call options with different maturities, where $\sigma_{BS}(\bar{C}_{i,t})$ denotes the Black-Scholes implied volatility based on the call option $\bar{C}_{i,t}$. The left figures (a), (c) and (e) are calculated based on at-the-money call options and the right figures (b), (d) and (f) based on out-of-the-money call options.

of very high volatility, the volatility term structure tends to be decreasing.

In order to account for moneyness and maturity effects in the estimation, we scale the selected call option prices by their corresponding Black-Scholes vega. The resulting time series of the vega-scaled option prices are illustrated in Figure C.2. Moreover, Table 6.3 summarizes the mean and standard deviation of the regular and vega-scaled option prices. Note that the relative difference between the minimum and maximum standard deviation can be reduced from 270% for the regular option prices to 140% for the vega-scaled option prices.

As already mentioned earlier, the selected call option prices suffer from some outliers. More precisely, the difference of the Black-Scholes implied volatilities between the call options $\bar{C}_{1,431}$ and $\bar{C}_{5,431}$ on January 30, 2008 is 12.91% and the difference between the call options $\bar{C}_{3,467}$ and $\bar{C}_{7,467}$ on October 8, 2008 is 8.49%, which is extremely high compared to the remaining differences (see Figure 6.4). In both cases this large difference can be lead back to the relatively low Black-Scholes implied volatility of the out-of-the-money call

6. Empirical study

Table 6.3.: Regular vs. vega-scaled option prices.

	$\bar{C}_{1,t}$	$\bar{C}_{2,t}$	$\bar{C}_{3,t}$	$\bar{C}_{4,t}$	$\bar{C}_{5,t}$	$\bar{C}_{6,t}$	$\bar{C}_{7,t}$	$\bar{C}_{8,t}$
regular	15.579	28.123	41.294	55.937	7.373	17.34	29.004	43.161
SD	(9.598)	(12.135)	(16.504)	(19.246)	(7.251)	(10.714)	(14.697)	(18.534)
vega-scaled	0.145	0.168	0.183	0.192	0.081	0.112	0.132	0.149
SD	(0.086)	(0.081)	(0.080)	(0.067)	(0.064)	(0.066)	(0.065)	(0.061)

Notes: This table reports the sample mean and standard deviation of the regular and vega-scaled option prices for the selected S&P 500 index call options.

option. In order to prevent the UKF estimation from difficulties due to these outliers, we adjust the option prices $\bar{C}_{5,431}$ and $\bar{C}_{7,467}$. We therefore first calculate the maximal difference between the Black-Scholes implied volatilities of the time series $\bar{C}_{1,t}$ and $\bar{C}_{5,t}$ without $t = 431$, which is 3.08%, and of the time series $\bar{C}_{3,t}$ and $\bar{C}_{7,t}$ without $t = 467$, which is 2.92%. Then we choose $\bar{C}_{5,431}$ and $\bar{C}_{7,467}$ just large enough, so that these maximal differences of the Black-Scholes implied volatilities are not changed including the adjusted call option prices. The replacement of extreme values with less extreme values is called *winsorization* and a commonly used statistical technique.⁷

6.4. Estimation results

6.4.1. Heston model

Next, we will estimate the Heston model based on the data set described in the previous section. The parameters of the Heston model are estimated via the UKF based on the state-space representation

$$\begin{aligned}\sigma_i &= e^{-\kappa\Delta t}\sigma_{i-1} + \phi(1 - e^{-\kappa\Delta t}) + e^{-\kappa\Delta t}\lambda\sqrt{\sigma_{i-1}}\eta_i, \\ y_i &= v_{BS}^{-1}(\bar{C}_i) \odot \mathcal{C}_{Heston}(S_i, \sigma_i, K_i, \tau_i | \theta_2) + \epsilon_i,\end{aligned}$$

where $\eta_i \sim \mathcal{N}(0, \Delta t)$ for $i = 1, \dots, N$. Here $y_i = \bar{C}_i \odot v_{BS}^{-1}(\bar{C}_i)$ denotes the observed call option prices, which are scaled by their corresponding Black-Scholes vega. Moreover, $\mathcal{C}_{Heston}$ denotes the call option price formula for the Heston model with risk-neutral parameters $\theta_2 = [\phi^* \ \kappa^* \ \lambda \ \rho]$, where $\kappa^* = \kappa + \gamma$ and $\phi^* = \kappa\phi/(\kappa + \gamma)$. The corresponding option prices are calculated based on the characteristic function (3.24) together with the Carr-Madan formula (3.10), where we applied the adaptive Gauss-Kronrod rule as numerical integration technique. In order to account for heteroscedasticity in the measurement error, we model the conditional variance of ϵ_i by an univariate *autoregressive conditional*

⁷For more details about this technique see e.g. Barnett and Lewis [18].

heteroskedasticity (ARCH) model, so that

$$\begin{aligned}\epsilon_i &= \sigma_i z_i, \\ \sigma_i^2 &= \omega + \alpha \mu(\epsilon_{i-1}^2),\end{aligned}\tag{6.1}$$

where $\omega > 0$, $\alpha \geq 0$ and $z_i \sim \mathcal{N}(0, I_8)$ for $i = 1, \dots, N$. Furthermore, $\mu(\epsilon_{i-1}^2)$ denotes the sample mean of the squared measurement errors at time $i - 1$. Hence, the state-space representation of the Heston model can fully be parameterized by $\theta = [\phi^* \ \kappa^* \ \lambda \ \rho \ \gamma \ \omega \ \alpha]$.

Table 6.4.: Estimation results for the parameters of the Heston model.

	ϕ^*	κ^*	λ	ρ	γ	ω	α
Coef.	0.124	0.367	0.272	-0.805	-2.742	$3.276 \cdot 10^{-5}$	0.215
<i>t</i> -statistic	(10.28)	(8.65)	(33.39)	(-28.85)	(-9.88)	(22.96)	(20.61)

Notes: The reported *t*-statistics are calculated as $t = (\hat{\theta} - \theta_0)/SE(\hat{\theta})$, where we set the null hypothesis to $\theta_0 = 0$. Hence, large *t*-statistics indicate that the corresponding parameter is significantly different from zero. In particular, $|t| > 1.960$ corresponds to a 5% significance level, $|t| > 2.576$ to a 1% significance level and $|t| > 3.291$ to a 0.1% significance level.

The estimation results for the parameters of the Heston model are summarized in Table 6.4. First note that all parameters are significantly different from zero at a significance level of 0.1%. Hence, in particular, the variance risk premium parameter γ is significantly negative, which results in an average variance risk premium of -4.01% .⁸ Accordingly, our estimate for the average volatility risk-premium is between the findings of Chernov and Ghysels [54] (-0.37%) and Pan [130] (-10.4%). The estimate for the long-term variance under the risk-neutral measure $\mathbb{Q}$ corresponds to an average volatility level of 35.21% , which together with the estimate for γ results in an average volatility level of 12.10% under the statistical measure $\mathbb{P}$. Therefore, the average volatility under $\mathbb{Q}$ is three times large than under $\mathbb{P}$. Next, our estimate for the risk-neutral mean-speed of reversion is relatively low with $\kappa^* = 0.367$, which is very similar to the findings of Chernov and Ghysels [54]. The corresponding estimate under the statistical measure is given by $\kappa = 3.109$, which is higher than the risk-neutral counterpart due to the negative variance risk premium. Our estimation result for the correlation parameter is with $\rho = -0.805$ relatively low compared to the previous findings in the literature. This can mainly be explained by the strong skew shape of the Black-Scholes implied volatilities implied by our option price data. Finally, our estimate for the volatility of the variance process is with $\lambda = 0.272$ very similar to the findings of Pan [130] and Jones [104] but relatively low if we consider that the sample period includes several market shocks.

As a byproduct of the UKF, we are also able to compute the evolution of the volatility under the statistical measure, which is illustrated in Figure 6.6(a). Furthermore, Figure 6.6(b) illustrates the corresponding evolution of the variance risk premium. In addition, both figures also show the evolution of the S&P 500 index during that time. Similarly as

⁸The average variance risk premium is calculated based on Heston's [95] specification of the variance risk-premium, which is $\gamma\sigma_t$. If we replace the current variance σ_t by its log-term average $\phi = \phi^*\kappa^*/(\kappa^* - \gamma)$, the average variance risk premium can be calculated as $\gamma\phi$.

6. Empirical study

Figure 6.6.: Volatility and variance risk premium estimates for the S&P 500 index based on the Heston model. These figures illustrate the evolution of the volatility (a) and the variance risk premium (b), which are estimated via the UKF based on the Heston model. For comparison reasons, we also illustrate the evolution of the S&P 500 index during that time. Recall that the volatility is defined as $\sqrt{\sigma_t}$ and the variance risk premium as $\gamma\sigma_t$.

for the Black-Scholes implied volatilities, the filtered volatility is low during normal market conditions and high during turbulent market conditions. Note that the volatility already starts to increase in mid-2007 and thus over one year before the peak of the subprime mortgage crisis. Hence, the filtered volatility is an indicator for current and future market conditions. Further note that the volatility and the S&P 500 index tend to move into opposite directions, which reflects the leverage effect. Moreover, the variance risk premium is always negative with values of less than -80% during the subprime mortgage crisis, which is due to the negative estimation result for the variance risk premium parameter γ .

Next, Figure 6.7 compares the empirical volatility term structure implied by the observed option prices (red line) with the volatility term structure implied by the estimation result of the Heston model (blue line).⁹ As already mentioned earlier, the empirical volatility term structure tends to be increasing during normal market conditions and decreasing during turbulent market conditions, as e.g. shortly after the bankruptcy of Lehman Brothers in September 2008. Due to the limited flexibility of the Heston model to capture the time series dynamics of the volatility term structure, the deviations between the empirical volatility term structure and the volatility term structure implied by the estimation results of the Heston model are partly large, especially, during the turbulent periods 2001 – 2003 and 2008 – 2009. Therefore it is not surprising that the Heston model has some difficulties to explain the observed option prices during that turbulent time. This can also be seen from the single measurement errors, which are illustrated in Figure C.3. For completeness, Table C.4 summarizes the mean and standard deviation of the single measurement errors.

In order to check the parameter stability of the Heston model, we repeatedly estimate

⁹Recall that the volatility term structure of the Heston model is increasing if $\sigma_t < \phi^*$ and decreasing if $\sigma_t > \phi^*$.

Figure 6.7.: Volatility term structure of S&P 500 index options and corresponding Heston model estimates. These figures compare the volatility term structure of at-the-money call options (a) and out-of-the-money call options (b) with the corresponding volatility term structure implied by the estimation result of the Heston model. The volatility term structure is determined by the difference between the Black-Scholes implied volatilities of long dated options (maturity class τ_4) and short dated options (maturity class τ_1). A positive difference indicates an increasing volatility term structure and a negative difference a decreasing volatility term structure.

the Heston model based on an increasing data set. We therefore increase the sample size after each estimation by four additional observation days, which approximately cover one month. While the underlying sample period always starts on November 3, 1999, the end of the sample period ranges from December 28, 2005 to September 30, 2009. The corresponding estimation results for the Heston parameters are illustrated in Figure 6.8.¹⁰ In addition, Figure 6.8 also shows the evolution of the S&P 500 index, which is indicated by the green line. First note that all Heston parameters are pretty unstable during the analyzed time period. The only exception is the long-term average ϕ^* , which is stable until the beginning of 2008 and then strongly increasing. Further note that the risk-neutral mean-speed of reversion κ^* decreases continuously. The volatility of the variance process λ increases during the subprime mortgage crisis, which is in line with the high uncertainty during that time. However, the evolution of the correlation parameter ρ is implausible since the correlation gets less negative during the subprime mortgage crisis, which implies that negative returns are less likely during that turbulent time. Interestingly, the variance risk premium parameter γ , which can be interpreted as a proxy variable for risk aversion, is stable until the beginning of 2007. Subsequently, γ increases until the bankruptcy of Lehman Brothers in September 2008, whereupon γ at first strongly decreases and then recovers at a slightly higher level. Due to the connection between the variance risk premium and risk aversion, the evolution of γ can be interpreted in a way that the market participants at first get less risk averse until the bankruptcy of Lehman Brothers, whereupon the risk aversion strongly increases and then reverts back to a slightly lower

¹⁰The complete estimation results are also summarized in Table C.5 and the corresponding t -statistics in Table C.6.

6. Empirical study

Figure 6.8.: Evolution of the Heston model parameters before and during the subprime mortgage crisis. These figures illustrate the evolution of the Heston model parameters (blue line) before and during the subprime mortgage crisis. We therefore repeatedly estimate the Heston model based on the data set described in Chapter 6.3. More precisely, we increase the sample size after each estimation by four additional observation days, which approximately cover one month. While the underlying sample period always starts on November 3, 1999, the end of the sample period, which is labeled on the x-axis, ranges from December 28, 2005 to September 30, 2009. Additionally, we contrast the estimation results of the Heston model parameters to the evolution of the S&P 500 index, which is indicated by the green line.

level. Finally, the evolution of ϕ and κ , which are calculated based on the parameters ϕ^* , κ^* and γ , mainly reflect the evolution of γ .¹¹ But the decrease/increase of ϕ/κ after the bankruptcy of Lehman Brothers is not intuitive since the volatility is rather increasing than decreasing during that time.

6.4.2. 2-Factor SV model

Now, we perform the same analysis as in the previous section for the 2-Factor SV model. Hence, the parameters of the 2-Factor SV model are estimated via the UKF based on the state-space representation

$$\begin{aligned} \begin{bmatrix} \sigma_i \\ \theta_i \end{bmatrix} &= \begin{bmatrix} e^{-\kappa_1 \Delta t} \\ e^{-\kappa_2 \Delta t} \end{bmatrix} \odot \begin{bmatrix} \sigma_{i-1} \\ \theta_{i-1} \end{bmatrix} + \begin{bmatrix} \phi_1 (1 - e^{-\kappa_1 \Delta t}) \\ \phi_2 (1 - e^{-\kappa_2 \Delta t}) \end{bmatrix} + \begin{bmatrix} e^{-\kappa_1 \Delta t} \lambda_1 \\ e^{-\kappa_2 \Delta t} \lambda_2 \end{bmatrix} \odot \sqrt{\begin{bmatrix} \sigma_{i-1} \\ \theta_{i-1} \end{bmatrix}} \odot \eta_i, \\ y_i &= v_{BS}^{-1}(\bar{C}_i) \odot \mathcal{C}_{2FSV}(S_i, x_i, K_i, \tau_i | \theta_2) + \epsilon_i, \end{aligned} \quad (6.2)$$

where $\eta_i \sim \mathcal{N}(0, \Delta t I_2)$ and $x_i = [\sigma_i \ \theta_i]^\top$ for $i = 1, \dots, N$. Furthermore, $\mathcal{C}_{2FSV}$ denotes the call option price formula for the 2-Factor SV model, which is parameterized by $\theta_2 = [\kappa_1^* \ \kappa_2 \ \phi_2 \ \lambda_1 \ \lambda_2 \ \rho]$. The corresponding option prices are calculated based on the characteristic function (3.43) together with the Carr-Madan formula (3.10), where we applied the caching technique of Chapter 5.3 together with the trapezoid rule. In order to account for heteroscedasticity in the measurement error, we model ϵ_i again by the univariate ARCH model specified in (6.1). Hence, the state-space representation of the 2-Factor SV model can fully be parameterized by $\theta = [\kappa_1 \ \kappa_1^* \ \kappa_2 \ \phi_1 \ \phi_2 \ \lambda_1 \ \lambda_2 \ \rho \ \omega \ \alpha]$.

Table 6.5.: Estimation results for the parameters of the 2-Factor SV model.

	κ_1	κ_1^*	κ_2	ϕ_1	ϕ_2	λ_1	λ_2	ρ	ω	α
Coef.	22.38	7.98	1.33	0.025	0.036	0.945	0.234	-0.664	$0.602 \cdot 10^{-5}$	0.017
t -statistic	(95.89)	(122.02)	(33.37)	(28.75)	(43.62)	(67.21)	(29.82)	(-114.59)	(41.99)	(15.73)

Notes: The reported t -statistics are calculated as $t = (\hat{\theta} - \theta_0)/SE(\hat{\theta})$, where we set the null hypothesis to $\theta_0 = 0$. Hence, large t -statistics indicate that the corresponding parameter is significantly different from zero. In particular, $|t| > 1.960$ corresponds to a 5% significance level, $|t| > 2.576$ to a 1% significance level and $|t| > 3.291$ to a 0.1% significance level.

Before we analyze the estimation results for the 2-Factor SV model, which are summarized in Table 6.5, recall that the Heston model and the 2-Factor SV model do have the same asset price dynamics under the statistical measure. First note that all parameters are significantly different from zero at a significance level of 0.1%. Moreover, the estimates for the mean-speed of reversion as well as for the volatility of the variance process are extremely high with values of $\kappa_1 = 22.38$ and $\lambda_1 = 0.945$ compared to the corresponding estimation results for the Heston model, where $\kappa = 3.109$ and $\lambda = 0.272$. In addition, also the long-term average of the variance is with $\phi_1 = 0.025$ higher than the corresponding estimate for the Heston model, where $\phi = 0.015$. These large differences between the two

¹¹Recall that ϕ and κ are defined as $\phi = \phi^* \kappa^* / (\kappa^* - \gamma)$ and $\kappa = \kappa^* - \gamma$, respectively.

6. Empirical study

models can be explained by the higher flexibility of the 2-Factor SV model under the risk-neutral measure, where the central tendency of the variance is modeled via the process θ_t . Furthermore, the parameters κ_1 and κ_1^* of the 2-Factor SV model can be estimated without any restriction, whereas the Heston parameters κ and κ^* as well as ϕ and ϕ^* are linked via the variance risk premium parameter γ . Further note that high values of κ_1 and λ_1 do have contrary effects. While high values of κ_1 reduce the variability of the variance process, high values of λ_1 increase the variability of σ_t . Next, our estimation result for the correlation parameter is with $\rho = -0.664$ slightly less negative than our estimate for the Heston model but very similar to the corresponding results of Jones [104] and Van der Ploeg [133]. In particular, a negative estimate of ρ is in line with the leverage effect. The estimation results for the mean-speed of reversion as well as for the volatility of the risk aversion process θ_t are relatively low with values of $\kappa_2 = 1.33$ and $\lambda_2 = 0.234$, respectively. In addition, the long-term average of the risk aversion is specified by $\phi_2 = 0.036$ and thus higher than the long-term average of the variance under the statistical measure. Note that ϕ_2 also determines the long-term average of the variance under the risk-neutral measure. Hence, similarly as for the Heston model, the long-term average of the variance under $\mathbb{Q}$ is higher than under $\mathbb{P}$. This results in a negative average variance risk premium in the amount of -1.46%, which is less negative than the corresponding estimate for the Heston model.¹²

Figures 6.9(a) – 6.9(b) illustrate the evolution of the filtered volatility, risk aversion and variance risk premium, which are contrasted to the evolution of the S&P 500 index during that time. In addition, Figure 6.9(d) compares the evolution of the filtered variance and risk aversion. Again, the evolution of the filtered volatility clearly reflects the different market conditions of the analyzed time period with low levels of volatility during calm periods and high levels of volatility during turbulent periods. Moreover, the S&P 500 index and the filtered volatility tend to move into opposite directions, which can be explained by the leverage effect. During turbulent market conditions, the volatility estimates based on the 2-Factor SV model are considerably higher than the corresponding estimates based on the Heston model. For example, at the peak of the subprime mortgage crisis, the maximum volatility estimate of the 2-Factor SV model is approximately 70% compared to 55% of the Heston model. Similarly as for the variance, also the risk aversion process θ_t is low during calm periods and high during turbulent periods. In particular, θ_t already increases over one year before the peak of the subprime mortgage crisis and thus can be interpreted as a proxy variable for “investor fear gauge”.¹³ The advantage of the risk aversion process as an indicator for the market sentiment is its lower variability compared to the volatility. This makes it easier to identify changing market conditions, which can be seen in Figure 6.9(d). Whereas the evolution of the variance risk premium for the 2-Factor SV model looks very similar to the one of the Heston model, the variance risk premium estimates based on the 2-Factor SV model are partly much lower than the corresponding estimates based on the Heston model. Interestingly, the variance risk premium estimates based on the 2-Factor SV model are positive from mid-2004 to the beginning of 2007 with

¹²The variance risk premium of the 2-Factor SV model is defined as $\kappa_1(\phi_1 - \sigma_t) - \kappa_1^*(\theta_t - \sigma_t)$. Hence, the average variance risk premium can be calculated if we replace σ_t and θ_t by their long-term averages ϕ_1 and ϕ_2 , which results in $-\kappa_1^*(\phi_2 - \phi_1)$.

¹³This naming was introduced by Whaley [154, 155] to describe the properties of the VIX index.

Figure 6.9.: Volatility, risk aversion and variance risk premium estimates for the S&P 500 index based on the 2-Factor SV model. These figures illustrate the evolution of the volatility (a) and the variance risk premium (b), which are estimated via the UKF based on the 2-Factor SV model. In addition, figure (c) illustrates the evolution of the risk aversion, which is contrasted to the evolution of the variance in (d). Recall that the volatility is defined as $\sqrt{\sigma_t}$ and the variance risk premium as $\kappa_1(\phi_1 - \sigma_t) - \kappa_1^*(\theta_t - \sigma_t)$.

a maximum variance risk premium of 32.11% on January 24, 2007. This indicates that the market participants were getting more and more risk seeking during that time. In particular, the positive estimates for the variance risk premium can be interpreted as a speculative bubble indicator.

Figure 6.10 compares the volatility term structure of the S&P 500 index options (red line) with the corresponding 2-Factor SV model estimates (blue line) for at-the-money and out-of-the-money options, respectively. As before, the volatility term structure is determined by the difference between the Black-Scholes implied volatilities of long dated options within maturity class τ_4 and short dated options within maturity class τ_1 . In contrast to the Heston model, the 2-Factor SV model is well able to capture the time series behavior of the empirical volatility term structure, especially, during turbulent market conditions, where the volatility term structure is sometimes strongly decreasing. The superior performance of the 2-Factor SV model in this context can be explained by the risk aversion process θ_t , which allows for a more flexible modeling of the volatility

6. Empirical study

Figure 6.10.: Volatility term structure of the S&P 500 index options and corresponding 2-Factor SV model estimates. These figures compare the volatility term structure of at-the-money call options (a) and out-of-the-money call options (b) with the corresponding volatility term structure implied by the estimation result of the 2-Factor SV model. The volatility term structure is determined by the difference between the Black-Scholes implied volatilities of long dated options (maturity class τ_4) and short dated options (maturity class τ_1). In particular, a positive difference indicates an increasing volatility term structure and a negative difference a decreasing volatility term structure.

term structure. In particular, a decreasing volatility term structure within the framework of the 2-Factor SV model can be captured if $\sigma_t > \theta_t$ whereas the Heston model requires that $\sigma_t > \phi^*$. Thus, high values of ϕ^* , as e.g. found in the empirical study of the previous section, make it more difficult for the Heston model to capture a decreasing volatility term structure. Nevertheless, also the 2-Factor SV model has some difficulties to explain the call option prices during turbulent market conditions, which can be seen from the single measurement errors illustrated in Figure C.4. For completeness, Table C.7 summarizes the mean and standard deviation of the single measurement errors.

Similarly as for the Heston model, we also check the parameter stability of the 2-Factor SV model by repeatedly estimating the model parameters based on a continuously increasing data set. The corresponding estimation results are illustrated in Figure 6.11, which also shows the evolution of the S&P 500 index during that time. Even though no parameter is completely stable over the whole period, at least some parameters are pretty stable with only small changes, as for example, the long-term averages of the variance and risk aversion ϕ_1 and ϕ_2 respectively. Interestingly, the long-term average of the risk aversion ϕ_2 slightly decreases until the bankruptcy of Lehman Brothers in September 2008 and then suddenly increases, which can be interpreted in a way that the average risk aversion level of market participants increases during the subprime mortgage crisis. Also the volatility of the risk aversion λ_2 is pretty stable and only slightly increases during the analyzed time period. In contrast, the volatility of the variance λ_1 is only stable until the bankruptcy of Lehman Brothers and then increases considerably, which reflects the increased variability of the variance during that time. Similarly, also the risk-neutral mean-speed of reversion of the variance κ_1^* as well as the correlation ρ are pretty stable until mid-2008 and then

Figure 6.11.: Evolution of the 2-Factor SV model parameters before and during the subprime mortgage crisis. These figures illustrate the evolution of the 2-Factor SV model parameters (blue line) before and during the subprime mortgage crisis. We therefore repeatedly estimate the 2-Factor SV model based on the data set described in Chapter 6.3. More precisely, we increase the sample size after each estimation by four additional observation days, which approximately cover one month. While the underlying sample period always starts on November 3, 1999, the end of the sample period, which is labeled on the x-axis, ranges from December 28, 2005 to September 30, 2009. Additionally, we contrast the estimation results of the 2-Factor SV model parameters to the evolution of the S&P 500 index, which is indicated by the green line.

increase and decrease, respectively. The decreasing correlation results in a higher negative skewness of the return distribution, so that negative returns are more likely during these turbulent market conditions. In contrast, the increase of κ_1^* at the peak of the subprime mortgage crisis is not intuitive since this increase reduces the variability of the variance process under the risk-neutral measure. A possible explanation for this behavior could be that the increase in κ_1^* partly offsets the strong increase in λ_1 . Finally, the mean-speed of reversion of the variance κ_1 as well as of the risk aversion κ_2 are unstable. This is due to the general difficulty to estimate the mean-speed of reversion under the statistical measure as found in the Monte-Carlo study of Chapter 5. Nevertheless, both κ_1 and κ_2 decrease at the peak of the subprime mortgage crisis, which reflects the higher variability of the variance and risk aversion during turbulent market conditions.

6.5. Forecasting performance

6.5.1. Setup

In the following section, we will assess the out-of-sample forecasting performance of the Heston model and the 2-Factor SV model within the framework of an empirical study. We therefore will perform Value-at-Risk (VaR) forecasts for the S&P 500 index during the subprime mortgage crisis for forecasting horizons of 1, 2, 3 and 4 weeks together with confidence levels of 95%, 99% and 99.9%. In order to better classify these results, we will also perform VaR forecasts based on two commonly used VaR models, namely the Black-Scholes model, which also relies on option-implied information, and the GARCH model, which relies on historical information.

Our forecasting period starts on January 4, 2006 and ends on October 28, 2009, which encompasses 200 weekly observations of the S&P 500 index as well as corresponding call options. In particular, this period includes the subprime mortgage crisis, which is a challenging forecasting period. The total forecasting period is further subdivided into a *precrash period*, which starts on January 4, 2006 and ends on December 5, 2007, and a *crash period*, which starts on December 12, 2007 and ends on October 28, 2009. This subdivision has the objective to check if the forecasting performance of the different VaR models deteriorate during the subprime mortgage crisis, and if so, to which extent. The subdivision in precrash and crash period was simply achieved by dividing the total period into two equally sized subperiods each with 100 weekly observations. Of course, this subdivision is somehow arbitrary. But, on the one hand, there is no clear agreement in the literature when the subprime mortgage crisis exactly started. On the other hand, our crash period includes all major events of the subprime mortgage crisis, which begins with the bailout of Bear Stearns by the *Federal Reserve System* (Fed) together with JPMorgan Chase on March 14, 2008 followed by the bailout of Fannie Mae and Freddie Mac by the Fed on July 3, 2008 and then reaching its climax with the bankruptcy of Lehman Brothers on September 15, 2008. Furthermore, all VaR calculations are only based on the information available at the respective forecasting date, which is also referred to as

out-of-sample forecasting. In order to account for parameter instability, we reestimate the GARCH, Heston and 2-Factor SV model every four weeks anew based on an enlarged data set, which also includes the four additional observation days while keeping the beginning of the data set always fixed on November 3, 1999. In contrast, the Black-Scholes model only requires the average Black-Scholes implied volatilities at the respective forecasting date, so that no reestimation based on an enlarged data set is necessary.

Before we describe the results of the empirical study in more detail, we will first introduce the risk-measure Value-at-Risk. Subsequently, we will describe the underlying Value-at-Risk forecasting models.

6.5.2. Value-at-Risk

In the late 1970s and 1980s some large financial institutions began to use internal risk models. Among those financial institutions was also the investment bank J.P. Morgan, which developed the risk measure *Value-at-Risk*. [68, p. 7 f.]. Meanwhile, this risk measure is a standard in risk management. Also the regulatory capital requirements for financial institutions, which are imposed by the *Basel Committee on Banking Supervision*, are based on the risk-measure VaR [19, 20].

In order to define the VaR, we consider a long position in the asset S_t . Hence, the future profit or loss of S_t within the period $[t, t+\tau]$ is given by

$$X_{t,t+\tau} = S_{t+\tau} - S_t .$$

Following Tasche [147] the VaR of the position S_t with a confidence level of α and a holding period of τ years is defined as the maximum loss, which is not exceeded with a probability of (at least) α [147, p. 2] More formally, the VaR can also be defined as

$$VaR_\alpha(X_{t,t+\tau}) = q_\alpha(-X_{t,t+\tau}) , \quad (6.3)$$

where

$$q_\alpha(X) = \inf \{x \in \mathbb{R} : \mathbb{P}[X \leq x] \geq \alpha\} ,$$

denotes the α -quantile of the random variable X . The VaR definition (6.3) is most general and also accounts for discontinuous profit and loss distributions. Note that if the profit and loss distribution is continuous, the VaR can also be calculated as

$$VaR_\alpha(X_{t,t+\tau}) = -q_{1-\alpha}(X_{t,t+\tau}) ,$$

which is illustrated in Figure 6.12.

Instead of calculating the VaR based on a profit or loss distribution, one can alternatively calculate the VaR based on the corresponding return distribution. The VaR can then be interpreted as the maximum percentage loss, which is not exceeded with a probability of (at least) α . The continuously compounded return of the position S_t within the period

Figure 6.12.: The Value-at-risk for a continuous profit and loss distribution.

$[t, t + \tau]$ is defined as

$$r_{t,t+\tau} = \ln(S_{t+\tau}/S_t) ,$$

so that the VaR with a confidence level of α is given by

$$VaR_{\alpha}(r_{t,t+\tau}) = -q_{1-\alpha}(r_{t,t+\tau}) , \quad (6.4)$$

where we assumed a continuous return distribution. If the VaR (6.4) is exceeded, i.e.

$$\bar{r}_{t,t+\tau} < q_{1-\alpha}(r_{t,t+\tau}) ,$$

where $\bar{r}_{t,t+\tau}$ denotes the realization of the random variable $r_{t,t+\tau}$, we call this a *VaR exception*. If the VaR model is correctly specified, the total number of VaR exceptions x should amount to $(1 - \alpha) \cdot 100\%$ of all VaR forecasts N .¹⁴

In order to assess the accuracy of our VaR models, we will apply the test of Kupiec [118], which checks if the proportion of VaR exceptions is consistent with the chosen confidence level. Hence, the corresponding null hypothesis is given by

$$H_0 : \alpha = \hat{\alpha} ,$$

where $\hat{\alpha} = 1 - x/N$. According to Kupiec [118], this null hypothesis can be tested via the following likelihood ratio test:

$$LR = -2 \ln [\alpha^{N-x}(1 - \alpha)^x] + 2 \ln [\hat{\alpha}^{N-x}(1 - \hat{\alpha})^x] . \quad (6.5)$$

The test statistic (6.5) is asymptotically χ^2 distributed with one degree of freedom.

¹⁴Note that neither a too loose nor a too conservative VaR model is desirable. On the one hand, too small VaR forecasts could imply that a bank does not hold enough capital to cover possible future losses, which can lead to financial distress or bankruptcy. On the other hand, too high VaR forecasts could imply that banks hold unnecessarily much capital, which in turn is unprofitable [43, p. 9].

6.5.3. Forecasting models

Before we introduce our different forecasting models, we make the simplifying assumption that the expected return $r_{t,t+\tau}$ is always zero, i.e.

$$\mathbb{E}[r_{t,t+\tau}] = 0 , \quad (6.6)$$

which is a common assumption in the literature (see e.g. Brownlees and Gallo [45]). According to Danielsson and Zigrand [64], this assumption can be justified by two arguments. On the one hand, the forecasting horizon of risk models is usually very short, so that the drift term is negligible compared to the volatility during that time. On the other hand, so far there exists no obvious way to obtain an accurate estimate of the drift term. In particular, Merton [126, p. 333] argues that a negative estimate of the drift term is a biased-low estimate, which would be very likely for our specific data period. Based on this simplifying assumption, we now introduce our different VaR models.

Black-Scholes model

Our first forecasting model is based on the Black-Scholes model, which assumes that the asset price evolves as

$$dS_t = \mu S_t dt + \sigma S_t dW_t . \quad (6.7)$$

Following Shreve [144, p. 147 f.], the SDE (6.7) has the closed-form solution

$$S_{t+\tau} = S_t \exp\{(\mu - 1/2 \sigma^2)\tau + \sigma (W_{t+\tau} - W_t)\} ,$$

so that the continuously compounded return for the period $[t, t+\tau]$ is given by

$$r_{t,t+\tau} = (\mu - 1/2 \sigma^2)\tau + \sigma (W_{t+\tau} - W_t) .$$

Due to assumption (6.6), we set the drift to $\mu = 1/2 \sigma^2$, so that $r_{t,t+\tau}$ is distributed as

$$r_{t,t+\tau} | \mathcal{F}_t \sim \mathcal{N}(0, \tau \sigma^2) . \quad (6.8)$$

Note that the annual variance σ^2 is scaled by the forecasting horizon τ or, equivalently, the annual volatility σ is scaled by the square-root of the forecasting horizon $\sqrt{\tau}$. The latter is also called *square-root-of-time rule* and is often applied in the practice of volatility forecasting.¹⁵ In contrast to the constant volatility assumption of the Black-Scholes model, we repeatedly estimate the volatility parameter σ based on the currently observable call option prices. More precisely, we use the average of the Black-Scholes implied volatilities

¹⁵One example for an application of the square-root-of-time rule is the volatility forecasting model of Riskmetrics, which is commonly applied in the banking industry. Moreover, also the Basel Committee on Banking Supervision makes use of this rule, even under Basel 3. It's worth mentioning that the square-root-of-time rule can lead to a systematic underestimation of future volatility and hence also risk if the returns are not normally distributed [64].

6. Empirical study

of all call options for the volatility estimate at time t , so that σ in (6.8) is replaced by

$$\bar{\sigma}_{BS,t} = \frac{1}{8} \sum_{i=1}^8 \sigma_{BS}(\bar{C}_{i,t}) ,$$

which completes the specification of our Black-Scholes forecasting model.

GARCH model

Before we discuss our GARCH forecasting model, we shortly introduce the GARCH model of Bollerslev [33].¹⁶ The GARCH model, which is also referred to as GARCH(p,q) model, describes the evolution of the conditional variance of some residual ϵ_t as follows:

$$\sigma_t^2 = \omega + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^p \beta_i \sigma_{t-i}^2 .$$

The constraints $\omega > 0$, $\alpha_i \geq 0$ and $\beta_i \geq 0$ ensure that the variance is always strictly positive. The residual ϵ_t is defined as

$$\epsilon_t = \sigma_t z_t , \quad (6.9)$$

where z_t is a sequence of independent and identically distributed random variables with zero mean and unit variance. Hence, the conditional mean and variance of the residual (6.9) are given by

$$\mathbb{E}[\epsilon_t | \mathcal{F}_{t-1}] = 0 \quad \text{and} \quad \text{Var}[\epsilon_t | \mathcal{F}_{t-1}] = \sigma_t^2 ,$$

where $\mathcal{F}_t$ denotes the sigma-algebra generated by ϵ_t .

Based on the zero drift assumption (6.6), we can now model the evolution of the return as

$$r_t = \sigma_t z_t , \quad (6.10)$$

where z_t is now a sequence of independent standard normally distributed random variables. For the evolution of the conditional variance, we assume a GARCH(1,1) model, which is

$$\sigma_t^2 = \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2 . \quad (6.11)$$

We made the particular choice of the GARCH(1,1) model not only because of its simplicity but also due to its proven superior forecasting performance compared to more complex specifications, as e.g. shown by Hansen and Lunde [124].¹⁷ Now, if we iterate (6.11) one time step forward, we obtain the variance forecast σ_{t+1}^2 for the period $[t, t+1]$. Since our VaR forecasting horizon will not always correspond to the time spacing of the data sample, we also apply the square-root-of-time rule in order to get a variance forecast for

¹⁶For a detailed treatment of GARCH models, we refer to Bollerslev et al. [35].

¹⁷Further note that the GARCH(1,1) model with the particular choice of $\alpha = 0.04$ and $\beta = 0.96$ for one day ahead forecasts and $\alpha = 0.03$ and $\beta = 0.97$ for one month ahead forecasts is essentially the volatility forecasting model of Riskmetrics, which is commonly used in the banking industry.

the period $[t, t + \tau]$.¹⁸ Therefore, the return for the period $[t, t + \tau]$ is approximately distributed as

$$r_{t,t+\tau} \approx \mathcal{N}\left(0, \frac{\tau}{\Delta t} \sigma_{t+1}^2\right),$$

where we assumed that the time spacing of the data sample is equidistant with interval length Δt .¹⁹

Heston and 2-Factor SV model

Finally, we will use the Heston model and the 2-Factor SV model for the purpose of VaR forecasting. Recall that both models assume the same statistical asset price dynamics

$$\begin{aligned} dS_t &= \mu S_t dt + \sqrt{\sigma_t} S_t dW_{1,t}, \\ d\sigma_t &= \kappa(\phi - \sigma_t)dt + \lambda\sqrt{\sigma_t} dW_{2,t}, \end{aligned} \tag{6.12}$$

where $dW_{1,t} dW_{2,t} = \rho dt$. However, due to the additional risk aversion process of the 2-Factor SV model, the estimation results for the parameters in (6.12) are generally different for the two models. In contrast to the Black-Scholes model, the SDE (6.12) has no closed-form solution, so that the return distribution cannot be stated in closed-form. However, the characteristic function of the log asset price is known in closed-form. Hence, the return distribution can be obtained via Fourier inversion, which is described in Appendix B.3 in more details. Unfortunately, since the SDE (6.12) has no closed-form solution, we can only approximately impose the zero drift condition (6.6) by setting $\mu = 1/2 \sigma_t$.

6.5.4. Results

In the following, we discuss the forecasting performance of the different VaR models. More precisely, we check if the number of VaR exceptions is in line with the chosen confidence level by means of the test of Kupiec [118]. The number of VaR exceptions for all confidence levels and forecasting horizons as well as the corresponding p -values are summarized in Tables 6.6 – 6.8 for the precrash, crash and total period, respectively.²⁰ Similar as in Chapter 2.5.4, we call a VaR model *slightly accurate* indicated by one asterisk if $0.01 \leq p < 0.05$, *accurate* indicated by two asterisk if $0.05 \leq p < 0.1$ and *highly accurate* indicated by three asterisk if $p \geq 0.1$. Furthermore, VaR models with p -values smaller than 0.01 will be referred to as *inaccurate*.

First, the GARCH model is able to forecast the 95% VaR with high accuracy during the precrash period, irrespective of the forecasting horizon. Accordingly, the square-root-of-

¹⁸Brummelhuis and Kaufmann [46] showed that the square-root-of-time rule applied to the GARCH(1,1) model works remarkably well.

¹⁹The normalization of the forecasting horizon by Δt is due to the fact that σ_{t+1} is a volatility forecast for Δt years and not as the Black-Scholes implied volatilities for one year.

²⁰Recall that, given the null hypothesis is true, the p -value denotes the probability of obtaining a test statistic, which is at least as “extreme” as the realized test statistic. In particular, large p -values indicate that the null hypothesis is true.

Table 6.6.: Value-at-Risk forecasting performance – Precrash period (January 4, 2006 – December 5, 2007).

Forecasting horizon (weeks)	1			2			3			4		
	95%	99%	99.9%	95%	99%	99.9%	95%	99%	99.9%	95%	99%	99.9%
Var level ($\alpha \cdot 100\%$)												
GARCH	5***	4*	3	6***	4*	1**	8***	4*	2	4***	2***	1**
<i>p</i> -value	(1.00)	(0.02)	(0.00)	(0.66)	(0.02)	(0.09)	(0.20)	(0.02)	(0.00)	(0.64)	(0.38)	(0.09)
Black-Scholes	4***	3***	0***	4***	2***	0***	5***	1***	0***	4***	0***	0***
<i>p</i> -value	(0.64)	(0.10)	(0.65)	(0.64)	(0.38)	(0.65)	(1.00)	(1.00)	(0.65)	(0.64)	(0.16)	(0.65)
Heston	5***	3***	2	5***	3***	2	6***	2***	1**	4***	2***	0***
<i>p</i> -value	(1.00)	(0.10)	(0.00)	(1.00)	(0.10)	(0.00)	(0.66)	(0.38)	(0.09)	(0.64)	(0.38)	(0.65)
2-Factor SV	4***	0***	0***	3***	0***	0***	2***	0***	0***	1*	0***	0***
<i>p</i> -value	(0.64)	(0.16)	(0.65)	(0.32)	(0.16)	(0.65)	(0.12)	(0.16)	(0.65)	(0.03)	(0.16)	(0.65)

Notes: This table summarizes the number of VaR exceptions of the GARCH, Black-Scholes, Heston and 2-Factor SV model for forecasting horizons of 1, 2, 3 together with 4 weeks and confidence levels of 95%, 99% and 99.9%. In order to test if the VaR with a confidence level of α is violated more or less than $(1 - \alpha) \cdot 100\%$ of the time, we apply the LR test of Kupiec [118], which is described in Chapter 6.5.2. The reported *p*-values correspond to the test statistic (6.5), where $0.01 \leq p < 0.05$ is indicated by *, $0.05 \leq p < 0.1$ by ** and $p \geq 0.1$ by ***.

Table 6.7.: Value-at-Risk forecasting performance – Crash period (December 12, 2007 – October 28, 2009).

Forecasting horizon (weeks)	1			2			3			4		
	95%	99%	99.9%	95%	99%	99.9%	95%	99%	99.9%	95%	99%	99.9%
VaR level ($\alpha \cdot 100\%$)												
GARCH <i>p</i> -value	8*** (0.20)	3*** (0.10)	2 (0.00)	8*** (0.19)	4* (0.02)	3 (0.00)	12 (0.005)	6 (0.00)	4 (0.00)	13 (0.00)	7 (0.00)	4 (0.00)
Black-Scholes <i>p</i> -value	4*** (0.64)	1*** (1.00)	1** (0.09)	5*** (0.98)	3*** (0.10)	2 (0.00)	6*** (0.62)	3*** (0.10)	2 (0.00)	8*** (0.18)	5 (0.00)	2 (0.00)
Heston <i>p</i> -value	4*** (0.64)	1*** (1.00)	1** (0.09)	5*** (0.98)	3*** (0.10)	1** (0.09)	6*** (0.62)	2*** (0.36)	2 (0.00)	8*** (0.18)	5 (0.00)	2 (0.00)
2-Factor SV <i>p</i> -value	5*** (1.00)	1*** (1.00)	0*** (0.65)	5*** (0.98)	2*** (0.37)	0*** (0.66)	6*** (0.62)	2*** (0.36)	0*** (0.66)	11* (0.01)	3** (0.095)	1** (0.09)

Notes: This table summarizes the number of VaR exceptions of the GARCH, Black-Scholes, Heston and 2-Factor SV model for forecasting horizons of 1, 2, 3 together with 4 weeks and confidence levels of 95%, 99% and 99.9%. In order to test if the VaR with a confidence level of α is violated more or less than $(1 - \alpha) \cdot 100\%$ of the time, we apply the LR test of Kupiec [118], which is described in Chapter 6.5.2. The reported *p*-values correspond to the test statistic (6.5), where $0.01 \leq p < 0.05$ is indicated by *, $0.05 \leq p < 0.1$ by ** and $p \geq 0.1$ by ***.

Table 6.8.: Value-at-Risk forecasting performance – Total period (January 4, 2006 – October 28, 2009).

Forecasting horizon (weeks)	1			2			3			4		
	95%	99%	99.9%	95%	99%	99.9%	95%	99%	99.9%	95%	99%	99.9%
VaR level ($\alpha \cdot 100\%$)												
GARCH	13*** (0.35)	7 (0.006)	5 (0.00)	14*** (0.21)	8 (0.00)	4 (0.00)	20 (0.00)	10 (0.00)	6 (0.00)	17* (0.03)	9 (0.00)	5 (0.00)
p -value												
Black-Scholes	8*** (0.50)	4*** (0.21)	1*** (0.20)	9*** (0.75)	5** (0.07)	2* (0.02)	11*** (0.72)	4*** (0.21)	2* (0.02)	12*** (0.50)	5** (0.07)	2* (0.02)
p -value												
Heston	9*** (0.74)	4*** (0.21)	3 (0.00)	10*** (0.99)	6* (0.02)	3 (0.00)	12*** (0.51)	4*** (0.21)	3 (0.00)	12*** (0.50)	7 (0.005)	2*** (0.02)
p -value												
2-Factor SV	9*** (0.74)	1*** (0.43)	0*** (0.53)	8*** (0.51)	2*** (0.99)	0*** (0.53)	8*** (0.52)	2*** (0.99)	0*** (0.53)	12*** (0.50)	3*** (0.49)	1*** (0.20)
p -value												

Notes: This table summarizes the number of VaR exceptions of the GARCH, Black-Scholes, Heston and 2-Factor SV model for forecasting horizons of 1, 2, 3 together with 4 weeks and confidence levels of 95%, 99% and 99.9%. In order to test if the VaR with a confidence level of α is violated more or less than $(1 - \alpha) \cdot 100\%$ of the time, we apply the LR test of Kupiec [118], which is described in Chapter 6.5.2. The reported p -values correspond to the test statistic (6.5), where $0.01 \leq p < 0.05$ is indicated by *, $0.05 \leq p < 0.1$ by ** and $p > 0.1$ by ***.

Figure 6.13.: Evolution of the GARCH parameters. These figures illustrate the evolution of the parameters α (a) and β (b) of the GARCH model specified in (6.10) - (6.11). The x-axis indicates the respective end date of the S&P 500 index return time series, which always starts on November 03, 1999, based on which the GARCH parameters are estimated. Furthermore, the y-axis denotes the corresponding estimation results for the parameters α and β .

time rule, which in our context scales one week volatility forecasts in up to four week volatility forecasts, seems to work well. Except for the two week ahead VaR forecasts, the forecasting performance deteriorates with increasing confidence level of the VaR. For example, VaR forecast with a confidence level of 99% and forecasting horizons of 1, 2 or 3 weeks are only slightly accurate. This bad forecasting performance for large confidence levels is not surprising since we assumed that the returns are normally distributed. However, as found in Chapter 6.3, the distribution of the S&P 500 index returns is slightly negatively skewed and fat tailed. Hence, the normality assumption will result in an underestimation of extreme events, which in turn results in too many VaR exceptions for large confidence levels. Interestingly, the forecasting performance for the 99.9% VaR is inaccurate for forecasting horizons of 1 and 3 weeks but accurate for forecasting horizons of 2 and 4 weeks. Unfortunately, the VaR forecasting performance of the GARCH model deteriorates in the crash period, where only one week ahead VaR forecasts with confidence levels of 95% and 99% as well as two week ahead VaR forecasts with a confidence level of 95% are highly accurate. Especially, if the forecasting horizon is larger than two weeks, all VaR forecasts are inaccurate. The bad forecasting performance of the GARCH model during the crash period can be explained by its parameter instability, which is illustrated in Figure 6.13. While the GARCH parameters α and β are pretty stable during the precrash period, they get strongly instable during the crash period. In particular, the bankruptcy of Lehman Brothers on September 15, 2008 results in a subsequent increase/decrease of the parameter α/β in the amount of approximately 0.1. Recall that our GARCH model describes the evolution of the conditional return variance as

$$\sigma_{t+1}^2 = \omega + \alpha r_t^2 + \beta \sigma_t^2, \quad (6.13)$$

where the time- t variance $\sigma_t^2 = \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2$ is only calculated based on past information. Therefore, the variance forecast (6.13) can be interpreted as a weighted average of current and past information. The increase/decrease of the parameter α/β at

6. Empirical study

the peak of the crisis can be interpreted in a way that current information becomes more important and past information less important.

Next, the VaR forecasting performance of the Black-Scholes model is remarkably well during the precrash period since all VaR forecasts are highly accurate, irrespective of the confidence level or forecasting horizon. Unfortunately, also the forecasting performance of the Black-Scholes model deteriorates during the crash period. However, despite this challenging forecasting period, the Black-Scholes model is still able to perform highly accurate VaR forecasts for confidence levels of 95% and 99% together with forecasting horizons of 1, 2 and 3 weeks. Also the 95% four week ahead VaR forecasts are highly accurate and the 99.9% one week ahead VaR forecasts are still accurate. Interestingly, the Black-Scholes model performs well also for large confidence levels, in particular, during the precrash period. This is quite surprising since, just like the GARCH model, also the Black-Scholes model relies on the normality assumption for the return distribution, which should be inadequate for large confidence levels. This phenomenon can be explained by a negative variance risk premium, whereby volatility estimates under the risk-neutral measure, as e.g. Black-Scholes implied volatilities, are consistently higher than volatility estimates under the statistical measure, as e.g. GARCH-type volatility estimates.²¹ As a result, also the VaR forecasts of the Black-Scholes model get larger since the quantile of a normal distribution is scaled by its volatility.²² Furthermore, the deviation between risk-neutral and statistical volatility estimates gets larger during times of high volatility. Again, this can be explained by the variance risk premium, which decreases during turbulent market conditions.²³ Hence, the VaR forecasts of the Black-Scholes model get even more conservative during periods of high volatility.

Similarly as the Black-Scholes model, also the Heston model has a good VaR forecasting performance during the precrash period. More precisely, the Heston model is able to perform highly accurate VaR forecasts for confidence levels of 95% and 99%, irrespective of the forecasting horizon. In contrast, the forecasting performance for a confidence level of 99.9% is mixed with inaccurate one and two week ahead VaR forecasts but accurate three week ahead and highly accurate four week ahead VaR forecasts. During the crash period, the VaR forecasting performance of the Heston model is also very similar to the one of the Black-Scholes model, except that the Heston model is also able to perform accurate 99.9% two week ahead VaR forecasts. Interestingly, the all over forecasting performance of the Black-Scholes model is slightly better than the one of the Heston model, even though the Black-Scholes model only relies on risk-neutral information and, in addition, assumes a simpler return distribution. A possible explanation for the inferior forecasting performance of the Heston model could be that the corresponding pricing kernel is too restrictive in order to account for time-varying preferences. The need for a more flexible modeling of preferences can be seen in Figure 6.8 from the strong time variation of the variance risk premium parameter γ , which reflects the preferences in the Heston model. Hence, a given estimate of γ represents an average value for the preferences in the un-

²¹See Chapter 3.4.5 for more details about this reasoning.

²²Note that if $X \sim \mathcal{N}(\mu, \sigma^2)$ then the α -quantile of X is given by $\mu + \sigma \cdot q_\alpha(\tilde{X})$, where $\tilde{X} \sim \mathcal{N}(0, 1)$.

²³The behavior of the variance risk premium during different market conditions is discussed in Chapter 6.4.

derlying sample period, which might deviate from the current preferences. Accordingly, the attempt to risk-adjust the risk-neutral density of the Heston model by virtue of a historical average variance risk premium parameter might result in a biased real-world density.

Finally, the VaR forecasting performance of the 2-Factor SV model is extraordinary since it is able to perform highly accurate VaR forecasts for almost all forecasting horizons and confidence levels, both during the precrash and crash period. During the precrash period, the VaR forecasting performance of the 2-Factor SV model is highly accurate for almost all confidence levels and forecasting horizons. The only exception is the 95% four week ahead VaR, which is only slightly accurate. Moreover, during the crash period, the 2-Factor SV model is able to perform highly accurate VaR forecasts for forecasting horizons upto three weeks, irrespective of the confidence level. In addition, four week ahead VaR forecasts are still accurate for confidence levels of 99% and 99.9% and slightly accurate for a confidence level of 95%. Most notably, if we consider the whole forecasting period, the 2-Factor SV model is the only model, which is able to perform highly accurate VaR forecasts for all confidence levels and forecasting horizons. The superior forecasting performance of the 2-Factor SV model can be attributed to its ability to account for time-varying preferences, which in turn allows for a more accurate estimation of the real-world density. The need for this additional flexibility compared to the Heston model can be seen from Figure 6.9(c), which illustrates the evolution of the risk aversion process θ_t .

Figure 6.14.: Evolution of the 99% one week ahead VaR forecast for the GARCH (a), Black-Scholes (b), Heston (c) and 2-Factor SV model (d).

6. Empirical study

In order to illustrate the differences of the analyzed VaR models, Figure 6.14 exemplarily shows the evolution of the different 99% one week ahead VaR forecasts together with the realized returns, so that the VaR exceptions can be identified. Furthermore, Figure 6.15 contrasts these different VaR forecasts. First, the VaR forecasts of the GARCH model are too loose during turbulent market conditions resulting in too many VaR exceptions during that time. Moreover, the GARCH model reacts too slowly and too little to changing market conditions, which further explains the inferior forecasting performance of the GARCH model. In contrast, the VaR models based on option-implied information are more conservative during the subprime mortgage crisis and, in addition, react faster and strong to changing market conditions, which can be attributed to the forward-looking information content of option prices. Further note that the VaR forecasts of the Black-Scholes model and the Heston model are pretty similar, even though the Black-Scholes model is slightly more conservative during normal market conditions and, in addition, reacts stronger to the bankruptcy of Lehman Brothers in September 2008. Finally, the 2-Factor SV model is slightly more conservative than the Heston model and the Black-Scholes model just one year before the bankruptcy of Lehman Brothers. Furthermore, the 2-Factor SV model is the most conservative VaR model during normal market conditions and reacts strongest to the bankruptcy of Lehman Brothers.

Figure 6.15.: Comparison of the 99% one week ahead VaR forecasts.

Altogether, VaR models based on option-implied information perform remarkably well, both during normal and turbulent market conditions. Thereby the 2-Factor SV model has the overall best forecasting performance, followed by the Black-Scholes model and the Heston model. In contrast, the GARCH model, which only relies on historical information, has the worst VaR forecasting performance, especially, during turbulent market conditions.

7. Conclusion

In the theory of financial economics, asset prices can be determined by computing the expectation of the “preference-weighted” payoff profile. Thereby the payoff profile of the underlying asset is weighted by some pricing kernel, which reflects the risk preferences of the representative investor over different states of the world. In the consumption-based model, the pricing kernel is specified by the utility function of the representative investor. In contrast, the pricing kernel of an option pricing model corresponds to the ratio of risk-neutral to statistical asset price dynamics and can be obtained by Girsanov’s theorem. Economic theory suggests that pricing kernels of risk averse investors are monotonically decreasing in wealth. Interestingly, empirical studies suggest that pricing kernels tend to be S-shaped and thus exhibit an increasing region. Furthermore, empirical pricing kernels change their shape through time, which indicates that risk preferences are time-varying. However, none of the standard option pricing models, such as e.g. the Heston model, is able to capture S-shaped pricing kernels. Also the consumption-based model together with common utility functions, such as power utility or exponential utility, cannot reproduce the shape of empirical pricing kernels. Therefore, this phenomenon is also called “pricing kernel puzzle”.

Option prices aggregate the forward-looking views of market participants about the future development of the underlying based on all currently available information. This forward-looking information can be extracted via theoretical option pricing models. Empirical studies suggest that the forward-looking information contained in option prices, such as e.g. Black-Scholes implied volatilities or entire risk-neutral densities, do have a superior forecasting performance compared to forecasts based on historical information, such as e.g. time series of historical stock returns. Unfortunately, the risk-neutral density also contains risk preferences and thus is a biased estimator of future asset prices unless market participants are risk-neutral. However, the risk-neutral density can be adjusted for risk preferences by means of pricing kernels, so that the resulting real-world density is also based on forward-looking information. Some empirical studies suggest that forward-looking real-world densities do have a better forecasting performance than their unadjusted risk-neutral counterpart. The crucial point of this approach is that the supposed pricing kernel used for the risk-adjustment must adequately describe the current risk preferences. Otherwise, also the forward-looking real-world density is a biased estimator of future asset prices. This problem was highlighted within this thesis by an empirical study based on Credit Default Swap data. Thereby, the risk-neutral density estimates were adjusted for risk preferences via a utility transformation and a more flexible calibration transformation. On the one hand, we found that both pricing kernels exhibit strong parameter instability, which indicates that the underlying preferences are

7. Conclusion

time-varying. On the other hand, the two transformation methods did not improve the forecasting performance compared to the unadjusted risk-neutral density.

Affine processes offer a great flexibility in modeling asset prices while keeping valuation problems, such as option pricing, analytically tractable. One of the key properties of affine processes is that their characteristic function is known up to the solution of a system of Riccati equations, which can either be solved in closed-form or via some numerical method. This allows to compute option prices via Fourier inversion techniques. A commonly used Fourier inversion technique in this context is the Carr-Madan formula, which allows to compute option prices via the *Fast Fourier transform* (FFT). However, the speed advantage of the FFT compared to a direct numerical integration is only given if we need to calculate option prices for a large number of different strike prices. In contrast, for a moderate number of strike prices, as sufficient in most applications, a direct numerical integration is much faster. Moreover, the Carr-Madan formula comprises an additional free parameter, which is also called the damping coefficient. A smart choice for the damping coefficient can facilitate the numerical integration.

The Heston model is a stochastic volatility model, which describes the asset price dynamics by an affine process. Hence, option prices can be calculated based on the characteristic function of the underlying by means of Fourier inversion techniques. The Heston model is able to capture most of the stylized facts of asset returns, such as fat tails, the leverage effect and volatility clustering. Also the commonly observed smile or skew shape of the volatility surface implied by option prices can be realized by the Heston model. Since the volatility is modeled by a stochastic process, the Heston model incorporates a variance risk premium. Heston [95] motivated his choice for the variance risk premium by Breeden's [40] consumption-based capital asset pricing model, which relates the variance risk premium to the risk preferences of market participants. In this connection, a negative variance risk premium is in line with risk averse market participants. Empirical studies suggest that the variance risk premium is negative. Furthermore, there is some empirical evidence that the variance risk premium is time-varying, which can only inadequately be captured by the Heston model. In addition, the pricing kernel of the Heston model cannot capture the S-shape, which is typically found in the empirical literature. To overcome these problems, we propose a new 2-Factor *stochastic volatility* (SV) model, which is based on the Heston model and also describes the asset price dynamics via an affine process. Hence, option prices can be calculated based on the characteristic function of the underlying. The basic idea of the 2-Factor SV model is to describe the risk preferences of market participants by an additional stochastic process called the "risk aversion process". This risk aversion process is assumed to be uncorrelated to the statistical asset price dynamics and only enters the risk-neutral volatility dynamics. Thereby, the 2-Factor SV model is able to capture changes in option prices, which are only due to changes in risk preferences as opposed to changes in the statistical asset price dynamics. On the one hand, the 2-Factor SV model allows for a more flexible volatility term structure than the Heston model since it is also able to capture reversing volatility term structures, such as a first increasing and then decreasing volatility term structure. On the other hand, the 2-Factor SV model is able to produce S-shaped pricing kernels. Besides, the variance risk premium dynamic of the 2-Factor SV model is economically intuitive. More precisely, during normal

market conditions and thus low levels of volatility, a negative variance risk premium is less likely, which indicates that market participants are rather risk seeking. Conversely, during turbulent market conditions and thus high levels of volatility, a negative variance risk premium is more likely, which indicates that market participants are rather risk averse.

The statistical and risk-neutral asset price dynamics of an option pricing model can be estimated by virtue of the unscented Kalman filter. The unscented Kalman filter is an extension of the classical Kalman filter, which allows to account for nonlinearities in the filtering equations as often encountered in financial applications. Heart of the unscented Kalman filter is the unscented transformation, which is a numerical method in order to approximate the mean and variance of a possibly nonlinear transformed random variable. This approximation method was developed by Julier and Uhlmann [106] and is at least accurate up to the second order Taylor approximation. Thereby, the unscented Kalman filter achieves a higher filtering accuracy than the commonly used extended Kalman filter, which solely relies on a first order Taylor approximation of the state-space representation. In order to estimate the parameters of an option pricing model, we first need to put the model into state-space form. This can be achieved via an Euler discretization of the statistical asset price dynamics. Then the parameters of an option pricing model can be estimated via the unscented Kalman filter together with the *Quasi-maximum likelihood* (QML) method. Under certain regularity conditions, which involve that the conditional mean and variance of the measurements are correctly specified, the QML estimator is weakly consistent and asymptotically normally distributed. Accordingly, we are also able to perform statistical inference for the estimated parameters of an option pricing model based on the corresponding standard error estimates. A sparse parameterization of the measurement error can be achieved via the concept of vega-scaling, where option prices are divided by their corresponding Black-Scholes vega. The resulting time series of vega-scaled option prices is much more comparable across different strikes and maturities.

The ability of the unscented Kalman filter together with the QML method to accurately estimate the parameters of the Heston model was shown within this thesis by means of a thorough Monte-Carlo study. However, we found that it is more challenging to estimate the parameters of the 2-Factor SV model, which is due to the additional risk aversion process. Nevertheless, most parameters of the 2-Factor SV model can be estimated sufficiently well within this estimation framework. Furthermore, we found that too small values for the scaling parameter λ of the unscented transformation together with the adaptive Simpson rule as numerical integration technique for the option price calculation can result in a discontinuous log-likelihood function. To overcome this problem, we suggest to use rather large values for the scaling parameter λ together with the adaptive Gauss-Kronrod rule as numerical integration technique. The latter is also much faster than the adaptive Simpson rule. In addition, we developed and applied a caching technique for the estimation of the 2-Factor SV model together with the unscented Kalman filter. This caching technique exploits that once the characteristic function of the 2-Factor SV model is evaluated for a sufficiently dense integration grid, option prices can be computed for arbitrary values of the underlying, variance and risk aversion, respectively. Hence, our caching technique offers a great speed advantage in connection with the unscented Kalman filter.

7. Conclusion

In the empirical part of this thesis, we estimated the Heston model and the 2-Factor SV model via the unscented Kalman filter based on S&P 500 index options. The underlying data period ranges from November 3, 1999 to October 28, 2009 and thus includes several market shocks like the burst of the dot-com bubble, the terror attacks 9/11 and the subprime mortgage crisis. For both models, we obtained reasonable and highly significant estimation results. In line with the previous literature, our estimates for the variance risk premium of the Heston model are negative. Similarly, also the variance risk premium of the 2-Factor SV model is negative during turbulent market conditions. However, during the calm market period from mid-2004 to mid-2007, the variance risk premium of the 2-Factor SV model is slightly positive. Due to the connection between the variance risk premium and risk aversion, this can be interpreted as an indication of risk seeking market behavior during that time. From mid-2007 the variance risk premium turns again negative and the risk aversion process continuously increases, which reflects the growing fear of the market participants during the subprime mortgage crises. Moreover, we found that the risk aversion process has a lower variability compared to the volatility process, which makes it easier to identify changing market conditions. In addition, the 2-Factor SV model fits the empirical volatility term structure much better than the Heston model and also features a higher parameter stability. Finally, we performed out-of-sample *Value-at-Risk* (VaR) forecasts for the period of the subprime mortgage crises based on the estimation results of the Heston model and the 2-Factor SV model. The VaR refers to the maximum loss of a position, which is not exceeded for a given confidence level and forecasting horizon. For our empirical study, we choose confidence levels of 95%, 99% and 99.9% together with forecasting horizons of 1, 2, 3 and 4 weeks. In order to better classify these results, we also performed VaR forecasts based on two commonly used VaR models, namely the Black-Scholes model, which also relies on option-implied information, and the GARCH model, which relies on historical information. The accuracy of our VaR forecasts was checked via the test of Kupiec [118], which tests if the VaR with a confidence level of α is violated more or less than $(1 - \alpha) \cdot 100\%$ of the time. As a result, the 2-Factor SV model has the best forecasting performance, followed by the Black-Scholes, the Heston and the GARCH model. In particular, the 2-Factor SV model is the only one, which is able to perform highly accurate VaR forecasts for all confidence levels and forecasting horizons despite the challenging forecasting period. The superior forecasting performance of the 2-Factor SV model can be attributed to its incorporation of time-varying risk preferences and its S-shaped pricing kernel.

Further research can be conducted in different directions. From a theoretical point of view, the 2-Factor SV model could be extended by adding a jump process to the asset price dynamics, which would further improve the fit for short dated options. The necessity to add jumps to the asset price dynamics can be seen from the high estimate for the volatility parameter of the variance process, which was pretty close to one. Furthermore, it would be interesting to analyze the pricing kernel properties of this 2-Factor SV jump model. Concerning the estimation framework, we could consider alternative specifications for the measurement error, which are also able to capture the correlations between the measurement errors and thus are more adequate. This can be achieved, for example, via multivariate GARCH models. From an empirical point of view, we could make a more detailed analysis about how the pricing kernel of the 2-Factor SV model changes

through time. Moreover, we could use alternative pricing kernel specifications, such as the four-parameter orthogonal polynomial specification of Rosenberg and Engle [138] or the non-parametric approach of Aït-Sahalia and Lo [1], and compare the corresponding results with the one for the 2-Factor SV model. Along the same lines, we could also check if our estimation results for the variance risk premium based on the 2-Factor SV model are in line with alternative estimation methods for the variance risk premium (see e.g. Bakshi and Kapadia [12] or Bollerslev et al. [36]). Moreover, in accordance with Bollerslev et al. [36], the risk aversion process of the 2-Factor SV model could be used as an explanatory variable in order to explain future stock returns via a linear regression model. Thereby, it would be interesting to check if the risk aversion process does have a higher predictive power than more traditional explanatory variables in this context, such as e.g. the price-earnings ratio, credit spreads or industrial production. Finally, the 2-Factor SV model could also be estimated for other markets, such as e.g. the German market. In this connection, it would be interesting to check if other markets also exhibit positive variance risk premia some time before the subprime mortgage crises. In addition, we could examine the correlation between the risk aversion processes of different markets.

8. Zusammenfassung

In vielen Bereichen der Finanzwirtschaft ist eine präzise Vorhersage der Preisentwicklung von Vermögenswerten, wie z.B. Aktien oder Anleihen, von großer Bedeutung. Neben der Bewertung von Optionen und der Portfoliotheorie ist hier insbesondere das Risikomanagement zu nennen. Zwar kann in der Regel keine genaue Aussage darüber getroffen werden, welchen Preis ein Vermögenswert zukünftig erreicht. Jedoch kann mit Hilfe einer entsprechenden Verteilungsschätzung die erwartete Preisentwicklung, Volatilität oder die Wahrscheinlichkeit einzelner Szenarien beurteilt werden. Im Risikomanagement ist vor allem das Szenario eines großen Verlustes bzw. dessen Eintrittswahrscheinlichkeit von Interesse. Das wohl bekannteste Maß zur Bewertung von Risiken ist der Value-at-Risk, welcher den maximalen Verlust einer Position bei vorgegebenem Konfidenzniveau und Prognosehorizont angibt (siehe z.B. Tasche [147]).

Um die Verteilung zukünftiger Preise oder Renditen zu schätzen, verwendet man für gewöhnlich ein ökonometrisches oder finanzmathematisches Modell, welches auf Basis von historischen Daten geschätzt wird. Dementsprechend spiegelt das Schätzergebnis die Charakteristik der Vergangenheit wider. Wenn sich allerdings zum Ende des historischen Datensatzes ein Strukturbruch ereignet, dann erfasst das Prognosemodell die aktuelle Marktsituation nicht oder nur unzureichend. So verwundert es nicht, dass die Risikomodelle von Banken große Probleme hatten, das Risiko während der Subprimekrise zu messen. Exemplarisch kann an dieser Stelle Charles Prince, der ehemaligen CEO der Citigroup, zitiert werden, der Abschreibungen in Milliardenhöhe mit den Worten “our risk models simply did not work” [32] kommentierte.

Eine Alternative zu vergangenheitsorientierten Prognosen auf Basis von historischen Daten bieten Optionspreise, die alle momentan am Markt verfügbaren Informationen über die zukünftige Preisentwicklung des Underlyings enthalten. Diese in die Zukunft schauenden Informationen, welche insbesondere auch die aktuelle Marktlage widerspiegeln, können mit Hilfe von Optionspreismodellen extrahiert werden. Empirische Studien belegen, dass die in Optionspreisen enthaltenen Informationen, wie z.B. Black-Scholes implizite Volatilitäten, eine bessere Prognosequalität haben als Vorhersagen auf Basis von historischen Daten (siehe Christoffersen et al. [57] für einen Literaturüberblick).

Ein Optionspreismodell besteht aus zwei Komponenten, der statistischen und der risikoneutralen Dynamik. Während die statistische Dynamik die Charakteristik des tatsächlichen Preisverlaufs eines Underlyings beschreibt, ist die risiko-neutrale Dynamik sowohl vom tatsächlichen Preisverlauf als auch von den Risikopräferenzen der Marktteilnehmer bestimmt. Insbesondere ist die risiko-neutrale Dynamik relevant für die Bewertung von

Optionen. Das Verhältnis von risiko-neutraler zu statistischer Dynamik, welches auch als “Pricing Kernel” bezeichnet wird, veranschaulicht die unterstellten Risikopräferenzen des Optionspreismodells. Der Name “Pricing Kernel” wird vor allem im Kontext von sogenannten “Consumption-based Models” oder zu deutsch “konsumbasierten Modellen” verwendet, die das intertemporale Konsum- und Investitionsentscheidungsproblem eines repräsentativen Investors beschreiben. Maßgeblich für die Struktur des Pricing Kernels im konsumbasierten Modell ist die angenommene Nutzenfunktion. Gewöhnliche Nutzenfunktionen implizieren Pricing Kernels, die mit dem Konsumniveau monoton fallen. Dementsprechend ist der repräsentative Investor bereit, für Auszahlungen in schlechten Umweltzuständen (niedriges Konsumniveau) mehr zu zahlen als für Auszahlungen in guten Umweltzuständen (hohes Konsumniveau). Ein Beispiel für Vermögenswerte, die in schlechten Umweltzuständen zu Auszahlungen führen, sind Versicherungen [59, p. 13 f.].

Um die in Optionspreisen enthaltenen, in die Zukunft schauenden Informationen mit Hilfe von Optionspreismodellen zu extrahieren, muss sowohl die risiko-neutrale Dynamik als auch der Pricing Kernel auf angemessene Art und Weise modelliert werden. Die empirischen Eigenschaften der risiko-neutralen Dynamik sind wohl bekannt und werden üblicherweise in Form von Black-Scholes impliziten Volatilitäten beschrieben (siehe z.B. Cont et al. [61]). Insbesondere sind moderne Optionspreismodelle, wie z.B. das stochastische Volatilitätsmodell von Heston [95], in der Lage diese Eigenschaften abzubilden. Hingegen sind die empirischen Eigenschaften von Pricing Kernels wenig erforscht. Dies kann hauptsächlich darauf zurückgeführt werden, dass die Risikopräferenzen von Marktteilnehmern nicht direkt beobachtbar und deshalb auch schwer zu schätzen sind. Allerdings deuten einige empirische Studien darauf hin, dass Pricing Kernels typischerweise S-förmig sind bzw. einen Bereich aufweisen, indem sie monoton zunehmen (siehe z.B. Ait-Sahalia und Lo [1], Jackwerth [99] und Rosenberg und Engle [138]). Da die Pricing Kernels üblicher Optionspreismodelle und konsumbasierter Modelle allerdings nicht in der Lage sind, diese typische Charakteristik abzubilden, wird dieses Phänomen auch als “Pricing Kernel Puzzle” bezeichnet (siehe z.B. Brown and Jackwerth [44]). Zudem ist die Form von Pricing Kernels normalerweise nicht konstant, sondern ändert sich über die Zeit, was ein Hinweis auf zeitvariierende Risikopräferenzen ist (siehe z.B. Rosenberg und Engle [138] und Grith et al. [89]).

Neben einem passenden Optionspreismodell benötigt man ein geeignetes Schätzverfahren, um die in Optionspreisen enthaltenen Informationen zu extrahieren. Zudem genügt es für Prognosezwecke nicht nur, die risiko-neutrale Dynamik zu schätzen, da diese durch die Risikopräferenzen der Marktteilnehmer verzerrt ist. Eine Möglichkeit, die statistische und risiko-neutrale Dynamik eines Optionspreismodells gemeinsam zu schätzen, liefert der Kalman Filter. Der Kalman Filter ist ein Algorithmus, um den Zustand eines zeitvariierenden Systems auf Basis von verrauschten Beobachtungen zu schätzen. Ursprünglich wurde der Kalman Filter für ingenieurwissenschaftliche Anwendungen, wie z.B. der GPS (Global Positioning System) Navigation, genutzt. Inzwischen findet der Kalman Filter auch Anwendung im Kontext von ökonometrischen Fragestellungen. Im Gegensatz zum klassischen Kalman Filter kann der von Julier und Uhlmann [106] entwickelte Unscented Kalman Filter auch auf nicht-lineare Systeme angewendet werden, wodurch dessen Einsatzmöglichkeiten deutlich vielseitiger sind. So kann der Unscented Kalman Filter auch

zur Schätzung von Optionspreismodellen verwendet werden, was erstmalig von Bakshi et al. [11] und Bakshi and Wu [16] vorgeschlagen wurde.

Der erste Beitrag dieser Dissertation ist die Entwicklung eines neuen Optionspreismodells, das sowohl die empirischen Eigenschaften der risiko-neutralen Dynamik als auch die tatsächliche Charakteristik von Pricing Kernels berücksichtigt. Basis dieses neuen 2-Faktor stochastischen Volatilitäts (SV) Modells ist das Heston Modell, welches um einen Risikoaversionsprozess erweitert wird. Dieser Risikoaversionsprozess geht allerdings nur in die risiko-neutrale Dynamik und nicht in die statistische Dynamik ein. Dadurch können Änderungen von Optionspreisen abgebildet werden, welche ausschließlich auf veränderte Risikopräferenzen zurückzuführen sind und nicht durch Änderungen in der statistischen Dynamik verursacht werden. Insbesondere ist dieses Modell in der Lage, S-förmige Pricing Kernels und zeitvariierende Risikopräferenzen abzubilden. Nicht zuletzt hat das 2-Faktor SV Modell auch eine flexiblere Volatilitätsstrukturkurve als das Heston Modell, wodurch u.a. auch Zeitreihen von Optionspreisen besser erklärt werden können.

Der zweite Beitrag dieser Arbeit besteht aus einer umfassenden Monte-Carlo Studie, in der die Schätzgenauigkeit des Unscented Kalman Filter für das Heston Modell und das 2-Faktor SV Modell überprüft wird. Dabei zeigt sich, dass die Parameter des Heston Modells sehr genau geschätzt werden können. Hingegen ist die Schätzung des 2-Faktor SV Modells auf Grund des zusätzlichen Risikoaversionsprozesses deutlich schwieriger. Gleichwohl können die meisten Parameter des 2-Faktor SV Modells gut geschätzt werden. Zudem werden im Rahmen der Monte-Carlo Studie numerische Schwierigkeiten beleuchtet, welche bei der Schätzung von Optionspreismodellen in Verbindung mit dem Unscented Kalman Filter auftreten können. Genauer kann eine ungünstige Parametrisierung des Unscented Kalman Filters zu einer nicht stetigen Log-Likelihood-Funktion führen, falls die adaptive Simpson-Regel für die Berechnung von Optionspreisen verwendet wird. Außerdem wird eine spezielle Speichertechnik vorgestellt, welche die Schätzung des 2-Faktor SV Modells mit dem Unscented Kalman Filter erheblich beschleunigt.

Der dritte Beitrag dieser Dissertation ist eine zweiteilige empirische Studie. Im ersten Teil der empirischen Studie werden das Heston Modell und das 2-Faktor SV Modell auf Basis von S&P 500 Index Optionen in Verbindung mit dem Unscented Kalman Filter geschätzt. Die zugrundeliegende Datenperiode reicht vom 3. November 1999 bis zum 28. Oktober 2009 und umfasst somit verschiedene Marktschocks, wie das Platzen der Internetblase, die Terroranschläge vom 11. September und die Subprimekrise. Die Schätzergebnisse für beide Modelle sind hoch signifikant und nehmen zudem plausible Werte an. Besonders hervorzuheben sind hier die Schätzergebnisse für die Varianz-Risikoprämie, welche Rückschlüsse auf die Risikopräferenzen von Marktteilnehmern zulässt. Ähnlich wie in der vorangegangenen Literatur (siehe z.B. Chernov und Ghysels [54], Pan [130] und Jones [104]) wurde für das Heston Modell eine negative Varianz-Risikoprämie ermittelt, was auf risiko-averse Marktteilnehmer schließen lässt. Auch die Varianz-Risikoprämie des 2-Faktor SV Modells ist während turbulenter Marktphasen negativ. Hingegen ist die Varianz-Risikoprämie des 2-Faktor SV Modells während der ruhigen Marktphase von Mitte 2004 bis Mitte 2007 leicht positiv, was auf risikofreudige Marktteilnehmer hindeutet. Ab Mitte 2007 wird die Varianz-Risikoprämie wieder negativ und der Risikoaver-

8. Zusammenfassung

sionsprozess nimmt stetig zu, was sich durch die wachsende Angst der Marktteilnehmer während der Subprimekrise erklären lässt. Außerdem zeigt sich, dass der Risikoaversionsprozess eine geringere Schwankungsintensität als der Varianzprozess hat, wodurch eine veränderte Marktlage einfacher zu erkennen ist. Nicht zuletzt kann das 2-Faktor SV Modell die empirische Volatilitätsstrukturkurve besser als das Heston Modell erklären und hat zudem eine höhere Parameterstabilität.

Im zweiten Teil der empirischen Studie wird untersucht, in welcher Qualität das 2-Faktor SV Modell durch die Berücksichtigung zweitvariierender Risikopräferenzen bzw. durch den S-förmigen Pricing Kernel die in Optionspreisen enthaltenen Informationen extrahiert. Um dies zu überprüfen wurden Value-at-Risk-Prognosen für den Zeitraum der Subprimekrise durchgeführt. Hierbei wird der Value-at-Risk mit Prognosehorizonten von 1, 2, 3 und 4 Wochen und Konfidenzniveaus von 95%, 99% und 99,9% für den S&P 500 Index berechnet. Als Referenzmodell wird neben dem Heston Modell und dem Black-Scholes Modell auch das GARCH-Modell untersucht, das im Gegensatz zu den übrigen Prognosemodellen ausschließlich historische Informationen verwendet. Hauptergebnis dieser Untersuchung ist, dass das 2-Faktor SV Modell die beste Prognosequalität hat, gefolgt vom Black-Scholes, vom Heston und vom GARCH-Modell. Insbesondere ist das 2-Faktor SV Modell das einzige Modell, das in der Lage ist, für alle Prognosehorizonte und alle Konfidenzniveaus genaue Value-at-Risk-Prognosen zu erstellen - trotz der herausfordernden Prognoseperiode.

A. Supplementary theory and notation

A.1. Properties of the multivariate normal distribution

In accordance to Särkkä [140], we will describe some important properties of the multivariate normal distribution.

Definition A.1 (Multivariate normal distribution). *A random variable $X \in \mathbb{R}^n$ is multivariate normally distributed with mean $\mu \in \mathbb{R}^n$ and variance $\Sigma \in \mathbb{R}^{n \times n}$ if the probability density function of X is given by*

$$f_X(x) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x - \mu)^\top \Sigma^{-1}(x - \mu)\right\} ,$$

where $|\Sigma|$ denotes the determinant of Σ .

A multivariate normally distributed random variable with mean μ and variance Σ will be denoted by $X \sim \mathcal{N}(\mu, \Sigma)$.

Lemma A.1 (Joint distribution of normal random variables). *Let $X \in \mathbb{R}^m$ and $Z \in \mathbb{R}^n$ be independent and multivariate normally distributed random variables with $X \sim \mathcal{N}(\mu, \Sigma_1)$ and $Z \sim \mathcal{N}(0, \Sigma_2)$, respectively. Moreover, define*

$$Y = AX + Z ,$$

where A is a $n \times m$ matrix. Then the joint distribution of X and Y is given by

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu \\ A\mu \end{bmatrix}, \begin{bmatrix} \Sigma_1 & \Sigma_1 A^\top \\ A\Sigma_1 & A\Sigma_1 A^\top + \Sigma_2 \end{bmatrix}\right) .$$

Lemma A.2 (Conditional distribution of normal random variables). *Let $X \in \mathbb{R}^m$ and $Y \in \mathbb{R}^n$ be multivariate normally distributed random variables with joint distribution*

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_1 & \Sigma_3 \\ \Sigma_3^\top & \Sigma_2 \end{bmatrix}\right) .$$

Then the marginal and conditional distribution of X and Y are given by

$$\begin{aligned} X &\sim \mathcal{N}(\mu_1, \Sigma_1) , \\ Y &\sim \mathcal{N}(\mu_2, \Sigma_2) , \end{aligned}$$

$$\begin{aligned} X|Y &\sim \mathcal{N}(\mu_1 + \Sigma_3 \Sigma_2^{-1} (Y - \mu_2), \Sigma_1 - \Sigma_3 \Sigma_2^{-1} \Sigma_3^\top) , \\ Y|X &\sim \mathcal{N}(\mu_2 + \Sigma_3^\top \Sigma_1^{-1} (X - \mu_1), \Sigma_2 - \Sigma_3^\top \Sigma_1^{-1} \Sigma_3) . \end{aligned}$$

A.2. Multi-dimensional Taylor series expansion

Following Simon [145, p. 23], we will explain the short notation for the multi-dimensional Taylor series. Let $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be a nonlinear transformation with $f \in C^\infty$. For ease of notation, we first start with the 2-dimensional case and expand f around $\bar{x} \in \mathbb{R}^2$, which is

$$\begin{aligned} f(x) = & f(\bar{x}) + \left. \frac{\partial f(x)}{\partial x_1} \right|_{x=\bar{x}} \tilde{x}_1 + \left. \frac{\partial f(x)}{\partial x_2} \right|_{x=\bar{x}} \tilde{x}_2 + \\ & \frac{1}{2!} \left(\left. \frac{\partial^2 f(x)}{\partial x_1^2} \right|_{x=\bar{x}} \tilde{x}_1^2 + 2 \left. \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} \right|_{x=\bar{x}} \tilde{x}_1 \tilde{x}_2 + \left. \frac{\partial^2 f(x)}{\partial x_2^2} \right|_{x=\bar{x}} \tilde{x}_2^2 \right) + \\ & \frac{1}{3!} \left(\left. \frac{\partial^3 f(x)}{\partial x_1^3} \right|_{x=\bar{x}} \tilde{x}_1^3 + 3 \left. \frac{\partial^3 f(x)}{\partial x_1^2 \partial x_2} \right|_{x=\bar{x}} \tilde{x}_1^2 \tilde{x}_2 + 3 \left. \frac{\partial^3 f(x)}{\partial x_1 \partial x_2^2} \right|_{x=\bar{x}} \tilde{x}_1 \tilde{x}_2^2 + \left. \frac{\partial^3 f(x)}{\partial x_2^3} \right|_{x=\bar{x}} \tilde{x}_2^3 \right) + \\ & \dots , \end{aligned}$$

where $\tilde{x} = x - \bar{x}$. This can also be expressed more compactly as

$$\begin{aligned} f(x) = & f(\bar{x}) + \left(\tilde{x}_1 \frac{\partial}{\partial x_1} + \tilde{x}_2 \frac{\partial}{\partial x_2} \right) f(x) \Big|_{x=\bar{x}} + \frac{1}{2!} \left(\tilde{x}_1 \frac{\partial}{\partial x_1} + \tilde{x}_2 \frac{\partial}{\partial x_2} \right)^2 f(x) \Big|_{x=\bar{x}} + \\ & \frac{1}{3!} \left(\tilde{x}_1 \frac{\partial}{\partial x_1} + \tilde{x}_2 \frac{\partial}{\partial x_2} \right)^3 f(x) \Big|_{x=\bar{x}} + \dots . \end{aligned}$$

Now, we apply this short notation to the n -dimensional case, so that

$$\begin{aligned} f(x) = & f(\bar{x}) + \left(\tilde{x}_1 \frac{\partial}{\partial x_1} + \dots + \tilde{x}_n \frac{\partial}{\partial x_n} \right) f(x) \Big|_{x=\bar{x}} + \\ & \frac{1}{2!} \left(\tilde{x}_1 \frac{\partial}{\partial x_1} + \dots + \tilde{x}_n \frac{\partial}{\partial x_n} \right)^2 f(x) \Big|_{x=\bar{x}} + \\ & \frac{1}{3!} \left(\tilde{x}_1 \frac{\partial}{\partial x_1} + \dots + \tilde{x}_n \frac{\partial}{\partial x_n} \right)^3 f(x) \Big|_{x=\bar{x}} + \dots . \end{aligned}$$

Furthermore, if we define the operator $D_{\tilde{x}}^k f$ as

$$D_{\tilde{x}}^k f = \left(\sum_{i=1}^n \tilde{x}_i \frac{\partial}{\partial x_i} \right)^k f(x) \Big|_{x=\bar{x}} ,$$

the n -dimensional Taylor series expansion can be written as

$$f(x) = f(\bar{x}) + D_{\bar{x}}^1 f + \frac{1}{2!} D_{\bar{x}}^2 f + \frac{1}{3!} D_{\bar{x}}^3 f + \dots$$

A.3. Runge-Kutta method

The *Runge-Kutta method* is a family of numerical algorithms, which can be used to approximate the solution of an *ordinary differential equation* (ODE). In the following, we consider the initial value problem

$$y'(t) = f(t, y(t)) , \quad y(t_0) = y_0 , \quad (\text{A.1})$$

and try to approximate the solution of $y \in C^1([t_0, T], \mathbb{R}^d)$ on a discrete grid $t_0 < t_1 < \dots < t_n = T$. For simplicity, we assume that the discrete grid is equidistant, so that $\tau = t_i - t_{i-1}$ for $i = 1, \dots, n$. Furthermore, we will indicate the dependence of y on t in the subscript and set $t_i = i$.

At first, we can write the solution of the ODE (A.1) in integral form as

$$\begin{aligned} y_T &= \int_{t_0}^T f(s, y_s) ds + y_0 \\ &= \sum_{i=0}^{n-1} \Delta y(t_{i+1}, t_i) + y_0 , \end{aligned}$$

where

$$\Delta y(t_{i+1}, t_i) = \int_{t_i}^{t_{i+1}} f(s, y_s) ds . \quad (\text{A.2})$$

Even though the function f is known, we cannot calculate $\Delta y(t_{i+1}, t_i)$ since y is only known at $t = t_0$. To overcome this problem, we approximate $\Delta y(t_1, t_0)$ based on the knowledge of y_0 , so that also y_1 can be approximated. Then we continue to approximate y_2 via the same procedure and so forth. This iterative approach is the basic idea for many numerical ODE solver algorithms such as e.g. the Runge-Kutta method. Before we explain the Runge-Kutta method in more detail, we will start with the most basic numerical ODE solver – the *Euler method*.

Following Deuffhard and Borneman [66, pp. 131 f.], the *Euler method* is a very simple approach based on the Taylor approximation. Thereby, the change (A.2) is approximated by a first order Taylor approximation, which is

$$\Delta y(t_{i+1}, t_i) \approx f(t_i, y_i) \tau .$$

Since y_0 is known, the solution of (A.1) can be approximated recursively via

$$\hat{y}_{i+1} = \hat{y}_i + f(t_i, \hat{y}_i) \tau \quad \text{for } i = 0, \dots, n-1 , \quad (\text{A.3})$$

A. Supplementary theory and notation

where we set $\hat{y}_0 = y_0$.

To assess the accuracy of the Euler method, we need to introduce some further notation. First, the solution of the ODE (A.1) can also be expressed as

$$y_t = \Phi(t, t_0)y_0 ,$$

where Φ is called the *evolution* of the ODE [66, p. 45]. Hence, the numerical approximation for the solution of y_t on the grid $t_0 < t_1 < \dots < t_n = T$ can be expressed as

$$y_i = \Psi(t_i, t_0)y_0 \quad \text{for } i = 1, \dots, n ,$$

where Ψ denotes the *discrete evolution* [66, p. 130]. Following Deuffhard and Borneman [66, pp. 131ff.], we can now introduce the notion of *consistency*.

Definition A.2 (Consistency). *The discrete evolution Ψ is consistent of order p if the consistency error*

$$\epsilon(t, x, \tau) = \Phi(t + \tau, t)x - \Psi(t + \tau, t)x$$

fulfills the relation

$$\epsilon(t, x, \tau) = \mathcal{O}(\tau^{p+1}) \quad \text{for } \tau \rightarrow 0 .$$

The order of consistency is a measure for the accuracy of the numerical method Ψ . In particular, consistency is a necessary condition that the numerical solution converges to the true solution.¹ The Euler method is consistent of order $p = 1$ [66, pp. 133f.]. To achieve higher orders of consistency the Euler approach can be extended by using Taylor approximations of higher order. Unfortunately, this will also necessitate to calculate derivatives of f , which can get quite complex for higher orders [66, pp. 140].

The *Runge-Kutta method* is a derivative free alternative to the Euler method. Following Deuffhard and Borneman [66, p. 141ff.], we will illustrate the underlying idea exemplarily for the *2-stage Runge-Kutta method*. First note that the change (A.2) can also be approximated by the *midpoint rule*, which is

$$\Delta y(t_{i+1}, t_i) = \tau f(t_i + \frac{\tau}{2}, y_{i+\tau/2}) + \mathcal{O}(\tau^3) .$$

Unfortunately, we cannot approximate the solution of (A.1) iteratively via

$$\Delta y(t_{i+1}, t_i) \approx \tau f(t_i + \frac{\tau}{2}, y_{i+\tau/2}) , \tag{A.4}$$

since (A.4) cannot be calculated based on the knowledge of y_i . However, since f is multiplied by τ , it is sufficient to approximate $y_{i+\tau/2}$ up to terms of order $\mathcal{O}(\tau^2)$ to maintain the order of accuracy in (A.4) [66, p. 142]. Hence, we can simply approximate $y_{i+\tau/2}$ via the Euler method (A.3), which is

$$y_{i+\tau/2} = y_i + \frac{\tau}{2} f(t_i, y_i) + \mathcal{O}(\tau^2) . \tag{A.5}$$

¹For more details see Deuffhard and Borneman [66, Satz 4.10].

In particular, we can now plug (A.5) into (A.4), so that

$$\Delta y(t_{i+1}, t_i) \approx \tau f(t_i + \frac{\tau}{2}, y_i + \frac{\tau}{2} f(t_i, y_i)) ,$$

which is still accurate up to order $\mathcal{O}(\tau^3)$. Now, the solution of (A.1) can be approximated recursively via

$$\left. \begin{aligned} \hat{y}_{i+1} &= \hat{y}_i + \tau k_2 \\ k_1 &= f(t_i, \hat{y}_i) \\ k_2 &= f(t_i + \frac{\tau}{2}, \hat{y}_i + \frac{\tau}{2} k_1) \end{aligned} \right\} \quad \text{for } i = 0 \dots n-1 .$$

This second order Runge-Kutta method was first proposed by Runge [139] and is consistent of order $p = 2$ [66, p. 142].

We can extend this approach to the so called *s-stage Runge-Kutta method*, where the function f is evaluated s times at each iteration. Thereby the evaluations of f are chosen such that the corresponding s -stage Runge-Kutta method has a possibly high order of consistency.² In particular, for $s \leq 4$ the number of evaluations equals the order of consistency, i.e. $s = p$. However, if we require an order of consistency, which is higher than $p = 4$, the number of evaluations is larger than the order of consistency, i.e. $s > p$ [66, p. 162]. This is one reason why the *4-stage Runge-Kutta method* is a commonly used version and therefore also called the *classical Runge-Kutta method*. In particular, this Runge-Kutta version is often used in the finance literature in order to solve the Riccati equations (3.2), which determine the characteristic function of an affine process [72, 132]. For completeness, we also state the *4-stage Runge-Kutta method*, which is given by

$$\left. \begin{aligned} \hat{y}_{i+1} &= \hat{y}_i + \tau \left(\frac{1}{6} k_1 + \frac{1}{3} k_2 + \frac{1}{3} k_3 + \frac{1}{6} k_4 \right) \\ k_1 &= f(t_i, \hat{y}_i) \\ k_2 &= f(t_i + \frac{\tau}{2}, \hat{y}_i + \frac{\tau}{2} k_1) \\ k_3 &= f(t_i + \frac{\tau}{2}, \hat{y}_i + \frac{\tau}{2} k_2) \\ k_4 &= f(t_i + \tau, \hat{y}_i + \tau k_3) \end{aligned} \right\} \quad \text{for } i = 0, \dots, n-1 ,$$

and refer to Deuffhard and Borneman [66, p. 146ff.] for a derivation.

If the ODE (A.1) involves *stiffness*, the (explicit) Runge-Kutta method can get numerically inefficient.³ This is due to the fact that in such a case, we have to choose a very small step-size τ in relation to the smoothness of the solution of the ODE. Otherwise the Runge-Kutta method will be numerically unstable. In the context of affine processes, this problem can arise if we consider multiple mean-reverting processes with extremely differ-

²For more details about the s -stage Runge-Kutta see e.g. Deuffhard and Borneman [66, p. 142ff.]

³Following Gear [82, p. 209], “stiff differential equations are differential equations with greatly differing time constants”, so that some components of the solution decay much faster than the others [97, p. 102].

A. Supplementary theory and notation

ent mean reversion rates [97, p. 101]. As shown by Huang and Yu [97], it is numerically more efficient to use implicit Runge-Kutta methods in the presence of stiffness. For a detailed treatment of the implicit and explicit Runge-Kutta methods, we refer to Butcher [48].

B. Additional calculations

B.1. Supplement to the proof of Lemma 3.1

In the following we will show how to calculate the integral

$$\int_0^t \frac{1 - e^{-\sqrt{d}s}}{1 - ge^{-\sqrt{d}s}} ds . \quad (\text{B.1})$$

More precisely, we will first split the integral (B.1) into

$$\underbrace{\int_0^t \frac{1}{1 - ge^{-\sqrt{d}s}} ds}_{a)} - \underbrace{\int_0^t \frac{e^{-\sqrt{d}s}}{1 - ge^{-\sqrt{d}s}} ds}_{b)} ,$$

and then integrate each part separately:

ad a): By change of variables, where we set $u = e^{\sqrt{d}s}$, it follows that

$$\int_0^t \frac{1}{1 - ge^{-\sqrt{d}s}} ds = \frac{1}{\sqrt{d}} \int_1^{e^{\sqrt{d}t}} \frac{1}{u - g} du .$$

Another change of variable with $x = u - g$ leads to

$$\begin{aligned} \frac{1}{\sqrt{d}} \int_1^{e^{\sqrt{d}t}} \frac{1}{u - g} du &= \frac{1}{\sqrt{d}} \int_{1-g}^{e^{\sqrt{d}t}-g} \frac{1}{x} dx \\ &= \frac{1}{\sqrt{d}} \ln \left(\frac{e^{\sqrt{d}t} - g}{1 - g} \right) . \end{aligned}$$

ad b): The second integral can also be calculated by change of variable, where we first set $u = e^{-\sqrt{d}s}$ so that

$$\int_0^t \frac{e^{-\sqrt{d}s}}{1 - ge^{-\sqrt{d}s}} ds = -\frac{1}{\sqrt{d}} \int_1^{e^{-\sqrt{d}t}} \frac{1}{1 - gu} du .$$

B. Additional calculations

Again, by a second change of variables with $x = 1 - gu$, we get

$$\begin{aligned} -\frac{1}{\sqrt{d}} \int_1^{e^{-\sqrt{d}t}} \frac{1}{1-gu} du &= \frac{1}{\sqrt{d}g} \int_{1-g}^{1-ge^{-\sqrt{d}t}} \frac{1}{x} dx \\ &= \frac{1}{\sqrt{d}g} \ln \left(\frac{1-ge^{-\sqrt{d}t}}{1-g} \right) . \end{aligned}$$

Now, we only need to subtract $b)$ from $a)$, which is

$$\begin{aligned} \int_0^t \frac{1-e^{-\sqrt{d}s}}{1-ge^{-\sqrt{d}s}} ds &= \frac{1}{\sqrt{d}} \ln \left(\frac{e^{\sqrt{d}t}-g}{1-g} \right) - \frac{1}{\sqrt{d}g} \ln \left(\frac{1-ge^{-\sqrt{d}t}}{1-g} \right) \\ &= \frac{1}{\sqrt{d}} \ln \left(\frac{1}{e^{-\sqrt{d}t}} \frac{1-ge^{-\sqrt{d}t}}{1-g} \right) - \frac{1}{\sqrt{d}g} \ln \left(\frac{1-ge^{-\sqrt{d}t}}{1-g} \right) \\ &= \frac{1}{\sqrt{d}} \left(\sqrt{d}t + \ln \left(\frac{1-ge^{-\sqrt{d}t}}{1-g} \right) \right) - \frac{1}{\sqrt{d}g} \ln \left(\frac{1-ge^{-\sqrt{d}t}}{1-g} \right) \\ &= t + \frac{g-1}{\sqrt{d}g} \ln \left(\frac{1-ge^{-\sqrt{d}t}}{1-g} \right) . \end{aligned}$$

B.2. Discretization of the Cox-Ingersoll-Ross process

In accordance to van der Ploeg [134, p. 76 f.], we will derive a discretization for the Cox-Ingersoll-Ross process

$$d\sigma_t = \kappa(\phi - \sigma_t)dt + \lambda\sqrt{\sigma_t}dW_t . \quad (\text{B.2})$$

First, define

$$\tilde{\sigma}_t = e^{\kappa t} \sigma_t ,$$

and consider the differential of $\tilde{\sigma}_t$. By Itô's lemma,

$$\begin{aligned} d\tilde{\sigma}_t &= \kappa e^{\kappa t} \sigma_t dt + e^{\kappa t} \kappa(\phi - \sigma_t)dt + e^{\kappa t} \lambda \sqrt{\sigma_t} dW_t \\ &= e^{\kappa t} \kappa \phi dt + e^{\kappa t} \lambda \sqrt{\sigma_t} dW_t . \end{aligned} \quad (\text{B.3})$$

Next, we integrate (B.3) over the interval $[t_{i-1}, t_i]$ with $t_i - t_{i-1} = \Delta t$:

$$\tilde{\sigma}_i - \tilde{\sigma}_{i-1} = \kappa \phi \int_{t_{i-1}}^{t_i} e^{\kappa u} du + \lambda \int_{t_{i-1}}^{t_i} e^{\kappa u} \sqrt{\sigma_u} dW_u , \quad (\text{B.4})$$

where we set $i = t_i$. Finally, we multiply equation (B.4) by $e^{-\kappa t_i}$ and solve for σ_i , which is

$$\begin{aligned}
 \sigma_i &= e^{-\kappa \Delta t} \sigma_{i-1} + \kappa \phi \int_{t_{i-1}}^{t_i} e^{-\kappa(t_i-u)} du + \lambda \int_{t_{i-1}}^{t_i} e^{-\kappa(t_i-u)} \sqrt{\sigma_u} dW_u \\
 &= e^{-\kappa \Delta t} \sigma_{i-1} + \phi (1 - e^{-\kappa \Delta t}) + \lambda \int_{t_{i-1}}^{t_i} e^{-\kappa(t_i-u)} \sqrt{\sigma_u} dW_u \\
 &\approx e^{-\kappa \Delta t} \sigma_{i-1} + \phi (1 - e^{-\kappa \Delta t}) + e^{-\kappa \Delta t} \lambda \sqrt{\sigma_{i-1}} (W_i - W_{i-1}) \\
 &= e^{-\kappa \Delta t} \sigma_{i-1} + \phi (1 - e^{-\kappa \Delta t}) + e^{-\kappa \Delta t} \lambda \sqrt{\sigma_{i-1}} \sqrt{\Delta t} \epsilon_i ,
 \end{aligned} \tag{B.5}$$

where $\epsilon_i \sim \mathcal{N}(0, 1)$. In contrast to the classical Euler approximation, where (B.2) is approximated as

$$\sigma_i = \sigma_{i-1} + \kappa(\phi - \sigma_{i-1})\Delta t + \lambda\sqrt{\sigma_{i-1}}\sqrt{\Delta t}\epsilon_i ,$$

equation (B.5) captures the drift term in closed-form.

B.3. Return distribution

Let X be a random variable with characteristic function $\phi_X(u)$. Following Wu [157], the pdf of X is then given by

$$f_X(x) = \frac{1}{\pi} \int_0^\infty e^{-iux} \phi_X(u) du , \tag{B.6}$$

and the cdf of X by

$$F_X(x) = \frac{1}{2} + \frac{1}{2\pi} \int_0^\infty \frac{e^{iux} \phi_X(-u) - e^{-iux} \phi_X(u)}{iu} du . \tag{B.7}$$

Now, the continuous return of the asset price process S_t within the period $[t, t+\tau]$ is defined as

$$r_{t,t+\tau} = \ln(S_{t+\tau}) - \ln(S_t) , \tag{B.8}$$

so that

$$\begin{aligned}
 F_{r_{t,t+\tau}}(x) &= \mathbb{P}[r_{t,t+\tau} \leq x] \\
 &= \mathbb{P}[\ln(S_{t+\tau}) - \ln(S_t) \leq x] \\
 &= \mathbb{P}[\ln(S_{t+\tau}) \leq x + \ln(S_t)] \\
 &= F_{\ln(S_{t+\tau})}(x + \ln(S_t)) .
 \end{aligned} \tag{B.9}$$

For affine option pricing models, we generally only know the characteristic function of the log asset price $\ln(S_t)$. Hence, the return pdf and cdf can be recovered via equation (B.6) and (B.7), respectively, together with equation (B.9).

C. Additional figures and tables

C.1. Value-at-Risk forecasts for CDS spreads

The following Tables C.1 - C.3 are complementary to the empirical study of Chapter 2.5.4 and summarize the number of *Value-at-Risk* (VaR) exceptions out of 171 VaR forecasts for the 3, 4 and 5 year CDS spreads for various companies. We therefore supposed a forecasting horizon of one week and a confidence level of 95%. In order to test if the VaR with a confidence level of α is violated more or less than $(1 - \alpha) \cdot 100\%$ of the time, we apply the LR test of Kupiec [118]. The reported p -values correspond to the test statistic LR (6.5), where $0.01 \leq p < 0.05$ is indicated by *, $0.05 \leq p < 0.1$ by ** and $p \geq 0.1$ by ***.

Table C.1.: Value-at-Risk forecasts for 3-year CDS spreads.

Company	Utility transformation	Calibration transformation	No transformation
<i>Adecco</i>	3*	2*	3*
<i>p-value</i>	(0.03)	(0.01)	(0.03)
<i>BMW</i>	9***	6***	8***
<i>p-value</i>	(0.88)	(0.35)	(0.85)
<i>Deutsche Telekom</i>	2*	5***	2*
<i>p-value</i>	(0.01)	(0.18)	(0.01)
<i>Linde</i>	4**	3*	5***
<i>p-value</i>	(0.08)	(0.03)	(0.18)
<i>Lufthansa</i>	3*	3*	3*
<i>p-value</i>	(0.03)	(0.03)	(0.03)
<i>Thyssen</i>	7***	5***	8***
<i>p-value</i>	(0.58)	(0.18)	(0.85)
<i>Valeo</i>	8***	5***	7***
<i>p-value</i>	(0.85)	(0.18)	(0.58)
<i>VW</i>	6***	5***	6***
<i>p-value</i>	(0.35)	(0.18)	(0.35)

Table C.2.: Value-at-Risk forecasts for 4-year CDS spreads.

Company	Utility transformation	Calibration transformation	No transformation
<i>Adecco</i>	3*	4**	4**
<i>p-value</i>	(0.03)	(0.08)	(0.08)
<i>BMW</i>	12***	8***	12***
<i>p-value</i>	(0.25)	(0.85)	(0.25)
<i>Deutsche Telekom</i>	3*	8***	5***
<i>p-value</i>	(0.03)	(0.85)	(0.18)
<i>Linde</i>	7***	3*	8***
<i>p-value</i>	(0.58)	(0.03)	(0.85)
<i>Lufthansa</i>	4**	3*	4**
<i>p-value</i>	(0.08)	(0.03)	(0.08)
<i>Thyssen</i>	9***	6***	9***
<i>p-value</i>	(0.88)	(0.35)	(0.88)
<i>Valeo</i>	13***	10***	13***
<i>p-value</i>	(0.15)	(0.62)	(0.15)
<i>VW</i>	7***	8***	9***
<i>p-value</i>	(0.58)	(0.85)	(0.88)

Table C.3.: Value-at-Risk forecasts for 5-year CDS spreads.

Company	Utility transformation	Calibration transformation	No transformation
<i>Adecco</i>	7***	7***	7***
<i>p-value</i>	(0.58)	(0.58)	(0.58)
<i>BMW</i>	20	15*	21
<i>p-value</i>	(0.00)	(0.04)	(0.00)
<i>Deutsche Telekom</i>	12***	18	13***
<i>p-value</i>	(0.25)	(0.00)	(0.15)
<i>Linde</i>	13***	10***	13***
<i>p-value</i>	(0.15)	(0.62)	(0.15)
<i>Lufthansa</i>	6***	6***	7***
<i>p-value</i>	(0.35)	(0.35)	(0.58)
<i>Thyssen</i>	15*	13***	16*
<i>p-value</i>	(0.04)	(0.15)	(0.02)
<i>Valeo</i>	15*	17	16*
<i>p-value</i>	(0.04)	(0.00)	(0.02)
<i>VW</i>	14**	15*	14**
<i>p-value</i>	(0.08)	(0.04)	(0.08)

C.2. Data analysis

The following figures are complementary to the descriptive analysis of the S&P 500 index and the corresponding call options performed in Chapter 6.3. First, Figure C.1(a) illustrates the annualized dividend payments of the S&P 500 index, which are provided by *OptionMetrics Ivy DB US*. Furthermore, *OptionMetrics Ivy DB US* also provides the “riskfree rate” for the US market, which is illustrated in Figure C.1(b).¹ In order to account for moneyness and maturity effects in the estimation, we scaled the selected call option prices by their corresponding Black-Scholes vega. The resulting time series of the vega-scaled option prices are illustrated in Figure C.2.

(a)

(b)

Figure C.1.: Dividends for the S&P 500 index (a) and riskfree rate for the U.S. market (b).

¹The “riskfree rate” provided by *OptionMetrics Ivy DB US* is derived from the LIBOR (*London Interbank Offered Rate*) and settlement prices of Euro/Dollar futures.

Figure C.2.: Vega-scaled S&P 500 index options.

C.3. Estimation results for the Heston model

The following figures and tables complement the estimation results of the Heston model based on the unscented Kalman filter, which were presented in Chapter 6.4.1. First, Table C.4 summarizes the mean and standard deviation of the measurement errors for the Heston model. Moreover, Figure C.3 illustrates the time series of the single measurement errors for the Heston model. In addition, Tables C.5 and C.6 summarize the estimation results and corresponding t -statistics for the parameters of the Heston model before and during the subprime mortgage crises.

Table C.4.: Measurement errors for the Heston model.

	ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5	ϵ_6	ϵ_7	ϵ_8
Coef.	0.004	0.008	0.007	0.003	0.000	0.004	0.005	0.002
SD	(0.034)	(0.026)	(0.022)	(0.021)	(0.032)	(0.024)	(0.020)	(0.020)

Notes: This table summarizes the mean and standard deviation of the single measurement errors of the state-space representation for the Heston model.

Figure C.3.: Measurement errors for the Heston model.

Table C.5.: Evolution of the Heston model parameters (Part 1/2).

Date	ϕ^*	κ^*	λ	ρ	γ	ω	α
28-Dec-2005	0.064	0.989	0.273	-0.828	-3.629	$2.869 \cdot 10^{-5}$	0.121
25-Jan-2006	0.065	0.945	0.269	-0.828	-3.578	$2.780 \cdot 10^{-5}$	0.126
22-Feb-2006	0.066	0.909	0.266	-0.821	-3.622	$2.683 \cdot 10^{-5}$	0.131
22-Mar-2006	0.067	0.884	0.266	-0.811	-3.648	$2.618 \cdot 10^{-5}$	0.134
19-Apr-2006	0.069	0.828	0.262	-0.806	-3.588	$2.565 \cdot 10^{-5}$	0.140
17-May-2006	0.069	0.796	0.258	-0.806	-3.533	$2.571 \cdot 10^{-5}$	0.141
14-Jun-2006	0.069	0.799	0.257	-0.812	-3.483	$2.558 \cdot 10^{-5}$	0.140
12-Jul-2006	0.069	0.807	0.259	-0.814	-3.560	$2.595 \cdot 10^{-5}$	0.134
09-Aug-2006	0.069	0.802	0.258	-0.819	-3.580	$2.545 \cdot 10^{-5}$	0.135
06-Sep-2006	0.069	0.808	0.258	-0.819	-3.631	$2.529 \cdot 10^{-5}$	0.134
04-Oct-2006	0.069	0.804	0.257	-0.823	-3.656	$2.482 \cdot 10^{-5}$	0.135
01-Nov-2006	0.069	0.776	0.253	-0.824	-3.633	$2.431 \cdot 10^{-5}$	0.138
29-Nov-2006	0.070	0.751	0.250	-0.828	-3.596	$2.388 \cdot 10^{-5}$	0.141
27-Dec-2006	0.070	0.759	0.249	-0.836	-3.604	$2.347 \cdot 10^{-5}$	0.143
24-Jan-2007	0.070	0.747	0.246	-0.843	-3.609	$2.308 \cdot 10^{-5}$	0.144
21-Feb-2007	0.071	0.697	0.234	-0.859	-2.732	$3.453 \cdot 10^{-5}$	0.108
21-Mar-2007	0.072	0.683	0.239	-0.849	-2.676	$3.481 \cdot 10^{-5}$	0.105
18-Apr-2007	0.071	0.685	0.240	-0.852	-2.714	$3.493 \cdot 10^{-5}$	0.102
16-May-2007	0.072	0.672	0.238	-0.854	-2.732	$3.413 \cdot 10^{-5}$	0.106
13-Jun-2007	0.072	0.675	0.237	-0.861	-2.712	$3.382 \cdot 10^{-5}$	0.106
11-Jul-2007	0.072	0.681	0.237	-0.867	-2.676	$3.397 \cdot 10^{-5}$	0.105
08-Aug-2007	0.071	0.686	0.238	-0.869	-2.580	$3.420 \cdot 10^{-5}$	0.101
05-Sep-2007	0.067	0.752	0.245	-0.858	-2.353	$3.698 \cdot 10^{-5}$	0.090
03-Oct-2007	0.067	0.757	0.247	-0.857	-2.424	$3.736 \cdot 10^{-5}$	0.088
31-Oct-2007	0.069	0.746	0.248	-0.859	-2.474	$3.762 \cdot 10^{-5}$	0.089
28-Nov-2007	0.068	0.754	0.252	-0.851	-2.303	$3.815 \cdot 10^{-5}$	0.084
26-Dec-2007	0.070	0.732	0.251	-0.854	-2.416	$3.853 \cdot 10^{-5}$	0.088
23-Jan-2008	0.070	0.741	0.255	-0.849	-2.409	$3.882 \cdot 10^{-5}$	0.086
20-Feb-2008	0.069	0.761	0.258	-0.845	-2.502	$3.965 \cdot 10^{-5}$	0.079
19-Mar-2008	0.069	0.753	0.259	-0.846	-2.464	$3.957 \cdot 10^{-5}$	0.080
16-Apr-2008	0.071	0.729	0.256	-0.853	-2.254	$3.986 \cdot 10^{-5}$	0.082
14-May-2008	0.072	0.716	0.254	-0.860	-2.150	$4.051 \cdot 10^{-5}$	0.082
11-Jun-2008	0.074	0.690	0.253	-0.863	-2.058	$4.044 \cdot 10^{-5}$	0.086
09-Jul-2008	0.075	0.681	0.253	-0.867	-1.962	$4.062 \cdot 10^{-5}$	0.084
06-Aug-2008	0.076	0.672	0.252	-0.870	-2.006	$4.165 \cdot 10^{-5}$	0.082
03-Sep-2008	0.077	0.651	0.251	-0.874	-1.933	$4.198 \cdot 10^{-5}$	0.083
01-Oct-2008	0.079	0.643	0.263	-0.843	-3.215	$3.720 \cdot 10^{-5}$	0.134
29-Oct-2008	0.081	0.624	0.273	-0.813	-3.343	$3.337 \cdot 10^{-5}$	0.174
26-Nov-2008	0.081	0.629	0.279	-0.801	-3.047	$3.291 \cdot 10^{-5}$	0.181
24-Dec-2008	0.084	0.595	0.276	-0.805	-2.846	$3.237 \cdot 10^{-5}$	0.188
21-Jan-2009	0.089	0.552	0.275	-0.806	-2.601	$3.195 \cdot 10^{-5}$	0.194
18-Feb-2009	0.092	0.533	0.277	-0.803	-2.617	$3.194 \cdot 10^{-5}$	0.194
18-Mar-2009	0.090	0.544	0.278	-0.805	-2.609	$3.217 \cdot 10^{-5}$	0.192
15-Apr-2009	0.094	0.515	0.279	-0.799	-3.036	$3.186 \cdot 10^{-5}$	0.201
13-May-2009	0.101	0.474	0.279	-0.800	-3.071	$3.182 \cdot 10^{-5}$	0.202
10-Jun-2009	0.104	0.459	0.278	-0.802	-2.988	$3.273 \cdot 10^{-5}$	0.203
08-Jul-2009	0.108	0.436	0.277	-0.801	-2.965	$3.281 \cdot 10^{-5}$	0.204
05-Aug-2009	0.112	0.415	0.275	-0.803	-2.893	$3.272 \cdot 10^{-5}$	0.207
02-Sep-2009	0.120	0.381	0.272	-0.807	-2.678	$3.272 \cdot 10^{-5}$	0.215
30-Sep-2009	0.124	0.367	0.272	-0.805	-2.742	$3.276 \cdot 10^{-5}$	0.215

Notes: This table summarizes the estimation results of the parameters for the Heston model before and during the subprime mortgage crises. We therefore repeatedly estimate the Heston model based on the data set described in Chapter 6.3. More precisely, we increase the sample size after each estimation by four additional observation days, which approximately cover one month. While the underlying sample period always starts at November 3, 1999, the end of the sample period ranges from December 28, 2005 to September 30, 2009.

Table C.6.: Evolution of the Heston model parameters (Part 2/2).

Date	ϕ^*	κ^*	λ	ρ	γ	ω	α
28-Dec-2005	26.512	16.817	24.944	-26.216	-8.060	15.266	10.160
25-Jan-2006	28.022	14.623	24.966	-28.453	-8.549	15.201	10.172
22-Feb-2006	14.860	10.153	26.243	-25.229	-10.847	11.327	8.924
22-Mar-2006	37.842	44.585	23.254	-17.004	-19.490	74.021	13.153
19-Apr-2006	20.724	14.149	27.227	-28.036	-11.407	13.405	11.350
17-May-2006	22.505	15.785	27.157	-28.092	-11.551	16.209	12.722
14-Jun-2006	0.515	0.463	1.627	-0.665	-0.399	0.357	4.217
12-Jul-2006	23.041	14.877	24.019	-28.021	-5.178	16.454	11.404
09-Aug-2006	15.920	9.824	26.174	-27.331	-10.429	19.057	12.791
06-Sep-2006	18.298	9.840	18.217	-13.982	-14.291	11.044	10.119
04-Oct-2006	22.000	15.250	29.116	-30.366	-20.003	16.134	12.641
01-Nov-2006	25.033	15.805	30.589	-67.323	-6.992	14.625	10.908
29-Nov-2006	13.560	9.184	34.753	-60.507	-21.421	17.053	12.359
27-Dec-2006	6.695	5.009	1.862	-1.579	-5.728	1.098	2.060
24-Jan-2007	20.951	14.890	27.843	-29.107	-9.122	15.397	12.733
21-Feb-2007	13.534	9.212	32.064	-29.307	-11.366	24.194	9.828
21-Mar-2007	17.579	11.947	24.123	-17.185	-36.218	12.187	7.361
18-Apr-2007	19.188	13.546	31.155	-28.681	-21.458	16.970	10.179
16-May-2007	19.547	14.668	29.429	-29.528	-7.693	18.064	9.547
13-Jun-2007	22.735	15.713	29.292	-28.184	-15.235	69.962	15.062
11-Jul-2007	23.373	15.847	26.429	-28.590	-19.257	40.919	12.081
08-Aug-2007	18.924	13.331	28.984	-28.200	-9.763	21.550	10.492
05-Sep-2007	12.152	8.798	27.742	-23.970	-3.100	19.024	9.753
03-Oct-2007	6.402	3.662	9.386	-21.991	-5.404	6.758	11.897
31-Oct-2007	20.737	14.453	29.024	-28.494	-5.503	20.261	9.093
28-Nov-2007	47.832	38.785	36.789	-51.415	-34.999	19.310	9.574
26-Dec-2007	19.762	13.393	31.751	-30.341	-10.324	18.445	8.511
23-Jan-2008	29.262	25.274	33.856	-31.999	-19.742	23.515	10.107
20-Feb-2008	3.696	2.434	36.217	-27.575	-3.055	2.954	1.431
19-Mar-2008	13.852	10.872	15.349	-8.378	-5.955	5.394	3.160
16-Apr-2008	26.556	19.734	31.949	-30.466	-4.783	24.868	10.731
14-May-2008	19.748	13.664	31.256	-30.736	-4.911	19.248	7.636
11-Jun-2008	19.619	13.914	35.557	-32.751	-9.181	26.556	9.420
09-Jul-2008	18.307	12.776	31.250	-30.085	-4.948	20.121	8.664
06-Aug-2008	17.282	12.234	38.827	-92.310	-28.558	22.190	8.495
03-Sep-2008	18.729	13.574	32.826	-31.535	-4.639	22.237	9.000
01-Oct-2008	53.531	56.741	37.721	-36.145	-14.025	25.563	13.444
29-Oct-2008	12.729	8.456	10.830	-19.053	-2.176	18.798	16.141
26-Nov-2008	17.314	13.033	32.051	-32.603	-9.578	37.719	19.230
24-Dec-2008	23.936	18.378	36.100	-34.749	-21.723	25.474	17.493
21-Jan-2009	13.261	10.065	38.023	-49.007	-12.112	20.063	17.326
18-Feb-2009	18.204	12.143	33.881	-31.222	-14.698	15.885	15.768
18-Mar-2009	14.064	10.769	34.921	-35.212	-11.512	21.036	17.323
15-Apr-2009	25.567	27.060	28.715	-22.732	-13.388	22.897	18.058
13-May-2009	17.685	14.363	42.992	-39.376	-9.696	21.371	16.340
10-Jun-2009	21.183	18.747	52.528	-49.294	-36.170	36.293	23.551
08-Jul-2009	4.485	3.668	33.966	-28.384	-6.596	18.436	17.004
05-Aug-2009	11.971	5.766	6.850	-3.774	-5.458	1.332	3.127
02-Sep-2009	14.207	12.029	34.965	-33.748	-11.915	22.721	18.502
30-Sep-2009	10.275	8.647	33.389	-28.849	-9.881	22.960	20.614

Notes: This table summarizes the t -statistics of the estimation results for the Heston model, which are listed in Table C.5. The reported t -statistics are calculated as $t = (\hat{\theta} - \theta_0)/SE(\hat{\theta})$, where we set the null hypothesis to $\theta_0 = 0$. Hence large t -statistics indicate that the corresponding parameter is significantly different from zero. In particular, $|t| > 1.960$ corresponds to a 5% significance level, $|t| > 2.576$ to a 1% significance level and $|t| > 3.291$ to a 0.1% significance level.

C.4. Estimation results for the 2-Factor SV model

The following figures and tables complement the estimation results of the 2-Factor SV model based on the unscented Kalman filter, which were presented in Chapter 6.4.2. First, Table C.7 summarizes the mean and standard deviation of the measurement errors for the 2-Factor SV model. Moreover, Figure C.4 illustrates the time series of the single measurement errors for the 2-Factor SV model. In addition, Tables C.8 and C.9 summarize the estimation results and corresponding t -statistics for the parameters of the 2-Factor SV model before and during the subprime mortgage crises.

Table C.7.: Measurement errors for the 2-Factor SV model.

	ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5	ϵ_6	ϵ_7	ϵ_8
Coef.	0.009	0.009	0.007	0.006	0.007	0.008	0.006	0.005
SD	(0.037)	(0.030)	(0.025)	(0.023)	(0.035)	(0.028)	(0.024)	(0.022)

Notes: This table summarizes the mean and standard deviation of the single measurement errors of the state-space representation for the 2-Factor SV model.

Figure C.4.: Measurement errors of the 2-Factor SV model.

Table C.8.: Evolution of the 2-Factor SV model parameters (Part 1/2).

Date	κ_1	κ_1^*	κ_2	ϕ_1	ϕ_2	λ_1	λ_2	ρ	ω	α
28-Dec-2005	23.760	6.345	1.657	0.021	0.035	0.806	0.199	-0.634	$0.500 \cdot 10^{-5}$	0.010
25-Jan-2006	23.895	6.332	1.639	0.021	0.035	0.804	0.197	-0.633	$0.494 \cdot 10^{-5}$	0.010
22-Feb-2006	23.987	6.331	1.632	0.021	0.035	0.804	0.197	-0.632	$0.490 \cdot 10^{-5}$	0.010
22-Mar-2006	24.238	6.313	1.652	0.020	0.035	0.804	0.197	-0.631	$0.487 \cdot 10^{-5}$	0.010
19-Apr-2006	24.371	6.276	1.632	0.020	0.035	0.800	0.195	-0.631	$0.483 \cdot 10^{-5}$	0.010
17-May-2006	24.454	6.258	1.611	0.020	0.034	0.797	0.193	-0.630	$0.481 \cdot 10^{-5}$	0.010
14-Jun-2006	24.310	6.237	1.605	0.020	0.034	0.794	0.194	-0.633	$0.479 \cdot 10^{-5}$	0.010
12-Jul-2006	24.644	6.184	1.618	0.020	0.034	0.796	0.194	-0.633	$0.478 \cdot 10^{-5}$	0.009
09-Aug-2006	24.591	6.183	1.597	0.020	0.034	0.791	0.193	-0.635	$0.475 \cdot 10^{-5}$	0.009
06-Sep-2006	24.504	6.186	1.582	0.020	0.034	0.788	0.191	-0.637	$0.473 \cdot 10^{-5}$	0.009
04-Oct-2006	24.084	6.204	1.557	0.020	0.034	0.779	0.188	-0.641	$0.471 \cdot 10^{-5}$	0.010
01-Nov-2006	24.145	6.194	1.542	0.019	0.033	0.776	0.187	-0.641	$0.466 \cdot 10^{-5}$	0.010
29-Nov-2006	24.167	6.184	1.533	0.019	0.033	0.773	0.186	-0.641	$0.460 \cdot 10^{-5}$	0.010
27-Dec-2006	23.903	6.222	1.507	0.019	0.033	0.768	0.184	-0.644	$0.458 \cdot 10^{-5}$	0.010
24-Jan-2007	23.840	6.211	1.497	0.019	0.033	0.763	0.182	-0.646	$0.454 \cdot 10^{-5}$	0.010
21-Feb-2007	23.147	5.830	1.437	0.021	0.033	0.752	0.196	-0.641	$0.451 \cdot 10^{-5}$	0.014
21-Mar-2007	23.635	5.629	1.533	0.022	0.033	0.756	0.209	-0.639	$0.462 \cdot 10^{-5}$	0.013
18-Apr-2007	23.576	5.648	1.509	0.021	0.033	0.753	0.208	-0.640	$0.462 \cdot 10^{-5}$	0.013
16-May-2007	23.518	5.668	1.485	0.021	0.033	0.749	0.206	-0.641	$0.463 \cdot 10^{-5}$	0.013
13-Jun-2007	23.496	5.696	1.479	0.021	0.033	0.748	0.205	-0.641	$0.471 \cdot 10^{-5}$	0.013
11-Jul-2007	23.263	5.817	1.450	0.021	0.033	0.750	0.203	-0.644	$0.479 \cdot 10^{-5}$	0.013
08-Aug-2007	23.532	5.800	1.469	0.021	0.032	0.755	0.204	-0.644	$0.475 \cdot 10^{-5}$	0.013
05-Sep-2007	23.719	5.811	1.480	0.021	0.032	0.766	0.204	-0.643	$0.485 \cdot 10^{-5}$	0.013
03-Oct-2007	23.829	5.840	1.497	0.021	0.032	0.769	0.204	-0.644	$0.484 \cdot 10^{-5}$	0.014
31-Oct-2007	24.126	5.843	1.510	0.021	0.032	0.776	0.205	-0.644	$0.485 \cdot 10^{-5}$	0.014
28-Nov-2007	23.694	5.880	1.499	0.022	0.032	0.788	0.206	-0.643	$0.489 \cdot 10^{-5}$	0.014
26-Dec-2007	24.139	5.732	1.556	0.021	0.032	0.783	0.214	-0.645	$0.597 \cdot 10^{-5}$	0.013
23-Jan-2008	24.382	5.724	1.570	0.021	0.032	0.791	0.219	-0.644	$0.599 \cdot 10^{-5}$	0.012
20-Feb-2008	24.693	5.708	1.583	0.021	0.032	0.797	0.220	-0.643	$0.600 \cdot 10^{-5}$	0.012
19-Mar-2008	24.663	5.707	1.598	0.021	0.032	0.797	0.221	-0.644	$0.600 \cdot 10^{-5}$	0.012
16-Apr-2008	24.992	5.813	1.617	0.021	0.032	0.808	0.222	-0.644	$0.618 \cdot 10^{-5}$	0.011
14-May-2008	25.190	5.799	1.625	0.021	0.032	0.811	0.221	-0.645	$0.608 \cdot 10^{-5}$	0.012
11-Jun-2008	25.210	5.791	1.629	0.021	0.032	0.810	0.222	-0.646	$0.611 \cdot 10^{-5}$	0.012
09-Jul-2008	25.408	5.769	1.646	0.021	0.032	0.816	0.223	-0.645	$0.612 \cdot 10^{-5}$	0.011
06-Aug-2008	25.526	5.756	1.644	0.021	0.032	0.818	0.222	-0.645	$0.606 \cdot 10^{-5}$	0.011
03-Sep-2008	25.584	5.821	1.646	0.022	0.032	0.821	0.221	-0.646	$0.609 \cdot 10^{-5}$	0.011
01-Oct-2008	24.118	6.508	1.505	0.022	0.033	0.837	0.214	-0.652	$0.582 \cdot 10^{-5}$	0.015
29-Oct-2008	22.209	7.115	1.408	0.023	0.034	0.854	0.209	-0.658	$0.569 \cdot 10^{-5}$	0.017
26-Nov-2008	23.517	7.700	1.355	0.023	0.035	0.906	0.218	-0.655	$0.563 \cdot 10^{-5}$	0.017
24-Dec-2008	22.881	7.888	1.339	0.023	0.036	0.910	0.227	-0.660	$0.563 \cdot 10^{-5}$	0.018
21-Jan-2009	21.760	7.917	1.309	0.024	0.036	0.900	0.230	-0.665	$0.565 \cdot 10^{-5}$	0.019
18-Feb-2009	21.815	7.864	1.335	0.024	0.036	0.906	0.232	-0.664	$0.560 \cdot 10^{-5}$	0.020
18-Mar-2009	21.638	7.912	1.327	0.024	0.036	0.909	0.231	-0.665	$0.564 \cdot 10^{-5}$	0.019
15-Apr-2009	21.808	8.031	1.344	0.024	0.036	0.920	0.233	-0.665	$0.569 \cdot 10^{-5}$	0.019
13-May-2009	21.820	7.993	1.336	0.024	0.036	0.924	0.234	-0.665	$0.567 \cdot 10^{-5}$	0.019
10-Jun-2009	22.280	7.995	1.352	0.024	0.036	0.930	0.235	-0.664	$0.577 \cdot 10^{-5}$	0.018
08-Jul-2009	22.042	7.960	1.329	0.024	0.036	0.937	0.233	-0.664	$0.590 \cdot 10^{-5}$	0.018
05-Aug-2009	22.138	7.905	1.319	0.024	0.036	0.937	0.233	-0.663	$0.591 \cdot 10^{-5}$	0.018
02-Sep-2009	22.298	8.007	1.330	0.025	0.036	0.945	0.233	-0.664	$0.604 \cdot 10^{-5}$	0.017
30-Sep-2009	22.384	7.982	1.332	0.025	0.036	0.945	0.234	-0.664	$0.602 \cdot 10^{-5}$	0.017

Notes: This table summarizes the estimation results of the parameters for the 2-Factor SV model before and during the subprime mortgage crises. We therefore repeatedly estimate the 2-Factor SV model based on the data set described in Chapter 6.3. More precisely, we increase the sample size after each estimation by four additional observation days, which approximately cover one month. While the underlying sample period always starts at November 3, 1999, the end of the sample period ranges from December 28, 2005 to September 30, 2009.

Table C.9.: Evolution of the 2-Factor SV model parameters (Part 2/2).

Date	κ_1	κ_1^*	κ_2	ϕ_1	ϕ_2	λ_1	λ_2	ρ	ω	α
28-Dec-2005	20.270	22.581	21.103	23.050	39.578	46.028	21.517	-67.697	19.317	8.352
25-Jan-2006	22.993	28.865	41.235	23.340	43.092	48.701	21.176	-82.264	19.482	8.867
22-Feb-2006	25.617	35.574	26.180	21.861	43.154	34.796	21.234	-78.396	20.015	8.493
22-Mar-2006	43.000	45.887	20.886	24.236	42.945	68.800	21.409	-92.401	19.672	9.872
19-Apr-2006	19.058	16.723	22.506	21.665	41.530	28.433	20.098	-74.300	16.417	8.535
17-May-2006	15.296	24.041	40.909	23.545	42.122	28.818	22.025	-69.394	19.395	8.169
14-Jun-2006	84.928	51.185	116.185	24.217	44.130	49.819	20.733	-101.632	23.170	9.963
12-Jul-2006	36.662	47.727	85.911	23.905	49.279	59.694	24.282	-97.445	27.544	9.560
09-Aug-2006	10.864	15.208	8.005	23.513	33.989	17.442	17.699	-49.268	17.463	9.600
06-Sep-2006	10.829	26.892	20.306	24.991	43.331	23.180	20.144	-60.042	26.064	8.832
04-Oct-2006	26.346	20.962	7.867	11.105	15.510	59.521	23.355	-43.854	5.549	2.273
01-Nov-2006	15.333	47.444	36.915	24.484	44.141	31.165	21.955	-80.595	22.007	9.143
29-Nov-2006	20.437	28.302	30.448	24.644	44.261	31.570	21.843	-81.499	20.889	10.071
27-Dec-2006	16.189	22.048	17.456	24.528	38.647	30.630	21.093	-71.883	20.171	8.950
24-Jan-2007	96.632	89.720	70.276	24.940	47.975	49.047	23.206	-101.978	20.571	10.241
21-Feb-2007	68.478	21.824	15.342	30.557	32.914	60.381	19.954	-97.789	17.797	10.554
21-Mar-2007	37.502	47.784	21.559	28.334	41.452	41.050	22.196	-104.210	14.019	10.698
18-Apr-2007	38.697	8.630	8.614	22.887	34.034	32.062	10.896	-98.613	19.947	5.419
16-May-2007	20.919	18.702	22.974	31.574	36.069	33.903	24.372	-83.209	22.502	11.654
13-Jun-2007	27.705	31.615	19.172	30.269	37.621	58.503	22.047	-93.363	18.667	11.026
11-Jul-2007	15.554	31.544	20.501	28.181	35.159	24.465	21.230	-59.223	20.210	10.418
08-Aug-2007	33.017	18.213	14.001	29.170	37.247	34.288	19.682	-102.497	19.254	11.711
05-Sep-2007	37.119	28.948	24.714	29.144	39.277	53.343	22.209	-107.546	20.405	9.770
03-Oct-2007	49.881	34.697	40.910	30.210	41.685	40.952	24.698	-92.066	21.817	11.862
31-Oct-2007	33.828	57.600	38.792	30.725	41.147	67.123	25.283	-107.394	25.467	13.116
28-Nov-2007	18.749	24.835	19.287	31.753	38.540	32.370	21.936	-79.006	19.843	11.084
26-Dec-2007	18.765	17.183	15.682	26.958	33.599	29.288	21.727	-73.957	20.757	10.789
23-Jan-2008	27.439	40.314	57.121	29.270	37.820	40.852	24.336	-87.381	25.828	11.327
20-Feb-2008	39.049	9.706	6.847	29.710	18.996	58.663	22.391	-54.211	9.679	7.976
19-Mar-2008	57.187	32.302	33.261	28.015	39.286	58.007	27.934	-119.289	23.704	10.839
16-Apr-2008	13.763	39.933	9.649	27.225	41.630	39.170	23.812	-97.218	12.896	6.162
14-May-2008	12.219	9.058	6.367	28.183	26.309	25.469	12.093	-37.485	9.271	9.794
11-Jun-2008	18.303	19.424	15.808	28.027	37.126	31.665	21.677	-78.108	19.506	9.653
09-Jul-2008	10.737	7.776	6.078	19.810	30.502	10.312	14.049	-46.083	22.994	8.612
06-Aug-2008	20.675	73.130	23.124	28.898	40.250	36.263	26.671	-83.307	19.558	9.952
03-Sep-2008	13.249	8.754	7.576	25.291	32.115	14.717	13.682	-63.757	27.462	10.149
01-Oct-2008	20.923	21.078	17.347	29.088	32.574	31.266	22.793	-78.713	15.634	11.525
29-Oct-2008	35.246	44.141	129.278	29.212	40.547	96.272	26.245	-136.590	28.706	15.270
26-Nov-2008	32.629	24.975	25.096	26.248	34.956	57.415	25.401	-96.344	20.564	11.835
24-Dec-2008	7.197	2.447	2.895	3.954	6.745	7.245	12.259	-26.778	5.986	8.311
21-Jan-2009	38.123	25.403	12.870	27.026	34.527	41.545	22.812	-93.205	24.472	12.440
18-Feb-2009	31.020	30.955	59.515	28.923	44.011	43.077	26.589	-92.924	23.895	14.348
18-Mar-2009	22.688	30.856	18.411	27.773	36.876	34.529	25.794	-86.156	19.683	12.820
15-Apr-2009	4.072	5.079	0.855	8.606	6.517	4.026	26.426	-16.391	3.278	3.256
13-May-2009	35.275	38.038	32.444	26.985	41.940	42.575	27.250	-97.853	20.009	12.924
10-Jun-2009	36.153	39.324	13.623	26.993	44.176	46.392	20.621	-118.335	23.178	12.440
08-Jul-2009	4.400	6.211	30.124	14.022	19.673	6.434	27.221	-33.235	9.720	2.802
05-Aug-2009	33.649	27.428	24.924	26.886	37.852	40.348	26.384	-93.624	21.447	13.443
02-Sep-2009	54.110	34.750	27.101	27.563	42.127	48.277	26.618	-102.420	21.699	14.628
30-Sep-2009	95.891	122.023	33.369	28.754	43.617	67.213	29.820	-114.589	41.989	15.726

Notes: This table summarizes the t -statistics of the estimation results for the 2-Factor SV model, which are listed in Table C.8. The reported t -statistics are calculated as $t = (\hat{\theta} - \theta_0)/SE(\hat{\theta})$, where we set the null hypothesis to $\theta_0 = 0$. Hence, large t -statistics indicate that the corresponding parameter is significantly different from zero. In particular, $|t| > 1.960$ corresponds to a 5% significance level, $|t| > 2.576$ to a 1% significance level and $|t| > 3.291$ to a 0.1% significance level.

D. A modified Heston model

Similarly as Bollerslev et al. [36], we modify the Heston model in order to account for time-varying risk aversion. But in contrast to them, we will model the variance risk premium indirectly by assuming that the risk-neutral long-term average of the variance ϕ^* is not constant but time-varying. The reason for this different approach is that the supposed variance risk premium dynamics γ_t of Bollerslev et al. [36] together with Heston's [95] original specification of κ^* and ϕ^* results in¹

$$\kappa_t^* = \kappa + \gamma_t \quad \text{and} \quad \phi_t^* = \kappa\phi/(\kappa + \gamma_t) . \quad (\text{D.1})$$

However, the specification (D.1) does not ensure that the corresponding risk-neutral variance dynamics are always well defined. In particular, if $\gamma_t < -\kappa$, the parameters κ_t^* and ϕ_t^* get negative. To circumvent this problem, we suppose to directly model the risk-neutral long-term average of the variance ϕ^* as a Cox-Ingersoll-Ross process, so that the risk-neutral asset price dynamics of the modified Heston model are given by

$$\begin{aligned} dS_t &= rS_t dt + \sqrt{\sigma_t} S_t d\widetilde{W}_{1,t} , \\ d\sigma_t &= \kappa_t^*(\phi_t^* - \sigma_t)dt + \lambda_1 \sqrt{\sigma_t} d\widetilde{W}_{2,t} , \\ d\phi_t^* &= \kappa_2^*(\phi - \phi_t^*)dt + \lambda_2 \sqrt{\phi_t^*} d\widetilde{W}_{3,t} , \end{aligned}$$

where $\widetilde{W}_{3,t}$ is uncorrelated to $\widetilde{W}_{1,t}$ and $\widetilde{W}_{2,t}$. Hence, if we impose the condition (D.1), the dynamics of the variance risk premium parameter are indirectly modeled as

$$\gamma_t = \kappa \left(\frac{\phi}{\phi_t^*} - 1 \right) . \quad (\text{D.2})$$

Further note that equation (D.2) together with (D.1) determines the dynamics of κ_t^* since

$$\kappa_t^* = \kappa + \gamma_t .$$

As long as the Feller condition $2\kappa_2^*\phi > \lambda_2^2$ is satisfied, ϕ_t^* stays positive, so that also κ_t^* is well defined. In particular, the variance risk premium is negative, and thus the representative investor risk averse, if $\phi_t^* > \phi$. Conversely, the variance risk premium is positive, and thus the representative investor risk seeking, if $\phi_t^* < \phi$. Moreover, if we assume that ϕ_t^* evolves deterministically, option prices can be calculated as in the Heston model.

¹Recall from Chapter 3.4 that the risk-neutral mean-speed of reversion κ^* and the risk-neutral long-term average of the variance ϕ^* are defined as $\kappa^* = \kappa + \gamma$ and $\phi^* = \kappa\phi/(\kappa + \gamma)$, respectively.

Bibliography

- [1] Y. AÏT-SAHALIA and A. LO (2000). „Nonparametric risk management and implied risk-aversion“. In: *Journal of Econometrics* 94 (1-2). Pp. 9–51.
- [2] H. ALBRECHER et al. (2007). „The little heston trap“. In: *Wilmott Magazine*. Pp. 83–92.
- [3] F. ALONSO, R. BLANCO, and G. RUBIO (2009). „Option-implied preferences adjustments, density forecasts, and the equity risk premium“. In: *Spanish Economic Review* 11 (2). Pp. 141–164.
- [4] I. ANAGNOU-BASIOUDIS et al. (2005). „Forecasting accuracy of implied and GARCH-based probability density functions“. In: *Review of Futures Markets* 14 (1). Pp. 41–66.
- [5] T. ANDERSEN and L. BENZONI (2009). „Realized volatility“. In: *Handbook of financial time series*. Ed. by T. ANDERSEN et al. Springer.
- [6] T. ANDERSEN, T. BOLLERSLEV, and F. DIEBOLD (2009). „Parametric and nonparametric volatility measurement“. In: *Handbook of financial econometrics*. Ed. by Y. AÏT-SAHALIA and L. HANSEN. North Holland.
- [7] T. ANDERSEN and O. BONDARENKO (2007). „Construction and interpretation of model-free implied volatility“. In: *Volatility as an asset class*. Ed. by I. NELKEN. RISK Publication.
- [8] T. ANDERSEN et al. (2003). „Modeling and forecasting realized volatility“. In: *Econometrica* 71 (2). Pp. 579–625.
- [9] A. ANG et al. (2006). „The cross-section of volatility and expected returns“. In: *Journal of Finance* 61 (1). Pp. 259–299.
- [10] M. ATTARI (2004). *Option pricing using Fourier transforms: A numerically efficient simplification*. Working Paper, pp. 1–7.
- [11] G. BAKSHI, P. CARR, and L. WU (2008). „Stochastic risk premiums, stochastic skewness in currency options, and stochastic discount factors in international economies“. In: *Journal of Financial Economics* 87 (1). Pp. 132–156.
- [12] G. BAKSHI and N. KAPADIA (2003). „Delta-hedged gains and the negative market volatility risk premium“. In: *Review of Financial Studies* 16 (2). Pp. 527–566.
- [13] G. BAKSHI and N. KAPADIA (2003). „Volatility risk premium embedded in individual equity options: Some new insights“. In: *Journal of Derivatives* 11 (1). Pp. 45–54.

- [14] G. BAKSHI and D. MADAN (2000). „Spanning and derivative-security valuation“. In: *Journal of Financial Economics* 55 (2). Pp. 205–238.
- [15] G. BAKSHI and D. MADAN (2007). *Investor heterogeneity, aggregation, and the non-monotonicity of the aggregate marginal rate of substitution in the price of market-equity*. Working paper, pp. 1–37.
- [16] G. BAKSHI and L. WU (2010). „The behavior of risk and market prices of risk over the Nasdaq bubble period“. In: *Management Science* 56 (12). Pp. 2251–2264.
- [17] P. BALDUZZI, S. DAS, and S. FORESI (1998). „The central tendency: A second factor in bond yields“. In: *Review of Economics and Statistics* 80 (1). Pp. 62–72.
- [18] V. BARNETT and T. LEWIS (1994). *Outliers in statistical data*. Third edition. John Wiley & Sons.
- [19] BASEL COMMITTEE ON BANKING SUPERVISION (2009). *Revisions to the Basel II market risk framework*. Consultative document.
- [20] BASEL COMMITTEE ON BANKING SUPERVISION (2010). *Basel III: A global regulatory framework for more resilient banks and banking systems*. Consultative document.
- [21] D. BATES (1996). „Jumps and stochastic volatility: Exchange rate processes implicit in Deutsche Mark options“. In: *Review of Financial Studies* 9 (1). Pp. 69–107.
- [22] D. BATES (2000). „Post-’87 crash fears in the S&P 500 futures option market“. In: *Journal of Econometrics* 94 (1-2). Pp. 181–238.
- [23] D. BATES (2003). „Empirical option pricing: A retrospection“. In: *Journal of Econometrics* 116 (1-2). Pp. 387–404.
- [24] D. BATES (2012). „U.S. stock market crash risk, 1926 - 2010“. In: *Journal of Financial Economics*. forthcoming.
- [25] L. BENZONI, P. COLLIN-DUFRESNE, and R. GOLDSTEIN (2011). „Explaining asset pricing puzzles associated with the 1987 market crash“. In: *Journal of Financial Economics* 101 (3). Pp. 552–573.
- [26] J. BERKOWITZ (2001). „Testing density forecasts with applications to risk management“. In: *Journal of Business and Economic Statistics* 19 (4). Pp. 465–474.
- [27] N. BINGHAM and R. KIESEL (2004). *Risk-neutral valuation*. Second Edition. Springer.
- [28] F. BLACK (1976). „Studies of stock price volatility changes“. In: *Proceedings from the American Statistical Association, Business and Economic Statistics section*, pp. 177–181.
- [29] F. BLACK and M. SCHOLES (1973). „The pricing of options and corporate liabilities“. In: *Journal of Political Economy* 81 (3). Pp. 637–654.
- [30] B. BLAIR, S.-H. POON, and S. TAYLOR (2001). „Forecasting S&P 100 volatility: The incremental information content of implied volatilities and high frequency index returns“. In: *Journal of Econometrics* 105 (1). Pp. 5–26.

- [31] R. BLISS and N. PANIGIRTZOGLU (2004). „Option-implied risk aversion estimates“. In: *Journal of Finance* 59 (1). Pp. 407–446.
- [32] A. BODINE and C. NAGEL (2008). *Quants gone wild – the subprime crises*. Online article. <http://247wallst.com/2008/03/27/quants-gone-wild/>.
- [33] T. BOLLERSLEV (1986). „Generalized autoregressive conditional heteroskedasticity“. In: *Journal of Econometrics* 31 (3). Pp. 307–327.
- [34] T. BOLLERSLEV (2010). „Glossary to ARCH (GARCH)“. In: *Volatility and time series econometrics: Essays in honor of Robert F. Engle*. Ed. by T. BOLLERSLEV, J. RUSSELL, and M. WATSON. Oxford University Press.
- [35] T. BOLLERSLEV, R. ENGLE, and D. NELSON (1994). „ARCH models“. In: *Handbook of econometrics*. Ed. by R. ENGLE and D. MCFADDEN. North Holland.
- [36] T. BOLLERSLEV, M. GIBSON, and H. ZHOU (2011). „Dynamic estimation of volatility risk premia and investor risk aversion from option-implied and realized volatilities“. In: *Journal of Econometrics* 160 (1). Pp. 235–245.
- [37] T. BOLLERSLEV, G. TAUCHEN, and H. ZHOU (2009). „Expected stock returns and variance risk premia“. In: *Review of Financial Studies* 22 (11). Pp. 4463–4492.
- [38] T. BOLLERSLEV and J. WOOLDRIDGE (1992). „Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances“. In: *Econometric Reviews* 11 (2). Pp. 143–172.
- [39] T. BOLLERSLEV and H. ZHOU (2006). „Volatility puzzles: A simple framework for gauging return-volatility regressions“. In: *Journal of Econometrics* 131 (1-2). Pp. 123–150.
- [40] D. BREEDEN (1979). „An intertemporal asset pricing model with stochastic consumption and investment opportunities“. In: *Journal of Financial Economics* 7 (3). Pp. 265–296.
- [41] D. BREEDEN and R. LITZENBERG (1978). „Prices of state-contingent claims implicit in option prices“. In: *Journal of Business* 51 (4). Pp. 621–651.
- [42] M. BRITTEN-JONES and A. NEUBERGER (2000). „Option prices, implied price processes and stochastic volatility“. In: *Journal of Finance* 55 (2). Pp. 839–866.
- [43] C. BROOKS and G. PERSAND (2003). „Volatility forecasting for risk management“. In: *Journal of Forecasting* 22 (1). Pp. 1–22.
- [44] D. BROWN and J. JACKWERTH (2012). „The pricing kernel puzzle: Reconciling index option data and economic theory“. In: *Contemporary studies in economics and financial analysis: Derivative securities pricing and modelling*. Ed. by J. BATTEN and N. WAGNER. forthcoming. Emerald Group.
- [45] C. BROWNLEES and G. GALLO (2010). „Comparison of volatility measures: A risk management perspective“. In: *Journal of Financial Econometrics* 8 (1). Pp. 29–56.
- [46] R. BRUMMELHUIS and R. KAUFMANN (2004). *Time Scaling for GARCH(1,1) and AR(1)-GARCH(1,1) processes*. Working paper, pp. 1–48.

- [47] T. BUSCH, B. CHRISTENSEN, and M. NIELSEN (2011). „The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets“. In: *Journal of Econometrics* 160 (1). Pp. 48–57.
- [48] J. BUTCHER (2003). *Numerical methods for ordinary differential equations*. Fifth Edition. New Jersey: Prencise Hall.
- [49] P. CARR and D. MADAN (1998). „Towards a theory of volatility trading“. In: *Volatility: New estimation techniques for pricing derivatives*. Ed. by R. JARROW. RISK Publication.
- [50] P. CARR and D. MADAN (1999). „Option valuation using the fast Fourier transform“. In: *Journal of Computational Finance* 2 (4). Pp. 61–73.
- [51] P. CARR and L. WU (2009). „Variance risk premiums“. In: *Review of Financial Studies* 22 (3). Pp. 1311–1341.
- [52] F. CHABI-YO, R. GARCIA, and E. RENAULT (2008). „State dependence can explain the risk aversion puzzle“. In: *Review of Financial Studies* 21 (2). Pp. 973–1011.
- [53] M. CHERNOV (2007). „On the role of risk premia in volatility forecasting“. In: *Journal of Business & Economic Statistics* 25 (4). Pp. 411–426.
- [54] M. CHERNOV and E. GHYSELS (2000). „A study towards a unified approach to the joint estimation of objective and risk neutral measures for the purpose of options valuation“. In: *Journal of Financial Economics* 56 (3). Pp. 407–458.
- [55] B. CHRISTENSEN and N. PRABHALA (1998). „The relation between implied and realized volatility“. In: *Journal of Financial Economics* 50 (2). Pp. 125–150.
- [56] P. CHRISTOFFERSEN, S. HESTON, and K. JACOBS (2011). *A GARCH option model with variance-dependent pricing kernel*. Working paper, pp. 1–45.
- [57] P. CHRISTOFFERSEN, K. JACOBS, and B. CHANG (2012). „Forecasting with option implied information“. In: *Handbook of economic forecasting*. Ed. by G. ELLIOTT and A. TIMMERMANN. Vol. 2. North-Holland.
- [58] K. CHUNG (2001). *A course in probability theory*. Third edition. Academic Press.
- [59] J. COCHRANE (2005). *Asset pricing*. Revisted edition. Princeton University Press.
- [60] R. CONT (2001). „Empirical properties of asset returns: Stylized facts and statistical issues“. In: *Quantitative Finance* 1 (2). Pp. 223–236.
- [61] R. CONT, J. DA FONSECA, and V. DURRLEMAN (2002). „Stochastic models of implied volatility surfaces“. In: *Economic Notes* 31 (2). Pp. 361–377.
- [62] J. COX, J. INGERSOLL, and S. ROSS (1985). „A theory of the term structure of interest rates“. In: *Econometrica* 53 (2). Pp. 385–407.
- [63] Q. DAI and K. SINGLETON (2000). „Specification analysis of affine term structure models“. In: *Journal of Finance* 55 (5). Pp. 1943–1978.

- [64] J. DANIELSSON and J.-P. ZIGRAND (2006). „On time-scaling of risk and the square-root-of-time rule“. In: *Journal of Banking & Finance* 30 (10). Pp. 2701–2713.
- [65] R. DAVIDSON and J. MACKINNON (1993). *Estimation and inference in econometrics*. Oxford University Press.
- [66] P. DEUFLHARD and F. BORNEMANN (2008). *Numerische Mathematik 2 – Gewöhnliche Differentialgleichungen*. Third Edition. de Gruyter.
- [67] P. DOBRÁNSZKY (2009). *Option pricing using numerically evaluated characteristic functions*. Presentation on the 3rd conference on numerical methods in finance, Paris.
- [68] K. DOWD (2002). *An introduction to market risk measurement*. Chichester and New York: Wiley & Sons.
- [69] J. DRIESSEN, P. MAENHOUT, and G. VILKOV (2009). „The price of correlation risk: Evidence from equity options“. In: *Journal of Finance* 64 (3). Pp. 1377–1406.
- [70] J. DUARTE and C. JONES (2007). *The price of market volatility risk*. Working paper, pp. 1–59.
- [71] D. DUFFIE (2001). *Dynamic asset pricing theory*. Third edition. Princeton University Press.
- [72] D. DUFFIE and R. KAN (1996). „A yield-factor model of interest rates“. In: *Mathematical Finance* 6 (4). Pp. 379–406.
- [73] D. DUFFIE, J. PAN, and K. SINGLETON (2000). „Transform analysis and asset pricing for affine jump diffusions“. In: *Econometrica* 68 (6). Pp. 1343–1376.
- [74] B. ERAKER (2004). „Do stock prices and volatility jump? Reconciling evidence from spot and option prices“. In: *Journal of Finance* 59 (3). Pp. 1367–1404.
- [75] P. FACKLER and R. KING (1990). „Calibration of option-based probability assessments in agricultural commodity markets“. In: *American Journal of Agricultural Economics* 72 (1). Pp. 73–83.
- [76] E. FAMA and K. FRENCH (1992). „The cross-section of expected stock returns“. In: *Journal of Finance* 47 (2). Pp. 427–465.
- [77] S. FIGLEWSKI (2010). „Estimating the implied risk neutral density for the U.S. market portfolio“. In: *Volatility and time series econometrics: Essays in honor of Robert F. Engle*. Ed. by T. BOLLERSLEV, J. RUSSELL, and M. WATSON. Oxford University Press.
- [78] D. FILIPOVIĆ and E. MAYERHOFER (2009). „Affine diffusion processes: Theory and applications“. In: *Advanced financial modelling*. Ed. by H. ALBRECHER, W. RUNGGALDIER, and W. SCHACHERMAYER. de Gruyter.
- [79] J. FLEMING, B. OSTDIEK, and R. WHALEY (1995). „Predicting stock market volatility: A new measure“. In: *Journal of Futures Markets* 15 (3). Pp. 265–302.
- [80] G. FOLLAND (1999). *Real analysis*. John Wiley & Sons.

- [81] J. GATHERAL (2006). *The volatility surface – A practitioner’s guide*. John Wiley & Sons.
- [82] G. GEAR (1971). *Numerical initial value problems in ordinary differential equations*. Prentice-Hall.
- [83] H. GEMAN, N. EL KAROUI, and J. ROCHET (1995). „Changes of numéraire, changes of probability measure and option pricing“. In: *Journal of Applied Probability* 32 (2). Pp. 443–458.
- [84] J. GIL-PELAEZ (1951). „Note on the inversion theorem“. In: *Biometrika* 38 (3-4). Pp. 481–482.
- [85] P. GIOT (2005). „Implied volatility indexes and daily Value-at-Risk models“. In: *Journal of Derivatives* 12 (4). Pp. 54–64.
- [86] P. GIOT and S. LAURENT (2007). „The information content of implied volatility in light of the jump/continuous decomposition of realized volatility“. In: *Journal of Futures Markets* 27 (4). Pp. 337–359.
- [87] L. GRAFAKOS (2004). *Classical and modern Fourier analysis*. Pearson Education.
- [88] W. GREENE (2003). *Econometric analysis*. John Wiley & Sons.
- [89] M. GRITH, W. HÄRDLE, and J. PARK (2009). *Shape invariant modelling pricing kernels and risk aversion*. Working paper, pp. 1–30.
- [90] J. HAMILTON (1994). *Time series analysis*. Princeton University Press.
- [91] P. HANSEN, Z. HUANG, and H. SHEK (2012). „Realized GARCH: A joint model of returns and realized measures of volatility“. In: *Journal of Applied Econometrics*. forthcoming.
- [92] A. HARVEY (2003). *Forecasting, structural time series models and the Kalman filter*. Cambridge: University Press.
- [93] F. HAYASHI (2000). *Econometrics*. Princeton University Press.
- [94] T. HENS and C. REICHLIN (2012). *Three solutions to the pricing kernel puzzle*. Working paper, pp. 1–34.
- [95] S. HESTON (1993). „A closed-form solution for options with stochastic volatility with applications to bond and currency options“. In: *Review of Financial Studies* 6 (2). Pp. 327–343.
- [96] B. HOLMQUIST (1988). „Moments and cumulants of the multivariate normal distribution“. In: *Stochastic Analysis and Applications* 6 (3). Pp. 273–278.
- [97] S. HUANG and J. YU (2007). „On stiffness in affine asset pricing models“. In: *Journal of Computational Finance* 10 (3). Pp. 99–123.
- [98] J. HULL (2009). *Options, futures and other derivatives*. Seventh edition. Pearson Prentice Hall.
- [99] J. JACKWERTH (2000). „Recovering risk aversion from option prices and realized returns“. In: *Review of Financial Studies* 13 (2). Pp. 433–451.

- [100] J. JACKWERTH (2004). *Option-implied risk-neutral distributions and risk aversion*. Research Foundation of AIMR.
- [101] J. JACKWERTH and M. RUBINSTEIN (1996). „Recovering probability distributions from option prices“. In: *Journal of Finance* 51 (5). Pp. 1611–1631.
- [102] G. JIANG and Y. TIAN (2005). „Model-free implied volatility and its information content“. In: *Review of Financial Studies* 18 (4). Pp. 1305–1342.
- [103] G. JIANG and Y. TIAN (2007). „Extracting model-free volatility from option prices: An examination of the VIX index“. In: *Journal of Derivatives* 14 (3). Pp. 35–60.
- [104] C. JONES (2003). „The dynamics of stochastic volatility: Evidence from underlying and options markets“. In: *Journal of Econometrics* 116 (1-2). Pp. 181–224.
- [105] S. JULIER (2002). „The scaled unscented transformation“. In: *Proceedings of the American Control Conference*. Pp. 4555–4599.
- [106] S. JULIER and J. UHLMANN (1997). „A new extension of the Kalman filter to nonlinear systems“. In: *Proceedings of SPIE*. Pp. 182–193.
- [107] S. JULIER, J. UHLMANN, and H. DURRANT-WHYTE (1995). „A new approach for filtering nonlinear systems“. In: *Proceedings of the American Control Conference*. Pp. 1628–1632.
- [108] C. KAHL (2007). *Modelling and simulation of stochastic volatility in finance*. Dissertation, pp. 1–220.
- [109] C. KAHL and R. LOARD (2006). „Why the rotation count algorithm works“. In: *Tinbergen Institute Discussion Paper*. Pp. 1–31.
- [110] C. KAHL and R. LOARD (2007). „Optimal Fourier inversion in semi-analytical option pricing“. In: *Journal of Computational Finance* 10 (1). Pp. 1–30.
- [111] C. KAHL and R. LOARD (2010). „Complex logarithms in Heston-like models“. In: *Mathematical Finance* 20 (4). Pp. 671–694.
- [112] R. KALMAN (1960). „A new approach to linear filtering and prediction problems“. In: *Transactions of the ASME, Journal of Basic Engineering*. Pp. 35–45.
- [113] R. KALMAN (1963). „New methods in Wiener filtering theory“. In: *Proceedings of the first symposium of engineering applications of random function theory and probability*. Ed. by J. BOGDANOFF and F. KOZIN. John Wiley & Sons, pp. 270–388.
- [114] B. KANG and T. KIM (2006). „Option-implied preferences: An extension to wider classes of utility functions“. In: *Journal of Financial Markets* 9 (2). Pp. 180–198.
- [115] F. KILIN (2011). „Accelerating the calibration of stochastic volatility models“. In: *Journal of Derivatives* 18 (3). Pp. 7–16.
- [116] S. KOOPMAN, B. JUNGBACKER, and E. HOL (2005). „Forecasting daily variability of the S&P 100 stock index using historical, realized and implied volatility measurements“. In: *Journal of Empirical Finance* 12 (3). Pp. 445–475.

- [117] S. KOU (2002). „A jump-diffusion model for option pricing“. In: *Management Science* 48 (8). Pp. 1086–1101.
- [118] P. KUPIEC (1995). „Techniques for verifying the accuracy of risk management models“. In: *Journal of Derivatives* 3 (2). Pp. 73–84.
- [119] D. LANDO (2004). *Credit risk modeling*. Princeton University Press.
- [120] R. LEE (2004). „Option pricing by transform methods: Extensions, unification and error control“. In: *Journal of Computational Finance* 7 (3). Pp. 51–86.
- [121] A. LEWIS (2001). *A simple option formula for general jump-diffusion and other exponential levy processes*. Working paper, pp. 1–24.
- [122] X. LIU et al. (2007). „Closed-form transformations from risk-neutral to real-world distributions“. In: *Journal of Banking & Finance* 31 (5). Pp. 1501–1520.
- [123] D. LOUZIS, S. XANTHOPOULOS-SISINIS, and A. REFENES (2011). *The role of high frequency intra-daily data, daily range and implied volatility in multi-period Value-at-Risk forecasting*. Working paper, pp. 1–39.
- [124] A. LUNDE and P. HANSEN (2005). „A forecast comparison of volatility models: Does anything beat a GARCH(1,1)?“. In: *Journal of Applied Econometrics* 20 (7). Pp. 873–889.
- [125] B. MANDELBROT (1963). „The variation of certain speculative prices“. In: *Journal of Business* 36 (4). Pp. 394–419.
- [126] R. MERTON (1980). „On estimating the expected return on the market – An explanatory investigation“. In: *Journal of Financial Economics* 8 (4). Pp. 323–361.
- [127] S. MIKHAILOV and U. NÖGEL (2003). „Hestons stochastic volatility model implementation, calibration and some extensions“. In: *Wilmott magazine*. Pp. 74–79.
- [128] S. MIXON (2007). „The implied volatility term structure of stock index options“. In: *Journal of Empirical Finance* 14 (3). Pp. 333–354.
- [129] P. NYBERG and A. WILHELMSSON (2010). „Volatility risk premium, risk aversion and the cross-section of stock returns“. In: *Financial Review* 45 (4). Pp. 1079–1100.
- [130] J. PAN (2002). „The jump-risk premia implicit in options: Evidence from an integrated time-series study“. In: *Journal of Financial Economics* 63 (1). Pp. 3–50.
- [131] L. PERRIDON and M. STEINER (2004). *Finanzwirtschaft der Unternehmung*. Vahlen.
- [132] M. PIAZZESI (2009). „Affine term structure models“. In: *Handbook of financial econometrics*. Ed. by Y. AÏT-SAHALIA and L. HANSEN. North Holland.
- [133] A. VAN DER PLOEG (2004). *Multifactor volatility models: Evidence from stock and option markets*. Working paper, pp. 1–44.
- [134] A. VAN DER PLOEG (2006). *Stochastic volatility and the pricing of financial derivatives*. Dissertation, pp. 1–340.

- [135] S.-H. POON and C. GRANGER (2003). „Forecasting volatility in financial markets: A review“. In: *Journal of Economic Literature* 41 (2). Pp. 478–539.
- [136] P. PROTTER (2005). *Stochastic integration and differential equations*. Springer.
- [137] R. RITCHEY (1990). „Call option valuation for discrete normal mixtures“. In: *Journal of Financial Research* 13 (4). Pp. 285–296.
- [138] J. ROSENBERG and R. ENGLE (2002). „Empirical pricing kernels“. In: *Journal of Financial Economics* 64 (3). Pp. 341–372.
- [139] C. RUNGE (1895). „Über die numerische Auflösung von Differentialgleichungen“. In: *Mathematische Annalen* 1895. Pp. 167–178.
- [140] S. SÄRKKÄ (2011). *Bayesian estimation of time-varying systems*. Lecture notes.
- [141] M. SCHMELZLE (2010). *Option pricing formulae using Fourier transform: Theory and application*. Working paper, pp. 1–86.
- [142] R. SEYDEL (2009). *Tools for computational finance*. Fourth edition. Springer.
- [143] M. SHACKLETON, S. TAYLOR, and P. YU (2010). „A multi-horizon comparison of density forecasts for the S&P 500 using index returns and option prices“. In: *Journal of Banking & Finance* 34 (11). Pp. 2678–2693.
- [144] S. SHREVE (2004). *Stochastic calculus for finance II*. Springer.
- [145] D. SIMON (2006). *Optimal state estimation*. John Wiley & Sons.
- [146] K. SINGLETON (2006). *Empirical dynamic asset pricing: Model specification and econometric assessment*. Princeton University Press.
- [147] D. TASCHE (2002). *Expected shortfall and beyond*. Working paper, pp. 1–18.
- [148] S. TAYLOR (2005). *Asset price dynamics, volatility, and prediction*. Princeton University Press.
- [149] S. TAYLOR, P. YADAV, and Y. ZHANG (2010). „The information content of implied volatilities and model-free volatility expectations: Evidence from options on individual stocks“. In: *Journal of Banking & Finance* 34 (4). Pp. 871–881.
- [150] E. WAN and R. VAN DER MERWE (2000). „The unscented Kalman filter for nonlinear estimation“. In: *Proceedings of Symposium 2000 on Adaptive Systems for Signal Processing, Communication and Control*. Pp. 153–158.
- [151] E. WAN and R. VAN DER MERWE (2001). „The unscented Kalman filter“. In: *Kalman filtering and neural networks*. Ed. by SIMON HAYKIN. John Wiley & Sons.
- [152] Y.-H. WANG (2009). „The impact of jump dynamics on the predictive power of option-implied densities“. In: *Journal of Derivatives* 16 (3). Pp. 9–22.
- [153] T. WATANABE (2012). „Quantile forecasts of financial returns using the realized GARCH models“. In: *Japanese Economic Review* 63 (1). Pp. 68–80.
- [154] R. WHALEY (2000). „The investor fear gauge“. In: *Journal of Portfolio Management* 26 (3). Pp. 12–17.
- [155] R. WHALEY (2009). „Understanding the VIX“. In: *Journal of Portfolio Management* 35 (3). Pp. 98–105.

- [156] H. WHITE (1982). „Maximum likelihood estimation of misspecified models“. In: *Econometrica* 50 (1). Pp. 1–25.
- [157] L. WU (2012). *From characteristic functions and Fourier transforms to PDFs/CDFs and option prices*. Lecture notes. <http://faculty.baruch.cuny.edu>.
- [158] J. ZHU (2009). *Applications of Fourier transform to smile modeling – Theory and implementation*. Springer.
- [159] A. ZIEGLER (2007). „Why does implied risk aversion smile?“ In: *Review of Financial Studies* 20 (3). Pp. 860–904.

Acknowledgment

First and foremost I want to thank my supervisor Professor Dr. Rüdiger Kiesel for the excellent assistance and guidance throughout the last three years. It was always a great pleasure working with him. Rüdiger always took his time to discuss current issues of my thesis and supported me in all matters. Furthermore, he encouraged me to meet other research fellows, both in Germany and abroad, which allowed me to get new insights and research ideas.

For the helpful comments and support while completing this thesis, I am also deeply indebted to my second advisor Professor Dr. Gunter Löffler.

I also want to express my appreciation for the research training group 1100 at the University of Ulm. Special mention must be made in this context to Professor Dr. Karsten Urban, speaker of the research training group 1100, who is always fully engaged in this institution and its scholars. Moreover, I want to thank my colleagues for the great time we had together. I am also deeply grateful to the Deutsche Forschungsgemeinschaft (DFG) for the financial support.

Additionally, I would like to address my thanks to Dr. Gary Kendall, founder of CDO², for supporting my research in the field of credit risk.

For the research stay in December 2010 at the Collaborative Research Center 649 “Economic Risk” at the Humboldt-Universität zu Berlin, I would like to thank Professor Dr. Wolfgang Härdle.

In April 2011 I had the chance to visit Professor Dr. Liuren Wu at the Baruch College in New York. His great experience in my research field allowed me to make great progress during this time. Therefore I want to thank him very much.

My very special thanks go to Sabine Häge, Dr. Stefanie Schürle and Dr. Josef Schürle for proofreading my dissertation.

Finally, I want to thank my parents, who gave me the opportunity to study. Moreover, I want to thank Sabine and my family for the love and support.

Ulm, November 2012

Florentin Rahe

Declaration

I hereby declare that this thesis was performed and written on my own and that references and resources used within this work have been explicitly indicated.

I am aware that making a false declaration may have serious consequences.

Ulm, June 9, 2013

Signature:

Curriculum vitae

Personal Information

Name: Florentin Markus Sebastian Rahe
Marital status: single
Citizenship: German
Date of birth: 09/18/1980
Place of birth: Stuttgart, Germany

Education

03/2009 - 05/2012 **University of Ulm**, Doctoral Studies in Financial Mathematics, Specializations: Credit and Market Risk, Major advisor: Professor Dr. Rüdiger Kiesel
10/2002 - 02/2009 **University of Hohenheim**, Diploma in Economics, Specializations: Accounting and Finance, Banking, Statistics and Econometrics, Average Grade: 1.4
09/2006 - 05/2007 **University of Connecticut**, Master of Science in Mathematics, Specializations: Probability Theory, Stochastic Processes, Financial Mathematics, Major advisor: Professor Ron C. Blei, GPA: 4.01
10/2005 - 08/2006 **University of Stuttgart**, Mathematics, Specialization: Analysis
10/2002 - 08/2004 **University of Hohenheim**, Pre-Diploma in Economics

Research and Teaching Experience

04/2008 - 07/2008, 10/2007 - 02/2008 **University of Hohenheim**, Department of Accounting and Finance, Tutor for the course "Investment and Finance", Professor Dr. Dirk Hachmeister
10/2005 - 07/2006, 05/2004 - 03/2005 **University of Hohenheim**, Department of Accounting and Finance, Student assistant, Professor Dr. Dirk Hachmeister

Work Experience

06/2012 – present	DZ Bank , Frankfurt, Department of Risk Controlling, Referent
04/2005 – 09/2005	Landesbank Baden-Württemberg , Stuttgart, Department of Risk Management Leverage Finance, Internship
11/2002 – 04/2004	Behr GmbH & Co. KG , Stuttgart, Department of Controlling, Student trainee

Awards and Scholarships

01/2010	GENO-Wissenschaftspreis 2009 for my diploma-thesis: <i>“Prognose des Value-at-Risk mit multivariaten GARCH-Modellen”</i> .
10/2009	Best diploma graduate in Economics at the University of Hohenheim (3. Price).
03/2009 – 02/2012	German Research Society (DFG) Scholarship for doctoral studies
09/2006 – 05/2007	Fulbright Scholarship Program (Travel grant)
09/2006 – 05/2007	Baden-Württemberg Scholarship Program
09/2006 – 05/2007	Exchange Program of the state Baden-Württemberg