Delft University of Technology
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft Institute of Applied Mathematics

Forecasting the implied volatility surface in risk-management applications

A thesis submitted to the
Delft Institute of Applied Mathematics
in partial fulfillment of the requirements

for the degree

MASTER OF SCIENCE
in
APPLIED MATHEMATICS

by

Marcel Theodorus Petrus van Dijk

Delft, the Netherlands
November 2017

MSc THESIS APPLIED MATHEMATICS

"Forecasting the implied volatility surface in risk-management applications"

Marcel Theodorus Petrus van Dijk

Delft University of Technology

Daily Supervisors

Prof. Dr. Ir. C.W. Oosterlee

Dr. C.S.L. de Graaf

Responsible Professor

Prof. Dr. Ir. C.W. Oosterlee

Other thesis committee members

Dr. Ir. C. Kraaikamp

November 2017

Delft, the Netherlands

Abstract

In risk-management, one typically simulates many states of the market using models that are in line with historical data, also known as real-world models. For example, new regulations require insurance companies to value their position on a 1-year horizon. Insurance companies issue guarantees that need to be valued according to market expectations, instead of historical data. By calibrating option pricing models to market prices or equivalently, the implied volatility surface, one obtains market consistent values for these guarantees. Currently, it is common practice to assume that the parameters of these option pricing models are constant, i.e. the calibrated parameters from time $t = 0$ are used, as the option prices at $t = 1$ are unknown. However, empirical data shows that the parameters are not constant and depend on the state of the market. In this research, we propose regression models that predict the calibrated parameters, given a set of market variables such as the VIX index and risk-free interest rates. When these market variables are included in the real-world simulation, one is able to predict the calibrated parameters and consequently the option prices which are in line with the simulated state of the market. By performing a regression we are able to predict the out-of-sample implied volatility surfaces accurately. Moreover, the impact on the Solvency Capital Requirement has been evaluated for different points in time. The impact depends on the initial state of the market and varies from -46% to $+52\%$.

Keywords: Option pricing, risk management, recalibration, implied volatility surface, Heston model.

Acknowledgements

This thesis has been submitted for the degree of Master of Science in Applied Mathematics at Delft University of Technology, the Netherlands. The academic supervisor of this project was prof. dr. ir. C.W. Oosterlee of the Numerical Analysis group at the Delft Institute of Applied Mathematics. The research has been conducted under the supervision of dr. C.L.S. de Graaf at Ortec Finance, a consultancy company which aims at improving investment decision-making by providing consistent solutions for advice and risk management through a combination of market knowledge, mathematical models and information technology.

First of all, I would like to thank Kees Oosterlee and Kees de Graaf for their guidance and enthusiasm throughout this project. Moreover, I would like to thank Cor Kraaikamp for being part of the examination committee. Furthermore, I would like to thank Ortec Finance for the opportunity and resources to write this thesis and I would like to thank my colleagues at Ortec Finance for the pleasant working environment. Finally, I want to thank my friends and family for their support during my studies.

Contents

1 Introduction

The implied volatility surface holds essential information about prices of financial securities. This surface quantifies the market value of European call and put options for a range of different strikes and maturities. It is common practice to derive market expectations from these option values, as they are frequently traded and their pay-off depends on future states of the market. First, a model that describes the market is established and its parameters are defined such that the implied volatility surface produced by the model matches the one observed on the market as good as possible. This process is called calibration and a calibrated model can be used to price exotic options, which are less frequently traded. When selling options, it is crucial to accurately determine the model parameters, otherwise traders will make profits at the expense of the writer of the option. A model that describes the implied volatility surface accurately is called a risk-neutral model.

As mentioned before, the implied volatility surface quantifies the expectations of the market with respect to a certain asset. When time progresses and new information about the asset is revealed, the implied volatility surface will change accordingly. We therefore observe an implied volatility surface that changes over time and depends on the state of the market. Consequently, the parameters of the risk-neutral option price model require recalibration over time as well. This can be a cumbersome procedure, as it typically involves a local optimizer, by which one only finds a local optimum which depends on the initial values. In order to find the global optimum (or at least a close approximation), one can sample many initial values or implement a global optimizer. Both solutions are time-consuming and therefore undesirable.

Another issue with respect to recalibration of the implied volatility surface arises when one wishes to evaluate option prices at time $t > 0$, which is common in risk-management applications. When assessing the risk associated with, for example, a balance sheet, one typically simulates the market up to time T using models that are in line with historical data, also known as a real-world measure. The value of the balance sheet is evaluated in every simulated state of the market. Hereafter, the distribution of the balance sheet at time T can be derived. When the balance sheet contains options or other products that involve risk-neutral valuation, the risk-neutral measure at time T is required, which depends on the implied volatility surface at time T. It is, however, very difficult to realistically simulate the evolution of the implied volatility surface, as it quantifies the market's expectations, which depend on many factors. Moreover, even if the implied volatility surface is realistically simulated, a calibration procedure is still required for each simulated path, which leads to impractical computational times. To this end, the parameters of the risk-neutral model are usually assumed to be constant and implied by the implied volatility surface at $t = 0$. This assumption is not in line with historical observations. Indeed, it is highly unlikely that the expectations of a stable market will coincide with the expectations of a market subjected to a financial crisis. Consequently, as the implied volatility surface is a quantification of the future expectations of the market, it is highly unlikely that these markets will have the same risk-neutral measure. It would be more in line with historical observations to include a recalibration process in the real-world simulation, such that the risk-neutral measure can be recalibrated along each simulated trajectory. This way, one conditions the risk-neutral measure at time T on the simulated state of the market, instead of

today's market.

Currently, there is not much literature available regarding this issue. For example, Bauer et al. (2010) and Feng et al. (2017) do not even mention the recalibration process. The main explanation for the lack of literature is that the recalibration process is in violation with the underlying assumptions of the risk-neutral market. One of the main assumptions of the risk-neutral market is that all positions can be hedged. Hence, one should be able to hedge a future position, using the parameters of the current implied volatility surface. However, this is no longer true when the implied volatility surface is recalibrated after every change in the market. So in terms of the theoretical option price dynamics, the recalibration process is irrelevant. But in terms of the actual option price dynamics, the recalibration process does appear to be relevant, according to historical data. For example, Ayache (2016) states that is important that one hedges against changes of (theoretically) constant parameters, as one is not hedging against changes of the underlying process, but against changes in the option price. Since the actual dynamics of the option price are subjected to the recalibration of the risk-neutral measure, it appears to be more accurate to take changes of theoretically constant parameters into account, even if this is in violation with the underlying assumptions of the risk-neutral market.

This thesis is closely related to Singor et al. (2014), where a linear relationship between the VIX index and the parameters of the Heston model is proposed. By including the VIX index in the real-world simulation, one is able to condition the Heston parameters on the simulated state of the market, without the need of the implied volatility surface. The prediction of the implied volatility surface according to the linear regression model, was already relatively accurate. However, other market variables and other regression models were not considered. The main objective of this thesis will be therefore be to improve this model by considering more market indices and regression models. These models aim at improving the accuracy of the original model, while preserving the advantages. Moreover, we will implement the models in the main application: risk-management.

This thesis can be separated into four parts. The first part consists of chapter 2, in which we will discuss the basic properties of the real-world and risk-neutral markets. Moreover, we will stress the importance of finding the relationship between these markets. The second part will be dedicated to improving the model proposed in Singor et al. (2014), by including more market indices and regression models. These models include popular machine learning approaches, which are becoming more and more relevant. The theory and results of these models are discussed in chapter 3. Hereafter, we will discuss the justification of violating the theoretical assumptions of the risk-neutral market. By means of a hedge test, we will investigate whether it is more accurate to consider recalibrated parameters when evaluating future option prices. Chapter 4 describes the theory behind this test and its results. The fourth and final part consists of the regression models' main application: risk-management. This will be the topic of chapter 5. In chapter 6, we will finalize this thesis by pointing out the main findings of this research and topics for future studies.

2

Real-world and risk-neutral markets

When dealing with financial models, it is crucial to define the appropriate measure that characterizes the market. This measure does not only separate the possible from the impossible events, it also defines the probability of possible events occurring. In general, we can distinguish two important measures: the real-world and risk-neutral measure.

The real-world measure is a reflection of the *actual* market expectations and it is mostly used in risk management. For instance, to determine the risk associated with a certain portfolio, the market can be simulated under the real-world measure. Each trajectory will give rise to a different portfolio value and by combining all simulated scenarios one is able to assess the portfolio's sensitivity to changes in the market.

When pricing a derivative, one needs to consider the risk-neutral measure, rather than the real-world measure. A market characterized by the risk-neutral measure ignores individual risk preferences and the expected return on all assets is equal to the risk-free rate. These assumptions are not necessarily realistic, but they do allow for a market where the *unique* price of a derivative can be established.

In this chapter we will discuss some important concepts of the (risk-neutral) option market such as the implied volatility surface, as well as different methodologies to price derivatives. Moreover, we will explain the importance of finding the relationship between real-world and risk-neutral markets.

2.1. Mathematical notation and definitions

Throughout this thesis we will assume that the real-world market is defined on a finite horizon $[0,T]$ and characterized by the probability space $(\Omega,\mathscr{F},\mathbb{P})$, with sample space Ω containing all possible outcome elements ω. Moreover, $\mathscr{F}$ is a σ-algebra on Ω, which can be thought of as every possible combination of outcome elements and $\mathbb{P}:\mathscr{F}\to[0,1]$ is the probability measure that maps these events into probabilities. Finally, natural filtration $\{\mathscr{F}_t, t\in[0,T]\}$ can be thought of as all available information up to time t.

We will assume there exists a measure $\mathbb{Q}$ equivalent to $\mathbb{P}$, meaning that these measures agree on events with probability 0, but they may not agree on events with probability >0. The measure $\mathbb{Q}$ is called risk-neutral when the discounted expected value of every bounded self-financing portfolio[1] under $\mathbb{Q}$ is equal to its initial value,

$$\Pi_0 = e^{-rT}\mathbb{E}^{\mathbb{Q}}[\Pi_T|\mathscr{F}_0], \tag{2.1}$$

with risk-free rate r and portfolio value Π_t at time t. Throughout this thesis we will assume the risk-free rate r to be constant, which justifies the discount factor in (2.1). Using this definition, one can prove that under risk-neutral measure $\mathbb{Q}$, the discounted price process of any (risky) asset is a martingale with respect to the natural filtration. In other words, for the discounted price process $\tilde{S}_t = \{e^{-rt}S_t, t\in[0,T]\}$, the following needs to hold

$$\mathbb{E}^{\mathbb{Q}}\left[\tilde{S}_t\,|\mathscr{F}_s\right] = \tilde{S}_s \qquad \forall 0\le s\le t\le T. \tag{2.2}$$

[1] A self-financing portfolio is defined as a portfolio without investments or withdrawals, besides the investment at $t=0$.

In order to understand the necessity of the risk-neutral probability measure, we first need to introduce the concept of arbitrage. Arbitrage refers to a portfolio with value Π_t at time t and initial value Π_0, such that for some $T > 0$

$$P(\Pi_T \geq \Pi_0) = 1 \qquad \text{and} \qquad P(\Pi_T > \Pi_0) > 0. \tag{2.3}$$

In other words, arbitrage is a situation where profit is possible, but loss is impossible. When arbitrage exists in the market, we will assume traders to take advantage of this situation, until the market is free of arbitrage. Hence, it is reasonable to assume an arbitrage-free market and therefore we need to be able to establish arbitrage-free derivative values. These assumptions have led to the fundamental theorems of asset pricing.

Theorem 1 (First Fundamental Theorem of Asset Pricing). *If a market is characterized by a risk-neutral measure $\mathbb{Q}$, equivalent to the real-world measure $\mathbb{P}$, then it is arbitrage-free.*

This theorem states that if we want to price derivatives without creating arbitrage, we need to consider a risk-neutral measure. In other words, we need a measure such that equation (2.1) holds.

Theorem 2 (Second Fundamental Theorem of Asset Pricing). *Consider a risk-neutral market. The market is complete*[2] *if and only if the risk-neutral measure is unique.*

Thus, by assuming a complete market, one obtains a unique risk-neutral measure $\mathbb{Q}$, leading to unique derivative prices. The proofs of these theorems are omitted in this thesis, but they are discussed in almost every derivative pricing textbook, for example Duffie (1996).

Following the notation of Kenyon et al. (2015) we will introduce the *observed* real-world measure $\mathbb{A} : \mathscr{F}^{'} \to [0,1]$, defined on the *observed* space $(\Omega^{'}, \mathscr{F}^{'})$. This measure is defined according to historically realized market observations, hence $\Omega^{'} \subseteq \Omega$ and $\mathscr{F}^{'} \subseteq \mathscr{F}$. Consequently, $\mathbb{A}$ may not be equivalent to risk-neutral measure $\mathbb{Q}$. Since we are only able to develop a real-world measure based on past observations, we will consider $\mathbb{A}$ as real-world measure throughout this thesis, unless otherwise specified.

2.2. Difference between the risk-neutral and real-world measure

Even though we use the risk-neutral measure to price derivatives, its assumptions regarding the dynamics of the underlying assets (such as risk-neutral drift) do not coincide with the empirical evidence provided by the real-world market. To illustrate this difference, we will provide an example where the dynamics of a real-world stock index $S_t^{\mathbb{P}}$ are modelled using a geometric Brownian motion under probability measure $\mathbb{P}$,

$$\mathrm{d}S_t^{\mathbb{P}} = \mu S_t^{\mathbb{P}} \mathrm{d}t + \sigma S_t^{\mathbb{P}} \mathrm{d}W_t^{\mathbb{P}}, \qquad S^{\mathbb{P}}(0) = S_0, \tag{2.4}$$

where μ and σ are respectively the expected return and volatility of the index and $W_t^{\mathbb{P}}$ is a Wiener process under the real-world measure $\mathbb{P}$. Moreover, we assume a risk premium on this (risky) asset in addition to the risk-free rate. Thus,

$$\mu = r + \mu_p, \tag{2.5}$$

with $\mu_p > 0$. Under these assumptions we can derive an analytical expression for the price process,

$$S_t^{\mathbb{P}} = S_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t^{\mathbb{P}}}. \tag{2.6}$$

The conditional expectation of the discounted price process under $\mathbb{P}$ is given by

$$\begin{aligned}
\mathbb{E}^{\mathbb{P}}\left[\tilde{S}_t \,|\mathscr{F}_s\right] &= \mathbb{E}^{\mathbb{P}}\left[e^{-rt} S_t \,|\mathscr{F}_s\right] \\
&= e^{-rt}\mathbb{E}^{\mathbb{P}}\left[S_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t^{\mathbb{P}}} \,|\mathscr{F}_s\right] \\
&= e^{-rt} S_0 e^{(\mu - \frac{1}{2}\sigma^2)t}\mathbb{E}^{\mathbb{P}}\left[e^{\sigma W_t^{\mathbb{P}}} \,|\mathscr{F}_s\right] \\
&= S_0 e^{(\mu - r - \frac{1}{2}\sigma^2)t} e^{\frac{1}{2}\sigma^2(t-s) + \sigma W_s^{\mathbb{P}}} \\
&= e^{(\mu - r)(t-s)} e^{-rs} S_0 e^{(\mu - \frac{1}{2}\sigma^2)s + \sigma W_s^{\mathbb{P}}} \\
&= e^{\mu_p (t-s)} \tilde{S}_s \neq \tilde{S}_s \qquad \forall 0 \leq s < t \leq T.
\end{aligned} \tag{2.7}$$

[2] A market is considered to be complete if every risk position can be hedged and every security can be exchanged.

We may conclude that the discounted price process is not a martingale, hence $\mathbb{P}$ is not a risk-neutral measure. Using this measure to price derivatives would lead to arbitrage opportunities. So even though this model reflects the expected real-world dynamics of the index, we cannot use it to price derivatives. To this end, we will change from measure $\mathbb{P}$ to $\mathbb{Q}$ by applying Girsanov's theorem:

$$\mathrm{d}S_t^{\mathbb{Q}} = rS_t^{\mathbb{Q}}\mathrm{d}t + \sigma S_t^{\mathbb{Q}}\mathrm{d}W_t^{\mathbb{Q}}, \qquad S^{\mathbb{Q}}(0) = S_0. \tag{2.8}$$

Note that we assume the initial values of the processes to be equal (since they are known). The measures $\mathbb{Q}$ and $\mathbb{P}$ are equivalent due to the properties of Girsanov's theorem. Moreover, the conditional expectation can be written as

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}}\left[\tilde{S}_t \,|\mathscr{F}_s\right] &= \mathbb{E}^{\mathbb{Q}}\left[e^{-rt}S_t \,|\mathscr{F}_s\right] \\
&= e^{-rt}\mathbb{E}^{\mathbb{Q}}\left[S_t \,|\mathscr{F}_s\right] \\
&= S_0 e^{-rt} e^{(r-\frac{1}{2}\sigma^2)t}\mathbb{E}^{\mathbb{Q}}\left[e^{\sigma W_t^{\mathbb{Q}}} \,|\mathscr{F}_s\right] \\
&= S_0 e^{\frac{1}{2}\sigma^2 t} e^{\frac{1}{2}\sigma^2(t-s)+\sigma W_s^{\mathbb{Q}}} \\
&= e^{-rs} S_0 e^{(r-\frac{1}{2}\sigma^2)s+\sigma W_s^{\mathbb{Q}}} \\
&= \tilde{S}_s \qquad \forall 0 \le s \le t \le T.
\end{aligned} \tag{2.9}$$

Hence, the discounted price process is a martingale, indicating the risk-neutrality of $\mathbb{Q}$, leading to an arbitrage-free pricing method according to the first fundamental theorem of asset pricing. The relation between the real-world and risk-neutral measure might seem trivial in this case (change μ into r), but this relation gets more complicated when more involved asset processes are assumed (we will describe some of these processes in section 2.4). Extensive research has been performed investigating the relation between $\mathbb{Q}$ and $\mathbb{P}$, see for example de Vincent-Humphreys and Noss (2012). Contrary to the aforementioned research, this thesis will not directly focus on the relationship between the real-world and risk-neutral probability density functions, but rather on the relationship between real-world indices and the parameters of risk-neutral models. These risk-neutral models can be used to price a variety of different derivatives, but we limit our research to stock index options[3]. In sections 2.3-2.5 we will discuss different option pricing models and how to calibrate them.

Finally, most of our models will be characterized by measure $\mathbb{Q}$, hence we will remove the superscripts throughout this thesis for notational convenience.

2.3. Black-Scholes implied volatility surface

The model proposed in Black and Scholes (1973) plays an important role when it comes to option pricing. In their model, Black and Scholes assume the underlying asset S_t to follow a Geometric Brownian Motion,

$$\mathrm{d}S_t = rS_t\mathrm{d}t + \sigma S_t\mathrm{d}W_t, \qquad S(0) = S_0. \tag{2.10}$$

With risk-free interest rate r, spot price of the underlying asset S_0 and volatility σ. Using this model, one is able to value options in closed form. For example, the price of a European option at time t for strike price K and maturity T is given by

$$\begin{cases} C^{\mathrm{BS}}(t,\sigma,r,S_t,K,T) = \omega[(S_t\mathscr{N}(\omega d_1) - Ke^{-r(T-t)}\mathscr{N}(\omega d_2)], \\ d_1 = \frac{\log(\frac{S_t}{K})+(r-\frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}, \\ d_2 = d_1 - \sigma\sqrt{T-t}. \end{cases} \tag{2.11}$$

With the cumulative standard normal distribution $\mathscr{N}(\cdot)$, $\omega = 1$ for a call option and $\omega = -1$ for a put option. The Black-Scholes equation is widely used to convert (European) option prices to the Black-Scholes implied volatility σ_{imp}, which we will simply refer to as implied volatility throughout this thesis. The implied volatility is defined as the volatility for which the Black-Scholes price matches the market price:

$$C^{\mathrm{Market}} = C^{\mathrm{BS}}\left(t,\sigma_{\mathrm{imp}},r,S_t,K,T\right), \tag{2.12}$$

[3] A stock index option is a financial product which pay-off depends on the uncertain movement of one or multiple stock indices.

where C^{Market} is the option price on the market. By combining the implied volatilities of different strikes and maturities we obtain the implied volatility surface. The volatility is assumed to be constant for different strikes and maturities in the model of Black and Scholes, hence, the implied volatility surface should be flat. However, on the market we observe implied volatility surfaces as in figure 2.1 and therefore the Black-Scholes model fails to describe the market realistically. Thus, if we want to price other (exotic)

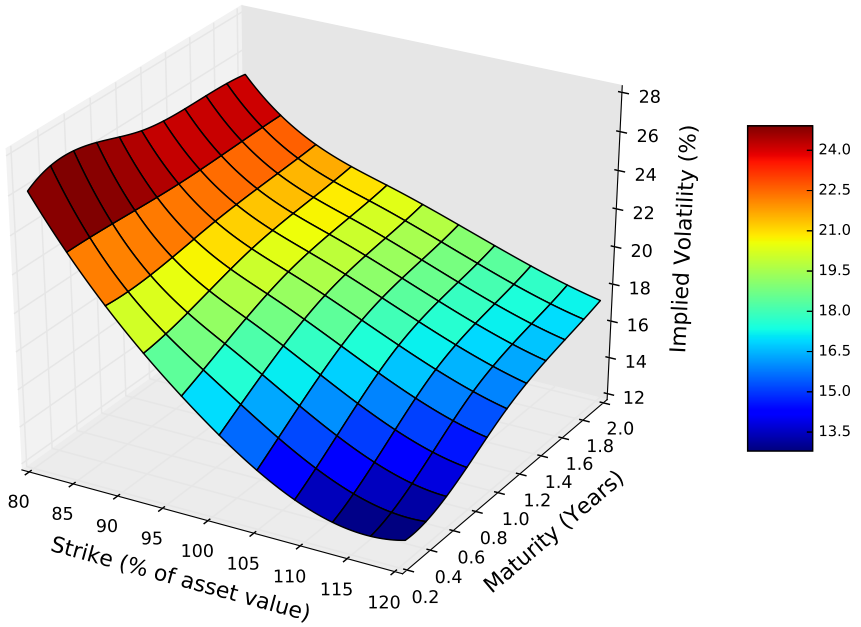

Figure 2.1: Implied volatility surface S&P-500 options, March 30th 2012.

options in a fair way, we need to assume models able to describe these surfaces. There exists a wide range of such models, of which we will give an overview in the next section.

2.4. Overview of option price models

In order to replicate the implied volatility surface observed on the market, one needs to assume a model different from the Black-Scholes model. Usually, these models involve stochastic differential equations (SDEs) to describe important underlying processes, such as asset price, asset volatility and interest rates. Note that these processes are characterized by the risk-neutral measure, rather than the real-world measure.

In general, we can distinguish three different groups: local volatility models, stochastic volatility models (possibly with jumps) and local-stochastic volatility models. In this section we will briefly describe the assumptions, advantages and disadvantages of these models, complemented by some examples we consider relevant for our research.

2.4.1. Local volatility models

Local volatility models assume a deterministic volatility function consistent with the observed market volatilities. The underlying SDE is assumed to be

$$\mathrm{d}S_t = \mu(t)S_t\mathrm{d}t + \sigma_{\text{LV}}(t,S_t)S_t\mathrm{d}W_t, \qquad S(0) = S_0, \tag{2.13}$$

with the deterministic local volatility function $\sigma_{\text{LV}}(t,S_t)$.

In general, these models are able to describe the observed implied volatility surface perfectly, but they are very unstable over time, meaning that the volatility function often needs recalibration, leading to hedging strategies that perform worse than Black-Scholes hedging strategies[4]. In other words, even

[4]In chapter 4 we will discuss the details of hedging strategies

though these models are perfectly capable of describing the current implied volatility surface, they fail to capture information about the dynamics of the underlying asset. In this section we will review a few examples of local volatility functions.

Dupire's local volatility
The idea of local volatility functions was first considered in Dupire (1994). This research proposed a volatility function consistent with observed option prices, that still allowed for hedging strategies and pricing without arbitrage. Dupire's local volatility function is based on the Fokker-Planck equation and it incorporates the derivative of option prices observed on the market, both in maturity and strike direction:

$$\sigma_{\mathrm{LV}}^2(T,K) = \frac{\dfrac{\partial C^{\mathrm{Market}}}{\partial t}}{\dfrac{1}{2}K^2\dfrac{\partial^2 C^{\mathrm{Market}}}{\partial K^2}}. \tag{2.14}$$

Dupire's equation is among the most famous local volatility functions due to its direct link with the observed market prices. This allows for a non-parametric structure, which greatly simplifies the calibration procedure.

Mixture of densities
In Brigo and Mercurio (2002) a general class of analytically tractable models was proposed, based on a given mixture of densities. They considered N diffusion processes following

$$\mathrm{d}S_t^i = \mu S_t^i \mathrm{d}t + v_i(t,S_t^i)\mathrm{d}W_t, \qquad i = 1,\ldots,N, \tag{2.15}$$

with initial conditions S_0^i equal to S_0. By assuming some regularity conditions for $v_i(t,S_t^i)$, they were able to define the local volatility function as a weighted average of these processes

$$\sigma_{\mathrm{LV}}(t,S_t) = \sqrt{\frac{\sum_{i=1}^N \lambda_i v_i^2(t,S_t)p_t^i(S_t)}{\sum_{i=1}^N \lambda_i S_t^2 p_t^i(S_t)}}. \tag{2.16}$$

The processes are characterized by their weights λ_i and density functions p_t^i. In particular they considered a mixture of log-normal densities, where the option price can be seen as a weighted average of Black-Scholes option prices.

Local volatility with stochastic parameters
In order to overcome the poor hedging performances of local volatility models, Alexander and Nogueira (2004) constructed a general framework for parametric local volatility models that incorporates stochasticity. By allowing the parameters of the local volatility model to evolve stochasticly over time, they were able to adjust the Greeks (sensitivities to changes in the market), which improved the hedging strategies of these models. In this thesis we will adopt a similar approach for parameters of a stochastic volatility model.

2.4.2. Stochastic volatility models (with jumps)

Stochastic volatility models assume the volatility to be driven by a stochastic process, rather than a deterministic function. These assumptions lead to a system of SDEs:

$$\begin{cases} \mathrm{d}S_t = \mu_S(t,S_t)\mathrm{d}t + \sigma_S(t,v_t)S_t\mathrm{d}W_t^S + J_S(t,S_t,v_t)\mathrm{d}N_t, & S(0) = S_0, \\ \mathrm{d}v_t = \mu_v(t,v_t)\mathrm{d}t + \sigma_v(t,v_t)\mathrm{d}W_t^v + J_v(t,S_t,v_t)\mathrm{d}N_t, & v(0) = v_0, \\ \langle \mathrm{d}W_t^S, \mathrm{d}W_t^v \rangle = \rho \mathrm{d}t. \end{cases} \tag{2.17}$$

The stochasticity in these models can come from the jump process N_t and/or the Wiener processes W_t^S and W_t^v.

In general, stochastic volatility models assume a more realistic underlying process than local volatility models, leading to improved hedging strategies, for example. However, they do not always succeed in replicating the implied volatility surfaces observed on the market. One can reduce this problem by introducing more complicated models. However, these models are still not perfect in terms of accuracy, but they do increase computational complexity. In this section, we will briefly discuss the properties of a few relevant stochastic volatility models.

Heston model
A well-known stochastic volatility model is proposed in Heston (1993), which assumes

$$\begin{cases} \mathrm{d}S_t = rS_t\mathrm{d}t + \sqrt{v_t}S_t\mathrm{d}W_t^S, & S(0) = S_0, \\ \mathrm{d}v_t = \kappa(\bar{v} - v_t)\mathrm{d}t + \gamma\sqrt{v_t}\mathrm{d}W_t^v, & v(0) = v_0, \\ \langle \mathrm{d}W_t^S, \mathrm{d}W_t^v \rangle = \rho\mathrm{d}t. \end{cases} \tag{2.18}$$

The volatility follows a so-called CIR process (constructed in Cox et al. (1985)), which is a mean-reverting process and prevents negative volatility. In this model, κ is the speed of mean reversion coefficient, i.e. the coefficient that determines the rate at which v_t is pulled back towards its long term mean $\bar{v}$, γ is the volatility of the volatility and ρ is the correlation between the Wiener processes W_t^S and W_t^v. The Heston model is known for its analytical tractability and its ability to reproduce most implied volatility surfaces. However, the model has difficulties with short-term maturities.

Time-dependent Heston model
In Benhamou et al. (2010) a general framework is proposed for the Heston model with time-dependent parameters. By comparing the Heston model with time-dependent parameters to an equivalent perturbed process, they were able to identify the characteristics of the model for any time-dependent (deterministic) process of the parameters. The price for put and call options can be approximated by the price according to the Black-Scholes equation plus an adjustment. The main disadvantage of this approach is the presence of complicated integrals. An involved integration procedure is required, in order to approximate these integrals, unless the parameters are assumed to follow a convenient, but unrealistic, function (for example piecewise constant). This leads to a complex calibration procedure and higher approximation errors.

Stochastic volatility with jumps
Besides pure diffusion models such as the Heston model, there also exists a wide range of jump-diffusion models. One of the first models to include both stochastic volatility and jumps was proposed in Bates (1996). The underlying asset is assumed to follow

$$\begin{cases} \mathrm{d}S_t = (r - \lambda\bar{\mu})S_t\mathrm{d}t + \sqrt{v_t}S_t\mathrm{d}W_t^S + (e^J - 1)S_t\mathrm{d}N_t, & S(0) = S_0, \\ \mathrm{d}v_t = \kappa(\bar{v} - v_t)\mathrm{d}t + \gamma\sqrt{v_t}\mathrm{d}W_t^v, & v(0) = v_0, \\ \langle \mathrm{d}W_t^S, \mathrm{d}W_t^v \rangle = \rho\mathrm{d}t. \end{cases} \tag{2.19}$$

N_t is a Poisson process with intensity λ and the jump amplitude is given by J. This model takes into account the fear of sudden jumps in the underlying asset, which should improve the accuracy of the model. There are multiple variations within the jump-diffusion models, for example by introducing a jump process in the volatility SDE. However increasing the number of jump processes in the model increases the number of parameters as well, which complicates the calibration procedure.

2.4.3. Local-stochastic volatility models

Local-stochastic volatility models combine the previous two models by assuming a partly deterministic, partly stochastic volatility function. The general SDE representation of these models is

$$\begin{cases} \mathrm{d}S_t = \mu_S(t,S_t)\mathrm{d}t + \sigma_{\mathrm{LV}}(t,S_t)\sigma_S(t,v_t)\mathrm{d}W_t^S + J_S(t,S_t,v_t)\mathrm{d}N_t, & S(0) = S_0, \\ \mathrm{d}v_t = \mu_v(t,v_t)\mathrm{d}t + \sigma_v(t,v_t)\mathrm{d}W_t^v + J_v(t,S_t,v_t)\mathrm{d}N_t, & v(0) = v_0, \\ \langle \mathrm{d}W_t^S, \mathrm{d}W_t^v \rangle = \rho\mathrm{d}t. \end{cases} \tag{2.20}$$

These models are able to reproduce the current implied volatility surfaces as well as the future dynamics of the underlying asset. However, they usually do not have closed-form solutions, making calibration and evaluation even more challenging.

Heston model combined with Dupire's equation
One of the first models exploiting the idea of combining local and stochastic volatility was proposed in Jex et al. (1999). They assumed a combination of the Heston model and Dupire's equation,

$$\begin{cases} \mathrm{d}S_t = rS_t\mathrm{d}t + \sqrt{X(t,S_t)}\sqrt{v_t}S_t\mathrm{d}W_t^S, & S(0) = S_0, \\ \mathrm{d}v_t = \kappa(\bar{v} - v_t)\mathrm{d}t + \gamma\sqrt{v_t}\mathrm{d}W_t^v, & v(0) = v_0, \\ \langle \mathrm{d}W_t^S, \mathrm{d}W_t^v \rangle = \rho\mathrm{d}t, \\ X(T,K)^2 = \frac{\sigma_{\mathrm{LV}}^2(T,K)}{\mathbb{E}[v_t^2 | S(T)=K]}, \end{cases} \tag{2.21}$$

where $\sigma^2_{\text{LV}}(K,T)$ is defined as in equation (2.14). The aim of this model is to combine the advantages of both the Heston and Dupire's model. However, this approach does require evaluation of a conditional expectation, which introduces additional computational complexity. Several studies have been dedicated to the efficient evaluation of this conditional expectation, see for example van der Stoep et al. (2014). Even though results continue to improve regarding this issue, the methods are still based on costly simulation approaches, because there is no closed-form expression for the expectation.

Lipton's universal volatility model
The model proposed in Lipton (2002) aims to improve accuracy even more by adding a jump component to (2.21). Adding more degrees of freedom, naturally leads to improved results, but this comes at the cost of a more involved model. Complexity plays an even bigger role in this model: it does not only suffer from the problems of the previous model (calculation of the conditional expectation), it also adds more complexity by adding jump parameters.

2.4.4. Choosing the optimal model

Every model described in this section has its advantages and disadvantages, giving rise to the difficult decision of choosing the right one. To the end, there is always a trade-off between accuracy and complexity: by assuming a highly accurate model one increases computational complexity, but by assuming a rather simple model one loses the realistic representation. Moreover, one could question the applicability of complex models. Most models improve the accuracy by allowing more degrees of freedom, which naturally leads to improved in-sample results, but not necessarily to improved out-of-sample results.

With the above-mentioned considerations in mind, we decided upon the Heston model for our application. The Heston model has a rather realistic representation of the market (by assuming stochastic volatility) and due to its analytical tractability it remains fast to evaluate and calibrate. We acknowledge that this is a highly problem-dependent decision, hence we do not claim the Heston model to be optimal in general.

2.5. Calibration of the Heston model

Calibrating a model is equivalent to finding the set of parameters, such that the implied volatility produced by the model matches the implied volatility observed on the market as good as possible. In this section we will explain in more detail how to perform this procedure applied to the Heston model.

Before a model can be calibrated, one has to be able to determine European option prices for different maturities and strikes, given a set of parameters Ω^{Heston}. During calibration, many option values have to be evaluated for different parameter sets. Hence, a fast and robust pricing algorithm is desirable. Due to its analytical tractability, a variety of fast pricing methods exist for the Heston model. Throughout this thesis we will use the method constructed in Carr and Madan (1999), which determines the Fourier transform of the option price and uses the fast Fourier transform to evaluate this expression. A detailed overview of this pricing process can be found in appendix A.

After obtaining the option prices, it is possible to calibrate the Heston model by comparing the market prices to the calculated prices. However, it is difficult to define a good loss function when comparing low priced options to high priced options. It is much more stable, in fact, to compare the implied volatilities observed on the market to the implied volatilities of the model. In other words, one needs to find the volatility such that the price according to the Black-Scholes equation matches the price produced by the model

$$C^{\text{BS}}\left(t,\sigma^{\text{Heston}},r,S_t,K,T\right) = C^{\text{Heston}}\left(t,\Omega_t^{\text{Heston}},S_t,K,T\right). \tag{2.22}$$

Unfortunately, we cannot derive an analytical expression for σ^{Heston} from (2.22). Hence, we need to extract σ^{Heston} numerically, for which we will use the well-known Newton-Raphson method. The Newton-Raphson method is an iterative method that is able to find the roots of an equation $f(\sigma)$ by applying the following process:

$$\begin{cases} \sigma_0 = \sigma_{\text{init}}, \\ \sigma_{i+1} = \sigma_i - \frac{f(\sigma_i)}{f'(\sigma_i)}. \end{cases} \tag{2.23}$$

If we define

$$f(\sigma^{\text{Heston}}) = C^{\text{BS}}\left(t, \sigma^{\text{Heston}}, r, S_t, K, T\right) - C^{\text{Heston}}\left(t, \Omega_t^{\text{Heston}}, S_t, K, T\right), \tag{2.24}$$

$f'(\sigma^{\text{Heston}})$ is given by

$$f'(\sigma^{\text{Heston}}) = \frac{\partial C^{\text{BS}}}{\partial \sigma^{\text{Heston}}} = S_t N'(d_1)\sqrt{T-t}. \tag{2.25}$$

With the standard normal density function $N'(\cdot)$ and d_1 defined as in (2.11). After defining σ_{init}, we can extract σ^{Heston} by applying (2.24) and (2.25) to (2.23). Note that $f'(\sigma^{\text{Heston}})$ is strictly positive, leading to fast convergence and a unique solution.

Finally, we need to minimize the difference between the implied volatilities produced by the model and the implied volatilities observed on the market, by adjusting the model parameters. First, one needs to identify the parameter search space. For the Heston parameters we assume:

$$\begin{aligned}\Omega^{\text{Search}} &= \Omega_\kappa^{\text{Search}} \times \Omega_{v_0}^{\text{Search}} \times \Omega_{\bar{v}}^{\text{Search}} \times \Omega_\gamma^{\text{Search}} \times \Omega_\rho^{\text{Search}} \\ &= [0,10] \times [0,1] \times [0,1] \times [0,2] \times [-1,1].\end{aligned} \tag{2.26}$$

After defining the search space, one way of finding the calibrated parameters at time t is by minimizing the sum of squared errors:

$$\Omega_t^{\text{Heston}} = \underset{\Omega \in \Omega^{\text{Search}}}{\arg\min} \left(\sum_K \sum_T \left(\sigma^{\text{Market}}(t,K,T) - \sigma^{\text{Heston}}(t,\Omega,K,T) \right)^2 \right). \tag{2.27}$$

This can be solved by, for example, the Levenberg-Marquardt least-squares algorithm. This algorithm finds a local minimum, so it is best to sample different initial conditions, repeat the process and keep the best results. The results will give, for each time-step t, the optimal parameters of the Heston model: $\Omega_t^{\text{Heston}} = \{\kappa_t, v_{(0,t)}, \bar{v}_t, \gamma_t, \rho_t\}$.

2.6. Evolution of the Heston model parameters

In the previous section we described how to find the set of calibrated parameters for the Heston model. In practice, it is very common to recalibrate these parameters after changes in the implied volatility surface. Thus, with an ever-changing market, these parameters evolve over time, despite the fact that they are assumed to be constant in the Heston model. The calibration procedure has been applied to monthly implied volatility surfaces of the S&P-500 index, from January 2006 until February 2017 and the results can be found in figure 2.2. During our calibration procedure we assumed κ to be constant; an unrestricted κ led to unstable results and did not improve the accuracy. The other parameters, however, do not appear to be constant over time, which gives rise to multiple issues.

One of these issues is of practical nature: as soon as the market and the accompanied implied volatility surface changes, the model requires recalibration, which can be a troublesome procedure. In order to apply a (local) minimization algorithm, one needs to specify the initial parameter values. If chosen properly, these initial values will lead to the global optimum. However, due to the non-linearity of the minimization problem, it is more likely to end up in a local minimum, especially with an increasingly complex model. In order to find the global optimum (or at least a close approximation) one can either implement a global optimization algorithm or re-sample the initial conditions and keep the best results. Both solutions can be time-consuming, which is undesirable.

The other issue is related to risk-management. Typically, one simulates many real-world paths to assess the sensitivity to the market of a portfolio, balance sheet, etc. In many cases, a risk-neutral valuation is required nested inside these simulations, for example if the portfolio or balance sheet contains options. Consequently, the future implied volatility surface for each trajectory needs to known, such that the Heston parameters can be calibrated accordingly. It is, however, very complicated to simulate the implied volatility surface, as it quantifies the market's expectations, which depend on many factors. Moreover, even if the implied volatility surface is modelled, one still needs to perform a costly calibration procedure. Performing this calibration procedure in each of the, say, 10.000 simulated trajectories leads to huge computational times. It is therefore common practice to assume constant parameters and by

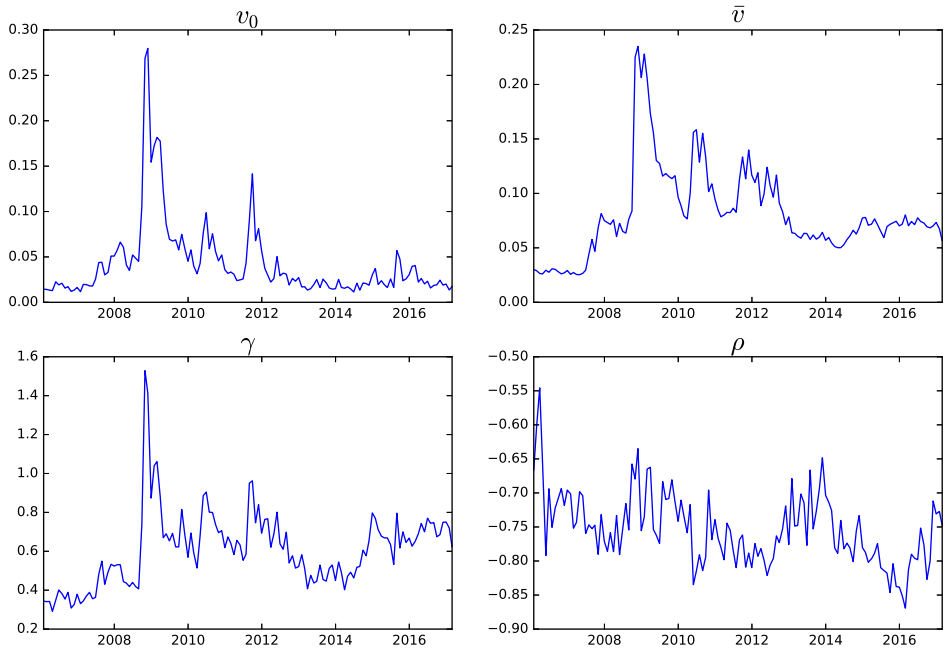

Figure 2.2: Evolution of the Heston model parameters

doing so, one ignores future implied volatility surfaces.

Figure 2.2 clearly shows that the parameters are time-dependent, which is in contrast with the assumptions of the Heston model. Hence, it seems natural to adopt the time-dependent Heston model discussed in section 2.4.2. This model does not assume constant parameters and therefore appears to be more in line with historical observations. To remain arbitrage-free, this model must be calibrated according to the implied volatility surface at $t = 0$. Consequently, one obtains a term structure in line with today's option market. However, the actual development of the market will not necessarily follow this term structure and it will be adjusted as soon as the implied volatility surface changes, just like the Heston parameters. Adopting the time-dependent Heston model will therefore not solve the aforementioned issues, as its parameters still depend on future implied volatility surfaces and will change over time. Moreover, there is currently no risk-neutral model completely invariant to changes in the implied volatility surface, hence the parameters of all risk-neutral models will change over time. This leads to the conclusion that the issues caused by the dynamic behaviour of the implied volatility surface cannot be solved by adopting a different risk-neutral model. Instead, this research aims at investigating the relationship between the Heston parameters and observable market indices. By including the market indices in the real-world simulation, we will be able to directly forecast the set of Heston parameters within each simulated trajectory. This way, one is able to condition the risk-neutral measure on the simulated state of the market, without the need of a simulated implied volatility surface.

It is, however, important to stress the different assumptions with respect to the real-world and risk-neutral market. Risk-neutral valuations will be performed under the Heston model, which assumes constant parameters. However, in real-world simulations we will assume the Heston parameters to change over time according to the simulated state of the market, similar to figure 2.2. One could argue that this approach is invalid, since we are violating the assumptions of the risk-neutral market. To this end, we will discuss a justification of this approach in chapter 4, by means of a hedge test.

2.7. Evolution of the implied volatility surface

The previous section discussed the dynamic behaviour of the Heston model parameters. This dynamic behaviour is directly linked to the dynamic behaviour of the risk-neutral market, which is quantified by the implied volatility surface. So in order to improve our understanding of the evolution of the Heston

model parameters, we will review some research regarding the dynamics of the implied volatility surface in this section.

Mixon (2002) investigated whether macroeconomic factors are able to explain changes in the implied volatility surface. By applying principal component analysis to a variety of (observable) factors, they concluded that most of the variation within the implied volatility surface of the S&P-500 index can be explained by three unobservable factors[5]. Moreover, they investigated which market factors accounted for most of the explanatory power. They concluded that a large proportion of the changes in the implied volatility surface is caused by the path of the underlying index, both for long-term and short-term options.

Cont et al. (2003) proposed a method to jointly simulate the underlying asset and the implied volatility surface. First they performed a Karhunen-Loève decomposition[6] on the implied volatility surface, to obtain the factors describing this surface. Hereafter, they proposed a stochastic process to model the evolution of these factors. By linking the processes of the implied volatility surface to the process of the underlying asset, they were able to jointly simulate future trajectories of the underlying asset and the accompanied implied volatility surface. This approach could be used to determine future expected option prices. However, as we stated before in section 2.6, this methodology requires a calibration procedure for every simulated trajectory, leading to high computational times.

Audrino and Colangelo (2010) proposed a slightly different approach to model the dynamics of the implied volatility surface. This research assumed an initial regression function and aimed at updating the residuals between the forecasted and observed implied volatility surfaces. The magnitude of the residuals were estimated by regression trees and a variety of different market indices. They concluded that the magnitude of the residuals were mainly driven by strike, maturity and closing price of the underlying index. Adding interest rates for several maturities slightly improved the accuracy, but adding even more explanatory variables did not lead to significant improvements. In chapter 3 we will apply a comparable model to the parameters of the Heston model, called gradient boosted regression trees.

One can think of the implied volatility surface as a quantification of the market's expectations, as option prices depend on the future state of the market. The above-mentioned researches show that these expectations are linked to the current state of the market. For example, during a financial crisis the expectations with respect to future states of the market will be relatively uncertain, leading to high implied volatilities. This could explain the correlation between v_0, $\bar{v}$ and γ (see figure 2.2), as they are all subjected to the same expectations. This is one of the most important features we will use in this research, because it allows us to predict the values of the Heston parameters, conditioned on the state of the market. Consequently, this allows us to condition risk-neutral valuations at time $t > 0$ on the simulated state of the market, instead of today's market.

[5]The first three principal components.
[6]The equivalent of principal component analysis for random surfaces.

3

Predicting the Heston parameters

In this chapter we will describe the details of the VIX Heston model, as proposed in Singor et al. (2014), which imposes constraints on the Heston model parameters. These constraints will give us multiple advantages, some regarding the calibration and simulation purposes, at the cost of decreasing accuracy. Moreover, we will propose multiple extensions of the VIX Heston model, which aim at increasing the accuracy, while preserving the advantages of the original model. The models will be compared by the accuracy of the forecasted Heston parameters and the accuracy of the accompanied implied volatility surfaces.

3.1. VIX Heston model

In the previous chapter we stated the importance of linking the real-world to the risk-neutral measure dynamics. This is not a new concept, as several studies already incorporate real-world indices in risk-neutral market calibrations. In particular, the VIX index appears in multiple studies. The VIX index is a volatility measure for the S&P-500 index, calculated by the Chicago Board Options Exchange (CBOE). It represents the weighted average of the implied volatility smile with infinite strike range and one month maturity. Further details regarding the VIX and its calculation can be found in CBOE (2015).

Attractive properties of the VIX index are its direct link to the implied volatility surface and its ability to be incorporated in real-world simulations. Consequently, extracting information from the VIX index has been the topic of multiple studies. For example, in Duan and Yeh (2012) a particle-filter based calibration method is proposed, which implements the S&P-500 index value and the VIX term structure simultaneously.

In this section, we will focus on the methodology presented in Singor et al. (2014), where the development of the Heston model parameters for the S&P-500 index options and the VIX index have been analysed. In this paper, a correlation between the VIX index and the Heston parameters was found, which is visually represented in figure 3.1. After analysis, it was concluded:

- The initial volatility $\sqrt{v_{0,t}}$ and the volatility of the volatility γ_t are highly correlated to the VIX index, with a correlation coefficient of 0.99 and 0.76, respectively.
- The long term volatility $\sqrt{\bar{v}_t}$ appears to be correlated to the VIX index trendline (estimated by a Kalman filter) with a correlation coefficient of 0.74.

To this end, the following restrictions were imposed on the Heston model parameters:

$$\Omega_t^{\text{Heston}}(X) = \begin{cases} \kappa_t & = \kappa, & \kappa \in \mathbb{R}_+, \\ v_{0,t} & = \left(a_{v_0} \cdot \text{VIX}_t + b_{v_0}\right)^2, & a_{v_0}, b_{v_0} \in \mathbb{R}, \\ \bar{v}_t & = \left(a_{\bar{v}} \cdot \text{VIX}_{\text{filter}_t} + b_{\bar{v}}\right)^2, & a_{\bar{v}}, b_{\bar{v}} \in \mathbb{R}, \\ \gamma_t & = a_\gamma \cdot \text{VIX}_t + b_\gamma, & a_\gamma, b_\gamma \in \mathbb{R}, \\ \rho_t & = \rho, & \rho \in [-1, 1], \end{cases} \tag{3.1}$$

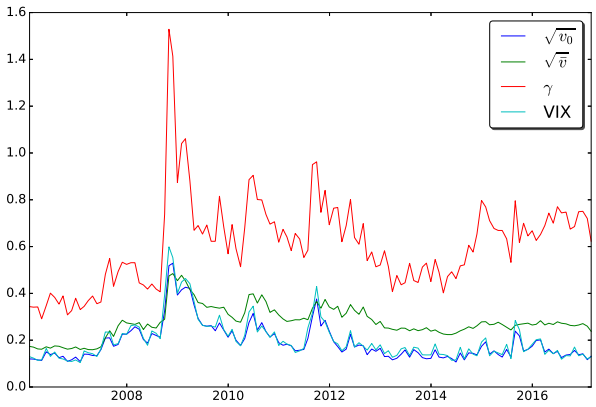

Figure 3.1: Development Heston parameters and VIX index.

where both the mean reversion parameter κ_t and the correlation coefficient ρ_t are assumed to be constant over time. The constant ρ assumption is justified by the fact that ρ displays a mean reverting pattern and it can therefore be approximated by its long-term mean. The constant κ assumption is justified by observations in Gauthier and Rivaille (2009). They argue that the effect on the implied volatility surface of increasing κ is similar to decreasing γ. Thus, allowing κ to change over time, unnecessarily overcomplicates the model. Moreover, numerical experiments show that an unrestricted κ leads to unstable results.

One is able to calibrate the constraint parameters through a procedure similar to (2.27), by changing the parameter set from Ω_t^{Heston} to $X = \{\kappa, a_{v_0}, b_{v_0}, a_{\bar{v}}, b_{\bar{v}}, a_\gamma, b_\gamma, \rho\}$ and summing over all points in time:

$$X = \underset{X^{\text{S}} \in X^{\text{Search}}}{\arg\min} \left(\sum_t \sum_K \sum_T \left(\sigma^{\text{Market}}(t,K,T) - \sigma^{\text{Heston}}(t, \Omega_t^{\text{Heston}}(X^{\text{S}}), K, T) \right)^2 \right), \tag{3.2}$$

where the search space is defined as

$$\begin{aligned} X^{\text{Search}} &= X_\kappa^{\text{Search}} \times X_{a_{v_0}}^{\text{Search}} \times X_{b_{v_0}}^{\text{Search}} \times X_{a_{\bar{v}}}^{\text{Search}} \times X_{b_{\bar{v}}}^{\text{Search}} \times X_{a_\gamma}^{\text{Search}} \times X_{b_\gamma}^{\text{Search}} \times X_\rho^{\text{Search}} \\ &= [0,10] \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times [-1,1]. \end{aligned} \tag{3.3}$$

By introducing additional parameters, the calibration procedure is no longer time-dependent. In other words, only one optimization procedure is required for the entire data set. This is in contrast with the procedure described in section 2.5, where each implied volatility surface is calibrated individually.

After the constraint parameters have been identified, this approach is able to deal with the challenges discussed in section 2.6. The Heston parameters can be calibrated through (3.1) by observing the VIX index, instead of performing a costly optimization procedure. Moreover, the VIX index can be implemented in real-world simulations, unlike the implied volatility surface. This way, one is able to recalibrate the Heston parameters within each simulated trajectory.

The results in Singor et al. (2014) were promising, but also left room for improvement. The R-squared error[1] (R^2) of the unrestricted model was 0.97, while the R^2 error of the VIX Heston model turned out to be 0.92. Moreover, the R^2 errors of different out-of-sample tests varied from 0.83 to 0.88. Hence, imposing the constraints of (3.1) led to a loss in accuracy. To this end, we will suggest some extensions of the VIX Heston model throughout this chapter. The extended models aim at improving the accuracy, while preserving the advantages of the original VIX Heston model.

3.2. New market indices

The VIX Heston model only considers the VIX index as explanatory variable, as it is highly correlated to the Heston parameters. However, the explanatory power of other market features has not been tested.

[1] A statistical measure indicating how well the observed outcomes are replicated by the outcomes of the model, with 1 corresponding to a perfect fit.

Therefore, in order to improve prediction accuracy, several additional market factors will be included. The main requirements for the new features are their ability to explain changes in the Heston parameters, their observability in the market and their ability to be simulated.

SKEW index

Multiple studies have shown the relationship between the implied volatility surface skewness and the Heston parameters, such as Janek et al. (2011) and Gauthier and Rivaille (2009). Gatheral (2006) even derives an analytical approximation in terms of the parameters for both short-dated and long-dated skewness. These results are also verified by our own numerical experiments. Figure 3.2 shows the relationship between ρ and the skewness of the observed implied volatility, which in this case is equal to the slope of the linear line connecting the implied volatility at 80% strike level and 120% strike level at one year maturity. This appears to be a good explanatory variable for ρ. However, it is very difficult to directly

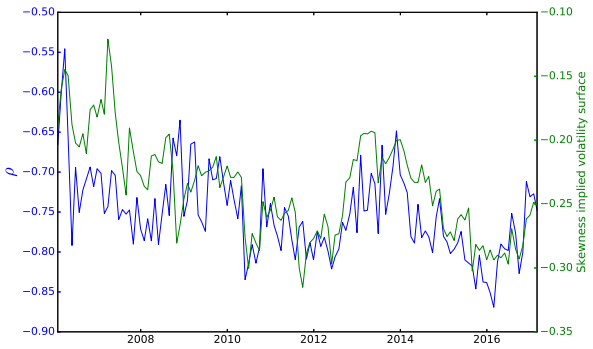

Figure 3.2: ρ vs. skewness of the implied volatility surface.

simulate the implied volatility surface, since it reflects the market's expectations, which depend on many market features (as explained in the previous section). In conclusion, this feature cannot be considered in the extended models, as it fails to meet one of the main requirements: it cannot be incorporated in simulations.

Fortunately, the CBOE has other measures, besides the VIX index. One of these measures is called the SKEW index. While the VIX index represents the market volatility, the SKEW index corresponds to the market skewness. The SKEW index is an indicator for the probability of a large loss, as its calculations involve out-of-the money S&P-500 options. Just like the VIX index, the SKEW index is observable on the market (due to the calculations of the CBOE) and easier to simulate (since it is a single index). In other words, it meets the requirements and can therefore be implemented in the extended models. We refer the reader to CBOE (2010) for further details regarding the SKEW index.

Return

Hibbert et al. (2008) shows there is a negative relation between returns of the S&P-500 index and changes in the implied volatility, which is directly linked to the Heston parameters. Hence, we expect there is a relationship between the returns of the S&P-500 index and the parameters of the Heston model as well. Moreover, one can imagine that the actual value of the S&P-500 index also has a relationship with the Heston model parameters, which is for example discussed in Mixon (2002). Typically, when the asset price is high, volatility is low and vice versa. This raises the belief that the S&P-500 index itself might be able to predict (some of the) changes in the parameters as well.

Interest rates

Hamrita and Trifi (2011) examined the relationship between interest rates, exchange rates and stock prices. They did not find a significant relationship between the exchange rates and the interest rates. However, they did discover a significant relationship between interest rates and stock prices. We already explained how the S&P-500 index might be related to the parameters of the Heston model. Following the same reasoning, one can include the (risk-free) interest rates as explanatory variables. For completeness, we will include the interest rates for several maturities: 3 months, 1 year and 2 years. These will be referred to as r_{90}, r_{360} and r_{720}, respectively. Furthermore, we expect the term spread to reflect the future

evolution of the market. The long-term interest rate is related to the long-term market expectations and the short-term interest rate is an indicator of the current state of the market, hence their difference reflects the expected evolution of the market. Thus, $r_d = r_{720} - r_{90}$ will also be considered as explanatory variable.

3.3. Regression models

In the VIX Heston model, a linear relationship between the VIX index and the Heston parameters is assumed. The main argument for this assumption is the high correlation between the index and parameters. However, different relationships were never explored. In this section we will discuss some other regression models that estimate the Heston parameters, possibly in a non-linear approach.

A common issue with regression models that have too many degrees of freedom is overfitting, in which case the model relies too heavily on the data set. When there is noise present in the observed data points (or when data is scarce), the regression model will include these noise terms, which can lead to unstable out-of-sample results. To prevent this phenomenon, we will adopt a K-fold cross-validation approach in our optimization procedures. In this approach, the data set is split into k parts. Hereafter, a regression is performed using only $k-1$ parts and the error is based on the kth part that was left out. This process is repeated until each of the k sets has been left out and the validation error will be equal to the average of all individual errors. This should prevent overfitting, as it only depends on out-of-sample errors.

3.3.1. Polynomial regression

A polynomial relationship seems to be the most logical extension of the VIX Heston model, as we do not want to lose the linear relationship that seems to be present. The main idea of polynomial regression is mapping the features into a polynomial feature space. For example, mapping two features into a second degree polynomial:

$$\phi_2\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 \\ x_2 \\ x_1 x_2 \\ x_1^2 \\ x_2^2 \end{bmatrix}. \tag{3.4}$$

After mapping the features into the desired polynomial degree, one can perform a simple linear regression to find the optimal parameters.

A major problem when using multiple features and polynomial regression is the dimensionality. For example, transforming eight features into a second degree polynomial, already gives a dimensionality of 44. It is very likely that not all transformed features will add explanatory power. To this end, we will sort the features based on their correlation with the target variables (in this case the Heston parameters). Hereafter, only the n_{feats} features with the highest correlation will be selected as explanatory variables.

To determine the optimal constraints for the polynomial regression, a two-stage calibration procedure is required. The first stage is qualitative, rather than quantitative. The objective of this stage is to identify:

- The market factors that predict the parameters best.
- The optimal polynomial degree.
- The optimal number of features used in the regression.

The optimization procedure follows a K-fold cross-validation approach as described above. The combination of market features, polynomial degree and number of components with the lowest cross-validation error will be considered optimal.

After the optimal set-up has been identified, we obtain constraints for the Heston model parameters,

similar to (3.1). In general, these constraints will be defined as

$$\Omega_t^{\text{Heston}}(X_{\text{Ex}}) = \begin{cases} \kappa_t & = \kappa, & \kappa \in \mathbb{R}_+, \\ v_{0,t} & = \left[c_{\{0,v_0\}} + \sum_{j=1}^{n_{\text{feats},v_0}} c_{\{j,v_0\}} \cdot P_{\{t,v_0\}}(j) \right]^2, & c_{\{i,v_0\}} \in \mathbb{R}, \\ \bar{v}_t & = \left[c_{\{0,\bar{v}\}} + \sum_{j=1}^{n_{\text{feats},\bar{v}}} c_{\{j,\bar{v}\}} \cdot P_{\{t,\bar{v}\}}(j) \right]^2, & c_{\{i,\bar{v}\}} \in \mathbb{R}, \\ \gamma_t & = c_{\{0,\gamma\}} + \sum_{j=1}^{n_{\text{feats},\gamma}} c_{\{j,\gamma\}} \cdot P_{\{t,\gamma\}}(j), & c_{\{i,\gamma\}} \in \mathbb{R}, \\ \rho_t & = \rho & \rho \in [-1,1], \end{cases} \tag{3.5}$$

where $P_{\{t,p\}}$ denotes the jth component for parameter p at time t with weight $c_{\{j,p\}}$. In the first stage, the regression components $P_{\{t,p\}}$ have already been identified for each parameter p. In the second stage, the component weights $c_{\{j,p\}}$ are optimized. The weights are considered optimal when the implied volatility surfaces generated by the model match the surfaces observed on the market as good as possible. To this end, we will perform a minimization procedure similar to (3.2),

$$X_{\text{Ex}} = \underset{X^{\text{S}} \in X^{\text{Search}}}{\arg\min} \left(\sum_t \sum_K \sum_T \left(\sigma^{\text{Market}}(t,K,T) - \sigma^{\text{Heston}}(t, \Omega_t^{\text{Heston}}(X^{\text{S}}), K, T) \right)^2 \right). \tag{3.6}$$

The search space X^{Search} depends on the components identified in the first stage. Similar to the unrestricted calibration procedure discussed in section 2.5, we will use the Levenberg-Marquardt least-squares algorithm to evaluate (3.6), with initial parameters based on the regression performed in the first stage. After performing the calibration procedure, one obtains the optimal weights for each component, which fully defines the characteristics of the model.

Finally, note that the VIX Heston model given in (3.1) is a special case of the current set-up. However, the polynomial regression approach allows for more explanatory variables. Potentially, this could lead to more accurate results, but it is a disadvantage from a computational point of view, as more parameters are involved in the calibration procedure.

3.3.2. Multi-output support vector regression

The literature regarding regression models is vast. In the past few years the so-called machine learning models have gained a lot of attention and popularity. These models are characterized by their ability to 'learn' a regression function based on training data and update this function as soon as new data becomes available. As the amount of available data is growing rapidly, these models are becoming more and more relevant.

Support vector regression is one of these machine learning algorithms, which is well-known and used in multiple studies predicting financial time-series, see for example Cao and Tay (2003) or Law and Shawe-Taylor (2017). Its attractive properties include accurate predictions in small data sets and dependence on only a few hyper parameters, making it relatively easy to identify the optimal set-up. The idea behind this model was first introduced in Cortes and Vapnik (1995), applied to the classification problem. Using the same principles, a regression based technique was proposed by Drucker et al. (1997).

In this section we will describe the intuition behind this approach, for more information regarding the history and derivation of support vector regression we refer the reader to Smola and Schölkopf (2004). Suppose we want to fit training samples $\{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)\}$ to a function f of the form

$$f(\mathbf{x}) = \phi(\mathbf{x})^T w + b, \tag{3.7}$$

with weight vector w, bias term b and feature map $\phi(\mathbf{x})$, which can correspond to a polynomial function, for example. Note that $\mathbf{x}_i$ can be multi-dimensional (which in our case corresponds to multiple market features), while $y_i \in \mathbb{R}$. In support vector regression, the goal is to define f such that each observed training sample falls inside the so-called ϵ-tube. This is visualized in figure 3.3. Moreover, in order to ensure robustness, we wish to create a flat model. In other words, this model aims at minimizing the size of w, giving

$$\begin{aligned} &\underset{w,b}{\text{minimize}} && \frac{1}{2}\|w\|^2, \\ &\text{subject to} && \begin{cases} y_i - \phi(\mathbf{x}_i)^T w - b \le \epsilon, \\ \phi(\mathbf{x}_i)^T w + b - y_i \le \epsilon, \end{cases} \quad i \in \{1, \ldots, N\}. \end{aligned} \tag{3.8}$$

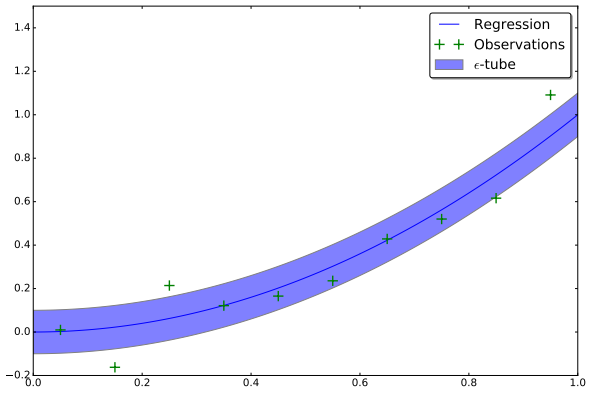

Figure 3.3: ϵ-tube for a polynomial kernel, with $\epsilon = 0.1$.

In some cases, however, this procedure does not yield a feasible solution. For example, when the data contains too much noise and/or when ϵ is too small. Moreover, in terms of overfitting, it might be wise to allow certain errors. To this end, we will introduce slack variables ξ_i and ξ_i^*, which can be thought of as the distance between y_i and the ϵ-tube. This way, the minimization procedure will always yield a solution. Of course, we wish to minimize the distance between the data set and the ϵ-tube, therefore we obtain

$$\begin{aligned} &\underset{w,b}{\text{minimize}} \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*), \\ &\text{subject to} \quad \begin{cases} y_i - \phi(\mathbf{x}_i)^T w - b & \leq \epsilon + \xi_i, \\ \phi(\mathbf{x}_i)^T w + b - y_i & \leq \epsilon + \xi_i^*, \\ \xi_i, \xi_i^* & \geq 0, \end{cases} \qquad i \in \{1,\dots,N\}. \end{aligned} \tag{3.9}$$

with positive regularization parameter $C \geq 0$, which corresponds to the importance of deviations from the ϵ-tube. Models with large values of C will focus on minimizing these deviations, while ignoring the flatness of w, resulting in more complex regression functions. Conversely, models with small values of C will lead to 'simpler' regression functions. One must find the optimal balance between the values of ϵ and C to prevent overfitting.

By using Lagrange multipliers, it can be shown that this minimization procedure is equivalent to the dual problem

$$\begin{aligned} &\underset{\boldsymbol{\alpha},\boldsymbol{\alpha}^*,b}{\text{maximize}} \quad -\frac{1}{2}\sum_{i,j=1}^{N}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\phi(\mathbf{x}_i)^T\phi(\mathbf{x}_j) - \epsilon\sum_{i=1}^{N}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{N} y_i(\alpha_i - \alpha_i^*), \\ &\text{subject to} \quad \begin{cases} \sum_{i=1}^{N}(\alpha_i - \alpha_i^*) = 0, \\ \alpha_i, \alpha_i^* \in [0, C], \end{cases} \qquad i \in \{1,\dots,N\}. \end{aligned} \tag{3.10}$$

Moreover, the prediction function can be written as follows

$$f(\mathbf{x}) = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)\phi(\mathbf{x}_i)^T\phi(\mathbf{x}_j) + b. \tag{3.11}$$

A key feature of this approach is the presence of the kernel function $\phi(\mathbf{x}_i)^T\phi(\mathbf{x}_j)$. This allows for mappings with unknown $\phi(\mathbf{x})$, but known dot product, such as the radial basis or sigmoid function. Finally, using the Karush-Kuhn-Tucker conditions, one is able to show that the coefficients corresponding to data-points inside the ϵ-tube are 0. In other words, the solution only depends on the data-points that lie outside the ϵ-tube, the so-called support vectors. Hence, the dimensionality of the solution does not depend on the size of the data set, but on the number of support vectors.

In this thesis we will use a modified version of this algorithm proposed in Xu et al. (2013), called multi-output least-squares support vector regression. This algorithm is an extension of the least-squares

support vector regression, which aims at minimizing *all* errors, instead of only minimizing the errors outside the ϵ-tube. This leads to a faster optimization procedure[2] and fewer hyper parameters, since ϵ does not need to specified. Moreover, the multi-output extension takes correlation between target variables into account. This might be beneficial, as the Heston parameters appear to be highly correlated. The remainder of this section will discuss the specifics of this approach.

Notation
In the following description we will use a few different functions, in line with the notation of Xu et al. (2013),

- For matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, define trace($\mathbf{A}$) as the sum of its diagonals,

$$\text{trace}(\mathbf{A}) = \sum_{i=1}^{n} \mathbf{A}(i,i). \tag{3.12}$$

  The notation $\mathbf{A}(c,d)$ refers to the element in the cth row and dth column of matrix $\mathbf{A}$.
- For matrix $\mathbf{A} \in \mathbb{R}^{k \times l}$, define repmat($\mathbf{A}, p, q$) as the $p \times q$ block matrix, with each element equal to $\mathbf{A}$.
- For matrices $\mathbf{A}_1, \ldots, \mathbf{A}_s \in \mathbb{R}^{k \times l}$ define blockdiag($\mathbf{A}_1, \ldots, \mathbf{A}_s$) as the block diagonal matrix equal to $\mathbf{A}_1$, $\ldots$, $\mathbf{A}_s$ on its diagonal blocks and 0 elsewhere.

Moreover, we define $\mathbf{1}_k = [1, \ldots, 1]^T \in \mathbb{R}^k$ and $\mathbf{1}_{k \times l} = [\mathbf{1}_k, \ldots, \mathbf{1}_k]^T \in \mathbb{R}^{k \times l}$. Similarly, we define $\mathbf{0}_k$ and $\mathbf{0}_{k \times l}$.

Now consider a set of training samples $\{(\mathbf{x}_1, \mathbf{y}_1), \ldots, (\mathbf{x}_N, \mathbf{y}_N)\}$ similar to the regular support vector regression, with the key difference that $\mathbf{y}_i \in \mathbb{R}^m$. Analogously to the original support vector regression we assume a feature mapping ϕ, which transforms the explanatory variables $\mathbf{x}_i$ (in our case the market features) into the desired feature space,

$$\phi : \mathbb{R}^{n_f} \mapsto \Phi, \tag{3.13}$$

where n_f denotes the number of explanatory variables and $\Phi \subseteq \mathbb{R}^{n_\phi}$ corresponds to the feature space with dimension n_ϕ. Moreover, it is assumed that the weight vectors for every target variable can be written as a combination of a shared term $\mathbf{w}_0$ and an individual term $\mathbf{v}_i$,

$$\mathbf{w}_i = \mathbf{w}_0 + \mathbf{v}_i, \qquad i \in \{1, \ldots, m\}, \tag{3.14}$$

where $\mathbf{w_0}, \mathbf{v_i} \in \mathbb{R}^{n_\phi}$ for $i \in \{1, \ldots, m\}$. It is expected that the more the target variables are correlated, the larger $\mathbf{w}_0$ will be compared to $\mathbf{v}_i$. Now consider the following matrices

$$\begin{cases} \mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_m] \in \mathbb{R}^{n_\phi \times m}, \\ \mathbf{W} = [\mathbf{w}_0 + \mathbf{v}_1, \ldots, \mathbf{w}_0 + \mathbf{v}_m] \in \mathbb{R}^{n_\phi \times m}, \\ \mathbf{b} = [b_1, \ldots, b_m]^T \in \mathbb{R}^m, \\ \Xi = [\xi_1, \ldots, \xi_m] \in \mathbb{R}^{N \times m}, \\ \mathbf{Z} = [\phi(\mathbf{x}_1), \ldots, \phi(\mathbf{x}_N)] \in \mathbb{R}^{n_\phi \times N}, \\ \mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_N]^T \in \mathbb{R}^{N \times m}. \end{cases} \tag{3.15}$$

These correspond to the individual weight, total weight, bias, slack, feature and observation matrices, respectively. Hence, the prediction function f can be written as

$$f(\mathbf{x}) = \phi(\mathbf{x})^T \mathbf{W} + \mathbf{b}. \tag{3.16}$$

Multi-output least-squares support vector regression
Similar to the original support vector regression, we wish to minimize the size of the weight vectors and the error between prediction and observation, leading to

$$\begin{aligned} \underset{\mathbf{w}_0, \mathbf{V}, \mathbf{b}}{\text{minimize}} \quad & J(\mathbf{w}_0, \mathbf{V}, \Xi) = \frac{1}{2}\mathbf{w}_0^T \mathbf{w}_0 + \frac{1}{2}\frac{\lambda_1}{m}\text{trace}(\mathbf{V}^T\mathbf{V}) + \frac{\lambda_2}{2}\text{trace}(\Xi^T \Xi), \\ \text{subject to} \quad & \mathbf{Y} = \mathbf{Z}^T\mathbf{W} + \text{repmat}(\mathbf{b}^T, N, 1) + \Xi. \end{aligned} \tag{3.17}$$

[2] Support vector regression results in a convex quadratic system, while the least-squares approach corresponds to a convex linear system.

The Lagrangian function corresponding to this problem is equal to

$$\mathscr{L}(\mathbf{w}_0,\mathbf{V},\mathbf{b},\mathbf{A}) = J(\mathbf{w}_0,\mathbf{V},\Xi) - \text{trace}(\mathbf{A}^T(\mathbf{Z}^T\mathbf{W} + \text{repmat}(\mathbf{b}^T,N,1) + \Xi - \mathbf{Y})), \tag{3.18}$$

with

$$\mathbf{A} = [\alpha_1,\ldots,\alpha_N] \in \mathbb{R}^{N\times m}, \tag{3.19}$$

the matrix of Lagrange multipliers. Following the Karush-Kuhn-Tucker conditions, we obtain

$$\begin{cases} \frac{\partial\mathscr{L}}{\partial\mathbf{w}_0} \Rightarrow \mathbf{w}_0 = \sum_{i=1}^m Z\alpha_i, \\ \frac{\partial\mathscr{L}}{\partial\mathbf{V}} \Rightarrow \mathbf{V} = \frac{m}{\lambda_1}\mathbf{Z}\mathbf{A}, \\ \frac{\partial\mathscr{L}}{\partial\mathbf{b}} \Rightarrow \mathbf{A}^T\mathbf{1}_N = \mathbf{0}_N, \\ \frac{\partial\mathscr{L}}{\partial\Xi} \Rightarrow \mathbf{A} = \lambda_2\Xi, \\ \frac{\partial\mathscr{L}}{\partial\mathbf{A}} \Rightarrow \mathbf{Z}^T\mathbf{W} + \text{repmat}(\mathbf{b}^T,N,1) + \Xi - \mathbf{Y} = \mathbf{0}_{N\times m}. \end{cases} \tag{3.20}$$

From (3.20) we can derive the relation

$$\mathbf{w}_0 = \frac{\lambda_1}{m}\sum_{i=1}^m \mathbf{v}_i, \tag{3.21}$$

which is proportional to the average of all individual terms. Hence, (3.17) is trying to find the optimal balance between the size of the individual weight vectors and the closeness of these vectors to their average. This way, the algorithm takes the correlation between different components into account.

Hereafter, from (3.20), one is able to eliminate $\mathbf{W}$ and Ξ, resulting in the following system of equations

$$\begin{bmatrix} \mathbf{0}_{Nm\times m} & \mathbf{P}^T \\ \mathbf{P} & \mathbf{H} \end{bmatrix}\begin{bmatrix} \mathbf{b} \\ \alpha \end{bmatrix} = \begin{bmatrix} \mathbf{0}_m \\ \mathbf{y} \end{bmatrix}, \tag{3.22}$$

with

$$\begin{cases} \mathbf{P} = \text{blockdiag}(\mathbf{1}_N,\ldots,\mathbf{1}_N) \in \mathbb{R}^{Nm\times m}, \\ \mathbf{H} = \text{repmat}(\mathbf{Z}^T\mathbf{Z},m,m) + \frac{1}{\lambda_2}\mathbf{I}_{Nm} + \frac{m}{\lambda_1}\text{blockdiag}(\mathbf{Z}^T\mathbf{Z},\ldots,\mathbf{Z}^T\mathbf{Z}) \in \mathbb{R}^{Nm\times Nm}, \\ \alpha = [\alpha_1^T,\ldots,\alpha_N^T] \in \mathbb{R}^{Nm}, \\ \mathbf{y} = [\mathbf{y}_1^T,\ldots,\mathbf{y}_N^T] \in \mathbb{R}^{Nm}. \end{cases} \tag{3.23}$$

Xu et al. (2013) discussed the difficulties of solving this system, as the obtained matrix is not positive definite. To this end, they propose to rewrite the system, leading to

$$\begin{bmatrix} \mathbf{P}^T\mathbf{H}^{-1}\mathbf{P} & \mathbf{0}_{Nm\times Nm} \\ \mathbf{0}_{m\times m} & \mathbf{H} \end{bmatrix}\begin{bmatrix} \mathbf{b} \\ \mathbf{H}^{-1}\mathbf{P}\mathbf{b} + \alpha \end{bmatrix} = \begin{bmatrix} \mathbf{P}^T\mathbf{H}^{-1}\mathbf{y} \\ \mathbf{y} \end{bmatrix}, \tag{3.24}$$

This system can be solved by performing the following steps:

1. Solve η and ν from $\mathbf{H}\eta = \mathbf{P}$ and $\mathbf{H}\nu = \mathbf{y}$.
2. Determine optimal solution: $\begin{cases} \mathbf{b}^* = (\mathbf{P}^T\eta)^{-1}\eta^T\mathbf{y}, \\ \alpha^* = \nu - \eta\mathbf{b}. \end{cases}$

The systems in step 1 can be solved efficiently with, for example, Cholesky decompositions, as $\mathbf{H}$ is positive definite. Finally, we can rewrite the regression function as

$$\begin{aligned} f(\mathbf{x}) &= \phi(\mathbf{x})^T\mathbf{W} + \mathbf{b} \\ &= \text{repmat}\left(\sum_{i=1}^m\sum_{j=1}^N \alpha_{i,j}^*\phi(\mathbf{x})^T\phi(\mathbf{x_j})\right) + \frac{m}{\lambda_1}\sum_{j=1}^N \alpha_j\phi(\mathbf{x})^T\phi(\mathbf{x_j}) + \mathbf{b}^{*T}. \end{aligned} \tag{3.25}$$

Again, similar to the original support vector regression, this function only depends on the dot product $\phi(\mathbf{x})^T\phi(\mathbf{x_j})$. Thus, one only needs to specify $\phi(\mathbf{x})^T\phi(\mathbf{x_j})$, which allows for more involved kernel functions. For more information regarding this approach and the difference between the regular support vector regression, we refer the reader to Xu et al. (2013).

3.3.3. Gradient boosted regression trees

In this section we will present another well-known machine learning algorithm: gradient boosted regression trees. As indicated by its name, this algorithm is based on regression trees. Before we will introduce the concept of gradient boosted regression trees, we will first discuss the specifics of an individual regression tree.

Regression trees

A regression tree is an algorithm which aims at splitting a data set into two or more subsets. The process of splitting is done in a most significant approach, based on the input variables. After splitting the data set multiple times, one obtains the final subsets, called leafs in regression tree terminology. When a new data point is observed, one must find its corresponding leaf, by propagating it through the predefined regression tree. The prediction of this new observation is equal to the average over all previously observed data-points belonging to that leaf. We will provide an example in order to visualize this process.

Suppose we want to predict v_0 based on the VIX index and the long-term interest rate r_{720}. Firstly, this requires a training set containing samples of v_0 with the corresponding VIX and r_{720}. Based on this training set, a regression tree is defined. A visualization of a possible tree can be found in figure 3.4. Friedman (1979) discusses how the shape of these trees can be determined and its methodologies are still

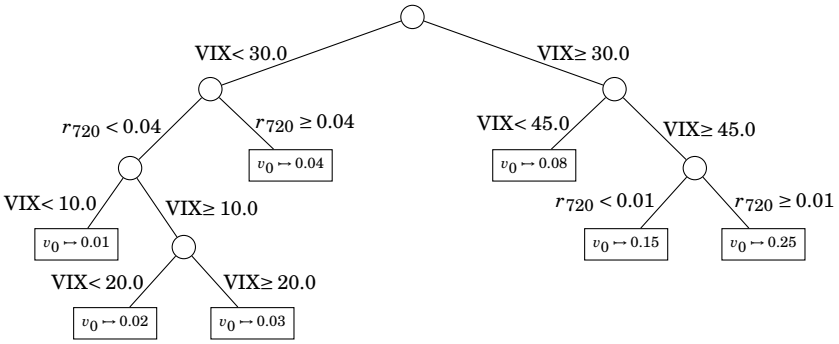

Figure 3.4: A possible realization of a regression tree, which predicts v_0 based on the VIX index and long-term interest rate.

relevant.

The first step is to identify the splitting rule. In other words, one needs to determine which feature (in this example VIX or r_{720}) and what value is used to split the data set. Ideally, the split should improve the accuracy as much as possible. To this end, Friedman (1979) proposes a mean squared error approach. Suppose we are at node i, containing subset J of the entire data set. We can estimate a linear regressor using all explanatory variables X_j and target variables y_j belonging to subset $j \in J$

$$f(X_j) = A^T X_j + b, \qquad j \in J, \tag{3.26}$$

with residuals

$$r_j = y_j - f(X_j). \tag{3.27}$$

If the data set is split into a left and right subset J_r and J_l, based on feature p, then the linear estimator for both subsets can be redetermined by adjusting the component for feature p. We obtain

$$f^p(X_j) - f(X_j) = \begin{cases} A_l(p)X_j(p) + b_l, & \text{if } j \in J_l, \\ A_r(p)X_j(p) + b_r, & \text{if } j \in J_r. \end{cases} \tag{3.28}$$

The goal of the splitting procedure is to minimize the residuals as much as possible. To this end, expression (3.28) is fitted to the residuals of (3.27) by a linear regressor. Feature p and associated splitting value v_p with the lowest squared error are considered optimal. For example, in case of the first node of figure 3.4 we obtain $p = \text{VIX}$ and $v_p = 30.0$.

The next step is determining when to stop the splitting process. If the amount of leafs is equal to the amount of data points, then one obtains a perfect fit on the training set. It is, however, questionable if

this will lead to robust results, as we might be overfitting. There are several ways to prevent this. A naive approach is to define a minimum amount of samples required for a split (n_{split}), a minimum amount of samples to become a leaf (n_{leaf}) or a maximum depth (n_{depth}). However, Friedman (1979) argues that this naive approach is suboptimal, as one only wishes to perform a split if it adds accuracy. Instead, he proposes to first determine a full tree (according to the naive approach) and consequently remove nodes that do not improve the model enough. In the gradient boosted regression trees model this is not as important, since it only consists of shallow trees with low depth. Nevertheless, it is important to identify the optimal n_{split}, n_{leaf} and n_{depth}.

Finally, the predictions for each leaf are simply defined as the average value of all training points belonging to that leaf. Newly observed variables are propagated through the tree and their predictions are equal to the value of the corresponding leaf. For example, the regression tree in figure 3.4 would predict $v_0 = 0.03$ in case of a data point with VIX $= 25.0$ and $r_{720} = 0.02$.

Gradient boosted regression trees
Regression trees are rarely used individually due to their limited structure: a single regression tree will always return a step-function. Instead, most tree based approaches involve multiple regression trees. The gradient boosted regression trees algorithm as proposed in Friedman (2001) is an example of such an approach. The principles of gradient boosting can be applied to any regression model, but it is most commonly used in combination with regression trees.

Again, we assume a training set of samples $\{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)\}$, which is used to determine regression function $f(\cdot)$. With gradient boosting one first estimates a shallow base learner $f_0(\cdot)$. In case of gradient boosted regression trees this corresponds to a tree with low depth. The residuals of this base learner are defined as

$$r_{0,i} = f_0(\mathbf{x}_i) - y_i, \qquad i \in \{1, \ldots, N\}. \tag{3.29}$$

Typically, the residuals of the base learner are large, as it is only a shallow representation of the data set. Hereafter, this base learner is updated according to the gradient of the error function. For example, if one assumes a squared error function, this gradient simply corresponds to the residuals of the base learner. Hence, a new (shallow) regression tree is estimated on the data set $\{(\mathbf{x}_1, r_{0,1}), \ldots, (\mathbf{x}_N, r_{0,N})\}$. This regression tree aims at predicting the residuals as defined in (3.29). The base learner is updated according to the newly estimated regression tree $f_1^r(\cdot)$ and learning rate η

$$f_1(\mathbf{x}_i) = f_0(\mathbf{x}_i) + \eta f_1^r(\mathbf{x}_i), \qquad i \in \{1, \ldots, N\}. \tag{3.30}$$

This process can be repeated on the newly defined residuals. After n_{est} repetitions the process stops and we obtain

$$f(\mathbf{x}_i) = f_0(\mathbf{x}_i) + \eta \sum_{j=1}^{n_{\text{est}}} f_j^r(\mathbf{x}_i), \qquad i \in \{1, \ldots, N\}. \tag{3.31}$$

For generalization purposes, Friedman (2002) suggests to use only a random part of the entire data set at each iteration. By using only a random fraction s of the entire data set in the estimation of each individual regression tree, one prevents overfitting and improves out-of-sample accuracy.

In conclusion, the regression trees require optimization of the hyper parameters n_{split}, n_{leaf} and n_{depth}. Moreover, the gradient boosting approach demands optimization of n_{est}, η and s, for optimal out-of-sample accuracy. These hyper parameters will be estimated with a K-fold cross validation approach as discussed at the start of this section. Finally, note that this algorithm is only able to estimate one variable at a time. Therefore, each target variable (in our case the Heston parameters) requires an individual regressor.

3.4. Results

In this section we will present the results of the regression models proposed in section 3.3, using the market features discussed in 3.2. We will compare the accuracy of all three regression models to the original VIX Heston model discussed in section 3.1.

3.4.1. Data

The data set will be split into two subsets: a training set and a test set. The training set will be used to determine the optimal set-up of each model. The test set will assess the accuracy of each model and test for overfitting.

The training set contains daily implied volatility surfaces of the S&P-500 European put and call options from January 2006 until February 2014. Each implied volatility surface contains 5 different strike levels (80%, 90%, 100%, 110% and 120% of S_0) and maturities (0.25, 0.5, 1, 1.5 and 2 years). For computational purposes, we will use monthly data to train the VIX Heston, polynomial regression and multi-output least-squares support vector regression. The daily data set will be used to train the gradient boosted regression trees, as it is more difficult to train and therefore requires more samples. The test set contains monthly implied volatility surfaces of the S&P-500 European options from March 2014 until February 2017. This set will only be used to assess the accuracy of the trained regression models.

Finally, to assess the robustness of the approaches, the models will be applied to monthly implied volatility surfaces of the FTSE-100 (United Kingdom) and STOXX-50 (Europe) as well. The training set includes data from October 2010 until June 2015 and the test set contains data from July 2015 until February 2017. We will not be able to accurately train the gradient boosted regression trees approach for the UK and Europe data sets, as they only contain monthly data. Therefore, we will only apply the gradient boosted regression trees to the US data set.

3.4.2. Optimal set-up

Before we are able to compare the different regression models, the optimal set-up for each approach must be identified. These set-ups include the trained hyper parameters and the optimal market features used to predict the Heston parameters. The optimal set-ups are identified according to the K-fold cross validation methods discussed in section 3.3. In this section, we will discuss some key properties of each model. The results below are based on the US data set. The results of the UK and Europe data set can be found in appendix B.

VIX Heston

The proposed regression models will be compared to the VIX Heston model proposed in Singor et al. (2014). Hence, we require the optimal regression coefficients of this model. These are estimated according to (3.6), giving

$$\Omega_t^{\text{Heston}}(X) = \begin{cases} \kappa_t & = 1.0, \\ v_{0,t} & = (0.0140 + 0.0090 \cdot \text{VIX}_t)^2, \\ \bar{v}_t & = \left(0.0957 + 0.0087 \cdot \text{VIX}_{\text{filter}_t}\right)^2, \\ \gamma_t & = 9.6479 \cdot 10^{-5} + 0.0270 \cdot \text{VIX}_t, \\ \rho_t & = -0.7294. \end{cases} \tag{3.32}$$

Polynomial regression

The optimal polynomial degree, number of features and market features are determined in the first stage of the polynomial regression model. These are identified by a K-fold cross validation approach and the results are presented in table 3.1. In this case, it appears that the polynomial regression model is

Parameter	Market features	Polynomial degree	Number of features
v_0	VIX	1	1
$\bar{v}$	$\text{VIX}_{\text{filter}}$, r_{720}	1	2
γ	VIX	1	1

Table 3.1: Optimal set-up polynomial regression model.

almost similar to the VIX Heston model. The main difference lies in the prediction of $\bar{v}$, which uses r_{720} in addition to $\text{VIX}_{\text{filter}}$. Moreover, no degree higher than 1 is considered optimal. Hence, the linear relationship found in Singor et al. (2014) appears to be optimal. The regression coefficients belonging to

these features can be estimated with the calibration procedure described in section 3.3.1. This yields

$$\Omega_t^{\text{Heston}}(X) = \begin{cases} \kappa_t & = 1.0, \\ v_{0,t} & = (0.0152 + 0.0089 \cdot \text{VIX}_t)^2, \\ \bar{v}_t & = \left(0.1461 + 0.0074 \cdot \text{VIX}_{\text{filter}_t} - 0.01394 \cdot r_{720}\right)^2, \\ \gamma_t & = 0.0874 + 0.0231 \cdot \text{VIX}_t, \\ \rho_t & = -0.7398. \end{cases} \tag{3.33}$$

This model proposes a negative relationship between $\bar{v}$ and r_{720}, which is an intuitively appealing result. Typically, when the implied volatility is high, the market is unstable. As a result, traders are willing to accept lower risk-free rates. Conversely, low implied volatility indicates a stable market. Hence, traders are more inclined to invest their money in risky assets, unless the risk-free rate is high enough.

Multi-output least-squares support vector regression
An attractive property of the multi-output least-squares support vector regression approach is the presence of only a few hyper parameters, as all Heston parameters are estimated simultaneously. The most important input variables are the type of kernel and market features. The algorithm also depends on the two regularization parameters, but these do not appear to be as important in this case.

A linear kernel is considered optimal in this approach, similar to the VIX Heston and polynomial regression models. Moreover, the algorithm is most accurate when using the market features VIX, $\text{VIX}_{\text{filter}}$ and r_{720}, which is similar to the polynomial regression model. The estimated regularization parameters are equal to

$$\lambda_1 = 10^{-8}, \qquad \lambda_2 = 10^{-10}. \tag{3.34}$$

Moreover, the Heston parameters κ and ρ are constant and equal to the mean of all previously observed data points.

Gradient boosted regression trees
The gradient boosted regression trees algorithm contains many hyper parameters, as discussed in section 3.3. It is computationally demanding to simultaneously tune all these parameters. We therefore propose a stage-wise approach.

First some initial n_{split}, n_{leaf}, n_{depth}, n_{est}, η and s are defined. The optimal market features are identified according to the K-fold cross validation, using these initial estimates. Then, the optimal n_{split}, n_{leaf}, n_{depth} are identified according to the previously optimized market features. Hereafter, n_{est} is determined in a similar way and finally s is estimated. This process has been repeated for different values of η, but the final accuracy appears to be almost unaffected by the learning rate.

In terms of market features, this approach has similar results to the polynomial regression and multi-output least-squares support vector regression, which can be seen in table 3.2. Despite the fact that this

Parameter	Market features	n_{split}	n_{leaf}	n_{depth}	n_{est}	η	s
v_0	VIX, $\text{VIX}_{\text{filter}}$	5	1	3	140	0.1	0.3
$\bar{v}$	$\text{VIX}_{\text{filter}}$, r_{720}	2	2	1	310	0.05	0.05
γ	VIX, $\text{VIX}_{\text{filter}}$, r_{720}	28	2	2	790	0.05	0.1

Table 3.2: Optimal set-up gradient boosted regression trees.

approach does not predefine a linear/non-linear relationship between the Heston parameters and market features, it still selects the VIX index and interest rates as most important. This indicates the robustness of the selected market features.

Again, the Heston parameters κ and ρ are constant and equal to the mean of all previously observed data points.

3.4.3. Accuracy

In this section we will compare the previously described models in terms of accuracy. Ideally, the models produce implied volatility surfaces equal to the ones observed on the market. In reality, this will be impossible due to limitations of the Heston model and errors in the Heston parameter predictions. To this end, we will first perform an unrestricted calibration procedure for each time-step as described in section 2.5. The unrestricted model can be thought of as optimal and its errors are caused by limitations of the Heston model. Consequently, error differences between the regression models and the unrestricted model are the result of suboptimal Heston parameter predictions.

As described in section 3.4.1, we have separated the data set in a training set and a test set. The training set has identified the optimal set-ups, as discussed in section 3.4.2. In this section, these set-ups are used to predict the implied volatility surfaces of the test set. The accuracy will be assessed by comparing the predicted to the observed implied volatility surfaces, according to the following error measures

$$\begin{cases} \mathrm{SSE}(M) = \sum_t \sum_K \sum_T \left(\sigma^{\mathrm{Market}}(t,K,T) - \sigma^{\mathrm{Heston}}(t,\Omega_t^{\mathrm{Heston}}(M),K,T)\right)^2, \\ \mathrm{MAE}(M) = \frac{1}{N_\sigma} \sum_t \sum_K \sum_T \left|\sigma^{\mathrm{Market}}(t,K,T) - \sigma^{\mathrm{Heston}}(t,\Omega_t^{\mathrm{Heston}}(M),K,T)\right|, \\ R^2(M) = 1 - \frac{\mathrm{SSE}(M)}{\sum_t \sum_K \sum_T \left(\sigma^{\mathrm{Market}}(t,K,T) - \overline{\sigma}^{\mathrm{Market}}\right)^2}, \\ R^2_{\mathrm{Min}}(M) = \min_t \left\{R^2(M)_t : t \in [t_{\min}, t_{\max}]\right\}, \end{cases} \tag{3.35}$$

where $\Omega_t^{\mathrm{Heston}}(M)$ is defined as the predicted parameter set of regression model M, N_σ is the total number of observed implied volatilities and $\overline{\sigma}^{\mathrm{Market}}$ is the average of all observed implied volatilities. Moreover, we will assess the models' prediction accuracy with respect to the individual Heston parameters. For parameter p we determine

$$\mathrm{MSE}_p(M) = \frac{1}{N_p} \sum_t \left(p_{\mathrm{opt}}(t) - p(M,t)\right), \tag{3.36}$$

with $p_{\mathrm{opt}}(t)$ and $p(M,t)$ defined as the parameter according to the unrestricted and regressions models at time t, respectively. Moreover, N_p denotes the total number of data points. The results of these models are displayed in table 3.3.

Model	SSE	MAE	R^2	R^2_{Min}	MSE_{v_0}	$\mathrm{MSE}_{\bar{v}}$	MSE_{γ}
VIX Heston	0.2286	0.0124	0.8948	0.8159	$1.374 \cdot 10^{-5}$	$3.008 \cdot 10^{-4}$	0.0670
Polynomial Regression	0.2567	0.0139	0.8819	0.7573	$1.282 \cdot 10^{-5}$	$1.034 \cdot 10^{-4}$	0.0533
Multi-output SVR	**0.1471**	**0.0105**	**0.9291**	**0.8404**	$\mathbf{1.802 \cdot 10^{-6}}$	$\mathbf{7.164 \cdot 10^{-5}}$	**0.0317**
GB regression trees	0.2510	0.0135	0.8845	0.6913	$4.425 \cdot 10^{-6}$	$1.747 \cdot 10^{-4}$	0.0326
Unrestricted	*0.0173*	*0.0034*	*0.9916*	*0.9811*	-	-	-

Table 3.3: Out-of-sample accuracy of the regression models according to the error measures defined in (3.35) and (3.36).

The multi-output support vector regression appears to outperform the other regression models according to all error measures. The multi-output distinguishes itself from the other approaches by taking correlation between the parameters into account. This results in more accurate parameter predictions and consequently in more accurate implied volatility surface predictions. The evolution of the predicted parameters compared to the optimal parameters is represented in figure 3.5. Despite the increased accuracy, the regression models still lose accuracy compared to the unrestricted model. On average, there is an error of 0.003 between the implied volatility and the unrestricted Heston model. The multi-output support vector regression has an average absolute error of 0.01. So on average we introduce an additional error of 0.007, by implementing the regression model.

In terms of the implied volatility surface, the VIX Heston model is more accurate than polynomial regression and gradient boosted regression trees. However, in terms of parameter predictions, it is the worst performing model. This phenomenon can be explained the degrees of freedom in the Heston model. For example, the effect of an underestimated v_0 can be (partially) nullified by an overestimated $\bar{v}$. This

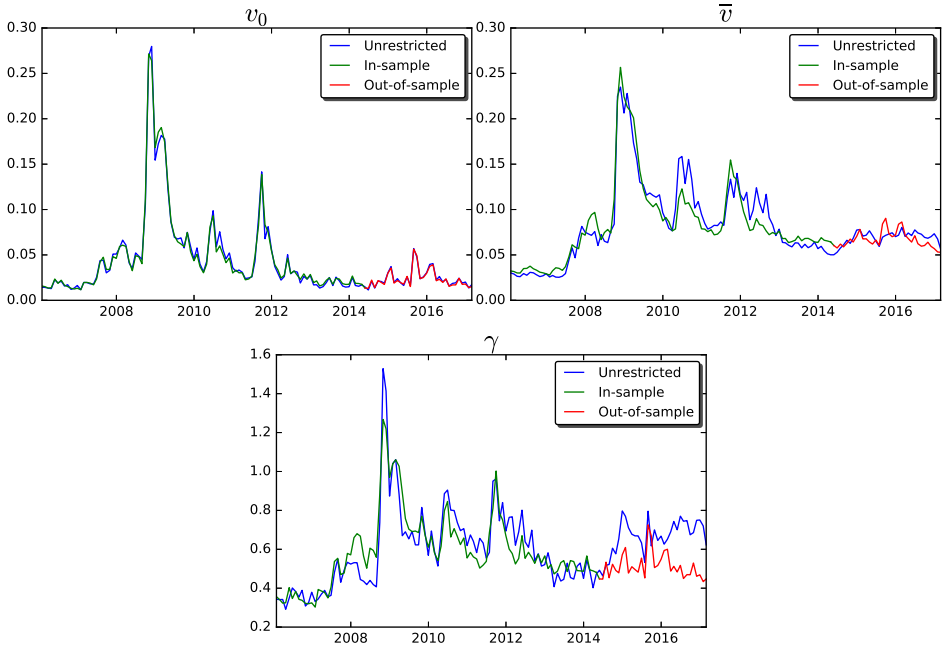

Figure 3.5: Unrestricted versus multi-output least-squares support vector regression predicted parameters over time.

way, one obtains more accurate implied volatility surface predictions, despite the more inaccurate Heston parameter predictions. One could, however, question the robustness of this phenomenon: are the more accurate predictions the result of a 'lucky' guess or do they appear structurally over time? At this point we can therefore not distinguish between the VIX Heston model, polynomial regression and gradient boosted regression trees on the US data set.

The out-of-sample results of the UK and Europe data sets can be found in appendix B. Similar to the US data set, the multi-output support vector regression approach is considered optimal in both the UK and Europe data set. In the Europe data set, the VIX Heston model is superior to polynomial regression in terms of implied volatility surface prediction, despite the fact that it is outperformed by the other models in terms of parameter predictions. This indicates a more structural superiority of the VIX Heston model over polynomial regression. However, as the differences between the models are relatively small, it remains difficult to draw conclusions from these results.

Finally, we would like to discuss the predictions with respect to γ. In the US data set, it appears to be quite difficult to make an accurate prediction of γ (see figure 3.5), as the out-of-sample correlation to the VIX index is much lower than the in-sample correlation (0.86 in-sample versus 0.32 out-of-sample). In the UK and Europe data sets, this phenomenon does not seem to be present and consequently the predictions of γ are much more accurate (see figures B.1 and B.2). So there appears to be another, yet unknown, factor driving γ in the US data set, which is absent in the UK and Europe data set.

In summary, we conclude that the multi-output support vector regression model predicts the Heston parameters most accurately in all data sets. This also leads to the most accurate implied volatility surface predictions in all data sets. In the UK and Europe data sets, error differences are small and it is difficult to distinguish the models. However, based on the US data set, we recommend the multi-output support vector regression, as it produces the most accurate predictions and is the most robust among different data sets.

3.4.4. Improving the accuracy

The accuracy results in tables 3.3, B.1 and B.2 indicate that the parameter predictions of the regression models are suboptimal. Indeed, the forecasted implied volatility surfaces do not match the ones observed on the market as good as the unrestricted approach. Using these parameters when pricing options could

lead to pricing errors and arbitrage opportunities. To this end, we wish to improve the accuracy of these predictions by allowing stochastic errors, in addition to the previously discussed regression models. The stochastic components can, for example, be analysed with the unscented Kalman filter. By doing so, one obtains an accurate estimate of the Heston parameters, while decreasing calibration time. Moreover, the solution no longer depends on the initial parameters, which removes the need for re-sampling initial values to obtain the most accurate results. The details of the unscented Kalman filter and how it can be combined with the regression models proposed in this chapter are discussed in appendix C. However, note that it is difficult to implement this approach in risk-management applications, as it requires the implied volatility surface.

4

Hedge test

In the previous chapter we discussed how to determine the risk-neutral measure (and the Heston parameters) conditioned on the market indices as seen on that specific date. By doing so, we implicitly assume that the risk-neutral measure is time-dependent, since the market indices change over time. This assumption, however, leads to a paradox; on one hand, we accept the Heston model and its assumptions, including constant parameters. But on the other hand we desire to produce market consistent prices at any point in time, hence we recalibrate the (constant) model parameters as soon as the observed implied volatility surface changes. This raises the following question: 'Is it justified to assume a recalibrated risk-neutral measure in (real-world) simulations, even when this is in conflict with the assumptions of the underlying risk-neutral pricing model?' This leads to an interesting and difficult discussion, one which has been discussed from different point of views in the existing literature.

In Rebonato (2004) this topic is treated extensively throughout multiple chapters, mostly from a trader's point of view. In his argument, the input of a model should reproduce future implied volatility surfaces as similar as possible to the ones encountered in the market. In other words, the best models and calibration methods are the ones that require as little future recalibration as possible. As true as this statement might be, to this day there is no model completely invariant to changes in the implied volatility surface, hence recalibration will always be required in order to remain market consistent and arbitrage-free. Regarding the concept of arbitrage, Rebonato explains that it is often forgotten what arbitrage-free prices guarantee in practice. He states:

> "A constant-volatility model, a jump–diffusion model and a local-volatility model can all be constructed to be arbitrage-free, yet they all predict very different present and future option prices, and the trader who knew which one was the *true* model could make unlimited profits at the expense of the users of the other (arbitrage-free) models."

He therefore recommends that today's option prices should be accurately, but not perfectly, recovered, even if this leads to a theoretical arbitrage. For example, a hedger[1] is not necessarily interested in the current option price, but in future option prices. A model that perfectly replicates today's implied volatility surface, but fails at generating reliable future implied volatility surfaces (such as local-volatility models) will therefore not be useful to a hedger. Instead, the hedger prefers a model that more accurately captures the dynamics of the implied volatility surface (and equivalently the option price), even if it fails to capture the current implied volatility surface and leads to a theoretical arbitrage. This, however, does require knowledge of the future state of the market, something of which we cannot be certain (unless the market behaves similarly to historical observations).

Ayache (2015) has a more philosophical point of view. He argues that the constant parameters subjected to recalibration cannot be described by stochastic differential equations, since these processes are also parameter-dependent. The parameters of these newly defined processes are subjected to recalibration as well and should therefore be described by other stochastic processes. This leads to a never-ending

[1] The principle of hedging will be explained in section 4.1.

loop of processes to describe newly introduced parameters. Instead, he claims that a regime-switching model is perhaps the only solution to the recalibration problem, possibly with an infinite amount of regimes. Regarding the difficulty of infinitely many regimes he says:

> "The whole idea is to stop thinking that this infinity of regimes is progressively revealed over time [...] and to start thinking that this infinity is virtually present from the beginning"

He therefore claims in Ayache (2016) that one should hedge against changes of all parameters, even when the model assumes some of them to be constant. This out-of-model hedging (as Rebonato (2004) calls it) is important because one is not hedging against changes in the underlying process but against changes of the option price. However, with his philosophical point of view, he does not discuss how this regime-switching model and the out-of-model hedges are implemented in practice.

A practical point of view regarding these out-of-model hedges is presented in Alexander et al. (2009). They acknowledge that recalibration is "one of the main sources of model risk in option pricing models". Using historical data and a Delta-Gamma hedge, they show that ignoring recalibration results in poor hedging performances. To this end, they assumed the recalibrated parameters to follow stochastic processes[2], which efficiently improved the hedging performance. This way, they implicitly show that one should consider a recalibrated risk-neutral measure, even when the assumption of dynamic parameters is in violation with the underlying risk-neutral pricing model.

In summary, the aforementioned points of view all agree upon one thing: none of the existing option models is able to fully capture the underlying process, because it is impossible to know the future state of the market and therefore every model requires recalibration. Moreover, they all discuss the effect of recalibration on hedging strategies and they argue that it is important to take recalibration into account when creating a portfolio. Yet, they disagree on how to implement this in practice, to summarize

- Rebonato (2004) suggests risk-neutral pricing models and calibration methods that require as little future recalibration as possible.
- Ayache (2015) thinks of the recalibration process as a regime-switching model, possibly with an infinite amount of regimes and states that it is important to hedge against changes of the option price, even if this violates the assumptions of the risk-neutral pricing model.
- Alexander et al. (2009) assumes that the parameters of the risk-neutral pricing method can be described by stochastic differential equations. By exploiting correlations between the processes, they propose an out-of-model hedge that takes changes of these parameters into account.

In this chapter, we will investigate the effect of different assumptions with regard to the risk-neutral measure (such as recalibration) on certain hedging strategies. First, we will explain the basic idea of hedging. Hereafter, we will discuss different market models and hedging strategies we will implement, some of which are similar to Alexander et al. (2009). Finally, the strategies will be applied to historical and simulated data and we will discuss the results and their implications.

4.1. Principles of hedging

The concept of hedging is important and it plays a crucial role in option pricing methods (and therefore the risk-neutral measure). The idea of hedging is simple: we wish to create a portfolio without risk. In practice, it can be quite challenging to create a truly risk-neutral portfolio, which we will show in this chapter. But first, in order to illustrate the hedging principle, we will consider a simplified example.

In this example, we will assume an initial stock price $S_0 = 10$. Furthermore, at time $T = 1$ the stock has probability 0.6 to take value $S_1^{(1)} = 12$ and probability 0.4 to take value $S_1^{(2)} = 6$. Finally, assume the existence of a European call option on the stock with unknown initial value C_0, maturity $T = 1$ and strike price $K = 10$. Consequently, the option price at time $T = 1$ is equal to its pay-off: $C_1^{(1)} = 2$ and $C_1^{(2)} = 0$. Figure 4.1 visualizes this example in the form of a binary tree. Now, we wish to create a hedge portfolio containing a long position in Δ stocks and a short position in the option. In order to be risk-neutral, the portfolio needs to attain the same value in both scenarios:

$$\Pi_1^{(1)} = \Pi_1^{(2)},$$

[2] which is in conflict with the discussion of Ayache (2015)

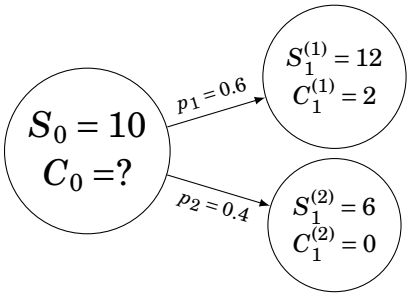

Figure 4.1: Binary tree describing the simplified market.

$$\begin{aligned} \Delta S_1^{(1)} - C_1^{(1)} &= \Delta S_1^{(2)} - C_1^{(2)}, \\ \Delta 12 - 2 &= \Delta 6, \\ \Delta &= \frac{1}{3}. \end{aligned} \tag{4.1}$$

In both cases, the final value of the portfolio is equal to 2 and it is therefore risk-neutral. Now assume there exists another risk-neutral asset in the market (for example a bank account or government bond) with return r, while the hedge portfolio has return r_Π. If we assume the two returns are not equal, there are two possibilities:

- $r > r_\Pi$: By taking a short position in the portfolio and a long position in the risk-neutral asset one earns $r - r_\Pi$. Since both assets are free of risk, this leads to an arbitrage opportunity.
- $r < r_\Pi$: Similarly, a long position in the portfolio and a short position in the risk-neutral asset leads to arbitrage.

Consequently, under the assumption of an arbitrage-free market, the portfolio has to grow with the risk-neutral rate. If we assume $r = 0.01$, the initial value of the portfolio must be equal to

$$\Pi_0 = \Pi_1 e^{-rt} = 2e^{-0.01} = 1.98. \tag{4.2}$$

This leads to the unique arbitrage-free option price

$$\begin{aligned} \Pi_0 &= \Delta S_0 - C_0, \\ 1.98 &= \frac{1}{3}10 - C_0, \\ C_0 &= 1.35. \end{aligned} \tag{4.3}$$

Remember that we can also write the option price as the discounted expected pay-off under the risk-neutral measure:

$$\begin{aligned} C_0 &= e^{-rt}\mathbb{E}^{\mathbb{Q}}[C_1], \\ 1.35 &= 2e^{-0.01}q_1 + 0e^{-0.01}q_2, \\ q_1 &= 0.67 \Rightarrow q_2 = 0.33. \end{aligned} \tag{4.4}$$

Note that

$$\mathbb{P} = (p_1, p_2) \neq (q_1, q_2) = \mathbb{Q}, \tag{4.5}$$

which exactly illustrates the difference between the real-world and risk-neutral measure. For example, suppose there are two traders: trader A owns a speculative portfolio (which is not hedged) and trader B does not want to be subjected to risk and therefore hedges his portfolio. Trader A will be interested in the real-world probabilities, since they will tell him something about the actual expectations of the stock and therefore his portfolio. However, trader B does not care whether the stock moves up or down, since his portfolio is protected against these movements. Thus, he will only be interested in the risk-neutral probabilities to value the option and determine his hedging strategy.

This example demonstrates the principles of hedging, but is by no means realistic. In reality, we do not know the possible states of the world at $T = 1$, which complicates the hedging strategy. But the main idea remains unchanged: we would like to create a risk-neutral portfolio using a number of stocks (possibly combined with other financial instruments). In the next sections we will discuss different hedging strategies based on this principle and how they are implemented under more realistic market assumptions.

4.2. Hedging strategies

In the previous section we treated a simplified example to illustrate the principles of hedging. With these principles in mind we will derive two different hedging strategies: Delta and Delta-Vega hedging. The former hedges against random movements in the underlying asset price (similar to the simplified example), while the latter hedges against changes in both asset price and volatility.

4.2.1. Delta hedging

The Delta hedging strategy aims at protecting the portfolio against changes of the underlying asset by holding an amount of Δ stocks (similar to the example in section 4.1). The portfolio is given by

$$\begin{cases} \Pi_t = -C_t + \Delta_{C_t} S_t + B_t, \\ \Pi_0 = 0, \end{cases} \tag{4.6}$$

where B_t denotes a bank account which grows with risk-free rate r. Now, similar to the simplified example, we must define Δ_{C_t} such that it protects the portfolio against changes in the underlying asset. Hence, we require that

$$\langle \mathrm{d}\Pi_t, \mathrm{d}S_t \rangle = 0. \tag{4.7}$$

The operator $\langle \cdot, \cdot \rangle$ refers to the covariation between two processes. The most relevant properties of this operator are discussed in appendix D.1. Following Bakshi et al. (1997), we can derive the so-called minimized variance Δ_{C_t},

$$\begin{aligned} 0 &= \langle \mathrm{d}\Pi_t, \mathrm{d}S_t \rangle, \\ 0 &= -\langle \mathrm{d}C_t, \mathrm{d}S_t \rangle + \Delta_{C_t} \langle \mathrm{d}S_t, \mathrm{d}S_t \rangle + \langle \mathrm{d}B_t, \mathrm{d}S_t \rangle, \\ \langle \mathrm{d}C_t, \mathrm{d}S_t \rangle &= \Delta_{C_t} \langle \mathrm{d}S_t, \mathrm{d}S_t \rangle, \\ \Delta_{C_t} &= \frac{\langle \mathrm{d}C_t, \mathrm{d}S_t \rangle}{\langle \mathrm{d}S_t, \mathrm{d}S_t \rangle}. \end{aligned} \tag{4.8}$$

By assuming this strategy, one obtains a portfolio which has no covariation with the underlying asset. In other words, changes of the underlying asset will have no direct or indirect (through correlations) effects on the portfolio.

This strategy is intuitively appealing. Suppose a financial institution (the writer) sells a European call option at time $t = 0$. The writer does not know in advance whether the stock value will rise or fall and wishes to be protected against a possible loss. If the stock value falls, the final pay-off will equal zero and the writer does not need any stocks to cover this loss (since there is no loss). On the other hand, if the stock value rises, the writer possibly faces a major loss. By buying a number of stocks beforehand, the writer will be able to cover this loss, since the previously bought stocks have risen in value as well. Thus, the higher the probability of a positive pay-off, the more stocks the writer will buy to cover a possible loss. Consequently, in case of a call option we expect Δ_{C_t} to behave as follows

$$\begin{cases} \Delta_{C_t} \approx 0, & \text{If the option is out-of-the-money}^3. \\ \Delta_{C_t} \approx 1, & \text{If the option is in-the-money}^4. \\ \Delta_{C_t} \approx \frac{1}{2}, & \text{If the option is at-the-money}^5. \end{cases}$$

In case of the European call option, this will lead to a buy-high sell-low strategy: we buy stocks when the price rises (because higher stock value means higher probability of a positive pay-off) and sell stocks when the price drops (lower stock value means lower probability of a positive pay-off). The losses induced by this unfavourable strategy should (on average) be covered by the initial option price.

[3] A call/put option is called out-of-the-money when it has no pay-off according to the current market.
[4] A call/put option is called in-the-money when it has a positive pay-off according to the current market.
[5] Boundary between in-the-money and out-of-the-money.

4.2.2. Delta-Vega hedging

In the previous section we treated the Delta hedging strategy, which hedges against changes of the underlying asset. In practice, a portfolio is subjected to other risks as well. For example, the portfolio value might change when the volatility increases/decreases. To this end, we will implement a Delta-Vega strategy. Unlike the Delta-hedge, the Delta-Vega strategy hedges against changes in both asset price and volatility. In general, the portfolio will be defined as,

$$\begin{cases} \Pi_t = -C_t + \bar{\Delta}^{(1)}_{C_t} S_t + \bar{\Delta}^{(2)}_{C_t} \tilde{C}_t + B_t, \\ \Pi_0 = 0, \end{cases} \tag{4.9}$$

where $\tilde{C}_t$ represents the value of another option (called option B from this point onwards), which depends on *the same underlying market factors*, but with different contract details (for example different strike/maturity). The strategy is characterized by the so-called hedging ratios:

- The amount of stocks held at time t, $\bar{\Delta}^{(1)}_{C_t}$.
- The amount of options B held at time t, $\bar{\Delta}^{(2)}_{C_t}$.

The purpose of $\bar{\Delta}^{(1)}_{C_t}$ is to remove all randomness associated with the asset price and the purpose of $\bar{\Delta}^{(2)}_{C_t}$ is to remove all randomness associated with the volatility that is not correlated with the asset price. Thus, we require

$$\begin{cases} \langle \mathrm{d}\Pi_t, \mathrm{d}S_t \rangle = 0, \\ \langle \mathrm{d}\Pi_t, \mathrm{d}W^v_t \rangle = 0, \end{cases} \tag{4.10}$$

where W^v_t is the Brownian motion independent from S_t, driving random changes in the volatility. By substituting the portfolio as defined in (4.9) into the first condition, we can derive

$$\begin{aligned} 0 &= \langle \mathrm{d}\Pi_t, \mathrm{d}S_t \rangle, \\ 0 &= -\langle \mathrm{d}C_t, \mathrm{d}S_t \rangle + \bar{\Delta}^{(1)}_{C_t} \langle \mathrm{d}S_t, \mathrm{d}S_t \rangle + \bar{\Delta}^{(2)}_{C_t} \langle \mathrm{d}\tilde{C}_t, \mathrm{d}S_t \rangle + \langle \mathrm{d}B_t, \mathrm{d}S_t \rangle, \\ \bar{\Delta}^{(1)}_{C_t} \langle \mathrm{d}S_t, \mathrm{d}S_t \rangle &= \langle \mathrm{d}C_t, \mathrm{d}S_t \rangle - \bar{\Delta}^{(2)}_{C_t} \langle \mathrm{d}\tilde{C}_t, \mathrm{d}S_t \rangle, \\ \bar{\Delta}^{(1)}_{C_t} &= \frac{\langle \mathrm{d}C_t, \mathrm{d}S_t \rangle}{\langle \mathrm{d}S_t, \mathrm{d}S_t \rangle} - \bar{\Delta}^{(2)}_{C_t} \frac{\langle \mathrm{d}\tilde{C}_t, \mathrm{d}S_t \rangle}{\langle \mathrm{d}S_t, \mathrm{d}S_t \rangle}. \end{aligned} \tag{4.11}$$

The definition of $\bar{\Delta}^{(2)}_{C_t}$ can be derived from the second condition,

$$\begin{aligned} 0 &= \langle \mathrm{d}\Pi_t, \mathrm{d}W^v_t \rangle, \\ 0 &= -\langle \mathrm{d}C_t, \mathrm{d}W^v_t \rangle + \bar{\Delta}^{(1)}_{C_t} \langle \mathrm{d}S_t, \mathrm{d}W^v_t \rangle, + \bar{\Delta}^{(2)}_{C_t} \langle \mathrm{d}\tilde{C}_t, \mathrm{d}W^v_t \rangle + \langle \mathrm{d}B_t, \mathrm{d}W^v_t \rangle, \\ \langle \mathrm{d}C_t, \mathrm{d}W^v_t \rangle &= \bar{\Delta}^{(2)}_{C_t} \langle \mathrm{d}\tilde{C}_t, \mathrm{d}W^v_t \rangle, \\ \bar{\Delta}^{(2)}_{C_t} &= \frac{\langle \mathrm{d}C_t, \mathrm{d}W^v_t \rangle}{\langle \mathrm{d}\tilde{C}_t, \mathrm{d}W^v_t \rangle}. \end{aligned} \tag{4.12}$$

In conclusion, the optimal strategy is given by

$$\begin{cases} \bar{\Delta}^{(2)}_{C_t} = \dfrac{\langle \mathrm{d}C_t, \mathrm{d}W^v_t \rangle}{\langle \mathrm{d}\tilde{C}_t, \mathrm{d}W^v_t \rangle}, \\ \bar{\Delta}^{(1)}_{C_t} = \dfrac{\langle \mathrm{d}C_t, \mathrm{d}S_t \rangle}{\langle \mathrm{d}S_t, \mathrm{d}S_t \rangle} - \bar{\Delta}^{(2)}_{C_t} \dfrac{\langle \mathrm{d}\tilde{C}_t, \mathrm{d}S_t \rangle}{\langle \mathrm{d}S_t, \mathrm{d}S_t \rangle}. \end{cases} \tag{4.13}$$

This strategy is intuitively similar to Delta hedging. If the volatility rises, we expect the asset price to fluctuate more. These fluctuations will lead to higher Delta hedging costs, because our buy-high sell-low strategy will be more pronounced. To this end, the writer will buy a certain amount of option B beforehand. Option B is subjected to the same volatility as the hedged option, hence the value of option B will increase if the volatility increases, which should cover the additional losses induced by the Delta strategy. Similarly, the converse is true when volatility decreases. This way, one is indifferent to changes in both asset price and volatility.

4.3. Market models

In the previous sections we discussed the Delta and Delta-Vega hedging strategies. In this section we will apply these strategies to different market environments. The dynamics of these markets are closely related to the dynamics of the option price, which will influence the hedging strategies. To this end, we need to derive the option price dynamics under the different market assumptions.

It is important to stress that we will simulate the market factors (such as asset price and volatility) under a real-world measure for every $t \in [0,T]$, because we are interested in the actual development of the market. These simulations will define the state of the market for every $t \in [0,T]$, which include stock value and volatility at time t, for example. Furthermore, depending on the state of the market, the risk-neutral measure will be defined for every $t \in [0,T]$. Hereafter, we will be able to establish the option prices and hedging strategies for each $t \in [0,T]$ according to these risk-neutral measures. In order to determine the hedging strategies, one needs to know the option price dynamics. These dynamics will, just like the hedging strategy, be defined under the risk-neutral measure. Note that these dynamics do not necessarily coincide with the actual option price dynamics.

This test will allow us to review certain features of the real-world and risk-neutral measure (such as drift and volatility) and their effect on the hedging strategies. In other words, we will be able to investigate the hedging performances in case the option dynamics implied by the risk-neutral measure differ from the actual option dynamics simulated by the real-world measure. For instance, in the simplified example we showed the irrelevance of the real-world probabilities, but the possible (real-world) values of S_t at time $t = 1$ did influence the hedging strategy. In the next sections we will discuss which real-world features are important for hedging strategies and which features are not. But for now, we will ignore the real-world measure and derive the hedging ratios under the risk-neutral measure.

4.3.1. Black-Scholes

Before determining the option price dynamics, one first needs to assume an appropriate model for the underlying asset. First, we will consider the Black-Scholes model. Recall,

$$\mathrm{d}S_t = rS_t\mathrm{d}t + \sigma S_t \mathrm{d}W_t^S. \tag{4.14}$$

We will assume this relation during the derivations of this section.

Option price dynamics

We can write the option value at time t as

$$C_t^{\mathrm{BS}} \equiv C(t, S_t, \sigma, r, K, T). \tag{4.15}$$

Applying Ito's lemma will define the option price dynamics under the Black-Scholes model

$$\mathrm{d}C_t^{\mathrm{BS}} = \frac{\partial C}{\partial t}\mathrm{d}t + \frac{\partial C}{\partial S_t}\mathrm{d}S_t + \frac{1}{2}\frac{\partial^2 C}{\partial S_t^2}\left\langle \mathrm{d}S_t, \mathrm{d}S_t \right\rangle. \tag{4.16}$$

Substituting (4.14) into (4.16) leads to

$$\begin{aligned}\mathrm{d}C_t^{\mathrm{BS}} &= \frac{\partial C}{\partial t}\mathrm{d}t + \frac{\partial C}{\partial S_t}\left(rS_t\mathrm{d}t + \sigma S_t \mathrm{d}W_t^S\right) + \frac{1}{2}\frac{\partial^2 C}{\partial S_t^2}\sigma^2 S_t^2 \mathrm{d}t \\ &= \left(\frac{\partial C}{\partial t} + rS_t\frac{\partial C}{\partial S_t} + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 C}{\partial S_t^2}\right)\mathrm{d}t + \sigma S_t \frac{\partial C}{\partial S_t}\mathrm{d}W_t^S. \end{aligned} \tag{4.17}$$

This shows that the option dynamics can be separated into two parts

- A deterministic part, associated with the $\mathrm{d}t$ term.
- A random part, associated with stochastic changes in the stock price, $\mathrm{d}W_t^S$.

Delta hedging
In case of the Black-Scholes model, Δ_{C_t} is defined as

$$\Delta_{C_t^{\mathrm{BS}}} = \frac{\langle \mathrm{d}C_t^{\mathrm{BS}}, \mathrm{d}S_t \rangle}{\langle \mathrm{d}S_t, \mathrm{d}S_t \rangle} = \frac{\sigma^2 S_t^2 \mathrm{d}t \frac{\partial C}{\partial S_t}}{\sigma^2 S_t^2 \mathrm{d}t} = \frac{\partial C}{\partial S_t}, \tag{4.18}$$

where we have used the covariation properties based on Ito's calculus, as discussed in appendix D.1. This will lead to the following portfolio dynamics

$$\begin{aligned}
\mathrm{d}\Pi_t &= -\mathrm{d}C_t + \Delta_{C_t^{\mathrm{BS}}} \mathrm{d}S_t + \mathrm{d}B_t \\
&= -\mathrm{d}C_t + \frac{\partial C}{\partial S_t} \mathrm{d}S_t + \mathrm{d}B_t \\
&= -\left(\frac{\partial C}{\partial t} + rS_t \frac{\partial C}{\partial S_t} + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 C}{\partial S_t^2} \right) \mathrm{d}t - \sigma S_t \frac{\partial C}{\partial S_t} \mathrm{d}W_t^S + \frac{\partial C}{\partial S_t} \left(rS_t \mathrm{d}t + \sigma S_t \mathrm{d}W_t^S \right) + rB_t \mathrm{d}t \\
&= \left(-\frac{\partial C}{\partial t} - \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 C}{\partial S_t^2} + rB_t \right) \mathrm{d}t.
\end{aligned} \tag{4.19}$$

The portfolio dynamics only depend on the deterministic $\mathrm{d}t$ term. In other words, the Delta strategy removes all randomness from the portfolio. Moreover, the portfolio has no dependence on $\frac{\partial C_t}{\partial S_t}$. Thus, a Delta hedged portfolio is risk-free and does not depend on changes of the underlying asset price, if the assumptions of the Black-Scholes model are not violated.

Delta-Vega hedging
Volatility is considered constant in the Black-Scholes framework, hence, theoretically it is redundant to hedge against changes in volatility. However, historical data shows that the constant volatility assumption is violated in general. The Black-Scholes Delta-Vega hedging strategy can be thought of as a way to overcome the hedging errors caused by the constant volatility assumption.

Note that $\langle \mathrm{d}C_t^{\mathrm{BS}}, \mathrm{d}W_t^v \rangle = 0$, due to the constant volatility assumption. Therefore, we will slightly adjust the strategy by following Kurpiel and Roncalli (1998), which replaces the covariance operator by $\frac{\partial C}{\partial \sigma}$. Substituting the derivative operator in equation (4.13) gives

$$\begin{cases}
\bar{\Delta}^{(2)}_{C_t^{\mathrm{BS}}} = \dfrac{\partial C_t / \partial \sigma}{\partial \tilde{C}_t / \partial \sigma}, \\
\bar{\Delta}^{(1)}_{C_t^{\mathrm{BS}}} = \dfrac{\partial C_t}{\partial S_t} - \bar{\Delta}^{(2)}_{C_t^{\mathrm{BS}}} \dfrac{\partial \tilde{C}_t}{\partial S_t}.
\end{cases} \tag{4.20}$$

Option price sensitivities
In the above derivations, we have seen that the hedging ratios depend on the option price sensitivities with respect to asset price and volatility. Thus, one needs to be able to determine those sensitivities before implementing these strategies. Fortunately, under the Black-Scholes model, there are analytical expression for these quantities,

$$\begin{cases}
\dfrac{\partial C_t}{\partial S_t} = \mathcal{N}[d_1], \\
\dfrac{\partial C_t}{\partial \sigma} = S_t \mathcal{N}'[d_1] \sqrt{T-t}, \\
d_1 = \dfrac{\log\left(\frac{S_t}{K}\right) + \left(r + \frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{T-t}}.
\end{cases} \tag{4.21}$$

Here $\mathcal{N}[\cdot]$ and $\mathcal{N}'[\cdot]$ are, respectively, the cumulative distribution and the probability density function of the standard normal distribution.

4.3.2. Heston

Similar to the Black-Scholes derivation, we can consider the Heston model

$$\begin{cases}
\mathrm{d}S_t = rS_t \mathrm{d}t + \sqrt{v_t} S_t \mathrm{d}W_t^1, \\
\mathrm{d}v_t = \kappa(\bar{v} - v_t)\mathrm{d}t + \gamma\sqrt{v_t} \mathrm{d}W_t^2, \\
\langle \mathrm{d}W_t^1, \mathrm{d}W_t^2 \rangle = \rho \mathrm{d}t.
\end{cases} \tag{4.22}$$

Option price dynamics
The option value at time t can be expressed as

$$C_t^{\text{Heston}} \equiv C(t, S_t, v_t, r, \Omega^{\text{Heston}}, K, T). \tag{4.23}$$

Under the assumption of two independent Brownian motions W_t^S and W_t^v, defined through

$$\begin{cases} W_t^1 = W_t^S, \\ W_t^2 = \rho W_t^S + \sqrt{1-\rho^2} W_t^v, \end{cases} \tag{4.24}$$

we can write the Heston option dynamics as

$$\begin{aligned} \mathrm{d}C_t^{\text{Heston}} = &\left(\frac{\partial C}{\partial t} + r S_t \frac{\partial C}{\partial S_t} + \kappa(\bar{v} - v_t) \frac{\partial C}{\partial v_t} + \frac{1}{2} v_t S_t^2 \frac{\partial^2 C}{\partial S_t^2} + \frac{1}{2} \gamma^2 v_t \frac{\partial^2 C}{\partial v_t^2} + \rho \gamma v_t S_t \frac{\partial^2 C}{\partial S_t \partial v_t} \right) \mathrm{d}t \\ &+ \left(\sqrt{v_t} S_t \frac{\partial C}{\partial S_t} + \rho \gamma \sqrt{v_t} \frac{\partial C}{\partial v_t} \right) \mathrm{d}W_t^S + \gamma \sqrt{v_t (1-\rho^2)} \frac{\partial C}{\partial v_t} \mathrm{d}W_t^v. \end{aligned} \tag{4.25}$$

The derivation of these dynamics is similar to the derivation under the Black-Scholes model and is discussed in appendix D.2. Again, we can separate these dynamics into different parts

- A deterministic part, associated with the $\mathrm{d}t$ term.
- A random part, associated with stochastic changes in the stock price, $\mathrm{d}W_t^S$.
- A random part, associated with stochastic changes in the stock volatility, $\mathrm{d}W_t^v$.

Delta hedging
The correlation between asset price and volatility in the Heston model slightly changes the Delta hedging procedure. Taking this correlation into account gives

$$\begin{aligned} \Delta_{C_t^{\text{Heston}}} &= \frac{\langle \mathrm{d}C_t^{\text{Heston}}, \mathrm{d}S_t \rangle}{\langle \mathrm{d}S_t, \mathrm{d}S_t \rangle} \\ &= \frac{v_t S_t^2 \frac{\partial C}{\partial S_t} \mathrm{d}t + \rho \gamma v_t S_t \frac{\partial C}{\partial v_t} \mathrm{d}t}{v_t S_t^2 \mathrm{d}t} \\ &= \frac{\partial C}{\partial S_t} + \rho \gamma \frac{1}{S_t} \frac{\partial C}{\partial v_t}. \end{aligned} \tag{4.26}$$

Now, we are able to derive the portfolio dynamics (the full derivation can be found in appendix D.3)

$$\begin{aligned} \mathrm{d}\Pi_t = &- \left(\frac{\partial C}{\partial t} + \kappa(\bar{v} - v_t - r\rho\gamma) \frac{\partial C}{\partial v_t} + \frac{1}{2} v_t S_t^2 \frac{\partial^2 C}{\partial S_t^2} + \frac{1}{2} \gamma^2 v_t \frac{\partial^2 C}{\partial v_t^2} + \rho \gamma v_t S_t \frac{\partial^2 C}{\partial S_t \partial v_t} - r B_t \right) \mathrm{d}t \\ &- \gamma \sqrt{v_t (1-\rho^2)} \frac{\partial C}{\partial v_t} \mathrm{d}W_t^v. \end{aligned} \tag{4.27}$$

It follows that we are unable to remove all randomness from the portfolio, using just the stock as hedge instrument. Although this portfolio is insensitive to changes in asset price, it is not protected against changes in asset volatility. In order to remove all randomness, one needs to use an additional financial instrument (for example, by implementing the Delta-Vega strategy).

Delta-Vega hedging
In the above derivation we have seen that the Delta strategy is unable to create a portfolio completely free of randomness. In order to remove the randomness associated with volatility, we will implement the Delta-Vega hedging strategy. In this case, the hedging ratios are given by

$$\begin{aligned} \bar{\Delta}^{(2)}_{C_t^{\text{Heston}}} &= \frac{\langle \mathrm{d}C_t^{\text{Heston}}, \mathrm{d}W_t^v \rangle}{\langle \mathrm{d}\tilde{C}_t^{\text{Heston}}, \mathrm{d}W_t^v \rangle} \\ &= \frac{\gamma \sqrt{v_t (1-\rho^2)} \frac{\partial C}{\partial v_t} \mathrm{d}t}{\gamma \sqrt{v_t (1-\rho^2)} \frac{\partial \tilde{C}}{\partial v_t} \mathrm{d}t} \end{aligned}$$

$$= \frac{\partial C / \partial v_t}{\partial \tilde{C} / \partial v_t}, \tag{4.28}$$

and

$$\begin{aligned}
\bar{\Delta}^{(1)}_{C_t^{\text{Heston}}} &= \frac{\langle \mathrm{d}C_t^{\text{Heston}}, \mathrm{d}S_t \rangle}{\langle \mathrm{d}S_t, \mathrm{d}S_t \rangle} - \bar{\Delta}^{(2)}_{C_t^{\text{Heston}}} \frac{\langle \mathrm{d}\tilde{C}_t^{\text{Heston}}, \mathrm{d}S_t \rangle}{\langle \mathrm{d}S_t, \mathrm{d}S_t \rangle} \\
&= \frac{\partial C}{\partial S_t} + \rho\gamma \frac{1}{S_t} \frac{\partial C}{\partial v_t} - \bar{\Delta}^{(2)}_{C_t^{\text{Heston}}} \left(\frac{\partial \tilde{C}}{\partial S_t} + \rho\gamma \frac{1}{S_t} \frac{\partial \tilde{C}}{\partial v_t} \right) \\
&= \frac{\partial C}{\partial S_t} - \bar{\Delta}^{(2)}_{C_t^{\text{Heston}}} \frac{\partial \tilde{C}}{\partial S_t}.
\end{aligned} \tag{4.29}$$

Substituting these definitions in (4.22) gives

$$\begin{aligned}
\mathrm{d}\Pi_t = \Bigg(&- \frac{\partial C}{\partial t} - \frac{1}{2} v_t S_t^2 \frac{\partial^2 C}{\partial S_t^2} - \frac{1}{2} \gamma^2 v_t \frac{\partial^2 C}{\partial v_t^2} - \rho\gamma v_t S_t \frac{\partial^2 C}{\partial S_t \partial v_t} \\
&+ \bar{\Delta}^{(2)}_{C_t^{\text{Heston}}} \left(\frac{\partial C}{\partial t} + \frac{1}{2} v_t S_t^2 \frac{\partial^2 C}{\partial S_t^2} + \frac{1}{2} \gamma^2 v_t \frac{\partial^2 C}{\partial v_t^2} + \rho\gamma v_t S_t \frac{\partial^2 C}{\partial S_t \partial v_t} \right) + r B_t \Bigg) \mathrm{d}t,
\end{aligned} \tag{4.30}$$

which we prove in appendix D.4. Note that the random components have disappeared from the portfolio. Hence, a Delta-Vega strategy fully hedges a portfolio, in case the market does not violate the Heston model assumptions. Moreover, the deterministic and stochastic parts of the portfolio dynamics do not contain any $\frac{\partial C_t}{\partial S_t}$ or $\frac{\partial C_t}{\partial v_t}$ terms. Hence, the Delta-Vega strategy does not depend on the deterministic and stochastic changes of the asset price and volatility, at least in terms of the first derivative.

Option price sensitivities

The Heston hedging strategies require sensitivities of the option price with respect to asset price and volatility. But unlike the Black-Scholes model, they cannot be expressed analytically. Thus, we need to approximate these sensitivities by adjusting the FFT algorithm. In appendix A we already explained how one can approximate the option price using this algorithm. This approximation involves a finite sum of differentiable components. Consequently, the derivative of this sum is equal to the sum of its derivatives,

$$\begin{cases}
\frac{\partial C_t}{\partial S_t} \approx \frac{e^{-ak}}{\pi} \Re \left\{ \sum_{j=1}^{N} e^{-iz_j k} \psi_T \left(z_j, \Omega_t \right) \Delta z \frac{\alpha + 1 + iz_j}{S_t} \right\}, \\
\frac{\partial C_t}{\partial v_t} \approx \frac{e^{-ak}}{\pi} \Re \left\{ \sum_{j=1}^{N} e^{-iz_j k} \psi_T \left(z_j, \Omega_t \right) \Delta z \frac{1}{\gamma^2} \left(\frac{1 - e^{-D(T-t)}}{1 - G e^{-D(T-t)}} \right) (\kappa - i\rho\gamma(z - (\alpha+1)i) - D) \right\},
\end{cases} \tag{4.31}$$

where the components are defined as in appendix A.

4.3.3. Dynamic Heston

In the previous chapters we have discussed the dynamic behaviour of the Heston parameters. This behaviour is caused by changes in option prices or equivalently, the implied volatility surface. Hence, in order to simulate a market more in line with these historical observations, we assume the Heston parameters to follow certain stochastic processes. We have already seen that the Heston model parameters are highly correlated with each other, therefore we will assume

$$\begin{cases}
\mathrm{d}\bar{v}_t = \kappa_{\bar{v}} (\bar{v}_{\text{Mean}} - \bar{v}_t) \mathrm{d}t + a_{\bar{v}} \bar{v}_t \left(\rho_{\bar{v}} \mathrm{d}W_t^2 + \sqrt{1 - \rho_{\bar{v}}^2} \mathrm{d}W_t^{\bar{v}} \right), \\
\mathrm{d}\gamma_t = \kappa_{\gamma} (\gamma_{\text{Mean}} - \gamma_t) \mathrm{d}t + a_{\gamma} \gamma_t \left(\rho_{\gamma} \mathrm{d}W_t^2 + \sqrt{1 - \rho_{\gamma}^2} \mathrm{d}W_t^{\gamma} \right),
\end{cases} \tag{4.32}$$

where both parameters follow a mean-reverting path, W_t^2 is defined as in (D.5) and the Brownian motions are all independent. Note that $\rho_{\bar{v}}$ and ρ_{γ} indicate the correlation with v_t, which we expect to be high, based on historical observations. Consequently, v_t, $\bar{v}_t$ and γ_t will follow highly correlated paths.

These assumptions lead to the paradox we discussed at the beginning of this chapter. On one hand we assume the Heston model to price options, which assumes $\bar{v}_t$ and γ_t to be constant. However, according to historical data $\bar{v}_t$ and γ_t change over time, which influences the *actual* option price dynamics and therefore the hedging performance.

To deal with these kinds of problems, Alexander et al. (2009) proposed an out-of-model hedge adjustment, which we will follow in this section. Out-of-model hedging refers to strategies that hedge against changes of constant/deterministic parameters. One of the most well-known out-of-model hedges is the Black-Scholes Delta-Vega hedge we discussed in section 4.3.1. From a theoretical point of view, out-of-model hedging is redundant, because these parameters do not add additional randomness to the model. However, in practice these parameters do not always behave as the model expects them to, as we have seen in the previous chapters. Therefore, out-of-model hedging can be thought of as a way to overcome the deficiencies of a model by violating some theoretical assumptions. Thus, the following derivations will not be in accordance with the Heston model, but they can be thought of as the option price dynamics under the assumption that $\bar{v}_t$ and γ_t evolve stochastically.

Option price dynamics: dynamic Heston
Under these assumptions we can write the option price as

$$C_t^{\text{Dynamic}} \equiv C\left(t, S_t, v_t, r, \bar{v}_t, \gamma_t, \kappa, \rho, K, T\right). \tag{4.33}$$

Applying Ito's lemma leads to

$$\mathrm{d}C_t^{\text{Dynamic}} = \frac{\partial C}{\partial t}\mathrm{d}t + \frac{\partial C}{\partial S_t}\mathrm{d}S_t + \frac{\partial C}{\partial v_t}\mathrm{d}v_t + \frac{\partial C}{\partial \bar{v}_t}\mathrm{d}\bar{v}_t + \frac{\partial C}{\partial \gamma_t}\mathrm{d}\gamma_t + \frac{1}{2}\sum_{p_t}\sum_{q_t}\frac{\partial^2 C}{\partial p_t \partial q_t}\langle \mathrm{d}p_t, \mathrm{d}q_t\rangle, \tag{4.34}$$

with $p_t, q_t \in \{S_t, v_t, \bar{v}_t, \gamma_t\}$. For notational purposes we will rewrite this expression as

$$\mathrm{d}C_t^{\text{Dynamic}} = c_1\mathrm{d}t + c_2\mathrm{d}W_t^S + c_3\mathrm{d}W_t^v + c_4\mathrm{d}W_t^{\bar{v}} + c_5\mathrm{d}W_t^{\gamma}. \tag{4.35}$$

By substituting (4.22) and (4.32) into (4.34), we obtain

$$\begin{cases} c_1 = \frac{\partial C}{\partial t} + rS_t\frac{\partial C}{\partial S_t} + \kappa(\bar{v}_t - v_t)\frac{\partial C}{\partial v_t} + \kappa_{\bar{v}}(\bar{v}_{\text{Mean}} - \bar{v}_t)\frac{\partial C}{\partial \bar{v}_t} + \kappa_{\gamma}(\gamma_{\text{Mean}} - \gamma_t)\frac{\partial C}{\partial \gamma_t} \\ \qquad + \frac{1}{2}\sum_{p_t}\sum_{q_t}\frac{\partial^2 C}{\partial p_t \partial q_t}\langle \mathrm{d}p_t, \mathrm{d}q_t\rangle, \\ c_2 = \sqrt{v_t}S_t\frac{\partial C}{\partial S_t} + \rho\gamma_t\sqrt{v_t}\frac{\partial C}{\partial v_t} + \rho\rho_{\bar{v}}a_{\bar{v}}\bar{v}_t\frac{\partial C}{\partial \bar{v}_t} + \rho\rho_{\gamma}a_{\gamma}\gamma_t\frac{\partial C}{\partial \gamma_t}, \\ c_3 = \gamma_t\sqrt{v_t(1-\rho^2)}\frac{\partial C}{\partial v_t} + \rho_{\bar{v}}a_{\bar{v}}\bar{v}_t\sqrt{1-\rho^2}\frac{\partial C}{\partial \bar{v}_t} + \rho_{\gamma}a_{\gamma}\gamma_t\sqrt{1-\rho^2}\frac{\partial C}{\partial \gamma_t}, \\ c_4 = a_{\bar{v}}\bar{v}_t\sqrt{1-\rho_{\bar{v}}^2}\frac{\partial C}{\partial \bar{v}_t}, \\ c_5 = a_{\gamma}\gamma_t\sqrt{1-\rho_{\gamma}^2}\frac{\partial C}{\partial \gamma_t}. \end{cases} \tag{4.36}$$

Note that the Heston option price dynamics can be derived as a special case of the dynamic Heston model, when the mean reversion rate and volatility of $\bar{v}_t$ and γ_t are zero.

Delta hedging
In case of the dynamic Heston model, we can define the Δ_{C_t} as

$$\begin{aligned} \Delta_{C_t^{\text{Dyn}}} &= \frac{\left\langle \mathrm{d}C_t^{\text{Dyn}}, \mathrm{d}S_t \right\rangle}{\langle \mathrm{d}S_t, \mathrm{d}S_t\rangle} \\ &= \frac{c_2\sqrt{v_t}S_t\mathrm{d}t}{v_t S_t^2\mathrm{d}t} \\ &= \frac{c_2}{\sqrt{v_t}S_t} \\ &= \frac{\partial C}{\partial S_t} + \rho\gamma\frac{1}{S_t}\frac{\partial C}{\partial v_t} + \rho\rho_{\bar{v}}a_{\bar{v}}\bar{v}_t\frac{1}{\sqrt{v_t}S_t}\frac{\partial C}{\partial \bar{v}_t} + \rho\rho_{\gamma}a_{\gamma}\gamma_t\frac{1}{\sqrt{v_t}S_t}\frac{\partial C}{\partial \gamma_t}. \end{aligned} \tag{4.37}$$

Again, the Heston model strategy can be seen as a special case of the dynamic Heston model, if the correlation or volatility of $\bar{v}_t$ and γ_t is equal to zero. Implementing this strategy leads to the following portfolio dynamics

$$\mathrm{d}\Pi_t = -\mathrm{d}C_t + \Delta_{C_t^{\text{Dyn}}}\mathrm{d}S_t + \mathrm{d}B_t$$

$$= -c_1 \mathrm{d}t - c_2 \mathrm{d}W_t^S - c_3 \mathrm{d}W_t^v - c_4 \mathrm{d}W_t^{\bar{v}} - c_5 \mathrm{d}W_t^\gamma + \frac{c_2}{\sqrt{v_t} S_t} \left(r S_t \mathrm{d}t + \sqrt{v_t} S_t \mathrm{d}W_t^S \right) + r B_t \mathrm{d}t$$
$$= \left(-c_1 + c_2 \frac{r}{\sqrt{v_t}} + r B_t \right) \mathrm{d}t - c_3 \mathrm{d}W_t^v - c_4 \mathrm{d}W_t^{\bar{v}} - c_5 \mathrm{d}W_t^\gamma. \tag{4.38}$$

Similar to the Heston model, we are only able to remove the randomness from the portfolio associated to changes in S_t.

Delta-Vega hedging

We can define the Dynamic Heston Delta-Vega hedge ratios as

$$\begin{aligned} \bar{\Delta}^{(2)}_{C_t^{\mathrm{Dyn}}} &= \frac{\left\langle \mathrm{d}C_t^{\mathrm{Dyn}}, \mathrm{d}W_t^v \right\rangle}{\left\langle \mathrm{d}\tilde{C}_t^{\mathrm{Dyn}}, \mathrm{d}W_t^v \right\rangle} \\ &= \frac{c_3 \mathrm{d}t}{\tilde{c}_3 \mathrm{d}t} \\ &= \frac{c_3}{\tilde{c}_3}, \end{aligned} \tag{4.39}$$

and

$$\begin{aligned} \bar{\Delta}^{(1)}_{C_t^{\mathrm{Dyn}}} &= \frac{\left\langle \mathrm{d}C_t^{\mathrm{Dyn}}, \mathrm{d}S_t \right\rangle}{\langle \mathrm{d}S_t, \mathrm{d}S_t \rangle} - \bar{\Delta}^{(2)}_{C_t^{\mathrm{Dyn}}} \frac{\left\langle \mathrm{d}\tilde{C}_t^{\mathrm{Dyn}}, \mathrm{d}S_t \right\rangle}{\langle \mathrm{d}S_t, \mathrm{d}S_t \rangle} \\ &= \frac{c_2}{\sqrt{v_t} S_t} - \frac{\tilde{c}_2 c_3}{\sqrt{v_t} S_t \tilde{c}_3}, \end{aligned} \tag{4.40}$$

where $\tilde{c}_2$ and $\tilde{c}_3$ are the coefficients related to option B and they are defined by (4.36). Substituting these definitions results in the following portfolio dynamics

$$\begin{aligned} \mathrm{d}\Pi_t &= -\mathrm{d}C_t + \bar{\Delta}^{(1)}_{C_t^{\mathrm{Dyn}}} \mathrm{d}S_t + \bar{\Delta}^{(2)}_{C_t^{\mathrm{Dyn}}} \mathrm{d}\tilde{C}_t + \mathrm{d}B_t \\ &= -c_1 \mathrm{d}t - c_2 \mathrm{d}W_t^S - c_3 \mathrm{d}W_t^v - c_4 \mathrm{d}W_t^{\bar{v}} - c_5 \mathrm{d}W_t^\gamma + \left(\frac{c_2}{\sqrt{v_t} S_t} - \frac{\tilde{c}_2 c_3}{\sqrt{v_t} S_t \tilde{c}_3} \right) \left(r S_t \mathrm{d}t + \sqrt{v_t} S_t \mathrm{d}W_t^S \right) \\ &\quad \frac{c_3}{\tilde{c}_3} \left(\tilde{c}_1 \mathrm{d}t + \tilde{c}_2 \mathrm{d}W_t^S + \tilde{c}_3 \mathrm{d}W_t^v + \tilde{c}_4 \mathrm{d}W_t^{\bar{v}} + \tilde{c}_5 \mathrm{d}W_t^\gamma \right) + r B_t \mathrm{d}t \\ &= \left(-c_1 + \frac{c_3}{\tilde{c}_3} \tilde{c}_1 + \frac{\tilde{c}_2 r}{\sqrt{v_t}} - \frac{\tilde{c}_2 c_3 r}{\sqrt{v_t} \tilde{c}_3} + r B_t \right) \mathrm{d}t + \left(\frac{c_3}{\tilde{c}_3} \tilde{c}_4 - c_4 \right) \mathrm{d}W_t^{\bar{v}} + \left(\frac{c_3}{\tilde{c}_3} \tilde{c}_5 - c_5 \right) \mathrm{d}W_t^\gamma. \end{aligned} \tag{4.41}$$

The only randomness left in this portfolio is associated with changes in $\bar{v}_t$ and γ_t, but note that

$$\begin{cases} \rho_{\bar{v}} \uparrow 1 \Rightarrow c_4, \tilde{c}_4 \downarrow 0, \\ \rho_\gamma \uparrow 1 \Rightarrow c_5, \tilde{c}_5 \downarrow 0. \end{cases} \tag{4.42}$$

So high correlation between v_t, $\bar{v}_t$ and γ_t will lead to less randomness in the Dynamic Heston Delta-Vega portfolio. In the previous chapters we have already seen that the correlation between these parameters is very high, hence we expect this strategy to work relatively well.

However, even when the correlations are equal to one, we do not expect a perfect hedge. This strategy only takes stochastic changes into account and not deterministic changes. Recall the Heston Delta portfolio dynamics from equation (4.27). Even with a correlation equal to one, this strategy still depends deterministically on $\frac{\partial C_t}{\partial v_t}$. Hence, its dynamics still depend on changes in volatility. This is contrast with the Heston Delta-Vega portfolio dynamics given in equation (4.30), where all $\frac{\partial C_t}{\partial S_t}$ and $\frac{\partial C_t}{\partial v_t}$ terms disappeared, both in the deterministic and stochastic parts. So, in order to be able to fully hedge against changes of $\bar{v}_t$ and γ_t, one needs to include two additional options, such that both the deterministic and stochastic dependencies are removed, which we will show below.

Full hedge dynamic Heston

In the above derivation we discussed the limited properties of the Delta-Vega hedge in a dynamic Heston market: we can only protect the portfolio against the changes of $\bar{v}_t$ and γ_t that are correlated with v_t and not against the deterministic and uncorrelated part. A full hedge in a dynamic Heston market requires three options and the underlying asset,

$$\begin{cases} \Pi_t = -C_t + \tilde{\Delta}_t^{(1)} S_t + \tilde{\Delta}_t^{(2)} \tilde{C}_t + \tilde{\Delta}_t^{(3)} \bar{C}_t + \tilde{\Delta}_t^{(4)} \hat{C}_t, \\ \Pi_0 = 0. \end{cases} \tag{4.43}$$

In this case we require no covariation between all random processes, thus

$$\begin{cases} \langle \mathrm{d}\Pi_t, \mathrm{d}S_t \rangle = 0, \\ \langle \mathrm{d}\Pi_t, \mathrm{d}W_t^{v} \rangle = 0, \\ \langle \mathrm{d}\Pi_t, \mathrm{d}W_t^{\bar{v}} \rangle = 0, \\ \langle \mathrm{d}\Pi_t, \mathrm{d}W_t^{\gamma} \rangle = 0. \end{cases} \tag{4.44}$$

Substituting these conditions in (4.43) gives

$$\begin{bmatrix} \langle \mathrm{d}S, \mathrm{d}S \rangle & \langle \mathrm{d}C_2, \mathrm{d}S \rangle & \langle \mathrm{d}C_3, \mathrm{d}S \rangle & \langle \mathrm{d}C_4, \mathrm{d}S \rangle \\ 0 & \langle \mathrm{d}C_2, \mathrm{d}W^{v} \rangle & \langle \mathrm{d}C_3, \mathrm{d}W^{v} \rangle & \langle \mathrm{d}C_4, \mathrm{d}W^{v} \rangle \\ 0 & \langle \mathrm{d}C_2, \mathrm{d}W^{\bar{v}} \rangle & \langle \mathrm{d}C_3, \mathrm{d}W^{\bar{v}} \rangle & \langle \mathrm{d}C_4, \mathrm{d}W^{\bar{v}} \rangle \\ 0 & \langle \mathrm{d}C_2, \mathrm{d}W^{\gamma} \rangle & \langle \mathrm{d}C_3, \mathrm{d}W^{\gamma} \rangle & \langle \mathrm{d}C_4, \mathrm{d}W^{\gamma} \rangle \end{bmatrix} \begin{bmatrix} \tilde{\Delta}_t^{(1)} \\ \tilde{\Delta}_t^{(2)} \\ \tilde{\Delta}_t^{(3)} \\ \tilde{\Delta}_t^{(4)} \end{bmatrix} = \begin{bmatrix} \langle \mathrm{d}C_1, \mathrm{d}S \rangle \\ \langle \mathrm{d}C_1, \mathrm{d}W^{v} \rangle \\ \langle \mathrm{d}C_1, \mathrm{d}W^{\bar{v}} \rangle \\ \langle \mathrm{d}C_1, \mathrm{d}W^{\gamma} \rangle \end{bmatrix},$$

$$\begin{bmatrix} v_t S_t^2 \mathrm{d}t & \tilde{c}_2 \sqrt{v_t} S_t \mathrm{d}t & \bar{c}_2 \sqrt{v_t} S_t \mathrm{d}t & \hat{c}_2 \sqrt{v_t} S_t \mathrm{d}t \\ 0 & \tilde{c}_3 \sqrt{v_t} S_t \mathrm{d}t & \bar{c}_3 \sqrt{v_t} S_t \mathrm{d}t & \hat{c}_3 \sqrt{v_t} S_t \mathrm{d}t \\ 0 & \tilde{c}_4 \sqrt{v_t} S_t \mathrm{d}t & \bar{c}_4 \sqrt{v_t} S_t \mathrm{d}t & \hat{c}_4 \sqrt{v_t} S_t \mathrm{d}t \\ 0 & \tilde{c}_5 \sqrt{v_t} S_t \mathrm{d}t & \bar{c}_5 \sqrt{v_t} S_t \mathrm{d}t & \hat{c}_5 \sqrt{v_t} S_t \mathrm{d}t \end{bmatrix} \begin{bmatrix} \tilde{\Delta}_t^{(1)} \\ \tilde{\Delta}_t^{(2)} \\ \tilde{\Delta}_t^{(3)} \\ \tilde{\Delta}_t^{(4)} \end{bmatrix} = \begin{bmatrix} c_2 \sqrt{v_t} S_t \\ c_3 \sqrt{v_t} S_t \\ c_4 \sqrt{v_t} S_t \\ c_5 \sqrt{v_t} S_t \end{bmatrix},$$

$$\begin{bmatrix} \sqrt{v_t} S_t & \tilde{c}_2 & \bar{c}_2 & \hat{c}_2 \\ 0 & \tilde{c}_3 & \bar{c}_3 & \hat{c}_3 \\ 0 & \tilde{c}_4 & \bar{c}_4 & \hat{c}_4 \\ 0 & \tilde{c}_5 & \bar{c}_5 & \hat{c}_5 \end{bmatrix} \begin{bmatrix} \tilde{\Delta}_t^{(1)} \\ \tilde{\Delta}_t^{(2)} \\ \tilde{\Delta}_t^{(3)} \\ \tilde{\Delta}_t^{(4)} \end{bmatrix} = \begin{bmatrix} c_2 \\ c_3 \\ c_4 \\ c_5 \end{bmatrix},$$

$$\begin{bmatrix} \tilde{\Delta}_t^{(1)} \\ \tilde{\Delta}_t^{(2)} \\ \tilde{\Delta}_t^{(3)} \\ \tilde{\Delta}_t^{(4)} \end{bmatrix} = \begin{bmatrix} \sqrt{v_t} S_t & \tilde{c}_2 & \bar{c}_2 & \hat{c}_2 \\ 0 & \tilde{c}_3 & \bar{c}_3 & \hat{c}_3 \\ 0 & \tilde{c}_4 & \bar{c}_4 & \hat{c}_4 \\ 0 & \tilde{c}_5 & \bar{c}_5 & \hat{c}_5 \end{bmatrix}^{-1} \begin{bmatrix} c_2 \\ c_3 \\ c_4 \\ c_5 \end{bmatrix}. \tag{4.45}$$

By implementing this strategy, the portfolio is protected against stochastic and deterministic changes of S_t, v_t, $\bar{v}_t$, γ_t. This portfolio is quite complicated, containing positions in four different options, and therefore it might not be practical in reality. Nevertheless, we will include it in our tests to assess the impact of hedging against changes of $\bar{v}_t$ and γ_t.

Option price sensitivities

Similar to the Heston model, we need to approximate the option price sensitivities by using the FFT algorithm. In case of the dynamic hedging model, two additional sensitivities are required, namely the sensitivities with respect to $\bar{v}_t$ and γ_t. The approximation of these sensitivities is similar to the ones already derived for the Heston model. However, the dependence of γ_t in the price approximation is rather complicated, hence we propose a numerical approximation:

$$\begin{cases} \frac{\partial C_t}{\partial \bar{v}_t} \approx \frac{e^{-\alpha k}}{\pi} \Re \left\{ \sum_{j=1}^{N} e^{-i z_j k} \psi_T\left(z_j, \Omega_t\right) \Delta z \frac{\kappa}{\gamma_t^2} \left\{ (T-t)(\kappa - i\rho\gamma_t u - D) - 2\log\left[\frac{1 - G e^{-D(T-t)}}{1-G} \right] \right\} \right\}, \\ \frac{\partial C_t}{\partial \gamma_t} \approx \frac{C(t, S_t, v_t, r, \bar{v}_t, \gamma_t + \delta_\gamma, \kappa, \rho, K, T) - C(t, S_t, v_t, r, \bar{v}_t, \gamma_t - \delta_\gamma, \kappa, \rho, K, T)}{2\delta_\gamma}, \end{cases} \tag{4.46}$$

where $\delta_\gamma > 0$ is a small constant. Moreover, the sensitivities with respect to asset price and volatility are defined as in (4.31).

4.4. Hedging in practice

In the previous sections we have derived some hedging strategies and option price dynamics for different market environments. These derivations were all performed under the assumptions that the hedging ratios are updated in continuous time. In practice, it is of course impossible to buy/sell the stocks/options after every infinitesimal change in the market. A discretization is therefore required, in which we will rebalance the hedge portfolio after every time-step Δt. In this section we will only discuss how to implement the Delta-Vega strategy. A Delta strategy (or any other hedging strategy, for that matter) can be implemented in a similar way.

First we will discretize the interval $[0,T]$, where $t=0$ and $t=T$ respectively denote the starting point and maturity of the option. We propose the following discretization

$$\begin{cases} t_0 = 0, \\ t_N = T, \\ \Delta t = \frac{T}{N-1}, \\ t_{i+1} = t_i + \Delta t, \qquad i \in \{0,\dots,N-1\}. \end{cases} \tag{4.47}$$

Initially, S_0, v_0 and the risk-neutral parameters are known. These parameters define the option values and the hedging ratios, given by (4.9) in case of Delta-Vega hedging.

At time $t=0$ the writer sells the option with value C_0 and, after buying the necessary amount of stocks/options according to the hedging strategy, there will be a certain amount of cash left. Thus, in case of a Delta-Vega strategy, the initial bank account is given by

$$B_0 = C_0 - \bar{\Delta}^{(1)}_{C_0} S_0 - \bar{\Delta}^{(2)}_{C_0} \tilde{C}_0. \tag{4.48}$$

Note that, by substituting this definition,

$$\Pi_0 = 0. \tag{4.49}$$

Now suppose we are at time t_i, with $i \in \{0,\dots,N-1\}$, and we have adjusted the hedging ratios according to the hedging strategy. The portfolio value is given by

$$\Pi_i = -C_i + \bar{\Delta}^{(1)}_{C_i} S_i + \bar{\Delta}^{(2)}_{C_i} \tilde{C}_i + B_i. \tag{4.50}$$

The hedging ratios will not be adjusted in the interval (t_i, t_{i+1}), hence the portfolio value before rebalancing at time t_{i+1} must be equal to

$$\Pi_{i+1} = -C_{i+1} + \bar{\Delta}^{(1)}_{C_i} S_{i+1} + \bar{\Delta}^{(2)}_{C_i} \tilde{C}_{i+1} + e^{r\Delta t} B_i. \tag{4.51}$$

At time t_{i+1}, we will rebalance the portfolio. In other words, we will buy/sell a certain amount of stock/options, such that our portfolio agrees with hedging strategy. This will have an effect on the amount of cash we are holding, thus

$$B_{i+1} = \left(\bar{\Delta}^{(1)}_{C_i} - \bar{\Delta}^{(1)}_{C_{i+1}}\right) S_{i+1} + \left(\bar{\Delta}^{(2)}_{C_i} - \bar{\Delta}^{(2)}_{C_{i+1}}\right) \tilde{C}_{i+1} + e^{r\Delta t} B_i. \tag{4.52}$$

After rebalancing, the portfolio value can be expressed as

$$\Pi_{i+1} = -C_{i+1} + \bar{\Delta}^{(1)}_{C_{i+1}} S_{i+1} + \bar{\Delta}^{(2)}_{C_{i+1}} \tilde{C}_{i+1} + B_{i+1}. \tag{4.53}$$

Rebalancing does not have an effect on the total value of the portfolio, we are just reallocating our resources from one asset to another. Hence, the expressions in equations (4.51) and (4.53) should be equal. This concludes the hedging process at time t_{i+1} and it will be repeated at time $t_{i+2},\dots,t_N$.

Ideally the portfolio value is equal to zero for every point in time, because the initial portfolio value is equal to zero. Every deviation from zero will be thought of as a hedge error. Over the entire life of the option we desire the mean and standard deviation of this error to be as close as possible to zero. To this end, we introduce the following error measures for simulation j

$$E^{(j)}_{\text{Mean}} = \frac{1}{N+1} \sum_{i=0}^{N} \Pi^{(j)}_{\frac{i}{N}T}, \qquad E^{(j)}_{\text{Std}} = \sqrt{\frac{1}{N} \sum_{i=0}^{N} \left(\Pi^{(j)}_{\frac{i}{N}T} - E^{(j)}_{\text{Mean}}\right)^2}. \tag{4.54}$$

The overall hedging performance of the M simulations can be judged by

$$\begin{cases} \bar{E}_{\text{Mean}} = \frac{1}{M}\sum_{j=1}^{M} E^{(j)}_{\text{Mean}}, \\ \bar{E}_{\text{Std}} = \frac{1}{M}\sum_{j=1}^{M} E^{(j)}_{\text{Std}}. \end{cases} \tag{4.55}$$

The hedge errors can be interpreted as follows

- The error mean indicates the hedge performance on average. A mean close to zero indicates the strategy is performing well on average. A positive/negative mean indicates the strategy is over/under-hedging on average.
- The standard deviation indicates how much the individual realisations deviate from the error mean. Ideally we want a standard deviation equal to zero, which implies a stable strategy. A high standard deviation can be caused by a misspecification in the hedging strategy (either wrong parameters or the wrong strategy in general) or by rebalancing too infrequently, in which case Δt needs to be decreased.

In the current set-up, these error measures are random variables, as they follow from Monte Carlo simulations. To this end, we will analyse the stability of the error measures across the simulated trajectories by the standard error

$$\text{SE}(E) = \frac{\sqrt{\frac{1}{M-1}\sum_{j=1}^{M}\left(\bar{E} - E^{(j)}\right)^2}}{\sqrt{M}}, \tag{4.56}$$

with error measure E, its mean $\bar{E}$ and its simulated trajectories $E^{(j)}$.

Finally, we have summarized the procedure described in this section in algorithm 1, where the realisations of S_t, v_t and the risk-neutral parameters follow from historical data or simulations.

Algorithm 1 Delta-Vega Hedge

procedure DELTA-VEGA STRATEGY
  Observe S_0, v_0 and the risk-neutral parameters at time $t = 0$.
  Determine C_0 and $\tilde{C}_0$.
  Calculate $\bar{\Delta}^{(1)}_{C_0}$ and $\bar{\Delta}^{(2)}_{C_0}$ using (4.13).
  Determine the initial bank account B_0 according to (4.48).
  Set the initial portfolio value Π_0 equal to zero.

  for $i \in \{0, \ldots, N-1\}$ **do**
    Observe S_{i+1}, v_{i+1} and the risk-neutral parameters at time $t = t_{i+1}$.
    Determine C_{i+1} and $\tilde{C}_{i+1}$.
    Calculate $\bar{\Delta}^{(1)}_{C_{i+1}}$ and $\bar{\Delta}^{(2)}_{C_{i+1}}$ using (4.13).
    Update the bank account B_{i+1} according to (4.52).
    Calculate the new portfolio Π_{i+1} value using either (4.51) or (4.53).

  Determine the hedge errors of this realization, $E^{(j)}_{\text{Mean}}$ and $E^{(j)}_{\text{Std}}$, with (4.54).

4.5. Simulation results

In this section we will present the results of the hedging strategies in different market environments. First, we will simulate the (real-world) market realizations according to the Black-Scholes, Heston and dynamic Heston models. The simulation processes are explained in appendix E. Hereafter, the hedging performance in these different markets will be assessed by applying the hedging strategies discussed in the previous section (see algorithm 1).

In practice, a hedger is subjected to more risks than just changes in option price. One of these risks is the so-called model risk: errors due to model misspecifications. In this section we investigate the impact of several model misspecifications. To be more precise, we will discuss

- The impact of misspecified parameters. For example, what happens if the volatility implied by the risk-neutral market does not coincide with the realized volatility.
- The impact of using the wrong model altogether. For instance, what is the effect on the hedging performance if one assumes the Heston model, but the market follows the dynamic Heston model.

Analysing the impact of these misspecifications will give us more insight in the hedging performance of historical data. One can think of these tests as a way to understand why certain hedging strategies succeed and others may fail.

4.5.1. General assumptions

In this section we assume the writer hedges a short position in a European call option with maturity $T = 1.0$ and strike $K = 50$. Moreover, at the starting date of the option we assume

$$\begin{cases} S_0 = 49, & v_0 = 0.05, & \bar{v}_0 = 0.1, & \gamma_0 = 0.7, \\ \kappa = 1.0, & \rho = -0.75, & r = 0.01, & \sigma = 0.215, \end{cases} \tag{4.57}$$

with implied volatility σ, which depends on the Heston parameters. The option price C_0 (according to both the Black-Scholes formula and the Heston FFT algorithm) is equal to 4.0. In case of a Delta-Vega hedge we will use an additional option depending on the same parameters, but with maturity $\tilde{T} = 2.0$ and strike $\tilde{K} = 50$. This leads to an implied volatility $\tilde{\sigma}$ of 0.231 and option price $\tilde{C}_0$ equal to 6.35.

4.5.2. Black-Scholes market

First, we will consider a market simulated with the real-world Black-Scholes model. We will apply the Black-Scholes Delta strategy to the simulated paths with different Δt to assess its hedging performance. The simulation is performed under the real-world measure, thus

$$\mathrm{d}S_t = \mu \mathrm{d}t + \sigma S_t \mathrm{d}W_t^S, \tag{4.58}$$

using σ as specified in (4.57) and $\mu = 0.1$. Note that μ does not have to be equal to r, since the portfolio dynamics derived in (4.19) do not depend on the real-world drift. After all, we are hedging against changes of the underlying asset, hence deterministic changes should not influence the performance. On the other hand, the hedging strategy does depend on σ. It is therefore important that the implied volatility is equal to the realized volatility. In terms of the simplified example of section 4.1, this is similar to changing the possible outcomes of S_1. An increased volatility would lead to more extreme asset values at $T = 1$, which leads to a different hedging strategy and option value. By expecting a higher/lower volatility, we will protect ourselves more (less) against random movements of the asset price, which will lead to over-hedging (under-hedging). Thus, in a well-specified model, the risk-neutral and real-world σ are equal.

The results after applying using $M = 1000$ simulations are presented in table 4.1.

Frequency[6]	$\bar{E}_{\text{Mean}}$	$\bar{E}_{\text{Std}}$
Once per week	-0.008 *(0.0078)*	0.163 *(0.0027)*
Once per day	0.011 *(0.0034)*	0.079 *(0.0013)*
Ten times per day	0.000 *(0.0012)*	0.025 *(0.0004)*

Table 4.1: Black-Scholes Delta hedge errors. The standard errors of the estimates are given between the parentheses.

These results are in line with the portfolio dynamics derived in (4.19), where we showed that the portfolio is hedged against changes of the asset price. In the Black-Scholes market, asset price is the only dynamic variable, hence this portfolio is fully hedged. In this case, the mean hedge error is close to zero. Moreover, by increasing the rebalance frequency, we decrease the standard deviation, but the mean remains (mostly) unchanged. In other words, on average we will always have the same hedge error, but increasing the number of rebalances will decrease the deviation from this average.

A visualisation of the strategy for two possible realizations is shown in figures 4.2 and 4.3. These

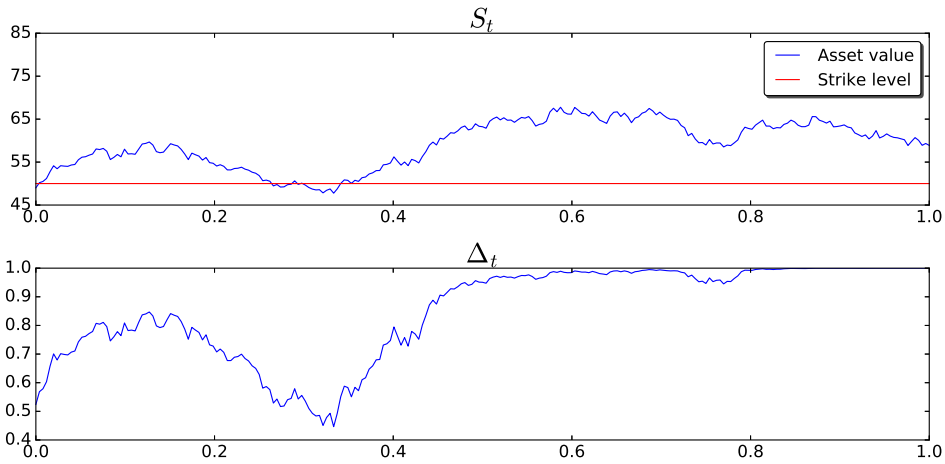

Figure 4.2: Black-Scholes Delta strategy when S_t ends in-the-money.

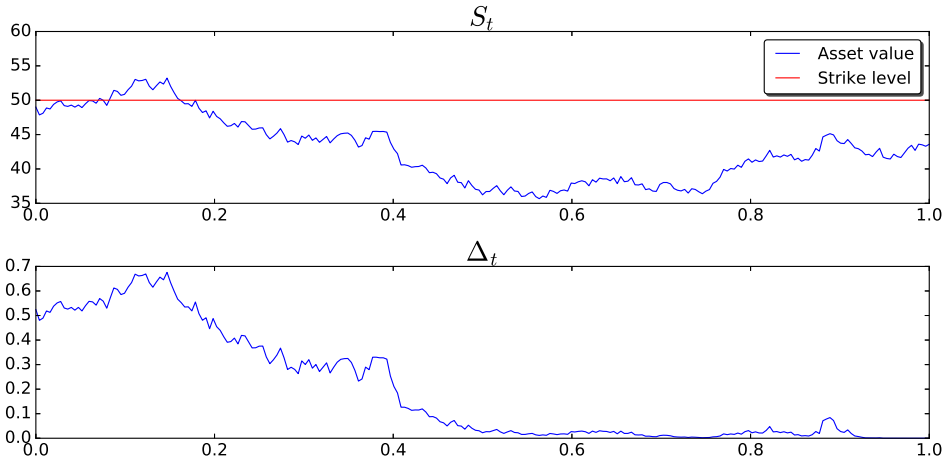

Figure 4.3: Black-Scholes Delta strategy when S_t ends out-of-the-money.

figures represent the expectations we discussed in the previous sections. When the probability of a positive pay-off is high, then the hedger should buy more stocks to cover a possible loss. In this case, a high value of S_t coincides with a higher probability of a positive pay-off, hence high values of S_t lead to high values of Δ_t. Note that this relationship becomes more pronounced closer to maturity, because there is an even lower probability that S_t decreases below K. Contrary, the probability of a positive pay-off decreases when S_t has a low value. In that case, the hedger does not have to protected against a possible loss and Δ_t will be low.

One of the reasons this test works well is because the realized and implied volatility are equal. In practice, we do not know whether the volatility implied by the risk-neutral market is equal to volatility realized during the life of the option. To test the impact of this model risk, we performed the same test as before, but with risk-neutral/hedging volatility σ_{imp} different from the realized volatility σ. The results can be found in table 4.2.

It is clear that a misspecified volatility can lead to serious hedging errors. The strategy aims at protecting the portfolio from changes in asset price. By assuming a higher (lower) volatility, we will expect more (less) volatile paths which lead to too much (little) protection against the *actual* movements of S_t, as discussed above. This behaviour is reflected in the test results, where higher/lower volatility leads to an over/under-hedged portfolio. In practice, we cannot choose the implied and realized volatility, but they are

[6] Based on 52 weeks and 252 trading days per year.

Frequency	$\bar{E}_{\text{Mean}}$	$\bar{E}_{\text{Std}}$
$\sigma_{\text{imp}} = 0.8\sigma$:		
Once per week	-0.418 *(0.0087)*	0.284 *(0.0054)*
Once per day	-0.445 *(0.0054)*	0.264 *(0.0040)*
Ten times per day	-0.449 *(0.0035)*	0.257 *(0.0031)*
$\sigma_{\text{imp}} = 1.2\sigma$:		
Once per week	0.416 *(0.0077)*	0.292 *(0.0045)*
Once per day	0.417 *(0.0041)*	0.257 *(0.0030)*
Ten times per day	0.417 *(0.0025)*	0.251 *(0.0024)*

Table 4.2: Black-Scholes Delta hedge errors with misspecified σ. The standard errors of the estimates are given between the parentheses.

defined by the market and not necessarily equal. Thus, hedging errors in practice are always subjected to this model risk and this misspecification might explain why certain hedging strategies perform worse than others.

4.5.3. Heston market

Now, we will assume a market which follows the dynamics of a real-world Heston model, which is simulated by the simulation scheme discussed in appendix E. Girsanov's Theorem is able to transform the real-world Heston stochastic differential equations to the risk-neutral Heston model. By doing so, one can derive the following real-world Heston model

$$\begin{cases} \mathrm{d}S_t = \mu S_t \mathrm{d}t + \sqrt{v_t} S_t \mathrm{d}W_t^1, \qquad S(0) = S_0, \\ \mathrm{d}v_t = \kappa^*(\bar{v}^* - v_t)\mathrm{d}t + \gamma\sqrt{v_t}\mathrm{d}W_t^2, \qquad v(0) = v_0, \\ \langle \mathrm{d}W_t^1, \mathrm{d}W_t^2 \rangle = \rho \mathrm{d}t. \end{cases} \tag{4.59}$$

Note that the real-world κ^* and $\bar{v}^*$ might differ from the risk-neutral κ and $\bar{v}$. Similar to μ, we do not expect these parameters to influence the hedging performance. Indeed, the Heston Delta-Vega portfolio dynamics derived in equation (4.30) do not depend on κ and $\bar{v}$[7], because they only influence the deterministic behaviour of the volatility. So, without loss of generality, we will assume

$$\kappa^*, \bar{v}^* = \kappa, \bar{v}. \tag{4.60}$$

On the other hand, v_0, γ and ρ are equal under the real-world and risk-neutral measure in a well-specified model. Misspecifying these parameters will lead to over/under-hedged portfolios, similar to a misspecified σ in the Black-Scholes model. Consequently, the Heston model is subjected to the same model risk as the Black-Scholes models, namely the risk of misspecified risk-neutral parameters.

First, we will perform the Delta hedge strategy. We will apply the Black-Scholes and Heston strategies in order to investigate the impact of assuming an incorrect model. In other words, what happens to the hedge performance when the hedging strategy assumes Black-Scholes but the market follows Heston dynamics. The results of the $M = 1000$ simulations are given in table 4.3.

On average, these strategies appear to perform relatively well, since the mean errors of both models are close zero. However, the strategies are less stable, indicated by high standard deviations. Increasing the rebalancing frequency slightly reduces the standard deviation, but it is not nearly as stable as Delta hedging in the Black-Scholes market. In other words, a trader implementing this strategy in a Heston market cannot rely on its performance. In some cases this strategy might over-hedge, while in other cases the portfolio is under-hedged. This is of course an undesirable feature and demonstrates the need for the Delta-Vega hedge. It is noteworthy, however, that the results of the Black-Scholes strategy are similar to the Heston strategy. The standard deviation of the Black-Scholes hedging strategy is slightly higher than in the Heston model, but the impact of a misspecified model is not nearly as big as the impact of misspecified parameters. This indicates the robustness of the Black-Scholes model, despite its unrealistic

[7]The Heston Delta strategy does depend on κ and $\bar{v}$, since it does not fully hedge against changes of volatility.

Frequency	$\bar{E}_{\text{Mean}}$	$\bar{E}_{\text{Std}}$
Black-Scholes:		
Once per week	0.030 *(0.0389)*	0.665 *(0.0105)*
Once per day	-0.087 *(0.0399)*	0.643 *(0.0095)*
Heston:		
Once per week	0.014 *(0.0332)*	0.569 *(0.0095)*
Once per day	-0.073 *(0.0348)*	0.543 *(0.0089)*

Table 4.3: Delta hedge errors for Black-Scholes and Heston. The standard errors of the estimates are given between the parentheses.

assumptions.

In order to overcome the unreliable performance of the Delta hedge portfolio, we will implement the Delta-Vega hedge strategy. In case of the Heston model, this strategy should remove all dependence on changes in the underlying asset and volatility. Again, the Black-Scholes and Heston strategies are both implemented, to assess the impact of model misspecification, the results are presented in table 4.4.

Frequency	$\bar{E}_{\text{Mean}}$	$\bar{E}_{\text{Std}}$
Black-Scholes:		
Once per week	0.018 *(0.0173)*	0.327 *(0.0047)*
Once per day	-0.037 *(0.0172)*	0.305 *(0.0039)*
Heston:		
Once per week	0.005 *(0.0030)*	0.095 *(0.0026)*
Once per day	0.000 *(0.0014)*	0.043 *(0.0012)*

Table 4.4: Delta-Vega hedge errors for Black-Scholes and Heston. The standard errors of the estimates are given between the parentheses.

In this test, the Delta-Vega strategy clearly outperforms the Delta strategy, in terms of standard deviation. Both models produce much more stable results when taking changes of volatility into account, especially the Heston model. In this case, the effect of assuming an incorrect model becomes more clear. On average, the Black-Scholes model performs similar to the Heston model, but the individual realisations are less stable and this problem cannot be solved by increasing the rebalancing frequency. In other words, assuming an incorrect model introduces additional risk which cannot be hedged, but is negligible on average in the current set-up.

4.5.4. Dynamic Heston market

Finally, we will consider the dynamic Heston market, which is most in line with historical observations. We will simulate this market by assuming $\bar{v}^*$ and γ in (4.59) to be time-dependent and defined as in (4.32). The performance of the Heston Delta-Vega, the dynamic Heston Delta-Vega and the dynamic Heston full hedge strategies in this market will be compared. The dynamic Heston hedging strategies take changes of $\bar{v}$ and γ into account (at least to some extent), hence we expect them to outperform the original Heston hedging strategies. The hedging strategies and simulation highly depend on the parameters of $\bar{v}$ and γ, which we need to define beforehand. Because there is a high correlation between v_t and the Heston parameters, we will assume

$$\begin{cases} \kappa_{\bar{v}} = 1.4, & \bar{v}_{\text{Mean}} = 0.1, & \alpha_{\bar{v}} = 0.8, & \rho_{\bar{v}} = 0.9, \\ \kappa_{\gamma} = 2.1, & \gamma_{\text{Mean}} = 0.7, & \alpha_{\gamma} = 1.0, & \rho_{\gamma} = 0.9. \end{cases} \tag{4.61}$$

In reality, these parameters cannot be chosen, but they are implied by the market. By analysing historical behaviour of $\bar{v}_t$ and γ_t, one is able to estimate their SDE parameters. In this case, however, we

will assume to know these parameters and use them when determining the hedging strategy. The performances of this set-up based on $M = 200$ simulations are given in table 4.5.

Frequency	$\bar{E}_{\text{Mean}}$	$\bar{E}_{\text{Std}}$
Delta-Vega Heston:		
Once per week	-0.339 *(0.0207)*	0.228 *(0.0084)*
Once per day	-0.354 *(0.0227)*	0.210 *(0.0080)*
Delta-Vega dynamic Heston:		
Once per week	-0.204 *(0.0097)*	0.137 *(0.0072)*
Once per day	-0.220 *(0.0071)*	0.114 *(0.0047)*
Full hedge dynamic Heston:		
Once per week	-0.044 *(0.0040)*	0.068 *(0.0042)*
Once per day	-0.051 *(0.0030)*	0.045 *(0.0022)*

Table 4.5: Hedge errors for Heston and dynamic Heston in a dynamic Heston market. The standard errors of the estimates are given between the parentheses.

These results show that it is beneficial to take parameter correlations into account when hedging, both in terms of mean error and standard deviation. While still not perfect, the dynamic Heston Delta-Vega is better able to remain risk-neutral on average and it deviates less from this average. The full hedge performs even better with a mean approximately equal to zero and a standard deviation equal to or lower than any of the previous strategies, despite the dynamic behaviour of the market.

The purpose of these hedging strategies is to replicate the value of an option. In case of the Heston Delta-Vega hedge, only S_t and v_t are allowed to change. By respecting the assumptions of the Heston model, we are not fully able to replicate the option value in the future. On the other hand, by assuming the dynamic Heston model, $\bar{v}_t$ and γ_t are allowed to change as well. Hedging strategies considering this dynamic behaviour, turn out to produce more accurate future option price estimates in the current set-up. This indicates the importance of taking dynamic parameters into account when determining future option prices, even when the assumptions of the underlying model are violated. But, do note that the comparison is not completely fair, since we specifically assume a dynamic Heston market. It is therefore no big surprise that a strategy taking these assumptions into account outperforms one that does not. In order to assess their true performance, we will apply these strategies to historical data in the next section.

4.6. Empirical results

In the previous section we have shown that it is beneficial for hedging portfolios to take changes of the Heston parameters into account. However, these observations are based on a controlled environment, where we know the behaviour of all market variables. When hedging in practice, the underlying assumptions are not always respected and the *true* parameters are unknown. To quantify the effect of these difficulties, we will test the hedging strategies on historical data[8] in this section. Three different strategies are considered:

1. A Heston Delta-Vega strategy with hedging ratios described (4.28) and (4.29).
2. A dynamic Heston Delta-Vega strategy. Thus, the hedging ratios are calculated according to (4.39) and (4.40).
3. A full dynamic Heston hedge where the hedging ratios are defined as in (4.45).

All hedging ratios will depend on the recalibrated risk-neutral parameters, hence $\bar{v}$, γ and ρ will vary over time, according to changes in the historically observed implied volatility surfaces. Moreover, the dynamic Heston Delta-Vega hedge depends on the correlation and volatility of $\bar{v}$ and γ, which we define

[8]The data set equals the daily data set discussed in 3.4.1.

as follows

$$\begin{cases} \rho_{\bar{v}} = \rho_{\gamma} = 0.95, \\ a_{\bar{v}} = \sqrt{\frac{\frac{1}{N}\sum_{i=-N}^{-1} \log\left(\frac{\bar{v}_{t_{i+1}}}{\bar{v}_{t_i}}\right)^2}{\Delta t}}, \\ a_{\gamma} = \sqrt{\frac{\frac{1}{N}\sum_{i=-N}^{-1} \log\left(\frac{\gamma_{t_{i+1}}}{\gamma_{t_i}}\right)^2}{\Delta t}}. \end{cases} \tag{4.62}$$

It can be quite challenging to determine the correlation between the Brownian motions driving the parameters, hence we assume it to be equal for each time-interval. Moreover, the volatility estimator only depends on past observations; the sum indices vary from $-N$ to -1 with $t_{-N} = -1$, where $t_0 = 0$ indicates the starting date of the option.

In this test we will hedge an at-the-money European call option with one year maturity on the S&P-500 index. All hedging strategies will be subjected to daily rebalances which are based on the parameters as seen on that date. The transaction costs will be excluded from this test, as we are interested in the performance of the hedging strategies with respect to changes in $\bar{v}$ and γ. Including transaction costs would for example increase the costs of the full dynamic Heston hedging strategy, as it involves more financial assets. This is would bias the results and we therefore exclude the transaction costs from this test.

The test will be repeated on a monthly basis from July 2006 until February 2013 and the performance will be assessed by the mean error and mean squared error during the life of the option,

$$E^{(j)}_{\text{Mean}} = \frac{1}{N+1}\sum_{i=0}^{N} \Pi^{(j)}_{\frac{i}{N}T}, \qquad E^{(j)}_{\text{MSE}} = \frac{1}{N+1}\sum_{i=0}^{N} \left(\Pi^{(j)}_{\frac{i}{N}T}\right)^2. \tag{4.63}$$

The time intervals of the hedging portfolios will overlap in this set-up, since the test is repeated on a monthly basis and the option maturity is one year. However, all strategies will depend on different initial conditions and will therefore perform differently, despite the overlapping time-intervals. The results of this test are graphically presented in figure 4.4. The hedging performances are similar to the simulation

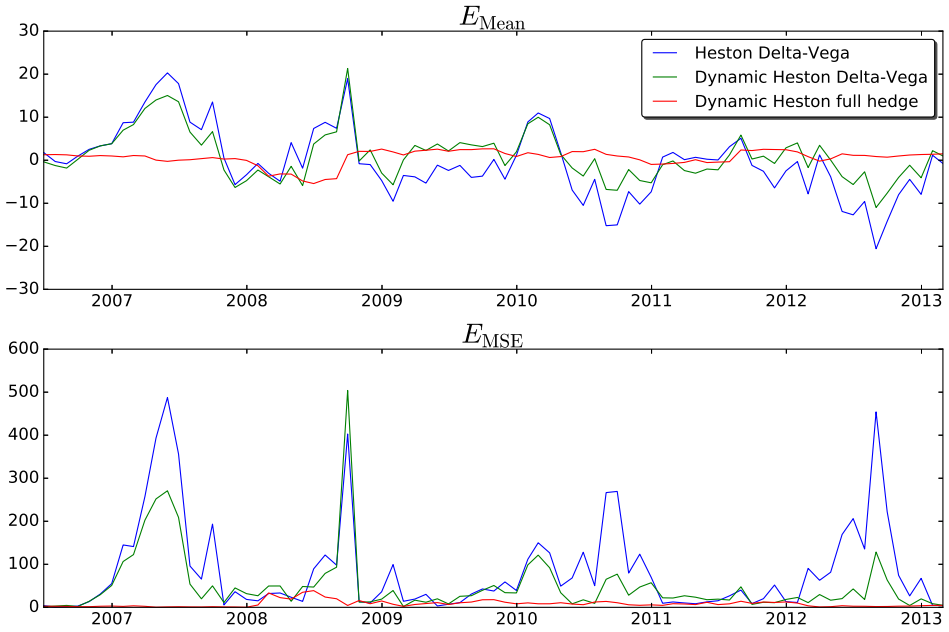

Figure 4.4: Mean error and Mean squared error for different hedging strategies performed on monthly historical data

results:

- The Heston Delta-Vega does not take changes of the parameters into account and appears to be the most unstable method. This strategy has the most and highest error 'peaks' and is therefore the most unreliable.

- The dynamic Heston Delta-Vega is still not perfect but appears to be more stable than the Heston Delta-Vega, the error 'peaks' happen less frequently and are less pronounced. The error can be minimized by optimizing the correlation and volatility of $\bar{v}_t$ and γ_t, but the optimization can only be performed afterwards, which is not the objective of this test.
- The dynamic Heston full hedge is the most stable out of the three strategies. It does not have any error 'peaks' and outperforms the other two strategies in most cases. Moreover, this strategy does not depend on additional parameters which can introduce an error if chosen poorly, such as in the dynamic Heston Delta-Vega hedge.

We can conclude that respecting the assumptions of the underlying model (in this case the Heston model) does not necessarily lead to more accurate future option prices. By taking recalibration of the parameters into account we were able to replicate option values more accurately both in a controlled (simulation) and uncontrolled (historical) environment.

4.7. Implications

Throughout this chapter we have demonstrated the effect of recalibration on hedging strategies. These strategies all aim at replicating the value of an option, which is driven by different market factors. The most important factors are asset price and volatility of the asset price, something which most risk-neutral pricing methods acknowledge. In practice, however, an option price is also affected by recalibration of (constant) parameters. By means of hedge tests, we have shown that the effect of recalibration is non-negligible. Hedging strategies taking recalibration into account significantly outperform ones that do not, both in simulation and empirical tests. This leads to the conclusion that recalibration is in fact an important factor when determining future option prices.

We started this chapter by raising the question: 'Is it justified to assume a recalibrated risk-neutral measure in (real-world) simulations, even when this is in conflict with the assumptions of the underlying risk-neutral pricing model?'. Throughout this chapter we have discussed this issue with respect to the Heston model.

In theory, only the asset price and volatility of the asset price are allowed to change according to the Heston model. Hence, one can argue that the recalibration process with respect to $\bar{v}$ and γ should not be included in real-world simulations, as this violates the assumptions of the Heston model. Instead, $\bar{v}$ and γ should be constant and their values are implied by the implied volatility surface at $t = 0$. This leads to a unique risk-neutral measure, one everyone agrees with. If one is interested in option prices at $t > 0$, for example in case of risk-management, the unique risk-neutral measure can be used. This way, everyone agrees on the risk associated with certain positions. From a regulatory point of view this might be beneficial, as it is important that the risk assessment of regulatory frameworks is unique. However, this approach does ignore the recalibration process. Indeed, it considers $\bar{v}$ and γ to be constant, which is not in line with historical observations. As we have shown throughout this chapter, ignoring the recalibration process can lead to serious errors with respect to future option prices.

By implementing the regression models proposed in chapter 3, one is able to include the recalibration process in a real-world simulation. This way, one obtains $\bar{v}$ and γ calibrated to the simulated state of the market. Intuitively this process can be explained as follows. If the market crashes, it is much more likely to observe a high $\bar{v}$ and γ, based on historical data. Conversely, if the market flourishes it is much more likely to observe a low $\bar{v}$ and γ. By implementing the recalibration process in a real-world simulation, one creates this behaviour in the simulated states of the market. Assuming $\bar{v}$ and γ are allowed to change over time will lead to more accurate future option prices, according to the hedge tests presented in this chapter. From a risk-management point of view this might be beneficial, as it is important to accurately assess the risk associated with certain positions. However, one does lose the unique interpretability of the risk-neutral measure. A key feature of the risk-neutral market is the fact that it is arbitrage-free and everyone agrees upon the market prices. By subjecting the implied volatility surface to recalibration one loses this uniqueness, as it becomes dependent on the underlying assumptions with respect to the evolution of the risk-neutral parameters. This way, different risk-managers might come to different conclusions, based on their perspective of the market.

In conclusion, it is difficult to give a decisive answer to the proposed research question, as it depends on the point of view. If one wishes to assess the risk involved with certain positions in the most accurate way, then the answer is yes, by violating the constant $\bar{v}$ and γ assumption of the Heston model one obtains more accurate future option prices, according to our simulation and empirical tests. However, if the objective is to create a unique framework where everyone agrees upon the risk associated with certain positions, then the answer is no, by implementing the recalibration process one subjects the implied volatility surface to the modeller's point of view and different practitioners might come to different conclusions.

5

Impact test

In the previous chapters we have discussed several models that forecast the risk-neutral measure conditioned on the (simulated) state of the market. The approaches have been tested in terms of prediction accuracy and a hedge test. In this chapter, however, we will not discuss the justification of the proposed methodologies. Instead, we will show the impact of these models in terms of risk management. More precisely, we will apply the models on an important quantity for insurers: the Solvency Capital Requirement.

On January 1 of 2016 the Solvency II directive came into effect with the purpose of reducing the probability that an insurer is unable to pay outstanding claims to its policyholders. One of the most important regulatory measures was the introduction of the Solvency Capital Requirement (SCR). The SCR is defined as the minimum amount of capital held by an insurer, such that it is able to pay its claims over a one-year horizon with a 99.5% probability. The regulator demands that the insurer's available capital must be greater than or equal to the SCR, which stresses the importance of this quantity.

According to the directive, an insurer can either use the standard formula or an internal model to determine the SCR. The standard formula assesses the insurer's risk towards several events[1] by applying deterministic shocks to the balance sheet. The SCR is calculated by taking correlations between these different risks into account. The standard formula is known to be relatively conservative and it generally fails to give a realistic reflection of the insurer's risk profile. An internal model, on the other hand, is tailored to the characteristics of the insurer. Internal models usually involve Monte Carlo simulations of the balance sheet at $t = 1$ and the SCR is defined as the 99.5% quantile of the loss function. Internal models are generally more accurate than the standard formula, however, it can be quite complicated and expensive to develop an internal model that satisfies the strict guidelines of the regulator. Consequently, most small to medium-sized insurance companies resort to the standard formula. Throughout this chapter we will only discuss the impact of a risk-neutral measure with time-dependent parameters on internal models, since the standard formula is fully defined and therefore unaffected by our models. For more information regarding the similarities and dissimilarities between the standard formula and internal models we refer the reader to Devineau and Loisel (2009).

As discussed above, the SCR highly depends on the current and future balance sheet of the insurer, which consists of assets and liabilities. Among others, the liabilities depend on the *fair* value of outstanding claims. These claims usually involve guarantees (also called embedded options), which are rights that can generate a profit but never a loss to the policyholder. The value of these guarantees is defined under the risk-neutral measure. In order to hedge against the risks involved with these claims, insurers often acquire (complex) option portfolios that require risk-neutral valuation as well. It is therefore of vital importance to correctly identify the risk-neutral measure when determining the SCR and this is where our models come into play. The SCR does not only depend on the value of the balance sheet at $t = 0$, it also depends on the distribution of the balance sheet value at $t = 1$. Hence, we need to define $\mathbb{Q}$ at time $t = 0$ as well as $t = 1$. The risk-neutral measure at $t = 0$ depends on the observed implied volatility surface and

[1] Such as changes in stock price, interest rate or mortality rate.

is therefore well-defined. However, the definition of the risk-neutral measure at $t = 1$ is debatable. None of the existing literature regarding the SCR mentions recalibration of the risk-neutral measure at $t = 1$ (see for example Bauer et al. (2010) or Feng et al. (2017)). But, as we have shown by means of a hedge test, ignoring recalibration has a significant impact on future option prices. We will therefore study the effect on the SCR of different assumptions regarding the risk-neutral measure at $t = 1$. We will compare a risk-neutral measure with time-dependent parameters, as proposed in the chapter 3, to a risk-neutral measure with constant parameters.

This chapter is structured as follows, first we will present the formal definition of the SCR. Hereafter we will discuss different approaches to calculate the SCR. Finally, we will discuss the impact of time-dependent versus constant parameters in terms of the SCR, based on historical data.

5.1. Solvency Capital Requirement

In this section we will present the formal definition of the Solvency Capital Requirement associated with a certain contract, following the notation of Feng et al. (2017), supplemented with examples to visualize the process.

5.1.1. Mathematical definition

Assume an insurer sells a policy with maturity T. We denote the policy's net income over the interval $[0,t]$ by A_t and it is defined by the cash flows generated under the real-world market. The mathematical definition is given by

$$A_t = \int_0^t e^{\mu(t-s)}\text{cashflow}(s,\Delta s)\text{d}s, \tag{5.1}$$

where μ is the expected return and cashflow$(s,\Delta s)$ denotes the generated cash flow over the interval $[s, s+\Delta s]$. Similarly, we define the policy's liabilities by L_t and they are given by the discounted expected cash flows under the risk-neutral measure over the interval $[t,T]$:

$$L_t = \mathbb{E}^{\mathbb{Q}_t}\left[\int_t^T e^{-r(s-t)}\text{cashflow}(s,\Delta s)\text{d}s \,\middle|\, \mathscr{F}_t\right], \tag{5.2}$$

with risk-free rate r. Note that a positive/negative cash flow corresponds to an income/liability for the insurer. Finally we define

$$N_t = A_t - L_t, \tag{5.3}$$

which can be thought of as the policy's net value at time t. The Solvency Capital Requirement is defined as the 99.5% Value-at-Risk (i.e. the 99.5% quantile) of the 1 year loss distribution under the real-world measure,

$$\text{SCR} = \text{VaR}_{0.995}\left(N_0 - \tilde{N}_1\right) := \inf\left\{x \,\middle|\, \mathbb{A}\left(N_0 - \tilde{N}_1 < x\right) > 0.995\right\}. \tag{5.4}$$

Here, $\tilde{N}_1$ is defined as the discounted value of N_1.

5.1.2. Example: Variable Annuities

A variable annuity is a well-known contract offered by insurers. With a variable annuity, the policyholder typically pays a premium and chooses how this premium is invested (in stocks, bonds, money market, etc.). In return, the policyholder receives a single or periodic payment from the insurer. The size of the payments depends on the performance of the investment fund during the so-called accumulation phase. The insurer also offers additional guarantees (or riders) in exchange for a fee, as an insurance against unforeseen losses. There exist different types of guarantees, for instance

- Guaranteed Minimum Accumulation Benefit (GMAB) guarantees a minimum payout on expiration of the annuity, for example the initial premium. If the fund value exceeds this amount, the policyholder receives a payment equal to the fund value.
- With the Guaranteed Minimum Withdrawal Benefit (GMWB) a policyholder can annually withdraw a certain percentage (usually around 7%) of the initial premium until the entire amount is recovered. The policyholder can continue to make these withdrawals, even if the fund value has been depleted.

- Guaranteed Minimum Death Benefit (GMDB) offers a payout if the policyholder dies during the lifetime of the variable annuity.

More information regarding different guarantees and how to price them can be found in Bauer et al. (2008). Note that these guarantees can also be combined to meet the demands of the policyholder, which can lead to complex products. It is therefore important for the insurer to identify the risks associated with these products and to be protected against these risks by holding some reserves. These reserves should, in case of lacking fund values, be able to pay the claims of policyholders and they must at least be equal to the SCR of the offered products. This indicates the importance of the SCR: inaccurate estimations could potentially lead to the default of an insurer during a financial crisis.

The remainder of this section will be dedicated to deriving the fund dynamics and SCR of two different variable annuities: one with a GMAB rider and the other with a GMWB rider.

Guaranteed Minimum Accumulation Benefit

In this example we will assume the fund only contains stocks, but the derivation is similar when different assets are combined. Denote the stock and fund value by S_t and F_t, respectively, and we define the initial premium as G. Finally, we assume the payout at maturity T is at least equal to the initial premium, in other words

$$\text{Payout}_T = \max(F_T, G). \tag{5.5}$$

The dynamics of the fund are very similar to the stock dynamics, except for the fee α which is deducted from the fund, as payment to the insurer. This fee can be thought of as a dividend yield. According to Milevsky and Salisbury (2001), we therefore obtain

$$\mathrm{d}F_t = \mathrm{d}S_t \frac{F_t}{S_t} e^{-\alpha t}, \qquad F_0 = G. \tag{5.6}$$

The specific dynamics depend on the assumptions regarding S_t. In this chapter we will assume the Black-Scholes model under the observed real-world measure $\mathbb{A}$, giving

$$\mathrm{d}F_t^{\mathbb{A}} = (\mu - \alpha) F_t^{\mathbb{A}} \mathrm{d}t + \sigma F_t^{\mathbb{A}} \mathrm{d}W_t^{\mathbb{A}}, \qquad F_0 = G. \tag{5.7}$$

Moreover, a risk-neutral Heston model will be implemented, leading to

$$\begin{cases} \mathrm{d}F_t = (r - \alpha) F_t \mathrm{d}t + \sqrt{v_t} F_t \mathrm{d}W_t^F, & F_0 = G, \\ \mathrm{d}v_t = \kappa(\bar{v} - v_t)\mathrm{d}t + \gamma\sqrt{v_t}\mathrm{d}W_t^v, \\ \langle \mathrm{d}W_t^S, \mathrm{d}W_t^v \rangle = \rho \mathrm{d}t. \end{cases} \tag{5.8}$$

The income is generated by the accumulated fees, hence

$$A_t = \int_0^t \alpha F_s^{\mathbb{A}} e^{\mu(t-s)} \mathrm{d}s. \tag{5.9}$$

The liabilities, on the other hand, depend on the final value of the fund:

- If $F_T \geq G$: the policyholder receives F_T and the insurer has no liabilities.
- If $F_T < G$: the policyholder receives G and the liabilities of the insurer are equal to $G - F_T$.

Moreover, the insurer continues to claim future fees, hence, according to (5.2) we can write the liabilities as

$$\begin{aligned} L_t &= \mathbb{E}^{\mathbb{Q}_t}\left[e^{-r(T-t)} \max(G - F_T, 0) - \int_t^T e^{-r(s-t)} \alpha F_s \mathrm{d}s \,\middle|\, \mathscr{F}_t \right] \\ &= e^{-r(T-t)} \mathbb{E}^{\mathbb{Q}_t}[\max(G - F_T, 0) | \mathscr{F}_t] - \mathbb{E}^{\mathbb{Q}_t}\left[\int_t^T e^{-r(s-t)} \alpha F_s \mathrm{d}s \,\middle|\, \mathscr{F}_t \right] \\ &= \text{Put}(F_t, G) - \int_t^T e^{-r(s-t)} \alpha \mathbb{E}^{\mathbb{Q}_t}[F_s | \mathscr{F}_t] \mathrm{d}s \\ &= \text{Put}(F_t, G) - \int_t^T e^{-r(s-t)} \alpha F_t e^{(r-\alpha)(s-t)} \mathrm{d}s \end{aligned}$$

$$
\begin{aligned}
&= \text{Put}(F_t, G) - \alpha F_t \int_t^T e^{-\alpha(s-t)} \mathrm{d}s, \\
&= \text{Put}(F_t, G) + F_t \left(e^{-\alpha(T-t)} - 1 \right),
\end{aligned} \tag{5.10}
$$

where $\text{Put}(F_t, G)$ denotes the value of a European put option on the fund at time t with strike price G and dividend yield α. Now, we can substitute these definitions in (5.4) to obtain

$$
\begin{aligned}
\text{SCR} &= \text{VaR}_{0.995} \left(N_0 - e^{-r} N_1 \right) \\
&= \text{VaR}_{0.995} \left(A_0 - L_0 - e^{-r} (A_1 - L_1) \right) \\
&= \text{VaR}_{0.995} \left(0 - \text{Put}(F_0, G) - F_0 \left(e^{-\alpha T} - 1 \right) - e^{-r} \left[\int_0^1 \alpha F_s e^{\mu(1-s)} \mathrm{d}s - \text{Put}(F_1, G) - F_1 \left(e^{-\alpha(T-1)} - 1 \right) \right] \right) \\
&= \text{VaR}_{0.995} \left(e^{-r} \left[\text{Put}(F_1, G) + F_1 \left(e^{-\alpha(T-1)} - 1 \right) - \int_0^1 \alpha F_s e^{\mu(1-s)} \mathrm{d}s \right] \right) - \text{Put}(F_0, G) - F_0 \left(e^{-\alpha T} - 1 \right) \\
&= e^{-r} \text{VaR}_{0.995} \left(\text{Put}(F_1, G) - g(F_1, 1) \right) - \text{Put}(F_0, G) + g(F_0, 0).
\end{aligned} \tag{5.11}
$$

With

$$
g(F_t, t) = \int_0^t \alpha F_s e^{\mu(t-s)} \mathrm{d}s + F_t \left(1 - e^{-\alpha(T-t)} \right), \tag{5.12}
$$

which can be thought of as the sum of realized and expected fees. In this case, the SCR depends on the real-world distribution of F_1, which determines $g(F_1, 1)$ and influences the risk-neutral valuation of $\text{Put}(F_1, G)$. The distributions of $g(F_1, 1)$ and $\text{Put}(F_1, G)$ can be obtained by means of a Monte Carlo simulation, which takes the following steps

- First, the net policy value N_0 is determined, according to definition (5.3).
- Hereafter, the fund value F_t along with other explanatory variables are simulated according to the real-world measure up to time $t = 1$.
- Then, the values of $\text{Put}(F_1, G)$ and $g(F_1, 1)$ are evaluated for each trajectory. The value of $g(F_1, 1)$ can be obtained directly from the trajectory of F_t, but $\text{Put}(F_1, G)$ requires a risk-neutral valuation for which the risk-neutral measure at time $t = 1$ is required.
- Finally, the simulated values are combined to construct the loss distribution. The SCR corresponds to the 99.5% quantile of this distribution.

In order to clarify this process, we will discuss two possible scenarios of F_t, along with their income and liabilities.

Consider a Black-Scholes real-world measure as defined in (5.7), with $F_0 = 1000$ and $\alpha = 0.01$. Moreover, assume that the VIX index follows (5.48), with $\text{VIX}_0 = 20.7$. The process parameters are equal to (5.49). At times $t = 0$ and $t = 1$, we will value the European put option according to the Heston model defined in (5.8), with

$$
\bar{v} = 0.08, \qquad \gamma = 0.55, \qquad r = 0.04, \qquad T = 5, \qquad G = 1000. \tag{5.13}
$$

Initially, we can determine the net policy value:

$$
\begin{aligned}
N_0 &= -\text{Put}(F_0, G) + g(F_0, 0) \\
&= -\text{Put}(F_0, G) + F_0 \left(1 - e^{-\alpha T} \right) + \int_0^0 \alpha F_s e^{-s\mu} \mathrm{d}s \\
&= -83.7,
\end{aligned} \tag{5.14}
$$

where the put option is priced according to the FFT algorithm described in appendix A.13. Hereafter, the fund value and the other explanatory variables are simulated according to the real-world measure. Two possible scenarios of the fund value and the accompanied VIX index can be found in figure 5.1.

Clearly, both scenarios will give different net policy values at $t = 1$,

$$
N_1^{(i)} = -\text{Put}(F_1^{(i)}, G) + g(F_1^{(i)}, 1). \tag{5.15}
$$

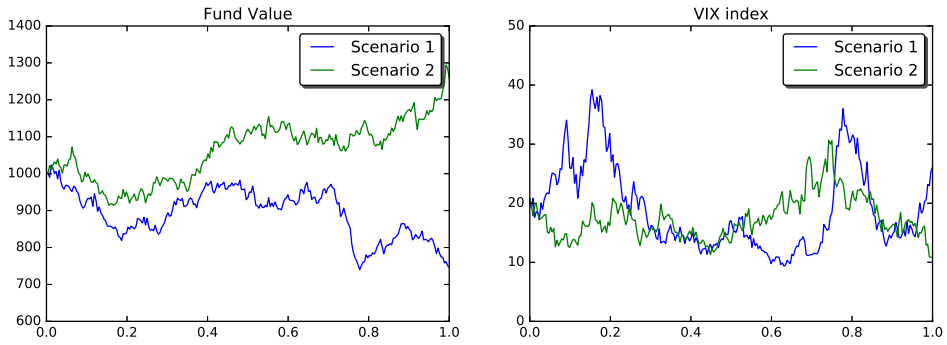

Figure 5.1: Possible scenarios of the fund value and VIX index over a 1-year horizon

The sum of realized and expected fees, $g(F_1^{(i)},1)$, can be calculated directly from the trajectories,

$$\begin{cases} g(F_1^{(1)},1) = \int_0^1 \alpha F_s^{(1)} e^{\mu(t-s)} \mathrm{d}s + F_1^{(1)}\left(1-e^{-\alpha(T-1)}\right) = 38.4, \\ g(F_1^{(2)},1) = \int_0^1 \alpha F_s^{(2)} e^{\mu(t-s)} \mathrm{d}s + F_1^{(2)}\left(1-e^{-\alpha(T-1)}\right) = 60.2. \end{cases} \tag{5.16}$$

Fees are relative to the fund value, hence high fund values lead to higher realized and expected fees, which explains the values of $g(F_1^{(i)},1)$. The put option depends on the risk-neutral measure at time $t=1$. As discussed before, we can assume $\bar{v}$ and γ to be either constant or dependent on the state of the market. In case of constant parameters, the realized put values are equal to

$$\begin{cases} \mathrm{Put}(F_1^{(1)},G) = 211.9, \\ \mathrm{Put}(F_1^{(2)},G) = 75.9. \end{cases} \tag{5.17}$$

The fund value of the first scenario is much more likely to end up below the guarantee of $G=1000$, as it is already below this level. Consequently, the expected liabilities at $t=1$ are much higher in this scenario. The 1-year losses are now given by

$$\begin{cases} l^{(1)} = N_0 - e^{-r}N_1^{(1)} = N_0 + e^{-r}\mathrm{Put}(F_1^{(1)},G) - e^{-r}g(F_1^{(1)},1) = 83.0, \\ l^{(2)} = N_0 - e^{-r}N_1^{(2)} = N_0 + e^{-r}\mathrm{Put}(F_1^{(2)},G) - e^{-r}g(F_1^{(2)},1) = -68.6. \end{cases} \tag{5.18}$$

These examples show that increasing fund values lead to gains (or a negative loss in this case) and decreasing fund values lead to losses. Now, by generating many simulations, the distribution of l can be derived and the SCR corresponds to the 99.5% quantile of this distribution.

If we assume a VIX-dependent $\bar{v}$ and γ, then, according to the VIX Heston model discussed in chapter 3, we obtain

$$\begin{cases} (v_1^{(1)},\gamma_1^{(1)}) = (0.067, 0.703), \\ (v_1^{(2)},\gamma_1^{(2)}) = (0.059, 0.294). \end{cases} \tag{5.19}$$

This affects the expected liabilities at $t=1$ and consequently the losses become

$$\begin{cases} l^{(1)} = 63.8, \\ l^{(2)} = -84.5. \end{cases} \tag{5.20}$$

The difference between the two approaches is around 20% for these trajectories, which already indicates the impact of different assumptions regarding the risk-neutral measure at $t=1$.

Guaranteed Minimum Withdrawal Benefit

The fund dynamics of the GMWB rider are very similar to the GMAB dynamics, except for one key feature: the GMWB allows for withdrawals during the lifetime of the product. We will model the withdrawal behaviour according to the static model presented in Milevsky and Salisbury (2005). This model

assumes maximum withdrawals from the policyholder until the initial premium has been recovered. This behaviour corresponds to

$$\begin{cases} \mathrm{d}F_t = \frac{\mathrm{d}S_t}{S_t} F_t e^{-\alpha t} - w\mathrm{d}t, \qquad F_0 = G, \\ w = \frac{G}{T}. \end{cases} \tag{5.21}$$

Substituting this definition in a real-world Black-Scholes model gives

$$\mathrm{d}F_t^{\mathbb{A}} = (\mu - \alpha)F_t^{\mathbb{A}}\mathrm{d}t - w\mathrm{d}t + \sigma F_t^{\mathbb{A}}\mathrm{d}W_t^{\mathbb{A}}, \qquad F_0 = G. \tag{5.22}$$

Or equivalently, for a risk-neutral Heston model,

$$\begin{cases} \mathrm{d}F_t = (r - \alpha)F_t\mathrm{d}t - w\mathrm{d}t + \sqrt{v_t}F_t\mathrm{d}W_t^F, \quad F_0 = G, \\ \mathrm{d}v_t = \kappa(\bar{v} - v_t)\mathrm{d}t + \gamma\sqrt{v_t}\mathrm{d}W_t^v, \\ \langle \mathrm{d}W_t^S, \mathrm{d}W_t^v \rangle = \rho\mathrm{d}t. \end{cases} \tag{5.23}$$

Due to withdrawals of the policyholder, it might occur that the fund is depleted before maturity. In that case, the withdrawals are paid by the insurer instead of the fund. To this end, we will introduce two different hitting times. The first hitting time corresponds to the first time the fund value hits zero under the real-world measure

$$\tau_1 = \inf\{t : F_t^{\mathbb{A}} \leq 0, t \geq 0\}. \tag{5.24}$$

The second hitting time is similarly defined, but under the risk-neutral measure,

$$\tau_2 = \inf\{t : F_t \leq 0, t \geq 0\}, \tag{5.25}$$

where $F_t^{\mathbb{A}}$ and F_t correspond to the fund value at time t under the real-world and risk-neutral measure, respectively. Both hitting times play an important role in the evaluation of this product. There are a few possible scenarios

- When $\tau_1 \geq t$, the fund has not been depleted yet at time t and all withdrawals have been paid by the fund. Moreover, the insurer can collect its fees and the income is equal to

$$A_t = \int_0^t \alpha F_s^{\mathbb{A}} e^{\mu(t-s)}\mathrm{d}s. \tag{5.26}$$

- When $\tau_1 < t$, the fund has been depleted before time t and the withdrawals after τ_1 have been paid by the insurer, leading to the following income

$$A_t = \int_0^{\tau_1} \alpha F_s^{\mathbb{A}} e^{\mu(t-s)}\mathrm{d}s - \int_{\tau_1}^t w e^{\mu(t-s)}\mathrm{d}s. \tag{5.27}$$

Combining these possibilities gives

$$A_t = \int_0^{\tau_1 \wedge t} \alpha F_s^{\mathbb{A}} e^{\mu(t-s)}\mathrm{d}s - \int_{\tau_1 \wedge t}^t w e^{\mu(t-s)}\mathrm{d}s, \tag{5.28}$$

where $\tau_1 \wedge t$ is defined as the minimum of τ_1 and t. The liabilities can be defined in a similar way, by distinguishing the following scenarios

- When $\tau_1 < t$, the fund has been depleted before time t and all future withdrawals must be paid by the insurer. Hence, the liabilities are given by

$$L_t = \int_t^T w e^{-r(s-t)}\mathrm{d}s. \tag{5.29}$$

- When $\tau_1 \geq t$, the fund has not been depleted yet and some future liabilities are paid by the fund. Furthermore, if $\tau_2 \geq T$, all withdrawals are paid by the fund and there are no additional costs for the insurer. However, if $\tau_2 < T$, some future withdrawals are paid by the insurer. Similar to the income, we can describe this as

$$L_t = \mathbb{E}^{\mathbb{Q}_t}\left[\int_{\tau_2 \wedge T}^T w e^{-r(s-t)}\mathrm{d}s - \int_t^{\tau_2 \wedge T} \alpha F_s e^{-r(s-t)}\mathrm{d}s \middle| \mathscr{F}_t \right]. \tag{5.30}$$

We can rewrite these expressions into

$$\begin{aligned} L_t &= \int_t^T we^{-r(s-t)}\mathrm{d}s - \mathbb{1}(\tau_1 > t)\mathbb{E}^{\mathbb{Q}_t}\left[\int_t^T \mathbb{1}(\tau_2 > s)(\alpha F_s + w)e^{-r(s-t)}\mathrm{d}s \middle| \mathscr{F}_t\right] \\ &= \frac{w}{r}\left(1 - e^{-r(T-t)}\right) - \mathbb{1}(\tau_1 > t)\mathbb{E}^{\mathbb{Q}_t}\left[\int_t^T \mathbb{1}(\tau_2 > s)(\alpha F_s + w)e^{-r(s-t)}\mathrm{d}s \middle| \mathscr{F}_t\right], \end{aligned} \tag{5.31}$$

where $\mathbb{1}(\cdot)$ is defined as the indicator function which is equal to 1 when its argument is true and 0 otherwise. Substituting these definitions in (5.4) will define the SCR for a variable annuity with the Guaranteed Minimum Withdrawal Benefits rider. The calculation of the SCR requires, similar to the GMAB case, a risk-neutral valuation. However, unlike the GMAB, this conditional expectation does not have an analytical solution and one must resort to different methodologies for its evaluation. In the next section we will discuss how to efficiently and accurately determine this conditional expectation.

5.2. Evaluating the conditional expectation

In the previous section we have discussed how to determine the SCR for different variable annuity riders. Both variable annuity riders required a risk-neutral valuation at time $t = 1$, since the liabilities are by definition conditional expectations under the risk-neutral measure. Often there is no analytical expression for L_t, thus the evaluation of the conditional expectation requires an approximation. In this section we will review two different methodologies that can approximate this conditional expectation: Nested Monte Carlo simulation and the Least-Squares Monte Carlo method.

5.2.1. Nested Monte Carlo simulation

A common approach to evaluate expectations is the Monte Carlo method, which simulates many paths of the random variables in question and evaluates the outcome for each trajectory i. For example, in order to evaluate the expectation of a variable annuity with the GMWB rider at time $t = 0$, we need to evaluate

$$L_0^{(i)} = \int_0^T \mathbb{1}(\tau_2 > s)\left(\alpha F_s^{(i)} + w\right)e^{-rs}\mathrm{d}s. \tag{5.32}$$

The expectation is approximated by the mean of all M simulated trajectories

$$\mathbb{E}^{\mathbb{Q}_0}\left[\int_0^T \mathbb{1}(\tau_2 > s)(\alpha F_s + w)e^{-rs}\mathrm{d}s \middle| \mathscr{F}_0\right] \approx \frac{1}{M}\sum_{i=1}^{M} L_0^{(i)}. \tag{5.33}$$

This principle is applied in Nested Monte Carlo simulation as well, but with $t > 0$. The key difference between Monte Carlo simulations at $t = 0$ and $t > 0$ are the initial values (such as the initial value of F_t). In case of $t = 0$, the initial values can be observed and are therefore deterministic. But when $t > 0$, the initial values are unknown and become random variables that require a (real-world) simulation. Consequently, the Monte Carlo approximations as described above are nested within each simulated trajectory. We therefore have to distinguish between outer (real-world) and inner (risk-neutral) simulations. The outer simulations determine the initial values of the risk-neutral valuation at time t. The inner simulations evaluate the conditional expectation, by means of a Monte Carlo simulation. A visualization of this procedure can be found in figure 5.2.

This method is intuitively appealing, but does experience some drawbacks. The main issue of this method is the computational complexity. Suppose an accurate estimation of the SCR requires 10.000 outer simulations and each simulation requires another 10.000 inner simulations to accurately estimate the conditional expectation. This leads to a total of 10.000.000 inner simulations, which can be very time consuming. Moreover, Bauer et al. (2010) shows that this approach is biased. To this end, a more efficient and unbiased method called Least-Squares Monte Carlo was developed, which we will discuss in the next section.

5.2.2. Least-Squares Monte Carlo method

The Least-Squares Monte Carlo method was first proposed by Longstaff and Schwartz (2001) for the valuation of American options. However, Bauer et al. (2010) was the first to implement it in an SCR context, to the best of our knowledge. The main purpose of the Least-Squares Monte Carlo algorithm is

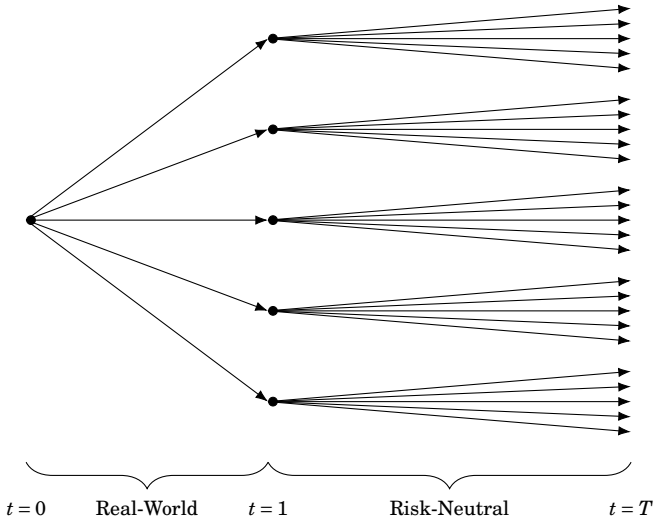

Figure 5.2: Schematic overview of Nested Monte Carlo simulation.

to reduce the number of inner simulations, possibly even to 1 path. The accuracy of the inner estimates is drastically reduced by reducing the number of inner simulations, but by combining the results of all outer simulations, the inner errors are cancelled out. Below, we will describe this process in detail.

The Least-Squares Monte Carlo algorithm can be separated into three phases. In the first phase, regression variables are simulated. This phase is very similar to the Nested Monte Carlo approach, but the number of inner simulations is drastically reduced. The conditional expectation, however, is still approximated according to the inner simulations with (5.33), possibly containing a large error. This phase is graphically presented in figure 5.3.

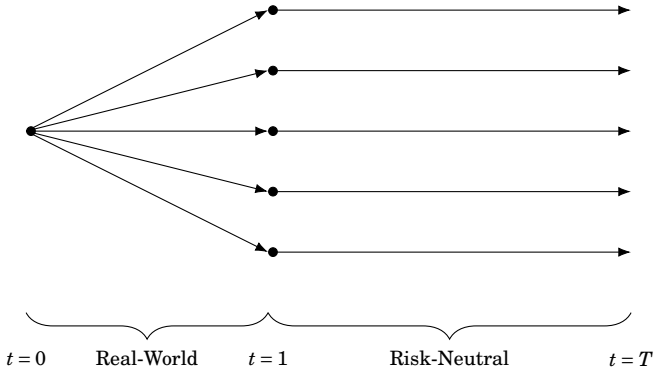

Figure 5.3: First phase of the Least-Squares Monte Carlo method.

In the second phase a regression function is constructed using the simulated conditional expectations from the first phase. The conditional expectation is approximated by a number of basis functions $\phi_n(F_t, Y_t)$,

$$\mathbb{E}^{\mathbb{Q}_t}[\cdot|\mathscr{F}_t] \approx f(F_t, Y_t) := \sum_{n=1}^{N} a_n \phi_n(F_t, Y_t), \tag{5.34}$$

where Y_t is defined as the vector of explanatory variables at time t, which are necessary to estimate the conditional expectation besides the fund value F_t. The idea is to find the set of coefficients a_n and basis functions ϕ_n, such that the conditional expectation is replicated as accurately as possible. For the basis functions, we will first consider regular polynomials

$$\phi_n(F_t, Y_t) = F_t^{n_1} \prod_{i=1}^{y} Y_t(i)^{n_{i+1}}. \tag{5.35}$$

With $Y_t(i)$, the ith component of the explanatory variables vector at time t and n_j the degree of explanatory variable j. We will compare the performance of the regular polynomial to the Legendre polynomial,

$$\begin{cases} \phi_n(F_t, Y_t) = \mathscr{L}(F_t, n_1) \prod_{i=1}^{y} \mathscr{L}(Y_t, n_{i+1}), \\ \mathscr{L}(x, n_i) = \frac{1}{2^{n_i} n_i!} \frac{\mathrm{d}^{n_i}}{\mathrm{d}x^{n_i}} [(x^2-1)^{n_i}]. \end{cases} \tag{5.36}$$

The optimal coefficients $\hat{a}$ are defined such that the sum of squared errors is minimized

$$\hat{a} = \underset{a \in \mathbb{R}^{y+1}}{\arg\min} \left\{ \sum_{i=1}^{M} \left(L_t^{(i)} - \sum_{n=1}^{N} a_n \phi_n \left(F_t^{(i)}, Y_t^{(i)} \right) \right)^2 \right\}, \tag{5.37}$$

where $L_t^{(i)}$ is the Monte Carlo approximation of trajectory i, evaluated in the first phase. Since the model is a linear combination of basis functions, we can determine these coefficients by

$$\hat{a} = (X^T X)^{-1} X^T \overline{L}_t, \tag{5.38}$$

with

$$X^T = \begin{bmatrix} \phi_1(F_t^{(1)}, Y_t^{(1)}) & \cdots & \phi_1(F_t^{(M)}, Y_t^{(M)}) \\ \vdots & & \vdots \\ \phi_N(F_t^{(1)}, Y_t^{(1)}) & \cdots & \phi_N(F_t^{(M)}, Y_t^{(M)}) \end{bmatrix}, \quad \overline{L}_t = \begin{bmatrix} L_t^{(1)} \\ \vdots \\ L_t^{(M)} \end{bmatrix}. \tag{5.39}$$

Besides finding the optimal coefficients, it is also crucial to correctly identify the optimal polynomial degree. If the polynomial degree is too low, the regression might fail to capture the complexity of the conditional expectation. On the other hand, the polynomial degree cannot be set too high due to large errors in the Monte Carlo approximations. We have illustrated this principle in figure 5.4. In this figure we have estimated a Black-Scholes put option for a range of asset values, using the Least-Squares Monte Carlo approach with different polynomial degrees. The lower polynomial degrees seem to fail to capture

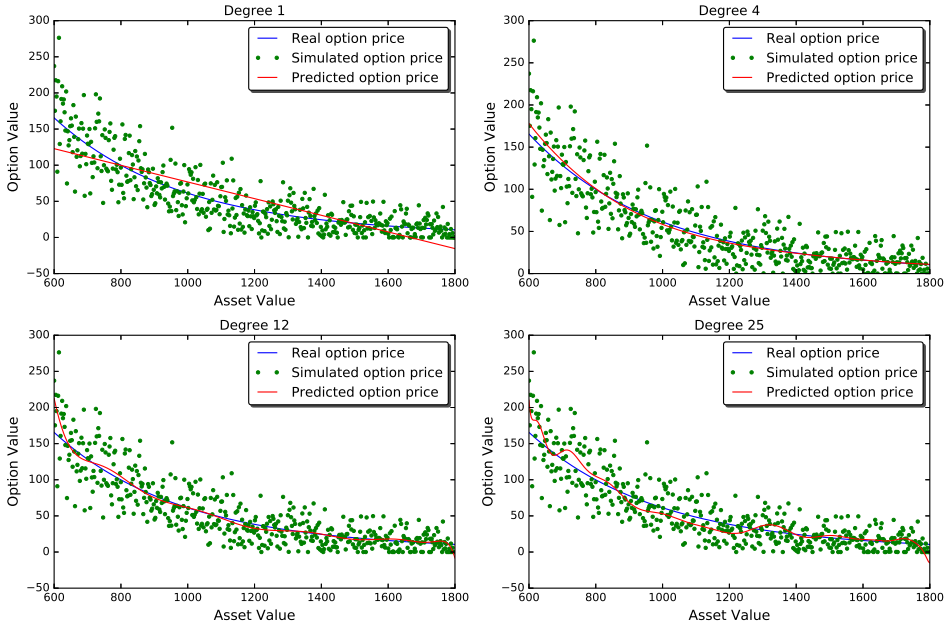

Figure 5.4: Underfitting versus overfitting Least-Squares Monte Carlo for a Black-Scholes Put option with: Strike = 1000, $T = 9$, $r = 0.04$, $\sigma = 0.2$, 500 outer points and 20 inner simulations per outer point.

the shape of the put option value and the higher polynomial degrees are too dependent on the errors of the Monte Carlo approximation. Thus, one must test different degrees in order to find the optimal balance between underfitting and overfitting, with for example a trial-and-error procedure. In case of figure 5.4 this yields

$$f(F_t) = a_1 + a_2 F_t + a_3 F_t^2 + a_4 F_t^3 + a_5 F_t^4. \tag{5.40}$$

In the final phase, another real-world simulation is executed. But, contrary to the Nested Monte Carlo approach, no inner simulations are required to approximate the conditional expectation. Instead, the conditional expectation is approximated using $f(F_t, Y_t)$, which has been defined in the second phase of the algorithm. We have visualized the final phase of the algorithm in figure 5.5.

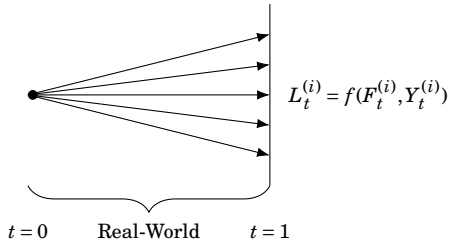

Figure 5.5: Final phase of Least-Squares Monte Carlo

Finally, it is important to optimize the resource allocation or, in other words, the ratio between outer and inner simulations. If the number of *total* inner simulations is bounded by, say, 10.000, there are many possibilities, some of which we have graphically presented in figure 5.6. Again, we have plotted the Least-Square Monte Carlo estimation of a Black-Scholes put option for a range of asset values. The regression function follows (5.40).

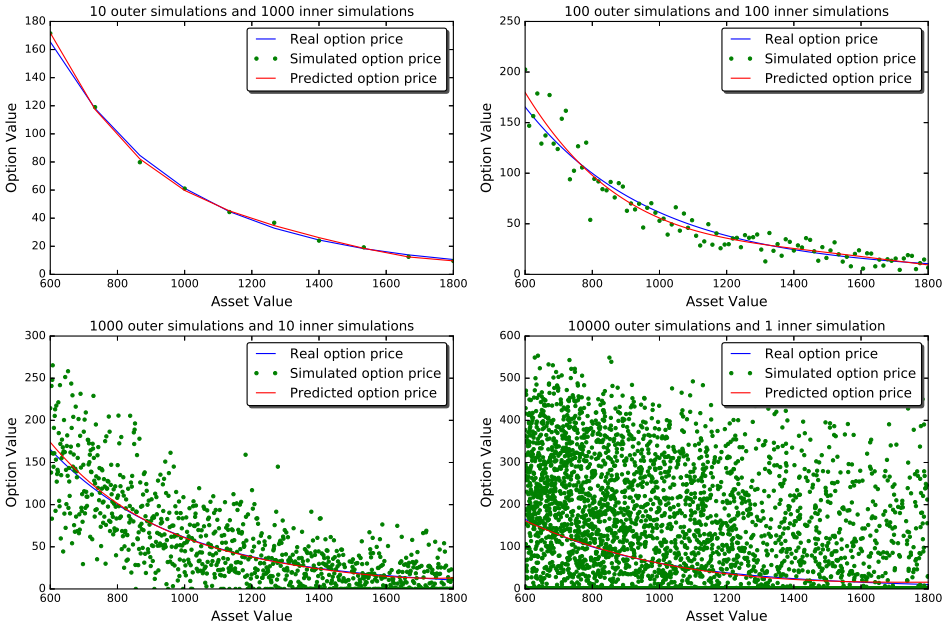

Figure 5.6: Different resource allocations Least-Square Monte Carlo, for a Black-Scholes Put option, with Strike = 1000, $T = 9$, $r = 0.04$ and $\sigma = 0.2$.

Note that,

- If we use 10.000 outer simulations and 1 inner simulation per outer simulation, there are huge errors in the inner Monte Carlo approximations. However, this set-up will also give more regression points able to nullify these errors.
- If 10 outer simulations are used then the number of inner simulations can be increased to 1.000 per outer simulation. This will give fewer regression points, but each point will have a more accurate inner Monte Carlo estimate.

In section 5.2.4 we will discuss some theoretical properties of the Least-Squares Monte Carlo algorithm that can be used to identify the theoretical optimal set-up. Moreover, in appendix F, a numerical test will be implemented to test these theoretical properties.

5.2.3. Alternative approaches

Above, we have discussed two possible methodologies that evaluate conditional expectations at $t > 0$: Nested Monte Carlo simulation and the Least-Squares Monte Carlo method. For completeness, we will present some alternative approaches that have been proposed to determine the SCR.

Replicating portfolio

The guarantees of variable annuities are often very similar to options, as we have seen with the GMAB rider and the European put option. To this end, Schrager (2008) suggests replicating portfolios to determine the liabilities of insurance products. The replicating portfolio consists of standard financial assets and it should match the value of liabilities as good as possible[2]. A real-world Monte Carlo simulation is performed and, for each trajectory, the value of the replicating portfolio is a predictor for $L_t^{(i)}$.

The main advantages of this approach are its efficiency and interpretability. Unfortunately, many features cannot be captured by standard financial assets, such as mortality rate and policyholder behaviour. It is therefore impossible to take all risk-drivers into account, which greatly reduces the accuracy for complex products.

Curve fitting

In Algorithmics (2011) an algorithm very similar to the Least-Squares Monte Carlo approach is discussed: Curve Fitting. In the first phase of this approach, only a few outer scenarios are generated and each outer scenario performs an accurate Monte Carlo simulation, with many inner scenarios. In the second phase, similar to Least-Squares Monte Carlo method, a regression function is fitted using these scenarios and in the final phase this function is used to determine the expected liabilities at $t = 1$. This procedure is very similar to the top left plot in figure 5.6.

Just like the Least-Squares Monte Carlo algorithm, Curve Fitting is able to incorporate many risk-drivers. However, complex products with many risk-drivers still require lots of outer scenarios and this approach will behave like Nested Monte Carlo simulation in terms of efficiency. Moreover, this algorithm is heavily dependent on the few outer scenarios generated in the first phase and it requires an expert opinion to select appropriate scenarios. The Least-Squares Monte Carlo approach does not suffer from these drawbacks, since it generates many outer scenarios.

Stochastic Grid Bundling Method

A modification of the least-squares method developed by Longstaff and Schwartz is proposed in Jain and Oosterlee (2015). This approach is called the Stochastic Grid Bundling Method (SGBM) and it was first implemented in an SCR environment by Feng et al. (2017). There are two key differences between the Stochastic Grid Bundling Method and Least-Squares Monte Carlo. First of all, the Stochastic Grid Bundling Method maps the domain into disjoint bundles and the regression is performed on each individual bundle instead of the entire domain. The second difference lies in the regression function. The regression of the Least-Squares Monte Carlo approach is based on the expected discounted liabilities, but the regression of the Stochastic Grid Bundling Method aims at replicating the option value and relies on the moments of the characteristic function. These adjustments should improve the accuracy compared to the Least-Squares Monte Carlo algorithm, at the cost of a (slightly) more demanding procedure.

In this chapter we are evaluating the impact of time-dependent risk-neutral parameters and, apart from prediction accuracy, we do not expect different outcomes between SGBM and Least-Squares Monte Carlo. To this end, we will only implement the Least-Squares Monte Carlo approach, as it is easier to interpret and more frequently used in the insurance business.

Sampling methods

The first phase of the Least-Squares Monte Carlo algorithm uses real-world simulations as initial values for the inner simulations. However, Cathcart (2012) shows that, even though this approach leads to an accurate overall performance, it is slightly less accurate in the tails of the distribution. He proposes sampling methods instead of real-world simulations. Compare, for example, sample points from a real-world Heston simulation to the quasi-random Sobol sampling method (see Sobol and Levitan (1999)),

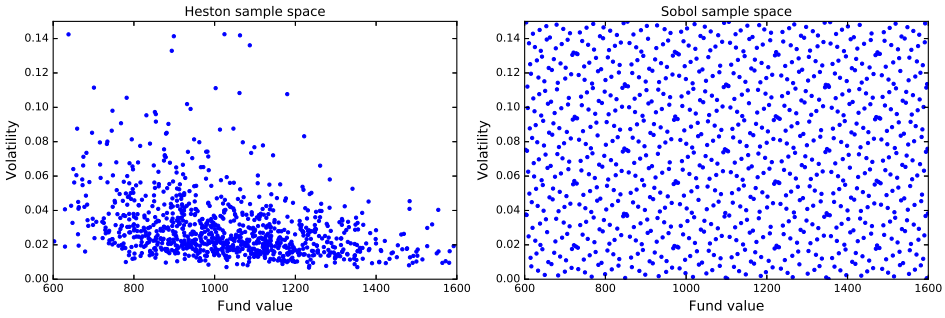

Figure 5.7: Heston versus Sobol sample points.

which is shown in figure 5.7. The sample points of the Heston model are densely clustered, which can be beneficial because it assigns more weight and thus more accuracy to realizations that are more likely to occur. On the other hand, fewer points are assigned to the tails of the distribution, making it more difficult to give an accurate tail prediction. Many important quantities in risk-management rely on the tails, remember for example the SCR which depends on the 99.5% quantile of the loss distribution. It can therefore be advantageous to redistribute the sampling points. The Sobol sampling method, for example, maximizes the distance between points across the entire domain. This way, more points are assigned to the tail of the distribution and its predictions should consequently be more accurate. It will, however, result in less accurate estimates over the entire distribution, because the accuracy in other parts of the domain is reduced. So, depending on the region of interest, one can decide upon different sampling methods.

5.2.4. Error analysis

In this section we will discuss some theoretical properties of the Least-Squares Monte Carlo algorithm. There are many degrees of freedom in this algorithm, as discussed in section 5.2.2. For example, the type of basis function, sampling methods and resource allocation all influence the stability and accuracy of this approach. To this end, we will discuss the theoretical properties of the different set-ups in this section.

In general, we can think of the Monte Carlo observations as a combination of the optimal regression function and an error term

$$\overline{L}_t = Xa^* + \epsilon, \tag{5.41}$$

where $\overline{L}_t$ and X are defined as in (5.39) and the optimal regression coefficients and vector of error terms are denoted by a^* and ϵ, respectively. The optimal regression coefficients are defined as

$$a^* = \underset{a\in\mathbb{R}^{y+1}}{\arg\min}\left\{\sum_{i=1}^{M}\left(L^*\left(F_t^{(i)},Y_t^{(i)}\right)-\sum_{n=1}^{N}a_n^*\phi_n\left(F_t^{(i)},Y_t^{(i)}\right)\right)^2\right\}, \tag{5.42}$$

where $L^*\left(F_t^{(i)},Y_t^{(i)}\right)$ are defined as the true expected liabilities for explanatory variables $\left(F_t^{(i)},Y_t^{(i)}\right)$. Moreover, ϵ is assumed to have independent identically distributed terms with

$$\mathbb{E}[\epsilon] = \overline{0}, \qquad \mathbb{E}[\epsilon\epsilon^T] = \sigma_\epsilon^2 I. \tag{5.43}$$

The objective of the algorithm is to minimize the distance between the optimal regression coefficients a^* and the observed regression coefficients $\hat{a}$ as defined in (5.38),

$$\begin{aligned}\hat{a} &= (X^TX)^{-1}X^T\overline{L}_t\\ &= (X^TX)^{-1}X^T\left(Xa^*+\epsilon\right)\\ &= a^* + (X^TX)^{-1}X^T\epsilon.\end{aligned} \tag{5.44}$$

[2]These portfolios are very similar to the hedging portfolio discussed in the previous chapter.

Given X (which is in our case equal to the basis functions corresponding to set of explanatory variables at $t = 1$), this $\hat{a}$ has a mean equal to a^*, as ϵ has mean 0. Moreover, the variance-covariance matrix is equal to

$$\begin{aligned}
\text{Var}(\hat{a}) &= \mathbb{E}[\hat{a}\hat{a}^T] \\
&= \mathbb{E}\left[a^* + (X^TX)^{-1}X^T\epsilon\left((X^TX)^{-1}X^T\epsilon\right)^T\right] \\
&= \mathbb{E}\left[(X^TX)^{-1}X^T\epsilon\epsilon^TX(X^TX)^{-1}\right] \\
&= (X^TX)^{-1}X^T\mathbb{E}\left[\epsilon\epsilon^T\right]X(X^TX)^{-1} \\
&= (X^TX)^{-1}X^T\sigma_\epsilon^2IX(X^TX)^{-1} \\
&= \sigma_\epsilon^2(X^TX)^{-1}X^TX(X^TX)^{-1} \\
&= \sigma_\epsilon^2(X^TX)^{-1}. \qquad (5.45)
\end{aligned}$$

Hence, the variance of the estimate depends on the variance of the (inner) Monte Carlo estimates σ_ϵ^2 and the matrix $(X^TX)^{-1}$, which depends on the sampling method and type of basis function. In Cathcart (2012) it is argued that, ideally, the covariance between the regression coefficients equal 0, as this reduces the variance of the final estimates. To this end, one might consider orthogonal polynomials, such as the Legendre polynomial discussed in section 5.2.2. A polynomial sequence, $\phi_n(x)$ with $n \in \{0,1,\dots\}$ is considered orthogonal on interval $[a,b]$ when

$$\int_a^b \phi_n(x)\phi_m(x)\mathrm{d}x = \begin{cases} 0 & \text{if } m \neq n, \\ c_n & \text{if } m = n, \end{cases} \qquad (5.46)$$

with constant c_n. Hence, if the explanatory variables are evenly spaced across $[a,b]$ and the number of outer samples converges to infinity, then $(X^TX)^{-1}$ will converge to a diagonal matrix. The Legendre polynomial is orthogonal on the interval $[-1,1]$. By transforming the features

$$x \mapsto 2\frac{x - x_{\min}}{x_{\max} - x_{\min}} - 1, \qquad (5.47)$$

the Legendre basis functions should estimate an $\hat{a}$ with lower covariance between its components than the regular polynomial, as long as the explanatory variables are evenly spaced across the domain. This also implies that Sobol sampling should result in more stable predictions than real-world simulation sampling, as it is more evenly spaced across the domain, which can be seen in figure 5.7.

Finally, there is the trade-off between inner and outer samples. Increasing the number of inner samples results in a lower inner Monte Carlo variance σ_ϵ^2, but increasing the number of outer samples results in lower (co)variance between the regression coefficients through $(X^TX)^{-1}$. In Broadie et al. (2015) a theoretical analysis of this problem is presented. In this analysis they show that the asymptotic bias is unaffected by the resource allocation, but the asymptotic variance is minimized when the maximum amount of outer samples and only a single inner sample is used. Hence, in terms of the mean-squared error (which can be decomposed into a bias and variance term), it is theoretically optimal to maximize the amount of outer samples.

5.3. Impact analysis

In the previous section we have presented many algorithms capable of calculating the SCR. For a number of reasons discussed in section 5.2.3, we will implement the Least-Squares Monte Carlo method. The main purpose of this test is to show the impact on the SCR of different assumptions regarding the risk-neutral measure at $t = 1$. We will compare a risk-neutral measure with constant parameters, which we will refer to as the original risk-neutral measure, to a risk-neutral measure with time-dependent parameters as defined in chapter 3, which will be referred to as the time-dependent risk-neutral measure. In this section, the parameters of the time-dependent risk-neutral measure will follow the VIX Heston model. The test will be performed on two different variable annuities: Guaranteed Minimum Accumulation Benefit and Guaranteed Minimum Withdrawal Benefit.

In this section, we will compare three scenarios. Each scenario will have different initial values and consequently the initial expected liabilities, L_0, will differ as well. In this part we will assume a no-arbitrage fee, i.e. a fee for which L_0 is equal to zero. The general contract details of the variable annuities can be found in table 5.1 and the initial values of the different scenarios accompanied with the fair premiums are presented in table 5.2.

Parameter	Value
F_0	1000
G	1000
w	7%
T_{GMAB}	10
T_{GMWB}	14.28
r	0.04

Table 5.1: General contract details of the GMAB and GMWB riders.

Parameter	Scenario 1	Scenario 2	Scenario 3
v_0	0.04	0.01	0.27
$\bar{v}_0$	0.08	0.025	0.24
γ_0	0.55	0.05	1.4
α_{GMAB}	0.0174	0.0057	0.0345
α_{GMWB}	0.0137	0.0041	0.0284

Table 5.2: Initial values and fair premiums of the GMAB and GMWB riders.

5.3.1. Implementation

The implementation of a risk-neutral measure with constant versus time-dependent parameters is very similar in the Least-Squares algorithm, but does differ on some key aspects, which we will discuss below.

First phase

The first phase starts with generating sample points of the explanatory variables at $t = 1$. These sample points can either be generated by a real-world simulation or a sampling method, as we have discussed in section 5.2.2. We will briefly review the implementation of both approaches.

In our test we will assume the fund value to follow a Geometric Brownian Motion with initial value $F_0 = 1000$, so in case of the GMAB and GMWB riders, the fund value will follow (5.7) and (5.22), respectively. The VIX index will be modelled simultaneously, following a mean-reverting path

$$\begin{cases} \mathrm{d}\text{vix}_t &= \kappa_{\text{vix}}(\text{vix}_{\text{Mean}} - \text{vix}_t) + \gamma_{\text{vix}}\text{vix}_t^{\lambda_{\text{vix}}}\left(\rho_{\text{vix}}\mathrm{d}W_t^S + \sqrt{1-\rho_{\text{vix}}^2}\mathrm{d}W_t^{\text{vix}}\right), \\ \text{VIX} &= 100 \cdot \text{vix}. \end{cases} \tag{5.48}$$

The process parameters are estimated with the Generalized Method of Moments (see Hansen (1982)),

$$\sigma = 0.21, \quad \mu = 0.05, \quad \kappa_{\text{vix}} = 4.964, \quad \text{vix}_{\text{Mean}} = 0.207, \quad \gamma_{\text{vix}} = 1.859, \quad \lambda_{\text{vix}} = 1.271. \tag{5.49}$$

Moreover, we set $\rho_{\text{vix}} = -0.75$, which is in line with observations of the risk-neutral market. The stochastic differential equations will be discretized by the Milstein scheme. For each simulated trajectory i, the parameters of the original risk-neutral measure at time $t = 1$ are defined as

$$\begin{cases} F_1^{(i),\mathbb{Q}} = F_1^{(i),\mathbb{A}}, \\ v_1^{(i)} = \left(0.0140 + 0.0090 \cdot \text{VIX}_1^{(i)}\right)^2, \\ \bar{v}_1^{(i)} = \bar{v}_0, \\ \gamma_1^{(i)} = \gamma_0. \end{cases} \tag{5.50}$$

The time-dependent risk-neutral measure will follow the VIX Heston model derived in chapter 3,

$$\begin{cases} F_1^{(i),\mathbb{Q}} = F_1^{(i),\mathbb{A}}, \\ v_1^{(i)} = \left(0.0140 + 0.0090 \cdot \text{VIX}_1^{(i)}\right)^2, \\ \bar{v}_1^{(i)} = \left(0.0957 + 0.0087 \cdot \text{VIX}_{\text{filter}_1}^{(i)}\right)^2, \\ \gamma_1^{(i)} = 9.6479 \cdot 10^{-5} + 0.0270 \cdot \text{VIX}^{(i)}. \end{cases} \tag{5.51}$$

The VIX filter component is as defined in section 3.2. By simulating many paths, one obtains a sample set of explanatory variables at $t = 1$, consistent with the real-world measure.

In case of Sobol sampling, we do not need to generate real-world trajectories. Instead, we generate sample points using a predefined algorithm, which is, in case of Sobol sampling, available in many software packages. We will, however, slightly adjust the algorithm. Due to correlations, it is unlikely to simultaneously obtain a high VIX index and a high fund value. Therefore, we will split the domain into several bundles and assign points to those more likely to occur. In doing so, we ensure that the grid still contains over 99.9% of the real-world simulations, but the accuracy is increased in the regions of interest, for example the tail of the distribution. In case of the GMAB and GMWB riders, the sample points are distributed as in figure 5.8.

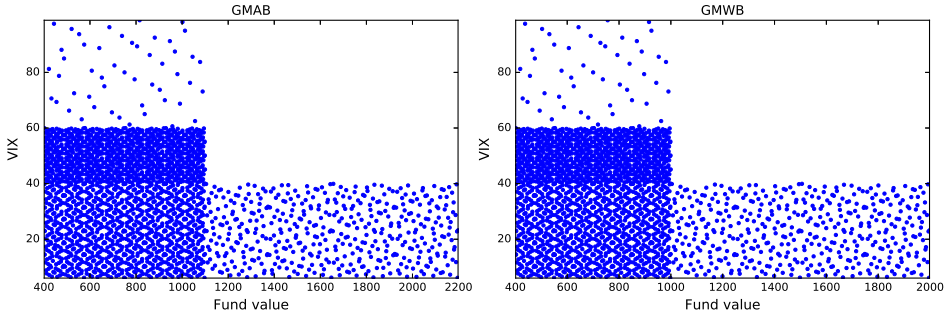

Figure 5.8: Adjusted Sobol sampling for VIX index and fund value.

The time-dependent risk-neutral measure also requires the VIX trend component, besides the fund value and VIX index. The trend is highly correlated to the VIX index and it is therefore disadvantageous to generate a 3-dimensional Sobol sample space, as we would lose the correlation structure. To this end, we will first generate a 2-dimensional Sobol sample space, similar to figure 5.8. The trend component will then be sampled according to the VIX index. Based on simulated data we have defined upper and lower boundaries, such that 99.9% of the trend component is captured within the intervals between those boundaries,

$$l_{\text{filter}} = 3 + 0.4 \cdot \text{VIX}, \qquad u_{\text{filter}} = 16 + \text{VIX}. \tag{5.52}$$

Finally, each sample point in figure 5.8 will be assigned a 1-dimensional Sobol sample space between the upper and lower boundary of (5.52). This way, we obtain a 3-dimensional sample space where the VIX is realistically correlated to its trend component. A visualization of the VIX index sampled according to figure 5.8 and its accompanied $\text{VIX}_{\text{filter}}$ sample points is given in figure 5.9. Again, the parameters of the risk-neutral measure at time $t = 1$ are given by (5.50) and (5.51).

The risk-neutral simulation at $t = 1$ is similar for both risk-neutral measures, apart from the different parameters, as they both follow the Heston model (see (5.8) and (5.23)). The simulation scheme for the Heston model is described in appendix E.

During this phase, Simpson's rule will be used to approximate the integral expressions presented in section 5.1.2,

$$\int_{t_0}^{t_1} h(t,\cdot)\mathrm{d}t \approx \frac{\Delta t}{3} \sum_{j=1}^{N/2} \left[h(t_{2j-2},\cdot) + 4h(t_{2j-1},\cdot) + h(t_{2j},\cdot)\right], \tag{5.53}$$

with $t_i = t_0 + i\Delta t$ and $\Delta_t = \frac{t_1 - t_0}{N}$.

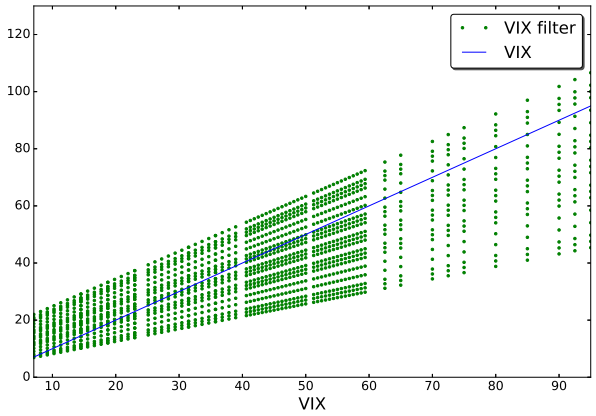

Figure 5.9: VIX and the accompanied $\text{VIX}_{\text{filter}}$ sample points.

Second phase
In the second phase of the algorithm a regression is performed in the form of basis functions and their weights. The basis functions will depend on the explanatory variables at $t = 1$ and these will differ for the two risk-neutral measures. The original risk-neutral measure only requires two explanatory variables, the fund value and volatility at time $t = 1$. All other parameters do not change and will therefore not be able to explain differences between outcomes. The time-dependent measure will require two additional explanatory variables, $\bar{v}$ and γ. These parameters are allowed to change in the real-world simulation and might therefore be able to explain different outcomes of the risk-neutral valuation at time $t = 1$. For example, in case of a regular polynomial basis function, the basis functions will have the following shape

$$\begin{cases} \phi_{n,\text{original}} = F_1^{n_1} v_1^{n_2}, \\ \phi_{n,\text{time-dependent}} = F_1^{n_1} v_1^{n_2} \bar{v}_1^{n_3} \gamma_1^{n_4}. \end{cases} \tag{5.54}$$

Final phase
The final phase starts with a real-world simulation identical to the one described in the first phase. Hereafter, the loss function can be evaluated,

$$\begin{aligned} l^{(i)} &= N_0 - e^{-r} N_1^{(i)} \\ &= A_0 - L_0 - e^{-r}(A_1^{(i)} - L_1^{(i)}) \\ &= e^{-r} L_1^{(i)} - e^{-r} A_1^{(i)} - L_0, \end{aligned} \tag{5.55}$$

where L_t and A_t are defined as in section 5.1.2. Moreover, $L_1^{(i)}$ is obtained from the regression function constructed in the second phase, $A_1^{(i)}$ is evaluated by the Simpson's rule applied to the real-world trajectory and L_0 is estimated with a risk-neutral Monte Carlo simulation at $t = 0$. From these simulated losses we will be able to derive its distribution and with it, the 99.5% quantile necessary to determine the SCR.

5.3.2. Set-up Least-Squares Monte Carlo method

As described in section 5.2.2, the Least-Squares Monte Carlo method has some degrees of freedom, such as the type of basis function, resource allocation and sampling method. In this chapter, we wish to assess the impact on the SCR as accurately as possible. To this end, we wish to optimize the accuracy of the Least-Squares Monte Carlo method. Appendix F describes the analyses we performed to determine the optimal set-up for this test. A brief overview of this set-up can be found in table 5.3. Note that these findings are in line with the analyses of section 5.2.4, except for the type of basis function which did not appear to have a significant impact on the accuracy.

5.3.3. Guaranteed Minimum Accumulation benefit

After testing the optimal set-up of the Least-Squares Monte Carlo algorithm, we are able to assess the impact of assuming time-dependent risk-neutral parameters at $t = 1$. For 100.000 real-world simulations,

Feature	Optimal set-up
Basis function	Regular polynomial
Polynomial degree	4
Sampling method	Sobol sampling (see figure 5.8)
Outer simulations original $\mathbb{Q}$	200.000
Inner simulations original $\mathbb{Q}$	1
Outer simulations time-dependent $\mathbb{Q}$	500.000
Inner simulations time-dependent $\mathbb{Q}$	1

Table 5.3: Optimal set-up LSMC algorithm.

the loss function as defined in (5.55) is evaluated and from these trajectories we are able to construct a probability density function. This process is repeated for three different scenarios, whose specifics can be found in table 5.1. The scenarios represent three different initial markets, corresponding to a low, average and high implied volatility surface level. The premium fees α are defined such that the expected liabilities are 0 at $t = 0$. Hence, the loss function of all three scenarios only depends on the realized income and expected liabilities at $t = 1$. In figure 5.10 we have graphically represented the impact on the probability density function of the loss distribution for the different scenarios. Moreover, the Solvency Capital Requirements associated with these distributions can be found in table 5.4.

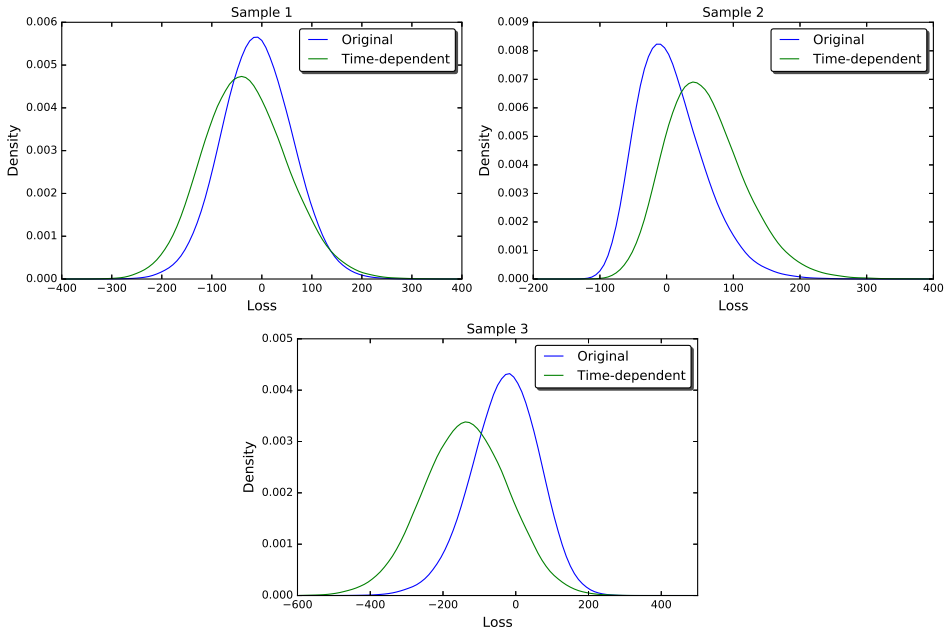

Figure 5.10: Probability density functions of the 1-year loss distribution for a variable annuity with the GMAB rider, under the original and time-dependent risk-neutral measure. Scenario 1 = average initial volatility, Scenario 2 = low initial volatility, Scenario 3 = high initial volatility.

	Original	Time-dependent
Scenario 1	170.1	195.5
Scenario 2	166.9	237.5
Scenario 3	178.8	150.8

Table 5.4: Solvency Capital Requirement of the scenarios for the original and time-dependent risk-neutral measure.

The impact on the probability density functions and the Solvency Capital Requirements is substantial

and we wish to highlight a few noteworthy features.

The loss distribution under the original risk-neutral measure appears to be centred around 0, independent of the initial conditions. The 1-year loss is defined as the difference between the policy value at $t = 0$ and $t = 1$. On average, the policy value will not change significantly if the risk-neutral parameters stay the same. Therefore, the loss distribution must be centred around 0, as long as the initial risk-neutral parameters do not change. The loss distribution under the time-dependent risk-neutral measure, on the other hand, heavily depends on the risk-neutral parameters at $t = 0$. Consider scenario 2 for example. Initially, the volatility, $\bar{v}$ and γ are relatively low, resulting in low initial expected liabilities. However, according to the mean-reverting VIX index, these parameters are more likely to increase over time and with them, the expected liabilities. Consequently, the 1-year losses are much higher compared to those under the original risk-neutral measure, which still assumes the relatively low initial parameters at $t = 1$. In figure 5.10, this effect is clearly visible where the loss function is shifted to the right. Conversely, the 1-year losses under the time-dependent risk-neutral measure in scenario 3 are much lower, as the expected liabilities are more likely to decrease. In conclusion, when the initial volatility is low (high), we can expect a higher (lower) SCR under the time-dependent risk-neutral measure.

Besides the shifted mean, the loss distribution under the time-dependent risk-neutral measure also tends to have heavier tails, which is especially visible in scenario 1. This is caused by the fact that $\bar{v}$ and γ depend on the state of the market, which results in more extreme losses (or gains). If, for example, the market crashes, $\bar{v}$ and γ are likely to increase. This will generate even higher expected liabilities, resulting in even higher losses. However, if the market flourishes, $\bar{v}$ and γ tend to be much lower, leading to lower expected liabilities and lower losses (or higher gains). This feature is present in all scenarios of figure 5.10, but is best visible in scenario 1, where the probability of an extreme loss as well as the probability of an extreme gain are higher under the time-dependent risk-neutral measure. Consequently, the SCRs under the two risk-neutral measures are not necessarily equal, not even when the initial conditions are equal to the average market conditions (such as in scenario 1).

Historical analysis

Finally, in order to give a more broad overview of the impact, we will determine the SCR of a variable annuity with the GMAB rider for multiple points in time. The contract details presented in table 5.1 will remain unchanged, but the initial parameters will depend on historical data. For computational purposes, α is assumed to be constant and equal to 0.01. In this test we will compare four different risk-neutral measures

1. A time-dependent risk-neutral measure where all parameters depend on the simulated state of the market.
2. A risk-neutral measure where F_1 and v_1 depend on the simulated market and the risk-neutral parameters are equal to the parameters as observed on $t = 0$. This measure is equivalent to the original risk-neutral measure we have previously defined.
3. A risk-neutral measure where F_1 and v_1 depend on the simulated market and the risk-neutral parameters are equal to the parameters as observed on $t = 1$. We will refer to this measure as the future risk-neutral measure.
4. A risk-neutral measure where F_1 and v_1 depend on the simulated market and the risk-neutral parameters are equal to the *realized* regression model predictions at $t = 1$ of figure 3.5. This measure is different from the time-dependent risk-neutral measure, as it depends on the realized state VIX, instead of the simulated VIX. We will refer to this measure as the future VIX risk-neutral measure.

So far, we have applied the first two risk-neutral measures in our analyses. The latter two can only be applied on historical data (otherwise, the observed parameters at $t = 1$ are undefined) and are merely added for explanatory purposes. Ideally, the SCR under the future and the future VIX risk-neutral measures is equal. The difference between these measures is caused by prediction errors of the regression model. Hence, the difference between the SCRs under these measures will be an indication of the accuracy of the regression models.

In figure 5.11 the results under the different risk-neutral measures are displayed. The difference between the original and time-dependent risk-neutral measure is also summarized in table 5.5.

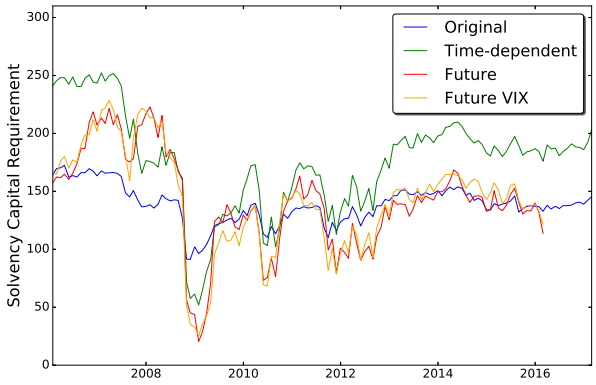

Figure 5.11: SCR over time under the different risk-neutral measures.

	Value	%
Mean absolute difference	41.1	28.7%
Maximum absolute difference	85.3	52.0%

Table 5.5: Difference in SCR between original and time-dependent risk-neutral measure.

The impact on the SCR is significant, with a maximum absolute difference of over 50%. Moreover, the difference appears to be structural over time, with a mean absolute difference of almost 30%. The results in figure 5.11 also contain some stylized facts we have already seen in figure 5.10 and we will discuss them by distinguishing different time periods:

- 2005-2007: The volatility in these years was relatively low and this translates to a somewhat higher SCR under the original measure and an even higher SCR under the time-dependent measure, similar to scenario 2. The SCR under the future measure is rising, due to higher expected liabilities at $t = 1$, indicating that more volatile times are coming.
- Early 2008: The market has not crashed yet, but volatility is starting to increase, resulting in a smaller difference between the time-dependent and original measures, which is comparable to scenario 1. The future measure, however, takes the fact that the market will crash into account. Hence, the SCR is the highest under the future measure.
- Late 2008-2012: During these years, several spikes occurred in the implied volatility surface, which will increase the initial liabilities and therefore the expected loss will decrease, analogously to scenario 3. This results in lower SCRs during this period. Moreover, note that the SCRs under the future risk-neutral measure are lowest during these highly volatile periods, since this measure depends on the realized market at $t = 1$, which has returned to its less volatile state. Consequently, the expected liabilities at $t = 1$ and the SCR are the lowest under the future risk-neutral measure.
- 2013-2017: This period is comparable to 2005-2007, apart from the fact that the volatility is approximately constant throughout these years. This translates into an almost equal SCR prediction under the original and future risk-neutral measures.

Moreover, the difference between the future and future VIX measures is small. This means that the expected liabilities at $t = 1$ are almost equal under both measures, indicating the accuracy of the regression model, at least for the realized states of the market. Under the assumption that the simulated markets behave similar to historical observations, this means that expected liabilities at $t = 1$ under the time-dependent measure will be in line with the simulated states of the market.

Finally, we wish to highlight an interesting feature. In figure 5.12 we have plotted the difference between the SCR under the original and time-dependent risk-neutral measures and the development of $\bar{v}$ over time. The impact of assuming time-dependent risk-neutral parameters at $t = 1$ appears to be proportional to $\bar{v}$. This indicates an almost linear relationship between the initial parameters and the impact on the SCR. The only difference between the two approaches is the definition of the risk-neutral measure at time $t = 1$, resulting in different expected liabilities at $t = 1$. All other components are equal in both approaches. Hence, the difference between the SCRs is exactly equal to the discounted difference

between the 99.5% quantiles of the expected liabilities at $t = 1$, which can be found in the right image of figure 5.12.

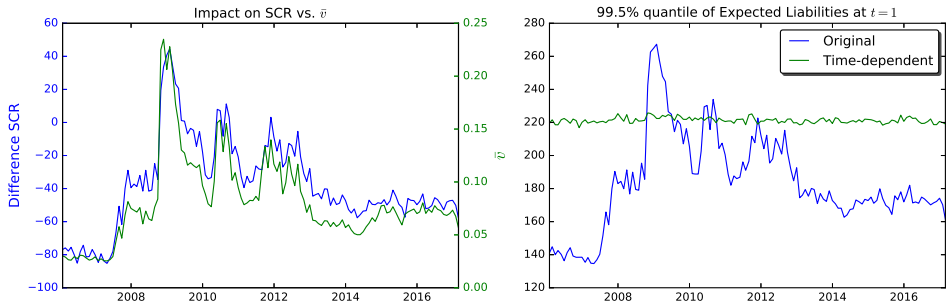

Figure 5.12: Difference between the SCR under the original and time-dependent risk-neutral measure versus $\bar{v}$.

At $t = 1$, the original risk-neutral measure determines the expected liabilities with the initial $\bar{v}$ and γ. Consequently, the expected liabilities at $t = 1$ heavily depend on the initial parameters, which is clearly visible in figure 5.12. On the other hand, the expected liabilities under the time-dependent measure are approximately independent of the initial parameters, because $\bar{v}$ and γ will converge to their long term mean, according to the mean-reverting VIX-index. Therefore, the 99.5% quantile of the expected liabilities' distribution at $t = 1$ is almost constant over time, under the time-dependent risk-neutral measure. The difference between these quantiles, or equivalently the impact on the SCR, will therefore be proportional to the initial parameters.

5.3.4. Guaranteed Minimal Withdrawal Benefit

Similar to the GMAB rider, we will assess the impact on the SCR of the different assumptions regarding the risk-neutral measure at $t = 1$. Again, we will distinguish three cases that correspond to different initial states of the market. The general contract details of this variable annuity and the three different initial market conditions can be found in tables 5.1 and 5.2, respectively. Applying the Least-Squares Monte Carlo approach to the different scenarios resulted in the distributions as shown in figure 5.13. Table 5.6 contains the corresponding Solvency Capital Requirements.

	Original	Time-dependent
Scenario 1	189.0	196.2
Scenario 2	192.1	213.3
Scenario 3	190.9	178.8

Table 5.6: Solvency Capital Requirement of the scenarios for the original and time-dependent risk-neutral measure.

The results of these tests are very similar to the GMAB case; the distributions under the original risk-neutral measure are centred around 0 and the time-dependent distributions are shifted, depending on the initial conditions. Moreover, the tails of the time-dependent distributions are heavier, due to dependence between the market and the risk-neutral parameters, as explained in section 5.3.3. The effects are, however, less pronounced than the effects on a variable annuity with the GMAB rider. Possibly, because some liabilities have already been deducted from the fund. Nevertheless, these results show that recalibration has a structural effect on different insurance products, when determining the SCR.

Historical analysis

We finalize our analysis by evaluating the SCR over time, based on historical data. The general contract details in table 5.1 are the same for each point in time, but the initial market conditions will change, depending on the historical implied volatility surfaces. Again, we will assume a constant fee α equal to 0.01 for each point in time. The results are summarized in table 5.7 and figures 5.14 and 5.15.

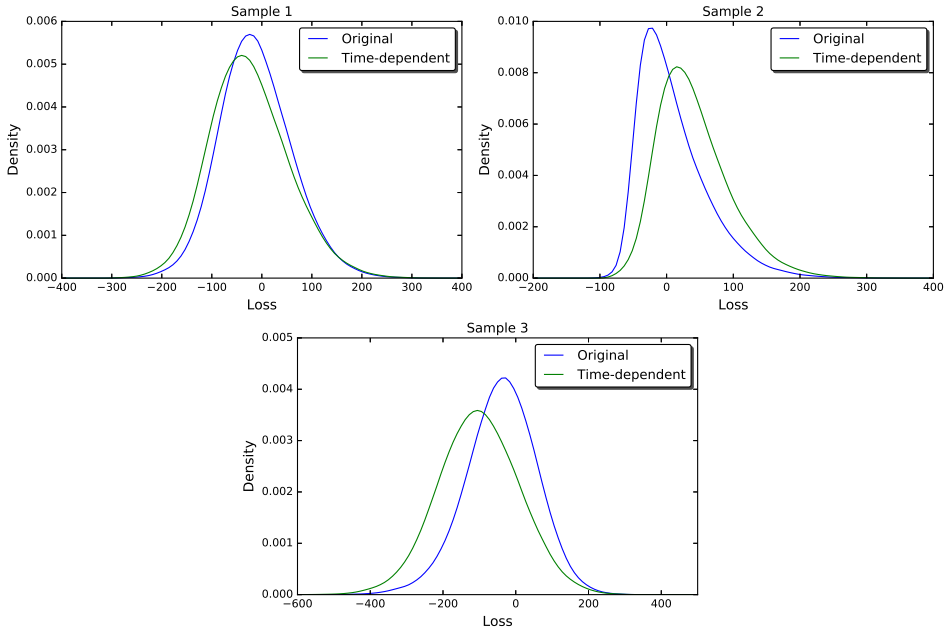

Figure 5.13: Probability density functions of the 1-year loss distribution for a variable annuity with the GMWB rider, under the original and time-dependent risk-neutral measure. Scenario 1 = average initial volatility, Scenario 2 = low initial volatility, Scenario 3 = high initial volatility.

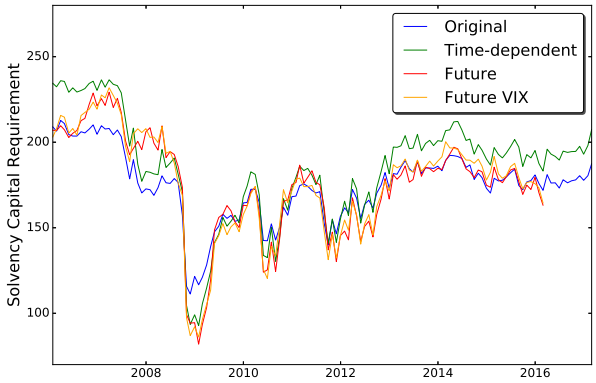

Figure 5.14: SCR and expected liabilities under the original and time-dependent risk-neutral measure.

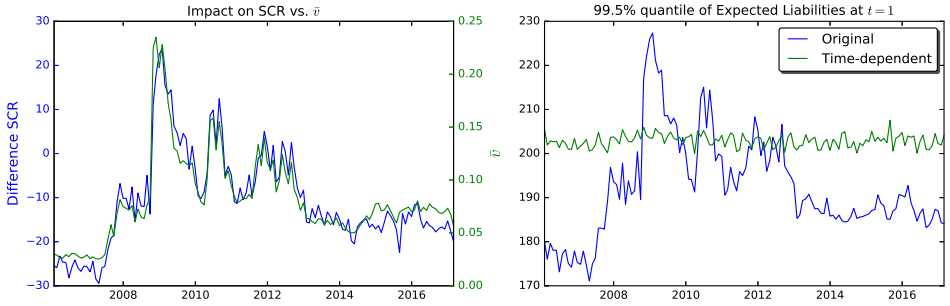

Figure 5.15: Difference between the SCR under the original and time-dependent risk-neutral measure versus $\bar{v}$.

The development of the Solvency Capital Requirement of the GMWB rider is comparable to the GMAB rider, but the effect is less pronounced, which is in line with our previous observations. Again, the 99.5% quantiles of the expected liabilities at $t = 1$ are proportional to the risk-neutral parameters in case of the original risk-neutral measure and almost constant in case of the time-dependent risk-neutral measure.

	Value	%
Mean absolute difference	13.5	7.6%
Maximum absolute difference	29.5	20.5%

Table 5.7: Difference in SCR between original and time-dependent risk-neutral measure.

Hence, the impact on the SCR is dependent on $\bar{v}$, similar to the GMAB rider.

6

Conclusion and outlook

6.1. Summary and conclusion

Throughout this thesis we have investigated the predictability of the Heston parameters and the effect of implementing the recalibration process in risk-management applications. After introducing the theoretical assumptions of the risk-neutral and real-world measure, we explained the importance of predicting the Heston parameters. The advantages are twofold. First of all, it simplifies the calibration procedure. Secondly, the regression models provide a way to implement the recalibration process in real-world simulations. Currently, it is common to assume constant Heston parameters throughout real-world simulations, however this is not in line with historical observations. By implementing the regression models, one is able to condition the Heston parameters on the simulated state of the market.

In chapter 3 we discussed the VIX Heston model. As the VIX index is highly correlated to the Heston parameters, a linear model was proposed to predict the Heston parameters. This model leads to calibration and simulation advantages, but suffers from accuracy loss. To this end several regression models were proposed, aimed at preserving the advantages of the VIX Heston model, while increasing prediction accuracy. It turned out to be difficult to significantly improve the VIX Heston model, as the linear model was already relatively accurate. However, the multi-output support vector regression approach appeared to be the most robust methodology throughout the different data sets. The multi-output support vector regression approach distinguishes itself from the other models by taking correlation between the Heston parameters into account, which is an important feature in this case, as the Heston parameters are highly correlated to each other. The other models (polynomial regression and gradient boosted regression trees) did improve the accuracy on a parameter level. However, this did not lead to improved implied volatility surface predictions, due to the degrees of freedom in the Heston model: misspecifications in one parameter can be (partially) nullified by misspecification in another parameter. One can argue that this is an undesirable feature, in which case the polynomial regression and gradient boosted regression trees are preferred over the VIX Heston model. Finally, all regression models selected the VIX index and risk-free interest rates as explanatory variables. This indicates the robustness of these features, as each regression model assumes a different relationship between the market variables and Heston parameters.

Chapter 4 aims at answering the research question 'Is it justified to assume a recalibrated risk-neutral measure in (real-world) simulations, even when this is in conflict with the assumptions of the underlying risk-neutral pricing model?' The Heston model assumes constant $\bar{v}$ and γ, hence recalibration of these parameters in a real-world simulation is in violation with the theoretical assumptions. To this end, we proposed a hedge test where different hedging strategies were compared: a strategy that does not take changes of $\bar{v}$ and γ into account and two strategies that do take changes of $\bar{v}$ and γ into account. The latter produced the most accurate future option prices in simulation as well as empirical tests. This led to the conclusion that it is justified to recalibrate the Heston parameters if one is interested in the most accurate future option prices. However, the underlying assumptions cannot be violated if one is interested in uniqueness of the risk assessment (for example in a regulatory framework), as the recalibration process is subjected to the modeller's point of view, while the implied volatility surface is unique.

The implementation of the VIX Heston model in a real-world simulation is discussed in chapter 5. In this chapter, the impact of implementing the recalibration process is assessed in terms of the Solvency Capital Requirement. An essential part of the Solvency Capital Requirement is the risk-neutral valuation of the expected liabilities at $t = 1$, which requires the risk-neutral parameters at $t = 1$. One can assume the risk-neutral parameters to be constant and implied by the implied volatility surface at $t = 0$. However, by implementing the regression models one is able to recalibrate the Heston parameters at $t = 1$, conditioned on the simulated state of the market. We assessed the Solvency Capital Requirement according to the Least-Squares Monte Carlo algorithm of two different insurance products: GMAB and GMWB. The two different approaches (constant and recalibrated parameters) did not result in the same Solvency Capital Requirement. Based on historical data, the mean absolute difference between the approaches was 28.7% and 7.6% for the GMAB and GMWB riders, respectively.

6.2. Future research

There are still many topics that require future research. In what follows, we present a list of possible topics.

In figure 3.5 it can be seen that the out-of-sample accuracy of γ in the US data set is low. The in-sample correlation between the VIX index and γ is 0.86, but the out-of-sample correlation is only 0.32. It might therefore be relevant to investigate the cause of this phenomenon and adjust the regression model accordingly, such that prediction accuracy with respect to γ is improved.

During our research we did not find any accurate predictor for ρ, despite its pronounced relationship with the implied volatility surface (see figure 3.2). Hence, we assumed a constant ρ throughout our research, which is justified by its mean reverting behaviour. However, the implied volatility surface predictions will become more accurate, if one is able to accurately predict ρ.

So far, we have investigated three different regression models. Of course, this is only a small fraction of all available models. The development of these models, especially machine learning algorithms, is an active field of research. In future research it can therefore be advantageous to investigate the accuracy of other regression models, for example neural networks.

This research is limited to the Heston model and it is unclear whether the parameters of other models are as predictable as the Heston parameters. Future research could focus on the predictability of parameters of different risk-neutral models.

The current impact in chapter 5 test is somewhat simplified. These simplifications must be removed, if one wishes to perform a realistic risk-assessment. For example, the asset price follows a geometric Brownian motion, which is unable to capture all characteristics of the market. Moreover, the recalibration process is modelled by the VIX Heston model. In chapter 3 we showed that this model is outperformed by the multi-output support vector regression in terms of accuracy. Hence, in order to develop a more accurate risk-assessment, we recommend a real-world simulation that includes the risk-free rates, such that the multi-output support vector regression can be implemented.

Finally, we recommend investigating the implementation of the recalibration process in multi-horizon applications. In chapter 5 we only focused on the expected liabilities at $t = 1$. In some risk-management applications, one is interested in multiple future dates. Under the assumption of constant parameters, one can simulate a single set of risk-neutral scenarios. This set can then be used to define the regression functions at all future dates, for example with the Least-Square Monte Carlo method. This approach does not work in case of recalibrated parameters, as the risk-neutral measure is adjusted at every time-step. Hence, separate risk-neutral scenario sets are required for each date of interest, which is computationally more demanding. To this end, it is important that an efficient approach is developed that can efficiently deal with multi-horizon applications.

Bibliography

Alexander, C. and Nogueira, L. M. Hedging with Stochastic Local Volatility. *ISMA Centre Discussion Papers in Finance*, 2004.

Alexander, C., Kaeck, A., and Nogueira, L. M. Model Risk Adjusted Hedge Ratios. *The Journal of Futures Markets*, 29, no. 11, pp. 1021–1049, 2009.

Algorithmics. Curve fitting for calculating SCR under Solvency II: Practical insights and best practices from leading European Insurers. *White Paper*, 2011.

Andersen, L. Efficient Simulation of the Heston Stochastic Volatility Model. *Journal of Computational Finance*, 11, no. 3, pp. 1–42, 2008.

Audrino, F. and Colangelo, D. Semi-parametric forecasts of the implied volatility surface using regression trees. *Statistics and Computing*, 20, no. 4, pp. 421–434, 2010.

Ayache, E. *The Medium of Contingency: An Inverse View of the Market*. Palgrave Macmillan, 2015.

Ayache, E. Response to Johnson : A random sample versus the radical event. *Finance and Society*, 2, no. 2, pp. 205–216, 2016.

Bakshi, G., Cao, C., and Chen, Z. Empirical Performance of Alernative Option Pricing Models. *Journal of Finance*, 52, no. 5, pp. 2003–2049, 1997.

Bates, D. S. Bates Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options. *The Review of Financial Studies*, 9, no. 1, pp. 69–107, 1996.

Bauer, D., Bergmann, D., and Reuss, A. Solvency II and nested simulations–a least-squares Monte Carlo approach. In *Proceedings of the 2010 ICA congress*, 2010.

Bauer, D., Kling, A., and Russ, J. A Universal Pricing Framework for Guaranteed Minimum Benefits in Variable Annuities. *ASTIN Bulletin*, 38, no. 2, pp. 621–651, 2008.

Benhamou, E., Gobet, E., and Miri, M. Time dependent Heston model. *SIAM Journal on Financial Mathematics*, 1, pp. 289–325, 2010.

Black, F. and Scholes, M. The pricing of options and corporate liabilities. *The Journal of Political Economy*, 81, no. 3, pp. 637–654, 1973.

Brigo, D. and Mercurio, F. Lognormal-mixture dynamics and calibration to market volatility smiles. *International Journal of Theoretical and Applied Finance*, 5, no. 4, pp. 427–446, 2002.

Broadie, M., Du, Y., and Moallemi, C. C. Risk Estimation via Regression. *Operations Research*, 63, no. 5, pp. 1077–1194, 2015.

Cao, L. J. and Tay, F. E. H. Financial Time Series Forecasting. *IEEE Transactions on neural networks*, 14, no. 6, pp. 1506–1518, 2003.

Carr, P. and Madan, D. Option valuation using the fast Fourier transform. *Journal of Computational Finance*, 2, no. 1990, pp. 1–18, 1999.

Cathcart, M. J. *Monte Carlo Simulation Approaches to the Valuation and Risk Management of Unit-Linked Insurance Products with Guarantees*. PhD thesis, Heriot-Watt University, 2012.

CBOE. The CBOE Skew Index. *Exchange Organizational Behavior Teaching Journal*, 2010.

CBOE. The CBOE Volatility Index - VIX: The powerful and flexible trading and risk managment tool from the Chicago Board Options Exchange. *White Paper*, 2015.

Chen, S., Zhou, Z., and Li, S. An efficient estimate and forecast of the implied volatility surface: A nonlinear Kalman filter approach. *Economic Modelling*, 58, pp. 655–664, 2016.

Cont, R., da Fonseca, J., Durrleman, V., Fonseca, D., and José. Stochastic Models of Implied Volatility Surfaces. *Economic Notes*, 31, no. 2, pp. 361–377, 2003.

Cortes, C. and Vapnik, V. Support-Vector Networks. *Machine Learning*, 297, no. 3, pp. 273–297, 1995.

Cox, J., Ingersoll Jr, J., and Ross, S. A theory of the term structure of interest rates. *Econometrica*, 53, no. 2, pp. 385–407, 1985.

de Vincent-Humphreys, R. and Noss, J. Estimating probability distributions of future asset prices: empirical transformations from option-implied risk-neutral to real-world density functions. Bank of England Working Papers, 2012.

Devineau, L. and Loisel, S. Risk aggregation in Solvency II: How to converge the approaches of the internal models and those of the standard formula? *Bulletin Français d'Actuariat*, 9, no. 18, pp. 107–145, 2009.

Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A., and Vapnik, V. Support Vector Regression Machines. *Neural Information Processing Systems*, 9, pp. 155–161, 1997.

Duan, J.-C. and Yeh, C.-Y. Price and Volatility Dynamics Implied by the VIX Term Structure Price and Volatility Dynamics Implied by the VIX Term Structure. NUS RMI Working Paper No. 11/05, 2012.

Duffie, D. *Dynamic Asset Pricing Theory*. Princeton University Press, 1996.

Dupire, B. Pricing with a Smile. *Risk*, 7, pp. 18–20, 1994.

Feng, Q., Feng, R., and Oosterlee, C. W. Efficient Calculation of Solvency Capital Requirement for Guaranteed Minimum Withdrawal Benefit. Working paper, 2017.

Friedman, J. H. A Tree-Structured Approach to Nonparametric Multiple Regression. *Smoothing Techniques for Curve Estimation*, pages 5–22, 1979.

Friedman, J. H. Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29, no. 5, pp. 1189–1232, 2001.

Friedman, J. H. Stochastic Gradient Boosting. *Computational Statistics & Data Analysis*, 38, no. 4, pp. 367–378, 2002.

Gatheral, J. *The Volatility Surface: A Practitioner's Guide*. John Wiley & Sons Ltd, 2006.

Gauthier, P. and Rivaille, P.-Y. H. Fitting the smile, Smart parameters for SABR and Heston, 2009. URL `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1496982`.

Gavin, H. P. The Levenberg-Marquardt method for nonlinear least squares curve-fitting problems, 2017. URL `http://een.iust.ac.ir/profs/Farrokhi/NeuralNetworks/Levenberg-Marquardt/lm.pdf`.

Hamrita, M. E. and Trifi, A. The Relationship between Interest Rate, Exchange Rate and Stock Price: A Wavelet Analysis. *International Journal of Economics and Financial Issues*, 1, no. 4, pp. 220–228, 2011.

Hansen, L. P. Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, 50, no. 4, pp. 1029–1054, 1982.

Heston, S. L. A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *The Review of Financial Studies*, 6, no. 2, pp. 327–343, 1993.

Hibbert, A. M., Daigler, R. T., and Dupoyet, B. A behavioral explanation for the negative asymmetric return-volatility relation. *Journal of Banking and Finance*, 32, no. 10, pp. 2254–2266, 2008.

Jain, S. and Oosterlee, C. W. The Stochastic Grid Bundling Method : Efficient Pricing of Bermudan Options and their Greeks. *Applied Mathematics and Computation*, 269, no. 1, pp. 412–431, 2015.

James W. Cooley and Tukey, J. W. An Algorithm for the Machine Calculation of Complex Fourier Series. *Mathematics of Computation*, 19, no. 90, pp. 297–301, 1965.

Janek, A., Kluge, T., Weron, R., and Wystup, U. FX smile in the Heston model. *Statistical Tools for Finance and Insurance*, pages 133–162, 2011.

Jex, M., Henderson, R., and Wang, D. Pricing exotics under the smile. *Risk*, pages 72–75, 1999.

Kenyon, C., Green, A., and Berrahoui, M. Which measure for pfe? the risk appetite measure, a, 2015. URL `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2703965`.

Kurpiel, A. and Roncalli, T. Option hedging with stochastic volatility, 1998.

Law, T. and Shawe-Taylor, J. Practical Bayesian support vector regression for financial time series prediction and market condition change detection Practical Bayesian support vector regression for financial time series prediction and market. *Quantitative Finance*, pages 1403–1416, 2017. ISSN 1469-7688. doi: 10.1080/14697688.2016.1267868.

Lipton, A. The vol smile problem. *Risk*, pages 61–65, 2002.

Longstaff, F. A. and Schwartz, E. S. Valuing American Options by Simulation : A Simple Least-Squares Approach. *The Review of Financial Studies*, 14, no. 1, pp. 113–147, 2001.

Milevsky, M. A. and Salisbury, T. S. The Real Option to Lapse a Variable Annuity: Can Surrender Charges Complete the Market. In *Conference Proceedings of the 11th Annual International AFIR Colloquium*, 2001.

Milevsky, M. A. and Salisbury, T. S. Financial valuation of guaranteed minimum withdrawal benefits. *Insurance: Mathematics and Economics*, 38, no. 1, pp. 21–38, 2005.

Mixon, S. Factors Explaining Movements in the Implied Volatility Surface. *Journal of Futures Markets*, 22, no. 10, pp. 915–937, 2002.

Rebonato, R. *Volatility and Correlation 2nd Edition*. John Wiley & Sons Ltd, 2004.

Schrager, D. Replicating Portfolios for Insurance Liabilities. *Actuarial Sciences*, pages 57–60, 2008.

Singor, S. N., Boer, A., and Oosterlee, C. W. Modeling the dynamics of equity index option implied volatilities in a real world scenario set, 2014. URL `http://www.ortec-finance.com/-/media/Files/Research/Working-Papers/Methodological/Modeling-the-dynamics-of-equity-index-option-implied-volatilities-in-a-real-world-scenario-set.ashx`.

Smola, A. J. and Schölkopf, B. A tutorial on support vector regression. *Statistics and Computing*, 14, pp. 199–222, 2004.

Sobol, I. M. and Levitan, Y. U. L. A Pseudo-Random Number Generator for Personal Computers. *Computers and Mathematics with Applications*, 37, pp. 33–40, 1999.

Teixeira, B. O. S., Tôrres, L. A. B., Aguirre, L. A., and Bernstein, D. S. On unscented Kalman filtering with state interval constraints. *Journal of Process Control*, 20, no. 1, pp. 45–57, 2010.

van der Stoep, A. W., Grzelak, L. A., and Oosterlee, C. W. The Heston Stochastic-Local Volatility Model: Efficient Monte Carlo Simulation. *International Journal of Theoretical and Applied Finance*, 17, no. 7, pp. 1–30, 2014.

Wan, E. A. and van der Merwe, R. The unscented Kalman filter for nonlinear estimation. *Technology*, 5, pp. 153–158, 2000.

Xu, S., An, X., Qiao, X., Zhu, L., and Li, L. Multi-output least-squares support vector regression machines. *Pattern Recognition Letters*, 34, no. 9, pp. 1078–1084, 2013.

A

Pricing under the Heston model: FFT method

When calibrating the Heston model, one needs to price many options with different strikes and maturities. Hence, an efficient pricing method is desirable. To this end we will implement the FFT method proposed in Carr and Madan (1999). In this section we will give a brief overview of the method. Further details and analysis can be found in Carr and Madan (1999).

Fourier transform of an option price

This method operates in log space, so let k denote the log of strike price K. Furthermore, let $C_T(k)$ be the value of a call option with maturity T and strike price e^k . Moreover, let $q_T(s)$ be the risk-neutral density of the log price s_T. Finally, define the characteristic function $\phi_T(u)$ of this risk-neutral density by:

$$\phi_T(u) \equiv \int_{-\infty}^{\infty} e^{ius} q_T(s) \mathrm{d}s. \tag{A.1}$$

The value of an option is equal to the discounted value of the expected pay-off. For a European call option, this is equivalent to

$$C_T(k) \equiv \int_{k}^{\infty} e^{-rT}(e^s - e^k) q_T(s) \mathrm{d}s. \tag{A.2}$$

Note that $C_T(k)$ is not square-integrable: $C_T(k) \to S_0$ as $k \to -\infty$. To this end we will modify $C_T(k)$ into a square-integrable function:

$$c_T(k) \equiv e^{\alpha k} C_T(k), \qquad \alpha > 0. \tag{A.3}$$

It is expected that $c_T(k)$ is square-integrable for a range of positive values of α. The proper choice of α will be discussed later in this section. Now that we have a square-integrable function, we can consider its Fourier transform by

$$\psi_T(z) = \int_{-\infty}^{\infty} e^{izk} c_T(k) \mathrm{d}k. \tag{A.4}$$

If we are able to develop an expression for $\psi_T(z)$, we can use it to determine the price of a call option by using the inverse transform:

$$C_T(k) = \frac{e^{-\alpha k}}{\pi} \Re \left\{ \int_0^{\infty} e^{-izk} \psi_T(z) \mathrm{d}z \right\}. \tag{A.5}$$

We can obtain an analytical expression for $\psi_T(z)$ by using its definition:

$$\begin{aligned}
\psi_T(z) &= \int_{-\infty}^{\infty} e^{izk} c_T(k) \mathrm{d}k \\
&= \int_{-\infty}^{\infty} e^{izk} \int_{k}^{\infty} e^{\alpha k} e^{-rT}(e^s - e^k) q_T(s) \mathrm{d}s \mathrm{d}k \\
&= \int_{-\infty}^{\infty} e^{-rT} q_T(s) \int_{-\infty}^{s} \left(e^{s + k(\alpha + iz)} - e^{k(1+\alpha+iz)} \right) \mathrm{d}k \mathrm{d}s
\end{aligned}$$

$$
\begin{aligned}
&= \int_{-\infty}^{\infty} e^{-rT} q_T(s) \left(\frac{e^{s(1+\alpha+iz)}}{\alpha+iz} - \frac{e^{s(1+\alpha+iz)}}{1+\alpha+iz} \right) \mathrm{d}s \\
&= \frac{e^{-rT}}{\alpha^2+\alpha-z^2+i(2\alpha+1)z} \int_{-\infty}^{\infty} q_T(s) e^{i(z-i(\alpha+1))s} \mathrm{d}s \\
&= \frac{e^{-rT}\phi_T[z-(\alpha+1)i]}{\alpha^2+\alpha-z^2+i(2\alpha+1)z}.
\end{aligned} \tag{A.6}
$$

Now, by substituting (A.6) into (A.5), we obtain an integral expression for the value of a European call option. In the next section we will describe the numerical scheme to evaluate this expression. Note that for $\alpha = 0$ there is a singularity when $z = 0$, which justifies the modification of $C_T(k)$ by $\exp(\alpha k)$ or something similar. It is possible to derive an upper bound for α by looking at the characteristic function, but in practice we experienced that $\alpha = 1$ leads to stable results.
Finally, we still have the (model-dependent) characteristic function $\phi_T(u)$. In log space, this function is known for the Heston model, given its initial conditions $(\log[S_0], v_0)$:

$$
\begin{aligned}
\phi_T(u;\log[S_0],v_0) = \quad & \exp\left(iu\log[S_0] + iur(T-t_0) + \frac{v_0}{\gamma^2}\left(\frac{1-e^{-D(T-t_0)}}{1-Ge^{-D(T-t_0)}}\right)(\kappa - i\rho\gamma u - D)\right) \\
& \exp\left(\frac{\kappa\bar{v}}{\gamma^2}\left\{(T-t_0)(\kappa - i\rho\gamma u - D) - 2\log\left[\frac{1-Ge^{-D(T-t_0)}}{1-G}\right]\right\}\right),
\end{aligned} \tag{A.7}
$$

with

$$
D = \sqrt{(\kappa - i\rho\gamma u)^2 + (u^2+iu)\gamma^2}, \tag{A.8}
$$

$$
G = \frac{\kappa - i\rho\gamma u - D}{\kappa - i\rho\gamma u + D}. \tag{A.9}
$$

Applying the fast Fourier transform

In the previous section we deducted a pricing formula for a European call option (see equation (A.5)). Even though we derived an analytical expression for $\psi_T(k)$, we still need to integrate over an infinite domain. There is no analytical expression for this integral, hence we need to discretize the domain into N points:

$$
C_T(k) \approx \hat{C}_T(k) = \frac{e^{-\alpha k}}{\pi} \Re\left\{ \sum_{j=1}^{N} e^{-iz_j k} \psi_T(z_j) \Delta z \right\}. \tag{A.10}
$$

With the distance between two discretized points Δz and $z_j = \Delta z(j-1)$. We are mainly interested in the log strike values k near 0, hence we will apply a linear spacing with size λ:

$$
k_p = -b + \lambda(p-1), \qquad p = 1,\ldots,N. \tag{A.11}
$$

Where $b = \frac{1}{2}\lambda N$, this leads to log strikes values ranging from $-b$ to b. Substituting (A.11) into (A.10) gives us

$$
\begin{aligned}
\hat{C}_T(k_p) = \quad & \frac{e^{-\alpha k_p}}{\pi} \Re\left\{ \sum_{j=1}^{N} e^{-iz_j[-b+\lambda(p-1)]} \psi_T(z_j)\Delta z \right\} \\
= \quad & \frac{e^{-\alpha k_p}}{\pi} \Re\left\{ \sum_{j=1}^{N} e^{-i\lambda\Delta z(j-1)(p-1)} e^{ibz_j} \psi_T(z_j)\Delta z \right\}, \quad \text{for } p = 1,\ldots,N.
\end{aligned} \tag{A.12}
$$

With the approximation written in this form, we can apply the fast Fourier transform. The fast Fourier transform is an algorithm that is able to efficiently compute sums of the form

$$
w(k) = \sum_{j=1}^{N} e^{-i\frac{2\pi}{N}(j-1)(k-1)} x(j), \qquad \text{for } k = 1,\cdots,N, \tag{A.13}
$$

with $N = 2^m$, for some constant m. The algorithm was proposed in James W. Cooley and Tukey (1965) and it reduces the complexity from $O(N^2)$ to $O(N\log_2[N])$, by using recurrences in the exponent. For details regarding this algorithm we refer the reader to James W. Cooley and Tukey (1965). From (A.13) we observe the constraint

$$
\lambda\Delta z = \frac{2\pi}{N}. \tag{A.14}
$$

Hence, refining the integration approximation grid (by decreasing Δz) leads to a larger log strike spacing. This means there is a possibility that no strike price close to the desired strike price is calculated, which

could lead to large errors. To this end, we will use Simpson's weighting rule in our summation weight, allowing us to increase Δz (which decreases λ) without losing accuracy. Applying Simpson's weighting rule results into

$$\hat{C}_T(k_p) = \frac{e^{-\alpha k_p}}{\pi} \Re \left\{ \sum_{j=1}^{N} e^{-i\frac{2\pi}{N}(j-1)(p-1)} e^{ibz_j} \psi_T(z_j) \frac{\Delta z}{3} \left(3 + (-1)^j - \delta_{j-1}\right) \right\}, \qquad \text{for } p = 1, \ldots, N. \tag{A.15}$$

with $\delta_0 = 1$ and $\delta_i = 0$ otherwise. The summation for each p in (A.15) is efficiently computed using the fast Fourier transform. Finally, if one wants to determine the price for a specific strike $k^* = \log(K^*)$, an interpolation on the calculated k_p is needed. A cubic-spline interpolation turned out to be highly accurate, after performing numerical experiments.

Error Analysis

In this section we will present an analysis regarding the error of the approximation (which can also be found in Carr and Madan (1999)), complemented by our findings regarding the optimal parameters.
The first part of the error is caused by imposing an upper-bound on the infinite integral. By summing up to z_N we neglect the upper tail of the integral. To analyse the behaviour of this error we note that the modulus of ϕ_t is bounded by $\mathbb{E}[S_T^{\alpha+1}]$. Hence, we know

$$|\psi_T(z)|^2 \leq \frac{\mathbb{E}[S_T^{\alpha+1}]}{(\alpha^2 + \alpha - z^2)^2 + (2\alpha+1)^2 z^2} \leq \frac{A}{z^4}, \tag{A.16}$$

for some constant A. Equivalently

$$|\psi_T(z)| \leq \frac{\sqrt{A}}{z^2}. \tag{A.17}$$

It follows that the upper tail of the integral is bound by

$$\int_a^\infty |\psi_T(z)| < \frac{\sqrt{A}}{a}. \tag{A.18}$$

By using this bound we can ensure the truncation error of the approximation to be smaller than ϵ, by setting

$$z_N = a > \frac{\exp(-\alpha k)\sqrt{A}}{\pi \epsilon}. \tag{A.19}$$

The second part of the error is caused by discretizing the integral domain. We used Simpson's rule for determining the weight of each discretized part. It is well-known that the error of Simpson's rule is $O(N^{-4})$. Thus, by doubling the number of points, we decrease this part of the error by a factor of 16.

After numerical experiments, we decided upon the following parameters. For strike levels near S_0 we noticed convergence when using $N = 2^{14}$ and $\Delta z = 0.1$. This automatically leads to $z_N = 2^{13}$ and $\lambda \approx 3.8 \cdot 10^{-3}$. Options far out of the money (strike levels at 50% of S_0) with short maturities ($T = 3$ months) required more points to be accurate, namely $N = 2^{16}$, resulting in $z_N = 2^{15}$ and $\lambda \approx 9.6 \cdot 10^{-4}$. However, increasing N comes at the cost of computational time, hence we only increase N when necessary.

B

Regression model results: UK and Europe

B.1. Optimal set-up

B.1.1. VIX Heston

UK

$$\Omega_t^{\text{Heston}}(X) = \begin{cases} \kappa_t & = 1.0, \\ v_{0,t} & = (-0.0014 + 0.0096 \cdot \text{VFTSE}_t)^2, \\ \bar{v}_t & = \left(0.0590 + 0.0110 \cdot \text{VFTSE}_{\text{filter}_t}\right)^2, \\ \gamma_t & = 0.2556 + 0.0206 \cdot \text{VFTSE}_t, \\ \rho_t & = -0.6858. \end{cases} \tag{B.1}$$

Europe

$$\Omega_t^{\text{Heston}}(X) = \begin{cases} \kappa_t & = 1.0, \\ v_{0,t} & = (0.0013 + 0.0094 \cdot \text{VIX}_t)^2, \\ \bar{v}_t & = \left(0.0518 + 0.0100 \cdot \text{VIX}_{\text{filter}_t}\right)^2, \\ \gamma_t & = 0.0571 + 0.0252 \cdot \text{VIX}_t, \\ \rho_t & = -0.6471. \end{cases} \tag{B.2}$$

B.1.2. Polynomial regression

UK

$$\Omega_t^{\text{Heston}}(X) = \begin{cases} \kappa_t & = 1.0, \\ v_{0,t} & = (-0.0014 + 0.0096 \cdot \text{VFTSE}_t)^2, \\ \bar{v}_t & = \left(0.0590 + 0.0110 \cdot \text{VFTSE}_{\text{filter}_t}\right)^2, \\ \gamma_t & = 0.2556 + 0.0206 \cdot \text{VFTSE}_t, \\ \rho_t & = -0.6858. \end{cases} \tag{B.3}$$

Europe

$$\Omega_t^{\text{Heston}}(X) = \begin{cases} \kappa_t & = 1.0, \\ v_{0,t} & = (0.0015 + 0.0094 \cdot \text{VFTSE}_t)^2, \\ \bar{v}_t & = \left(0.0530 + 0.0098 \cdot \text{VFTSE}_{\text{filter}_t} + 0.0094 \cdot r_{720}\right)^2, \\ \gamma_t & = 0.0543 + 0.0254 \cdot \text{VFTSE}_t, \\ \rho_t & = -0.6454. \end{cases} \tag{B.4}$$

B.1.3. Multi-output support vector regression

UK

Explanatory variables: VFTSE, $\text{VFTSE}_{\text{filter}}$, r_{720}

$$\lambda_1 = 10^{-8}, \qquad \lambda_2 = 10^{-10}. \tag{B.5}$$

Europe
Explanatory variables: VSTOXX, $\text{VSTOXX}_{\text{filter}}$

$$\lambda_1 = 10^{-8}, \qquad \lambda_2 = 10^{-10}. \tag{B.6}$$

B.2. Accuracy results

UK

Model	SSE	MAE	R^2	R^2_{Min}	MSE_{v_0}	$\text{MSE}_{\bar{v}}$	MSE_{γ}
VIX Heston	0.0577	0.0093	0.9340	0.7401	$3.723 \cdot 10^{-5}$	$\mathbf{7.155 \cdot 10^{-5}}$	0.0039
Polynomial Regression	0.0577	0.0093	0.9340	0.7401	$3.723 \cdot 10^{-5}$	$\mathbf{7.155 \cdot 10^{-5}}$	0.0039
Multi-output SVR	**0.0547**	**0.0091**	**0.9374**	**0.7413**	$\mathbf{2.483 \cdot 10^{-4}}$	$8.273 \cdot 10^{-5}$	**0.0023**
Unrestricted	*0.0126*	*0.0045*	*0.9855*	*0.9690*	-	-	-

Table B.1: UK Out-of-sample accuracy of the regression models according to the error measures defined in (3.35) and (3.36).

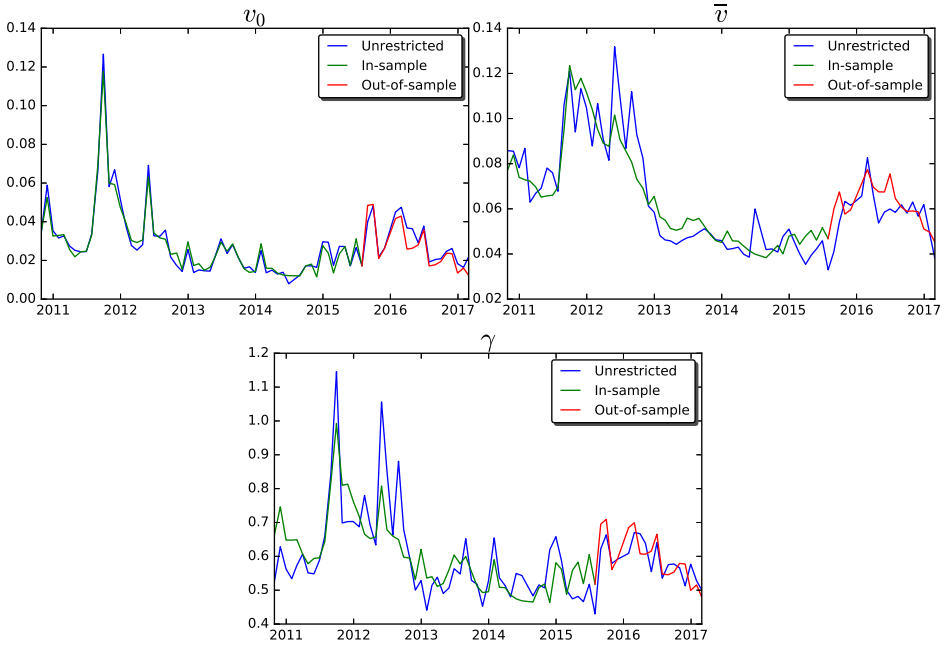

Figure B.1: Unrestricted versus multi-output least-squares support vector regression predicted parameters over time of the UK data set.

Europe

Model	SSE	MAE	R^2	R^2_{Min}	MSE_{v_0}	$\text{MSE}_{\bar{v}}$	MSE_{γ}
VIX Heston	0.0433	0.0078	0.9389	0.7693	$4.231 \cdot 10^{-5}$	$1.064 \cdot 10^{-4}$	0.0073
Polynomial Regression	0.0446	0.0082	0.9371	0.7293	$4.217 \cdot 10^{-5}$	$\mathbf{6.807 \cdot 10^{-5}}$	0.0076
Multi-output SVR	**0.0376**	**0.0074**	**0.9469**	**0.7859**	$\mathbf{2.067 \cdot 10^{-5}}$	$1.345 \cdot 10^{-4}$	**0.0038**
Unrestricted	*0.0154*	*0.0051*	*0.9779*	*0.9190*	-	-	-

Table B.2: Europe Out-of-sample accuracy of the regression models according to the error measures defined in (3.35) and (3.36).

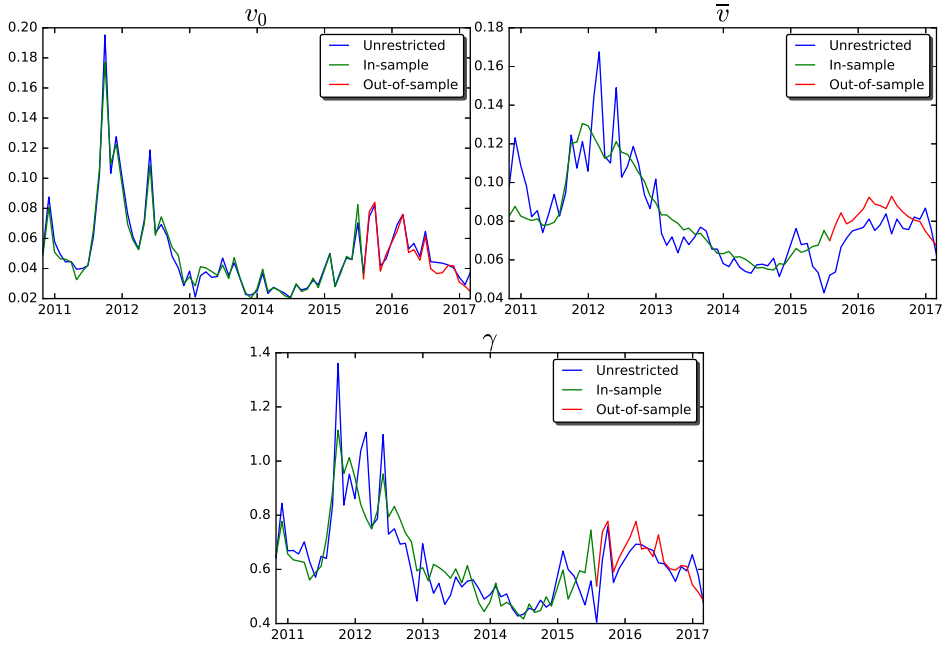

Figure B.2: Unrestricted versus multi-output least-squares support vector regression predicted parameters over time of the Europe data set.

C

Improved calibration procedure

In chapter 3 we proposed several approaches to predict the Heston parameters, given a set of market features. Unfortunately, these predictions did not always coincide with the optimal parameters of the unrestricted model (see for example figure 3.5). When valuing derivatives, these suboptimal parameters can lead to pricing errors and arbitrage opportunities. To this end, we aim to improve the accuracy of these predictions, by allowing the parameters to be driven by a stochastic component, in addition to the previous estimations. These stochastic components can be thought of as prediction errors. After identifying the stochastic processes describing these errors, we will be able to update the parameter predictions according to the observed implied volatility surface.

In this chapter, the dynamics of the Heston parameters will be represented by stochastic differential equations (SDEs). In general, parameter p will follow

$$\mathrm{d}p_t = \mu_p(t, p_t)\mathrm{d}t + \sigma_p(t, p_t)\mathrm{d}W_t^p. \tag{C.1}$$

The drift term $\mu_p(t, p_t)$ and volatility component $\sigma_p(t, p_t)$ can be any bounded time-dependent dynamic process. For example, the drift term can coincide with one of the prediction models proposed in chapter 3 and the volatility component can correspond to the prediction error. Moreover, we assume a correlation between parameter p and q:

$$\left\langle \mathrm{d}W_t^p, \mathrm{d}W_t^q \right\rangle = \Gamma_{\{p,q\}}\mathrm{d}t. \tag{C.2}$$

The objective of this chapter is to identify the hidden state of the market (quantified by the Heston parameters) given some measurement (in our case the implied volatility surface). One of the most famous methods that solves these kinds of problems is the Kalman filter. The Kalman filter is known for its ability to combine historical data with knowledge of the system in an optimal recursive approach. However, the Kalman filter assumes a linear relation between the hidden state and observed measurement, an assumption which is violated in most applications. To this end, several extensions of the Kalman filter have been introduced, dealing with non-linear relations. In this chapter we will consider the unscented Kalman filter (proposed in Wan and van der Merwe (2000)), which is able to deal with non-linear relations and is known for its speed and accuracy. For example, Chen et al. (2016) used the unscented Kalman filter to calibrate the parameters of the SABR model, where they assumed mean reverting SDEs to describe the parameters. This way, they were able to efficiently estimate the parameters and give a six minute forecast of the out-of-the-money implied volatility surface. We will adopt a similar approach for the regression models proposed in chapter 3. Throughout this chapter, we will describe the mathematical details of the unscented Kalman filter and explain how it is combined with the regression models. Finally, we will discuss the results of the model, by comparing it to a benchmark model.

C.1. Unscented Kalman filter

The unscented Kalman filter is a method that is able to filter (hidden) state variables from a (noisy) measurement. The state variables are driven by dynamic stochastic processes and the measured variable

depends on the state variables. In general, the link between state variables Ω_{t_k} and observed measurements σ_{t_k} at time t_k can be represented as follows,

$$\begin{cases} \Omega_{t_k} &= f(\Omega_{t_{k-1}}, t_{k-1}, a_{t_{k-1}}) + w_{t_{k-1}}, \\ \sigma_{t_k} &= h(t_k, \Omega_{t_k}, a_{t_k}) + v_{t_k}. \end{cases} \tag{C.3}$$

With state function $f(\cdot)$, measurement function $h(\cdot)$ and auxiliary input a_{t_k}. Moreover, the random component $w_{t_{k-1}}$ and measurement noise v_{t_k} are assumed to be Gaussian distributed.

This model is analysed by recursively updating the parameters in a two-stage procedure, a forecast step and a measurement-update step. The forecast step samples different parameter sets (based on the previous state variables and covariance), in order to capture the *true* mean and covariance of the state variables. The samples are called sigma points and they provide a prior estimate of the parameters, covariance and measurement. The measurement-update step establishes a posterior prediction of the state variables by adding the Kalman gain (which depends on the observed measurement) to the prior estimate. The details of this process are discussed in this section.

C.1.1. Sigma points

A crucial step of the unscented Kalman filter is identifying the sample set of state variables, also called sigma points. Assume that the parameter set $\Omega_{k-1|k-1}$ and covariance matrix $P_{k-1|k-1}$ of time-step $k-1$ are known. In order to analyse the behaviour of the parameters at time-step k, a set of sigma points and their associated weights are defined recursively:

$$\left(\overline{\Omega}_{k-1|k-1}, \overline{W}_{k-1}\right) = \left\{ \left(\Omega^j_{k-1|k-1}, W^j_{k-1}\right) \middle| j = 0, \ldots, 2n \right\}, \tag{C.4}$$

where n is the number of parameters. The sigma points are defined as follows,

$$\begin{cases} \Omega^0_{k-1|k-1} = \Omega_{k-1|k-1}, \\ \Omega^j_{k-1|k-1} = \Omega^0_{k-1|k-1} + \sqrt{n+\lambda}\left(\sqrt{P_{k-1|k-1}}\right)_j, & j = 1, \ldots, n, \\ \Omega^{j+n}_{k-1|k-1} = \Omega^0_{k-1|k-1} - \sqrt{n+\lambda}\left(\sqrt{P_{k-1|k-1}}\right)_j, & j = 1, \ldots, n. \end{cases} \tag{C.5}$$

With $\lambda > -n$, a predefined constant and $\left(\sqrt{P_{k-1|k-1}}\right)_j$, the jth row of matrix $\sqrt{P_{k-1|k-1}}$. The square root in equation (C.5) denotes the Cholesky square root. The Cholesky square root of a matrix is defined as the lower-triangular matrix A that satisfies

$$A := \sqrt{P_{k-1|k-1}} \Leftrightarrow P_{k-1|k-1} = AA^T. \tag{C.6}$$

This operator is uniquely defined for positive definite matrices[1]. However, covariance matrices are only positive semi-definite[2], thus the Cholesky square root is only defined when the covariance matrix is of full rank. When the covariance matrix is not positive definite (usually due to numerical errors), one can perform the transformation

$$P_{k-1|k-1} \rightarrow P_{k-1|k-1} + \epsilon I, \qquad \text{with } \epsilon = 10^{-15} + \left| \min_i \lambda_i \right|, \tag{C.7}$$

with identity matrix I and eigenvalues λ_i. The transformed matrix has only positive eigenvalues and is symmetric (because the untransformed matrix is symmetric), hence it is a positive definite matrix. The transformed matrix is not equal to the true covariance matrix and this introduces an additional error. However, if the lowest eigenvalue is close to zero (which it generally is), this error is negligible. Finally, the associated weights of the sigma points are defined through

$$\begin{cases} W^0_{k-1} = & \frac{\lambda}{\lambda+n}, \\ W^j_{k-1} = & \frac{1}{2(\lambda+n)}, \qquad j = 1, \ldots, 2n, \end{cases} \tag{C.8}$$

such that $\sum_{j=0}^{2n} W^j_{k-1} = 1$.

[1] a symmetric matrix with solely positive eigenvalues
[2] a symmetric matrix with solely non-negative eigenvalues

For notational purposes we will define a function $\Psi(\cdot)$ that simulates the process of equations (C.4)-(C.8) and returns the sigma points with their associated weights:

$$\left(\overline{\Omega}_{k-1|k-1}, \overline{W}_{k-1}\right) = \Psi(\Omega_{k-1|k-1}, P_{k-1|k-1}, n, \lambda). \tag{C.9}$$

C.1.2. Forecast step

In the forecast step, a prior parameter set is estimated using the results of the previous time step and information about the system. First an initial state needs to be assumed, with known mean $\mu_0 = \mathrm{E}[\Omega_0]$ and covariance $P_{0|0} = \mathrm{E}\left[(\Omega_0 - \mu_0)(\Omega_0 - \mu_0)^T\right]$. The sigma points and their associated weights at time-step k are defined recursively, as discussed in section C.1.1,

$$\left(\overline{\Omega}_{k-1|k-1}, \overline{W}_{k-1}\right) = \Psi(\Omega_{k-1|k-1}, P_{k-1|k-1}, n, \lambda). \tag{C.10}$$

Evaluating (C.10), results in a set of sigma points defined at time-step $k-1$. To obtain a prediction of the sigma points at time-step k, we propagate them through the state function $f(\cdot)$ as defined in (C.3),

$$\Omega^j_{k|k-1} = f\left(\Omega^j_{k-1|k-1}, t_{k-1}, a_{t_{k-1}}\right), \qquad j = 0, \ldots, 2n. \tag{C.11}$$

The prior estimate of the parameter set is equal to the weighted average of the propagated sigma points,

$$\hat{\Omega}_{k|k-1} = \sum_{j=0}^{2n} W^j_{k-1} \Omega^j_{k|k-1}. \tag{C.12}$$

Similarly, we define the prior covariance matrix as

$$\hat{P}_{k|k-1} = \sum_{j=0}^{2n} W^j_{k-1} \left(\Omega^j_{k|k-1} - \hat{\Omega}_{k|k-1}\right)\left(\Omega^j_{k|k-1} - \hat{\Omega}_{k|k-1}\right)^T + \Sigma_{w_{t_{k-1}}}, \tag{C.13}$$

where $\Sigma_{w_{t_{k-1}}}$ is defined as the variance-covariance matrix of $w_{t_{k-1}}$, the stochastic component of the state process. Using these prior estimates, we are able to update the sigma points,

$$\left(\overline{\Omega}_{k|k-1}, \overline{W}_{k}\right) = \Psi(\hat{\Omega}_{k|k-1}, \hat{P}_{k|k-1}, n, \lambda). \tag{C.14}$$

Applying the updated sigma points to the measurement function $h(\cdot)$ leads to the prior measurement estimate,

$$\begin{cases} \sigma^j_{k|k-1} & = h\left(\Omega^j_{k|k-1}, t_k, a_{t_k}\right), \quad j = 0, \ldots, 2n, \\ \hat{\sigma}_{k|k-1} & = \sum_{j=0}^{2n} W^j_k \sigma^j_{k|k-1}. \end{cases} \tag{C.15}$$

C.1.3. Measurement-update step

So far, the random components of (C.3) have not been included in the estimates. The purpose of the measurement-update step is to identify these random components, conditioned on the observed measurement σ^{Obs}_k. The estimates are then updated accordingly. The magnitude of the update depends on the so-called Kalman gain

$$K_k = \hat{P}^{\Omega,\sigma}_{k|k-1}\left(\hat{P}^{\sigma}_{k|k-1}\right)^{-1}, \tag{C.16}$$

with the covariance between measurements $\hat{P}^{\sigma}_{k|k-1}$ and the covariance between the parameter set and measurements $\hat{P}^{\Omega,\sigma}_{k|k-1}$,

$$\begin{cases} \hat{P}^{\sigma}_{k|k-1} & = \sum_{j=0}^{2n} W^j_k \left(\sigma^j_{k|k-1} - \hat{\sigma}_{k|k-1}\right)\left(\sigma^j_{k|k-1} - \hat{\sigma}_{k|k-1}\right)^T + \Sigma_{v_{t_k}}, \\ \hat{P}^{\Omega,\sigma}_{k|k-1} & = \sum_{j=0}^{2n} W^j_k \left(\Omega^j_{k|k-1} - \hat{\Omega}_{k|k-1}\right)\left(\sigma^j_{k|k-1} - \hat{\sigma}_{k|k-1}\right)^T. \end{cases} \tag{C.17}$$

Where $\Sigma_{v_{t_k}}$ is defined as the variance-covariance matrix of the measurement noise v_{t_k}. Finally, the posterior estimates are given by

$$\begin{cases} \Omega_{k|k} & = \hat{\Omega}_{k|k-1} + K_k\left(\sigma^{\mathrm{Obs}}_k - \hat{\sigma}_{k|k-1}\right), \\ \sigma_{k|k} & = h\left(\Omega_{k|k}, t_k, a_{t_k}\right), \\ P_{k|k} & = \hat{P}_{k|k-1} - K_k \hat{P}^{\sigma}_{k|k-1} K^T_k. \end{cases} \tag{C.18}$$

The process of forecasting and updating the parameter set is repeated until the final time-step has been reached.

C.1.4. Interval constrained parameters

We know that the Heston model parameters are only defined on certain intervals: v_0, $\bar{v}$ and γ are always non-negative, while $|\rho|$ is always less than or equal to one. However, if we apply the algorithm described above, we cannot impose any interval constraints on the parameters. In other words, it might occur that a value outside of its domain is assigned to one or more of the Heston model parameters. To this end, we need to adjust both forecast and measurement-update step. Teixeira et al. (2010) discusses multiple strategies to handle interval constraints on the parameters of an unscented Kalman filter. These strategies transform the sigma points, the posterior estimates or both. We decided upon a strategy that enforces the sigma points as well as the posterior estimates to be interval constrained. In this section we will give a brief description of the necessary adjustments, which are based on Teixeira et al. (2010).

In the forecast step it suffices to restrict the sigma points, since the prior estimates are weighted averages of the sigma points. However, we do not want to change the procedure when all constraints are satisfied. Thus, the sigma points are redefined as follows,

$$\begin{cases} \Omega^0_{k-1|k-1} &= \Omega_{k-1|k-1}, \\ \Omega^j_{k-1|k-1} &= \Omega^0_{k-1|k-1} + \theta^j_{k-1}\left(\sqrt{P_{k-1|k-1}}\right)_j, \quad j = 1,\dots,n, \\ \Omega^{j+n}_{k-1|k-1} &= \Omega^0_{k-1|k-1} - \theta^{j+n}_{k-1}\left(\sqrt{P_{k-1|k-1}}\right)_j, \quad j = 1,\dots,n, \end{cases} \tag{C.19}$$

where, for parameters $p = 1,\dots,n$ and sigma points $j = 1,\dots,2n$,

$$\theta^j_{k-1} = \min_{p\in\{1,\dots,n\}}(\Theta_{k-1}(j,p)), \tag{C.20}$$

$$\Theta_{k-1}(j,p) = \begin{cases} \sqrt{n+\lambda} & \text{if } H_{k-1}(j,p) = 0, \\ \min\left(\sqrt{n+\lambda}, \frac{u_p - \Omega_{k-1|k-1}(p)}{H_{k-1}(j,p)}\right) & \text{if } H_{k-1}(j,p) > 0, \\ \min\left(\sqrt{n+\lambda}, \frac{l_p - \Omega_{k-1|k-1}(p)}{H_{k-1}(j,p)}\right) & \text{if } H_{k-1}(j,p) < 0, \end{cases} \tag{C.21}$$

$$H_{k-1}(j,p) = \begin{cases} \sqrt[*]{P_{k-1|k-1}}(j,p) & \text{if } 1 \le j \le n, \\ -\sqrt[*]{P_{k-1|k-1}}(j,p) & \text{if } n+1 \le j \le 2n. \end{cases} \tag{C.22}$$

With u_p and l_p defined as the upper and lower bound for parameter p, respectively. The notation $A(c,d)$ refers to the element in the cth row and dth column of matrix A. The weights associated to the sigma points are adjusted as well:

$$\begin{cases} W^0_{k-1} &= b_{k-1}, \\ W^j_{k-1} &= a_{k-1}\theta^j_{k-1} + b_{k-1}, \quad j = 1,\dots,2n., \end{cases} \tag{C.23}$$

with

$$\begin{cases} a_{k-1} &= \frac{2\lambda-1}{2(\lambda+n)\left(\sum_{j=1}^{2n}\theta^j_{k-1} - (2n+1)\sqrt{\lambda+n}\right)}, \\ b_{k-1} &= \frac{1}{2(\lambda+n)} - \frac{2\lambda-1}{2\sqrt{\lambda+n}\left(\sum_{j=1}^{2n}\theta^j_{k-1} - (2n+1)\sqrt{\lambda+n}\right)}. \end{cases} \tag{C.24}$$

For notational purposes we will refer to this procedure as $\Psi_{\text{IC}}\left(\Omega_{k-1|k-1}, P_{k-1|k-1}, n, \lambda, u, l\right)$. An example of the transformation procedure is graphically represented in figure C.1. Note that if the constraints are satisfied before the transformation, then $\Psi_{\text{IC}}(\cdot) \equiv \Psi(\cdot)$. However, if the constraints are violated, then $\Psi_{\text{IC}}(\cdot)$ transforms the sigma points such that the constraints are satisfied. Moreover, the weights are transformed along with the sigma points, such that the transformed covariance reflects the true covariance. In conclusion, substituting $\Psi_{\text{IC}}(\cdot)$ for $\Psi(\cdot)$ in equations (C.10) and (C.14), enforces interval constraints during the forecast step with minimal adjustments.

If the market is going through severe changes, the Heston parameters might fall outside their domain after the measurement-update step. To this end, we need to add a transformation for each parameter p:

$$\Omega_{k|k}(p) \to \begin{cases} l_p & \text{if } l_p > \Omega_{k|k}(p), \\ \Omega_{k|k}(p) & \text{if } l_p \le \Omega_{k|k}(p) \le u_p, \\ u_p & \text{if } u_p < \Omega_{k|k}(p). \end{cases} \tag{C.25}$$

Adding this transformation after equation (C.18) will ensure the interval constraints are not violated during the measurement-update step.

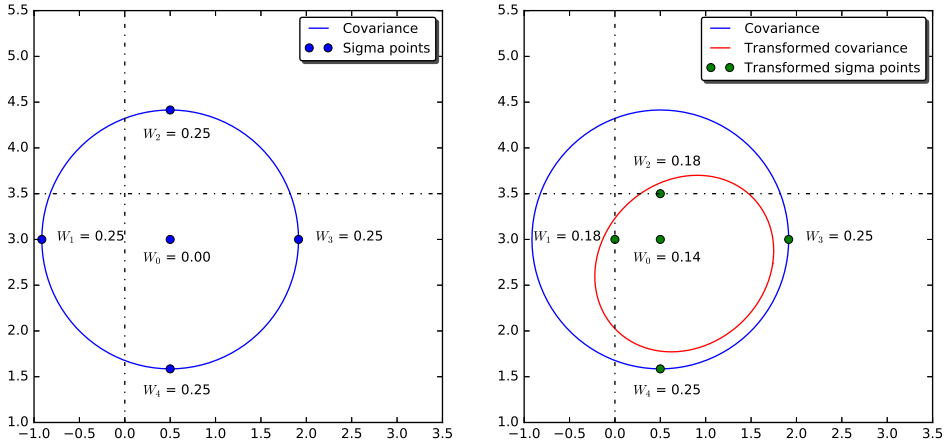

Figure C.1: An example of transformed sigma points, where $x = [0.5, 3]$, $P = I_{2\times 2}$, $x_1 \in [0, \infty)$, $x_2 \in (-\infty, 3.5]$ and $\lambda = 0$. The dashed lines represent the constraints.

C.2. State-space representation Heston model

The Heston model needs to be written in state-space form, like equation (C.3), before the unscented Kalman filter can be applied. To this end, we will derive the state-space representation of the Heston model in this section.

In this case, the hidden state is represented by the Heston parameters and they depend on the observed implied volatility surface. In section 2.5 we already described how the implied volatility surface can be obtained, given a set of Heston parameters, which is equivalent to the measurement function. Thus,

$$h(t_k, \Omega_{t_k}, a_{t_k}) = \sigma^{\text{Heston}}\left(t_k, \Omega_{t_k}^{\text{Heston}}, K, T\right). \tag{C.26}$$

The state function $f(\cdot)$, as defined in (C.3), is directly linked to the parameter dynamics given in (C.1). In general, there does not exist a closed-form solution of these SDEs, hence the state function is discretized,

$$\begin{aligned} f(\Omega_{t_{k-1}}^{\text{Heston}}, t_{k-1}, a_{t_{k-1}}) &= \Omega_{t_{k-1}}^{\text{Heston}} + \boldsymbol{\mu}(t_{k-1}, \Omega_{t_{k-1}}^{\text{Heston}}) \cdot (t_k - t_{k-1}), \\ w_{t_{k-1}} &= \sqrt{\Sigma_{\Omega_{t_{k-1}}}} Z_{w_{t_{k-1}}}, \\ \Sigma_{w_{t_{k-1}}} &= \Sigma_{\Omega_{t_{k-1}}}, \end{aligned} \tag{C.27}$$

where $\boldsymbol{\mu}(t_{k-1})$ is the vector consisting of the drift terms for all parameters, as defined in (C.1). Moreover, $\Sigma_{\Omega_{t_{k-1}}}$ is the variance-covariance matrix of the parameters at time-step t_{k-1} and $Z_{w_{t_{k-1}}}$ is a vector with independent standard normal distributed components. Again, the $\sqrt{\cdot}$ operator denotes the Cholesky square root. The components of the variance-covariance matrix for parameters p and q are given by

$$\begin{aligned} \Sigma_{\Omega_{t_{k-1}}}(p,q) &= \text{Cov}\left[p_{t_k}; q_{t_k} | p_{t_{k-1}}, q_{t_{k-1}}\right] \\ &= \text{Cov}\left[p_{t_{k-1}} + \mu_p(t_{k-1}, p_{t_{k-1}}) \cdot (t_k - t_{k-1}) + \int_{t_{k-1}}^{t_k} \sigma_p(s, p_s) \mathrm{d}W_s^p; \right. \\ &\quad \left. q_{t_{k-1}} + \mu_q(t_{k-1}, q_{t_{k-1}}) \cdot (t_k - t_{k-1}) + \int_{t_{k-1}}^{t_k} \sigma_q(s, q_s) \mathrm{d}W_s^q \,\middle|\, p_{t_{k-1}}, q_{t_{k-1}} \right] \\ &= \text{Cov}\left[\int_{t_{k-1}}^{t_k} \sigma_p(s, p_s) \mathrm{d}W_s^p; \int_{t_{k-1}}^{t_k} \sigma_q(s, q_s) \mathrm{d}W_s^q \,\middle|\, p_{t_{k-1}}, q_{t_{k-1}} \right] \\ &\approx \text{Cov}\left[\sigma_p(t_{k-1}, p_{t_{k-1}}) \mathcal{N}_p(0, t_k - t_{k-1}); \sigma_q(t_{k-1}, q_{t_{k-1}}) \mathcal{N}_q(0, t_k - t_{k-1}) \,\middle|\, p_{t_{k-1}}, q_{t_{k-1}} \right] \\ &= \sigma_p(t_{k-1}, p_{t_{k-1}}) \sigma_q(t_{k-1}, q_{t_{k-1}}) \text{Cov}[\mathcal{N}_p(0, t_k - t_{k-1}); \mathcal{N}_q(0, t_k - t_{k-1})] \\ &= \sigma_p(t_{k-1}, p_{t_{k-1}}) \sigma_q(t_{k-1}, q_{t_{k-1}}) \Gamma_{\{p.q\}}(t_k - t_{k-1}), \end{aligned} \tag{C.28}$$

where $\sigma_p(t, p_t)$ is defined as in (C.1) and $\Gamma_{\{p.q\}}$ represents the correlation between parameters p and q. The variance of parameter p can be retrieved by substituting $p_{t_{k-1}} = q_{t_{k-1}}$ into (C.28). Finally, we assume

the measurement noise to be

$$\begin{aligned} v_{t_k} &= \sqrt{\Sigma_{\sigma_{t_k}}} Z_{v_{t_k}}, \\ \Sigma_{v_{t_k}} &= \Sigma_{\sigma_{t_k}} = \begin{bmatrix} \sqrt{\delta} & 0 & \cdots & 0 \\ 0 & \sqrt{\delta} & 0 & \cdots \\ \vdots & \ddots & \ddots & \\ 0 & \cdots & 0 & \sqrt{\delta} \end{bmatrix}. \end{aligned} \tag{C.29}$$

With (constant) noise term δ and standard normal distributed vector $Z_{v_{t_k}}$ with independent components. In summary, the state-space representation of the Heston model is defined as

$$\begin{aligned} \Omega_{t_k}^{\text{Heston}} &= \Omega_{t_{k-1}}^{\text{Heston}} + \boldsymbol{\mu}(t_{k-1}, \Omega_{t_{k-1}}^{\text{Heston}}) \cdot (t_k - t_{k-1}) + \sqrt{\Sigma_{\Omega_{t_{k-1}}}} Z_{w_{t_{k-1}}}, \\ \sigma_{t_k}^{\text{Market}} &= \sigma^{\text{Heston}}\left(t_k, \Omega_{t_k}^{\text{Heston}}, K, T\right) + \sqrt{\Sigma_{\sigma_{t_k}}} Z_{v_{t_k}}. \end{aligned} \tag{C.30}$$

This representation is analysed as described in algorithm 2.

Algorithm 2 Calibrating the Heston parameters with an Unscented Kalman Filter

1: **procedure** UNSCENTED KALMAN FILTER
2: Define initial mean μ_0, parameter set $\Omega_{0|0}$ and correlation matrix $P_{0|0}$.
3: **for** $k \in \{1, \ldots, N\}$ **do**
4: Evaluate $\Psi_{\text{IC}}\left(\Omega_{k-1|k-1}, P_{k-1|k-1}, n, \lambda, u, l\right)$, as defined in (C.19)-(C.24), to determine the sigma points and their associated weights.
5: Propagate the sigma points through state function $f(\cdot)$, as defined in (C.27).
6: Determine the prior parameter set $\Omega_{k|k-1}$ and covariance $P_{k|k-1}$ with (C.12) and (C.13).
7: Update the sigma points according to the prior estimates with $\Psi_{\text{IC}}\left(\Omega_{k|k-1}, P_{k|k-1}, n, \lambda, u, l\right)$.
8: Determine the prior measurement following (C.15), with $h(\cdot)$ as defined in (C.26).
9: Evaluate the Kalman gain through (C.16).
10: Obtain the posterior estimates $\Omega_{k|k}$, $P_{k|k}$ with (C.18).
11: Transform the parameter set $\Omega_{k|k}$ with (C.25), such that the boundary constraints are satisfied.
12: Determine the posterior measurement estimate $\sigma_{k|k}$ by substituting the transformed parameter set $\Omega_{k|k}$ in (C.18).

C.3. Calibrating auxiliary parameters

The approach described in the previous sections contains many auxiliary parameters that require calibration beforehand, for example λ, δ and $\Gamma_{p,q}$ for each parameter combination p and q. The calibration procedure is similar to the VIX Heston model calibration. Fixing the auxiliary parameters fully defines the Heston model parameters, $\Omega_{k|k}^{\text{Heston}}(X_{\text{aux}})$, for each step k. Thus, one can find the optimal auxiliary parameters by performing a minimization procedure,

$$X_{\text{aux}} = \underset{X^{\text{Search}}}{\arg\min} \left(\sum_k \sum_K \sum_T \left(\sigma^{\text{Market}}(t_k, K, T) - \sigma^{\text{Heston}}(t_k, \Omega_{k|k}^{\text{Heston}}(X^{\text{Search}}), K, T) \right)^2 \right). \tag{C.31}$$

Again, we use the Levenberg-Marquardt least-squares algorithm to find the (local) optimum. In this application, it can be difficult for the algorithm to find a 'good' local optimum, due to a highly oscillating error function. However, sampling many initial values still leads to a good approximation of the global optimum, as is described in, for example, Gavin (2017).

C.4. Results

In this section we will present results of the unscented Kalman filter approach, in terms of accuracy and calibration time. We will distinguish two models: models using the regression functions presented in chapter 3 and a model that does not utilize any of the regression models. The latter serves as a benchmark model: it shows whether adding market information will improve the accuracy. Finally, we will

discuss the advantages of this approach, in terms of calibration time.

Throughout this entire chapter we will assume $\Delta t = 1$ month, for computational purposes. Moreover, the data set is separated into a training and test set, as discussed in section 3.4.1. The training set is used to perform the calibration procedure described in section C.3 and the test set is used to evaluate the accuracy of the models.

C.4.1. Benchmark model

Recall the general SDE representation for parameter p

$$\mathrm{d}p_t = \mu_p(t, p_t)\mathrm{d}t + \sigma_p(t, p_t)\mathrm{d}W_t^p. \tag{C.32}$$

Throughout the entire analysis we will assume for each parameter

$$\sigma_p(t, p_t) = \alpha_p p_t, \tag{C.33}$$

with α_p the constant volatility of parameter p. Moreover, the benchmark model assumes a mean reverting drift component,

$$\mu_p(t, p_t) = \kappa_p(\bar{p} - p_t). \tag{C.34}$$

This mean reverting process is in line with historical observations, as all parameters appear to be evolving around a certain long term mean. Using these assumptions, we can represent the stochastic processes as

$$\begin{cases} \mathrm{d}v_{0,t} &= \kappa_{v_0}(\bar{v_0} - v_{0,t})\mathrm{d}t + \alpha_{v_0} v_{0,t}\mathrm{d}W_t^{v_0}, \\ \mathrm{d}\bar{v}_t &= \kappa_{\bar{v}}(\bar{\bar{v}} - \bar{v}_t)\mathrm{d}t + \alpha_{\bar{v}}\bar{v}_t\mathrm{d}W_t^{\bar{v}}, \\ \mathrm{d}\gamma_t &= \kappa_\gamma(\bar{\gamma} - \gamma_t)\mathrm{d}t + \alpha_\gamma\gamma_t\mathrm{d}W_t^{\gamma}, \\ \mathrm{d}\rho_t &= \kappa_\rho(\bar{\rho} - \rho_t)\mathrm{d}t + \alpha_\rho\rho_t\mathrm{d}W_t^{\rho}. \end{cases} \tag{C.35}$$

A drawback of this approach is that it results in many auxiliary parameters that require calibration beforehand. However, the auxiliary parameters only require one calibration procedure for the entire data set, as described in section C.3. The results of this calibration procedure can be found in appendix C.4.5.

C.4.2. Hybrid model

In the previous section we discussed a model where the Heston parameters are assumed to follow a mean reverting path. This model does not depend on any market data other than the implied volatility surface. In this section, we propose a hybrid model, which implements the results of section 3 into the unscented Kalman filter. This way, we derive relevant information from both the implied volatility surface and the market indices[3]. In this section we will make the following assumptions:

- The deterministic increments of $v_{0,t}$, $\bar{v}_t$, γ_t are equal to the increments of the regression models proposed in chapter 3.
- The correlation coefficient ρ follows a mean reverting path.
- All parameters are driven by a stochastic noise term equal to (C.33).

This translates to

$$\begin{array}{ccc} \text{Regression model} & & \text{Unscented Kalman filter} \\ v_{0,t}^* = f_{v_0}(t,\cdot) & \rightarrow & \mathrm{d}v_{0,t} = \mathrm{d}v_{0,t}^* + \alpha_{v_0} v_{0,t}\mathrm{d}W_t^{v_0}, \\ \bar{v}_t^* = f_{\bar{v}}(t,\cdot) & \rightarrow & \mathrm{d}\bar{v}_t = \mathrm{d}\bar{v}_t^* + \alpha_{\bar{v}}\bar{v}_t\mathrm{d}W_t^{\bar{v}}, \\ \gamma_t^* = f_\gamma(t,\cdot) & \rightarrow & \mathrm{d}\gamma_t = \mathrm{d}\gamma_t^* + \alpha_\gamma\gamma_t\mathrm{d}W_t^{\gamma}, \\ \rho_t^* = \bar{\rho} & \rightarrow & \mathrm{d}\rho_t = \kappa_\rho(\bar{\rho} - \rho_t)\mathrm{d}t + \alpha_\rho\rho_t\mathrm{d}W_t^{\rho}, \end{array} \tag{C.36}$$

where $f_p(t,\cdot)$ is defined as the regression model for parameter p at time t. We will test this approach for all regression models proposed in chapter 3. After discretization, we define the deterministic increments for parameter p as

$$\mathrm{d}p_{t_k}^* = f_p(t_k,\cdot) - f_p(t_{k-1},\cdot). \tag{C.37}$$

[3] In this case, the relevant market features include the VIX index and the long-term mean r_{720}.

The auxiliary parameters are calibrated as described in section C.3 and the optimal auxiliary parameters of the benchmark model are estimated as

$$\begin{cases} \alpha_{v_0} = 0.0031, & \alpha_{\bar{v}} = 8.668 \cdot 10^{-5}, & \alpha_\gamma = 0.0051, & \alpha_\rho = 0.0051, & \kappa_{v_0} = 0.1369, \\ \kappa_{\bar{v}} = 0.1153, & \kappa_\gamma = 0.1765, & \kappa_\rho = 0.0204, & \bar{v}_0 = 0.0517, & \bar{\bar{v}} = 0.1043, \\ \bar{\gamma} = 0.6786, & \bar{\rho} = -0.7495, & \Gamma_{v_0,\bar{v}} = 0.9981, & \Gamma_{v_0,\gamma} = 0.9609, & \Gamma_{v_0,\rho} = -0.08, \\ \Gamma_{\bar{v},\gamma} = 0.9331, & \Gamma_{\bar{v},\rho} = 0.02, & \Gamma_{\gamma,\rho} = 0.22, & \lambda = 1.0932, & \delta = 2.1224 \cdot 10^{-8}. \end{cases} \tag{C.38}$$

The optimal auxiliary parameters of the hybrid model are estimated as

$$\begin{cases} \alpha_{v_0} = 4.205 \cdot 10^{-10}, & \alpha_{\bar{v}} = 2.381 \cdot 10^{-8}, & \alpha_\gamma = 5.905 \cdot 10^{-5}, & \alpha_\rho = 1.696 \cdot 10^{-6}, & \kappa_\rho = 0.0147, \\ \bar{\rho} = -0.7495, & \Gamma_{v_0,\bar{v}} = 0.9981, & \Gamma_{v_0,\gamma} = 0.9119, & \Gamma_{v_0,\rho} = 0.0000, & \Gamma_{\bar{v},\gamma} = 0.9345, \\ \Gamma_{\bar{v},\rho} = 0.0000, & \Gamma_{\gamma,\rho} = 0.0000, & \lambda = 1.2008, & \delta = 1.000 \cdot 10^{-14}. \end{cases} \tag{C.39}$$

The noise terms of the hybrid model are significantly smaller than the noise terms of the benchmark model. This can be explained by the more accurate drift terms, which are based on the regression models instead of a mean reverting process. The other hyper parameters are, as expected, comparable to the benchmark model.

C.4.3. Accuracy

After identifying the optimal set-up for both the benchmark and hybrid models, one is able to estimate the Heston parameters and the accompanied implied volatility surfaces of the test set. Using the error measures defined in (3.35) and (3.36) we will assess the performance of the benchmark model and compare it to the hybrid models, which are based on the regression functions established in chapter 3. The results of the US data set can be found in table C.1 and figure C.2. The results of the Europe and UK data sets can be found in section C.4.5.

Model	SSE	MAE	R^2	MSE_{v_0}	$\text{MSE}_{\bar{v}}$	MSE_γ	MSE_ρ
Benchmark	0.0491	0.0057	0.9763	$5.598 \cdot 10^{-5}$	$1.187 \cdot 10^{-4}$	0.0119	0.0022
VIX Heston	0.0198	0.0034	0.9905	$1.207 \cdot 10^{-6}$	$5.053 \cdot 10^{-6}$	0.0012	$7.003 \cdot 10^{-4}$
Polynomial Regression	0.0192	0.0034	0.9907	$1.409 \cdot 10^{-6}$	$5.940 \cdot 10^{-6}$	0.0012	$6.923 \cdot 10^{-4}$
Multi-output SVR	0.0186	0.0034	0.9911	$1.503 \cdot 10^{-6}$	$8.054 \cdot 10^{-6}$	0.0014	$6.914 \cdot 10^{-4}$
GB regression trees	0.0189	0.0034	0.9909	$7.549 \cdot 10^{-7}$	$6.462 \cdot 10^{-6}$	0.0010	$6.954 \cdot 10^{-4}$
Unrestricted	*0.0173*	*0.0034*	*0.9916*	-	-	-	-

Table C.1: Out-of-sample accuracy of the unscented Kalman filter according to the error measures defined in (3.35) and (3.36).

The accuracy of the hybrid models is comparable to the unrestricted model, as the mean absolute error and sum of squared errors are almost equal. Moreover, it seems that adding market information, by implementing the regression models, leads to improved performance. In every data set[4], the benchmark model is outperformed by all hybrid models in terms of accuracy. The benchmark model assumes mean reverting drift terms, hence all prior estimates will predict convergence to the long term mean, which does not always coincide with the actual market development. Consequently, the prior errors of this model will be relatively large, making it more difficult to make an accurate posterior estimation. The hybrid models, however, do incorporate development of the market. This leads to more accurate prior and posterior predictions.

The robustness of this approach is noteworthy. Not only in terms of different data sets, but also in terms of different regression models. The unscented Kalman filter produces accurate estimates for all hybrid models in all data sets, despite differences in the regression models. We can conclude that small errors in the prior predictions (regression models) have little to no effect on the posterior estimates, as these differences are thought of as noise and nullified by the Kalman gain, which depends on the observed implied volatility surface.

[4]US, UK and Europe

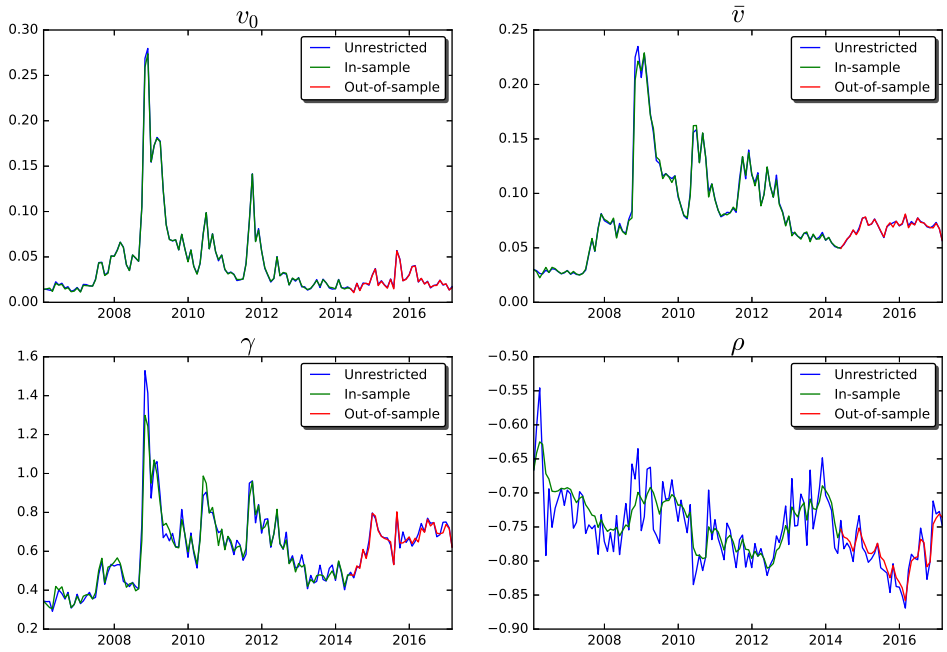

Figure C.2: Development of the Heston parameters according to the unrestricted model and the hybrid unscented Kalman filter.

Finally, this approach appears to be insusceptible for overfitting: the in-sample accuracy is comparable to the out-of-sample accuracy. This phenomenon is caused by two key features of the hybrid model. First of all, the use of regression models, with fairly accurate out-of-sample predictions (see chapter 3), improve the accuracy in-sample as well as out-of-sample. Moreover, the posterior predictions use the *observed* implied volatility surface, which contains all the information required for calibration, as the Heston parameters are defined according to this surface. However, the unscented Kalman filter is not powerful enough to rely solely on the implied volatility surface, which is indicated by the results of the benchmark model.

C.4.4. Calibration time

The unscented Kalman filter approach calibrates over 3 times faster than the Levenberg-Marquardt minimizer in the current set-up. In the current model, we estimate four Heston parameters and we assume κ to be constant. One can imagine that the calibration time increases (decreases), when more (fewer) parameters are estimated. To this end, we will explore the relation between calibration time and the number of calibrated parameters. In this test we will calibrate x parameters according to the unscented Kalman filter approach and the Levenberg-Marquardt minimizer, with $x \in \{1, \ldots, 5\}$. The other parameters are fixed. The calibration times of the US data set are graphically presented in figure C.3.

By increasing the number of parameters, one increases the size of the search space of the Levenberg-Marquardt minimizer. Consequently, more function evaluations are required to find the optimum, which increases calibration time. As the number of parameters grows, this effect becomes even more pronounced, since the size of the search space increases exponentially. This explains the non-linear behaviour of the Levenberg-Marquardt algorithm in figure C.3.

On the other hand, the calibration time of the unscented Kalman filter approach increases linearly with the number of calibrated parameters. Adding one parameter leads to two additional sigma points, resulting in a linear behaviour, as can be seen in figure C.3. Moreover, the loss of accuracy is negligible for all models. Thus, we can conclude that the unscented Kalman filter approach is not only faster than the Levenberg-Marquardt minimizer, its efficiency is even more pronounced in complex models (with more parameters).

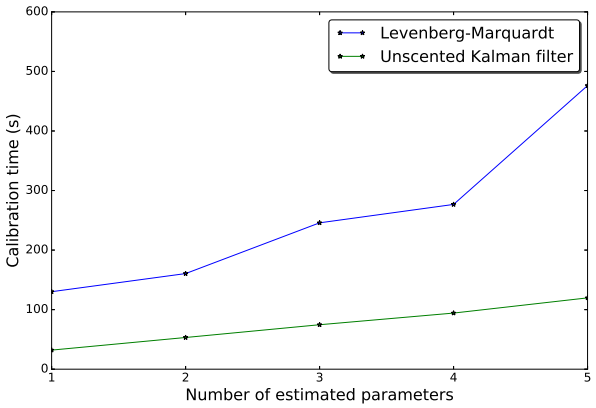

Figure C.3: Calibration times for different number of US Heston parameters with the unscented Kalman filter approach and Levenberg-Marquardt minimizer.

C.4.5. Results Europe and UK Data sets

UK

The optimal auxiliary parameters of the hybrid model are estimated as

$$\begin{cases} \alpha_{v_0} = 5.287 \cdot 10^{-7}, & \alpha_{\bar{v}} = 2.708 \cdot 10^{-7}, & \alpha_{\gamma} = 6.939 \cdot 10^{-5}, & \alpha_{\rho} = 2.976 \cdot 10^{-5}, & \kappa_{\rho} = 0.0471, \\ \bar{\rho} = -0.6400, & \Gamma_{v_0,\bar{v}} = 0.9577, & \Gamma_{v_0,\gamma} = 0.9091, & \Gamma_{v_0,\rho} = 0.2000, & \Gamma_{\bar{v},\gamma} = 0.9316, \\ \Gamma_{\bar{v},\rho} = -0.1000, & \Gamma_{\gamma,\rho} = -0.1000, & \lambda = 1.1956, & \delta = 1.160 \cdot 10^{-13}. & \end{cases} \tag{C.40}$$

The regression results of this set-up are presented in table C.2 and figure C.4.

Model	SSE	MAE	R^2	MSE_{v_0}	$\text{MSE}_{\bar{v}}$	MSE_{γ}	MSE_{ρ}
VIX Heston	0.0137	0.0046	0.9843	$1.506 \cdot 10^{-6}$	$5.086 \cdot 10^{-6}$	$8.875 \cdot 10^{-4}$	$2.898 \cdot 10^{-4}$
Polynomial Regression	0.0137	0.0046	0.9843	$1.506 \cdot 10^{-6}$	$5.086 \cdot 10^{-6}$	$8.875 \cdot 10^{-4}$	$2.898 \cdot 10^{-4}$
Multi-output SVR	0.0135	0.0046	0.9846	$1.423 \cdot 10^{-6}$	$5.296 \cdot 10^{-6}$	$9.006 \cdot 10^{-4}$	$2.810 \cdot 10^{-4}$
Unrestricted	*0.0126*	*0.0045*	*0.9854*	-	-	-	-

Table C.2: UK out-of-sample accuracy of the unscented Kalman filter according to the error measures defined in (3.35) and (3.36).

Europe

The optimal auxiliary parameters of the hybrid model are estimated as

$$\begin{cases} \alpha_{v_0} = 5.287 \cdot 10^{-7}, & \alpha_{\bar{v}} = 2.708 \cdot 10^{-7}, & \alpha_{\gamma} = 6.939 \cdot 10^{-5}, & \alpha_{\rho} = 2.976 \cdot 10^{-5}, & \kappa_{\rho} = 0.0471, \\ \bar{\rho} = -0.6400, & \Gamma_{v_0,\bar{v}} = 0.9577, & \Gamma_{v_0,\gamma} = 0.9091, & \Gamma_{v_0,\rho} = 0.2000, & \Gamma_{\bar{v},\gamma} = 0.9316, \\ \Gamma_{\bar{v},\rho} = -0.1000, & \Gamma_{\gamma,\rho} = -0.1000, & \lambda = 1.1956, & \delta = 1.150 \cdot 10^{-13}. & \end{cases} \tag{C.41}$$

The regression results of this set-up are presented in table C.3 and figure C.5.

Model	SSE	MAE	R^2	MSE_{v_0}	$\text{MSE}_{\bar{v}}$	MSE_{γ}	MSE_{ρ}
VIX Heston	0.0176	0.0053	0.9752	$1.534 \cdot 10^{-6}$	$9.245 \cdot 10^{-6}$	0.0018	$5.189 \cdot 10^{-4}$
Polynomial Regression	0.0176	0.0053	0.9752	$1.341 \cdot 10^{-6}$	$7.867 \cdot 10^{-6}$	0.0016	$4.828 \cdot 10^{-4}$
Multi-output SVR	0.0160	0.0051	0.9775	$1.142 \cdot 10^{-6}$	$8.934 \cdot 10^{-6}$	0.0015	$4.671 \cdot 10^{-4}$
Unrestricted	*0.0154*	*0.0051*	*0.9779*	-	-	-	-

Table C.3: Europe out-of-sample accuracy of the unscented Kalman filter according to the error measures defined in (3.35) and (3.36).

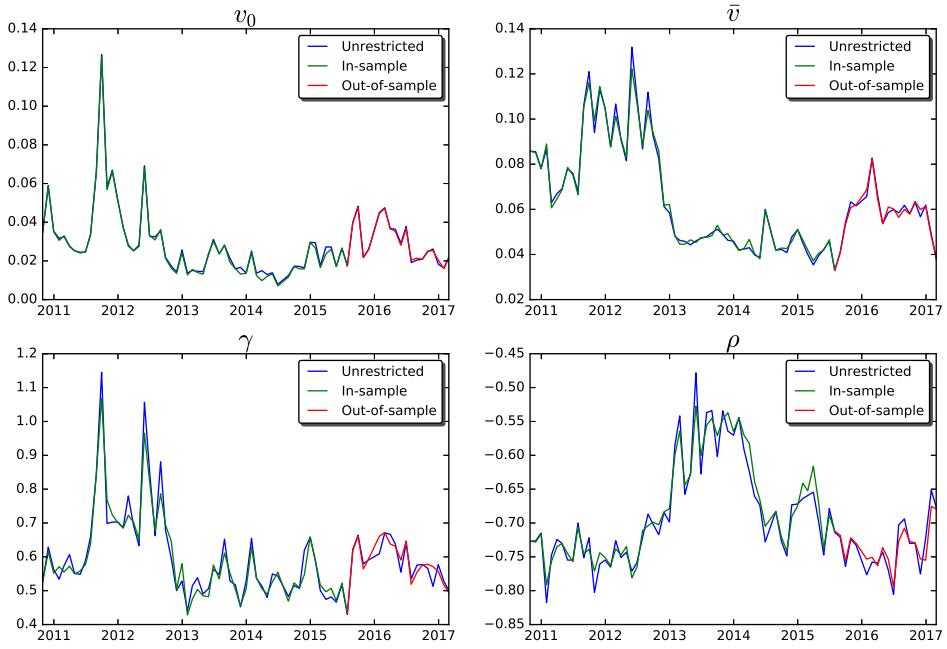

Figure C.4: Development of the UK Heston parameters according to the unrestricted model and the hybrid unscented Kalman filter.

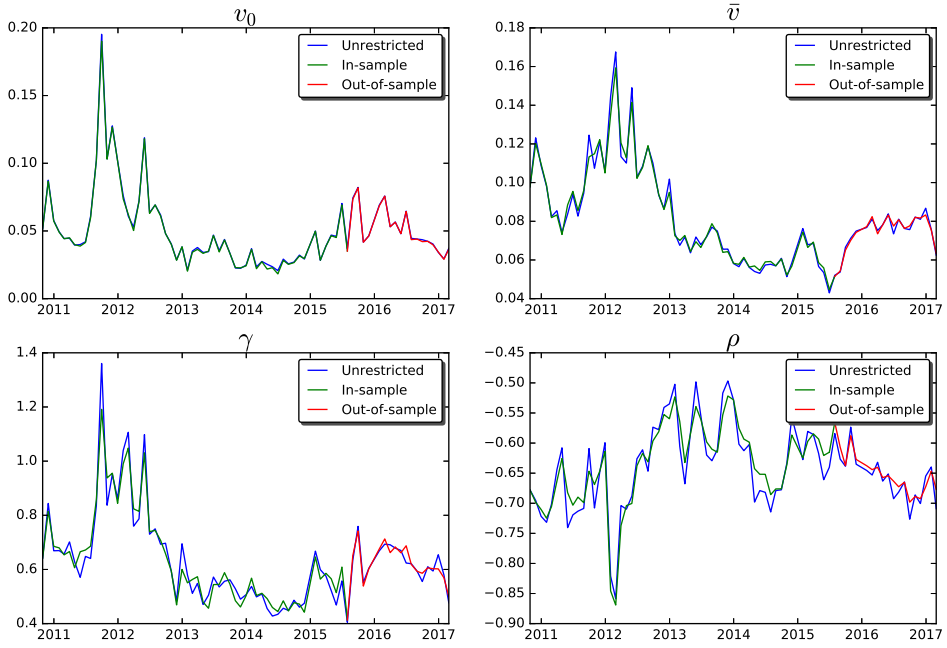

Figure C.5: Development of the Europe Heston parameters according to the unrestricted model and the hybrid unscented Kalman filter.

D

Hedging ratios

D.1. Covariation

Formally, the covariance of two processes X_t and Y_t is defined as

$$\langle X,Y\rangle_t = \lim_{||P||\to 0}\sum_{k=1}^{P_n}(X_{t_k}-X_{t_{k-1}})(Y_{t_k}-Y_{t_{k-1}}). \tag{D.1}$$

Where P is the partition of interval $[0,t]$ and P_n is the number of elements in this partition. Hence,

$$||P||\to 0 \Rightarrow P_n\to\infty. \tag{D.2}$$

This operator has the following properties

- $\langle cX,Y\rangle_t = c\langle X,Y\rangle_t$, for constant c and processes X_t and Y_t.
- $\langle X+Z,Y\rangle_t = \langle X,Y\rangle_t + \langle Z,Y\rangle_t$, for processes X_t, Y_t and Z_t.
- $\langle \mathrm{d}W_t^1, \mathrm{d}W_t^1\rangle = \mathrm{d}t$, for Brownian motion W_t^1.
- $\langle \mathrm{d}W_t^1, \mathrm{d}W_t^2\rangle = \rho\mathrm{d}t$, for Brownian motions W_t^1 and W_t^2 with correlation ρ.
- $\langle \mathrm{d}W_t^1, \mathrm{d}t\rangle = 0$, for Brownian motion W_t^1.

The first two properties can be easily derived from the operator's definition and latter three properties follow from the definition of the Brownian motion.

D.2. Option price dynamics

We can apply Ito's lemma to define the Heston option price dynamics:

$$\mathrm{d}C_t^{\text{Heston}} = \frac{\partial C}{\partial t}\mathrm{d}t + \frac{\partial C}{\partial S_t}\mathrm{d}S_t + \frac{\partial C}{\partial v_t}\mathrm{d}v_t + \frac{1}{2}\frac{\partial^2 C}{\partial S_t^2}\langle \mathrm{d}S_t,\mathrm{d}S_t\rangle + \frac{1}{2}\frac{\partial^2 C}{\partial v_t^2}\langle \mathrm{d}v_t,\mathrm{d}v_t\rangle + \frac{\partial^2 C}{\partial S_t\partial v_t}\langle \mathrm{d}S_t,\mathrm{d}v_t\rangle. \tag{D.3}$$

Now we can substitute (4.22) into (D.3), to obtain

$$\begin{aligned}
&= \frac{\partial C}{\partial t}\mathrm{d}t + \frac{\partial C}{\partial S_t}\left(rS_t\mathrm{d}t + \sqrt{v_t}S_t\mathrm{d}W_t^1\right) + \frac{\partial C}{\partial v_t}\left(\kappa(\bar v - v_t)\mathrm{d}t + \gamma\sqrt{v_t}\mathrm{d}W_t^2\right) + \frac{1}{2}\frac{\partial^2 C}{\partial S_t^2}v_tS_t^2\mathrm{d}t \\
&\quad + \frac{1}{2}\frac{\partial^2 C}{\partial v_t^2}\gamma^2 v_t\mathrm{d}t + \frac{\partial^2 C}{\partial S_t\partial v_t}\rho\gamma v_tS_t\mathrm{d}t \\
&= \left(\frac{\partial C}{\partial t} + rS_t\frac{\partial C}{\partial S_t} + \kappa(\bar v - v_t)\frac{\partial C}{\partial v_t} + \frac{1}{2}v_tS_t^2\frac{\partial^2 C}{\partial S_t^2} + \frac{1}{2}\gamma^2 v_t\frac{\partial^2 C}{\partial v_t^2} + \rho\gamma v_tS_t\frac{\partial^2 C}{\partial S_t\partial v_t}\right)\mathrm{d}t \\
&\quad + \sqrt{v_t}S_t\frac{\partial C}{\partial S_t}\mathrm{d}W_t^1 + \gamma\sqrt{v_t}\frac{\partial C}{\partial v_t}\mathrm{d}W_t^2.
\end{aligned} \tag{D.4}$$

We can slightly rewrite this expression by assuming two independent Brownian motions W_t^S and W_t^v defined through

$$\begin{cases} W_t^1 = W_t^S, \\ W_t^2 = \rho W_t^S + \sqrt{1-\rho^2} W_t^v. \end{cases} \tag{D.5}$$

Substituting this dependence structure in (D.4) gives the option price dynamics under the Heston model

$$\begin{aligned} \mathrm{d}C_t^{\text{Heston}} = &\left(\frac{\partial C}{\partial t} + rS_t \frac{\partial C}{\partial S_t} + \kappa(\bar{v} - v_t) \frac{\partial C}{\partial v_t} + \frac{1}{2} v_t S_t^2 \frac{\partial^2 C}{\partial S_t^2} + \frac{1}{2} \gamma^2 v_t \frac{\partial^2 C}{\partial v_t^2} + \rho\gamma v_t S_t \frac{\partial^2 C}{\partial S_t \partial v_t} \right) \mathrm{d}t \\ &+ \left(\sqrt{v_t} S_t \frac{\partial C}{\partial S_t} + \rho\gamma \sqrt{v_t} \frac{\partial C}{\partial v_t} \right) \mathrm{d}W_t^S + \gamma \sqrt{v_t(1-\rho^2)} \frac{\partial C}{\partial v_t} \mathrm{d}W_t^v. \end{aligned} \tag{D.6}$$

D.3. Delta hedge portfolio dynamics

Using the option price dynamics given in (4.25) and the definition of Δ_t given by (4.26), we can derive the Heston Delta hedge portfolio dynamics

$$\begin{aligned} \mathrm{d}\Pi_t &= -\mathrm{d}C_t + \Delta_{C_t^{\text{Heston}}} \mathrm{d}S_t + \mathrm{d}B_t \\ &= -\mathrm{d}C_t + \left(\frac{\partial C}{\partial S_t} + \rho\gamma \frac{1}{S_t} \frac{\partial C}{\partial v_t} \right) \mathrm{d}S_t + \mathrm{d}B_t \\ &= -\left(\frac{\partial C}{\partial t} + rS_t \frac{\partial C}{\partial S_t} + \kappa(\bar{v} - v_t) \frac{\partial C}{\partial v_t} + \frac{1}{2} v_t S_t^2 \frac{\partial^2 C}{\partial S_t^2} + \frac{1}{2} \gamma^2 v_t \frac{\partial^2 C}{\partial v_t^2} + \rho\gamma v_t S_t \frac{\partial^2 C}{\partial S_t \partial v_t} \right) \mathrm{d}t \\ &\quad - \left(\sqrt{v_t} S_t \frac{\partial C}{\partial S_t} + \rho\gamma \sqrt{v_t} \frac{\partial C}{\partial v_t} \right) \mathrm{d}W_t^S - \gamma \sqrt{v_t(1-\rho^2)} \frac{\partial C}{\partial v_t} \mathrm{d}W_t^v \\ &\quad + \left(\frac{\partial C}{\partial S_t} + \rho\gamma \frac{1}{S_t} \frac{\partial C}{\partial v_t} \right) \left(rS_t \mathrm{d}t + \sqrt{v_t} S_t \mathrm{d}W_t^S \right) + rB_t \mathrm{d}t \\ &= -\left(\frac{\partial C}{\partial t} + \kappa(\bar{v} - v_t - r\rho\gamma) \frac{\partial C}{\partial v_t} + \frac{1}{2} v_t S_t^2 \frac{\partial^2 C}{\partial S_t^2} + \frac{1}{2} \gamma^2 v_t \frac{\partial^2 C}{\partial v_t^2} + \rho\gamma v_t S_t \frac{\partial^2 C}{\partial S_t \partial v_t} - rB_t \right) \mathrm{d}t \\ &\quad - \gamma \sqrt{v_t(1-\rho^2)} \frac{\partial C}{\partial v_t} \mathrm{d}W_t^v. \end{aligned} \tag{D.7}$$

D.4. Delta-Vega hedge portfolio dynamics

For notational purposes we first define

$$b_1(C_t) = \frac{\partial C}{\partial t} + rS_t \frac{\partial C}{\partial S_t} + \kappa(\bar{v} - v_t) \frac{\partial C}{\partial v_t} + \frac{1}{2} v_t S_t^2 \frac{\partial^2 C}{\partial S_t^2} + \frac{1}{2} \gamma^2 v_t \frac{\partial^2 C}{\partial v_t^2} + \rho\gamma v_t S_t \frac{\partial^2 C}{\partial S_t \partial v_t}, \tag{D.8}$$

$$b_2(C_t) = \sqrt{v_t} S_t \frac{\partial C}{\partial S_t} + \rho\gamma \sqrt{v_t} \frac{\partial C}{\partial v_t}, \tag{D.9}$$

$$b_3(C_t) = \gamma \sqrt{v_t(1-\rho^2)} \frac{\partial C}{\partial v_t}. \tag{D.10}$$

Note that

$$\mathrm{d}C_t = b_1(C_t)\mathrm{d}t + b_2(C_t)\mathrm{d}W_t^S + b_3(C_t)\mathrm{d}W_t^v. \tag{D.11}$$

Now, by substituting the definitions we can derive

$$\begin{aligned} \mathrm{d}\Pi_t &= -\mathrm{d}C_t + \bar{\Delta}^{(1)}_{C_t^{\text{Heston}}} \mathrm{d}S_t + \bar{\Delta}^{(2)}_{C_t^{\text{Heston}}} \mathrm{d}\tilde{C}_t + \mathrm{d}B_t \\ &= -\mathrm{d}C_t + \left(\frac{\partial C}{\partial S_t} - \bar{\Delta}^{(2)}_{C_t^{\text{Heston}}} \frac{\partial \tilde{C}}{\partial S_t} \right) \mathrm{d}S_t + \bar{\Delta}^{(2)}_{C_t^{\text{Heston}}} \mathrm{d}\tilde{C}_t + \mathrm{d}B_t \\ &= -b_1(C_t)\mathrm{d}t - b_2(C_t)\mathrm{d}W_t^S - b_3(C_t)\mathrm{d}W_t^v + \left(\frac{\partial C}{\partial S_t} - \bar{\Delta}^{(2)}_{C_t^{\text{Heston}}} \frac{\partial \tilde{C}}{\partial S_t} \right) \left(rS_t \mathrm{d}t + \sqrt{v_t} S_t \mathrm{d}W_t^S \right) \\ &\quad + \bar{\Delta}^{(2)}_{C_t^{\text{Heston}}} \left[b_1(\tilde{C}_t)\mathrm{d}t + b_2(\tilde{C}_t)\mathrm{d}W_t^S + b_3(\tilde{C}_t)\mathrm{d}W_t^v \right] + rB_t \mathrm{d}t. \end{aligned} \tag{D.12}$$

First, we collect the $\mathrm{d}t$ terms

$$\begin{aligned}
b_4(C_t,\tilde{C}_t) &:= -b_1(C_t) + \bar{\Delta}^{(2)}_{C_t^{\text{Heston}}} b_1(\tilde{C}_t) + rS_t\frac{\partial C}{\partial S_t} - rS_t\bar{\Delta}^{(2)}_{C_t^{\text{Heston}}}\frac{\partial \tilde{C}}{\partial S_t} + rB_t \\
&= -\frac{\partial C}{\partial t} - rS_t\frac{\partial C}{\partial S_t} - \kappa(\bar{v}-v_t)\frac{\partial C}{\partial v_t} - \frac{1}{2}v_tS_t^2\frac{\partial^2 C}{\partial S_t^2} - \frac{1}{2}\gamma^2 v_t\frac{\partial^2 C}{\partial v_t^2} - \rho\gamma v_tS_t\frac{\partial^2 C}{\partial S_t\partial v_t} \\
&\quad + \bar{\Delta}^{(2)}_{C_t^{\text{Heston}}}\left(\frac{\partial \tilde{C}}{\partial t} + rS_t\frac{\partial \tilde{C}}{\partial S_t} + \kappa(\bar{v}-v_t)\frac{\partial \tilde{C}}{\partial v_t} + \frac{1}{2}v_tS_t^2\frac{\partial^2 \tilde{C}}{\partial S_t^2} + \frac{1}{2}\gamma^2 v_t\frac{\partial^2 \tilde{C}}{\partial v_t^2} + \rho\gamma v_tS_t\frac{\partial^2 \tilde{C}}{\partial S_t\partial v_t}\right) \\
&\quad + rS_t\frac{\partial C}{\partial S_t} - rS_t\bar{\Delta}^{(2)}_{C_t^{\text{Heston}}}\frac{\partial \tilde{C}}{\partial S_t} + rB_t \\
&= -\frac{\partial C}{\partial t} - \frac{1}{2}v_tS_t^2\frac{\partial^2 C}{\partial S_t^2} - \frac{1}{2}\gamma^2 v_t\frac{\partial^2 C}{\partial v_t^2} - \rho\gamma v_tS_t\frac{\partial^2 C}{\partial S_t\partial v_t} + \kappa(\bar{v}-v_t)\left(-\frac{\partial C}{\partial v_t} + \frac{\frac{\partial C}{\partial v_t}}{\frac{\partial \tilde{C}}{\partial v_t}}\frac{\partial \tilde{C}}{\partial v_t}\right) \\
&\quad + \bar{\Delta}^{(2)}_{C_t^{\text{Heston}}}\left(\frac{\partial C}{\partial t} + \frac{1}{2}v_tS_t^2\frac{\partial^2 C}{\partial S_t^2} + \frac{1}{2}\gamma^2 v_t\frac{\partial^2 C}{\partial v_t^2} + \rho\gamma v_tS_t\frac{\partial^2 C}{\partial S_t\partial v_t}\right) + rB_t \\
&= -\frac{\partial C}{\partial t} - \frac{1}{2}v_tS_t^2\frac{\partial^2 C}{\partial S_t^2} - \frac{1}{2}\gamma^2 v_t\frac{\partial^2 C}{\partial v_t^2} - \rho\gamma v_tS_t\frac{\partial^2 C}{\partial S_t\partial v_t} \\
&\quad + \bar{\Delta}^{(2)}_{C_t^{\text{Heston}}}\left(\frac{\partial C}{\partial t} + \frac{1}{2}v_tS_t^2\frac{\partial^2 C}{\partial S_t^2} + \frac{1}{2}\gamma^2 v_t\frac{\partial^2 C}{\partial v_t^2} + \rho\gamma v_tS_t\frac{\partial^2 C}{\partial S_t\partial v_t}\right) + rB_t. \qquad \text{(D.13)}
\end{aligned}$$

Similarly, we can collect the $\mathrm{d}W_t^S$ terms

$$\begin{aligned}
b_5(C_t,\tilde{C}_t) &:= -b_2(C_t) + \bar{\Delta}^{(2)}_{C_t^{\text{Heston}}} b_2(\tilde{C}_t) + \sqrt{v_t}S_t\frac{\partial C_t}{\partial S_t} - \sqrt{v_t}S_t\bar{\Delta}^{(2)}_{C_t^{\text{Heston}}}\frac{\partial \tilde{C}_t}{\partial S_t} \\
&= -\sqrt{v_t}S_t\frac{\partial C}{\partial S_t} - \rho\gamma\sqrt{v_t}\frac{\partial C}{\partial v_t} + \frac{\frac{\partial C}{\partial v_t}}{\frac{\partial \tilde{C}}{\partial v_t}}\left(\sqrt{v_t}S_t\frac{\partial \tilde{C}}{\partial S_t} + \rho\gamma\sqrt{v_t}\frac{\partial \tilde{C}}{\partial v_t}\right) + \sqrt{v_t}S_t\frac{\partial C_t}{\partial S_t} - \sqrt{v_t}S_t\frac{\frac{\partial C}{\partial v_t}}{\frac{\partial \tilde{C}}{\partial v_t}}\frac{\partial \tilde{C}_t}{\partial S_t} \\
&= 0. \qquad \text{(D.14)}
\end{aligned}$$

Finally, we collect the $\mathrm{d}W_t^v$ terms

$$\begin{aligned}
b_6(C_t,\tilde{C}_t) &:= -b_3(C_t) + \bar{\Delta}^{(2)}_{C_t^{\text{Heston}}} b_3(\tilde{C}_t) \\
&= -\gamma\sqrt{v_t(1-\rho^2)}\frac{\partial C}{\partial v_t} + \frac{\frac{\partial C}{\partial v_t}}{\frac{\partial \tilde{C}}{\partial v_t}}\gamma\sqrt{v_t(1-\rho^2)}\frac{\partial \tilde{C}}{\partial v_t} \\
&= 0. \qquad \text{(D.15)}
\end{aligned}$$

Thus, we can conclude that

$$\begin{aligned}
\mathrm{d}\Pi_t &= b_4(C_t,\tilde{C}_t)\mathrm{d}t \\
&= \left(-\frac{\partial C}{\partial t} - \frac{1}{2}v_tS_t^2\frac{\partial^2 C}{\partial S_t^2} - \frac{1}{2}\gamma^2 v_t\frac{\partial^2 C}{\partial v_t^2} - \rho\gamma v_tS_t\frac{\partial^2 C}{\partial S_t\partial v_t}\right. \\
&\quad \left. + \bar{\Delta}^{(2)}_{C_t^{\text{Heston}}}\left(\frac{\partial C}{\partial t} + \frac{1}{2}v_tS_t^2\frac{\partial^2 C}{\partial S_t^2} + \frac{1}{2}\gamma^2 v_t\frac{\partial^2 C}{\partial v_t^2} + \rho\gamma v_tS_t\frac{\partial^2 C}{\partial S_t\partial v_t}\right) + rB_t\right)\mathrm{d}t. \qquad \text{(D.16)}
\end{aligned}$$

E

Simulation Procedures

In this chapter we will describe the simulation procedures to generate Black-Scholes, Heston and dynamic Heston markets. The markets will be simulated for interval $[0,T]$, which will be discretize into N time-intervals of size Δt, hence

$$N = \frac{T}{\Delta t}. \tag{E.1}$$

E.1. Black-Scholes

The Black-Scholes asset dynamics can be expressed analytically,

$$S_t = S_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t^S}. \tag{E.2}$$

Hence, simulating the Black-Scholes market is equivalent to generating Brownian motion paths distributed as

$$W_t^S \sim \mathcal{N}(0,t). \tag{E.3}$$

Here $\mathcal{N}(a,b)$ is a normal distribution with mean a and variance b.

E.2. Heston

In this section we will follow the Quadratic Exponential scheme proposed in Andersen (2008) (see page 19 with $\gamma_1 = 0.5$, $\gamma_2 = 0.5$). First we will rewrite the Heston dynamics

$$\begin{cases} \mathrm{d}X_t = (r - \frac{1}{2}v_t)\mathrm{d}t + \sqrt{v_t}\left(\rho \mathrm{d}W_t^v + \sqrt{1-\rho^2}\mathrm{d}W_t^S\right), \\ \mathrm{d}v_t = \kappa(\bar{v} - v_t)\mathrm{d}t + \gamma\sqrt{v_t}\mathrm{d}W_t^v. \end{cases} \tag{E.4}$$

With $X_t = \log(S_t)$. The integral expression is given by

$$\begin{cases} X_{t+\Delta t} = X_t + \int_t^{t+\Delta t} (r - \frac{1}{2}v_s)\mathrm{d}s + \rho \int_t^{t+\Delta t} \sqrt{v_t}\mathrm{d}W_t^v + \sqrt{1-\rho^2}\int_t^{t+\Delta t} \sqrt{v_s}\mathrm{d}W_s^S, \\ v_{t+\Delta t} = v_t + \int_t^{t+\Delta t} \kappa(\bar{v} - v_s)\mathrm{d}s + \gamma \int_t^{t+\Delta t} \sqrt{v_s}\mathrm{d}W_s^v. \end{cases} \tag{E.5}$$

By performing an Euler discretization one is able to evaluate these expressions. But too large values of Δt might produce negative values for v_t, which is impossible. However, as proven in Cox et al. (1985), v_t follows a non-central chi squared distribution defined as

$$\begin{cases} v_{t+\Delta t} \sim c_{\Delta t}\chi^2(d, \lambda(t, v_t)), \\ c_{\Delta t} = \frac{\gamma^2}{4\kappa}(1 - e^{-\kappa\Delta t}), \\ d = \frac{4\kappa\bar{v}}{\gamma^2}, \\ \lambda(t, v_t) = \frac{4\kappa e^{-\kappa\Delta t}}{\gamma^2(1-e^{-\kappa\Delta t})} v_t. \end{cases} \tag{E.6}$$

With $\chi^2(d,\lambda(t,v_t))$ the non-central chi-squared distribution with d degrees of freedom and non-centrality $\lambda(t,v_t)$. To prevent negative values of v_t we will draw $v_{t+\Delta t}$ from this distribution. Now, given v_t and $v_{t+\Delta t}$, we can derive

$$v_{t+\Delta t} = v_t + \int_t^{t+\Delta t} \kappa(\bar{v} - v_s)\mathrm{d}s + \gamma \int_t^{t+\Delta t} \sqrt{v_s}\mathrm{d}W_s^v$$

$$\Leftrightarrow,$$

$$\int_t^{t+\Delta t} \sqrt{v_s}\mathrm{d}W_s^v = \frac{1}{\gamma}\left(v_{t+\Delta t} - v_t - \int_t^{t+\Delta t} \kappa(\bar{v} - v_s)\mathrm{d}s\right). \tag{E.7}$$

By discretizing we can approximate this expression,

$$\int_t^{t+\Delta t} \sqrt{v_s}\mathrm{d}W_s^v \approx \frac{1}{\gamma}\left(v_{t+\Delta t} - v_t - \kappa\left(\bar{v} - \frac{v_t + v_{t+\Delta t}}{2}\right)\Delta t\right). \tag{E.8}$$

Substituting this in the integral expression for X_t and discretizing gives

$$\begin{aligned} X_{t+\Delta t} &\approx X_t + \int_t^{t+\Delta t} (r - \frac{1}{2}v_s)\mathrm{d}s + \frac{\rho}{\gamma}\left(v_{t+\Delta t} - v_t - \kappa\left(\bar{v} - \frac{v_t + v_{t+\Delta t}}{2}\right)\Delta t\right) + \sqrt{1-\rho^2}\int_t^{t+\Delta t} \sqrt{v_s}\mathrm{d}W_s^S \\ &\approx X_t + \left(r - \frac{v_t + v_{t+\Delta t}}{4}\right)\Delta t + \frac{\rho}{\gamma}\left(v_{t+\Delta t} - v_t - \kappa\left(\bar{v} - \frac{v_t + v_{t+\Delta t}}{2}\right)\Delta t\right) + \sqrt{\frac{v_t + v_{t+\Delta t}}{2}\Delta t(1-\rho^2)}Z^S. \end{aligned} \tag{E.9}$$

With independent standard normal distributed random variable Z^S. In conclusion, given v_t and S_t, one can generate $v_{t+\Delta t}$ and $S_{t+\Delta t}$ as follows,

Algorithm 3 Simulating a Heston market

procedure HESTON SIMULATION
  Define S_0 and v_0.
  Set $X_0 = \log(S_0)$.

  for $i \in \{0,\dots,N-1\}$ **do**
    Draw $v_{t+\Delta t}$ from $c_{\Delta t}\chi^2(d,\lambda(t,v_t))$ as defined in (E.6).
    Draw Z^S from an independent standard normal distribution.
    Calculate $X_{t+\Delta t}$ according to (E.9).
    Set $S_{t+\Delta t} = e^{X_{t+\Delta t}}$.

E.3. Dynamic Heston

We can rewrite the dynamic Heston model as follows

$$\begin{cases} \mathrm{d}X_t = (r - \frac{1}{2}v_t)\mathrm{d}t + \sqrt{v_t}\left(\rho\mathrm{d}W_t^v + \sqrt{1-\rho^2}\mathrm{d}W_t^S\right), \\ \mathrm{d}v_t = \kappa(\bar{v}_t - v_t)\mathrm{d}t + \gamma_t\sqrt{v_t}\mathrm{d}W_t^v, \\ \mathrm{d}\bar{v}_t = \kappa_{\bar{v}}(\bar{v}_{\text{Mean}} - \bar{v}_t)\mathrm{d}t + a_{\bar{v}}\bar{v}_t\left(\rho_{\bar{v}}\mathrm{d}W_t^v + \sqrt{1-\rho_{\bar{v}}^2}\mathrm{d}W_t^{\bar{v}}\right), \\ \mathrm{d}\gamma_t = \kappa_\gamma(\gamma_{\text{Mean}} - \gamma_t)\mathrm{d}t + a_\gamma\gamma_t\left(\rho_\gamma\mathrm{d}W_t^v + \sqrt{1-\rho_\gamma^2}\mathrm{d}W_t^\gamma\right). \end{cases} \tag{E.10}$$

Note that we can still generate v_t and S_t similar to the previous section, only with time-dependent parameters. The difficulty lies in simulating $\bar{v}_t$ and γ_t, such that they are correlated to v_t. First, we will discretize the processes

$$\begin{cases} v_{t+\Delta t} \approx v_t + \kappa\left(\bar{v}_t - \frac{v_t + v_{t\Delta t}}{2}\right)\Delta t + \gamma_t\sqrt{\frac{v_t + v_{t\Delta t}}{2}\Delta t}Z^v, \\ \bar{v}_{t+\Delta t} \approx \bar{v}_t + \kappa(\bar{v}_{\text{Mean}} - \bar{v}_t)\Delta t + a_{\bar{v}}\rho_{\bar{v}}\bar{v}_t\sqrt{\Delta t}Z^v + a_{\bar{v}}\bar{v}_t\sqrt{\Delta t(1-\rho_{\bar{v}}^2)}Z^{\bar{v}}, \\ \gamma_{t+\Delta t} \approx \gamma_t + \kappa(\gamma_{\text{Mean}} - \gamma_t)\Delta t + a_\gamma\rho_\gamma\gamma_t\sqrt{\Delta t}Z^v + a_\gamma\gamma_t\sqrt{\Delta t(1-\rho_\gamma^2)}Z^\gamma. \end{cases} \tag{E.11}$$

Where Z^v, $Z^{\bar{v}}$ and Z^γ are independent standard normal distributed random variables. Now we are able to derive an approximation for Z_v, given $v_{t+\Delta t}$,

$$\sqrt{\Delta t} Z^v \approx \frac{1}{\gamma_t \sqrt{\frac{v_t + v_{t+\Delta t}}{2}}} \left(v_{t+\Delta t} - v_t - \kappa \left(\bar{v}_t - \frac{v_t + v_{t+\Delta t}}{2} \right) \Delta t \right). \tag{E.12}$$

This approximation can be substituted in (E.11), which ensures the correlation between v_t, $\bar{v}_t$ and γ_t. In summary, the following steps must be taken to simulate a dynamic Heston market,

Algorithm 4 Simulating a dynamic Heston Market

procedure DYNAMIC HESTON SIMULATION
  Define S_0, v_0, $\bar{v}_0$ and γ_0.
  Set $X_0 = \log(S_0)$.

  for $i \in \{0, \ldots, N-1\}$ **do**
    Draw $v_{t+\Delta t}$ from $c_{\Delta t} \chi^2(d, \lambda(t, v_t))$ as defined in (E.6).
    Draw Z^S, $Z^{\bar{v}}$, Z^γ from independent standard normal distributions.
    Calculate $X_{t+\Delta t}$ according to (E.9).
    Set $S_{t+\Delta t} = e^{X_{t+\Delta t}}$.
    Determine Z^v according to (E.12).
    Calculate $\bar{v}_{t+\Delta t}$ and $\gamma_{t+\Delta t}$ using (E.11).

F

Accuracy Least-Squares Monte Carlo

F.1. Guaranteed Minimal Accumulation Benefit

The GMAB rider has an analytical solution, as we have shown in (5.10), which makes it a perfect test case for the Least-Squares Monte Carlo method. We will determine the expected liabilities according to the FFT algorithm described in appendix A and compare it to the Least-Squares Monte Carlo approach discussed throughout this chapter. By doing so, the accuracy of the Least-Squares Monte Carlo method can be assessed. In addition, we will be able to identify the algorithm's optimal set-up, which depends on the resource allocation, sampling method and type of basis function. The accuracy will be evaluated by four error measures,

$$\begin{cases} E_{\text{Mean}} = \frac{1}{M}\sum_{i=1}^{M}\left|l_{\text{pred}}^{(i)} - l_{\text{real}}^{(i)}\right|, \\ E_{\text{Std}} = \frac{1}{\sqrt{M-1}}\sqrt{\sum_{i=1}^{M}\left(l_{\text{pred}}^{(i)} - l_{\text{real}}^{(i)}\right)^2}, \\ E_{99.5} = l_{\text{pred}}^{[99.5]} - l_{\text{real}}^{[99.5]}, \\ E_{\text{tail}} = \frac{1}{0.005M}\sum_{i\in[99.25,99.75]}\left|l_{\text{pred}}^{[i]} - l_{\text{real}}^{[i]}\right|, \end{cases} \tag{F.1}$$

where $l^{(i)}$ indicates the loss of trajectory i and $l^{[j]}$ the loss of the j% quantile. These error measures can be thought of as

- E_{Mean} and E_{Std}: The mean absolute error and standard deviation, to test the stability across the entire distribution.
- $E_{99.5}$: Error on the 99.5% quantile, to test the error associated with the SCR.
- E_{tail}: The mean absolute error of the tail, to test the stability in the tail of the distribution.

In order to optimize the Least-Squares Monte Carlo algorithm, we will compare different set-ups:

- Regular polynomial basis functions (P) versus Legendre polynomial basis functions (L).
- Real-world sampling (Rw) versus Sobol sampling (So).
- Different resource allocations, where the *total* number of inner simulations is bounded. In case of the original risk-neutral measure, the total number of inner simulations is bounded by 200.000. To achieve similar accuracy, the time-dependent measure requires more sample points due to more volatile outcomes. Hence, we bound the number of inner simulations by 500.000 for the time-dependent risk-neutral measure. We will test different numbers of outer samples (R) and inner samples per outer sample (N), such that the total number of inner simulations does not exceed these bounds.

A few possible combinations and their errors are displayed in table F.1 for the original risk-neutral measure and table F.2 for the time-dependent risk-neutral measure. The contract details are given in table 5.1 and the initial market values follow scenario 1 of table 5.2. Moreover, all tests are performed with a fourth degree polynomial, which appeared to be optimal in all cases. The results of the other scenarios can be found in section F.1.1.

Set-up	E_{Mean}	E_{Std}	$E_{99.5}$	E_{tail}
L, So, R_1, N_1	0.9021	1.3215	**-0.6004**	**0.6986**
P, So, R_1, N_1	0.9021	1.3215	**-0.6004**	**0.6986**
L, Rw, R_1, N_1	**0.8293**	**1.2192**	3.5665	3.3367
P, Rw, R_1, N_1	**0.8293**	**1.2192**	3.5670	3.3370
L, So, R_2, N_2	1.7631	1.6940	1.2448	1.0234
P, So, R_2, N_2	1.7631	1.6940	1.2448	1.0234
L, Rw, R_2, N_2	1.6793	2.4735	-1.2260	0.8201
P, Rw, R_2, N_2	1.6793	2.4735	-1.2260	0.8201

Table F.1: Results of LSMC for different set-ups: Original risk-neutral measure GMAB. $R_1 = 200.000$ outer samples, $N_1 = 1$ inner sample, $R_2 = 20.000$ outer samples and $N_2 = 10$ inner samples.

Set-up	E_{Mean}	E_{Std}	$E_{99.5}$	E_{tail}
L, So, R_1, N_1	1.1742	**1.0822**	-0.0309	0.4627
P, So, R_1, N_1	**1.0573**	1.1014	**0.0292**	**0.3168**
L, Rw, R_1, N_1	2.1364	4.3668	3.6342	4.1385
P, Rw, R_1, N_1	2.0986	4.4141	4.2426	4.7362
L, So, R_2, N_2	1.2769	2.0972	1.3171	0.8684
P, So, R_2, N_2	1.2565	2.0915	1.3509	0.8639
L, Rw, R_2, N_2	1.5674	5.0697	-2.0135	2.5386
P, Rw, R_2, N_2	1.5411	5.0699	-1.9694	2.4902

Table F.2: Results of LSMC for different set-ups: Time-dependent risk-neutral measure GMAB. $R_1 = 200.000$ outer samples, $N_1 = 1$ inner sample, $R_2 = 20.000$ outer samples and $N_2 = 10$ inner samples.

The results appear to be inconclusive: different set-ups are considered optimal according to different error measures. Moreover, the different scenarios do not seem to agree on the optimal set-up. Therefore, the most robust set-up in terms of tail accuracy will be considered optimal.

First of all, there does not appear to be an optimal performing basis function according to this test. In theory, the Legendre polynomial should outperform the regular polynomial, due to its orthogonality. However, in case of the original risk-neutral measure, the difference of the basis functions is negligible. In case of the time-dependent measure, the results vary between the different scenarios. In some set-ups, the Legendre polynomial is more accurate, while it is outperformed by the regular polynomial in other set-ups. Since there is no clear optimal basis function, we will consider the regular polynomial in future analyses, as it is computationally more efficient to evaluate.

In the tail, Sobol sampling is more stable than real-world simulation sampling in this test, which is especially visible in the time-dependent risk-neutral measure results. This behaviour was expected, as described in section 5.2.3 and according to the findings of Cathcart (2012). Moreover, in some cases Sobol sampling even outperforms real-world simulation sampling in terms of overall accuracy. This is possibly the result of lower covariance between the least-squares regression parameters, as described in section 5.2.4.

Finally, according to this test it appears to be advantageous to sample as many outer points as possible. By sampling many outer points, the dependence on the inner errors is reduced, which produces the most stable solution, see for example figure 5.6. This is in line with the findings of Broadie et al. (2015), as discussed in section 5.2.4.

F.1.1. Accuracy results scenarios 2 and 3

Scenario 2

Set-up	E_{Mean}	E_{Std}	$E_{99.5}$	E_{tail}
L, So, R_1, N_1	1.3481	1.8438	1.0507	0.8955
P, So, R_1, N_1	1.3481	1.8438	1.0507	0.8955
L, Rw, R_1, N_1	1.2843	1.7203	4.3514	4.7482
P, Rw, R_1, N_1	1.2843	1.7203	4.3514	4.7482
L, So, R_2, N_2	**1.1902**	1.0300	**0.8742**	0.9167
P, So, R_2, N_2	**1.1902**	1.0300	**0.8742**	0.9167
L, Rw, R_2, N_2	1.2924	1.6475	**-0.0033**	**0.4624**
P, Rw, R_2, N_2	1.2924	1.6475	**-0.0033**	**0.4624**

Table F.3: Results of LSMC for different set-ups scenario 2: Original risk-neutral measure GMAB. $R_1 = 200.000$ outer samples, $N_1 = 1$ inner sample, $R_2 = 20.000$ outer samples and $N_2 = 10$ inner samples.

Set-up	E_{Mean}	E_{Std}	$E_{99.5}$	E_{tail}
L, So, R_1, N_1	1.2733	1.0714	-0.3997	0.8292
P, So, R_1, N_1	**1.1393**	**1.0828**	**-0.2637**	0.6802
L, Rw, R_1, N_1	1.6458	2.9598	-0.4833	**0.5713**
P, Rw, R_1, N_1	2.0853	2.9501	-1.2389	1.2006
L, So, R_2, N_2	1.2365	1.9761	1.0927	0.8840
P, So, R_2, N_2	1.2159	1.9712	1.1271	0.8773
L, Rw, R_2, N_2	2.0585	5.5933	5.2360	4.9830
P, Rw, R_2, N_2	2.0623	5.5944	5.2243	4.9695

Table F.4: Results of LSMC for different set-ups scenario 2: Time-dependent risk-neutral measure GMAB. $R_1 = 200.000$ outer samples, $N_1 = 1$ inner sample, $R_2 = 20.000$ outer samples and $N_2 = 10$ inner samples.

Scenario 3

Set-up	E_{Mean}	E_{Std}	$E_{99.5}$	E_{tail}
L, So, R_1, N_1	1.8241	1.9000	-0.3290	0.5360
P, So, R_1, N_1	1.8241	1.9000	-0.3290	0.5360
L, Rw, R_1, N_1	**1.3040**	1.9162	0.7978	0.4726
P, Rw, R_1, N_1	**1.3040**	1.9162	0.7978	0.4726
L, So, R_2, N_2	1.8929	**1.7394**	**0.0261**	**0.1476**
P, So, R_2, N_2	1.8929	**1.7394**	**0.0261**	**0.1476**
L, Rw, R_2, N_2	1.4651	3.2432	1.1996	1.0515
P, Rw, R_2, N_2	1.4651	3.2432	1.1996	1.0515

Table F.5: Results of LSMC for different set-ups scenario 3: Original risk-neutral measure GMAB. $R_1 = 200.000$ outer samples, $N_1 = 1$ inner sample, $R_2 = 20.000$ outer samples and $N_2 = 10$ inner samples.

Set-up	E_{Mean}	E_{Std}	$E_{99.5}$	E_{tail}
L, So, R_1, N_1	0.9397	**1.1332**	-0.5277	0.5396
P, So, R_1, N_1	**0.8781**	1.1603	-0.3215	**0.4111**
L, Rw, R_1, N_1	1.6104	2.9706	-5.4341	5.1993
P, Rw, R_1, N_1	1.7128	2.9754	-5.9395	5.6953
L, So, R_2, N_2	0.9842	2.4713	**-0.0184**	0.5111
P, So, R_2, N_2	0.9707	2.4640	0.0235	0.5127
L, Rw, R_2, N_2	1.8073	7.4200	1.6028	1.6835
P, Rw, R_2, N_2	1.8055	7.4216	1.6233	1.7093

Table F.6: Results of LSMC for different set-ups scenario 3: Time-dependent risk-neutral measure GMAB. $R_1 = 200.000$ outer samples, $N_1 = 1$ inner sample, $R_2 = 20.000$ outer samples and $N_2 = 10$ inner samples.

F.2. Guaranteed Minimum Withdrawal Benefit

It is more difficult to assess the accuracy of the Least-Squares Monte Carlo approximation in case of a variable annuity with the GMWB rider, as there is no analytical solution available. To this end, we will only test the stability of the algorithm on a certain set of quantiles. First, we will obtain the loss distribution according to the Least-Squares Monte Carlo algorithm, which uses the optimal set-up identified in section 5.3.3. From this distribution, the set of quantiles and the associated parameters (such as fund value and volatility at $t = 1$) can be determined. Hereafter, the *true* liabilities corresponding to these parameter sets will be evaluated according to a Monte Carlo approximation with 100.000 simulations. If the method is stable, we expect the Monte Carlo evaluations to be approximately equal to the Least-Squares Monte Carlo predictions. Table F.7 summarizes the results of this analysis.

Quantile	Original: LSMC	Original: MC	Time-dependent: LSMC	Time-dependent: MC
0.5%	-193.3	-190.5 *(0.5450)*	-211.5	-212.8 *(0.3604)*
10.0%	-99.3	-98.9 *(0.5345)*	-118.7	-120.3 *(0.4494)*
25.0%	-60.9	-59.4 *(0.5779)*	-78.5	-77.9 *(0.4400)*
50.0%	-16.0	-13.7 *(0.5638)*	-29.5	-29.1 *(0.4534)*
75.0%	33.8	35.6 *(0.5469)*	25.6	26.5 *(0.5460)*
90.0%	81.9	82.2 *(0.5558)*	79.7	78.5 *(0.5079)*
99.0%	174.0	172.2 *(0.5939)*	180.5	180.4 *(0.5249)*
99.25%	183.6	181.9 *(0.6369)*	190.3	186.4 *(0.7396)*
99.5%	196.7	196.1 *(0.5477)*	204.2	204.9 *(0.4637)*
99.75%	220.8	221.0 *(0.5277)*	228.2	228.3 *(0.4802)*

Table F.7: Stability analysis of a variable annuity with the GMWB rider for different risk-neutral measures. The standard errors of the estimates are given between the parentheses.

Based on these results we can assume that the approach is stable, since no major errors appear on any of the quantiles, especially at the upper-tail, which contains additional regression points. Moreover, similar results were achieved for the other scenarios (see tables F.8 and F.9), which indicates the robustness of the Least-Squares Monte Carlo approach.

F.2.1. Accuracy results scenarios 2 and 3

Quantile	Original: LSMC	Original: MC	Time-dependent: LSMC	Time-dependent: MC
0.5%	-69.4	-72.2 *(0.0867)*	-61.3	-61.1 *(0.2560)*
10.0%	-42.9	-42.2 *(0.1288)*	-19.1	-18.5 *(0.2734)*
25.0%	-27.5	-25.7 *(0.1774)*	2.4	4.1 *(0.3292)*
50.0%	-1.1	0.2 *(0.2436)*	32.7	32.8 *(0.3645)*
75.0%	37.2	36.0 *(0.2871)*	71.6	71.3 *(0.4488)*
90.0%	81.0	79.9 *(0.3502)*	114.1	112.6 *(0.5083)*
99.0%	174.7	175.6 *(0.4164)*	200.0	199.5 *(0.6469)*
99.25%	184.9	185.7 *(0.4252)*	209.7	209.3 *(0.5076)*
99.5%	199.9	202.3 *(0.3733)*	222.0	221.4 *(0.4768)*
99.75%	223.2	225.3 *(0.3750)*	243.1	242.8 *(0.5795)*

Table F.8: Stability analysis of a variable annuity with the GMWB rider for different risk-neutral measures of scenario 2. The standard errors of the estimates are given between the parentheses.

Quantile	Original: LSMC	Original: MC	Time-dependent :LSMC	Time-dependent: MC
0.5%	-328.0	-319.2 *(1.2513)*	-399.3	-399.4 *(0.5559)*
10.0%	-163.4	-164.4 *(1.0581)*	-241.7	-245.9 *(0.5781)*
25.0%	-101.1	-101.7 *(0.9776)*	-176.3	-179.1 *(0.5525)*
50.0%	-37.1	-36.0 *(0.9223)*	-102.4	-102.6 *(0.5781)*
75.0%	25.3	27.4 *(0.8601)*	-26.4	-26.0 *(0.8252)*
90.0%	80.1	82.7 *(0.7918)*	42.1	43.9 *(0.6536)*
99.0%	176.2	178.8 *(0.6942)*	160.2	163.3 *(0.5424)*
99.25%	185.0	188.0 *(0.6974)*	171.7	173.9 *(0.7486)*
99.5%	198.7	201.3 *(0.6752)*	187.2	186.5 *(0.5418)*
99.75%	220.8	223.1 *(0.6664)*	212.8	212.0 *(0.5349)*

Table F.9: Stability analysis of a variable annuity with the GMWB rider for different risk-neutral measures of scenario 3. The standard errors of the estimates are given between the parentheses.