On the modelling of nested risk-neutral stochastic processes with applications in insurance

S. N. Singor, A. Boer, J. S. C. Alberts & C. W. Oosterlee

Published online: 09 Oct 2017.

Submit your article to this journal ⎘

View related articles ⎘

View Crossmark data ⎘

Routledge
Taylor & Francis Group

ARTICLE

Check for updates

On the modelling of nested risk-neutral stochastic processes with applications in insurance

S. N. Singor[a,b], A. Boer[a], J. S. C. Alberts[a] and C. W. Oosterlee[b,c]

[a]Ortec Finance, Rotterdam, The Netherlands; [b]Delft University of Technology, Delft, The Netherlands; [c]CWI - National Research Institute for Mathematics and Computer Science, Amsterdam, The Netherlands

ABSTRACT

We propose a modelling framework for risk-neutral stochastic processes nested in a real-world stochastic process. The framework is important for insurers that deal with the valuation of embedded options and in particular at future points in time. We make use of the class of State Space Hidden Markov models for modelling the joint behaviour of the parameters of a risk-neutral model and the dynamics of option market instruments. This modelling concept enables us to perform non-linear estimation, forecasting and robust calibration. The proposed method is applied to the Heston model for which we find highly satisfactory results. We use the estimated Heston model to compute the required capital of an insurance company under Solvency II and we find large differences compared to a basic calibration method.

1. Introduction

In post-crisis risk management, both high-quality real-world and risk-neutral scenario models are required. Obviously, when a real-world model (Steehouwer 2005) is not able to model tail behaviour properly, the Solvency Capital Requirement (SCR) under Solvency II for insurance companies may be under- or over-valued. An advanced risk-neutral model (Brigo and Mercurio 2007; Hull 2006) is required to model relevant empirical facts in option market data.

We present a modelling framework to compute values of options in insurance that start at a future point in time. Such a modelling framework is important for ex-ante risk management for insurance companies (Zenios and Ziemba 2007). Insurance products such as unit-linked, variable-annuity and profit sharing contain guarantees that are valued using risk-neutral valuation techniques. Since valuation via Monte Carlo simulations (Glasserman 2004) is the market standard, this gives rise to a so-called nested Monte Carlo simulation problem, see Figure 1. Nested simulations for valuation recently gained increased interest in risk management applications of insurers (Bauer, Reuss, and Singer 2012). Instead of solving the nested simulation problem (Bauer, Bergmann, and Kiesel 2010; Cramwinckel, Singor, and Varbanescu 2015; Bauer, Bergmann, and Reuss 2010), here we propose
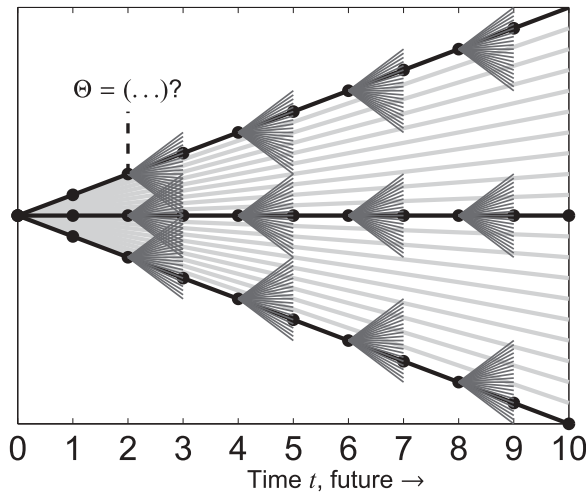

Figure 1. The dark grey lines refer to risk-neutral scenarios and the black (and light grey) lines refer to real-world scenarios. The challenge is to specify a risk-neutral valuation model, i.e. the model parameters Θ, in each real-world node.

a different modelling framework for the risk-neutral models to compute *future option values*. In order to compute future option values at given monitoring dates, we need a risk-neutral model that is consistent with the generated real-world dynamics of interest rates and implied volatility (IV) at each monitoring date.

In Bauer, Reuss, and Singer (2012), a mathematical framework is provided for the derivation of the required risk capital under the European regulatory framework Solvency II. Different alternatives for the numerical implementation based on nested simulations are reviewed. The model parameters should vary in each scenario in order to obtain consistency with the real-world scenarios. The calibration to the simulated real-world dynamics of option market prices would be too time consuming to perform for each real-world scenario set and hence avoiding such calibrations is desired in practice. We specify and calibrate a simulation model for the model parameters, so that calibration time is reduced.

In Feng et al. (forthcoming) and related articles (Stein, 2013; Stein, 2014), the computation of exposure distributions at a future point in time is discussed. These distributions are required for the computation of a credit valuation adjustment and potential future exposure, that are relevant for banks. Nested simulations are avoided by deriving an efficient approximation technique based on regression. First, risk-neutral scenarios are generated, that are used to calculate option values via regression. Second, real-world scenarios are generated and linked to the earlier regressed option values for the risk-neutral scenarios. Our methodology is different, we directly forecast the model parameters under the real-world measure and hence we embed the risk-neutral measure dynamics in the real-world measure dynamics.

The problem of modelling the dynamics of equity index option prices (often in terms of IV) has been investigated intensively in the academic literature. The focus is however typically on the empirical facts in IV surfaces and not on the link with an underlying

risk-neutral model that can be used for out-of-sample valuation. The latter is relevant to be able to determine the future value of embedded options and is therefore considered in this article. In Homescu (2011), a survey of methodologies is carried out for constructing IV surfaces at a certain time t. In Gatheral and Jacquier (2014), the Stochastic Volatility Inspired (SVI) parametrization method for equity option IVs is discussed. In Cont and Da Fonseca et al. (2002), a study to the dynamics of the IV surfaces is performed and it is shown that they are driven by a small number of random factors. A factor model is proposed, which is compatible with the empirical observations. In Carr and Liuren (2010), a partial differential equation (PDE) approach is presented to model the dynamics of IVs.

We make use of the class of the well-known *State Space Hidden Markov* (SSHM) models (Doucet and Johansen, 2009), that provide us a general modelling framework. SSHM models gained interest in the last 20 years due to the increased computational power and the improved estimation techniques. SSHM models are applied in many applications like navigation, time series analysis or robotic motion planning and control. By using the class of SSHM models, we can connect the dynamics of the model parameters of a risk-neutral model to the dynamics of the option market prices. By modelling the model parameters as stochastic processes under the real-world measure, we enforce consistency with the option market price dynamics and the risk-neutral model can be used for (out-of-sample) valuation at a monitoring date.

During the calibration of the SSHM models, we reduce the dimension of the calibration problem to improve the numerical stability of the model parameters with respect to the option market data, i.e. we wish to avoid overfitting issues. Numerical stability is an important requirement in practice, because stable balance sheet valuations are desired. We reduce the dimension by distinguishing static and dynamic model parameters, where static model parameters do not change over time. This is also beneficial for *present value* calibrations (and valuations) because only the dynamic part needs to be calibrated.

We apply the proposed framework to the risk-neutral Heston model (Heston, 1993), which is popular in practice due to its flexibility in modelling IV structures, and use S&P-500 index option data for calibration. Although the Heston model can be well calibrated to market option data, we show that some industry standard numerical techniques suffer from numerical instability, which results in undesired noisy calibrations. Using the proposed modelling framework, we reduce the 5 risk factors, i.e. the mean reversion, initial variance, long-term variance, volatility of variance and correlation, of the Heston model to one single hidden-risk factor. The calibration fit to the market data is highly satisfactory with respect to accuracy and stability.

The remainder of this article is organized as follows. In Section 2, we outline the problem and formulate the mathematical framework. In Section 3, we propose the modelling framework for predicting the risk-neutral model parameters under the real-world probability measure. In Section 4, we apply the proposed methodology to the Heston equity model and we show the impact of the proposed methodology by valuing a unit-linked insurance product. We conclude the study in Section 5.

2. Methodology

2.1. *Risk management context*

In risk management, insurance companies are required to compute present and future embedded option values (Bauer, Bergmann, and Reuss, 2010). Embedded options are rights in (life) insurance policies or pension contracts that may provide a profit to policy holders, but never a loss. The preferred valuation method is by means of risk-neutral Monte Carlo simulations (Glasserman, 2004). Computing option values at present time $t = 0$ is straightforward, but the computation of future time option values is more involved. At present time, we can download option market data, calibrate the model parameters $\Theta \in \mathbb{R}^{d_\Theta}$ and compute the embedded option values. The calibration step, i.e. to determine the model parameters given the option market data, is more complicated at a future time. Future option values are relevant for ex-ante risk management. In risk management, one forecasts the balance sheet using real-world scenarios and in each time step one determines the option value using risk-neutral Monte Carlo simulations, see Figure 1. The combination of these two simulations is called a nested simulation.

2.2. *Mathematical framework*

2.2.1. *Risk-neutral valuation*

Let $(\Omega, \mathcal{F}, \mathbb{Q})$ be a probability space, where Ω represents the space of all possible states in the financial market, $\mathbb{Q}$ is the so-called risk-neutral probability measure and the filtration $\{\mathcal{F}_u\}_{u \in [0, T_u]}$ represents all information about the financial market up to time u where $T_u \in \mathbb{R}_+$ is the maturity of the longest-term option in the option portfolio. We assume that investors can trade continuously in a frictionless financial market. We introduce a d_Z-dimensional Markov process $\{Z_{u,\Theta}\}_{u \in [0, T_u]}$, to model the uncertainty in the financial market required for option valuation, where $Z_{u,\Theta}$ evolves under $\mathbb{Q}$ according to the stochastic differential equation (SDE) in Equation (1) and is parametrized by a d_Θ-dimensional vector $\Theta \in \mathbb{R}^{d_\Theta}$,

$$dZ_{u,\Theta} = \mu_{Z,\Theta}(Z_{u,\Theta})\,du + \sigma_{Z,\Theta}(Z_{u,\Theta})\,dW_u^Z, \quad \left(Z_{0,\Theta} \in \mathbb{R}^{d_Z}, 0 \leq u \leq T_u\right). \tag{1}$$

Here, $\mu_{Z,\Theta} : \mathbb{R}^{d_Z} \mapsto \mathbb{R}^{d_Z}$ denotes the drift process, $\sigma_{Z,\Theta} : \mathbb{R}^{d_Z} \mapsto \mathbb{R}^{d_Z \times d_Z}$ the volatility process and W^Z are (correlated) standard Wiener processes.

We consider the valuation of tradable options in the market and embedded options in insurance liabilities. We let V denote a set of d_V tradable options and analogously we let $\tilde{V}$ represent a set of $d_{\tilde{V}}$ embedded options that are part of the balance sheet of an insurance company. We are interested in the valuation of V and $\tilde{V}$ given the risk-neutral model $Z_{u,\Theta}$ with model parameters Θ. For notational convenience, we define the valuation mapping of V as $\eta : \mathbb{R}^{d_\Theta} \mapsto \mathbb{R}^{d_V}$. That is, the function η gives the values of the options in V with respect to the risk-neutral model $Z_{u,\Theta}$, given a set of model parameters Θ. Analogously, we define the mapping $\tilde{\eta}$ as the valuation function of the embedded options $\tilde{V}$. We note that prices for V are observable in the market, but this is not the case for the embedded options $\tilde{V}$.

Let us denote the pay-off functions of V and $\tilde{V}$, respectively, by $p_V : \mathbb{R}^{D_Z} \times \mathbb{R} \mapsto \mathbb{R}$ and $p_{\tilde{V}} : \mathbb{R}^{D_Z} \times \mathbb{R} \mapsto \mathbb{R}$, so that the future cash flows of the options are given by, respectively,

$$C_{u,V} = p_V(u, Z_{u,\Theta}), \quad C_{u,\tilde{V}} = p_{\tilde{V}}(u, Z_{u,\Theta}).$$

The most well-known example is an equity put option with pay-off function $C_{u,V} = \max(K - S_{T_u}, 0)$, where K denotes the strike level, T_u the option maturity and S_u the equity index. The cash flows $C_{u,\tilde{V}}$ are in practice generated by the actuarial system of an insurer. Under $\mathbb{Q}$, option payments can be computed as expected discounted cash flows with respect to a chosen numéraire process, e.g. the bank account $\{B_{u,r}\}_{u\in[0,T_u]}$, with $B_{u,r} = e^{\int_0^u r_v \, dv}$, where $\{r_u\}_{u\in[0,T_u]}$ is the risk-neutral risk-free interest rate process. We furthermore define the discount factor by $B_{u,r}^{-1} = \frac{1}{B_{u,r}}$ (the inverse bank account). In the risk-neutral world, all individuals are indifferent to risk and expect to earn on all assets a return equal to the instantaneous risk-free short rate. The risk-neutral option value is then computed as the expected value under the risk-neutral measure:

$$\eta(\Theta) = \mathbb{E}^{\mathbb{Q}}\left[\int_0^{T_u} B_{u,r}^{-1} C_{u,V} \, du\right], \quad \tilde{\eta}(\Theta) = \mathbb{E}^{\mathbb{Q}}\left[\int_0^{T_u} B_{u,r}^{-1} C_{u,\tilde{V}} \, du\right].$$

Given a set of observed option prices V, we would like to find the parameters Θ such that $\eta(\Theta)$ is the best approximation of V. We define the calibration function $\chi : \mathbb{R}^{d_V} \mapsto \mathbb{R}^{d_\Theta}$ by

$$\chi(V) = \arg \min_{\Theta \in \mathcal{D}_\Theta} \| V - \eta(\Theta) \|, \tag{2}$$

where $\mathcal{D}_\Theta \subset \mathbb{R}^{d_\Theta}$ denotes the relevant domain of Θ. So, $\Theta = \chi(V)$ is the best fit to V with respect to norm $\| \cdot \|$. We will use the Euclidean norm, which in particular means that all option values are weighted equally, but note that it can be useful to apply weights for extreme in or out of the money options. This calibration function is computed using numerical optimization techniques, because closed form solutions usually do not exist.

Once we have fitted the model parameters to the data, i.e. when we have computed $\chi(V)$, the values of the embedded options readily follow from

$$\tilde{\eta}(\chi(V)) = \mathbb{E}^{\mathbb{Q}}\left[\int_0^{T_u} B_{u,r}^{-1} C_{u,\tilde{V}} \, du\right]. \tag{3}$$

In Section 3.4, we discuss the numerical stability of this function, and in particular how it relates to η.

2.2.2. Nested simulation

We let $\mathbb{P}$ denote the real-world (physical) probability measure and $\{Y_t\}_{t\in[0,T_t]}$ with $T_t \in \mathbb{R}_+$ be a d_Y-dimensional Markov process to model the uncertainty of the real-world financial market. We let T_t denote the horizon of the real-world simulation and t

denotes time in the real-world simulation. The market variables, i.e. the interest rate curve R and option instruments V, that are required for calibrating Θ follow hence stochastic processes: $\{R_t\}_{t\in[0,T_t]}$ and $\{V_t\}_{t\in[0,T_t]}$. Since the model parameters Θ of $Z_{u,\Theta}$ are connected to these stochastic market variables as shown in Equation (2), they cannot be fixed given the uncertainty of the real-world market. Therefore, we model them by a stochastic process $\{\Theta_t\}_{t\in[0,T_t]}$.

This means that at each time $t \in [0, T_t]$ a risk-neutral model is parametrized, as

$$\Theta_t \mapsto \left\{Z_{u,\Theta_t}\right\}_{u\in[0,T_u]}.$$

By modelling the stochastic process Θ_t, we generalize the nested simulation framework in Bauer, Reuss, and Singer (2012) because we also take the IV risk factor into account. In Bauer, Reuss, and Singer (2012), the parameters in Θ are kept constant during the real-world scenario simulation. The proposed modelling framework leads to a more realistic modelling of market IVs and model parameters of a risk-neutral model. We note that the embedded option values also follow a stochastic process $\left\{\tilde{V}_t\right\}_{t\in[0,T_t]}$ given the link with Θ in Equation (3), and the IV of an equity index option is obtained by solving for the volatility parameter of the Black-Scholes formula.

During numerical computation, we simulate a stochastic process given a discretization in the time domain $t \in [0, T_t]$; the random sample is called a scenario set. This can be applied to generate $N_{\mathbb{P}}$ independent paths of the real-world variables Y_t^j with $j = 1, \ldots, N_{\mathbb{P}}$. However, in order to calculate $\eta(\Theta_t)$ using $N_{\mathbb{Q}}$ Monte Carlo simulations given a scenario set for Θ_t, we have in turn to simulate $\left\{Z_{u,\Theta_t^{(j)}}\right\}_{u\in[0,T_u]}$ for each $t \in [0, T_t]$ and scenario $j = 1, \ldots, N_{\mathbb{P}}$ of Θ_t. This is a *nested simulation*. The scenarios for $\left\{Z_{u,\Theta_t^{(j)}}\right\}_{u\in[0,T_u]}$ are the *inner* scenarios, those for Y_t the *outer* scenarios. In the nested modelling structure in Figure 1, the dark grey lines refer to the stochastic process $\left\{Z_{u,\Theta_t}\right\}_{u\in[0,T_u]}$ given a scenario set of Θ_t and the black (and light grey) lines refer to the real-world stochastic process $\{Y_t\}_{t\in[0,T_t]}$.

There are two procedures that make nested simulations expensive. First, analytic formulas often do not exist to compute the option values V_t. Applying analytic formulas is advantageous regarding the computation time, but restrictive when expanding the model set-up to hybrid models (Grzelak and Oosterlee, 2011), that are generally required to obtain a high quality-of-fit with the data. In other words, analytic formulas can only be derived for specific cases, such as the Black-Scholes and Heston models for call/put options, but are often not available for hybrid models such as the Heston Hull-White model. Therefore, Monte Carlo simulations (Glasserman, 2004) are often used to compute the risk-neutral option values V_t, which makes the mapping η expensive to evaluate. Second, the calibration of the model parameters Θ_t (see Equation (2)) for each time t and scenario $j = 1, \ldots, N_{\mathbb{P}}$ can generally not be done in closed form. Numerical optimization techniques that can be used for calibration are generally expensive and they will be more time consuming in case of using Monte Carlo simulations.

We do not aim to improve computation times by approximating η by an inexpensive function. A relevant approximation method is the Least-Square Monte Carlo (LSMC) (Bauer, Bergmann, and Reuss, 2010; Jain and Oosterlee, 2015). Another way to reduce computation time is to make use of High Performance Computing (HPC). In Cramwinckel, Singor, and Varbanescu (2015), a GPU framework for accelerating nested simulations is proposed for simulating equity index option values.

In the proposed modelling framework, which we discuss in Section 3, we avoid the numerical calibration procedure to obtain Θ_t given the market variables Y_t, see Equation (2). We make use of the class of SSHM models to directly model the stochastic process Θ_t. The observable process V_t is subsequently modelled via the valuation mapping $\eta(\Theta_t)$. We also discuss numerical stability of the model parameters with respect to changes in the market option data for calibration.

3. Modelling framework

In this section, we explain the proposed modelling framework. In Section 3.1, we first elaborate on the SSHM models. In Section 3.2, we discuss the estimation of SSHM models. In Section 3.3, we discuss our proposed modelling framework. An important requirement is to improve numerical stability of the model parameters Θ with respect to the option market data, which we discuss in Section 3.4.

3.1. SSHM models

SSHM models provide a general and flexible framework for modelling time-series in a broad range of applications. A thorough introduction into the field can be found in the books (Doucet and Johansen, 2009; Cappé, Moulines, and Tobias, 2009). These models provide a flexible framework in financial modelling for stochastic processes. The most popular SSHM model is the linear Gaussian model, which can be estimated in closed form using the well-known Kalman filter (Durbin and Koopman, 2012). However, in most cases, the Kalman filtering method cannot be applied due to the non-linear behaviour of the observable process.

For notation[1], we let $X_n := X_{t_n}$ and $X_{1:n} := \{X_k\}_{k=t_1}^{t_n}$ (and similar for V). In an SSHM modelling structure, we consider a hidden state process $X_n \in \mathcal{X}$ that evolves according to

$$X_1 \sim f_{1,X,\psi}(X_1), \quad X_n \mid (X_{n-1} = x_{n-1}) \sim f_{X,\psi}(x_n | x_{n-1}), \tag{4}$$

where X_n is the state at time t_n, $f_{1,X,\psi}$ a probability density function (PDF) and $f_{X,\psi}$ describes the transition from x_{n-1} to x_n, where x_n is a realization of X_n, with unknown parameters $\psi \in \Psi \subset \mathbb{R}^{d_\psi}$. The hidden states are conditionally independent.

We are interested in the hidden states $X_{1:N}$, but in our case we can only observe the $V_n \in \mathcal{V}$ process for $n = 1, \ldots, N$. Conditional on $X_{1:n}$, the observations V_n are assumed to be independent and their marginal densities are given by

$$V_n \mid (X_n = x_n) \sim f_{V,\psi}(v_n | x_n), \tag{5}$$

where $f_{V,\psi}$ describes the transition from x_n to v_n, with v_n a realization of V_n. The non-linear relation between x_n and v_n makes the SSHM model involved. In general, the PDFs $f_{X,\psi}$ and $f_{V,\psi}$ are chosen to be well-known distribution functions. We assume that all information in the observed data is explained by the underlying hidden process $X_{1:N}$, so that specification of the correlation between X_n and V_n is redundant. Such an assumption is standard in classical state space models. An overview of alternative representations is provided in Murphy (2002).

3.2. Estimation

State inference given the observed process $V_{1:N} = v_{1:N}$ and fixed parameters ψ, we are interested in inferring the states $X_{1:N} = x_{1:N}$. In a Bayesian framework, inference of $X_{1:n}$ given a realization $v_{1:n}$ of $V_{1:n}$ relies on the posterior distribution $f_\psi(x_{1:n}|v_{1:n})$:

$$f_\psi(x_{1:n}|v_{1:n}) = \frac{f_\psi(v_{1:n}, x_{1:n})}{f_\psi(v_{1:n})} = \frac{f_\psi(v_{1:n}|x_{1:n})f_\psi(x_{1:n})}{f_\psi(v_{1:n})}, \tag{6}$$

with $f_\psi(v_{1:n}|x_{1:n})$ the likelihood, $f_\psi(x_{1:n})$ the prior, $f_\psi(v_{1:n})$ the evidence and $f_\psi(v_{1:n}, x_{1:n})$ is often referred to as the complete data likelihood. The prior, likelihood and evidence (also known as the marginal likelihood) are defined by

$$f_\psi(x_{1:n}) = f_{1,X,\psi}(x_1) \prod_{k=2}^{n} f_{X,\psi}(x_k|x_{k-1}), \tag{7}$$

$$f_\psi(v_{1:n}|x_{1:n}) = \prod_{k=1}^{n} f_{V,\psi}(v_k|x_k), \tag{8}$$

$$f_\psi(v_{1:n}) = \int f_\psi(v_{1:n}, x_{1:n}) \, dx_{1:n}. \tag{9}$$

For the linear Gaussian model, the posterior distribution is a Gaussian distribution whose mean and covariance can be computed in closed form (Durbin and Koopman, 2012). For most non-linear non-Gaussian models, it is not possible to compute these distributions in closed-form and we need to employ numerical methods.

We are interested in the filtering and marginal likelihood computations, i.e. the sequential approximation of the distributions $\{f_\psi(x_{1:n}|v_{1:n})\}_{n\geq1}$ and marginal likelihoods $\{f_\psi(v_{1:n})\}_{n\geq1}$, because directly solving Equation (6) is problematic due to the high dimensionality. In this sequential computation, one often relies on the so-called prediction and update equations (Doucet and Johansen, 2009); the recursion satisfies the marginal distribution $f_\psi(x_n, v_{1:n})$. The prediction and update equations are respectively given by

$$\text{Prediction}: \quad f_\psi(x_n|v_{1:n-1}) = \int f_{X,\psi}(x_n|x_{n-1})f_\psi(x_{n-1}|v_{1:n-1})dx_{n-1}, \tag{10}$$

$$\text{Update}: \quad f_\psi(x_n|v_{1:n}) = \frac{f_{V,\psi}(v_n|x_n)f_\psi(x_n|v_{1:n-1})}{f_\psi(v_n|v_{1:n-1})}, \tag{11}$$

where

$$f_{\psi}(v_n | v_{1:n-1}) = \int f_{V,\psi}(v_n | x_n) f_{X,\psi}(x_n | x_{n-1}) f_{\psi}(x_{n-1} | v_{1:n-1}) \, dx_{n-1:n}. \tag{12}$$

If we can compute $f_{\psi}(x_{1:n} | v_{1:n})$ and thus $f_{\psi}(x_n | v_{1:n})$ sequentially, then the evidence $f_{\psi}(v_{1:n})$ can also be evaluated recursively using

$$f_{\psi}(v_{1:n}) = f_{\psi}(v_1) \prod_{j=2}^{n} f_{\psi}(v_n | v_{1:n-1}),$$

where the $f_{\psi}(v_n | v_{1:n-1})$ are computed by Equation (12).

Once the filtering stage is completed, one may smooth the filtered state process. *Filtering* is the estimation of the distribution of the current state x_n based upon the observations received up until current time $v_{1:n}$. *Smoothing* implies estimating the distribution of the state x_n given all observations up to some later time $v_{1:N}$, i.e. $f_{\psi}(x_n | v_{1:N})$. In general, smoothing is computationally more challenging than filtering. The trajectory estimates obtained by such methods, as a result of the additional information available, tend to be smoother than those obtained by filtering.

3.2.1. *Non-linear Hidden Markov models*

Unfortunately, in most applications the SSHM model is non-linear and non-Gaussian. Therefore, various numerical approximations were developed over the years. The well-known Extended Kalman Filter (EKF) and Unscented Kalman Filter (UKF) were introduced shortly after the Kalman Filter. In general, the UKF can acquire more accurate estimation results than the EKF, but can lead to serious errors for non-Gaussian distributions (Simon et al., 2001; Simon, 2006). Since we want our approach to be as general as possible, we make use of so-called *Sequential Monte Carlo* (SMC) methods.

SMC methods represent a set of flexible and powerful simulation-based methods which provide samples distributed approximately according to posterior distributions and facilitate the approximate calculation of $f_{\psi}(v_{1:n})$. The main idea behind SMC methods is to obtain a large collection of weighted random samples, named particles, whose empirical distribution converges to the distribution we wish to sample from, the posterior distribution $f_{\psi}(x_{1:n} | v_{1:n})$. For this reason, SMC methods are also referred to as *Particle Filters* in the filtering context. They have become a popular class of methods for inference in non-linear non-Gaussian state space models. An overview of the theory and applications of different SMC methods can be found in Doucet and Johansen (2009).

3.2.2. *Combined state and parameter inference*

Because the transitional PDFs in Equations (4) and (5) are parametrized by ψ, we are also interested in inferring these parameters given the observed process $V_{1:N} = v_{1:N}$. The Expectation-Maximization (EM) algorithm (Dempster, Laird, and Rubin, 1977) is an efficient method for combined state and parameter inference. The EM is an iterative method for maximizing the likelihood $f_{\psi}(v_{1:n})$ in Equation (9). This method is useful when it is not possible to evaluate and optimize this likelihood directly. The method can

be used for combined state and parameter inference in SSHM models. The algorithm is based on the insight that the auxiliary function

$$Q(\psi, \psi') = \int \log f_\psi(v_{1:n}, x_{1:n}) f_{\psi'}(x_{1:n} | v_{1:n}) \, dx_{1:n}, \tag{13}$$

may be used as a surrogate model for $f_\psi(v_{1:n})$, because increasing $Q(\psi, \psi')$ forces an increase of $f_\psi(v_{1:n})$ (Cappé, Moulines, and Tobias, 2009). The EM algorithm is initialized by $\psi_0 \in \Psi$ and $x_{1:N}^{(0)}$ and iterates between an expectation (E) step, which facilitates state inference, and a maximization (M) step for parameter inference:

$$\text{E} - \text{step} : \text{Conditional on } \psi_{k-1} \text{ compute } Q(\psi, \psi_{k-1}), \tag{14}$$

$$\text{M} - \text{step} : \text{Conditional on } x_{1:N}^{(k)} \text{ compute } \psi_k = \arg \max_{\psi \in \Psi} Q(\psi, \psi_{k-1}). \tag{15}$$

This EM-algorithm generates a sequence $\left\{ \psi_k, x_{1:N}^{(k)} \right\}_{k=0}^{N_{EM}}$, which converges to a stationary point of the likelihood for $N_{EM} \to \infty$, with the number of iterations $N_{EM} \in \mathbb{N}_+$. We solve the M-step of the EM algorithm using optimization techniques from the Matlab library. In particular, we use the fmincon algorithm in Matlab, which is a constrained nonlinear optimization algorithm. The advantage of this Matlab function is that one can impose constraints on the parameters to be optimized. One can choose between a number of algorithms, but we have used the interior point algorithm.

Note that since the computation of the E-step includes a complicated multi-dimensional integral, we can approximate it by using Monte Carlo integration. Assuming we can sample from $f_{\psi_{k-1}}(x_{1:N} | v_{1:N})$ using a particle filtering method, we replace the E-step by the simulation of $N_{par} \in \mathbb{N}_+$ realizations $\left\{ X_{1:N}^j \right\}_{j=1}^{N_{par}}$ from $f_{\psi_{k-1}}(x_{1:N} | v_{1:N})$ and the computation of

$$\tilde{Q}_k(\psi) = \frac{1}{N_{par}} \sum_{j=1}^{N_{par}} \log f_\psi(X_{1:N}^j, v_{1:N}).$$

This leads to the Monte Carlo EM algorithm (MCEM). Unfortunately, a drawback of this approach is that it requires the number of particles N_{par} to grow with each new iteration of the algorithm $k \leq N_{EM}$ (Cappé, Moulines, and Tobias, 2009). Besides this, we need to sample a whole new set of realizations of the hidden states $\left\{ X_{1:N}^j \right\}_{j=1}^{N_{par}}$ at each iteration, that are not re-used in later iterations. The Stochastic Approximation EM (SAEM) algorithm (Delyon, Lavielle, and Moulines, 1999) makes more efficient use of the simulated variables by replacing $\tilde{Q}_k(\theta)$ with a stochastic averaging procedure

$$\hat{Q}_k(\psi) = (1 - \gamma_k)\hat{Q}_{k-1}(\psi) + \gamma_k \left(\frac{1}{N_{par}} \sum_{j=1}^{N_{par}} \log f_\psi(X_{1:N}^j, v_{1:N}) \right), \tag{16}$$

where $\{\gamma_k\}_{k \geq 0}$ is a decreasing sequence of weights that satisfy $\sum \gamma_k = \infty$ and $\sum \gamma_k^2 < \infty$. As $k \to \infty$, the SAEM algorithm converges to a local maximum of the likelihood function (Delyon, Lavielle, and Moulines, 1999). The computational advantage of the SAEM algorithm is especially significant in problems where maximization is much cheaper than simulation.

The MCEM and SAEM methods cannot be applied directly, because for non-linear non-Gaussian SSHM models it is not possible to directly sample from the posterior distribution. Using a particle filter method for sampling from the posterior distribution leads to an SMC-analogue of the previous methods, the PSEM method described in Schön, Wills, and Ninness (2011). If we take the SAEM approach, it is sufficient to generate a single sample each iteration.

In Kuhn and Lavielle (2004) it is shown that for convergence of the SAEM algorithm, it is not necessary to sample exactly from the posterior distribution. We can also sample from a family of so-called Markov kernels $\{M_\psi(x_{1:N}|x'_{1:N})\}_{\psi \in \psi}$ on $\mathcal{X}^N$ that leaves the family of posterior distributions invariant. Assume that we have such a family and let in iteration $k \leq N_{EM}$ of the SAEM method $x_{1:N}[k-1]$ be the previous draw from the Markov kernel. We then sample $X_{1:N}[k] \sim M_{\psi_{k-1}}(x_{1:N}|x_{1:N}[k-1])$ and update $\hat{Q}$ according to

$$\hat{Q}_k(\psi) = (1 - \gamma_k)\hat{Q}_{k-1}(\psi) + \gamma_k \log f_\psi(X_{1:N}[k], \nu_{1:N}). \tag{17}$$

The next approximation of ψ is then obtained by maximizing this quantity w.r.t. $\psi \in \Psi$ (the M-step) using conditional particle filters (CPF) as the required Markov kernel. We use the SAEM method using CPF, which is referred to as CPF-SAEM (Lindsten, 2013; Lindsten, Jordan, and Schön, 2014). In this method, the Markov kernel is constructed by running a SMC sampler in which one particle trajectory $x'_{1:N}$ is specified a priori, a so-called CPF. We can think of this reference trajectory as guiding the simulated particles to a relevant region of the state space. The path $x'_{1:N}$ is ensured to survive all re-sampling steps.

3.3. Modelling concept

In Section 2, we discussed the concept of nested simulation. The process Y_t models the uncertainty of the real-world financial market, and consists amongst others of the processes R_t and V_t. To avoid a numerical calibration procedure to obtain Θ_t, see Equation (2), we model the process Θ_t directly. Given the process Θ_t, we compute the observable process by $V_t = \eta(\Theta_t)$ and the embedded option values by $\tilde{V}_t = \tilde{\eta}(\Theta_t)$.

Modelling the joint behaviour of possibly a large number of variables in Θ_t without any constraints is generally problematic, because estimating an underlying model may result in a high-dimensional problem. To reduce the dimension of the problem, we assume the d_Θ-dimensional process $\{\Theta_t\}_{t \in [0,T_l]}$ to be generated by a latent (hidden) d_X-dimensional process $\{X_t\}_{t \in [0,T_l]} \in \mathcal{X}$, with

$$\Theta_t = \Upsilon(X_t), \quad \Upsilon : \mathbb{R}^{d_X} \mapsto \mathbb{R}^{d_\Theta}. \tag{18}$$

The dimension of the calibration problem is controlled by $d_X \leq d_\Theta$. We aim to vary the model parameters that are most significant with respect to the calibration fit, hence by keeping d_X as small as possible. The parameters of Υ (including X_t) are still determined

by maximizing the calibration fit to the option market data V_t. This is beneficial for interpretation and can improve numerical stability.

When $d_X \leq d_\Theta$, we achieve an effective dimension-reduction, since the model parameters Θ_t are then driven by a lower-dimensional process X_t. Several ways of defining the mapping Υ exist and three requirements for Υ should be taken into account. First, the mapping should lead to a satisfactory fit to the market data. Second, the mapping should lead to a realistic representation of the model parameters. Third, the mapping should lead to numerically stable results. A basic assumption is a linear mapping between X_t and Θ_t in Equation (19), i.e.

$$\Theta_t = \Upsilon(X_t) := a_\Upsilon X_t + b_\Upsilon, \quad \left(a_\Upsilon \in \mathbb{R}^{d_X \times d_\Theta}, b_\Upsilon \in \mathbb{R}^{d_\Theta} \right), \tag{19}$$

where a_Υ and b_Υ are unknown parameters. The mapping in Equation (18) can be split into a dynamic part X_t and static part $\{a_\Upsilon, b_\Upsilon\}$. That is, the static part does not change over time and is calibrated once. Calibration to available market data then only takes place with respect to the dynamic part (while the static part is kept fixed). This reduces computation times of calibration, because (by construction) the parameter space is smaller. In Section 4, we estimate these dynamic and static parts for the Heston model.

The mapping Υ should be specified in such a way that Θ_t belongs to domain $\mathcal{D}_\Theta \subset \mathbb{R}^{d_\Theta}$. For example, the correlation matrix should be positive definite, the correlation parameters should be in $[-1, 1]$ and volatility parameters should be non-negative. To ensure that Υ maps values of X_t to the correct domain for Θ_t, we truncate values that are outside the domain $\mathcal{D}_\Theta$, i.e.

$$\Theta_t \in \left[\mathcal{D}_{\Theta,min}, \mathcal{D}_{\Theta,max} \right],$$

where the minimum and maximum domains are, respectively, denoted by $\mathcal{D}_{\Theta,min} \in \mathbb{R}^{d_\Theta}$ and $\mathcal{D}_{\Theta,max} \in \mathbb{R}^{d_\Theta}$.

Using this approach, we compute the process V_t by the following composite function:

$$V_t = \eta(\Upsilon(X_t)), \quad (\Theta_t = \Upsilon(X_t)). \tag{20}$$

Next, we let the hidden states X_t evolve through time via the transition PDF $f_{X,\psi}$ and given a realization $X_t = x_t$, the option price $V_t = v_t$ is computed via Equation (20). We model the (calibration) error between the valuation function η and the observed process V_t as a stochastic process by means of the PDF $f_{V,\psi}$. This is required because the underlying risk-neutral model is generally not able to perfectly fit the market option data. The Black-Scholes model can for example only model flat IV surfaces, whereas the IV typically varies per strike and maturity. Second, we introduce an error by reducing the number of degrees of freedom in the risk-neutral model for calibration: $d_X \leq d_\Theta$. A basic assumption for $f_{V,\psi}$ would be a normal distribution. In this case, the error $v_t - \eta(\Upsilon(x_t))$ is symmetric around zero.

The proposed SSHM model is given by

$$\text{State}: \quad X_1 \sim f_{1,X,\psi}(X_1), \quad X_n \,|\, (X_{n-1} = x_{n-1}) \sim f_{X,\psi}(x_n | x_{n-1}), \tag{21}$$

$$\text{Observation}: \quad V_n \,|\, (X_n = x_n) \sim f_{V,\psi}(v_n | \eta(\Upsilon(x_n))), \tag{22}$$

The SSHM model defined in Equations (21),(22) is non-linear due to the non-linearity of the valuation function η. Hence, analytic solutions do not exist. We condition on the state x_n (as in Equation (5)), but we incorporate the valuation function η in the PDF $f_{V,\psi}$, which we highlight by conditioning on $\eta(\Upsilon(x_n))$, $n \leq N$.

Given the observed process $V_{1:N} = v_{1:N}$, we wish to infer states $X_{1:N} = x_{1:N}$ and parameters ψ, where the static part of Υ (i.e. a_Υ and b_Υ) are part of ψ. We use the EM algorithm (see Section 3.2) for combined state and parameter inference. That is, the dynamic part of Υ, the hidden states X_t, is determined via the expectation step and the fixed model parameters ψ are determined via the maximization step. We aim to keep the dimension d_X as low as possible, but by optimizing the fit to the market data via maximizing the Q-function in Equation (13).

3.4. Numerical stability

We wish to achieve a satisfactory fit to the option market data V by using as few parameters as possible, but we also wish to achieve stable valuations of the embedded options $\tilde{V}$, i.e. we wish to avoid overfitting issues. For notational convenience, we drop the subscript t in this subsection. Numerical stability is important in practice because the calibrated model parameters are used by insurers to value the (out-of-sample) embedded options on their balance sheet. If changes in the market data imply undesired large changes in the model parameters, this can result in noisy valuations of embedded options. Overfitting occurs when a model has too many parameters relative to the number of observations or has an incorrect model structure. Overfitting may lead to poor predictive performance, as it typically results in excessive reactions to minor fluctuations in the option market data used for calibration.

On the other hand, a modelling setup that does not react to changes in the market data at all may be numerically very stable, but will not result in realistic embedded option values. We therefore also require that our modelling setup fits market prices well enough. So, there is a trade-off to be made. Below we will focus on the numerical stability, but we emphasize that the quality-of-fit should also be checked when evaluating the modelling approach.

In Section 2.2, we have explained that the model parameters are determined via the calibration function χ in Equation (2). These calibrated model parameters are then used to compute the embedded option values via Equation (3). By applying the chain rule to the composite function $\tilde{\eta} \circ \chi$ in Equation (3) we have

$$D_{\tilde{\eta} \circ \chi} = D_{\tilde{\eta}} D_\chi,$$

where D denotes is the Jacobian matrix. Hence, the sensitivity of $\tilde{\eta}$ to the option market data V can be decomposed to the sensitivity of $\tilde{\eta}$ to Θ and the sensitivity of Θ to V.

We use the operator norm $\| \cdot \|^{op}$ to quantify numerical stability, where we recall that the operator norm is defined as

$$\| A \|^{op} = \sup \left\{ \frac{\| Ax \|}{\| x \|} \ : \ x \in \mathbb{R}^n, x \neq 0 \right\},$$

for any linear map $A : \mathbb{R}^n \to \mathbb{R}^m$, and any $n, m > 0$. We use the Euclidean norm to compute $\| \cdot \|$. We also recall that the operator norm equals the largest singular value of A (Atkinson and Han, 2005).The operator norm quantifies the direction that is amplified the most by A.

A small change in the market option prices δV will result in a change in embedded option values $D_{\tilde{\eta} \circ \chi} \delta V$, up to first order. The operator norm therefore provides an upper bound to how strong embedded option values change as a result of a small change in observed market prices, and is therefore a measure of numerical stability. Using the chain rule and the Cauchy-Schwartz inequality, we obtain an upper bound of $\| D_{\tilde{\eta} \circ \chi} \|^{op}$:

$$\| D_{\tilde{\eta} \circ \chi} \|^{op} = \| D_{\tilde{\eta}} D_{\chi} \|^{op} \leq \| D_{\tilde{\eta}} \|^{op} \| D_{\chi} \|^{op}.$$

The sensitivity of $\tilde{\eta}$ to $\Theta = \chi(V)$ follows from the actuarial system of an insurer (see Section 2.2). We focus on the sensitivity of the calibration function χ in Equation (2), as we assume the actuarial system of the insurance company as given. More concretely, to keep $\| D_{\tilde{\eta} \circ \chi} \|^{op}$ as small as possible, we require the operator norm $\| D_{\chi} \|^{op}$ to be as small as possible for stable calibrations of Θ.

Since we have no closed form expression for χ, and numerical calibration of its derivative is expensive, we relate D_{χ} to D_{η} for which we do not have to perform numerical optimizations. In order to do this, we assume that η is locally an embedding near Θ, so in particular D_{η} has maximal rank. This is a very modest assumption because a zero vector of D_{η} would be a salient over-parametrization: a change in parameter would not lead to *any* change in option values up to first order.

We proceed to decompose the tangent space $T_{\eta(\Theta)}$ into the image of $D_{\eta(\Theta)}$, im $D_{\eta(\Theta)}$, and its ortho-complement, $(\text{im } D_{\eta(\Theta)})^{\perp}$. Let us denote the orthogonal projection onto im $D_{\eta(\Theta)}$ by π_{η}

$$\pi_{\eta}(\text{im } D_{\eta(\Theta)})^{\perp} = 0, \quad \pi_{\eta}(x) = x, \ \forall x \in \text{im} D_{\eta(\Theta)}.$$

Since $D_{\eta(\Theta)}$ has a trivial kernel, the following defines a pseudo-inverse of $D_{\eta(\Theta)}$:

$$D_{\eta(\Theta)}^{+} = (D_{\eta(\Theta)}^{T} D_{\eta(\Theta)})^{-1} D_{\eta(\Theta)}^{T}.$$

This is also a left-inverse of $D_{\eta(\Theta)}$ and moreover $D_{\eta(\Theta)} D_{\eta(\Theta)}^{+}$ is the orthogonal projection on the image of $D_{\eta(\Theta)}$, i.e.

$$D_{\eta(\Theta)} D_{\eta(\Theta)}^{+} = \pi_{\eta}.$$

We now claim that

$$D_{\chi(V)} = D_{\eta(\Theta)}^{+}, \quad (\Theta = \chi(V)). \tag{23}$$

To see this, consider a first-order perturbation δV of the option market data V. This decomposes into

$$\delta V = \pi_{\eta}(\delta V) + (\delta V)^{\perp}, \quad (\delta V)^{\perp} := \delta V - \pi_{\eta}(\delta V) \in (\text{im} D_{\eta})^{\perp}.$$

Let us write, $\delta\Theta = D^+_{\eta(\Theta)}(\delta V)$, so that

$$D_{\eta(\Theta)}\delta\Theta = D_{\eta(\Theta)}D^+_{\eta(\Theta)}(\delta V) = \pi_\eta(\delta V).$$

Using this, we have up to first order that

$$
\begin{aligned}
\chi(V + \delta V) \quad &= \arg\min_{\Theta \in \mathcal{D}_\Theta} \parallel V + \delta V - \eta(\Theta) \parallel = \arg\min_{\Theta \in \mathcal{D}_\Theta} \parallel \\
&V + \delta V - D_{\eta(\Theta)}\delta\Theta - \eta(\Theta - \delta\Theta) \parallel + \mathcal{O}^2(\delta V) \\
&= \arg\min_{\Theta \in \mathcal{D}_\Theta} \parallel V + (\delta V)^\perp - \eta(\Theta - \delta\Theta) \parallel + \mathcal{O}^2(\delta V).
\end{aligned}
\tag{24}
$$

Now, to determine for which Θ the minimum is attained, we observe that

$$
\begin{aligned}
\frac{d}{d\Theta}\big|_{\Theta + \delta\Theta} \parallel V + (\delta V)^\perp - \eta(\Theta - \delta\Theta)\parallel^2 &= -2\Big(V + (\delta V)^\perp - \eta(\Theta)\Big)D_{\eta(\Theta)} \\
&= -2(V - \eta(\Theta))D_{\eta(\Theta)} \\
&= \frac{d}{d\Theta}\big|_\Theta \parallel V - \eta(\Theta)\parallel^2.
\end{aligned}
$$

Since the last term equals zero for $\Theta = \chi(V)$, it follows that the minimum in Equation (24) is attained at $\Theta = \chi(V) + \delta\Theta$, and so

$$\chi(V + \delta V) = \chi(V) + \delta\Theta + \mathcal{O}^2(\delta V),$$

which proves Equation (23).

Note that this implies that the singular values of $D_{\chi(V)}$ are equal to the reciprocals of the singular values of $D_{\eta(\Theta)}$ (which is a general fact about pseudo-inverses). Consequently,

$$\parallel D_\chi \parallel^{op} = \sigma_{D_\chi, max} = \frac{1}{\sigma_{D_\eta, min}},$$

where $\sigma_{D_\chi, max}$ and $\sigma_{D_\eta, min}$ are, respectively, the largest singular value of D_χ and smallest singular value of D_η. We conclude that for numerical stability of the calibration part, we would like $\sigma_{D_\eta, min}$ to be as large as possible. In Section 4, we compute $\sigma_{D_\eta, min}$ for the Heston model.

3.4.1. *Dimension reduction*
We use the parametrization Υ in Equation (18) to reduce the dimension of Θ. More precisely, we use the following adjusted calibration function

$$\chi^\Upsilon(V) = \Upsilon\left(\arg\min_{x \in \mathbb{R}^{d_X}} \parallel V - \eta(\Upsilon(x)) \parallel,\right).$$

We assume that Υ is an embedding, and therefore $\eta \circ \Upsilon$ as well, so that we can apply exactly the same arguments as used above to prove Equation (23), but now applied to

$\eta \circ \Upsilon$ instead of η. We omit the points in which derivatives are taken for brevity. This results in

$$D_{\chi^\Upsilon} = D_\Upsilon \, D\left(\arg\min_{x \in \mathbb{R}^{d_X}} \| \, V - \eta(\Upsilon(x)) \, \|, \right) = D_\Upsilon \, (D_{\eta \circ \Upsilon})^+.$$

We argue that using Υ to parameterize Θ *generically* improves numerical stability. To this end, observe that $D_\eta^+ D_\eta = I$, and so

$$\begin{aligned} D_{\chi^\Upsilon} &= D_\eta^+ D_\eta D_\Upsilon \, (D_{\eta \circ \Upsilon})^+ = D_\eta^+ D_{\eta \circ \Upsilon} (D_{\eta \circ \Upsilon})^+, \\ &= D_\chi \pi_{\eta \circ \Upsilon}, \end{aligned} \tag{25}$$

where we have denoted the orthogonal projection operator onto the image of $D_{\eta \circ \Upsilon}$ by $\pi_{\eta \circ \Upsilon}$. Equation (25) can be used to compare the operator norms of D_χ and $D_{\chi \circ \Upsilon}$. On the one hand, we have that

$$\begin{aligned} \| \, D_\chi \|^{op} =\| \, D_\chi \circ \pi_\eta \|^{op} &= \sup\left\{ \frac{\| \, D_\chi \pi_\eta x \, \|}{\| \, x \, \|} \; : \; x \in \mathbb{R}^{d_X}, x \neq 0 \right\} \\ &= \sup\left\{ \frac{\| \, D_\chi x \, \|}{\| \, x \, \|} \; : \; x \in \operatorname{im} D_\eta, x \neq 0 \right\}, \end{aligned} \tag{26}$$

whereas we also have

$$\begin{aligned} \| \, D_{\chi^\Upsilon} \|^{op} &= \sup\left\{ \frac{\| \, D_\chi \pi_{\eta \circ \Upsilon} x \, \|}{\| \, x \, \|} \; : \; x \in \mathbb{R}^{d_X}, x \neq 0 \right\} \\ &= \sup\left\{ \frac{\| \, D_\chi x \, \|}{\| \, x \, \|} \; : \; x \in \operatorname{im} D_{\eta \circ \Upsilon}, x \neq 0 \right\}. \end{aligned} \tag{27}$$

Since $\operatorname{im} D_{\eta \circ \Upsilon} \subseteq \operatorname{im} D_\eta$, we can conclude from Equations (26) and (27) that

$$\| \, D_{\chi^\Upsilon} \|^{op} \leq \| \, D_\chi \|^{op}. \tag{28}$$

Since the intention of using Υ is to reduce the dimension of the parameter space, we can safely assume that $\operatorname{im} D_{\eta \circ \Upsilon}$ is a strict subset of $\operatorname{im} D_\eta$, in practice in fact of high co-dimension. Under this assumption, Equation (28) will *generically be a strict inequality*. The only case in which this does not hold, is when the right-singular vector of D_χ, corresponding to the largest singular value, lies in the image of $D_{\eta \circ \Upsilon}$. Obviously, this is generically not the case for a subspace of positive co-dimension.

The extent to which $\| \, D_{\chi^\Upsilon} \|^{op}$ is smaller than $\| \, D_\chi \|^{op}$, and so the extent to which we have improved numerical stability using Υ as parametrization, of course depends very much on the details of the situation. In the special case that we use only a single dimension to parameterize Θ, i.e. $d_X = 1$, we can make the operator norm $\| \, D_{\chi^\Upsilon} \|^{op}$ more explicit. Let us now consider this univariate case.

We use Equation (27) to compute the operator norm, so suppose $x \in \operatorname{im} D_{\eta \circ \Upsilon}$. This is of the form $x = \lambda D_\eta D_\Upsilon$ for some non-zero $\lambda \in \mathbb{R}$, so

$$D_{\chi^\Upsilon} x = D_\Upsilon \, (D_{\eta \circ \Upsilon})^+ x = D_\Upsilon \, (D_\eta A)^+ \, (\lambda D_\eta D_\Upsilon) = \lambda D_\Upsilon,$$

because $(D_\eta D_\Upsilon)^+ D_\eta D_\Upsilon = 1$. It follows that

$$\frac{\| D_{\chi^\Upsilon} x \|}{\| x \|} = \frac{|\lambda| \| D_\Upsilon \|}{|\lambda| \| D_\eta D_\Upsilon \|} = \frac{\| D_\Upsilon \|}{\| D_\eta D_\Upsilon \|},$$

for any $x \in \operatorname{im} D_{\eta \circ \Upsilon}$. By Equation (27) it then follows that

$$\| D_{\chi^\Upsilon} \|^{op} = \frac{\| D_\Upsilon \|}{\| D_\eta D_\Upsilon \|}. \tag{29}$$

Given the formulation of Υ in Equation (19), we have

$$D_\Upsilon = a_\Upsilon \in \mathbb{R}^{d_X \times d_\Theta}.$$

In Section 4, we will use the expression in Equation (29) to quantify numerical stability of the proposed method.

4. Application

In this section, we perform a calibration to historical equity option market data. We use the Heston model for numerical experiments. We calibrate by using a basic calibration method (*Basic*), wherein we calibrate the full parameter space Θ, and the proposed method (*Proposed*) in, respectively, Sections 4.1 and 4.2. We compare the results and we also analyse the numerical stability. In Section 4.3, we compare the calibrated models by computing the present and future embedded option values of a fictive insurance company.

4.1. *The Heston model*

Under the Heston model, we model the stock index S_u, the variance process v_u and the money market account B_u under the risk-neutral measure $\mathbb{Q}$ by

$$\begin{cases} dS_u &= (r-q)S_u \, du &+ \sqrt{v_u} S_u \, dW_u^S, & S_0 = 1, \\ dv_u &= \kappa_v(\bar{v} - v_u) \, du &+ \sigma_v \sqrt{v_u} \, dW_u^v, & v_0 \geq 0, \\ dB_u &= rB_u \, du, & B_0 = 1, \end{cases} \tag{30}$$

with $Z_u = \{S_u, v_u, B_u\}$ and $u \in [0, T_u]$, r denotes the risk-free interest rate, q the dividend yield, κ_v the speed of mean reversion, $\bar{v}$ the long-term variance level, σ_v the volatility of variance and ρ the correlation between the Wiener processes dW^S and dW^v.

We consider the dividend yield to be zero and the interest rate r is equal to the observed interest rate in the market, such that we have $\Theta = \{\kappa_v, v_0, \bar{v}, \sigma_v, \rho\}$. The volatility parameter σ_v affects the kurtosis (peak) of the probability distribution function of equity (log) returns. The lower the volatility of variance parameter, the higher the kurtosis (peak). The correlation parameter ρ affects the skewness of the probability distribution function of equity (log) returns. The lower the correlation parameter (between the stock index and the variance process), the higher the skewness, i.e. heavy tails to the left. Hence, these parameters affect the modelling of IVs in the strike dimension. In equity, ρ is often negative. The Heston model enables semi-analytical

pricing of equity index put/call options via the so-called COS method (Fang and Oosterlee, 2008). We use this method for valuation.

4.2. The calibration data

We use IV surfaces of S&P-500 equity index options. Multiple S&P-500 index options are quoted with different strike levels and maturities, but we use the most liquid option data. For calibration, we use strikes $K = \{0.8, 1, 1.2\}$ and maturities $T = \{0.25, 1, 2\}$, with $d_K = 3$ and $d_T = 3$ and hence $d_V = 9$. For notation, we use $V_t^{i,j}$ with $i = 1, \ldots, d_K$ and $j = 1, \ldots, d_T$ to refer to the option value with strike K_i and maturity T_j. We define a discrete set of (historical) times $\mathcal{T} = \{t_1, t_2, \ldots, t_N\}$ for which we can observe the process V_t for $t \in \mathcal{T}$, where N denotes the number of observations. We assume that the index $n = 1, \ldots, N$ refers to time $t_n \in \mathcal{T}$. We use historical data available from June 2005 until March 2014 with a monthly frequency, which leads to $N = 106$ observations.

In the literature, much research has been performed to study the empirical behaviour of IV surfaces, see for example (Cont and Da Fonseca et al., 2002) and the references therein. It turns out that for a given maturity, the IVs increase when relative strike levels decrease.[2] Long maturing options are less volatile than short maturing options. The volatility of the IV is level dependent. That is, the higher the (average) level of the IV, the more skewed the IV surface and hence more volatile. This may be due to the fact that investors sell their call options and buy put options for protection in case of high volatility, i.e. in times of crisis.

Given these empirical facts, IV surfaces tend to dynamically change over time. That is, the skew or smile in the strike dimension and the term structure of IVs changes dynamically over time. In times of crisis the (average) IV level is typically higher, the mean reversion from short-term IVs to long-term IVs is faster and the IV surface is more skewed. Opposite IV structures hold in less volatile periods. The VIX index is a popular measure of the IV of short maturing S&P 500 index options. It represents a measure of the market's expectation of equity market volatility over the next 30-day period.

4.3. The basic calibration approach

In the basic calibration approach, we calibrate the full parameter space Θ_t at each historical time t by using the calibration function χ in Equation (2). We let $\eta_{i,j}$ refer to the option with strike K_i and maturity T_j with $i = 1, \ldots, d_K$ and $j = 1, \ldots, d_T$. We set the calibration bounds equal to $\mathcal{D}_{\Theta,min} = \{0.01, 0.001, 0.001, 0.1, -1\}$ and $\mathcal{D}_{\Theta,max} = \{5, 0.5, 0.5, 1, -0.1\}$. We specify small lower limits for the volatility parameters to avoid numerical difficulties and the upper limits are specified based on expert opinion. We note that for each time we (independently) calibrate 5 parameters to obtain the best fit with the market data, so in total we calibrate $5 \times 106 = 530$ parameters.
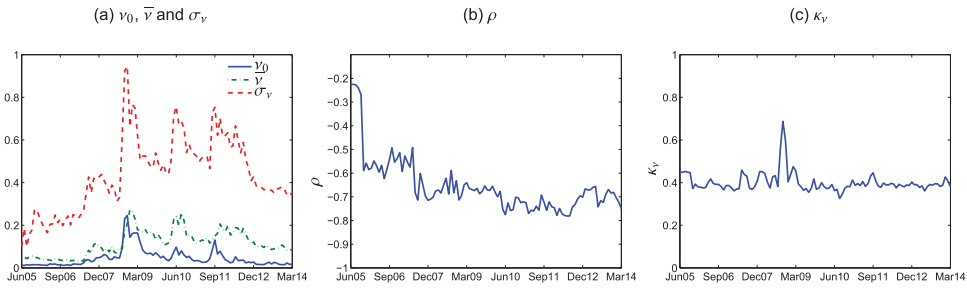
Figure 2. Model parameters Θ_t.

Table 1. Quality-of-fit of Heston model.

	$K = 0.8$			$K = 1.0$			$K = 1.2$		
	$T = 0.25$	$T = 1$	$T = 2$	$T = 0.25$	$T = 1$	$T = 2$	$T = 0.25$	$T = 1$	$T = 2$
Av. SSE ($\times 10^{-5}$)	0.25	0.15	0.15	0.35	0.26	0.04	0.08	0.09	0.05
Max. SSE ($\times 10^{-4}$)	0.49	0.07	0.11	0.16	0.13	0.02	0.20	0.09	0.09

4.3.1. *Quality-of-fit*

The Heston model fits well to the available data. The R^2 errors are > 0.99 for all times $t \in \mathcal{T}$, see Figure 2(a), which is highly satisfactory. To gain insight in the quality-of-fit of separate parts of the option market data, we show in Table 1 the sum squared errors (SSE) per option and for all $t \in \mathcal{T}$.

The fit is worse for shorter maturing options. This is due to the fact that the skew in the maturity dimension from the Heston model decreases linearly (Gatheral, 2006), whereas the market data contains more curved behaviour. The Heston model is hence not powerful enough to model such behaviour.

4.3.2. *Numerical stability of Θ*

The calibrated model parameters Θ are visualized in Figure 2

The parameters κ_v and ρ show a constant, but spiky, behaviour around 0.4 and -0.7, respectively. The parameters v_0, $\bar{v}$, σ_v show common behaviour, see Figure 2(a), with correlations

$$\begin{pmatrix} 1 & & \\ \rho_{v_0,\bar{v}} & 1 & \\ \rho_{v_0,\sigma_v} & \rho_{\bar{v},\sigma_v} & 1 \end{pmatrix} = \begin{pmatrix} 1 & & \\ 0.65 & 1 & \\ 0.76 & 0.90 & 1 \end{pmatrix}. \tag{31}$$

The determinant of this correlation matrix is close to zero 0.08. The determinant of a correlation matrix approaches zero when some of the variables are strongly correlated. The long-term variance level shows lagged behaviour relative to the initial variance, the lag one cross-correlation is 0.75.

We recall from Section 3.4 that $\| D_\eta \|^{op}$ measures the numerical stability of V to Θ and $\| D_\chi \|^{op}$ measures the numerical stability of Θ to V. The operator norm is equal to the maximum singular value and, in particular, we have showed that the singular values of D_χ are equal to the reciprocals of the singular values of D_η. This means that, the

(a) Singular values of D_η (b) Singular values of D_χ (c) The VIX index
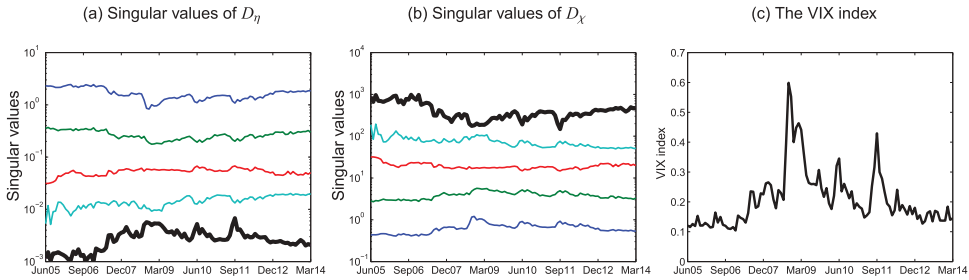
Figure 3. Numerical stability.

higher the minimum singular value $\sigma_{\eta,min}$ of D_η, the lower the maximum singular value $\sigma_{\chi,max}$ of D_χ and hence the better the numerical stability.

In Figure 3(a), we show the singular values of D_η, where we emphasize the minimum singular value $\sigma_{\eta,min}$ in black. In Figure 3(b), we show the singular values of D_χ, where we emphasize the maximum singular value $\sigma_{\chi,max}$ in black. For reference, we show the VIX index in Figure 3(c). Using the VIX index we gain insight in the dynamic behaviour of numerical stability.

We find that there is less variation in the Heston model in volatile periods and vice versa. This means that the Heston model is less sensitive to market changes in volatile periods. There is a significant negative correlation between the market volatility, represented by the VIX index, and maximum singular value of D_χ. The Pearson correlation coefficient over this period is -0.71.

In times of a low VIX index, the IV surfaces in the market are typically relatively flat. The Heston model has too many degrees of freedom in this case and is therefore sensitive to changes in the market data, (Heston, 1993). In times of a high VIX index, the IV surfaces are typically not flat. In this case, more degrees of freedom in the Heston model are used to obtain a good fit. Therefore, in times of high values of the VIX index, the Heston model is less sensitive to changes in the market data.

4.4. The proposed calibration approach

In this section, we apply the proposed methodology in Section 3 to the Heston equity model in Equation (30). We note that the methodology can be applied to other models as well. By using the proposed methodology we aim at reducing the number of parameters in the risk-neutral model to improve stability, but in such a way that the fit to the option market data is satisfactory.

4.4.1. Model specification

We use the mapping Υ in Equation (19) to reduce the dimension of Θ. We consider a 1-dimensional hidden state process, i.e. $d_X = 1$ and vectors $a_\Upsilon, b_\Upsilon \in \mathbb{R}^5$. To model the hidden state process, X_t and the observable process V_t we need to specify the PDFs $f_{X,\psi}$ and $f_{V,\psi}$. For both PDFs, we consider a (simple) normal distribution and an (advanced) non-normal distribution. This results in 4 cases, which we describe below.

We consider the Ornstein-Uhlenbeck (OU) and Cox-Ingersoll-Ross (CIR) models (Brigo and Mercurio, 2007) for modelling the hidden process $X_{1,t}$ and $X_{2,t}$, which are, respectively, presented by the following SDEs:

$$
\begin{aligned}
dX_{1,t} &= \kappa_{1,X}\left(\bar{X}_1 - X_{1,t}\right)dt &+ \sigma_{1,X}\,dW^{X_1}(t), & X_{1,0} \in \mathbb{R}, \\
dX_{2,t} &= \kappa_{2,X}\left(\bar{X}_2 - X_{2,t}\right)dt &+ \sigma_{2,X}\sqrt{X_{2,t}}\,dW^{X_2}(t), & X_{2,0} \geq 0,
\end{aligned}
\tag{32}
$$

where $\kappa_{,X} \in \mathbb{R}_+$ is the speed of mean reversion, $\bar{X}_{.} \in \mathbb{R}_+$ a long-term state parameter and $\sigma_{,X} \in \mathbb{R}_+$ a volatility parameter (with $\cdot = 1, 2$). The transitional PDFs for, respectively, the Vasicek and CIR model are known in analytical form:

$$
f_{X,\psi}(x_{1,n}|x_{1,n-1}) = f_{norm}\left(x_{1,n}; x_{1,n-1}e^{-\kappa_{1,X}\Delta t} + \frac{\bar{X}_1}{\kappa_{1,X}}\left(1 - e^{-\kappa_{1,X}\Delta t}\right), \frac{\sigma_{1,X}^2}{2\kappa_{1,X}}\left(1 - e^{-2\kappa_{1,X}\Delta t}\right)\right),
\tag{33}
$$

$$
f_{X,\psi}(x_{2,n}|x_{2,n-1}) = \frac{1}{C}f_{nccs}\left(x_{2,n}/C; d, \lambda_{X,t}\right),
\tag{34}
$$

where f_{norm} denotes the PDF of a normal distribution and f_{nccs} denotes the PDF of a Non-Central Chi-Squared (NCCS) distribution where

$$
C = \frac{\sigma_{2,X}^2(1 - e^{-\kappa_{2,X}\Delta t})}{4\kappa_{2,X}}, \quad d = \frac{4\kappa_{2,X}\bar{X}_2}{\sigma_{2,X}^2} \quad \text{and} \quad \lambda_{X,t} = \frac{4\kappa_{2,X}x_{2,n-1}e^{-\kappa_{2,X}\Delta t}}{\sigma_{2,X}^2(1 - e^{-\kappa_{2,X}\Delta t})},
\tag{35}
$$

with $\Delta t = t_n - t_{n-1} = \frac{1}{12}$, because we use a monthly frequency of the historical data.

We consider a normal and a skewed normal distribution for modelling the observable process V_t conditional on the hidden process X_t. The two choices for modelling the observable process denoted by $V_{1,t}$ and $V_{2,t}$ are given by the following PDFs:

$$
f_{V_1,\psi}(v_{1,n}|x_{.,n}) = f_{norm}\left(v_{1,t}; \eta(\Upsilon(x_{.,t})), \sigma_{1,V}\right),
\tag{36}
$$

$$
f_{V_2,\psi}(v_{2,n}|x_{.,n}) = \frac{2}{\sigma_{2,V}}f_{norm}\left(\frac{v_{2,t} - \eta(\Upsilon(x_{.,t}))}{\sigma_{2,V}}; 0, 1\right)F_{norm}\left(\alpha_V\left(\frac{v_{2,t} - \eta(\Upsilon(x_{.,t}))}{\sigma_{2,V}}\right); 0, 1\right),
\tag{37}
$$

with $\cdot = 1, 2$ and $\alpha_V \in (-1, 1)$. The skewed normal PDF equals the normal PDF when $\alpha_V = 0$. In the normal distribution, the error between the observed option price and the model option price are centred symmetrically around zero. Under the skewed normal distribution, we are able to model a skewed distribution of these errors via the parameter α_V. This is for example relevant if the Heston model is only able to model certain combinations of K and T of the option market data. We note that negative option values are truncated to zero.

We consider four cases in which we use either the OU or the CIR model for the state process, and either the normal or the skewed normal distributions for the errors of the option prices with respect to the fit. Summarizing, we consider the following 4 cases:

• **Case I**: The state process follows the PDF in Equation (33) and the observable process follows the PDF in Equation (36). The model parameters to be estimated are

$\psi = \{a_Y, b_Y, \kappa_{1,X}, \bar{X}_1, \sigma_{1,X}, \sigma_{1,V}\}$. This results in the estimation of 120 parameters (including the hidden states for $t \in \mathcal{T}$).

• **Case II**: The state process follows the PDF in Equation (33) and the observable process follows the PDF in Equation (37). The model parameters to be estimated are $\psi = \{a_Y, b_Y, \kappa_{1,X}, \bar{X}_1, \sigma_{1,X}, \sigma_{2,V}, \alpha_V\}$. This results in the estimation of 121 parameters (including the hidden states for $t \in \mathcal{T}$).

• **Case III**: The state process follows the PDF in Equation (34) and the observable process follows the PDF in Equation (36). The model parameters to be estimated are $\psi = \{a_Y, b_Y, \kappa_{2,X}, \bar{X}_2, \sigma_{2,X}, \sigma_{1,V}\}$. This results in the estimation of 120 parameters (including the hidden states for $t \in \mathcal{T}$).

• **Case IV**: The state process follows the PDF in Equation (34) and the observable process follows the PDF in Equation (37). The model parameters to be estimated are $\psi = \{a_Y, b_Y, \kappa_{2,X}, \bar{X}_2, \sigma_{2,X}, \sigma_{2,V}, \alpha_V\}$. This results in the estimation of 121 parameters (including the hidden states for $t \in \mathcal{T}$).

In each case, we hence use a much smaller number of parameters than in the basic calibration method where we used 530 parameters. We reduce the number of parameters by a factor ≈ 4.4.

Next, in Section 3.2, we have discussed the estimation of SSHM models. We use the EM algorithm in combination with particle filtering. We have implemented the numerical methods in Matlab. In particle filtering, the estimation results depend on the EM iterations N_{EM}, the number of particles N_{par} and the initial parameter settings. Unfortunately, it is beforehand difficult to specify N_{EM} and N_{par}. Ideally, we specify those values as large as possible to guarantee convergence, but this would slow down the estimation process. There is hence a trade-off to be made. We take $N_{EM} = 50$, $N_{par} = 10$ and set the initial parameters equal to:

$$a_Y = (0.0, 0.0, 0.0, 0.0, 0.0), \quad b_Y = (0.35, 0.1, 0.1, 0.4, -0.7),$$
$$\sigma_{\cdot,X} = \sigma_{\cdot,V} = \kappa_{\cdot,X} = \bar{X}_{\cdot,X} = 0.1, \quad \alpha_V = 0,$$

with $\cdot = 1, 2$. We find that this parameter setting leads to satisfactory results for all experiments.

4.4.2. *Estimated parameters*
The estimated state processes $X_t = x_t$ are shown in Figure 4 for Cases I-IV.

We observe similar estimations. The estimated processes for Cases III and IV are positive due to the underlying CIR process; the NCCS distribution has a positive domain. It turns out that the state processes are highly correlated with the VIX index; the Pearson correlations are

$$\text{Case I} = 0.884, \quad \text{Case II} = 0.922, \quad \text{Case III} = 0.929, \quad \text{Case VI} = 0.936. \tag{38}$$

In Case IV, we find the highest correlation with the VIX index. The estimated parameters ψ are given in Table 2.

The estimated skewness parameter α_V is negative for Cases II and IV, so that the PDF $f_{V,\psi}$ is negatively skewed. This means that market prices are overestimated by the
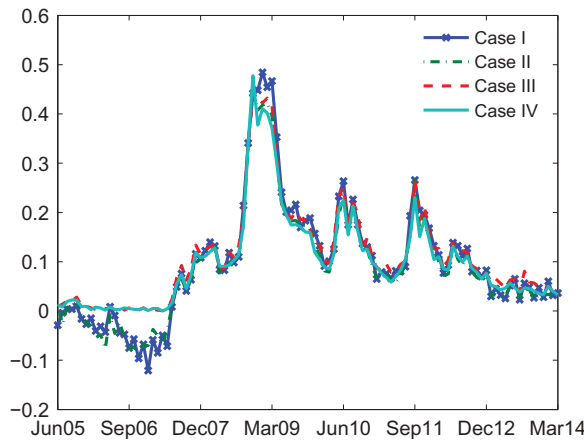
Figure 4. Estimated state process.

Table 2. Overview of model parameters ψ.

	κ_X	$\bar{X}$	σ_X	σ_V	a_V
Case I	0.8379	0.1048	0.1539	0.0050	–
Case II	0.8813	0.0835	0.1364	0.0039	−0.0991
Case III	0.3600	0.1162	0.3506	0.0043	–
Case IV	0.3387	0.1230	0.3461	0.0039	−0.3450

	a_1	a_2	a_3	a_4	a_5	b_1	b_2	b_3	b_4	b_5
Case I	0.1210	0.2725	0.3122	0.3625	−0.06300	0.1914	0.0132	0.1232	0.3475	−0.6647
Case II	0.1100	0.3529	0.3631	0.4043	−0.07100	0.2302	0.0114	0.1133	0.3779	−0.6751
Case III	0.1300	0.3182	0.5336	0.3988	−0.06400	0.2678	0.0111	0.0750	0.3970	−0.6780
Case IV	0.1400	0.3852	0.4296	0.6328	−0.09000	0.4584	0.0111	0.0562	0.4132	−0.6820

Heston model, i.e. the model prices are (on average) higher than the market prices. The PDF $f_{V,\psi}$ corrects for this via the negative α_V parameter.

The estimated long-term states $\bar{X}$ are comparable, but the $\bar{X}$ is highest for Case IV. The speed of mean reversion and correlation parameters are almost constant, i.e. the corresponding elements in a_Υ are approximately zero. The estimated ρ and κ_v are comparable although Cases III and IV estimate some higher mean reversion parameters than Cases I and II. The parameters v_0, $\bar{v}$ and σ_v change dynamically over time because the corresponding elements in the vector a_Υ are non-negative and hence these parameters are directly connected to the dynamic part x_t. We observe that the elements in a_Υ of v_0, $\bar{v}$ and σ_v are all positive which indicates positive correlations between v_0, $\bar{v}$ and σ_v. This coincides with the correlations computed in Equation (31).

4.4.3. *Quality-of-fit*
The R^2-error for Cases I–IV and the basic calibration method is visualized in Figure 5(a,b).

We observe in Figure 5(a) highly satisfactory results, i.e. $R^2 > 0.93$ and by removing the spike at October 2008 we have $R^2 > 0.98$ for all time points. We note that the data quality of October 2008 is bad due to illiquidity in the market. These results are
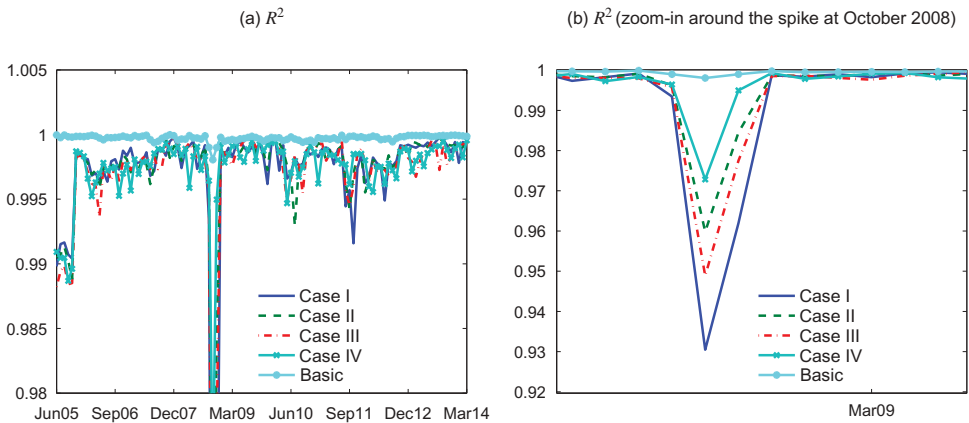
(a) R^2 (b) R^2 (zoom-in around the spike at October 2008)

Figure 5. Quality-of-fit.

obtained for $d_X = 1$. The results typically improve when we increase the dimension d_X; for $d_X = 5$ the results would be similar to the basic calibration approach since all model parameters are used in that case.

The largest error is when the VIX index is highest, i.e. when there is turmoil in the market. In times of market stress, the IV surface is typically more skewed due to (amongst others) the higher demand for put options for protection. Typically, in those periods, correlations are different as compared to more quiet periods. In the proposed methodology, the correlation parameter is estimated by a constant and hence there is not enough flexibility to model an extreme skew pattern.

The Cases I and II perform worst. This is due to the underlying normal distribution for the state process X_t, which is not capable of modelling the heavy tails in the data. We find in Figure 5(b) that the Cases III and IV perform better than Cases I and II, respectively, because the underlying CIR process is able to model the heavy tails in the data. Case IV performs best given the fit to the available data. The mean $R^2 = 0.9974$ is slightly smaller than the R^2 of the benchmark model $R^2 = 0.9997$. Therefore, we favour Case IV. We obtain a similar fit as the basic calibration method, but by using a much smaller parameter space. We recall that we reduce the number of parameters by a factor of 4.4 (from 530 to 121 in case of Case IV).

To gain insight in the performance of separate parts of the option surface, we show in Table 3 the sum squared errors (SSE) per option type and for all $t \in \mathcal{T}$. The results are shown for Case IV.

The overall fit to the market data is still satisfactory. Since we use a much smaller parameter space, we obviously lose flexibility with respect to the basic calibration method. We observe that the model is less capable of calibrating the longer maturing options.

Table 3. Quality-of-fit of the Heston model.

	$K = 0.8$			$K = 1.0$			$K = 1.2$		
	$T = 0.25$	$T = 1$	$T = 2$	$T = 0.25$	$T = 1$	$T = 2$	$T = 0.25$	$T = 1$	$T = 2$
Av. SSE ($\times 10^{-4}$)	0.07	0.09	0.23	0.35	0.10	0.17	0.02	0.13	0.18
Max. SSE ($\times 10^{-3}$)	0.31	0.17	0.21	0.37	0.24	0.10	0.03	0.15	0.21

From Table 2, we observe that the dynamic part, i.e. the non-zero element of a_Υ, is related to the parameters v_0, $\bar{v}$ and σ_v. The parameters κ_v and ρ are approximately constant. This parametrization of Θ is able to model the height of the term structure of IVs (via v_0 and $\bar{v}$), but with approximately a fixed mean reversion κ_v. This might be restrictive in extreme market circumstances. We are able to model dynamic skews in the IV surface via σ_v, but perhaps not well enough in extreme events since the parameters v_0, $\bar{v}$ and σ_v are linked to the same (single) risk driver x_t. Since the correlation parameter is approximately fixed over time, we model fixed skews in the IV surface. This can be restrictive in extreme cases, because skewness is important in high volatile periods; see Figures 5a and 5b.

4.4.4. Numerical stability

By using the proposed methodology, we also aim at improving the numerical stability, see Section 3.4. Since we assume $d_X = 1$ we have 1 singular value (instead of 5 in Section 4.1). We compute the operator norm of D_{χ^Υ} by using Equation (29) and is illustrated in Figure 6. For comparison, we also show the operator norm of D_χ from the basic calibration method in Section 4.1 denoted by *Basic*. We show the results of Case IV, but the results for Cases I–IV are similar.

We observe a major improvement of numerical stability with factors up to 1000 with respect to the *Basic* method. Contrary to the basic calibration method, there is more variation in the Heston model in volatile periods and vice versa. This is due to the lower number of risk drivers, i.e. $d_X = 1 \leq 5 = d_\Theta$. This means that the Heston model under the proposed method is more sensitive to market changes in volatile periods. There is a significant positive correlation between the market volatility (represented by the VIX index) and the (maximum) singular value of D_{χ^Υ}. The Pearson correlation coefficient between $\| D_{\chi^\Upsilon} \|^{op}$ and the VIX index over this period is 0.89. Despite the variation in volatile periods, the *Proposed* method is for all $t \in \mathcal{T}$ more stable than the *Basic* method.
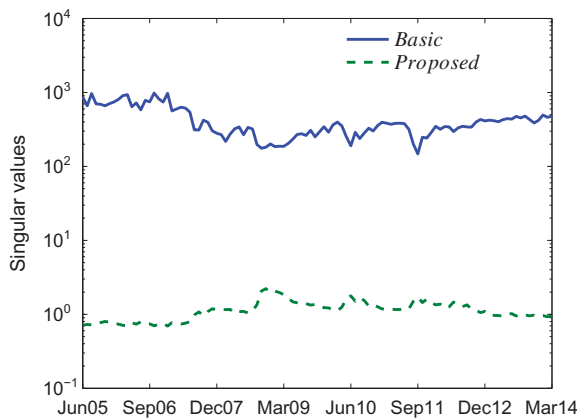

Figure 6. Numerical stability of the χ and χ^Υ mappings.

4.4.5. *Matching empirical facts*

We use the Υ mapping in Equation (19) to reduce the dimension of the Θ space. As mentioned in Section 3.3, the mapping consists of a static part and a dynamic part X_t, which are estimated as is described in Section 4.2. We have showed that the proposed method leads to a highly satisfactory fit, obtained with numerical stability. We aim to get insight in the ability of the proposed method to model the relevant empirical facts in the data. The results described below are generated using the settings from Case IV.

In Figure 7, we show the IV surfaces belonging to $x_{min} = 0$, $x_{av} = 0.15$ and $x_{max} = 0.5$, i.e. the surface belonging to the minimum, average and maximum values of the hidden states $X_t = x_t$.

In the market data, we observe that the level of the IV surface is high in times of turmoil, i.e. in times of a high VIX index. The estimated hidden states are highly correlated with the VIX index, see Equation (38). We also observe that when the level of the IV surface is high, then there is more skew for the short-term options. This is also observed in historical option market data.

To gain more insight in the match with the empirical facts we perform a principal component analysis (PCA) to the historical S&P-500 equity index IV as seen in the market and the corresponding IVs generated by the proposed model. We investigate whether the PCA factors are comparable. PCA factors are orthogonal (zero correlation) linear combinations of a time series that explain the largest part of the total variance. These factors are determined by assigning a weight (loading) on each of the input time series. The weights are determined such that the resulting factors describe the largest part of the joint movements (correlations) of the input time series.

We perform a PCA to the historical S&P-500 equity index IV market data and to the IV surfaces calibrated by the *Proposed* model. We use the option maturities 0.25, 0.50, 1.00, 1.50 and 2.00 years and strike levels 80%, 90%, 100%, 110% and 120%. It turns out that the first PCA component of the historical data accounts for 97.6% and the first PCA component of the *Proposed* model 98.0%. The first components, that account for most of the variance, are very similar, so that the proposed model is able to model the dynamics in the historical IV surfaces rather well. The first PCA factors are shown in Figure 8.

We observe indeed a satisfactory resemblance of the first PCA factor. The most significant skews are observed for out-of-the-money strike levels (for put options) and short maturing options. We find an almost flat/linear IV structure for the in-the-money strike levels (for put options) and long maturing options.
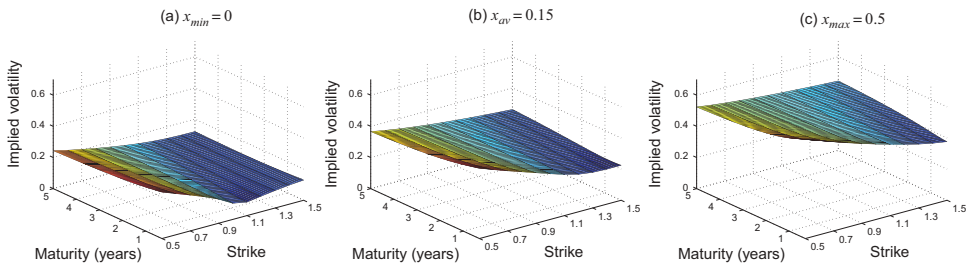

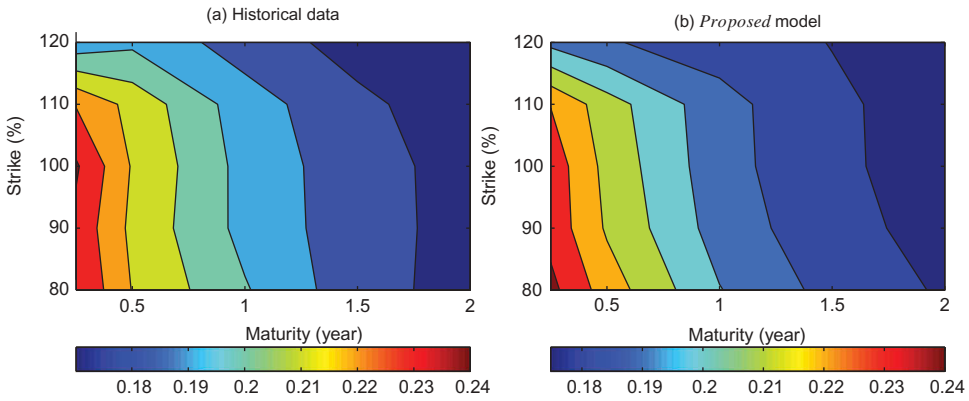**Figure 7.** IV structures in the maturity dimension.
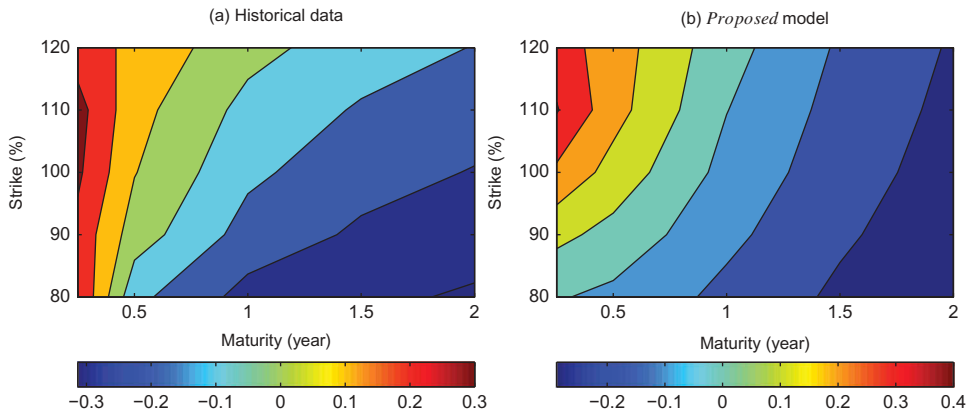
Figure 8. First PCA factor.

Figure 9. Second PCA factor.

The second PCA factors account for 1.5% and 1.4% of the total variance of, respectively, the historical data and the *Proposed* model. The second PCA factors are shown in Figure 9.

Although the IV pattern is similar for different maturities and strikes, we observe some deviations in the second PCA factors. We note however that, the second PCA factor only contributes for a small part to the total variance, so that these deviations are less important.

4.5. *Impact study*

In Sections 4.1 and 4.2, we have calibrated the *Basic* and *Proposed* methods to option market data. We found that the calibration fit is highly satisfactory for both methods. However, we have showed that for the *Proposed* method the model parameters are less sensitive to changes in the option market data. Therefore, we favour the *Proposed* method. In this section, we compare both methods by valuing an embedded option of an insurance company. That is, we compute the composite functions $\tilde{\eta} \circ \chi$ and $\tilde{\eta} \circ \chi^{\Upsilon}$,

which we discussed in Section 3.4. In Section 4.3.1 we focus on computing present embedded option values, and in Section 4.3.2 we compute future embedded option values.

We assume a fictive insurer XYZ which sells unit-linked products[3] of which we consider a simplified version. We assume the unit-linked product is of European type, where policyholders capital is invested in a single equity series. The unit-linked product then becomes an equity index put option. We consider a 120% in-the-money put option with maturity equal to 10-years. The embedded option value is relevant to the insurers' balance sheet.

4.5.1. *Computing present embedded option values*

For comparison, we use the *Basic* and *Proposed* methods to compute the embedded option value of insurance company XYZ, which we respectively denote by $\tilde{V}^B$ and $\tilde{V}^P$. For the proposed method, we use Case IV.

We first consider the computation of present embedded option values, and an option notional of 500. We use the operator norm to measure numerical stability, see Section 3.4. In Figure 10(a), we compare the embedded option values $\tilde{V}^B$ and $\tilde{V}^P$. We gain insight in the *Basic* and *Proposed* models in Figure 10(b) by computing the maximum singular values of $\| D_{\tilde{\eta} \circ \chi} \|^{op}$ and $\| D_{\tilde{\eta} \circ \chi^\gamma} \|^{op}$ (see Section 3.4).

From Figure 10, we observe that differences between the basic and proposed methods can be substantial. For the embedded option value, we observe differences up to ∈25, which is around 17% relative difference.

It depends on how large the embedded option value is with respect to the total liability value to gain insight in the impact to the surplus, for example. The surplus is the difference between the assets and the liabilities on the insurer's balance sheet and is an important quantity of an insurer. The larger the embedded option value with respect to the total liability value, the larger the impact of the differences between *Basic* and *Proposed* to the surplus. Considering the balance sheet in Table 4, the impact of a 17% difference in the embedded option value would result in relative difference of 20% in the surplus value.
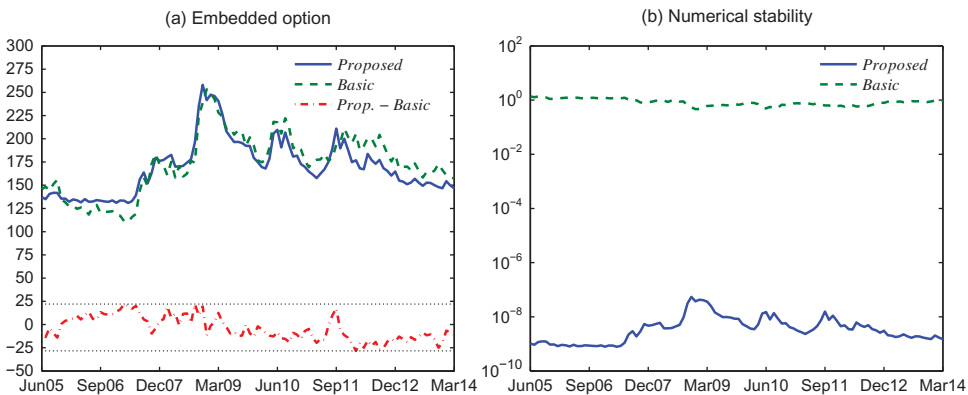

Figure 10. Comparison of the *Basic* and *Proposed* models.

We observe in Figure 10(b) that, as expected, the *Proposed* method is more stable than the *Basic* method.

4.5.2. *Computing future embedded option values*

We compare the proposed methodology (*Proposed*) with a basic modelling approach (*Basic*), which is often applied in practice, to compute future embedded option values. The starting balance sheet in Euros of insurance company XYZ is summarized in Table 4. We assume a perfectly matching asset portfolio for the guaranteed liability cash flows, an asset portfolio (denoted by A) and an embedded option (denoted by $\tilde{V}$). The notional of the embedded option is determined in such a way that the embedded option value is equal to €100 for the *Basic* and *Proposed* models.

To project the balance sheet for $t \geq 0$, the asset process A_t and the embedded option process $\tilde{V}_t$ are relevant. The surplus (available capital) S_t is computed by the difference of the value of the asset and the value of the liabilities:

$$S_t = A_t - \tilde{V}_t. \tag{39}$$

Other balance sheet items are neglected in this simplified example.

We show the impact of the *Basic* and *Proposed* methods to the embedded option value, the surplus of the balance sheet and the required capital. Special attention is devoted to the $t = 1$-year horizon because this is important for Solvency II computations. Under *Proposed* we again consider the estimated model Case IV from Section 4.2 for the model parameters. Under *Proposed*, we forecast the model parameters Θ_t using the transitional PDF $f_{X,\psi}$ in combination with Equation (19).

Under the *Basic* method, valuation is based on a fixed parameter setting (as in (Bauer, Reuss, and Singer, 2012), (Feng et al., forthcoming)), so independent of $t \geq 0$: $\Theta_t = \Theta_0$. For comparison, we use the same parameter settings as the initial parameters of Case IV:

$$\Theta_0 = \{\kappa_v, v_0, \bar{v}, \sigma_v, \rho\}_0 = a_\Upsilon x_0 + b_\Upsilon = \{0.4630, 0.0236, 0.0702, 0.4338, -0.6830\},$$

where a_Υ and b_Υ are found in Table 2. Using the *Proposed* and *Basic* valuation models, we are able to compare the approximation of the embedded option $\tilde{V}_t^P$ and $\tilde{V}_t^B$. Via Equation (39) we can also compare the surplus values which we denote by S_t^P and S_t^B.

In this stylized example, we assume that $Y_t = \{A_t, r_t, X_t\}$ models the uncertainty in the real-world financial market. Therefore, these processes are modelled under the $\mathbb{P}$ measure, see Section 2.2. We assume that the return equity portfolio A_t is modelled by a Black-Scholes Vasicek model. So, we assume a fixed volatility, but stochastic interest

Table 4. Balance sheet of XYZ at $t = 0$.

Assets		Liabilities	
Matching portfolio €700			
Equity	€200	Embedded option	€100
		Surplus	€100
Total	€900	Total	€900

rates generated by the Vasicek model (Brigo and Mercurio, 2007). The state process X_t follows the estimated CIR process of Case IV from Section 4.2.

$$dA_t = \mu_A A_t dt + \sigma_A A_t dW_t^A, \quad \left(A_0 = 200, \mu_A = 0.06, \sigma_A = 0.2\right) \tag{40}$$

$$r_t = \kappa_r(\bar{r} - r_t)dt + \sigma_r dW_t^r, \quad (r_0 = 0.02, \bar{r} = 0.04, \kappa_r = 0.02, \sigma_r = 0.01) \tag{41}$$

$$X_t = \kappa_X(\bar{X} - X_t)dt + \sigma_X \sqrt{X_t}dW_t^X,$$
$$(X_0 = 0.0326, \bar{X} = 0.1230, \kappa_X = 0.3387, \sigma_X = 0.3461). \tag{42}$$

Since we are interested in computing the embedded option values $\tilde{V}_t$, we only consider the estimated PDF $f_{X,\psi}$ from Case IV to generate the model parameters Θ via Equation (19). The PDF $f_{V,\psi}$ is not needed in this experiment.

We remark that the *Basic* model only depends on the stochastic interest rates, because the model parameters are kept fixed during the simulation. Besides the stochastic interest rates, IV risk is also taken into account for the *Proposed* model because the model parameters follow a stochastic process. We assume the following correlation matrix Σ between the Wiener processes (W_t^A, W_t^r, W_t^X):

$$\Sigma = \begin{pmatrix} 1 & & \\ \rho_{A,r} & 1 & \\ \rho_{A,X} & \rho_{r,X} & 1 \end{pmatrix} = \begin{pmatrix} 1 & & \\ 0.10 & 1 & \\ -0.60 & -0.05 & 1 \end{pmatrix}. \tag{43}$$

The joint modelling of X_t and r_t is important, because both risk drivers affect the embedded option value. We assume $\rho_{A,X}$ is negative, because the process X_t is highly correlated with IV, which in turn are negatively correlated with the asset returns. In the numerical experiments, we vary the correlation $\rho_{r,X}$ between -0.8 and 0.7 to gain insight in the impact. We generate $N_{\mathbb{P}} = 10,000$ real-world scenarios of (A_t, r_t, X_t) with a $T_t = 1$-year horizon. For each time point, we value the Unit-Linked option under the *Proposed* and *Basic* methods using the Heston pricing formula (Fang and Oosterlee, 2008).

4.5.3. The embedded option process

In Figure 11(a,b), we compare the embedded option process for the *Proposed* and *Basic* methods. In Figure 11(a), density plot is shown. In Figure 11(b), insight is provided in the (right) tail risk. We compute tail risk by means of Value-at-Risk (VaR), i.e. we compute the value $\bar{V}$ such that $\mathbb{P}(\tilde{V}_1 \leq \overline{V}) = \alpha$, where we vary $\alpha \in [0.9, 1]$ (the right tail is relevant). The grey-filled area in Figure 11 is computed by varying the correlation parameter $\rho_{r,X}$ between -0.8 and 0.7.

We observe that the PDF of *Proposed* has a wider range of embedded option values, i.e. the embedded option process is more volatile under the *Proposed* method. This is due to the extra stochastic process for modelling the equity volatility part of the Heston model. Due to the wider PDF, the right tail is more heavier in the *Proposed* method, which results in differences up to $\in 30$. Varying the correlation parameter $\rho_{X,r}$ results in different valuations of the *Proposed* method. The upper line of the filler area belongs to $\rho_{X,r} = 0.7$ and the lower line to $\rho_{X,r} = -0.8$.
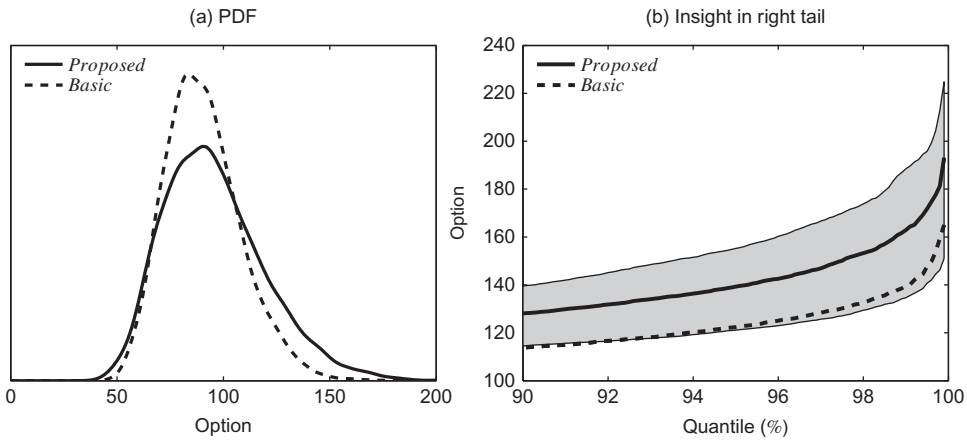
Figure 11. Comparison of the embedded option values.

4.5.4. *The surplus process*

In Figure 12(a,b), we compare the surplus process S_t for *Proposed* and *Basic*. In Figure 12(a), a density plot is shown and in Figure 12(b) insight is provided in the tail risk. We again compute tail risk by means of Value-at-Risk (VaR), i.e. we compute the value $\bar{S}$ such that $\mathbb{P}(S_1 \leq \bar{S}) = \alpha$, where we vary $\alpha \in [0, 0.1]$ (the left tail is relevant). The grey-filled area in Figure 12(b) is again computed by varying the correlation parameter $\rho_{r,X}$ between -0.8 and 0.7.

We again observe that the PDF of the *Proposed* method has a wider range of values, although the impact is smaller compared to the embedded option process. Due to the wider PDF, the left tail is more heavier using the *Proposed* method, which results in differences up to €25. The choice of the correlation $\rho_{A,X}$ is crucial in this case. When A_t decreases, then X_t increases due to the negative correlation, so that the value of the embedded option $\tilde{V}_t$ increases. The latter results in a decrease of the surplus.
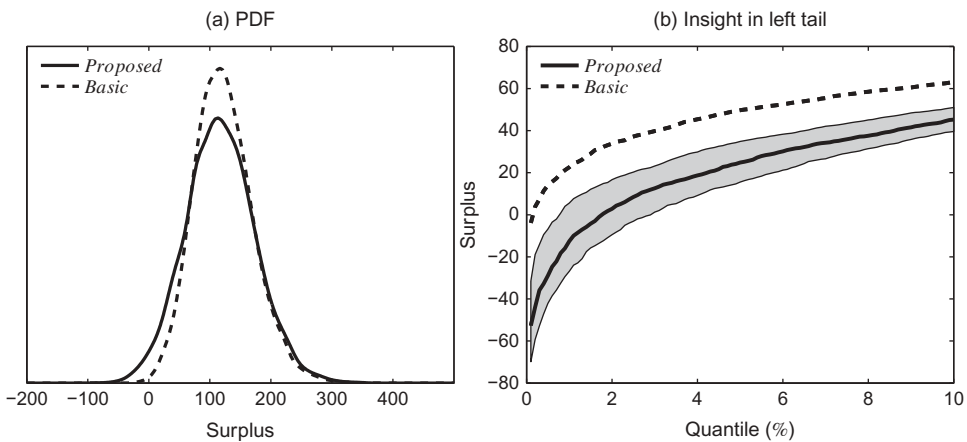

Figure 12. Comparison of the surplus.

4.5.5. The required capital

Last, we analyse the impact of computing the required capital (RC), which is comparable to the SCR under Solvency II. Insurers use internal models to compute required capital for internal steering and/or reporting to the regulator. The required capital is the amount of capital the insurer must hold against unforeseen losses during a one-year period. The required capital is the VaR of a loss-function on a certain horizon (often the 1-year horizon). A common loss function (see (Bauer, Bergmann, and Reuss, 2010)) is

$$L_t = S_{t-\Delta}(1 + R_{t-\Delta,\Delta}) - S_t \quad \text{for} \quad (\Delta = 1, t \geq \Delta),$$

with $R_{t-\Delta,\Delta}$ the Δ-year risk-free rate in year $t - \Delta$. This loss function will be used in our experiments. Based on a one-year horizon and for a certain confidence level $\alpha \in [0, 1]$, the RC, RC_α, is computed by:

$$\mathbb{P}(L_1 \geq RC_\alpha) \leq 1 - \alpha, \quad (L_1 = S_0(1 + R_{0,1}) - S_1), \tag{44}$$

where we set $R_{0,1} = 0.025$ is the one-year risk-free rate at $t = 0$. Hence, the probability that the loss over one year exceeds the RC is less or equal to $1 - \alpha$. In practice, this confidence level is often set to 99.5%.

In Figure 13(a,b), we compare the loss function for *Proposed* and *Basic*. In Figure 13(b), the PDF is illustrated of the loss function. In Figure 13(b), insight is provided in the right tail of the distribution. The grey-filled area in Figure 13(b) is again computed by varying the correlation parameter $\rho_{r,X}$ between -0.8 and 0.7.

We again observe that the PDF of the *Proposed* method has a wider range of values compared to the *Basic* method. Due to the wider PDF, the right tail is more heavy in the *Proposed* method, which results in differences up to €40.

In Table 5, we give an overview of numerical values of the required capital for different confidence levels. Insurance companies with an higher credit rating are obligated to compute their required capital with a lower confidence level α. The values
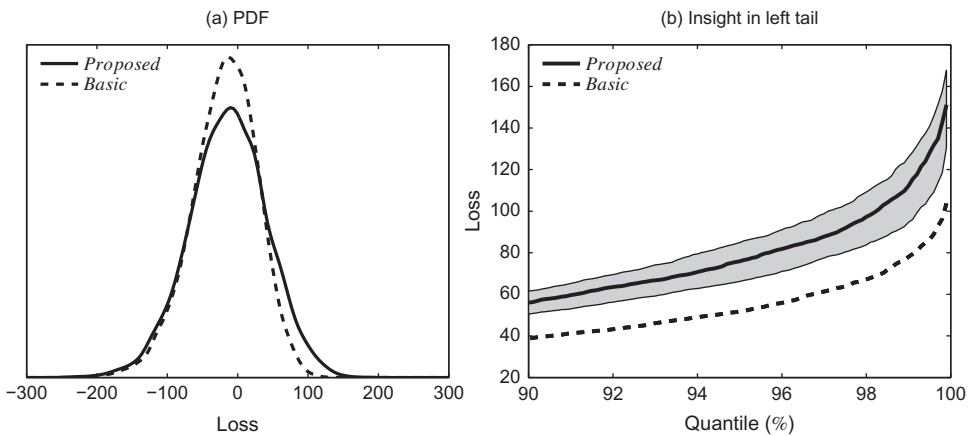

Figure 13. Comparison of the loss function.

Table 5. Comparison of the required capital (in Euro).

	99.5%	97.5%	95.0%	90.0%
Proposed	121(0.81)	92(1.08)	74(1.34)	56(1.78)
Basic	86(1.16)	64(1.53)	51(1.92)	37(2.67)

in parenthesis are the solvency ratio, i.e. the ratio between the surplus and required capital.

We observe that the absolute difference in RC between *Proposed* and *Basic* becomes smaller when the confidence value α decreases from 99.5% to 90.0%. For $\alpha = 99.5\%$, the difference is ∈35, which is 35% of the surplus. The differences in the solvency ratio is even larger: 81% for *Proposed* and 116% for *Basic*.

These differences show the relevance of taking the IV risk factor into account, i.e. to apply the *Proposed* method. We note however that the impact of the differences between *Basic* and *Proposed* heavily depends on the initial balance sheet settings and in particular the value of the embedded option portfolio with respect to the total liability value, which is in this simplified example 14% (100/700).

5. Conclusions

We presented a method for modelling risk-neutral models in a real-world scenario model to perform nested Monte Carlo simulations. This is important for ex-ante risk management applications for insurance companies. In such applications, an insurer is required to compute their embedded option values at a future point in time. We make use of the well-known SSHM models, which provide for a flexible modelling framework. By introducing a hidden state process, we are able to reduce the dimension of the calibration problem. In this way, we reduce the computation time and we improve the numerical stability of the model parameters with respect to option market data. The latter is desired in practice for transparency and stable valuations of embedded options.

We apply the proposed method to the well-known risk-neutral Heston model. Although the Heston model consists of five model parameters, i.e. the mean reversion, initial variance, long-term variance, volatility of variance correlation parameters, we show that in our proposed methodology a one-dimensional state process already results in highly satisfactory calibration results. That is, we obtain a maximum dimension reduction. We also show that numerical stability of the model parameters with respect to option market data is greatly improved. We measure numerical stability by means of the operator norm.

To show the relevance of our method, we compare the estimated Heston model from the proposed methodology to a basic parameter setting. We use both models to compute the present and future values of a (simplified) unit-linked product of a fictive insurer. The results differ substantially and especially the tails of the distributions differ, which are important in practice. Given the large differences in this simplified case study, we advice to use the proposed methodology for calibration, valuation and simulation.

Notes

1. We use similar notation as in Doucet 2009 tutorial (Doucet and Johansen, 2009).
2. Possible explanations for this phenomenon are the (see (Kamal and Gatheral, 2010)) negative correlation between asset returns and volatility changes (leverage effect). Big jumps in the asset (spot) price tend to be downwards rather than upwards.
3. Unit-linked product are guaranteed investment return products.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Atkinson, K., and W. Han. 2005. "Linear Operators on Normed Spaces." In *Theoretical Numerical Analysis: A Functional Analysis Framework*, 51–111. Berlin: Springer.

Bauer, D., A. Reuss, and D. Singer. 2012. "On the Calculation of the Solvency Capital Requirement Based on Nested Simulations." *Astin Bulletin* 42 (02): 453–499.

Bauer, D., D. Bergmann, and A. Reuss. 2010. "Solvency II and Nested Simulations–A Least-Squares Monte Carlo Approach." Proceedings of the 2010 ICA Congress, Syndey.

Bauer, D., D. Bergmann, and R. Kiesel. 2010. "On the Risk-Neutral Valuation of Life Insurance Contracts with Numerical Methods in View." *Astin Bulletin* 40 (01): 65–95. doi:10.2143/AST.40.1.2049219.

Brigo, D., and F. Mercurio. 2007. *Interest Rate Models-Theory and Practice: With Smile, Inflation and Credit*. Berlin: Springer.

Cappé, O., E. Moulines, and R. Tobias. 2009. *Inference in Hidden Markov Models*. Berlin: Springer.

Carr, P., and L. Wu. 2010. "A New Simple Approach for For Constructing Implied Volatility Surfaces." *Preprint*. http://papers.ssrn.com/sol3/papers.cfm?abstract id=1701685, 2010.

Cont, R., and J. da Fonseca. 2002. "Dynamics of Implied Volatility Surfaces." *Quantitative Finance* 2: 45–60. doi: 10.1088/1469-7688/2/1/304.

Cramwinckel, J., S. Singor, and A. L. Varbanescu. 2015. "Fins: A Framework for Accelerating Nested Simulations on Heterogeneous Platforms." In *European Conference on Parallel Processing*, 246–257. Berlin: Springer.

Delyon, B., M. Lavielle, and E. Moulines. 1999. "Convergence of a Stochastic Approximation Version of the EM Algorithm." *The Annals of Statistics* 27 (1): 94–128.

Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. "Maximum Likelihood from Incomplete Data via the EM Algorithm." *Journal of the Royal Statistical Society. Series B (Methodological)* 39 (1): 1–38.

Doucet, A., and A. M. Johansen. 2009. "A Tutorial on Particle Filtering and Smoothing: Fifteen Years Later." *Handbook of Nonlinear Filtering* 12: 656–704.

Durbin, J., and S. J. Koopman. 2012. *Time Series Analysis by State Space Methods*. Number 38. Oxford: Oxford University Press.

Fang, F., and C. W. Oosterlee. 2008. "A Novel Pricing Method for European Options Based on Fourier-Cosine Series Expansions." *SIAM Journal on Scientific Computing* 31 (2): 826–848. doi:10.1137/080718061.

Feng, Q., S. Jain, P. Karlsson, D. Kandhai, and C. W. Oosterlee. forthcoming. "Efficient Computation of Exposure Profiles on Real-World and Risk-Neutral Scenarios for Bermudan Swaptions." *Journal of Computational Finance*.

Gatheral, J. 2006. *The Volatility Surface: A Practitioner's Guide*. Vol. 357. Hoboken, NJ: John Wiley & Sons.

Gatheral, J., and A. Jacquier. 2014. "Arbitrage-Free SVI Volatility Surfaces." *Quantitative Finance* 14 (1): 59–71. doi:10.1080/14697688.2013.819986.

Glasserman, P. 2004. *Monte Carlo Methods in Financial Engineering*. Vol. 53. Berlin: Springer.

Grzelak, L. A., and C. W. Oosterlee. 2011. "On the Heston Model with Stochastic Interest Rates." *SIAM Journal on Financial Mathematics* 2 (1): 255–286. doi:10.1137/090756119.

Heston, S. L. 1993. "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options." *Review of Financial Studies* 6 (2): 327–343. doi:10.1093/rfs/6.2.327.

Homescu, C. 2011. "Implied Volatility Surface: Construction Methodologies and Characteristics." *arXiv Preprint arXiv:1107.1834*.

Hull, J. C. 2006. *Options, Futures, and Other Derivatives*. Harlow: Pearson Education.

Jain, S., and C. W. Oosterlee. 2015. "The Stochastic Grid Bundling Method: Efficient Pricing of Bermudan Options and Their Greeks." *Applied Mathematics and Computation* 269: 412–431. doi:10.1016/j.amc.2015.07.085.

Kamal, M., and J. Gatheral. 2010. "Implied Volatility Surface." In *Encyclopedia of Quantitative Finance*. Hoboken, NJ: John Wiley & Sons.

Kuhn, E., and M. Lavielle. 2004. "Coupling a Stochastic Approximation Version of EM with an MCMC Procedure." *ESAIM: Probability and Statistics* 8: 115–131. doi:10.1051/ps:2004007.

Lindsten, F. 2013. "An Efficient Stochastic Approximation EM Algorithm Using Conditional Particle Filters." Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference On, 6274–6278, IEEE, Vancouver, BC, May 26–31.

Lindsten, F., M. I. Jordan, and T. B. Schön. 2014. "Particle Gibbs with Ancestor Sampling." *Journal of Machine Learning Research* 15 (1): 2145–2184.

Murphy, K. P. 2002. "Dynamic Bayesian networks: representation, inference and learning." PhD thesis, Berkeley: University of California.

Schön, T. B., A. Wills, and B. Ninness. 2011. "System Identification of Nonlinear State-Space Models." *Automatica* 47 (1): 39–49. doi:10.1016/j.automatica.2010.10.013.

Simon, D. 2006. *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*. Hoboken, NJ: John Wiley & Sons.

Simon S. H., et al. 2001. *Kalman Filtering and Neural Networks*. Wiley Online Library. Hoboken, NJ: John Wiley & Sons.

Steehouwer, H. 2005. "Macroeconomic Scenarios and Reality: A Frequency Domain Approach for Analyzing Historical Time Series and Generating Scenarios for the Future." PhD thesis, VU University Amsterdam. http://hdl.handle.net/1871/9058.

Stein, H. J. 2013. "Joining Risks and Rewards." *SSRN Electronic Journal*. Available at SSRN 2368905. doi:10.2139/ssrn.2368905.

Stein, H. J. 2014. "Fixing Underexposed Snapshots–Proper Computation of Credit Exposures under the Real World and Risk Neutral Measures." *Technical Report*. December, 1–23. Bloomberg LP.

Zenios, S. A., and W. T. Ziemba. 2007. *Handbook of Asset and Liability Management: Applications and Case Studies*. Vol. 2. Amsterdam: Elsevier.