

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

Rüdiger Kiesel^{a,*}, Florentin Rahe^b

^a *University of Duisburg-Essen, 45117 Essen, Germany*
Centre of Mathematics for Applications, University of Oslo, Oslo, Norway

^b *Ulm University, 89081Ulm, Germany*
PhD Training Programme 1100

Abstract

We present a two-factor option-pricing model, which parsimoniously captures the difference in volatility persistences under the historical and risk-neutral probabilities. The model generates an S-shaped pricing kernel that exhibits time-varying risk aversion. We apply our model for two purposes. First, we analyze the risk preference implied by S&P500 index options during 2001–2009 and find that risk-aversion level strongly increases during stressed market conditions. Second, we apply our model for Value-at-Risk (VaR) forecasts during the subprime crisis period and find that it outperforms several leading VaR models.

Keywords: Pricing kernel, Option pricing, Implied risk premium, Value-at-Risk forecast

JEL: G12, G13, G14

[☆]The authors thank N.H. Bingham for fruitful discussions and comments. Furthermore, the authors are thankful to two anonymous referees for valuable suggestions.

*Corresponding author

Email addresses: ruediger.kiesel@uni-due.de (Florentin Rahe), f.rahe@gmx.de (Florentin Rahe)

1. Introduction

It is well-known that the Black and Scholes (1973) model exhibits several biases. Most prominent among these are its failure to reproduce the option-implied volatility smirk and the absence of a leverage effect. One-factor stochastic volatility (SV) models such as Heston (1993) or Hull and White (1987) address some of the biases, but fall short in capturing these effects simultaneously. In particular, Christoffersen et al. (2008, 2009) point out that one-factor volatility models are not able to simultaneously capture the level and the slope of option-implied volatility smirks demanded by the data for a given parametrization. To address this shortcoming, they introduce two-factor volatility models consisting of the short-run and long-run volatility components under the historical measure \mathbb{P} , and obtain the corresponding dynamic under the risk-neutral measure \mathbb{Q} using a standard measure change. In their studies, the volatility processes remain structurally the same under \mathbb{P} and \mathbb{Q} , albeit with certain parameter changes.

It is important to note that such a standard model, as implemented in Christoffersen et al. (2008), does not improve the (\mathbb{P} -) fit on return data compared to a GARCH(1,1). However, it performs substantially better when fitted using additional information on option prices (\mathbb{Q} -fit). This indicates that the persistent two-factor volatility dynamic is more relevant under the measure \mathbb{Q} .

We illustrated this effect in Figure 1. Using the Unscented Kalman filter (UKF) we jointly estimate the statistical (\mathbb{P}) and risk-neutral (\mathbb{Q}) asset price dynamics from the S&P500 index and corresponding call options for the period 2006 to 2009 (see §3.1 and §3.2 for details). In the left panel we show the changes in the S&P500 index together with the movement of the long-term variance process of the Heston model (for the formal definition see equation (4.1.3)). While the \mathbb{P} long-term variance level ϕ is relatively stable, the \mathbb{Q} long-term variance level ϕ^* increases rapidly at the outset of the financial crisis.

The same effect occurs for a typical 2-factor SV model (see §4.1.5 for a specification). As illustrated in the right-hand panel of Figure 1 the long-term variance level also remains fairly stable under \mathbb{P} (illustrated in terms of the parameter ϕ) and increases significantly under the

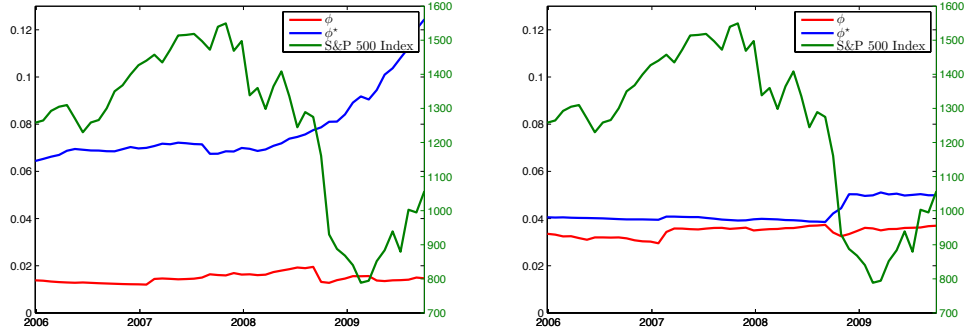


Figure 1: Comparison of the \mathbb{P} and \mathbb{Q} long-term variance dynamic for the Heston model (left panel) and for the alternative 2-factor SV (right panel) during the period 2006 to 2009. The red line exhibits the \mathbb{P} -dynamic (ϕ) and the blue line the \mathbb{Q} -dynamic (ϕ^*). The estimations are based on a time series of the S&P500 index and option prices and are done using the Unscented Kalman filter, see §3.1 and §3.2 for details.

pricing measure \mathbb{Q} (see parameter ϕ^*).

Motivated by these facts we propose a model that extends the stochastic volatility model of Heston (1993) by an additional “long-run target variance process”, which allows to specify the difference of the historical \mathbb{P} and the pricing \mathbb{Q} dynamics of the underlying price process explicitly. As this process significantly determines the sign and magnitude of the model’s variance risk premium, which is closely related to the risk preferences of market participants (see Bakshi and Kapadia (2003), Bollerslev et al. (2011) and Heston (1993)), we are able to capture changes of risk aversion. These changes are then a source for price variations without any changes in the underlying historical probability distribution.

In the spirit of Christoffersen et al. (2013), we introduce independent adjustments for different risk sources. In particular, we not only separate equity and variance risk, but split the variance risk into a short-run and an additional long-run target variance process. However, the innovative aspect of our model is that the additional long-run target variance process only acts under \mathbb{Q} and is not observable under \mathbb{P} . This is in contrast to the typical 2-factor volatility models which specify short- and long-run volatility components under \mathbb{P} and \mathbb{Q} . Our approach also allows one to use a more parsimonious set of parameters.

Our new option-pricing model also incorporates the stylized facts of pricing kernels. Empirical studies suggest that pricing kernels are typically S-shaped, or rather exhibit a monotonically increasing area (see Aït-Sahalia and Lo (2000), Jackwerth (2000) and Rosenberg and Engle (2002)). This phenomenon is also called the “pricing kernel puzzle” (see Brown and Jackwerth (2012) and Hens and Reichlin (2013)), since neither standard option-pricing models nor consumption-based models are typically able to capture this property. In addition, the shape of pricing kernels is usually not constant, but changes over time, which indicates that risk preferences are time-varying (see Rosenberg and Engle (2002) and Grith et al. (2013)). As we obtain a variance-dependent pricing kernel, which, dependent on the level of the long-run target variance process, exhibits the familiar S- or U-shapes, we are able to capture both of these properties.

We apply our framework for two purposes. First, our model and its variance risk premium will be used in order to analyze the risk preferences of market participants through time. As pointed out above the long-run target variance process is a suitable measure for this. The volatility implied by option prices is often used as an alternative measure for analyzing market sentiment.¹ The advantage of the long-run target variance process as an indicator for market sentiment is its lower variability compared to the implied volatility. This makes it easier to identify changing market conditions.

Second, we use our model to extract the forward-looking information content of option prices. As Andersen and Bondarenko (2007) put it this information “(...) include[s] but vastly exceed[s] the information contained in historical returns”. Recently, this option-implied information was used in the context of density forecasting (see Christoffersen et al. (2012) for an overview). For this, pricing kernels are used to “translate” risk-neutral densities into real-world densities. The crucial point of this approach is that the supposed pricing kernel must adequately describe the current risk preferences. Otherwise, the resulting real-world

¹The most prominent example in this context is the VIX index of the Chicago Board Options Exchange (CBOE), which quantifies the volatility of S&P500 index options and is also called the “investor fear gauge” (see Whaley (2000, 2009)).

density is a biased estimator of future asset prices. However, our model is well-suited to circumvent this issue.

We use S&P500 index options with an underlying data period 2001 to 2009 for our applications, which includes several market shocks like the burst of the dot-com bubble, the terror attacks 9/11 and the subprime mortgage crisis. As in the previous literature (Chernov and Ghysels (2000), Pan (2002) and Jones (2003)), we find a negative variance risk premium. However, during the calm market period from mid-2004 to mid-2007 the variance risk premium of our model is slightly positive, which indicates risk-seeking market behaviour during that time. From mid-2007 the variance risk premium turns negative and the long-run target variance process continuously increases, which reflects the growing fear of market participants during the subprime mortgage crisis. Also the evolution of the pricing kernel reflects the different market conditions of the period. The typical S-shaped pricing kernel before the crisis becomes more and more pronounced until the bankruptcy of Lehman Brothers and then turns into a U-shape. Again, this evolution first indicates increasing willingness to carry risk, which turns into risk aversion after the outbreak of the financial crisis.

In the second part of our empirical study, we analyze the quality of the forward-looking information embedded in option prices extracted by our model. We perform out-of-sample Value-at-Risk (VaR) forecasts for the S&P500 index during the period of the subprime mortgage crisis. In order to better classify the corresponding forecasting results, we also perform VaR forecasts based on five alternative VaR models, namely an alternative 2-factor SV^2 , the Heston, the Heston-Nandi, and the Black-Scholes model, which also rely on option-implied information, and the GARCH model, which relies on historical information only. We find that the 2-factor SV models have the best forecasting performance, followed by the Black-Scholes, the Heston, the Heston-Nandi and the GARCH model. In particular, the two 2-factor SV models are the only ones able to perform highly accurate VaR forecasts for all confidence levels (95%, 99% and 99.9%) and forecasting horizons (1, 2, 3 and 4 weeks), despite the

²We are thankful to a referee who suggested this additional comparison.

challenging forecasting period.

The remainder of this paper is structured as follows. The next section formalizes our model specification. Then Section 3 outlines the applied estimation method, the data and the estimation results. Section 4 discusses the forecasting performance of the analyzed VaR models, and is followed by Section 5, which concludes the paper.

2. The 2-factor stochastic volatility model

This section introduces the 2-factor stochastic volatility (SV) model. We extend the Heston (1993) model by a long-term variance target process, which is unrelated to the statistical asset-price dynamics but part of the risk-neutral asset-price dynamics.

2.1. Statistical dynamics

We fix a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ satisfying the usual conditions and let S_t denote the asset price at time t , which could represent a single stock or an index. The asset-price dynamics under the statistical measure³ \mathbb{P} are specified as follows:

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

where $W_i = (W_{i,t})_{t \geq 0}$, $i = 1, 2, 3$ denote $(\mathcal{F}_t)_{t \geq 0}$ -adapted Brownian motions. The dynamics of the assets' variance $(\sigma_t)_{t \geq 0}$ is modeled by a Cox-Ingersoll-Ross (CIR) process⁴, where κ_1 controls for how fast the current variance returns to its long-term average ϕ_1 . Furthermore, the volatility of σ_t (the so-called vol of vol) is specified by the parameter λ_1 . In order to accommodate the leverage effect, the Brownian motions W_1 and W_2 are correlated via $dW_{1,t} dW_{2,t} = \rho dt$. The third process $(\theta_t)_{t \geq 0}$, the long-run target variance process, will capture changes in asset prices that are only due to changes in risk preferences and not

³The statistical measure is sometimes called “objective” or “real-world” measure.

⁴We assume that the standard restrictions on the coefficients of the process are satisfied.

to changes in the statistical asset-price dynamics. Therefore, the long-run target variance process is uncorrelated with the asset price S_t and its variance σ_t . The latter can simply be achieved by supposing that the Brownian motion W_3 is uncorrelated with W_1 and W_2 and that the defining stochastic differential equations are not coupled. Just like the variance process, the long-run target variance process also is a mean-reverting CIR process⁴, with long-term average ϕ_2 , mean-speed of reversion κ_2 and volatility λ_2 . Since the long-run target variance θ_t is independent of S_t and σ_t , the 2-factor SV model has the same statistical asset-price properties as the Heston model.

2.2. Risk-neutral dynamics

We use independent Brownian motions $\overline{W}_i = (\overline{W}_{i,t})_{t \geq 0}$, $i = 1, 2, 3$ to construct the correlated Brownian motions W_1 and W_2 via Cholesky's transform from \overline{W}_1 and \overline{W}_2 and choose $\overline{W}_3 = W_3$. We then derive the risk-neutral asset-price dynamics of the 2-factor SV model by virtue of Girsanov's theorem. For this we define the following Brownian motions with drift:

$$d\widehat{W}_{1,t} = d\overline{W}_{1,t} + \gamma_{1,t}dt ,$$

$$d\widehat{W}_{2,t} = d\overline{W}_{2,t} + \gamma_{2,t}dt ,$$

$$d\widehat{W}_{3,t} = d\overline{W}_{3,t} + \gamma_{3,t}dt .$$

The processes $\gamma_i = (\gamma_{i,t})_{t \geq 0}$, $i = 1, 2, 3$ are assumed to be adapted to the filtration $(\mathcal{F}_t)_{t \geq 0}$ and specify the market price of risk of S_t , σ_t and θ_t , respectively. In the following, we will assume that the long-run target process θ_t carries no market price of risk, so $\gamma_3 = 0$. Hence, the change from the statistical measure \mathbb{P} to an equivalent martingale measure \mathbb{Q} can be performed via the Radon-Nikodým derivative

$$Z_t = \exp \left\{ - \int_0^t \sum_{i=1}^2 \gamma_{i,u} d\overline{W}_{i,u} - \frac{1}{2} \int_0^t \sum_{i=1}^2 \gamma_{i,u}^2 du \right\} .$$

By further imposing the condition

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

the risk-neutral asset-price dynamics are given by

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

where $\widetilde{W}_{1,t} = \sqrt{1 - \rho^2} \widehat{W}_{1,t} + \rho \widehat{W}_{2,t}$, $\widetilde{W}_{2,t} = \widehat{W}_{2,t}$ and $\widetilde{W}_{3,t} = \widehat{W}_{3,t}$. From (1) we see that the equity risk premium is, as in the Heston model, driven by the market price of risk process $\frac{\mu - r}{\sqrt{\sigma_t}}$ (i.e. risk aversion to market risk) and a variance risk premium (risk aversion to variance). We define the market price of variance risk as

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

which includes the long-run target variance process. This makes the effects of the long-run target variance process on the variance risk premium explicit.

The risk-neutral variance dynamics are given by

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

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

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

The long-run target process θ_t is now part of the variance process, and replaces the long-term average ϕ_1 by the central tendency θ_t . Duffie et al. (2000) propose a similar specification to model asset prices. Furthermore, Bates (2012) supposes a 2-factor SV model with jumps, but assumes that θ_t stays constant over intra-daily intervals in order to keep the model computationally tractable. However, none of these authors uses the second volatility factor θ_t only under the risk-neutral measure. Further note that 2-factor models of this form are

also applied in the interest-rate literature, where one factor reflects the central tendency of the short rate (see e.g. Balduzzi et al. (1998), Dai and Singleton (2000), and for a textbook account Brigo and Mercurio (2006)).

2.3. Characteristic function

In the context of option pricing, we are typically interested in the characteristic function of the log-asset price, as we are able to compute option prices by Fourier inversion methods (see Schmelzle (2010) for an overview). The log-asset price of the 2-factor model follows an affine process, so the corresponding characteristic function can be stated up to the solution of a Riccati equation, which is summarized in the next proposition (see Appendix A for a proof).

Proposition 2.1. *Let S_t , σ_t and θ_t be the risk-neutral log-asset-price dynamics of the 2-factor SV model specified by (3). Then the characteristic function of $\ln(S_T)$ is given by*

$$\mathbb{E}[e^{iu \ln(S_T)} \mid \mathcal{F}_t] = e^{\phi(T-t,u) + \psi_1(T-t,u) \ln(S_t) + \psi_2(T-t,u) \sigma_t + \psi_3(T-t,u) \theta_t} , \quad (4)$$

where

$$\phi(t, u) = iurt + \kappa_2 \phi_2 \int_0^t \psi_3(s, u) ds ,$$

$$\psi_1(t, u) = iu ,$$

$$\psi_2(t, u) = \frac{1}{\lambda^2} (c - \sqrt{d}) \frac{1 - e^{-\sqrt{d}t}}{1 - ge^{-\sqrt{d}t}} ,$$

and $\psi_3(t, u)$ is the solution of

$$\partial_t \psi_3(t, u) = \kappa_1^* \psi_2(t, u) - \kappa_2 \psi_3(t, u) + 1/2 \lambda_2^2 \psi_3(t, u)^2 , \quad \psi_3(0, u) = 0, \quad (5)$$

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

The characteristic exponents $\psi_1(t, u)$ and $\psi_2(t, u)$ are the same as in the Heston model. For applications, the ordinary differential equation (ODE) (5) has to be solved numerically, commonly done by the Runge-Kutta method (see Duffie and Kan (1996) and Piazzesi (2009)), which we also apply.

2.4. Pricing kernel

The term “pricing kernel” is originally used in the context of consumption-based models, which offer a very general framework to price any kind of asset.⁵ Within this approach the price of an asset corresponds to the expectation of the “preference-weighted” payoff profile. More precisely, the payoff profile of an asset is weighted by a pricing kernel based on the utility function of the representative investor, and thus reflects his risk preferences over different states of the world. Common utility functions imply pricing kernels which are monotonically decreasing with the consumption level. Accordingly, a representative investor is willing to pay more for payoffs in bad states of the world (low consumption level) than in good states of the world (high consumption level). Approaches to estimate the pricing kernel directly are discussed in Jackwerth (2000), Grith et al. (2013), and Grith et al. (2016). Hens and Reichlin (2013) present a theoretical explanation for the fact that pricing kernels may have increasing parts.

In the context of option-pricing models the pricing kernel is defined as the ratio of risk-neutral to statistical asset-price dynamics. Following Grith et al. (2013), p. 371, we formally define the pricing kernel as

$$PK(x) = \frac{f_{ret}^{\mathbb{Q}}(x)}{f_{ret}^{\mathbb{P}}(x)}, \quad (6)$$

where $f_{ret}^{\mathbb{Q}}$ denotes the risk-neutral and $f_{ret}^{\mathbb{P}}$ the statistical return density. For the 2-factor SV model these densities can be computed by means of Fourier inversion of the characteristic function of the log-asset price (see Appendix D for more details).

The pricing kernel of the 2-factor SV model is illustrated in Figure 2 for different levels of the long-run target variance process θ_t while keeping the remaining parameters fixed. The parameter constellations used are as follows:

- Case 1: $\theta_t < \phi_1 < \phi_2$ (low risk aversion)
- Case 2: $\phi_1 < \theta_t < \phi_2$ (medium risk aversion)

⁵See Cochrane (2005) for an introduction to consumption-based models.

- Case 3: $\phi_1 < \phi_2 < \theta_t$ (high risk aversion)
- Case 4: $\phi_1 < \phi_2 \ll \theta_t$ (extremely high risk aversion)

The parameterization of the 2-factor SV model reflects a calm market period, where the current level of the variance σ_t is lower than its long-term average ϕ_1 . Moreover, ϕ_1 is supposed to be lower than the long-term average ϕ_2 of the long-run target variance process, in line with the empirical observation that risk-neutral volatilities on average tend to be larger than realized volatilities (see Jackwerth and Rubinstein (1996) and Bollerslev et al. (2011)). Except for the last case, where the level of risk aversion is extremely high, the shape of the pricing kernel has the typical S-form found in the literature (see Jackwerth (2000) and Rosenberg and Engle (2002)). In particular, the lower the current level of risk aversion, the more pronounced the S-shape of the pricing kernel. By contrast, for high levels of current risk aversion, the pricing kernel becomes U-shaped. Further note that the pricing kernel with $\theta_t = 0.20^2$ (Case 1) looks similar to pricing kernel estimates given previously by Jackwerth (2004), p. 57, Fig. 11; Ait-Sahalia and Lo (2000), p. 36, Fig. 3; and Rosenberg and Engle (2002), p. 361, Fig. 5.

2.5. Variance risk premium and relative risk aversion

The variance risk premium is defined as the difference between the statistical and risk-neutral variance drift term⁶, which is

$$\begin{aligned} VRP(\sigma_t, \theta_t) &= [\kappa_1 (\phi_1 - \sigma_t) - \kappa_1^* (\theta_t - \sigma_t)], \\ &= [\kappa_1 (\phi_1 - \theta_t) + \Delta\kappa (\sigma_t - \theta_t)], \end{aligned} \tag{7}$$

where $\Delta\kappa = \kappa_1^* - \kappa_1$. Hence, the variance risk premium depends on the parameters κ_1 and $\Delta\kappa$ as well as on the current level of the long-run target variance process θ_t compared to the

⁶Carr and Wu (2009) define the (return) variance risk premium as the difference between implied and realised variance. In their context it relates to the correlation of realised quadratic variation and the pricing kernel.

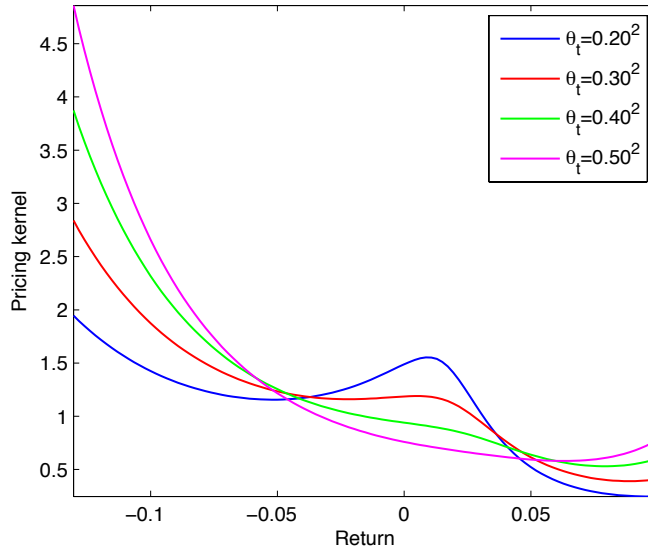


Figure 2: Pricing kernel of the 2-factor SV model. This figure illustrates the pricing kernel of the 2-factor SV model for different values of the long-term target variance θ_t . The pricing kernels are computed as the ratio of risk-neutral to statistical return density based on the following parameterisation: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

long-term average of the variance ϕ_1 and the current level of the variance σ_t , respectively. The sign of the variance risk premium can be determined via the boundary

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

obtained by setting (7) equal to zero and then solving for θ_t . Since (7) is decreasing in θ_t , the variance risk premium is positive if $\theta_t < b(\sigma_t)$ and negative if $\theta_t > b(\sigma_t)$. Furthermore, the boundary $b(\sigma_t)$ is decreasing in σ_t if $\Delta\kappa < 0$. Empirical studies suggest that the mean speed of reversion of the variance is usually smaller under the risk-neutral measure than under the statistical measure (see e.g. Bakshi and Wu (2010), Bates (2012) and Pan (2002)), so that it is reasonable to assume that $\Delta\kappa < 0$. Consequently, during normal market conditions (low levels of σ_t) a negative variance risk premium is less likely since the boundary for θ_t to cross is high. Conversely, during turbulent market conditions (high levels of σ_t) a negative variance risk premium is more likely since the boundary for θ_t to cross is low.

Following Bollerslev et al. (2011), we can link the variance risk premium and the coefficient of risk aversion for the representative investor within the standard intertemporal asset pricing framework. From the specification of the market price of variance risk, we find⁷

$$\left\langle \frac{dm_t}{m_t}, d\sigma_t \right\rangle = -\lambda_1 \sigma_t \gamma_{2,t} dt, \quad (8)$$

where m_t denotes the pricing kernel, or marginal utility of wealth for the representative investor. Assuming the representative investor has a power utility function with risk-aversion parameter γ ,

$$U_t = e^{-\delta t} \frac{W_t^{1-\gamma}}{1-\gamma},$$

where δ denotes a constant subjective time discount rate, and in equilibrium the investor holds the market portfolio, then marginal utility equals

$$m_t = e^{-\delta t} W_t^\gamma.$$

Using Itô's lemma implies

$$\left\langle \frac{dm_t}{m_t}, d\sigma_t \right\rangle = -\gamma \lambda_1 \sigma_t \rho dt. \quad (9)$$

⁷ $\langle \cdot, \cdot \rangle$ denotes the quadratic covariation, see Bingham and Kiesel (2004), §5.3.2.

Putting equations (8) and (9) together, yields

$$\gamma\rho = \gamma_{2,t} = \frac{\kappa_1(\phi_1 - \sigma_t) - \kappa_1^*(\theta_t - \sigma_t)}{\lambda_1\sqrt{\sigma_t}}. \quad (10)$$

So we see

$$\gamma = \frac{VRP(\sigma_t, \theta_t)}{\rho\lambda_1\sqrt{\sigma_t}}. \quad (11)$$

As ρ is negative, a positive/negative variance risk premium indicates risk-seeking/risk-averse market participants. Hence, the 2-factor SV model implies that market sentiment is rather relaxed during normal market conditions and rather stressed during turbulent market conditions. Bollerslev et al. (2011) and Carr and Wu (2009) provide evidence that the variance risk premium is time-varying.

3. Estimation

In the following section, we jointly estimate the statistical and risk-neutral asset-price dynamics of the 2-factor SV model based on a time series of option prices. From the corresponding estimation results we derive conclusions about the risk aversion inherent in the market, during both normal and turbulent market conditions. Moreover, we analyze the parameter stability of the 2-factor SV model.

3.1. Estimation method

To jointly estimate the risk-neutral and statistical asset-price dynamics of an option-pricing model we use the Unscented Kalman filter (UKF). The UKF is an algorithm for estimating the state of a time-varying system indirectly observed through noisy measurements (Julier and Uhlmann (1997), Särkkä (2011), and Wan and van der Merwe (2001)). The heart of the UKF is the unscented transformation, an approximation method for determining the mean and variance of a transformed random variable. In contrast to the classical Kalman filter, the UKF allows for nonlinearities in the filtering equations, which makes it very suitable for financial applications (see Bakshi and Wu (2010), Bakshi et al. (2008), and Christoffersen et al. (2014)).

The basis for each filtering technique is a state-space representation describing the dynamics of the underlying system. In the context of option pricing, the state-space representation describes the statistical dynamics of the underlying asset-price process. This asset-price process is linked to observed option prices by a theoretical option-pricing model. In the following, we assume that the asset-price process is more generally given by a m -dimensional stochastic process $X_t = [X_{1,t} \dots X_{m,t}]^\top$ with statistical dynamics

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t, \quad (12)$$

where $\mu : \mathbb{R}^m \rightarrow \mathbb{R}^m$ specifies the drift and $\sigma : \mathbb{R}^m \rightarrow \mathbb{R}^{m \times m}$ the volatility of the process. We summarize the parameters of $\mu(x)$ and $\sigma(x)$ in the parameter vector θ_1 .

For the purpose of estimation, we use an equidistant time grid $t_1 < \dots < t_N$ with $t_i - t_{i-1} = \Delta t$. Then the state equation can be obtained by an Euler discretization of the statistical asset-price dynamics (12), which is

$$X_i = X_{i-1} + \mu(X_{i-1})\Delta t + \sigma(X_{i-1})\sqrt{\Delta t}\eta_i. \quad (13)$$

Here, η_i denotes a normally distributed random variable with zero mean and variance specified by a m -dimensional identity matrix I_m , i.e. $\eta_i \sim \mathcal{N}(0, I_m)$, and the time index t_i corresponds to i .

Next, the measurement equation relates the state vector X_i to observable variables. As measurements we use a time series of observed option prices $\bar{\mathcal{O}}_1, \dots, \bar{\mathcal{O}}_N$, where $\bar{\mathcal{O}}_i = [\bar{\mathcal{O}}_{1,i} \dots \bar{\mathcal{O}}_{n,i}]^\top$ is a $n \times 1$ vector of option prices with strike prices $K_i = [K_{1,i} \dots K_{n,i}]^\top$ and maturities $\tau_i = [\tau_{1,i} \dots \tau_{n,i}]^\top$. However, for estimation purposes it is better to first divide the option prices $\bar{\mathcal{O}}_i$ by their corresponding Black-Scholes vegas $v_{BS}(\bar{\mathcal{O}}_i)$, which is

$$y_i = v_{BS}^{-1}(\bar{\mathcal{O}}_i) \odot \bar{\mathcal{O}}_i. \quad (14)$$

Here, \odot denotes the Hadamard product, defined as the entrywise product of two matrices, i.e. $(A \odot B)_{i,j} = (A)_{i,j} \cdot (B)_{i,j}$. The background of this procedure is that the variance of time series of vega-scaled option prices is better comparable across different strike prices and maturities, allowing for a simpler specification of the measurement error (see Bakshi and

Wu (2010) and Bakshi et al. (2008)). Based on this data transformation, the measurement equation now becomes

$$y_i = v_{BS}^{-1}(\bar{\mathcal{O}}_i) \odot \mathcal{O}_{Model}(X_i, K_i, \tau_i | \theta_2) + \epsilon_i . \quad (15)$$

Here, $\mathcal{O}_{Model}(X_i, K_i, \tau_i | \theta_2)$ denotes an option-pricing model with parameters θ_2 , whose risk-neutral dynamics are derived from the statistical dynamics (12). In order to account for heteroscedasticity, we model the conditional variance of the measurement error $\epsilon_i = [\epsilon_{1,i}, \dots, \epsilon_{n,i}]^\top$ by the following ARCH-type model:

$$\epsilon_i = \tilde{\sigma}_i z_i , \quad \tilde{\sigma}_i^2 = \omega + \alpha \frac{1}{n} \sum_{j=1}^n \epsilon_{j,i-1}^2 . \quad (16)$$

Here, $\omega > 0$, $\alpha \geq 0$ and $z_i \sim \mathcal{N}(0, I_n)$. So, the state-space representation can fully be parameterized by the parameter vector $\theta = [\theta_1 \ \theta_2 \ \omega \ \alpha]$, which can be estimated by the Quasi-Maximum Likelihood method described in Appendix E.

3.2. Data

Our empirical study is based on Wednesday's weekly closing prices of the S&P500 index and corresponding call options.⁸ We analyze the time period from November 3, 1999 to October 28, 2009, which includes several market shocks such as the bursting of the dot-com bubble in 2000, the terror attacks 9/11 in 2001 and the subprime mortgage crisis starting in 2007.

First, Figure 3(a) gives an overview of the evolution of the S&P500 index for the underlying time period, where the several market shocks are reflected by sudden falls in the index prices. Figure 3(b) illustrates the corresponding index returns, which exhibit low levels of volatility during the calm period 2004 – 2007 and high levels of volatility before and after this calm period. Moreover, Figures 3(c) and 3(d) show a kernel density estimation and a Quantile-Quantile plot of the index returns, which both indicate that the distribution of the

⁸The data are drawn from OptionMetrics Ivy DB US and were provided by the Collaborative Research Center 649 – Economic Risk at the Humboldt-Universität zu Berlin.

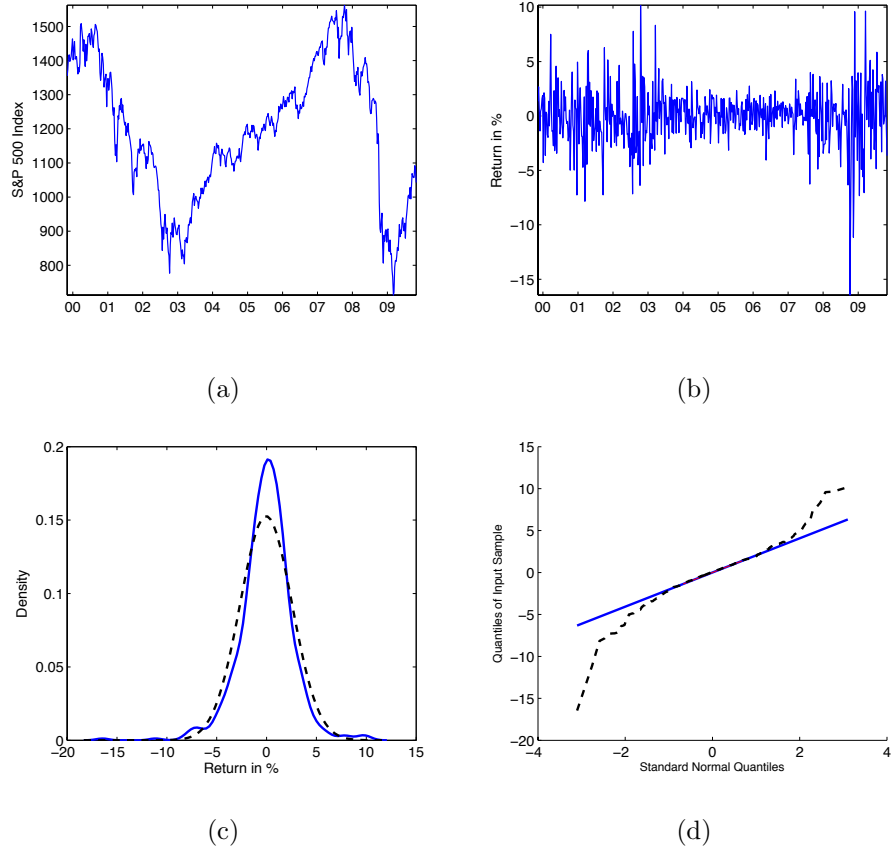


Figure 3: Statistical properties of the S&P500 index. These figures illustrate the evolution of the S&P500 index (a) and the corresponding index returns (b). Moreover, figure (c) shows a kernel density estimation of the index returns (solid line), which is contrasted to the normal distribution (dotted line). The latter is specified by the sample mean and variance of the index returns. In addition, figure (d) shows the Quantile-Quantile plot of the index returns.

index returns is fat-tailed and slightly negatively skewed. These observations are in line with the sample skewness and kurtosis of the S&P500 index returns, summarized in Table 1.

Table 1: Descriptive statistics for the S&P500 index returns.

N	min	max	mean	SD	skewness	kurtosis
521	-16.4508	10.1824	-0.0503	2.6125	-0.4910	7.6359

Notes: This table reports the summary statistics of the weekly S&P500 index returns. All figures are calculated based on percentage returns.

From the wide spectrum of available S&P500 index options, we choose call options according to the following strike choice function:

$$K_{j,i} = \lfloor S_i/25 \rfloor \cdot 25 + j \cdot 25 \quad \text{for } j = 1, 2 ,$$

where $\lfloor \cdot \rfloor$ denotes the floor function and S_i the S&P500 index at time t_i . Hence, the resulting strike prices $K_{1,i}$ and $K_{2,i}$ correspond to (near) at-the-money ($j = 1$) and out-of-the-money ($j = 2$) call options. In addition, for each strike level we choose four call options with average maturities of 22, 45, 89 and 148 days. Altogether, this particular choice results in a call option price time series $\bar{C}_i = [\bar{C}_{1,i} \dots \bar{C}_{8,i}]^\top$ with maturities $\tau_i = [\tau_{1,i} \dots \tau_{4,i} \tau_{1,i} \dots \tau_{4,i}]^\top$ and strike prices $K_i = [K_{1,i} \dots K_{1,i} K_{2,i} \dots K_{2,i}]^\top$.⁹

As we have a small number of outliers in the option data we apply the standard win-sorization technique to adjust for outliers.¹⁰

3.3. State-space representation of the 2-factor SV model

The state-space representation of the 2-factor SV model is derived as in Section 3.1, with two exceptions. First, we do not include the dynamics of the underlying S_t in the state equation. This facilitates the estimation, since the drift term μ is typically difficult

⁹In some special cases option prices with the required strike price are not traded, so that the corresponding closest available strike price is chosen instead.

¹⁰For more details on this technique see e.g. Barnett and Lewis (1994).

to estimate and, in addition, not relevant for our purposes. Second, we use a semi-analytic solution for the dynamics of the variance and the long-run target variance, which both follow CIR processes (see Appendix F for a derivation). Taking these two adjustments into account, the state-space representation of the 2-factor SV model can be specified via

$$\begin{bmatrix} \sigma_i \\ \theta_i \end{bmatrix} = \begin{bmatrix} e^{-\kappa_1 \Delta t} \\ e^{-\kappa_2 \Delta t} \end{bmatrix} \odot \begin{bmatrix} \sigma_{i-1} \\ \theta_{i-1} \end{bmatrix} + \begin{bmatrix} \phi_1 (1 - e^{-\kappa_1 \Delta t}) \\ \phi_2 (1 - e^{-\kappa_2 \Delta t}) \end{bmatrix} + \begin{bmatrix} e^{-\kappa_1 \Delta t} \lambda_1 \\ e^{-\kappa_2 \Delta t} \lambda_2 \end{bmatrix} \odot \sqrt{\begin{bmatrix} \sigma_{i-1} \\ \theta_{i-1} \end{bmatrix}} \odot \eta_i, \quad (17)$$

$$y_i = v_{BS}^{-1}(\bar{C}_i) \odot \mathcal{C}_{2\text{FSV}}(X_i, K_i, \tau_i | \theta_2) + \epsilon_i,$$

where $\eta_i \sim \mathcal{N}(0, \Delta t I_2)$ and $X_i = [S_i \ \sigma_i \ \theta_i]^\top$. Furthermore, $\mathcal{C}_{2\text{FSV}}$ denotes the call-option price formula for the 2-factor SV model, which is parameterized by $\theta_2 = [\kappa_1^* \ \kappa_2 \ \phi_2 \ \lambda_1 \ \lambda_2 \ \rho]$. The corresponding option prices are calculated based on the characteristic function (4) and the Fourier inversion approach of Carr and Madan (1999). To account for heteroscedasticity in the measurement error, we model ϵ_i by the univariate ARCH model (16). Hence, the state-space representation of the 2-factor SV model can be fully parameterized by $\theta = [\kappa_1 \ \kappa_1^* \ \kappa_2 \ \phi_1 \ \phi_2 \ \lambda_1 \ \lambda_2 \ \rho \ \omega \ \alpha]$.

3.4. Estimation results

We compare the estimation results of the 2-factor SV (2FSV) model with an alternative 2-factor SV (A2FSV) model (specified below in §4.1.5) and the Heston model (specified below in §4.1.3). The parameters of all models are estimated as described in §3.1. Figure 4 shows that the 2-factor SV model dominates the alternative 2-factor SV model and the Heston model in terms of log-likelihood values \mathcal{L} . The corresponding parameter estimation results are summarized in Table 2.

First note that all parameters are significantly different from zero at a significance level of 0.1%. While both two-factor models have a low error-term variability (captured in the parameter α), the error-term variability is substantially higher for the Heston model. This fact again emphasises the need for a second volatility component.

All three models show negative values for the correlation parameter ρ which is in line with the negatively skewed return distribution observed in the data.

Comparing the 2-factor models we see that under the historical measure the 2-factor SV has relatively high estimates for the mean speed of reversion as well as for the volatility of the variance process σ_t , with values of $\kappa_1 = 22.38$ and $\lambda_1 = 0.945$. However, while high values of κ_1 reduce the variability of the variance process, high values of λ_1 increase it. So, these estimation results have contrary effects. For the alternative 2-factor SV these two parameters are lower, but as stated above the path behaviour of the first volatility component will be similar due to the contrary effects of the parameters. Strikingly, in the risk-neutral setting the speed of mean reversion κ_1^* of the alternative 2-factor SV increases to a level higher than the corresponding parameter in the 2-factor SV model. As the volatility of the volatility process λ_1 remains unchanged this implies less variability in the volatility process for the alternative 2-factor SV model.

The estimation for the parameters of the long-term volatility components show that the volatility of the long-term variance λ_2 is almost the same for both models. However, the speed of mean-reversion of this component changes for the alternative 2-factor SV model between the historical and the risk-neutral probability measures from a higher level to a level comparable to the 2-factor SV model. As this parameter is only relevant for the 2-factor SV model under the risk-neutral measure this supports our modelling approach to only consider the long-term component under the risk-neutral measure.

Focusing on the 2-factor SV model we see that the estimation results for the mean speed of reversion as well as for the volatility of the long-run target process θ_t are relatively low, with values of $\kappa_2 = 1.33$ and $\lambda_2 = 0.234$, respectively. Hence, in comparison to the variance process, the long-run target process is less volatile and more persistent. The long-term average of the long-run target process specified by $\phi_2 = 0.036$ is higher than the long-term average of the variance process specified by $\phi_1 = 0.025$, both under the statistical measure \mathbb{P} . In addition, the parameter ϕ_2 also determines the long-term average of the variance under the risk-neutral measure \mathbb{Q} . So, the long-term average of the variance under \mathbb{Q} is higher than under \mathbb{P} , giving a negative average variance risk premium in the amount of -1.46%.¹¹

¹¹Recall that the variance risk premium of the 2-factor SV model is defined as $\kappa_1(\phi_1 - \sigma_t) - \kappa_1^*(\theta_t - \sigma_t)$. Hence, the average variance risk premium can be calculated if we replace σ_t and θ_t by their long-term averages

Table 2: Estimation results for the parameters of the SV models

Heston	κ	κ^*	-	-	ϕ	ϕ^*	λ	-	ρ	ω	α	\mathcal{L}
Coef.	3.11	0.37	-	-	0.015	0.124	0.272	-	-0.805	$3.276 \cdot 10^{-5}$	0.215	$1.6352 \cdot 10^4$
<i>t</i> -stat.	n.a.	(8.65)	-	-	n.a.	(10.28)	(33.39)	-	(-28.85)	(22.96)	(20.61)	-
2FSV	κ_1	κ_1^*	κ_2	-	ϕ_1	ϕ_2	λ_1	λ_2	ρ	ω	α	\mathcal{L}
Coef.	22.38	7.98	1.33	-	0.025	0.036	0.945	0.234	-0.664	$0.602 \cdot 10^{-5}$	0.017	$1.8981 \cdot 10^4$
<i>t</i> -stat.	(95.89)	(122.02)	(33.37)	-	(28.75)	(43.62)	(67.21)	(29.82)	(-114.59)	(41.99)	(15.73)	-
A2FSV	κ_1	κ_1^*	κ_2	κ_2^*	ϕ	ϕ^*	λ_1	λ_2	ρ	ω	α	\mathcal{L}
Coef.	7.38	9.61	4.94	0.79	0.037	0.050	0.666	0.237	-0.789	$0.7030 \cdot 10^{-5}$	0.0240	$1.8765 \cdot 10^4$
<i>t</i> -stat.	(4.11)	(12.65)	(5.80)	(6.65)	(13.87)	(11.33)	(30.50)	(18.55)	(-33.48)	(7.63)	(12.89)	-

Notes: The reported *t*-statistics are calculated as $t = (\hat{\theta} - \theta_0) / SE(\hat{\theta})$, using the null hypothesis $\theta_0 = 0$. So, large *t*-statistics indicate that the corresponding parameter is significantly different from zero. In particular, $|t| > 1.960$ corresponds to a 5% significance level, $|t| > 2.576$ to a 1% level and $|t| > 3.291$ to a 0.1% level.

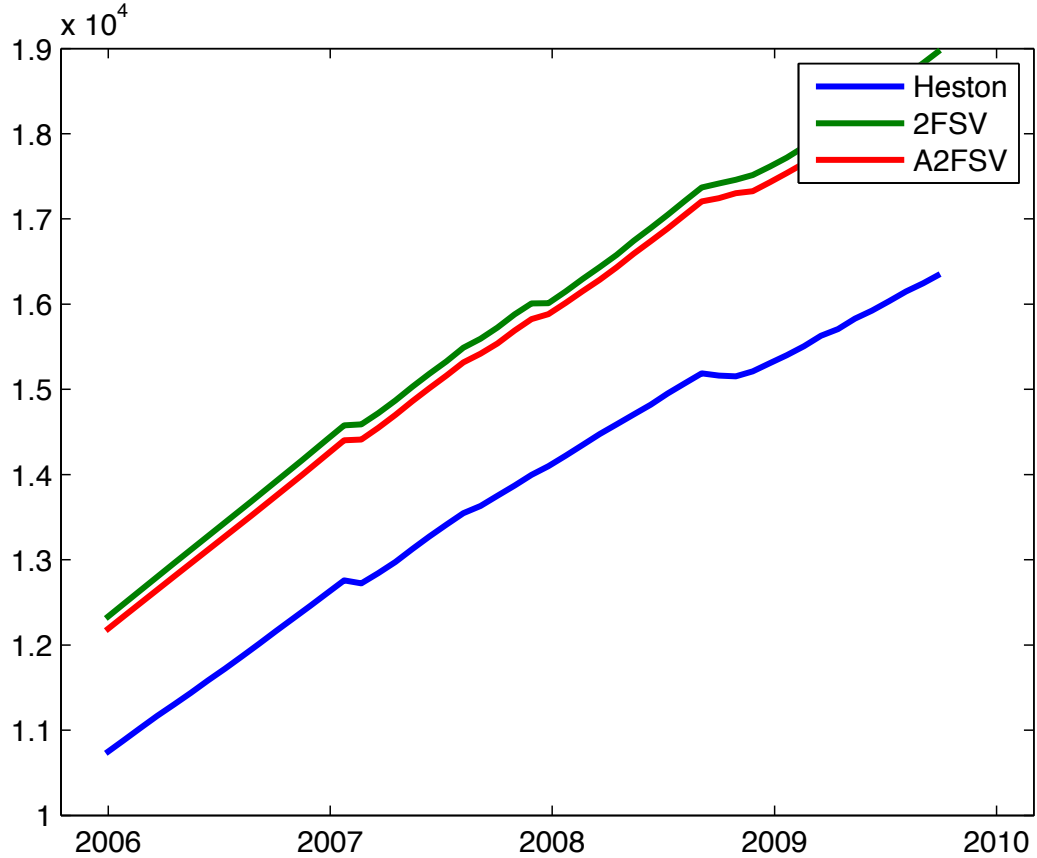


Figure 4: Comparison of Log-likelihood values \mathcal{L} for the Heston model, 2-factor SV model and alternative 2-factor SV model

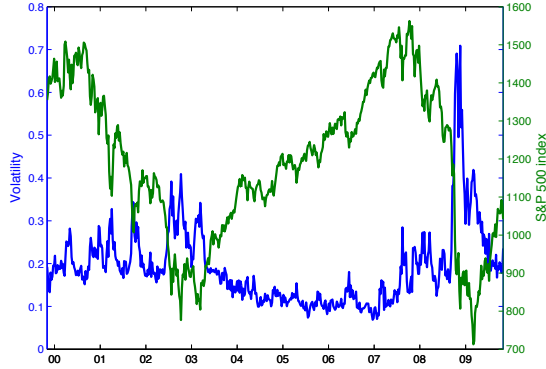
Figures 5(a) – 5(c) illustrate the evolution of the volatility, long-run target variance and variance risk premium, which are contrasted to the evolution of the S&P500 index during that time. In addition, Figure 5(d) compares the evolution of the variance and long-run target variance. The evolution of the volatility clearly reflects the different market conditions of the analyzed time period with low volatility during calm periods and high volatility during turbulent periods. Similarly, the long-run target variance is also low during calm periods and high during turbulent periods. The advantage of the long-run target process as an indicator

ϕ_1 and ϕ_2 , which results in $-\kappa_1^*(\phi_2 - \phi_1)$.

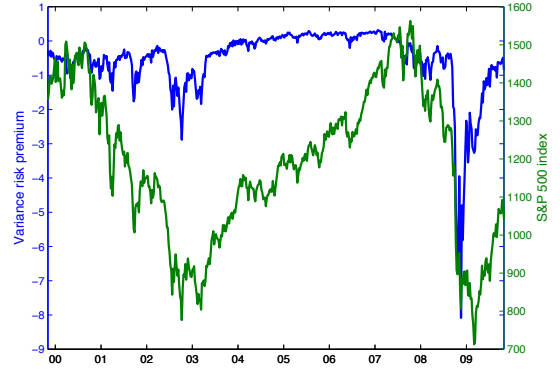
for the market sentiment is its lower variability compared to the variance process. This makes it easier to identify changing market conditions, which can also be seen in Figure 5(d). Interestingly, the variance risk premium estimates are positive from mid-2004 to the beginning of 2007, with a maximum variance risk premium of 32.11% on January 24, 2007. This indicates that the market participants are risk-seeking during that time.

In order to account for parameter instability in the context of the forecasting application, we repeatedly re-estimate the 2-factor SV model based on an increasing data set. To do so we increase the sample size after each estimation by four additional observation days, which approximately cover one month. While the underlying sample period always starts on November 3, 1999, the end of the sample period ranges from December 28, 2005 to September 30, 2009. The corresponding estimation results are illustrated in Figure 6, which also shows the evolution of the S&P500 index during that time.

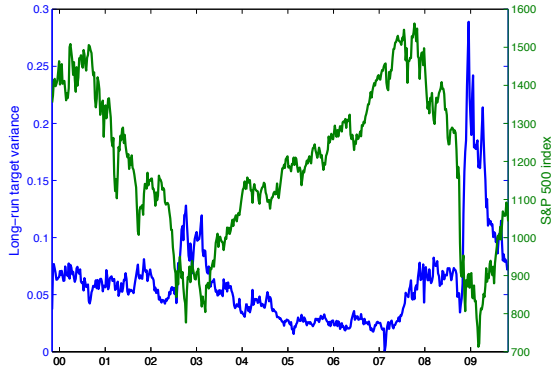
Although all parameters show some variation over the whole period, some parameters are stable, such as the long-term averages of the variance and long-run target process ϕ_1 and ϕ_2 , which only exhibit a small variation. Interestingly, the long-run target variance level of market participants ϕ_2 continuously decreases until the bankruptcy of Lehman Brothers in September 2008 and then suddenly increases. Also, the volatility of the long-run target process λ_2 is stable and only slightly increases during the time period considered. By contrast, the volatility of the variance λ_1 is only stable until the bankruptcy of Lehman Brothers and then increases considerably, which reflects the increased variability of the variance during that time. Similarly, the risk-neutral mean-speed of reversion of the variance κ_1^* and the correlation ρ are stable until mid-2008 and then increase and decrease, respectively. The decreasing correlation results in a higher negative skewness of the return distribution, so that negative returns are more likely during these turbulent market conditions. By contrast, the increase of κ_1^* at the peak of the subprime mortgage crisis is not intuitive, since this increase reduces the variability of the variance process under the risk-neutral measure. A possible explanation for this behavior could be that the increase in κ_1^* partly offsets the strong increase in λ_1 . Finally, the mean-speed of reversion of the variance κ_1 and the risk aversion κ_2 are unstable. This is due to the general difficulty in estimating the mean-speed of reversion



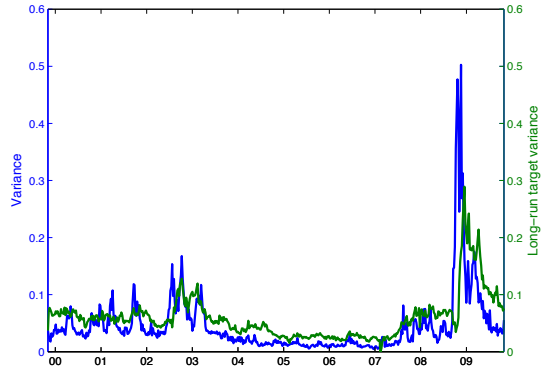
(a)



(b)



(c)



(d)

Figure 5: Volatility, long-run target variance and variance risk premium estimates for the S&P500 index based on the 2-factor SV model. These figures illustrate the evolution of the volatility (a) and the variance risk premium (b), which are estimated via the UKF based on the 2-factor SV model. In addition, figure (c) illustrates the evolution of the long-run target variance, which is also contrasted to the evolution of the variance in figure (d).



Figure 6: Evolution of the 2-factor SV model parameters before and during the subprime mortgage crisis. These figures illustrate the evolution of the parameters of the 2-factor SV model (blue line) before and during the subprime mortgage crisis. Additionally, we contrast these estimation results to the evolution of the S&P500 index (green line).

under the statistical measure. Nevertheless, both parameters decrease at the peak of the subprime mortgage crisis, which reflects the higher variability of the variance and long-run target variance during turbulent market conditions.

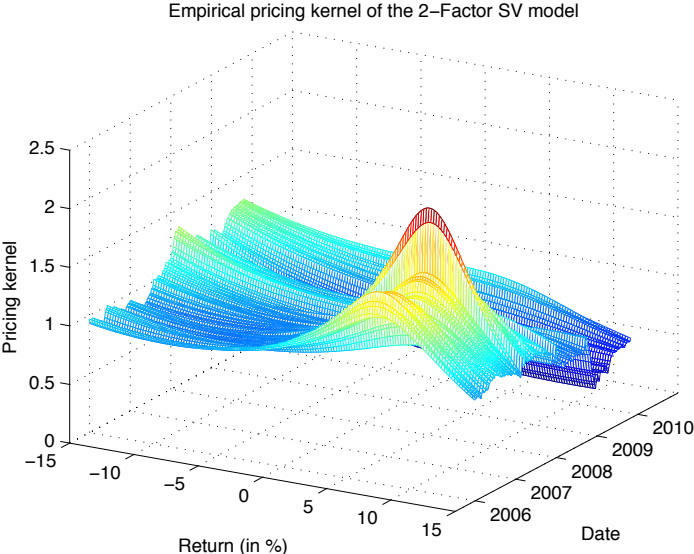


Figure 7: Empirical Pricing Kernel of the 2-factor SV Model

We conclude this section with a discussion of the evolution of the pricing kernel as illustrated in Figure 7. Again, the different market conditions are clearly reflected in the shape of the pricing kernel. More precisely, before the crisis the pricing kernel was S-shaped, as typically found in the literature, with the strongest characteristic on February 21, 2007. Due to the connection between the level of risk aversion and the shape of the pricing kernel (see Section 2.4) this indicates that market participants became more and more risk seeking until the beginning of 2007. Thereafter, the S-shape got less pronounced, reflecting the increasing fear of the market participants, ultimately turning into a U-shape from October 1, 2008 – i.e. shortly after the bankruptcy of Lehman Brothers. The pricing kernel returned to an S-shape from July 15, 2009 on, which can be interpreted that market sentiment started to relax from there on.

4. Value-at-Risk forecasting

In the following section, we assess the forecasting performance of the 2-factor SV model¹². We perform out-of-sample Value-at-Risk (VaR) forecasts for the S&P500 index with forecasting horizons of 1, 2, 3 and 4 weeks at confidence levels of 95%, 99% and 99.9%. Our forecasting period starts on January 4, 2006 and ends on October 28, 2009, encompassing 200 weekly observations of the S&P500 index and the corresponding call options. The total forecasting period is further subdivided into a precrash period, from January 4, 2006 to December 5, 2007, and a crash period, from December 12, 2007 to October 28, 2009. The crash period includes all major events of the subprime mortgage crisis, which begins with the bailout of Bear Stearns by the Federal Reserve System (Fed) together with JPMorgan Chase on March 14, 2008 followed by the bailout of Fannie Mae and Freddie Mac by the Fed on July 3, 2008 and then reaching its climax with the bankruptcy of Lehman Brothers on September 15, 2008. Both the precrash and crash period each encompass 100 weekly observations. This subdivision of the forecasting period has the objective of checking whether the forecasting performance deteriorated during the subprime mortgage crisis, and if so, to what extent. In order to better classify the forecasting performance of the 2-factor SV model, we also perform VaR forecasts based on five alternative models, namely the Heston, the Heston-Nandi, and the Black-Scholes model, which also rely on option-implied information and the GARCH model, which relies on historical return information. In addition, we investigate an alternative version of the 2-factor SV model, where the short- and long-run volatility components are present in the volatility dynamics under the statistical measure. In order to account for parameter instability, we re-estimate the GARCH, Heston, Heston-Nandi and 2-factor SV models every four weeks anew based on an enlarged data set, which also includes the four additional observation days while keeping the beginning of the data set always fixed on November 3, 1999. By contrast, the Black-Scholes model only requires the average Black-Scholes implied volatilities at the respective forecasting date, so that no re-estimation based

¹²For a comparison of option pricing performance between the Heston and the two stochastic volatility models see AppendixC

on an enlarged data set is necessary.

4.1. Forecasting models

Before we describe the forecasting models in more detail, we first make the simplifying assumption that the expected return over the forecasting period is zero. According to Daníelsson and Zigrand (2006), this assumption can be justified by two arguments. On the one hand, the forecasting horizon of risk models is usually very short, so that the drift term is negligible compared to the volatility during that time. On the other hand, there exists so far no obvious way to obtain an accurate estimate of the drift term. In particular, Merton (1980), p. 333, argues that a negative estimate of the drift term is a biased-low estimate, which would be very likely for our specific data period. Based on this simplifying assumption, we introduce the different VaR models, whose key features are summarized in Table 3.

4.1.1. Black-Scholes model

Our first forecasting model is the Black-Scholes model, which assumes that the asset price evolves as a geometric Brownian motion with constant volatility σ . Hence, the log return over the period $[i, i+\tau]$, written $r_{i,i+\tau}$, is distributed as

$$r_{i,i+\tau} \sim \mathcal{N}(0, \tau\sigma^2) , \quad (18)$$

which already accounts for the zero drift assumption. In contrast to the constant volatility assumption of the Black-Scholes model, we repeatedly estimate the volatility parameter σ based on all currently observable call-option prices. So, we simply build the average of the corresponding Black-Scholes implied volatilities, and σ in (18) becomes

$$\bar{\sigma}_{BS,i} = \frac{1}{8} \sum_{j=1}^8 \sigma_{BS}(\bar{C}_{j,i}) .$$

4.1.2. GARCH model

In the GARCH model the return is specified as

$$r_i = \sigma_i z_i , \quad (19)$$

Model	Return distribution	Data basis	Forecasting type
GARCH	Normal distribution	historical + current returns	\mathbb{P} -Backward-looking
Black-Scholes	Normal distribution	current option prices	\mathbb{Q} -Forward-looking
Heston	Heston dynamics	historical + current index and option prices	\mathbb{Q} -Forward-looking + pricing kernel (constant risk preferences)
Heston-Nandi	Heston-Nandi dynamics	historical + current index and option prices	\mathbb{Q} -Forward-looking + pricing kernel (constant risk preferences)
2-factor SV	Heston dynamics	historical + current index and option prices	\mathbb{Q} -Forward-looking + pricing kernel (time-varying risk preferences)
Alternative 2-factor SV	2-factor SV dynamics	historical + current index and option prices	\mathbb{Q} -Forward-looking + pricing kernel (time-varying risk preferences)

Table 3: Overview of the Value-at-Risk forecasting models.

with z_i a standard normal random variable. For the evolution of the conditional return variance, we use a GARCH(1,1) model,

$$\sigma_i^2 = \omega + \alpha r_{i-1}^2 + \beta \sigma_{i-1}^2 . \quad (20)$$

We chose GARCH(1,1) not only for its simplicity but also its proven superior forecasting performance compared to more complex specifications (see e.g. Lunde and Hansen (2005)). Further note that GARCH(1,1) with the particular choice of $\alpha = 0.04$ and $\beta = 0.96$ for one-day-ahead forecasts and $\alpha = 0.03$ and $\beta = 0.97$ for one-month-ahead forecasts is essentially the volatility forecasting model of Riskmetrics, commonly used in the banking industry. By iterating (20) one time step forward, we obtain the variance forecast for the period $[i, i+1]$. Since our VaR forecasting horizon will not always correspond to the time spacing of the data sample, we also apply the square-root-of-time rule in order to get a variance forecast for the period $[i, i+\tau]$.¹³ Consequently, the return for the period $[t, t+\tau]$ is approximately normally distributed conditional on the variance:

$$r_{i,i+\tau} \approx \mathcal{N}\left(0, \frac{\tau}{\Delta t} \sigma_{i+1}^2\right) .$$

Here, we assumed that the time spacing of the data sample is equidistant with interval length Δt .

4.1.3. Heston and 2-factor SV model

Our next two VaR models are the Heston and the 2-factor SV model, which both assume that the asset price under the statistical measure evolves as

$$dS_t = \mu S_t dt + \sqrt{\sigma_t} S_t dW_{1,t} , \quad (21)$$

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

¹³Alternatively, the variance forecast for time-horizons larger than one time step can be obtained by further iterations of (20) (see McNeil et al. (2005), p. 159, for an explicit formula). However, similar to the findings in Brummelhuis and Kaufmann (2007), §4 and §6, in our empirical study the square-root-of-time rule worked much better.

with $dW_{1,t}dW_{2,t} = \rho dt$. However, due to the additional long-run target process of the 2-factor SV model, the estimation results for the parameters in (21) are generally different for the two models. In contrast to the Black-Scholes model, the stochastic differential equation (21) has no closed-form solution, so that the return distribution cannot be stated in closed form. However, the characteristic function of the log-asset price is known in semi-closed form, so that the return distribution can be obtained via Fourier inversion.¹⁴ Unfortunately, since the stochastic differential equation (21) has no closed-form solution, we can only approximately impose the zero drift condition by setting $\mu = 1/2 \sigma_t$.

4.1.4. Heston-Nandi model

The Heston-Nandi model is a one-factor stochastic volatility model where the variance is modeled by a GARCH-type process. In the following we consider the GARCH(1,1) version of this model which contains the classical Heston model as a continuous-time limit. The asset price under the statistical measure for this single lag version is given by

$$\begin{aligned}\ln(S_t) &= \ln(S_{t-1}) + r + \lambda\sigma_t + \sqrt{\sigma_t}\epsilon_t, \\ \sigma_t &= \omega + \beta_1\sigma_{t-1} + \alpha_1(\epsilon_{t-1} - \gamma_1\sqrt{\sigma_{t-1}})^2.\end{aligned}$$

The parameter λ determines the return premium being proportional to the variance σ_t . The parameters ω , α_1 , β_1 and γ_1 specify the GARCH(1,1) process which is stationary with finite mean and variance if $\beta_1 + \alpha_1\gamma_1^2 < 1$. The kurtosis of the return distribution is controlled by α_1 and the skewness by γ_1 . In particular, positive values for α_1 and γ_1 imply a negative correlation between return and variance and thus reflect the leverage effect. The characteristic function of the log-asset price is known in closed form so that the return distribution can be obtained via Fourier inversion. For the purpose of the Value-at-Risk forecasting we impose the zero drift condition $r = -\lambda\sigma_t$.

¹⁴See Appendix D for more details.

4.1.5. Alternative 2-factor SV model

Our last VaR model is an alternative version of the 2-factor SV model, where we also introduce a long-run component of the volatility process under the statistical measure. So the analysis of this model allows one to quantify the additional value of time-varying long-run volatility under the statistical measure. We assume that the asset price under the statistical measure evolves as

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

where $dW_{1,t} dW_{2,t} = \rho dt$. Furthermore, $W_{3,t}$ is uncorrelated to $W_{1,t}$ and $W_{2,t}$. The characteristic function of the log-asset price under \mathbb{P} is already specified in Proposition 2.1, where r needs to be replaced by μ and κ_1^* by κ_1 . Again, the characteristic function can be used to calculate the return distribution via Fourier inversion. In doing so, we approximately impose the zero-drift condition by setting $\mu = 1/2 \sigma_t$.

The risk-neutral dynamics of this alternative 2-factor SV model are assumed to have the same structure as the 2-factor SV model. However the supposed risk premia – and hence also the pricing kernel – are different. The derivation of the risk-neutral dynamics of the alternative 2-factor SV model is described in Appendix B.

4.2. Forecasting results

We now assess the out-of-sample forecasting performance of the VaR models by applying the test of Kupiec (1995), which checks whether the proportion of VaR exceptions is consistent with the chosen confidence level $\alpha\%$. Instead of calculating the VaR based on a profit or loss distribution, we can alternatively calculate the VaR based on the corresponding return distribution. Then the $\alpha\%$ -VaR can be interpreted as the maximum percentage loss, which is not exceeded with a probability of (at least) $\alpha\%$ within a specific forecasting period τ . If the realized return within this period is smaller than its VaR forecast, we call this a VaR exception. For a correctly specified VaR model, the total number of VaR exceptions should

VaR-exceptions:	1	2	3	4	5	6	7	8	> 9
Basel II “traffic-light”:	Green			Yellow				Red	
Chosen categorization:	Green				Yellow	Orange	Red		

Table 4: Two categorizations for the accuracy of VaR models

amount to $(100 - \alpha)\%$ of all VaR forecasts, which is also the null hypothesis of the Kupiec (1995) test statistic. We categorize the forecasting performance of the analyzed VaR models by means of the corresponding p -value of Kupiec’s test statistic. In doing so, we call a VaR model “slightly accurate”, indicated by one asterisk, if $0.01 \leq p < 0.05$, “accurate”, indicated by two asterisks, if $0.05 \leq p < 0.1$ and “highly accurate”, indicated by three asterisks, if $p \geq 0.1$. Furthermore, VaR models with p -values smaller than 0.01 are referred to as “inaccurate”. This categorization can also be motivated by the traffic-light approach of Basel II, which specifies as to when banks VaR models are inaccurate and if so which penalty surcharge is associated. Table 4 compares the traffic-light approach of Basel II with the chosen categorization of this empirical study, where a confidence level of 99% together with 250 VaR forecasts is supposed.

The number of VaR exceptions for all confidence levels and forecasting horizons and the corresponding p -values are summarized in Tables 5 – 7 for the total, precrash and crash period, respectively.

4.2.1. GARCH model

The GARCH model is able to forecast the 95% VaR with high accuracy during the pre-crash period, irrespective of the forecasting horizon. Accordingly, the square-root-of-time rule, which in our context scales one-week volatility forecasts to up to four-week volatility forecasts, works well. However, the forecasting performance deteriorates with increasing confidence level. For example, VaR forecast with a confidence level of 99% and forecasting horizons of 1, 2 or 3 weeks are only slightly accurate. This bad forecasting performance for large confidence levels is not surprising, since we assumed that returns are normally

Table 5: Value-at-Risk forecasting performance – Total period (January 4, 2006 – October 28, 2009).

Forecasting horizon (weeks)	1			2			3			4		
VaR level ($\alpha \cdot 100\%$)	95%	99%	99.9%	95%	99%	99.9%	95%	99%	99.9%	95%	99%	99.9%
GARCH	13***	7	5	14***	8	4	20	10	6	17*	9	5
p -value	(0.35)	(0.006)	(0.00)	(0.21)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.03)	(0.00)	(0.00)
Black-Scholes	8***	4***	1***	9***	5**	2*	11***	4***	2*	12***	5**	2*
p -value	(0.50)	(0.21)	(0.20)	(0.75)	(0.07)	(0.02)	(0.72)	(0.21)	(0.02)	(0.50)	(0.07)	(0.02)
Heston	9***	4***	3	10***	6*	3	12***	4***	3	12***	7	2***
p -value	(0.74)	(0.21)	(0.00)	(0.99)	(0.02)	(0.00)	(0.51)	(0.21)	(0.00)	(0.50)	(0.005)	(0.02)
Heston-Nandi	13***	6*	4	12***	7*	5	15***	7*	3	15***	8	4
p -value	(0.35)	(0.02)	(0.00)	(0.52)	(0.01)	(0.00)	(0.12)	(0.01)	(0.00)	(0.11)	(0.00)	(0.00)
2-factor SV	9***	1***	0***	8***	2***	0***	8***	2***	0***	12***	3***	1***
p -value	(0.74)	(0.43)	(0.53)	(0.51)	(0.99)	(0.53)	(0.52)	(0.99)	(0.53)	(0.50)	(0.49)	(0.20)
Alternative 2-factor SV	7***	1***	0***	9***	1***	0***	8***	2***	0***	7***	2***	0***
p -value	(0.30)	(0.43)	(0.53)	(0.75)	(0.43)	(0.53)	(0.52)	(0.99)	(0.53)	(0.33)	(0.98)	(0.53)

Notes: This table summarizes the number of VaR exceptions of the GARCH, Black-Scholes, Heston, Heston-Nandi, 2-factor SV and alternative 2-factor SV model for forecasting horizons of 1, 2, 3 and 4 weeks together with confidence levels of 95%, 99% and 99.9%. In order to test whether the VaR with a confidence level of $\alpha\%$ is violated more or less than $(1 - \alpha)\%$ of the time, we apply the test of Kupiec (1995). The reported p -values correspond to the test statistic of Kupiec (1995), where $0.01 \leq p < 0.05$ is indicated by *, $0.05 \leq p < 0.1$ by ** and $p > 0.1$ by ***.

Table 6: Value-at-Risk forecasting performance – Precrash period (January 4, 2006 – December 5, 2007).

Forecasting horizon (weeks)	1				2				3				4			
	95%	99%	99.9%	99.9%	95%	99%	99.9%	99.9%	95%	99%	99.9%	99.9%	95%	99%	99.9%	99.9%
VaR level ($\alpha \cdot 100\%$)	5***	4*	3	3	6***	4*	1**	1**	8***	4*	2	2	4***	2***	1**	99.9%
GARCH	(1.00)	(0.02)	(0.00)	(0.00)	(0.66)	(0.02)	(0.09)	(0.02)	(0.20)	(0.02)	(0.00)	(0.00)	(0.64)	(0.38)	(0.09)	(0.09)
p -value	4***	3***	0***	0***	4***	2***	0***	0***	5***	1***	0***	0***	4***	0***	0***	0***
Black-Scholes	(0.64)	(0.10)	(0.65)	(0.65)	(0.64)	(0.38)	(0.65)	(0.65)	(1.00)	(1.00)	(0.65)	(0.65)	(0.64)	(0.16)	(0.65)	(0.65)
p -value	5***	3***	2	2	5***	3***	2	2	6***	2***	1**	1**	4***	2***	0***	0***
Heston	(1.00)	(0.10)	(0.00)	(0.00)	(1.00)	(0.10)	(0.00)	(0.00)	(0.66)	(0.38)	(0.09)	(0.09)	(0.64)	(0.38)	(0.65)	(0.65)
p -value	7***	4*	3	3	7***	3***	2	2	8***	3***	1	1	6***	3***	1**	1**
Heston-Nandi	(0.39)	(0.02)	(0.00)	(0.00)	(0.39)	(0.10)	(0.00)	(0.00)	(0.20)	(0.10)	(0.09)	(0.09)	(0.66)	(0.10)	(0.09)	(0.09)
p -value	4***	0***	0***	0***	3***	0***	0***	0***	2***	0***	0***	0***	1*	0***	0***	0***
2-factor SV	(0.64)	(0.16)	(0.65)	(0.65)	(0.32)	(0.16)	(0.65)	(0.65)	(0.12)	(0.16)	(0.65)	(0.65)	(0.03)	(0.16)	(0.65)	(0.65)
p -value	4***	0***	0***	0***	4***	0***	0***	0***	2***	0***	0***	0***	1*	0***	0***	0***
Alternative 2-factor SV	(0.64)	(0.16)	(0.65)	(0.65)	(0.64)	(0.16)	(0.65)	(0.65)	(0.12)	(0.16)	(0.65)	(0.65)	(0.03)	(0.16)	(0.65)	(0.65)
p -value	4***	0***	0***	0***	4***	0***	0***	0***	2***	0***	0***	0***	1*	0***	0***	0***

Notes: This table summarizes the number of VaR exceptions of the GARCH, Black-Scholes, Heston, Heston-Nandi, 2-factor SV and alternative 2-factor SV model for forecasting horizons of 1, 2, 3 and 4 weeks together with confidence levels of 95%, 99% and 99.9%. In order to test whether the VaR with a confidence level of $\alpha\%$ is violated more or less than $(1 - \alpha)\%$ of the time, we apply the test of Kupiec (1995). The reported p -values correspond to the test statistic of Kupiec (1995), where $0.01 \leq p < 0.05$ is indicated by *, $0.05 \leq p < 0.1$ by ** and $p > 0.1$ by ***.

Table 7: Value-at-Risk forecasting performance – Crash period (December 12, 2007 – October 28, 2009).

Forecasting horizon (weeks)	1		2		3		4		
	95%	99%	99.9%	95%	99%	99.9%	95%	99%	
VaR level ($\alpha \cdot 100\%$)	8*** (0.20)	3*** (0.10)	2 (0.00)	8*** (0.19)	4* (0.02)	3 (0.00)	12 (0.005)	4 (0.00)	99% (0.00)
GARCH	8***	3***	2	8***	4*	3	12	4	7
p -value	(0.20)	(0.10)	(0.00)	(0.19)	(0.02)	(0.00)	(0.005)	(0.00)	(0.00)
Black-Scholes	4***	1***	1**	5***	3***	2	6***	2	8***
p -value	(0.64)	(1.00)	(0.09)	(0.98)	(0.10)	(0.00)	(0.62)	(0.00)	(0.18)
Heston	4***	1***	1**	5***	3***	1**	6***	2	8***
p -value	(0.64)	(1.00)	(0.09)	(0.98)	(0.10)	(0.09)	(0.62)	(0.00)	(0.18)
Heston-Nandi	6***	2***	1**	5***	4*	3	7***	2	9**
p -value	(0.66)	(0.38)	(0.09)	(0.98)	(0.02)	(0.00)	(0.36)	(0.02)	(0.08)
2-factor SV	5***	1***	0***	5***	2***	0***	6***	0***	11*
p -value	(1.00)	(1.00)	(0.65)	(0.98)	(0.37)	(0.66)	(0.62)	(0.36)	(0.01)
Alternative 2-factor SV	3***	1***	0***	5***	1***	0***	6***	0***	6***
p -value	(0.32)	(1.00)	(0.65)	(0.98)	(0.99)	(0.66)	(0.62)	(0.36)	(0.60)

Notes: This table summarizes the number of VaR exceptions of the GARCH, Black-Scholes, Heston, Heston-Nandi, 2-factor SV and alternative 2-factor SV model for forecasting horizons of 1, 2, 3 and 4 weeks together with confidence levels of 95%, 99% and 99.9%. In order to test whether the VaR with a confidence level of $\alpha\%$ is violated more or less than $(1 - \alpha)\%$ of the time, we apply the test of Kupiec (1995). The reported p -values correspond to the test statistic of Kupiec (1995), where $0.01 \leq p < 0.05$ is indicated by *, $0.05 \leq p < 0.1$ by ** and $p > 0.1$ by ***.

distributed. However, as seen in §3.2, the distribution of the S&P500 index returns is slightly negatively skewed and fat-tailed. So, the normality assumption results in an underestimation of extreme events, which in turn results in too many VaR exceptions for large confidence levels. Also, the VaR forecasting performance of the GARCH model deteriorates in the crash period, where only one-week-ahead VaR forecasts with confidence levels of 95% and 99% as well as two-week-ahead VaR forecasts with a confidence level of 95% are highly accurate. Especially, for forecasting horizon larger than two weeks, all VaR forecasts are inaccurate. The bad forecasting performance of the GARCH model during the crash period can be explained by its parameter instability, which is illustrated in Figure 8. While the GARCH parameters α and β are stable during the pre-crash period, they become strongly unstable during the crash period. In particular, at the beginning of the crash period the parameters α and β increase and decrease significantly, respectively. Recall that the one-week variance forecast of the GARCH model can be obtained by iterating (20) one time step forward. Hence, the one-week variance forecast is a weighted average of current and past information. So, the increase of α and the decrease of β can be interpreted as saying that during turbulent market conditions current information becomes more important and past information less important.

4.2.2. Black-Scholes model

The VaR forecasting performance of the Black-Scholes model is remarkably good during the pre-crash period as all VaR forecasts are highly accurate, irrespective of the confidence level and forecasting horizon. As for the GARCH model, the forecasting performance of the Black-Scholes model deteriorates during the crash period. However, despite this challenging forecasting period, the Black-Scholes model is still able to deliver highly accurate VaR forecasts for confidence levels of 95% and 99% together with forecasting horizons of 1, 2 and 3 weeks. Interestingly, the Black-Scholes model performs well also for large confidence levels, in particular, during the pre-crash period. This is quite surprising since just like the GARCH model the Black-Scholes model also relies on the normality assumption for the return distribution, which should be inadequate for large confidence levels. This phenomenon

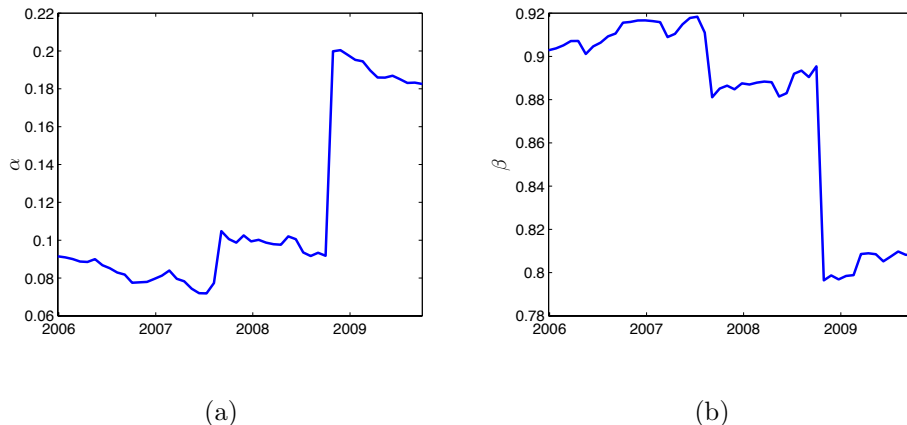


Figure 8: Evolution of the GARCH parameters. These figures illustrate the evolution of the parameters α (a) and β (b) of the GARCH model specified in (19) - (20). The x-axis indicates the respective end date of the S&P500 index return time series, which always starts on November 03, 1999, based on which the GARCH parameters are estimated. Furthermore, the y-axis denotes the corresponding estimation results for the parameters α and β .

can be explained by a negative variance risk premium, which results in higher volatility estimates under the risk-neutral measure than under the statistical measure. As a result, the Black-Scholes model is generally more conservative than the GARCH model. Furthermore, the deviation between risk-neutral and statistical volatility estimates gets larger during times of high volatility. So, the VaR forecasts of the Black-Scholes model become even more conservative during turbulent market conditions.

4.2.3. Heston model

The VaR forecasting performance of the Heston model is very similar to that of the Black-Scholes model, with slight disadvantages during the pre-crash period and slight advantages during the crash period. But the overall forecasting performance of the Black-Scholes model is better than of the Heston model, even though the Black-Scholes model only relies on risk-neutral information and, in addition, assumes a simpler return distribution. A possible explanation for the inferior forecasting performance of the Heston model might be that the corresponding pricing kernel is too restrictive to account for time-varying risk preferences.

The need for a more flexible modeling of risk preferences can be seen in Figure 9 from the strong time variation of the variance risk premium parameter $\Delta\kappa$, which reflects the risk preferences in the Heston model. So, a given estimate of $\Delta\kappa$ represents an average value for the risk preferences in the underlying sample period, which might deviate from the current risk preferences. Accordingly, the attempt to risk-adjust the risk-neutral density of the Heston model by virtue of a historical average variance risk premium parameter might result in a biased real-world density.

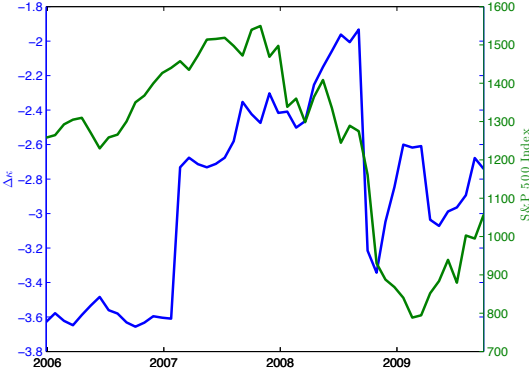


Figure 9: Evolution of the variance risk premium parameter $\Delta\kappa$ of the Heston model.

4.2.4. *Heston-Nandi model*

As expected, the Heston-Nandi model performs overall better than the basic GARCH model. For the pre-crash period its performance is similar to the Heston model. During the crash period it performs worse than the Heston model for the higher confidence levels (99% and 99,9%). This may be due to our parameter setting, which implies a symmetric return distribution. In summary, we find that the Heston-Nandi model performs slightly worse than the Heston model.

4.2.5. *2-factor SV model*

The 2-factor SV model performs remarkably well during the pre-crash period with highly accurate VaR forecasts for almost all confidence levels and forecasting horizons. The only exception is the 95% four-week-ahead VaR, which is only slightly accurate. Moreover, during

the crash period, the 2-factor SV model is able to perform highly accurate VaR forecasts for forecasting horizons up to three weeks, irrespective of the confidence level. Most notably, if we consider the total forecasting period, the 2-factor SV model is able to perform highly accurate VaR forecasts for all confidence levels and forecasting horizons. The superior forecasting performance of the 2-factor SV model compared to the models above is due to its ability to account for time-varying preferences, which in turn allows for a more accurate estimation of the real-world density.

4.2.6. Alternative 2-factor SV model

The VaR forecasting performance of the alternative 2-factor SV model during the pre-crash period is almost identical to the one of the 2-factor SV model. Only for the 95% two-week-ahead VaR the alternative 2-factor SV model results in one additional VaR exception compared to the 2-factor SV model. Also during the crash period, the alternative 2-factor SV model is able to perform highly accurate VaR forecasts for all forecasting horizons and confidence levels. Hence, for long forecasting horizons (four-weeks) the forecasting performance of the alternative 2-factor SV model is slightly better than the one of the 2-factor SV model. Like the 2-factor SV model, the alternative 2-factor SV model is able to perform highly accurate VaR forecasts for all forecasting horizons and confidence levels. The forecasting performance of the alternative 2-factor SV suggests that for long forecasting horizons under stressed market conditions the additional “time-varying long-run variance” under the statistical measure becomes relevant. During calm market conditions and for shorter forecast horizons the more parsimonious 2-factor model seems to be preferable.

5. Conclusion

We presented a new option-pricing model, called the 2-factor SV model, which allows one to capture the empirical properties of pricing kernels. In particular, our model can produce S-shaped and time-varying pricing kernels. The basis of the 2-factor SV model is an additional “long-run target process”, which acts only under the risk-neutral measure and describes the central tendency of the risk-neutral variance. Moreover, the long-run target

process also significantly determines the sign and magnitude of the model's variance risk premium and thus also the supposed risk preferences.

We applied our framework in order to analyze the risk preferences of market participants invested in S&P500 index options from 2001 to 2009. In line with economic theory, we found that risk aversion increases during stressed market conditions and relaxes during normal market conditions. Interestingly, some time before the subprime mortgage crisis, we found evidence for risk-seeking behavior in terms of a positive variance risk premium and an increasing S-shape characteristic of the pricing kernel. In addition, the long-run target process allows one to easily identify changing market conditions and thus might be an interesting quantity for analysts and central banks.

Moreover, we extracted the forward-looking information from S&P500 index options and then performed out-of-sample Value-at-Risk (VaR) forecasts during the period of the subprime mortgage crisis. In order to better classify the corresponding forecasting results, we also performed VaR forecasts based on five alternative VaR models, namely the Heston, Heston-Nandi, Black-Scholes, and an alternative 2-factor model, which also rely on option-implied information, and a GARCH model, which relies on historical information only. We found that the 2-factor SV models have the best forecasting performance, followed by the Black-Scholes, the Heston, the Heston-Nandi and the GARCH model. In particular, the 2-factor SV models are the only ones able to perform highly accurate VaR forecasts for all confidence levels (95%, 99% and 99.9%) and forecasting horizons (1, 2, 3 and 4 weeks), despite the challenging forecasting period. These results highlight the value of option price data for forecasting, but also the need to account for time-varying risk aversion when dealing with this kind of information.

AppendixA. Characteristic function of the 2-factor SV model

Before we derive the characteristic function of the 2-factor SV model, we first need to define and characterize affine processes. Following (Kahl, 2007, pp. 13 ff.), let $W_t =$

$[W_{1,t} \dots W_{d,t}]^\top$ be a d -dimensional standard Brownian motion defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Next, define the d -dimensional stochastic process $X_t = [X_{1,t} \dots X_{d,t}]^\top$ via the following stochastic differential equation:

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

The drift vector $\mu : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is assumed to be continuous differentiable and the diffusion matrix $\sigma : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$ to be twice continuous differentiable. Now, an affine process can be defined as follows (Kahl, 2007, p. 15):

Definition AppendixA.1 (Affine process). *A process X_t is called affine if the covariance matrix $\sigma(x)\sigma(x)^\top$ and the drift $\mu(x)$ are affine in x . That is,*

$$\sigma(x)\sigma(x)^\top = a + \sum_{i=1}^d x_i \alpha_i , \tag{A.1}$$

$$\mu(x) = b + \sum_{i=1}^d x_i \beta_i ,$$

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

As first shown by Duffie et al. (2000), the characteristic function of an affine process X_t is known up to the solution of a system of Riccati equations and can be specified via the following theorem (Filipović and Mayerhofer, 2009, p. 127):¹⁵

Theorem AppendixA.1. *Suppose X_t is an affine process with drift and covariance matrix specified by (A.1). Then the \mathcal{F}_t -conditional characteristic function of X_T is exponential affine in X_t for all $t \leq T$. That is, there exist \mathbb{C} - and \mathbb{C}^d -valued functions $\phi(t, u)$ and $\psi(t, u)$ with jointly continuous t -derivatives, such that*

$$\mathbb{E}[e^{iu^\top X_T} | \mathcal{F}_t] = e^{\phi(T-t, u) + \psi(T-t, u)^\top X_t} , \tag{A.2}$$

¹⁵See Filipović and Mayerhofer (2009) for a proof.

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

$$\begin{aligned}\partial_t \phi(t, u) &= \frac{1}{2} \psi(t, u)^\top a \psi(t, u) + b^\top \psi(t, u), \\ \partial_t \psi_i(t, u) &= \frac{1}{2} \psi(t, u)^\top \alpha_i \psi(t, u) + \beta_i^\top \psi(t, u), \quad 1 \leq i \leq d,\end{aligned}\tag{A.3}$$

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

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

The functions ϕ and ψ will also be called characteristic exponents. For some special cases, as e.g. the Heston model, the ordinary differential equations (ODE)s (A.3) can be solved explicitly. Unfortunately, in many cases this is not possible, so that numerical methods need to be applied in order to approximate the solution. Duffie et al. (2000) propose to use the Runge-Kutta method in this context.

The next lemma shows that the 2-factor model belongs to the class of affine processes.

Lemma AppendixA.1. *Let S_t , σ_t and θ_t be the risk-neutral asset-price dynamics of the 2-factor SV model specified by (3). Then $X_t = [\ln(S_t) \ \sigma_t \ \theta_t]^\top$ is an affine process with drift and covariance specified by (A.1), where*

$$b = \begin{bmatrix} r \\ 0 \\ \kappa_2 \phi_2 \end{bmatrix}, \quad \beta_1 = 0, \quad \beta_2 = \begin{bmatrix} -1/2 \\ -\kappa_1^* \\ 0 \end{bmatrix}, \quad \beta_3 = \begin{bmatrix} 0 \\ \kappa_1^* \\ -\kappa_2 \end{bmatrix},\tag{A.4}$$

$$a = 0, \quad \alpha_1 = 0, \quad \alpha_2 = \begin{bmatrix} 1 & \lambda_1 \rho & 0 \\ \lambda_1 \rho & \lambda_1^2 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \alpha_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \lambda_2^2 \end{bmatrix}.\tag{A.5}$$

Proof. A process with drift and covariance specified by (A.1) is by Definition AppendixA.1 affine. Hence, we only need to check if the resulting dynamics coincide with $X_t = [\ln(S_t) \ \sigma_t \ \theta_t]^\top$,

which is equivalent to

$$\begin{aligned}
d\ln(S_t) &= (r - 1/2\sigma_t)dt + \sqrt{\sigma_t}d\widetilde{W}_{1,t} , \\
d\sigma_t &= \kappa_1^*(\theta_t - \sigma_t)dt + \lambda_1\sqrt{\sigma_t}d\widetilde{W}_{2,t} , \\
d\theta_t &= \kappa_2(\phi_2 - \theta_t)dt + \lambda_2\sqrt{\theta_t}d\widetilde{W}_{3,t} .
\end{aligned} \tag{A.6}$$

If we plug (A.4) in $\mu(x) = b + \sum_{i=1}^3 x_i\beta_i$, it is apparent that the resulting drift term is the same as in (A.6). Moreover, the covariance matrix of (A.6) is given by

$$dX_t dX_t = \begin{bmatrix} \sigma_t & \lambda_1\rho\sigma_t & 0 \\ \lambda_1\rho\sigma_t & \lambda_1^2\sigma_t & 0 \\ 0 & 0 & \lambda_2^2\theta_t \end{bmatrix} dt . \tag{A.7}$$

Again, if we plug (A.5) in $\sigma(x)\sigma(x)^\top = a + \sum_{i=1}^3 x_i\alpha_i$, we can see that the resulting covariance matrix coincides with (A.7). \square

Now, we are able to proof Proposition 2.1.

Proof. According to Lemma AppendixA.1, X_t is an affine process, so that the characteristic function of $\ln(S_T)$ can be calculated by virtue of Theorem AppendixA.1. In particular, the ODEs of the characteristic exponents in (4) are simply functions of the affine drift and covariance specification (A.4) and (A.5), respectively. Since we are interested in the characteristic function of $\ln(S_T)$, we make the particular choice of $u = [u \ 0 \ 0]^\top$. Hence, the ODEs for the characteristic exponents are given by

$$\partial_t\phi(t) = r\psi_1(t) + \kappa_2\phi_2\psi_3(t) , \tag{A.8}$$

$$\partial_t\psi_1(t) = 0 , \tag{A.9}$$

$$\partial_t\psi_2(t) = -1/2\psi_1(t) - \kappa_1^*\psi_2(t) + 1/2\psi_1(t)^2 + \lambda_1\rho\psi_1(t)\psi_2(t) + 1/2\lambda_1^2\psi_2(t)^2 , \tag{A.10}$$

$$\partial_t\psi_3(t) = \kappa_1^*\psi_2(t) - \kappa_2\psi_3(t) + 1/2\lambda_2^2\psi_3(t)^2 , \tag{A.11}$$

subject to the boundary conditions $\phi(0) = 0$, $\psi_1(0) = iu$, $\psi_2(0) = 0$ and $\psi_3(0) = 0$. For better clarity, we omitted the dependence of the characteristic exponents on u . Due to the boundary condition $\psi_1(0) = iu$, the solution of the second ODE (A.9) is simply $\psi_1(t) = iu$. Following (Kahl, 2007, p. 17), the third ODE (A.10) can be written as

$$\frac{dy}{dt} = c_1 y^2 + c_2 y + c_3 , \quad (\text{A.12})$$

where we set $y = \psi_2(t)$, $c_1 = \frac{1}{2}\lambda^2$, $c_2 = iu\lambda\rho - \kappa^*$ and $c_3 = \frac{1}{2}(iu + u^2)$. The quadratic form of this ODE allows us to write (A.12) as (Kahl, 2007, p. 18)

$$\frac{dy}{dt} = c_1 (y - \tilde{c}_2^\mp) (y - \tilde{c}_3^\mp) , \quad (\text{A.13})$$

with

$$\tilde{c}_2^+ = \frac{-c_2 + \sqrt{d}}{2c_1} \quad \text{and} \quad \tilde{c}_3^- = \frac{-c_2 - \sqrt{d}}{2c_1} , \quad (\text{A.14})$$

or

$$\tilde{c}_2^- = \frac{-c_2 - \sqrt{d}}{2c_1} \quad \text{and} \quad \tilde{c}_3^+ = \frac{-c_2 + \sqrt{d}}{2c_1} . \quad (\text{A.15})$$

The coefficients \tilde{c}_2^\pm and \tilde{c}_3^\mp , where the superscript indicates the choice of + or - in the nominator, can be obtained by comparison of the coefficients of the equations (A.12) and (A.13). This results in a quadratic equation, wherefore we have two possible solutions. As shown by Albrecher et al. (2007) and Kahl and Loard (2006, 2010), the choice (A.15) is numerically more stable. Therefore, we choose the coefficients (A.15), which will subsequently simply be denoted by \tilde{c}_2 and \tilde{c}_3 .

Next, the ODE (A.13) can be solved by separation of variables, which is

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

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

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

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

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

and hence

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

Finally, the first ODE (A.8) can simply be solved by integrating the differential with respect to t , which is

$$\begin{aligned} \phi(t) &= \int_0^t [r\psi_1(s) + \kappa^* \phi^* \psi_2(s)] ds , \\ &= iurt + \kappa^* \phi^* \frac{1}{\lambda^2} \left(c - \sqrt{d} \right) \int_0^t \frac{1 - e^{-\sqrt{d}s}}{1 - g e^{-\sqrt{d}s}} ds , \\ &= iurt + \kappa^* \phi^* \frac{1}{\lambda^2} \left(c - \sqrt{d} \right) \left[t + \frac{g-1}{\sqrt{d}g} \ln \left(\frac{1 - g e^{-\sqrt{d}t}}{1 - g} \right) \right] \\ &= iurt + \kappa^* \phi^* \frac{1}{\lambda^2} \left[t \left(c - \sqrt{d} \right) - 2 \ln \left(\frac{1 - g e^{-\sqrt{d}t}}{1 - g} \right) \right] . \end{aligned} \tag{A.16}$$

Unfortunately, $\psi_3(t, u)$ cannot be calculated in closed form, wherefore we only state the corresponding ODE at this point. Based on the numerical solution of (A.11), the ODE (A.8) can then be solved via numerical integration. \square

Appendix B. Risk-neutral asset-price dynamics of the alternative 2-factor SV model

In the following we derive the risk-neutral asset-price dynamics of the alternative 2-factor SV model. Similarly as in Section 2.2, we start by defining the following Brownian motions

with drift:

$$d\widehat{W}_{1,t} = d\overline{W}_{1,t} + \gamma_{1,t}dt ,$$

$$d\widehat{W}_{2,t} = d\overline{W}_{2,t} + \gamma_{2,t}dt ,$$

$$d\widehat{W}_{3,t} = d\overline{W}_{3,t} + \gamma_{3,t}dt ,$$

where $\overline{W}_{1,t}$, $\overline{W}_{2,t}$ and $\overline{W}_{3,t}$ are independent Brownian motions. The processes $\gamma_{1,t}$, $\gamma_{2,t}$ and $\gamma_{3,t}$ are assumed to be adapted to the filtration $(\mathcal{F}_t)_{t \geq 0}$ and specify the market price of risk of S_t , σ_t and θ_t , respectively. Analogously to the Heston model, we impose the following condition:

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

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

$$dS_t = rS_t dt + \sqrt{\sigma_t}S_t \widetilde{W}_{1,t} , \quad (\text{B.1})$$

$$d\sigma_t = [\kappa_1(\theta_t - \sigma_t) - \lambda_1\sqrt{\sigma_t}\gamma_{2,t}] dt + \lambda_1\sqrt{\sigma_t}d\widetilde{W}_{2,t} , \quad (\text{B.2})$$

$$d\theta_t = [\kappa_2(\phi - \theta_t) - \lambda_2\sqrt{\theta_t}\gamma_{3,t}] dt + \lambda_2\sqrt{\theta_t}d\widetilde{W}_{3,t} , \quad (\text{B.3})$$

where $\widetilde{W}_{1,t} = \sqrt{1 - \rho^2}\widehat{W}_{1,t} + \rho\widehat{W}_{2,t}$, $\widetilde{W}_{2,t} = \widehat{W}_{2,t}$ and $\widetilde{W}_{3,t} = \widehat{W}_{3,t}$. For the market price of variance risk $\gamma_{2,t}$ we impose the following condition:

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

Hence, $\gamma_{2,t}$ is given by

$$\gamma_{2,t} = \frac{(\kappa_1 - \kappa_1^*)(\theta_t - \sigma_t)}{\lambda_1\sqrt{\sigma_t}} . \quad (\text{B.4})$$

Combining (B.2) with (B.4) results in the following risk-neutral variance dynamics:

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

Similarly as for the market price of variance risk, we impose the following condition for $\gamma_{3,t}$:

$$\kappa_2(\phi - \theta_t) - \lambda_2\gamma_{3,t}\sqrt{\theta_t} = \kappa_2^*(\phi^* - \theta_t).$$

Hence, $\gamma_{3,t}$ is given by

$$\gamma_{3,t} = \frac{\kappa_2(\phi - \theta_t) - \kappa_2^*(\phi^* - \theta_t)}{\lambda_2\sqrt{\theta_t}} . \quad (\text{B.5})$$

Combining (B.3) with (B.5) results in the following risk-neutral long-run target variance dynamics:

$$d\theta_t = \kappa_2^*(\phi^* - \theta_t)dt + \lambda_2\sqrt{\theta_t}d\widetilde{W}_{3,t} ,$$

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

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

AppendixC. Option Pricing Comparison

We use the data set as described in §3.2 and perform an in-sample analysis for the Heston, the two 2-factor SV (2FSV) model and the alternative 2-factor SV (A2FSV) model. In order to compare the pricing effects related to the risk-neutral parameters, we estimate the historical parameter simultaneously. Thus differences in the pricing performance arises only from the risk-neutral parameters.

We performed an in-sample analysis for at-the-money (ATM) and out-of-the money (OTM) options for for various times to maturity LZ_i in $i = 1, 2, 3, 4$ ¹⁶ and strikes K_1 (ATM) and K_2 (OTM).¹⁷ We compared the performance in terms of the mean error of all options in our sample and the means-squared error (MSE). As expected, both 2-factor SV models out-perform the Heston model. While the mean error criterion favours the 2-factor SV model, both models show roughly the same performance in terms of the MSE.

¹⁶The average maturities are 22, 45, 89 and 148 days.

¹⁷We use the following strike choice function $K_i = \lfloor S/25 \rfloor \cdot 25 + i \cdot 25$, $i = 1, 2$. For the time period considered the average moneyness of the (OTM) option is 96%.

Table C.8: Comparison of Pricing Errors

Pricing errors		ATM-options (K_1) various times to maturities (LZ)				OTM-options (K_2) various times to maturities (LZ)			
Model	Metric	LZ_1/K_1	LZ_2/K_1	LZ_3/K_1	LZ_4/K_1	LZ_1/K_2	LZ_2/K_2	LZ_3/K_2	LZ_4/K_2
Heston	Mean	0.38	1.17	1.41	0.68	0.03	0.58	0.87	0.25
2FSV	Mean	0.01	0.18	0.21	0.40	0.00	0.04	-0.07	-0.05
A2FSV	Mean	-0.26	-0.07	0.09	0.38	-0.29	-0.25	-0.24	-0.14
Heston	MSE	10.64	15.21	21.15	28.72	7.37	11.67	18.37	26.81
2FSV	MSE	9.53	14.30	18.23	21.89	6.64	11.85	17.09	20.54
A2FSV	MSE	9.34	14.01	18.28	22.24	6.78	11.84	17.21	20.82

Notes: Comparison of option pricing performance of the SV models

AppendixD. Return distribution

Let X be a random variable with characteristic function $\phi_X(u)$. Following Wu (2012), the probability density function of X is then given by

$$f_X(x) = \frac{1}{\pi} \int_0^\infty e^{-iux} \phi_X(u) du , \quad (\text{D.1})$$

and the cumulative distribution function of X by

$$F_X(x) = \frac{1}{2} + \frac{1}{2\pi} \int_0^\infty \frac{e^{iux} \phi_X(-u) - e^{-iux} \phi_X(u)}{iu} du . \quad (\text{D.2})$$

Now, the continuous return of the asset-price process S_t within the period $[t, t+\tau]$ is defined as

$$r_{t,t+\tau} = \ln(S_{t+\tau}) - \ln(S_t) , \quad (\text{D.3})$$

so that

$$\begin{aligned} F_{r_{t,t+\tau}}(x) &= \mathbb{P}[r_{t,t+\tau} \leq x] = \mathbb{P}[\ln(S_{t+\tau}) - \ln(S_t) \leq x] \\ &= \mathbb{P}[\ln(S_{t+\tau}) \leq x + \ln(S_t)] = F_{\ln(S_{t+\tau})}(x + \ln(S_t)) . \end{aligned} \quad (\text{D.4})$$

For affine option-pricing models, we generally only know the characteristic function of the log-asset price $\ln(S_t)$. Hence, the return pdf and cdf can be recovered via equation (D.1) and (D.2), respectively, together with equation (D.4).

AppendixE. Unscented Kalman Filter

The Unscented Kalman Filter (UKF) is an extension of the classical Kalman Filter and allows for a nonlinear state-space representation

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

$$y_i = g(x_i, \epsilon_i | \theta_2) , \quad (\text{E.2})$$

where $f : \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}^m$ and $g : \mathbb{R}^m \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ are possibly nonlinear functions. The state equation (E.1) parameterized by θ_1 describes the dynamics of the possibly unobservable

state x_i ($n \times 1$) and the measurement equation (E.2) parameterized by θ_2 relates the state of the system to observations y_i ($m \times 1$). For ease of notation, we summarize the parameters of the state-space representation in the vector $\theta = [\theta_1 \ \theta_2]$. Moreover, the disturbance terms η_i ($m \times 1$) and ϵ_i ($n \times 1$) as well as the initial state x_0 ($m \times 1$) are assumed to be normally distributed with

$$\eta_i \sim \mathcal{N}(0, P) , \quad \epsilon_i \sim \mathcal{N}(0, Q) \quad \text{and} \quad x_0 \sim \mathcal{N}(m_0, R_0) . \quad (\text{E.3})$$

Based on these assumptions, the UKF approximates the distribution of the system via

$$\begin{matrix} x_i \\ y_i \end{matrix} \bigg|_{\mathcal{I}_{i-1}} \approx \mathcal{N} \left(\begin{bmatrix} m_i^-(\theta) \\ n_i^-(\theta) \end{bmatrix}, \begin{bmatrix} R_i^-(\theta) & C_i^\top(\theta) \\ C_i(\theta) & S_i(\theta) \end{bmatrix} \right) , \quad (\text{E.4})$$

where $\mathcal{I}_i = \{y_1, \dots, y_i\}$ denotes the available information at time i . The vectors $m_i^-(\theta)$ and $n_i^-(\theta)$ specifying the conditional mean, and the matrices $C_i(\theta)$, $R_i^-(\theta)$ and $S_i(\theta)$ specifying the conditional variance, depend on the parametrization of the state-space representation. Following (Särkkä, 2011, p. 55 ff.), the parameters specifying (E.4) can be calculated recursively via the following algorithm.

Algorithm AppendixE.1 (Unscented Kalman filter). *Let (E.1) – (E.2) be the underlying state-space representation and assume that the disturbance terms η_i and ϵ_i as well as the initial state x_0 are normally distributed as specified in (E.3). Then the state x_i given all observations up to time i is approximately normally distributed with*

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

where $\mathcal{I}_i = \{y_1, \dots, y_i\}$ denotes the observations up to time i . Furthermore, the conditional mean m_i and the conditional variance R_i can be calculated recursively as follows:

First, define the augmented state vector $\tilde{x}_{i-1} = [x_{i-1}^\top \ \eta_i^\top \ \epsilon_i^\top]^\top$, which is of dimension $d = 2m + n$, and approximate the conditional mean $\mathbb{E}[\tilde{x}_{i-1} | \mathcal{I}_{i-1}]$ and variance $\text{Var}[\tilde{x}_{i-1} | \mathcal{I}_{i-1}]$

by

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

Moreover, implicitly define the matrices A ($m \times d$), B ($m \times d$) and C ($n \times d$), such that

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

Next, build the sigma vectors $\mathcal{X}_{i-1} = [\tilde{x}_{i-1,0} \dots \tilde{x}_{i-1,2d}]$ according to

$$\begin{aligned} \tilde{x}_{i-1,0} &= \tilde{m}_{i-1}, \\ \tilde{x}_{i-1,j} &= \tilde{m}_{i-1} + \alpha\sqrt{d}\sigma_{i-1,j} \quad \text{for } j = 1, \dots, d, \\ \tilde{x}_{i-1,j} &= \tilde{m}_{i-1} - \alpha\sqrt{d}\sigma_{i-1,j-d} \quad \text{for } j = d+1, \dots, 2d, \end{aligned} \tag{E.5}$$

where $\sigma_{i-1,j}$ is the j^{th} row of the matrix square root $\sqrt{\tilde{R}_{i-1}}$. The scaling parameter α , with $0 < \alpha \leq 1$, determines the spread of the sigma points around the mean of \tilde{x}_{i-1} . The corresponding weights for these sigma vectors are specified by

$$w_0^\mu = \lambda/(d + \lambda), \tag{E.6}$$

$$w_0^\Sigma = \lambda/(d + \lambda) + (1 - \alpha^2 + \beta), \tag{E.7}$$

$$w_i^\mu = w_i^\Sigma = 1/(2d + 2\lambda) \quad \text{for } i = 1, \dots, 2d, \tag{E.8}$$

where $\lambda = d(\alpha^2 - 1)$. The vectors w_i^μ and w_i^Σ denote the weights for the mean and variance approximation, respectively. Furthermore, the parameter β allows one to incorporate prior knowledge about the distribution of \tilde{x}_{i-1} .¹⁸

Finally, define the transformed sigma vectors

$$x_{i,j}^- = f(A\tilde{x}_{i-1,j}, B\tilde{x}_{i-1,j})$$

$$y_{i,j}^- = g(x_{i,j}^-, C\tilde{x}_{i-1,j})$$

and proceed with the following steps:

1. Prediction step:

$$m_i^- = \sum_{j=0}^{2d} w_j^\mu x_{i,j}^-,$$

$$R_i^- = \sum_{j=0}^{2d} w_j^\Sigma (x_{i,j}^- - m_i^-)(x_{i,j}^- - m_i^-)^\top.$$

¹⁸For example, if X is normally distributed, $\beta = 2$ would be the optimal choice. See Julier (2002) for more details about the parameterization of the unscented transformation.

2. Update step:

$$m_i = m_i^- + C_i^\top S_i^{-1} (y_i - n_i^-) ,$$

$$R_i = R_i^- - C_i^\top S_i^{-1} C_i ,$$

where

$$n_i^- = \sum_{j=0}^{2d} w_j^\mu y_{i,j}^- ,$$

$$C_i = \sum_{j=0}^{2d} w_j^\Sigma (x_{i,j}^- - m_i^-) (y_{i,j}^- - n_i^-)^\top ,$$

$$S_i = \sum_{j=0}^{2d} w_j^\Sigma (y_{i,j}^- - n_i^-) (y_{i,j}^- - n_i^-)^\top .$$

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

The distribution of the measurement y_i conditional on the information \mathcal{I}_{i-1} can be approximated via

$$y_i | \mathcal{I}_{i-1} \approx \mathcal{N}(n_i^-(\theta), S_i(\theta)) .$$

So the log-likelihood function for observing $\bar{y}_1, \dots, \bar{y}_N$ can be approximated by

$$\mathcal{L}(\theta) = -\frac{nN}{2} \ln(2\pi) - \frac{1}{2} \sum_{i=1}^N \ln(|S_i(\theta)|) - \frac{1}{2} \sum_{i=1}^N \xi_i^\top S_i(\theta)^{-1} \xi_i , \quad (\text{E.9})$$

where $\xi_i = \bar{y}_i - n_i^-(\theta)$ denotes the prediction error. Now, the unknown parameter vector θ_0 can be estimated via the Quasi-maximum likelihood (QML) method, where (E.9) is maximized with respect to θ .

AppendixF. Discretization of the Cox-Ingersoll-Ross process

In accordance with (van der Ploeg, 2006, p. 76 f.), we derive a discretization for the Cox-Ingersoll-Ross process

$$d\sigma_t = \kappa(\phi - \sigma_t)dt + \lambda\sqrt{\sigma_t}dW_t . \quad (\text{F.1})$$

First, define

$$\tilde{\sigma}_t = e^{\kappa t} \sigma_t ,$$

and consider the differential of $\tilde{\sigma}_t$. By Itô's lemma,

$$\begin{aligned} d\tilde{\sigma}_t &= \kappa e^{\kappa t} \sigma_t dt + e^{\kappa t} \kappa (\phi - \sigma_t) dt + e^{\kappa t} \lambda \sqrt{\sigma_t} dW_t \\ &= e^{\kappa t} \kappa \phi dt + e^{\kappa t} \lambda \sqrt{\sigma_t} dW_t . \end{aligned} \tag{F.2}$$

Next, we integrate (F.2) over the interval $[t_{i-1}, t_i]$ with $t_i - t_{i-1} = \Delta t$:

$$\tilde{\sigma}_i - \tilde{\sigma}_{i-1} = \kappa \phi \int_{t_{i-1}}^{t_i} e^{\kappa u} du + \lambda \int_{t_{i-1}}^{t_i} e^{\kappa u} \sqrt{\sigma_u} dW_u , \tag{F.3}$$

where we set $i = t_i$. Finally, we multiply equation (F.3) by $e^{-\kappa t_i}$ and solve for σ_i , which is

$$\begin{aligned} \sigma_i &= e^{-\kappa \Delta t} \sigma_{i-1} + \kappa \phi \int_{t_{i-1}}^{t_i} e^{-\kappa(t_i-u)} du + \lambda \int_{t_{i-1}}^{t_i} e^{-\kappa(t_i-u)} \sqrt{\sigma_u} dW_u \\ &= e^{-\kappa \Delta t} \sigma_{i-1} + \phi (1 - e^{-\kappa \Delta t}) + \lambda \int_{t_{i-1}}^{t_i} e^{-\kappa(t_i-u)} \sqrt{\sigma_u} dW_u \\ &\approx e^{-\kappa \Delta t} \sigma_{i-1} + \phi (1 - e^{-\kappa \Delta t}) + e^{-\kappa \Delta t} \lambda \sqrt{\sigma_{i-1}} (W_i - W_{i-1}) \\ &= e^{-\kappa \Delta t} \sigma_{i-1} + \phi (1 - e^{-\kappa \Delta t}) + e^{-\kappa \Delta t} \lambda \sqrt{\sigma_{i-1}} \sqrt{\Delta t} \epsilon_i , \end{aligned} \tag{F.4}$$

where $\epsilon_i \sim \mathcal{N}(0, 1)$. In contrast to the classical Euler approximation, where (F.1) is approximated as

$$\sigma_i = \sigma_{i-1} + \kappa (\phi - \sigma_{i-1}) \Delta t + \lambda \sqrt{\sigma_{i-1}} \sqrt{\Delta t} \epsilon_i ,$$

equation (F.4) captures the drift term in closed form.

References

- Aït-Sahalia, Y., Lo, A., 2000. Nonparametric risk management and implied risk-aversion. *Journal of Econometrics* 94, 9–51.
- Albrecher, H., Mayer, P., Schoutens, W., Tistaert, J., 2007. The little Heston trap. *Wilmott Magazine*, 83–92.
- Andersen, T., Bondarenko, O., 2007. Construction and interpretation of model-free implied volatility. In: Nelken, I. (Ed.), *Volatility as an asset class*. RISK Publication.
- Bakshi, G., Carr, P., Wu, L., 2008. Stochastic risk premiums, stochastic skewness in currency options, and stochastic discount factors in international economies. *Journal of Financial Economics* 87, 132–156.
- Bakshi, G., Kapadia, N., 2003. Delta-hedged gains and the negative market volatility risk premium. *Review of Financial Studies* 16, 527–566.
- Bakshi, G., Wu, L., 2010. The behavior of risk and market prices of risk over the Nasdaq bubble period. *Management Science* 56, 2251–2264.
- Balduzzi, P., Das, S., Foresi, S., 1998. The central tendency: A second factor in bond yields. *Review of Economics and Statistics* 80, 62–72.
- Barnett, V., Lewis, T., 1994. *Outliers in statistical data*. John Wiley & Sons, third edition.
- Bates, D., 2012. U.S. stock market crash risk, 1926 - 2010. *Journal of Financial Economics* 105, 229–259.
- Bingham, N. H., Kiesel, R., 2004. *Risk-neutral valuation*. Springer, Second Edition.
- Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. *Journal of Political Economy* 81, 637–654.

- Bollerslev, T., Gibson, M., Zhou, H., 2011. Dynamic estimation of volatility risk premia and investor risk aversion from option-implied and realized volatilities. *Journal of Econometrics* 160, 235–245.
- Brigo, D., Mercurio, F., 2006. *Interest Rate Models - Theory and Practice*, 2nd Edition. Springer Finance. Springer, New-York.
- Brown, D., Jackwerth, J., 2012. The pricing kernel puzzle: Reconciling index option data and economic theory. In: Batten, J., Wagner, N. (Eds.), *Contemporary studies in economics and financial analysis: Derivative securities pricing and modelling*. Emerald Group, forthcoming.
- Brummelhuis, R., Kaufmann, R., 2007. Time scaling for Garch(1,1) and AR(1)-Garch(1,1) processes. *Journal of Risk* 9.
- Carr, P., Madan, D., 1999. Option valuation using the fast Fourier transform. *Journal of Computational Finance* 2, 61–73.
- Carr, P., Wu, L., 2009. Variance risk premiums. *Review of Financial Studies* 22, 1311–1341.
- Chernov, M., Ghysels, E., 2000. A study towards a unified approach to the joint estimation of objective and risk neutral measures for the purpose of options valuation. *Journal of Financial Economics* 56, 407–458.
- Christoffersen, P., Dorion, C., Jacobs, K., Karoui, L., 2014. Nonlinear Kalman filtering in affine term structure models. *Management Science*, 2248–2268.
- Christoffersen, P., Heston, S., Jacobs, K., 2009. The shape and term structure of the index option smirk: Why multifactor stochastic volatility models work so well. *Management Science* 55, 1914–1932.
- Christoffersen, P., Heston, S., Jacobs, K., 2013. Capturing option anomalies with a variance-dependent pricing kernel. *Review of Financial Studies* 26, 1962–2006.

- Christoffersen, P., Jacobs, K., Chang, B., 2012. Forecasting with option implied information. In: Elliott, G., Timmermann, A. (Eds.), *Handbook of economic forecasting*. Vol. 2. North-Holland.
- Christoffersen, P., Jacobs, K., Ornathanalai, C., Wang, Y., 2008. Option valuation with long-run and short-run volatility components. *Journal of Financial Economics* 90, 272–297.
- Cochrane, J., 2005. *Asset pricing*. Princeton University Press, revised edition.
- Dai, Q., Singleton, K., 2000. Specification analysis of affine term structure models. *Journal of Finance* 55, 1943–1978.
- Daniélsson, J., Zigrand, J.-P., 2006. On time-scaling of risk and the square-root-of-time rule. *Journal of Banking & Finance* 30, 2701–2713.
- Duffie, D., Kan, R., 1996. A yield-factor model of interest rates. *Mathematical Finance* 6, 379–406.
- Duffie, D., Pan, J., Singleton, K., 2000. Transform analysis and asset pricing for affine jump diffusions. *Econometrica* 68, 1343–1376.
- Filipović, D., Mayerhofer, E., 2009. Affine diffusion processes: Theory and applications. In: Albrecher, H., Runggaldier, W., Schachermayer, W. (Eds.), *Advanced financial modelling*. de Gruyter.
- Grith, M., Härdle, W., Krätschmer, V., 2016. Reference dependent preferences and the EPK puzzle. *Review of Finance*, 1–30.
- Grith, M., Härdle, W., Park, J., 2013. Shape invariant modelling pricing kernels and risk aversion. *Journal of Financial Econometrics* 11, 370–399.
- Hens, T., Reichlin, C., 2013. Three solutions to the pricing kernel puzzle. *Review of Financial Studies* 17, 1065–1098.

- Heston, S., 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies* 6, 327–343.
- Hull, J., White, A., 1987. The pricing of options on assets with stochastic volatility. *Journal of Finance* 42, 281–300.
- Jackwerth, J., 2000. Recovering risk aversion from option prices and realized returns. *Review of Financial Studies* 13, 433–451.
- Jackwerth, J., 2004. Option-implied risk-neutral distributions and risk aversion. *Research Foundation of AIMR*.
- Jackwerth, J., Rubinstein, M., 1996. Recovering probability distributions from option prices. *Journal of Finance* 51, 1611–1631.
- Jones, C., 2003. The dynamics of stochastic volatility: Evidence from underlying and options markets. *Journal of Econometrics* 116, 181–224.
- Julier, S., 2002. The scaled unscented transformation. *Proceedings of the American Control Conference*, 4555–4599.
- Julier, S., Uhlmann, J., 1997. A new extension of the Kalman filter to nonlinear systems. *Proceedings of SPIE*, 182–193.
- Kahl, C., 2007. Modelling and simulation of stochastic volatility in finance. *Dissertation*.
- Kahl, C., Loard, R., 2006. Why the rotation count algorithm works. *Tinbergen Institute Discussion Paper*, 1–31.
- Kahl, C., Loard, R., 2010. Complex logarithms in Heston-like models. *Mathematical Finance* 20, 671–694.
- Kupiec, P., 1995. Techniques for verifying the accuracy of risk management models. *Journal of Derivatives* 3, 73–84.

- Lunde, A., Hansen, P., 2005. A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? *Journal of Applied Econometrics* 20, 873–889.
- McNeil, A., Frey, R., Embrechts, P., 2005. *Quantitative Risk Management: Concepts, Techniques, and Tools*. Princeton University Press.
- Merton, R., 1980. On estimating the expected return on the market – an explanatory investigation. *Journal of Financial Economics* 8, 323–361.
- Pan, J., 2002. The jump-risk premia implicit in options: Evidence from an integrated time-series study. *Journal of Financial Economics* 63, 3–50.
- Piazzesi, M., 2009. Affine term structure models. In: Aït-Sahalia, Y., Hansen, L. (Eds.), *Handbook of financial econometrics*. North Holland.
- Rosenberg, J., Engle, R., 2002. Empirical pricing kernels. *Journal of Financial Economics* 64, 341–372.
- Särkkä, S., 2011. Bayesian estimation of time-varying systems. Lecture notes.
- Schmelzle, M., 2010. Option pricing formulae using Fourier transform: Theory and application. Working paper.
- van der Ploeg, A., 2006. Stochastic volatility and the pricing of financial derivatives. Dissertation.
- Wan, E., van der Merwe, R., 2001. The unscented kalman filter. In: Haykin, S. (Ed.), *Kalman filtering and neural networks*. John Wiley & Sons.
- Whaley, R., 2000. The investor fear gauge. *Journal of Portfolio Management* 26, 12–17.
- Whaley, R., 2009. Understanding the vix. *Journal of Portfolio Management* 35, 98–105.
- Wu, L., 2012. From characteristic functions and Fourier transforms to pdfs/cdfs and option prices. Lecture notes, <http://faculty.baruch.cuny.edu>.